

2016

## OSGeo conference videos as a resource for scientific research: The TIB|AV Portal

Peter Lemmens

*TIB Hannover*

Margret Plank

*TIB Hannover*

Angelina Kraft

*TIB Hannover*

Britta Dreyer

*TIB Hannover*

Follow this and additional works at: <https://scholarworks.umass.edu/foss4g>

 Part of the [Cataloging and Metadata Commons](#), [Geographic Information Sciences Commons](#),  
and the [Scholarly Publishing Commons](#)

---

### Recommended Citation

Lemmens, Peter; Plank, Margret; Kraft, Angelina; and Dreyer, Britta (2016) "OSGeo conference videos as a resource for scientific research: The TIB|AV Portal," *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings*: Vol. 16 , Article 6.

DOI: <https://doi.org/10.7275/R5BC3WR6>

Available at: <https://scholarworks.umass.edu/foss4g/vol16/iss1/6>

This Paper is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact [scholarworks@library.umass.edu](mailto:scholarworks@library.umass.edu).

# OSGeo conference videos as a resource for scientific research: The TIB|AV Portal

Peter Löwe\*, Margret Plank, Kraft Angelina, Britta Dreyer

TIB Hannover, Germany – peter.loewe|margret.plank|angelina.kraft|britta.dreyer@tib.eu

**KEYWORDS:** Audiovisual content, Digital Object Identifiers, Citation, OSGeo, Preservation, Semantic Web, Information Mining, Citation, Altmetrics, Video Abstracts

## ABSTRACT:

*This paper reports on new opportunities for research and education in Free and Open Source Geoinformatics as a translational part of Open Science, enabled by a growing collection of OSGeo conference video recordings at the German National Library of Science and Technology (TIB). Since 2015, OSGeo conference recordings have been included to the collection sphere of TIB in information sciences. Currently, video content from selected national (FOSSGIS), regional (FOSS4G-NA) and global (FOSS4G) conferences is being actively collected. The annual growth exceeds 100 hours of new content relating to the OSGeo software projects and the OSGeo scientific-technical communities. This is seconded by retrospective acquisition of video material dating from past conferences, going back until 2002 to preserve this content, ensuring both long term availability and access. The audiovisual OSGeo-related content is provided through the TIB|AV Portal, a web-based platform for scientific audiovisual media providing state-of-the art multimedia analysis and retrieval. It implements the requirements by research libraries for reliable long term preservation. Metadata enhancement analysis provides extended search and retrieval options. Digital Object Identifiers (DOI) enable scientific citation of full videos, excerpts and still frames, use in education and also referral in social networks. This library-operated service infrastructure turns the audiovisual OSGeo-related content in a reliable source for science and education.*

## 1 The Challenge of Scientific Technical Information in OSGeo

OSGeo, the Open Source Geospatial Foundation, is an international non-profit and nongovernmental organization, to promote open access to geospatial data and sustainable and future-safe open source software development (AUTHOR & Neteler, 2014). Similar to the Apache Foundation, members contribute to foundation projects and community governance. Communication within OSGeo involves organisation committees, project maintainers, developers, application related topics and also education and reach-out. From the library perspective, much of this consists of scientific technical information (STI) and is conducted on alternative channels beyond the traditional journal-based scientific discussion. The advent of ubiquitous digital recording equipment and screen-cast-software has led to a steady rise of audiovisual content, such as lecture recordings or data animations. The first OSGeo-related video recordings date from 2002 (Open Source Free Software GIS - GRASS users conference, 2002). This trend continues and is likely to accelerate further due to the widespread use of handheld end-user devices to access such content, a diversification of application scenarios (“software mash-ups”) and the growing number of OSGeo-approved software projects. Until now, the majority of this content has been distributed via proprietary commercial Web 2.0 platforms (e.g. Youtube.com) or isolated academic websites. Access to and reuse of such content in both science and industry has been hampered. This is due to the lack of effective deep content retrieval based on rich-metadata, reliable scientific citation and long term preservation. These topics, while crucial for long term availability and reuse, are both beyond the business models of the currently still used Web 2.0 audiovisual platforms and also exceed the capabilities of the maintainers of conference websites. Since its beginning in 2002, the amount of Open Source Geospatial related video recordings has continued to grow and diversify.

