

CDS Videos - The new platform for CERN videos

Tuesday, 10 July 2018 16:40 (20 minutes)

CERN Document Server (CDS, cds.cern.ch) is the CERN Institutional Repository based on the Invenio open source digital repository framework. It is a heterogeneous repository, containing more than 2 million records, including research publications, audiovisual material, images, and the CERN archives. Its mission is to store and preserve all the content produced at CERN as well as to make it easily available to any outlet interested.

CDS aims to be the CERN's document hub. To achieve this we are transforming CDS into an aggregator over specialized repositories, each having its own software stack, with features enabled based on the repository's content. The aim is to enable each content producer community to have its own identity, both visually and functionally, as well as increased control on the data model and the submission, curation, management, and dissemination of the data. This separation is made possible by using the Invenio 3 framework.

The first specialized repository created is CDS Videos (videos.cern.ch). It has been launched in December 2017, and is the first step in the long-term project to migrate the entire CDS to the Invenio 3 framework.

CDS Videos provides an integrated submission, long-term archival and dissemination of CERN video material. It offers a complete solution for the CERN video team, as well as for any department or user at CERN, to upload video productions. The CDS Videos system will ingest the video material, interact with the transcoding server for generating web and broadcaster subformats, mint DOI persistent identifiers, generate embeddable code to be reused by any other website, and store the master files for long-term archival.

The talk will detail the software architecture of the CDS Videos as well as the infrastructure needed to run such a large-scale web application. It will present the technical solutions adopted, including the Python-based software stack (using among others Flask, IIIF, Elasticsearch, Celery, RabbitMQ) and the new AngularJS-based user interface which was exclusively designed for CDS Videos. It will also present our solution to a lossless migration of data: more than 5'000 videos from 1954 to 2017, summing up to 30TB of files, have been migrated from DFS to EOS in order to populate the CDS Videos platform. All this could be of high interest to other institutes wanting to reuse the CDS Videos open source code for creating their own video platform. Last but not least, the talk will detail how the user community at CERN and beyond can take advantage of the CDS Videos platform for creating and disseminating video content.

Primary authors: Mr COSTA, Flavio (CERN); Mr GABANCHO, Esteban (CERN); Mr GONZALEZ LOPEZ, Jose Benito (CERN); Mrs MARIAN, Ludmila (CERN); Mr TAROCCO, Nicola (CERN); Mr WITOWSKI, Sebastian (CERN)

Presenters: Mr COSTA, Flavio (CERN); Mr GABANCHO, Esteban (CERN); Mr GONZALEZ LOPEZ, Jose Benito (CERN); Mrs MARIAN, Ludmila (CERN); Mr TAROCCO, Nicola (CERN); Mr WITOWSKI, Sebastian (CERN)

Session Classification: Posters

Track Classification: Track 8 –Networks and facilities