INTRODUCING KIARA

USING KIARA TO IMPROVE RESEARCH TRANSPARENCY & SUPPORT DIGITAL LITERACY IN HISTORICAL RESEARCH

LUCA FEDERICO CERRA, LAUREN COETZEE, ELIANE SCHMID *

WHAT IS KIARA?

KIARA IS A DATA ORCHESTRATION TOOL DEVELOPED BY THE DHARPA PROJECT, WHICH ALLOWS USERS TO DOCUMENT AND CRITICALLY REFLECT ON THE PROCESS OF DATA CLEANING, PRE-PROCESSING, MANIPULATION AND CREATING A DATA LINEAGE WHICH ARCHIVES ALL MODIFICATIONS AND META DATA. INSPIRED BY "CHIAROSCURO", KIARA ILLUMINATES THE ENTIRE RESEARCH PROCESS IN DIGITAL HUMANITIES, NOT JUST THE INPUTS AND OUTPUTS, BUT ALSO THE PROCESS OF GETTING THERE, RE-THINKING THE BLACK BOXES OF HISTORICAL RESEARCH.

IN TODAY'S DATA-RICH ENVIRONMENT, IT IS IMPORTANT TO THINK ABOUT HOW THIS ABUNDANCE OF DATA GETS PROCESSED TO MAKE SENSE OF HISTORICAL DATA. NEW APPROACHES AND TOOLS ARE NECESSARY THAT CAPTURE COMPONENTS OF THE RESEARCH PROCESS, LEADING TO THE DHARPA PROJECT TO THINK ABOUT HOW THIS ACTIVE DECISION-MAKING COULD BE CONVERTED INTO A HISTORICAL SOURCE AND ANALYTICAL TOOL.

WITH THE EXPLOSION OF VARIOUS DIGITAL HUMANITIES TOOLS, A GAP IN THE EFFECTIVENESS OF SEVERAL TOOLS IS THAT THEY LACK THE MEANS TO ALLOW FOR CRITICAL REFLECTION ON THE MATERIALS, PROCESSING AND TOOLS INCORPORATED. WITH NEW APPROACHES FOR DATA ANALYSIS, INTERPRETATION, VISUALISATION AND CREATION OF DATA BEING A CORE ACTIVITY IN RESEARCH, KIARA ADDRESSES THIS GAP BY OFFERING A TECHNICAL FRAMEWORK FOR ORCHESTRATING SOLUTIONS TO THESE NEW ISSUES, IN AN INTELLIGIBLE MANNER THAT DOES NOT SIMPLY OVERLOOK THE "BLACK BOXES" IN THE USE OF THESE TOOLS. THIS IS ONE OF THE CORE PURPOSES OF KIARA AS A DIGITAL TOOL.

RATHER THAN FALLING INTO THE CATEGORY OF A "BLACK BOX" TOOL WHERE THE DATA PROCESS OF THE USER IS LOST ONCE THE PROJECT IS COMPLETED, KIARA PROMOTES INCREASED ORGANISATION, TRANSPARENCY AND ACCESSIBILITY IN THE WORLD OF DATA AND RESEARCH AND CREATES REPRODUCIBLE RESEARCH SOURCES, UNLIKE OTHER TOOLS WHICH OBSCURE THESE USER PROCESSES.

KIARA IN DIGITAL HISTORY

DOCUMENTING AND CRITIQUING NOT ONLY OUR SOURCES, BUT ALSO OUR ENGAGEMENT WITH THEM IS CRUCIAL IN CONDUCTING ETHICAL RESEARCH. THIS INCLUDES RECORDING THE RESEARCHER'S THOUGHT PROCESS, BUT ALSO THE INFLUENTIAL LITERATURE THEY WERE THINKING ABOUT WHEN THEY CHOSE TO ALTER OR ANALYSE THE DATA IN A PARTICULAR WAY. DIGITAL HUMANISTS SHOULD BE ABLE TO TRACK WHEN PROCESSES ARE SET IN MOTION, BECAUSE SCHOLARSHIP IS NEVER FINISHED, ALWAYS EVOLVING, AND KNOWING WHEN WE THOUGHT ABOUT SOMETHING AND IN WHAT ORDER CAN HELP US, AND OTHERS, RE-TRACE AND RE-EVALUATE OUR WORK. THUS, THE PROCESS OF WRITING HISTORY BECOMES A SOURCE ITSELF, MAKING THE RESEARCH MORE COMPREHENSIBLE TO EXTERNAL PEOPLE, AND ITS TRACEABILITY ADDS TRANSPARENCY AND REPLICABILITY TO THE RESEARCH.