Effective and efficient access to and mining of this distributed repository of audiovisual Scientific-Technical Information holds great potential for the monitoring of the state of the art in Geoinformatics, analysis of trend

---

\* Corresponding author

patterns and innovation, and also technology transfer to new fields. This is part of a larger challenge currently affecting academia regarding access to heterogeneous STI content as part of the general trend towards data-driven science.

## 2 The German National Library of Science and Technology

The German National Library of Science and Technology (TIB) ranks as one of the largest specialized libraries worldwide: Its mission is to comprehensively acquire and archive literature from around the world pertaining to all areas of engineering, architecture, chemistry, information technology, mathematics and physics. The TIB's information portal provides access to more than 160 million data sets from specialised databases, publishers and library catalogues. The TIB development branch and the Competence Centre for non-textual Materials drive application focused research and development towards innovative library services. This especially applies to the fields of scientific-technical information in the greater field of Open Science and other non-textual information. For this, tools and infrastructure are developed to support users in scientific work processes, enabling the easy publication, discoverability and long-term availability of scientific-technical information. One example is the DOI service to provide digital object identifiers (DOI names), because the long-term digital preservation of scientific objects is a major task nowadays. As a DOI registration agency for scientific content, the TIB offers a solution for the problem mentioned above. A DOI consists of a unique character string that identifies an entity in a digital environment - in other words: it identifies the object itself and not the place where it is located. The DOI system was started in 1998 by the International DOI Foundation (IDF) (IDF 2002) and was standardized as ISO 26324 in 2012. In 2004 the TIB, in collaboration with scientific institutes, developed an infrastructure model for the DOI registration, establishing a complete workflow for the referencing of research data. Five years later this work led to the founding of the globally oriented non-profit organisation DataCite, initiated by TIB. Further DataCite members include for example the British Library, the California Digital Library, the Library of ETH Zürich and the Australian National Data Service. DataCite offers an infrastructure that supports simple and effective methods of data citation, discovery and access.

## 3 The TIB|AV Portal Video Platform

Launched in April 2014, the TIB|AV-Portal ([av.tib.eu](http://av.tib.eu)) is a bilingual (English/German) web-based portal for audiovisual media that optimizes access to scientific videos in the field of science and technology. It was designed to overcome the limitations encountered in current commercial Web 2.0 video portals according to the requirements of a data-driven research library. By combining state-of-the-art multimedia retrieval techniques with semantic analysis, it provides content-based access to videos at the segment level and the ability to link data to new knowledge. The processing workflow of the video analysis includes structural analysis based on video shot detection, optical character recognition, automated speech-to-text transcription, and visual concept detection.

### 3.1 Metadata of the TIB|AV Portal and the Linked Open Data Context

The DataCite metadata schema, is based on the Dublin Core Metadata Standard (<http://dublincore.org/>), was adapted to describe and manage audiovisual content by formal, technical and content-describing authoritative metadata in a standardized manner (Lichtenstein et al., 2014). Automatically generated metadata are processed via linguistic and semantic analysis, i.e. named entities are identified, disambiguated and mapped to an authoritative knowledge base. This knowledge base consists of subject specific parts of the Integrated German Authority Files (GND), available as Linked Data Services of the German National Library. This allows for semantical enrichment of the data and context information concerning authors and institutions from external databases and leads to an improvement of the interoperability of the data (Sack and AUTHOR 2014). The English labels were gained by mapping GND entities onto other authority files like the DBpedia and the Library of Congress Subject headings. (Strobel 2014). These technologies improve access to scientific videos by e.g. enabling pinpoint search of individual video segments. Further content-based filter facets for search results enable the exploration of the increasing number of videos.

