TUTORIAL - Putting Historical Data in Context: how to use DSpace-GLAM

Andrea Bollini and Claudio Cortese

4Science S.r.l., Milano, Italy
{andrea.bollini, claudio.cortese}@4science.it

1 Instructors

1.1 Andrea Bollini

Andrea is Chief Technology and Innovation Officer (CTO / CTIO) at 4Science, with the responsibility of ensuring the use of the most efficient technologies to effectively achieve the results of each single project. Andrea is actively involved in various international open source and open standards communities, often with leading roles: DSpace committer, Deputy Leader of the euroCRIS' CERIF Task Group "CERIF" and member of the COAR Working Group "Next Generation Repositories". Chair, speaker and reviewer in important conferences, before joining 4Science Andrea worked for two Italian University Consortia, CILEA and CINECA, where he was responsible for the development, design and management of IT solutions and projects in the field of research, electronic publishing and open access repositories.

1.2 Claudio Cortese

Claudio is Project Manager and Business Analyst at 4Science. He has a PhD in archaeology and is lecturer of "Computer Applications in Archaeology" at the Catholic University of Milan. For several years he has been dealing with modeling, management and analysis of data for cultural heritage using a great variety of methods, standards and technologies: from relational databases to GIS, from Digital Libraries to architectures based on the Semantic Web. Before joining 4Science, he worked for two Italian University Consortia, CILEA and CINECA, mainly dealing with Digital Library Management Systems to preserve, use and distribute digital cultural resources. In the historical/archaeological field Claudio focused on the design and creation of databases for many Archaeological Missions and University Research Units. He constantly gives consultancies, lessons and courses to universities and public/private institutions.

2 Aims, scope and learnig objectives

The proposed tutorial will introduce attendees to DSpace-GLAM (Galleries, Libraries, Archives, Museums), the Digital Library Management System based on

DSpace (http://www.dspace.org) and DSpace-CRIS (https://wiki.duraspace.org/display/DSPACECRIS/DSpace-CRIS+Home), developed by 4Science for the management, analysis and preservation of digital cultural heritage, covering its functional and technical aspects.

DSpace-GLAM is an additional open-source configuration for the DSpace platform. It extends the DSpace data model providing the ability to manage, collect and expose data about every entity important for the cultural heritage domain, such as persons, organizations, projects, events, places, concepts and so on.

The extensible data model will be explained in depth, through examples and discussions with participants.

Other main topics will be DSpace-GLAM "components", relationships management and network analysis.

Finally 4Science new add-ons for digital cultural resources fruition and analysis (the IIIF - International Image Interoperability Framework - Image Viewer, the Audio/Video Streaming Module, the OCR Module and the CKAN integration) will be illustrated.

At the end of the tutorial the participants will be able to understand the DSpace-GLAM data model, to adapt it to their needs and to evaluate if DSpace GLAM fits the needs of their institution.

3 Relevance to TPDL 2017 and significance for the research field

In the last years humanities are witnessing a growth of available data, thanks to the increasing use of databases, electronic journals, digitization of cultural heritage and tools for data extraction and analysis. In this context scholars and cultural heritage professionals have to be able to correlate different data sources, to better investigate the articulation of historical phenomena and of the transformation processes that affected human history and culture. In the analysis of the digital data, indeed, it is essential that they are not considered in isolation but in conjunction with all the contextual information needed to answer the research questions. Historical data are often fragmentary, partial and biased, so frequently we can understand a content only in a contextual framework, analyzing its relationships within a global and multidimensional approach.

Since almost two decades Digital Libraries have been managing a variety of objects, like texts, audios, videos, blogs and so on, and they are still evolving to scale into Virtual Research Environments (VREs), integrating the entire extent of an institution's scholarship, including articles, theses, dissertations, journals and also research datasets.

In the Big Data Age, Digital Library Management Systems have to forcefully enter the research process to manage both qualitative and quantitative aspects of digital cultural heritage and allow researchers also to analyze data, highlighting and enhancing their relationships at different scales and to explain their interpretations about the important dimensions of variation and about the network of

contextual links that affect the historical sources. Therefore the flexibility of the data model and the availability of tools for analysis and interpretation become fundamental features for such systems.

DSpace-GLAM has been developed to face these challenges.

Using DSpace-GLAM, istitutions will be able to manage, analyze and preserve digital objects. together with historical, archaeological or other cultural datasets, relating them with other entities (persons, events, places, concepts, contexts, etc.) for describing the context of cultural objects and data, according to different granularity levels.

4 Full description

After an introduction focused on the theoretical and methodological foundations of the project (see section 3), the tutorial will be articulated in 5 main sections as follows:

1. Managing digital objects within DSpace:
   – the DSpace conceptual model: communities, collections, items, bundles, bitstreams
   – configuring metadata
   – the submission process
   –– manual submission and batch submission
   – curating items with additional add-ons
   –– upload images to the IIIF Image Server
   –– extract images from PDF
   –– create access images from RAW image formats
   –– extract text (hOCR) from images and send it to the annotation server
2. Extending the DSpace data model with DSpace-GLAM:
   – DSpace-GLAM main entities (Persons, Organizations, Projects, Events, Concepts, Fonds, Places)
   – defining, configuring and managing new entities (Dynamic Objects)
   – network visualization configuration
3. Integrating DSpace-GLAM and DSpace:
   – connecting items to new entities and creating inverse properties
4. Digital cultural resources fruition and sharing
   – search and browsing within communities and collections
   – using the 4Science add-ons to visualize and share images and videos
5. Dataset analysis
   – linking datasets to items and to every defined GLAM entity
   – using the open source 4Science integration with CKAN to analyze your datasets
   –– visualize data within graphs
   –– visualize data within maps

So, at the end of the tutorial, every aspect of the process needed to setup a Digital Library with DSpace-GLAM will have been addressed, using both open source components and dedicated add-ons (`http://www.4science.it/en/dspace-add-ons/`), ingesting different kinds of digital objects, according to the different metadata standards (e.g. Dublin Core, MODS, EAD, METS, etc.) required by everyone's own project.

5 Tutorial history

It is the first time we offer this tutorial and that we propose to present DSpace-GLAM

6 Format of the tutorial

3 hrs.

7 Target audience and expected prerequisite knowledge of the audience

The level of this tutorial is introductory. It is addressed to librarians, archivists, historians, archaeologists, researchers and to those who want to build their own digital library but do not want to write their own software nor buy a proprietary solution. No programming ability is required. Basic knowledge of digital libraries and repositories architectures and of the relational model, though not mandatory, can guarantee a better learning experience.

8 Citations to publications covered by the tutorial

- **Cortese C.**, 2017, in *AIUCD 2017 Conference*, (Big) Data Science and historical archaeological studies: a methodological, technological and cultural challenge, `https://www.slideshare.net/4Science/presentations`
- **Bollini A, Cortese C., Groppo E., Mornati S.**, 2017, in *IRCDL 2017 Conference*, Extending DSpace to fulfil the requirements of Digital Libraries for Cultural Heritage Management, `http://ircdl2017.unimore.it/papers/IRCDL_2017_paper_7.pdf`
- **Palmer D. T., Bollini A., Mornati S., Mennielli M.**, 2014, in *Procedia Computer Science - 12th International Conference on Current Research Information Systems, CRIS 2014 - Managing data intensive science - The role of Research Information Systems in releasing the digital agenda.*, 118-123, Dspace-CRIS@hku: Achieving visibility with a CERIF compliant open source system.

9 Resources required

Personal laptop, Video Projector, Internet Connection