The Clariah Media Suite: A Hybrid Approach to System Design in the Humanities

Melgar-Estrada, L.; Koolen, M.; Beelen, K.; Huurdeman, H.; Wigham, M.; Martinez-Ortiz, C.; Blom, J.; Ordelman, R.

DOI
10.1145/3295750.3298918

Publication date
2019

Document Version
Final published version

Published in
CHIIR'19

The CLARIAH Media Suite: A Hybrid Approach to System Design in the Humanities

Liliana Melgar-Estrada
University of Amsterdam
The Netherlands
melgar@uva.nl

Marijn Koolen
KNAW Humanities Cluster
The Netherlands
marijn.koolen@di.huc.knaw.nl

Kaspar Beelen
University of Amsterdam
The Netherlands
k.beelen@uva.nl

Hugo Huurdeman
University of Amsterdam
The Netherlands
h.c.huurdeman@uva.nl

Mari Wigham
Netherlands Institute for Sound and Vision
The Netherlands
mwigham@beeldengeluid.nl

Carlos Martinez-Ortiz
Netherlands eScience Center
The Netherlands
c.martinez@esciencecenter.nl

Jaap Blom
Netherlands Institute for Sound and Vision
The Netherlands
jblom@beeldengeluid.nl

Roeland Ordelman
University of Twente
The Netherlands
roeland.ordelman@utwente.nl

ABSTRACT

The practices of digital humanists are evolving, highly diversified and experimental. There is also a lack of agreement about whether or not digital humanists should have data and programming skills. Thus, their underlying needs for higher levels of flexibility and transparency may be contradicted by their explicit requests for user-friendly graphic user interfaces (GUIs), creating challenges for designing information systems in the digital humanities. This paper describes the experience of designing the Media Suite, which provides access to important Dutch audiovisual collections and is part of the Dutch infrastructure for digital humanities. We outline a solution to the conflicting needs of scholars, by combining a semi-traditional GUI with Jupyter Notebooks. This solution tackles the needs of both novice and advanced users in digital research methods in the humanities. This demonstration paper explains how the Media Suite and the Jupyter notebooks work together, and elaborates on the rationale behind the design choices. We also outline the implications this hybrid and extensible approach has for interface design for the information science and scholarly community.

CCS CONCEPTS

• **Human-centered computing** → *Command line interfaces*; *Graphical user interfaces*;

KEYWORDS

Complex Tasks; Research Information Systems; Digital Humanities

ACM Reference Format:
Liliana Melgar-Estrada, Marijn Koolen, Kaspar Beelen, Hugo Huurdeman, Mari Wigham, Carlos Martinez-Ortiz, Jaap Blom, and Roeland Ordelman. 2019. The CLARIAH Media Suite: A Hybrid Approach to System Design in the Humanities. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19), March 10–14, 2019, Glasgow, United Kingdom.* ACM, New York, NY, USA, Article 4, 5 pages. https://doi.org/10.1145/3295750.3298918

CHIIR '19, March 10–14, 2019, Glasgow, United Kingdom

ACM ISBN 978-1-4503-6025-8/19/03.
https://doi.org/10.1145/3295750.3298918

1 INTRODUCTION

The emergence of *Digital Humanities* is tied to an increasing data-driven approach to scholarship, in which the so-called *fourth paradigm* in science [7] meets the humanities, challenging their established epistemologies and methods [9]. Digital humanists address research questions by identifying patterns in textual and audio-visual sources (for example, by massively analyzing topics in digitized newspapers or color patterns in films [1]) or by extracting structured data from historical records to reconstruct objects, places, or events from the past.[2] Hence, information systems are no longer used only for searching for information sources, but for the direct analysis, manipulation and experimentation with humanities sources.

In this paper, we introduce a hybrid and extensible solution to the design of information systems to support digital humanists, combining a more traditional search interface with Jupyter Notebooks,[3] a flexible system for interactive data science.

