© 2022. This work is licensed under a Creative Commons "Attribution 4.0 International" license.

Introducing the DHARPA Project: An Interdisciplinary Lab to Enable Critical DH Practice

Angela R. Cunningham[1], Helena Jaskov[1], Sean Takats[1], Lorella Viola[1], and with contributions from Markus Binsteiner and Mariella de Crouy-Chanel[1]

[1]Luxembourg Centre for Contemporary and Digital History (C^{2}DH), University of Luxembourg

In this article, we introduce software under development by the Digital History Advanced Research Projects Accelerator (DHARPA), an interdisciplinary team of researchers and developers working to enable best practices in the humanities through technology. We argue that the strength and appeal of historical inquiry lies largely in the relationship between scholars and their sources, a connection in which the former engage with the latter through critical assessment, contextualisation and documentation. However, we also contend that these symbiotic processes themselves need to be evaluated and recorded. While digital tools and techniques have been accused of alienating historians from their materials, critically informed digital methodologies can also fortify and extend this relationship. To that end, we are building software that will enable users to not only apply best practices to their sources but also to allow them to write the history of those interactions, thus providing material for self-reflection and critique. In this article, having laid out our goal, we describe how the modular and data-centric design of our software's backend and the interactive documentary capabilities of its frontend operationalize a critical epistemology centered on the scholar-source relationship. We continue with a discussion of our team's internal dynamics in creating this software, and conclude with an invitation to readers to contribute to the process through commentary and testing.

Keywords: digital critical literacy, encoding criticism, digital transparency, digital traceability; reproducibility

1 Introduction

At the level of research outputs, the Digital History Advanced Research Projects Accelerator (DHARPA) responds to the ways that technology has already reshaped

historians' work by developing tools and methods. But more broadly speaking, the project's conception and execution draw on and contribute to long-running debates surrounding the digital humanities (DH). *The Chronicle of Higher Education* offers an insight into the violent nature of such debates by characterizing the dispute between the proponents and opponents of the DH as a "war" (Da, 2019). While introducing Nan Z. Da's scathing assessment of the application of computational methods to literary interpretation, *The Chronicle* also resurfaced past articles across nearly a decade that catalogued the promises and the pitfalls of DH (Brennan, 2017; Conrad, 2014; Fitzpatrick, 2011). The critiques raised in these articles and others are now, sadly, quite familiar.[1] It has been argued that in the flood of digital data and techniques, humanists have lost their way, blinded by the spectacle of technology and the glittering grant money attached to the latest academic trend (Allington et al., 2016). Even DH proponents have suggested that a surfeit of data and push-button tools have led to a false sense of mastery (Tenen, 2016; Weingart, 2011),[2] the unproblematized (mis)application of methodologies that originated in other disciplines (Drucker, 2011), and stolen or abandoned scholarly agency (Grimshaw, 2018; Liu, 2012). Poetic metaphors like "black boxes" and "mechanical turks"[3] evoke a naïve and starry-eyed reliance on poorly understood computer-enabled techniques (Diakopoulos, 2014; Noble, 2018; Pasquale, 2016): at best, critics charge, digital humanists have allowed themselves to become dependent on technologies of which they cannot or will not examine the inner workings, or, worse, they have let themselves believe in the specious autonomy of the machine (Tenen, 2016). In short, digital intermediation has broken the essential connection between scholars and their sources, a loss that can feel profoundly dehumanizing.

We all know the critiques, and many of us have been unsettled by the rupture. But how do we as humanists take back ownership of and responsibility for the work that we do increasingly – and some would say unavoidably – with digital data, digital tools and digital methodologies? How do we shine a light on the interior workings of technological and intellectual black boxes? This article introduces the software currently under development by the Digital History Advanced Research Projects Accelerator (DHARPA), an interdisciplinary team working to address these issues.[4] While DHARPA is multifaceted, with an interest in collaboration, pedagogy, and infrastructure, here we will focus on the software at the foundation of our project, and particularly on aspects of that software which speak to the need to (re)build a critically-aware connection between scholar and sources in a digital world. In the following section, we will lay out our ideas about DH best practice. Then, in describing our "Virtual Research Environment" we will outline how we are operationalizing

[1] In this article we will focus largely on epistemological issues. For more discussions on the economic and political issues that are associated with the digital humanities as an institutional entity, see the various contributions to the C21 conference under the telling header "The Dark Side of Digital Humanities" (Chun, 2013; Jagoda, 2013; Raley, 2013).

[2] Weingart (2011) critiques the push-button mentality with popular network analysis software, while Tenen (2016) observes that "some tools encourage intellectual laziness by obscuring methodology". Tenen, however, also points out that "more often, it is not the tool but rather a mode of lazy thinking that is at fault".

