UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2024-085
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 5/12/2024 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN INFORMATIQUE

by

XENGIE CHENG DOAN
Born on 15 March 1994 in California, (United States of America)

A framework for user-centered, legal-ethical collective consent models: genomic data sharing

Dissertation defence committee

Dr Gabriele LENZINI, dissertation supervisor
Associate Professor, Université du Luxembourg

Dr Annika SELZER
Fraunhofer-Institut für sichere Informationstechnologie, Darmstadt
Head of "IT Law and Interdisciplinary Privacy Research" at Fraunhofer SIT

Dr Gilbert FRIDGEN, Chairman
Professor, Université du Luxembourg

Dr Jessica EYNARD
Associate Professor, University of Toulouse Capitole, France

Dr Christian FISCH, Vice Chairman
Associate Professor, Université du Luxembourg

Abstract

Sharing genomic data can be useful for personalized medicine, advancing research, or uncovering genetic ancestry. However, there are also unintended and unexplored risks because genomic data can implicate identifiable collectives from genetic relatives (in the past and future) to genetic minorities (those sharing a rare disease or trait). Individual rights to consent are insufficient, and a collective consent process to inform groups and individuals and safeguard their rights should be put into place. While collective consent has been mandated for Indigenous groups, the same process is not offered for other collectives. To study how to build a collective consent process, I used methods from computer science (requirements engineering), privacy (contextual integrity), HCI (user studies), and governance (European Union (EU) regulations and bioethics) to center both end-users and business users in a compliant, ethical collective consent framework. I analyzed the EU legal and ethical regulations and guidelines to characterize gray areas and conflicts regarding consent. Then, I assessed public privacy and consent policies of leading direct-to-consumer genetic testing companies – combining well-established contextual integrity analyses with analysis of user-relevant governance terms and risk and benefit information. This analysis revealed that a majority of information about genetic data sharing was vague, confusing, and not framed in a useful way to the user, with no collective risks and benefits explained, which does not align with the EU General Data Protection Regulation (GDPR). Subsequently, I tested the adaptability of contextual integrity methodologies in eliciting and validating policies for businesses to improve their documentation through mixed-method interviews. Employees revealed that the method would be useful for more specific, structured, and complete information flows for future audits and documentation, as well as helping employees write documentation and analyze the quality, which overall can increase communication between teams. In parallel, I also tested using Business Process Model and Notation (BPMN) modeling for consent processes, developing requirements, creating artefacts, and piloting a validation study with employee interviews. Participants reported that it offered a useful visual overview and helped to identify conflicts and analyze compliance processes. I then surveyed and asked potential end-users to rank the most useful and engaging features of different mediums (video, infographic, comic, plain text, newsletter) to better understand their needs, goals, and desires for consent management and their ranking of attention-grabbing elements. Users revealed that they wanted quick, relevant, and understandable information to make a consent decision, which they preferred to be stored digitally on a centralized app or platform. Supporting the prioritization of information, elements like structure, step-by-step design, and readability were the most highly ranked individually, and present in the winning infographic medium. My work offers a framework to build collective consent through increased transparency, user-centered consent management, and methods with real-world business applicability. Given the many unexplored challenges regarding collective consent for a general population, the specific gaps, methods, and user perceptions I have characterized could significantly advance our understanding of how to build collective consent to address existing needs.

Acknowledgements

In his acceptance speech in 1997, Fred Rogers took a moment for the whole audience to think of the people who loved them into being. I would like to offer space for you, the reader, to also think of those people.

For me, I would like to thank everyone who has loved this thesis, the student that I am, and the person I wan to be into being. I never expected to return to academia after my master's degree and working, or to make it to the end of the PhD. Perhaps my ignorance helped me begin. I have more people in mind than I can express in this section, and however small the gesture may be, I am so grateful. I carry a part of you with me, even if we don't talk as much, or we talk every day, or you have passed on.

Specifically, thank you for Gabriele for welcoming me into the group, helping me grow, and pushing me beyond my comfort zone. I have truly enjoyed chatting about random topics as well as research, and really appreciate your open approach to life. I am glad I took the leap to Luxembourg to join your group.

Thank you to current and past members of the IRiSC group for a lovely working environment, I am sorry if I emailed you too many times before deciding to join the lab or chatted too much in the office and distracted you. To Monica for the emotional support and thesis feedback even while on pregnancy leave. To my daily supervisors: Arianna for her support for much of my PhD.; Marietjie for being the best officemate and helping to read my thesis while in a different university in South Africa.

Thank you to LeADS (H2020 Grant Agreement ID 956562) for the many travel opportunities, I especially look fondly upon my research stay at SSSA with Francesca and at Tellu with Liubov, who were so generous with their time. Thank you as well to the University of Luxembourg for all their support, especially all the adminstrative staff who answered all my many questions.

I would also like to thank all my co-authors who shared their ideas, challenged me to write better, and gave me their time. I am grateful to have found talented individuals to work with, through introduction or sliding into Twitter DMs. Many chapters were truly enhanced with your help, and taught me so much about interdisciplinary work.

To my strong support network, I could not have done it without you. To my family for keeping in touch through a 9 hour time difference, coming to visit me in Europe, and spoiling me when I'm home. To my cat for emotionally supporting me by being the loveliest thing alive. To my friends for seeing the best in me and my abilities, and listening to my worries when I spiral. I would be nowhere without my group-chats to keep me grounded, especially with people who have been through academia. To my LeADS kin, Barbara and Mitisha, I would have perished without you.

Last but not least, I would like to acknowledge how the privileged I am to study in peace in times of war, genocide, and oppression around the world. To the people of Palestine, Sudan, Congo, Ukraine, and all the countless others I cannot recall, I hope to never falter in solidarity with the oppressed. I hope one day we will all be free. As Mariame Kaba said, hope is a discipline. I hope we keep building towards that future in spite of how the sky feels like it is falling down. I hope in some small way, this work helps you see a new type of future and imagine your own.

Contents

List of Figures

List of Tables

Acronyms

ADPC Advanced Data Protection Control. 28, 116

BPMN Business Process Model and Notation. vi, 19, 20, 65–69, 71–80, 113–115, 117, 122

CGRCO Chief Governance, Risk and Compliance Officer. 52, 59–62, 66, 69, 70, 73–76, 78, 79, 114, 117

CI Contextual Integrity. vi, viii, 32–34, 36–40, 45, 48, 50–53, 56, 61–63, 112–115, 117

CIOMS Council for International Organizations of Medical Sciences. 6

CJEU Court of Justice of the European Union. 14, 15

CMPs consent management platforms. 26

DC Dynamic Consent. 8, 9, 12, 13, 29, 30, 105, 118, 119

DGA Data Governance Act. 23, 27, 28, 83, 104

DNA Deoxyribonucleic acid. 32, 40, 43

DPA Data Protection Authority. 15, 78

DPCCMs Data Protection and Consenting Communication Mechanisms. 28

DPO Data Protection Officer. 76

DPV Data Privacy Vocabulary. 9, 12, 28

DRIP Declaration on the Rights of Indigenous Peoples. 3, 4

DTC Direct to Consumer. 20, 32–35, 39, 42, 47, 48, 109, 112, 122

DTCGTC Direct to Consumer Genetic Testing Company. 19

DTCGTCs Direct to Consumer Genetic Testing Companies. 109–112

EARs Easy Approach to Requirements Syntax. 66, 69

eConsent electronic consent. 18

EDPB European Data Protection Board. 5, 6, 8, 24, 111, 118, 121

EDPS European Data Protection Supervisor. 26, 121

EHDS European Health Data Space. 7, 19, 21, 117, 119–121

eIDAS electronic identification trust services. 6

EU European Union. 1, 2, 4, 6, 9, 15, 18–20, 22, 23, 26, 28, 30, 64, 65, 75, 88, 107, 109–112, 117, 119–122

FE Flourishing Ethics. 27

GDPR General Data Protection Regulation. 2, 4–9, 14, 15, 18, 19, 23–27, 32–35, 40, 41, 46, 48, 51–53, 59, 61, 62, 64–67, 75–79, 85, 107, 110–112, 117–121

GRC Governance, Risk and Compliance. 54, 60, 62, 73

H3Africa Human Heredity and Health in Africa. 3, 10

HCI Human-Computer Interaction. 1, 109, 115

HDAB Health Data Access Body. 120

IAB Interactive Advertising Bureau. 15

IC Informed Consent. 4–7, 9, 65, 71, 105, 109, 110, 113–117

ICT information and communication technologies. 53

ISO International Organization for Standardization. 9, 12, 22, 27, 30, 66, 77

IT Information Technology. 5

NHS UK National Health Service. 8

NoK Next of Kin. 54, 65, 68, 69, 71, 75, 119

ORDL Open Digital Rights Language. 12

P Participant. 59–61, 73–79, 93–96, 98–102, 105–107

PEOU Perceived Ease of Use. 53, 60, 71, 72, 77, 113, 114

PPAs Personalized Privacy Assistants. 29

PU Perceived Usefulness. 53, 60, 71–73, 77, 78, 113, 114

RE Requirements Engineering. vi, 65–69, 71, 113, 115

RQ Research Question. viii, 20, 102, 103, 109

SDT self-determination theory. 27

SME Small Medium Enterprise. 2, 20, 66, 68, 78, 79, 109, 113, 115, 117

SMEs Small Medium Enterprises. 2, 18, 19, 66, 78, 79, 111, 114

TAM Technology Acceptance Model. 53, 57, 66, 68, 79, 113–115

TCF Transparency and Consent Framework. 15

TPB Theory of Planned Behavior. 53, 113

TRA Theory of Reasonable Action. 53

UML Unified Modeling Language. 66, 67, 77

UN United Nations. 3, 4

US United States. 34

USD United States dollar. 32

UX user experience. 75

W3C World Wide Web Consortium. 9, 28

WP Working Party. 5, 62, 83

WP29 Article 29 Working Party. 5

Chapter 1

Introduction

1.1 Background

Does consent work in the modern world? How can we imagine a useful consent for data sharing?

Especially in the case of sensitive health data used for biomedical research, clinical research, or to provide services, how can we ensure people are informed and in control over their own information? Consent is one of the key methods for people to be informed of where their data is going, how it is being used, how they may be affected, and what their rights are. This is especially complex for the case of genetic data, which connects an individual to genetic relatives or genetic minorities (e.g, people with specific mutations in their genes). Collective data may have far reaching benefits (e.g., better health for family members) and consequences (e.g., unintended consequences for future generations), but the current system of consent ignores collectives – mainly targeting individuals at the time of data collection.

Collective consent is an already existing model of consent (which will be discussed in more detail in Section 1.1.1.3) that incorporates decision making for upholding collective wellbeing and interests by consulting group governance structures, such as tribal leaders [103]. However, while collective consent has been established in relation to indigenous rights, others are not afforded the same community considerations, although the benefits and harms may affect the whole collective. How can similar principles of formalized informed collective consent apply to other groups, such as families or genetic minorities when sharing DNA data?

This thesis aims to combine and create a framework for collective consent using methods from computer science (requirements engineering), privacy (contextual integrity), Human-Computer Interaction (HCI) (user studies), and governance (European Union (EU) regulations, bioethics) to characterize the gaps in digital collective

consent and develop and test tools to enhance the transparency of consent processes and informational transparency.

The reality is also that large and small organizations that implement consent hold most of the power over end-users, defining the scope of what end-users can do. Thus, this work targets two audiences: end-users and users in Small Medium Enterprises (SMEs). End-users are the typical target of user studies as the final recipients of the product, whose needs and desires are often neglected in the design process. Also important are the employees at a Small Medium Enterprise (SME) who also need the ability to understand their responsibilities and translate user research into business processes to improve their product. Especially for a complex case such as consent for health data in the EU with different privacy, regulatory, and ethical concerns, it can be difficult to align goals within employees. As one SME can provide consent services to many customers, even the use-case explored in this thesis can begin to illustrate the mutually supportive and synergistic cycle of this framework.

I explore genetic data sharing as a use-case because it offers a concrete example of sensitive collective data. It is one of the examples of sensitive data in the EU under the General Data Protection Regulation (GDPR), and can offer concrete examples of unique collective data that can also be applied to other types of health data or collective information, such as social media data. In the course of this research, I also investigated health data in general (e.g., for legal consent gaps) and similar types of health data due to use-case constraints (e.g., video data for a collective industry partner use case). As a subset of health data, the overall findings can still apply to genetic data, and hopefully could be applied to more types of shared data. As such, I introduce consent from both a bioethical and privacy perspective to better understand the field, which will be further elaborated in subsequent chapters.

1.1.1 Consent

1.1.1.1 Ethical Consent

While many may think of digital cookie consents due to their ubiquity on websites, the concept of consent was formally mandated as a bioethical tool as a reaction to medical human rights violations reported in World War II [348] and later from the US Tuskeegee Syphilis trials [63]. These led to ethical guidelines for human subject research established in the Declaration of Helsinki [15] in 1964 and the Belmont Report in 1979 [254], respectively. The Declaration of Helsinki offers descriptive guidelines for what constitutes informed consent to ensure that it respects individuals in cases of medical research involving human subjects. Overall, the document follows principles such as doing no harm, respect for autonomy, and centering the patient's best interests. The Helsinki Report is followed by the Declaration of Taipei in 2002 [16], which offers additional guidelines for research on health databases, big data, and biobanks. These declarations focus on the biomedical context, not on specific types of health data. In the case of genetic data, it is a type of data often stored in biobanks, used for big data analyses and, thus, stored in databases, and the Declaration of Taipei may also apply if used for medical research purposes. The Belmont Report [254] builds on principles of autonomy of the individual, beneficence to oblige institutions to maximize benefits and do no harm, and justice for relative benefits and risk to individuals and institutions. It also stipulates conditions for informed consent but is more general as it applies to all research involving human subjects.

This bioethical justification for consent does not include non-Western cultural beliefs or critiques about the concept of individual autonomy. First, principles of Western bioethics have often been exported to different cultural contexts without accounting for the differences, leading to complications in the ethical reasoning and misalignment between ethical issues in practice and in theory [75, 86]. For example, principalism and autonomy may value individual autonomy in Western bioethics, while more collectivist cultures may have a more "*familial and communal idea of autonomy*" [75] that takes into consideration the greater good or relational impacts. This can make it unsuitable to apply Western bioethics in such cultures. In addition, in many countries, doctors have a greater power over patients and individual autonomy is minimized. These cultures may also have long histories of medicine that do not follow the Hippocratic oath. In non-Western cultures, it is often accepted to, "*prioritize the consent of community leaders or the head of family – usually men – over the voluntary and free consent of the individual.*" [86]. Ekmekci et al. explain, "*Self-construal in Asian and Ubuntu cultures is the key to understanding their decision making perspective. Asian and Ubuntu ethics encourage interdependent self-construal, while liberal ethics are built upon independent self-construal.*" To navigate these cultural differences, guidelines for community engagement and cultural sensitivity have been drafted, for example, the Human Heredity and Health in Africa (H3Africa) policy framework [351] or documents for indigenous rights that will be discussed later in Section 1.1.1.1. This way, *ethical imperialism*, in which one set of ethical norms are forcefully imposed despite being ill-fitting, can be avoided [188].

Many other concepts of autonomy exist, such as relational autonomy, which considers that individual autonomy is impacted or entwined by relational surroundings [313, 84] while still centering the most impacted individual. Stoljar write that, "*Taking relational autonomy seriously suggests that in addition to securing informed consent, health care providers have an important role to play in promoting patient autonomy. Providers must be alert to the social conditions that affect patients' capacities for autonomous reasoning*" [313]. In that example, the doctor would be aware of the cultural or familial norms that impact their decision making and help counsel them. These notions do not suggest removing individual informed consent, but augmenting best practices to include relational considerations from healthcare, institutions, and more. Health data and genetic data specifically are clear examples of relational data, as they may pose shared risks and benefits for related individuals or those with rare genetic diseases.

Collective Rights While the Belmont Report and Declaration of Taipei are grounded in individual autonomy, some mention of groups or collectives can be found. For example, in the Declaration of Helsinki familial or relational considerations are given, "*Although it may be appropriate to consult family members or community leaders, no individual capable of giving informed consent may be enrolled in a research study unless he or she freely agrees*" [15][1]. However, this is still framed in the context of individual rights superseding all others. This raises questions about the usefulness and appropriateness of such strict individualism, especially in the case of genetic data where risks and benefits may impact genetic relatives.

Multiple documents that have come after the Declaration of Helsinki specify collective rights for indigenous groups, such as the United Nations (UN) Declaration on the Rights of Indigenous Peoples (DRIP) in 2007 which states in Article 1 that,

[1]Section B22 DECLARATION OF HELSINKI Ethical Principles for Medical Research Involving Human Subjects (2002)

"*Indigenous peoples have the right to the full enjoyment, as a collective or as individuals, of all human rights and fundamental freedoms as recognized in the Charter of the United Nations, the Universal Declaration of Human Rights and International human rights law*." In Australia, they followed the UNDRIP to create national standards with the Australian Institute of Aboriginal and Torres Strait Islander Studies Code of Ethics for Aboriginal and Torres Strait Islander Research [4] which specifies collective rights and collective and individual consent. The Taiwanese Indigenous Peoples Basic Law mandates that researchers must obtain the consent of the individual and collective consent of the indigenous community for any work involving them [141]. This contrasts with the EU, which endorsed the United Nations Declaration on the Rights of Indigenous Peoples in that year (2007), but has no EU wide guidance. Previously, the European Commission released a working document in 1998 [333], but it only discusses supporting indigenous peoples for sustainable development and reducing poverty. In the Americas, 35 member countries, including the United States, have signed the American Declaration on the Rights of Indigenous Peoples [11]. In it, Article 6 describes collective rights. However, they do not specifically state that there is a right to collective consent as in other countries. Article 13 states that redress must be given if action is taken, "*without their free, prior, and informed consent or in violation of their laws, traditions, and customs*," which may fall under collective governance and decision making. Relying on redress can fail to properly protect consent and can be seen in practice, as indigenous rights regarding consent were violated in Canada and Guatemala [355] and the US [68]. This brings to mind an infamous lawsuit in the US wherein researchers failed to obtain informed consent from the Havasupai for sensitive research including mental health and population studies (given the small size of the population, not only can it be highly revealing, but it is also taboo in the Havasupai Indigenous culture) [68]. While there was no legal precedent set since the Havasupai tribe won in a settlement, the judgment changed how "informed" the people involved needed to be, and the types of research purposes that should be pursued. The researchers obtained broad consent for a diabetes study in 1989 and then used it to study schizophrenia, migration, and inbreeding. A member of the tribe discovered this in 2003 and subsequently filed a lawsuit with her tribe. Broad consent requires that a participant give consent to certain stated uses and any future uses that are unknown at the time. In this case, the secondary purposes were unjust and harmful according to the vulnerable indigenous tribe. Subsequently, this has led to more awareness about indigenous collaboration and inadequate consent [105]. While some international or national collective rights may be outlined, how they are respected is still contested.

1.1.1.2 Legal Consent for Data Sharing and Processing

Shifting from bioethical consent for biomedical or health purposes, consent is also a requirement for lawful data sharing and processing in many jurisdictions, including the EU. Similar to the free, prior, and informed consent in international declarations, one of the currently active, core EU regulations regarding Informed Consent (IC) is the GDPR. The GDPR applies to any situations when processing or sharing personal data, which is defined as any information that related to an individual who can be identified (directly or indirectly). Consent is one of the legal grounds for processing personal data. This regulation governs any personal data, which can be used to identify a natural person (a "data subject") directly or indirectly. The data controller (the party in charge of the data that determines the purposes and processing) must collect consent that is: "a clear, affirmative action" and "freely given,

specific, informed and unambiguous" (Art. 4(11)); easily withdrawn (Art. 7(3)); presented in an intelligible and easily accessible form using clear and plain language (Art. 7(2)); explicitly given for biomedical and genome data categorized as sensitive data (Art. 9); transparent in terms of completeness, comprehensibility, and accessibility of the information disclosures (Art. 12, 13 and 14 GDPR); and compliant with the principles of data protection by design and by default (Art. 25 GDPR) [336]. The processing of sensitive data has stricter requirements (Art. 9), one of which is consent. On the other hand, there are other legal bases for sensitive and non-sensitive data, such as if the processing is part of a contract with the data subject or legitimate interest (which may include marketing, Information Technology (IT) security, etc.). The data processor is the legal term for the entity that is analyzing or processing data, but is not responsible for collecting legal consent. There may be multiple data controllers, depending on how the responsibility for overseeing data processing and overall goals are set out.

IC is also closely tied to GDPR transparency requirements. In addition to the main body of the GDPR, GDPR recitals offer more clarity on how different articles should be interpreted. Recital 39 of the GDPR clarifies the principles of data processing and the goals of transparency: "*It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed ... any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used*" [336]. This aligns with the concept of information transparency, which is defined by how complete the information provided by firms is regarding their business activities [342]. In addition to the text of the GDPR, the European Data Protection Board (EDPB) also endorses the Article 29 Working Party Guidelines on Transparency [243] and Consent [245], which aims to guide the interpretation of the GDPR in more detail. They offer specific guidelines for implementation, such as that consent cannot be hidden in long policies, and that the data controller must ensure that consent should data subjects to easily identify who the controller is and to understand what they are agreeing to. To do that, the controller "*must clearly describe the purpose for data processing for which consent is requested*.[2]" To accomplish this, they also suggest using layered information techniques to display high level and granular information [245]. The consent guidelines then point to the guidelines for transparency which goes into detail about each of the GDPR requirements. For example, they state, "*For complex, technical or unexpected data processing, Article 29 Working Party (WP29)'s position is that as well as providing the prescribed information under Articles 13 and 14 [...], controllers should also separately spell out in unambiguous language what the most important consequences of the processing will be*.[3]" They also have clear examples of how to avoid vague qualifying terms and suggestions for transparency-enhancing methods like cartoons, infographics, pop-up notices, or appropriate methods given the circumstances (digital, in-person, age of audience, etc.). There are many possible tools to enhance transparency and many concepts to explain, which is acknowledged by the WP, "*There is an inherent tension in the GDPR between the requirements on the one hand to provide the comprehensive information to data subjects which is required under the GDPR, and on the other hand do so in a form that is concise, transparent, intelligible and easily accessible*.[4]" Following the principles of accountability and fairness, the WP unequivocally states that the responsibility falls on the data controller to, "*undertake their own analysis of the nature,*

[2]Page 14 Article 29 Working Party (WP) Guidelines on Consent under Regulation 2016/679
[3]Page 7 Article 29 WP Guidelines on transparency under Regulation 2016/679
[4]Page 18 Article 29 WP Guidelines on transparency under Regulation 2016/679

circumstances, scope and context of the processing of personal data which they carry out and decide, within the legal requirements of the GDPR and taking account of the recommendations in these Guidelines particularly at paragraph 36 below, how to prioritise information which must be provided to data subjects and what are the appropriate levels of detail and methods for conveying the information.[5]" [243].

Requirements by Data Type Legal consent also varies by data type and purpose. In this work, we focus on consent for primary uses, which must undergo a complex web of considerations in the EU based on the context. Recital 33 of the GDPR regarding scientific research acknowledges that not all purposes for processing may be known at the time of collection, so if "recognised ethical standards" are followed and if participants can choose to consent "*only to certain areas of research or parts of research projects*" then it would be an exemption from specific consent [336]. Recital 51 delves into sensitive personal data and sets an exemption from consent for using sensitive data for medical research for "the public interest" or for "compliance with legal obligations." This consent exemption for research purposes must be accompanied by appropriate safeguards (Article 6(1)(f); Recitals 47, 157 GDPR). However, the safeguards are not defined and there are many legal instruments (e.g., the Declaration of Helsinki, the Declaration of Taipei) that may apply in different cases [311]. These different instruments give people different rights (but not as many as consent), which Staunton et al. have explored more in their paper. The EDPB also clarified that processing health data for reliability and safety purposes may not require consent, but processing solely for research purposes may fall under consent, legitimate interest, or public interest [33]. If data is used for a clinical trial, the Clinical Trials Regulation also applies [338]. In this, consent is an ethical safeguard for participating in a clinical trial that may be complemented by the legal consent for data processing per the GDPR, but they are not necessarily the same [74]. If the data is digital, it may make sense to give consent electronically. The possibility of signing electronically must be satisfied by the electronic identification trust services (eIDAS) in the EU [339] for how to collect a proper signature. However, national regulations may have more specific restrictions for electronic consent if it is allowed, such as checking the ID of the person signing in Belgium to confirm the identity. However, Swiss clinical trials are the only ones that require a handwritten signature, though electronic tools can be used elsewhere in the consent process [74]. This is just an overview of some of the many considerations involved when using different types of data for different primary purposes, which one must investigate for a specific use-case each time.

Tensions between legal and ethical consent EU regulations and ethical guidelines can differ or conflict, which leads to complicated scenarios when the data is used for research. The GDPR sets very high requirements for specific IC while broad consent may be accepted under certain conditions [213, 125, 20], as in Recital 33 of the GDPR where future purposes are yet unknown. While contested [51, 127], broad consent is often used in biobanks [127, 213] and for genomic data [125, 20], and some of the specific requirements of Recital 33 might be hard to offer (e.g., the right to opt out of certain categories). In addition, while the GDPR refers to ethical safeguards, they keep it open to interpretation. In international ethical guidelines (e.g., The Declaration of Helsinki, Council for International Organizations of Medical Sciences (CIOMS) guidelines), research ethics committees play an important role and

[5]Page 18 Article 29 WP Guidelines on transparency under Regulation 2016/679

are often used to help justify exceptions to consent [108]. Even more problematic, the European Health Data Space (EHDS) proposal was accepted in April 2024 and will go through the adoption process until the fall of 2024, after which it will be published and go into effect. The EHDS will cover all material from humans [335]. While the current draft says it builds upon the GDPR, it reduces transparency requirements in favor of oversight from national health data access bodies and mentioned opt-out consent for primary and secondary purposes (Art. 8(e)(h), 48a EHDS draft [6]). This contrasts the GDPR's requirement for affirmative, opt-in consent. While many things will be open to states to interpret and implement, the EHDS at the current stage does not center data subject's rights [196]. The EDHS will be an area for much future research.

1.1.1.3 Models of consent

Here I will examine the different models and implementations of IC. Models of consent stem from biomedical research and are considered to be, "*study specific consent that is given by research participants after being adequately informed about the aims, benefits, and burdens of a particular study.*" [361]. Wiertz also posits that different models of consent have been established to address different consent challenges and address different ethical assumptions, such as self-determination, public good, and autonomy [361]. I have included collective consent, though Wiertz does not. The main benefits of different models of consent are displayed in Figure 1.1, and will be discussed in this section.

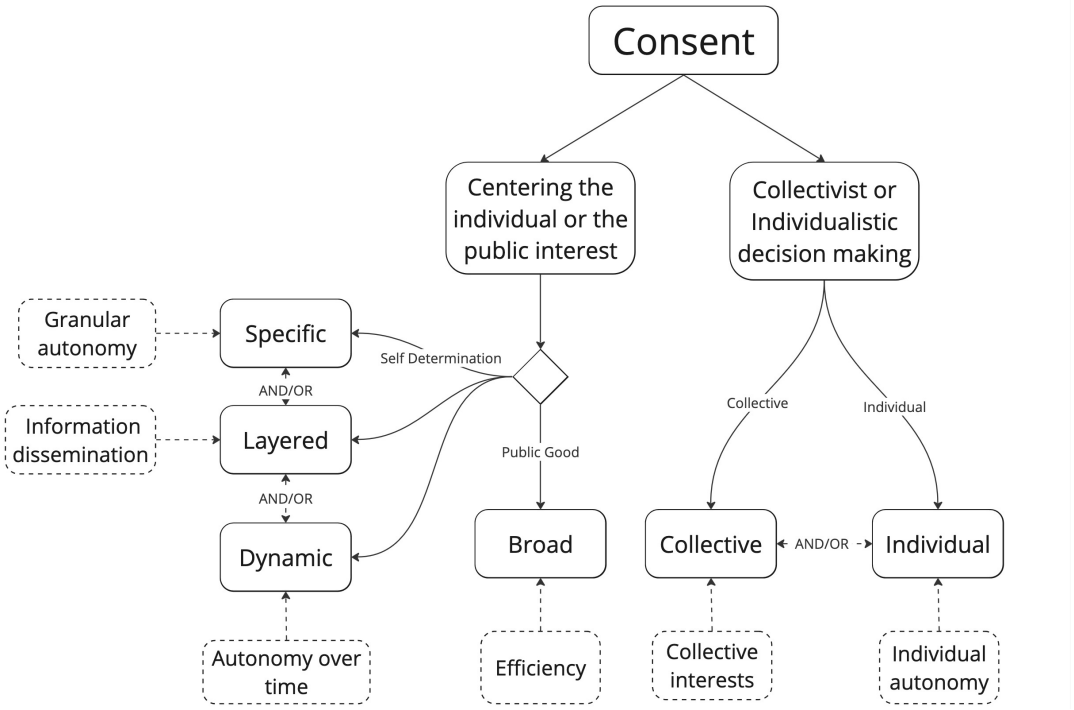

FIGURE 1.1: Consent models theoretically categorized by ethical assumptions from [361], including an extension for collective consent. The dotted squares indicate the consent challenges the model of consent seeks to address. Consent models may also be used together or separately and are shown in the AND/OR relationships.

[6] European Parliament legislative resolution of 24 April 2024 on the proposal for a regulation of the European Parliament and of the Council on the European Health Data Space (COM(2022)0197 – C9-0167/2022 – 2022/0140(COD))

Broad consent Broad consent has been widely used in many studies and biobanks. One large project is the 100,000 Genomes Project, a large clinical research initiative created to sequence whole genomes from UK National Health Service (NHS) patients. Consent was a large concern in this study, and due to challenges with specific consent, participants were asked for broad consent. For example, "*participants were at some points asked to make broad rather than specific decisions; for example, they could decide whether to receive [additional findings] or not, but they could not pick and choose what these [additional findings] might be*." 1337 participants were surveyed, with 24 follow-up interviews. From this, it was revealed that most individuals were satisfied with their experience [20].

Broad consent is also used by the Lausanne Institutional Biobank, a hospital-based biobank. In this model, "*individuals consent to the broad, open-ended use of their samples and medical data in research, including genome analyses, as long as these projects have been approved by the local Institutional Review Board. In addition, BIL participants could accept or not to be re-contacted if clinically actionable findings are found in research*" [21]. From interviews with biobank recruiters and participants, it was found that most participants relied "*substantially on the perception of the recruiter and/or the institution. Therefore, the decision to participate or not in a biobank may depend as much on participants' past experiences and beliefs [...] as on information and participants' assessment of risks and benefits*" [21].

There have been several proponents that say that it is logistically feasible for governance to enact and sufficient to satisfy GDPR requirements, given that broad consent is in-depth [125] [127] [213] and opponents that say relative costs, benefits of public research, and the concept of minimal risk for participants warrants traditional (study specific) consent [51]. For health research, Recital 33 of the GDPR offers some exemptions since "*[i]t is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research in accordance with recognized ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose*." However, the EDPB suggests that if applicable, participants should be asked to re-consent to further data processing – a requirement which may be more suitable for other models of consent [59] [108].

Dynamic Consent Dynamic Consent (DC) is a broad category of consent that involves an electronic platform to facilitate the process between participants and researchers, data processors, and governance. DC may cover many types of specific implementations, all tackling the issue from different angles and offering different innovations.

The EnCoRe project's DC was a pilot to propose a solution for privacy-preserving data sharing and withdrawing consent with the concept of "wrapping" the participant's consent preferences in "sticky policies" with a trusted authority managing access to the data [157]. These policies are attached to data as they are shared for data processing and uses new homomorphic encryption and a trust authority that can revoke decryption rights, effectively locking the information in the file so it may not be used if participants withdraw access to the data [151].

DC platforms can be used for other purposes, for example, CTRL in Australia [123] created an interdisciplinary team that included, "*two consumer representatives and a patient advocate [...] a clinical geneticist; genetic counselors; bioethicists; a patient and community education and engagement professional; health technology and data management experts and [researchers]*." They helped guide the design of CTRL to integrate

genomics with healthcare services by building a DC tool offering participants greater control over their data and data sharing, data processing, and access to researchers. They offer more fine grained controls (including easy changes to consent), access to results, and ways to message or contact researchers via the platform. Instead of building a legal compliance architecture, they focused on user requirements, user interviews, and data sharing using the Global Alliance for Genomics's Health Data Use Ontology [175].

Another DC platform for a well tested longitudinal health study, CHRIS in South Tyrol, Italy [28], has also built a custom platform to better engage users. Although opponents criticize DC as "*too complicated, too resource heavy, and to lend itself to consent fatigue*" compared to broad consent [328], broad consent was used in this study after consultation with participants who felt it was burdensome to continue to consent for specific purposes. They used a DC platform to respond to the needs of researchers, participants, and the study itself, which has been running for more than ten years and improved the relationship between researchers and patients in the community, as well as specific governance mechanisms and communication tools [28, 199]. More in-depth, participants agreed to broad consent for the collection and use of data and samples within the scientific areas of interest of the CHRIS study and the limits of the informed consent. As time went on, participants would receive information about new studies through an app, webpage, or newsletter and may be asked to re-consent.

One possibility is also building a usable DC management platform with underlying standardized consent vocabularies and receipt specification systems that help to record and automate the record-keeping of decision-making [143]. The EU project TRAPEZE [35] has integrated and helped to develop a GDPR compliant ontology. This helps to record consent for accountability purposes and transparently show how individuals' data is processed along with a consent management platform that allows participants to control their data sharing and view the data processing [163]. The OWL2 (Web Ontology Language, https://www.w3.org/TR/owl2-overview/) is a World Wide Web Consortium (W3C) standard [35] that may be integrated with the W3C's Data Privacy Vocabulary (DPV) [240] community group that has GDPR vocabularies and covers data uses and purposes as well [277]. This was happening along with more user-testing of graphical dashboard DC with pilot scenarios from the government, telecom, and financial sectors [35]. This vocabulary can also be extended with the Global Alliance for Genomics's Health Data Use Ontology [238]. Consent receipts also comply with International Organization for Standardization (ISO) standard 27560 on Privacy technologies — Consent record information structure [3], and standardized consent structure with ISO standard 29184 on Privacy Notices and Consent [1, 239]

While individual implementations may be successful as certain aspects true "dynamic" consent may still be difficult to implement. A review paper broke DC into defined several sub-requirements: dynamic permissions, dynamic education, and dynamic preferences, and compared different implementations of DC based on those criteria. They identified different parts of the IC process and set specific criteria, such as in the dynamic permissions there must be an online portal to share information and allow for changes, or in the dynamic education inclusion of timely notices about research opportunities, or in dynamic preferences participants must be allowed to indicate their preference for how to receive information. They noted that most studies fulfilled most of the criteria, but some were still lacking, especially in the education section. In addition, most relied on broad consent which may decrease a participant's autonomy by not allowing for more granular participation in specific studies. Since each DC is tailored for specific projects or purposes that may not fully

fulfill all the criteria laid out and/or be applicable to other uses, some criteria such as dynamic preferences were not met [69]. These would be useful requirements to keep in mind for future iterations to be truly "dynamic".

Tiered Consent Tiered consent encompasses three different sub-concepts that all form a personalized approach to consent [43]. Tiers refer to categories that would be meaningful for participants, such as purposes of testing or the impact of the test results. Then the tiered consent is also layered so that the information is extendable upon request, so essential information is given to all participants while more detailed information is shared for those who actively seek it. Layering for notices is a concept that is also discussed by the Article 29 Working Party in their Guidelines on Transparency [243]. Last is the concept of stages, or step-wise informed consent that may include an initial sharing of information before consent, then a follow-up that may require separate discussions before consent can be given. This model was developed in reaction to personalized genetic testing services that involve complex genetic and health information given to a participant who may not fully understand the risks, benefits, and concepts – and the study itself may not know the full scope of the research to be carried out [43]. Tiered consent is a user-centered perspective breaking up a large amount of information into tiers, layers, and stages for when participants may make decisions about consenting to data sharing or genetic testing with specific information likely needed to make more informed decisions about whether or not to consent.

Tiered consent has been used in Africa with the H3Africa Consortium, a large collection of projects applying genomics to African populations for disease purposes. Of 13 projects, 5 used a tiered consent model, 1 used specific consent, and 7 used broad consent [221]. Another project used tiered consent for electronic health record data in research, called iCONCUR. It was piloted in 4 outpatient clinics of an academic medical center with 394 patients who agreed to participate. It was found that "[p]articipants indicated that having granular choices for data sharing was appropriate and that they liked being informed about who was using their data for what purposes, as well as about outcomes of the research." [160]

Tiered consent seems to be less resource-heavy for studies that may want a middle-ground between study-specific consent and broad consent. It may also offer participants options that are desired, as a study that compared tiered consent to traditional or binary consent found that "*most participants are willing to publicly release their genomic data; however, a significant portion prefers restricted release*" [206]. After all participants were shown the options, 53% chose public data release compared to the 83% before being shown traditional, binary, and tiered consent. More participants chose restricted release or opted out entirely[206].

Collective Consent Group or collective consent is a broad description that generally includes a group or collective of individuals all involved in the consent process. This is elaborated by "*1) One could mean by group consent, the unanimous consent of individuals who all share a common culture. Presumably, individuals could make their individual decisions based not only on the basis of their self-interest but also on what impact they perceive it might have on their cultural group (for example, the case of African Americans and sickle cell testing). 2) Alternatively, group consent could mean the consent of a collection of individuals who share not only a common culture but also a common decision-making process*" [288]. Group consent is not a replacement for individual consent, but may increase understanding of risks and benefits to the community, help researchers be

more accountable to more relevant parties, and allow researchers to create rapport with other stakeholders [138]. Theoretically, collective consent can be delineated into two actions with different dimensions, collective decision-making and collective consent [132]. For this thesis, we focus more on collective consent, and discuss decision-making only insofar as the information given along with consent to make the decision.

This is especially relevant for genetic health data since it can "*uniquely identify individuals [...] thus providing excellent biometric information (that is, a genomic 'fingerprint').*" [36]. Only 75 statistically independent single nucleotide polymorphisms, a marker of genetic variation, are needed to identify a unique individual across the global population [183]. Thus, even sharing 'aggregate' data and not individual-level data can be a risk to individual privacy. It may also pose a risk of identification for relatives of the individual without their consent. For example, in the Golden State Killer case, DNA from the crime scene was analyzed with an open-source genealogical website that uses aggregated data for people to search for relatives. A family tree was constructed and the crime scene DNA was attributed to be a 4th cousin [359] to a member in the database. A fourth cousin only shares one set of great-great-great grandparents with another individual and may be implicated without their consent. This begs the question of what group rights exist or should exist, and if group consent might address some of these issues.

There are several use cases where group consent has been heavily championed. One is population genetics and another is research with indigenous populations. In population genetics, it is argued that the population itself is the subject and may face negative consequences such as discrimination or negative stigmas. Also, the population of interest may have a governance or leadership structure of its own and should be consulted [114]. Similarly, this is also relevant to indigenous populations or groups with non-Western cultural governance and leadership. For example, "*[i]n the case of informed consent, recognising mechanisms of collective consent that provide a more robust process for evaluating the outcomes of research for the community (issues of external ethicality) would be welcomed by many indigenous communities*" [138].

However, collective consent is difficult to put into practice and is still debated because defining the group to request consent to and using or building a collective decision-making system is complex. One researcher suggests using group consent where governance structures already exist for collective consent [150]. However, it still can be difficult to understand how formalized such structures must be – would an elected tribal leader or a de facto family head count? Another scholar suggests making the requirement for group involvement less stringent, such as through procedures of consultation [80] to ethically advance research while keeping logistical constraints in mind. However, the consultation process addresses the group decision-making, but not the consent process itself. After more than 20 years, the debate continues on how to establish groups [103]. The Model Law on Health Data Governance is a draft that aims to set a global standard for health data legislation, including guidelines for collective rights given that there is a governing structure, but sets the burden on individual states to define what collectives are and what governing structures would be acceptable [39]. In this thesis, groups are not delineated using either definition, but based on context if given a specific use-case.

1.1.1.4 Consent Challenges in Practice

Health data is a sensitive type of data and thus has some unique complications and solutions for transparency since there is a tension between sharing health data to

advance science and securing and maintaining the privacy of sensitive data. There has been increased digitalization of many types of health data due to advances in genome sequencing, personalized medicine, electronic health records, and more. The transition from biomedical consent to digital consent for legal data processing has exacerbated existing challenges and created new ones.

To align with individual preferences and increase public trust and engagement, individuals should be able to see and manage the access and use permissions they have given regarding their personal data over time [44]. Not only should data subjects be able to audit their own and a data controller and processor's actions, but data controllers need to be auditable as organizations. Data controllers should have a clear record of actions to be lawful, for example, data flows, privacy and security measures, or consent management. For example, they could use a consent receipt specification that is machine-readable, such as KANTARA [143], which has been cited in the ISO/IEC 29184:20208 standard for Online Privacy Notices and Consent [1] or the DPV (https://w3c.github.io/dpv/dpv/) [240] and Open Digital Rights Language (ORDL) (https://www.w3.org/TR/odrl-vocab/), are machine readable policy languages that can be used for consent processes (as discussed in Section 1.1.1.3).

For data controllers, it can be useful and legally required to be able to track and share how privacy by design was used in the governance model. However, more transparency may not necessarily help the end-users, as increased demand for user cognition can lead to overload and it should be appropriately framed to communicate accurately to users [173]. The transparency and understandability of consent notices and information about data governance is often seen as a trade-off through a transparency paradox, in which increased transparency may decrease understandability, and increased understandability may require decreased transparency [225]. Moreover, design choices can impact data sharing, as the way health data sharing information was presented to users can help individuals better understand their options and indicate their preferences (e.g., privacy-protective defaults or user-relevant terms) [146]. Johansson et al. conducted interviews and discrete choice experiments to define the key types and levels of information for health data sharing for research purposes, which can be used for future projects (e.g., the granularity of the data controller should encompass the type of company, but more detail is not required). Some design choices can also manipulate them into mindless data sharing (e.g., privacy-invasive defaults) [5]. For example, dark patterns, one method of manipulating design choices, in cookie consent are incredibly pervasive. In a study that tested how design choices affect user decisions, it was found that offering more granular control of consent on the first page decreased consent by 8-20 percentage points [229]. This higher transparency allowed the user to make more mindful decisions. However, the understandability of consent is another issue. There are often barriers to understanding the text which led to many efforts to make the reading level easier and more accessible to a larger portion of the population [368]. One review also found that, in addition to text, verbal communication helped improve understanding [321].

While navigating the tension between transparency and understandability seems impossible, research involving users in the design process has shown to be fruitful. For example, in the study by Johansson et al.[146], they involved many patients and experts over time to develop governance terms and definitions that would be the most useful to explain health data sharing for research. The aforementioned CTRL DC platform was evaluated in a follow-up study where the platform was offered as

a secondary consent method and patients could sign up voluntarily and were monitored for 22 months. Patients of varying ages, educational levels, and sex showed a positive reaction, without evidence of 'consent fatigue' with high completion rates and showing that they took advantage of more granular choices for data sharing, receiving information, and contact frequency [122]. The CHRIS DC platform also had a high involvement with the community of South Tyrol, using multiple patient engagement techniques including surveys, field observations, and interviews [199]. Data has been collected over the 10 years and continuous improvements have been made; for example, study assistants reported that using a video decreased the time needed to explain the study from 20 minutes to 5 minutes [248]. User involvement is on the rise, with a review showing increased user satisfaction across many studies [170]. Another review showed that increased satisfaction, ease of use, and a positive impact on sales were felt, however, the methods greatly varied and should be chosen with care to maximize the benefits relative to the costs [350]. In the health field, patient design is on the rise as well [67, 204] with many studies regarding consent [126, 264, 97], including DC examples mentioned before [199, 123].

1.1.2 Datafication

As we move from the world where consent was originally conceived for medical procedures to the age of big data, algorithms, and AI, the context for data sharing also changes. Now, one gives consent (sometimes) to be datafied. In 2013, Mayer-Schönberger and Cukier wrote that data is a "*significant corporate asset, a vital economic input, and the foundation of new business models. It is the oil of the information economy*." Any and all parts of life are transformed into data to create value in a process called "datafication" [202]. While one might consider health data a common type of data, the things this includes greatly expands from information with known outcomes (e.g., smoking status) to any information that may be useful for health reasons (e.g., a yet unstudied region in DNA, the time of day one visits the doctor, etc.). Not only is datafication an economy, but Zuboff's work places it within surveillance capitalism – a complex economic model of surveillance based on new methods of extracting data and knowledge production [374]. An important point is that the end goal of surveillance capitalism is to gain power, and the means is technology, which itself is neutral. Zuboff points to the history of surveillance capitalism through Google's use of behavioral data to match advertisements to users to become profitable, in that behavioral data are, "*more than what is required for product and service improvements*." This becomes part of a new revenue stream to sell and shape predictive behaviors, where users are "*a means to profits in a new behavioral futures market in which users are neither buyers nor sellers nor products*." [373] The sociopolitical situation from which surveillance capitalism arose was largely unregulated in the 2001s and the USA also had a great interest in surveillance (post 9/11) [304], so "they ignored privacy norms and laws, [...] and meaningless mechanisms of notice and consent to accumulate decision rights" [374, 19]. They also understood that they were not a search engine, but Larry Page is credited with answering the question, "What is Google?" by saying, "*if we did have a category, it would be personal information*." [85]

What are data, information, personal information, and datafication? From a scholarly point of view, data is often defined as the "*raw material for information, and information is the raw material for knowledge*" [372]. Following Zin's survey of 57 researchers for their definitions of data, information, and knowledge for the Information Sciences field, data surrounding objects were considered as, "*sets of signs that*

represent empirical stimuli or perceptions." Information is a, "*set of signs, which represent empirical knowledge*," while knowledge is, "*a set of signs that represent the meaning of thoughts that the individual justifiably believes that they are true*." [372] Within this framework, the GDPR defines personal information as a subset of information which are, "relating to an identified or identifiable natural person" [336]. According to a ruling by the General Court of the European Union (a constituent court of the Court of Justice of the European Union (CJEU)), the boundaries between pseudonymized personal and anonymized non-personal information should be determined on a contextual basis [92]. If a third party is given information by the data controller that is "effectively anonymized" in that the third party cannot reasonably de-identify the pseudonymized data, it can be considered anonymous for that context. This contrasts with opinions from a technical perspective because any information is impossible to anonymize and future-proof (e.g., inferences made with increased data or more sophisticated algorithms) and de-personalization techniques may always have a risk of failure [98]. However, the court decided that if the information is sufficiently pseudonymized so the recipient cannot reasonably infer identity, then it is sufficient to be non-personal information. This tension between the court ruling and the inadequacy of technical data protection measures can cause legal uncertainties which undermine data protection laws [187]. In a later section, we discuss genetic privacy and challenges to sufficiently anonymize genomic data (see Section 1.1.3), which may complicate applying this case in genomics.

1.1.3 Privacy, Consent, and Datafication

In the face of datafication and surveillance capitalism, where does the aforementioned steamrolled privacy and consent stand? First, we explore informational privacy as it relates to consent and sharing information, then raise challenges related to collective genetic data and non-individualistic theories of privacy.

Informational Privacy While many types and definitions of privacy exist (for example, Floridi distinguishes between physical, decisional, mental, and informational privacy for computer ethics [99]), for the purposes of this thesis, the key type we explore is informational privacy. Informational privacy was proposed by Westin in the 1960s as, "*the claim of an individual to determine what information about himself or herself should be known to others*." [357]. In the digital age, it can cover all aspects of "*personal data that is both stored in and communicated between electronic databases [...] and personal information communicated between parties*" [322]. As informational theory has evolved, some scholars have taken a restricted access approach (where protected, private zones restrict information) [8, 107] and the control approach (where someone has control over their personal information) [101, 260, 357]. Both approaches have benefits and drawbacks, and the Restricted Access/Limited Control Theory attempts to bridge the two approaches while adding a distinction between privacy management (via some individual control) and privacy as a concept (defined by restricted access) [219, 323]. In this, privacy norms help to define zones of protection in addition to limited individual controls, since once someone shares medical data, a multitude of qualified healthcare personnel could access it if given the rights [323]. Norms are also part of Nissenbaum's theory of contextual integrity [226]. This theory defines privacy as aligning with contextual norms in different spheres or contexts, and information must follow norms for appropriateness (e.g., sharing health data with a doctor and not a salesperson) and norms of distribution (i.e., how the information is shared or restricted) [226]. Nissenbaum's theory emphasizes the need

to extend privacy norms, which can be explicit privacy laws or informal privacy policies, to any context where information is flowing [226].

Consent can be seen in different lenses based on the theory of privacy involved: a normative method to control distribution under the theory of contextual integrity, or an act of self-determination via expressing control or restricting access. In addition, it can also signal privacy preferences and offer a formal notice of the potential consequences to users [65].

Consent is a pillar in data protection law and privacy theories though it is critiqued for lacking "informedness" and not being "freely given" due to the information asymmetries between the technological providers and data subjects [286]. This is often referred to as the consent paradox, and one scholar suggests that such critiques, and subsequent efforts to mitigate information asymmetries, keep consent relevant and useful [25]. Bergemann suggests that the GDPR acknowledges the pitfalls of consent and tries to reform it by stating that it must be freely given and informed, along with recitals and guidelines to clarify best practices. Many challenges to consent were explored in 1.1.1.4. Even as Solove critiques consent and proposes to give it less legitimacy, he also writes that one of the duties is to "obtain consent appropriately" [308].

The EU is also interested in improving consent and protecting privacy, as can be seen in the court case between Interactive Advertising Bureau (IAB) Europe and the Belgian Data Protection Authority (DPA), which went to the CJEU. Historically, the frameworks for implementing GDPR compliant digital consent were written by the IAB Europe in 2018, which was founded mainly by advertising technology companies. Research has shown that the Transparency and Consent Framework (TCF) to collect cookie consent on websites was nudging users, storing cookies without consent, or storing cookies before any decision is made [200]. In 2022, a decision was made that a central part of their framework enabling publishers to sell ad space on their websites, the TCF, is unlawful under EU data privacy laws on two bases - personal data stored within the cookie and the lack of joint controller responsibilities from the IAB [53]. Both consent pop-ups and the data collected to be auctioned off to advertisers were found illegal. IAB Europe then responded by saying that the framework itself was not banned and that they are seeking to reform the process, and they had since released a new version of the TCF. In 2022, IAB Europe appealed the decision, arguing against the claims they were processing personal data as joint controllers. Since then, the TCF seems to have moved to another legal basis for processing data, legitimate interest, which is often conflated with consent [220]. In 2024, the CJEU decided that the cookie data could be considered personal data under certain circumstances, for example if associated with other data (e.g., IP addresss) [334]. Second, they decided that IAB Europe has "reasonable means" to identify a data subject from the given data since others participating in TCF are required to provide it. Third, the court found that IAB had a vested interest in digital marketing, and determines the means of processing through the TCF framework, thus could be considered a joint controller. However, this does not extend to further processing of the data by third parties (e.g., website owners). Through the example of complex cookie consents, the EU clarified the concept of "personal data" and "joint controllership" under the GDPR to help safeguard personal rights.

