The archaeological excavation report of Rigny: an example of an interoperable logicist publication

Pierre-Yves Buard, Béatrice Markhoff, Olivier Marlet, Xavier Rodier, Elisabeth Zadora-Rio

To cite this version:

Pierre-Yves Buard, Béatrice Markhoff, Olivier Marlet, Xavier Rodier, Elisabeth Zadora-Rio. The archaeological excavation report of Rigny: an example of an interoperable logicist publication. CIDOC 2018, Sep 2018, Heraklion, Greece. hal-01892412

HAL Id: hal-01892412
https://hal.archives-ouvertes.fr/hal-01892412
Submitted on 10 Oct 2018

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.
The archaeological excavation report of Rigny: an example of an interoperable logicist publication
CIDOC 2018 Heraklion, Crete, Greece

The archaeological excavation report of Rigny: an example of an interoperable logicist publication

Pierre-Yves Buard¹, Béatrice Markhoff², Olivier Marlet³, Xavier Rodier³, Elisabeth Zadora-Rio³

¹ Pôle Document Numérique de la Maison de la Recherche en Sciences Humaines, Université de Caen Basse-Normandie, Esplanade de la Paix - Campus 1, 14000 Caen, France
² Laboratoire d'Informatique Fondamentale et Appliquée de Tours, Université de Tours, 64, Avenue Jean Portalis, 37200 Tours, France
³ UMR7324 CITERES – Laboratoire Archéologie et Territoires, Université de Tours, CNRS, 33-35 allée Ferdinand de Lesseps, 37200 Tours, France

olivier.marlet@univ-tours.fr

Abstract. The logicist programme, which was initiated in the 1970s by J.C. Gardin, aims to clarify the reasoning processes in the field of archeology and to explore new forms of publication, in order to get over the growing imbalance between the flood of publications and our capacities of assimilation. The logicist programme brings out the cognitive structure of archaeological constructs, which establishes a bridge between empirical facts, or descriptive propositions, at one end of the argumentation, and interpretative propositions at the other end. This alternative form of publication is designed to highlight the chain of inference and the evidence on which it stands. In the case of the logicist publication of the archaeological excavation in Rigny, our workflow can provide different levels of access to the content, allowing both speed-reading and in-depth consultation. Both the chains of inference and the ArSol database containing the field records that provide evidence for the initial propositions are visualized in a diagram structure. We rely on CIDOC CRM entities for ensuring the semantic interoperability of such publications within the Linked Open Data. Inference chains are mapped to CRMinf and ArSol records are mapped to CRM, CRMSci and CRMArcheo. Moreover, as part of the work carried out by the French Huma-Num MASA Consortium, a project is underway to allow the building of logicist publications starting from a graphical interface for describing the structure and content of propositions.

Keywords: publication, archaeology, logicism, inference

Introduction

It is now widely recognized that the number of papers currently published in archaeology is such that we are unable to read more than a very small fraction of the literature relevant to our research interests. Instead, we consult some of it, following our own selection strategies. The paradox is that while we are perfectly aware of this phenomenon, we continue to write as if our works were to be read, without any attempt to redraft them in the alternative perspective, that is consultation. Digital publishing as such does not solve the problem: it makes it worse.

A special challenge when consulting publications is to quickly assess their relevance to our work, and to do so, to have easy access to the scientific reasoning and the provenance of the diverse items (observations, comparisons, references) on which it is based.

In the pages that follow, we will use the forthcoming digital publication of the archaeological excavation of the settlement and church in Rigny (Indre-et-Loire, France) as a test-case to show that a publishing workflow, based on the principles of the logicist programme, the principles of which we will recall, can provide different levels of information retrieval, allowing both speed-reading and in-depth consultation.

1 The archaeological excavation of the medieval settlement and church in Rigny (Indre-et-Loire, France)
The summer excavation, which took place around the medieval church in Rigny (Indre-et-Loire, France) from 1986 to 1999, was started with a very small group of volunteers and became from 1989 onwards a training excavation for students in archaeology at the University of Tours (Fig. 1). The aim of the excavation was to retrace the formation and transformations of a parish centre and to study the population buried in the cemetery. The excavation revealed an occupation from the 7th to the 19th century (Zadora-Rio, Galinié et al. 1992, Zadora-Rio, Galinié et al. 1995 and Zadora-Rio, Galinié et al. 2001).