[3] The Mechanical Turk was a clever illusion, devised in the late 18th century and exhibited well into the 19th, which was composed of a mannequin dressed in traditional Ottoman costume and ersatz machinery in full view and, hidden from the audience, the human chess master who was in truth responsible for winning the majority of games the apparatus played against sundry famous opponents (Wikipedia contributors, 2022).

[4] The DHARPA project has been generously funded by the Luxembourg National Research Fund, grant 13307816. `https://www.fnr.lu/projects/digital-history-advanced-research-projects-accelerator/`.

these ideas through a modular and data-centric backend and an interactive frontend capable of documenting scholarly processes. Through this work, we hope to turn the digital into an opportunity not only to regain but also deepen our relationship with sources through technology.

2 Enabling and encouraging best (digital) humanities and history practice

History as a discipline is already home to a set of best practices. No historian worthy of the name would think of *not* critiquing their sources in light of their content and the context in which they developed, nor would any good historian fail to take copious notes or keep track of citations. Yet, the actual enactment of this best practice often goes unrecorded. Historians rarely document their research methods and are often reluctant to share their experience on how they identified and collected their sources in the first place (Beals, 2013; Putnam, 2016). Among many historians, there is often little or no documentation of the scholarly engagement that turned their sources into data nor of how they used these data to produce analyses and obtain insights (Viola and Fiscarelli, 2021). A similar charge may be leveled against practitioners of other humanities (and indeed of scientific and technological disciplines), where the influence of the researcher on the research is rarely acknowledged (Marcus, 2021). On the other side of the digital scholarship/traditional scholarship divide, some advocates for computer-supported methodologies laud technology's ability to make such scholarly engagement with sources more dynamic, with interactive on-the-fly functionality allowing the scholar more direct and exploratory involvement with their sources (Travis, 2015). Yet the outputs of this sort of playful exploration are likewise often unmoored from any accounting of the processes that created them. In both instances, the traditional and the digital, there is rarely any narrative of the researcher's interventions, any material for self-reflection, nor much evidence to counter the depiction of the humanities as esoteric. Arguably, both digital and traditional ways of doing history and humanities research may be accused of being like the Mechanical Turk, with the decisions and actions made by the researcher hidden from view and only the well-oiled and seemingly autonomous product – the article, the book, the conference presentation – on display.

In the unfortunate tendencies to simply "push the execute button" or to present scholarly work as the product of similarly automatic and obvious processes lie other dangers. Disassociating the humanities scholar, scientist or programmer from the technologies that they build and employ downplays the need to understand the cultural, philosophical and historical contexts in which digital tools and operations originated (Dobson, 2019; Hitchcock, 2013): this is akin to studying history without regard to historiography. In portraying computational and computer-aided methods as objective, unassailable, and observer- or user-independent – an attitude that Daston (1992) has described as a belief in "instrumental objectivity" – we risk perpetuating "the god trick" (Haraway, 1988; Sheppard, 2001), and thus supporting objectivist and positivist understandings that are 'fundamentally at odds with approaches to humanities scholarship premised on constructivist principles' (Drucker, 2011). As we find ourselves in an increasingly digital world, it is all the more important to incorporate critical theory into the digital (Berry and Fagerjord, 2017).

At the heart of DHARPA is the desire to encourage and enable good historical and humanities practice – or indeed the good practice of other disciplines that a scholar may wish to call upon – but also to extend that best practice through technology. We

argue that we need to document and critique our sources, but also our engagement with them. For example, how to best record a researcher's thought process? What literature were they thinking about when they chose to alter or analyse the data in a particular way? We argue that digital humanists should be able to keep track of when they set processes 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. DHARPA aims to encourage self-reflective appreciation of how the application of expertise can work in tandem with technology to produce knowledge, and through documenting academic labour, to make its value evident, to make it transparent, and to keep it honest and accountable. Digital humanists, we contend, should be able not only to look inside the box to see its inner workings, but to place themselves as the expert inside the machine.

3 DHARPA- practicing critical digital humanities through interdisciplinary collaboration

Writing in an era of boundary-drawing, Liu (2012) called on digital humanists 'to start reflecting on the wider net of relations of digital research'. A decade later, there is still a need not only to self-reflexively address the tools we use, the sources we apply them to, and - as we have argued here - how we work in relationship with them, but also how we work in relationship with each other. To that end, we now turn to describing our interdisciplinary team and how we have collaborated to bring the best practices of our own fields into application within the Virtual Research Environment (VRE).

Our project builds on the lessons of past digital humanities projects by operating on a longer time frame and with a team that mixes not only disciplinary expertise but also encourages collaboration across a wide range of skill sets and experience. DHARPA is supported by the Luxembourg National Research Fund under a scheme that is intended to promote high-risk, high-reward research, providing an unusually long period of performance of five years, plus ongoing institutional support beyond the initial five years. Embedded within the Luxembourg Centre for Contemporary and Digital History (C^2DH), currently the largest digital history center with 120 full-time equivalent staff, the DHARPA team is composed of four developers, six academics (including PhD students, postdoctoral researchers, and faculty), and an administrative assistant. The composition of the development team is explicitly modeled on the teams that developed Zotero, Tropy, and PressForward, all long-running and co-designed projects. Each of us brings our own expertise and opinions to the table, drawn from years of experience in computer programming, data engineering, data visualization, linguistics, geography, and various strains of history.