Genetic Privacy Challenges As a specific category of health data, genomic privacy is especially fraught due to the uniqueness of collective data, "*that is the intrusion on group or collectivity privacy interests*." [9] Though one database may be anonymized, it has been shown that individuals can be re-identified using other public databases,

or data that has been partially redacted could be uncovered. Using only publicly available demographic data with anonymous profiles from a genome project, individuals' names could be found [318]. With more information like genome sequences, not only individuals could be identified but also relatives who had not shared any genetic information [89]. In addition to the issue of reidentification, there is the privacy issue of phenotypic inference, where genomic data could be used to infer a phenotype (or outward sign of genome expression) such as disease information. Even though a genome may mask an area of risk, it could be inferred by public genomic data [230].

Group privacy is often used to discuss collective information, where a group is "a collective of individuals who are culturally or ethnically related, where shared genetic characteristics are either likely or possible" [9]. In this, families are generally included but not the focus; rather defined groups with demographic labels like Native American, Ashkenazi Jew, etc. are the focus of the group.

However, in the age of big data, defined groups are not the only ones at risk of genetic privacy violations. While many theories of privacy exist to address the datafication and arbitrary aggregation of groups by algorithms, we look at two in the following section.

Networked Privacy Boyd writes is a concept where the entities involved and boundaries between private and public are unclear [40], as with genetic data. This contrasts with the individual or defined group or list privacy controls. Boyd writes, "*Through this [genetic] test, I learned information about myself, but I also learned information about members of my family. Furthermore, by choosing to subject my DNA to this testing process, I didn't just reveal data about myself; I gave away data that provides insights into my mother, brother, grandparents, and even children that I don't yet have.*" Not only is the data networked in and of itself, genetic data is also an example of networked data where the aggregation of information leads to new insights [18]. Similar to other types of big data, genetic databases could be combined with other databases to reveal more information about individuals or groups. This is further developed in a study about social media data, wherein "Networked privacy invokes the constellation of audience dynamics, social norms, and technical functionality that affect the processes of information disclosure, concealment, obscurity, and interpretation within a networked public. [...] Achieving privacy requires that people have an understanding of and influence in shaping the context in which information is being interpreted." [198] This has been used as one of the explanations for the privacy paradox [129], which showed that young adults were concerned about privacy yet apathetic about privacy violations because they felt it was out of their control.

Similar to the other theories of privacy discussed, there are spheres with different contexts, but now they are blurred, and restricting access or controlling information may be insufficient for the complexity of the blurring spheres. Instead, "*it requires meaningful control over the networked contexts in which the information flows. In other words, achieving privacy requires that people have an understanding of and influence in shaping the context in which information is being interpreted. This can be done by co-constructing the architecture of the systems, or it can be done by embedding meaning and context into the content itself.*" [198]

Collective Privacy Instead of the premise of unclear boundaries between public and private, collective privacy focuses on the types of groups created by big data. As

more and more genetic data is collected and used in big data analytics, it may transform from known groups (genetic relatives, genetic minorities) to much more abstracted aggregations of minor signals for the creation of a group. In this case, group privacy is also insufficient [193, 194]. Mantelro writes that, traditionally, "*group privacy considers groups that are based on stable and socially recognized relationships between individuals, although they can be informal in nature (e.g., love affairs, priest-penitent relationships) or last only for a certain time (e.g., marital relationships, association)*" [193] (p.144). With big data, "clusters of individuals" are more accurate to describe the type of group. Mantelro then offers the definition that, "*collective privacy can be described as the right to limit the potential harms to the group itself that can derive from invasive and discriminatory data processing [...] The source of concern is not the lack of secrecy and intimacy, which represents the object of group privacy [30], but the unfair and harmful use of data that is processed by using modern analytics*," which are relevant even when there are conflicting opinions between members [193] (p.148). This offers an interesting point of view for collectives with varying opinions (non-aggregative). Mantelro writes that for these cases that the core societal values influence the nature of the interest such as environmental protection, equality, or freedom [193].

While networked privacy focuses on information and control, collective privacy focuses more on reducing harm from discrimination or surveillance. For the purposes of this thesis, both perspectives will be considered.

1.2 Thesis Overview

1.3 Summary

Sharing health data can offer many benefits, from more personalized care to a better understanding of human health. For example, genetic testing can help people to determine which medication may be most effective or those with a family history of a disease to understand their predictive risk. However, it is also sensitive and sharing it could have many risks, which is especially true for genetic data. For example, the data might reveal family member's conditions, or a genome might lead to personalized insurance premiums as well as personalized medicine. These potential harms should be addressed on an individual, group, and governance layer to protect privacy and security while also maximizing the benefits.

One of the ways to address the individual and group level is through collective consent, to offer relevant individuals a framework to consent to health data sharing. Collective consent comes from indigenous bioethics, in which indigenous tribes fought for their right to consent to biomedical research as a community, not just as individuals. It has been used in research partnerships with indigenous groups to improve stakeholder involvement instead of treating indigenous populations as test subjects.

Though it has been proposed, no digital collective consent (wherein multiple individuals consent in different via different governance structures such as families or tribal leader) exists yet as there are complex issues such as the conflicting right not to know and the duty to report, the mechanism of decision making, and governance structure of the collective are undefined. Generally, different legal-ethical considerations, technical properties such as interoperability, transparency, scalability, and core concepts such as trust, that all must work together to support digital collective consent.

As more and more genetic data are collected, stored, and shared in addition to different metadata, the genomic privacy risks also increase. The more datasets that are available, the more re-identifiable someone can be. For example, combining different datasets (e.g., public, controlled access, population level patterns) can re-identify participants from anonymized databases, and nonparticipants may also have their genomes and traits exposed. With the rise in electronic health records, genomic data, and digital recordings, individual consent is insufficient to address the highly interconnected world. Both the benefits and risks should be shared between affected individuals, and everyone should have a say in data sharing that could equally help or harm them.

Consent is not easy for an individual to understand and may be even more challenging for a collective. There may be different levels of digital literacy, reading level, or data management styles (e.g., someone who wants granular controls vs. someone who wants to consent once) from the user. From the system, there may be different levels of informational transparency, integrations with storage and access controls, and design. Overall, the institution that asks for consent is also a consideration for participants and can lead to trust evaluations that may influence their decision to collective consent or not. For example, if MyHeritage has historically had a data breach, it might decrease trust in participants regardless of how transparent and usable the consent management is.

Dynamic consent is a flexible existing model of consent that offers a patient-centered platform for electronic consent (eConsent) management and interactions with the research project and has been suggested to be compatible with collective consent. However, I have not found any that has studied group or collective decision making for genetic data in a digital context. For example, what level of transparency would be appropriate while also preserving the anonymity of individuals in a collective who share a genetic disease? What attributes are most relevant for an end-user?

In my thesis work, I address the issue of collective consent processes for primary uses of health data sharing by focusing on genetic data and more general health information, such as video data when it will be used for welfare purposes. This type of data is highly sensitive and has unique privacy risks as well as benefits. I will tackle the problem from a computer science, privacy, EU data protection, and biomedical lens and will be analyzing both the notice and consent, as there must be sufficient notice for "informed consent" under the GDPR, which takes a data protection approach to consent. I target end-users and SMEs as my audience, first to better understand end-user needs and second to understand employee needs and possible tools to improve their business processes for end-users.

1.4 Research Objective and Questions

Given the challenges in consent and informational privacy, two of the main focuses of this work will be informational transparency and user-centered design of information to address the "informed" consent issue. This also ties to the unique challenge of collective genetic data, such as communication of risks and potential harms or navigating the balancing of different opinions in a collective. Information also must be relevant and useful, or else it will fall into the transparency paradox. Thus, user-centered design is another key factor that allows for iterative research, enabling a system that addresses user needs. Given the limited amount of prior research on digital collective consent for undefined groups, where empirical research is not

available, the process should be taken from the ground up. Instead of focusing on innovative technology that may not be useful or usable, we will focus on consent needs and goals of potential end-users as well as from SMEs that may be motivated to implement a framework for more transparent, user-centered collective consent that abides by EU regulations. To be useful in the EU necessitates being legally-ethically aligned with EU regulations such as the GDPR. The scope of this work does not cover the EHDS for primary consent, which may change the landscape of health data sharing after it goes into force (at the time of writing it is not yet officially published).

Research Objective: To explore the challenges in collective consent and develop a framework of tools and methods to enhance user-centered, transparent collective consent to help end users and SMEs.

To fulfill the objective, we ask:

Research Question 1 How is notice and consent for genomic data sharing implemented in companies and research settings, and what are open challenges for collective consent?

Research Question 2 What methods, tools, and frameworks can help address the challenges in digital collective consent in SMEs?

Research Question 3 What components can enhance informational transparency in informed consent for users?

1.5 Thesis Overview

This thesis has both empirical and theoretical contributions. The introduction provides a novel overview of the challenges in collective consent integrating views from data privacy and bioethics. Then we dive into EU-specific legal and ethical consent issues surrounding the interplay between hard and soft law, ethics, and how technical solutions such as dynamic consent or standardized vocabularies can accommodate such complexity. Then we investigate the current status of company practices dealing with genetic data, analyzing the informational transparency and user-relevancy of popular Direct to Consumer Genetic Testing Company (DTCGTC) privacy and consent policies from a GDPR and privacy lens. From the analytical methods used, we wanted to see if they could be used *ex ante* in a business, so we applied those methods with a medium-sized welfare-tech company in Norway and collected employee perceptions about usefulness and usability. Following the work about application for businesses, we developed a collective consent BPMN and tested it with the same company to better understand how BPMNs may actually help companies, as it has not been applied and studied with consent before. We then end with a general user study to better characterize user needs for consent, perceptions of different mediums for communication, and consent management desires. This can offer a framework from which to develop more informational transparency, user-centered consent management, and methods with real-world applicability.

	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
RQ1	×	×			
RQ2			×	×	×
RQ3			×	×	

TABLE 1.1: This table displays the RQ and associated chapters that address them in this thesis

1.6 Thesis Organization

Chapter 2: Legal-ethical challenges to consent and technical solutions In this chapter, I examine the legal-ethical uncertainties regarding consent in the EU, such as the individual and collective data subject in the case of genetic data, consent as an ethical safeguard vs. legal obligation, and specificity of informed consent from a legal and ethical point of view. Then privacy assistants, standards and vocabularies, and dynamic consent that might address the legal uncertainty within guidelines and regulations are discussed. This chapter contributes to answering Research Question 1.

Chapter 3: Transparency and relevancy of public privacy and consent policies from Direct to Consumer (DTC) Genetic Testing Companies I then investigated the status quo of individual consent for genomic data sharing with DTC genetic testing companies, which have millions of customers worldwide. This chapter contributes to Research Question 1. I analyzed the transparency of information, the user relevance, and the risks and benefits stated using methods from privacy (the contextual integrity framework), as well as from governance with user-relevant terms from prior research, and surveyed risks and benefits (especially noting gaps in collective risks).

Chapter 4: Information flow and policy elicitation and validation with contextual integrity for collective data sharing I take one of the methods used in the previous chapter, contextual integrity, to test if can be applied to analyze and improve policies from the business side with a SME. We engaged with a medium-sized health tech company in Norway to see how useful and easy to use contextual integrity for privacy policy writing and analysis would be for employees. This chapter helps address Research Questions 2 and 3.

Chapter 5: Visualizing and analyzing collective consent processes with BPMNs In addition to more transparent information flows, this chapter explores if consent processes developed using BPMN notation can be more understandable, communicable, useful, and easy to use for a medium-sized health tech company in Norway. Employees were shown a collective consent BPMN solution and their perception of usefulness and ease of use were investigated to predict potential use of the technology. This helps address Research Questions 2 and 3.

Chapter 6: User goals for consent and engaging design elements from 5 different mediums This chapter surveys a general adult German user to better understand their needs, desires, and attitudes towards different mediums for consent (video, infographic, text, newsletter, and comic) and consent management. Users wanted

quickly understandable, structured consent notices and preferred the infographic to other mediums. They also preferred to manage their consent and revocation centrally. This uses individual consent as an example because users are more familiar with it, but could be applied to collective consent to help address Research Question 2.

Chapter 7: Discussion I discuss the findings across the thesis organized by the initial research question, with a summary of the different contributions. Then I offer a graphical summary and discuss limitations. Lastly, I dive into the future work and extrapolate on the possible outlook of the work in the context of collective data issues and the EHDS.

Chapter 8: Conclusion I reflect on this research project and how different research questions have contributed to different layers within consent processes, from regulatory, governance, business, and user spheres to contribute towards a framework for dynamic collective consent.

1.7 Contributions

Publications as main author

- Doan, Xengie Cheng, Annika Selzer, Arianna Rossi, Wilhelmina Maria Botes, and Gabriele Lenzini. "Conciseness, interest, and unexpectedness: User attitudes towards infographic and comic consent mediums." In Web Conference Companion Volume (ACM). ACM. 2022.
- Doan, Xengie, Marcu Florea, and Sarah E. Carter. "Legal-Ethical Challenges and Technological Solutions to e-Health Data Consent in the EU." In HHAI 2023: Augmenting Human Intellect, pp. 243-253. IoS Press, 2023.
- Doan, Xengie, Arianna Rossi, Marietjie Botes, and Annika Selzer. "Comparing Attitudes Toward Different Consent Mediums: Semistructured Qualitative Study." JMIR Human Factors 11 (2024): e53113.
- Doan, Xengie, Fatma Sümeyra Doğan, and Arianna Rossi. (In press)."Analysis of Transparency and User-relevancy of DTC Company Policies." Privacy Symposium 2024.

Chapter 2

Legal Ethical Consent in EU

This chapter adapted from: Doan, Xengie, Marcu Florea, and Sarah E. Carter. "Legal-Ethical Challenges and Technological Solutions to e-Health Data Consent in the EU." HHAI 2023: Augmenting Human Intellect. IOS Press, 2023. 243-253.

Author Contributions: Conceptualization, Investigation, Writing - original draft, Writing, review & editing

Abstract: e-Health data is sensitive and consenting to the collection, processing, and sharing involves compliance with legal requirements, ethical standards, and appropriate digital tools. We explore two legal-ethical challenges: 1) What are the scope and requirements of digital health data consent? 2) What are the legal-ethical reasons for obtaining consent beyond the GDPR's legal basis, and how might such consent be obtained? We then propose human-centered solutions to help navigate standards of ethical and legal consent across the EU, purposefully addressing those use cases to compensate for human difficulties in managing consent without clear guidelines. These solutions–including ISO standards, ontologies, consent mechanisms, value-centered privacy assistants, and layered dynamic consent platforms–complement and aid humans to help upholding ethical and rigorous consent.

2.1 Introduction

As mentioned in the main Introduction section 1.1.1, consent stems from medical consent as an ethical concept (hereafter called "ethical consent") from the 1940s after World War II in the Nuremberg Code [348]. Informed consent was based on the

principle of respect for autonomy and the dignity of persons. Respect for autonomy is enshrined in ethical guidelines such as the Belmont Report [276, 254] and the Declaration of Helsinki [15]. This understanding of respect for autonomy has been operationalized as *informed consent* [23], which requires that participants in medical research are informed in a sufficiently comprehensive and understandable manner (such as with a notice or information sheet) and that they are not manipulated [54]. Ethics also interacts with the legal dimensions of consent. With a transition to health data sharing using digital technologies (hereafter referred to as "e-health"), a rise in the accessibility of genetic testing, and the advent of AI technologies moving health data away from an exclusively medical context, ethics' interplay with laws has generated more complexities.

In the EU, the GDPR sets rather strict conditions for obtaining consent (hereafter referred to as "legal consent") [336]. In addition, health data are considered special under the GDPR because of their sensitivity, requiring more stringent protections and conditions for processing, including explicit consent (Art. 9 GDPR). The GDPR aims to protect the identified or identifiable natural persons whose data is processed ("data subject") by regulating the processing of personal data and reconciling individual control over one's data with other rights and interests at stake. However, the rules are complex and full of gray areas, making it difficult to comply with for researchers and companies and exercising rights requires much effort on the part of the data subject who must read, understand, and decide on the processing of their personal data.

Additionally, the newly enacted Data Governance Act (DGA) aims to improve data sharing in the EU by creating a harmonized framework for data exchanges and data governance [337]. The DGA introduces new concepts such as data intermediaries and data cooperatives and regulates the voluntary sharing of data for "altruistic purposes" and the provision of services assisting individuals in giving and withdrawing consent. Thus, the new regulation increases the uncertainty regarding how altruism interferes with the GDPR's legal grounds for processing personal data or how data intermediaries or cooperatives will enable individuals to express their privacy choices [307].

As such legal-ethical complexities build in e-health data consent, consent management solutions need to adapt to address the interests of data subjects and data controllers. In particular, such tools should address these interests in a human-centered and responsible manner to maximize human autonomy, promote lawfulness, and increase the transparency of data sharing and associated rights.

In this chapter, I repeat some of the key arguments regarding consent from the Introduction and extend into specific legal-ethical uncertainties with input from my co-authors. We discuss two legal-ethical issues: 1) What are the legal-ethical challenges regarding the scope and requirements for e-health data consent? 2) What are the legal-ethical reasons for obtaining consent beyond its role as a legal basis? Then, we describe available technological solutions and our work in the sphere. Our position is that the implementation of more interoperable, value-centered, and dynamic tools can assist humans in obtaining appropriately ethical and rigorous e-health data consent by helping them navigate legal-ethical uncertainties and challenges.

2.2 (Un)defined Requirements for Consent

2.2.1 Upholding Autonomy as an Ethical Principle for Digital Health Data

The rise of e-health data sharing has further muddied ethical debates regarding *how to*, *when to*, and *from whom* to obtain consent in order to uphold autonomy. While paper-based consent was debated in terms of understandability and transparency, data-collecting digital medicine devices add unique challenges. They contain often long and jargon-filled user agreements and introduce a layer of consent between company and patient [164]. For example, some smartphone mobile health ("mHealth") apps allow health data traditionally reserved for the doctor and patient to be accessible for other commercial purposes, such as third-party marketing [191]. Apps may also lack privacy policies and terms of agreements altogether, and if present, are difficult to read and comprehend [271]. Unlike more rigid consent practices in medicine, this data is given with a click of a privacy permission request on a smartphone, often with little understanding of what is being given [159]. From an ethical perspective, this raises concerns about how informed an individual's consent is and the skepticism that it upholds the principle of autonomy.

2.2.2 Who is the Genetic Data Subject?

Health data is sensitive due to the familial and interconnected nature of the data. Here, we focus on a highly connected, identifiable, and digitalized subset of health data, genomic data. With millions sequencing their DNA due to increased accessibility [195], distributed privacy risks have become even greater [36, 89]. Although there are varying guidelines across countries for notifying family members that their shared genetic data is being processed [320], from an ethical perspective, Minari et al. [214] argue for a form of family-group consent for genetic data processing due to shared risks.

Legally, a genetic group as a data subject is considered in guidelines but any current or future enforcement is unclear. The EDPB guidelines on genetic data state that data subjects can be families [244], however this has not been used in any way the author could find. Legal debates over individual and collective enforcement under the GDPR have developed [171] due to the possibility of conflicting rights, such as the right to object to processing from one individual [172], the *right not to know* [90, 109], or the right to process their data. There is little existing guidance on how to resolve such conflicts and is an area of active debate [166, 172]. It has been argued that managing conflicts is feasible [26] with the GDPR as a starting point. In addition, different countries have various approaches to weighing the rights of all parties based on the context and existing rulings (e.g., the right to privacy of the deceased is overruled by the right to health of the living [255]) and laws. This may also help data minimization by limiting data sharing and access to only well-justified cases [26]. While ethical and legal guidelines may allow for a collective interpretation, this challenges the status quo of individual consent and further complicates any possible GDPR enforcement.

2.2.3 Specificity in Consent: Purposes and Controllers

While consent must be "*specific, informed, and freely given*" (Art. (4)(11) GDPR), guidelines around purpose specificity from a legal-ethical perspective are unclear or contradictory. First, specificity is arguably at odds with broad consent models used in biomedical research. Broad consent refers to consent for specific and general future

purposes, while specific consent refers to consent for an explicit purpose. From an ethical perspective, broad consent could be acceptable if the individual is provided sufficient knowledge to be informed [54] –although whether current e-health consent meets these criteria and upholds autonomy is debatable. Second, though data protection law mandates specific consent, issues regarding interpretation remain. For example, too narrow an interpretation of specificity may lead to frequent re-consent from data subjects and induce consent fatigue [56]. Also, Recital 33 GDPR acknowledges that it is often impossible to identify all purposes of personal data processing for scientific research at the time of data collection and offers a solution of consent for "*certain areas of scientific research when in keeping with recognized ethical standards*". However, this is not reflected in the text of the GDPR itself, which advocates for specificity, and is open to different interpretations of scope and application [110, 125, 316].

For consent to be valid, the data subject should also be given the identity of the entity that decides the means and purpose of the processing ("data controller") (Recital 42 GDPR). However, this can be difficult to identify at the time of initial collection and while transparency regarding the type of controller (e.g., academic, pharmaceutical, etc.) impacts the data subject's decision to share their data [212], the exact identity of the controller is not always required (Art. 13(1)(a)(e) GDPR). The text of the GDPR is not clear on the elements that must be provided across different sections. While Article 13(1)(a) GDPR and Recital 42 GDPR require the identity of the data controller to be disclosed, Article 13(1)(e) suggests that the entities that process personal data can be clustered based on relevant criteria by referring to information about recipients or "categories of recipients". The notion of "recipient" (Art. 4(9) GDPR) can include third parties, but also controllers and processors, rendering the contents for the obligation to inform uncertainly. The GDPR states that the data subject should be informed of the recipients or categories of recipients of their data (Art. 13), but does not clarify how this affects the obligation to disclose the identity of data controllers. The new DGA [337] further complicates the recipient's identity. In complex data-sharing environments, it is unclear whether re-consent should be asked when additional persons become involved in the data processing. Under the GDPR, if the legal basis was legitimate interest, a contract, or vital interests the data can be used for another purpose if it is deemed compatible with the original purpose. If the legal basis was consent and the purpose changes, they should request new consent. In this case, some may fall under Art. 11 GDPR, where personal data is not required (e.g., if they receive pseudonymized data) and cannot re-establish identity to request re-consent. Such persons include those who process personal data on behalf of the data controller ("data processors"), additional data controllers, or new third parties that become involved after initial consent. These parties may fall under Art 14(4) GDPR, where appropriate measures to protect the data subjects should be in place, such as a public portal to check the status of research efforts and offer means to exercise rights, such as the right to opt-out.

In summary, the debates about how to uphold autonomy as an ethical principle, if a genetic data subject under the GDPR can be collective, and how specific consent is regarding purposes and the data controllers lack clear guidelines for e-health consent.

2.3 Consent is Relevant, Even When Not the Legal Basis

Even when consent is not the legal basis for processing data (Art. 6(1)(c)-(f) GDPR), such as legitimate interest, or the data falls under an exception for processing health data (Art. 9(2)(i)(j) GDPR), such as public health, ethical consent is relevant due to the possible legal consequences as an ethical standard or safeguard. Such data processing is subject to a balancing exercise based on the proportionality of interests and rights of the data subject and processor, which often requires the implementation of safeguards to help protect rights. Recital 33 of the GDPR refers to "*recognized ethical standards*" but lacks details or references. Perhaps this is due to how, depending on the project, varying subsets of EU legislation, such as the Clinical Trials Regulation [338] or EU Charter of Fundamental Rights [332], or international standards like the Declaration of Helsinki [15], may apply. Art. 9(2)(j) GDPR provides that health data can be processed for scientific research purposes based on Union or Member State law provided that appropriate safeguards are in place. Next, Article 89 of the GDPR also requires safeguards when data is processed for research purposes, but again the text lacks any definitions. In another instrument in the EU [338] or outside the EU [16, 90, 15], consent is a condition for participation in biomedical research. Commenting on the safeguards in the GDPR, Staunton et. al. [311] argues that ethical requirements such as consent and transparency could serve as safeguards to help inform the data subject of their rights.

The distinction between consent for research and consent for processing personal data must also be clearly communicated, and possibly combined through consent management platforms (CMPs). The EDPB [33] differentiates between the functions wherein *consent for participation in research* protects human dignity and the right to integrity of individuals while *consent for processing of personal data* is a requirement connected to the right to protection of personal data. The European Data Protection Supervisor (EDPS) also notes this separation and argues that informed consent can function as a safeguard for data subject rights' in medical research [316], and which comes from a long history of protection of a patient's bodily and psychological integrity and dignity in a doctor-patient relationship founded in ethics [90, 15] and the contractual relationship between the parties. The Commission DG Research & Innovation Guidance suggests that consent for lawfully processing personal data and informed consent for research could be integrated with CMPs to increase transparency to the data subject when it is difficult to identify the purpose of the research [91]. While CMPs can more transparently communicate processes and rights to data subjects, they can also confuse data subjects. If individuals provide their informed consent to participation in biomedical research, it might come as a surprise that the ground for processing personal data is not consent. If the distinction is not made clear, it might give the false impression that the data subject is in control. For example, data subjects do not have the right to withdraw consent (Art. 7(3) GDPR) or the right to data portability (Art. 20 GDPR) when consent is not the legal basis for processing data. Therefore, consent as a safeguard should be clearly explained to data subjects and differentiated from consent as a legal basis. Overall, it remains unclear whether the proposed CMPs would comply with the current requirements of legal consent or act as a legal safeguard, and if this could extend to cases for general health data sharing and not only for biomedical research.

From an ethical standpoint, one can argue that ethical consent in e-health should always be part of health data collection to uphold autonomy. The Belmont Report argues for respect for autonomy as a critical component of upholding human dignity through Principlism [24, 254]. However, autonomy and ethical consent have been

critiqued by bioethicist Onora O'Neill [232] who argued that this interpretation diminishes trust, wherein doctors value legal compliance more than true empathetic communication. It has also been critiqued for undervaluing collective concerns in favor of individual considerations. For example, one individual's consent to sharing genetic data may implicate genetic relatives without their knowledge through data breaches [228], yet despite these shared risks, only one individual consented. Despite these concerns about the Belmont Report's operationalization of autonomy into individual informed consent, few would argue that respect for autonomy is not worth upholding — but it may require a different ethical justification. Respect for autonomy can also be rooted in Flourishing Ethics (FE) [45, 46], which proposes that the pursuit and promotion of human flourishing is the ultimate ethical "good." This stems from Aristotle, who believed that all creatures have a "nature" that, through the pursuit of perfection, leads to flourishing. FE brings together a group of related understandings in computer and information ethics that have this idea of human flourishing as their primary ethical concern [99, 154, 218, 360]. In FE, "autonomy" is viewed as a requirement for human beings to flourish; we must be able to craft our lives one choice at a time in order to reach our full potential [45]. In psychology, theories exploring psychological well-being such as self-determination theory (SDT) translate philosophical understandings of human flourishing and autonomy [278]. SDT postulates that designing autonomous digital interactions requires interfaces or assistive technologies to promote both a user's sense of agency and consistency with a user's values, goals, and sense of purpose [249, 278]. In the case of health data, notice and consent (though flawed) allow an exercise of autonomy on the flow of their data to shape an increasingly important aspect of modern life: one's digital footprint. Forgoing this control, or coercing it, could undermine human flourishing [218].

2.4 Technological Solutions to Ethical-Legal Challenges

In this section, we identify and discuss technological solutions that we believe can help tackle the above ethical-legal challenges in e-health consent. We provide a critical analysis of existing solutions and our work towards more dynamic, transparent, and value-centered consent. These solutions center collaborations between technology and humans (data controllers, processors, subjects) to promote agency and value-centered choices.

2.4.1 Technical Standards, Ontologies, and Mechanisms for Consent

To address the lack of guidelines for specific consent and purpose specification in Sec. 2.2.3, we look towards technical standards from the ISO and ontologies built by expert communities. To address the role of consent even when it is not the legal basis from Sec. 2.3, consent mechanisms with options to object to data processing by legitimate interest will be analyzed. Overall, the development and adoption of these technologies can create a more interoperable ecosystem of consent.

First, ISO/IEC 29184 describes the structure and content of online consent and privacy notices to collect and process personally identifiable individual data. It outlines how to communicate transparent and understandable information about the data collection and processing, as well as how to obtain consent to be "*fair, demonstrable, transparent, unambiguous and revocable*" [1]. It has also been shown to enable compliance with the GDPR [239] and could be compatible with the DGA, which

requires the development of an altruistic consent form at the EU level available in an electronic, machine-readable form using a modular, customizable approach for specific sectors and purposes (Art. 25 DGA). However, as a closed standard with licensing fees, adoption by individuals or institutions with fewer resources may be difficult.

Second, the ISO standard suggests using consent policies based on standardized semantic vocabularies, such as the W3C's DPV [240], which can also aid in the specificity of consent. The semantic web is an effort from W3C to make the internet machine-readable and enable a web of linked data with vocabularies, query languages, and more. The DPV is an ontology about the use and processing of personal data with terms including processing purposes. Ontologies can be extended for different use cases (e.g., a GDPR compliant extension of the DPV [277]) or mapped to other compliant ontologies [78, 237]. They could also create "parts of research projects" to offer categories instead of single choices (Recital 33 GDPR). As an open-source technology, organizations can contribute and help address their use cases or map the logic of DPV to other ontologies. For example, terms in the DPV can be mapped with concepts in the Data Use Ontology [238]. Created by the Global Alliance for Genomics and Health, the Data Use Ontology addresses data sharing after consent and increases the FAIRness (ability to be findable, accessible, interoperable, and reusable) [175]. Other health ontologies [83, 152, 343] could be mapped and connected to standardize consent, data sharing, e-health records, and other health processes. While more work is required to make ontologies such as DPV applicable to more health data-sharing situations, we can envision an interoperable future for privacy, consent, data sharing, and legal compliance based on extensive open vocabularies. This can also help automate the activity of data intermediaries or cooperatives regulated under the DGA.

Third, these standards and ontologies can be communicated through the web using Data Protection and Consenting Communication Mechanisms (DPCCMs), containing the "*communication of data, metadata, information, preferences, or/and decisions related to data protection or/and consenting between different actors*" that can be used on the web or apps [139]. Examples include Do Not Track or Advanced Data Protection Control (ADPC). ADPC is more complex than a binary track or do not track, and can express the specific purpose along with the consent decision and object to processing based on legitimate interest. These technologies could also have a role in compliance with data protection law as it offers a way to object to processing when consent is not the legal ground for processing. Furthermore, ADPC could incorporate more complex values such as consent preferences, thereby enabling personalization across platforms using ADPC. However, some challenges still remain. There is no standardized process for developing the vocabularies regarding the values in ADPC, and adoption of such DPCCMs remains challenging, as with the obsolescence of Platform for Privacy Preferences Project [293]. DPCCMs could contribute to a more central ecosystem of consent to facilitate purpose specification, processing entities, and privacy profiles. While ADPC could enable increased autonomy through the ability to object to processing when the legal basis is a legitimate interest, the rights could still be obscured if the consent mechanism is not widely adopted. Similarly, while guidelines for specific consent are part of ISO standards and ontologies can increase specificity, a more unified and interoperable consent ecosystem requires social factors to gain traction outside the scope of this PhD thesis.

2.4.2 Ethical and User-Friendly Privacy Assistants

We can also consider technological assistants to better promote autonomy, value-centered choices, and human flourishing in smartphone mHealth settings (Secs. 2.2.1 and 2.3). When apps are collecting health data, Personalized Privacy Assistants (PPAs) could help empower humans to navigate consent permissions and promote more autonomous action. Smartphone PPAs use a decision tree to ask the user a series of privacy preference questions and determine their privacy preference profile. From this profile, PPAs provide the user with privacy setting notifications and privacy setting recommendations for the apps on their phone [185]. For users who use mHealth apps, such recommendations could remind them of their privacy preferences when they may have "clicked-through" permission settings when downloading the app. Infrastructure for PPAs for the Internet of Things has also been proposed, and such a system could help users manage data collected by complex, multi-system health sensors or medical devices by giving them similar notifications and recommendations [70].

A related system under development is the Value-centered Privacy Assistant [48], which aims to promote value-centered choices at the root of human wellbeing and flourishing [249, 278]. Profiles are based on how a user's personal values are involved in their app selection and privacy decision-making by mapping values onto acceptable data collection practices [49]. Notifications occur before downloading an app from the app store. These notifications serve as "selective friction" to warn users when they may be downloading an app that conflicts with their value set as determined by their profile. It also recommends alternative applications with similar functions that are more consistent with a user's values.

Both systems encourage users to exercise their autonomy when engaging with mHealth apps by assisting them with data privacy decisions that they may otherwise quickly click through or struggle to comprehend the privacy policies and terms of agreement [159, 271]. This more judicious use of collaborative technologies, by promoting autonomy, furthers human well-being, and flourishing in the e-health data space [278].

2.4.3 Layered User-centered Dynamic Consent

Last, DC can incorporate the above technologies, enable autonomy, and increase the specificity of consent regarding the genetic data subject (individual and collective), data processing purposes, and processing entities in Sec. 2.2. DC is a model of consent and digital platform centering data subjects in facilitating consent over time [123, 158, 199]. DC can request specific consent over time as data processing or the controller changes and be layered in terms of the type of consent (specific or broad) or information, allowing users to choose the depth of information and type of consent they prefer. On such systems, perhaps specific consent could be the default, with a layered approach that first shows key information and then offers more detailed information with broad consent as a secondary option. It could be personalized based on the data controller and data subject's legal jurisdiction, privacy preferences, and values—with value-centered privacy assistants to help make decisions. Similarly, collective-specific consent could be a default for genetic data sharing unless the individual chooses broad consent, and in the case of conflicting rights specific and granular rights can be carried out and relayed to the data collector to resolve issues more transparently.

As a platform, DC can also incorporate ISO standards, consent ontologies, and shared consent mechanisms. If multiple DC platforms use interoperable ontologies and/or consent mechanisms, a more unified consent management system could be envisioned. However, work from the technological side is needed to suit more use cases and larger societal challenges stand in the way of adopting shared technologies. Collective DC has been proposed [214, 253] but more research must be done in order to implement it well. This ties into future chapters in this thesis regarding tools and methods to increase transparency and user-centered design in consent and user studies regarding key elements of consent [82].

2.5 Conclusion

Technological solutions provide frameworks for managing the challenges of ethical-legal consent for e-health data, especially when human needs are centered and not legal-ethical compliance. However, the legal uncertainties must be decided by legal experts, courts, and cases. This process often takes years, especially for newer concepts of technologies (e.g., collective consent wherein legal experts are beginning to debate). The goal of suggesting tools and frameworks such as consent standards, ontologies, privacy assistants, and layered (collective and/or individual) dynamic consent is to enhance autonomy in the meantime. We focus more on ethics and the principle of autonomy, which is more flexible and general than legal challenges, especially as these could apply within the EU as well as outside. While these may also address legal challenges, we as scholars cannot determine compliance or mandate requirements. Some technologies still require further development to truly address current challenges, and the Author is researching legal, ethical, and technical aspects in their future work. From this, the wider adoption of these solutions could not only tackle ongoing legal-ethical ambiguity within the EU but also lay the foundation for cross-border health data transfers between different countries. Despite differing guidelines and requirements, a united technological front and deployment of human-centered tools for e-health management could help provide the basis for greater communication, understanding, and harmonization between jurisdictions.

Chapter 3

Transparency and User-relevancy of Consent Policies

This chapter is accepted and in press as: Doan, Xengie, Fatma Sümeyra Doğan, and Arianna Rossi. (In press).“Analysis of Transparency and User-relevancy of DTC Company Policies.” Privacy Symposium 2024.

Author Contributions: Conceptualization, Investigation, Formal analysis, Methodology, Software, Validation, Visualization, Investigation, Writing - original draft, Writing, review & editing

Abstract: Privacy policies often fail to uphold the goals of transparency – for individuals to understand the processing of their data and exercise their rights in a user-centered manner – which may lead to misalignment between privacy expectations and practices. Direct-to-consumer (DTC) genetic companies, expected to grow to more than 2.7 billion USD by 2032 in Europe, process sensitive data with many risks. We selected six leading DTC genetic companies and examined their EU privacy and research consent policies to answer: 1) How vague, confusing, or complete are information flows?; 2) Are they aligned with GDPR transparency requirements?; 3) How relevant is the information to users?; 4) What risk/benefit information is available? This study identified 62 flows for sharing genetic data and found that 81% were vague and 37% were contextually distinct and confusing. Consequently, GDPR transparency requirements may not be met. Qualitatively, information was not user-relevant and lacked collective risks of sharing data. We then offer specific suggestions to enhance user-centered transparency in policies and to use contextual integrity as a tool to assess, audit, and share data practices.

3.1 Introduction

Privacy policies have a long reputation of being time-consuming [205], unreadable [93], difficult to comprehend for non-legal experts [268, 140], and unusable as a decision-making tool [144]. While this may be a deliberate strategy to be legally comprehensive and accurate, especially in unclear situations for future-proofing [27], as one of the most public and comprehensive views into data processing practices, how can privacy policies be improved?

We take a multi-pronged approach to assess privacy and consent policies and suggest improvements. To assess information flows, we used Nissenbaum's privacy theory of CI which states that privacy can be defined by contextual norms dictating their transmission and how appropriate the information flows are [226]. As the context changes, so do the privacy risks. This theory also offers a framework to analyze specific parameters of an information flow (sender, recipient, transmission principle, attribute, and subject) and published methodologies to analyze transparency [300]. Second, we look to user-centered terms developed by Johansson et al. [146] in an extensive governance study to understand what users want to know in order to make informed decisions for research data sharing. We adapt the results to the context of companies. Last, we survey the risks and benefits of data sharing. This is especially important because the types of DTC companies we will investigate are genetic testing companies handling sensitive data.

Sequencing Deoxyribonucleic acid (DNA) has become more and more accessible with DTC genetic testing companies offering ancestry, wellness, and community services. By 2021, approximately 26 million people have taken an at-home ancestry test worldwide, and the European market is expected to grow to more than 2.7 billion United States dollar (USD) by 2032 [270]. Although genetic data are a special category of personal data under the EU's GDPR and should be subject to a high standard of legal and technical protection [336], in practice companies may not be transparent about their data processing activities as required in Art. 12-13 GDPR, or the risks involved. Once this information is in the hands of companies, risky events such as data breaches [228], re-identification based on combining public datasets [89], data sharing with law enforcement [279], or discrimination by insurance companies [147] may increase while consumers assume companies are ethical [17].

We selected and examined 6 market-leading DTC genetic companies' privacy and research consent (hereafter also referred to as "consent") policies for information transparency and user-relevancy. We were interested in the most publicly available and complete descriptions of data flows and practices, which are often the privacy policies and research consent policies for sharing data beyond the scope of the contract for additional research (internally and/or with third parties such as academics, companies, and/or individuals) secondary research. We identified genetic information flows from privacy and consent policies to answer: 1) How vague, confusing, or complete are information flows?; 2) How aligned with GDPR transparency requirements are existing information flows?; 3) How relevant is the information to users?; 4) What risk and benefit information is given and where is it located?

Our results show that 1) more than half of the identified 59 information flows used vague terms, 17% were missing information such as recipient, and 37% were bloated with up to 14 different reasons for sharing the data in one flow 2) from the previous analysis, information flows were not always aligned with GDPR transparency requirements; 3) most companies lacked user-relevant contextual information as described in [146]; 4) the communication about risks and benefits varies greatly across companies (1 to 6 types of risks shared) and lacks any collective risks. This confirms a pattern of non-transparent and non-compliant public information in policies in the context of sensitive genetic data. Then we suggest possible strategies for more transparent, user-relevant policies and using a CI analysis as part of audits by data controllers and regulators.

3.2 Related Work

3.2.1 Analyzing privacy policies using contextual integrity

While other tools and frameworks exist to assess the information in privacy policies, we chose to use CI [299]. In it, privacy is defined by Nissenbaum as the contextual norms dictating appropriate data transmission [226]. Sharing genetic data with your doctor and with your insurance agent has different appropriateness and privacy contexts, even if they are both third parties. Thus, CI can offer nuanced insights into the sociotechnical nature of privacy and data sharing.

The theory includes a framework for analyzing information flows to audit the data processing activities. It can determine, in a structured way, how personal data is shared and for what purposes to help audit the risks (thereby implementing a necessary step for data protection by design [32]). It can also be used to verify the correct implementation of transparency obligations: as these are meant to enable individuals to shape reliable expectations about the use of their data, the information must be presented in a useful and understandable manner. This is why Article 12 of the GDPR sets user-centered transparency requirements that encompass the "quality, accessibility and comprehensibility of the information" [243] of the data processing practices and the data subjects' rights. This means that communications, be it privacy policies, consent forms, or other instruments for exercising data rights, should be tailored to the specific informational needs and abilities of the intended audience, as well as subject to empirical tests to demonstrate their effectiveness [243, 274]. There is a movement within legal communication away from lengthy documents full of legal jargon and instead towards using information design elements [247] to address the needs and abilities of the intended audiences (who are not only legal experts) [252] and enhance the readability and comprehensibility of information [273].

In the case of DTC genetic testing companies, the contextual norms are obscured and often conflated, leading to privacy issues. Much of previous work analyzing the privacy policies of DTC genetic testing companies has a United States (US) focus, and researchers have found that their policies lack transparency and adequate protections against harms [104, 131, 60], especially in relation to data sharing with law enforcement [263, 305]. This was shown in a 2015 study [137], where the declared data practices of DTC genetic testing companies were assessed via the CI approach. Huang et al. analyzed the privacy policies to assess the different contexts and privacy issues. They identified three different contexts for data sharing (i.e., users involved in the online genetic company community, consumers utilizing genetic tests for reasons of ancestry, and consumers who have taken genetic tests who additionally participate in research activities) and related privacy issues. However, the previous study did not investigate specific elements of transparency or user-relevancy to improve the policies for users (e.g., to clarify contexts in information flows) that we are interested in.

3.2.2 Privacy risks and harms of sharing genetic data

Protecting genetic data requires increased caution because the potential harms to the data subject, and even relatives is great. For example, data breaches could affect living and future relatives [192]. Moreover, discrimination and stigmatization [124, 106] may occur. from the availability of genetic data. For example, the inference of Alzheimer's could be used by insurance companies in order to charge higher premiums [358, 148]. Other risks include possible re-identification from anonymous data, which was proved possible through the cross-referencing of two free databases in 2013 [120]. With the rise of more genetic databases and digital health information (both research and commercial), it can be easier to cross-reference and use the aggregate information to infer individual or family details, such as disease risks [89]. In the age of big data, privacy is networked – especially genetic data which may reveal information about similar genetic relatives [165, 40, 102]. Companies may not adequately inform customers of how their individual decisions may affect their families and descendants in the status quo.

3.2.3 GDPR transparency requirements and implementations

The disclosure of information in privacy notices should be leveraged to not only fulfill legal obligations but also to communicate useful, actionable information to the individuals and groups impacted by the data practices. Recital 39 of the GDPR clarifies this point, "*It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed ... any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used*" [243]. This aligns with the concept of information transparency, which is defined by how complete the information provided by firms is regarding their business activities [342]. While privacy policies have many issues [93, 268, 140, 144], they are a window into internal data practices of a company publicly available to regulators, researchers, and consumers alike.

3.2.4 User-centered decision making and privacy expectations

Johansson et al. [146] carried out discrete choice experiments with laypeople and experts to identify types of information relevant for research data sharing. Their results showed that high level categories of data user (e.g., pharmaceutical company/entity, academic research, technical company, or national authority), were most useful to people. Some elements go beyond legal requirements for transparency under the GDPR, or offer guidance for how to frame the information in such guidelines (e.g., by category of data processor and not the exact identity). By centering user research to determine elements of transparency, organizations can follow the legal principles underlying transparency with the guidance of empirical work.

3.3 Research Questions

Stemming from gaps in understanding the DTC genetic website's informational transparency, compliance with GDPR transparency requirements, and the quality of user-relevant information to enhance decision making, we collected privacy and consent policies on company websites regarding genetic data sharing to answer: **RQ1**: How transparent are information flows described in the available policies in terms of: including all parameters, using vague language, and parameter bloating (i.e., more than 2 concepts in one parameter)? **RQ2**: How aligned are the information flows with GDPR transparency requirements? **RQ3**: How aligned with existing user-relevant information categories for research [146] (i.e., data user, data collector, reason for data user, information and consent, and ethical review)? **RQ4**: How many and what types of risks and benefits are stated across privacy and consent policies?

3.4 Methods

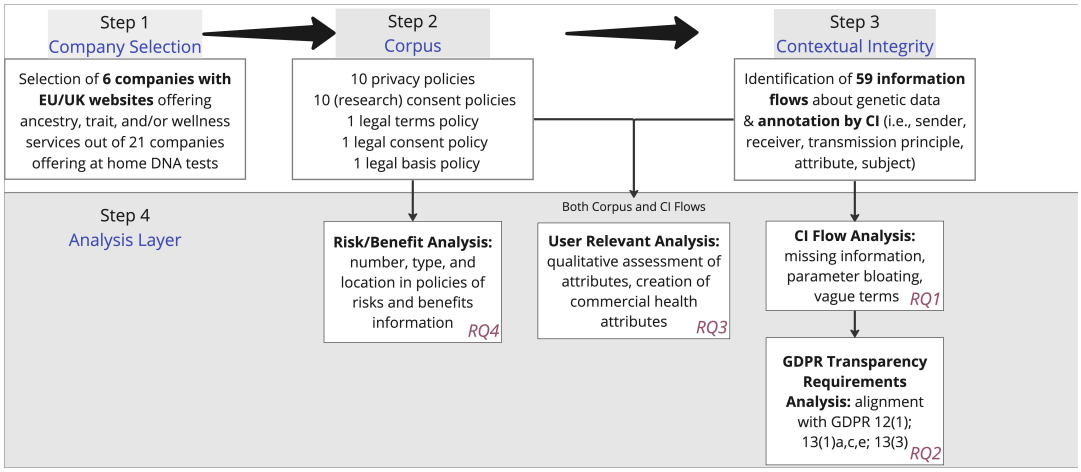

FIGURE 3.1: A process map of the steps to identifying genetic data flows and analyses used from 6 company websites and their corpus.

3.4.1 Company criteria and the corpus of text

We chose six of the most widely used DTC genetic testing companies with similar services of ancestry, trait, and/or wellness reports (excluding medical genetic

disease, paternity, microbiome, and full genome tests) with EU/UK websites (excluding business-to-business services or services requiring a medical professional). These companies include global market leaders and EU specific providers to try to gain a representative sample of companies. AncestryDNA reported over 25 million global customers [12], while 23andMe reported over 14 million [2] on their websites. MyHeritage and Family Tree DNA followed with more than 2 million customers each in 2018 [267]. The number of of TellMeGen customers were unavailable, but they have been in business since 2014 and were created in Spain and may have an EU specific strategy. Other company headquarters are across the USA (23andMe, AncestryDNA, and Family Tree DNA), UK (LivingDNA), and Israel (MyHeritage).

We accessed the EU or UK English websites' privacy policies and research consent policies from each company from December 2022 to January 2023. One exception was TellMeGen, whose policies were named differently. Thus, we analyzed their pages titled: "legal terms," which contains conditions for usage of the website and services offered, and "legal consent," which is given before entering the contract. Information on external pages was not included as they were linked outside the policies and would expand the scope of this research too greatly, and most pages would refer to the privacy policy for more information. In the end, our corpus included 10 privacy policies, 9 research consent, 1 group project (research carried out by non-academics) consent, 1 legal terms, 1 legal consent document (Step 2 in Fig.3.1).

Parameter	**Description**
Sender	Entity who shares or transfers data
Recipient	Entity who receives information
Transmission Principle	Terms and conditions wherein transfers should occur, including descriptions of how information is collected/used
Attribute	Information type
Subject	Subject of information flow (usually implied)

TABLE 3.1: CI framework from Shvartzshnaider et al. [300]

3.4.2 Contextual integrity

CI is a theory that posits that privacy is the appropriateness of information as defined by how they align with existing, legitimized norms for a given social contexts [226]. For example, sharing genealogy information with a doctor may be appropriate, while sharing the same information with law enforcement could be a violation of privacy. This is an example of how different recipients affect privacy. Using the CI framework, information flows are broken down into sender, recipient, transmission principle, attribute, and subject [300] (Table 3.1. Without specifying all five parameters, the context is too ambiguous to assess the implications. We followed methods from Shvartzshnaider et al. [300] to perform deductive qualitative coding (Step 3 Fig. 3.1) and assess the quality of CI flows with missing information, parameter bloating, and vague terms analyses (Step 4 in Fig. 3.1).

To code the corpus, we began by identifying information flows regarding genetic data and co-code the respective parameters (see Fig. 3.2).

We *[sender]* work with **other companies** *[recipient]* **when providing and marketing the Services** *[transmission principle]*. As a result, these companies will have access to or otherwise process **your** *[subject]* **data, including some of your Personal Information** *[attribute]*, in their systems. These companies are subject to contractual obligations governing privacy, data security, and confidentiality consistent with applicable laws. These companies and the Personal Information they may have access to include our:
Laboratory partners (such as your **DNA**); **DNA test shipping providers** (such as **name, shipping address, and phone number**); **Payment processors** (such as **Payment Information**); **Cloud services infrastructure providers** (Ancestry's web and mobile services are cloud-based services; **all your data resides with our cloud service vendors**); **Biological sample storage facilities** (such as **Biological Sample and DNA test kit code**);**Vendors that assist us in marketing and consumer research analytics, fraud prevention, and security** (such as **email address**); **Communications infrastructure providers** (such as **name and email address**); and, **Vendors that help us provide some Member Services functions**, like phone support or survey tools (such as **Account Information or name or email address**)

FIGURE 3.2: Example of annotated CI flow from AncestryDNA.