2 BACKGROUND

2.1 The infrastructure project

The Dutch CLARIAH infrastructure project[4] aims to build the Dutch part of the European infrastructures DARIAH[5] (Digital Research Infrastructure for the Arts and Humanities) [5], and CLARIN[6](Common Language Resources and Technology Infrastructure) [10]. The main target user groups of the CLARIAH infrastructure are humanities and social sciences scholars interested in Dutch history and society—taking into account that most sources are in Dutch language. All tools developed in the infrastructure are open source, to benefit the broader digital humanities community, as well as software developers. CLARIAH consists of three working groups centered around data types and the needs of scholars working with them: 1. textual data (linguists), 2. structured data (socio-economic historians), and

[1] E.g. *Mining Shifting Concepts Through Time* project (https://www.esciencecenter.nl/project/mining-shifting-concepts-through-time-shico) and the *Sensoring Moving Image Archive* project (http://sensorymovingimagearchive.humanities.uva.nl/)
[2] E.g. in the projects *Golden agents* (https://www.huygens.knaw.nl/golden-agents/?lang=en) and *Time machine* (https://timemachine.eu/)
[3] http://jupyter.org/
[4] https://www.clariah.nl.
[5] https://www.dariah.eu/
[6] https://www.clarin.eu/

3. audio-visual data (media historians and oral historians). This paper focuses on the third group, which includes a diverse group of scholars from disciplines such as media studies (including film and television studies), oral and political history.

2.2 User requirements

Following the steps for user-centered software design indicated by Toms [20], we evaluated first the requirements of this user group in different ways: via naturalistic observations, personal interviews, design sessions, and workshops (partly described in [12]). We found that, beyond generic functionality such as searching and browsing, the most important scholarly requirements were:

(a) Access to the original sources. Due mostly to copyright and privacy, most of the relevant collections (e.g., the Dutch television archive, or collections of oral history interviews) had previously only been made partially available. The scholars wanted access to the actual media content, and to the complete archival metadata, to preserve the principles of (meta)data transparency (respecting the original metadata from the providers as much as possible), and provenance (mostly for "traceability" to the original source).
(b) *Flexible* and *user-friendly* interfaces that support most phases in the process of simultaneous exploration and analysis of very complex and big, mixed media collections (e.g., the Dutch newspaper archive together with the Dutch television and radio archive).

The first prototype of the so-called "Media Suite," (Section 5.1), appeared after the initial requirements analysis phase.

3 PROBLEM DESCRIPTION

Large-scale infrastructure projects in the humanities and social sciences, such as the aforementioned DARIAH, CLARIN, or the Australian NuHi,[7] aim to provide solutions for preservation and access to collections and data necessary for scholarly research [22].

One of the main challenges of building these infrastructures is catering to methodologically diverse humanities disciplines [9], while at the same time supporting individual research projects which use specific methods. This is what van Zundert [22] has called "the generalization paradox." The requirements in Section 2.2 lead to different types of challenges in designing the infrastructure for the audio-visual group: data governance, sustainable development, user-friendliness, and the need for personalized work spaces [13].

Another important challenge is supporting the scholars' diverse information and data skills. Most scholars only had experience with using graphic user interfaces (GUI), but some used programming and command line tools for data manipulation. This diversity is not uncommon, because, even though the humanities started to experiment with computational approaches in the 1940s [11], the training in data-related work of humanists differs greatly between disciplines and institutions [15]. Besides, it is not yet agreed whether or not digital humanists should have coding skills [15]. Thus, we were faced with another paradox: most users request a low-complexity, user-friendly GUI-based environment, but their underlying needs require more exposure and transparency of the metadata (requirement *a)* above), as well as flexible querying and data manipulation (requirement *b)*). We refer to this as the "flexibility vs. complexity" paradox. Figure 1 indicates that the simpler the system (e.g., a standard GUI), the less flexible and transparent it is for scholarly work. While the most flexible and transparent way to manipulate the data (i.e., using APIs directly) would represent a high level of complexity for those less familiar with the "literacy of code" [23]. There is also evidence that just providing APIs for working with cultural heritage data may not be enough for humanists [5], for the same reasons mentioned above. Therefore, none of the pre-existing infrastructure service models (atomic services like CLARIN [10], or GUI-based VREs like CENDARI [2]) alone seemed to meet the needs of scholars working with audio-visual data. We concluded that we had to implement a hybrid solution, which combines a GUI with support for programmatic access to the APIs.

