



**HAL**  
open science

# Publishing open-access bibliographical data on Ancient Greek and Latin texts: challenges, constraints, progression

Julie Giovacchini, Laurent Capron

## ► To cite this version:

Julie Giovacchini, Laurent Capron. Publishing open-access bibliographical data on Ancient Greek and Latin texts: challenges, constraints, progression. *Journal of Data Mining and Digital Humanities*, 2021, Atelier Digit\_Hum, 10.46298/jdmdh.7530 . hal-03242823v2

**HAL Id: hal-03242823**

**<https://hal.science/hal-03242823v2>**

Submitted on 1 Oct 2021

**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.

## Publishing open-access bibliographical data on Ancient Greek and Latin texts: challenges, constraints, progression<sup>1</sup>

Julie Giovacchini<sup>1\*</sup>, Laurent Capron<sup>1</sup>

1 Centre Jean Pépin, UMR8230 CNRS-ENS-PSL

\*Corresponding author: Julie Giovacchini [julie.giovacchini@cnrs.fr](mailto:julie.giovacchini@cnrs.fr)

### Abstract

We present here both some of our thoughts on methodology in relation to the specific constraints that complexify the ways of structuring and accessing bibliographical data in the Sciences of Antiquity, and the solutions adopted by the IPhIS-CIRIS project for dealing with these constraints. The project began in 2014 in a general scientific environment that was still being standardised and structured, with digital bibliographical resources in this disciplinary field becoming increasingly numerous, although of uneven quality and hard to access and/or private.

### keywords

database; open access; bibliography; Sciences of Antiquity; semantic web; Nakala

### INTRODUCTION

The publication of bibliographical data on Classical and Late Antiquity<sup>2</sup> has a considerable history already, since it dates back to the early twentieth century with the publication in 1927 of the two volumes of an important retrospective bibliography entitled *Dix Années de Bibliographie Classique* (1914-1924) [Ten years of Classical bibliography, 1914-1924], by Jules Marouzeau<sup>3</sup>. In keeping with the evolving needs specific to each discipline with regard to Antiquity, a number of specialised bibliographies have flourished with the intention of providing a comprehensive list of works published in their respective fields. Some of these bibliographies have since made use of electronic media for their circulation, continuing to maintain the paper version as their reference, while others have been digital from the outset. Thus each has its own classification scheme and indexing system, and sometimes even its own canonical forms of authority. Since most of these resources are proprietary, it is not their natural vocation to be mutually compatible: at best, certain practices tend towards a standardisation of formats and classification schemes. The work required in terms of referencing the data to make it interoperable has not yet been completed, although numerous initiatives demonstrate a desire to use common repositories. The IPhIS-CIRIS project (*Information Philologique – Savoirs Antiques* [philological information - ancient knowledge], Centre Jean Pépin, CNRS-ENS) is participating more particularly in defining a set of reference data of names of authors and titles of ancient works to be fed into a database of editions of ancient texts. When drawing up these

---

<sup>1</sup> We would like to express our gratitude to Mrs Katherine Parsons who has translated this paper from French to English.

<sup>2</sup> This is the expression we shall use in the following pages to refer to all the intellectual productions written originally in Ancient Greek and/or Latin during a period stretching from the 8th century BC to the early 8th century AD.

<sup>3</sup> [Hilbold, 2019]

files we come up against a number of difficulties: concurrent forms for the same entry, inexistent forms of the state of certain works in the existing repositories, duplicates, homonymic forms of titles or authors, dubious attributions of works to a particular author, controversial attributions to various authors, etc. Deciding on the forms of these tables and the links between them created a number of technical difficulties from the outset; these are characteristic of Antiquity and have not yet been fully dealt with in the computer modelling.

## I STATE OF PLAY

### 1.1 A blossoming of initiatives

There is today a large quantity of bibliographical resources in the field of Classical and Late Antiquity that is available - partly or fully - on the Internet. Mention may be made of *L'Année Philologique*, published annually since 2017 by Brepols, after having for a long time been distributed by Les Belles-Lettres: not only is this bibliography the only one to attempt to cover the group of disciplines encompassing Classical and Late Antiquity - it is also the oldest and largest bibliography listing studies on these authors and editions of texts, and a fee is charged for access. The other resources are specialised bibliographies in the form of regularly updated databases. There are many examples<sup>4</sup>, but we may mention here:

- *Base d'Information Bibliographique en Patristique* (free access)<sup>5</sup>;
- *Bibliographie papyrologique en ligne* (free access)<sup>6</sup>;
- *L'Année épigraphique* (paywall)<sup>7</sup>;
- *Syri.ac: An annotated bibliography of Syriac resources online* (free access)<sup>8</sup>;
- *Répertoire des sources philosophiques antiques* (free access)<sup>9</sup>;
- *Droits Antiques* (free access)<sup>10</sup>.

New formats of shared bibliographical resources are springing up all the time, using software specifically designed for organising bibliographies. It is now possible, for example, to share collections and libraries in Zotero<sup>11</sup>. In addition, we are also seeing the appearance of digital libraries of ancient editions and manuscripts: each library has its own criteria for managing its collections of manuscripts and ancient editions, developing tools in-house that not only allow consultation but also, more often than not, prevent or restrict the downloading of data.

Mention should also be made of the massive but disorganised (and of very uneven quality) digitalisation work carried out by GoogleBooks / GoogleScholars, archive.org, HathiTrust (with restrictions on access from outside the US) projects, and the development of academic networks (such as academia.edu) for following up topics and researchers. These are new approaches to the publication and circulation of bibliographical data that - whatever we think of them - correspond to new uses or new expectations in academic circles.

Thus there is a host of initiatives in this field, but we are still at an experimental stage in terms of digitalising and structuring bibliographical data. There are a number of problems: firstly the

---

<sup>4</sup> The corresponding section can be viewed in [Babeu, 2011 p. 9-12].

<sup>5</sup> <http://www4.bibl.ulaval.ca/bd/bibp/recherche.html>; consulted on 23 April 2021.

<sup>6</sup> <http://www.aere-egke.be/BP/>; consulted on 23 April 2021.

<sup>7</sup> <http://www.anneepigraphique.msh-paris.fr/>; consulted on 23 April 2021.

<sup>8</sup> <http://syri.ac/>; consulted on 23 April 2021.

<sup>9</sup> <http://rspa.vjf.cnrs.fr/01/>; consulted on 23 April 2021.

<sup>10</sup> <https://www2i.misha.fr/flora/jsp/indexA.jsp>; consulted on 23 April 2021.

<sup>11</sup> For example, the Greek section at the IRHT has created a public group: [https://www.zotero.org/groups/57663/irht\\_section\\_grecque/library](https://www.zotero.org/groups/57663/irht_section_grecque/library); consulted on 23 April 2021.

large number of proprietary resources, and secondly the fact that the good quality open resources are still not well known outside digital humanities circles. Demand for the production of open data in fact involves producing data that meets not only the required technical standards but also scientific demands, always seeking to deliver data that is accessible to a wide public, whether or not its members are technologically informed. Thus it is necessary to bind together the political and scientific issue of access with the technical issues. This demands that we do not compromise on excellent scientific quality, ergonomics, ease of use, and good practices ensuring the continued existence of the tools, and it is sometimes difficult to maintain our stand.

### **1.2 The inexistent common portal for bibliography on Antiquity**

In an ideal world, we might imagine that all these databases could be compiled into one vast bibliography of Antiquity, but so far there is very little interoperability among them. The problem is firstly one of quality, because of the absence of convergence among the strictly bibliographical structuring systems, such as the Dublin Core, which is used more than extensively in describing standardised bibliographical data, and the systems that describe ancient sources which use specific and sometimes contradictory characters. True interoperability among bibliographical databases would require a standardisation not only of the choices applied in describing the bibliographical resources themselves (articles, monographs, editions) but also of their material sources (manuscripts, papyri, inscriptions) and their textual sources (standardisation of titles and authors). It is important to underline the effort made by the Project *Bibliissima*<sup>12</sup> to work on interoperability of a large variety of databases concerning history of books and libraries in the Middle Age and the Renaissance, including research tools for ancient authors and texts.