The automatically generated and time-coded authoritative metadata of each video are stored as RDF data (<https://www.w3.org/RDF/>) in a Triple-Store and are openly published as data dumps according to linked-data standards for further use RDF export data dumps (<http://av.tib.eu/opendata/>).

### 3.2 Features and Services

Each video within the TIB|AV Portal is registered by a digital object identifier (DOI). A DOI name clearly identifies the video, akin to the use of ISBN in books. In addition, the TIB|AV Portal offers a time-based citation link, enabling a citable DOI to be displayed for each video segment using the open standard media fragment identifier (MFID).

This approach enables users to use and share references to subsections and still-frames of a video. A visual table of contents provides a quick overview of the video facilitating access to individual video sequences. Content-based filter facets for search results enable the exploration of the increasing number of video assets. The keyword search is not only performed within authoritative metadata but also within metadata from video analysis, giving different term weight when searching. These techniques allow the users to search more efficiently and find content that otherwise would remain hidden. Producers of scientific films, such as the OSGeo communities, can upload their videos to the TIB|AV-Portal free of charge. Once the quality of the video has been checked, it is published in a legally watertight manner: indexed according to international standards; transcribed; digitally preserved; and given a DOI name. This ensures an optimal discoverability of scientific films.

### 3.3 OSGeo content acquisition

Within the TIB|AV Portal, films from the technical fields introduced in section 2 are currently collected in both German and English language, including recordings of conference presentations, panel discussions, recordings of experiments, e.g. microscopic images, modelling, simulations and presentations of specific software. Most of the videos are published under a Creative Commons licence (<https://creativecommons.org/>).

Starting in 2015 the TIB subject librarians added video recordings from selected OSGeo conferences to the collection sphere of TIB. This had been preceded by the successful acquisition trial for the historic USA-CERL GRASS GIS 1987 promotional video (Inman 1987) during the development phase of the AV-portal for testing purposes (AUTHOR & AUTHOR, 2014).

The acquisition rates for OSGeo conferences indicate that each event will yield at least 30 hours of scientific footage: The FOSSGIS 2015 conference produced 32 hours of footage, while the FOSS4G 2014 event generated 39 hours. Currently, three OSGeo conferences on national, regional and global level are being actively collected by TIB (FOSSGIS, FOSS4G-NA and FOSS4G), which results in an annual amount of over 100 hours of scientific footage.

From 2015 to 2016, the acquisition of OSGeo conference video content required a harvesting step by TIB personnel, to access videos from third-party web portals, where they had been published in the first place. As a result of an active dialogue with OSGeo Foundation, this procedure could be changed to a proactive model. Starting with the FOSS4G 2016 conference, TIB will be immediately provided with the video recordings by the conference organizers, including all available metadata. This allows TIB to start the content analysis process immediately, to publish the content in the TIB|AV Portal as early as possible, honoring possible embargo requests by the respective conferences.

## 4 Scientific Potential

From the perspective of science and the OSGeo community, the TIB|AV Portal has established a reliable and non-commercial infrastructure to use and reference audiovisual content, ranging from full films to subsections or still-frames. Currently, a range of usage scenarios are emerging. Two of them will be described here.

### 4.1 Reference of audiovisual content in Web 2.0 portals

Each film which is published in the TIB|AV Portal is assigned a DOI as a unique persistent identifier. Web browsers can resolve DOI, being a Universal Resource Locator, similar to HTTP. From the perspective of long term availability, DOI are superior to HTTP-links as they can by design never expire. A DOI consists of a unique character string that identifies an entity in a digital environment: It identifies the object itself, not the place where it is located. If the object is moved to another location (i.e. the URL has changed) the only requirement is to update the URL in the underlying database. This ensures that the DOI persistently resolves to the location of the object (Paskin 2002). For information sharing within a scientific technical community such as

OSGeo, this is a significant benefit, as DOI-links can both be embedded in markup code such as HTML and Wiki-content and do not need to be reassessed for obsolescence due to broken or expired links. Figure 1 shows the DOI-based demonstrator for thematic excerpts from the AV-Portal which was set up within the OSGeo Wiki in 2015.