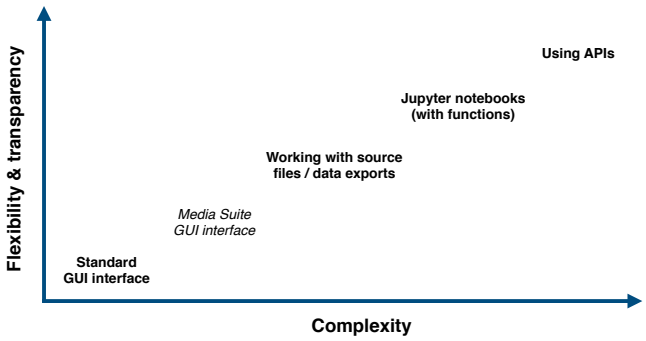

Figure 1: Degree of flexibility vs complexity in data provision for scholarly work.

4 RELATED WORK

There is a growing interest among scientists and data scientists in using Jupyter Notebooks. Querying the Scopus database with the term "Jupyter notebooks," with no restrictions, shows an increasing number of publications either using or referring to them since 2015. But this is mostly the case in scientific domains where, as [18] illustrates, they have become "a robust tool for scientists to share code, associated computation, and documentation," properly aligned with the FAIR principles (Findable, Accessible, Interopeable, Reusable) for scholarly digital research objects [25]. Jupyter Notebooks are also used in industry, for example, at Netflix, where they have become "the de facto standard for quick prototyping and exploratory analysis."[21]. However, except for presentations at some conferences[8], their use in the humanities disciplines is scarcely reported.

[7] https://huni.net.au/

[8] See for example the workshop at the Benelux 2018 Conference: http://2018.dhbenelux.org/workshops/#delpher)

Figure 2: Example of transparency features in the MS.

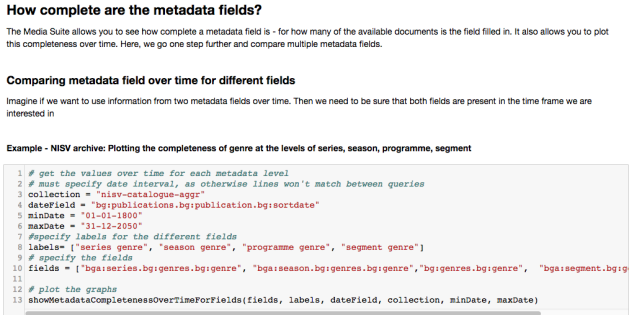

Figure 3: Example of a parameterizable Jupyter notebook that complements the MS.

5 IMPLEMENTATION APPROACH

5.1 The Resulting GUI

The Media Suite[9] is a virtual research environment (VRE) [4] where access to data and tools is authorized using a federated authentication mechanism. The indexes are built on Elasticsearch[10], offering the following features accessible via the GUI:

(1) *A metadata inspector*, which shows the metadata completeness of the different collections and completeness analysis features (e.g., per metadata field, or for data enrichments), Figure 2.
(2) *Traditional Boolean and faceted search and browsing features.*[11] We augmented this with flexible metadata selection possibilities per provider (e.g., the creation of custom facets for searching).
(3) *Basic visualizations* for data exploration, such as time lines of number of hits per year;
(4) *Personal work space and "user projects"*, [12] which allow users to create their own project spaces to store collections of bookmarks, saved queries, and user annotations created by the annotation tool included in the Media Suite [1];
(5) *Automatic enrichment services and data.* Currently, this includes automatic speech recognition (ASR), both for bulk processing and indivisual collections [14];
(6) *Basic export facilities* of user bookmarks and annotations.

5.2 Beyond the GUI: Jupyter notebooks

To move beyond the GUI, and so overcome the "flexibility vs complexity paradox", we integrate Jupyter Notebooks, a web-based UI for interactively writing and running code as well as visualizing outputs, into the research environment. Jupyter Notebooks are applications to create and share documents that contain live code, equations, visualizations and narrative text. They allow users to run Python (and use about 40 more programming languages) in a web browser, and surround their code with cells that include narrative text in Markdown [16].

For the Media Suite, a cloud-based service for the notebooks was offered. This service: (1) facilitates their use, since users do not have to install any extra software to run the notebooks, and (2) adds more security, since the notebooks become available only to users who have passed the authentication step of the Media Suite. Based on user needs elicited via the different studies described in the previous section, we populated the notebooks with different pre-built functions. These allow the users to specify parameters in the code and accept input values at runtime [21]. We then prepared queries and visualizations to enable our users to fulfill some of the tasks that the Media Suite could not handle.

