Analysis of Transparency and User-relevancy of DTC Company Policies*

Xengie Doan[1][0000−0002−8245−1555], Fatma Sümeyra Doğan[2][0000−0002−6774−0228], and Arianna Rossi[1,3][0000−0002−4199−5898]

[1] SnT, University of Luxembourg, 2 Avenue de l'Université, Esch-sur-Alzette, Luxembourg xengie.doan@uni.lu
[2] Jagiellonian University, Gołębia 24, Kraków, Poland
[3] LIDER Lab, Dirpolis, Sant'Anna School of Advanced Studies, Pisa, Italy

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.

Keywords: DTC genetic testing · transparency · contextual integrity · consent · human-centered design · privacy

1 Introduction

Privacy policies have a long reputation of being time-consuming [34], unreadable [12], difficult to comprehend for non-legal experts [23, 46], and unusable as a decision-making tool [24]. While this may be a deliberate strategy to be legally comprehensive and accurate, especially in unclear situations for future-proofing [6], 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 contextual integrity (CI) which states that privacy can be defined

* Supported by LeADS Grant Agreement ID 956562 and Luxembourg FNR Grant Agreement IS/14717072

by contextual norms dictating their transmission and how appropriate the information flows are [36]. 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 [54]. Second, we look to user-centered terms developed by Johansson et al. [25] 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 direct-to-consumer (DTC) companies we will investigate are genetic testing companies handling sensitive data.

Sequencing 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 USD by 2032 [47]. While genetic data is a special category of personal data under the EU's General Data Protection Regulation (GDPR) and should be subject to a high standard of legal and technical protection (GDPR) [3], 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 [38], re-identification based on combining public datasets [11], data sharing with law enforcement [50], or discrimination by insurance companies [26] may increase while consumers assume companies are ethical [5].

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, 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 [25]; 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.

2 Related Work

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 [53]. In it, privacy is defined by Nissenbaum as the contextual norms dictating appropriate data transmission [36]. 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 [7]). 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" [39] 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 [39, 49]. There is a movement within legal communication away from lengthy documents full of legal jargon and instead towards using information design elements [40] to address the needs and abilities of the intended audiences (who are not only legal experts) [42] and enhance the readability and comprehensibility of information [48].

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 US focus, and researchers have found that their policies lack transparency and adequate protections against harms [9, 15, 21], especially in relation to data sharing with law enforcement [44, 55]. This was shown in a 2015 study [22], 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.

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 [33]. Moreover, discrimination and stigmatization [16, 19] 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 [27, 60]. Other risks include possible re-identification from anonymous data, which was proved possible through the cross-referencing of two free databases in 2013 [18]. 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 [11]. In the age of big data, privacy is networked – especially genetic data which may reveal information about similar genetic relatives [8, 14, 30]. Companies may not adequately inform customers of how their individual decisions may affect their families and descendants in the status quo.

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*" [39]. This aligns with the concept of information transparency, which is defined by how complete the information provided by firms is regarding their business activities [59]. While privacy policies have many issues [12, 23, 24, 46], they are a window into internal data practices of a company publicly available to regulators, researchers, and consumers alike.

2.4 User-centered decision making and privacy expectations

Johansson et al. [25] 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 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 [25] (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?

4 Methods

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 [2], while 23andMe reported over 14 million [1] on their websites. MyHeritage and Family Tree DNA followed with more than 2 million customers each in 2018 [45]. 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 App. Fig.6).

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 1: CI framework from Shvartzshnaider et al. [54]

4.2 Contextual integrity

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 [36]. 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 [54] (Table 1. Without specifying all five parameters, the context is too ambiguous to assess the implications. We followed methods from Shvartzshnaider et al. [54] to perform deductive qualitative coding (Step 3 in App. Fig. 6) and assess the quality of CI flows with missing information, parameter bloating, and vague terms analyses (Step 4 in App. Fig. 6).

To code the corpus, we began by identifying information flows regarding genetic data and co-code the respective parameters (see Fig. 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. [54]. 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.. 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).

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**)

Fig. 1: Example of annotated CI flow from AncestryDNA.

Parameter bloating identifies any flows with two or more discrete entries per parameter which can cause confusion from the lack of directness (Fig. 2). 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.

Fig. 2: Example of a CI flow with parameter bloating from AncestryDNA text from Fig.1 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. [6]

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 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 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.

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. [25] (Table 3) to analyze the quality of the information within policies and information flows. Mappings to CI elements are included in Table 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 [25] 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. 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: Data sharing attributes from Johansson et al. [25]

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. 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.