3.4.2.1 CI flow and parameter analysis

To answer RQ1 about the transparency of CI flows, Author 1 used the following methods from Shvartzshnaider et al. [300]. While the validity of information and parameter bloating may be subjective, in confusing cases Author 1 consulted Author 2 for higher reliability. This was analyzed using materials available in the appendix (App. 3.9), R, and MaxQDA project `https://www.maxqda.com/`.

Missing information identifies if parameters are missing in a CI flow to gauge informativeness. Even one missing parameter can be confusing to an individual because one premise of CI is that the flow must be complete to understand the full context (e.g., the recipient of the data may greatly affect the appropriateness of the context but it is missing).

Parameter bloating identifies any flows with two or more discrete entries per parameter which can cause confusion from the lack of directness (Fig. 3.3). For example, if one information flow contains multiple transmission principles that span marketing, research, and providing ancestry services to the customer, then the customer has to consider multiple contexts without clear understanding of what is actually taking place.

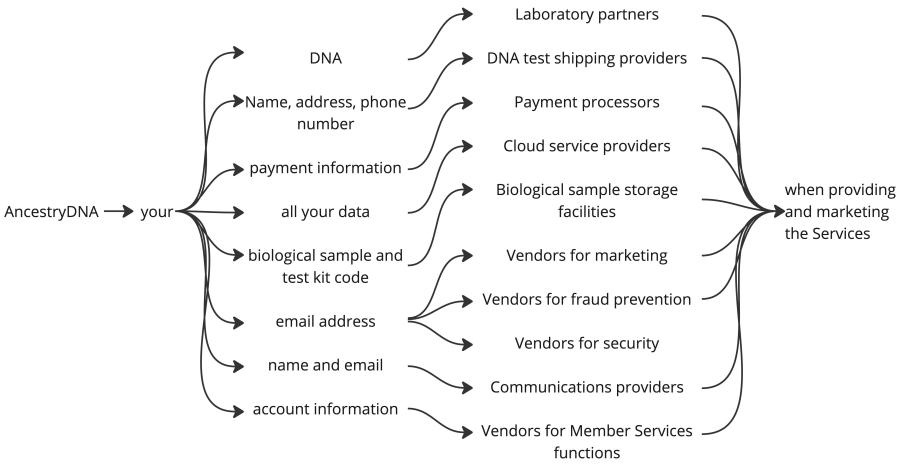

FIGURE 3.3: Example of a CI flow with parameter bloating from AncestryDNA text from Fig.3.2 with parameter bloating

Vague terms analysis identifies CI flows using vague terms that imply ambiguous and unspecific information. This is performed by matching the text of CI flows to a vague terms list from work by Bhatia et al., which identified common vague

terms and categorized them based on lexical categories: numeric quantifiers such as most or some, modal terms such as may, likely, and possibly, generalization terms such as mostly, normally, or primarily, and conditional terms such as sometimes, as necessary, and depending. [27]

3.4.3 GDPR transparency requirements mapping

CI Flow Element	**GDPR Transparency Requirement**	**User Relevant Attribute**
Sender	Data controller (Art. 13(1)a) and concise, transparent, clear language (Art. 12(1))	Health Information Collector
Recipient	Recipients of personal data (Art. 13(1)e) and concise, transparent, clear language (Art. 12(1))	Health Information Collector, Data User
Transmission Principle	Purpose of processing (Art. 13(1)c, secondary purposes (Art. 13(3)) and concise, transparent, clear language (Art. 12(1))	Reason for Data Use
Attribute	concise, transparent, clear language (Art. 12(1))	

TABLE 3.2: Mapping or relevant attributes from the CI flow, GDPR transparency requirements, and user-relevant information.

To address RQ2, Author 1 analyzed the GDPR transparency requirements to identify 6 legal requirements that could be extracted from CI flow parameters or analyses (See Table 3.2). Art. 12(1) states that information should be "concise, transparent, intelligible and easily accessible form, using clear and plain language," of which can be assessed by parameter bloating (conciseness), missing information (completeness), and vague terms analysis (clarity). Art. 13(1)a requires "the identity and the contact details of the controller," which is usually the sender of the information or the recipient if the data subject is sending their information to the company. Art. 13(1)e requires information on "the recipients or categories of recipients of the personal data," which could include third parties or other individuals in the context of online genetic communities. The purpose of processing (Art. 13(1)c), whether the personal data is a requirement to enter a contract (Art. 13(2)e), and if the controller intends to process the data for secondary purposes (Art. 13(3)) can be identified in the transmission principle of CI flows. While the presence or absence of relevant GDPR requirements could be systematically evaluated, we are not assessing compliance.

3.4.4 User relevant information

To answer RQ3 about user-relevant information, Author 1 looked to empirically derived results from a study by Johansson et al. [146] (Table 3.3) to analyze the quality of the information within policies and information flows. Mappings to CI elements are included in Table 3.2. CI flows and relevant portions of the policies based on

keywords (e.g., review, ethical, opt-out, opt-in, purpose, goal, etc.) were used. Coding began with existing attributes from research [146] with a bottom-up approach to address missing attributes related to the novel commercial context. Author 1 tried to be objective in adherence to Johansson et al.'s work and uncertainties were discussed with Author 2, though not co-coded. Some new codes in the "*reason for data use*" category include subcategories like: "*performing a contracted service*," "*unspecified research*," "*secondary DNA service*" to expand the codebook to address DTC genetic testing use-cases. Coding was performed by Author 1 in MaxQDA and the codebook is available in App. 3.9.

Attributes	**Levels**
Health Information Collector	Who collects the information? (technological company, academic research provider, etc.)
Data User	Who is the recipient of the data? (technological or pharmaceutical company, academic research project, etc.)
Reason for Data Use	Why the data user wants access? (develop a new product or service, advertise, etc.)
Information and Consent	Will the participant be informed? (Not informed, informed and opt-out, etc.)
Review of Data Sharing	Is there a review of data access and how is the decision made? (No review, review of transfer, etc.)

TABLE 3.3: Data sharing attributes from Johansson et al. [146]

3.4.5 Stated risks and benefits

To answer RQ3, Author 1 identified sections in the corpus referring to risks and benefits throughout the privacy and consent policies, which often were separate from CI flows, by using keywords such as "risk," "harm," "benefit,", and "compensation". Using MaxQDA, each section was coded for the type of risk or benefit using a bottom-up approach to map types of specific risks (e.g., exposing your family, lost sample) and benefits (e.g., financial compensation, altruistic benefit) to broader categories (e.g., re-identification, data breach, etc.) based on the types of harms they fell into. Then the number and variation of risks and benefits were analyzed through MaxQDA (codebook available in App. 3.9). To minimize subjectivity, categories of risks and benefits were coded with terms found in the corpus; for example, the term for third parties stems from TellMeGen's text "*undesired third parties (for example, health care provider service companies or insurance companies)*," and using common categories of risks from literature such as "data breach" to describe a sample being lost or stolen.

3.5 Results

3.5.1 Contextual integrity information flows

To answer RQ1, we identified 62 information flows across the companies' publicly available privacy or consent policies directly or indirectly referencing the transfer of genetic data (see Table 3.4. While about 30% of information flows are specific and uniquely about the sharing of genetic data, the others are unclear. About 35% mention genetic data in addition to other (less or equally sensitive) personal information, about 30% only write personal information in general, and 5% specify anonymized

DNA data or are missing the information. Due to the difficult nature of anonymizing genetic information, it was included in our analysis [365, 120]).

Company	Policy Type	# Flows	# Vague	# Missing information	# Bloated
23andMe					
	Privacy Policy	8	7	1 (transmission principle)	4
	Research Consent	2	2		
AncestryDNA					
	Privacy Policy	7	6		5
	Research Consent	1	1		
FamilyTreeDNA					
	Privacy Policy	11	10	2 (attribute, recipient)	5
	Group Project	3	3	1 (sender)	
LivingDNA					
	Privacy Policy	7	5	2 (transmission principle)	2
	Research Consent	1	1		
MyHeritage					
	Privacy Policy	12	10	4 (recipient x3, recipient/attribute/subject)	3
	Research Consent	1			
TellMeGen					
	Legal Terms	2	1		
	Legal Consent	7	4		1
Total		**62**	**50** (81%)	**10** (16%)	**20** (32%)

TABLE 3.4: The number of information flows with genetic data are reported for each company based on which policy it was found in with the total across all companies. The number of flows, the number containing vague terms, missing information, and parameter bloating are also shown per company policy with totals across companies at the bottom, along with the percentage of overall flows.

Missing information Using the CI analysis of missing information, 10 (16%) of flows had missing information, ranging from the transmission principle (n=3), recipient (n=4), attribute (n=1), and for one – recipient, attribute, and subject were all missing. Missing recipients were usually regarding data sharing for research or processing sensitive data for multiple purposes (including but not limited to research purposes). Second most common were CI flows with no transmission principles regarding third parties such as law enforcement or unspecified service providers. For example, "*We may share information with our professional advisers including lawyers, accountants and insurance advisers. We do not routinely share genetic information with our professional advisers, but it would be possible that this could happen, for example if court proceedings relating to genetic data were to be brought against us*" (LivingDNA). This one non-exhaustive example does not explain the conditions wherein sharing occurs, thus we annotated it as missing.

Missing information and GDPR requirements To address RQ2, we mapped the respective CI analyses to relevant GDPR requirements and identified misalignments. From the missing information analysis, 17% of information flows lack the information items mandated by the transparency obligation because they are incomplete (Art. 12(1)). In addition, 5 information flows did not identify the recipient of data

(Art. 13(1)e), 5 did not clearly state the transmission principle which relates to purpose of processing from Art. 13(1)c and the eventuality of further processing from Art. 13(3).

Parameter bloating Parameter bloating was found in 37% of information flows. Fig. 3.3 shows an example of parameter bloating from AncestryDNA. This statement has one sender, one subject, 8 attributes, 10 recipients, and one broad transmission principle. Overall, we found that statements across companies had one sender, up to 9 attributes, up to 8 recipients, up to 14 transmission principles, and one subject.

Parameter bloating and GDPR requirements More than two discrete parameters (e.g., more than two types of recipients, more than two transmission principles) can be indicative of a lack of conciseness (Art. 12(1)), showing that 37% of information flows should be reformulated to be more direct.

Vague terms 81% of flows across companies contained one or more vague terms. Of the vague terms used in Fig. 3.4 we can see that across all companies modal words (e.g., *may, can, possibly*) were used in 69% of all surveyed companies' information flows. This is followed by numeric quantifiers (e.g., *certain, some, and various*) at 33%, condition terms (e.g., *depending, as needed, appropriate*) at 16%. Lastly, only MyHeritage used any generalization terms such as *normally, usually, or primarily* (8%).

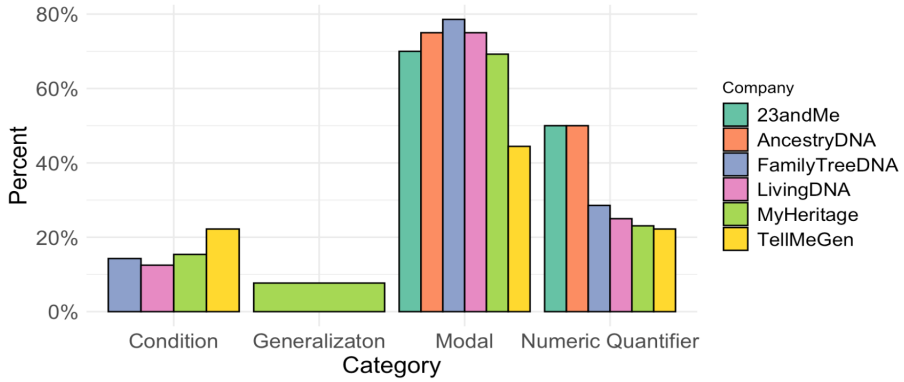

FIGURE 3.4: The percent of CI flows with vague terms by lexical category and company

Vague terms and GDPR requirements From this mapping of predetermined vague terms deriving from [27], 81% of information flows may not be using clear language (Art. 12(1)) due to the prevalence of vague terms.

3.5.2 User relevant information

To answer RQ3 regarding a qualitative analysis of the attributes reported in Table 3.3 within the corpus, we found that information flows and relevant parts of the corpus (e.g., "Purpose" sections) were generally confusing or lacking important information that would be relevant to a user to make decisions about their health data sharing in this commercial context.

Data collector The DTC genetic testing company most closely corresponds to a technological company data collector, as the customer enters a contract to use a service and sends their data for analysis. However, this was inferred, and they may also fall under an academic research project if research consent is given.

Data user Instead of identifying the recipient in flows by name or business category, the data user is a higher-level category that reveals the types of recipients relevant to the consumers (e.g., technological or pharmaceutical company). For instance, in Fig. 3.3, the information flow from AncestryDNA's privacy policy shows many technological companies of various natures. Some new categories derived from the privacy and consent policies were *healthcare professionals*, *law enforcement*, and *other users via genealogy services*. When looking at the informed consent for the further use of genetic data for research, data users such as a technological company, pharmaceutical company, national authority, or academic research project arise. In the informed consent document for research, some companies do not distinguish the various types of research being conducted, for example, "*On some research, we will collaborate with leading academics and scientists.*" (LivingDNA).

Reason for data use 34% of information flows had more than one reason for data use per flow. Overall, "providing a contracted service" was the most common reason for data use at 35.8%, followed by "developing a new product or service" (15.6%), and "unspecified research" (12.6%). This new category had to be created because documents failed to specify the purpose of research (e.g., investigating a government initiative, developing a new product or service, etc.). An example of such an instance is "*Perform statistical, scientific, and historical research*" (FamilyTreeDNA).

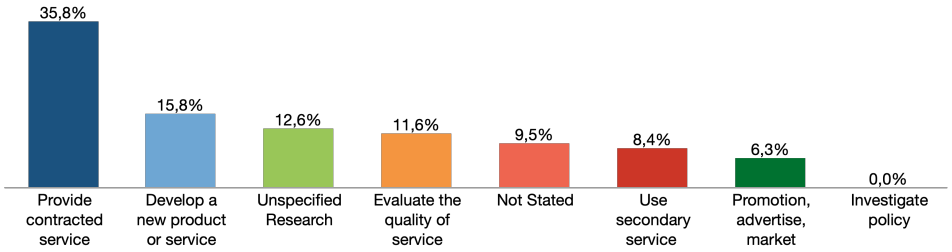

FIGURE 3.5: Percentage of coded segments with reasons for data use across companies

Information and consent Information (if individuals will be informed when information is shared and used in a new context) and consent (opt-in or opt-out) areuser-relevant attribute that are not very clear within flows. While privacy policies describe the purposes of data use and can be taken as information, they are often embedded in bloated transmission principles. For example, MyHeritage has an information flow about service providers that describes 8 different service providers, each with their own information and consent for each (some wherein information is given and some wherein both information and consent are given).

Information and consent also varied across companies for sample storage, with 23andMe, AncestryDNA and FamilyTreeDNA including information and opt-in consent within the privacy policy. AncestryDNA for instance writes, "*after our laboratory partner has processed your Biological Sample, you can consent to its storage in our bio bank for future testing at your option [...] If you do not consent to the storage of your Biological Sample, we will destroy your sample.*" While the biological sample can be used for other

purposes if the individual provides consents to it, the information also includes an opt-out consent for storage but it is unclear if storage itself entails any additional sharing and using the data in a new context (e.g., by the storage facility). Other companies do not allow an opt-out: TellMeGen destroys samples after two months, LivingDNA destroys samples after 10 years, and Family Tree DNA stores samples for future testing.

Review of data sharing Only 23andMe and AncestryDNA mention any review when the data is shared in a new context with regards to a type of ethical review in their informed consent policies. Of the two, 23andMe writes it is overseen "*by an independent ethics review board (also called an Institutional Review Board or "IRB").*" On the other hand, AncestryDNA will "*review all research requests for Biological and DNA Samples (as described below)*" but it never describes the type of review (e.g., using national laws, using corporate guidelines).

Some companies use technical or legal jargon that is too broad to determine if their protocols for review of data sharing are in place. For example, Living DNA mentions both"*The data [...] will only be used for ethically and scientifically approved research. Careful safeguards, in line with ISO:27001, are designed to ensure the confidentiality of your data and samples.*." They refer to an international technical standard for information security management, but they do not share relevant data about oversight.

3.5.3 Stated risks and benefits

Regarding RQ4, first we report the number and type of risk (App. Table 3.5). Across all companies, a total of 37 risks were identified, spanning 7 categories (identification, undesired third parties, inaccurate results, secondary findings, data breaches, discomfort, and general catch-all risks). **Identification** encompasses several subcategories that have to do with third parties revealing hidden information: re-identification of the customer (e.g., finding a full name), exposing the family (e.g., finding relatives' names), and reveal of phenotype (e.g., finding traits from the DNA such as disease status). **Undesired third parties** includes data sharing without consent to law enforcement and insurance agencies (which may lead to discrimination). **Inaccurate results** refers to the accuracy of methods of analysis used to determine ancestry, wellness, or other reports given to customers. **Secondary findings** refers to additional research that may be done on the data that leads to new findings (e.g., disease information from ancestry services). **Data breaches** covers both physical samples and any data stored on servers (e.g., genetic data, passwords, etc.). **Discomfort** refers to the feelings that being asked personal questions may bring up, and **general** covers any catch-all "unforeseen future risks". The most commonly stated risk is identification, followed by general risks, secondary findings, and data breaches all tied for second as shown in Fig. 3.6. Different companies would mention different subsets of risks, with some companies trying to be more exhaustive (e.g., TellMeGen notes 6 out of 7 types of risks, while LivingDNA only mentions the risk of secondary findings).

The number benefits is lower than risks: 5 benefits over two categories (indirect data altruism or general benefits). The benefits section in consent policies usually states that it is free to participate, that there will be no financial compensation (except 23andMe mentioning the possibility of financial compensation via cash or charitable donations), that results of the research studies will help research but are not provided to the participants (except 23andMe mentioning that they *may* inform individuals of research findings). Indirect data altruism is the most common benefit

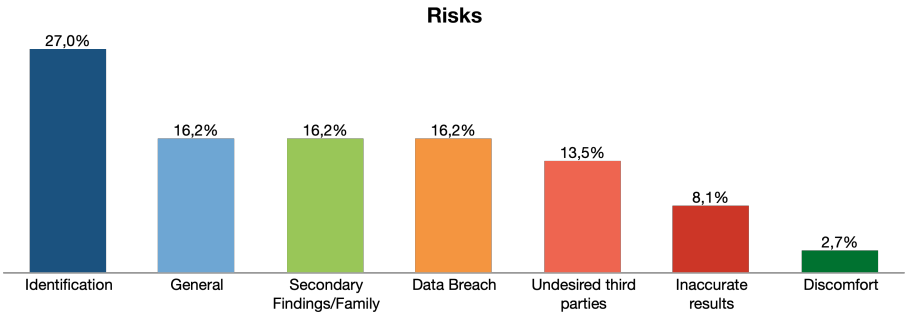

FIGURE 3.6: Risks across companies in descending order, based on percentage of coded segments

and very broadly defined, for example, "*Your participation may help advance scientific or medical knowledge*," (AncestryDNA), "*[...] this will assist academics and researchers to better understand the human species, learn or confirm certain facts and make predictions about future trends. Thus, the more Personal Information for Research is contributed to the Project, the better any potential results*" (MyHeritage), and "*Sometime in the future, you or others, including people who share your ancestry or health characteristics, may benefit indirectly from 23andMe Research discoveries, such as those that improve 23andMe product or services offerings or contribute to ways to prevent and treat disease*" (23andMe).

Interestingly, AncestryDNA includes a section that does not concern customer benefits, but financial benefits to companies or employees from secondary research activities: "*In some instances, AncestryDNA receives compensation from Collaborators who work on the Project. Some of the researchers who are employees of AncestryDNA also have a significant amount of stock or other ownership in AncestryDNA or Ancestry.com.*" The impact on consumer decision making is unclear as there are no monetary values listed and it is posed as a possibility.

3.5.3.1 Location of information

Dedicated sections for risks and benefits were clearly stated in the informed consent policies for secondary research, while risks in privacy policies were found in data security or methodological sections (See Tab. 3.5). For example, MyHeritage states risks of data security issues and that "*[...] while our reasonable security program is designed to manage data security risks and thus help prevent data security incidents and breaches, it cannot be assumed that the occurrence of any given incident or breach results from our failure to implement and maintain reasonable security*," and TellMeGen and LivingDNA mentioned a risk of error in results.

In the privacy policies, the framing is about minimizing liability for the company. They are often passive, as with MyHeritage's statement about security risks place the onus on the individual, as TellMeGen writes: " *[...] in the event that you decide to share your genetic information with health care professionals, you run the risk that said information can become part of your medical history and, because of this, could be in the future accessible to undesired third parties (for example, health care provider service companies or insurance companies). [...] you acknowledge that you will: [...] assume responsibility for the possible consequences.*" This contrasts with the framing in informed consent for research policies; for example, AncestryDNA writes, "*When Biological Samples are physically transferred from us to Collaborators, there is a potential risk that the samples could be lost or stolen while in transit or storage. We take precautions to reduce the likelihood that*

this will happen and your Biological Samples are not transferred with your name or contact information" which includes information about mitigation strategies to decrease risks to the customer.

3.6 Discussion

3.6.1 Informational opaqueness

Regarding RQ1, our findings of vague, incomplete, bloated, and generally confusing statements support previous research about the difficulty reading and understanding privacy policies [137, 93, 144, 300] and lack of transparency into a company's practices. This can lead to customer misinterpretation [17] and incorrect privacy expectations [161]. This is compounded by the findings from RQ4, wherein the risk information was insufficient. This can enhance privacy expectation misalignment. Customers may be concerned about privacy but decide to use the service while not fully grasping the consequences (e.g., challenges to anonymizing data or family implications) [280]. This is then reflected when data breaches and subsequent lawsuits occur. For example, MyHeritage's privacy policies states, "*while our reasonable security program is designed to manage data security risks and thus help prevent data security incidents and breaches, it cannot be assumed that the occurrence of any given incident or breach results from our failure to implement and maintain reasonable security.*" While legally sound, it may not meet customer expectations. A member of a class action lawsuit against MyHeritage due to a data breach in 2018 [228] stated that she would not have used the services if she had known that the necessary precautions were not in place [310]. In October 2023, 23andMe reported a data leak as well. Consequently, multiple class action lawsuits have sprung up in response to the lack of adequate privacy, security, and data breach notification procedures [149]. If companies want to address customer expectations and reassure them, they should prioritize addressing relevant consumer concerns about risks.

Regarding the results for RQ2, transparency requirements were presumably not met in all publicly available information flows, especially with missing data controllers in Art. 13(1)a (missing sender), recipients of personal data in Art. 13(1)c (missing recipient), or the purpose of processing in Art. 13(1)e and if the data would be used for secondary purposes in Art 13(3) (missing transmission principle). Arguably, parameter bloating of 14 different transmission principles in one data flow may not correspond to concise and clear language due to the presence of vague terms. 81% of flows used one or more vague terms from the list provided in [27]. However, companies may have more complete information but present it using vague language to cover multiple cases or any changes that might occur to help decrease the number of updates that would have to be made if it were more specific.

Article 29 WP guidelines on transparency offer examples of how to move away from vague language, recommending changing "*may*," to "*will*" when possible [243] and being more specific about the purpose of processing. For example, instead of "*We may use your personal data for research purposes*" wherein the research is unclear, they suggest "*We will retain and evaluate information on your recent visits to our website and how you move around different sections of our website for analytic purposes to understand how people use our website so that we can make it more intuitive*" to clarify the type of data, analysis, and purpose. This could be combined with CI analyses such as the vague terms analysis to ensure adherence, missing information analysis to ensure that all the needed components are present, and parameter bloating analysis to ensure that the statements are direct and regarding one context. Overall, this could

contribute to users' understanding of their data processing and subsequent rights. On the other hand, it may be too cumbersome to audit or replace all data flows with more direct language because it would need frequent updates; otherwise, policies would often risk becoming out of date.

Auditing and editing policies can be streamlined with algorithmic tools to scan policies and assess the presence or absence of transparency requirements [118] and of GDPR requirements [10, 330]. This can decrease the time needed to identify relevant information within policies. However, this would still require expert analysis to assess compliance, usability, and transparency. Such a transparency-enhancing process would also fulfill one of the necessary steps to ensure data protection by design (Article 25 GDPR) [336].

3.6.2 Lack of relevant transparency

The quality of information for user-relevant attributes (RQ3) that can contribute to providing understandable accountability, safeguards, and information about what rights individuals can exercise under the GDPR is important for building useful policies in the future and was lacking in existing policies.

A common issue was a vague or missing type of data collector and reasons for data use, which can also lead to confusion and misalignment between customer expectations and reality. This is especially non-transparent in consent policies. 23andMe states a possible overlap in their research consent wherein "*Some 23andMe Research may be sponsored by or conducted in collaboration with third parties, such as non-profit organizations, pharmaceutical companies, or academic institutions whose additional expertise and/or resources can help 23andMe make important discoveries*." However, individuals are generally more positive towards academic and non-profit institutions [17] compared to pharmaceutical companies and should have the information and ability to only consent to certain types of purposes by certain types of data collectors. In addition, while genetic data for research is often shared in the hopes of benefiting society [17, 161], some employees may be benefiting financially (as shown in the results). While the statements in consent policies may be vague about who they partner with, in 2018, 23andMe partnered with pharmaceutical company GlaxoSmithKline for a 4-year partnership to develop for-profit drugs [100]. Participants may not know that if they consented for research purposes, they were also giving a pharmaceutical company their data for for-profit purposes. Academic and non-profit research may also convey an image of altruism and lead individuals to believe that [161], while academic partnerships can lead to for-profit interests in the very same research outcomes [177, 76]. To address this, some additional user-relevant attributes might include whether the data will be publicly available (as opposed to only available within the company), or what the financial motivations are. This could add to the previously discussed ethical review. Thus, customers can more easily match their expectations with reality to make informed decisions.

The ethical review of data sharing is especially interesting as a method for sharing the governance structures in place to support individual choices about their data processing. Presumably by experts, it would alleviate some of the individual burden of protecting one's data. Such third-party oversight, especially if independent, can be a useful safeguard the transfer and use of sensitive data by other parties for research purposes (either internal research or with third parties). However, this may not completely safeguard the information if the company's data practices are still opaque.

3.6.3 Collective risks and harms

Overall the risk and benefit analysis (RQ4) found a lack of communication around collective risks and benefits in the public policies. This is especially concerning as such potential harms may not only impact the individual customer, but also their current, past, and future genetic relatives. The risk of re-identification and identification via genetic relatives has an increasing likelihood of occurrence as larger datasets are collected [120]. In the US, the Golden State Killer suspect was identified from a third cousin who was a distant relative from the 19th century [292], whose data were uploaded on the public genealogy database GEDMatch, whose purpose is for people to find relatives for personal reasons. Some DTC genetic testing companies offer genealogy communities as well, and the possible entanglement with law enforcement is questionable. For instance, FamilyTreeDNA works directly with law enforcement in the USA [279]. Though similar collaborations are not yet reported in the EU, they may nevertheless impact European citizens or EU residents. An open question concerns whether law enforcement should be allowed to use data that was originally contracted to help consumers understand their ancestry and health to search for possible criminals. No privacy or consent policies mention the risks to genetic family members [102] or suggest individuals to discuss the genetic test with their families to make a group decision. This lack of common deliberation and the silence around this issue can be especially harmful since privacy is contextual and networked [40], so even if one individual tries to protect their privacy, another person's privacy choices may directly impact them. Conversely, individuals may not share the benefits of taking a DTC genetic test or understand how indirect research done with their data may enrich their lives. Such relational risks should be clearly stated to give the impacted individuals and their family members a chance to understand the shared consequences and decisions together.

While collective notice and consent are important, there are no clear guidelines for implementation yet. A study was conducted that found no global consensus on how to give notice [320]. While safeguards were suggested, it is not required to mention collective risks or safeguards. Thus, it is very uncommon and not surprising that the DTC genetic testing companies sampled did not mention any. Historically, collective consent has been used in biomedical research to respect indigenous cultures and their collectivist governance structures, wherein communities are consulted and asked for consent [227, 250, 130] due to indigenous legal requirements. While this research has mostly been performed in person, a dynamic digital platform for collective dynamic consent in Australian Aboriginal and Torres Strait Islander communities has been positively discussed as a solution [253], moving collective consent into a more digital age. The need for collective family consent is especially discussed by legal and ethical scholars as well [214, 166]. However, translating such proposals into actual safeguards and digital tools is still an ongoing subject of research.

3.7 Limitations

This study had limitations based on the corpus chosen and methodologies. We only assess 6 companies, which might not be very representative. However, We chose to examine public privacy and consent policies, and did not include other pages. We were solely interested in the text and the quality of the information in the privacy and informed consent policy. We analyzed the policies from an expert perspective, and internally companies may have more information. Some results were coded by

a single author and should be extended and replicated to increase reliability, and should include user studies to validate our findings.

This was preliminary work to survey general practices and test the suitability of the methodology, not intended to be an indictment of any companies. We did not reach out to companies to clarify their policies and justify their choices. Future work is being carried out to validate the method of CI analysis and research collective notice and consent.

3.8 Conclusion

We were motivated by the lack of user-centered transparency about companies data practices and how to methodologically improve them. It can be especially risky when sensitive genetic data is processed, as with our use-case. We contributed empirical data about the informational transparency and user-relevancy of privacy and research consent policies from 6 leading DTC genetic testing companies. We identified 62 information flows across companies about genetic data sharing, wherein 81% used vague terms, 16% were missing information such as the recipient, and 32% were bloated with up to 8 different recipients, 9 data types, and 14 reasons for data sharing. This demonstrates a lack of informational transparency and a possible misalignment with GDPR transparency requirements. The quality of information within flows was not very relevant to users – for example with "unspecified research" as the purpose of processing present in 12% of flows. The reported risks were framed as individual, even when collective, and varied widely by company. We suggest using CI to audit and develop policies with user-relevant terms and better risks communication to increase usable transparency and decrease misalignment between customer expectations and actual practices. We hope this work encourages more nuanced, human-centered policies.

3.9 Appendix

Link to codebook: http://tinyurl.com/2j85rzaa. Other data is available from the corresponding author on reasonable request.

Company	Policy Type	# and type of risks	# and type of benefits
23andMe			
	Privacy Policy	0	0
	Research Consent	6 (discomfort, identification/expose family, data breach, identification/re-identification, identification/phenotype, unknown	2 (indirect)
AncestryDNA			
	Privacy Policy	0	0
	Research Consent	9 (data breach, identification/re-identification, data breach/sample lost, secondary findings, undesired third parties/law enforcement, undesired third parties/insurance, identification/expose family, undesired third parties/insurance/discrimination, unknown)	1 (indirect)
FamilyTreeDNA			
	Privacy Policy	0	0
	Group Project	1 (general)	1 (general)
LivingDNA			
	Privacy Policy	1 (inaccurate results)	0
	Research Consent	2 (secondary findings, data breach/sample lost)	1 (indirect)
MyHeritage			
	Privacy Policy	4 (data breach; DNA match: secondary findings, identification/re-identification, unknown)	0
	Research Consent	2 (identification/re-identification, secondary findings	1 (indirect)
TellMeGen			
	Legal Terms	1 (inaccurate results)	0
	Legal Consent	5 (secondary findings, inaccurate results, general, undesired third parties/insurance, identification/phenotype)	0
Total		**31**	**5**

TABLE 3.5: The number and type of risks per company separated by policy type.

Chapter 4

Contextual Integrity for Privacy Policy Elicitation and Validation

Author Contributions: Conceptualization, Investigation, Formal analysis, Methodology, Software, Validation, Visualization, Investigation, Writing - original draft, Writing, review & editing

Abstract: Understandable and verifiable privacy policies can be difficult for end-users to understand and for businesses to write, especially when information flows include complex data sharing. Typically, data-sharing information flows reside in lengthy privacy policies full of legal jargon and vague terms, which hinders both communication among different stakeholders and compliance verification of privacy, security, and legal requirements. What can help is following a rigorous methodology to analyze information flows to ensure understandable and usable privacy policies for stakeholders. Thus, the problem shifts to what is an effective, and usable methodological framework for the task. The contextual integrity (CI) theory of privacy offers a framework to analyze privacy policies, but its effectiveness and usability for developing better policies for organizations needs evidence. This work studies the CI framework to elicit and validate better information flows for privacy policies in a real-world use case. 13 employees across a welfare tech company in

Norway have been interviewed to assess the usefulness of CI for communication, GDPR validation, and perceived usability. Participants felt the method was useful for representing and analyzing information flows due to the categorization with discrete parameters and structure, which enhanced their understanding, scoping, and quality of information. They confirmed its usefulness for writing better documentation, auditing, and better communication with experts or customers. It was also thought to be usable as a tool for businesses. This pilot offers promising empirical evidence for CI methods in general to develop and validate complex information flows for organizations.

4.1 Introduction

As data sharing information flows become more complex in a highly interconnected world, accurate, complete, valid, understandable, and transparent privacy policies can be more difficult to produce and maintain. This can negatively impact businesses, increasing the burden of communication between different departments and with customers and partners, increasing the effort involved in information requests and audits, or making it difficult to elicit security requirements. It can also be a detriment to end-users, who do not receive a clear privacy policy for how their data is being used, when, and why. Privacy policies are part of the external perception of a company, and may impact their market position.

Privacy policies have been critiqued for being unreadable [93], time-consuming [205], incomprehensible to non-legal experts [140, 268], and useless as a decision-making tool [144]. Although vague legal terminology may be a deliberate strategy to be legally comprehensive and accurate, especially in unclear situations for future-proofing [27], these may not be fully in the spirit of the GDPR. The GDPR sets transparency requirements in Art. 12-13. It states that information must be presented in an "intelligible and easily accessible manner, through the use of clear and plain language" (Art. 7(2)) and Recital 39 of the GDPR clarifies this point: "It should be transparent to natural persons that personal data concerning them are collected, used, consulted, or otherwise processed and to what extent the personal data are or will be processed ... any information and communication relating to the processing of those personal data be easily accessible and easy to understand" [243].

To address this challenge, research methods could be adapted to help organizations break down the different elements of a complex data-sharing workflow and also analyze those elements for quality, such as completeness, directness, and alignment with legal requirements. One method of analyzing privacy policies comes from the theory of CI from Nissenbaum [226], considers privacy from a sociotechnical lens and has been used to analyze publicly available privacy policies [137, 300]. In those cases, it was used *ex post* to evaluate privacy policies for quality (e.g., vagueness, conciseness, directly) and alignment with privacy norms with general suggestions for improvements. However, would it be useful and feasible to apply CI directly with organizations *ex ante* to create more understandable and verifiable data sharing information flows and privacy policies?

This led to our research questions: (RQ1) Would the CI method be useful for creating more understandable and verifiable data sharing flow elicitation and validation for businesses? (RQ2) Would the CI method be usable in a complex business scenario?

We offer empirical data to help answer the question – we performed a mixed methods study with both a questionnaire and semistructured questions. We interviewed 13 employees of a medium-sized Norwegian welfare-technology company to better understand the perceived usefulness for better analysis, communication, and validation along with usability in their work. This organization was chosen because they have high standards for privacy and security with in interest in improving privacy with support from Author 2, the Chief Governance, Risk and Compliance Officer (CGRCO). They also have an applicable use case with complex and sensitive data flows with many data subjects (See Use case 4.3).

4.2 Related Work

4.2.1 Privacy Policies

While the GDPR addresses both organizations and individuals, privacy policies can fail to be helpful to end-users. They may be forced to use a service, and thus must agree regardless of a privacy policy [140] or agree without sufficient understanding [268]. However, research is also being done to transform privacy policies and test what is most useful [369] and to better understand from a business side the challenges [297]. While privacy policies can be enhanced for greater user understanding (or use privacy assistants), it may be more valuable to do it in the policy design phase so both business and user can take advantage of it. The task of establishing and maintaining privacy policies is time-consuming in complex environments and is often executed for the sake of compliance. The value of the policies for the business is limited in this case, and the business misses the opportunity of using the policies as tools for governance, management, communication, marketing, and audit, placing the burden of understanding heavy and poorly structured policy documents on an end-user. Since the businesses themselves manage personal data of many users, the privacy, transparency, and security of the data processing by design are key. When business starts to perceive privacy policies as public reports on the privacy measures and as a decision-making tool for the users, the quality and value of the policies will improve, reducing the efforts on the user side.

4.2.2 Contextual Integrity

Multiple methods exist to analyze privacy policy information, such as privacy requirements engineering [14], machine learning [363, 371], and the CI framework [299]. Although interesting, requirements engineering follows a broader goal-based method, which resulted in 12 goals and many more sub-classifications; whereas the CI analysis focuses on data sharing information flows and uses 5 parameters. Machine learning is useful for general analysis or policies but still requires expert analysis, so it would be more helpful *ex post*. CI also offers a theoretical grounding in addition to analytical methods. Nissenbaum uses the CI theory of privacy to define that "privacy" must align with existing, legitimized norms of privacy. For example, sharing video streams with a doctor may be appropriate, while sharing the same information with the company for internal research and development could be a violation of privacy. The methods break down data sharing information flows into sender, recipient, transmission principle, attribute, and subject. This offers a clear method for evaluating data flows in privacy policies or other documents, and performing quality analyses. One can analyze if a parameter is missing within a flow,

if terms are vague, complete with all five parameters, and direct (e.g., the transmission principle should not include a long list of all purposes of processing if they are distinct activities) [300].

While CI has been used for multiple purposes, such as to increase granularity and better detect and enforce data leakage protection systems for businesses [301] or as a privacy assistant [296], its use for businesses to help write and audit information flows for privacy policies has yet to be studied. The inspection of data flows, evaluation of completeness (if all five parameters are present) and quality (if vague terms are used, if there are too many discrete statements in one parameter, etc.) have not been researched in collaboration with organizations to develop better data flows. This may aid in data protection by design [32] to clarify information flows and implement necessary technical privacy or security measures. Additionally, Article 12 (GDPR) sets user-centered transparency requirements that encompass the "quality, accessibility and comprehensibility of the information" wherein the CI method could be used to analyze and improve information flows of the data processing practices[243].

4.2.3 Technology Acceptance Model

Using the Theory of Reasonable Action (TRA) and Theory of Planned Behavior (TPB) from psychology, the Technology Acceptance Model (TAM) has been used to explain behaviors towards adopting technology in the information systems in the information and communication technologies (ICT) field. Acceptance of the technology (commonly interpreted as use) is a three-stage process, where external factors such as the system design trigger a user's motivations which include *Perceived Ease of Use (PEOU)* and *Perceived Usefulness (PU)*. These can then inform the response, or acceptance of the technology. PU is posited to have a greater influence on acceptance, while PEOU interacts with PU, and they both influence the *Attitude Towards Using* the technology, which then informs their *Intention to Use*, which cumulates in usage behavior [71].

Since its inception in 1989, it has been extended, critiqued, and used. TAM is the model most used to explain and predict the use of the system [176] in research in many domains. Some studies have found it to be predictive of use, while others do not[179]. In a review of 20 healthcare TAM studies, it was found to be a strong predictor of uptake, but it is not without drawbacks [136]. In the case of the healthcare TAM reviews, they believed that a missing variable was related to the healthcare context, which would enrich the studies to be more predictive [136]. TAM has been critiqued for its choice in variables (or lack thereof), leading to many extensions. For example, TAM2 added 5 variables preceding PEOU and PU [347]. Other researchers have added beliefs that would affect the *Attitude Towards Using* [155] or the *Intention to Use* [326].

TAM is still widely used and accepted, with 170 publications in 2020 [87, 217]. Due to its long history, the methods have been validated. Instead of looking at new methods Moody used the original TAM to develop a theoretical model for validating design methods and offered questions that would address PEOU, PU, and *Intention to Use*. The Method Evaluation Model (MEM) is meant to evaluate of system design methods and modelling languages [217]. These have been used in other studies due to their reliability and validity [234, 346].

4.3 Methods

Use case We base our pilot study on a use case with a Norwegian welfare-technology company on one of their services with a complex information flow. The company is contracted by the local government (e.g., municipalities) to help provide services for home-care patients who do not require hospitalization but still require care for various reasons (Figure 4.1). The services include active safety for elderly patients and patients with dementia through alarm response and location tracking, planned supervision for remote care visits, continuous supervision through events from cameras and various sensors such as medicine dispensers, door locks, and bed sensors, chronic disease monitoring through remote communication with medical devices and questionnaires. The company offers web and mobile applications with vetted third parties (medication dispensing machines, video surveillance, etc.), and more. They can also involve the patient's next of Next of Kin (NoK) to help the patient and decrease the burden on healthcare professionals. NoK can be family members or community members who agree to help with responsibilities, living with or near the patient. Thus, video data may involve streaming video and audio of a patient, NoK, and/or healthcare professional for different purposes.

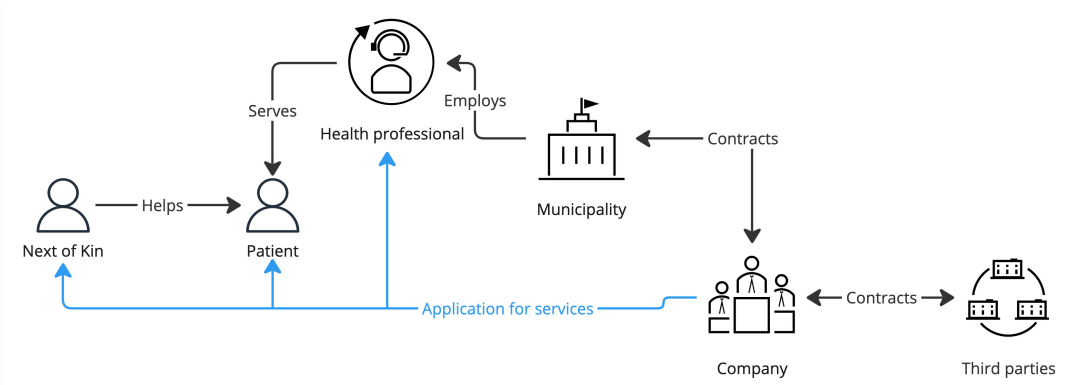

FIGURE 4.1: Overview of stakeholders in the use case with the Norwegian company

Not only is the flow of information incredibly complex with multiple third-parties, but consent can enter a gray area. Patient consent is implicit in order to access the services, while NoK consent is usually blanket consent given to the municipality. In cases where the data may be sensitive (e.g., video data), this consent may be uncertain because sensitive data is subject to higher protections, including explicit consent or another legal basis (Art. 9 GDPR). If the data flows are vague and it is unclear exactly who, why, and when individuals have access to data, it would be even harder to examine compliance.

4.3.1 Participant Demographics

We interviewed 13 employees across different teams at a mid-size Norwegian welfare-technology company (approximately 23% of the total employees) (Table 4.1). We used convenience sampling within the criteria of roughly even participation across DevOps, Product and Design, and Direct and Partner Sales, and Governance, Risk and Compliance (GRC) who were available within the time frame of work hours from November 22 - December 1, 2023. The interviews were approximately 30 minutes long and recorded with Microsoft Teams, either in person or online. For online participants, the notice, consent, and questionnaire were shared and collected using

Microsoft Forms, while the paper forms were used in person. Audio recording was only stored locally. The questions asked are shown in Table 4.2.

TABLE 4.1: Demographic distribution of participants (n=13)

Category		Count (n)	Percent
Sex	Male	10	76,9%
	Female	3	23,1%
Years of Experience (total)	0-5	1	7,7%
	6-10	1	7,7%
	11-15	6	46,2%
	16-20	1	7,7%
	21-25	1	7,7%
	25+	3	23,1%
Department	DevOps	3	23,1%
	Product	4	30,8%
	Direct Sales	4	30,8%
	Partner Sales	1	7,7%
	GRC	1	7,7%

Participant roles in relation to the consent processes included:

- **Product and Design** Product employees work on customer and market insights and user stories, specify consent functionality requirements and solution design, and prepare product documentation.
- **DevOps** Developers work on technical implementation of requirements from the Product team.
- **Direct Sales** Direct Sales employees work closely with agreements and customers and ensure customer success in onboarding and application of products.
- **Partner Sales** Partner Sales team works on agreements and relationships with product resellers.
- **GRC** The Governance, Risk, and Compliance (GRC) team is part of Management and responsible for the quality management system, including understanding external security and privacy requirements and embedding them into internal policies and guidelines for products and departments.

Participants included three heads of the department and ranged from 0-26 years of experience at the Company, with an average of 7 years. When possible, senior and junior members of the same team were interviewed.

TABLE 4.2: Demographic questions asked to participants surrounding job title, years of experience, and familiarity with consent on a 5-point Likert scale.

	Question
Q1	What is your current job title?
Q2	How many years of experience do you have in your field (overall)?
Q3	On a scale of 1-5 from Rarely to Always, how often do you interact with consent in your work (any part of the consent process like informed notice and consent, legal compliance, giving consent, etc.)?

4.3.2 Interview Protocol

Participants were presented with information about the study and asked to consent to data collection. Then the author presented the method in a PowerPoint to explain the theory of contextual integrity and the method used to improve privacy policies. An example of creating a contextual integrity data flow from existing information from a data processing agreement was shared. The presentation included the use case described in Use case 4.3 and information about CI theory and methods described in Related Work 4.2. CI was introduced by Nissenbaum in 2009 [226], theorizing that privacy is provided by appropriate information flows that conform with contextual norms, and privacy is not just about control over information but is a socio-technical phenomenon that evolves. To help assess the appropriateness, CI breaks down information flows into 5 parameters: the actors (senders, recipients, and subjects), data type (attribute), and condition (transmission principle). All five must be present to understand the context, and changing a parameter can affect appropriateness (e.g., sharing health data with a doctor vs. a salesperson). Once annotated, they can be compared with other versions for incompleteness, vagueness, conciseness, and more. [300]. To demonstrate the process, an excerpt from a data processing description used for data processing agreements, privacy policies, and data processing protocols was used. It showed information broken down by GDPR category (e.g., Purpose of Processing, Processing Activities, Data Subjects, etc.). This was analyzed using the CI framework. It was shown that there was missing information (e.g., the purpose of processing fits into the transmission principle but does not indicate the subject, recipient, attribute, or sender), vague terms (e.g., authorized personnel), and possible misalignment with GDPR transparency requirements due to vague and unclear information flows. Lastly, participants were shown a complete CI information flow with a sender, recipient, attribute (with information that it is a special category under the GDPR), transmission principle (including security measures and one purpose of processing), and subject (choosing only the NoK from multiple data subjects) (Figure 4.2). They were told to imagine multiple data flows for different subjects, data types, etc. Due to resource constraints, participants did not test the method themselves, but could ask questions and envisioned themselves using it for their roles.

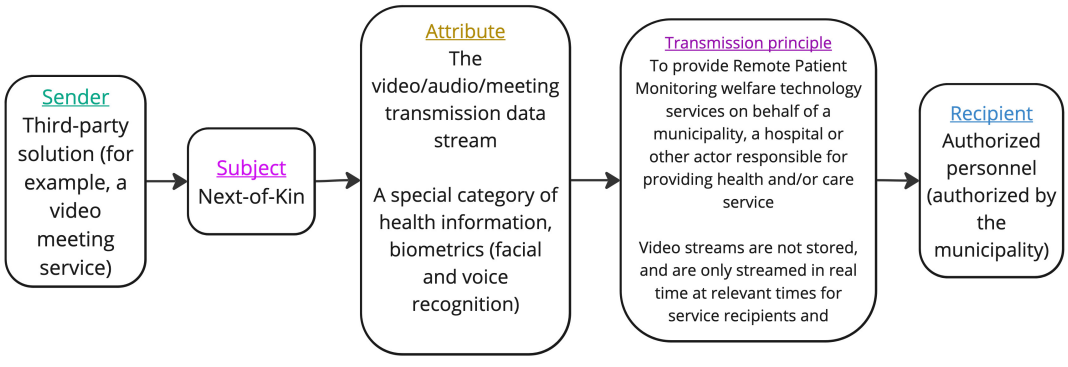

FIGURE 4.2: Next of Kin CI information flow (shown to participants as a table)

Participants were then asked to complete a questionnaire with demographic questions about their job titles, years of experience, and familiarity with privacy policies. These were Q1-Q3, while Q4 is part of a parallel study we will discuss in the next chapter. Q5-7 assess Perceived Ease of Use (Table 4.4) and Q8-15 assess Perceived

Usefulness (Table 4.3.2) using a 5-point Likert scale (Strongly Agree, Agree, Undecided, Disagree, Strongly Disagree) based on methods by Daniel Moody [217]. This is based on the TAM to evaluate perceived efficacy as a precursor to actual adoption. Questions were framed both positively and negatively to reduce unconscious responses. To increase the depth of answers, Q16-18 were open-ended questions to explain their reasoning for previous questions about usability (regarding analyzing and representing information, GDPR validation, and communication), while Q19 allowed for any additional comments or concerns (Table 4.4).

TABLE 4.3: Perceived usefulness Likert scale and open-ended questions from [217] with optional open-ended questions

	Question
Q8	I believe that this method would reduce the effort required to document privacy policies.
Q9	Privacy policies represented using this method would be more difficult for stakeholders to understand.
Q10	This method would make it easier to verify whether GDPR requirements are met.
Q11	Overall, I found the method to be useful.
Q12	Using this method would make it more difficult to maintain privacy policies.
Q13	Overall, I think this method does not provide an effective solution to the problem of analyzing and representing information in privacy policies.
Q14	Overall, I think this method is an improvement to the standard privacy policy.
Q15	Using this method would make it easier to communicate privacy issues to end users.
Q16	Please explain why or why you do not think this method provides an effective solution to the problem of analyzing and representing information in privacy policies.
Q17	Please explain why or why not you think this method would make it easier to verify whether GDPR requirements are met.
Q18	Please explain why or why not you think this method would make it easier to communicate privacy issues to end users.
Q19	Please describe any suggestions, concerns, or feedback from this method.

TABLE 4.4: Perceived ease of use questions derived from [217]

	Question
Q5	I found the procedure for applying the method complex and difficult to follow.
Q6	I found the method easy to learn.
Q7	I found the rules of the method clear and easy to understand.

4.3.3 Analysis

The data was analyzed in R (questionnaire answers) and MaxQDA (open-ended answers). The negatively framed Likert scale questions were adjusted in R, then the median was calculated since the data is ordinal. After all the interviews were finished, the Author 1 used the General Inductive Approach for evaluation data [329]. This method was chosen because of its ability to establish transparent and defensible ties between the research objectives and the summary of qualitative findings derived from the interviews. It also helps to develop a model to explain the underlying phenomena. Unlike the grounded theory approach [314] which also uses an inductive strategy, the steps differ in that coding is combined in the general inductive approach. In addition, it's goal is a theory, while the general inductive approach may result in a model or framework that the categories and themes fall into. The general inductive approach presumes research objectives, such as the research questions we posed earlier.

