HAL
open science

An interdisciplinary approach to Cultural Heritage Data Management - Building scientificity and dealing with heterogenous LOD

Loic Jeanson

To cite this version:

Loic Jeanson. An interdisciplinary approach to Cultural Heritage Data Management - Building scientificity and dealing with heterogenous LOD. Linked Past IV: Views from inside the LOD-cloud, Dec 2018, Mainz, Germany. hal-02490609

HAL Id: hal-02490609

https://hal.science/hal-02490609v1

Submitted on 25 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An interdisciplinary approach to Cultural Heritage Data Management

Building scientificity and dealing with heterogenous LOD

Loïc Jeanson
Université de Nantes, Centre François Viète
Centrale Nantes, Laboratoire des Sciences du Numérique de Nantes
loic.jeanson@ls2n.fr — +33 (0)2 40 37 69 56

Abstract

The ReSeed project develops tools and methods to support the works of heritage professionals and ultimately the heritage decision process. Gathering around a common table institutions, the project connects Engineers (IT and KM) and Social Scientists (historians, curators, heritage experts) in order to produce a heritage knowledge management and decision supporting tool. This poster will present the LOD aspects of the works of the ReSeed project since its beginning in November 2016 in order to structure data, information and knowledge about one of its case study : the astronomical observatory of the Pic du Midi. This 150 years old scientific station located at 1877m on the top of the Pyrenees in the south-west of France hosted and still hosts biological, meteorological, and of course astronomical experiments. Such a place where history of science and techniques meets a local social history within a national and international history of scientists and institutions offers a great opportunity to deal with complexity. This poster will focus on two aspects: 1) on a tension we had to resolve on the integrity (of the heritage, of the source documents, and maybe of our model) and 2) on our solutions to deal with complexity and heterogeneity while trying to integrate data and information from geometrical sources as much as more textual sources. The project's work is driven by its case studies and although we hope to deliver a more general solution, some aspects remain largely improvable and may seem at the time tailored to our needs.

An interdisciplinary project

The team aiming at developping tools for Heritage study and expertise had to be interdisciplinary in order to adress the multiple (computationals, epistemologic, technologicals, practical...) challenges. In details, the project gathers Mechanical Engineers, Knowledge Management Experts, IT Experts (from Centrale Nantes, Technical Uni. of Compiègne, Technical Uni. of Troyes and a company : DeltaCAD), alongside with Historians, Museum Curators and Cultural Heritage Experts (from Uni. of Nantes, the *Musée des Arts et Métiers* (Arts and Crafts Museum), the french Ministry of Culture and a company : MCC Heritage).

Three case studies

We chose to develop our tools with specific applications in mind. In order to adress various spectrums, the following use cases were selected :

- The Observatory of the Pic du Midi, in the Pyrenées : a 150yo. site with several activities (e.g. astronomy, meteorology, aerology, botanics).
- A collection of meridian circles, built within a range of 13 years, as similar as possible, and spread out within France.
- A physical mock-up of a metallic house, built around 1890.

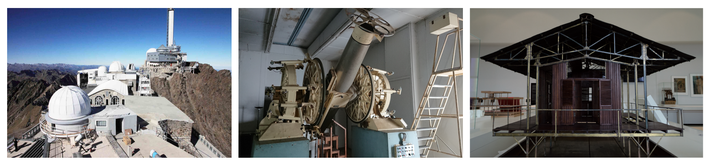

Figure 1: Pic du Midi Observatory. 2011. Photo P. Bastide — A Meridian Circle in Paris Observatory. 2018. Photo L. Jeanson — Mock-up of the Iron House for the Colonies by Moreau frères. Ca. 1890. Musée des arts et métiers. inv. 12002. Photo O. Delarozière

Introduction

Cultural Heritage Study uses, historical methods, tools, results and concepts. As for historians, the spectrum of data encompassed is extremely diverse. The digital/computational revolution leading to new tools and approaches called digital humanities had so far a limited impact on the Cultural Heritage Study and Expertise in France and, to some extent at the international level (at our knowledge, ICOMOS panels usages seem to remain mostly unaffected). Furthermore, the complexity of the sites and objects, especially the Scientific and Technical Heritage creates the need for tools to efficiently enlighten cultural digital object's qualities. The ReSeed Project is part of this dynamic. The acronym ReSeed can be translated and expanded as *Semantic Reverse-Engineering for Digial Cultural Heritage*.

A long list of motivations

The sources involved in Cultural Heritage Study could vary from an audio recording of an interview, to a note in the margins of an document in archives, or a 3D scan, or a diagram. And the fields experts could be an historian, a logistician in a museum or an architect or a restorer, a chemist specialist in ancient non-ferrous alloys,... This very large heterogeneity of data and actors, of points of interests and of domain related vocabularies are both a huge challenge and the very source of the richness and nuances required by Heritage Experts to work around Cultural Heritage Objects. Semantic Web technologies enable the use of standardized linked vocabularies and data without constraining it. We decided to use this kind of technologies in order to structure our data to be able to face the large heterogeneity. This use of technologies led us to newer problems for us : How will we be able to integrate this heterogeneous data ? How can we keep the standard practices of the field with this new way of data management ?

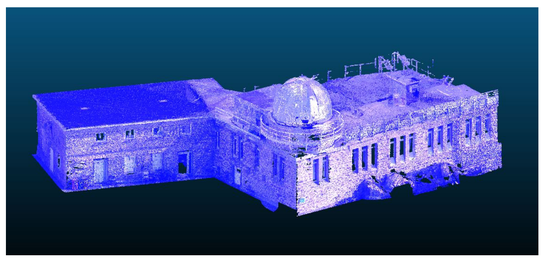

Figure 2: Point cloud of a building at the *Observatoire du Pic du Midi*, from 3D scanning using FARO Focus3D

Making the model readable

What is *(3D) modelling* ?