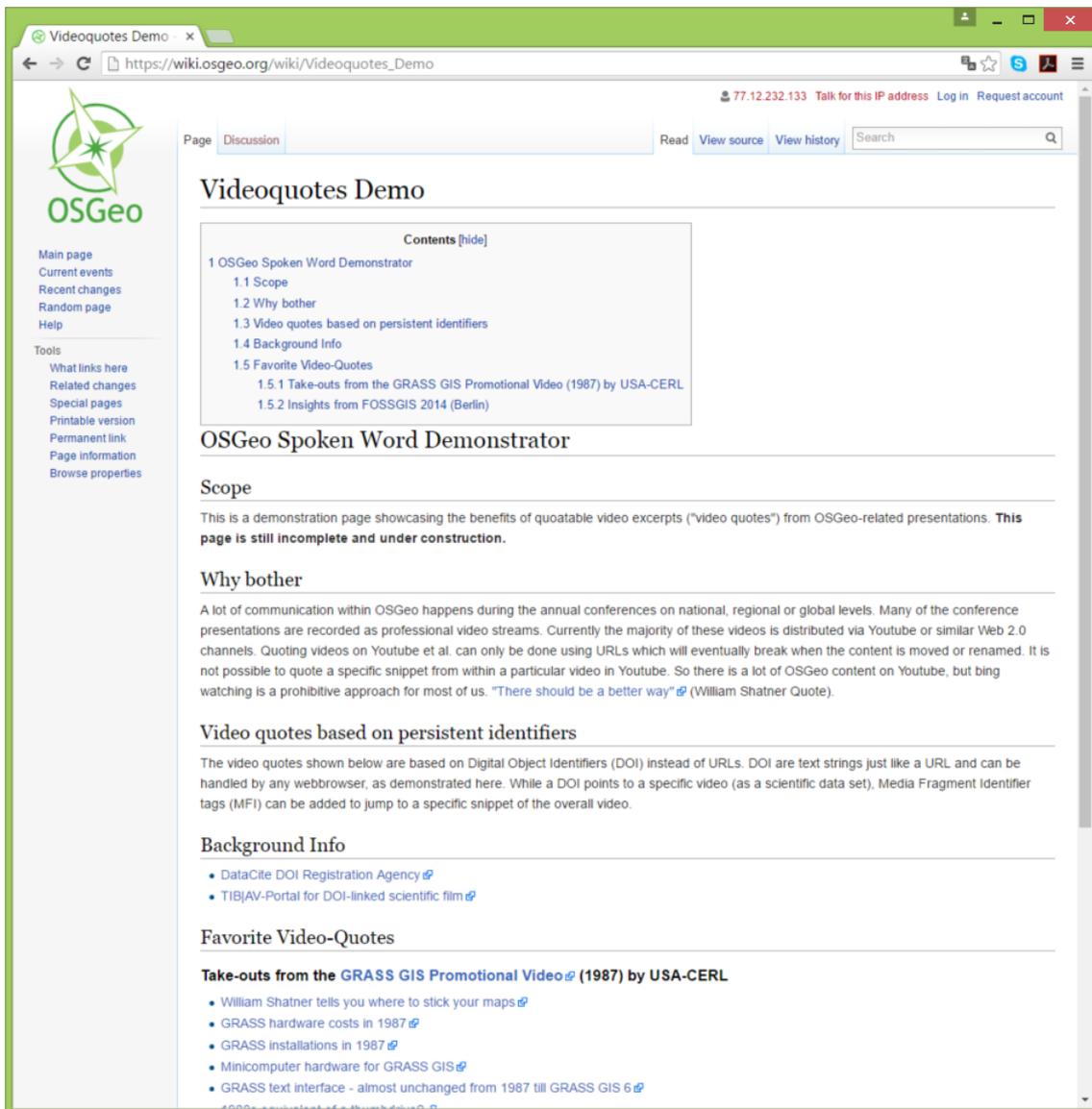


Figure 1: OSGeo videoquote demonstrator based on DOI-links into the TIB|AV Portal (OSGeo 2016).