To perform the qualitative analysis, Author 1 and a master's student cleaned the interview transcripts. Microsoft's transcription software generated text automatically, then it was edited for grammar, filler words, and repeating clauses. Then close reading of the text took place, after which interesting text was identified by Author 1. After some collaborative coding with the master's student, we independently labeled codes. Discrepancies in coding were discussed and resolved. We collaboratively refined the categories to define the hierarchy. The codebook is available on request.

4.4 Results

4.4.1 Perceived Usefulness for Better Data Flows

The median perceived usefulness of all participants over 8 questions in the usefulness category was 2, or "agree" on the Likert scale as displayed on Figure 4.3. The Sales department varied the most, with some ranking questions as strongly disagree (a participant in Partner Sales) and strongly agree (a participant in Direct Sales), while DevOps and GRC department ranked questions as "strongly agree" more frequently than others (GRC for 6 out of 8 questions, and multiple members of DevOps for 7 out of 8 questions).

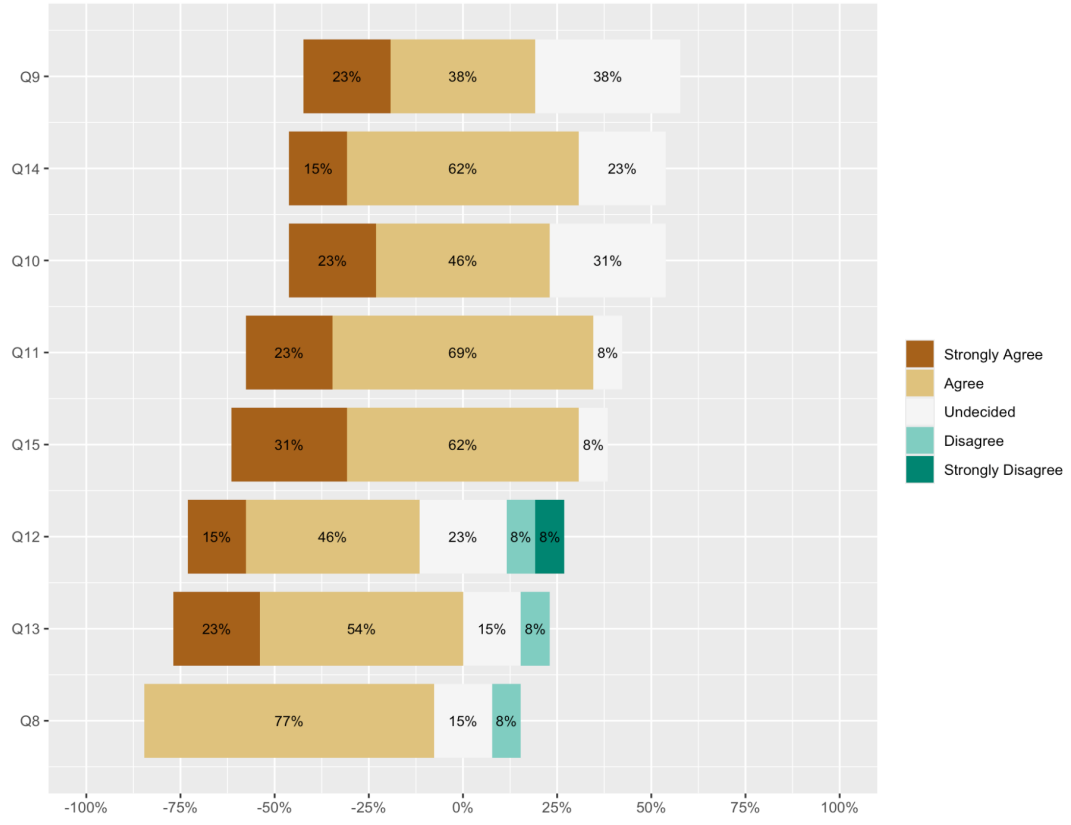

FIGURE 4.3: Perceived usefulness scores by question on a Likert scale where 1 is for *strongly agree* and 5 for *strongly disagree*, arranged by most positive to least positive responses (total n=13)

4.4.1.1 Auditing and Documentation

The method was useful for analyzing information, with coded segments across participants revealing that it helped with scoping (n=3, 25%), quality of information (n=2, 17%), and auditing (n=5, 58%). One participant stated how it would help with scoping and auditing, "*I have been in a lot of risk analysis with the customers, and on the customer side of it. It's like you get lost in analysis, suddenly you're analyzing a whole other scope. So, it will be a better method to stay at the problem or the focus*" (Participant (P)6). Other aspects of auditing could also be applicable, as P7 said, "*I think in terms of representing information, it gives us more of an overview of GDPR compliance and gap assessment. It is easy to perform to see missing elements*".

The method also helped with representing information. Specifically, coded segments identified across participants stated that the method offered useful categorization of information (n=16, 38%), structure (n=12, 29%), enhanced understanding (n=9, 21%), and a writing format for information flows (n=5, 11%). The contextual integrity framework was often mentioned as helpful for the categorization of information (e.g., sender, receiver, subject, etc.), structure (due to the tabular format), and overall understanding, with P4 saying, "*It was a smart way of categorizing the different elements and clearly distinguishing who is the subject, which role or person you are considering in the privacy impact for. So I thought it seemed a smart and effective way of splitting information into different categories to reduce the complexity of grasping the privacy considerations*." In addition to grasping the privacy considerations, the framework would help participants write information flows if required, "*I like having a structure and a method. For me, it would be very difficult to write something if I'm just writing text*" (P9).

4.4.1.2 GDPR Verification

10 participants (77%) thought that it would aid in GDPR compliance, though most were unfamiliar with the GDPR and unsure of how helpful it would be. The CGRCO, with GDPR experience, did think that the method would be useful because, "*it will be easier because then I will know how detailed to be, and it will force me to explain it for all subjects. I see that in the example, the sentences are either too generic or too vague to be used by auditors to validate GDPR compliance. We had [an audit] recently, and they struggled to find and understand the processing activities despite [the pre-existing] documents*." 3 participants declined to answer.

4.4.1.3 Communication with End Users

74% of participants (n=11) agreed that presenting information with this method would be helpful for communication, while one participant said no (instead saying it would be more useful for experts), and the last one was undecided. This revealed a nuanced idea of "end-user" that different participants defined as experts, non-experts, customers (e.g., the municipality), and end-users (which included them as a reader). Participants in both Direct Sales have particularly interesting opinions about how this method could serve as an important intermediate step, which could be shared with experts if needed. P12 said "*It's not necessary for the end users to know our method, but if we use the method, I think it would be much more clear, specific, and easy to understand for end users. If the end user is a person responsible for security and privacy issues within the municipality, then it could be good for them*." Another participant from direct Sales could see its use with customers to share when they request detailed information. "*It's the thing we struggle most with the customers and the customers' customer. They always ask, how do you process data in your solution, what do we need to*

document, and how do we agree to the terms? It's a large part of our day-to-day work. I always call [our CGRCO]. But when you do the fragmentation, it's easier to communicate because it's written in a way that makes sense for most people, regardless of your competence in the area." (P8). A member of Partner Sales thought it was best used for domain experts because "*[Privacy policies] do have relevance for companies like us because it regulates how we are supposed to work*". The CGRCO explained that it could be helpful for end-users, "*As an end user, I would like to assess the risk of this data processing. I can see more granularity when the flow is divided. In the example, I see that there are high risk special categories, but it's not clear the actual risk. If we were a company trying to misuse this, we could do some unauthorized things within this policy because it's so vague.*"

4.4.2 Perceived Ease of Use

The median perceived usefulness of all participants over 8 questions in the usefulness category was 2, or "agree" on the Likert scale. The different Sales branches varied the most, with the same member in Partner Sales ranking some questions as "strongly disagree" and the same member of Direct Sales ranking the method with "strongly agree" as for the PU questions. Similarly, the GRC department ranked questions as "strongly agree" more frequently than others, for 2 out of 3 PEOU questions.

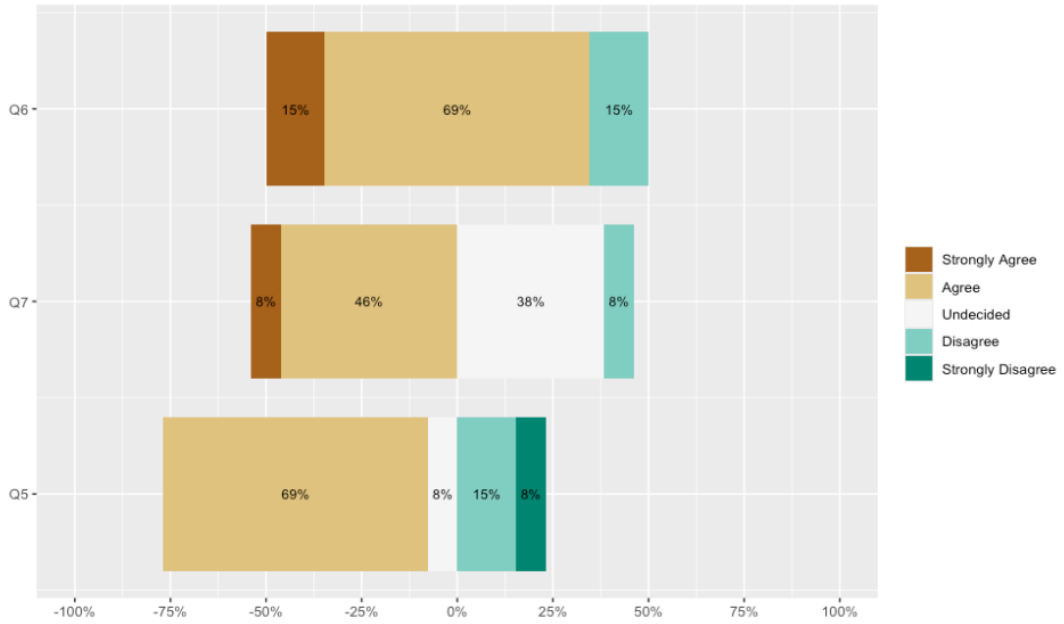

FIGURE 4.4: Perceived ease of use scores by question on a Likert scale where 1 is for *strongly agree* and 5 for *strongly disagree*, arranged by most positive to least positive responses (total n=13)

No open-ended questions specifically asked about perceived ease of use, but answers revealed coded segments about practical issues and improvements. Across all participants, 11 coded segments about practical issues arose. 27% were confused by the terms presented (n=3) and 73% needed more exposure to the method to make more conclusive remarks (n=8). Given that the demonstration of the method took about 15-20 minutes, participants shared concern about the gap between perceived ease of use and actual ease of use. Coded segments about anticipated implementation challenges were also shared, such as defining terms across diverse stakeholders (n=4, 40%), proper audience fit given many internal and external audiences (n=2,

20%), appropriate detail of content (n=2, 20%), and training needed (n=2, 20%) to implement the process across teams. P7 stated the terms and definitions would be challenging, "*For example, authorized personnel for me is something with approval [...] but I see that they use it more like health personnel that are allowed to provide this service to patients. So, this is really about quality assurance of what product managers are writing and thinking with different lenses. Now I'm the CGRCO and read it and it looks like a GDPR definition. If I read it as a patient, is something missing?*."

4.4.3 Suggestions and Sociotechnical Concerns

Suggestions for improvement covered both ease of use and usefulness, such as explicitly using GDPR terminology in data flows (n=2), offering high-level summaries for the reader (n=1), and showing the relationship of data flows to the application or product (n=1). GDPR requirements themselves can sometimes be vague, so P11 suggested that "*The GDPR requirements have to be translated by the municipality (data controller) using this same method, and then the [...] CGRCO sits and compares, and only in that way would you be able to validate*." Instead of focusing on improved information flows for end-user communication, P13 preferred a summary, "*There's a lot of information in the policies. When you click the box, you're hardly spending any time on it. But when you go back, you need the details. So when you accept, you need to have a very high level structured executive summary*."

The sociotechnical category arose from participants' contextual considerations of how this method would be useful. These include trust in the government (n=1, 33%), limits to understanding privacy policies (n=2, 67%), and the debated usefulness of privacy policies to end-users (n=2, 67%). While a clear privacy policy may be present, P10 argues, "*I may not even have the brain power to understand. I would trust that the municipality works for me and my benefit and that they have people who have gone through this rigorously, and that way it works*." P13 also thought that privacy policies were more of a legal obligation than a tool for end-users, "*In many cases, these policies aren't really read. They're just something you need to comply with. If we were to read them, we would be dead until we could do what we wanted to do, so basically what people need is to check the box*."

4.5 Discussion

Participants agreed that the method was useful for analyzing and representing information (median Likert scale of 2). **Auditing** was the most commonly stated use, with participants seeing immediate uses in risk assessment, appropriate scoping when discussing with customers, and gap assessments. This was supported by the way the information was represented. Most participants noted the **structure** with discrete parameters made it easy to see empty cells, break down information into CI parameters, and evaluate privacy considerations for specific information flows (e.g., certain subjects, certain attributes, or data types). Additionally, some participants thought the method would make it easier for **documentation** since it offered a framework for how to translate their work into an information flow or privacy policy since they had descriptions for what information to include in each parameter. The CGRCO also thought this would be useful for validating applicable GDPR requirements and sharing with external auditors. It could also use GDPR terms and tie into the security practices used for categories of data sharing (e.g., sensitive video data).

However, information flows only cover a portion of the information on privacy policies (e.g., right to deletion, etc.), so other GDPR requirements must be considered for a complete privacy policy. The company would also have to develop best practices to harmonize terms and definitions across teams, level of detail needed, and invest in time to train and improve the process.

Participants also agreed the method would be easy to use (mean Likert scale of 2.4), with variance across departments. Some members of the Product team "disagreed" or "strongly disagreed," which may indicate that some departments would have more difficulty or resistance to implementation. However, such resistance is natural. Product-specific privacy policies are usually based on product specifications, which are detailed and structured to allow testing and validation of product requirements, including privacy. Privacy policy is a subset of product specification, and structuring it in the same manner is efficient. From this perspective, the method will require additional efforts from Product while reducing efforts of end users, customers, and GRC and audit functions.

In addition, participants offered unique insights into potential issues. The CGRCO and others thought that clearly defining terms across diverse stakeholders (e.g., product, security and compliance, patients, municipalities, etc.) would require strategy and effort. The method itself does not offer guidelines for unified definitions. Thus, training and best practices should complement the CI method.

If CI flows were shared on a privacy policy, participants thought it would be helpful for communication with end-users, but it might not impact them as greatly as internal stakeholders, experts, or customers. This is interesting because some participants brought up how the main responsibility (e.g., consent, reviewing privacy policies) should fall on the government (e.g., municipality) and not on the user. This may reflect trust in the government, as Norway has a good reputation with residents and ensures a high level of trust [57], which may not be true in other countries. This may be interesting to explore, as it would offer additional safeguards in addition to individual autonomy to help ensure lawful and ethical data practices.

One suggestion was to layer information when asked for consent; end-users could be given high-level summaries if they do not wish to evaluate every aspect of the privacy policy and to have the privacy policy available for any future use. Though it is an additional effort, layered information is recommended by the Article 29 WP guidelines on transparency [243] and designers who have surveyed the field [285]. Layered consent coupled with a digital platform, such as dynamic consent, has already been effectively implemented and researched [123, 43, 37]. However, it requires a significant effort, as municipalities in Norway currently use paper consent forms, and one example reviewed by Author 1 used broad consent for unspecified purposes. Despite a concerted effort by the government to modify their consent processes, the company could improve its communication, documentation, and validation of data sharing practices.

4.6 Limitations and Conclusion

Limitations include the small sample size and scope of the study, which only asked for self-reported perceptions. Our sample may be biased due to the convenience sample, which is skewed towards men and those with many years of experience. Future work should iterate on this, testing implementation, increasing the sample size, and sampling more equitably across demographics to be more generalizable to other types of companies outside Norway. The Author had the opportunity to

visit and interview most employees in person, and extending the study to other organizations was beyond the resources for this research.

Overall, writing complete, transparent, and understandable data flows for privacy policies is not an easy task, especially with complex networks of data sharing (See Use case 4.3). This can be compounded by legal uncertainty, where some practices may or may not be compliant. These issues compound when information flows cannot be traced, verified, and explained to ensure compliance or that proper security measures are implemented. From our research, application of a CI framework – with clear sender, recipient, attribute (data type), transmission principle (purpose of processing and other details), and subject (who the data refers to) was well received by the company and could be useful for enhancing communication and auditing, as well as reasonably easy to use for the employees. If adopted, this would offer a promising direction for businesses to have more clear internal practices that they can share with relevant parties, or adapt as needed. However, the pressure to accept any terms due to necessity still exists. A patient could not opt-out if they felt the data sharing flows were not secure if only one provider exists. While businesses themselves are concerned with audits from third parties or the government, especially for sensitive data sharing, data controllers (such as municipalities) should have the competency and care to evaluate data sharing agreements and those they employ. Norwegian welfare technology companies bid for a contract from the government, so to differentiate themselves they seem to have high competency and care, as the sensitive video and audio are only streamed for short periods in this example, and not used internally for improving services or quality assurance. However, data breaches still occur. While enhancing understanding, validation, and writing of data flows may help to decrease risks and harms, it also exists in a network of trust and safety nets, which all need to be strong to protect privacy.

Chapter 5

User Perceptions on Potential Adoption of Consent BPMNs for SMEs

Author Contributions: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Investigation, Writing - original draft, Writing, review & editing

Abstract: Informed consent processes for complex cases with multiple stakeholders and can be difficult to understand, analyze, and document, especially with various legal and technical requirements in the EU. Businesses may struggle to integrate their system and user requirements with those from the GDPR. Textual requirements documentation can make it difficult to identify conflicts or inconsistencies, communicate across teams, and map to requirements. Instead, a process model may be more suitable to enhance understanding, problem-solving, and communication. One such standardized tool to visualize business processes is the Business Process Model and

Notation (BPMN), developed to be understandable to both domain experts and end-users, yet expressive enough to capture complex workflows. While commonly used in businesses, they have not been researched for consent specifically. We were interested in applying BPMNs in a collective consent requirements engineering process for a mid-sized Norwegian welfare technology company. A patient consents to home monitoring with video surveillance along with a next of kin (NoK), any person(s) who agrees to help care for the patient. As video data is a sensitive category under the GDPR, implicit consent, and explicit consent coincide with multiple responsible parties in a complex processes. We presented a collective consent BPMN to 13 employees to assess the usefulness of consent BPMNs for tracing consent, identifying gaps and GDPR validation, and giving high-level overviews. Participants felt the method was useful for those purposes, with some seeing concrete uses for their work, like identifying key steps. It was also perceived to be usable as a tool for businesses. This pilot offers promising empirical evidence for consent BPMNs for businesses.

5.1 Introduction

Giving IC to share, store, and process data is an important part of data protection and privacy. The purpose of IC is to provide an overview of data processing risks and benefits, leading to an autonomous decision that also expresses one's privacy preferences [65] which abides by EU regulations. The GDPR oversees the processing of personal data (any data which can identify a natural person or persons), which may include performing a contracted service, compliance, IC, or more (Art. 6). Within the GDPRs, high standards for IC are set, it must be "freely given, specific, informed and unambiguous" (Art. 4(11)); easily withdrawn (Art. 7(3)); presented in an intelligible and easily accessible form using clear and plain language (Art. 7(2)); explicitly given for sensitive data (Art. 9); transparent in terms of completeness, comprehensibility, and accessibility of the information disclosures (Art. 12, 13 and 14 GDPR); and compliant with the principles of data protection by design and by default (Art. 25 GDPR) [336].

Organizations can find ICs challenging to trace and implement to those standards, especially in complex data sharing scenarios with multiple parties and data types. For example, defining a consent process across diverse teams may lead to different definitions of "transparent" or "complete," or integrating privacy and data protection by design may be difficult without interdisciplinary teams. This can be compounded by legal gray areas without clear guidelines, which leave it up to companies to define and defend their practices. Take video data where a NoK is helping to care for a patient. The legal basis for processing the information can vary because one party is receiving a service and the other party should explicitly consent (Article 7 GDPR) or satisfy other legal bases (Article 6 GDPR). Sensitive data also has different conditions (Article 9 GDPR) [336], and is not always clear upon initial consent (e.g., which patients should enroll in remote home monitoring or require NoK assistance). When consent intertwines in a collective process, as with the patient and NoK, the complexity of complete, transparent, understandable, and legal IC increases.

To address this, we incorporate RE, an iterative process that gathers, analyzes, documents, and validates requirements to better match user and system goals [309] with BPMN to represent, assess, and communicate the consent process [233]. The RE

process can help clearly define high-level non-functional requirements such as privacy, usability, transparency, and compliance using the Easy Approach to Requirements Syntax (EARs) to give structure and constrain the scope [201]. Requirements can be derived from legal regulations, organizational goals, user studies, and more. Then, BPMNs can be used to develop, test, and improve consent processes to satisfy the defined requirements. BPMN is a popular modeling language for businesses [349] and has been applied to privacy-enhancing technologies [258] and developing clinical pathways [287], but research has not yet studied its application to consent.

We report how we defined requirements, developed BPMN consent processes, and tested the method with 13 employees through a questionnaire and semi-structured interview questions adapted from the TAM [217] in a medium-sized Norwegian welfare technology company ("Company") to better understand perceptions of 1) How useful are BPMNs for increasing understanding, problem-solving (i.e., GDPRs validation), and communication of consent, and 2) How easy to use is the process is for employees in a SME? This organization was chosen for its high standards in privacy and security and its interest in improving privacy and managing collective health data flows. This reflects the needs for effective methods for complex collective data flows, and we received support from Author 2, the CGRCO, to carry out research. We discuss our lessons learned about the use of consent BPMNs for SMEs, which include its usefulness for understandable documentation, identifying key issues, and providing necessary professional resources to integrate RE and BPMNs into daily work. These lessons offer insights into the pros and cons of consent BPMNs for researchers or similar SMEs looking to improve the business practices regarding complex data sharing and consent processes.

5.2 Related Work

5.2.1 BPMNs and RE

Modeling is a powerful tool that can enhance the understanding of processes, identify inconsistencies, test for conflicts, and help stakeholders communicate [119]. When they are based on a machine readable standard language, as with BPMNs, then they can be shared and combined with other languages into executable processes to clarify and expedite the development process [55]. While other process modeling languages exist, BPMNs are a widely recognized ISO standard and used in businesses such as IBM, Microsoft, and more [349]. It was developed over 6 years by industry professionals [266] and is now maintained by the Object Management Group, with specifications, certifications, and many other resources (https://www.omg.org/bpmn/).

BPMNs were developed to be a modeling language for business processes that is formalized, expressive, and understandable to experts and end-users using standardized visual representations of actors, events, logic, etc. [64]. It is similar to activity diagrams in Unified Modeling Language (UML), describing complex workflows, creating simulations, and testing the execution; however, it offers a higher-level overview. While UML captures object oriented software processes for development, BPMNs capture processes oriented business workflows for a wider audience of business analysts, developers, and other stakeholders. BPMNs can also be mapped to XML and be executed in Business Process Execution Language [235]. A business process diagram contains flow objects (events, gateways), connected by connecting objects (sequence flows, message flows, associations) within swimlanes

(pools and lanes), and can contain artifacts (data objects, annotations, groups) (See Appendix 5.8.1).

Although BPMNs are widely used, there are still challenges. In Recker's work, they describe several challenges from 3 years of interviews, surveys, and analysis across businesses. One is the lack of training, with few process modeling staff even in large organizations, and little formal training. Another is the lack of formal support for business rules, which may necessitate supporting documentation. Then scoping and decomposition is a large issue because there are no default solutions. Lastly, the flexible interpretation of many elements can create confusion [266]. Quality issues often arose from the misuse of BPMN elements or concepts such as the splits and joins, message flow, consistent labeling, and scoping. [178] Each domain will have their own challenges to manage and try to incorporate into BPMNs, as in healthcare it seems the many entities (patients, medical devices, healthcare provider, etc.) pose challenges to useful proper decomposition [257].

While BPMNs have their challenges, they have been useful for enhancing the quality of the RE process [73, 331, 233, 34]. Odeh mapped the RE steps to different BPMN utilities: It can be useful for understanding the scope, identifying challenges, and communicating during the elicitation phase. In the specification phase, it can be useful for visually describing textual requirements and increasing richness. Then in the validation phase, it can aid in inspection, verifying the 4Cs (correctness, completeness, consistency, and clarity), evaluating necessity, and ensuring feasibility [233].

While they might be suitable to address the business issues, how are they received in practice? They must compete with other modeling languages, like UML, or existing systems. A review found that usability tests were few [298]. One study found that UML activity diagrams were similarly effective, efficient, and satisfying to business administration students as BPMNs [29]. This contradicts studies that show that BPMNs are more usable than UML [231], however they are not directly comparable. The former study used graduate students with little prior experience to compare UMLs and BPMNs with different tests, while the latter worked with an existing company who was modernizing its architecture and wanted a modeling language to give context to existing systems. In the latter study, they note that BPMNs do not satisfy all requirements needed by the organization, but overall were ranked higher. In another study [73], BPMNs were found to be most useful for junior analysts, while senior analysts had experience in other systems and did not find it useful.

While BPMNs have not been studied with consent, other models have been explored for consent. First, it is important to distinguish between modeling consent in the computer science term and consent models as bioethical conceptualizations of how to address the needs of informed consent. In 1988, the first paper to use the words "model" and "consent" categorized informed consent into a one-time event, or a process and dialogue. [181] Current models include specific consent[72], broad consent [125], dynamic consent [158], tiered consent[43], and more. These pose theoretical frameworks for collecting informed consent, whereas the implementation of these models has no standards and can vary widely. So, previous work has used different modeling languages with different goals. One study modeled consent using UMLs and focused on access control policies in the health domain, from which patient consent can be derived from these rules [275]. Another work developed a new, goal based modeling language with clear mappings to GDPR requirements. While their thesis work does incorporate BPMNs, they use it to check GDPR compliance

and assume existing BPMNs or existing, large businesses with complete privacy policies to build them from [272].

Thus we were interested in exploring the RE process using BPMNs for consent using the TAM and MEM in the context of a SME to gather evidence of their thought processes.

5.2.2 Research Questions

This led to our research questions (RQs):

- RQ1: Would consent BPMNs developed be useful for understandable and verifiable consent processes?
- RQ2: Would consent BPMNs be usable in a business scenario?

5.3 Methods

5.3.1 Research setting

Our use case is based on a real-world example from the Company that helps to provide services to a patient by involving their NoK. NoK can be family or community members who agree to help with various responsibilities (e.g., medication, symptom monitoring, visits) for the patient. The service existed in the past, with ample documentation and user feedback. However, it is not currently active. The Company was contracted by municipalities to help provide services for home care patients who do not require hospitalization but are monitored for various reasons (e.g., dementia). The Company offers mobile applications for the patient, NoK, and healthcare professionals in an ecosystem with vetted third parties (e.g., medication dispensing machines, video surveillance). The municipality contacts the patient and NoK to both consent in-person and give responsibilities related to the patient's care to the NoK as directed by a healthcare professional. This enters a collective dimension because the patient and NoK should both give specific consent to sensitive data processing, such as video surveillance, that involves both parties. In Norway, patient consent is implicit to access health services, while NoK consent is broad consent given to the municipality.

Through interviews with product managers, video data was identified as an interesting data type to study due to its complex legal gray areas. It is used by healthcare professionals to visually check the patient without having to make a home visit. Due to the general nature of video surveillance, these data may involve recordings of a patient, NoK, and/or healthcare professional and be sent to third parties or the Company. The NoK consent may be uncertain because sensitive data are subject to higher protections, including explicit specific consent or another legal basis (Art. 9 GDPR). If consent processes are vague or undetermined, it would be even harder to track and examine compliance.

Building on this real-world scenario, we proposed a fictional NoK web or mobile application to develop BPMNs. The consent process would be part of dynamic consent, which offers a digital platform for users to manage consent and permissions over time, as well as to access information surrounding their care such as privacy policies, healthcare professional notes, or the terms of consent [158]. The consent process within the application would be integrated with a centralized national health platform (Helsenorge [1], which already exists and is well known in Norway),

[1] https://www.helsenorge.no/

which healthcare personnel and municipalities would also use. For example, digital consent would be collected by the municipality. Then metadata about the consent would be available on Helsenorge for authorized parties to access for confirming validity, expiration, access control restrictions, and more. These would then be used and displayed to users through the application to increase transparency, with options to update any settings (e.g., withdrawing consent). In this application, the patient and NoK would have access to manage their consent (view, give, or revoke consent) for the sharing of data, as with video data.

5.3.2 Derivation of consent process requirements

Author 1 discussed with Author 2, the CGRCO, about the company and use cases. Then, Author 1 informally interviewed the Head of DevOps, software developers, Product managers, and Sales employees to better understand their specific work and challenges. Author 1 was also given access to the relevant internal documentation, which included past user interviews, presentations, and specifications. From this, Author 1 derived high-level goals: *users should be more informed and in control; and the consent process should be more traceable and efficient* and confirmed it with Author 2, the CGRCO. The requirements were developed using EARs [201] and included: *the consent management shall provide traceable and understandable records to users; the consent process shall use privacy by design; the consenting parties shall be informed and under no duress to consent or refuse.* The prototyping of BPMNs was an iterative process aimed to satisfy the requirements, taking place over several weeks. The EARs and BPMN tools were actively applied to structure and visualize the prototype and facilitate improvement discussions. We created artefacts to address the initial collection of consent, re-consent, and collective consent. Of these, we only tested the most complex BPMN, modeling collective consent.

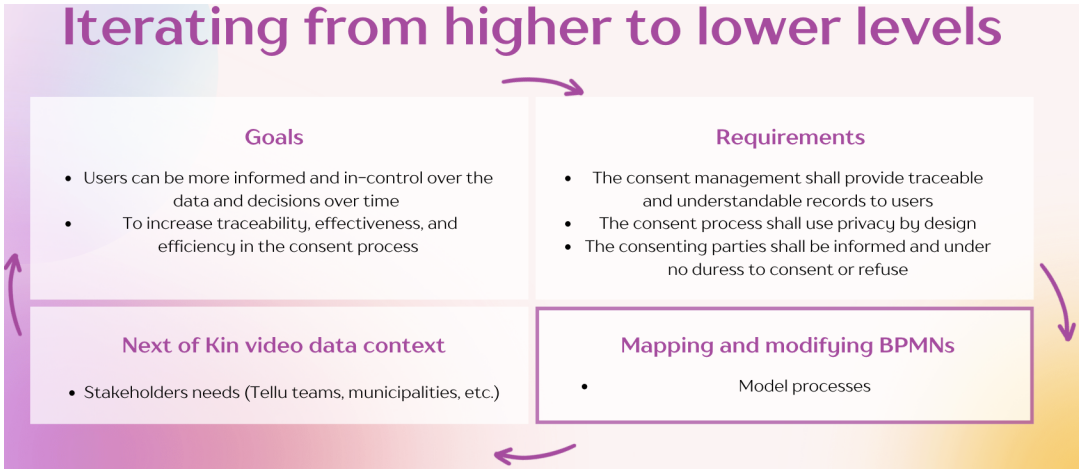

FIGURE 5.1: Explanation of the RE process shown to participants. The goals and requirements reflect those found throughout this process.

5.3.3 Research participants

We interviewed 13 employees across the DevOps, Product, and Sales (Direct and Partner) departments and Management in a midsize Norwegian welfare-tech company (approximately 23% of the total employees) (see Figure 5.2). Author 2 recommended equal representation of the relevant teams and departments for better

diversity of the background and privacy competence. Interviews were subject to availability as they were conducted during work hours, either in person or online, from November 22 - December 1, 2023. The duration was approximately 30 minutes. Participants included juniors and seniors from 0-26 years of experience with the Company, with an average of 7 years. Demographic questions are displayed in Table 5.2 and years of experience with the Company was provided by the CGRCO.

Participant roles in relation to the consent processes included:

- **Product** Product employees work on customer and market insights and user stories, specify consent functionality requirements and solution design, and prepare product documentation.
- **DevOps** Developers work on technical implementation of requirements from the Product team.
- **Direct Sales** Direct Sales employees work closely with agreements and customers and ensure customer success in onboarding and application of products.
- **Partner Sales** Partner Sales team works on agreements and relationships with product resellers.
- **GRC** The Governance, Risk, and Compliance (GRC) team is part of Management, responsible for the quality management system, including understanding external security and privacy requirements and embedding them in internal policies and guidelines for products and departments.

TABLE 5.1: Demographic distribution of participants (n=13)

Category		Count (n)	Percent
Sex	Male	10	76,9%
	Female	3	23,1%
Years of Experience (total)	0-5	1	7,7%
	6-10	1	7,7%
	11-15	6	46,2%
	16-20	1	7,7%
	21-25	1	7,7%
	25+	3	23,1%
Department	DevOps	3	23,1%
	Product	4	30,8%
	Direct Sales	4	30,8%
	Partner Sales	1	7,7%
	GRC	1	7,7%

TABLE 5.2: Demographic questions asked to participants surrounding job title, years of experience, and familiarity with consent on a 5-point Likert scale.

	Question
Q1	What is your current job title?
Q2	How many years of experience do you have in your field (overall)?
Q3	On a scale of 1-5 from Rarely to Always, how often do you interact with consent in your work (any part of the consent process like informed notice and consent, legal compliance, giving consent, etc.)?

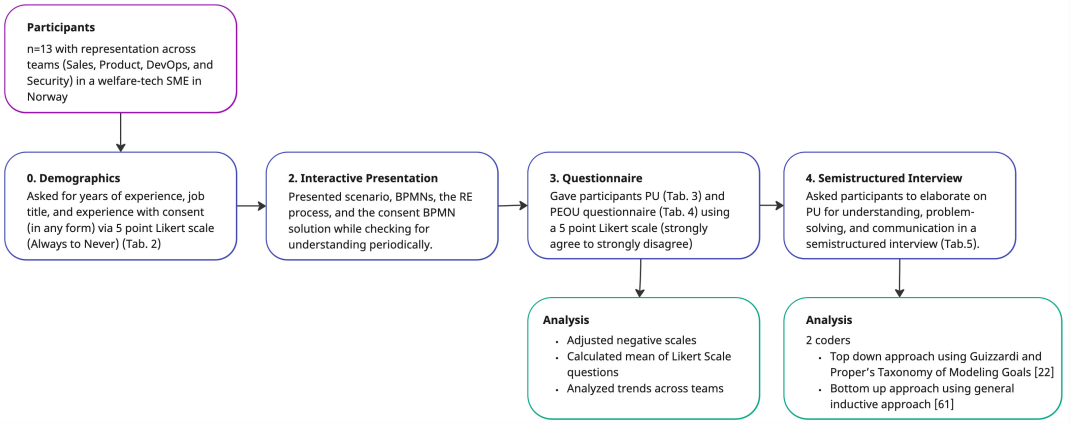

FIGURE 5.2: A diagram showing how participants were chosen based on the team in the company (n=13). Step-wise methods for the interview and questionnaire are outlined: the demographic questions (See Table 5.2), interactive presentation, questionnaire, and interviews(See Tables 5.3, 5.4, 5.5,). The analysis methods for questionnaire and interview data [119] and [329] follows.

5.3.4 Interview Procedure

Participants were presented with information about the study and asked to consent. They were asked for information on job titles, years of experience, and familiarity with consent. Then, Author 1 began an interactive, 1:1 presentation explaining the BPMN method, how the method fits into an iterative RE workflow, and the collective consent use case with the proposed BPMN solution (Step 2 in Figure 5.2).

Proposed consent solution The BPMN shows how the NoK app would gather consent decisions for video data from the patient and NoK in a network of responsibilities from different stakeholders across swimlanes shown in Figure 5.3. The green color code corresponds to transparency requirements, the blue consent requirements, and the purple privacy requirements. The flow begins from the NoK app, where consent is verified by accessing the centralized Norwegian national health data sharing platform (Helsenorge) [2]. Then it would check the necessity and purpose of video data based on the recommendation of healthcare personnel on Helsenorge. If the data are necessary for care, the patient does not need to give consent because it's required for the service. The NoK must give their IC given the sensitive data type with proper notice. Alternatively, if video data is not necessary for the patient's care but could ease or enhance the healthcare personnel's work, both should explicitly opt-in to video data. In this case, consent must be given within 30 days, or validity expires without active opt-in. If the NoK does not consent, the rejection is forwarded to Company personnel, while if the patient does not consent, it ends the service. The app would help users manage their IC, privacy policy, healthcare personnel recommendations, and other transparency elements.

After checking for understanding, participants filled out a questionnaire and elaborated on specific questionnaire questions about usefulness with a semi-structured interview (Table 5.4 with time for open feedback. Overall, 12 PU questions (9 Likert scale questions and 3 semi-structured questions), 3 PEOU Likert scale questions, and 1 open-ended question were asked. Participants were given a questionnaire using a Likert scale [81] with questions adapted from Moody [217] to measure responses to

[2](https://www.helsenorge.no/)

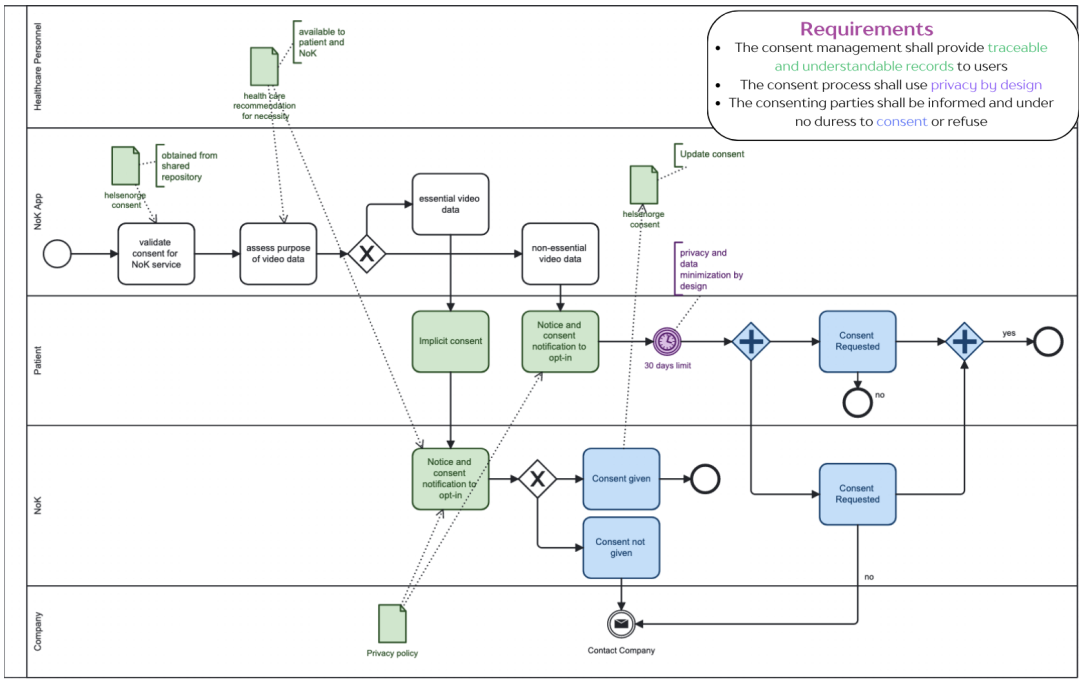

FIGURE 5.3: Collective consent BPMN with requirements mapping shown to participants. Green elements are for transparency, purple for privacy, and blue for consent.

RQ1-2. The 5-point Likert scale went from 1 to 5, corresponding to Strongly Agree, Agree, Undecided, Disagree, and Strongly Disagree. PU and PEOU questions were shuffled and framed positively and negatively. Questions are detailed in Table 5.5 and Table 5.3.

- Perceived Usefulness (PU): What utility does the subject think such a model can provide [81], measured using [217] guidelines.
- Perceived Ease of Use (PEOU): How usable does the subject think of the model [81], measured using [217].

The 3 PEOU (Table 5.5) and 9 PU (Table 5.3) questions were shuffled and framed both positively and negatively to reduce unconscious answers. The answers varied along a 5-point Likert scale from 1 to 5, corresponding to Strongly Agree, Agree, Undecided, Disagree, and Strongly Disagree. To increase the depth of answers, we added open-ended questions for the key PU questions with time for any additional comments (Table 5.5). Quotes in the paper have been paraphrased from participant answers to help protect their identity and account for any grammatical issues.

The Likert scale questionnaire answers were analyzed in R, while open ended answers were coded MaxQDA (`https://www.maxqda.com/`). The negatively framed Likert scale questions were adjusted in R and used for averages and plotting. Open-ended answers were coded a bottom up, general inductive approach to find core issues in the text related to the research objectives [329] and top down approach from the descriptive model goals (understanding, problem-solving, and communication) from Guizzardi and Proper's Taxonomy of Modeling Goals [119]. Themes or keywords were often created *in vivo* using text from participants. Author 1 and a master's student reviewed a subset of the semi-structured interviews and co-coded the text in person. Ambiguous text was clarified upon mutual consensus. Themes or keywords were created *in vivo* using text from participants.

TABLE 5.3: Perceived usefulness (PU) Likert scale questions adapted from Moody [2003] [217]

	Question
Q4	I believe that this method would reduce the effort required to document privacy policies.
Q7	Overall, I think this method is an improvement to current consent documentation.
Q9	Consent processes represented using this method would be more difficult for stakeholders to understand.
Q10	This method would make it easier to verify whether consent processes are compliant with GDPR requirements or goals for the app.
Q11	Overall, I think this method provides an effective solution to the problem of verifying consent processes' GDPR compliance.
Q12	Overall, I found the method to be useful.
Q13	Overall, I think this method does not provide an effective solution to the problem of representing consent processes.
Q14	Overall, I think this method does not provide an effective solution to the problem of communicating the consent processes to different stakeholders.
Q15	Using this method would make it more difficult to maintain consent processes.

TABLE 5.4: Perceived usefulness (PU) semi-structured and open-ended questions adapted from Moody [2003] [217]

	Question
Q16	Please explain why or why you do not think this method provides an effective solution to the problem of analyzing and representing consent processes.
Q17	Please explain why or why not you think this method would make it easier to verify whether GDPR requirements are met.
Q18	Please explain why or why not you think this method would make it easier to communicate consent processes to end users.
Q19	Please describe any suggestions, concerns, or feedback from this method.

5.4 Results

5.4.1 Perceived Usefulness

Overall, participants thought the PUrated an "agree" (median of 2) (Figure 5.4). No participants strongly disagreed about the usefulness of the method. Interestingly, there were some differences in scores across some teams. Two participants regularly disagreed with the PUand were the only participants throughout the interview to feel this way. Conversely, for the "strongly agree," more individuals across the Sales, Product, DevOps, and the singular CGRCO in GRC chose "strongly agree". BPMNs were generally well received, with 12 out of 13 participants (92%) reporting that the collective consent BPMNs were useful for documenting and assessing complex consent processes. Participant 5 (P5) said, "*I support the process visualizations like this. I think that's very good to explain and document these kind of processes. I'm familiar with most of it, but not with the exact symbols and notifications in the method*." Two participants also specifically pointed out the usefulness of the integrated, dynamic consent process shown, such as, "*In this case, patient and the next of kin involving the settings in a central portal like Helsenorge seems to be a huge step forward from many consent processes within healthcare today, which I think you know a lot is happening on paper*." (P4) From the top down and approach to coding from [119], the Understanding, problem-solving, and Communication category came about; while the bottom up approach led to the Professional Resources and Relationship with Consent categories. In the following sections, they will be discussed in more detail.

TABLE 5.5: Perceived ease of use (PEOU) questions adapted from Moody [2003][217]

	Question
Q5	I found the procedure for applying the method complex and difficult to follow.
Q6	I found the method easy to learn.
Q8	I found the rules of the method clear and easy to understand.

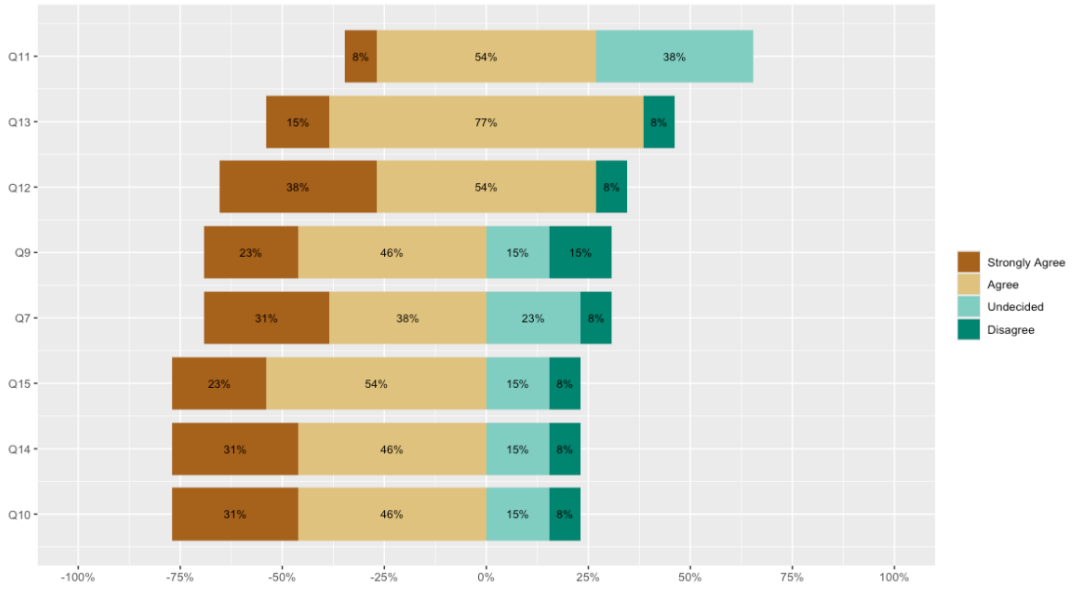

FIGURE 5.4: Perceived usefulness scores by question and participant's business area on a Likert scale where 1 is for *strongly agree* and 5 for *strongly disagree* (n=13)

5.4.1.1 Understanding

Understanding may help to clarify, or negotiate meaning [119]. In this case, clarity was the most important aspect. An overview of the consent process with clear steps was a key factor in helping participants better understand (n=16). (P2) stated, "*It gives a very dense overview of the flow. And you see how it flows on one page, you see the whole picture, you get the very quick overview of what is going on.*" It also helped to negotiate meaning because the value of a Tellu contact in case consent was denied was raised by a member of the Sales team who said, "*I would say don't contact us, we are already too busy with our customers*" (P8).

5.4.1.2 Problem-solving

Another modeling goal is problem-solving and decision making [119], and 4 participants (30%) thought that it was useful for identifying issues. Especially compared to a textual description, participants in different teams felt that it would help in their day to day work in different ways. The CGRCO stated, "*The problem of maintenance or understanding is even more clear here, because now you see the conflicts and the flows, and on the other side what compliance it is triggered by. Because if you only have consent from my patient, not next of kin, there may be consent but it's not formalized. So you may have conflicts that were not documented before. It helps show the problems. Now the maturity in the Norwegian market is low so they try to hide the problem.*" (P7) From DevOps, P1 felt a consent BPMN would be clearer to identify issues and their role in building the process as a developer, "*With requirements written in a paragraph sometimes it is hard*

to follow, implement, and to know what is missing. I would be able to see the process and know what to look for. Like if there is more than one NoK, or that the period is 30 days." In the Product team, as user experience (UX) designer felt that it was useful to see the overview of the process and possible conflicts, "*When you're signing up users to the service, you can track where they stopped, where the conflicts are, how they are bridged, when to provide that information, when to ask for you, agree, do you not agree. I would take this and convert it into a text in a linear story, and then use it. So this is a good resource for me as an interaction designer*." (P11). Interestingly, Sales participants did not share any specific examples of usefulness for their day to day work.

GDPR Validation One of the important aspects of problem-solving for business models in the EU is helping to decide if the process is legally compliant or not. 5 participants (38%) agreed that consent BPMNs would help with validating GDPR requirements, which aligns with one of the modeling goals, problem-solving [119]. However, 8 (62%) declined to answer with a definitive "yes" or "no" due to a lack of familiarity with the GDPR or with BPMN process design to feel comfortable commenting on the usefulness, but offered feedback nevertheless. The most useful feature was giving a process overview with key steps, requirements, and visual traceability (n=12). For example, "*It seems like it's a very efficient and proven way to collect the consent and ensure that the they are given before we can start using a service. All the steps are there and available to be documented. So I guess it will ease the process of verifying and getting the requirements*" (P4). The CGRCO pointed out how this consent process showed GDPR compliant specific consent for different parties, "*Dividing the current [broad] consent into singular consent items, it's very clear in this picture that the patient and next of kin have to think about specific consent as well. And having this interactive process that ends up in documentation, now we're talking about an overview for whom: for next of kin what consent I have given; or for us, if we have a valid consent to process the data. So in terms of [our company], this will definitely improve our digital compliance for next of kin*" (P7), in addition to the improved traceability and gap analysis mentioned in the previous section.

5.4.1.3 Communication

Participants explained that consent BPMNs would be useful for communication (n=15), explaining in coded segments that the visual aid (n=6) and more structured documentation (n=2) helped. P13 said, "*I think the idea of visualization is good because you know it, it kind of condenses the data into boxes. We tend to remember and process the information in pictures*." Additionally, a Product manager stated that it would help to justify and explain design choices, "*Because it's quite difficult to explain why it should be that step and why you need it*" (P10).

However, it may need to be tailored to different audiences and be accompanied by supporting materials. Since the BPMN shown was part of an internal, iterative process for developing better systems, to be shown to external stakeholders the diagram should be a finalized and simplified version to be shown. P8 stated, "*Using this chart would actually be good, but I think if we can cut some of that out so it's just the necessary one, it'd be good. Because when you bring up more possibilities, we may have this situation where it creates uncertainty and more questions, and that we're not able to answer those questions makes it more complicated. So you need to lock it down, and whatever you're presenting, do it in a way that reassures*." This includes more narrative text to accompany the visualization, like, "*It might need a wording description also in combination with this graphic presentation*" (P12).

The full BPMN was perceived to be the most useful for those who require a high-level overview (designers, decision makers, analysts, customers) and those working with the consent processes (DevOps, CGRCO) (n=3). P1 in DevOps has experience with similar diagrams and thought it might have a learning curve for others: "*Yes, it's helpful for communication if you know what flow diagrams are. If you don't, it might be harder since they are unfamiliar with the format and it takes getting used to*." P12 in Sales sees more use for the contracting municipalities than the end-users, saying, "*Maybe not to patients or people like that, but for people working in the municipalities it could be easier when you have a drawing there*."

