

Jasmijn Van Gorp, Liliana Melgar Estrada and Julia Noordegraaf

# Involving Users in Infrastructure Development: Methodological Reflections From the Research Pilot Projects Using the CLARIAH Media Suite

## Abstract

In this contribution we introduce a compilation of articles that reflect on the use of the CLARIAH Media Suite and its impact on research methods, as they were conducted during the CLARIAH research pilot projects (2017-18). The pilot projects fit in a co-development approach in which users are involved in infrastructure development from the start. We discuss how feedback of the pilot researchers was incorporated in the iterative development and testing of the Media Suite.

## Keywords

tools, digital humanities, archives, collaboration, research methods

The project *Images for the Future* (2007-2014) provided funding for the massive digitisation of audiovisual heritage in The Netherlands.<sup>1</sup> More than 200,000 hours of film, video and audio have been digitised. In addition, approximately 38,000 hours of digital-born material are ingested directly into the archives of Netherlands Institute for Sound and Vision on an annual basis.<sup>2</sup> The wealth of sources that have become digitally available, and the emergence of new tools to analyse those collections at scale, have fuelled expectations on methodological innovations that can lead to new breakthroughs in media scholarship.<sup>3</sup> At the same time, several humanities scholars have pointed out that such innovations require careful analysis and consideration of concrete user needs. In 2012, Joris van Zundert argued that large-scale, generic infrastructure solutions might in fact hinder such innovations.<sup>4</sup> Instead, he pleaded for smaller-scale solutions for research with digital sources and methods focused on user needs of scholars and knowledge exchange.<sup>5</sup> Relatedly, Sabine Lenk highlighted the importance of collaboration between the owners of digital collections and the scholars using them.<sup>6</sup> Responding to the large-scale digitisation of the *Images for the Future* project, she emphasised that scholars can bring their knowledge

and expertise and as such may meaningfully contribute to the accessibility and usability of the digital sources.

We followed this plea for involving users and collection owners in infrastructure development when building the Media Suite<sup>7</sup> in the context of the CLARIAH project<sup>8</sup>, that aims to develop the Dutch national infrastructure for digital humanities research. The Media Suite provides a virtual research environment for accessing and researching audiovisual sources and related contextual sources. In 2017-18, the development of the Media Suite included its testing and use in six research pilot projects that were selected in an open competition.<sup>9</sup> The aim of the pilots was to involve the researchers during the period of one year in the co-development of the infrastructure together with the information technology experts and the other scholars in the CLARIAH Media Studies team.<sup>10</sup> To implement this co-development process, we asked the research pilot project teams to write so-called demonstration scenarios, in which they outline all the steps in a research process focusing on one specific theme. These scenarios, then, were the sources that provided concrete requirements for the development of the Media Suite. The process of going back and forth between the description of a concrete research process and the data and tools made available via the Media Suite invited methodological reflection. Two of the six pilot project teams have published their methodological reflections during their projects.<sup>11</sup> In this dossier in *TMG Journal for Media History*, the researchers of the four other projects share their thoughts on the methodological dimensions of doing media scholarly research with the CLARIAH Media Suite sources and tools.

The CLARIAH Media Suite was officially launched in April 2019. It is a virtual research environment that provides access to audiovisual collections and the tools to search, visualise, enrich, and analyse them. The Media Suite provides online access to, among others, the entire audiovisual collection of The Netherlands Institute for Sound and Vision<sup>12</sup>, the Eye Film Museum Jean Desmet collection (included by UNESCO in their Memory of the World register)<sup>13</sup> and the oral history interview collections maintained by DANS<sup>14</sup>. Previously, these collections could only be accessed separately or on location because of copyright and privacy issues. The Media Suite works as a data aggregator of those collections, and as a research environment in which scholars can access both the metadata and the media content itself. Researchers can log in to this environment using their institutional affiliation credentials, and then search, explore, browse, store their selections, analyse and annotate, and visualise the outcomes of their analysis in a personal workspace.

