

Towards 'Stakeholder Readiness' in the CLARIAH Media Suite: Future-Proofing an Audio-Visual Research Infrastructure

Roeland Ordelman, Jasmijn Van Gorp, Willemien Sanders, Mari Wigham, Rana Klein, Jaap Blom, Willem Melder, Martijn van de Donk, Johannes Wassenaar, Frank Sträter, Christian Olesen, Victor de Boer, Mary-Joy van der Deure, Norah Karrouche, Jasper Keijzer, Asli Ozgen-Havekotte, Alexander Badenoch, Megan Phipps, Julia Noordegraaf¹

¹Netherlands Institute for Sound & Vision, University of Amsterdam, Utrecht University, Free University Amsterdam, University of Twente

Abstract

Digital heritage and cultural data are fundamental to research in the Social Sciences and Humanities (SSH) Tasovac et al. (2020). Research institutes and cultural heritage organisations are collaborating in cross-sector digital infrastructure projects to make the distributed and typically siloed, large, and heterogeneous multimedia *heritage data* available to scholars, seamlessly on their personal computers in the laboratory or at home. As such, SSH researchers are the main stakeholders in the development of online research environments. This development involves the implementation of sustainable applications and services for fine-grained qualitative or quantitative analysis of these data in all of its forms, including scholarly information activities or 'primitives' Blanke and Hedges (2013), Pacheco (2022), Unsworth (2000), such as searching, browsing, collecting, comparing, referring, annotating, analysing, representing and publishing.

The CLARIAH Media Suite¹ is an example of an online application or 'virtual research environment' that aims to facilitate these activities for scholars working with 'media data' in the Netherlands, such as radio and television broadcasts, and film, including a varied set of contextual data (e.g., program guides, viewer ratings, film posters), historical newspapers, and oral history interviews. The Media Suite has been developed in the course of several projects over the past years and the number of available 'media' data sets is growing steadily in terms

¹ <https://mediasuite.clariah.nl>

of quantity and types. These projects have in common a close collaboration between scholars, archivists, computer scientists and developers.

One of the key decisions in building the Media Suite in the very beginning, was to put these stakeholders, and specifically their 'scholarly workflow' central to its development, shifting focus from just providing 'closed tools' with a specific functionality, to providing 'recipes' that may combine several functionalities Martinez-Ortiz et al. (2017). This forced us to 'think modular': which steps does a scholarly workflow consist of and what APIs and (UI) components do we need to develop to support this work? During the years many scholars have used and assessed the various Media Suite functionalities for their research. In addition to scholarly publications, they provided us with useful feedback. For example, a collection building component can be combined with a collection metadata inspection component; a search component should include a data visualization component (e.g., histograms of query results); and a media browser component in turn includes an annotation component.

Some of these components are 'custom made' for the Media Suite, such as the collection search that is specific for the data collections from various institutions. Others can be provided by the larger CLARIAH digital infrastructure in the Netherlands that Media Suite is part of. An example is the annotation component that uses standardized annotation formats (i.e. W3C Web Annotations) that a scholar may want to use in multiple research environments, including in the Media Suite as well as in an external application that allows for fine-grained annotation of sound and images, such as ELAN Auer et al. (2010).

A special category of components that the Media Suite is using consists of infrastructure and data management services, typically provided on an infrastructure level. A crucial service to allow scholars to work with closed heritage and research data is a national service that Media Suite takes advantage of for authentication (who is this person) & authorisation (what is this person allowed to access), to enable access to data sets with restricted access (e.g., due to copyright or privacy concerns). More recently, services for facilitating 'computational humanities' are increasingly asked for, ranging from environments to 'program with data' (e.g., Jupyter Notebooks and Hubs) to on-demand analysis tools such as speech recognition and computer vision, and computation facilities for training AI models.

While developing the Media Suite and providing it to scholars for doing research, it quickly became evident that the implementation of digital workflows on the one hand and documentation and guidance on the other, are two sides of the same coin. The Media Suite team put a lot of effort in making available FAQs, tutorials and curated playlists in a special Media Suite 'Learn' section, based on use cases. Apart from keeping this information up-to-date, the challenge is to develop a sustainable model for creating, storing and curating these materials. Also in this domain, a 'modular approach' focusing on small and connectable building blocks of information, disseminated on established platforms such as Zenodo², may be the way to go.

By putting the scholarly workflow at the heart of our thinking about the Media Suite we aim for so called 'stakeholder readiness' instead of 'technology readiness'. How well digital tools fit in the digital workflows and common practices of the stakeholders, the scholars, determines the success of those tools, rather than the maturity of the (underlying) digital technology alone. As such, we took the scholarly workflow as our central focus for analysing and prioritising the

² <https://zenodo.org/>

feedback we have received about the Media Suite.

As the current project, CLARIAH-PLUS, finishes at the end of this year, we are confident we can fix the prioritised issues in the Media Suite upgrades scheduled for this year. At the same time, we will consider the design of the Media Suite for the longer term development: in order to best support scholarly workflows, which workflow components should remain central to the Media Suite itself, and which components can better be reused from elsewhere, such as the CLARIAH infrastructure?

For both short term fixes and long term design, our aim is to improve stakeholder readiness. This leaves us with essential questions: how to keep the infrastructure up to speed with the evolving user needs of a diverse and changing digital humanities community, and how to evaluate how we are doing? In this presentation, we discuss how we applied the concept of scholarly workflow to process and prioritise feedback on the Media Suite, and how we will use it to guide the modular redesign of the Media Suite. We will also address the question of how to measure stakeholder readiness, in terms of how well digital tools fit the scholarly workflow.

Acknowledgement This work was made possible by the CLARIAH-PLUS project funded by NWO (Grant 184.034.023).

References

- Eric Auer, Albert Russel, Han Sloetjes, Peter Wittenburg, Oliver Schreer, Stefano Masnieri, Daniel Schneider, and Sebastian Tschöpel. Elan as flexible annotation framework for sound and image processing detectors. In *Seventh conference on International Language Resources and Evaluation [LREC 2010]*, pages 890–893. European Language Resources Association (ELRA), 2010.
- Tobias Blanke and Mark Hedges. Scholarly primitives: Building institutional infrastructure for humanities e-Science. *Future Generation Computer Systems*, 29(2):654–661, 2 2013. ISSN 0167-739X. doi: 10.1016/J.FUTURE.2011.06.006.
- Carlos Martinez-Ortiz, Roeland Ordelman, Marijn Koolen, JJ Noordegraaf, Liliana Melgar, Lora Aroyo, Jaap Blom, J van Gorp, E Baaren, KJ Beelen, et al. From tools to “recipes”: Building a media suite within the dutch digital humanities infrastructure clariah. *DH Benelux*, pages 78–81, 2017.
- André Pacheco. Digital humanities or humanities in digital: revisiting scholarly primitives. *Digital Scholarship in the Humanities*, 3 2022. ISSN 2055-7671. doi: 10.1093/lsc/fqac012.
- Toma Tasovac, Sally Chambers, and Erzsébet Tóth-Czifra. Cultural Heritage Data from a Humanities Research Perspective: A DARIAH Position Paper. Technical report, 2020. URL <https://hal.archives-ouvertes.fr/hal-02961317>.
- John Unsworth. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this, 2000.