How Was This Done? An attempt at Formalising and Memorising a Digital Asset's Making-Of

Iwona Dudek, Jean-Yves Blaise, Livio De Luca, Laurent Bergerot, Noémie Renaudin
UMR CNRS/MCC 3495 MAP
Marseilles, France
iwona.dudek, jean-yves.blaise, livio.deluca, laurent.bergerot, noemie.renaudin@map.cnrs.fr

***Abstract*— This paper introduces the early results of a research programme called MEMORIA that aims at developing an information system enabling the description, structuring and storage of digital outputs produced in the course of Heritage Architecture studies. Our objective is to memorize not only a given result – *i.e.* a digital asset - but its *making-of* - in other words to record and share with future generations a *work process* rather than solely its outcomes. Digital assets are on the one hand described by a set of "classic" parameters (e.g. format, authors, creation date, *etc*.) and on the other hand associated with a process (concept that should be understood as a chain of activities). Ultimately, the project investigates how a digital resource resulting from a human-birthed cognitive process can be associated with descriptors ensuring that all actions mobilised to produce the resource are recorded, and therefore ensuring a sort-of scientific traceability of the "final" digital document.**

***Index Terms*—Heritage architecture, Methodology, Work processes, Documentation.**

I. Motivation

Massive digitisation campaigns carried out over the past decades in the Digital Humanities community, and notably in the cultural and heritage sectors, do result in a positive move towards more content available for analysts, more cross-analyses made possible, more knowledge collected, structured and shared [1] and generally speaking more focus potentially put on these sectors thanks to end-line products or valuation initiatives. In parallel, the notion of open science [2] has paved its way into the work practices of scholars and collection holders, with today a real concern in the above mentioned sectors for interoperability, open access, standards, metadata (*etc*.). As a result, a number of initiatives such as [3] or [4] do offer an unprecedented opportunity for web users to access without charge to digital content. In addition, the notion of ***digital asset*** – *i.e.* a digital content *along with* the right to use it – has been pinpointed as a mean to create value by [5], thereby adding economy to science in terms of perspectives.

However, with more and more content digitised, more and more content e-published, digital humanities actors are now clearly confronted to a challenge initially picked up in information sciences: *Big Data* (big volumes of data, dynamically changing data, as well as high variety, highly heterogeneous data). This is visible in many research programmes where tools and formalisms stemming from the computer science side are confronted to analysis needs stemming from the Humanities side – literary analysis is a classic example [6], [7], [8].

In the digital heritage, more data sets, more data acquisition procedures, more heterogeneity appears at first glance as a positive move – yet actors very well know that it also means more choices to make in terms of tools and methods, and more interpretation steps to memorise if wanting to ensure scientific validity. A quick overview of research programmes focusing on digital heritage assets shows that, beyond the digitisation effort itself, three challenges are at this stage on top of the research agenda:

- developing standards / data models in order to facilitate information sharing – a typical example is the CIDOC Conceptual Reference Model [9],
- sensemaking in massive data sets – a challenge picked up prominently by the fields of Visual Analytics and by the data mining community [10],
- allowing for an open access to the data, and ultimately considering it as a repurposable material ([11]).

But in some application fields - and in particular in heritage architecture – the digitisation effort **most often leads to the creation of a new, highly interpretative, digital asset** (typically a virtual reconstruction). This new ***output*** of the analysis process is considered by many as an ***extension*** and an ***enrichment*** of the initial collection. Hence a fourth, emerging challenge: describing not only the asset as such ("this is a virtual model of building B created by actors A1 and A2, showing B at time T5"), but memorising the various steps undertaken by the analyst(s) in his/their move from raw data (archival material, 3D point clouds, *etc.*) to interpretation and knowledge. This is obviously key if the asset is to be of any help for future generations, and if we consider a digital asset as potentially of scientific value – what is a conclusion worth if one cannot explain how one reached that conclusion?