## 4.2 Acknowledging OSGeo audiovisual content as Research Data

The minting of DOI for videos in the TIB|AV Portal is handled through DataCite, an international not-for-profit organization (Datacite 2016). DataCite fosters the citation of research data and to establish easier access to research data on the Internet, increase acceptance of research data as legitimate, citable contributions to the scholarly record and support data archiving that will permit results to be verified and re-purposed for future study. This includes scientific technical information, such as the OSGeo conference videos as part of the scientific technical collection of TIB.

DataCite operates the Metadata Store (MDS), a Metadata Search engine (Datacite 2016), which allows queries on all DOI-referenced research data. This enables high-level queries for scientific journal publications, data publications, scientific software and audiovisual content. This includes contributions of the OSGeo community in all these fields. A sample search for OSGeo-related content is demonstrated in Figures 2, 3 and 4: A query about Mrs. Athina Trakas, a long term OSGeo advocate, results in multiple DOI references for audiovisual

content (Figure 2). Each such DOI-link refers to a dedicated landing page with metadata about the audiovisual data set, as shown for doi.org/10.5446/14749 (Figure 3). The landing page in turn provides a link to the respective video in the TIB|AV Portal (Figure 4). By using the open standard media Fragment identifier (MDID), individual segments of a video can also be cited, just as easy as a chapter or a page in a book. In order to cite a video or a video segment its DOI link is simply copied and pasted into a document. This demonstrates how audiovisual content is already being recognized, and has become citable, as scientific research data. The accessibility of the associated metadata thereby determines the visibility of both the described content and the institution that provides the data. To further increase their discoverability and reuse, video references can be cited and linked to context information of authors and institutions. Such databases include e.g. the ORCID system for author publications (ORCID Organization 2016), or Current research information systems (CRIS) such as VIVO (VIVO 2016).

By connecting video assets to dynamic digital content such as text publications and datasets the digital infrastructure is optimized and provides additional value to a scientist's reputation: Digital research outputs can be traced across the life-cycle from acquisition and preservation to access and reuse.

