# CLARIAH Media Suite: Enabling Scholarly Research for Audiovisual and Mixed Media Data Sets in a Sustainable, Distributed Infrastructure

Carlos Martinez-Ortiz[1], Roeland Ordelman[2/8], Marijn Koolen[3], Julia Noordegraaf[4], Liliana Melgar[2/4], Lora Aroyo[5], Jaap Blom[2], Victor de Boer[5], Willem Melder[2], Jasmijn Van Gorp[7], Eva Baaren[2], Kaspar Beelen[4], Norah Karrouche[5/6], Oana Inel[5], Iris van Vliet[7], Johan Oomen[2], Themis Karavellas[2], Johannes Wassenaar[2], Eduardo Silva Navarrete[2] and Thomas Poell[4]

[1]Netherlands eScience Center [2]Netherlands Institute for Sound and Vision [3]KNAW Humanities Cluster [4]University of Amsterdam [5]VU University [6]Erasmus University Rotterdam [7]Utrecht University [8]Twente University

## Abstract

The Media Suite[1] is one of the components of the Dutch research infrastructure CLARIAH[2], which aims to serve the needs of media scholars, or digital humanists in general, by providing access to large audiovisual collections, that are distributed across various content providers, and their contextual data. The goal is to provide a user-friendly infrastructure, with applications to work with that data in a variety of scholarly projects. This infrastructure consists of multiple data sets and tools, it is accessible via a web-portal and is tailored to the needs of specific scholarly uses. It serves the needs for working with audiovisual data collections and related mixed-media contextual sources that are maintained by cultural heritage institutions and knowledge institutions such as The Netherlands Institute for Sound and Vision, EYE Film Institute of the Netherlands and the National Library of The Netherlands (KB).

The CLARIAH Media Suite is an authenticated environment, where data and metadata are aggregated and where users, in a personal workspace, can explore, browse and store these data, manually annotate or automatically enrich subsets of data, visualize results and export bookmarks and annotations. The Media Suite was first released in April 2017, and a second version

[1] http://mediasuite.clariah.nl/
[2] https://www.clariah.nl/

was released in December 2017 ( Figure 1), the third release is planned in June 2018.

## Approach

In order to develop this environment in a sustainable way, that can be used and developed further after the project's lifetime, we need to carefully align the requirements of scholars with the context of the ecosystem the Media Suite needs to live in: an ICT infrastructure hosted and maintained by multiple institutions that in turn, adheres to a diverse set of institutional requirements with respect to, for instance, data access permissions and software development and maintenance.

Access to and use of the majority of media studies related collections is often restricted, especially when they concern audiovisual media, due to intellectual property rights (IPR) or privacy issues. The approach of the CLARIAH Media Suite to tackle these existing challenges of access is (1) to organise and implement a federated authentication mechanism (a login) to overcome access barriers (Figure 2, number 5) for which we use a SURFConext application, and (2) to provide mechanisms that enable researchers to work with tools and aggregated data *within* the infrastructure. We refer to this approach as "bringing the tools to the data", as opposed to "bringing the data to the tools". These objectives have been accomplished in the first versions of the Media Suite, allowing scholars to search and analyse audio-visual collections via a central workspace, thus, enabling *data intensive research* in the humanities. Figure 2 shows the main elements/building blocks that constitute the Media Suite research environment (Ordelman et al., 2018).

## Tools

The digital humanities community incorporates a wide diversity of scholars with different interests, goals, methods, and levels of expertise in working with information processing techniques and technologies. We addressed this challenge by (1) focussing on the similarities in research methods from different disciplines (e.g., De Jong, Ordelman & Scagliola, 2011; Bron, Van Gorp & de Rijke, 2016; Melgar et al., 2017), (2) by analyzing tools that support qualitative methods (Melgar & Koolen, 2018), and (3) by working

with scholars as co-developers in the process. The resulting functionalities are built in a modular approach that supports both flexible software development of components and user friendly interaction with assembled tools.
The Media Suite tools offer the core functionalities needed for performing scholarly research tasks with audio-visual media and contextual collections. The tools available in version 3 of the Media Suite enable researchers to access data/collections and perform tasks with them. These include: inspecting the collection's metadata structure and completeness, browsing, searching/exploring, and viewing the media objects (in most cases). The Media Suite also makes it possible for researchers to visualize patterns in the metadata, compare different queries or collections, annotate and enrich media items, and export data.

**Workspace**
The CLARIAH Media Suite provides a personal and collaborative workspace (see Figure 3), that stores private session user data such as bookmarks, (manual) annotations, and search sessions. It furthermore enables collaboration with other scholars. Forthcoming implementations will include automatic speech recognition services for corpora and personal collections.

**The Media Suite Demonstration**
The aim of our proposed demonstration at the 2018 DH Benelux conference is to introduce and demonstrate the latest version of the CLARIAH Media Suite: version 3 (to be launched at the end of June 2018).