A barrier to tailor infrastructures to scholarly work is that, traditionally, the design of information systems was mostly focused on the technical aspects, incorporating user input only at the very end of the chain. As Claire Warwick recounts, it was generally considered as simply not useful to ask users about what they needed ‘because they would not know’ and to ask them to provide feedback ‘because they would not care’.<sup>15</sup> The technologists were the ones who knew best what to do and how they would work. A more *user-centred design* methodology started to gain ground already from the mid-1980s onwards and moved the attention to the users’ needs as the starting point for designing information systems.<sup>16</sup> Franciska De Jong et al. explain as follows: ‘The typical ICT cycle starts with a user study and the analysis of the requirements that need to be fulfilled in a scenario of use or for a specific task for which users seek technology support.’<sup>17</sup> With the research pilots, we followed this approach to involve users and implemented a two-staged format for co-development.

First, the pilot scholars were asked to write a so-called ‘demonstration scenario’ document, which is a description of the context and processes of the users’ goals and research phases, in a specific *scenario of use* for an infrastructure which is ideal for the use case but is not in place yet. They follow the typical steps of research (from exploration and selection over contextualisation and analysis to publication) and describe their actions. In order to be specific, they refer to other existing facilities to provide examples of the envisioned tools and functionalities.<sup>18</sup> Through those demonstration scenarios, the information and system experts can understand how the system could support the scholars’ goals and can draft a list of requirements. For the development of the Media Suite, these requirements were documented in the format of a *user story* (see Fig. 1), allowing the developers to understand the goals and context of the request.<sup>19</sup> The requirements were subsequently ranked according to priority and grouped together in terms of technical dependencies, after which the developers could start implementing them according to priority. The pilot scholars, then, would make use of the system, providing feedback after every release that was incorporated to the iterative development and testing of the Media Suite.

Secondly, while doing their research, the research pilot scholars were also invited to reflect upon the affordances of the digital data and tools in the Media Suite in-the-making. The idea was to encourage a ‘tool criticism’ approach’.<sup>20</sup> That is, to invite scholars to be explicit about how they use a specific tool in their research process and evaluate what it brings and where its limitations lie. Reflection is typically done at the very end of a project, by revisiting the research process, the results and research notes. In the case of the research pilots, they could not only make use of their research

Als <gebruiker>	wil ik <functionaliteit>	zodat <meerwaarde>	gedurende <stage>
For whom? (use group)	Functionality (brief description)	Purpose of functionality (brief description)	Possible research stage to which the functionality applies
Researcher	I want to have insight in the provenance of information in the metadata fields	so I know the source and the information is more transparent	Unknown
Researchers	I would like to have the option of creating my own user space	So I can store, curate&manage, analyze, and create presentations based on my own c	ALL
Curators & Reser	I would like to have the option of creating a visual analysis of the collections/data sets on my user space	so I can have precise/real time information of the collections I am using	Analysis
Researchers	I would like to have the option of annotating the objects that are in my own user space	so I can perform data analysis for my research	Analysis
Researchers	I would like to have the option of adding codes to fragments of different media	so I can perform coding for the analysis of themes and discourse	Analysis
Researchers	I would like to have the option of improving transcriptions of audiovisual material from automatic speech recogno	so I can perform better quality analysis	Analysis
Researchers	I would like to have the option of synchronizing uploaded transcriptions of audiovisual material	so I can perform better quality analysis	Analysis
Researchers	I would like to be able to extract (manually or semi-automatically) structured data from different textual sources	so I can register these data for further analysis	Analysis
Researchers	I would like to analyze patterns from specific selected fragments and create new annotations	so I can focus on my close reading of specific parts of media works	Analysis
Researchers	I would like to connect online social media data to other media (e.g., radio, television programs)	so I can perform cross-media analysis	Analysis
Curators & Reser	I would like to have the option of uploading my own collections to my own user space	so I can perform cross-data analysis for my research and use different tools with my co	Unknown
Researchers (ma	I want to be able to continue my work over different working sessions and keep a registry of my previous sear	so I can work on my research during a long period of time	Unknown
Researchers & J	I want to be able to access the stored items I have bookmarked during browsing or searching collections with c	so I can use them as my corpus (or purchase requests in case of media producers...)	Exploration & Assembling
Curators & Reser	I want to be able to access my shared items	so I can know what items I have already shared	Unknown
Curators & Reser	I want to be able to access the annotations I created during browsing or searching collections	so I can use them together with my corpus	ALL
Curators & Reser	I want to be able to curate and manage my collections in my user space	so I can have control on editing the collections (e.g., naming)	Unknown
Curators & Reser	I want to be able to curate my own annotations in my user space	so I can have control on updating my personal taxonomies or code lists	Analysis
Researchers	I want to be able to create visualizations, data analyses, or "narratives" using functionalities and data available	so I can present the results of my analysis	Presentation
Curators & Reser	I want to be able to share my browsing/searching histories that are stored in my user space	so I can show to other people how I have performed a search/browsing task	Presentation (and maybe ex
ALL	I want to be able to find documentation, source code, clear guidelines on how to use the CLARIAH tools, colle	so I can obtain support from CLARIAH for my research	ALL