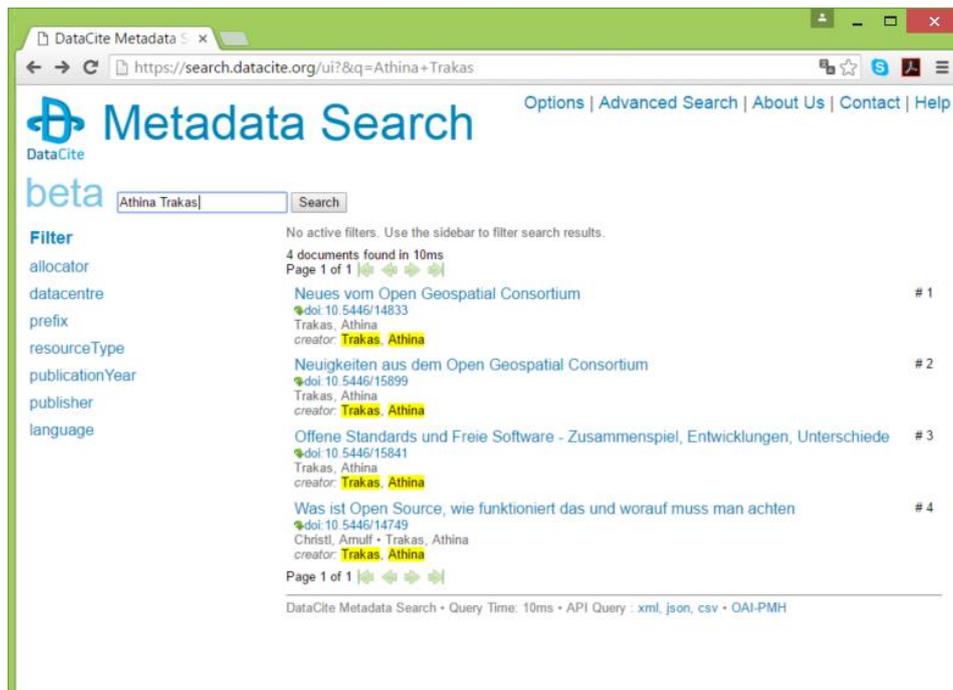


Figure 2: Query result from a DataCite search in the Metadata Store showing e.g. OSGeo-related content (Search “Athina Trakas”).

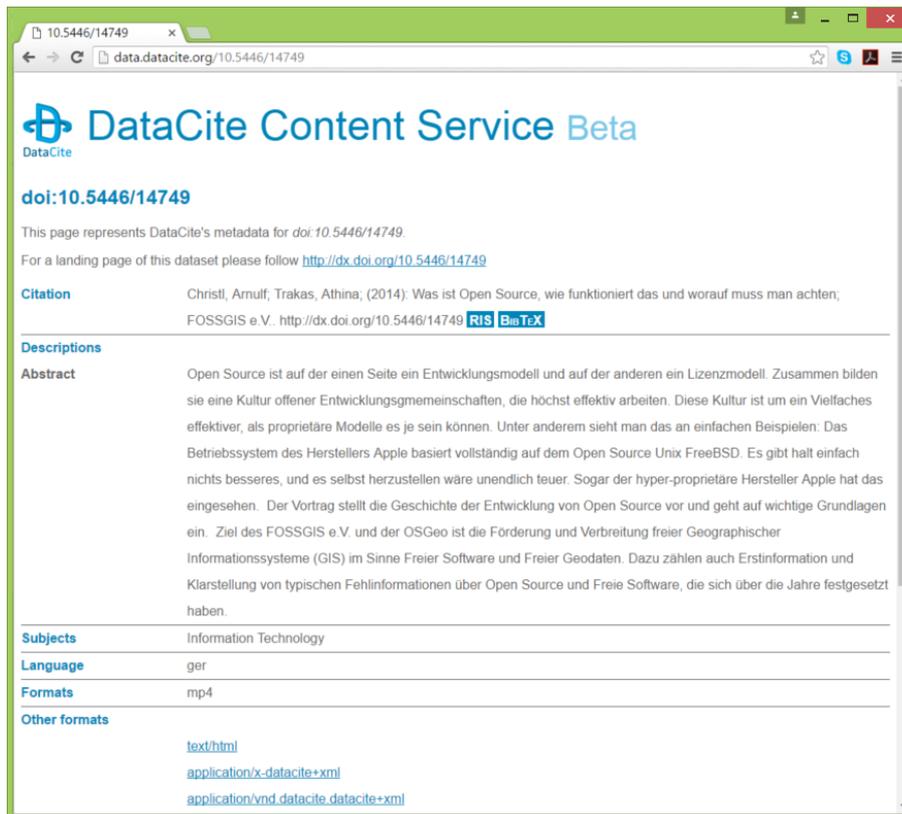


Figure 3: DataCite Content Service provides search results for OSGeo-related videos in the TIB|AV Portal. This is a function of the assigning of a persistent identifier, a Digital Object Identifier (DOI) to each video in the AV-Portal.