OUR APPROACH TO ANNOTATION DIFFERS FROM TYPICAL DIGITAL HUMANITIES FOCUS ON MEDIA ANNOTATION (E.G. SOURCES LIKE TEXTS, IMAGES, AUDIO, AND VIDEO), WHICH IN ITSELF IS OFTEN CONFLATED WITH THE PRODUCTION OF ANNOTATIONS THAT ARE BOTH PUBLICLY SHAREABLE AND WRITTEN IN COLLABORATION WITH OTHER RESEARCHERS. WE ARE NOT THE FIRST TO NOTE THIS DISTINCTION, BUT WE ARE ONE OF THE ONLY SUSTAINED ATTEMPTS TO ADDRESS THIS DEFICIT, PLACING THE EXPERT WITHIN THE MACHINE THROUGH DETAILED DOCUMENTATION OF THE RESEARCH PROCESS. ONE WAY WE CAN PLACE OURSELVES AS THE EXPERT INSIDE THE MACHINE IS THROUGH ANNOTATION, BY LEAVING TRACES OF OUR PRESENCE IN HUMANITIES DATA PRODUCTION, FROM ITS ORIGINAL CREATION – OFTEN MISLEADINGLY LABELED "ONBOARDING" OR "CLEANING" – OR THROUGH THE ACCRETION OF LAYERS WHICH SEEK TO ADD SIGNALS OR SUBTRACT NOISE THROUGH ACTIONS LIKE FILTERING OR VISUALISATION.