Figure 1. Sample of user stories as compiled by the WP5 team based on previous user studies and demonstration scenarios.

notes, but also compare the results with their initial demonstration scenarios and reflect on what they wanted to do at the start of the project, to the results and the obstacles encountered along the way. This second stage generated an additional set of requirements for adjusting the tools and interface for later versions of the Media Suite, which were intended to be implemented in the years after the pilot projects ended.

It is this second, reflective phase that forms the basis of the contributions in this dossier, which present and discuss the methodological aspects or questions that came to the fore when using the Media Suite. The pilot scholars meticulously show how they worked with the limitations and challenges as imposed by using a large-scale research infrastructure in development. In addition, they provide insight into the methods they designed for working with the then available functionalities of the Media Suite and suggest concrete routes for improvement. The contributions shed a unique light on the research practice in the digital humanities by providing real life and concrete examples of use of a specific environment for researching audiovisual data and contextual sources. By doing so, the dossier also responds to the recommendation of Fred Gibbs and Trevor Owens to incorporate concrete methodological examples to prospective users of digital tools.<sup>21</sup>

The four pilot scholars that contributed to this dossier all hail from different backgrounds and taut a variety of collection-expertise. Historian Susan Hogervorst, principal investigator of the CrossEWT project, investigated the topics discussed in oral history interviews conducted with World War II veterans. With her project, she also contributed to gain understanding on how the Media Suite could support historians and especially historians working with oral history collections. Her project is

centred around the accounts of eyewitnesses about World War II in the Netherlands since 1945, using a cross-media approach to working with newspapers, television documentaries, and oral history interviews simultaneously. In her reflection, she focuses on the difficulties for building her corpus during the initial phase in which the CKAN collection registry<sup>22</sup> and the Media Suite were built, and how she found workarounds and solutions to achieve her research goals. She ends her chapter with a reflection on the benefits of video annotation tools for oral history research and the wider challenges of using an infrastructure built for audiovisual collections for historical research.

Film scholar Christian Olesen was the principal investigator of the MIMEHIST project, which facilitated the integration of the UNESCO world heritage Jean Desmet collection into the Media Suite and simultaneously investigated the connections between films, posters, and Jean Desmet's business documents. In his contribution to this dossier, Olesen focuses on the methodological background and rationale for using and providing input for tailoring the existing annotation tool provided by the Media Suite to facilitate mixed media film analyses.<sup>23</sup> Enhancing the annotation tool was one of the aims in tackling the problem of establishing and exploring meaningful links between mixed media collections such as the Jean Desmet's collection.

