The Archaeological Excavation Report of Rigny: An Example of an Interoperable Logicist Publication

Olivier Marlet 1,*, Elisabeth Zadora-Rio 1,†, Pierre-Yves Buard 2,†, Béatrice Markhoff 3,† and Xavier Rodier 1,†

1 UMR7324 CITERES–Laboratoire Archéologie et Territoires, CNRS, Université de Tours, 37200 Tours, France; e.zadora.rio@gmail.com (E.Z.-R.); xavier.rodier@univ-tours.fr (X.R.)
2 Pôle Document Numérique de la Maison de la Recherche en Sciences Humaines, Université de Caen Basse-Normandie, 14000 Caen, France; pierre-yves.buard@unicaen.fr
3 Laboratoire d’Informatique Fondamentale et Appliquée de Tours, 37200 Tours, France; beatrice.markhoff@univ-tours.fr
* Correspondence: olivier.marlet@univ-tours.fr; Tel.: +33-247-361-506
† These authors contributed equally to this work.

Received: 23 January 2019; Accepted: 19 February 2019; Published: 22 February 2019

Abstract: The logicist program, which was initiated in the 1970s by J.C. Gardin, aims to clarify the reasoning processes in the field of archaeology and to explore new forms of publication, in order to overcome the growing imbalance between the flood of publications and our capacities of assimilation. The logicist program 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. 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 (Indre-et-Loire, France), 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 the International Committee for Documentation Conceptual Reference Model (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

1. 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 for an alternative mode of consumption, that is consultation. Digital publishing as such does not solve the problem. It makes it even worse if the transition from the paper to electronic support is not associated with a deep reorganisation of our modes of publication.

In recent years, the creation of web repositories such as Archaeology Data Service (http://archaeologydataservice.ac.uk), which gives access to all the electronic data-sets produced by British
excavations, has allowed publications such as Internet Archaeology (http://intarch.ac.uk) to provide links to their underlying data sets, but these achievements had little impact on the form of the publication, which remained otherwise unchanged.

The logicist program was developed by Gardin in the 1970’s in order to condense and schematize the architecture of scientific constructs. It provides a framework for such a transformation, which is especially suitable in the context of web publishing. It has been used in the context of the archaeology of techniques in The Arkeotek Journal (http://www.thearkeotekjournal.org/) and we will show in this paper that it can also be applied to the publication of excavation reports.

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). This publication will serve as a test-case to show that a publishing workflow, based on the principles of the logicist program, (detailed below), can provide different levels of information retrieval, allowing both speed-reading and in-depth consultation. We will present the structure of this publication and the tools used to facilitate its drafting and management. In addition to the publication, a complementary work was carried out to map the structure of the reasoning with an ontology adapted to the archaeological data and in particular to the scientific reasoning.

The publication tool described in this paper and which illustrates its arguments will be released in a production version to the public in the course of 2019 and be hosted at https://www.unicaen.fr/puc/rigny/.

2. 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 from 1989 onwards became a training excavation for students in archaeology at the University of Tours (Figure 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 [1–3].

Figure 1. The excavation of the cemetery and buildings around Rigny’s church.
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 Barker for the excavations at Wroxeter and Domen and by Biddle and Biddle for the urban excavations at Winchester [4]. 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 [5,6]. ArSol is designed for the management and processing of stratigraphic data and artifacts in the perspective of an analysis of the chronology and spatial organization of excavated sites.

3. Logicism and Digital Publication

The logicist program 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 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. This would constitute a tree structure, which gives a synoptic representation of the argumentation and enables a quick assessment of its content and logical framework [7] (pp. 244–273). The argumentation takes the form of a series of inference operations from \( \{ P_0 \} \) (initial propositions) to \( \{ P_n \} \) (final propositions) via intermediate propositions \( \{ P_i \} \) [8] (p. 19) (Figure 2). The inference rules are expressed as “if \( P_1 \), then \( P_2 \)” [9].