An important starting point for our collaboration was the collection of "user stories" from researchers. These plain-text descriptions of common operations or expectations when dealing with data, like recording and saving metadata, tracing data provenance, or validating input, allowed our developer team to infer the technical requirements that the backend was expected to provide. Throughout the process of building the VRE, we have had many discussions about the very nature of sources, how they are transformed into computer-legible data, and the ways that we in our varied positionalities interact with these data through technology (Cunningham et al., 2021) [5]. For instance, we

[5] The citation here is for a recording made of one of these discussions for presentation at an academic conference, using this polyvocal format to decenter one of us scholars as "the expert" and

have noted how problematic the term "dirty data" is, a term implying that there is a precisely recorded (and recordable) reality that we can distill if only given the right tools. Yet we have also needed to acknowledge that sources that have not been standardized and formatted in some way are useless to a block of code. Recalling the main argument of this article, we find that the scholar must intervene to make sources usable, whether as digitized data or as handwritten notes from an archival visit, and that it is best to keep a record of this manipulation. Such a realization has informed the functionality of all the elements of the DHARPA software.

As the software itself is envisioned to promote transparency and documentation in digital research, so is its development and the decision-making process meticulously documented. A small part of this is already publicly available via the project's GitHub account,[6] with more detailed publications in preparation. Browsing through this material, we can revisit our previous decisions, reflect on the work that we have accomplished, and make adjustments as we encounter challenges. For instance, an early debate that stands out in this documentation material is the question of what type of software we were going to build. Would it be a desktop app or a web application? Are we building a workflow execution tool or a workflow assembly tool? As the project continues to evolve, we have retained some flexibility on these issues by focusing first on the specification and development of a data orchestration backend. While our initial target remains a desktop application, we have also created the conditions for producing a future web app as well.

4 The Virtual Research Environment

To attain these goals, we at DHARPA are building a virtual research environment. Our software will be free and open-source, and to further encourage widespread and confident adoption the VRE will be available remotely and for downloadable local use; it will rely on a generalizable hosting infrastructure that ensures privacy and portability; and it will be supported by a long-term sustainability plan and training opportunities. The VRE will have modular design, allowing users to build and rebuild, and to run and rerun their own workflows or those they have adopted and adapted. The VRE's modules will include functionality for data ingestion, data standardization, textual analysis, network analysis, and geographical analysis, all brought within a seamless environment where work can flow between tasks from experimentation and modelling to presentation and dissemination. We anticipate that in the future this repertoire of core modules will be expanded by the community of DH scholars who would be able to create their own modules tailored to their specific research needs. The VRE will also be equipped with essential features for documenting the user's research process and tracing all the transformations which their data undergo, allowing the historian to write a self-reflexive meta-history of their relationship with their sources through the software. To describe in more detail how our VRE functions, we turn first to *kiara*, our backend, and then to *Lumy*, comprised of our frontend and the code necessary for the two pieces to communicate with each other. Figure 1 provides a simple schematic of how *kiara* and *Lumy* work together.

instead highlight the collaborative nature of our work.