We shall see below more specifically why true operability remains on a distant horizon with regard to the Sciences of Antiquity, and how we, for our part, have broached the issue. The problem is also one of quantity. Recent years have seen an inflation in both the bibliographical mass and the number of initiatives to make use of it, an immediate consequence of the bibliometric evaluation systems set up by the academic institutions. This burgeoning of scientific literature, even though it is a sign of the vitality of the research work being carried out, eventually wilts when faced with the defective referencing system that is in use. There is no longer any single bibliographical undertaking capable of coping with the entire quantity of scientific production<sup>13</sup>. We should also add that recent years have seen a renewal of scientific approaches in the field of Antiquity, with the emergence of areas of study devoted to cultures other than those of the Greco-Roman world in its strictest sense, and this in turn has led to numerous studies on the transmission and exchange of textual, literary, technical and philosophico-religious productions. However, the bibliographical tools that include these aspects are not yet particularly developed, and researchers are often faced with a fragmentation of resources, with access to each specialised bibliography creating yet another obstacle that needs to be overcome.

### **1.3 “Back to the source”**

Rather than bemoaning the fact that we are not able to embrace the galaxy of scientific publications related to Classical Antiquity, we have, with the IPhiS-CIRIS project, opted for

---

<sup>12</sup> <https://biblissima.fr/>; consulted on 23 April 2021.

<sup>13</sup> While we cannot entirely subscribe to the political hypotheses put forward by Claude Calame in [Calame, 2006], we cannot but subscribe to his visionary description, set out as early as 1999, of the editorial frenzy that took hold of the academic world in general and the Sciences of Antiquity in particular and produced a situation in which no researcher is now able to claim to have knowledge of the entire annual production in his/her disciplinary field, however confident he/she may be.

developing a tool in phase with these new methods, in terms of both content and ergonomics<sup>14</sup>. The first stage therefore involved refining this abundance of material in order to concentrate our efforts on a homogenous part of the publications. We chose to concentrate first on the editions of texts and then on those studies whose prime object is to identify authors and establish texts and their tradition. In doing so we introduced a new methodology: we chose to consider the text as the starting point for the architecture of our bibliographical database. However, there is nothing self-evident about identifying and referencing texts: there is no fixed form of titles<sup>15</sup> in Antiquity; many texts, in fragmentary form, have come down to us without any title, and others have been reworked and the various versions need to be differentiated, while others are no longer attributed with any certainty to a particular author, or were for a certain amount of time before scientific progress challenged their attribution, and so on.

Our aim - on a small yet ambitious scale - was therefore reformulated at the start of the IPhiS project in 2014 as being: 1) to supply as complete as possible a bibliography of the editions of Greek and Latin texts from the time printing began up to the present day, including in the different states in which they are known (in their original language or in translation; in partial or complete form; treated as part of a larger set; etc.); 2) to supply this bibliography in a structured format that would allow the exchange of data with other bibliographical databases, whether specialised in any particular discipline or not; 3) to provide, as far as possible, access to the resource by indicating the link to an open-access digitalisation of the edition or, if it exists, to the manuscripts used as the basis for the edition, to make it easier for our users to find their way around the global landscape of digital resources concerning Antiquity<sup>16</sup>. The project has given rise to a working database, IPhiS, which is not open to the public, and to a web interface for published data once it has been validated by the CIRIS<sup>17</sup> team.

## II CLASSICAL BIBLIOGRAPHY AND ITS PARTICULARITIES

### 2.1 What constitutes an ancient text

Ancient texts, which form the core of our project, are above all shifting realities: between the initial form of a text published as the work of a given author and the form in which it has come down to us there are quantities of intermediate forms which we lump together in the reference title of the text<sup>18</sup>. There are any number of possible causes for an early text being deformed into the version of it we know today: the text may have been reworked by its original author and we are in possession of several concurrent versions; the author of the text may be uncertain or under debate; the text may not have reached us in its original or complete form, either by choice (quoted by another author, for example) or by accident (a fragmentary, defective or badly copied

---

<sup>14</sup> We shall not set out again in detail here in the following pages the entire genesis of the project, which is well documented already. It was presented in detail at the Digit-Hum workshop in 2020 (<https://hdl.handle.net/10670/1.dpf0n>; consulted on 23 April 2021). We shall merely recall in this footnote all the active members of the project, all of whom contributed to the results presented in this text: Laurent Capron, Julie Giovacchini, Sébastien Grignon and Juliette Lemaire of the Centre Jean Pépin, and Bernard Weiss of the ARDIS unit (UPS2259) for implementing the computerised development of the various models.

<sup>15</sup> See our description of the data model selected and choice of text as the principal node in [Giovacchini et al., 2017 p. 6].

<sup>16</sup> It has to be underlined that, although our goal is to provide bibliographical data about editions of texts, we also try to give a direct link to open access databases of texts like Perseus (<http://www.perseus.tufts.edu/hopper/collection?collection=Perseus:collection:Greco-Roman>; consulted on 23 April 2021).

<sup>17</sup> <https://ciris.huma-num.fr/> consulted on 23 April 2021. The public interface is available in four languages (French, English, German and Latin). Since the data are bibliographical data, they cannot be translated and remain in their original language. The content block reproduces most of the time the table of contents of the book. In some cases, we provide some comments about the edition; these comments are always in French.

<sup>18</sup> See for example [Giovacchini, 2016 1 and 2].

manuscript), etc. However, all these forms refer to a single ideal conceptual philological object, which contains in it all the forms it has already been able to take and could still take in the future. The task of referencing these objects is ongoing, and no catalogue today can claim to be exhaustive or complete. Developments in the field of research also mean that this referencing cannot, by its very nature, be definitive, even if we seek to establish sustainable data: anyone could be caught out by an unexpected discovery that challenges an attribution or a title, or by the rediscovery of a text for which we thought we had reliable, constant information.

## 2.2 The problem of title

The question of the title of the ancient text is in itself a problem: in our efforts to draw up a list of titles we had to consider a number of realities that corresponded to the expression ‘title’. Some texts - such as plays and novels - really do have a title. It is less evident when we are dealing with scientific, philosophical or religious treatises, since they would be qualified as ‘(treatise) on a particular subject’, sometimes with a sub-title. This is then an explanatory title, but its form is not fixed and it may evolve, in copies that are made of it, by abbreviation, or by extension.

Other texts have no other title than that given by a subsequent publisher according to their genre (‘oration’, ‘speech’, ‘homily’, etc.): these texts are often passed down to us in a corpus, within which they are numbered; this order becomes ‘canonical’ until new manuscripts are discovered, challenging the order by adding new texts or removing others, and we are then faced with duplicate or triplicate numberings. Most of the poetry that comes down to us from Antiquity is only known in anthologies and collections within which the poems are numbered, but the same poem, since it is included in several collections, is given several identifiers: we then choose to consider the compilation as the reference text, and the poem as a constituent part rather than as a text in its own right.

Lastly, an incalculable number of texts have come down to us with no title at all, in isolation, sometimes only as a fragment: if their content is deemed important, they are given a title that quickly becomes the reference title. However, some of these titles are no more than designations of the accident as a result of which the text has come down to us: for example, what is designated as the ‘Palatine Anthology’ is a collection of poems which has come down to us via a manuscript found in the Palatine Library in Heidelberg<sup>19</sup>. We have sometimes chosen to propose a new, more meaningful title for a text which has a relatively meaningless title<sup>20</sup>. It is thus this multiplicity of situations that the field “title” covers. In our database we have made room for ‘aliases’ that retain the history of the titles, when they vary from one publisher to another, or from one piece of manuscript evidence to another, or simply because we give Greek texts a working title in Latin even though we are fully aware that the original title is the Greek one.

Since we always provide for each text a cross-reference towards other repertories of texts, we often notice that these repertories don’t use the same exact title for the same text. If there is no reason to reject all of these variant titles, we tend to keep the title given in the *Thesaurus Linguae Graecae Canon*<sup>21</sup> of texts or in the *Library of Latin Texts*<sup>22</sup> for classical texts, and the title given in the *Clavis Clavium*<sup>23</sup> for Christian texts.

---

<sup>19</sup> [Capron, 2016a]

<sup>20</sup> [Capron, 2016b]

<sup>21</sup> <http://stephanus.tlg.uci.edu/canon.php>; consulted on 23 April 2021.

<sup>22</sup> <http://clt.brepolis.net/llta/public/About.aspx?what=all>; consulted on 23 April 2021.