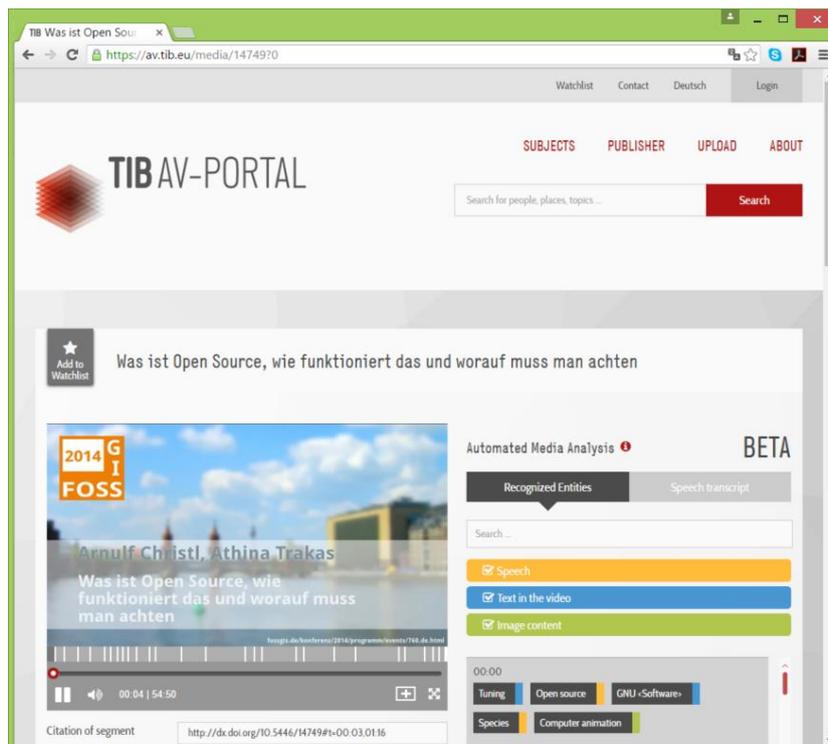


Figure 4: The landing page in the TIB|AV Portal for the conference video accessed via DOI: [doi.org/10.5446/14749](http://dx.doi.org/10.5446/14749).

## 5 A research data perspective on audiovisual OSGeo content

The statistics for the general growth of minted DOI for audiovisual content (Datacite 2016) indicate a long term growing trend of this kind of research data (Figure 5). This corresponds to the already encountered and projected growth rates for the TIB|AV Portal. In addition to this, for OSGeo, the amount of DOI-referenced content is expected to rise significantly in the near future, especially in the fields of geospatial data sets and scientific software.

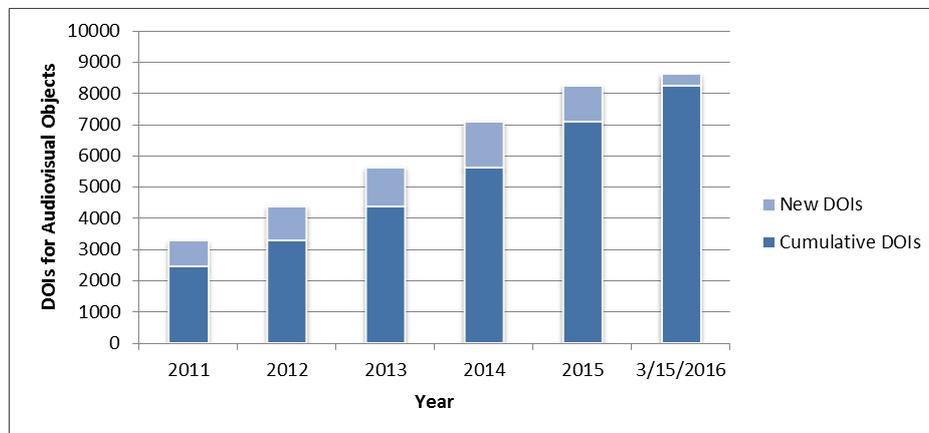


Figure 5: Yearly growth of minted DataCite DOI for audiovisual content (shown in dark blue column) and share of newly minted AV-DOI per year (light blue column).

The wealth of audiovisual content already generated by OSGeo exceeds both thematically (i.e. topics beyond TIB's acquisition foci, e.g.: Agriculture, Medicine) and qualitatively (i.e. multilingual content) the range of acquisition and customer services offered by TIB. A comprehensive acquisition strategy addressing the multilingual and multi-topical diversity of OSGeo's scientifically relevant output remains a challenge to be taken on by multiple research libraries. This will lead to added value for the OSGeo communities, libraries and their users, and also the general public. The experiences gained from the TIB|AV Portal can serve as a starting point towards a modular best-practice based approach for this.