[6] `https://github.com/DHARPA-Project`

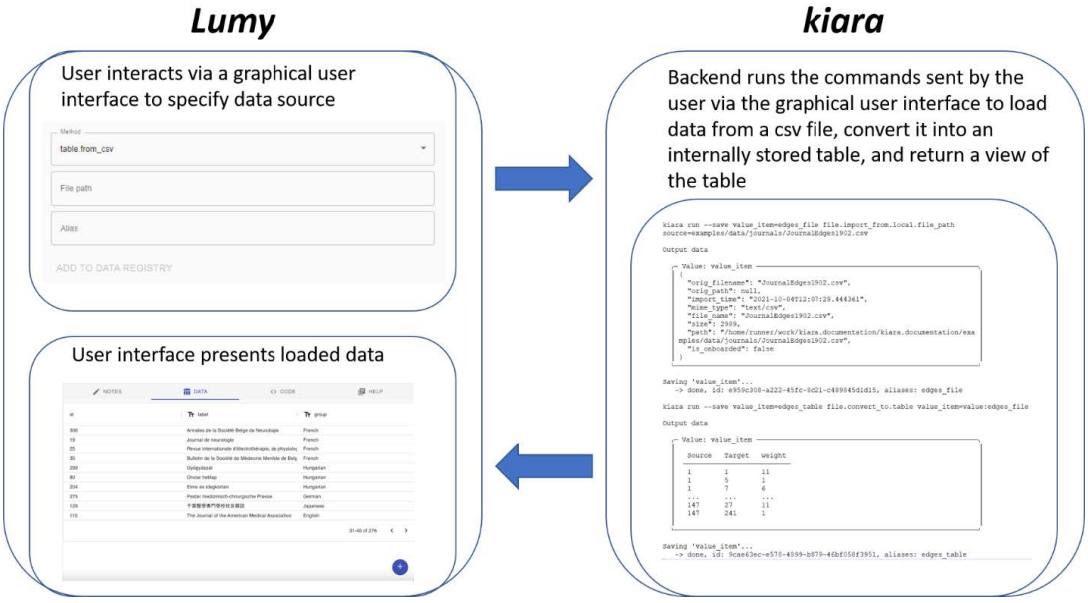

Figure 1: A schematic of how *Lumy* and *kiara* will work together with the scholar/user to complete an important task: bringing sources into the system and transforming them into computer-legible data.

4.1 kiara: backend modularity and a data-centric approach to encourage engaged exploration

Our VRE is built on top a backend, *kiara*, comprised of modules developed on the basis of collaboration between researchers and developers. *kiara* has been inspired by a variety of data pipeline and orchestration tools like airflow (`https://airflow.apache.org/`), prefect (`https://www.prefect.io/`), dagster (`https://dagster.io/`), snakemake (`https://snakemake.readthedocs.io/en/stable/`) and kedro (`https://kedro.readthedocs.io/en/stable/`).[7] What sets *kiara* apart from these tools is its strong focus on metadata management and interactive workflow execution. While most tools feature batch processing, *kiara* offers an interactive approach in addition to batch processing. This allows users to interact with their data at crucial steps within the workflow and to take control of these decisions in an interactive and user-friendly environment. *kiara* is a Python-based tool that can be navigated via the command line. Anyone with basic familiarity with Python can adapt existing Python code to *kiara* modules which can then be shared with other researchers. As shown in Figure 2, a *kiara* user obtains a list of available modules for a specific method by using the `<kiara operation list>` command and adding a keyword. In the following example, *kiara* displays the modules that belong to the network analysis group.

[7] For a survey of this software, see Ruf et al. (2021) as well as Matskin et al. (2021). Other comparable tools specializing in scalable workflow execution and data management are Ray (`https://docs.ray.io`) and Dask (`https://dask.org/`).

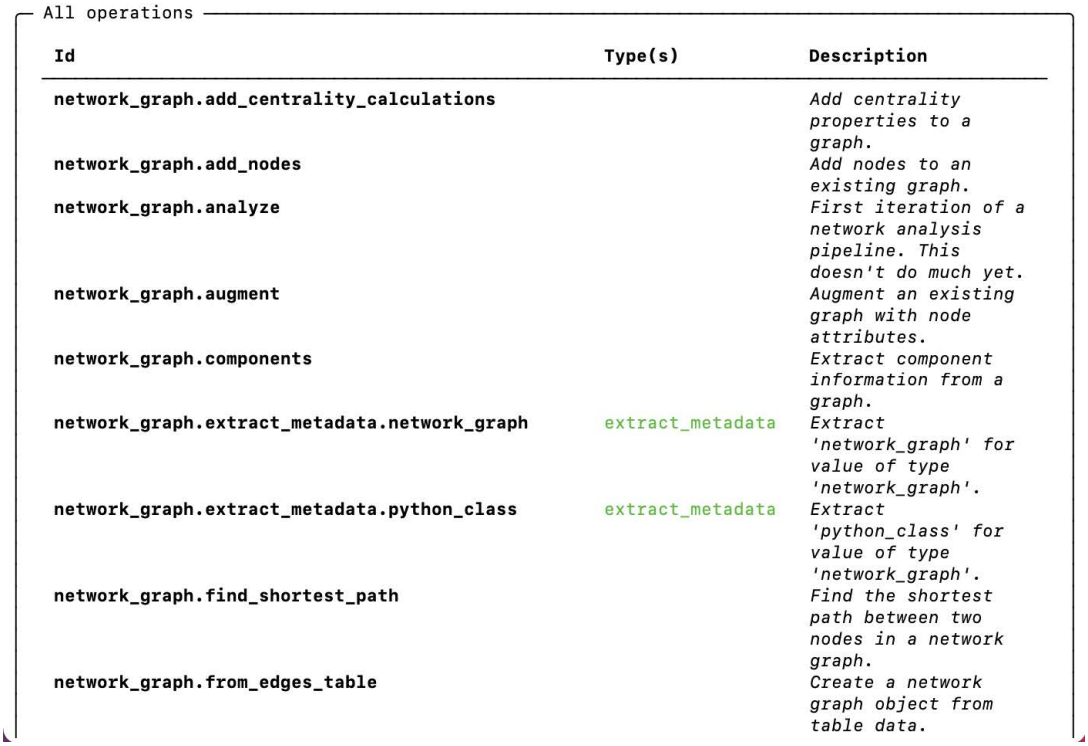

Figure 2: An excerpt from *kiara*'s collection of network analysis modules and their short descriptions. This output can be generated by executing the `<kiara list operations network>` command. A module-creating user can modify the description field and create categories within the "type" field to facilitate the search for desired module.