Media scholars Sabrina Sauer and Berber Hagedoorn, principal researchers of the NarDis project, investigate in their contribution how to trace back the creation of stories or narratives of disruptive events by television, newspapers and radio reports and how these are instilled with specific cultural-political meanings. They look at how researchers tell stories about two specific disruptive events, namely the assassination of Dutch politician Pim Fortuyn and the North Sea flood of 1953, based on nine user sessions with the Media Suite. They elaborately discuss the steps and challenges that their participants encountered when using the former DIVE+ recipe<sup>24</sup> (currently named Explore Tool) and an early version of the Comparative Search recipe (currently named Compare Tool) in the Media Suite. In their final section, they outline pathways for future functionalities, and make a strong plea for the incorporation of more collections and especially collections in linked open data format that can be used across tools.

Lastly, journalism scholars Mark Vallinga, Frank Harbers and Marcel Broersma, the research team of the ReSpoNs project, investigated how sports journalism in newspapers has changed in response to the rise of television between 1959 and 1989. They show how they worked with the first versions of the Compare tool in the Media Suite and with Delpher<sup>25</sup> as one of their main collections were newspapers. In their article, they compare challenges and pitfalls by juxtaposing the textual newspaper collection





Figure 2. Researchers at work with the CLARIAH Media Suite. Photo taken at CLARIAH Summerschool organised at Netherlands Institute for Sound and Vision in the Summer of 2018 (CC BY 2.0 Sebastiaan Ter Burg).

and the audiovisual television collection. They make recommendations on how an improved Compare Tool can do better justice to the specificities of these collections, and how this in the end could lead to cross-media research within one infrastructure. By the time of this publication, the CLARIAH Media Suite is fully functional and aimed to be sustained at least for the forthcoming decade. The pilot projects have been conducted with Media Suite version 2 and 3, whereas presently version 6.1 is available.<sup>26</sup> The majority of the needs envisioned by the research pilot scholars in their contributions, therefore, are already implemented in the new versions of the Media Suite. This included, for example, the option to do comparative research across the Radio and Television collections and newspaper collections, which is now possible via the Compare tool, and the integration of ASR (automatic speech recognition) transcriptions for a large part of the Radio and Television collection, which now can be searched in an equivalent of full text search for the newspapers in Delpher. Other wishes will be possible with the implementation of Jupyter Notebooks that will enable direct work with the data from these collections.<sup>27</sup> We invite all readers of *TMG Journal for Media History* to explore the methodological reflections in this dossier and to visit the Media Suite via [mediasuite.clariah.nl](https://mediasuite.clariah.nl) to dive into the various audiovisual collections and digital tools. Ideally, this will result in new use cases and user requirements that will inform the future improvement of the Media Suite as a core infrastructural facility for researchers studying audiovisual collections and related materials.

## Acknowledgements

The authors would like to thank the initiators and editors of the dossier, Frank Harbers and Marcel Broersma, copy editor Thunnis van Oort, the research pilot scholars who contributed to the dossier, and Eva Baaren for providing support to the research pilot scholars during their projects. This publication has been made possible by the support of the NWO-infrastructure project CLARIAH-PLUS.