The field recording was based on the model used in the Tours excavations, inspired itself by the recording system developed during the 1960s in Great Britain, especially by Philip Barker for the excavations at Wroxeter and Hen Domen and by Martin Biddle and Birthe Kjølbye Biddle for the urban excavations at Winchester (Galinié 2013). The computerization of the data was introduced in Rigny in 1990 in the form of a relational database, which constituted the first version of the ArSol database, developed since then by the Laboratoire Archéologie et Territoires, and upgraded by the addition of a second module for the processing of ceramics (Galinié et al. 2005; Husi, Rodier 2011). ArSol is designed for the management and processing of stratigraphical data and artifacts in the perspective of an analysis of the chronology and spatial organization of excavated sites.

Fig. 1. The excavation of the cemetery and buildings around Rigny's church

2 Logicism and digital publication

The logicist programme was developed by Jean-Claude Gardin in order to condense and to schematize the architecture of scientific constructs. From the start, its aim was twofold. The first was epistemological: the purpose was to make explicit the steps of the reasoning by distinguishing, on the one hand, the basic data (or "initial propositions"), and on the other hand, the inference operations carried out on these data to establish the interpretative hypotheses, so as to constitute a tree structure which gives a synoptic representation of the argumentation and enables a quick assessment of its relevance (Gardin 1979: 244-273). The argumentation takes the form of a series of inference operations from \{P0\} (initial propositions) to \{Pn\} (final propositions) via intermediate propositions \{Pi\} (Gardin et al. 1987: 19) (Fig. 2). The inference rules are expressed as "if p1, then p2" (Gardin 1989).
The second objective was editorial. Like all modelling, logicist structuring is a reduction, but it preserves all the constituent elements of the cognitive construction, freed from the rhetorical apparatus traditionally used in publications. It thus constitutes a means of reducing the imbalance observed between bibliographic production and our consuming capacities, and opens the way to a form of publication adapted to the growing preponderance of consultation over reading (Gardin 1999).

In the decades 1980-1990, logicist analysis was experimented in the field of archaeology, art history and history, but its dissemination remained very limited, because the exercise was long considered unattractive.

This first exploratory phase showed the epistemological value of logicist analysis, but it had no substantial influence on the designing of publications.

It is the development of information technologies that has enabled the non-linear reading possibilities offered by logicist schematizations to be exploited, while making them less ascetic thanks to a multimedia environment (Roux, Blasco 2004). The SCD format, acronym of Scientific Constructs and Data, was conceived by Valentine Roux (CNRS) and Philippe Blasco (Editions Epistèmes) for the digital publication of logicist rewritings and associated data in the domain of the archaeology of techniques. It was used first in the “Référentiels” collection (2003-2010), which consisted of short printed volumes accompanied by a CD-ROM containing the logicist schematizations, and later in the online journal Arkeotek (www.thearkeotekjournal.org), created by Valentine Roux in 2007 for the publication of papers and reference data in the archaeology of techniques.

Since 2011, thanks to the development of new web technologies, the SCD format has been entirely reprogrammed in XML TEI (Text Encoding Initiative) by the Pôle du Document Numérique (Digital Document Centre) of the Maison de la Recherche et des Sciences de l’Homme of Caen. It is in this XML TEI format, which offers new possibilities for navigation between text, logicist schematizations and online databases, that the publication of the Rigny excavation, initially prepared for publication in the Référentiels collection, is currently being processed (Zadora-Rio, Galinié et al., forthcoming).

3 The contribution of information technology to the accessibility of excavation data

It is well known that excavation is an irreversible process and does not allow the experiment to be repeated: once the excavation is finished, the field recording becomes the archaeologist’s primary source. In the 1980s, the development of computers allowed the development and even the multiplication of numerous databases for the recording of field data. The widespread use of the Internet and the improvement of information systems now enable access to these field databases and thus the possibility for any researcher to check the recorded data.
Thus, in 1996, the United Kingdom set up the Archaeology Data Service (ADS, http://archaeologydataservice.ac.uk), York University’s digital resource center providing access to all digitized archaeological records (databases, graphic documents, photographs, specialists reports, grey literature, etc.). In addition, the journal Internet Archaeology (http://intarch.ac.uk/) has set up online publications with hypertext links referring to the data used in the text, even if the general design of the papers remains traditional. In France, the digitization and online availability of field data is a later phenomenon. Among the early, local initiatives is the database of the Laboratoire Archéologie et territoires in Tours: ArSol (Archives du Sol, http://arsol.univ-tours.fr), which is on-line since 2014 (Fig. 3). Since 2013, the MASA Consortium (Mémoire des Archéologues et des Sites Archéologiques, https://masa.hypotheses.org), belonging to the Very Large Research Infrastructure Huma-Num (https://www.huma-num.fr), has set itself the objective of assisting archaeologists in digitizing and making available their excavation archives by applying good practices such as the FAIR principles. At the European level, these database publications are being developed thanks to the ARIADNE consortium and the implementation of a platform to access archaeological data (http://portal.ariadne-infrastructure.eu).