A *kiara* module consists of a metadata block and three core sections: input, output, and process. The input section specifies any required input, providing a short description of the expected properties in order to help users make informed decisions about the use of the module. The `<add_nodes>` module shown in the list above, for example, requires three input items: a graph object (the graph you wish to augment), a table (the nodes table from which you wish to extract the new nodes), and a string (the name of the column that holds the index of the nodes). The output section defines the expected output (in our example: the new augmented graph object), and the process section holds the Python code that is required to execute the module. Because *kiara* compartmentalizes into separate modules the workflow steps that would usually appear in a single Python script, they can be rearranged and chained together to create a range of workflow pipelines according to the researcher's needs.

Apart from the advantage of offering a flexible solution with a potential for further customization, this data-centric, modular approach also ensures a full recording and documentation of research data's creation and usage. In order to be able to keep track of the various transformations that the data undergoes in such a setting, *kiara* provides the ability to trace data ancestry. In the following example from the network analysis module, the user has requested the lineage of `<journals_graph>` (via the command `<kiara data explain --lineage journals_graph>`), and has received a record of the modules and inputs used to generate that graph. As can be inferred from Figure 3, the graph has been created from two different data sources. It has first been generated from an edges table and afterwards augmented with additional nodes and node attributes from a separate nodes table via the `<add_nodes>` module.

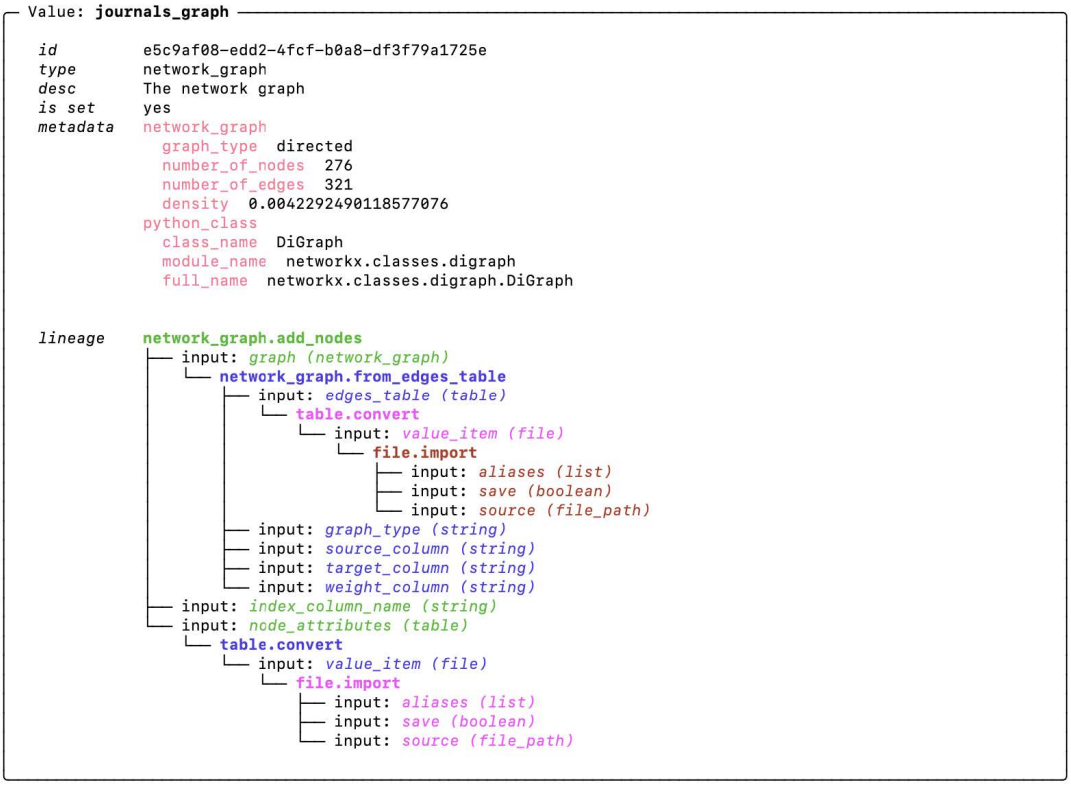

Figure 3: An example of *kiara*'s ability to record and recall the ancestry of a data value. In this example from the network analysis module, the user has requested the lineage of `<journals_graph>`, a directed graph of the citation connections among scientific journals, and has received a record of the modules and inputs used to calculate that graph.