SINCE THE FUTURE OF ARCHIVES IS INCREASINGLY DIGITAL, TRADITIONAL HANDS-ON ARCHIVAL RESEARCH HAS TRANSFORMED INTO A MORE EFFICIENT, DATA-DRIVEN PROCESS FACILITATED BY DIGITAL PHOTOGRAPHY. THIS SHIFT, WHILE ENHANCING ACCESSIBILITY, CHALLENGES RESEARCHERS TO MAINTAIN CRITICAL ENGAGEMENT WITH MATERIALS. TOOLS LIKE KIARA ADDRESS THESE CHALLENGES BY PROMOTING TRANSPARENCY AND REFLECTION, AUTOMATICALLY TRACKING SOURCES, VISUALIZATIONS, CODE, AND METADATA, THUS PROVIDING A COMPREHENSIVE DATA LINEAGE. BY COMBINING AUTOMATED AND MANUAL ANNOTATION CAPABILITIES, KIARA EMPOWERS HISTORIANS TO CREATE MORE ROBUST, TRANSPARENT, AND REFLECTIVE HISTORICAL NARRATIVES, RE-THINKING THE BLACK BOXES OF HISTORICAL RESEARCH.
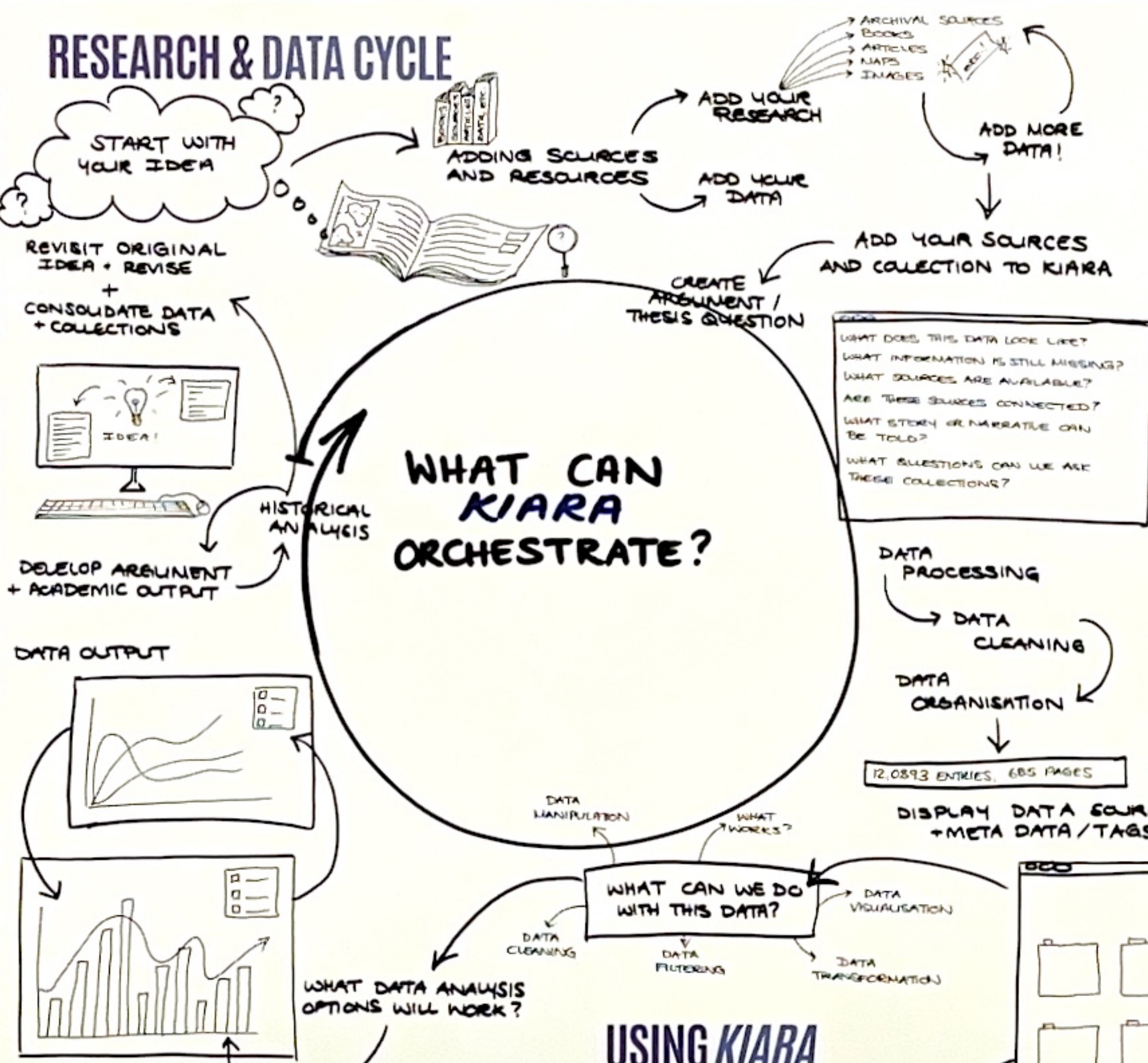
DHARPA PROJECT

THE DHARPA (DIGITAL HISTORY ADVANCED RESEARCH PROJECTS ACCELERATOR) PROJECT ASSESSES TECHNOLOGY'S IMPACT ON HISTORICAL RESEARCH AND EXPERIMENTS WITH HOW IT CAN RE-SHAPE HISTORY'S METHODOLOGICAL UNDERPINNINGS AS A SCIENTIFIC DISCIPLINE. ITS CORE FOCUS IS TO PROMOTE INCREASED TRANSPARENCY AND ACCESSIBILITY IN THE WORLD OF BIG DATA AND DIGITAL HISTORIES.

TO MAKE SENSE OF THE HISTORICAL DATA, DHARPA DEVELOPS NEW APPROACHES AND TOOLS, AIMING TO THINK ABOUT HOW TO TRANSFORM ACTIVE DECISION-MAKING INTO A HISTORICAL SOURCE AND MEANS OF ANALYSIS IN ITSELF. KIARA IS A KEY DEVELOPMENT IN ALLOWING FOR WORK TO BE BETTER ORGANISED, ACCESSIBLE AND MADE MORE TRANSPARENT, WORKING IN TANDEM WITH USERS TO DOCUMENT AND CAPTURE THE PROCESS OF RESEARCHING WITH DATA.