The functions that we included originated from our analysis of scholarly tasks. These were meant either to:

(1) *Extend existing GUI functionality.* For this purpose, we created different notebooks, which used the same name of the GUI features that they extend. For example, the MS offers an "inspector" tool for analyzing the metadata completeness of certain fields (Figure 2), but this can be done only for one field per collection in the GUI. The function integrated in the complementary "inspector" notebook allows for combining more than one metadata field from various collections (Figure 3). Because users may not know the technical labels of the metadata fields that should be used in the notebooks, the MS provides a "metadata dictionary," where the definitions of the metadata fields are included (Figure 2).
(2) *Add new functionalities.* For example, the Media Suite offers the time-labeled, automatic speech recognition (ASR) transcripts of one important part of the Netherlands Institute for Sound and Vision (NISV) collection, which can be searched, and used as interactive transcripts for the fragment-level navigation of individual resources. However, processing pipelines that take speech transcripts as input to analyse occurrences of named entities or to generate visualisations (e.g., word clouds) by applying natural language processing (NLP) tools, are not yet implemented, or are not feasible to implement given the flexibility that is required and the existence of libraries that are freely available. The "Analysis" notebook offers this flexibility: by using the *annotation* and *search APIs* in the notebook, the users can get the ASR transcripts for

[9] http://mediasuite.clariah.nl/
[10] https://www.elastic.co/products/elasticsearch
[11] Recent developments also include the simultaneous querying of Sparql end-points
[12] Some parts, like collaborative projects, are work in progress

their bookmarked resources, and generate word frequencies and related visualizations based on their analysis since the notebook has an integrated NLP toolkit.[13]

(3) *Experiment with new Media Suite functionalities.* Besides extending the Media Suite, the *extension* role of the notebooks facilitates the task of co-developing (experimenting and testing new functionalities with scholars), e.g., word clouds in this case, before actually implementing them in the GUI.
(4) *Provide dynamic collection overviews.* Finally, they also provide dynamic collection overviews that can be usefully deployed for both scholars and institutional archivists to answer questions such as: how many hours of television do we have in the archive? Are speech transcripts available for all news programmes? What percentage of the material is already digital? Are documentaries longer now than they were in the 70's? We further describe these and other advantages of the notebooks from an archival and scholarly perspective in Wigham et al. [24].

5.3 Initial Evaluation

The resulting hybrid version combining GUI and Jupyter Notebooks was informally evaluated during the first CLARIAH media studies Summer School, where circa forty scholars worked on eight group research projects using the Media Suite[14]. The scholars received explanation of the notebooks in a workshop, and had the opportunity to try out some examples during their working sessions, assisted by data scientists or developers, who also had knowledge of the collections. More formal evaluations are planned.

6 DEMONSTRATION OVERVIEW

We will demonstrate the combination of the Media Suite (MS)'s GUI and the Jupyter notebooks in this way:

(1) Via a walk-through of the GUI. We introduce the data that are available, and functionalities described in Section 5.1. Next, we illustrate how the MS's GUI works via a scholarly use case selected for the purpose, of a scholar investigating the representation of refugees in the Dutch mainstream media (radio, television, and newspapers). We show how the MS supports the different stages of scholarly work ([3, 12]), from collection overview and *Data criticism*, [6, 8] to then gathering a representative sample (i.e. a corpus), to the analysis (e.g. manual annotation) of that sample, concluding with further analysis and synthesis (exporting and interpreting the resulting corpus and their annotations).
(2) Via a walk-through of the Jupyter notebooks. First, we explain the service model (authentication and APIs), then show the types of notebooks that we offer, the specific functions that we have pre-built in the notebooks and how users can customize and extend them.
(3) Finally, we illustrate how the MS's GUI and the notebooks complement each other, which focuses attention on two scholarly tasks. First, *Data criticism* occurs when scholars are looking for overviews of the collection and need to observe the completeness of the collections' metadata. We show what levels of this type of preliminary analysis are supported by the MS GUI, and how the dedicated notebook for this task makes the analysis more complete and flexible. This illustrates the purpose of `Extending` existing functionality via the notebooks. Second, *Preliminary data analysis* occurs when scholars are looking for patterns in the portions of the collection(s) that they have selected. This illustrates the purpose of `Adding` new functionality via the notebooks.