<sup>23</sup> <https://clavis.brepols.net/clacla/Default.aspx>; consulted on 23 April 2021.. This database is a common portal for the previously only printed catalogues of texts *Clavis Patrum Graecorum*, *Clavis Patrum Latinorum*, *Bibliotheca Hagiographica Graeca* and *Bibliotheca Hagiographica Latina*: it is not a unified repository and every text keeps its former catalogue number.

## 2.3 Linking a text to an author

Another difficulty with referencing is that of creating a link between a text and an author. Two specific situations had to be taken into account in our bibliographical database: texts with no identified author, and texts whose presumed author is or has been disputed. In both cases, we must report the lack of precision or the ambiguity in our knowledge, without hiding anything.

### 2.3.1 Texts with no author

This first case appears simple, since it is possible to attribute a text with no author to an ‘anonymous’ author. However, the entry ‘anonymous’ in the table of authors would refer to just one person whose name has been lost! Our approach has therefore been to use collective headings in the ‘creator author’ field, making it possible to classify our texts in categories without concealing the loss of the name of their author, and providing the possibility of finding the text by means of a themed search. The headings and number of these collective entries are left completely up to us, but there is a tendency in academic circles to adopt common practices, and so we have tried to harmonise our headings with those used in other databases, particularly the Pinakes database<sup>24</sup> developed by the Greek section at the IRHT (CNRS - UPR 841)<sup>25</sup>. Thus these collective headings function as “supra-authors” and it would have been perfectly logical for us to attribute to each text, whether its author was known or not, a collective heading in addition to its possible author, in the same way as one would use markers or labels when organising index cards. We decided to only use this option in the case of unknown authors or for certain large-scale groupings in mainly early editions. Initially, we chose to use these names only in the case of texts without a known author, that is to say as an alternative generic author name. But we do not rule out modifying this parameter later and systematically assigning each text one or more of these names to allow a thematic search in the database.

### 2.3.2 Disputed attribution

The other situation is when texts have come down to us under the name of an author that subsequently proves to be incorrect, either specifically and deliberately – for example the many cases of religious authors whose work was banned but some of their texts continued to be copied under the name of another author who was authorised –, or as the result of an error in transmission caused by homonymy or confusion. Sometimes modern publishers of ancient texts took up such problems of attribution a while ago, and in some cases they have come up with a

Oeuvres :	titre	langue	époque
	<a href="#">De natura</a>	grec	vers 306-295 av. J.-C.
	<a href="#">Epistula ad Herodotum</a>	grec	4 <sup>er</sup> s. av. J.-C.
	<a href="#">Epistula ad Menoeceum</a>	grec	4e-3e s. av. J.-C.
	<a href="#">Epistula ad Pythoclem</a>	grec	4e-3e s. av. J.-C.
	<a href="#">Fragmentum De pietate</a>		vers 2e s. av. J.-C.
	<a href="#">Gnomologium Vaticanum Epicureum</a>	grec	4e-3e s. av. J.-C.
	<a href="#">Ratae sententiae</a>	grec	4e-3e s. av. J.-C.

Figure 1 : Reporting disputed attribution to an author (here, Epicurus).  
(<https://ciris.huma-num.fr/noticeauthor.php?langue=fr&id=785>, consulted on 21 April 2021).

<sup>24</sup> <https://pinakes.irht.cnrs.fr/>; consulted on 21 April 2021.

<sup>25</sup> List of our headings in the internal documentation for the database, produced by Juliette Lemaire: available at <https://ciris.huma-num.fr/manuel/doku.php?id=manuelciris:anonymes&do=login>; consulted on 21 April 2021. Collective headings in Pinakes are listed at: <https://pinakes.irht.cnrs.fr/mode-d-emploi.html>; consulted on 21 April 2021.

solution. Thus some texts no longer circulate under the wrong author’s name, and once the controversy has been deemed closed for several decades, we make no mention of this other than in a comment on the text. But much doubt still remains about many texts, and we felt it was necessary to report on this as a stage in academic research.

We therefore opted for creating the notion of ‘disputed attribution’ by mentioning specifically the bibliography involved in the controversy. It is therefore possible, in our database, to list a text under several authors and to mention which authors are accepted or rejected by modern publishers, with the corresponding bibliography. Apart from the fact that this provides additional access to the bibliographical data concerned, our solution makes use of the state of research and the importance of considering the uncertainty that sometimes surrounds the texts that have come down to us from Antiquity, by presenting such texts to the user on both the page for the author (Figure 1) and the page for the corresponding text (Figure 2).

### Fragmentum De pietate

Titre Fragmentum De pietate  
 Auteur [Epicurus](#), [Philosophica](#)  
 Langue  
 Époque vers 2e s. av. J.-C.  
 Mot-clés épicurisme, théologie

#### Références

Commentaire Fragment d'un texte d'inspiration nettement épicurienne traitant de la piété et du culte envers les dieux. Retrouvé à Oxyrhynchos en 1899. L'attribution du texte est controversée.  
 Controverse Attribution à [Epicurus](#) contestée

#### Editions :

titre	périodique	maison d'édition	année	premier contributeur
<a href="#">Textprobleme eines epikureischen Fragments über Götterverehrung (P. Oxy. 215)</a>	Rheinisches Museum für Philologie		1962	Wolfgang Schmid <a href="#">BNF - 127572</a>
<a href="#">Les Épicuriens</a>		Gallimard	2010	Daniel Delattre <a href="#">BNF - 121296701</a>

Figure 2: Reporting disputed attribution for a text.

(<https://ciris.huma-num.fr/noticetext.php?langue=fr&id=382>, consulted on 21 April 2021).

## 2.4 Geographical data

It is usual to designate certain Ancient authors by adding to their name that of a town or country to which they are attached by birth or the place of their main activity, particularly in cases of homonymy. Wherever possible we have preferred sequencing these two items of information, the one prosopographic and the other toponymic. However, this obviously supposes that the geographical data has been checked and cross-referenced against the historical geography repositories. Where this was known, we linked each author to one or more places, possibly specifying the event associated with the place (birth, education, activity, death). Here again we made it possible to weight the reliability of the historical data by adding an indication of whether the data is certain or disputed.

This specific data is displayed in an extension of the CIRIS application in a cartographic format<sup>26</sup> that we felt was more pertinent. For the time being, only the geographical data of the philosophical authors are being published; ultimately we hope to be able to provide this data for all the Ancient authors in the database<sup>27</sup>.

<sup>26</sup> <https://ciris.huma-num.fr/cartographie.php?langue=fr>; consulted on 23 April 2021. See also Fig. 12 supra. The specific publication of this geographical data has constituted an appendix to the project which we present in the report [Giovacchini et al., 2020].

<sup>27</sup> We have chosen to display our biographic data in a modern map, with contemporaneous place names. This choice is a pedagogical one, and is directly focused on the accessibility of our tool for non-scholar users and particularly high school students.



## 2.5 A bibliography in its environment

For the files of both titles and authors, we rely systematically whenever possible on repositories that have already been checked and found reliable, to which we direct users (Pinakes, VIAF, DataBNF, etc), but if appropriate we also propose our own repository, which has our own structuring. Once we had opted for acknowledging the rightful place of uncertain data on bibliographical and philological objects with rather complex contours, presenting a classification meant both respecting the standards without which it is impossible to share data as well as making decisions reaching well beyond straightforward indexing or alignment, and therefore supposed a degree of scientific risk-taking.

In the history of the IPhiS-CIRIS project, accessibility has played a structuring role. Initially, the project was constructed on the basis of an ambition to offer full open access, in reaction to the idiosyncratic and extremely constricting proprietary model of the *Année Philologique*<sup>28</sup>. Thus the notion of open access was directly correlated to the question of the structuring and display of the data. Opening up data is not simply the unequivocal gesture of lifting a barrier: rather it means from the outset thinking of the data as an element that may be shared or potentially improved or altered within a complex environment - what is often nowadays called a 'data ecosystem'. Which is tantamount to saying that accessibility is necessarily free of charge, but that is not its only feature. The project therefore exploits a dimension of structuring that is in fact already historically consubstantial with the Classical sources. Since the time of the very first publishers, Greco-Latin literary texts - because of their often incomplete or fragmentary state and the existence of successive versions that form overlays rather than replacements - possess systems of unique identifiers of varying complexity.