SINCE 2016, MANY DIGITAL HUMANITIES TOOLS HAVE EMERGED, BUT OFTEN LACK THE MEANS TO ALLOW FOR CRITICAL REFLECTION ON THE MATERIALS, PROCESSING AND TOOLS INCORPORATED. WITH THE MULTITUDE OF NEW APPROACHES FOR DATA ANALYSIS, INTERPRETATION, VISUALISATION AND CREATION OF DATA BEING A CORE ACTIVITY IN THE TEAM, A TECHNICAL FRAMEWORK WAS ESTABLISHED FOR ORCHESTRATING SOLUTIONS TO THESE NEW ISSUES, IN AN INTELLIGIBLE MANNER THAT DOES NOT SIMPLY OVERLOOK THE "BLACK BOXES" IN THE USE OF THESE TOOLS.

RESEARCH & DATA CYCLE

START WITH YOUR IDEA · ADDING SOURCES AND RESOURCES · ADD YOUR RESEARCH · ADD MORE DATA! · ADD YOUR DATA · ADD YOUR SOURCES AND COLLECTION TO KIARA · CREATE ARGUMENT / THESIS QUESTION · REVISIT ORIGINAL IDEA + REVISE + CONSOLIDATE DATA + COLLECTIONS · IDEA! · HISTORICAL ANALYSIS · DEVELOP ARGUMENT + ACADEMIC OUTPUT · DATA OUTPUT

WHAT CAN KIARA ORCHESTRATE?

WHAT DOES THIS DATA LOOK LIKE? · WHAT INFORMATION IS STILL MISSING? · WHAT SOURCES ARE AVAILABLE? · ARE THESE SOURCES CONNECTED? · WHAT STORY OR NARRATIVE CAN BE TOLD? · WHAT QUESTIONS CAN WE ASK THESE COLLECTIONS?

DATA PROCESSING → DATA CLEANING → DATA ORGANISATION · 12,089.3 ENTRIES, 685 PAGES · DISPLAY DATA SOURCES + META DATA / TAGS · DATA MANIPULATION · WHAT WORKS? · WHAT CAN WE DO WITH THIS DATA? · DATA CLEANING · DATA FILTERING · DATA VISUALISATION · DATA TRANSFORMATION · WHAT DATA ANALYSIS OPTIONS WILL WORK? · YOUR FILES YOUR COLLECTIONS

```
kiara is a data-orchestration framework; this is the command-line frontend for
it.
For more information, visit the kiara homepage:
https://dharpa.org/kiara.documentation .

Options
--config              -cnf    TEXT    A kiara config file (or folder
                                      containing one named
                                      'kiara.config').
--context             -ctx,-c TEXT    The name of the kiara context to
                                      use (or the path to a context
                                      file).
--pipelines           -p      TEXT    File(s) and folder(s) that contain
                                      extra pipeline definitions.
--plugin              -P      TEXT    Ensure the provided plugin
                                      package(s) are installed in the
                                      virtual environment.
--use-background-service -b           Always use the background service
                                      (start if not already running).
--version             -v              Show the version of kiara and
                                      installed plugins, then exit.
--runtime-info        -ri             Show current environment
                                      information, then exit.
--help                -h              Show this message and exit.

Commands
archive                Kiara archive related sub-commands.
context                Kiara context related sub-commands.
data                   Data-related sub-commands.
data-type              Information about available data types.
info                   Kiara config related sub-commands.
module                 Module-related sub-commands.
operation              Operation-related sub-commands.
pipeline               Pipeline-related sub-commands.
render                 Render-related sub-commands.
run                    Run a kiara operation.
```

```
Available operations
Id                          Type(s)      Description
assemble.tables                          Assemble a 'tables'
                                         value from multiple
                                         tables.
create.database.from.file   create.file  Create a database
                                         from a file.
create.database.from.file_bundle create.file  Create a database
                                         from a file_bundle.
create.database.from.table  create.table Create a database
                                         from a table.
create.database.from.tables create.tables Create a database
                                         value from a list
                                         of tables.
create.table.from.file      create.file  Create a table from
                                         a file, trying to
                                         auto-determine the
                                         format of said
                                         file.
create.table.from.file_bundle create.file_bundle Create a table
                                         from a file_bundle.
create.tables.from.file_bundle create.file_bundle Create a database
                                         from a file_bundle.
data.check_range                         Check whether a
                                         date falls within a
                                         specified date
                                         range.
data.extract_from_string                 Extract a date
                                         from a string.
download.file                            Download a single
                                         file from a remote
                                         location.
download.file.from.github                Download a single
                                         file from a github
                                         repo.
download.file.from.zenodo                Download a single
                                         file from a Zenodo
                                         record.
download.file_bundle                     Download a file
                                         bundle from a
                                         remote location.
download.file_bundle.from.github         Download a file
                                         bundle from a
                                         remote github
                                         repository.
```