## Notes

1. “Beelden voor de Toekomst”, Beeld en Geluid, accessed October 29, 2021, <https://www.beeldenvoordetoeekomst.nl/index.html>.
2. It is 8,000 hours of video, 30,000 hours of radio, adding up to 300 TB on an annual basis. This exact amount is based on unpublished slides of Netherlands Institute for Sound and Vision, “Inleiding tot Nederlands Instituut voor Beeld en Geluid” (Hilversum, 2018).
3. Tara McPherson, “Media Studies and the Digital Humanities,” *Cinema Journal* 48, no. 2 (2009): 119-123.
4. Van Zundert considers infrastructures for the humanities as large-scale digital networks that offer (general) digital tools to work with and provide access to digital collections. He specifically refers to CLARIAH as one of the examples of large-scale infrastructures.
5. Joris van Zundert, “If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities,” *Historical Social Research / Historische Sozialforschung*, 37, no. 3 (2012): 165–186, DOI: 10.2307/416636603.
6. Sabine Lenk, “Images for the Future: Will They Make TV Scholars Happy?” *Critical Studies in Television* 5, no. 2 (2012): 80-85, DOI: 10.7227/CST.5.2.11.
7. “About”, Media Suite, accessed October 29, 2021, <https://mediasuite.clariah.nl/>; Roeland Ordelman, Liliana Melgar, Jasmijn Van Gorp, Julia Noordegraaf, “Media Suite: Unlocking Audiovisual Archives for Mixed Media Scholarly Research,” *Selected papers from the CLARIN Annual Conference* (2018): 133-143.
8. The development of the Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH, [www.clariah.nl](http://www.clariah.nl)) was made possible by the CLARIAH Core (2014-2019) and CLARIAH Plus (2019-2024) projects funded by NWO under the Large Scale Research Infrastructure grant scheme.

9. The call for applications to this “Research Pilot” grant was launched in 2017; sixteen projects were awarded in the domains of socio-economic history, linguistics, media studies and oral history (<https://www.clariah.nl/en/projects?categories=research-pilots>). The six pilots that involved the CLARIAH Media Suite were NarDis (Sabrina Sauer and Berber Hagedoorn, RUG), ReSpoNs (Marcel Broersma, Frank Harbers and Mark Vallinga, RUG), MIMEHIST (Christian Olesen), M&M (Susan Aasman), DReAM (Toine Pieters and Berrie van der Molen, UU) and CrossEWT (Susan Hogervorst, EUR).
10. “WP5: Media Studies”, Clariah, accessed October 29, 2021, <https://www.clariah.nl/wp5-media-studies>; “People,” Clariah, accessed October 29, 2021, <https://www.clariah.nl/people>.
11. Susan Aasman, Liliana Melgar Estrada, Tom Sloomweg, and Rob Wegter, “Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History,” *VIEW Journal of European Television History and Culture* 7, no. 14 (31 december 2018): 73–87, DOI: 10.18146/2213-0969.2018.jethc154; Berrie van der Molen, Jasmijn Van Gorp and Toine Pieters, “Operationalizing ‘Public Debates’ Across Digitized Heterogeneous Mass Media Datasets in the Development and Use of the Media Suite,” *Selected papers from the CLARIN Annual Conference* (2018): 205-213, [http://www.ep.liu.se/ecp/article.asp?issue=159&article=021&volume=.](http://www.ep.liu.se/ecp/article.asp?issue=159&article=021&volume=)
12. “Netherlands Institute for Sound and Vision”, Clariah, accessed October 29, 2021, <http://mediasuitedata.clariah.nl/organization/netherlands-institute-for-sound-and-vision>.
13. “EYE (Film Institute Netherlands)”, Clariah, accessed October 29, 2021, <http://mediasuitedata.clariah.nl/group/eye-film-institute-netherlands>.
14. “Oral History (Aggregated collections)”, Clariah, accessed October 29, 2021, <http://mediasuitedata.clariah.nl/dataset/dans-oral-history>.
15. Claire Warwick, “Studying Users in Digital Humanities,” in *Digital Humanities in Practice*, ed. Claire Warwick, Melissa Terras and Julianne Nyhan (London: Facet, 2018): 1-22, DOI: 10.29085/9781856049054.002.
16. Elaine G. Toms, “User-Centered Design of Information Systems,” in *Encyclopedia of Library and Information Sciences* (3rd edition), ed. Marcia J. Bates and Mary Niles Maack (Boca Raton: CRC Press, 2015), DOI: /10.1081/E-ELIS3-120043525.
17. Franciska M.G. de Jong, Roeland J.F. Ordelman and Stef Scagliola, “Audio-visual Collections and the User Needs of Scholars in the Humanities: A Case for Co-Development,” *Proceedings of the 2nd Conference on Supporting Digital Humanities (SDH 2011)*, <https://research.utwente.nl/en/publications/audio-visual-collections-and-the-user-needs-of-scholars-in-the-hu>.