This paper introduces an ongoing research programme called MEMORIA, through which we try to preserve not only the ***result*** of a cognitive process (e.g. "*a Collada online 3D model showing a virtual reconstruction of Emperor Augustus' Tropaeum Alpium*") but the ***making-of*** this result (e.g. *How was the 3D data acquired? What pieces of generic knowledge were used?, What hypotheses were laid down?, What tools were used in the 3D modelling steps?, ...*).

The programme aims at investigating to which extent we can today preserve a ***cognitive process*** as such, since if and

only if we manage to do that can digital assets produced daily become a readable, sharable, scientifically relevant *digital heritage*. The approach builds on the idea that the creation of a digital asset results from a series of ***activities*** (ranging from in situ 3D data acquisition to analogical reasoning basing on archival material) that need to be recorded so as to allow for intersubjective verification. Activities can then be chained to form ***processes*** (successive moves from a raw cloud of points to a consistent meshed and textured 3D model for instance) that lead to the creation of an ***output*** (a new digital asset).

This research is conducted as multidisciplinary teamwork in a research unit focusing on heritage architecture analyses from various points of view (computer vision, geometric modelling, information systems, knowledge modelling, time-oriented data visualisation, *etc.*). Accordingly the outputs produced by the team members strongly vary in terms of format, of aim and target, and ultimately of role in the global understanding of how an artefact births and changes over time. We therefore will not claim that the MEMORIA approach is fully representative of our field of concern, but that it is tested on digital content that is heterogeneous enough to provide for a sound test field. Section 2 briefly comments on the notions of activities and processes while section 3 presents our main choices, and introduces the experimental platform itself.

II. Related Works

The notions of activity and process at the heart of this research are nothing but new or original: readers who once had to produce PERT or Gantt charts do have a clear idea of what is meant here, others will find in [12] a number of convincing examples of how these notions can be applied to real-life situations (SOPO diagrams, PlanningLines, Decision charts, *etc.*). An activity is basically a triplet of values: *who* does *what*, and *when*? Processes act as containers for an ordered chain of individual activities leading from *input* α to *output* Ω.

In short, the notions of activity and process are very present in the knowledge management community (see for instance [13] or [14]), but their application to heritage sciences remains challenging [15]. The MEMORIA initiative can be seen as an opportunity to question the applicability to historical sciences of formalisms that often birthed in the context of business and industry applied research.

III. The Memoria Approach

The project builds on four main notions: *outputs, activities, processes* and *sources/inputs*.

(a) Outputs are digital material/resources produced in the course of a study, whatever the media (still image, moving image, sound, *etc.*). Outputs can be an end result (e.g. a 3D model ready for publication on the web) but they can also be an intermediate outcome, ready for future use in alternative contexts (a 3D model needing texturing before web publication, and in the meantime ready for 3D printing). Ultimately, the project aims at associating a digital resource with a set of descriptors in order to trace all actions mobilised to produce the output, and therefore memorize not only the "final" digital document but also the main steps of the cognitive process that lead to its production. These descriptors are summarised in a visual metaphor illustrated on Figure 1, a visualisation that combines a thumbnail view (or icon) of the output and pseudo film perforations (perfs) thanks to which users have access to each standard parameter associated with the output (creator, date of creation, *etc.*).

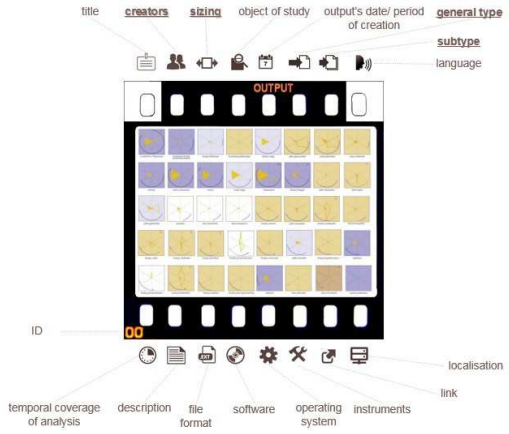

Fig. 1. The visual metaphor used as the user interface for *outputs*. White "perfs" correspond to blank fields (missing information).