![Figure 2. A schematization of the logical structure of scholarly papers in human sciences according to Gardin (1987) p.19](image)

The second objective was editorial. Like all modeling, 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 [11].

In the decades 1980–1990, logicist analysis was experimented with 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 [12]. The Scientific Constructs and Data (SCD) format was conceived by Roux (CNRS) and 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 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 Text Encoding Initiative (TEI) 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 [13].

4. 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 some of these field databases and thus the possibility for any researcher to check the recorded data available online.

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 more recent 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 online since 2014 (Figure 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. The FAIR principles (Findable, Accessible, Interoperable and Reusable data) imply in particular that; the data have metadata, that they are referenced in a sustainable way, that they are accessible on the web, that they respect the standards of the domain, that they are as reliable as possible and that they are distributed under a clear license [14]. At the European level, these archaeological 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), following the example of other European initiatives such as Europeana (https://www.europeana.eu/portal/fr).

The online access to the primary data, which allows the reader to get acquainted with the original field records, is a decisive step forward. It opens the way to new, shorter publications, more focused on the argumentation than on the dreary description of excavated features [15,16].
5. From Excavation Records to Logicist Publication

On-site field recording begins with the description of stratigraphic 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 backwards 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 (Figure 4).
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-excavation databases.

2. **Comparative data.** The comparative and observation data are considered as initial proposition \( \{P_0\} \), because the statement of similarity which forms the basis of the analogical reasoning, so common in archaeology, is never the result of a well-defined mathematical or logical procedure. As 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.” [10] (p. 235).

3. **Reference data.** The so-called reference data correspond to the background knowledge referring either to common sense [9] or 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 \( \{P_1, P_2, ..., P_i\} \) to the final propositions \( \{P_n\} \), they usually consist of 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 as those which are to be implemented in the Arkeotek Project [17], 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 a chrono-cultural context.
6. The Logicist Publication of the Rigny Excavations, and Beyond

6.1. 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 (Figure 5).

**Figure 5.** Diagram of the different reading levels of Rigny’s publication.

Block 1 contains the narrative, which gives a linear outline of the results and is connected by hypertext links to the logicist’s arguments from \( P_0 \) to \( P_n \) (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 (Figure 5).

Block 2 contains the logicist diagrams, which display the argumentation in the form of an inference tree developing from left to right (Figure 6). These diagrams, which provide a graphic overview of the argumentation, are interactive and allow access to the detailed argumentation (Figure 7). The diagrams are automatically produced through the XML TEI encoding of the texts. The digital publication also provides links to the bibliography, internal cross-references and the ArSol online database.

**Figure 6.** Extract from the logicist diagram in section 1 of Rigny’s publication, “The three churches of Rigny (from 7th/8th century to 1859)".
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 logistic diagrams, which give an overview of the inference trees as well as access to the detailed argumentation, are especially appropriate for a non-linear consultation of the publication.

6.2. Helping Building Logicist Trees

For the development of Rigny’s electronic publication, the Pôle du Document Numérique of the Maison de la Recherche et des Sciences de l’Homme (MRSH) of Caen has set up a tool to automatically generate the graph of inference in Scalable Vector Graphics (SVG) format from the source XML TEI file (Figure 6) [18]. To go one step beyond, within the MASA project the reverse is experimented. The online tool will allow the author to graphically construct the reasoning graph structure by linking the propositions to each other from the initial propositions (facts) 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 simultaneously interact with the graph or the TEI file, allowing easier progress in the design of the logicist publication.

Husi, CNRS coordinator of the research program on, “Medieval and modern potteries in the Middle Loire catchment area from the 6th to the 19th century” [19], plans to experiment with this application for the publication of the results of this program. 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 potteries, http://iceramm.univ-tours.fr).
The production of this new publication following the rules of logicist analysis will test the electronic publication process initiated in Rigny’s publication and it will be a step forward towards the constitution of logicist corpuses structured in data and interpretation rules [17].