5 Results

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 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 [18, 61]).

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. 2 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.

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 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.

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 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%).

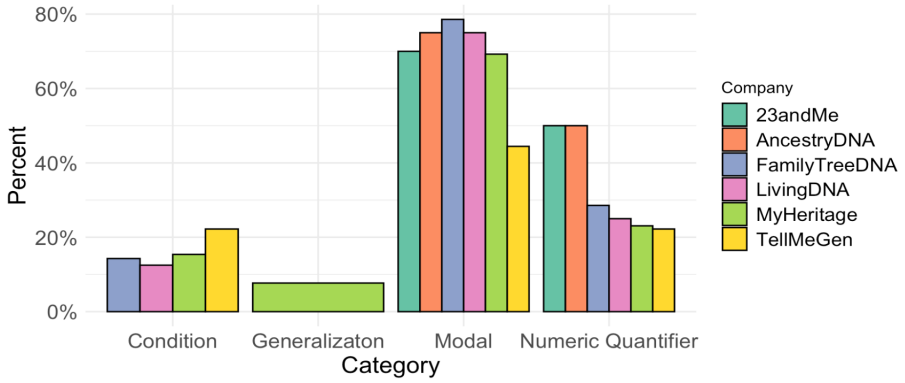

Fig. 3: The percent of CI flows with vague terms by lexical category and company

Vague terms and GDPR requirements From this mapping of predetermined vague terms deriving from [6], 81% of information flows may not be using clear language (Art. 12(1)) due to the prevalence of vague terms.

5.2 User relevant information

To answer RQ3 regarding a qualitative analysis of the attributes reported in Table 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. 2, 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).

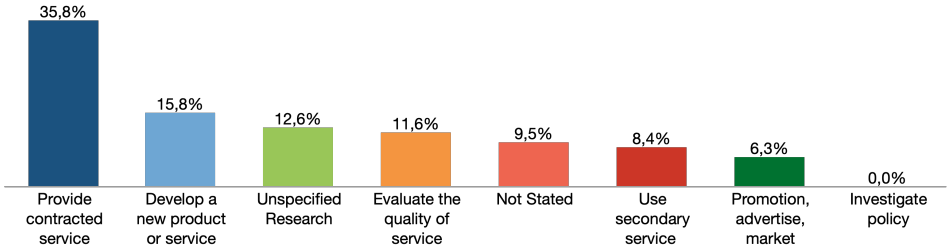

Fig. 4: Percentage of coded segments with reasons for data use across companies

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.

5.3 Stated risks and benefits

Regarding RQ4, first we report the number and type of risk (App. Table 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. 5. 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).

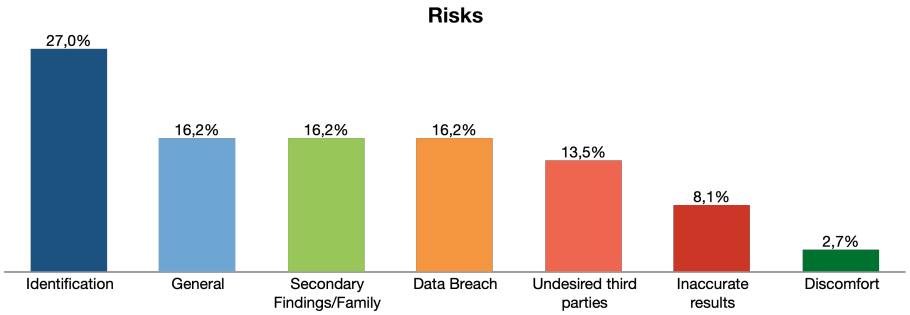

Fig. 5: Risks across companies in descending order, based on percentage of coded segments

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 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.

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 App. Tab. 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.

6 Discussion

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 [12,22,24,54] and lack of transparency into a company’s practices. This can lead to customer misinterpretation [5] and incorrect privacy expectations [29]. 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) [51]. 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 [38] stated that she would not have used the services if she had known that the necessary precautions were not in place [56]. 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 [28]. 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 [6]. 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 [39] 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 [17] and of GDPR requirements [4,58]. 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) [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 [5] 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 [5, 29], 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 [13]. 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 [29], while academic partnerships can lead to for-profit interests in the very same research outcomes [10, 32]. 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.