*Many non-classicists from academia and beyond still express surprise that classicists have been aggressively integrating computerized tools into their field for a generation. The study of Greco-Roman antiquity is, however, a data-intensive enterprise. Classicists have for thousands of years been developing lexica, encyclopedias, commentaries, critical editions, and other elements of scholarly infrastructure that are best suited to an electronic environment. Classicists have placed great emphasis on systematic knowledge management and engineering. The adoption of electronic methods thus reflects a very old impulse within the field of classics. The paper knowledge based on Greco-Roman antiquity is immense and well organized; classicists, for example, established standard, persistent citations schemes for most major authors, thus allowing us to convert nineteenth-century reference tools into useful electronic databases. Classicists are thus well prepared to exploit emerging digital systems. For many classicists, electronic media are interesting not (only) because they are new and exciting but because they allow us to pursue more effectively intellectual avenues than had been feasible with paper<sup>29</sup>.*

Since the time of the monuments of German erudition in the eighteenth and nineteenth centuries, Classicists have always considered manipulation of the corpora as constituting a consultation, far beyond cursive continuous reading. Consultation supposes transversality and navigation; the corpora are explored and refer back to each other in an inchoative method that can never end. This mode of appropriation is necessary in as much as ancient texts are not units that are closed in on themselves but rather are open samples that are always subject to revision and rereading and above all that serve as sources for other texts. By its very nature, one Greco-Latin literary text refers to other texts in the present state of transmission: one text transmits other texts and is itself transmitted by others. (By 'source' we mean here the material source as

---

<sup>28</sup> The general policy of CIRIS is readable online (<https://ciris.huma-num.fr/mentionslegales.php?langue=fr>; consulted on 23 April 2021) ; the model is one of a CC-BY-SA license.

<sup>29</sup> [Crane, 2004]

well as the historical and literary source, both meanings mingling fairly imperceptibly in the case of texts of the Greco-Latin Antiquity period<sup>30</sup>.)

Thus the very first step towards accessibility consists, as we have seen, of modelling this network of inter-generating texts, and making the text as a source the pivotal point or central table of the database. The second stage supposes adding to this central table a sufficient number of controlled identifiers to allow unambiguous navigation both inside and outside the database. These two gestures suppose a distancing from the usual structuring of a bibliography: departing from the traditional documentary model which constructs the reference around the modern edition and is not affected by fine granularity (fragment, extract, etc), considering the modern edition as the culminating point of a long historical process, one state of the text among others, neither the latest nor the ‘best’, and focusing attention on the source text identified as a sort of bibliographical invariable in its ideal form, whose actual occurrences are so many versions. This involves no more or less than applying the Lachmannian notion of archetype to the bibliography, albeit by shedding any naively realist posture<sup>31</sup>. It is supposed that a text was produced in a given period by a given author on a given subject; if the text in its empiric form is no longer accessible today, it has produced a certain number of historical avatars which all have in common a desire to be reproductions or indirect representations of the text, which must then be taken seriously as the focus, *raison d’être* and true purpose of these avatars. This is truly a matter of accessibility, since it is the only modelling that actually meets the expectations of the users of a bibliographical database; these users are primarily readers, as they read and carry out research on texts, but they are also authors who produce texts themselves. This departure from the usual bibliographical frame is also an interesting way of converging towards other schemes that are less strictly disciplinary and more generalist, with a view to strengthening the navigability of the data.

### III TOWARDS GLOBAL INTEROPERABILITY: OPENING UP DATA

#### 3.1 Semanticizing IPhIS

In 2019, the IPhIS database and its public web version (CIRIS) migrated from a local server on the Villejuif site to Huma-Num’s Very Large Research Infrastructure (TGIR). The conditions for this migration were excellent, offering guarantees of incomparable permanence compared with the constraints of local hosting. The operation, although highly beneficial for the project, was dependent on one medium-term condition: it had to be possible for the Isidore<sup>32</sup> search engine to harvest the essential CIRIS data. Since Isidore functions according to the principles of the web of linked data, we had to ‘translate’ our traditional relational database in such a way as to make it compatible with a number of semantic constraints, which meant converting the main IPhIS data into searchable resources meeting the semantic standards accepted by Isidore<sup>33</sup>.

---

<sup>30</sup> This is a problem that justifies the title of our job at the CNRS: the analysis of source, many examples of which are to be found in our team’s research logbook; see for example [Capron, 2016c] and [Grignon, 2016.].

<sup>31</sup> An expression of this type of realism forms a common theme running through recent experiments carried out on citing Ancient sources; all the models proposed are based on historical breakdowns and undisputed nomenclatures that are deemed historical fact, whereas they are actually no more than reconstructions. Thus the most advanced tool available to date in terms of citation, the CapiTainS environment (presented for example in [Clérice, 2017]), proposes alignments that are technically extremely advanced, constructed on the basis of the CTS (Canonical Text Services) model, which in turn uses a text identification system that does not dispute the pertinence of the titles being aligned. Although in theory this system is indeed able to allow title variants to be taken into account by distinguishing works from versions of works, in fact it only does so marginally, as it excludes for example the possibility of fuzzy matches, variable divisions, and intertextual conversion.

<sup>32</sup> <https://isidore.science/>; consulted on 23 April 2021.

<sup>33</sup> There is nothing original in this process of semanticisation on our part. Indeed it is a process that has become banal: items of digital data are adapted to the constraints of the data web so that they will comply with the

Isidore is a harvester and a tool of enrichment, but to be able to harvest and enrich it first needs to be able to recognise, identify and link. It is therefore necessary to propose data set out in a format it is able to interpret, and for the actual items of data to be described using metadata Isidore recognises. The Isidore documentation offers two options for preparing databases:

- Propose data using an XML flow of standardised metadata using the OAI-PMH protocol associated with metadata in Dublin Core format. This method is suitable for use with documentary databases, corpora, scientific archives and document/data libraries. For example, a tool such as Omeka offers OAI-PMH via a module.<sup>34</sup>
- Propose data using an XML sitemap flow pointing to webpages containing RDFa metadata. This method is suitable for research programme websites presenting corpora of documents or data, scientific blogs (but not Hypotheses.org), and webpages in general.<sup>35</sup>

We<sup>36</sup> selected the first option, specifically adapted to our case, and decided to do so by making use of another tool proposed by the Huma-Num grid of services: a Nakala<sup>37</sup> depository. This solution supplied us with an OAI-PMH depository generated automatically by our posting in Nakala without having to incorporate the data directly in a database that was already substantial enough; it also enabled us to select more closely which sets of data were most pertinent for harvesting purposes. This option nevertheless raised other problems, and not all of them have as yet been completely resolved. The principle of a Nakala depository is relatively simple: it is an autonomous uploading area in which items of data are associated either manually or using an API with metadata standardised according to the grids of Dublin Core vocabulary, which is currently the most frequently used model for producing bibliographical metadata<sup>38</sup>. The first question that had to be asked was therefore whether the fields in the various IPhis tables matched those of the Dublin Core vocabulary, with a view to being able to establish an equivalence between the two, this being a prerequisite for the possibility of interoperability between IPhis and Nakala. If we compare the IPhis data model with the Dublin Core vocabulary, we can see straight away that there are considerable convergences - which is perfectly natural and to be expected of a bibliographical database. The most important of these convergences is the adaptation in IPhis of an important distinction, namely differentiating between the creator of a resource and a contributor to a resource. It is a particular feature of ancient texts that their transmission involves a number of participants who, although they are not all authors strictly speaking, have nevertheless made a not inconsiderable contribution to the elaboration of a given text up to its final state: these participants include publishers, printers, translators and commentators. The distinction between creator and contributor was from the outset an elegant solution in the data model for differentiating sufficiently clearly the various stages in the transmission of the text; we preferred this to the more traditional distinction between ancient and modern authors which is based on a vague and relatively inoperative temporal difference as we wish to distinguish not between periods but between interventions

---

requirements of search engines, in a frame that is often institutional. We may note, for example, that the tool SKOS Play allowing the conversion of data into RDF/SKOS from an Excel spreadsheet was developed with funding from the Luxembourg State in order to produce public data in a semantic format (cf. [Francart, 2017]).