## 6 Conclusion

Nowadays, the provision of digital audiovisual content is an important communication tool for STI within and beyond OSGeo. With the diversifying and accelerating production of audiovisual content diversifying and accelerating, best practices are needed to address and solve this challenge on a global scale. This paper summarizes the current status and usage scenarios from the collecting effort undertaken by TIB to ensure the long term availability of OSGeo conference video recordings.

The audiovisual OSGeo-related content is provided through the TIB|AV Portal – a web-based platform for scientific audiovisual media featuring state-of-the art multimedia analysis and retrieval. Quality checked audiovisual content from the OSGeo communities is constantly being acquired for the portal as a part of TIB's mission to preserve relevant content in applied computer sciences for science, industry and the general public. Currently, the collection sphere includes video content from national (FOSSGIS), regional (FOSS4G-NA) and global (FOSS4G) conferences. The annual growth exceeds 100 hours of new content relating to the OSGeo software projects and the OSGeo scientific-technical communities.

As a research library, TIB provides reliable long term preservation for AV media, as well as metadata enhancement analysis with extended search and retrieval. Using DOI and a Media Fragment Identifier, a scientific citation of full videos, excerpts and still frames is provided. Scientists can therefore use video content e.g. in education and referral in social networks. Depositing AV media as research data in the TIB|AV Portal ensures that the requirements of Good Scientific Practice are met: Conference contributions can be listed in the researcher's citation record and other researchers will be able to find, re-use and cite published media. Thus, the library-operated service infrastructure TIB|AV Portal turns the audiovisual OSGeo-related content in a reliable source for science and education, making it a vital part of the development and heritage of the OSGeo software communities and scientific software development.

## Reference

- Löwe, P., Neumann, J., Plank, M., Ziedorn, F., Lazar, R., Westervelt, J., Inman R. (2014), GRASS GIS, Star Trek and old Video Tape – a reference case on audiovisual preservation for the OSGeo communities, FOSS4G 2014, doi: 10.13140/2.1.2983.6801
- Löwe, P., Neteler, M. (2014), Data Science: History repeated? – The heritage of the Free and Open Source GIS community, Geophysical Research Abstracts, Vol. 16, EGU2014-10027
- Berners-Lee, T. (2006), Linked Data-Design Issues. Available in : <http://www.w3.org/DesignIssues/LinkedData.html>
- DataCite Organisation (2016) DataCite: FIND, ACCESS, AND REUSE DATA <https://www.datacite.org> (Last checked 21.03.2016)
- DataCite (2016) Metadata Search. <http://search.datacite.org/ui> (Last checked 21.03.2016)
- Inman, R. (1987) GRASS - Geographic Resource Analysis Support System, Moving Pictures, US Army Construction Engineering Research Laboratory. doi: 10.5446/12963
- International DOI Foundation (DOI), (2012) The DOI system concept. In: DOI Handbook. doi: 10.1000/182
- Lichtenstein, A., AUTHOR., & Neumann, J. (2014), TIB's Portal for Audiovisual Media: Combining Manual and Automatic Indexing. Cataloging & Classification Quarterly. 52(5), 562-577. doi: 10.1080/01639374.2014.917135.
- Paskin, N., (2002) Digital Object Identifiers. Proceedings of ICSTI Seminar: Digital Preservation of the Record of Science. ICSTI 14, IOS Press, Oxford, pp 1-16
- Open Source Free Software GIS - GRASS users conference 2002, Trento <http://www.jus.unitn.it/services/arc/2002/0911/> (Last checked 21.03.2016)
- ORCID Organization (2016) ORCID Connecting Research and Researchers <http://orcid.org> (Last checked 21.03.2016)