What is called *3D modelling* can cover various realities. Mostly, it means two types of approaches and their combinations : 1) machine processing when producing a 3D model from measurements (by lasergrammetry, photogrammetry, tomography, etc.) or 2) human-processing when the producing from a 3D modelers or CAD software where the user direcly defines the needed geometries. The main drawback of the machine-made model is the level of details, that is only settings driven in softwares. It often results in a modelling too *rough* or too *heavy*, depending on human interest, with low added meaining. But in the mean time, it is repeatable and we can identify the source data and and the involved processes. Human-made models have a converse profile : the production process is rarely transparent, but the level of detail fits the human points of interest and the different components can be identified, named after some taxonomy. Hybrid processes for the production of models make it almost possible to combine the benefits of both worlds, except for sourcing the modelling process in details. The scientificity of a study depends on its ability to be criticism-resilient. Any text-based study render gives access to its sources and methods, an image-based render must be as transparent to have an equivalent scientific value.

A need to track

Our proposal in order to deal with the transparency upon the source data and the processing is multi-component

- In our triplestore, we use (misuse ?) named graphs. Every triple is contained inside of a named graph : a triple now has its own URI.
- Every triple comes with metadata. Our metadata bulimia is one of our problems : it adds noise on data, making the core data harder to grasp. This problem will not be covered here. The metadata is of several types, among which :
  - a status : `fact` or `hypothesis` : every triple coming from a source is a `fact`, and every information produced by a human or machine reasonning an `hypothesis`.
  - a sourcing process : every `fact` comes from a source and an extraction method, and every `hypothesis` comes from data and rules (heuristic, algorithmic, statistic, or less precisely formalised : practical rules).

Data Integration

Semantic Web technologies is well suited to data interoperability since it can structure virtually every datatype in an analoge manner.

What is to understand behind *Data heterogeneity*?

Heterogeneity in the data has several field dependent meanings. We here mean only the heterogeneity in nature of data. The physical description of an object can be provided by the use of a photography, a x-ray or an Infra-Red image, a 3D model or a text. Even a impulse response sound delivers informations about the physical feature of a room. All these types of documents store different natures of data. This heterogeneity in nature then produces other forms of heterogeneity : *same information?*, *same value?*, *same precision?*, *same meaning?*. Figure 3 shows an example of a diverging but close heterogeneity in nature. Distant from 40 years, they display the same type of object with different levels of completeness, accuracy and meaning.

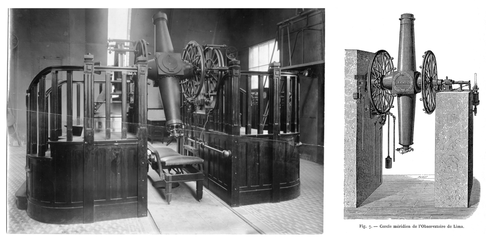

Figure 3: Meridian circle of São Januário, by Gautier. Ca. 1928. Archives of Museu de Astronomia e Ciências Afins, Rio de Janeiro, BR. — The Meridian circle of Lima (PE), by Eichens. ANDRÉ, C., ANGOT, A. in *L'astronomie pratique : les observatoires en Europe et en Amérique, Part IV, Observatoires de l'Amérique du Sud*. Paris : Gauthier-Villars, 1881

Integration quésako ?

Data integration also has field dependent meanings. We mean it as *the ability to link data through sources and to combine them*. In this regard, as soon as data is structured by Linked Open Vocabularies (LOV), we consider it integrated. The problem is the availability of vocabularies for any type of source. We focus here on *imagery* data (any 2D or 3D representation). The*3D ontology* is by far more complete, mapping the X3D file format. As for now we use it *as is*. For 2D pictures, we could find no equivalent vocabulary. The data structure is fortunately much less rich in 3D pictures, so we developed our own vocabulary.

Conclusions

This poster aimed at presenting our approach in terms of integration and scientificity. This approach, as any, has is share of shortcomings and among them, the biggest is the amount of created data and the cohabitation of nature-heterogenous data. This presentation did not cover any of those aspects. Also we did not -yet- opened our vocabularies. We are a mixed team coming from various backgrounds and we lack implication in LOD and LOV communities. We needed a solution to deal with our data in the short term, but are planning to try to be more involved in the ontology workshops in order to confront our approach and vocabularies.

Forthcoming Research

We raised another big problem breifly in this presentation when talking about the diversity of actors and of vocabularies. We are currently trying to formalize our approach and to tool up for implementation on this topic. The main goal is to regroup terms in a way, through clustering or by the use of some metrics, in order to modify the querying process to get broader results. Our research team is also member of the *consortium 3D*, national working group around methods and technologies for 3D representation. A discussion has to take place in order to improve and enrich our approach and solution. Another forthcoming work results from the effectiveness of the data integration. The developpement of a way to run data and pattern analysis in a graph composed of such heterogeneous datas and vocabularies is ongoing.

References

[1] 3d ontology website. `http://3dontology.org/`.

[2] Reseed project website. `http://www.reseed.ls2n.fr/`.

[3] M. Cotte, F. Laroche, L. Jeanson, and N. Bourgeois. Analysis of a site's integrity by 3D models and integrated database, Case Study: the Pic-du-Midi high-mountain observatory. In *ICOMOS*, New Delhi, India, December 2017. ICOMOS.

[4] L. Jeanson, M. Cotte, and F. Laroche. Indicateur(s) de qualité(s) de modèles numérique pour l'étude de l'intégrité physique d'un bien patrimonial. In *Jounrée d'études 'Le numérique : outil d'étude de la valeur patrimoniale*, May 2018.