A critical feature of the approach should therefore be a visual interface providing user-filtered accesses to the results of queries (outputs sorted by object of study, project, production process, creator, *etc.*). Progressive zooming is one of the solutions we are testing (Figure 2), but there is here clearly here a broader and still open challenge: bridging the gap between "classic" collection browsing modalities, with each individual item present in the display, and Visual Analytics solutions who introduce progressive filtering or collections reading mechanisms that will need to be tested.

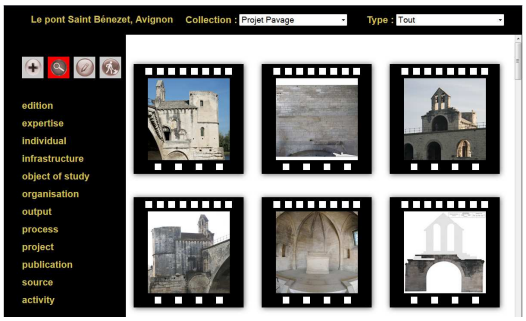

Fig. 2. Screenshot of the online test implementation.

(b) Activities are actions undertaken by one or several actors of the study in order to move from one point of the overall cognitive process to a new one. An activity is single out as such, and differentiated from a process, basing on the consistency of the method, tools and aim of a set of actions (e.g. *Translating three inventories in a row* is considered as one unique activity, but using these translations in order to *draw plans of buildings* is considered as another activity).

An activity is described by a set of standard parameters (creator, date and/or duration, inputs list, project id if relevant, external expertise called in, *etc.*). Activities are then grouped

by category (for instance steps dedicated to collecting data vs. steps dedicated to analysing data), with inside each category a specific hierarchy helping users to find their way in the reporting of "what they did" (e.g. activity desk-based research, sub-category handling of primary sources, Figure 3). Finally descriptors that are specific to one particular activity can be added (e.g. activity *laser scanning* > descriptor *measurement method*: optical triangulation, Figure 4).

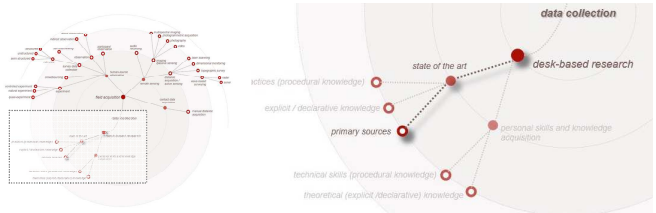

Fig. 3. The data collecting activities hierarchy (an activity identifies a series of actions mobilised to produce a digital resource).

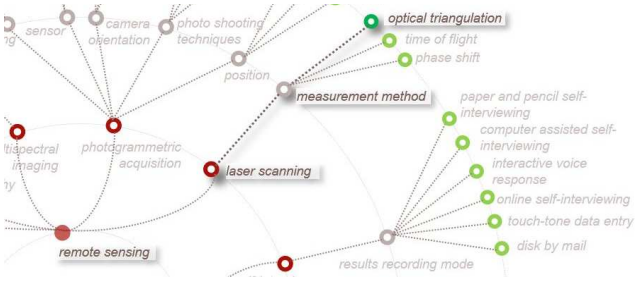

Fig. 4. Activity descriptors for the *data collecting* group of activities

(c) Processes act as containers for one or several activities. They are singled out basing on the fact that a new output has been created. In other words, each output is associated with a given process, containing one or more activities, and if this output is reused in the context of a new analysis task then the whole set of activities that lead to the creation of the output can be duplicated and memorised in the description of that new analysis task. A process keeps track of the institutional framework in which the work was carried out (organizations, projects, staff employed, *etc.*), as well as of primary sources used, or of the techniques and tools used. Each activity contributing to the process is represented by a multidimensional icon (Figure 5) that shows the activity's category (colour), and its position in the category's sub-hierarchy (glyph). Various processes can accordingly be compared visually, helping analysts and decision makers to get an overall and synthetic view of how outputs were produced (how many steps, what kind of steps, *etc.*).