7 DISCUSSION: EXPECTED IMPACT

A large part of the work of the digital humanities scholar depends on data manipulation (almost 80 percent of the work is data preparation [17]). This includes a wide variety of tasks that cannot be executed within the confines of a GUI. Therefore, some scholars view these interfaces as "golden cages", where rich data is only partially unlocked via pre-determined functionalities and pre-selected interactions. Certainly, the intrinsic value of these interfaces in the humanities is that they facilitate wider access and "close reading" [19] the resources, supporting user annotations, as well as collaboration. But, designing systems in a digital humanities context has an inherently experimental nature, and an intrinsic instability as scholarly methods are gradually changing. There is a call for system design in the humanities to be open and transparent, that is, where the "source code" can be the subject of analysis and critique. As van Zundert [22] indicated, "especially now that more digital tools are getting integrated into the methodology of humanities, the adequacy and validity of analyses depend to a certain extent on an adequate understanding of such specific rules." In this sense, Jupyter Notebooks allow testing and experimentation, with full potential for encouraging transparency in reading and sharing the code that builds some of the parts of the GUI. Moreover, the notebooks provide an extensible and more interactive research environment since the code represents open building blocks that users can reuse and or reconfigure. This hybrid solution allows a shift in scholarly practice by lowering the threshold for transparency of method and data processes. The challenges for system design are important, since we assume that as demand for this transparency grows in digital scholarship, hybrid and extensible solutions to the design of GUIs like the one we proposed here will be more needed. We anticipate that the creation of prepared functions in the notebooks will require a more direct connection between data scientists and traditional scholars for creating meaningful and common (yet parameterizable) functions that can support them in their data-related tasks. Finally, the hybrid solution we propose makes scholars more compelled to deal with the trade-offs between flexibility and complexity in system design.

8 ACKNOWLEDGEMENTS

The research for this paper was made possible by the CLARIAH-CORE project (www.clariah.nl) financed by NWO. The work described is the result of the collaboration between software developers and scholars of CLARIAH WP5.[15] We specifically acknowledge the contributions of Jonathan Blok and Willem Melder to this work.

[13] NLTK, see http://www.nltk.org/
[14] https://clariah.github.io/mediasuite-blog/blog/2018/10/01/Clariah-Media-Studies-Summer-School-report
[15] People involved are listed here: http://mediasuite.clariah.nl/documentation/faq/who-develops

REFERENCES

[1] Jaap Blom, Marijn Koolen, Liliana Melgar Estrada, Peter Boot, Ronald Haentjens Dekker, Christian Olesen, Susan Aasman, Norah Karrouche, and Rob Wegter. 2017. A Demonstration of Scholarly Web Annotation Support Using the W3C Annotation Data Model and RDFa. http://easychair.org/smart-program/BIGVID2017/2017-11-24.html

[2] Nadia Boukhelifa, Mike Bryant, Nataša Bulatović, Ivan Čukić, Jean-Daniel Fekete, Milica Knežević, Jörg Lehmann, David Stuart, and Carsten Thiel. 2018. The CENDARI Infrastructure. *Journal on Computing and Cultural Heritage* 11, 2 (April 2018), 1–20. https://doi.org/10.1145/3092906 arXiv: 1612.05239.

[3] Marc Bron, Jasmijn van Gorp, and Maarten de Rijke. 2016. Media studies research in the data-driven age: How research questions evolve. *Journal of the Association for Information Science and Technology* 67, 7 (2016), 1535–1554. https://doi.org/10.1002/asi.23458 jasist2015-bron-media.pdf.

[4] Christopher Brown. 2013. Implementing a virtual research environment (VRE). https://www.jisc.ac.uk/guides/implementing-a-virtual-research-environment-vre

[5] Jennifer Edmond, Vicky Garnett, and Agiatis Benardou. 2014. *So we've built it, but have they come? Investigating barriers and opportunities for API usage among the AHSS community.* Workshop Europeana Cloud WP1. Europeana Professional, The Hague, Netherlands. http://pro.europeana.eu/europeana-cloud/about-europeana-cloud

[6] Frederick W. Gibbs. 2016. New Forms of History: Critiquing Data and Its Representations. *The American Historian* (Feb. 2016). http://tah.oah.org/february-2016/new-forms-of-history-critiquing-data-and-its-representations/

[7] Tony Hey, Stewart Tansley, and Kristin Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery.* https://www.microsoft.com/en-us/research/publication/fourth-paradigm-data-intensive-scientific-discovery/

[8] Rik Hoekstra and Marijn Koolen. 2018. Data Scopes: towards Transparent Data Research in Digital Humanities – DH2018. Mexico. https://dh2018.adho.org/en/data-scopes-towards-transparent-data-research-in-digital-humanities/