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 [18]. In the US, the Golden State Killer suspect was identified from a third cousin who was a distant relative from the 19th century [52], 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 [50]. 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 [14] 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 [8], 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 [57]. 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 [20, 37, 41] 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 [43], moving collective consent into a more digital age. The need for collective family consent is especially discussed by legal and ethical scholars as well [31, 35]. However, translating such proposals into actual safeguards and digital tools is still an ongoing subject of research.

7 Limitations and Future Work

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.

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.

9 Appendix

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

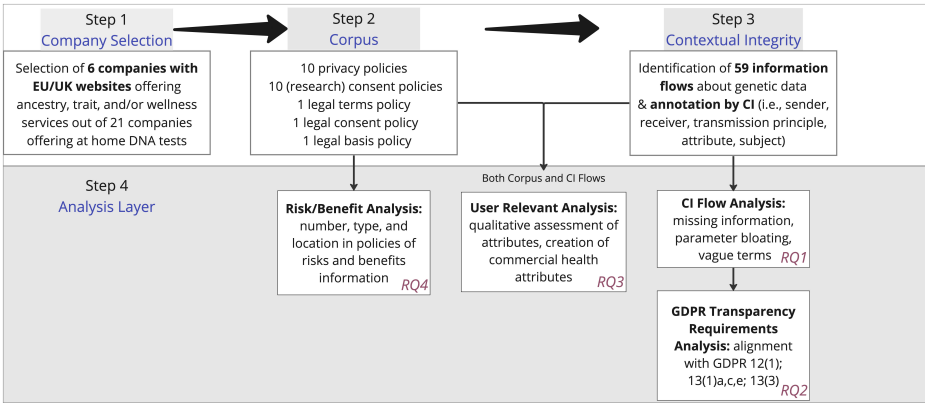

Fig. 6: A process map of the steps to identifying genetic data flows and analyses used from 6 company websites and their corpus.

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

References

1. 23andme for healthcare professionals, `https://medical.23andme.com/`

2. Company facts, https://www.ancestry.com/corporate/about-ancestry/company-facts
3. Regulation (EU) 2016/679 (General Data Protection Regulation), vol. 119 (Apr 2016), http://data.europa.eu/eli/reg/2016/679/oj/eng
4. Amaral, O., Abualhaija, S., Torre, D., Sabetzadeh, M., Briand, L.C.: Ai-enabled automation for completeness checking of privacy policies. IEEE Transactions on Software Engineering **48**(11), 4647–4674 (2021)
5. Baig, K., Mohamed, R., Theus, A.L., Chiasson, S.: " 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. pp. 1–13 (2020)
6. Bhatia, J., Breaux, T.D., Reidenberg, J.R., Norton, T.B.: A theory of vagueness and privacy risk perception. In: 2016 IEEE 24th International Requirements Engineering Conference (RE). pp. 26–35. IEEE (2016)
7. Board, E.D.P.: Guidelines 4/2019 on article 25 data protection by design and by default (Nov 2019)
8. Boyd, D.: Networked privacy. Surveillance & society **10**(3/4), 348 (2012)
9. Clayton, E.W., Evans, B.J., Hazel, J.W., Rothstein, M.A.: The law of genetic privacy: applications, implications, and limitations. Journal of Law and the Biosciences **6**(1), 1–36 (2019)
10. DeAngelis, C.D., Fontanarosa, P.B.: Impugning the integrity of medical science: the adverse effects of industry influence. Jama **299**(15), 1833–1835 (2008)
11. Erlich, Y., Shor, T., Pe'er, I., Carmi, S.: Identity inference of genomic data using long-range familial searches. Science **362**(6415), 690–694 (2018)
12. Fabian, B., Ermakova, T., Lentz, T.: Large-scale readability analysis of privacy policies. p. 18–25. WI '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3106426.3106427
13. Fox, M.: 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 (2018), accessed: 2023-02-5
14. Frizzo-Barker, J., Chow-White, P.A., Charters, A., Ha, D.: Genomic big data and privacy: challenges and opportunities for precision medicine. Computer Supported Cooperative Work (CSCW) **25**, 115–136 (2016)
15. Garner, S.A., Kim, J.: The privacy risks of direct-to-consumer genetic testing: A case study of 23andme and ancestry. Wash. UL Rev. **96**, 1219 (2018)
16. Garrison, N., Non, A.L.: Direct-to-consumer genomics companies should provide guidance to their customers on (not) sharing personal genomic information. The American Journal of Bioethics **14**(11), 55–57 (2014)
17. Grünewald, E., Pallas, F.: 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. pp. 636–646 (2021)
18. Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y.: Identifying personal genomes by surname inference. Science **339**(6117), 321–324 (2013)
19. Haeusermann, T., Fadda, M., Blasimme, A., Tzovaras, B.G., Vayena, E.: Genes wide open: Data sharing and the social gradient of genomic privacy. AJOB Empirical Bioethics **9**(4), 207–221 (2018)
20. Hayward, A., Sjoblom, E., Sinclair, S., Cidro, J.: A new era of indigenous research: Community-based indigenous research ethics protocols in canada. Journal of Empirical Research on Human Research Ethics **16**(4), 403–417 (2021)