kiara can also be used in combination with the open-source tool Streamlit (`https://streamlit.io/`) which opens up new possibilities of rapid app development and has the benefit of making *kiara* accessible within a community based software environment. Streamlit offers several advantages: on the one hand, it accelerates workflow prototyping, and on the other hand it can function as a handy tool to assist *kiara* module developers and the broader potential community of *kiara* contributors to create their own modules and pipelines. Streamlit enables contributors, who are familiar or beginners with Python, and who are not familiar with JavaScript, to very easily create user interface (UI) elements for *kiara* workflow steps. Streamlit's ease-of-use and flexibility provide module contributors with sufficient freedom on the arrangement of UI interactive and narrative components thereby enabling researchers to embed criticism into the UI. Moreover, Streamlit also makes it easy to create auto-generated UI elements linked to *kiara* modules, that can act as helpers for *kiara* module contributors, to facilitate developing and testing data science back-end processes. Such a development process becomes consequently more widely accessible to the broader community.

4.2 Lumy: frontend interactivity and documentation to promote holistic humanities practice

Much as *kiara* offers the ability to engage with data iteratively while helping to keep track of the history of those interactions, so too will *Lumy* encourage users to adopt

recursive, self-reflexive and reproducible work practices. Figure 4 presents a wireframe mock-up of what *Lumy*'s interface will look like for one part of the network analysis module. On the left side of the panel, the user will be able to interact with tabular data and interconnected visualizations simultaneously. On the right side of the panel, *Lumy* will pair the ability to record what the software is doing with the ability to record what the scholar is doing, cultivating the holistic practice of interweaving methodology, data, code, and metadata with a narrative of researcher's choices and actions. The user will be able to both snapshot a list of inputs and current parameters (all of which might be changed in a future iteration), and take their own notes, through a panel that will also be able to accommodate links, images, and references to relevant secondary literature.

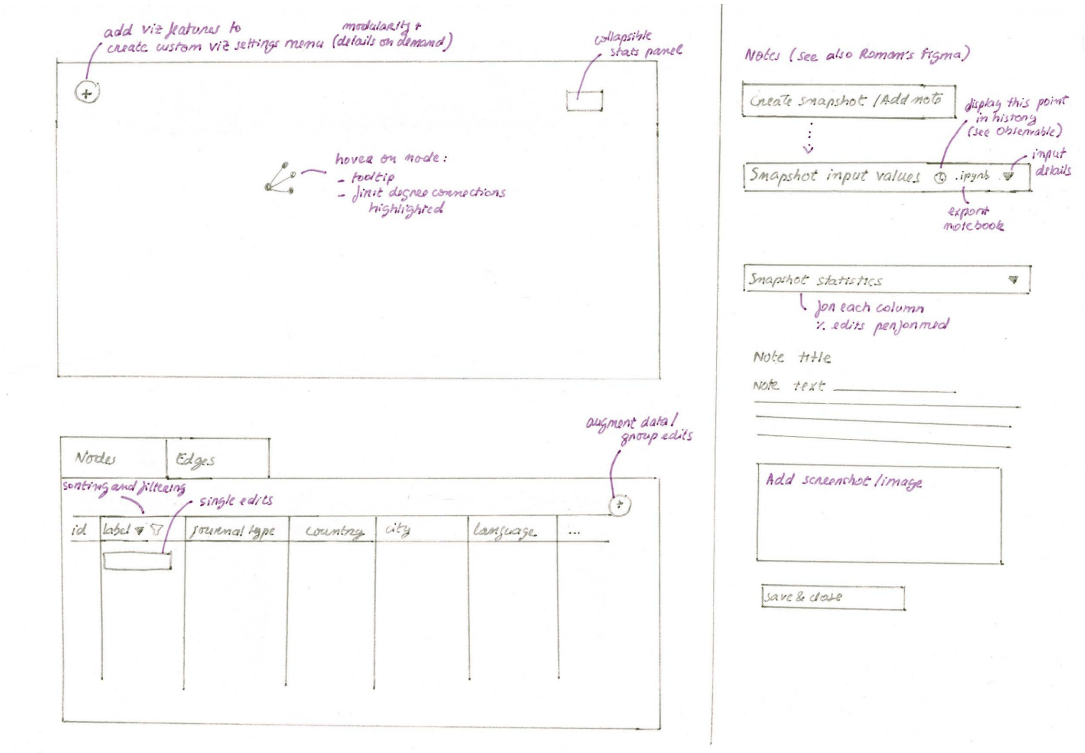

Figure 4: A wireframe outlining *Lumy*'s interface for network analysis. A visualization panel appears at top left and a data panel with tabular data (here separated into a node and an edge table) at bottom left, panels which will be connected to each other and interactive. To the right, the wireframe depicts buttons for taking a snapshot of the current parameters, taking narrative-style notes, and allowing the export of code to an html "codeview" document or a runnable Python notebook. The note panel will also allow the user to save screenshots and relevant citations. Buttons on other screens will also allow the user to view other citations and references that informed the implementation of the underlying methods.