<sup>34</sup> <http://info.omeka.net/build-a-website/manage-plugins/oai-pmh-repository/>; consulted on 23 April 2021.

<sup>35</sup> <https://documentation.huma-num.fr/isidore/>; consulted on 23 April 2021.

<sup>36</sup> The development and implementation work described in this section was carried out mainly by Bernard Weiss and Julie Giovacchini (ontology and mapping with the DCMI by Julie Giovacchini, scripts for automated supply to the Nakala API by Bernard Weiss); the tests were devised jointly by both these researchers.

<sup>37</sup> <https://www.nakala.fr/> consulted on 23 April 2021.

<sup>38</sup> This vocabulary is described and documented on the website of the Dublin Core Metadata Initiative: <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>; consulted on 23 April 2021.

on the texts. Thus the IPhIS data model allows the entry of an ‘ancient’ author as a contributor to a text on a par with a humanist or contemporary publisher.

### 3.2 Using Nakala to semanticize a database: limits and advantages

#### 3.2.1 Adapting to Nakala

For the actual uploading in Nakala, we had to overcome a number of relatively substantial technical constraints, which obliged us to delimit very clearly both what we wanted to display and our choice of metadata to accompany the display.

In the first place, Nakala is mainly intended to be used for uploading data in the form of files attached to metadata, and the main working interface is a visual interface that only allows the individual uploading of file after file. But our situation is very different: we want to post large sets of data in an automated way, by supplying the depository directly so that it updates itself at the same time as the database, which is designed to continue to receive data indefinitely. We therefore concentrated firstly on uploading the records of editions in the presentation proposed by CIRIS, as these constitute the most important and complete IPhIS data, leaving aside temporarily the direct display of the thesauruses of Ancient authors and texts. Our Nakala depository is envisaged as a place for displaying edition records which are considered as so many separate files, each comprising an image (PDF or screen capture of the page of the record in CIRIS) and a set of metadata (Figures 3 and 4<sup>39</sup>).

The screenshot displays a web interface for a file viewer. On the left, under 'Fichier', there is a file named 'CIRIS - site public.pdf'. The main area, 'Visualisation', shows navigation controls: 'Précédent', 'Suivant', and 'Télécharger' buttons. It also indicates 'Page: 1 / 1' and an 'Aller à la page' button. The URL is 'https://ciris.huma-num.fr/noticedition.php?langue=fr&id=45&export=print'. The record details are as follows:

date	ciris - 09/04/2021
titre	Irénée de Lyon - Contre les hérésies. Livre IV
contributeurs	Louis Doutreleau - éd. (BNF - 119005367) , Charles Mercier - éd. (BNF - cb130915660) , Bertrand Hemmerding - éd. (BNF - cb12019956x) , Adelin Rousseau - éd. et trad. (BNF - 11922866w)
maison d'édition	Éditions du Cerf
année	1965
volume	100
collection	Sources chrétiennes
pagination	2 vol.
type	Edition
ISSN	2-704-08123-X

At the bottom, the ID is: 10.34847/nkl.d08agei7/34b631e9ce6374762f2046246f1f463258faa9b4

Figure 3 : Example of file - CIRIS record 45

<sup>39</sup> The screen captures were generated at the time of the stage in the test of the procedure described in these pages, as the Nakala depository whose development we describe is not yet public.

Type	Image	
Description	Français	Edition -
Http://purl.org/dc/terms/publisher	Français	Éditions du Cerf
Http://purl.org/dc/terms/contributor	Français	Louis Doutreleau - éd.
Http://purl.org/dc/terms/date	Français	1965
Description	Français	Contient : t. 1 : introduction, notes justificatives, tables ; t. 2 : texte latin, traduction française en regard et fragments grecs en pied de page. Utilise les versions latine et arménienne du texte. Le texte latin a été âprement débattu entre les éditeurs - voir à ce propos l'article de B. Hemmerdinger (The Journal of Theological Studies, 1966, p. 308-326).
Http://purl.org/dc/terms/isversionof	Français	Aduersus haereses - Έλεγχος καὶ ἀνατροπή τῆς ψευδωνύμου γνώσεως - latin - 2e s. - texte complet variante - CPG: 1306

Figure 4 : Example of metadata - CIRIS record 45

Also, in order to automate the process and upload large quantities, we had to access Nakala via its API<sup>40</sup>. Using a script, we recuperated the necessary files and metadata in IPhis and then uploaded them in the API by using a ‘post’ request. This is the only way the depository can be updated automatically. The script, first developed in PHP, is still experimental, but once it has been stabilised, the next stage of the work will consist of proposing an open version in PHP and Python that can be adapted for use in other projects similar to ours. The Nakala API requires a particular formalism for the presentation of metadata, requiring us to adapt our requests to convert the fields in the IPhis database in such a way as to produce metadata in JSON format, which is compatible with Nakala. This part of the work was relatively delicate and called for a number of arbitrations that were at times somewhat frustrating. For a very straightforward record, limited for example to very little metadata in addition to the compulsory metadata, we arrive at the following form:

<sup>40</sup> <https://api.nakala.fr/doc>; consulted on 23 April 2021.

```

{
  "status": "published",
  "metas": [{
    "value": "Ermarco - Frammenti",
    "lang": "it",
    "typeUri": "http://www.w3.org/2001/XMLSchema#string",
    "propertyUri": http://nakala.fr/terms#title
  },{
    "value": {
      "givenname": "Hermarchus",
      "surname": "Philosophus",
      "orcid": null
    },
    "propertyUri": http://nakala.fr/terms#creator
  },{
    "value": null,
    "lang": null,
    "propertyUri": http://nakala.fr/terms#created
  },{
    "value": "CC-BY-4.0",
    "lang": null,
    "typeUri": "http://www.w3.org/2001/XMLSchema#string",
    "propertyUri": http://nakala.fr/terms#license
  },{
    "value": "http://purl.org/coar/resource_type/c_c513",
    "lang": null,
    "typeUri": "http://www.w3.org/2001/XMLSchema#anyURI",
    "propertyUri": http://nakala.fr/terms#type
  },{
    "value": "Francesca Longo Auricchio - éd. (VIAF - 7280456)",
    "lang": "it",
    "typeUri": "http://www.w3.org/2001/XMLSchema#string",
    "propertyUri": http://purl.org/dc/terms/contributor
  },{
    "value": "https://ciris.huma-num.fr/noticeedition.php?langue=fr&id=54",
    "lang": "fr",
    "typeUri": "http://www.w3.org/2001/XMLSchema#anyURI",
    "propertyUri": http://purl.org/dc/terms/relation
  }
  ]
}

```

Injecting this form into the API via a ‘post’ request produces the following display (Figure 5):

Ermarco - Frammenti		
<b>Titre</b>	Italien	Ermarco - Frammenti
<b>Auteur</b>		Hermarchus Philosophus
<b>Date de création</b>		Inconnue
<b>Licence</b>		Creative Commons Attribution 4.0 International (CC-BY-4.0)
<b>Type</b>		Image
<b>Http://purl.org/dc/terms/contributor</b>	Italien	Francesca Longo Auricchio - éd. (VIAF - 7280456)
<b>Http://purl.org/dc/terms/relation</b>	Français	<a href="https://ciris.huma-num.fr/noticedition.php?langue=fr&amp;id=54">https://ciris.huma-num.fr/noticedition.php?langue=fr&amp;id=54</a>

Figure 5: Example of metadata - CIRIS record 54.

Nakala relies mainly on two controlled vocabularies, Dublin Core and FOAF, and adds five compulsory metadata (type, license, creator, created, title) in its own namespace. As indicated above, it is theoretically possible to translate all the metadata for our bibliographic records using qualified Dublin Core – including relationships between texts, external sources and references, using the notion of relationship. Our initial mapping was organised as follows:

#### **IPhiS metadata**

title  
sub-title  
publisher  
type  
comment  
ISBN  
language  
year  
abstract  
key word  
ancient text  
source  
ancient author  
contributing author and his/her function  
external reference / access  
relationship between different editions  
periodical

#### **Qualified Dublin Core equivalent**

dc:title  
dc:title  
dc:publisher  
dc:description  
dc:description  
dc:identifier  
dc:language  
dc:date  
dc:abstract  
dc:subject  
dcterms:isVersionOf  
dc:source  
dc:creator  
dc:contributor  
dcterms:hasFormat  
dcterms:isReferencedBy  
dcterms:isPartOf

In practice we nevertheless came up against certain consistency issues, connected with the different intents of our database and the Nakala environment, and we had to partly adapt this mapping.