21. Hazel, J.W., Slobogin, C.: Who knows what, and when: a survey of the privacy policies proffered by us direct-to-consumer genetic testing companies. Cornell JL & Pub. Pol'y **28**, 35 (2018)
22. Huang, H.Y., Bashir, M.: Direct-to-consumer genetic testing: Contextual privacy predicament. Proceedings of the Association for Information Science and Technology **52**(1), 1–10 (2015)
23. Ibdah, D., Lachtar, N., Raparthi, S.M., Bacha, A.: "why should i read the privacy policy, i just need the service": A study on attitudes and perceptions toward privacy policies. IEEE access **9**, 166465–166487 (2021)
24. Jensen, C., Potts, C.: Privacy policies as decision-making tools: an evaluation of online privacy notices. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems. pp. 471–478 (2004)
25. Johansson, J.V., Shah, N., Haraldsdóttir, E., Bentzen, H.B., Coy, S., Kaye, J., Mascalzoni, D., Veldwijk, J.: Governance mechanisms for sharing of health data: An approach towards selecting attributes for complex discrete choice experiment studies. Technology in society **66**, 101625 (2021)
26. Joly, Y., Dupras, C., Pinkesz, M., Tovino, S.A., Rothstein, M.A.: Looking beyond gina: policy approaches to address genetic discrimination. Annual Review of Genomics and Human Genetics **21**, 491–507 (2020)
27. Joly, Y., Ngueng Feze, I., Simard, J.: Genetic discrimination and life insurance: a systematic review of the evidence. BMC medicine **11**, 1–15 (2013)
28. Jones, C.: Latest 23andme data claim would take leaked records to 5m (2023), `https://www.theregister.com/2023/10/19/latest_23andme_data_leak_takes/`
29. King, J.: " becoming part of something bigger" direct to consumer genetic testing, privacy, and personal disclosure. Proceedings of the ACM on Human-Computer Interaction **3**(CSCW), 1–33 (2019)
30. Klugman, C.M., Rodriguez, H.F.: Ethics of familial genetic genealogy: Solving crimes at the cost of privacy. DePaul J. Health Care L. **22**, 67 (2021)
31. Knoppers, B.M., Kekesi-Lafrance, K.: The genetic family as patient? American Journal of Bioethics **20**(6), 77–80 (Jun 2020). https://doi.org/10.1080/15265161.2020.1754505
32. Lehmann, L.S., Kaufman, D.J., Sharp, R.R., Moreno, T.A., Mountain, J.L., Roberts, J.S., Green, R.C.: Navigating a research partnership between academia and industry to assess the impact of personalized genetic testing. Genetics in medicine **14**(2), 268–273 (2012)
33. Majumder, M.A., Guerrini, C.J., McGuire, A.L.: Direct-to-consumer genetic testing: value and risk. Annual Review of Medicine **72**, 151–166 (2021)
34. McDonald, A.M., Cranor, L.F.: The cost of reading privacy policies. Isjlp **4**, 543 (2008)
35. Minari, J., Teare, H., Mitchell, C., Kaye, J., Kato, K.: The emerging need for family-centric initiatives for obtaining consent in personal genome research. Genome Medicine **6**(12), 118 (Dec 2014). https://doi.org/10.1186/s13073-014-0118-y
36. Nissenbaum, H.: Privacy in context. In: Privacy in Context. Stanford University Press (2009)
37. Norton, I.M., Manson, S.M.: Research in american indian and alaska native communities: navigating the cultural universe of values and process. Journal of consulting and clinical psychology **64**(5), 856 (1996)