For an overview of how different modules and submodules have been built into the larger scholarly process that the user has enacted, we will provide two functionalities. First, it will be possible to download and export a record of how the data in their current state were produced, as an immutable html document and as an editable and independently runnable Jupyter Notebook. Both the html "codeview" and Jupyter Notebook formats will document the interaction of scholar and source, as code plus notes, as machine process (via lineage tracing, metadata management, code view)

plus thought process (via note-taking function and annotation features), and allow this interaction to be reproduced to create the same output. Second, a more complex and iterative accounting of scholarly engagement will be captured and accessible in a work history pane in *Lumy*. This work history will allow the user to see the paths they have been on – including both those that have been abandoned and those that have been further pursued – and other paths that they could try. All the pieces of the relationship among scholar, source and software will again be preserved *in situ*, in a form that one will be able to review, change or revert to later as their thinking evolves. It will allow a researcher – whether the original user or another – to review former choices and decisions in their entire complexity.

To aid the user in making informed decisions, tools available in *Lumy* will be presented with informational popups describing how the tools were developed and tips for their usage with reference to established literature and techniques. This will help fulfill DHARPA's pedagogical intent to encourage best application of data engineering and textual, spatial or network analytical methodologies even among those who might be unfamiliar with them. It is also just one concrete example of how the VRE is drawing on the diversity of the project team.

5 Onwards

In the spirit of viewing scholarship as an iterative and never ending process rather than a product, we end our article not with a conclusion but a view towards the future of our project. We are looking forward to the moment when we will be able to run workshops to introduce students and our internal and external colleagues to our philosophy of technology-enabled and enhanced critical digital humanities scholarship. As proposed in the days of the original grant application, we aim to position our team into a rapid response lab and extend *Lumy* and *kiara*'s capabilities through new applications proposed by other scholars in the humanities and social sciences employing both qualitative and quantitative methodologies.

In the coming months, DHARPA will release a beta version of our initial modules. Although our frontend, backend and the connections between them are still works in progress, we alert the reader that portions of our code and documentation are already accessible via our GitHub account `https://github.com/DHARPA-Project`. The modularity of our software will allow external developers to build their own modules and workflows to extend *kiara* and *Lumy*. Just as good scholarship relies on peer review and constructive criticism, so too will our open-source VRE, therefore we encourage you to get in touch, share your ideas, and help us advance this project of shining a light on the too-often hidden cogs and gears of humanities research.

References

Daniel Allington, Sarah Brouillette, and David Golumbia. Neoliberal Tools (and Archives): A Political History of Digital Humanities. *Los Angeles Review of Books*, May 2016. URL `https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/`.

M H Beals. Record How You Search, Not Just What You Find: Thoughtfully Constructed Search Terms Greatly Enhance the Reliability of Digital Research,

June 2013. URL https://blogs.lse.ac.uk/impactofsocialsciences/2013/06/10/record-how-you-search-not-just-what-you-find/.

David M. Berry and Anders Fagerjord. *Digital Humanities: Knowledge and Critique in a Digital Age*. Polity, Cambridge, England, June 2017. ISBN 978-0-7456-9766-6.

Timothy Brennan. The Digital-Humanities Bust. *The Chronicle of Higher Education*, October 2017. ISSN 0009-5982. URL http://www.chronicle.com/article/The-Digital-Humanities-Bust/241424.

Wendy Hui Kyong Chun. The Dark Side of the Digital Humanities – Part 1, January 2013. URL https://www.c21uwm.com/2013/01/09/the-dark-side-of-the-digital-humanities-part-1/.

Kathryn Conrad. What the Digital Humanities Can't Do. *The Chronicle of Higher Education*, September 2014. URL http://www.chronicle.com/article/what-the-digital-humanities-cant-do/.

Angela R Cunningham, Markus Binsteiner, Helena Jaskov, and Lorella Viola. "Handmaiden to history"? A conversation on geography in a digital history center. In *American Association of Geographers Annual Conference*, virtual, 2021.

Nan Z. Da. The Digital Humanities Debacle. *The Chronicle of Higher Education*, 65(29), April 2019. URL https://www.chronicle.com/article/the-digital-humanities-debacle/.

Lorraine Daston. Objectivity and the escape from perspective. *Social Studies of Science*, 22(4):597–618, 1992. ISSN 0306-3127. URL https://www.jstor.org/stable/285456.

Nicholas Diakopoulos. Algorithmic accountability reporting: On the investigation of black boxes. Technical report, Tow Center for Digital Journalism, 2014. URL https://doi.org/10.7916/D8ZK5TW2.