DATA LINEAGE PROTOTYPE
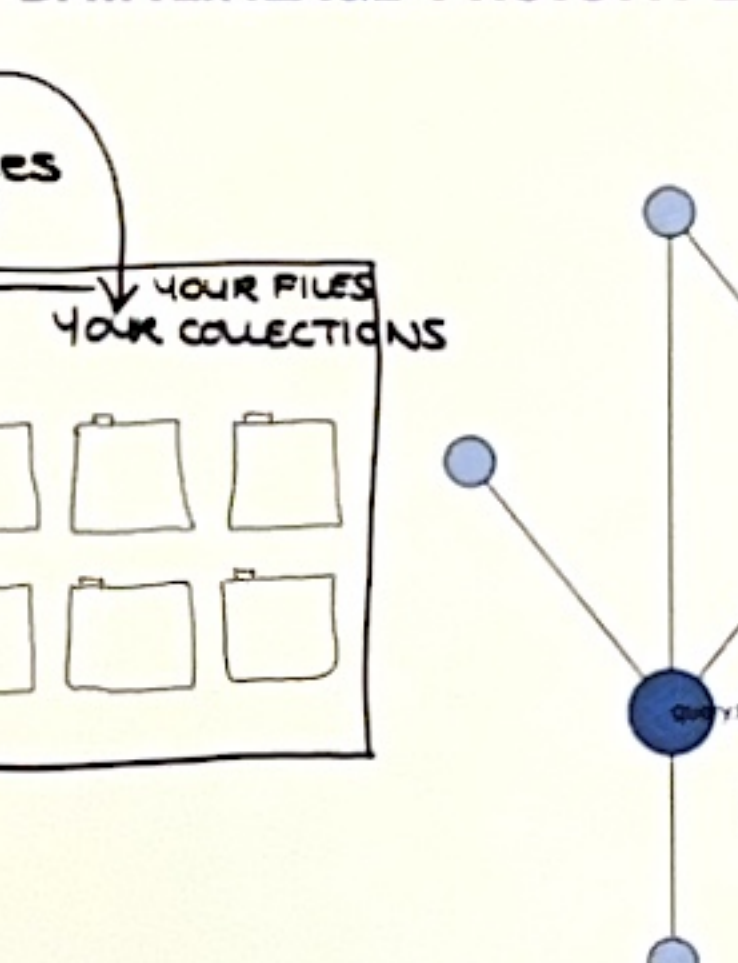

```
parse.date_array                         Create an array of
                                         date objects from
                                         an array of
                                         strings.
query.database                           Execute a sql query
                                         against a (sqlite)
                                         database.
query.table                              Execute a sql query
                                         against an (Arrow)
                                         table.
string_filter.tokens         filter      n/a
table.add_column                         Add a column to a
                                         table.
table.pick.column                        Pick one column
                                         from a table,
                                         returning an array.
table.filter.drop_columns    filter      n/a
table.filter.select_columns  filter      n/a
table.filter.select_rows     filter      n/a
tables.pick.column                       Pick one column
                                         from a table,
                                         returning an array.
tables.pick.table                        Pick one column
                                         from a table,
                                         returning an array.
```

