Future-saving audiovisual content for Data Science: Preservation of geoinformatics video heritage with the TIB|AV-Portal

Peter Löwe, Margret Plank, and Frauke Ziedorn

Technische Informationsbibliothek TIB, Development, Hannover, Germany (peter.loewe@tib.uni-hannover.de)

In data driven research, the access to citation and preservation of the full triad consisting of journal article, research data and -software has started to become good scientific practice. To foster the adoption of this practice the significance of software tools has to be acknowledged, which enable scientists to harness auxiliary audiovisual content in their research work.

The advent of ubiquitous computer-based audiovisual recording and corresponding Web 2.0 hosting platforms like Youtube, Slideshare and GitHub has created new ecosystems for contextual information related to scientific software and data, which continues to grow both in size and variety of content. The current Web 2.0 platforms lack capabilities for long term archiving and scientific citation, such as persistent identifiers allowing to reference specific intervals of the overall content.

The audiovisual content currently shared by scientists ranges from commented howto-demonstrations on software handling, installation and data-processing, to aggregated visual analytics of the evolution of software projects over time. Such content are crucial additions to the scientific message, as they ensure that software-based data-processing workflows can be assessed, understood and reused in the future.

In the context of data driven research, such content needs to be accessible by effective search capabilities, enabling the content to be retrieved and ensuring that the content producers receive credit for their efforts within the scientific community.

Improved multimedia archiving and retrieval services for scientific audiovisual content which meet these requirements are currently implemented by the scientific library community. This paper exemplifies the existing challenges, requirements, benefits and the potential of the preservation, accessibility and citability of such audiovisual content for the Open Source communities based on the new audiovisual web service TIB|AV Portal of the German National Library of Science and Technology. The web-based portal allows for extended search capabilities based on enhanced metadata derived by automated video analysis. By combining state-of-the-art multimedia retrieval techniques such as speech-, text-, and image recognition with semantic analysis, content-based access to videos at the segment level is provided. Further, by using the open standard Media Fragment Identifier (MFID), a citable Digital Object Identifier is displayed for each video segment.

In addition to the continuously growing footprint of contemporary content, the importance of vintage audiovisual information needs to be considered: This paper showcases the successful application of the TIB|AV-Portal in the preservation and provision of a newly discovered version of a GRASS GIS promotional video produced by US Army -Corps of Enginers Laboratory (US-CERL) in 1987. The video is provides insight into the constraints of the very early days of the GRASS GIS project, which is the oldest active Free and Open Source Software (FOSS) GIS project which has been active for over thirty years. GRASS itself has turned into a collaborative scientific platform and a repository of scientific peer-reviewed code and algorithm/knowledge hub for future generation of scientists [1].

This is a reference case for future preservation activities regarding semantic-enhanced Web 2.0 content from geospatial software projects within Academia and beyond.

References:
[1] Chemin, Y., Petras V., Petrasova, A., Landa, M., Gebbert, S., Zambelli, P., Neteler, M., Löwe, P.: GRASS GIS: a peer-reviewed scientific platform and future research