## References

**[Bron et al. (2016]** Marc M. Bron, Jasmijn van Gorp, and Maarten de Rijke, Media studies research in the data-driven age - How research questions evolve. *Journal of the Association for Information Science and Technology*, 67 (7), 2016.

**[de Jong et al. (2011)]** Franciska de Jong, Roeland Ordelman, and Stef Scagliola. *Audio-visual collections and the user needs of scholars in the humanities: a case for co-development.* In Proceedings of the 2nd Conference on Supporting Digital

Humanities (SDH 2011), Copenhagen, Denmark, 2011. Centre for Language Technology, Copenhagen.

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

**[Ordelman et al. (2018)]** Ordelman, R., Melgar, L., Martínez-Ortiz, C., Blom, J., Melder, W., Karavellas, T., Wassenaar, J., Silva Navarrete, E., Baaren, E.,van der Werf, L., de Boer, V., Aroyo, L., Noordegraaf, J., Karrouche, N., van Gorp, J., Poell, T., Beelen, K. *Enabling Scholarly Research for Distributed Audiovisual and Mixed Media Data Sets in a Sustainable Infrastructure*. Paper for the Digital Humanities Conference (DH2018), submitted for publication, Mexico city, Mexico, 2018.

**[Melgar, L., & Koolen, M. (2017)]** Liliana Melgar and Marijn Koolen. Audiovisual media annotation using Qualitative Data Analysis Software: a comparative analysis. *The Qualitative Report*, in print.

**[Martínez Ortiz et al. (2017)]** Martínez Ortiz, C., Ordelman, R., Koolen, M., Noordegraaf, J., Melgar Estrada, L., Aroyo, L., Poell, T., Blom, J., de Boer, V., Melder, W., van Gorp, J., Beelen, K., Baaren, E., Karrouche, N., Inel, O., Kiewik, R., Karavellas, T. *From Tools to "Recipes": Building a Media Suite within the Dutch Digital Humanities Infrastructure CLARIAH.* Presented at the Digital Humanities Benelux, Utrecht.