James E. Dobson. *Critical Digital Humanities: The Search for a Methodology*. Topics in the Digital Humanities. University of Illinois Press, Urbana, 2019. ISBN 978-0-252-05111-1.

Johanna Drucker. Humanities approaches to graphical display. *Digital Humanities Quarterly*, 005(1), March 2011. ISSN 1938-4122.

Kathleen Fitzpatrick. The Humanities, Done Digitally. *The Chronicle of Higher Education*, May 2011. URL http://www.chronicle.com/article/the-humanities-done-digitally/.

Mike Grimshaw. Towards a manifesto for a critical digital humanities: Critiquing the extractive capitalism of digital society. *Palgrave Communications*, 4(1):1–8, February 2018. ISSN 2055-1045. doi: 10.1057/s41599-018-0075-y. URL http://www.nature.com/articles/s41599-018-0075-y.

Donna Haraway. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, 14(3):575–599, 1988. ISSN 0046-3663. doi: 10.2307/3178066. URL http://www.jstor.org/stable/3178066.

Tim Hitchcock. Confronting the Digital. *Cultural and Social History*, 10(1):9–23, March 2013. ISSN 1478-0038. doi: 10.2752/147800413X13515292098070. URL https://doi.org/10.2752/147800413X13515292098070.

Patrick Jagoda. The Dark Side of the Digital Humanities – Part 3, January 2013. URL https://www.c21uwm.com/2013/01/09/the-dark-side-of-the-digital-humanities-part-3/.

Alan Liu. Where Is Cultural Criticism in the Digital Humanities? In Matthew K. Gold, editor, *Debates in the Digital Humanities*, pages 490–509. University of Minnesota Press, Minneapolis, 2012. URL https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/896742e7-5218-42c5-89b0-0c3c75682a2f.

Miranda Marcus. The God Trick in Data and Design, July 2021. URL https://towardsdatascience.com/the-god-trick-in-data-and-design-4ec71e19811.

Mihhail Matskin, Shirin Tahmasebi, Amirhossein Layegh, Amir H Payberah, Aleena Thomas, Nikolay Nikolov, and Dumitru Roman. A Survey of Big Data Pipeline Orchestration Tools from the Perspective of the DataCloud Project. In *International Conference on Data Analytics and Management in Data-Intensive Domains*, page 16, 2021.

Safiya Umoja Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, New York, 2018. ISBN 978-1-4798-4994-9 978-1-4798-3724-3.

Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, Cambridge, Massachusetts London, England, first harvard university press paperback edition edition, 2016. ISBN 978-0-674-97084-7 978-0-674-36827-9.

Lara Putnam. The transnational and the text-searchable: Digitized sources and the shadows they cast. *The American Historical Review*, 121(2):377–402, April 2016. ISSN 0002-8762. doi: 10.1093/ahr/121.2.377. URL https://doi.org/10.1093/ahr/121.2.377.

Rita Raley. The Dark Side of the Digital Humanities – Part 4, January 2013. URL https://www.c21uwm.com/2013/01/09/the-dark-side-of-the-digital-humanities-part-4/.

Philipp Ruf, Manav Madan, Christoph Reich, and Djaffar Ould-Abdeslam. Demystifying MLOps and Presenting a Recipe for the Selection of Open-Source Tools. *Applied Sciences*, 11(19):8861, 2021. doi: 10.3390/app11198861.

Eric Sheppard. Quantitative Geography: Representations, Practices, and Possibilities. *Environment and Planning D: Society and Space*, 19(5):535–554, October 2001. ISSN 0263-7758. doi: 10.1068/d307. URL https://doi.org/10.1068/d307.

Dennis Tenen. Blunt Instrumentalism: On Tools and Methods. In Matthew K Gold and Lauren F Klein, editors, *Debates in the Digital Humanities 2016*. University of Minnesota Press, Minneapolis, 2016. URL https://dhdebates.gc.cuny.edu/read/untitled/section/09605ba7-ca68-473d-b5a4-c58528f42619#ch09.

Charles B. Travis. *Abstract Machine: Humanities GIS*. Esri Press, Redlands, 2015. ISBN 978-1-58948-368-2.

Lorella Viola and Antonio Fiscarelli. From digitised sources to digital data: Behind the scenes of (critically) enriching a digital heritage collection. In A Webber, M Heerlien, E.G. Miracle, and K Wolstencroft, editors, *Proceedings of the International Conference Collect and Connect: Archives and Collections in a Digital Age*, 2021. URL http://ceur-ws.org/Vol-2810/paper5.pdf.

Scott Weingart. Demystifying networks, parts I & II. *Journal of Digital Humanities*, 1(1), 2011. URL http://journalofdigitalhumanities.org/1-1/demystifying-networks-by-scott-weingart/.

Wikipedia contributors. Mechanical Turk. *Wikipedia*, May 2022. URL https://en.wikipedia.org/w/index.php?title=Mechanical_Turk&oldid=1087252650.