5.4.1.4 Professional Resources Needed

While the the process was well received, some interesting caveats remained. First is investment in training (n=2), "textitIt requires skill, and it might be more than documentation because it requires more when you are diagramming to be clear on the process. You need someone who can design these well" (P1). This corresponds to the issue of proper model decomposition (n=13). The CGRCO, who must works on governance for the whole company would prefer an even higher level overview of all not to the consents, "*This one you take only video data, so it's a single one on one consent. But when we have one service and several consents in one, then we have a better overview of GDPR compliance. It needs mapping between the service and this single consent*" (P7). On the other hand, a director of business development and someone who often liaises with the municipalities (the customers) said, "*I support the the model and I think it's an effective solution. But I think maybe we're missing missing the the municipalities for example, but that's more into the the actual process and who's having the dialogue*" (P5). On the other hand, for individual usage, many would prefer a scoped down version for their roles and responsibilities. For example, "*I would ask my Data Protection Officer (DPO) too scope this down to what I'm building, and then of course the DPO needs to see if all the parts are sitting well together or is there conflict*" (P11). Then it would need to be updated and maintained as with any documentation, for example, "*One concern is you you need to make sure that you maintain it as well. Like when the app is changing. But that's that's always the thing with documentation. So at least visualizing it sort of helps because you see quite quickly what's there and not, so it should make it easier to to verify if it's still correct or not*." (P9). Lastly, a participant suggested using legal terminology to decrease the miscommunication of terms, "*Instead of using your own wordings, because then it's difficult because maintenance could be totally different between me and Jim and so on. But if you have a legislation and you can go back and look to what was the meaning of this word, then it's easier if you now use the wording and the meaning of the word*" (P12) which may point to efforts to standardization of definitions across diverse teams.

5.4.1.5 Relationship with Consent

P13 raised interesting issues of the current status of consent in the health space, which is tied to necessary care and can be burdensome on the individual. They share a real world example of a nursing home that is using blanket consent, "*When you accept to be part of this nursing home, you accept everything. End of story*." But they are trying to move towards individual and specific consent based on the needs of the patient, but not yet implemented. P7 also shared that, "*Now the maturity in the Norwegian market is low so they try to hide the problem [of consent].*" Instead of advocating for more specific consent, P13 argues is, "*It's not consent it's just ticking the legal requirement box*," they suggest, "*Some kind of a governing structure needs to*

make the consent systems or verify that they are OK. Like we are ISO 2701 certified, so it tells everyone the quality, and I don't need any detail. I that's why I'm saying that this is trying to make something very complex, simpler for someone that should never be involved in it." (P13)

5.4.2 Perceived Ease of Use

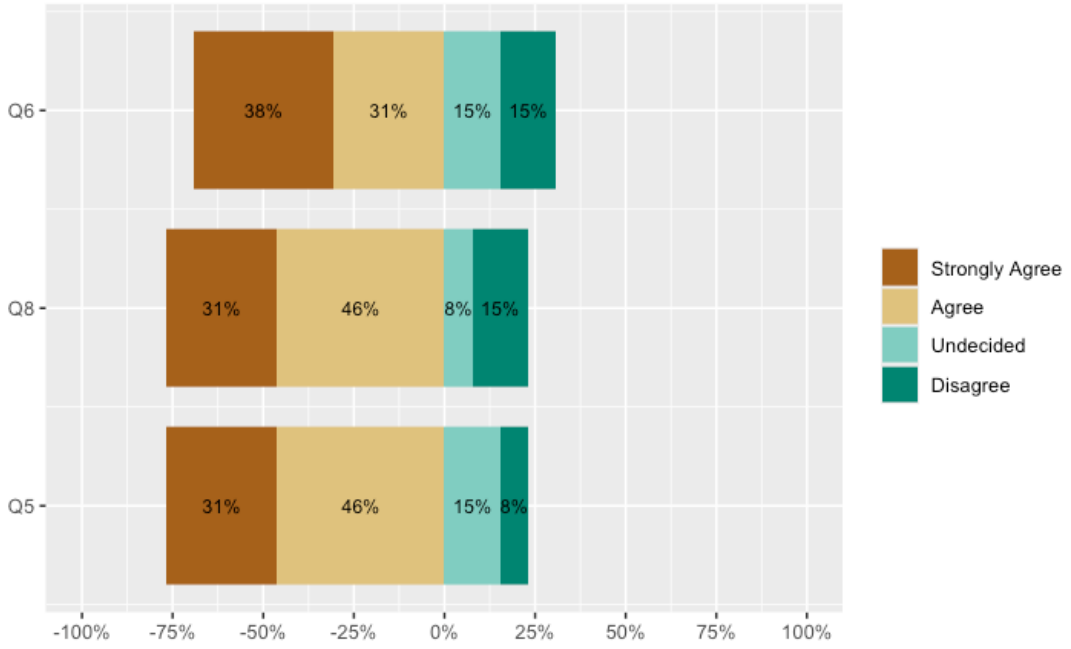

FIGURE 5.5: Perceived ease of use scores by question and participant's business area on a Likert scale where 1 is for *strongly agree* and 5 for *strongly disagree* (n=13)

The PEOU questions had a median of 2 as well, the agreeing with the usability of the method. Again, no one strongly disagreed with the method and rated it 5 (Figure 5.5). The PEOU showed slightly more homogenous variation between teams, with the Devops, Sales, GRC present in rating the method 1 (strongly agree), and Product and Sales rating the method 4 (disagree). While we did not specifically ask about the PEOU, participant answers revealed coded segments the usability. One participant in DevOps had experience using BPMNs in a previous company and shared that, "*Before I was spending quite some time using UML and then BPMNs. My experience is that teaching this activity diagram compared to the other types of diagrams like the UML, it was the easiest to use and to communicate, maybe together with use case diagrams, to communicate with different stakeholders.*"

5.5 Discussion

From both the questionnaire and semistructured interviews, participants perceived consent BPMNs to be a useful, easy to use, and promising tool for mapping and overview, analyzing, GDPR validation, understanding, explaining, and communicating complex consent processes. The authors note that the method appeared to be efficient, as the study participants did not need much time before they started to reflect on consent BPMNs or ask clarifying questions. This bodes well for technology acceptance, as both PU and PEOU were high according to the Likert scale averages

(2, or Agree) and the coded segments from the interviews where the participants qualified their reasoning. The questionnaire showed a variation in agreement between the teams, which may have some relationship with years of experience and familiarity with the legal requirements and challenges. The participants who *"disagreed"* with the method generally had fewer years of experience in the Company, while those with more years of experience or who often worked on consent issues in Management *"strongly agreed"*. This variation would be interesting to explore in more detail. In general, the participants agreed that BPMNs increased traceability and problem-solving in complex consent processes and could improve communication with high-level stakeholders. PU is a strong indicator of uptake [182, 136], and this can be used to improve future iterations. Processes should use legal terms or have clear definitions to reduce miscommunication across diverse teams. As identified previously, supporting documentation can address the lack of business rules or requirements and enhance BPMNs [266]. Lastly, training for proper decomposition into subprocesses was key to addressing employee needs. Some additional technical measures may help [284, 168, 197], or smaller fragments or sub-diagrams [261, 262, 364] could be useful next steps. Overall, the research is promising for consent BPMNs from the employee perspective, but it should be tested to understand how it scales for more consent processes and stakeholders in SMEs.

Recommendations for practitioners Participants mentioned their need for an expert to lead and translate consent processes into concrete checklists or approved actions from the CGRCO. This can be complex, as laws are purposefully vague and the interpretation of GDPR requirements can vary by jurisdiction, by trends in DPAs, and by the organization's capacity to reasonably comply. This can create a heavy burden on Management, who must evaluate, translate, and contextualize legal regulations or guidelines into practice. In the case of this SME, they have two qualified employees in privacy and security (Head of DevOps and CGRCO). More employees could be given the responsibility, especially in more junior roles; or privacy and security champions could be appointed to different teams to promote privacy and security issues. [319, 208] If there is a lack of expertise, then training time would have to be considered. This is also true for BPMNs and modeling; if it were to be implemented more widely, there would be a significant investment in training, adapting existing systems, or building new systems.

Reflections on the Case Study Before the case study, the Company acknowledged the importance of consent management and its value to the customers, however, the work was not prioritized due to the topic complexity, stakeholder dependencies, and market maturity. Clear responsibilities are a prerequisite for value creation in such situations. The majority of the company employees were aware of what consent is and its relevance for personal privacy and the Company, but may have misconceptions or confusion; relying heavily on the CGRCO. During the study, while some of the participants had limited consent experience in the Company, their personal experience and reflections were valuable for the study.

The larger context of consent's role was also brought up, both in relation to the current market and the role of governance. P13 shared an example of a nursing home that currently uses broad consent, but are trying to move towards individual and specific consent. However, the actual implementation can be slow. Author 2, the CGRCO, agrees that more granular consents and national consent governance are the way toward better privacy and control.

The study reveals that consent is complex and difficult to deal with, while it is often presented as a simple "checkbox". This leads to a gap between the regulations, expectations, and implementation, which results in confusion and low motivation: "*Some kind of governing structure needs to make consent systems or verify that they are OK. I don't need any detail, I that should never be involved in consent*." (P13) This is important to consider for more human-centered processes, which may mean moving away from improving business processes and instead trying to enact change from a higher level, such as collaborating with regulatory bodies.

5.6 Limitations and Future Work

Limitations include the small convenience sample and scope of the study, which only asked for self-reported perceptions and can only indicate actual acceptance and usage. While the TAM questions were validated by [217], the method itself may not accurately predict acceptance, and using a TAM extension might be more predictive of uptake. TAM relies on individual perceptions and ideas about their usefulness for their performance, which does not include the leaderships' strategy to decide strategies, goals, and methods. We also may not have reached theoretical saturation due to resource constraints, and the sample may be biased due to the convenience sample, which is skewed towards men and those with many years of experience which may indicate selection bias. Additionally, there is a selection bias due to those who accepted the interviews, they may have specific opinions or knowledge they want to share. The BPMN use-case was based on a best case scenario with working digital integrations, and considering the current widespread use of paper consents in Norway, the participant perceptions may be different than a realistic implementation of consent BPMNs. Future work can address such issues by extending this pilot study and continuing to test different iterations of consent BPMNs through implementation, sampling more employees in different companies, and moving from a small convenience sample to large random sampling or targeting specific demographics. This can make future work more generalizable to other types of companies outside Norway. For the scope of this paper, it was not possible to extend the study or sample sizes as the Author visited the company and interviewed most employees in person. It would also be interesting to test one of the decomposition methods from research compared to manual decomposition efforts for a SME, since its an ongoing challenge.

5.7 Conclusion

BPMNs are a powerful, flexible, user-friendly language to model complex business processes such as consent for sensitive data for SMEs. Participants felt it aided understanding, problem-solving and GDPR validation, and communication, and was usable. However, the flexibility of the BPMNs may make it difficult to implement consistently, reflecting previous research. Although it may be tempting to design the model to appeal to as many stakeholders as possible, it may be more prudent to focus on specific goals and stakeholders, e.g. the CGRCO for consent. In theory, a BPMN could cover a business process completely, but some may think it is too complex, while others think it is too simple. It is also important to weigh the pros and cons of investing resources into a new modeling system, especially for SMEs who may not have the bandwidth to assign modeling tasks or many automated processes that integrate easily with BPMNs. In addition, the organizational drive

to implement new technologies should not be ignored. An employee or team may independently use BPMNs, but it may create silos, replicated work, or incompatibilities with other documentation. It would be the most beneficial to have a BPMN Governance type system with executive sponsorship [349] to make sure people have the resources to make decisions, set goals, and set aside time.

5.8 Appendix

Data are available on reasonable request.

5.8.1 BPMN Symbols

BPMN notation guide available here: https://camunda.com/bpmn/reference/

Chapter 6

User consent goals, management, and preferences for mediums

This chapter adapted from: Doan, Xengie Cheng, Annika Selzer, Arianna Rossi, Wilhelmina Maria Botes, and Gabriele Lenzini. "Conciseness, interest, and unexpectedness: User attitudes towards infographic and comic consent mediums." In Web Conference Companion Volume (ACM). ACM. 2022.

and

Doan, Xengie, Arianna Rossi, Marietjie Botes, and Annika Selzer. "Comparing Attitudes Toward Different Consent Mediums: Semistructured Qualitative Study." JMIR Human Factors 11 (2024): e53113.

Author Contributions: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Investigation, Writing - original draft, Writing, review & editing

Abstract: As consent for data sharing evolves with the digital age, plain-text consent is not the only format in which information can be presented. However, designing a good consent form is highly challenging. The addition of graphics, video, and other mediums to use can vary widely in effectiveness; and improper use can be detrimental to users. This study aims to explore the expectations and experiences of adults toward consent given in infographic, video, text, newsletter, and comic forms in a health data sharing scenario to better understand the appropriateness of different mediums and identify elements of each medium that most affect engagement with the content. We set a data sharing scenario with a data trustee, and qualitatively investigated participants' expectations and opinions toward consent. We designed mock consent forms in infographic, video, text, newsletter, and comic versions. Semistructured interviews were conducted with adults who were interviewed about their expectations for consent and were then shown each consent medium and asked about engaging elements across mediums, preferences for consent mediums, and the value of document quality criteria. We transcribed and qualitatively co-coded to identify themes and perform analyses. We interviewed 24 users and identified different thematic archetypes based on participant goals, such as the Trust Seeker, who considered their own understanding and trust in organizations when making decisions. The infographic was ranked first for enhancing understanding, prioritizing information, and maintaining the proper audience fit for serious consent in health data sharing scenarios. In addition, specific elements such as structure, step-by-step organization, and readability were preferred engaging elements. We identified archetypes to better understand user needs and document elements that can be targeted to enhance user engagement with consent forms; this can help inform the design of more effective consent in the future.

6.1 Introduction

Consent takes on a pivotal role throughout EU regulations such as the General Data Protection Regulation (GDPR), but there are still challenges in digital consent and management. Digital decision-making about one's own data can be influenced or misled through online design choices (i.e., through so-called dark patterns [113, 22, 283, 153]), while the consent experience of most European users corresponds to nagging cookie consent requests to profiling and advertisement that induce consent fatigue while trying to access a needed service [229]. Decades of research in the biomedical domain show that study participants' consent can rarely be deemed actually informed [216], often due to the complexity of language [224] and lack of

health literacy [203], as well as to the lack of data literacy of the individuals [66]. Among other uses, health data can point to suitable candidates for clinical trials and contribute to scientific advancements, bringing benefits to public health (e.g., identifying hot spots of disease outbreaks to implement counter-measures) and to individuals (e.g., taking control of disease management and symptoms). However, health data is a special category of personal data that is protected by strict data processing rules, because information about an individual's health status and symptoms may be misused, resulting in discrimination and other harms.

Health data can be used to identify suitable candidates for clinical trials. The processing of such data holds great advantages for society (e.g., curing diseases) and individuals (directly or indirectly) (e.g., being able to participate in a clinical trial and possibly be cured of a disease). However, health data is considered a special category of personal data because the unauthorized disclosure of information about the health status or disease symptoms of an individual may result in discrimination and damage to the reputation of that person, hence the strict rules that are applicable for the processing of such data.

The free flow of high-quality data is but one element of the European single market that can only be realized by establishing formal mechanisms of trustworthy data governance, today proposed in the DGA proposal [337]. In this respect, data trustee models have increasingly been discussed, designed, or implemented [128, 169, 42, 295, 294] to act as an independent party between those who provide data and those who process that data. Data trustees offer independent data stewardship and are subject to the legal responsibility of guaranteeing that data sharing and use occur to the benefit of a specific group of people and organizations [128], as opposed to companies that unilaterally control the use and disclosure of people's personal information for their own exclusive benefit.

In a health research scenario, data trustees can assist in finding suitable participants for clinical trials in a privacy-friendly manner: individuals can transfer their data to a data trustee in exchange for various benefits (e.g., financial compensation, services), that passes them on to organizations that intend to carry out clinical trials (hereinafter "service providers"). In this and other cases, engaging individuals in a user-friendly consent experience is fundamental to enable them to meaningfully and freely signify their agreement or disagreement with a sense of satisfaction [95]. Readability and comprehensibility of consent notice are necessary but insufficient measures to determine whether consent is asked in a transparent manner that complies with legal obligations and ethical safeguards. Considering that the everyday Europeans' experience with consenting is the web cookie consent, often accused of being opaque and manipulative [22, 283, 153, 341, 200, 112, 207, 229], it becomes clear how consenting to data sharing, especially when it comes to sensitive data, must be designed in a different manner to enable a trustworthy data-informed economy.

Engaging individuals in a user-friendly consent experience is thus fundamental to enable them to meaningfully and freely make decisions with a sense of satisfaction [95] and agency. Improving the readability and comprehensibility of consent notices is one aspect of this, but research is also being done to explore visual communication techniques. Current research often focuses on the effect of multimedia on understanding [167, 236], which can have a varied effect based on different studies. Multimedia also spans many formats, and most studies reviewed for their effect on understanding compared 2-3 different formats [236]. The Article 29 WP also refers to visual design means, such as "cartoons, infographics, flowcharts", to enhance the comprehensibility of information, and specifically to "comics/cartoons, pictograms,

animations“ [243]. However, they do not offer further guidance about what mediums to use and for what purpose (e.g., how one might prioritize skimming while another might be better for complex information). Therefore, we experiment with five different mediums of consent in one study, building on studies researching the use of a comic [312, 38], video [13], infographic and illustrated text [354], and plain text as a control [303].

The objective of this study was to better profile user expectations and their attitudes towards different consent mediums, which included an infographic, video, text, newsletter, and comic. We specifically analyzed how different elements of consent mediums (e.g., narrative, color, audio) affected participant engagement to survey the different affordances of each medium. Each medium has their own strengths and weaknesses in representing various kinds of information and can achieve various informational goals (e.g., the video is low effort but not skimmable, while the text is skimmable but boring) [273]. Since which mediums to use and how participants would prefer different mediums, we compare multiple mediums in one study based on semi-structured interviews and deep-dive into participant motivations, expectations, and experiences. We also focused on the highest and lowest ranked mediums – comic and infographic and examined how various document design elements, such as graphical elements and the relationship between the writer and the reader, affected the engagement and experience of study participants.

The results hint at diverse goals across the participants; the elements of document design to make the information concise, structured, and appropriate for the audience; and the large influence of context on participant perception and expectations. Overall, the findings have implications for how to better design consent documents to address different general participant profiles using layering and to more effectively engage the audience with a suitable medium, especially in the digital health data sharing space to give more effective transparency to participants who are deciding whether to share sensitive data. The findings also show that individuals have formed expectations about how consent should look like and that the consent medium and its tone of voice should fit the target audience and the context, thus comics should be used cautiously. Infographics seem to be a better fit for biomedical contexts and it additionally allows strategic reading and enables understanding. Both mediums raise interest and attention because of their unconventionality, with a possible influence on user engagement. Considering all these elements may contribute to extending the conceptualization of user-centered transparency beyond text readability, graphical aids, and user interface elements. These insights can inform the design of consent requests to adequately engage different user groups in thoughtful decision-making.

6.2 Research Scenario

6.2.1 Use Case: Consent to Data Transfer

A clinical trial usually has very specific requirements for its participants, e.g., age, disease type, etc. Therefore, finding suitable candidates can be difficult and usually includes processing great quantities of possible participants' personal data before finding the few that meet the criteria. A data trustee can minimize the amount of personal data service providers can see, with a service provider applying to get only the personal data of individuals that fit the criteria of a proposed clinical trial. The data trustee contacts the relevant individuals, explains the reason, and asks them

if they consent to transfer their information to the service provider.[1] In case the individual gives its consent, the data trustee transfers its personal data to the service provider. Our use case focuses on the consent asked of individuals to transfer their personal data to a specific service provider. The consent a service provider would need from individuals to allow their participation in a clinical trial is not part of our use case.

6.3 Related Work

6.3.1 Informed consent and transparency requirements under the GDPR

The GDPR defines consent as any freely given, specific, informed, and unambiguous indication of the individual's wishes by which they signify agreement to the processing of their personal data (Art. 4(11)). Consent to the processing of sensitive information such as health data needs to also be explicit (Art. 9) [336].

Consent must also be intelligible for the average person (i.e., free from jargon and concise), expressed in clear and plain language (i.e., straightforward and concrete statements), and accessible (Art. 4(11) and 7(2)) [336, 245]. Audience fit figures among the user-centered requirements for consent to be informed: the logic of what information should be presented and how must derive from the identification of the audience needs (e.g., minors vs adults), also based on empirical studies [245]. Organizations have latitude as to how to present information to data subjects in the form of "written or oral statements, or audio or video messages", which can also be layered to respect the two-fold obligation of being concise and complete at once [245]. The organizations that are responsible for GDPR compliance shall also make the withdrawal of consent as easy as its provision (Art. 7(3)), otherwise, consent may be considered invalid [245].

However most existing informed decision-making solutions fail to reconcile theoretical demands with actual transparency. Conventional data privacy communication is characterized by lengthy, off-putting walls of complex jargon that impacts the readability, comprehensibility, navigability, and memorability of information [273]. In addition, it is often standard, vague, or boilerplate instead of customized to the different needs and abilities of the intended audiences [252] and the type of data and processing activity. Reaching beyond plain language, in the last years, there is a renewed attention (and quite some experimentation) towards document design criteria [352] that more holistically relate to language, writer-reader relationship, information design, and content.

6.3.2 Consent comprehensibility

Informed consent has traditionally been implemented to protect ethical-legal rights of participants, consisting or written forms collected for administrative purposes in medicine, where consent has been shown to focus on the needs of the researchers and institutions [367], so user-centered research is less common [162]. Similarly to the case of cookie consents: numerous studies have demonstrated that cookie banner UIs are often designed to extort users' agreement in a manipulative manner, thereby circumventing the tenets of the law described earlier [22, 283, 153, 341, 200, 112, 207, 229]. Whether informing online users or patients, it is difficult and time-consuming to establish an adequate level of being informed with respect to informed consent

[1] For further details on the data trustee model we based our interviews on, please refer to [295].

obligations. In clinical trials settings, a systematic review of 30 studies about informed consent found that when participants were informed about the clinical trial aims, risks, benefits, and more, only about half the content was understood [94].

This hints to the difficulty of finding an effective way to describe a research study in general and the relevant risks and benefits to achieve truly informed consent [186]. In addition, informed consent for health purposes often demands specialized data literacy to truly understand the statistics about health risks and benefits. Not only patients, but even doctors are susceptible to misinterpreting risk figures [111]. Thus health data should transparently and clearly state the relevance to an individual instead of using common statistics or terms.

Language	**How easy it is for people to understand the words**	**Design**	**The visual impact of the document and the way its design influences usability**
Directness	Using direct language to make clear who's doing what.	Legibility	Use of legible fonts and text layout.
Plain words	Extent to which the vocabulary is easily understood.	Graphic el.	Use of tables, bullet lists, graphs, charts, diagrams, etc.
Grammar	Conformity with the practice of good standard English.	Structure	Quality of the document's organization in relation to its. function.
Readability	Ease with which the reader can follow the argument of the text.	Impression	Attractiveness and approachability of the document's overall. appearance.
Relationship	**How far the document establishes a relationship with its users.**	**Content**	**How the content and the way it is organised deliver the document's purpose.**
Who from	Is it clear who is communicating?	Relevance	How relevant the content is to the recipient.
Contact	Whether there are clear contact points and means of contact.	Subject	Whether it is clear what the communication is about.
Audience fit	Appropriateness to the knowledge and skills of the users.	Action	Clarity about what action is required of the user.
Tone	Matching the style and language to the context.	Alignment	Compliance with the organization's intended aims and values.

TABLE 6.1: Document quality criteria elaborated by R. Waller [352]

Legal communication increasingly deviates from conventional lengthy, off-putting walls of legalese and makes use of information design elements [247] meant to enhance the readability, comprehensibility, navigability, and memorability of information [273], based on the different needs and abilities of the intended audiences [252]. The affordances offered by a document go beyond mere plain language criteria to embrace a whole set of best practices against which documents' benchmarking can be carried out [352] concerning language, writer-reader relationship, information design, and content (Table 6.1). When it comes to sensitive data sharing, the statutory requirement of transparency about data processing practices similarly applies to the information notices and consent requests that describe and ask user permission about such practices ("transparency by design") [274]. Asynchronous communication, in addition, entails risks of misunderstandings, as a professional is not present to clarify doubts, and risks of mindless consenting to data sharing which has been shown in other digital consent experiences (e.g., cookies)[156]. However, digital communication also offers new opportunities, i.e., experimenting with various media (i.e., e-mail, messages, webpages, videos, chatbots, etc.), interaction modes, scalability, [327] and timing [274].

6.3.3 Profiling with archetypes

Human-computer interaction research has used the persona technique (wherein potential users are given different profiles or personas with different goals and personalities based on demographic data) to better understand different users and needs and design fitting systems [256], but it is a lengthy process that is often used for IT systems, not the consent process. Personas or needs assessments have been conducted in relation to different demographics in health studies, but rather than focus on the informed consent aspect they focus on the health symptoms and how to address specific health-related needs [281]. One study used personas for patients with dementia for participatory design of a project to improve quality of life [134], which offers insights into how to design for different participant dementia related needs. Another study that began with focus groups to design multimedia consent focused on health-related needs, such as the mental state and capacity of patients with schizophrenia, breast cancer, or depression [145]. Another study looked at the

decisional needs for patients learning disabilities, such as more time to make decisions and non-verbal communication techniques [47]. Beyond health-related needs, we are interested more broadly in how the general adult population would interact with health consent and what elements would stand out and be engaging when making informed decisions. This aspect has not been studied, to the best of our knowledge, but would be important for understanding how to strategically create information disclosures for different goals. Thus we wanted to explore archetypes, which capture general profiles, instead of personas, which are a representation of imaginary individuals with specific population characteristics.

6.3.4 Multimedia tools for IC

The digitalization of data collection and use authorization allows for multimedia tools to be employed during the consent process and it can have a positive outcome on participants. Overall, a systematic review of multimedia consent for surgical procedures found increased patient satisfaction for usability and informational availability with multimedia consent, which included videos, interactive programs, and more [223]. However, for clinical trial consent, videos did not improve understanding [135]. Diving into the reasons that multimedia consent may be preferred to conventional text, one study found that comparing animated videos, slideshows with voice-over, comics, and text consent for medical practices found that a dual-channel approach combing audio with visuals helped participant understanding [167]. This study supported older research that repetition of information using different multimedia means increases retention [209], however, the specific elements of videos, comics, and text that contributed to effective communication in more general health consent were not studied - a gap that we intend to bridge with our work.

The use of comics was especially interesting as it combines textual and graphical means, uses a conversational style, and develops a narrative in a specific context allowing readers to identify with the depicted characters. Comics can also attract and retain attention by fighting notice fatigue [274], i.e. the habituation and alienation derived by the longstanding habit of experiencing inscrutable prose, and they have been used in contracts [121] and privacy policies [273]. To successfully bridge language barriers between scientists and indigenous populations in South Africa, in our previous work we created, tested, and refined a comic to ask consent for participation in a genomic research project [312, 38]. The study revealed that the population had specific expectations on how they wanted to be depicted in the comic to counter the exclusion and discrimination that happened in the past. The comic also increased general understanding of the research process and strengthened the ability to make fully informed consent decisions.

Other mediums to enhance consent engagement have also been studied in different contexts. Wang et al. [354] conducted a study comparing the use of infographics, comics, and illustrated text to communicate data-heavy information, such as graphics about renewable energy in Europe, finding that young academic students from different countries (aged 18-35) preferred comics, with the greatest understanding, engagement, and enjoyment of all mediums, while infographics performed best in aesthetics and exploration and were second to comics in the other dimensions.

6.3.5 Engagement with IC

Even in other domains, studies strive to understand how to achieve effective communication of complex information by analyzing participant engagement, understanding, and recall of the information. In this study engagement refers to time spent and fun reading a form and infographics, illustrated text, and data comics of complex economic data were tested [354]. They found that students from different countries (aged 18-35) preferred data comics due to enabling the greatest understanding, engagement, and enjoyment of all mediums, while the infographic performed best in aesthetics and exploration, and the illustrated text performed the worst. Since similar studies have not been performed on consent forms in a health scenario, we seek to study engagement as a factor of effective communication, as it may help understand what gains and retains attention within a complex digital attention economy.

Traditionally, engagement studies in biomedical consent refer to patient engagement with the research or biomedical process. Such engagement refers to participants interacting with the results of a study, updating information, or changing consent [62] [142] [123]. However, we are interested in participant motivations to consume the information in a consent form and give their initial and continued attention to a conventionally tedious process, taking inspiration from research about the attention economy [302, 58]. Can consent forms be interesting and attention grabbing? Other studies that mention patient engagement for health consent also take the latter approach where patient engagement refers to engagement with the research process in terms of results, updated information, or changing consent [62] [142].

6.3.6 Emotions in the consent process

Enjoyment and emotions are an integral part of human-computer interactions [41] and can offer rich insights in conjunction with usability studies [6]. In addition, other fields such as marketing also incorporate emotions to influence users [61][325]. Concerning consent, a study used emotions derived from Plutchik's emotion wheel [251] and reported that over 50% of users felt annoyed and indifferent from cookie consents [115], which may influence individual attitudes about consent in general. Another study investigated the influence of emotions on information processing and decision making in a clinical trial informed consent and found that fear significantly increases the average time spent reading the procedures and the benefits of participation [95]. Thus, emotions seem to impact engagement with consent processes and should be investigated.

6.4 Research Questions

The previous section has gathered evidence about the interplay between EU regulations for consent, transparent privacy and consent information, and and the use of archetypes and multimedia tools to enhance the experience. However, we lack understanding of user engagement regarding how different multimedia mediums affect participant consent experiences. Therefore, this study sought to answer the following research questions:

This study sought to answer the following research questions:

1. What are participants' general experiences with informed consent processes?

2. What are participants' expectations prior to exposure to different consent mediums?
3. What kind of goal-oriented archetypes can be created to better understand participant needs for consent?
4. What are participant preferences after exposure to the various consent mediums with respect to:
   (a) What were participants' rankings of consent mediums?
   (b) What elements reportedly influence their preference for mediums?
   (c) Based on document quality criteria [352], what elements did participants identify for each medium?
   (d) What were participant rankings of different engaging elements?
5. What kind of emotions do the top vs bottom ranked mediums trigger?
6. What kind of consent management (for giving of revoking consent) would participants want?

6.5 Methods

We carried out 24 semi-structured interviews in the autumn of 2021 in Germany. We created an interview guideline (See Appendix 6.11) which was validated prior to the interviews with three potential participants to ensure the clarity, comprehensibility, and precision of the questions.

6.5.1 Participants

We used word-of-mouth to find the participants in Germany, who were all German native speakers (as the interviewer's mother tongue is German). The demographic included adults from a cross-section of the German adult population by age, sex, and education level. The sample size and participant characteristics were based on a systemic review of unbiased citizen juries for health policies [315]. The 24 participants included 8 participants from the age ranges "18-30", "31-55", and "56-90". Within each age range, 4 men and 4 women were interviewed. Within each of the ages and sexes, there were 2 male and 2 female interviewees whose highest degree is a school-leaving certificate or a finished apprenticeship, and each 2 male and female interviewees whose highest degree is from a college or university. All participants were German native speakers and live in Germany. The interviews took an average of 60-75 minutes and the interviewees were offered 30 Euros compensation for their time. In Fig. 6.1 we show the key information of the structured interviews performed.

6.5.2 Study materials

We created an exemplary plain text form that asked consent for the transfer of personal data from a data trustee to a service provider (see Sec. 6.2.1 and Appendix 6.11), which included short sections on "Who are we?", "Which of your data do we process and where did we get it from?", "What happens if you agree?", "What exactly do we ask consent for today?", a section to accept or reject by signature, and a section about withdrawal of data. Author 1 designed 4 additional variations

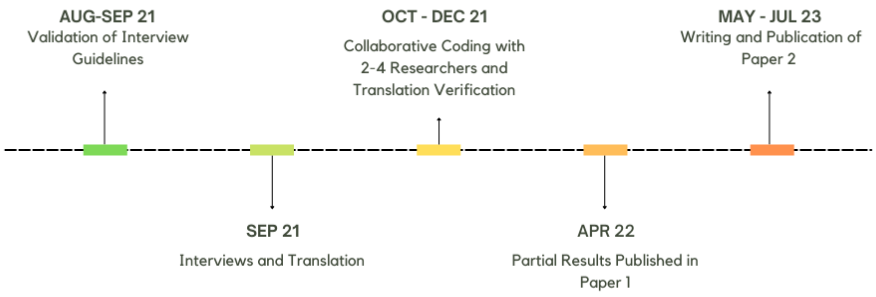

FIGURE 6.1: A timeline of key activities

in different mediums: newsletter, infographic, comic, and video including only the subsection "What happens if you agree?" of the consent form. Each medium had different engaging elements (e.g., numbered lists for a step-by-step format, bullet points for an open format) and were minimally adapted for the mediums (i.e., additional ellipses between comic text) for the purposes of the study. The video was suitable to test engaging elements such as audio and animations, the newsletter an open format, the infographic a step-by-step format, the text readability, and the comic a story element.

6.5.3 Study design

The interviews took place in German via an online video conference system and were documented by a summary transcription written right after each question and finalized right after each interview.

6.5.3.1 Use Case for Interviews

The participants were verbally presented with the fictional use case explained in Sec. 6.2.1 and were invited to imagine that they were a person who is contacted by a data trustee to obtain consent for the transfer of their data to a service provider who wants to carry out a clinical trial. We stressed that the data trustee only asks consent for the data transfer itself, not for participation in the clinical trial. The participants were also asked to read through the full, plain text version of the consent form we created and were invited to clarify any doubt with the researcher. We did so to ensure full understanding of the use case and to provide an example of a consent form that would be expected when giving consent within the use case, as we later on only presented the interviewees with a subsection of the consent form in different

FIGURE 6.2: A translated section of the infographic study material designed with a step-by-step format, color, and structured sections.

mediums since reading through all of the text multiple times would have taken too long.

6.5.3.2 Previous experiences and personal expectations about consent

We then asked participants about their previous experience with consent forms, then we showed them Plutchik's emotion wheel (Appendix Fig. 6.11) to choose one or more emotions at their leisure to describe how they felt during their past consent experiences (Q3, Q4) [251]. We then enquired about their expectations in regard to a consent form that would encourage them to engage with it (Q5, Q6-Q10).

6.5.3.3 Rankings of various consent forms, emotions, and meeting of expectations

After that, we showed them the subsection "What happens if I agree?" in different mediums (i.e., comic, infographic, plain text, newsletter, and video) in a random order. We asked them to rank the different forms according to their preference and clarify why, and whether they met their expectations (Q12). We stressed that we showed only a subsection of a complete consent form. We used Plutchik's emotion wheel (Appendix Fig. 6.11) to explore the interviewees' emotion(s) when shown the

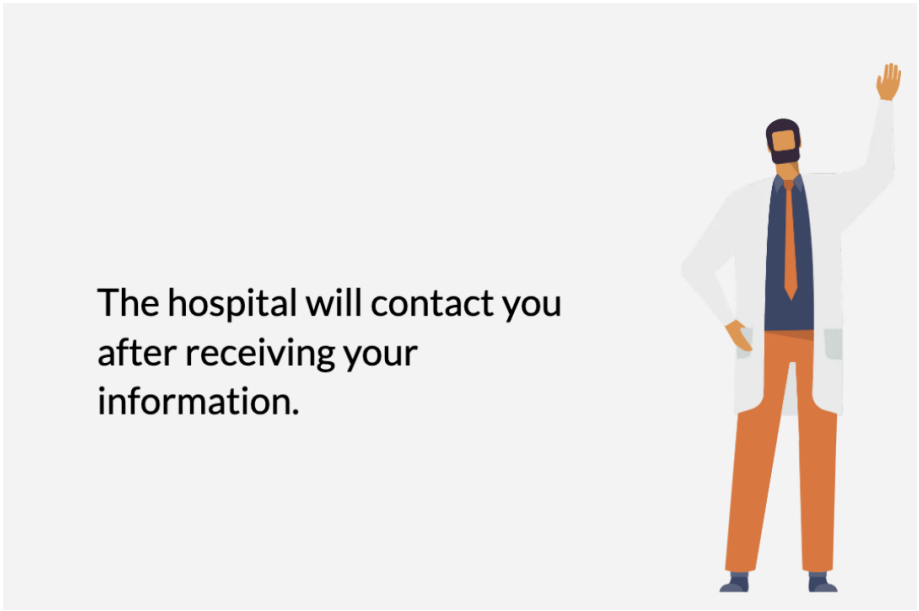

FIGURE 6.3: A translated section of the video study material designed with animation, color, and audio.

various designs. Furthermore, we asked if their expectations about consent engagement were met by the various consent forms (Q13-16).

6.5.4 Data Analysis

As the interviews were documented in German, to collaboratively analyze them with the non-German authors, anonymized answers were translated into English via DeepL https://www.deepl.com/translator and proof-read by Author 2 to ensure the translations' adherence to the original meaning. Such verification continued throughout the qualitative coding process in various sessions in November-December 2021 with the multidisciplinary team with Authors 1, 2, 3, and 4 (with expertise spanning data protection law, usable privacy, bioethics, bioinformatics, legal design), using the software MAXQDA https://www.maxqda.com/. We inductively and iteratively established a codebook over three 2-hour sessions of data labeling. The codebook combines a top-down approach with categories derived from the design, language, content, and relationship criteria for good documents [352] (see Table 6.1) and the Plutchik's emotions along with codes created from a bottom-up approach through analysis of the data (e.g., the concept of trust) (Appendix 6.11). To address the fact that we translated German interviews into English, we ensured that during the coding of the interviews both an English native speaker (Author 1) and a German native speaker (Author 2) were present. Furthermore, we made sure to look at both the English and the German version of the emotion wheel while coding the emotions and discussed any uncertainty.

Archetypes Participant consent expectations have been organized into archetypes depending on the exhibited salience of reported goals and relevant features. A table was created with the participant number, expected features, expected goals, expected behaviors to help group similar profiles.

FIGURE 6.4: A translated section of the comic study material designed with a story, color, and readability.

6.5.5 Ethical and Legal Considerations

The study design was authorized by the Research Ethics Committee at the University of Luxembourg (No. ERP 21-038 LeADS). We chose a summary transcription over a word-by-word-protocol to enable an easier anonymization of the interview documentation later on. Once manually anonymized, the transcripts were securely shared with the authors from the other organization.

6.6 Results

6.6.1 RQ1: Prior Experiences with Consent

Regarding RQ 1, we found that all participants had previous experience with consent before the study, with the majority citing consent in the context of healthcare and/or cookie banners (Q3-4). When asked how long they would engage with consent, 20 participants reported time estimates: more than half the participants (n=11) claimed they spent 1-5 minutes, two spent 30 seconds to one minute, five from 0 to 30 seconds, and two did not spend any time before consenting or rejecting (Q5). However, the time spent on consent also depends on contextual elements. For example, P19 said, "*With cookies, I immediately refuse as much as possible. At the doctor's office,*

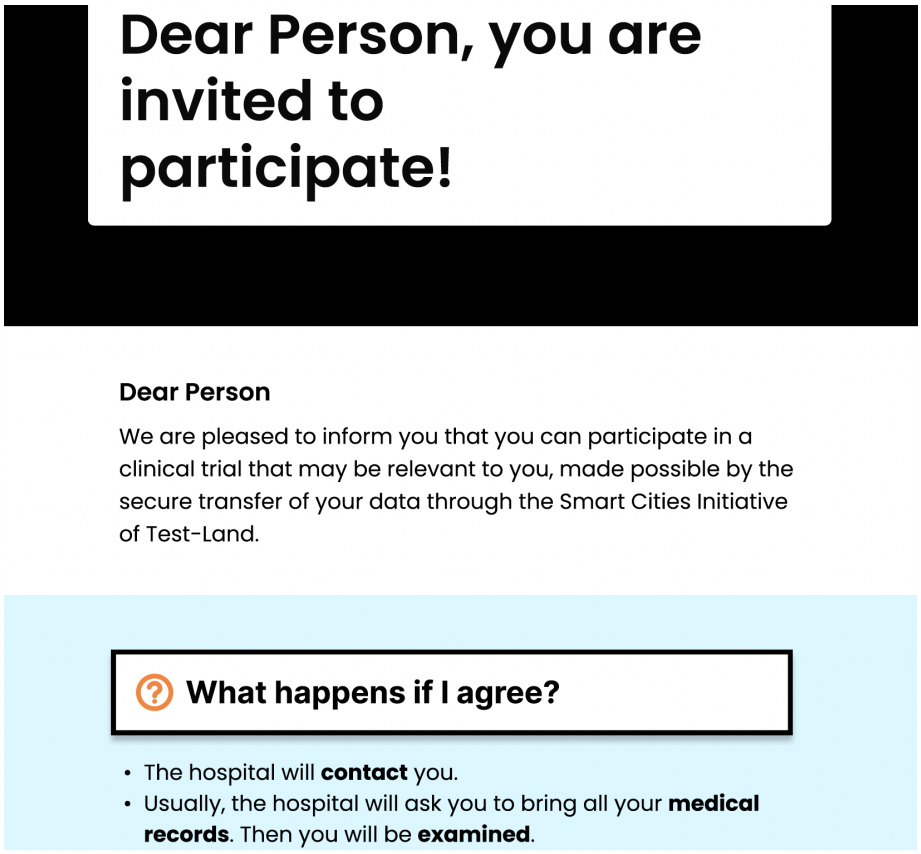

FIGURE 6.5: A translated section of the newsletter study material designed with an open format, color, and structured sections.

for example, I would read through a consent form twice [...] 5 minutes," while P9 said, "*[I]t depends on who asks it, accordingly I read more attentively or not. If it is something more important e.g. about my finances I read with more attention*." Most participants indicated they would spend "as much time as necessary to understand" (n=10), followed by "as little as possible to sign" (n=8), dependent on perceived trust (n=6) with P16 saying, "*At the doctor's office, I take little time because I have a lot of trust there. I skim over these consents briefly, taking maybe 30 seconds. On the Internet, I usually take a closer look*", "as much time needed just to skim" (n=5), and four based on other reasoning.

6.6.2 RQ2: Expectations for Consent

Concerning RQ 2a about users' expectations of the consent process, results yielded 148 segments coded according to the document criteria shown in Table 6.1 (Q6-Q10). Coded segments refer to the extracts of interview text where codes were applied, including overlaps in the text referring to multiple unique and relevant codes.

6.6.2.1 Design criteria

In terms of design criteria (n=36), graphic elements like bullet points, highlighting, and headings were most cited (n=19) as with, "*I would say that words or passages in bold type stick in my memory, so I would find it desirable, especially with long texts, [...] so that the most important information could be filtered out directly at a glance*" (P1). They were followed by structural elements like sections and organization (n=9), and impression (n=7).

6.6.2.2 Language criteria

Regarding language criteria (n=35), interviewees most valued textual directness and conciseness (n=21), for example, P10 said, "*When consent forms are particularly long and complicated, I feel like I'm being misled. In the example scenario, I find consent easy to understand. If I am to give consent in a stress-free way, I expect clear and pictorial language that clarifies what actually happens to the data*". Following that category was plain language (n=8), readability (n=5), and grammar (n=1), although interestingly two participants specifically expected technical terms, such as "*[...] The advantage of a few technical terms is that everything is easier to understand. At the same time, it takes longer to describe these terms in simple language. And that would take too long to read*" (P5).

6.6.2.3 Relationship criteria

As for the relationship criteria (n=18), or how the document establishes a relationship with the reader, the most cited one was audience fit (n=12), which refers to the appropriateness to the knowledge and skills of the users, such as "*From consents I expect that an average citizen can understand them*" (P17). Another relevant category was tone (n=5), which concerns how style and language match the context.

6.6.2.4 Content criteria

Regarding the expectations about the content of the consent form (n=12), participants predominantly mentioned how it is relevant to them (relevance, n=10), for example: "*[A]s an affected person, I would like to see a few examples to get a better understanding of what may be done with my data*" (P1), and what actions they can take e.g., withdrawal (n=2).

6.6.3 RQ3: Archetypes

Based on the previous results about participant expectations, desires, and needs, we organized existing patterns into three goal-oriented archetypes. Not all participants reported specific goals, while some participants reported multiple. Thus, the archetypes are based on grouping similar features (Fig. 6.6).

6.6.3.1 Goal-oriented archetypes

Fully Informed The most common goal explicitly reported by participants was understanding (n=14). This falls into the **Fully Informed** archetype that wanted relevant and fitting information to understand what they were consenting to. For example, "*As an affected person, I would like to see a few examples to get a better understanding of what may be done with my data*" Participant 1 (P1). The information must also be appropriate for them as an audience, saying, "*A simple explanation that everyone understands would be my preference*" (P12).

Record Keepers In addition to most participants wanting to understand, some specifically wanted to remember what they had agreed to (n=3) or to have a copy for their records (n=4). This is the **Record Keepers** archetype. For example, P13 had a clear idea of the elements they wanted to understand and retain a clear memory of, "*It needs to be clear to me what the consent is for, who it is from, and exactly what data is being processed for what purpose.*" Additionally, P4 stated, "*It doesn't matter to me*

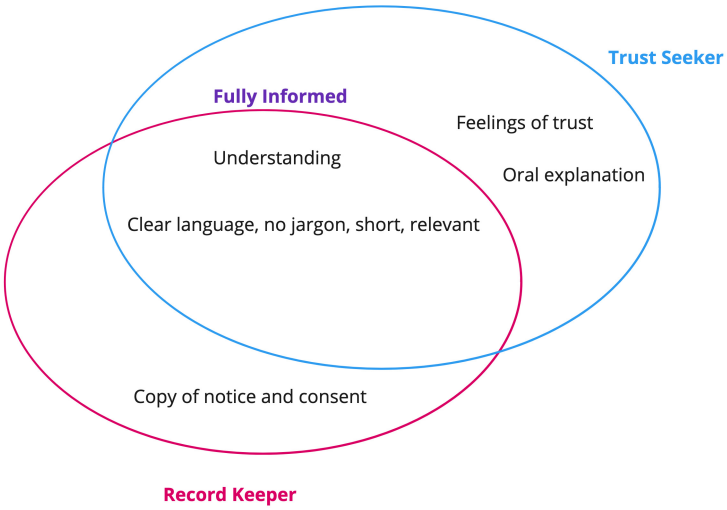

FIGURE 6.6: Venn diagram describing core goals and needs of archetypes

if it is paper or digital. The main thing is that I receive a copy of the text to which I have consented."

Trust seekers **Trust seekers** also seek understanding, but are cautious towards the system or desire a trustworthy system, with P3 saying, "*I must have the impression that the data trustee is a reliable company or that there is an expertise that proves that I can trust this data trustee,*" and P7 stating that they would rather avoid "*to invest time and read through stuff* and they'd rather *be able to trust - since I've already given my data [...] - and that my data will just be handled well.*"

When considered together, the archetypes lie on a spectrum (Fig. 6.7) where the Fully Informed archetype relies more on individual responsibility and capacity to make informed decisions, while Trust Seekers also consider the context of organizational reputation and trust in making their decisions. Additionally, the Record Keepers could be seen as individuals who want to manage their consent decision over time, while those who do not want records accept a one-time decision without copies of the consent.

Outliers In addition to finding patterns based on common goals, some individuals stood out for their unique consent desires, including more jargon (n=2) and wanting an oral explanation (by video or in-person) (n=6). The use of jargon seems to enable more time efficiency in some participants, such as P15 stated, "*If I had to choose between short technical language and simple but longer language that is easy for everyone to understand, I would choose the short technical language.*"

6.6.4 Top engaging elements per medium

The most frequent element ranked first was structure, followed by readability, colors, and step-by-step elements (tied), audio, and story with other (also tied) (see Fig. 6.8). The top element at rank 2 was also structure, and the top element at rank 3 was readability. Not all participants chose to elaborate on what "other" element they referred to, but when they did, a personal engagement (n=4) was most common. In rank 2 and three, structure, readability, and step-by-step were also the top engaging elements.

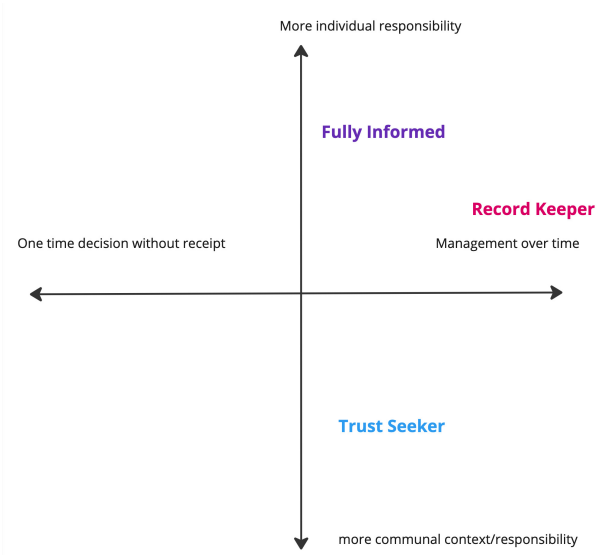

FIGURE 6.7: Goal-oriented archetypes placed on a axis to demonstrate different approaches to consent.

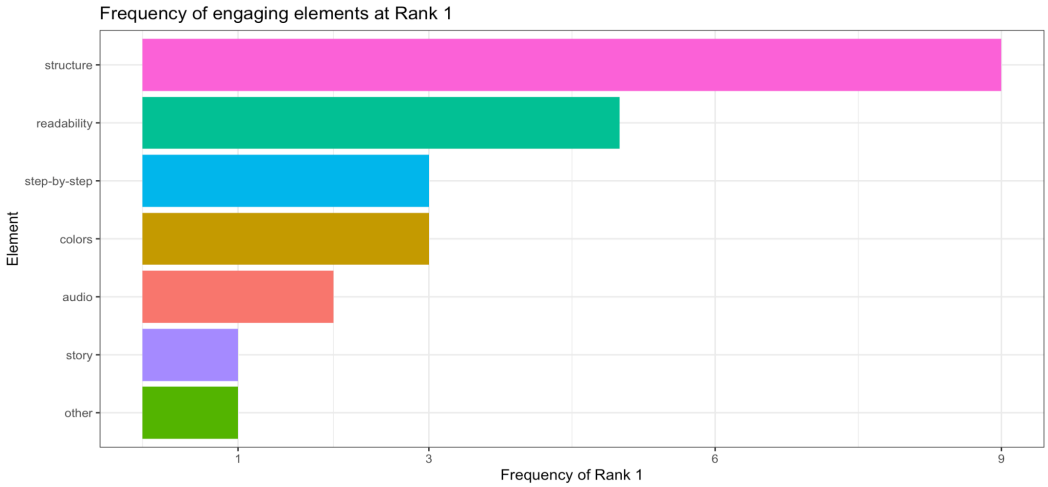

FIGURE 6.8: The frequency of each engaging element ranked first by participants.