[9] Rob Kitchin. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1 (Sept. 2014), 1–12. http://eprints.maynoothuniversity.ie/5364/

[10] Steven Krauwer and Erhard Hinrichs. 2014. The CLARIN Research Infrastructure: Resources and Tools for e-Humanities Scholars. http://dspace.library.uu.nl/handle/1874/307981

[11] Gerhard Lauer. 2014. Challenges for the Humanities: Digital Infrastructures. In *Facing the Future : European Research Infrastructures for the Humanities and Social Sciences*, Adrian Duşa, Dietrich Nelle, Günter Stock, and Gert G. Wagner (Eds.). Scivero Verlag, Berlin. https://edoc.bbaw.de/frontdoor/index/index/docId/2327

[12] Liliana Melgar Estrada, Marijn Koolen, Hugo Huurdeman, and Jaap Blom. 2017. A process model of time-based media annotation in a scholarly context. In *CHIIR 2017: ACM SIGIR Conference on Human Information Interaction and Retrieval.* Oslo.

[13] Roeland Ordelman, Carlos Martínez Ortíz, Liliana Melgar Estrada, Marijn Koolen, Jaap Blom, Willem Melder, Jasmijn van Gorp, Victor De Boer, Themistoklis Karavellas, Lora Aroyo, Thomas Poell, Norah Karrouche, Eva Baaren, Johannes Wassenaar, Oana Inel, and Julia Noordegraaf. 2018. Challenges in Enabling Mixed Media Scholarly Research with Multi-media Data in a Sustainable Infrastructure – DH2018. Mexico. https://dh2018.adho.org/en/challenges-in-enabling-mixed-media-scholarly-research-with-multi-media-data-in-a-sustainable-infrastructure/

[14] Roeland J.F. Ordelman and Adrianus J. van Hessen. 2018. Speech Recognition and Scholarly Research: Usability and Sustainability. In *CLARIN 2018 Annual Conference*, Inguna Skadina and Maria Eskevich (Eds.). 163–168.

[15] James O'Sullivan, Diane Jakacki, and Mary Galvin. 2015. Programming in the Digital Humanities. *Digital Scholarship in the Humanities* 30, suppl_1 (2015), i142–i147. https://doi.org/10.1093/llc/fqv042

[16] Fernando Perez and Brian E. Granger. 2015. Project Jupyter: Computational Narratives as the Engine of Collaborative Data Science. https://blog.jupyter.org/project-jupyter-computational-narratives-as-the-engine-of-collaborative-data-science-2b5fb94c3c58

[17] Gil Press. 2016. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

[18] Bernadette M. Randles, Irene V. Pasquetto, Milena S. Golshan, and Christine L. Borgman. 2018. Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. *arXiv:1804.05492 [cs]* (2018). http://arxiv.org/abs/1804.05492 arXiv: 1804.05492.

[19] Kathryn Schulz. 2011. The mechanic muse - what is distant reading? *The New York Times* (June 2011). http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html

[20] Elaine G. Toms. 2009. User-Centered Design of Information Systems. https://doi.org/10.1081/E-ELIS3-120043525

[21] Michelle Ufford, Matthew Seal, and Kyle Kelley. 2018. Beyond Interactive: Notebook Innovation at Netflix. https://medium.com/netflix-techblog/notebook-innovation-591ee3221233

[22] Joris van Zundert. 2012. If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities. *Historical Social Research / Historische Sozialforschung* 37, 3 (141) (2012), 165–186. http://www.jstor.org/stable/41636603

[23] Joris van Zundert and Ronald Haentjens Dekker. 2017. Code, scholarship, and criticism: When is code scholarship and when is it not? *Digital Scholarship in the Humanities* 32, suppl_1 (2017), i121–i133. https://doi.org/10.1093/llc/fqx006

[24] Mari Wigham, Liliana Melgar Estrada, and Roeland Ordelman. 2018. Jupyter Notebooks for generous archive interfaces. In *IEEE Big Data 2018: 3rd Computational Archival Science (CAS) Workshop.* Seattle, WA. https://dcicblog.umd.edu/cas/ieee-big-data-2018-3rd-cas-workshop/

[25] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Nature* 3 (March 2016), 160018. https://doi.org/10.1038/sdata.2016.18