18. See the original demonstration scenario as developed by Jasmijn Van Gorp for the AVResearcherXL project in CLARIN here: [http://dev.clarin.nl/sites/default/files/QuaMeRDES\\_M2\\_demonstrationscenario\\_final.pdf](http://dev.clarin.nl/sites/default/files/QuaMeRDES_M2_demonstrationscenario_final.pdf)
19. The user story format was ‘As a [role], I would like to have a [functionality] in order to [user need] during [stage].’ For example ‘As a researcher, I would like to create a time chart in order to get an overview of the collection during the exploration phase.’
20. Marijn Koolen, Jasmijn Van Gorp and Jacco van Ossenbruggen, “Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice,” *Digital Scholarship in the Humanities* 34, no. 2 (2018): 368-385.
21. Fred Gibbs and Trevor Owens, “Building Better Digital Humanities Tools: Towards Broader Audiences and User-Centred Designs,” *Digital Humanities Quarterly* 6, no. 2 (2021). <http://digitalhumanities.org/dhq/vol/6/2/000136/000136.html>.
22. “Search Data”, Clariah, accessed October 29, 2021, <https://mediasuitedata.clariah.nl/>.
23. Aasman et al., “Tales of a Tool Encounter”.
24. In the early versions of the Media Suite, we named tools ‘recipes’, a combination of functionalities that support the steps to accomplish a specific task.
25. “Delpher”, accessed October 29, 2021, <https://www.delpher.nl/>.
26. To get an overview of all the different changes and versions of the Media Suite, please consult the release notes: <https://mediasuite.clariah.nl/documentation/release-notes>.
27. Jupyter notebooks are open source web environments that contain live code, visualisations and narrative text (<https://jupyter.org/>). See also Mary Wigham, Liliana Melgar Estrada and Roeland J.F. Ordelman, “Jupyter Notebooks for Generous Archive Interfaces,” in *2018 IEEE International Conference on Big Data (Big Data): 3rd Computational Archival Science (CAS) Workshop*, <https://dcicblog.umd.edu/cas/ieee-big-data-2018-3rd-cas-workshop/>.

## Biographies

**Jasmijn Van Gorp** is assistant professor Audiovisual Data Studies at Utrecht University. Together with Julia Noordegraaf and Roeland Ordelman, she leads CLARIAH WP5 since 2019. She is working within the CLARIAH project from the start (2014) and co-developed its predecessor tools CoMeRDa (2012), AVResearcherXL (2013) and TROVe (2013).

**Liliana Melgar Estrada** is data specialist at the SKILLNET project at Utrecht University. Between 2015-2020 she was postdoctoral researcher at the University of Amsterdam and Utrecht University, conducted user studies for CLARIAH WP5 systematizing user requirements, supporting the pilot scholars and investigating scholarly annotations of audiovisual sources.

**Julia Noordegraaf** is professor of Digital Heritage at the University of Amsterdam, where she leads the digital humanities research program and lab Creative Amsterdam: An E-Humanities Perspective. Since 2014, she has been the leader of CLARIAH WP5 and has acted as board member of CLARIAH for the domain of audiovisual data and media studies.

**TMG Journal for Media History**

Volume 24 No (1/2)/2021

**DOI**<https://dx.doi.org/10.18146/tmg.809>**PUBLISHER**

Netherlands Institute for Sound and Vision

**COPYRIGHT**

Each article is copyrighted © by its author(s) and is published under license from the author(s). When a paper is accepted for publication, authors will be requested to agree with the Creative Commons Attribution 4.0 International License.