Thus the compulsory metadata in Nakala includes the notions of *creator* and *created* in respect of the date of creation. But these notions, in Nakala's internal logic as expressed in the API, apply not to the content of the data but to the data itself: the creator is the person who posted the data, and the date is the date of posting. If we maintain this idea in our display, the metadata then proves to be very poor, and harvesting is likely to prove of little pertinence, even though it is always possible to add further layers of metadata subsequently. For our type of bibliographical data, creating the possibility of carrying out a search on the basis of the name of the person uploading an item of data merely results in the creation of unnecessary noise and is probably a nuisance for users of both Nakala and Isidore.

We therefore decided to circumvent this constraint and indicate as creator the creator of the intellectual content – so, as far as we are concerned, still a creator author, but an ancient author. However, in this case it is impossible to associate a creation date with this creator. This is because the date format for an Ancient work cannot be expressed as D/M/YYYY as required by Nakala's API. At the time of posting we therefore systematically need to associate a zero value to the compulsory *created* item of metadata. These constraints are offset by the possibility of incorporating a clickable link in the metadata items leading directly to the record in the database. Thus whether by visiting the depository directly or as a result of harvesting by Isidore, the user will always have the possibility of rapid access to the full record in its original place of publication, as the last metadata item visible in Figure 5 shows.

### 3.2.2. *What we display in Nakala, and why*

This obviously raises the question of the necessary redundancy of accesses to an item of digital data. The purpose of conversion is to produce a display which will, we hope, increase the visibility of the database – but which cannot take the place of direct consultation of the database if full information is required. Therein lies all the ambiguity of the application of semantic technologies to digital objects that are already natively relational: the overlay of relational logic, thought out as navigation or an explicit path from one resource to another, by an object logic in which the fine description provides the key but does not set out the entire path; thus the description increases the visibility of the resource in certain search tools, but does not take the place of direct exploration of the resource itself. In other words, full display in Nakala would require the construction of a semantic clone of our database, and that would be pointless. Nakala is therefore used like a magnifying glass, with the intention of making it possible for Isidore to harvest certain strategic metadata in order to increase *in fine* the flow of visits not to the Nakala depository but to the database itself. It is in this logic that the date of creation of the data or its creator becomes a useless item of information, which must be replaced by the creator of the intellectual content likely to be of interest to the user.



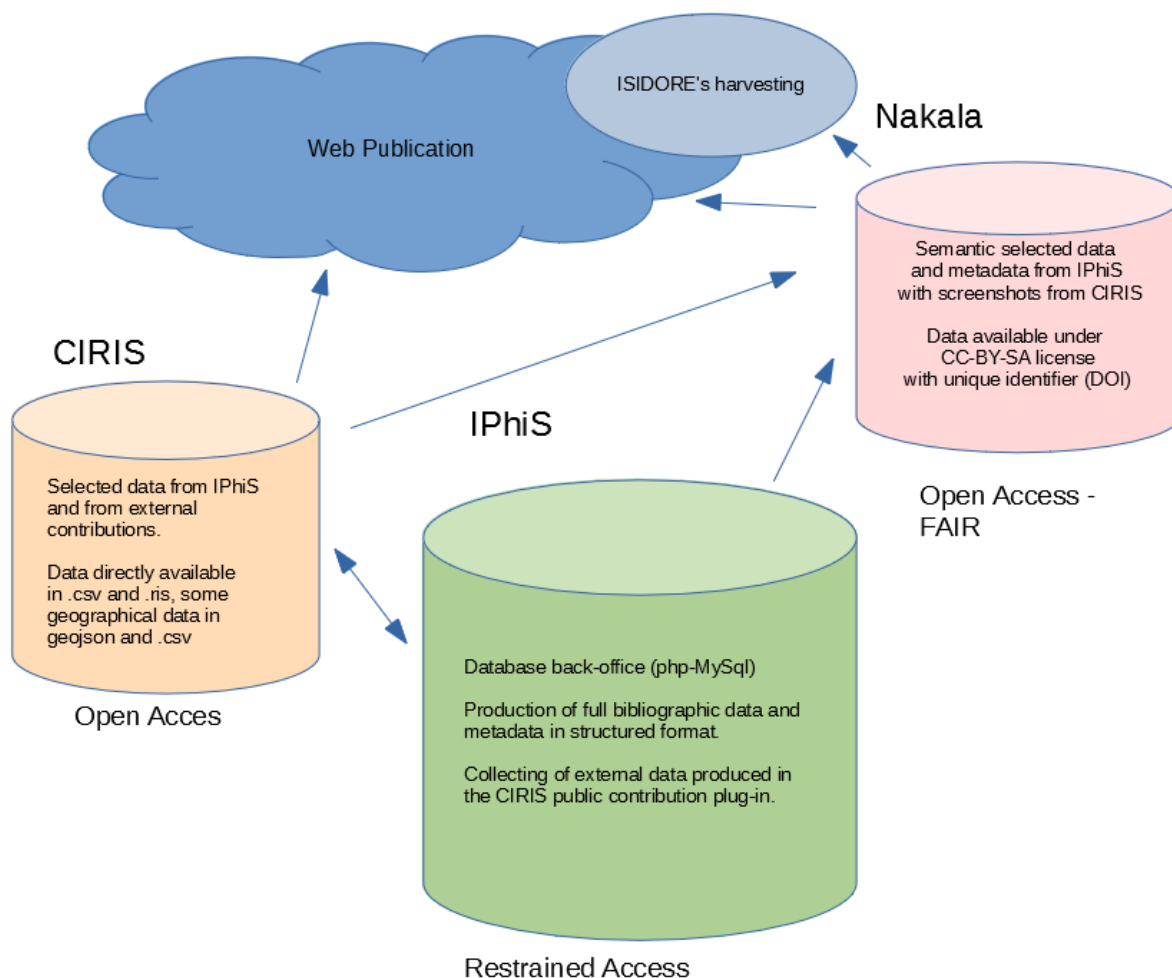


Figure 6: Complete process of publication and display.

These choices, summed up in Figure 6, which are still being implemented, are the final stage in a lengthy thinking process; before accepting these compromises it was necessary to start by exploring the hypothesis of systematicity. Would we have been able to display the complete set of IPHiS data directly in Nakala if we had wanted to? The answer is no, for the time being, because firstly if we had wanted to display not only the records for editions but also all the connected data in the IPHiS thesaurus, we would have fairly quickly reached the limit of the vocabularies accepted by the API, and secondly Nakala does not natively recognise the SKOS vocabulary. If we compare for example the IPHiS data and a set that is close not in terms of quantity but of type of object listed - the BNF data - we can see fairly well how DataBNF compensates for the shortcomings of qualified Dublin Core and FOAF with on the one hand certain elements of SKOS for the data in its thesauruses and on the other a specific ontology describing relations and objects specific to its resources and its display<sup>41</sup>. Similarly, a full IPHiS semantic display would suppose the creation of a specific ontology to express certain traits which, as far as we know, are not currently taken into account by the standard bibliographical

<sup>41</sup> Cf. the full outline of DataBNF at [https://data.bnf.fr/images/modele\\_donnees\\_2018\\_02.pdf](https://data.bnf.fr/images/modele_donnees_2018_02.pdf); consulted on 23 April 2021.

ontologies. This is more particularly the case for the matter of incomplete texts, noted as fragments or pieces, or collections of texts, anthologies or corpora; it is also, and perhaps even more so, the case for describing the complex progression in order to distinguish between a text's creator, copyist, translator and publisher throughout its history. It is possible to construct this ontology, and we are able to propose an overview of it here, produced using Protégé software<sup>42</sup> (Figures 7 to 11).