6.6.5 RQ5: Medium ranking and document criteria

6.6.5.1 RQ5(a): Ranking of Mediums

First, we report results about participants' ranking for their preferred consent form after being shown each medium in Table 6.9.

Then in following sections, each medium is discussed based on 1) the top 3 factors that influenced the ranking to address RQ5(b) and 2) the top 3 positive or negative document criteria adapted from Waller's document criteria to address RQ5(c).

Because a participant could share multiple influencing factors or document criteria we instead looked at the number of unique coded segments within their answer. Participants could share as few (though they were prompted to try to give at least one) or as many as they desired. The coded segments for influencing factors, such as the element of time, could be positive (time-saving) or negative (time-wasting). This was to help us identify the categories that were most important to participants.

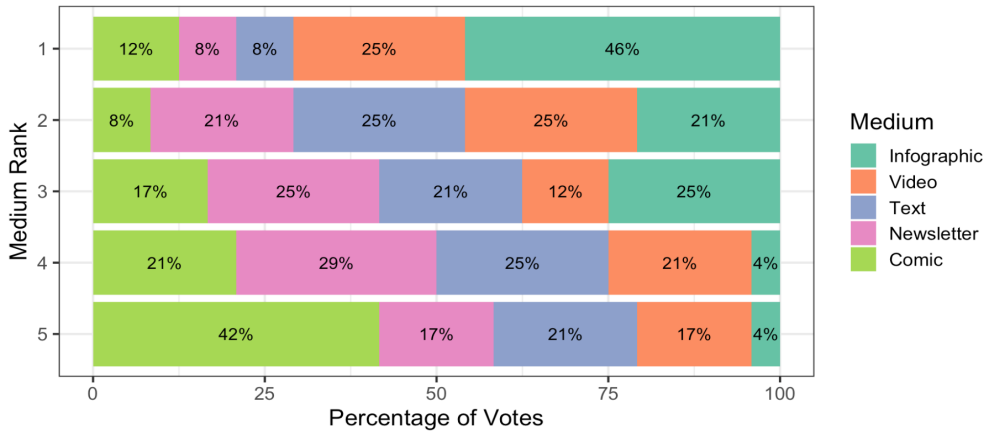

FIGURE 6.9: Participant ranking of mediums by percentage, where 1 corresponds to the first choice and 5 to the last choice

Then, we contextualize the data and report if important factors were positive or negative and respective coded segment counts in the detailed section.

6.6.5.2 Infographic Medium

Influencing Factors The infographic was strongly preferred, with one-third of participants citing understanding as positive driving factors with time and interest in close second. Elements such as the numbered bullet points, bold headings, and icons were referenced. For example, "*With the bullets, you know right away what each is about in the text written underneath. In general, this is easy to grasp*" (P3). The top three influencing factors were overall positive in contrast to further mediums below.

Document criteria The majority of positive document criteria concerns design criteria; in particular step-by-step elements, icons, bold headings, bullet points, and color. There were much fewer negatively received elements, also related to the design criteria: the overuse of color and icons and how large the infographic was. Participants had specific reactions to different icons, such as the hospital or medical professionals at the top and bottom that did not support any text, or specific icons that might seem manipulative: "*with the consent form, the 'thumbs up' graphic makes it look like I'm being preempted from making a decision*" (P14).

6.6.5.3 Video Medium

Influencing Factors The video ranked second, with almost one-third of participants reporting that it influenced their understanding, followed by time and effort. Understanding was largely positive, partially due to the format that, "*[...]]forced to watch it from beginning to end, so that you perceive the whole content*" (P15). On the other hand, time was slightly more positive than negative because while the majority of participants felt that the video saved time compared to reading, some felt it was inefficient compared to their reading speed, or wanted to review material but felt rewinding would be time-wasting. Saving effort was wholly positive with participants saying that it was more accessible, entertaining, or less attention draining while still being understanding. One minor interesting influencing factors unique to the video was a feeling of trust from the audio, with two participants mentioning that a human voice engendered confidence in the process.

Document Criteria More than half of the positive feedback about the video mentioned the audio element, followed by the sequential nature and use of animation and images. Less than one-fourth of participants liked the content which included the interplay between text and graphics and the story element. For example, "*What I like about the video is that [...] you see movements that show what you hear at the same time via audio*" (P3).

There were about half the number of negative elements than positive, and most were due to the video pacing. Some wanted it faster, while some wanted it slower. Interestingly, one participant noted: "*I have the feeling that with a video like this, people are rather uncritical of the content of the consent form. One is rather tempted to agree to something. If, for example, a button appeared after the video that allowed me to consent, I would probably consent*." (P17)

6.6.5.4 Text Medium

Influencing Factors The text ranked third. About one-third of participants indicated that interest and understanding were most influenced by the text. Interest was a complex influencing factor that was slightly more negative. Those stating it negatively influenced attention felt it was boring or lacked interest compared to other mediums. Participants who viewed it positively said that the text had a simple, clean layout allowing for quick skimming, and those who felt it was neutral felt like P21, who said,"*This is the format that I know and have simply accepted by now*'. Understanding was generally positive influencing factor, with many saying that it was clear, concise, and short, however some felt it was difficult to skim or the text was confusing or dry. Some participants also felt that it saved time by being short and concise.

Document Criteria The most cited positive elements of the text were use of clear sections, headlines, and bullet points. Positive elements were double that of negative elements, the majority also stemmed from design. Participants wanted more highlighting of key facts via color, bold, italicized, or underlined words. Less than one-third of participants also cited the negative impression the document gave them, for example, "*[...] it is still a bit boring and trivial, so you might not read it properly if you get it as a letter home, for example*" (P5).

6.6.5.5 Newsletter Medium

Influencing Factors The newsletter ranked fourth with more than a one quarter regarding prioritization, less than a quarter understanding, and 18% in interest. Prioritization and understanding were positive influencing factors, with participants saying that the bold words and ability to skip sections allowed them to roughly understand the contents because of the bold text to highlight the important information in sentences. However, interest was equally mixed, with the positive influence surrounding the bolded text and headers, while the negative influence was mainly attributed to the (unintentional) association with advertising spam. More than one-third of participants agreed with P3 who stated, "*It looks like a billboard or an election advertisement*." Though it had positive influencing factors, the negative interest likely had a large impact on the lower ranking of this form.

Document Criteria The newsletter's positive elements were largely regarding the design and use of structure, headings, bold text, sectioning, and the open format

for skimming. The negative elements also similarly mentioned the design criteria because it looked like advertising based on prior experiences. The use of color was also disliked because the black header was too strong and off-putting.

6.6.5.6 Comic Medium

Influencing Factors The comic ranked fifth and participants. The main influencing factor was understanding, with one third of participants mentioning it both positively and negatively. A slight majority cited a positive influence on understandability. Interest was generally a positive influencing factor because it was novel. Less than one fourth of coded segments showed that the comic had an overall negative influence on skimming, as the narrative driven step-by-step format made it hard to prioritize, re-read for specific elements, or gain a quick overview.

In addition, many participants explained their rankings with personal preferences indicating the audience fit was for or not suiting them, such as "*The information is all there but I would present this explanation to a child [...] at most*" (P4). This greatly influenced the ranking considering the positive affordances of the comic. For example "*I found the comic a bit inappropriate for the topic. Basically, the content or the message is better visualized by the little pictures, which may be better remembered but I don't like it*." (P20)

Document criteria Almost half of the positive feedback for the comic stemmed from the support of text by graphics, narrative elements, and illustrations. A third felt the tone and audience fit suited them. However, negative impressions were almost double of the positive ones, due mainly to the fact that audience fit and tone were unsatisfactory for more than half the participants. Participants said, "*I'm out of the age where I still like comics. [...] I don't feel like I'm being taken seriously as a customer with a consent form like this*" (P16). Other negative feedback arose from the impression and graphic elements concerning the execution of illustrations, legibility, and lack of structure.

6.6.6 Overview of all mediums

After reporting results for each individual medium of consent, we now present them together to give an overview of respective rank, influencing factors, and document criteria (Table 6.2).

Medium	Influencing Factors	Document Criteria	Rank
Infographic	(+) understanding, time, and interest	(+) numbered lists, icons, bold headings, and graphic elements (-) extraneous or leading icons	1
Video	(+)understanding and effort (+/-) time	(+) audio, step-by-step element, and interplay of text and graphics	2
Text	(+) understandable and time-saving uninteresting(-)	(+) structured layout (-) lacked highlighting elements	3
Newsletter	(+) prioritization and understanding (-) association with advertisements	(+) bolded key text, sections, and open format for skimming (-) advertisement impression	4
Comic	(+) understanding and interest (-) inappropriate fit for the context	(+) text and graphics (-) tone and audience fit	5

TABLE 6.2: Overview of top 3 influencing factors and document criteria per medium with overall participant ranking. (+) is an overall positive element, (-) is a negative element, and (-/+) is a mixed element.

6.6.7 RQ6: Emotions triggered by infographic and comic

In Figure 6.10 we compared the number of coded segments from overlapping emotions from comic (n=64) and infographic (n=52) that participants indicated on the emotion wheel (Appendix Figure 6.11, Q13-Q16). Anticipation, interest, acceptance, and surprise are the top 4 emotions present in the infographic, while in the comic the top 3 emotions are surprise, disapproval, interest, and distraction. For example, P8's emotions around the infographic: "*It is unusual, yet I like it because it is creative and surprising. When reading the infographic, I feel the emotions are attentive and trusting*" while P11 said, "*I feel surprised, confused, and dismissive because it would seem unserious to me.*" Anticipation and acceptance in the infographic (n=7) denote positive emotions and disapproval and distraction in the comic (n=9 and n=8 respectively) denote negative emotions. In fact, interest and distraction are opposites on the emotion wheel. Not all emotions were stark contrasts though, as the overlap in surprise for infographic (n=6) and comic (n=10) mainly concern the unconventional medium of consent, such as, "*My emotions are accepting, attentive, but also surprised because it is a new way of processing,*" (P15) about the infographic and, "*Looking at the comic I feel surprised and amazed in a positive sense*" (P18). Only 3 out of 10 participants meant surprise in a negative sense for the comic and 1 out of 6 for the infographic. Interest also denoted curiosity in our codebook, and this word was most often mentioned in this category in both the infographic and comic.

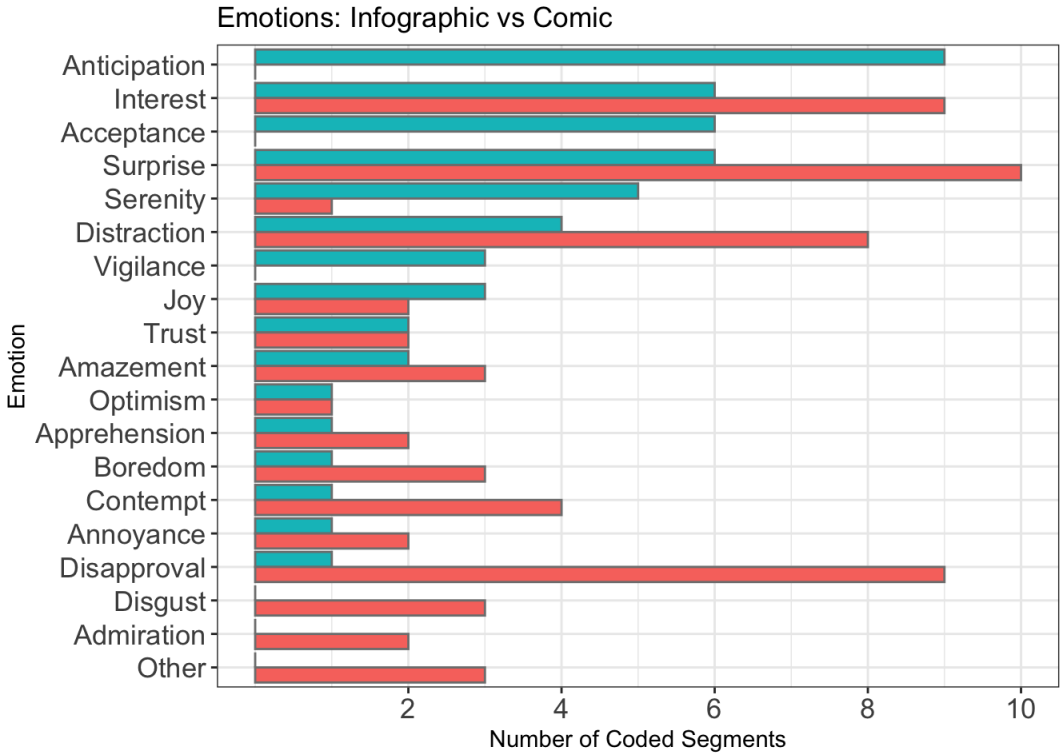

FIGURE 6.10: Emotions elicited by top and bottom ranked medium
Comic is shown in orange and infographic in blue.

6.6.8 RQ7: Consent Management and Revocation

Most particpants preferred a digital consent (n=12), while some preferred a physical document (n=3) and others did not have a preference. Following the question about

the medium, when offered the idea of a digital platform 20 participants preferred a type of online management platform, especially if it was a "one-stop-shop" where they could see an overview of all their consent and revoke them if needed (n=9). P7 stated, "*The best thing would be if I would have a matrix for each institution with multiple fields for each type of data and I could revoke and allow data processing for each of the data types separately*." To carry this out, participants had varying suggestions: a mobile app or platform (n=18), browser plugin, platform connected to an official online ID system (n=1), and a privacy manager that saved consent privacy preferences (n=1). 4 participants preferred using email, though most also shared sophisticated management systems such as a sortable code in the subject line, an email summary following browser tracking of consent decisions, or encrypted emails.

The management platform was also the most preferred way to revoke consent (n=22), while 2 participants preferred emails. When questioned about possible reasons for revoking consent, the abuse of the data via data breaches, improper storage, or misuse in the purpose or third party processors was the most common reason. 21 of the 24 interviewees reported it as a reason, followed by disinterest in the service (n=2) and privacy pessimism (n=1), where the participant would change their behavior instead of revoking consent because they felt it would be futile. Many participants had a sophisticated reasoning for their expected behavior, though it could not predict actual behavior. P18 shared, "*When I was 20, I consented to different things than I would consent to today. I would want to revoke such things. I would also insist on the final deletion of the data*," and P22 shared, "*In the event of data leaks or similar offenses, I would consider revoking my consents. However, it has never happened. In the case of spam, however, I have often revoked newsletters, for example.* "

6.7 Discussion

Concerning RQ1 about consent experiences, an element that emerged from the results is the importance of time: the majority of the participants clearly indicated that they spent only as long as necessary to understand, however almost half of them also mentioned that it should not take longer than one minute. That said, time spent on consent seems to be contextual and depends on the type of data and the entity asking to share such data.

It emerged that individuals are aware that they do not engage in attentive, word-by-word reading of consent forms, as they also expect consent (RQ2) to be short, concise, direct, and with elements that allow the visual prioritization of some information over others. Rather they engage in strategic reading [352] depending on their objective: as readers need to find "surface-level cues" to skim effectively, the consent document should include headings, bullet points, and highlights to help people navigate it efficiently and quickly grasp which information is more important than other. Our results confirm the findings of Schriver [289] that an informative document should enhance skimming and provide information in a time and context relevant to the needs and preferences of the reader. Moreover, as individuals read through the documents quickly, the information should be concise and essential, otherwise, the working memory becomes easily overloaded. However, conciseness is in contrast to the copious information that is required by transparency requirements. The relevance of content to the reader seems also crucial, as opposed to the provision of abstract and general information.

To answer RQ3, three archetypes, the Fully Informed, Record Keeper, and Trust Seeker were revealed following data from RQ2. All participants wanted a high level

of understanding to consent, with some valuing additional elements like copies for records or trust in the institutions.

To answer RQ4 (a-d), the best to worst ranked mediums were the infographic, video, text, newsletter, and comic. Within those, elements that allow the visual prioritization of certain content over others, like headings, bullet points, and highlights (i.e., "surface-level cues" [352]) that allow individuals to skim the document effectively and discern at first sight the most important information. Based on the ranking of engaging elements, participants preferred step-by-step documents (e.g., linear numbered lists with clear headings), instead of open or story-based formats. Structure, readability, and step-by-step elements are the top three engaging elements and could be easily integrated in most mediums. While our study only designed the infographic using four of the top engaging elements (i.e., structure, readability, color, and step-by-step element), other mediums like the text could also employ color and step-by-step elements instead of the open-format element. However, the tone and audience fit of mediums greatly influenced participant rankings even if affordances enhanced understanding or visual interest (e.g., comic and newsletter). Instead of prioritizing one medium over the other, there could be a greater focus on including the engaging elements wherever possible first and choosing the medium that suits the desired engaging elements and context.

For RQ5-6, we found that most participants wanted centralized consent management via digital platforms, and would revoke consent in the case of a data breach, improper storage, or misuse of the data.

6.7.1 Implications for Practice

First, the creation of data-informed archetypes can be used for better understanding the diverse needs of a population. Using data-related information as a self-determination instrument, individuals can receive contextualized information and concrete examples relevant to their specific needs (e.g., Fully Informed, Trust Seeking), rather than one-size-fits all terms. Archetypes (identified here and in future research) can support general audience tailoring for different goals. Different approaches to consent notices, for example, which some have preferred to take more individual responsibility while others consider contextual information such as their perceived trust in institutions, may reflect strategies to cope with how responsibility of consent decisions is individual while privacy is networks across the individual, responsible institutions, and more [129]. The information provided to such users might focus on building trust in organizations instead of specific data processing activities. Using archetypes to base user profiles could also be a way to customize their experience while not needing to customize every possibility, however more research is needed [246] to determine the actual benefit of tailoring information to the styles against the increased costs of its creation and implementation. Message customization for commercial and political marketing has already reached a high level of sophistication, and biomedical research institutions could use such techniques to enhance the user experience.

Second, different mediums can be targeted based on needed affordances (See Table 6.2) and layered to combine and reinforce complex information, e.g., through a combination of text, video, and infographics. Official guidance about data protection's transparency requirement implementation [243] portrays layering techniques as an appropriate means to achieve the requirement of full disclosure while allowing for prioritization and brevity. For example, summaries containing an overview of the main clauses can accompany the more comprehensive version and can be

browsed more easily at the time of consent and afterward; short videos and privacy icons can also constitute the first layer of a written notice [273]. Distributing information on separate mediums can additionally contribute to presenting the relevant information at an appropriate time: for example, the first layer with essential information can be displayed at the time of making a consent decision, while detailed information can always remain accessible on request, even later [285]. However, as more guidelines for image-heavy consents arise, [243, 273], testing and co-designing consent is key, otherwise a negative audience fit and context may be more harmful than plain text consent. This can be important to test for the intended audience, especially as comics have been a case study for cultural stigmas [189]. While have been suitable for or indigenous populations [38], some researchers are pushing for more serious comics (similar to serious games for education) [184] and the comic co-design process itself as a research practice [324].

Tailoring and Layering Approaches Moreover, in order to use data-related information as an actual self-determination instrument, individuals need to receive contextualized information and concrete examples that are relevant to their specific needs (e.g., Fully Informed archetype), rather than abstract, general terms. For such reasons, identified archetypes can support audience tailoring for different goals. The hereby defined archetypes highlight how different individuals approach consent notices, for example that some have prefer to take more individual responsibility while others consider contextual information such as their perceived trust in institutions. These may be strategies built to cope with the difficulty in decision making, wherein the responsibility of consent decisions is individual privacy is networks across the individual, responsible institutions, and more [129].

Layers could also be based on profiles, personas, or archetypes. The hereby defined archetypes highlight how different individuals approach consent notices, for example that some have prefer to take more individual responsibility while others consider contextual information such as their perceived trust in institutions. These may be strategies built to cope with the difficulty in decision making, wherein the responsibility of consent decisions is individual privacy is networks across the individual, responsible institutions, and more [129]. Layering based on these archetypes and third party ethical oversight may help users feel more secure in their decision. The designs of the consent form could therefore be informed by the goals in the identified archetypes, especially when it comes to the blueprint of the standardized consent document envisaged by the DGA [337]. Although usually consent notices are created to be as widely generalizable as possible, also due to the efforts for their creation, this approach risks alienating some user groups, such as the Jargon-Lover who may feel more precisely informed with technical terms in the notices (for instance, data protection and medical experts) or the Personal-Contact who wanted to be in conversation with a person. In addition, the Fully Informed archetype emerging from this study has shown that participants want relevant and appropriate consent documents based on their contextual needs and desires.

Layering could also be a way for users themselves to express their preferences and customize their experience by choosing their desired layers with their favorite modality to engage with consent content. Such an approach could leverage various preferences and accessibility needs(e.g., a blind user could hear the video), even though more research is needed [246] to determine the actual benefit of tailoring information to the styles against the increased costs of its creation and implementation. Customization may be time-consuming and profiling individual preferences may be inaccurate, so solutions to these possible issues should be formulated. Even though

personalized disclosures may seem burdensome, message customization commercial and political marketing have already reached a high level of sophistication. Moreover, with the rise of personal data spaces, proactively indicating personal preferences about consent modalities and data use permissions (e.g., the types of entities and specific purposes to and for which individuals wish to disclose their personal data) will become a reality (see e.g., [77].

Layering has been integrated with DC platforms. DC was built to leverage the benefits of digital communication for health research by using digital platforms to connect people and researchers and allow participants to see and change their consent and data sharing permissions dynamically. This consent arose from the Ensuring Consent and Revocation project (EnCoRe) [50, 215] as a technological response to challenges to the IC from the uncertainty of research re-use in biobanking [328]. Australia's CTRL [123], a dynamic consent platform with open-source code, incorporates multimedia (video, illustrated text, infographics), personalization options, and informational layering techniques. Building upon this, the layering could incorporate archetypes of general profiles to be tailored for different goals. Users of different ages may prefer different mediums, such as comics for younger audiences and videos for older audiences, or users with technical expertise could choose to see jargon.

Possible Manipulation Although we did not explicitly ask about undue influence on consent decisions and trust implications, participants clearly connected the two and more research is needed to better understand the deep connection. The infographic had a few complaints about specific graphics, with P14 saying that showing a "thumbs-up" icon was perceived as a manipulative way to preempt one into giving consent. Similarly for the video, P17 brought up that they might believe anything shown in the video and be inclined to give consent. Guidance on ethical nudging design [88, 269, 210], as well as research on dark patterns to avoid [229, 112] can help shed light on such thorny issues. This would also suggest that overall organizational governance and systemic changes should be carefully implemented to engender and maintain trust. Moving away from the rational person myth [290, 133], consent maximization using engaging elements is a delicate balance and increasing elements may only complicate matters.

6.7.2 Audience Fit and Context

Comparing the first and last ranked infographic and comic, audience fit had as much if not more of an impact as the more objective affordances each medium offered. Participants had specific expectations about consent forms, though they did not necessarily need to resemble a conventional plain text document and it may be dependent on personal preferences. Both mediums raised participants' interest and attention, which could be reflected in user engagement. Comics were considered appropriate by some, but mostly inappropriate (e.g., *childish*, *unserious*) due to the seriousness of medical settings and mainly caused negative emotions such as distraction and disapproval. Again, context is key. Contrary to our findings, Wermuth [356] found that comics proved useful in a medical setting with medical experts and patients (both adults and children). However, Wermuth paid attention to maintaining appropriate intent, style, tone, and emotional salience. Our prior work on consent to genomic research [38, 312] how comics can be deemed appropriate by certain communities and enhance their involvement in the process. Thus ensuring the correct depiction of the audience's culture, style and tone are critical aspects for the acceptance and

success of comics as a consent communication medium. In contrast, the infographic was more well received with positive emotions like anticipation, interest, and acceptance. The implications of emotions must not be ignored: positive first impressions may enhance attention and interest in the consent form and increase engagement, while negative ones may alienate potential readers right after their first impression and even before they engage with consent.

Hence, tone of voice and audience fit are important aspects that are rarely considered in the transparency of privacy information, whereas plain language accompanied by graphical elements and illustrations to support understanding and information navigation are nowadays recommended as best practices and their appreciation is reflected in our results [243, 273]. Privacy communication, as well as legal communication in general, has traditionally ignored the audience it intends to reach, by flattening the style to a communication made "by lawyers for lawyers" that focuses on the precision of the rules, rather than on the possibility of the intended audience to grasp their meaning and act upon such rules [247]. But as a user-centered design approach enters the realm of law, legal communication increasingly becomes permeable to considerations about the audience fit; see, e.g., the importance of tone of voice and the use of comics in contracts [121]. Comics or other mediums may still be unacceptable for some; however, layered approaches that present the information in complementary multimedia are increasingly being experimented with and could be meaningful to better tune legal-technical communication to different needs and preferences.

Trust has also been mentioned to explain certain reasoning, although it was not a specific focus of the study. For example, P19 said that they have different behaviors for cookies compared to doctor's office, P9 said they evaluate the person asking, or P16 distrusted the internet more than the doctor's office. The approach to consent may be influenced by the (perceived) trustworthiness of the institution asking it (e.g., doctors). A few participants also stressed that consent should not contain any misleading statement or any deceptive information. This reference to potential manipulation of one's own choices is also reflected in the ongoing lively discussion about digital consent, recalled in Sec. 4.2. Further research may elucidate what consent elements may increase trust in data disclosure.

From these observations, consent is a complex process that does not happen at a single point in time when the form is presented to individuals. Rather, it carries expectations derived by previous experiences and can trigger a whole set of emotions. Finding the right medium for certain audiences is not trivial, opening the question of whether standardization of consent is really possible.

6.8 Limitations

Though we strove to obtain a balanced age, education, and sex representation in our participants, they cannot be fully representative of the population. Our sample size was small for in-depth interviews, so the results cannot be generalized and should be combined with other studies. Our methods only concerned self-reported opinions so there may be a discrepancy between reported and observed preferences and behaviors. The study materials may have influenced them as well, as study materials were generated by Author 1 who is not a professional designer. Therefore, certain choices (e.g., the comic style) could have influenced participants' attitudes.

To suit each medium the base consent text had minor edits or added text (i.e., headers, ellipses). Before implementing consent mediums in line with applicable constraints, the relevant expertise should be included in the design and evaluation of each medium. Additionally, the research scenario with a data trustee has a simpler consent process compared to informed consent in clinical trials, which generally has greater ethical-legal safeguards. As a consequence, it would be important to test if our recommendations are reliable in various contexts and sufficiently fine-grained before applying them at scale.

6.9 Future Work

This raises interesting questions about how cultural contexts and prior experiences with mediums might affect the audience fit of different mediums, and it would be interesting to test across diverse populations. Perhaps other countries may rank infographics lower due to negative prior experiences

The influence of trust across different types of health consent and with different entities would also be interesting to study. For example, P19 said that they have different behaviors for cookies compared to doctor's office, or P9 said they evaluate the person asking, or P16 distrusting the internet more than the doctors office where there is lots of trust. Perhaps a doctor's office would be preferable to an unknown hospital clinic, or less trusted than a well known university research effort.

Other future work may include investigating emotions in terms of their intensity using another emotion wheel such as the Geneva emotion wheel https://www.unige.ch/cisa/gew to continue to develop a more "fun" or less "negative" consent experience.

6.10 Conclusion

In order to better understand the diversity of participant preferences, opinions, and emotions for informed health consent and the relevance of specific document criteria for engagement with various mediums (i.e., infographic, video, text, newsletter, and comic), this study interviewed 24 individuals. The results have informed the generation of archetypes and analyzed reported experiences with the different mediums to show how various document design elements reportedly affected their engagement and their overall experience with different mediums of consent. The different archetypes based on desired document features and goals and can help to create standardized consent documents that use layering to help address varying needs identified via archetypes. We also proposed recommendations for designing consent (multimedia) forms with the structure to promote prioritization such as headers, bullet points, and bold type within a contextually appropriate medium, such as an infographic or video that are seen by our participants as more attention-grabbing and appropriate than comics or a newsletter. The emotions the infographic triggered in participants were almost entirely positive and included interest and acceptance, whereas comics triggered feelings of interest in surprise, but also of disapproval and even disgust. Our results would be interesting to repeat in other countries and could lead to contextually designed consent that aligns with the GDPR and other EU regulations. Beyond this scenario, many others exist: the sharing, reuse, and analysis of data can encourage the use of data across *smart city actors* to improve the health of smart cities citizens during a pandemic or other health emergency, or to improve the educational system – to make a smart city even more liveable [295] and [294]. In

these cases, the complexity of personal data processing and disclosure, and the consent process itself can increase dramatically in comparison to the use case presented in this article. The findings reported here are meant to encourage further research to determine how to better involve individuals into designing useful, engaging consent to facilitate better data sharing in the data economy composed of increasingly complex data-informed services.

6.11 Appendix

Data (interview guidelines, full consent mediums, codebook) are available here: https://osf.io/4bvkx/?view_only=ec9b2396f7634248b2c4b66af2d1e34d

Emotion Wheel During the interviews, participants were shown the German version of the emotion wheel [2], and the English version is shown in Figure 6.11. Participants were asked to describe their emotional response to the mediums shown.

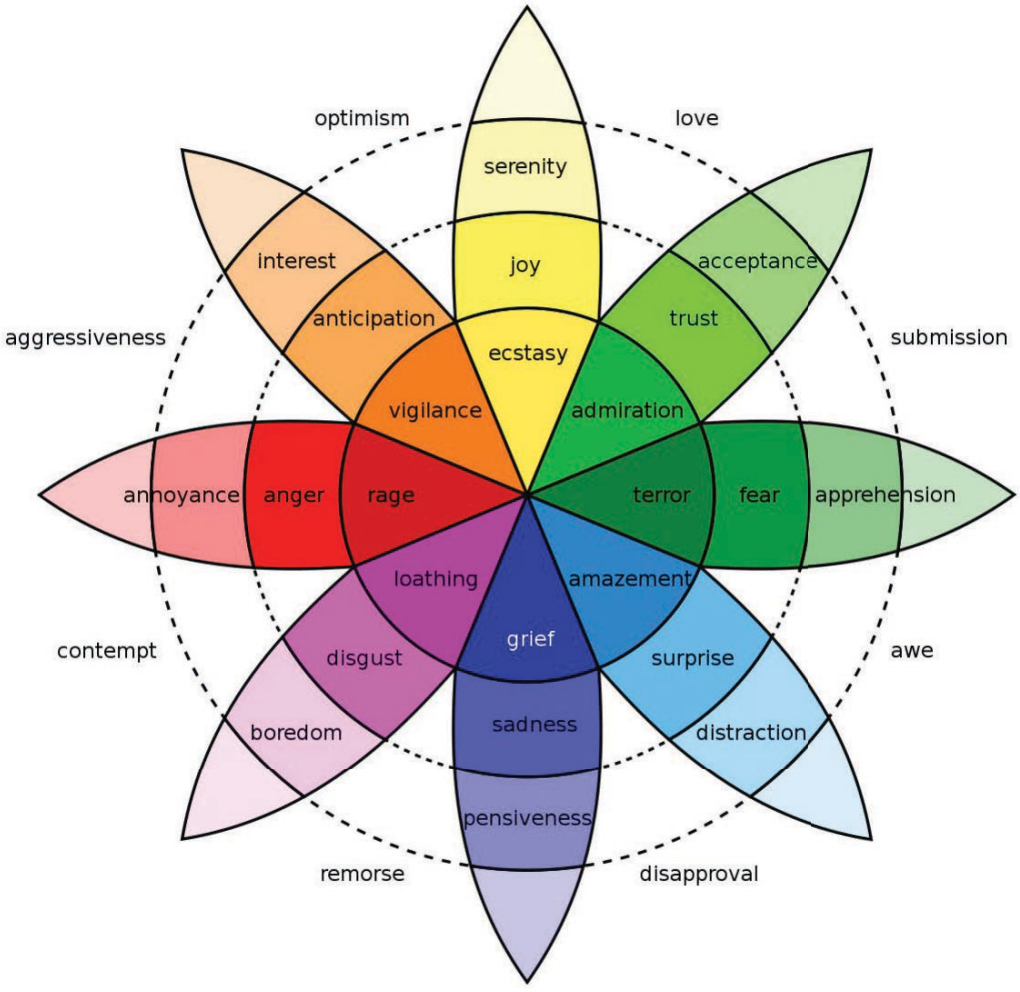

FIGURE 6.11: Plutchik's emotion wheel
Image sourced from Wikipedia, at https://en.wikipedia.org/wiki/Robert_Plutchik#/media/File:Plutchik-wheel.svg.

[2] https://de.wikipedia.org/wiki/Robert_Plutchik#/media/Datei:Plutchik-wheel_de.svg

Chapter 7

Summary and Discussion

7.1 Key Findings

Health data has emerged as a significant source of medical and research knowledge, benefiting individuals and related groups. Although there may be many shared benefits and risks, in the status quo, health data sharing remains largely individualistic (e.g., individual consent) and poorly communicated. This is because 1) much of the history of legal-ethical consent is built on individual autonomy and, in practice, collective notice and consent is ignored unless legally required, 2) consent processes are complex in the digital age with multiple parties and difficult to transparently record and understandably communicate, and 3) user-centered consent design is highly contextual and there are few studies in the field for shared health data.

My thesis work focused on the identification of collective consent gaps and the characterization of methods to enhance transparency and usability in collective digital consent. Such gaps in collective consent included the tensions between collective autonomy and individual autonomy, the legal gray-area surrounding EU guidelines for genomic data (wherein it is technically and legally collective data, but in practice it is not considered collective data), which often fail to uphold the stated informational transparency requirements. These findings, combined with the small amount of existing literature on collective consent led me to combine interdisciplinary methods in an innovative context to study how privacy, usability, and collective consent interact. I used these to empirically identify the current challenges regarding informational transparency and user-relevancy for consent in Direct to Consumer Genetic Testing Companies (DTCGTCs), characterize employee perceptions of IC methods to improve the design of consent processes and policy documentation, assess user needs and attitudes toward consent, and characterize user perceptions of engaging elements different consent mediums offered.

This work extends previous work on collective digital consent by adapting and testing methods from privacy [300, 226], requirements engineering [217, 119], HCI [352], and governance [146] in new contexts regarding genetic data and SME consent processes. This results in a new framework for implementing future prototypes to address consent challenges. I developed new datasets throughout the different studies - beginning with DTC genetic testing company's privacy and consent policies, understanding user needs for consent, consent management, and consent design using different mediums. The different spheres of the consent process studied in this thesis are shown in Figure 7.1 with their respective studies and RQs.

In this chapter, I briefly summarize the main findings from my work organized by the initial research question and discuss some key findings and contributions as defined by [366]. Then I discuss the limitations of my work and offer an illustrated summary of the framework based on all the findings. Finally, I speculate about the broad relevance and implications of my work in usable privacy and consent for

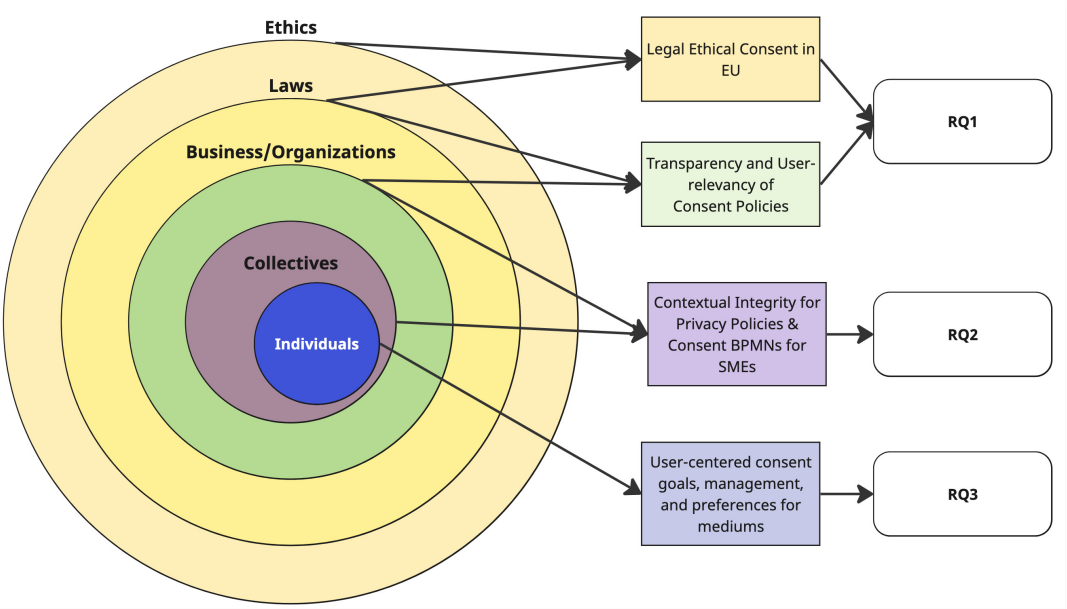

FIGURE 7.1: The different spheres present targeted by the thesis regarding the consent process and the specific studies.

collective data in the EU. Some sections of this chapter are excerpts or adapted from previous papers.

7.1.1 Collective consent is overlooked in the status quo

In this section, we discuss work addressing: ***RQ1:****How is notice and consent implemented for genomic data sharing in companies and research, and what are open challenges for collective consent?*

Looking at the practices set by leading DTCGTCs in Chapter 3, there are no requirements to disclose collective risks and benefits despite legal-ethical guidelines in the EU, such as transparency requirements – so DTCGTCs fail to mention any possible collective issues. In addition, they also fail informational transparency metrics (e.g., completeness, direct language) and are not framed towards customers. In the EU, consent can be divided into legal consent for processing personal data (as one of the legal bases) and ethical consent to align with bioethical guidelines [33] as discussed in Chapter 2. Consent has been proposed as an ethical safeguard for autonomy regardless of legal bases for processing data, especially regarding research with sensitive health data [316, 91] and should be expanded to more situations, such as commercially. In the case of DTCGTCs, they bypass any primary IC for sensitive data processing by using the legal basis that they are providing a contracted service and only ask for consent for secondary purposes (for research and sharing with external third parties). When looking at both the privacy policy and consent policies for secondary purposes, the text was often incomplete, vague, and too complex. The type and descriptions of risks and benefits varied greatly across the six different companies analyzed, from sharing only one general risk to sharing over 6 specific categories of risks. This is a key gap as privacy and consent policies offer a window into the internal data processing practices, which can both affect expert opinions on privacy and safety, as well as individual opinions on whether they will contract with a particular company or share data for secondary research purposes. The GDPR also sets high transparency requirements, which the companies may not meet through

failing these assessments. From analyzing these industry leaders' non-transparent policies, their success in the face of disregarding informational transparency and shared risks and benefits may indicate a general trend towards lack of care regarding these issues for other DTCGTCs, even with high GDPR requirements. There companies hold a large market share and seem to be able to weather data breaches, while SMEs may not. One example from Chapter 3 of shared risk and benefit information showed how it was framed to shield companies from any future lawsuits, not to reassure the reader that they would be protected. Internally, the companies likely have their own risk assessments and protocols to address different risks, but it is not shared. This contrasts with the goals of the GDPR's transparency requirements "*A central consideration of the principle of transparency outlined in these provisions is that the data subject should be able to determine in advance what the scope and consequences of the processing entails and that they should not be taken by surprise at a later point about the ways in which their personal data has been used. [...] data controllers should assess whether there are particular risks for natural persons involved in this type of processing which should be brought to the attention of data subjects. This can help to provide an overview of the types of processing that could have the highest impact on the fundamental rights and freedoms of data subjects concerning the protection of their personal data*." Sharing more information about existing protocols in a user-centered manner and including collective issues would be a step forward for informational transparency.

However, without clear guidelines, frameworks, and specific legal requirements, companies are unlikely to share such information. While EU regulations like the GDPR apply to all Member States, it also allows each Member State to enact it as they see fit. This is compounded by conflicting rights, for example, the *right to know* and the *right not to know* genomic information [52], which may be used for positive health decisions or have negative effects on one's mental state [174, 222]. Each country has specific laws surrounding confidentiality and privacy of patient information and exceptions, such as special genetic cases [345], which can lead to confusion in patients about the scope of medical confidentiality [282]. Although a guideline to suggest discussing information with family members or any type of informal collective notice and decision-making can be helpful, it would have to be tailored to each country and medical context, or else it might be more confusing than helpful. The EDPB acknowledges that there may be a collective data subject through genetic data [33], which may identify genetic families all at once. However, due to the history of individual consent and autonomy, the implementation of collective consent and collective rights is still highly debated. The conflict of different individual's rights is the most debated aspect, as how to balance different people's rights can be difficult and without much precedent [171, 172]. Regardless, DTCGTCs have been part of contentious cases in which information was unexpectedly used by law enforcement to find serial killers [359] or in which public family trees reveal surprise relatives [291], and it is surprising that there is little useful information about these difficult situations. In the latter case, one study shared how participants suggested, "*increased warnings pre-discovery and improved support post-discovery*" [291] upon learning about their true genetic parents. In contrast, it can be important to discuss information with relatives that could be useful, such as disease risk scores. There is literature that reveals that familial risk scores can be much more predictive than general population scores, even if the disease is not linked to a genetic cause [344]. However, the researchers also cautioned against using family risk too liberally, as it may increase undue worry. In these cases, it would be helpful to refer to more testing and genetic counseling. It has been suggested before that DTCGTCs "have a responsibility to provide support" to their customers, and one of the ways is through information

about genetic counseling [211]. This can make it difficult to offer specific guidelines on how to frame possible results and risks beyond directing customers to experts in genetic counseling. Though guidelines can be useful, there needs to be more research into this area to ensure usable steps for organizations.

While guidelines for shared data are not well studied, how to communicate risks to highlight understanding has been a longstanding research goal. DTCGTCs have been criticized for its portrayal of possible harms, especially after data breaches have been revealed. For example, MyHeritage's privacy policies say, "*while our reasonable security program is designed to manage data security risks and thus help prevent data security incidents and breaches, it cannot be assumed that the occurrence of any given incident or breach results from our failure to implement and maintain reasonable security.*" While legally the statement may be sound, it may not meet customer expectations. A woman who was part of a class action lawsuit against MyHeritage due to a data breach [228] stated that she would have not used the genetic services if she had known that the necessary precautions were not in place [310]. Risks could be improved by addressing general fears, using graphics, and sharing more information about risk management. Although consent policies may be framed to address participants using active language and with a clear subject (e.g., "you"), the risks are still varied in number and category across companies and do not address concerns usefully. For example, the company whose one stated risk of unexpected secondary findings would not address any of the top harms from a survey across 22 countries of attitudes towards genomic sharing [212]. The three most commonly feared potential harms from the study were: if friends or the government obtained information without consent, and the use of the information for marketing purposes. Also, while it is difficult to convey probabilities for genetic risks, the use of simple pie charts or '100 person diagrams' was preferred by participants in a study, although the most desired option was access to a health professional for questions [306]. Directly addressing common concerns using simple graphs and pictograms with professional support may be useful for DTC genetic testing websites to share risks more concretely and alleviate concerns. A solution from the EU is a Data Protection Impact Assessment [242] for the purpose of delineating those risks and devising appropriate mitigation measures for those processing special categories of data (e.g., genetic data) at scale (Art. 35(3) GDPR). Not only is it good internal practice that enables organizations to manage the substantial data protection risks that arise from genetic data processing, but some of the information could be communicated to customers to increase transparency.

Empirical Contributions

- Characterization of the gap between transparency practices of DTCGTCs and legal requirements from the GDPR

Theoretical

- Novel descriptive qualitative interdisciplinary synthesis from privacy, EU regulations, and biomedical perspectives surrounding collective digital consent.

Methodological

- Adapted, combined, and applied methods on user centered governance [146] with CI analysis [300] and risk/benefit information on DTCGTCs privacy and consent policies.

Dataset

- Developed an annotated dataset of policies related to the above methodology, available on open access repositories.

7.1.2 Interdisciplinary methods can enhance transparency and design for a better IC process

In this section we discuss findings from: ***RQ2:*** *What methods, tools, and frameworks can help address the challenges in digital collective consent?*

I identified a lack of informational transparency and user-centered framing as two of the challenges regarding collective digital consent and consent for genomic data sharing in general. Previous studies have analyzed how privacy policies for the sharing of health data are opaque and critiqued consent processes for being confusing and not protecting participants' rights. Although I used CI to analyze existing policies and it has been suggested for use in auditing, it had not yet been tested. I adapted the CI method to develop better notice (via privacy policies) in a welfare tech SME and studied employee perceptions of usefulness and ease of use (Chapter 4. In addition, I also use the RE design process to develop consent process BPMNs to address the transparency, privacy, and consent requirements for specific collective consent use cases with the same SME (Chapter 5. To address the issue of consent design I also surveyed potential users about their consent needs and tested user perception towards different mediums of consent (comic, video, text, infographic, newsletter). The study with the general public reveals several archetypes that could be used to characterize the different needs of the population, as well as the importance of specific elements across different mediums, such as graphic elements or stepwise design. The user studies within the SME revealed that both methods were useful for understanding, communicating and validating requirements, and were promising tools that employees would potentially adopt for future implementation and study.

Although many factors are likely to have an impact on employee perceptions of the CI and BPMN methods, I identified how it was useful for understanding, analysis and communication and was easy to use. Both the CI and BPMN were perceived to be well accepted through a questionnaire and semistructured interviews. These studies used the TAM, assessing employee perceptions of usefulness and ease of use to predict usage and acceptance of the technology. Therefore, it can be inferred that CI and consent BPMNs would be acceptable to use under the TAM. While no previous studies have been done in the same context, I look at similar studies to compare results. TAM and its extensions have been used in consent for health data sharing and related fields [370, 180, 79]. A study investigated patient perceptions of different consent methods and found no functional (PEOU, PU) or practical (perceived workload) differences between the three methods [362]. Another study regarding user acceptance of mobile medical platforms used the TAM and TPB with three additional three patient-centered factors: perceived convenience, perceived credibility, and perceived privacy risk, and found that perceived privacy risk, perceived credibility, and PEOU predicted the PU of such platforms [353]. Another study focused on IC providers and tested the usefulness of an IC document tool for project managers [317]. They used PU and PEOU from TAM in addition to the System Usability Scale for usability and perceived utility. They found that project managers generally thought the tool was useful, with room for improvement in compliance or usability. While not in consent, another study looked at nurses' acceptance of electronic

medical records using TAM and showed increased PU with the electronic medical records [7] but that PEOU was greatly affected by the self-efficacy of the user, the support of top management, and the quality of information. These results show that largely the core TAM theory can reveal useful information, while extensions can add more context. Similar to our studies, we also identified key PU elements and elicited improvements and sociotechnical concerns from the interview data. This suggests a promising start to using CI and BPMNs in SMEs.

Our work included core TAM elements, it may be interesting to evaluate extensions, such as the ones described above using perceived convenience, or to adapt findings of previous studies into new TAM extensions (e.g., support from top management). This is complex because many TAM extensions exist, and none fit the exact context but may have to be adapted. These changes to methodology may have to be validated and then used to evaluate users. It may offer insights into how the leadership's attitudes affect the implementation of a technology. While the interviewees in the studies included some members of leadership like the CGRCO and the head of DevOps, it would seek them out as a demographic. Individual-level acceptance is important, but more important may be how higher levels of leadership and governance can set best practices and the company culture. For example, BPMNs can be implemented in many ways, and a BPMN The governance system with executive sponsorship [349] may ensure enough resources to successfully implement a new system throughout the company.

Then the consent prototype phase, archetypes, and engaging consent medium design elements could be incorporated to meet the user's goals. To help define the goals from user interviews, we took inspiration from personas to better understand different users and needs and design fitting systems [256]. Research into personas for health focused on specific health-related needs like dementia [281, 134] or disabilities [145] but not on the overall IC process to strategically create information disclosures for different sub-demographics. The general profiles of archetypes revealed from user data showed how different demographics could receive contextualized information and concrete examples relevant to their specific needs (e.g., Fully Informed, Trust Seeking), rather than one-size-fits-all terms (Chapter 6). More personalized systems have been the subject of much debate, and CI could also follow suit. More research is needed [246] to determine the actual benefit of tailoring information to the styles against the increased costs of its creation and implementation. Using resources wisely was also the goal of investigating the most influential design elements of the mediums (comic, infographic, text, newsletter, and video) as well as the engaging elements that attracted their attention. That way, different mediums and elements could be used to target different IC goals. Structure, readability, and a step-by-step design were the top most engaging elements, while Table 6.2 in Chapter 6 offers an overview of positive and negative elements from each medium. The top-ranking medium, the infographic, ranked highly because it helped understanding, time-saving, and interest through the use of numbered lists, icons, bold headings, and graphic elements.

Empirical Contributions

- Tested CI methods for developing better policies and BPMNs for better consent processes and revealed insights into employee perceptions about PU, PEOU, and contextual elements.

- Characterized user goals via archetypes and affordances offered by different mediums by constituent parts (engaging elements, influencing factors, and document criteria [352]).

Methodological

- Tested and adapted design methods from RE and BPMNs, privacy and CI with evaluation methods from TAM and HCI for a novel collective consent business use case and showed promise for improving SME processes.

Dataset

- Built new datasets of interview data and artefacts, available on open access repositories or upon reasonable request.

7.1.3 Condensed, graphical, and contextually appropriate IC and consent management

Here we address how the thesis helped answer: ***RQ3:*** *What components can enhance informational transparency within the informed consent user experience?*

First, German adults were asked about their experiences with consent and their goals. The majority of the participants clearly indicated that they spent only as long as necessary to understand, however, almost half of them also mentioned that it should not take longer than one minute. That said, time spent on consent seems to be contextual and depends on the type of data and the entity asking to share such data (Chapter 6).