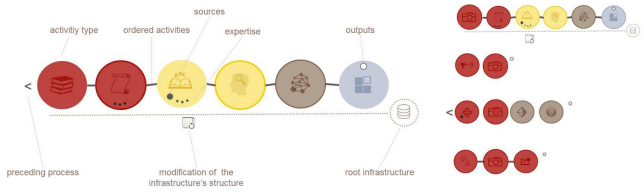

Fig. 5. The notion of process: a chain of activities mobilised to produce one or more outputs (including potential links to preceding processes).

A fruitful side-effect of the introduction of processes is the possibility this notion offers to spot methodological patterns (recurrent chains of activities) developed by a team, an individual, or in relation with a given type of data. As a consequence, with time, the notion of process can contribute to a better depiction of work processes in our field of concern, and ultimately to the creation of visual dashboards [15] helping actors to better explain, compare, quantify work processes.

(d) Sources and inputs

An activity can be based on one or more sources (external resources), as well as on outputs produced previously and described by a separate process. To avoid confusion the outputs used in an activity as initial root elements are called inputs.

Sources and inputs are pieces of data and information (digital or not) that are needed in order to start an activity or a process. They can be raw data (typically a 3D point cloud acquired in situ, or historical evidence retrieved from archives) or an intermediate outcome (e.g. a mesh built from the initial point cloud). Accordingly, they are described in a way that is quite similar to outputs, with a set of standard parameters (creator, date of creation, format, *etc.*) and will be searchable through an interface (Figure 6) that closely resembles to this that has been implemented for outputs (Figures 1- 2).

Fig. 6. A source is basically a piece of data or information, digital or not, needed to carry out one or more activities. A source is an "external" element on which an activity or process will base on. By contrast *inputs*, are intermediate results, reused in later MEMORIA activities.

Besides these four main elements, the MEMORIA platform allows for the storage of other pieces of information, about *infrastructures*, *actors*, *organisations*, *objects of study*, *expertise*, *context*, *etc.* as exemplified in Figure 7.

Fig. 7. Visual dashboards for supplementary pieces of information attached to outputs or activities.

The implementation we are currently developing (both for updating and browsing/searching the collection of outputs) combines an RDBMS and WebGL/Js modules built as pseudo-objects (in the sense of OOP programming). It is likely, depending on potential end-users feedbacks, to include also PDF outputs summarising in a one page "ready to print" document a visual overview of processes that are associated with a given digital output.

One of the acute issues we are facing in the MEMORIA programme is dealing with *existing* outputs: one thing is promoting a "new" preservation policy, another is coping with the digital assets that have *already* been produced. If we ignore, to a large extent, how they were produced, how can we integrate them into the MEMORIA approach? As an answer, the description of activities (as a hierarchy, with growing levels of details) and the corresponding visual interfaces are designed in order to allow for the recording (and searching) of "poorly" known work processes (for instance, recording the fact that a study bases on a 3D photogrammetric survey but not detailing the actual camera used, the date, *etc*.). Users are given a possibility to record processes as ordered chains (activities are known to have occurred in a precise order - Figure 8, left), or as disordered lists (*i.e.* a "bag of activities" (Figure 8, right).

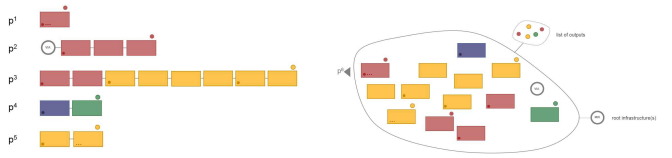

Fig. 8. An illustration of ordered and non-ordered processes.