The online access to the primary data, which allows the reader to get acquainted with the original field records, is a decisive step forward, which opens the way to new, shorter publications, more focused on the argumentation than on the dreary description of excavated features.

Fig. 3. ArSol, online database of the Laboratoire Archéologie et territoires: http://arsol.univ-tours.fr/
4 From excavation records to logicist publication

On-site field recording begins with the description of stratigraphical units, their grouping into hierarchical spatial entities (Features, Walls, Burials or Structures), their phasing, and then their functional, chronological and morphological interpretation according to an empirico-inductive approach. Logicist writing follows the reverse order, starting from the interpretation to reconstruct backward the chain of inferences bridging the gap between the conclusions \{P_n\} and the empirical data \{P_0\} at the other end of the argumentation, and bringing out the structure of the interpretative constructs which correlate empirical observations to lifestyles and social practices (Fig. 4).

![Diagram of empirico-inductive and hypothetico-deductive processes in archaeological excavation]  

**Fig. 4.** The logicist schematization of interpretative constructs applied to the case of archaeological excavation

It is important to point out that the initial propositions \{P_0\} belong to three categories:

1) **Observation data selected from excavation records.** These descriptive features used in the argumentation may concern either the intrinsic properties of archaeological entities (materials, form, etc.) or their relative chronology (stratigraphic relationships). They represent only a small selection of the field-recording and post-exavcation databases.

2) **Comparative data.** The comparative data are considered as initial proposition \{P_0\}, as well as the observation data, because the statement of similarity, which forms the basis of the analogical reasoning, is never the result of a well-defined mathematical or logical procedure. As J.C. Gardin has pointed out, it forms the basis of the "attribute transfer": "IF two artefacts or monuments X and Y are declared comparable, in view of certain common properties (shape, materials, ornaments, etc.), and that Y is endowed in addition with one or more known attributes (date, origin, function), THEN one is entitled to transfer to X the same attributes" (Gardin et al. 1987: 235).

3) **Reference data.** The so-called reference data correspond to the background knowledge referring either to the common sense (Gardin 1989) or to specialist knowledge. This latter category includes laboratory analyses (e.g. radiocarbon dating), as well as artifact dating when based on a typo-chronology established by other publications. Reference data are considered to be initial data \{P_0\}, as well as observation and comparison data, because they are not demonstrated in the publication.
In spite of the apparent diversity of the initial data-set in the case of an archaeological excavation, the rules of inference are relatively standardized: at all levels, from the successive intermediate propositions (P1, P2, ..., Pn) to the final propositions (Pn), they usually consist in assigning to one or several entities either a function (in the broadest meaning of the word), or a chronology (date, or time-span...), or an original morphology. Their compilation in a knowledge base, or inference-rules corpus, such are those which are to be implemented in the Arkeotek Project (Roux, Aussenac-Gilles 2010) would give the opportunity to test their degree of validity, and to discuss, for example, the material correlates which are considered necessary in order to assign to a building a function of storage, of dwelling or of place of worship, in such and such chrono-cultural context.

5 The logicist publication of the Rigny excavations, and beyond

5.1 The architecture of the publication

The current logicist publication of the Rigny excavations consists of several “blocks”, which provide different levels of access to the content, allowing both speed-reading and an in-depth consultation (Fig. 5).

![Diagram of the different reading levels of Rigny's publication](http://arsol.univ-tours.fr)

The first block contains the Narrative, which gives a linear outline of the results and is connected by hypertext links to the “logicists arguments” from {P0} to {Pn} (“Block 4”). It is designed for speed-reading, but it also allows for the assessment of the argumentation and the retrieval of the data on which it is based if the reader chooses to follow the links (Fig. 5).