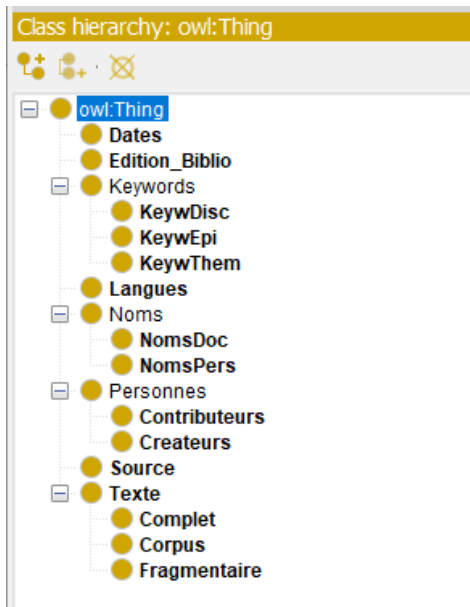


Figure 7: Hierarchy of classes in IPhiS ontology in Protégé.

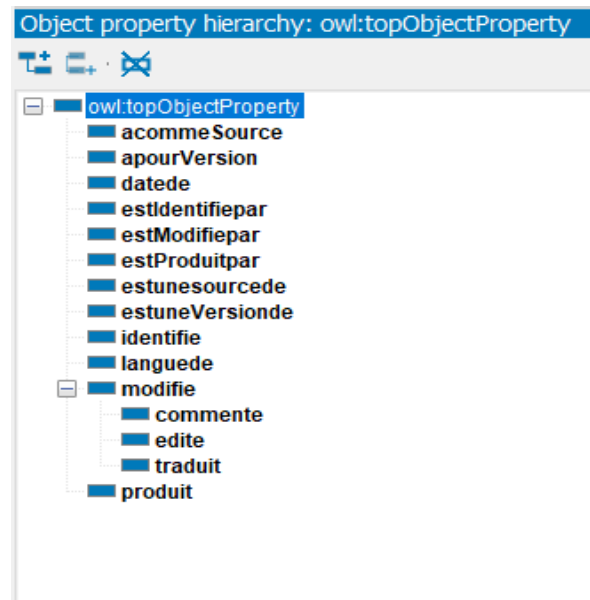


Figure 8: Hierarchy of properties of IPhiS ontology in Protégé.

<sup>42</sup> See <http://protegeproject.github.io/protege/> for installing and obtaining up-to-date documentation on Protégé, the open-source software developed by Stanford University; consulted on 23 April 2021.

```

542 <owl:Class rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Texte">
543   <rdfs:subClassOf>
544     <owl:Restriction>
545       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#apourVersion"/>
546       <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Edition_Biblio"/>
547     </owl:Restriction>
548   </rdfs:subClassOf>
549   <rdfs:subClassOf>
550     <owl:Restriction>
551       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#estIdentifiepar"/>
552       <owl:someValuesFrom>
553         <owl:Class>
554           <owl:intersectionOf rdf:parseType="Collection">
555             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Dates"/>
556             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#KeywThem"/>
557             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Langues"/>
558             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#NomsDoc"/>
559           </owl:intersectionOf>
560         </owl:Class>
561       </owl:someValuesFrom>
562     </owl:Restriction>
563   </rdfs:subClassOf>
564   <rdfs:subClassOf>
565     <owl:Restriction>
566       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#estModifiepar"/>
567       <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Contributeurs"/>
568     </owl:Restriction>
569   </rdfs:subClassOf>
570   <rdfs:subClassOf>
571     <owl:Restriction>
572       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#estProduitpar"/>
573       <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Createurs"/>
574     </owl:Restriction>
575   </rdfs:subClassOf>
576 </owl:Class>
577
578

```

Figure 9: IPhIS ontology, extracted from the description in RDF: semantics of Texts.

```

269 <!-- http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Edition_Biblio -->
270
271 <owl:Class rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Edition_Biblio">
272   <rdfs:subClassOf>
273     <owl:Restriction>
274       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#estIdentifiepar"/>
275       <owl:someValuesFrom>
276         <owl:Class>
277           <owl:intersectionOf rdf:parseType="Collection">
278             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Dates"/>
279             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#KeywThem"/>
280             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Langues"/>
281             <rdf:Description rdf:about="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#NomsDoc"/>
282           </owl:intersectionOf>
283         </owl:Class>
284       </owl:someValuesFrom>
285     </owl:Restriction>
286   </rdfs:subClassOf>
287   <rdfs:subClassOf>
288     <owl:Restriction>
289       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#estuneVersionde"/>
290       <owl:someValuesFrom rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Texte"/>
291     </owl:Restriction>
292   </rdfs:subClassOf>
293   <rdfs:subClassOf>
294     <owl:Restriction>
295       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#acommeSource"/>
296       <owl:allValuesFrom rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Source"/>
297     </owl:Restriction>
298   </rdfs:subClassOf>
299   <rdfs:subClassOf>
300     <owl:Restriction>
301       <owl:onProperty rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#estProduitpar"/>
302       <owl:allValuesFrom rdf:resource="http://www.semanticweb.org/giova/ontologies/2021/2/ontoIPHiS#Contributeurs"/>
303     </owl:Restriction>
304   </rdfs:subClassOf>
305 </owl:Class>

```

Figure 10: IPhIS ontology, extracted from the description in RDF: semantics of Editions.

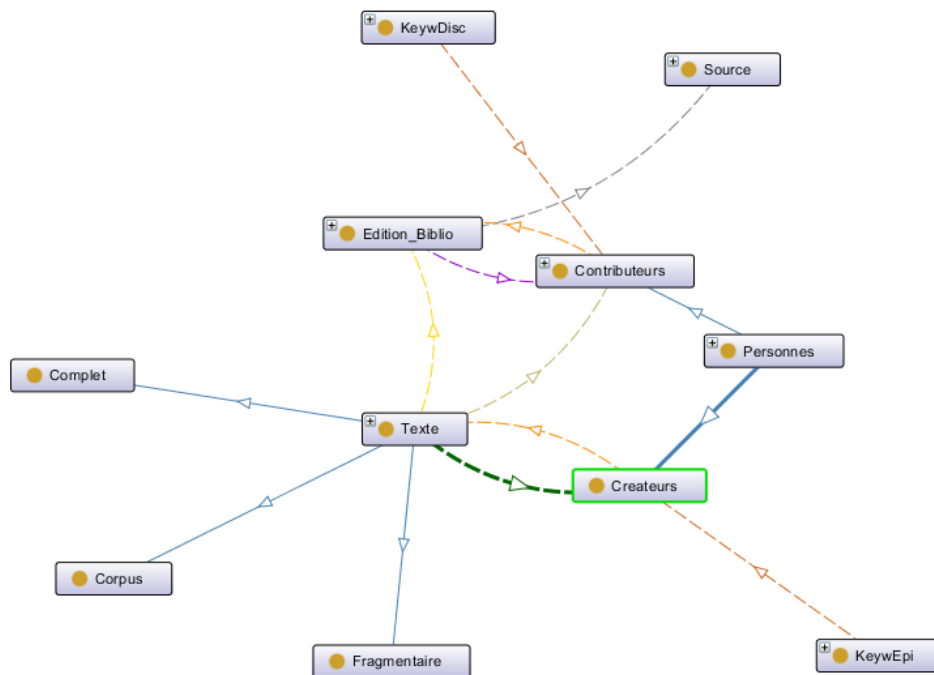


Figure 11: Partial graph of IPhIS ontology.

Implementation of this ontology is not immediately desirable, since it would lead to a very significant technical constraint by causing, apart from the creation of one or more namespaces, the need to attribute a considerable number of URI because of the growing size of the database and manage the negotiation of content between the HTML and RDF versions for each item of data. This is a very real constraint, and one to which experts in the bibliographical semantic web have been drawing attention for quite some time<sup>43</sup>.

## Conclusion

As we see it, it is not pertinent, because of the nature of our data on ancient texts, to only use tools such as Nakala to make sure that they are displayed and harvested. Our philological aims are as varied as they are complex; there is therefore always a constraint to either force their harmonisation so that it is not necessary to draw up over-heavy sharing schemas, or simplify them artificially so that they fit into categories that were not designed to accommodate them.

<sup>43</sup> “The attribution of URI specific to your set of data for secondary entities is thus an additional burden since you will have to maintain them; however, it is also a security feature in terms of the consistency of your set of data. It all depends on how much you trust the sets of data whose URI you re-use, and the ease of discerning which entities you are manipulating within these external artefacts. If you create your own URI for secondary entities, you can always link them to the others at a later stage (see A.7).” [trans.] [Bermès, Isaac & Poupeau, 2013 paragraph 49]. See: <https://www.cairn.info/le-web-semanticque-en-bibliotheque--9782765414179-page-69.htm>; consulted on 23 April 2021.