Finally, it is important to state that the MEMORIA approach can also help showing the evolution of methods, techniques and tools used over time, as well as types of activities mobilised to produce a given type of output, thematic migration of teams or individuals, history of cooperation, *etc*.. In that sense, the MEMORIA approach, and in particular its visualisation steps, can contribute to a better understanding buy decision makers of today's work processes, of temporal constraints connected to each and every activity, of technical bottlenecks, and more generally of planning difficulties.

IV. Conclusion

This paper's claim is that in some application fields, and in particular in heritage architecture studies, we have come to a point where we need to preserve not only a digital asset as such (format, creator, owner, *etc*.) but also its *making-of* (analogies, tools, decimation or simplification, sources? *etc*.). With (at least) 20 years of digital assets production behind us most actors in the field are now facing hundreds when not thousand s of "*semantically orphan*" digital content.

We introduce an experimental platform called MEMORIA thanks to wish we investigate a methodological framework aimed at memorising the various steps (represented as activities and processes) undertaken by human analysts in order to move from an initial state of knowledge to the production of one or several new digital assets. Although this research is still at an early stage, it underlines the potential benefits of enabling us to answer to the "*How was this done*?" question, and unveils a number of open challenges, for instance in terms of knowledge modelling (e.g. *What is the optimal level of detail in the description of activities/outputs?*) and of visualisation (e.g. *How do we handle a focus+context view over heterogeneous and massive collections?*).

To conclude, our vision is that the approach, primarily intended at ameliorating the readability and usability of digital assets over time for scientists or specialists at large, can also renew the way historical sciences can be presented, or "storytold", to the wide public, at a time when social networking and crowdsourcing significantly infiltrate practices of scientific communities.

References

[1] G. Schuller, "Designing universal knowledge", Lars Müller publishers, Baden 2009.

[2] Open science, blog-like resources curated by F. Piron

[3] Europeana project web portal

[4] 3D-ICONS project web portal

[5] A. J. Van Niekerk, "The Strategic Management of Media Assets; A Methodological Approach", Allied Academies, New Orleans Congress, 2006.

[6] R. Vuillemot, T. Clement, C. Plaisant, A. Kumar, "What's being said near 'Martha'? Exploring name entities in literary text collections", IEEE Symposium on Visual Analytics Science and Technology (VAST), 2009, pp.107-114.

[7] F. Wanner, J. Fuchs, D. Oelke, D.A. Keim, "Are my children old enough to read these books? Age suitability analysis", POLIBITS: Research journal on computer science and computer engineering with applications, 2011.

[8] S. Koch, J. Markus, M. Wörner, A. Müller, T. Ertl, "VarifocalReader – In-Depth Visual Analysis of Large Text Documents", IEEE Transactions on Visualization and Computer Graphics, Vol. 20, No. 12, 2014.

[9] <http://www.cidoc-crm.org>

[10] D. Keim, J. Kohlhammer, G. Ellis, F. Mansmann, "Mastering the Information Age – Solving Problems with Visual Analytics", 2010, online at <www.vismaster.eu>.

[11] J.Y. Blaise, I. Dudek, P. Bénistant, A. Durand, "Online 2D/3D graphic interfaces using XML 'repurposable' heritage contents", 12th International Conference on Computer Graphics, Visualisation and Computer Vision, WSCG 2004, Plzen 2004, pp. 39-46.

[12] W. Aigner, S. Miksch, H. Schumann, C. Tominski, "Visualization of Time-Oriented Data", Human-Computer Interaction Series: Springer, 2011.

[13] H. Ping Tserng, Yu-Cheng Lin, "Developing an activity-based knowledge management system for contractors", Automation in Construction 13, 2004, pp.781– 802.

[14] M.J. Eppler, "A Generic Framework for Information Quality in Knowledge-intensive Processes" Proceedings of the Sixth International Conference on Information Quality, 2001.

[15] J.Y. Blaise, I. Dudek, "Picturing what others know: towards a dashboard for interdisciplinarity", Proc. of the 14th International i-KNOW conference, 2014, Article no 15, ACM 2014.