USING KIARA

```
GephiFile = kiara.run_job('import.local.file', inputs={'path':'/Users/eliane.schmid/Desktop/Gephi_Kiara.csv'}, comment='upload file')

GephiFile = GephiFile['file']
GephiFile

        value_id          8542cc9b-46c2-41bd-a4db-3dd58bc36e18
        kiara_id          7ebeb4f6f-1d7c-4172-ac94-058d26db0b00

        data_type_info
                data_type_name      file
                data_type_config
                                    {
                                      "content_type": null
                                    }
                characteristics
                                    {
                                      "is_scalar": false,
                                      "is_json_serializable": false
                                    }
                data_type_class
                                    python_class_name      FileValueType
                                    python_module_name     kiara.data_types.included_core_types.filesystem
                                    full_name              kiara.data_types.included_core_types.filesystem.FileValueType

        destiny_backlinks   {}
        job_id              414bd03c-1886-492e-023c-4f93b1d7db47
        property_links
                                    {
                                      "metadata.file": "b751e472-aa31-4e85-8e3d-4ae97aa5a027",
                                      "metadata.python_class": "46831653-dba0-4452-0d73-5060b7eaecce"
                                    }
        value_created       2024-06-17 13:52:21.393082+02:00
        value_hash          zdpuAvMGpGa4fYUCbdqvCnqtVPLe5U0F8DhKghgg0FL65ed
        value_schema
                                    type              file
                                    type_config       {}
                                    default           __not_set__
                                    optional          False
                                    is_constant       False
                                    doc               The loaded files.
        value_size          3.79 KB
        value_status        -- set --
```


ADD WATCH FILE + KEEP UPDATING SOURCES / COLLECTION

```
GEPHITABLE = KIARA.RUN_JOB('CREATE.TABLE.FROM.FILE',
INPUTS={'FILE':GEPHIFILE, FIRST_ROW_IS_HEADER:TRUE}, COMMENT='CREATE TABLE')

TABLE = KIARA.RUN_JOB('IMPORT.TABLE.FROM.LOCAL_FILE_PATH',
INPUTS={'PATH':'/USERS/ELIANE.SCHMID/DESKTOP/GEPHI_KIARA.CSV', FIRST_ROW_IS_HEADER:TRUE}, COMMENT='COMMENT')
```

THE COMMENT COMMAND IS KEY IN KIARA - YOU CAN DOCUMENT FOR YOUR FUTURE SELF WHAT YOU DID AND WHY! OPERATION INFO ALLOWS YOU TO UNDERSTAND WHAT WORKS HOW!

```
kiara.retrieve_operation_info('create.table.from.file')

Author(s)
        Markus Binsteiner    markus@frkl.io

Context
                Tags       tabular
                Labels     package: kiara_plugin.tabular
                References source_repo: https://github.com/DHARPA-Project/kiara_plugin.tabular
                           documentation: https://DHARPA-Project.github.io/kiara_plugin.tabular/

Module type     create.table
Module config
                "source_type":  "file"
                "target_type":  "table"

Operation details
Documentation   Create a table from a file, trying to auto-determine the format of said file.
                Currently supported input file types:
                  • csv
                  • parquet

Inputs
field name       type    description                                    Required  Default
file             file    The source value (of type 'file').             yes       -- no defau
first_row_is_header boolean Whether the first row of a (csv)            no        -- no defau
                           file is a header row. If not
                           provided, kiara will try to
                           auto-determine. Ignored if not a
                           csv file.

Outputs
field name       type    description
table            table   The result value (of type 'table').
```

HOW DOES IT WORK?

KIARA USES A MODULAR APPROACH TO LET USERS RE-USE PROVEN ANALYTICAL WORKFLOWS, AND CREATE NEW ONES FROM EXISTING BUILDING BLOCKS. IT MANAGES RESEARCH DATA AND AUGMENTS IT WITH AUTOMATICALLY-, SEMI-AUTOMATICALLY-, AND MANUALLY-CREATED METADATA. KIARA CAN PERFORM ANY TASK POSSIBLE IN A PYTHON OR JUPYTER ENVIRONMENT WHILST CAPTURING METADATA, TIMESTAMPS, COMMENTS, PROCEDURES AND DOCUMENTATION THROUGHOUT THE PROJECT.

KIARA'S BUILT IN MODULES ALLOW FOR INTEGRATION OF VARIOUS PYTHON LIBRARIES FOR TASKS SUCH AS NETWORK ANALYSIS AND TOPIC MODELLING. SOME ADVANTAGES OF USING KIARA INCLUDE:

- MODULES CAN OFFER EXPLANATIONS FOR HOW THEY OPERATE.
- MODULES CAN BE CHAINED TOGETHER AND FORMED INTO PIPELINES, WITH ASSOCIATED LINEAGES OF THE DATA.
- DEMONSTRATE WHERE THIS DATA WAS SOURCED FROM, AND PROCESSING STEPS.
- VISUALISATION OF THE MODULE AND PIPELINE LOGIC, TRACING THE INPUTS, CHOICES, LINEAGES AND DATA FLOWS THROUGHOUT THE PROJECT.