The main goal was also impacted by different emotions, document criteria, and engaging elements (Chapter 6). The mediums ranked from best to worst were the infographic, video, text, newsletter, and comic. Among those, surface level cue elements that allow visual prioritization like headings, bullet points, and highlights [352]) enabled individuals to skim the document effectively and discern at first sight the most important information. Based on the ranking of engaging elements, participants preferred step-by-step documents (e.g., linear numbered lists with clear headings), instead of open or story-based formats. Structure, readability, and step-by-step elements are the top three engaging elements and could be easily integrated in most mediums. While our study only designed the infographic using four of the main engaging elements (i.e., structure, readability, color, and step-by-step element), other mediums like the text could also employ color and step-by-step elements instead of the open-format element. Importantly, context is key. Although the comic had positives regarding understanding and interest, it elicited feelings of disapproval and was deemed too "unserious" for health consent for adults, and would be better suited for children, the elderly, or non-native German speakers. On the other hand, the infographic performed the best because it improved understanding, saved time, and was interesting while eliciting feelings of anticipation, interest, acceptance, and surprise. Contrary to [354], which surveyed adult students, comics were not appropriate for the majority of German adults surveyed, who thought it would suit children, the elderly, or non-native German speakers better. Comics have been useful for communicating medical information with children [117], while another found that comics for medical IC especially increased understanding for children with lower comprehension scores [96]. Understanding was one of the positive affordances given by comics and can help give insight into how strong negative emotions and bad audience fit can override any benefits within a medium. Comics have

been a case study for cultural stigmas [189], and while they may have been suitable for an indigenous population [38], some researchers are pushing for more serious comics (similar to serious games for education) [184] and the comic co-design process itself as a research practice [324]. It is important to test and co-design consent with the intended audience, otherwise a negative audience fit and context may be more harmful than plain text consent.

Another key finding was how many participants wanted a centralized consent management platform to be able to see, manage, and track the consent they have been given. A digital management platform through a mobile app or website was the most desired way to manage their decisions, and participants also said they would like to revoke consent in this way. This indicates that individual informational transparency has limits. While informational transparency can address better engagement and understanding of specific ICs, over a lifetime they become unmanageable. Due to the many different consent management platforms and lack of centralization, other researchers have focused on interoperability and standardization. Existing browser-based consent mechanisms range from Do Not Track, Global Privacy Control, ADPC, and more [139] while consent apps and platforms may use custom or proprietary tools. Dynamic consent management platforms, even if not centralized, can be beneficial to building trust, carrying out IC over time, and providing feedback to participants [265, 123, 122, 199]. Building upon this, the previous archetype layering methods could incorporate archetypes of general profiles to be tailored for different goals. Users of different ages may prefer different mediums, such as comics for younger audiences and videos for older audiences, or users with technical expertise could choose to see jargon.

Empirical Contributions

- Insights into how user perceptions of IC mediums are influenced by emotions, and audience fit compared to medium affordances.
- Characterization of user desires for centralized consent management.

Dataset

- Developed new interview corpus datasets, available on open access repositories or upon reasonable request.

7.2 Limitations

Generalizability of User Studies The user studies were qualitative in-depth studies investigating specific contexts and participants, so the findings may not necessarily be extrapolated to other populations. Although the findings were useful for detailed insight into user perspectives, it would be important to expand the sample size and context for future work for external validation. It may also be interesting to replicate in other specific use cases, as the context was found to be highly important for the user studies, with employees' perceptions tied to their day-to-day responsibilities and the German adults' perceptions of comics heavily influencing the ranking contrary to more objective uses the medium provided. While exact replications are unlikely in the field, any extensions using this work should be repeated across different populations and more diverse demographics to increase reliability

and generalizability. For example, the German population had a very strong reaction to comics, and research on comics in other populations has shown more positive feedback (e.g., the San people in South Africa [38]). There may also be some biases in the design of the study materials for Chapter 6, as they were created by the author and not by professional artists or designers, which may have affected user perception. This also shows how expanding the types of mediums can require substantial resource investment, and it would be prudent to test co-designed mediums for the specific IC context to be efficient.

Evaluation of the Framework The thesis contribution cumulated in a framework for more transparent, user-centered collective digital consent based on various studies, so the natural next step would be to test the framework with a prototype combining the individual study contributions such as the CI for the consent notice, BPMNs for consent process design, and collective digital consent following design guidelines for management with an infographic. A longitudinal study of applying the method before and after would be able to test the effects of implementation, and in this case, it may be useful to continue testing the framework with the same company. Once that is done, it would be useful to iterate on key points of the framework and test the validity of the results at similar companies to understand how it applies to other contexts. This should be tested and improved for sufficient usefulness and ease of use as part of an iterative RE and UX design approach.

The evaluation may also benefit from more nuanced investigation into different sub-demographics or incorporating other spheres. Although our interviewees with the SME included the CGRCO and the head of DevOps, we did not specifically target leadership as a demographic while a similar study has shown that it might have an impact on the adoption of a technology [7]. If using the TAM, extensions should be evaluated, even if none were developed for exact context. In addition, investigating an economic sphere can enrich the current framework by examining the possible economic impacts of implementing such systems, and the legal sphere will need to be re-evaluated considering the anticipated regulation in the following section.

7.3 Graphical Summary of Key Contributions

Figure 7.2 shows a diagram of an iterative consent process from legal-ethical requirements to consent requirements and modeling to the consent prototype and evaluation, then repeating. The methods and main findings from the thesis that address different steps are shown below with the different consent process steps as they apply.

7.4 Future Work and Open Questions

In this section, I dive into areas of future work that have many open questions and opportunities for research. First, I discuss one of the most pressing issues – open challenges for collective consent, data sharing, and data management. While there may not be much research in the field to track how collective issues are evolving, evermore collective data will be processed or created. Then I discuss how the new EU regulation, the EHDS may affect consent for health data sharing. The current draft does explicitly rely upon consent, and it may conflict with the GDPR or the role of IC in EU data protection.

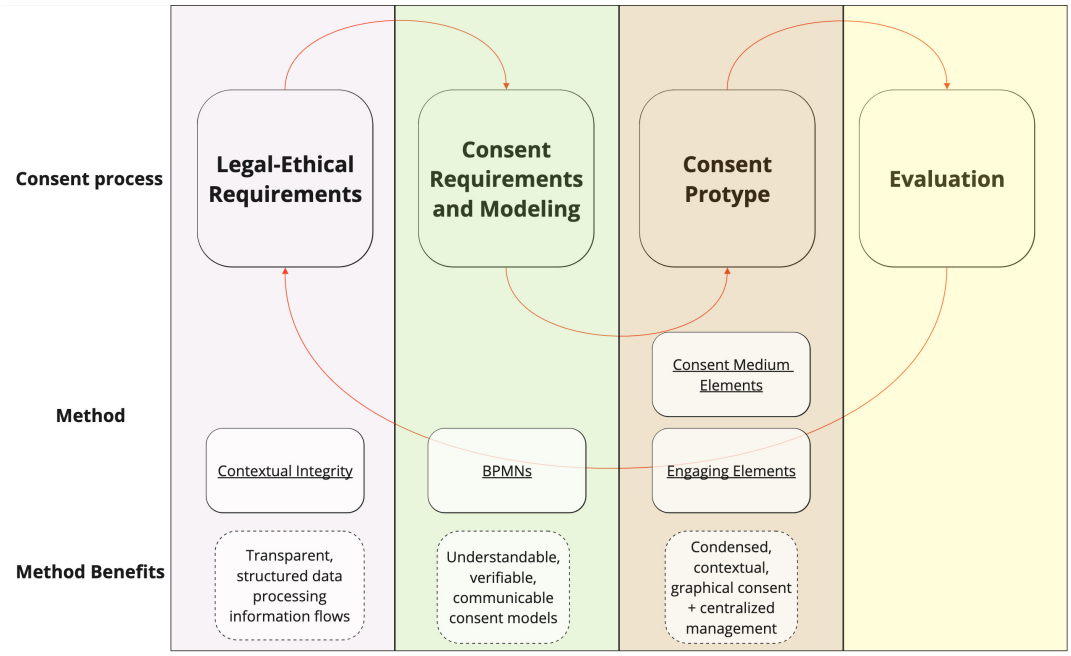

FIGURE 7.2: A diagram of an iterative consent process and the methods and main findings from the thesis that address different steps.

Collective Issues Many open questions for collective data management remain, from continuing to prototype collective consent, collective consent management, and applications for non-health data. Remaining challenges for a collective, dynamic consent include how this might apply to diffuse groups (such as genetic relatives), how to vote, and how to manage conflicts [171, 172]. The EDPB guidelines on genetic data state that data subjects can be families [244], but years after this guideline, collective data is still full of controversy. There are varying guidelines across countries for giving notice to family members that their shared genetic data is being processed [320], from an ethical perspective, Minari et al. [214] argue for a form of family-group consent for genetic data processing due to shared risks. However, it may also conflict with international bioethical guidelines with the *right not to know* [90, 109], where someone should not be informed of health information against their will. It has also been argued that managing conflicts between individual and collective rights would be feasible [26] with the GDPR as a starting point. In addition, different countries have various approaches to weighing the rights of all parties based on the context and existing rulings (e.g., the right to privacy of the deceased is overruled by the right to health of the living [255]) and laws. This may also help data minimization principles by limiting data sharing and access to only well-justified cases [26].

While many issues remain, it would be prudent to begin with specific use cases instead of focusing on the huge scope of the issue. Previously, scholars have considered DC for Aboriginal Australian and Torres Strait Islander's collective consent, saying, "*DC may have specific relevance for Indigenous people if it can support group decision making*" [253]. DC aims to use a digital consent platform to center data subjects in facilitating consent over time and communication with research projects [158], can incorporate the above technologies and enable autonomy and address issues in the unspecified data subject (both individual and collective) and the unspecific data processing purposes and processing entities. DC can ask for specific consent over time as data processing changes, since DC is a long term tool that can allow

for updates and re-consent in case the purpose changes in specification or the data controllers and processors, as well as based on individual preferences. Consent can also be layered, so not only could specific consent be used but also broad consent on the same platform. For example, specific consent could be the default but based on a layered approach showing key information to detailed information, broad consent could be chosen by the individual. Or, a personalized approach based on the individual's legal jurisdiction and privacy preferences could also be used. If applied in the biomedical research context, this can help to comply with Recital 33 and 42 GDPR, in cases when the purpose of processing or the identity of the entities involved becomes more clear during the processing life cycle. When the research projects for which the data is used are better delineated or new entities are involved in the processing of data, this information can be communicated to the data subject who can then exercise their rights accordingly. This can be facilitated by DC but technology cannot solve whether re-consent should be asked or if notice is enough, as that is still up to the discretion of the data controller. The scholars also acknowledge the remaining challenges of DC for the Aboriginal Australians and Torres Strait Islanders – poor electronic infrastructure for remote populations, lack of digital literacy and needed sovereignty, and unsolved governance issues, such as, "*Allowing a DC interface to support shared decision making in the Indigenous context would require careful consideration, amongst the community, of who should have access, and who is able to make changes and edit preferences* [253]. Outside of work with indigenous populations with established collective dynamics, other types of use-cases to continue work on could be with organizations that have a rough collective or group data sharing model, such as with the company interviewed in Chapter 4 and 5 which deals with a complex patient and NoK situation. It may also be interesting to expand out of genetic or health data, as social media data is an area where networked privacy was first conceived [40, 198] and has been studied in other aspects, such as how to mediate group decision making for privacy choices across social media networks [190]. There seem to be infinite open questions and it would be interesting to study as the field is still growing.

European Health Data Space The EHDS is a new regulation, proposed in 2022. The draft underwent some significant changes and in April 2024 the European Parliament and Council agreed upon the provisional text. At the time of writing in January 2025, the final version has not yet been officially published. Following official publication, it will undergo secondary legislation regarding implementation, delegation, and technical specifications by 2027. Then from 2027 to 2029, each Member State will prepare to implement the regulation.

It may completely change the landscape of consent for health data in the EU [241]. While the EHDS is said to be built upon the GDPR and previous regulations, the EHDS does not explicitly follow the strict, affirmative specific consent requirements for primary or secondary uses of health data in the current draft (Art. 8(e)(h) EHDS draft[1]). Instead, it relies on public interest as the legal basis processing primary health data for MyHealth@EU to enable cross-border data sharing, better connectivity, and efficient healthcare services in the EU (Recital 34 EHDS). The legal bases for processing secondary data may include but not be limited to consent, "*For the purpose of processing electronic health data for secondary use, one of the legal bases referred to in Article 6(1), points (a), (c), (e) or (f), of Regulation (EU) 2016/679 in conjunction with Article 9(2) thereof is required*" (Recital 52 EHDS draft) [340]. In the current

[1] see footnote 1

draft, Art. 10(1) states,"*Member States' laws may provide that natural persons have the right to opt out from the access to their personal electronic health data*." [340] However, this is not required. Recital 52 also states that, "*An easily understandable and accessible user-friendly mechanism to exercise that right to opt out should be provided for*," however, it may be overridden based on public interest. Art. 71 then covers the opt-out for processing of personal electronic health data for secondary use. This article writes that "*Member states shall provide*" this mechanism in an understandable and accessible format where data subjects can exercise their right at any time, indicating that this is required. However, this is also subject to exceptions, such as scientific public interest.

In addition, Member States may add additional rules for genetic data or other categories if they are stricter and more protective (Recital 52 EHDS draft) [340]. The draft states, "*This Regulation does not affect Member States' competences concerning the initial registration of personal electronic health data, such as making the registration of genetic data subject to the natural person's consent or other safeguards*" [340]. This may conflict with other countries' rules, creating more complex mazes of regulation instead of harmonizing rules across the EU as intended. Recital 18 acknowledges the different attitudes towards privacy across Member States and allows stricter opt-out of primary processing of data in such terms, saying, "*due to the different sensitivities in the Member States on the degree of patients' control over their health data, Member States should be able to provide for an absolute right to opt out from access to their personal electronic health data by anyone other than the original controller, without any possibility to override that opt-out in emergency situations. In such a case, Member States should establish the rules and specific safeguards regarding such opt-out mechanisms. Those rules and specific safeguards could also relate to specific categories of personal electronic health data, for example genetic data*." [340]

The EHDS would cover a wide range of data, "*Every hospital, every health professional, and every organization possessing health data is caught by this obligation. In Article 33 a great number of health data are listed, such as [electronic health records] and human genetic, genomic, and proteomic data*" [116, 335]. Since there may be different legal bases for processing primary health data, any existing affirmative consent infrastructure may also change. This is and subject to critique: "*Member States having an opt-in system for sharing digital health data based on the prior consent given by natural persons, are obliged to abandon this system. The reason for this is that in the EHDS Regulation the fundamental decision is taken that the opting-out principle applies to the sharing of these data*" [116].

Each Member State would have a Health Data Access Body (HDAB) that would act as an intermediary for data access, granting permits to third parties who wish the access the data for secondary use (beyond primary medical care). This process is predicted to be challenging, as legal health data experts are scarce and would vary greatly by Member State [259]. Although it is written that the EHDS is built on the GDPR and data access permits themselves must comply with GDPR in terms of data minimization and storage, the exact role of consent is unclear in the first draft: "*As Article 46 EHDS Regulation does not mention the consent of the patient concerned as a condition for granting a data permit, national laws requiring such permission are not in line with EU law and have to be disapplied, once Article 46 of the EHDS Regulation enters into force without being amended substantially*" [116]. In the second draft [241], opt-out is mentioned but the mechanism and transparency requirements are still unclear (Art. 8, Art. 48 EHDS draft [2]). Consent is one of the legal bases for processing

[2] see footnote 1

secondary data, as mentioned before, but there are also others that may apply. This is a stronger protection of data subject's rights and transparency into data processing than in the previous version [335] and was highly suggested by researchers and the EDPB and EDPS [31, 311]. It may still be lacking adequate protections as it depends on each Member State to impose stricter safeguards, or to to allow for opt-out of sharing primary health data outside of their direct provider. The actualization of the regulation is ongoing, and the second state of legislation is yet to begin. It will be important to track how different EU Member States approach the regulation [116] and attempt to harmonize at the EU level.

It seems that it will greatly change the roles of data subjects and instead place more emphasis on the data intermediaries in charge of health data. Although collective consent is still ethically justified, especially for genomic data, the overall role of health data consent may change greatly due to the shift from opt-in consent to possibly opt-out for primary uses and opt-out for secondary uses. This would change the timing of consent decisions to *ex post* instead of *ex ante*, and more research on how this may affect users is needed. For example, transparency measures implemented for opt-in consent before data collection may not impact users the same when used for opt-out consent after data collection. The future work of a user-centered, transparent health data digital consent fits with the GDPR, and may fit with the EHDS, but would have to be seriously (re)considered once more information is available.

Chapter 8

Conclusion

Using interdisciplinary methods, I combined user studies, literature reviews, and analysis of privacy policies for quantitative and qualitative data to demonstrate the usefulness of BPMNs and contextual integrity methods to investigate current challenges in legal-ethical consent in the EU and DTC genetic testing company practices, improve business privacy practices, and categorize consent needs and preferences in the general population. Together, this builds the foundation for more user-centered collective consent that addresses end-user needs as well as business needs, aligned with EU regulations and ethical standards for health data. These results offer a framework for collective consent across multiple angles – ethical, legal, governance/business, and end-users for better collective consent systems. Considering that there are only speculations [253] and no existing research on how to build a digital collective consent, this is an important step towards a more formalized system for collective decision making and consent for all people involved. In this way, the shared risks and benefits of health data can be holistically considered by all relevant parties, and informed decisions can be recorded. These results provide compelling evidence to support the idea that more informationally transparent, user-centered consent is possible for collective data in the EU. Therefore, studying collective consent can be critical to the EU data sharing economy, individual autonomy, and collective privacy, and helping to protect sensitive data while also making it as open as possible for people to benefit from the findings.

Bibliography

[1] Joint Technical Committee ISO/IEC JTC 1. *ISO/IEC 29184:2020*. (Accessed on 05/06/2024). URL: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/03/70331.html.

[2] 23andMe. *23andMe for Healthcare Professionals*. Accessed on 12/06/2022. URL: https://medical.23andme.com/.

[3] ISO/IEC Joint Technical Committee 1/SC 27. *ISO/IEC TS 27560:2023*. (Accessed on 05/10/2024). URL: https://www.iso.org/standard/80392.html.

[4] Australian Institute of Aboriginal and Torres Strait Islander Studies. *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*. (Accessed on 05/06/2024). 2020. URL: https://aiatsis.gov.au/sites/default/files/2020-10/aiatsis-code-ethics.pdf.

[5] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, et al. "Nudges for privacy and security: Understanding and assisting users' choices online". In: *ACM Computing Surveys (CSUR)* 50.3 (2017), pp. 1–41.

[6] Anshu Agarwal and Andrew Meyer. "Beyond usability: evaluating emotional response as an integral part of the user experience". In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. Chi Ea '09. New York, NY, USA: Association for Computing Machinery, 2009-04, pp. 2919–2930.

[7] Bakheet Aldosari, Sheema Al-Mansour, Hanan Aldosari, and Abdullah Alanazi. "Assessment of factors influencing nurses acceptance of electronic medical record in a Saudi Arabia hospital". In: *Informatics in Medicine Unlocked* 10 (2018), pp. 82–88.

[8] Anita L Allen. "Coercing privacy". In: *William & Mary Law Review* 40 (1998), p. 723.

[9] Sheri A Alpert. "Protecting medical privacy: challenges in the age of genetic information". In: *Journal of Social Issues* 59.2 (2003), pp. 301–322.

[10] Orlando Amaral, Sallam Abualhaija, Damiano Torre, Mehrdad Sabetzadeh, and Lionel C Briand. "AI-enabled automation for completeness checking of privacy policies". In: *IEEE Transactions on Software Engineering* 48.11 (2021), pp. 4647–4674.

[11] Organization of American States. *American Declaration on the Rights of Indigenous Peoples*. (Accessed on 01/06/2024). 2016. URL: https://indianlaw.org/sites/default/files/ADRIP%5C%201-17-17.pdf.

[12] Ancestry.com. *Company Facts*. Accessed on 12/07/2022. URL: https://www.ancestry.com/corporate/about-ancestry/company-facts.

[13] Holly Antal, H Timothy Bunnell, Suzanne M McCahan, Chris Pennington, Tim Wysocki, and Kathryn V Blake. "A cognitive approach for design of a multimedia informed consent video and website in pediatric research". In: *Journal of biomedical informatics* 66 (2017), pp. 248–258.

[14] Annie I Antón, Julia Brande Earp, and Angela Reese. "Analyzing website privacy requirements using a privacy goal taxonomy". In: *Proceedings IEEE Joint International Conference on Requirements Engineering*. IEEE. 2002, pp. 23–31.

[15] World Medical Association. "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects". In: *JAMA* 310.20 (2013-11), pp. 2191–2194.

[16] World Medical Association. *World Medical Association Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks*. (Accessed on 11/05/2023). 2016-10. URL: https://www.wma.net/what-we-do/medical-ethics/declaration-of-taipei/.

[17] Khadija Baig, Reham Mohamed, Anna-Lena Theus, and Sonia Chiasson. ""I'm hoping they're an ethical company that won't do anything that I'll regret": Users Perceptions of At-home DNA Testing Companies". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. ACM, 2020, pp. 1–13.

[18] Sophia Jeeyun Baik. "Layers of "networked privacy": context collapses across relations, technologies, institutions, and data". In: *AoIR Selected Papers of Internet Research* (2020-10).

[19] Yannis Bakos, Florencia Marotta-Wurgler, and David R Trossen. "Does anyone read the fine print? Consumer attention to standard-form contracts". In: *The Journal of Legal Studies* 43.1 (2014), pp. 1–35.

[20] Lisa M Ballard, Rachel H Horton, Sandi Dheensa, Angela Fenwick, and Anneke M Lucassen. "Exploring broad consent in the context of the 100,000 Genomes Project: a mixed methods study". In: *European journal of human genetics* 28.6 (2020), pp. 732–741.

[21] Gaia Barazzetti, Francesca Bosisio, Daria Koutaissoff, and Brenda Spencer. "Broad consent in practice: lessons learned from a hospital-based biobank for prospective research on genomic and medical data". In: *European Journal of Human Genetics* 28.7 (2020-07), pp. 915–924. ISSN: 1018-4813. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7316733/ (visited on 2022-03-15).

[22] Jan M Bauer, Regitze Bergstrøm, and Rune Foss-Madsen. "Are you sure, you want a cookie?–The effects of choice architecture on users' decisions about sharing private online data". In: *Computers in Human Behavior* 120 (2021), p. 106729.

[23] Tom L Beauchamp. "Informed consent: its history, meaning, and present challenges". In: *Cambridge Quarterly of Healthcare Ethics* 20.4 (2011), pp. 515–523.

[24] Tom L Beauchamp and O Rauprich. "Principlism". In: *Encyclopedia of Global Bioethics*. Ed. by H ten Have. Champlain: Springer, 2016, pp. 1–12.

[25] Benjamin Bergemann. "The consent paradox: Accounting for the prominent role of consent in data protection". In: *Privacy and Identity Management. The Smart Revolution: 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2. 2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers 12*. Springer, 2018, pp. 111–131.

[26] Iñigo De Miguel Beriain and Daniel Jove. "Is it possible to place limits on the self-determination of your own genetic data? Certainly, and there is an urgent need for it!" In: *BioLaw Journal-Rivista di BioDiritto* 1S (2021), pp. 209–222.

[27] Jaspreet Bhatia, Travis D Breaux, Joel R Reidenberg, and Thomas B Norton. "A theory of vagueness and privacy risk perception". In: *2016 IEEE 24th International Requirements Engineering Conference (RE)*. IEEE. 2016, pp. 26–35.

[28] Roberta Biasiotto, Peter P Pramstaller, and Deborah Mascalzoni. "The dynamic consent of the Cooperative Health Research in South Tyrol (CHRIS) study: broad aim within specific oversight and communication". In: *BioLaw Journal-Rivista di BioDiritto* 1S (2021), pp. 277–287.

[29] Dominik Q Birkmeier, Sebastian Klöckner, and Sven Overhage. "An empirical comparison of the usability of BPMN and UML activity diagrams for business users". In: *European Conference on Information Systems 2010 Proceedings*. 2010.

[30] Edward J Bloustein and Nathaniel J Pallone. *Individual and group privacy*. Routledge, 2018.

[31] European Data Protection Board. *EDPB-EDPS Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space*. (Accessed on 31/06/2024). URL: https://www.edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-032022-proposal%5C_en.

[32] European Data Protection Board. *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*. (Accessed on 31/06/2023). 2019-11. URL: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en.

[33] European Data Protection Board. *Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR)*. (Accessed on 07/05/2023). 2019-01. URL: https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers%5C_en.

[34] Narasimha Bolloju and Sherry XY Sun. "Benefits of supplementing use case narratives with activity diagrams—An exploratory study". In: *Journal of Systems and Software* 85.9 (2012), pp. 2182–2191.

[35] Piero A Bonatti, Luigi Sauro, and Jonathan Langens. "Representing consent and policies for compliance". In: *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2021, pp. 283–291.

[36] Luca Bonomi, Yingxiang Huang, and Lucila Ohno-Machado. "Privacy challenges and research opportunities for genomic data sharing". In: *Nature genetics* 52.7 (2020), pp. 646–654.

[37] Wendy Bos and Eline M Bunnik. "Informed consent practices for exome sequencing: An interview study with clinical geneticists in the Netherlands". In: *Molecular Genetics & Genomic Medicine* 10.3 (2022), e1882.

[38] Maria Wilhelmina Botes and Arianna Rossi. "Visualisation Techniques for Consent: Finding Common Ground in Comic Art with Indigenous Populations". In: *COnSeNT 2021 - 1st International Workshop on Consent Management in Online Services, Networks and Things*. Online: Ieee, 2021, pp. 292–297.

[39] Marietjie Botes, Paul Esselaar, and Donrich Thaldar. *Draft of Model law on health data governance*. (Accessed on 15/06/2024). 2024-05. URL: https://healthdataprinciples.org/pdfs/Model-Law-on-Health-Data-Governance-English.pdf.

[40] Danah Boyd. "Networked privacy". In: *Surveillance & society* 10.3/4 (2012), p. 348.

[41] Scott Brave and Cliff Nass. *Emotion In Human-computer Interaction*. 2nd ed. Stanford, California: CRC Press, 2007.

[42] Bundesdruckerei. *Datentreuhänder*. https://www.bundesdruckerei.de/de/loesungen/datentreuhaender. 2022.

[43] Eline M Bunnik, A Cecile J W Janssens, and Maartje H N Schermer. "A tiered-layered-staged model for informed consent in personal genome testing". In: *European Journal of Human Genetics* 21.6 (2013-06), pp. 596–601.

[44] Wylie Burke, Armand H Matheny Antommaria, Robin Bennett, Jeffrey Botkin, Ellen Wright Clayton, Gail E Henderson, Ingrid A Holm, Gail P Jarvik, Muin J Khoury, Bartha Maria Knoppers, et al. "Recommendations for returning genomic incidental findings? We need to talk!" In: *Genetics in Medicine* 15.11 (2013), pp. 854–859.

[45] Terrell Ward Bynum. "Flourishing ethics". In: *Ethics and Information Technology* 8.4 (2006), pp. 157–173.

[46] Terrell Ward Bynum. "The foundation of computer ethics". In: *Computers and Society* 30.2 (2000), pp. 6–13.

[47] Lois Cameron and Joan Murphy. "Obtaining consent to participate in research: the issues involved in including people with a range of learning and communication disabilities". In: *British Journal of Learning Disabilities* 35.2 (2007), pp. 113–120. ISSN: 1468-3156.

[48] Sarah E Carter. "A Value-Centered Exploration of Data Privacy and Personalized Privacy Assistants". In: *Digital Society* 1.27 (2022), pp. 1–24.

[49] Sarah E Carter, Ilaria Tiddi, and Dayana Spagnuelo. "A "Mock App Store" Interface for Virtual Privacy Assistants". In: *HHAI2022: Augmenting Human Intellect: Proceedings of the First International Conference on Hybrid Human-Artificial Intelligence*. Ed. by Stefan Schlobach, María Pérez-Ortiz, and Myrthe Tielman. IOS Press, 2022, pp. 266–268.

[50] Marco Casassa Mont, Siani Pearson, Sadie Creese, Michael Goldsmith, and Nick Papanikolaou. "A conceptual model for privacy policies with consent and revocation requirements". In: *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*. Springer, 2010, pp. 258–270.

[51] Timothy Caulfield and Jane Kaye. "Broad Consent in Biobanking: Reflections on Seemingly Insurmountable Dilemmas". In: *Medical Law International* 10.2 (2009), pp. 85–100.

[52] Ruth Chadwick, Mairi Levitt, and Darren Shickle. *The Right to Know and the Right Not to Know: Genetic Privacy and Responsibility*. 2nd ed. Cambridge Bioethics and Law. Cambridge University Press, 2014.

[53] Belgian Data Protection Authority (Litigation Chamber). *Decision on the merits 21/2022 of 2 February 2022: Complaint relating to Transparency & Consent Framework*. (Accessed on 05/06/2024). 2022. URL: https://www.autoriteprotectiondonnees.be/publications/decision-quant-au-fond-n-21-2022-en.pdf.

[54] James F Childress. "The Place of Autonomy in Bioethics". In: *The Hastings Center Report* 20.1 (1990), pp. 12–17.

[55] Michele Chinosi and Alberto Trombetta. "BPMN: An introduction to the standard". In: *Computer Standards & Interfaces* 34.1 (2012), pp. 124–134.

[56] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung. "The role of privacy fatigue in online privacy behavior". In: *Computers in Human Behavior* 81 (2018), pp. 42–51.

[57] Tom Christensen and Per Lægreid. "Trust in government: The relative importance of service satisfaction, political factors, and demography". In: *Public performance & management review* 28.4 (2005), pp. 487–511.

[58] Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. "The production of information in the attention economy". In: *Scientific reports* 5.1 (2015), pp. 1–6.

[59] Niamh Clarke, Gillian Vale, Emer P Reeves, Mary Kirwan, David Smith, Michael Farrell, Gerard Hurl, and Noel G McElvaney. "GDPR: an impediment to research?" In: *Irish Journal of Medical Science* 188 (2019), pp. 1129–1135.

[60] Ellen Wright Clayton, Barbara J Evans, James W Hazel, and Mark A Rothstein. "The law of genetic privacy: applications, implications, and limitations". In: *Journal of Law and the Biosciences* 6.1 (2019), pp. 1–36.

[61] Domenico Consoli. "A New Concept of Marketing: The Emotional Marketing". In: *Broad Research in Accounting, Negotiation, and Distribution* 1.11 (2010), pp. 52–59. ISSN: 20678177.

[62] David J Cook, Dennis M Manning, Diane E Holland, Sharon K Prinsen, Stephen D Rudzik, Véronique L Roger, and Claude Deschamps. "Patient engagement and reported outcomes in surgical recovery: effectiveness of an e-health platform". In: *Journal of the American College of Surgeons* 217.4 (2013), pp. 648–655.

[63] Esther Cuerda-Galindo, X Sierra-Valenti, E González-López, and F López-Muñoz. "Syphilis and human experimentation from World War II to the present: a historical perspective and reflections on ethics". In: *Actas Dermo-Sifiliográficas (English Edition)* 105.9 (2014), pp. 847–853.

[64] Bill Curtis, Marc I Kellner, and Jim Over. "Process modeling". In: *Communications of the ACM* 35.9 (1992), pp. 75–90.

[65] Bart Custers, Francien Dechesne, Wolter Pieters, Bart Schermer, and Simone Van Der Hof. "Consent and privacy". In: *The Routledge handbook of the ethics of consent*. Routledge, 2018, pp. 247–258.

[66] Catherine D'Ignazio. "Creative data literacy: Bridging the gap between the data-haves and data-have nots". In: *Information Design Journal* 23.1 (2017-07), pp. 6–18.

[67] Annette De Vito Dabbs, Brad A Myers, Kenneth R Mc Curry, Jacqueline Dunbar-Jacob, Robert P Hawkins, Alex Begey, and Mary Amanda Dew. "User-centered design and interactive health technologies for patients". In: *CIN: Computers, Informatics, Nursing* 27.3 (2009), pp. 175–183.

[68] Rex Dalton. "When two tribes go to war". In: *Nature* 430.6999 (2004), pp. 500–503.

[69] Fida K Dankar, Marton Gergely, Bradley Malin, Radja Badji, Samar K Dankar, and Khaled Shuaib. "Dynamic-informed consent: A potential solution for ethical dilemmas in population sequencing initiatives". In: *Computational and structural biotechnology journal* 18 (2020), pp. 913–921.

[70] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. "Personalized privacy assistants for the internet of things: Providing users with notice and choice". In: *IEEE Pervasive Computing* 17.3 (2018), pp. 35–46.

[71] Fred D Davis. "Perceived usefulness, perceived ease of use, and user acceptance of information technology". In: *MIS quarterly* (1989), pp. 319–340.

[72] Terry C Davis, Hans J Berkel, Randall F Holcombe, Sumona Pramanik, and Stephen G Divers. "Informed consent for clinical trials: a comparative study of standard versus simplified forms". In: *JNCI: Journal of the National Cancer Institute* 90.9 (1998), pp. 668–674.

[73] José Luis De la Vara and Juan Sánchez. "Improving requirements analysis through business process modelling: A participative approach". In: *International Conference on Business Information Systems*. Springer. 2008, pp. 165–176.

[74] Evelien De Sutter, Janos Meszaros, Pascal Borry, and Isabelle Huys. "Digitizing the informed consent process: a review of the regulatory landscape in the European Union". In: *Frontiers in Medicine* 9 (2022), p. 906448.

[75] Raymond De Vries and Leslie Rott. "Bioethics as Missionary Work: The Export of Western Ethics to Developing Countries". In: *Bioethics Around the Globe*. Ed. by Catherine Myser. 1st ed. Oxford University PressNew York, 2011-06, pp. 3–18.

[76] Catherine D DeAngelis and Phil B Fontanarosa. "Impugning the integrity of medical science: the adverse effects of industry influence". In: *Jama* 299.15 (2008), pp. 1833–1835.

[77] Laurens Debackere, Pieter Colpaert, Ruben Taelman, and Ruben Verborgh. "A Policy-Oriented Architecture for Enforcing Consent in Solid". In: *Companion Proceedings of the Web Conference 2022*. Lyon France: ACM, 2022-04, pp. 516–524.

[78] Christophe Debruyne, Jonathan Riggio, Olga De Troyer, and Declan O'Sullivan. "An Ontology for Representing and Annotating Data Flows to Facilitate Compliance Verification". In: *2019 13th International Conference on Research Challenges in Information Science (RCIS)*. IEEE. 2019, pp. 1–6.

[79] Zhaohua Deng, Ziying Hong, Cong Ren, Wei Zhang, Fei Xiang, et al. "What predicts patients' adoption intention toward mHealth services in China: empirical study". In: *JMIR mHealth and uHealth* 6.8 (2018), e9316.

[80] Neal W Dickert and Jeremy Sugarman. "Community consultation: not the problem—an important part of the solution". In: *The American Journal of Bioethics* 6.3 (2006), pp. 26–28.

[81] Ahmet Dikici, Oktay Turetken, and Onur Demirors. "Factors influencing the understandability of process models: A systematic literature review". In: *Information and Software Technology* 93 (2018), pp. 112–129.

[82] Xengie Doan, Annika Selzer, Arianna Rossi, Wilhelmina Maria Botes, and Gabriele Lenzini. "Context, Prioritization, and Unexpectedness: Factors Influencing User Attitudes About Infographic and Comic Consent". In: *Companion Proceedings of the Web Conference 2022*. 2022, pp. 534–545.

[83] Robert H Dolin, Liora Alschuler, Calvin Beebe, Paul V Biron, Sandra Lee Boyer, Daniel Essin, Elliot Kimber, Tom Lincoln, and John E Mattison. "The HL7 clinical document architecture". In: *Journal of the American Medical Informatics Association* 8.6 (2001), pp. 552–569.

[84] Edward S Dove, Susan E Kelly, Federica Lucivero, Mavis Machirori, Sandi Dheensa, and Barbara Prainsack. "Beyond individualism: Is there a place for relational autonomy in clinical practice and research?" In: *Clinical ethics* 12.3 (2017), pp. 150–165.

[85] Douglas Edwards. *I'm feeling lucky: The confessions of Google employee number 59*. Houghton Mifflin Harcourt, 2011.

[86] Perihan Elif Ekmekci and Berna Arda. "Interculturalism and informed consent: Respecting cultural differences without breaching human rights". In: *Cultura* 14.2 (2017), pp. 159–172.

[87] Mostafa Al-Emran and Andrina Granić. "Is it still valid or outdated? A bibliometric analysis of the technology acceptance model and its applications from 2010 to 2020". In: *Recent advances in technology acceptance models and theories* (2021), pp. 1–12.

[88] Bart Engelen. "Ethical criteria for health-promoting nudges: a case-by-case analysis". In: *The American journal of bioethics* 19.5 (2019), pp. 48–59.

[89] Yaniv Erlich, Tal Shor, Itsik Pe'er, and Shai Carmi. "Identity inference of genomic data using long-range familial searches". In: *Science* 362.6415 (2018), pp. 690–694.

[90] Council of Europe. *Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine*. Council of Europe, 1997. URL: https://rm.coe.int/168007cf98.

[91] European Commission DG Research & Innovation. *Ethics and data protection*. 2021-07. URL: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection%5C_he%5C_en.pdf.

[92] General Court of the European Union. *Case T-557/20, SRB v EDPS*. (Accessed on 11/06/2024). 2023-04. URL: https://curia.europa.eu/juris/document/document.jsf;jsessionid=6DEB111DC20DDDE1398DB9DE2787D763?text=&docid=272910&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=3454095.

[93] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. "Large-scale readability analysis of privacy policies". In: *Proceedings of the international conference on web intelligence*. Association for Computing Machinery, 2017, pp. 18–25.

[94] Matthew E Falagas, Ioanna P Korbila, Konstantina P Giannopoulou, Barbara K Kondilis, and George Peppas. "Informed consent: how much and what do patients understand?" In: *The American Journal of Surgery* 198.3 (2009-09), pp. 420–435. ISSN: 00029610.

[95] Rebecca A Ferrer, Jennifer Tehan Stanley, Kaitlin Graff, William MP Klein, Nina Goodman, Wendy L Nelson, and Silvia Salazar. "The effect of emotion on visual attention to information and decision making in the context of informed consent process for clinical trials". In: *Journal of Behavioral Decision Making* 29.2-3 (2016), pp. 245–253.

[96] Cristina Ferrer-Albero and Javier Díez-Domingo. "Does a comic style informed assent form improve comprehension for minors participating in clinical trials?" In: *Clinical Ethics* 16.1 (2021), pp. 37–45.

[97] Francesca Ferrucci, Manuele Jorio, Stefano Marci, Antonia Bezenchek, Giulia Diella, Cinzia Nulli, Ferdinando Miranda, Guido Castelli-Gattinara, et al. "A web-based application for complex health care populations: user-centered design approach". In: *JMIR human factors* 8.1 (2021), e18587.

[98] Michèle Finck and Frank Pallas. "They who must not be identified—distinguishing personal from non-personal data under the GDPR". In: *International Data Privacy Law* 10.1 (2020), pp. 11–36.

[99] Luciano Floridi. "Information ethics: On the philosophical foundation of computer ethics". In: *Ethics and information technology* 1.1 (1999), pp. 33–52.

[100] Maggie Fox. *Drug giant Glaxo teams up with DNA testing company 23andMe*. https://www.nbcnews.com/health/health-news/drug-giant-glaxo-teams-dna-testing-company-23andme-n894531. (Accessed on 02/05/2023)). 2018.

[101] Charles Fried. "Privacy [a moral analysis]". In: *Philosophical Dimensions of Privacy: An Anthology*. Ed. by Ferdinand David Schoeman. Cambridge University Press, 1984, pp. 203–222.

[102] Julie Frizzo-Barker, Peter A Chow-White, Anita Charters, and Dung Ha. "Genomic big data and privacy: challenges and opportunities for precision medicine". In: *Computer Supported Cooperative Work (CSCW)* 25 (2016), pp. 115–136.

[103] Agomoni Ganguli-Mitra. "Collective consent". In: *Ethical Issues in Governing Biobanks*. Routledge, 2016, pp. 121–130.

[104] Samual A Garner and Jiyeon Kim. "The privacy risks of direct-to-consumer genetic testing: A case study of 23andMe and Ancestry". In: *Wash. UL Rev.* 96 (2018), p. 1219.

[105] Nanibaa'A Garrison. "Genomic justice for Native Americans: impact of the Havasupai case on genetic research". In: *Science, Technology, & Human Values* 38.2 (2013), pp. 201–223.

[106] Nanibaa'A Garrison and Amy L Non. "Direct-to-consumer genomics companies should provide guidance to their customers on (not) sharing personal genomic information". In: *The American Journal of Bioethics* 14.11 (2014), pp. 55–57.

[107] Ruth Gavison. "Privacy and the Limits of Law". In: *The Yale law journal* 89.3 (1980), pp. 421–471.

[108] Eugenijus Gefenas, J Lekstutiene, V Lukaseviciene, M Hartlev, M Mourby, and KÓ Cathaoir. "Controversies between regulations of research ethics and protection of personal data: informed consent at a cross-road". In: *Medicine, Health Care and Philosophy* (2022), pp. 1–8.

[109] Scientific General Conference of the United Nations Educational and Cultural Organization. *Universal Declaration on the Human Genome and Human Rights*. (Accessed on 10/07/2023). URL: https://www.unesco.org/en/legal-affairs/universal-declaration-human-genome-and-human-rights.

[110] Association for German Supervisory Authorities. *Guidance on the interplay between recital 33 and the definition of consent in the GDPR*. 2019-04. URL: https://www.datenschutzkonferenz-%20online.de/media/dskb/20190405%5C_auslegung%5C_bestimmte%5C_bereiche%5C_wiss%5C_forschung.pdf.

[111] Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M Schwartz, and Steven Woloshin. "Helping Doctors and Patients Make Sense of Health Statistics". In: *Psychological Science in the Public Interest* 8.2 (2007-11), pp. 53–96. ISSN: 1529-1006.

[112] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. "Dark and Bright Patterns in Cookie Consent Requests". In: *Journal of Digital Social Research* 3.11 (2021-02), pp. 1–38. ISSN: 2003-1998.

[113] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. "The dark (patterns) side of UX design". In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. 2018, pp. 1–14.

[114] Henry T Greely. "Informed consent and other ethical issues in human population genetics". In: *Annual Review of Genetics* 35.1 (2001), pp. 785–800.

[115] Louise Gröndahl. *Public knowledge of digital cookies: Exploring the design of cookie consent forms*. 2020.

[116] Johan van de Gronden and Marc Veenbrink. "EHDS and Free Movement of Patients: What EU Intervention is Needed?" In: *European Journal of Health Law* 31.3 (2024), pp. 249–284.

[117] Petronella Grootens-Wiegers, Martine C de Vries, Mara M van Beusekom, Laura van Dijck, and Jos M van den Broek. "Comic strips help children understand medical research: targeting the informed consent procedure to children's needs". In: *Patient education and counseling* 98.4 (2015), pp. 518–524.

[118] Elias Grünewald and Frank Pallas. "TILT: a GDPR-aligned transparency information language and toolkit for practical privacy engineering". In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. 2021, pp. 636–646.

[119] Giancarlo Guizzardi and Henderik A Proper. "On understanding the value of domain modeling". In: *CEUR workshop proceedings*. Vol. 2835. Rheinisch Westfälische Technische Hochschule. 2021, pp. 51–62.

[120] Melissa Gymrek, Amy L McGuire, David Golan, Eran Halperin, and Yaniv Erlich. "Identifying personal genomes by surname inference". In: *Science* 339.6117 (2013), pp. 321–324.

[121] Helena Haapio, Robert deRooy, and Thomas D Barton. "New Contract Genres". In: *Data Protection/LegalTech. Proceedings of the 21th International Legal Informatics Symposium IRIS*. Bern: Editions Weblaw, 2018, pp. 455–460.

[122] Matilda A Haas, Evanthia O Madelli, Rosie Brown, Megan Prictor, and Tiffany Boughtwood. "Evaluation of CTRL: a web application for dynamic consent and engagement with individuals involved in a cardiovascular genetic disorders cohort". In: *European Journal of Human Genetics* 32.1 (2024), pp. 61–68.

[123] Matilda A Haas, Harriet Teare, Megan Prictor, Gabi Ceregra, Miranda E Vidgen, David Bunker, Jane Kaye, and Tiffany Boughtwood. "'CTRL': an online, Dynamic Consent and participant engagement platform working towards solving the complexities of consent in genomic research". In: *European Journal of Human Genetics* 29.4 (2021), pp. 687–698.

[124] Tobias Haeusermann, Marta Fadda, Alessandro Blasimme, Bastian Greshake Tzovaras, and Effy Vayena. "Genes wide open: Data sharing and the social gradient of genomic privacy". In: *AJOB Empirical Bioethics* 9.4 (2018), pp. 207–221.

[125] Dara Hallinan. "Broad consent under the GDPR: an optimistic perspective on a bright future". In: *Life Sciences, Society and Policy* 16.1 (2020-01), p. 1.

[126] Nao Hamakawa, Atsushi Kogetsu, Moeko Isono, Chisato Yamasaki, Shirou Manabe, Toshihiro Takeda, Kazumasa Iwamoto, Tomoya Kubota, Joe Barrett, Nathanael Gray, et al. "The practice of active patient involvement in rare disease research using ICT: experiences and lessons from the RUDY JAPAN project". In: *Research Involvement and Engagement* 7 (2021), pp. 1–15.

[127] Mats G Hansson, Joakim Dillner, Claus R Bartram, Joyce A Carlson, and Gert Helgesson. "Should donors be allowed to give broad consent to future biobank research?" In: *The lancet oncology* 7.3 (2006), pp. 266–269.

[128] Jack Hardinges, Peter Wells, Alex Blandford, Jeni Tennison, and Anna Scott. *Data trusts: lessons from three pilots*. 2019. URL: https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit?usp=sharing.

[129] Eszter Hargittai and Alice Marwick. ""What can I really do?" Explaining the privacy paradox with online apathy". In: *International journal of communication* 10 (2016), p. 21.

[130] Ashley Hayward, Erynne Sjoblom, Stephanie Sinclair, and Jaime Cidro. "A new era of indigenous research: Community-based indigenous research ethics protocols in Canada". In: *Journal of Empirical Research on Human Research Ethics* 16.4 (2021), pp. 403–417.

[131] James W Hazel and Christopher Slobogin. "Who knows what, and when: a survey of the privacy policies proffered by US direct-to-consumer genetic testing companies". In: *Cornell JL & Pub. Pol'y* 28 (2018), p. 35.

[132] Richard Healey. "From individual to collective consent: The case of Indigenous Peoples and UNDRIP". In: *International Journal on Minority and Group Rights* 27.2 (2019), pp. 251–269.

[133] Bert Heinrichs. "Myth or magic? Towards a revised theory of informed consent in medical research". In: *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*. Vol. 44. 1. Oxford University Press US. 2019, pp. 33–49.

[134] Niels Hendriks, Frederik Truyen, and Erik Duval. "Designing with dementia: Guidelines for participatory design together with persons with dementia". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2013, pp. 649–666.

[135] Brianna Hoffner, Susan Bauer-Wu, Suzanne Hitchcock-Bryan, Mark Powell, Andrew Wolanski, and Steven Joffe. ""Entering a clinical trial: Is it right for you?" A randomized study of the clinical trials video and its impact on the informed consent process". In: *Cancer* 118.7 (2012), pp. 1877–1883.

[136] Richard J Holden and Ben-Tzion Karsh. "The technology acceptance model: its past and its future in health care". In: *Journal of biomedical informatics* 43.1 (2010), pp. 159–172.

[137] Hsiao-Ying Huang and Masooda Bashir. "Direct-to-Consumer genetic testing: Contextual privacy predicament". In: *Proceedings of the Association for Information Science and Technology* 52.1 (2015), pp. 1–10.

[138] Maui Hudson. "Think globally, act locally: Collective consent and the ethics of knowledge production". In: *International Social Science Journal* 60.195 (2009), pp. 125–133.

[139] Soheil Human, Harshvardhan J Pandit, Victor Morel, Cristiana Santos, Martin Degeling, Arianna Rossi, Wilhelmina Botes, Vitor Jesus, and Irene Kamara. "Data Protection and Consenting Communication Mechanisms: Current Open Proposals and Challenges". In: *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2022, pp. 231–239.

[140] Duha Ibdah, Nada Lachtar, Satya Meenakshi Raparthi, and Anys Bacha. ""Why Should I Read the Privacy Policy, I Just Need the Service": A Study on Attitudes and Perceptions Toward Privacy Policies". In: *IEEE access* 9 (2021), pp. 166465–166487.

[141] Council of Indigenous Peoples. *The Indigenous Peoples Basic Law*. (Accessed on 01/06/2024). 2018. URL: https://law.moj.gov.tw/eng/lawclass/lawall.aspx?pcode=d0130003.

[142] Desislava Ivanova and Panagiotis Katsaounis. "Real-Time Dynamic Tiered e-Consent: A Novel Tool for Patients' Engagement and Common Ontology System for the Management of Medical Data". In: *Innovations in Digital Health, Diagnostics, and Biomarkers* 1.2 (2021), pp. 45–49.

[143] Harshvardhan J Pandit, Vitor Jesus, Shankar Ammai, Mark Lizar, and Salvatore D'Agostino. "Role of Identity, Identification, and Receipts for Consent". In: *Open Identity Summit 2021* (2021).

[144] Carlos Jensen and Colin Potts. "Privacy policies as decision-making tools: an evaluation of online privacy notices". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 2004, pp. 471–478.

[145] Holly B Jimison, Paul P Sher, Richard Appleyard, and Yvonne LeVernois. "The use of multimedia in the informed consent process". In: *Journal of the American Medical Informatics Association* 5.3 (1998), pp. 245–256.

[146] Jennifer Viberg Johansson, Nisha Shah, Eik Haraldsdóttir, Heidi Beate Bentzen, Sarah Coy, Jane Kaye, Deborah Mascalzoni, and Jorien Veldwijk. "Governance mechanisms for sharing of health data: An approach towards selecting attributes for complex discrete choice experiment studies". In: *Technology in Society* 66 (2021), p. 101625.

[147] Yann Joly, Charles Dupras, Miriam Pinkesz, Stacey A Tovino, and Mark A Rothstein. "Looking beyond GINA: policy approaches to address genetic discrimination". In: *Annual Review of Genomics and Human Genetics* 21 (2020), pp. 491–507.

[148] Yann Joly, Ida Ngueng Feze, and Jacques Simard. "Genetic discrimination and life insurance: a systematic review of the evidence". In: *BMC medicine* 11 (2013), pp. 1–15.