38. NortonLifeLock: Myheritage data breach exposes info of more than 92 million users (2018), https://us.norton.com/blog/emerging-threats/myheritage-data-breach-exposes-info-of-more-than-92-million-user
39. Party, A..D.P.W.: Guidelines on transparency under regulation 2016/679, 17/en wp260 rev.01 (Apr 2018), https://ec.europa.eu/newsroom/article29/redirection/document/51025, published: Online at
40. Passera, S.: Beyond the wall of contract text. visualizing contracts to foster understanding and collaboration within and across organizations (2017), https://aaltodoc.aalto.fi/bitstream/handle/123456789/27292/isbn9789526075280.pdf
41. Piquemal, N.: Free and informed consent in research involving native american communities. American Indian Culture and Research Journal **25**(1) (2001)
42. Potel-Seville, M., Talbourdet, E.: Empowering children to understand and exercise their personal data rights, p. 253–276. Ledizioni, Italy (2021)
43. Prictor, M., Huebner, S., Teare, H.J.A., Burchill, L., Kaye, J.: Australian aboriginal and torres strait islander collections of genetic heritage: The legal, ethical and practical considerations of a dynamic consent approach to decision making. Journal of Law, Medicine & Ethics **48**(1), 205–217 (2020). https://doi.org/10.1177/1073110520917012
44. Ram, N.: Genetic privacy after carpenter. Virginia Law Review **105**(7), 1357–1425 (2019)
45. Regalado, A.: More than 26 million people have taken an at-home ancestry test (Feb 2019), https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/
46. Reidenberg, J.R., Breaux, T., Cranor, L.F., French, B., Grannis, A., Graves, J.T., Liu, F., McDonald, A., Norton, T.B., Ramanath, R.: Disagreeable privacy policies: Mismatches between meaning and users' understanding. Berkeley Tech. LJ **30**, 39 (2015)
47. Research, E.: 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 (2022), accessed: 2022-12-5
48. Rossi, A., Ducato, R., Haapio, H., Passera, S.: When design met law: Design patterns for information transparency. Droit de la Consommation Consumenterecht : DCCR **122–123**, 79–121 (2019)
49. Rossi, A., Lenzini, G.: Transparency by design in data-informed research: A collection of information design patterns. Computer Law & Security Review **37**(105402), 1–22 (2020). https://doi.org/https://doi.org/10.1016/j.clsr.2020.105402
50. Saey, T.H.: What familytreedna sharing genetic data with police means for you (Feb 2019), https://www.sciencenews.org/article/family-tree-dna-sharing-genetic-data-police-privacy
51. Saha, D., Chan, A., Stacy, B., Javkar, K., Patkar, S., Mazurek, M.L.: User attitudes on direct-to-consumer genetic testing. In: 2020 IEEE European Symposium on Security and Privacy (EuroS&P). pp. 120–138. IEEE (2020)
52. Schwab, A.P., Luu, H.S., Wang, J., Park, J.Y.: Genomic privacy. Clinical chemistry **64**(12), 1696–1703 (2018)
53. Shvartzshnaider, Y., Apthorpe, N., Feamster, N., Nissenbaum, H.: Analyzing privacy policies using contextual integrity annotations. arXiv preprint arXiv:1809.02236 (2018)
54. Shvartzshnaider, Y., Apthorpe, N., Feamster, N., Nissenbaum, H.: 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, pp. 162–170 (2019)
55. Skeva, S., Larmuseau, M.H., Shabani, M.: Review of policies of companies and databases regarding access to customers' genealogy data for law enforcement purposes. Personalized medicine **17**(2), 141–153 (2020)
56. Spicer, C.: Myheritage class action lawsuit says dna reports exposed in data hack (Sep 2018), `https://topclassactions.com/lawsuit-settlements/lawsuit-news/myheritage-class-action-lawsuit-says-dna-reports-exposed-data-hack/`
57. Takashima, K., Maru, Y., Mori, S., Mano, H., Noda, T., Muto, K.: Ethical concerns on sharing genomic data including patients' family members. BMC medical ethics **19**, 1–6 (2018)
58. Torre, D., Abualhaija, S., Sabetzadeh, M., Briand, L., Baetens, K., Goes, P., Forastier, S.: An ai-assisted approach for checking the completeness of privacy policies against gdpr. In: 2020 IEEE 28th International Requirements Engineering Conference (RE). pp. 136–146. IEEE (2020)
59. Vaccaro, A., Madsen, P.: 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. pp. 145–156. Springer (2006)
60. Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T., et al.: The complete genome of an individual by massively parallel dna sequencing. nature **452**(7189), 872–876 (2008)
61. Wjst, M.: Caught you: threats to confidentiality due to the public release of large-scale genetic data sets. BMC medical ethics **11**, 1–4 (2010)