Further developments within a separate project will investigate their interoperability with other logicist publications, which could be achieved through a mapping with the CIDOC CRM. The Conceptual Reference Model of the International Committee for Documentation is an ontology dedicated to the field of cultural heritage. This is a standard that is increasingly used, particularly by the British Museum and the Bibliothèque National de France. Mapping the publication with CIDOC would give the opportunity to compare and to discuss the validity of the rules of inference implemented in different logicist publications.

7. Towards Semantic Interoperability: Mapping Logicist Publications to CIDOC CRM

Using web technologies, and among them the XML TEI, we provide archaeologists with quick and easy access to the content of a logicist publication, the underlying reasoning being 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 [20].

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 [8] (pp. 27–55) [10] (p. 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 extension [21], 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 CRMarcheo 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 \{P_0\} 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 (Figure 8).
Figure 8. Example of typologies of initial propositions \{P0\} and intermediate propositions.

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 observation data or comparison data, mapping could be:

\[
\text{S15}_\text{Observable\_Entity} \rightarrow \text{O11}_\text{was\_described\_by} \rightarrow \text{S6}_\text{Data\_evaluation (IsA I5}_\text{Inference\_Making IsA I1}_\text{Argumentation)} \rightarrow \text{J2}_\text{conclued\_that} \rightarrow \text{I2}_\text{Belief} \rightarrow \text{J4}_\text{that} \rightarrow \text{I4}_\text{Proposition\_Set}
\]

\text{I5}_\text{Inference\_Making} \rightarrow \text{J3}_\text{applies} \rightarrow \text{I3}_\text{Inference\_Logic}

For a proposition based on reference data, mapping could be:

\[
\text{E31}_\text{Document (IsA E73}_\text{Information\_Object)} \rightarrow \text{J7}_\text{is\_evidence\_for} \rightarrow \text{I7}_\text{Belief\_Adoption (IsA I1}_\text{Argumentation)} \rightarrow \text{J6}_\text{adopted} \rightarrow \text{I2}_\text{Belief} \rightarrow \text{J4}_\text{that} \rightarrow \text{I4}_\text{Proposition\_Set}
\]

For intermediate or final propositions, mapping could be:

\[
\text{I4}_\text{Proposition\_Set} \rightarrow \text{J4}_\text{is\_subject\_of} \rightarrow \text{I2}_\text{Belief} \rightarrow \text{J1}_\text{was\_premise\_for} \rightarrow \text{S8}_\text{Categorical\_hypothesis\_building (IsA I5}_\text{Inference\_Making IsA I1}_\text{Argumentation)} \rightarrow \text{J2}_\text{conclued\_that} \rightarrow \text{I2}_\text{Belief} \rightarrow \text{J4}_\text{that} \rightarrow \text{I4}_\text{Proposition\_Set}
\]

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 Figure 9, which shows that the Rigny logicist publication is represented by a I1_Argumentation, P14_carried_out_by Zadora-Rio, involving S4_Observation, I5_Inference_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 that allows the tags to be explicitly associated with the entities and properties of the CIDOC CRM and its extensions (Figure 10).

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

Figure 10. XML TEI file model using the International Committee for Documentation Conceptual Reference Model (CIDOC CRM) entities to identify tags.
This TEI XML file is then dynamically transformed into HTML for the narrative or into SVG for the diagrams. The CIDOC CRM metadata could easily be extracted from the TEI XML files for answering SPARQL queries when needed, by defining mappings using either the 3M tool [22], GRDDL [23] or SPARQL Generate [24]. The mapping of Rigny’s publication with CRMinf is an ancillary ongoing project that will not appear in the forthcoming publication.

8. Conclusions

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, including the XML format and the semantic level, have 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 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.

Author Contributions: Writing–original draft, O.M.; Writing–review & editing, E.Z.-R., P.-Y.B., B.M. and X.R.

Funding: This research received no external funding.

Acknowledgments: This research was carried out with the support of the MASA Consortium (https://masa.hypotheses.org/) from the TGIR HUMA-NUM (https://www.huma-num.fr/).

Conflicts of Interest: The authors declare no conflict of interest.

References


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).