Another block (Block 2) contains the logicist diagrams, which display the argumentation in the form of an inference tree developing from left to right (Fig.6). These diagrams, which provide a graphic overview of the argumentation, are interactive and allow access to the detailed argumentation (Fig. 7). The diagrams are automatically produced through the XML TEI encoding of the texts. The digital publication also provides links to the bibliography, to internal cross-references and to the ArSol online database.
The use of logicist analysis leads to a reduction of the published text, if compared to a classic monograph of excavation, without loss of content, and brings out the chains of inference bridging the gap between empirical facts, or descriptive propositions, at one end of the argumentation, and final propositions (or conclusions) at the other end. The highlighting of the interpretative processes allows their critical assessment and opens the way for different levels of reading, from speed-reading of the results to the in-depth examination of the evidence. The logicist diagrams, which give an overview of the inference trees as well as an access to the detailed argumentation, are especially appropriate for a non-linear consultation of the publication.
5.2 Helping building logicist trees

For the development of Rigny's electronic publication, the Pôle du Document Numérique of the MRSH of Caen has set up a tool to generate automatically the graph of inference in SVG (Scalable Vector Graphics) format from the source XML TEI file (see Fig. 6) (Buard, Zadora-Rio forthcoming). To go one step beyond, within the MASA project, the reverse is experimented: the online tool will allow author to graphically construct the reasoning graph structure by linking the propositions to each other from the initial propositions (evidences) to the final propositions (conclusions). From this inference graph, an XML TEI file containing the logicist analysis is generated. An XML editor can then be used to supplement the propositions with text, illustrations, or bibliographic references. The author can then simultaneously interact with the graph or the TEI file, allowing easier progress in the design of the logicist publication.

Philippe Husi (CNRS), coordinator of the research programme on "Medieval and modern ceramics in the Middle Loire catchment area from the 6th to the 19th century" (Husi 2013) plans to experiment this application for the publication of the results of this programme. This research is based on the material evidence from about forty sites excavated in the area under study, accessible online on the Iceramm network website (network about medieval and modern ceramics, http://iceramm.univ-tours.fr).

The production of this new publication following the rules of logicist analysis will allow to test the electronic publication process initiated in the Rigny's publication and it will be a step forward towards the constitution of "logicist corpses" structured in data and interpretation rules (Roux, Aussenac-Gilles 2013). Their interoperability could be achieved through a mapping with CIDOC CRM. It would give the opportunity to compare and to discuss the validity of the rules of inference implemented in different logicist publications.

6 Towards semantic interoperability: mapping logicist publications to CIDOC CRM

Using web technologies, and among them the XML TEI, we provide archeologists with a quick and easy access to the content of a logicist publication, the underlying reasoning being a part of this content. But our goal is also to provide to web applications the ability to deal with such knowledge. For the same purpose, we already extended the ArSol online web publication with a CIDOC CRM based SPARQL querying API (Marlet, Curet et al., 2016).

As Gardin has pointed out, the logicist schematizations can be compared to knowledge-based systems combining the data (stored in a Fact base) with rules of reasoning (stored in a Rule base) through an inference engine (Gardin et al. 1987 :27-55; Gardin 1987 :7). Due to the formalized structure of the logicist schematizations, the inference rules can be processed with the same tools as those used for the datasets. Thus, by mapping the logicist propositions with entities of the CIDOC CRM and in particular with those of the CRMinf extension1, we can ensure a semantic interoperability of this publication within the Linked Open Data.

To this end, we propose to map the Rigny publication to the CRMinf, the extension of the CIDOC CRM intended to be used as a global schema for integrating metadata about argumentation and inference-making, in the following way:

(i) The field records of the ArSol database are mapped to the basic CIDOC CRM plus CRMsci and CRMarcheoo extensions. Thus field data are potentially interoperable and can be available both as such and as evidence supporting scientific reasoning.

(ii) Inference chains are mapped to the CRMinf extension (embedded in the CRMsci).

According to the principles presented in Section 4, the initial propositions \{P0\} have been typed according to whether they are observation data, comparative data, or reference data corresponding to the background knowledge.

The propositions have been assigned to three categories according to what they are dealing with, either Function, Time or Morphology. “Function” is taken in the broadest meaning of the word (from assigning a function to a structure or a building to complex socio-cultural inferences). The category “Time” encompasses the propositions dealing with dates, relative chronology, or duration. “Morphology” refers to propositions concerning the original form of structures or buildings, architectural reconstructions or spatial partitions (Fig. 8).

1 Reference document: http://www.ics.forth.gr/isl/CRMext/CRMinf/docs/CRMinf-0.7.pdf
Initial propositions \{P0\} can be mapped to CRMinf elements depending on whether they are based on observation and comparison data or on reference data.