[149] Connor Jones. *Latest 23andMe data claim would take leaked records to 5M*. 2023. URL: https://www.theregister.com/2023/10/19/latest%5C_23andme_data%5C_leak%5C_takes/.

[150] Eric T Juengst. "Group identity and human diversity: Keeping biology straight from culture". In: *The American Journal of Human Genetics* 63.3 (1998), pp. 673–677.

[151] Georgios Kaissis, Alexander Ziller, Jonathan Passerat-Palmbach, Théo Ryffel, Dmitrii Usynin, Andrew Trask, Ionésio Lima Jr, Jason Mancuso, Friederike Jungmann, Marc-Matthias Steinborn, et al. "End-to-end privacy preserving deep learning on multi-institutional medical imaging". In: *Nature Machine Intelligence* 3.6 (2021), pp. 473–484.

[152] Dipak Kalra, Thomas Beale, and Sam Heard. "The openEHR foundation". In: *Studies in health technology and informatics* 115 (2005), pp. 153–173.

[153] Georgios Kampanos and Siamak F Shahandashti. "Accept All: The Landscape of Cookie Banners in Greece and the UK". In: *IFIP International Conference on ICT Systems Security and Privacy Protection*. Copenhagen: Springer, 2021, pp. 213–227.

[154] Nesibe Kantar and Terrell Ward Bynum. "Global ethics for the digital age – flourishing ethics". In: *Journal of Information, Communication and Ethics in Society* 19.3 (2021), pp. 329–344.

[155] Elena Karahanna, Detmar W Straub, and Norman L Chervany. "Information technology adoption across time: A cross-sectional comparison of pre-adoption and post-adoption beliefs". In: *MIS quarterly* (1999), pp. 183–213.

[156] Farzaneh Karegar, John Sören Pettersson, and Simone Fischer-Hübner. "The dilemma of user engagement in privacy notices: Effects of interaction modes and habituation on user attention". In: *ACM Transactions on Privacy and Security (TOPS)* 23.1 (2020), pp. 1–38.

[157] Jane Kaye, Catherine Heeney, Naomi Hawkins, Jantina de Vries, and Paula Boddington. "Data sharing in genomics — re-shaping scientific practice". In: *Nature Reviews Genetics* 10.5 (2009-05), pp. 331–335. (Visited on 2022-04-04).

[158] Jane Kaye, Edgar A Whitley, David Lund, Michael Morrison, Harriet Teare, and Karen Melham. "Dynamic consent: a patient interface for twenty-first century research networks". In: *European journal of human genetics* 23.2 (2015), pp. 141–146.

[159] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. "Privacy as part of the app decision-making process". In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. Ed. by Sussane Bødker, Steven Brewster, Patrick Baudisch, Michel Beaudouin-Lafon, and Wendy E Mackay. Paris: ACM, 2013, pp. 3393–3402.

[160] Hyeoneui Kim, Elizabeth Bell, Jihoon Kim, Amy Sitapati, Joe Ramsdell, Claudiu Farcas, Dexter Friedman, Stephanie Feudjio Feupe, and Lucila Ohno-Machado. "iCONCUR: informed consent for clinical data and bio-sample use for research". In: *Journal of the American Medical Informatics Association* 24.2 (2017), pp. 380–387.

[161] Jennifer King. ""Becoming Part of Something Bigger" Direct to Consumer Genetic Testing, Privacy, and Personal Disclosure". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–33.

[162] Helen Michelle Kirkby, Melanie Calvert, Heather Draper, Thomas Keeley, and Sue Wilson. "What potential research participants want to know about research: a systematic review". In: *BMJ Open* 2.3 (2012-01), e000509. ISSN: 2044-6055, 2044-6055.

[163] Sabrina Kirrane, Javier D Fernández, Wouter Dullaert, Uros Milosevic, Axel Polleres, Piero A Bonatti, Rigo Wenning, Olha Drozd, and Philip Raschke. "A scalable consent, transparency and compliance architecture". In: *The Semantic Web: ESWC 2018 Satellite Events: ESWC 2018 Satellite Events, Heraklion, Crete, Greece, June 3-7, 2018, Revised Selected Papers 15*. Springer. 2018, pp. 131–136.

[164] Craig M Klugman, Laura B Dunn, Jack Schwartz, and I Glenn Cohen. "The Ethics of Smart Pills and Self-Acting Devices: Autonomy, Truth-Telling, and Trust at the Dawn of Digital Medicine". In: *American Journal of Bioethics* 18.9 (2018), pp. 38–47. ISSN: 15360075.

[165] Craig M Klugman and Hector F Rodriguez. "Ethics of familial genetic genealogy: solving crimes at the cost of privacy". In: *DePaul Journal of Health Care Law* 22 (2021), p. 67.

[166] Bartha Maria Knoppers and Kristina Kekesi-Lafrance. "The genetic family as patient?" In: *The American Journal of Bioethics* 20.6 (2020), pp. 77–80.

[167] Stephanie A Kraft, Melissa Constantine, David Magnus, Kathryn M Porter, Sandra Soo-Jin Lee, Michael Green, Nancy E Kass, Benjamin S Wilfond, and Mildred K Cho. "A randomized study of multimedia informational aids for research on medical practices: Implications for informed consent". In: *Clinical Trials* 14.1 (2017), pp. 94–102.

[168] Jan Kubovy, Dagmar Auer, and Josef Küng. "Behavior-based Decomposition of BPMN 2.0 Control Flow." In: *ICEIS (3)*. 2014, pp. 263–271.

[169] Juergen Kuehling, Florian Sackmann, and Hilmar Schneider. *Datenschutzrechtliche Dimensionen Datentreuhänder*. 2020-11.

[170] Sari Kujala. "User involvement: a review of the benefits and challenges". In: *Behaviour & Information Technology* 22.1 (2003), pp. 1–16.

[171] Taner Kuru. "Genetic data: The Achilles' heel of the GDPR?" In: *European Data Protection Law Review* 7 (2021), p. 45.

[172] Taner Kuru and Iñigo de Miguel Beriain. "Your genetic data is my genetic data: Unveiling another enforcement issue of the GDPR". In: *Computer Law & Security Review* 47 (2022), p. 105752.

[173] Lauren F Laker, Craig M Froehle, Jaime B Windeler, and Christopher John Lindsell. "Quality and efficiency of the clinical decision-making process: Information overload and emphasis framing". In: *Production and Operations Management* 27.12 (2018), pp. 2213–2225.

[174] Graeme T Laurie. "Challenging medical-legal norms: the role of autonomy, confidentiality, and privacy in protecting individual and familial group rights in genetic information". In: *Journal of Legal Medicine* 22.1 (2001), pp. 1–54.

[175] Jonathan Lawson, Moran N Cabili, Giselle Kerry, Tiffany Boughtwood, Adrian Thorogood, Pinar Alper, Sarion R Bowers, Rebecca R Boyles, Anthony J Brookes, Matthew Brush, et al. "The Data Use Ontology to streamline responsible access to human biomedical datasets". In: *Cell Genomics* 1.2 (2021).

[176] Younghwa Lee, Kenneth A Kozar, and Kai RT Larsen. "The technology acceptance model: Past, present, and future". In: *Communications of the Association for information systems* 12.1 (2003), p. 50.

[177] Lisa Soleymani Lehmann, David J Kaufman, Richard R Sharp, Tanya A Moreno, Joanna L Mountain, J Scott Roberts, and Robert C Green. "Navigating a research partnership between academia and industry to assess the impact of personalized genetic testing". In: *Genetics in medicine* 14.2 (2012), pp. 268–273.

[178] Henrik Leopold, Jan Mendling, and Oliver Günther. "Learning from quality issues of BPMN models from industry". In: *IEEE software* 33.4 (2015), pp. 26–33.

[179] Long Li. "A critical review of technology acceptance literature". In: *Referred Research Paper* 4 (2010), p. 2010.

[180] Qingchuan Li. "Healthcare at your fingertips: The acceptance and adoption of mobile medical treatment services among Chinese users". In: *International journal of environmental research and public health* 17.18 (2020), p. 6895.

[181] Charles W Lidz, Paul S Appelbaum, and Alan Meisel. "Two models of implementing informed consent". In: *Archives of Internal Medicine* 148.6 (1988), pp. 1385–1389.

[182] Fenghsiu Lin, Chin-Wei Liu, and I-Hung Kuo. "Moderating Effect of Perceived Usefulness on the Relationship between Ease of Use, Attitude toward Use and Actual System Use." In: *International Journal of Organizational Innovation* 5.3 (2013).

[183] Zhen Lin, Art B Owen, and Russ B Altman. "Genomic research and human subject privacy". In: *Science* 305.5681 (2004), pp. 183–183.

[184] Stephanie B Linek and Markus Huff. "Serious comics: a new approach for science communication and learning". In: *INTED2018 Proceedings*. Iated. 2018, pp. 3883–3890.

[185] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun Zhang, Norman Sadeh, Alessandro Acquisti, and Yuvraj Agarwal. "Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions". In: *Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS)*. Ed. by Mary Ellen Zurko, Sunny Consolvo, and Matthew Smith. Denver: USENIX, 2016, pp. 27–41.

[186] Andrew Lloyd, Paul Hayes, Peter RF Bell, and A Ross Naylor. "The role of risk and benefit perception in informed consent for surgery". In: *Medical decision making* 21.2 (2001), pp. 141–149.

[187] Alexandre Lodie and Cedric Lauradoux. "Is it Personal data? Solving the gordian knot of anonymisation". In: *Privacy Symposium 2024*. 2024.

[188] Erich H Loewy. "Families, communities, and making medical decisions". In: *The Journal of Clinical Ethics* 2.3 (1991), pp. 150–153.

[189] Paul Lopes. "Culture and stigma: Popular culture and the case of comic books". In: *Sociological Forum*. Vol. 21. Springer. 2006, pp. 387–414.

[190] Juniper L Lovato, Antoine Allard, Randall Harp, Jeremiah Onaolapo, and Laurent Hébert-Dufresne. "Limits of individual consent and models of distributed consent in online social networks". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022, pp. 2251–2262.

[191] Federica Lucivero and Karin R Jongsma. "A mobile revolution for healthcare? Setting the agenda for bioethics". In: *Journal of Medical Ethics* 44.10 (2018), pp. 685–689.

[192] Mary A Majumder, Christi J Guerrini, and Amy L McGuire. "Direct-to-consumer genetic testing: value and risk". In: *Annual Review of Medicine* 72 (2021), pp. 151–166.

[193] Alessandro Mantelero. "From group privacy to collective privacy: towards a new dimension of privacy and data protection in the big data era". In: *Group privacy: new challenges of data technologies* (2017), pp. 139–158.

[194] Alessandro Mantelero. "Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection". In: *Computer law & security review* 32.2 (2016), pp. 238–255.

[195] Elaine R Mardis. "A decade's perspective on DNA sequencing technology". In: *Nature* 470.7333 (2011), pp. 198–203.

[196] Luca Marelli, Marthe Stevens, Tamar Sharon, Ine Van Hoyweghen, Martin Boeckhout, Ilaria Colussi, Alexander Degelsegger-Márquez, Seliem El-Sayed, Klaus Hoeyer, Robin van Kessel, et al. "The European health data space: Too big to succeed?" In: *Health policy* 135 (2023), p. 104861.

[197] Francisco Martins, Dulce Domingos, and Daniel Vitoriano. "Automatic Decomposition of IoT Aware Business Processes with Data and Control Flow Distribution." In: *ICEIS (2)*. 2019, pp. 516–524.

[198] Alice E Marwick and Danah Boyd. "Networked privacy: How teenagers negotiate context in social media". In: *New media & society* 16.7 (2014), pp. 1051–1067.

[199] Deborah Mascalzoni, Roberto Melotti, Cristian Pattaro, Peter Paul Pramstaller, Martin Gögele, Alessandro De Grandi, and Roberta Biasiotto. "Ten years of dynamic consent in the CHRIS study: informed consent as a dynamic process". In: *European Journal of Human Genetics* 30.12 (2022), pp. 1391–1397.

[200] Célestin Matte, Nataliia Bielova, and Cristiana Santos. "Do cookie banners respect my choice?: Measuring legal compliance of banners from IAB Europe's transparency and consent framework". In: *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE. 2020, pp. 791–809.

[201] Alistair Mavin, Philip Wilkinson, Adrian Harwood, and Mark Novak. "Easy approach to requirements syntax (EARS)". In: *2009 17th IEEE International Requirements Engineering Conference*. IEEE. 2009, pp. 317–322.

[202] Viktor Mayer-Schönberger and Kenneth Cukier. *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt, 2013.

[203] Colleen M McBride, Deborah Bowen, Lawrence C Brody, Celeste M Condit, Robert T Croyle, Marta Gwinn, Muin J Khoury, Laura M Koehly, Bruce R Korf, Theresa M Marteau, Kenneth McLeroy, Kevin Patrick, and Thomas W Valente. "Future Health Applications of Genomics: Priorities for Communication, Behavioral, and Social Sciences Research". In: *American Journal of Preventive Medicine* 38.5 (2010-05), pp. 556–565.

[204] Tara McCurdie, Svetlena Taneva, Mark Casselman, Melanie Yeung, Cassie McDaniel, Wayne Ho, and Joseph Cafazzo. "mHealth consumer apps: the case for user-centered design". In: *Biomedical Instrumentation & Iechnology* 46.s2 (2012), pp. 49–56.

[205] Aleecia M McDonald and Lorrie Faith Cranor. "The cost of reading privacy policies". In: *Isjlp* 4 (2008), p. 543.

[206] Amy L McGuire, Jill M Oliver, Melody J Slashinski, Jennifer L Graves, Tao Wang, P Adam Kelly, William Fisher, Ching C Lau, John Goss, Mehmet Okcu, et al. "To share or not to share: a randomized trial of consent for data sharing in genome research". In: *Genetics in Medicine* 13.11 (2011), pp. 948–955.

[207] Thomas Mejtoft, Erik Frängsmyr, Ulrik Söderström, and Ole Norberg. "Deceptive Design: Cookie Consent and Manipulative Patterns". In: *34th Bled eConference Digital Support from Crisis to Progressive Change: Conference Proceedings*. Bled, Slovenia: University of Maribor Press, 2021, pp. 393–404.

[208] Uta Menges, Jonas Hielscher, Laura Kocksch, Annette Kluge, and M Angela Sasse. "Caring Not Scaring-An Evaluation of a Workshop to Train Apprentices as Security Champions". In: *Proceedings of the 2023 European Symposium on Usable Security*. 2023, pp. 237–252.

[209] Don Menn. "Multimedia in education". In: *PC world* (1993), pp. M52–m60.

[210] Christian Meske and Ireti Amojo. "Ethical Guidelines for the Construction of Digital Nudges". In: *53rd Hawaii International Conference on Systems Sciences (HICSS)*. 2020, pp. 3928–3937.

[211] Anna Middleton, Álvaro Mendes, Caroline M Benjamin, and Heidi Carmen Howard. "Direct-to-consumer genetic testing: where and how does genetic counseling fit?" In: *Personalized medicine* 14.3 (2017), pp. 249–257.

[212] Anna Middleton, Emilia Niemiec, Barbara Prainsack, Jason Bobe, Lauren Farley, Claire Steed, James Smith, Paul Bevan, Natasha Bonhomme, Erika Kleiderman, et al. "'Your DNA, Your Say': global survey gathering attitudes toward genomics: design, delivery and methods". In: *Personalized medicine* 15.04 (2018), pp. 311–318.

[213] Rasmus Bjerregaard Mikkelsen, Mickey Gjerris, Gunhild Waldemar, and Peter Sandøe. "Broad consent for biobanks is best – provided it is also deep". In: *BMC Medical Ethics* 20.1 (2019-10), p. 71.

[214] Jusaku Minari, Harriet Teare, Colin Mitchell, Jane Kaye, and Kazuto Kato. "The emerging need for family-centric initiatives for obtaining consent in personal genome research". In: *Genome Medicine* 6.12 (2014-12), p. 118.

[215] Marco Casassa Mont, Siani Pearson, Gina Kounga, Yun Shen, and Pete Bramhall. "On the management of consent and revocation in enterprises: setting the context". In: *HP Laboratories, Technical Report HPL-2009-49* 11 (2009).

[216] Wanda Montalvo and Elaine Larson. "Participant Comprehension of Research for Which They Volunteer: A Systematic Review". In: *Journal of Nursing Scholarship* 46.6 (2014), pp. 423–431. ISSN: 1547-5069.

[217] Daniel L Moody. "The method evaluation model: a theoretical model for validating information systems design methods". In: *New Paradigms in Organizations, Markets and Society: Proceedings of the 11th European Conference on Information Systems (ECIS 2003)*. Ed. by Claudio Ciborra, Riccardo Mercurio, Marco De Marco, Marcello Martinez, and Andrea Carignani. European Conference on Information Systems 2003, ECIS 2003 ; Conference date: 16-06-2003 Through 21-06-2003. Department of Information Systems, London School of Economics, 2003, pp. 1–17.

[218] James H Moor. "Just consequentialism and computing". In: *Ethics and Information Technology* 1.1 (1999), pp. 65–69.

[219] James H Moor. "Towards a theory of privacy in the information age". In: *ACM Sigcas Computers and Society* 27.3 (1997), pp. 27–32.

[220] Victor Morel, Cristiana Santos, Viktor Fredholm, and Adam Thunberg. "Legitimate interest is the new consent – large-scale measurement and legal compliance of IAB Europe TCF paywalls". In: *Proceedings of the 22nd Workshop on Privacy in the Electronic Society*. 2023, pp. 153–158.

[221] Nchangwi Syntia Munung, Patricia Marshall, Megan Campbell, Katherine Littler, Francis Masiye, Odile Ouwe-Missi-Oukem-Boyer, Janet Seeley, DJ Stein, Paulina Tindana, and Jantina De Vries. "Obtaining informed consent for genomics research in Africa: analysis of H3Africa consent documents". In: *Journal of Medical Ethics* 42.2 (2016), pp. 132–137.

[222] Claire Murphy, Sarah Sturm, Meghan Juliana McKenna, and Kelly E Ormond. "The right not to know: Non-disclosure of primary genetic test results and genetic counselors' response". In: *Journal of Genetic Counseling* 33.4 (2024), pp. 875–887.

[223] Jean Nehme, Ussamah El-Khani, Andre Chow, Sherif Hakky, Ahmed R Ahmed, and Sanjay Purkayastha. "The use of multimedia consent programs for surgical procedures: a systematic review". In: *Surgical innovation* 20.1 (2013), pp. 13–23.

[224] Emilia Niemiec, Danya F Vears, Pascal Borry, and Heidi Carmen Howard. "Readability of informed consent forms for whole-exome and whole-genome sequencing". In: *Journal of Community Genetics* 9.2 (2018-04), pp. 143–151.

[225] Helen Nissenbaum. "A contextual approach to privacy online". In: *Daedalus* 140.4 (2011), pp. 32–48.

[226] Helen Nissenbaum. "Privacy in context". In: *Privacy in Context*. Stanford University Press, 2009.

[227] Ilena M Norton and Spero M Manson. "Research in American Indian and Alaska Native communities: navigating the cultural universe of values and process". In: *Journal of consulting and clinical psychology* 64.5 (1996), p. 856.

[228] NortonLifeLock. *MyHeritage data breach exposes info of more than 92 million users*. (Accessed on 14/11/2023). 2018. URL: https://us.norton.com/blog/emerging-threats/myheritage-data-breach-exposes-info-of-more-than-92-million-user.

[229] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence". In: *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020, pp. 1–13.

[230] Dale R Nyholt, Chang-En Yu, and Peter M Visscher. "On Jim Watson's APOE status: genetic information is hard to hide". In: *European Journal of Human Genetics* 17.2 (2009), pp. 147–149.

[231] Anna Gunhild Nysetvold and John Krogstie. "Assessing business process modeling languages using a generic quality framework". In: *Advanced Topics in Database Research, Volume 5*. IGI Global, 2006, pp. 79–93.

[232] Onora O'Neill. *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press, 2002-04.

[233] Yousra Odeh. "BPMN in engineering software requirements: an introductory brief guide". In: *Proceedings of the 9th International Conference on Information Management and Engineering*. 2017, pp. 11–16.

[234] Philipp Offermann, Olga Levina, Marten Schönherr, and Udo Bub. "Outline of a design science research process". In: *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology*. 2009, pp. 1–11.

[235] Chun Ouyang, Marlon Dumas, Arthur HM Ter Hofstede, and Wil MP Van der Aalst. "From BPMN process models to BPEL web services". In: *2006 IEEE International Conference on Web Services (ICWS'06)*. IEEE. 2006, pp. 285–292.

[236] Barton W Palmer, Nicole M Lanouette, and Dilip V Jeste. "Effectiveness of multimedia aids to enhance comprehension during research consent: A systematic review". In: *Irb* 34.6 (2012), p. 1.

[237] Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. "Legal ontology for modelling GDPR concepts and norms". In: *Legal Knowledge and Information Systems*. IOS Press, 2018, pp. 91–100.

[238] Harshvardhan J Pandit and Beatriz Esteves. "Enhancing Data Use Ontology (DUO) for health-data sharing by extending it with ODRL and DPV". In: *Semantic Web* Preprint (2023), pp. 1–26.

[239] Harshvardhan J Pandit and Georg Philip Krog. "Comparison of notice requirements for consent between ISO/IEC 29184: 2020 and General Data Protection Regulation". In: *Journal of Data Protection & Privacy* 4.2 (2021), pp. 193–204.

[240] Harshvardhan J Pandit, Axel Polleres, Bert Bos, Rob Brennan, Bud Bruegger, Fajar J Ekaputra, Javier D Fernández, Roghaiyeh Gachpaz Hamed, Elmar Kiesling, Mark Lizar, et al. "Creating a vocabulary for data privacy: The first-year report of data privacy vocabularies and controls community group (DPVCG)". In: *On the Move to Meaningful Internet Systems: OTM 2019 Conferences: Confederated International Conferences: CoopIS, ODBASE, C&TC 2019*. Springer. 2019, pp. 714–730.

[241] European Parliament. *European Parliament legislative resolution of 24 April 2024 on the proposal for a regulation of the European Parliament and of the Council on the European Health Data Space (COM(2022)0197 – C9-0167/2022 – 2022/0140(COD))*. (Accessed on 05/10/2024). 2024. URL: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0331_EN.html.

[242] Article 29 Data Protection Working Party. *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*. 2017-10. URL: https://ec.europa.eu/newsroom/just/document.cfm?doc%5C_id=47711.

[243] Article 29 Data Protection Working Party. *Guidelines on Transparency under Regulation 2016/679, 17/EN WP260 rev.01*. (Accessed on 05/06/2024). 2018-04. URL: https://ec.europa.eu/newsroom/article29/redirection/document/51025.

[244] Article 29 Working Party. *12178/03/EN WP 91 Working Document on Genetic Data*. (Accessed on 05/06/2024). 2004. URL: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2004/wp91%5C_en.pdf.

[245] Article 29 Working Party. *Guidelines 05/2020 on consent under Regulation 2016/679*. (Accessed on 05/06/2024). 2020-05. URL: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_en.

[246] Harold Pashler, Mark McDaniel, Doug Rohrer, and Robert Bjork. "Learning Styles: Concepts and Evidence". In: *Psychological Science in the Public Interest* 9.3 (2008-12), pp. 105–119. ISSN: 1529-1006.

[247] Stefania Passera et al. *Beyond the wall of contract text-visualizing contracts to foster understanding and collaboration within and across organizations*. Aalto University, 2017.

[248] Cristian Pattaro, Martin Gögele, Deborah Mascalzoni, Roberto Melotti, Christine Schwienbacher, Alessandro De Grandi, Luisa Foco, Yuri D'elia, Barbara Linder, Christian Fuchsberger, et al. "The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results". In: *Journal of Translational Medicine* 13 (2015), pp. 1–16.

[249] Dorian Peters, Rafael A Calvo, and Richard M Ryan. "Designing for motivation, engagement and wellbeing in digital experience". In: *Frontiers in psychology* 9 (2018), p. 797.

[250] Nathalie Piquemal. "Free and informed consent in research involving Native American communities". In: *American Indian Culture and Research Journal* 25.1 (2001).

[251] Robert Plutchik. "The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice". In: *American Scientist* 89.4 (2001), pp. 344–350. ISSN: 0003-0996.

[252] Marie Potel-Seville and Elisabeth Talbourdet. "Empowering children to understand and exercise their personal data rights". In: *Legal design perspectives: theoretical and practical insights from the field*. Italy: Ledizioni, 2021, pp. 253–276.

[253] Megan Prictor, Sharon Huebner, Harriet JA Teare, Luke Burchill, and Jane Kaye. "Australian Aboriginal and Torres Strait Islander collections of genetic heritage: the legal, ethical and practical considerations of a dynamic consent approach to decision making". In: *Journal of Law, Medicine & Ethics* 48.1 (2020), pp. 205–217.

[254] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *Belmont Report*. Text. (Accessed on 25/05/2023). 1979-04. URL: https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html.

[255] Garante per la Protezione dei Dati Personali. *Dati inerenti allo stato di salute - dati genetici, Cittadini e socie-tà dell'informazione*. it. (Accessed on 15/05/2023). 1999. URL: https://www.garanteprivacy.it/documents/10160/10704/996886.

[256] John Pruitt and Jonathan Grudin. "Personas: practice and theory". In: *Proceedings of the 2003 conference on Designing for user experiences*. 2003, pp. 1–15.

[257] Luise Pufahl, Francesca Zerbato, Barbara Weber, and Ingo Weber. "BPMN in healthcare: Challenges and best practices". In: *Information Systems* 107 (2022), p. 102013.

[258] Pille Pullonen, Jake Tom, Raimundas Matulevičius, and Aivo Toots. "Privacy-enhanced BPMN: enabling data privacy analysis in business processes models". In: *Software and Systems Modeling* 18 (2019), pp. 3235–3264.

[259] Paul Quinn, Erika Ellyne, and Cong Yao. "Will the GDPR Restrain Health Data Access Bodies Under the European Health Data Space (EHDS)?" In: *Computer Law & Security Review* 54 (2024), p. 105993.

[260] James Rachels. "Why privacy is important". In: *Philosophy & Public Affairs* (1975), pp. 323–333.

[261] Maryam Radgui, Rajaa Saidi, and Salma Mouline. "Extracting reusable fragments from business process using BPMN". In: *Second International Conference on the Innovative Computing Technology (INTECH 2012)*. IEEE. 2012, pp. 424–429.

[262] Maryam Radgui, Rajaa Saidi, Salma Mouline, M Radgui, R Saidi, and S Mouline. "A pattern for the decomposition of business processes". In: *Softw. Eng. Databases Expert Syst* (2012).

[263] Natalie Ram. "Genetic privacy after Carpenter". In: *Virginia Law Review* 105.7 (2019), pp. 1357–1425.

[264] S Raquel Ramos. "User-centered design, experience, and usability of an electronic consent user interface to facilitate informed decision-making in an HIV clinic". In: *CIN: Computers, Informatics, Nursing* 35.11 (2017), pp. 556–564.

[265] Henriette Rau, Lars Geidel, Martin Bialke, Arne Blumentritt, Martin Langanke, Wenke Liedtke, Sandra Pasewald, Dana Stahl, Thomas Bahls, Christian Maier, et al. "The generic Informed Consent Service gICS®: implementation and benefits of a modular consent software tool to master the challenge of electronic consent management in research". In: *Journal of Translational Medicine* 18 (2020), pp. 1–12.

[266] Jan Recker. "Opportunities and constraints: the current struggle with BPMN". In: *Business Process Management Journal* 16.1 (2010), pp. 181–201.

[267] Antonio Regalado. *More than 26 million people have taken an at-home ancestry test*. 2019-02. URL: https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/.

[268] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, and Rohan Ramanath. "Disagreeable privacy policies: Mismatches between meaning and users' understanding". In: *Berkeley Tech. LJ* 30 (2015), p. 39.

[269] Karen Renaud and Verena Zimmermann. "Ethical guidelines for nudging in information security & privacy". In: *International Journal of Human-Computer Studies* 120 (2018), pp. 22–35.

[270] Emergen Research. *Direct-to-Consumer Genetic Testing Market Size, Share | Industry Forecast by 2030*. https://www.emergenresearch.com/amp/industry-report/direct-to-consumer-genetic-testing-market. Accessed: 2022-12-5. 2022.

[271] Julie M Robillard, Tanya L Feng, Arlo B Sporn, Jen-Ai Lai, Cody Lo, Monica Ta, and Roland Nadler. "Availability, readability, and content of privacy policies and terms of agreements of mental health apps". In: *Internet interventions* 17 (2019), p. 100243.

[272] Marco Robol et al. "Consent modeling and verification: privacy regulations compliance from business goals to business processes". In: (2020).

[273] Arianna Rossi, Rossana Ducato, Helena Haapio, and Stefania Passera. "When Design Met Law: Design Patterns for Information Transparency". In: *Droit de la Consommation Consumenterecht : DCCR* 122–123 (2019), pp. 79–121.

[274] Arianna Rossi and Gabriele Lenzini. "Transparency by design in data-informed research: A collection of information design patterns". In: *Computer Law & Security Review*. Elsevier 37.105402 (2020), pp. 1–22.

[275] Chun Ruan and Sang-Soo Yeo. "Modeling of an Intelligent e-Consent System in a Healthcare Domain." In: *J. Univers. Comput. Sci.* 15.12 (2009), pp. 2429–2444.

[276] Kenneth John Ryan, Joseph V Brady, Robert E Cooke, Dorothy I Height, Albert R Jonsen, Patricia King, Karen Lebacqz, David W Louisell, Donald W Seldin, Eliot Stellar, and Robert H Turtle. *The Belmont Report: ethical principles and guidelines for the protection of human subjects of research*. Tech. rep. Washington DC: US Department of Health, Education, and Welfare, 1979.

[277] Paul Ryan, Harshvardhan J Pandit, and Rob Brennan. "A common semantic model of the GDPR register of processing activities". In: *Frontiers in Artificial Intelligence and Applications* (2020), pp. 251–254.

[278] Richard M Ryan, Randall R Curren, and Edward L Deci. "What humans need: Flourishing in Aristotelian philosophy and self-determination theory." In: *The Best within Us: Positive Psychology Perspectives on Eudaimonia.* Ed. by Alan S Waterman. American Psychological Association, 2013, pp. 57–75.

[279] Tina Hesman Saey. *What FamilyTreeDNA sharing genetic data with police means for you*. 2019-02. URL: https://www.sciencenews.org/article/family-tree-dna-sharing-genetic-data-police-privacy.

[280] Debjani Saha, Anna Chan, Brook Stacy, Kiran Javkar, Sushant Patkar, and Michelle L Mazurek. "User attitudes on direct-to-consumer genetic testing". In: *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE. 2020, pp. 120–138.

[281] Joni Salminen, Kathleen Wenyun Guan, Soon-Gyo Jung, and Bernard Jansen. "Use cases for design personas: A systematic review and new frontiers". In: *CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–21.

[282] Pamela Sankar, Susan Mora, Jon F Merz, and Nora L Jones. "Patient perspectives of medical confidentiality: a review of the literature". In: *Journal of general internal medicine* 18.8 (2003), pp. 659–669.

[283] Cristiana Santos, Arianna Rossi, Lorena Sanchez Chamorro, Kerstin Bongard-Blanchy, and Ruba Abu-Salma. "Cookie Banners, What's the Purpose? Analyzing Cookie Banner Text Through a Legal Lens". In: *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*. South Korea - online: Ieee, 2021, pp. 187–194.

[284] Emanuel Santos, Jaelson Castro, Juan Sánchez, and Oscar Pastor. "A Goal-Oriented Approach for Variability in BPMN." In: *WER*. 2010, pp. 17–28.

[285] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. "A design space for effective privacy notices". In: *Eleventh symposium on usable privacy and security (SOUPS 2015)*. 2015, pp. 1–17.

[286] Bart W Schermer, Bart Custers, and Simone Van der Hof. "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection". In: *Ethics and Information Technology* 16.2 (2014), pp. 171–182.

[287] Hubert Scheuerlein, Falk Rauchfuss, Yves Dittmar, Rüdiger Molle, Torsten Lehmann, Nicole Pienkos, and Utz Settmacher. "New methods for clinical pathways—business process modeling notation (BPMN) and tangible business process modeling (t.BPM)". In: *Langenbeck's archives of surgery* 397 (2012), pp. 755–761.

[288] Brian Schrag. "Research with groups: Group rights, group consent, and collaborative research". In: *Science and Engineering Ethics* 12.3 (2006), pp. 511–521.

[289] KA Schriver and J Frascara. "The rhetoric of redesign in bureaucratic settings". In: Champaign, IL: Common Ground Publishing LLC, 2015, pp. 173–184.

[290] Peter H Schuck. "Rethinking informed consent". In: *Yale Law Journal* (1994), pp. 899–959.

[291] Olivia Schuman, Caroline Beit, Jill Oliver Robinson, Whitney Bash Brooks, Amy L McGuire, and Christi Guerrini. ""The truth should not be hidden": Experiences and Recommendations of Individuals Making NPE Discoveries Through Genetic Genealogy Databases". In: *Genetics in Medicine* (2024), p. 101210.

[292] Abraham P Schwab, Hung S Luu, Jason Wang, and Jason Y Park. "Genomic privacy". In: *Clinical chemistry* 64.12 (2018), pp. 1696–1703.

[293] Ari Schwartz. "Looking back at P3P: lessons for the future". In: *Center for Democracy & Technology* (2009).

[294] Annika Selzer and Ingo J Timm. "Gestaltung eines Systems zum anonymen Datenaustausch – Gewährleistung angemessener Schutzmaßnahmen". In: *Datenschutz und Datensicherheit-DuD* 45.12 (2021), pp. 826–830.

[295] Annika Selzer and Ingo J Timm. "Potenziale anonymer Datenverarbeitungen nutzen – Ein Vorschlag für Smart Cities". In: *Datenschutz und Datensicherheit-DuD* 45.12 (2021), pp. 816–820.

[296] Gwen Shaffer. "Applying a contextual integrity framework to privacy policies for smart technologies". In: *Journal of Information Policy* 11 (2021), pp. 222–265.

[297] Murtaza Hussain Shaikh. "Analysis of Privacy Policy: Responses and Challenges."Experiences from Service Providers"". In: *International Journal of Innovation, Management and Technology* 3.1 (2012).

[298] Maria Shitkova. "On the usability of business process modelling tools-A review and future research directions". In: (2014).

[299] Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, and Helen Nissenbaum. "Analyzing privacy policies using contextual integrity annotations". In: *arXiv preprint arXiv:1809.02236* (2018).

[300] Yan Shvartzshnaider, Noah Apthorpe, Nick Feamster, and Helen Nissenbaum. "Going against the (appropriate) flow: A contextual integrity approach to privacy policy analysis". In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. Vol. 7. 2019, pp. 162–170.

[301] Yan Shvartzshnaider, Zvonimir Pavlinovic, Ananth Balashankar, Thomas Wies, Lakshminarayanan Subramanian, Helen Nissenbaum, and Prateek Mittal. "Vaccine: Using contextual integrity for data leakage detection". In: *The World Wide Web Conference*. 2019, pp. 1702–1712.

[302] Herbert A Simon et al. "Designing organizations for an information-rich world". In: *Computers, communications, and the public interest* 72 (1971), p. 37.

[303] Vanessa Watts Simonds, Eva Marie Garroutte, and Dedra Buchwald. "Health literacy and informed consent materials: designed for documentation, not comprehension of health research". In: *Journal of Health Communication* 22.8 (2017), pp. 682–691.

[304] G Alex Sinha. "NSA surveillance since 9/11 and the human right to privacy". In: *Loy. L. Rev.* 59 (2013), p. 861.

[305] Sevasti Skeva, Maarten HD Larmuseau, and Mahsa Shabani. "Review of policies of companies and databases regarding access to customers' genealogy data for law enforcement purposes". In: *Personalized medicine* 17.2 (2020), pp. 141–153.

[306] Amelia K Smit, Louise A Keogh, Jolyn Hersch, Ainsley J Newson, Phyllis Butow, Gabrielle Williams, and Anne E Cust. "Public preferences for communicating personal genomic risk information: a focus group study". In: *Health Expectations* 19.6 (2016), pp. 1203–1214.

[307] Daniel J Solove. "Introduction: Privacy self-management and the consent dilemma". In: *Harvard Law Review* 126 (2012), p. 1880.

[308] Daniel J Solove. "Murky consent: an approach to the fictions of consent in privacy law". In: *Boston University Law Review* 104.2 (2024).

[309] Ian Sommerville. "Systems engineering for software engineers". In: *Annals of Software Engineering* 6.1 (1998), pp. 111–129.

[310] Christina Spicer. *MyHeritage Class Action Lawsuit Says DNA Reports Exposed in Data Hack*. 2018-09. URL: https://topclassactions.com/lawsuit-settlements/lawsuit-news/myheritage-class-action-lawsuit-says-dna-reports-exposed-data-hack/.

[311] Ciara Staunton, Santa Slokenberga, and Deborah Mascalzoni. "The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks". en. In: *European Journal of Human Genetics* 27.8 (2019-08), pp. 1159–1167.

[312] Wilhelmina Maria Botes Steenberg. "Visual communication as a legal-ethical tool for informed consent in genome research involving the San community of South Africa". PhD thesis. University of South Africa, 2017.

[313] Natalie Stoljar. "Informed consent and relational conceptions of autonomy". In: *Journal of Medicine and Philosophy* 36.4 (2011), pp. 375–384.

[314] Anselm Strauss and Juliet Corbin. "Grounded theory methodology: An overview." In: *Handbook of qualitative research* (1994).

[315] Jackie Street, Katherine Duszynski, Stephanie Krawczyk, and Annette Braunack-Mayer. "The use of citizens' juries in health policy decision-making: a systematic review". In: *Social science & medicine* 109 (2014), pp. 1–9.

[316] European Data Protection Supervisor. *Preliminary Opinion on data protection and scientific research*. en. 2023-02. URL: https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific%5C_en (visited on 2023-02-22).

[317] Katherine A Sward, Rene Enriquez, Jeri Burr, Julie Ozier, Megan Roebuck, Carrie Elliott, and J Michael Dean. "Consent Builder: an innovative tool for creating research informed consent documents". In: *JAMIA open* 5.3 (2022), ooac069.

[318] Latanya Sweeney, Akua Abu, and Julia Winn. "Identifying Participants in the Personal Genome Project by Name". In: *SSRN* (2013).

[319] Mohammad Tahaei, Alisa Frik, and Kami Vaniea. "Privacy champions in software teams: Understanding their motivations, strategies, and challenges". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–15.

[320] Kyoko Takashima, Yuichi Maru, Seiichi Mori, Hiroyuki Mano, Tetsuo Noda, and Kaori Muto. "Ethical concerns on sharing genomic data including patients' family members". In: *BMC medical ethics* 19 (2018), pp. 1–6.

[321] Leonardo Tamariz, Ana Palacio, Mauricio Robert, and Erin N Marcus. "Improving the informed consent process for research subjects with low literacy: a systematic review". In: *Journal of general internal medicine* 28 (2013), pp. 121–126.

[322] Herman T Tavani. "Informational privacy: Concepts, theories, and controversies". In: *The handbook of information and computer ethics* (2008), pp. 131–164.

[323] Herman T Tavani and James H Moor. "Privacy protection, control of information, and privacy-enhancing technologies". In: *ACM Sigcas Computers and Society* 31.1 (2001), pp. 6–11.

[324] Rui Tavares, Mireia Alemany-Pagès, Sara Araújo, Neil Cohn, João Ramalho-Santos, and Anabela Marisa Azul. "Comics in Science and Health Communication: Insights From Mutual Collaboration and Framing a Research Practice". In: *International Journal of Qualitative Methods* 22 (2023), p. 16094069231183118.

[325] Ronald K Taylor. "Marketing Strategies: Gaining A Competitive Advantage Through The Use Of Emotion". In: *Competitiveness Review: An International Business Journal* 10.2 (2000-01), pp. 146–152. ISSN: 1059-5422.

[326] Shirley Taylor and Peter A Todd. "Understanding information technology usage: A test of competing models". In: *Information systems research* 6.2 (1995), pp. 144–176.

[327] Governance Team. *The Elements of Informed Consent. A Toolkit V2.0*. https://sagebionetworks.org/wp-content/uploads/2019/07/SageBio_EIC-Toolkit_V2_17July19_final.pdf. 2019-07.

[328] Harriet JA Teare, Megan Prictor, and Jane Kaye. "Reflections on dynamic consent in biomedical research: the story so far". In: *European Journal of Human Genetics* 29.4 (2021), pp. 649–656.

[329] David R Thomas. "A general inductive approach for analyzing qualitative evaluation data". In: *American journal of evaluation* 27.2 (2006), pp. 237–246.

[330] Damiano Torre, Sallam Abualhaija, Mehrdad Sabetzadeh, Lionel Briand, Katrien Baetens, Peter Goes, and Sylvie Forastier. "An AI-assisted approach for checking the completeness of privacy policies against GDPR". In: *2020 IEEE 28th International Requirements Engineering Conference (RE)*. IEEE. 2020, pp. 136–146.

[331] Adriana Unger, Mauro Spinola, and Marcelo Schneck de Paula Pessôa. "Requirements Engineering Approaches to derive Enterprise Information Systems from Business Process Management: a Systematic Literature Review." In: *Modellierung (Workshops)*. 2018, pp. 261–271.

[332] European Union. *Charter of Fundamental Rights of the European Union*. Legislative Body: EUMS. 2012-10. URL: http://data.europa.eu/eli/treaty/char_2012/oj/eng.

[333] European Union. *Indigenous peoples within the framework of the development cooperation of the Community and the Member States*. (Accessed on 05/06/2024). 1998. URL: https://cendoc.docip.org/collect/cendocdo/index/assoc/HASH0193.dir/Council%5C%20Resolution%5C%20Nov%5C%201998.pdf.

[334] European Union. *Judgment of the Court (Fourth Chamber) of 7 March 2024 (request for a preliminary ruling from the hof van beroep te Brussel - Belgium) – IAB Europe v Gegevensbeschermingsautoriteit*. (Accessed on 17/01/2025). 2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:62022CA0604.

[335] European Union. *Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space*. (Accessed on 05/06/2024). 2022. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0197.

[336] European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. en. Official Journal of the European Union, 2016-04. URL: http://data.europa.eu/eli/reg/2016/679/oj/eng (visited on 2023-02-22).

[337] European Union. *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*. Official Journal of the European Union, 2022. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022R0868.

[338] European Union. *Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC*. Vol. 158. Legislative Body: EP, CONSIL. Official Journal of the European Union, 2014-04. URL: http://data.europa.eu/eli/reg/2014/536/oj/eng.

[339] European Union. *Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC*. Vol. 257. (Accessed on 05/06/2024). Official Journal of the European Union, 2014-07. URL: http://data.europa.eu/eli/reg/2014/910/oj/eng.

[340] European Union. *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847*. (Accessed on 15/01/2025). 2024. URL: https://data.consilium.europa.eu/doc/document/PE-76-2024-INIT/en/pdf.

[341] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. "(Un)informed Consent: Studying GDPR Consent Notices in the Field". In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. London United Kingdom: ACM, 2019-11, pp. 973–990.

[342] Antonino Vaccaro and Peter Madsen. "Firm information transparency: Ethical questions in the information age". In: *Social Informatics: An Information Society for all? In Remembrance of Rob Kling: Proceedings of the Seventh International Conference on Human Choice and Computers (HCC7), IFIP TC 9, Maribor, Slovenia, September 21–23, 2006 7*. Springer. 2006, pp. 145–156.

[343] Jonathan Vajda, J Neil Otte, Cooper Stansbury, Frank J Manion, Elizabeth Umberfield, Yongqun He, Marcelline Harris, Jihad Obeid, Mathias Brochhausen, William D Duncan, et al. "Coordinated evolution of ontologies of informed consent". In: *ICBO* (2018).

[344] Morten Valberg, Mats Julius Stensrud, and Odd O Aalen. "The surprising implications of familial association in disease risk". In: *BMC Public Health* 18 (2018), pp. 1–9.

[345] Thierry Vansweevelt, Nils Broeckx, and Filip Dewallens. "Privacy and health in Belgium". In: *Privacy and Medical Confidentiality in Healthcare*. Edward Elgar Publishing, 2023, pp. 5–23.

[346] Rosa Velasquez, Claudia Negri-Ribalta, Rene Noel, and Oscar Pastor. "Exploring Understandability in Socio-technical Models for Data Protection Analysis: Results from a Focus Group". In: *International Conference on Conceptual Modeling*. Springer. 2023, pp. 263–273.

[347] Viswanath Venkatesh and Fred D Davis. "A theoretical extension of the technology acceptance model: Four longitudinal field studies". In: *Management science* 46.2 (2000), pp. 186–204.

[348] Jochen Vollmann and Rolf Winau. "Informed consent in human experimentation before the Nuremberg code". In: *Bmj* 313.7070 (1996), pp. 1445–1447.

[349] Mark Von Rosing, Henrik Von Scheel, and August-Wilhelm Scheer. *The Complete Business Process Handbook: Body of Knowledge from Process Modeling to BPM, Volume 1*. Vol. 1. Morgan Kaufmann, 2014.

[350] Karel Vredenburg, Ji-Ye Mao, Paul W Smith, and Tom Carey. "A survey of user-centered design practice". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2002, pp. 471–478.

[351] Jantina de Vries, Paulina Tindana, Katherine Littler, Michèle Ramsay, Charles Rotimi, Akin Abayomi, Nicola Mulder, and Bongani M Mayosi. "The H3Africa policy framework: negotiating fairness in genomics". In: *Trends in Genetics* 31.3 (2015), pp. 117–119.

[352] Rob Waller. *What makes a good document? The Criteria We Use*. 2011-04.

[353] Hailiang Wang, Jiaxin Zhang, Yan Luximon, Mingfu Qin, Ping Geng, and Da Tao. "The determinants of user acceptance of mobile medical platforms: An investigation integrating the TPB, TAM, and patient-centered factors". In: *International Journal of Environmental Research and Public Health* 19.17 (2022), p. 10758.

[354] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. "Comparing effectiveness and engagement of data comics and infographics". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–12.

[355] Tara Ward. "The right to free, prior, and informed consent: indigenous peoples' participation rights within international law". In: *Northwester Journal of International Human Rights* 10 (2011), p. 54.

[356] Cornelia Wermuth. "Creativity in the translation of medical comics". In: *Translation and Reality Forum: Theories and Realities in Translation*. Naples, Italy: Aracne, 2016, pp. 1–15.

[357] Alan F Westin. "Privacy and freedom". In: *Washington and Lee Law Review* 25.1 (1968), p. 166.

[358] David A Wheeler, Maithreyan Srinivasan, Michael Egholm, Yufeng Shen, Lei Chen, Amy McGuire, Wen He, Yi-Ju Chen, Vinod Makhijani, G Thomas Roth, et al. "The complete genome of an individual by massively parallel DNA sequencing". In: *nature* 452.7189 (2008), pp. 872–876.

[359] Ray A Wickenheiser. "Forensic genealogy, bioethics and the Golden State Killer case". In: *Forensic science international: Synergy* 1 (2019), pp. 114–125.

[360] Norbert Wiener. *The Human Use of Human Beings: Cybernetics and Society*. 2nd. Houghton Mifflin, 1950.

[361] Svenja Wiertz and Joachim Boldt. "Evaluating models of consent in changing health research environments". In: *Medicine, Health Care and Philosophy* 25.2 (2022), pp. 269–280.

[362] Mackenzie Klein Wilson, Amro Khasawneh, Amal Ponathil, Shraddhaa Narasimha, Sruthy Agnisarman, Brandon Welch, and Kapil Chalil Madathil. "A preliminary study investigating patients' perceptions of research consenting methods". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 63. 1. SAGE Publications Sage CA: Los Angeles, CA. 2019, pp. 1931–1935.

[363] Shomir Wilson, Florian Schaub, Frederick Liu, Kanthashree Mysore Sathyendra, Daniel Smullen, Sebastian Zimmeck, Rohan Ramanath, Peter Story, Fei Liu, Norman Sadeh, et al. "Analyzing privacy policies at scale: From crowdsourcing to automated annotations". In: *ACM Transactions on the Web (TWEB)* 13.1 (2018), pp. 1–29.

[364] Piotr Wiśniewski. "Decomposition of business process models into reusable sub-diagrams". In: *ITM Web of Conferences*. Vol. 15. EDP Sciences. 2017, p. 01002.

[365] Matthias Wjst. "Caught you: threats to confidentiality due to the public release of large-scale genetic data sets". In: *BMC Medical Ethics* 11 (2010), pp. 1–4.

[366] Jacob O Wobbrock and Julie A Kientz. "Research contributions in human-computer interaction". In: *interactions* 23.3 (2016), pp. 38–44.

[367] Antonia Xu, Melissa Therese Baysari, Sophie Lena Stocker, Liang Joo Leow, Richard Osborne Day, and Jane Ellen Carland. "Researchers' views on, and experiences with, the requirement to obtain informed consent in research involving human participants: a qualitative study". In: *BMC Medical Ethics* 21.1 (2020-10), p. 93. ISSN: 1472-6939.

[368] Daniel R Young, Donald T Hooker, and Fred E Freeberg. "Informed consent documents: increasing comprehension by reducing reading level". In: *IRB: Ethics & Human Research* 12.3 (1990), pp. 1–5.

[369] Fue Zeng, Qing Ye, Zhilin Yang, Jing Li, and Yiping Amy Song. "Which privacy policy works, privacy assurance or personalization declaration? An investigation of privacy policies and privacy concerns". In: *Journal of Business Ethics* (2022), pp. 1–18.

[370] Jiaxin Zhang, Yan Luximon, and Qingchuan Li. "Seeking medical advice in mobile applications: How social cue design and privacy concerns influence trust and behavioral intention in impersonal patient–physician interactions". In: *Computers in human behavior* 130 (2022), p. 107178.

[371] Sebastian Zimmeck and Steven M Bellovin. "Privee: An architecture for automatically analyzing web privacy policies". In: *23rd USENIX Security Symposium*. 2014, pp. 1–16.

[372] Chaim Zins. "Conceptual approaches for defining data, information, and knowledge". In: *Journal of the American society for information science and technology* 58.4 (2007), pp. 479–493.

[373] Shoshana Zuboff. "'We make them dance': surveillance capitalism, the rise of instrumentarian power, and the threat to human rights". In: *Human rights in the age of platforms* (2019), pp. 3–51.

[374] Shoshana Zuboff. *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: PublicAffairs, 2019.