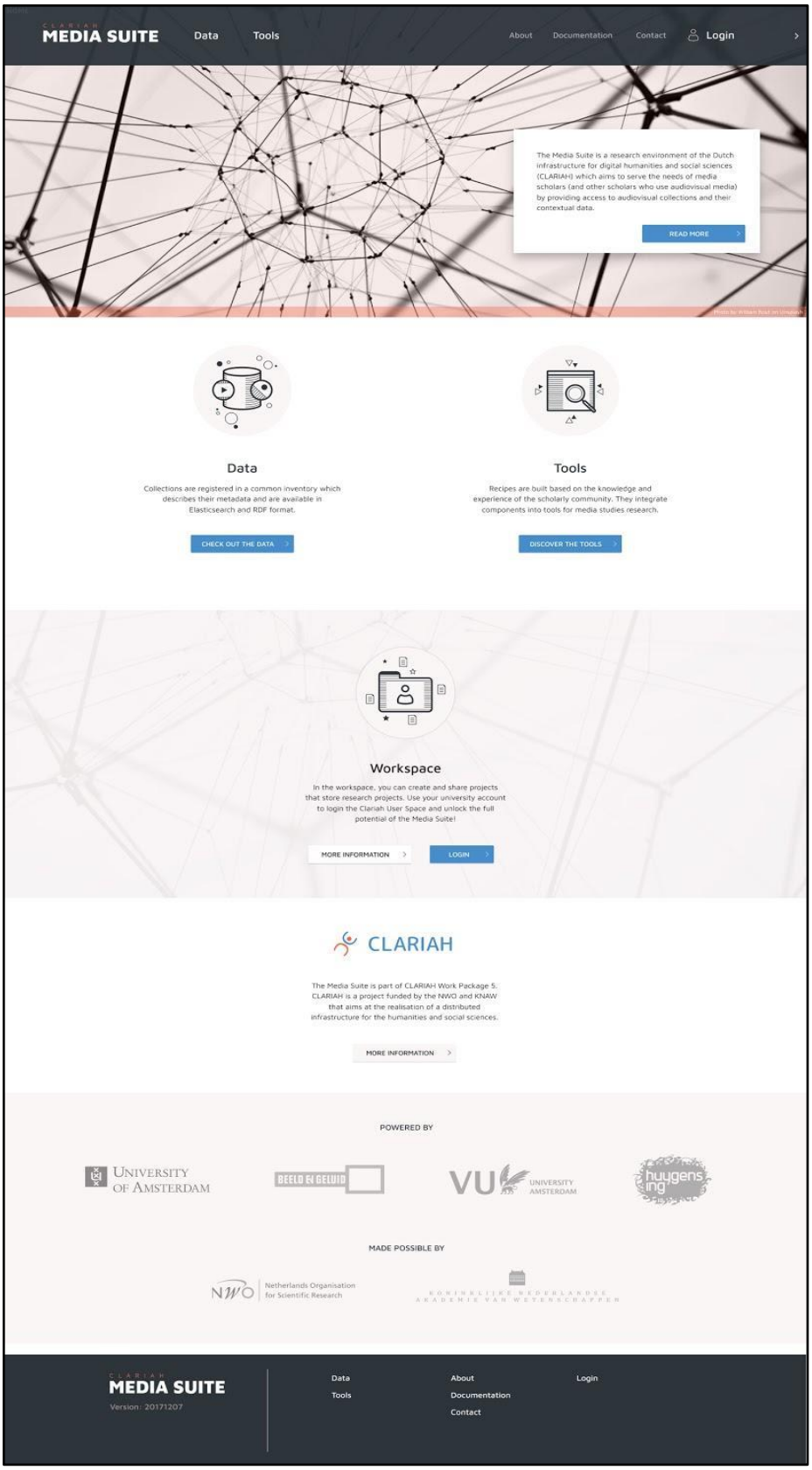


Figure 1. The CLARIAH Media Suite's homepage (version 2)

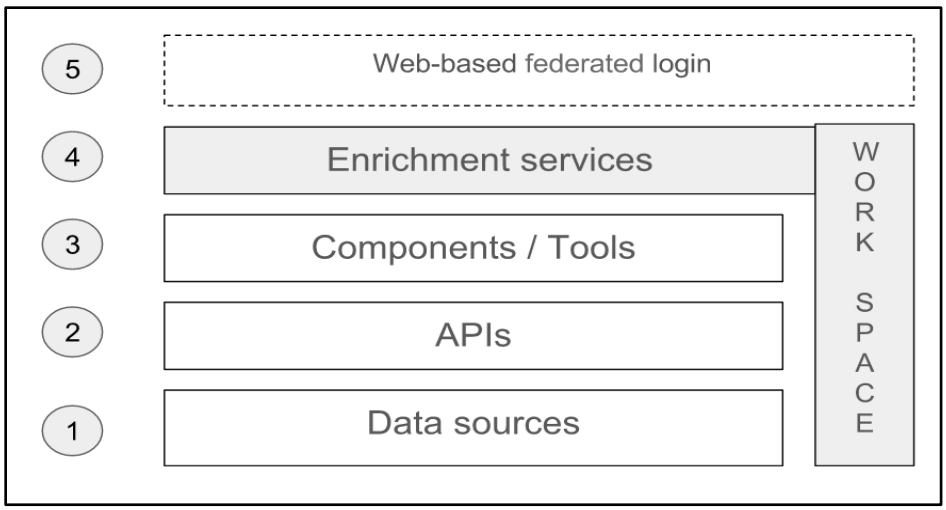

Figure 2. The building blocks of the CLARIAH Media Suite

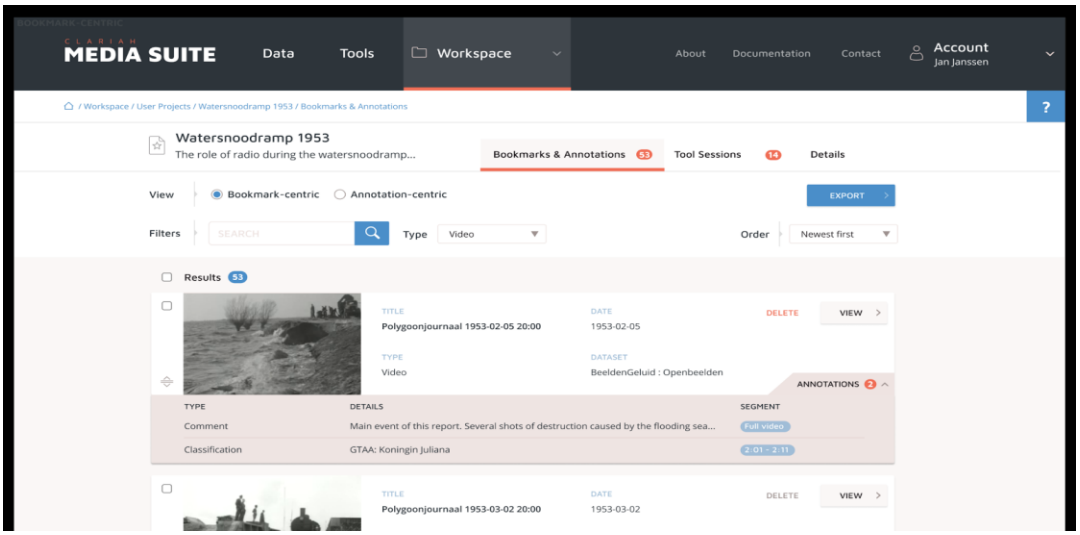

Figure 3. The CLARIAH Media Suite ‘Workspace’