THIS APPROACH ENHANCES TRANSPARENCY AND REPRODUCIBILITY IN DIGITAL HUMANITIES RESEARCH.

WHY USE KIARA?

BLACK BOXES IN RESEARCH, ESPECIALLY AI-RELATED, OFTEN EVOKE NEGATIVE PERCEPTIONS, SCEPTICISM AND AVOIDANCE. KIARA WAS DEVELOPED TO OPEN THE BLACK BOX OF DIGITAL HISTORICAL PRACTICE, REVEALING AND CONTRIBUTING TO THE INFORMATION ARCHITECTURE THAT STRUCTURES RESEARCH.

RAMSAY AND ROCKWELL (2012) ARGUE THAT SOURCE CODE PROVIDES INSIGHT INTO BLACK BOXES THEORETICAL ASSUMPTIONS, WITH OPEN-SOURCE PHILOSOPHY OFFERING NECESSARY TRANSPARENCY FOR PEER REVIEW. THIS PERSPECTIVE FRAMES BLACK BOXES NOT AS A TECHNICAL ISSUE, BUT AS A PROBLEM OF UNCHECKED, UNEVALUATED SOURCE CODES BY FELLOW HISTORIANS.

THE PRIMARY CHALLENGE IN DIGITAL HUMANITIES IS NO LONGER ABUNDANCE, OR EVEN DIGITISATION AND ACCESSIBILITY, BUT RATHER ONE OF INTELLIGIBILITY WITH TODAY'S DIGITAL ABUNDANCE, RESEARCH IS RECONFIGURED WITH THE DIGITISATION OF SOURCES AND INCORPORATION OF GREATER DATASETS AND ARCHIVES, NECESSITATING INCREASED CONCERN FOR THE UNDERLYING METHODS, PRACTICES AND PROCESSES THAT GENERATE DATA AND METADATA.

KEY CONTRIBUTIONS

HISTORICAL RESEARCH WITH DATA EXTENDS BEYOND INPUTS AND OUTPUTS, ENCOMPASSING THE ENTIRE PROCESS, INCLUDING THE IN BETWEEN, THE BLACK BOXES. THE ACTIVE DECISIONS TAKEN BY THE RESEARCHER APPLYING TOOLS/ FILTERS, EDITING OR VISUALISING OR CONCEPTUALISING ARE CRUCIAL ASPECTS OF METHODOLOGICAL REFLECTION. IT IS AN ACTIVE DECISION-MAKING PROCESS THAT GETS LOST IN THE REALM OF DIGITAL HUMANITIES.

THE SHIFT FROM AN "AGE OF SCARCITY" TO AN "AGE OF ABUNDANCE" IN DIGITAL DATA AND GREATER INCORPORATION OF THESE DATASETS INTO RESEARCH NECESSITATED CREATING A SPACE WHERE THESE CHANGES ARE APPROPRIATELY FLESHED OUT, MANAGED AND MONITORED. THE INCREASE IN DIGITAL HUMANITIES TOOLS IN THE LAST DECADE HIGHLIGHTED THE NEED FOR COMPREHENSIVE SOFTWARE THAT NOT ONLY STORES AND SORTS DATA, BUT ALSO ACTIVELY PROCESSES, CLEANS, FILTERS, EDITS, MANIPULATES AND COLLATES IT- ALL DONE IN KIARA.

THE WAY THAT DATA IS LAYERED, COLLATED AND MANIPULATED CAN BE AS SIGNIFICANT AS THE QUALITY AND TYPE OF DATA INTEGRATED INTO HISTORICAL RESEARCH. IN THIS WAY, KIARA IS NOT A NEW INNOVATION, RATHER A REMINDER OF THE ANALOGUE FOUNDATIONS OF HISTORICAL CRAFT: CAREFUL SOURCE TRACING, ANNOTATION, AND NOTE TAKING THAT SHOULD BE RE-INTEGRATED IN DAILY WORK IN THE DIGITAL AGE.

* DOCTORAL RESEARCHERS AT THE UNIVERSITY OF LUXEMBOURG. LUCAFEDERICO.CERRA@UNI.LU · LAUREN.COETZEE@UNI.LU · ELIANE.SCHMID@UNI.LU