For a proposition based on \textbf{observation data} or \textbf{comparison data}, mapping could be:

\[
\begin{align*}
S15_{\text{Observable\_Entity}} & \rightarrow O11_{\text{was\_described\_by}} \rightarrow S6_{\text{Data\_evaluation}} \quad (\text{IsA} \\
I5_{\text{Inference\_Making}} & \text{ IsA } I1_{\text{Argumentation}}) \rightarrow J2_{\text{conclued\_that}} \rightarrow I2_{\text{Belief}} \rightarrow J4_{\text{that}} \rightarrow \\
I4_{\text{Proposition\_Set}} & \\
I5_{\text{Inference\_Making}} & \rightarrow J3_{\text{applies}} \rightarrow I3_{\text{Inference\_Logic}}
\end{align*}
\]

For a proposition based on \textbf{reference data}, mapping could be:

\[
\begin{align*}
E31_{\text{Document}} \quad (\text{IsA} \quad E73_{\text{Information\_Object}}) & \rightarrow J7_{\text{is\_evidence\_for}} \rightarrow I7_{\text{Belief\_Adoption}} \quad (\text{IsA} \\
I1_{\text{Argumentation}}) & \rightarrow J6_{\text{adopted}} \rightarrow I2_{\text{Belief}} \rightarrow J4_{\text{that}} \rightarrow I4_{\text{Proposition\_Set}}
\end{align*}
\]

For intermediate or final propositions, mapping could be:

\[
\begin{align*}
I4_{\text{Proposition\_Set}} & \rightarrow J4_{\text{is\_subject\_of}} \rightarrow I2_{\text{Belief}} \rightarrow J1_{\text{was\_premise\_for}} \rightarrow \\
S8_{\text{Categorical\_hypothesis\_building}} \quad (\text{IsA} \quad I5_{\text{Inference\_Making}} \text{ IsA } I1_{\text{Argumentation}}) & \rightarrow J2_{\text{conclued\_that}} \rightarrow I2_{\text{Belief}} \rightarrow J4_{\text{that}} \rightarrow I4_{\text{Proposition\_Set}}
\end{align*}
\]

In this way, the reasoning stream is represented by the graph of I2_Belief instances related by S8_Categorical_hypothesis_building instances, which are also I5_Inference_Making instances by IsA entailments. It is important to notice that each I2_Belief instance has a I4_Proposition_Set instance which can be annotated with the kind of applied inference: functional inference, morphological inference or temporal inference.

This is illustrated by the diagram in Fig. 9, which shows that the Rigny logicist publication is represented by a I1_Augumentation, P14_carried_out_by Elisabeth Zadora-Rio, involving S4_Observation, I5_Inferenc_Making and I7_Belief_adoption instances, etc.
For the time being, the inference graph-structure and the texts of the logicist propositions are both represented in XML TEI files, structured in a way which allows the tags to be explicitly associated with the entities and properties of the CIDOC CRM and its extensions (Fig. 10).

Fig. 9. Diagram representing the CRMinf model as it can be used for Rigny’s publication

Fig. 10. XML-TEI file model using CIDOC-CRM entities to identify tags
The CIDOC CRM metadata could be easily extracted from the XML files for answering SPARQL queries when needed (by defining mappings using either the 3M tool\(^2\), GRDDL\(^3\) or SPARQL Generate\(^4\)).

**Conclusion**

The logicist framework set up by Jean-Claude Gardin in the 1970s was intended to provide a solution to overcome the growing imbalance between the flood of publications and our assimilation capacities, but it is only recently that the developments in computing and the web, the XML format and its widespread use, plus the semantic level, has allowed its implementation in publishing workflows. Thanks to the formal organization of arguments in logicist publications, they can be processed as datasets, that can be made available in the Linked Open Data, making them potentially interoperable. The presented current models and mappings are experimental and need to be applied to other logicist publications in order to assess their validity and their semantic interoperability efficiency. Moreover, the innovative aspects of Rigny's publication and, hopefully, of the forthcoming Iceramm’s publication, need to be discussed before our propositions can be implemented for further excavation publications. Let’s hope that Rigny's logicist publication will encourage other archaeologists to try the adventure.

**Bibliography**


http://citeres.univ-tours.fr/doc/lat/pecada/F2_17.pdf


---

\(^2\) Reference site: https://www.ics.forth.gr/isl/index_main.php?l=e&c=721

\(^3\) Reference site: https://www.w3.org/2001/sw/wiki/GRDDL

\(^4\) Reference site: https://ci.mines-stetienne.fr/sparql-generate/