Happily, there are other less ponderous solutions for displaying data such as ours in the best possible way. Some are already virtually in place, as a result of our initial choice to cross our own data with as many external repositories as possible. The alignment work carried out with the VIAF, the Pinakes database authorities, dataBNF, or even, in the case of our cartographic extension, several geographical repositories, may ultimately make it possible to envisage entering IPhIS data via channels that are not strictly philological, either via a general bibliography or by using access paths that cease to have anything to do with the base material and are semantic by nature from the outset - or at least readily convertible into a semantic logic. What is more, we have long given CIRIS users the possibility of downloading a large part of the data in open format for their personal use. This is the case not only for the thesauruses of texts by ancient authors, which are directly downloadable in \*.csv format in CIRIS, but also for the edition records that can be exported in RIS format, thereby making them compatible with most bibliographical management software. In the cartographic extension carried out using the open application uMap<sup>44</sup>, all the data for the maps can also be downloaded (Figure 12). Diversifying the ways of displaying our data does not in any way mean having to renounce full interoperability; quite the contrary, in fact: it means optimising this sharing by selecting, for each aspect of the items of data, the tool best suited to their display. It is important to take care in determining which item of data should be entrusted to which tool. Doing so takes advantage of the specific features of the items - their transmission is complex, they are charged with a scientific uncertainty that it is our duty to make known, and they represent a wealth of variety of types that it would be wrong to attempt to standardise.

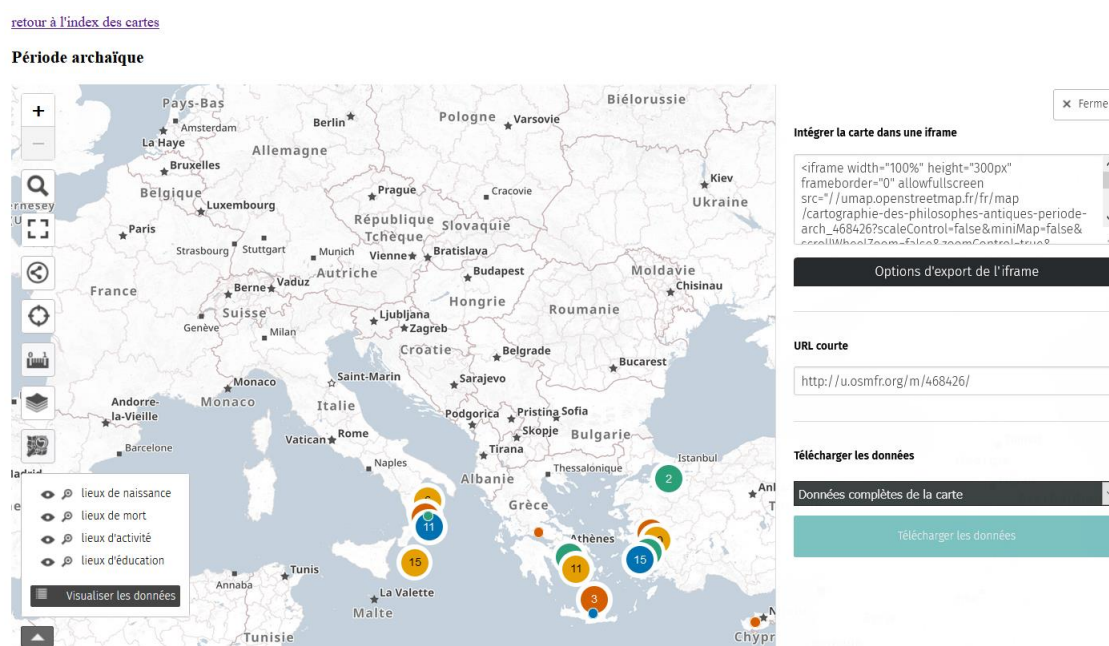


Figure 12: Access to geographical data in CIRIS.

An effort still needs to be made regarding the data contained in thesauruses. Here we feel the choice to retain a firm barrier between back- and front-office remains pertinent, since this is checked data that must be subjected to tight editorial control, under an expert eye. That is why, while contributions to the IPhIS database are allowed, in order to enable external users to

<sup>44</sup> <https://ciris.huma-num.fr/cartographie.php>; consulted on 23 April 2021.

incorporate bibliographic records<sup>45</sup>, there are no plans to open up the thesauruses of ancient authors and ancient texts to similar collaboration.

However, although IPhIS cannot be opened up to collaboration, this is possible in other frameworks. One of our desiderata for the coming years is an improvement in the interoperability of these thesauruses in order to incorporate them in what is currently the largest collaborative data reservoir in the world: WikiData. Ultimately, this is a way of thinking about opening up data that is nourished not by a desire for the visibility of a project but rather for the possibility of contributing to a collective, global improvement in knowledge in our field – which we believe ought to continue to be the main aim of truly open science.

## IV REFERENCES AND CITATIONS

- Babeu, A. 'Rome Wasn't Digitized in a Day': Building a Cyberinfrastructure for Digital Classicists, CLIR (blog), 2011. <https://www.clir.org/pubs/reports/pub150/>
- Bermès, E., Isaac, A., & Poupeau, G. Cas A : Publier des données en RDF. *Le Web sémantique en bibliothèque*, 2013 : 69-122. <http://www.cairn.info/le-web-semantique-en-bibliotheque--9782765414179-page-69.htm>
- Calame, C. Les sciences de l'Antiquité. Entre néolibéralisme et culture de supermarché : inflation bibliographique et désarroi méthodologique. *Les Cahiers du Centre de Recherches Historiques. Archives*. 2006; 37 (June): 67-87. <https://doi.org/10.4000/ccrh.3131>
- Capron, L. Le 'Fragmentum Muratorianum'. *L'information philologique* (blog). 2016a. <https://iphi.hypotheses.org/158>
- Capron, L. Sarmatas, Silvain et Théophile sont dans un codex (ou deux) : quand l'identification des textes sert la codicologie. *L'information philologique* (blog). 2016b. <https://iphi.hypotheses.org/418>
- Capron, L. De quoi l'Anthologie grecque est-elle le nom ? Histoire éditoriale d'un corpus. *L'information philologique* (blog). 2016c. <https://iphi.hypotheses.org/562>
- Clérice, T. Les outils CapiTainS, l'édition numérique et l'exploitation des textes. *Médiévales*. 2017;73: 115-31. <http://www.jstor.org/stable/44988401>
- Crane, G. Classics and the Computer: An End of the History. S. Schreibman, R. Siemens, and J. Unsworth (eds.) *A Companion to Digital Humanities*, Blackwell (Oxford), 2004. <http://www.digitalhumanities.org/companion/>
- Francart, T. Créer des référentiels SKOS/RDF à partir d'Excel. Sparna Blog (blog). 12 January 2017. <http://blog.sparna.fr/2017/01/12/creer-des-referentiels-skosrdf-a-partir-dexcel/>
- Giovacchini, J. 'Qu'est-ce que c'est que ça s'il vous plaît ? C'est le titre !' (1) and (2). *L'information philologique* (blog). 2016. <https://iphi.hypotheses.org/372> and <https://iphi.hypotheses.org/377>
- Giovacchini, J., Berra, A., Capron, L., Psilakis, C., Weiss, B., Langlais, P.-C. & Lemaire, J. Technical report on 'Carte des philosophes antiques' project. Labex HaStec, 2020. <https://hal.archives-ouvertes.fr/hal-03065915>
- Giovacchini, J., Weiss, B., Capron, L., Grignon, S. & Caye, P. Technical report on IPhIS project, 2017. <https://hal.archives-ouvertes.fr/hal-01509598>
- Grignon, S. 'Je ne suis pas [qu'] un numéro !' *L'information philologique* (blog). 2016. <https://iphi.hypotheses.org/345>
- Hilbold, I. Jules Marouzeau and *L'Année Philologique*: The Genesis of a Reform in Classical Bibliography. *History of Classical Scholarship*, 2019: 174-202. <https://www.hcsjournal.org/ojs/index.php/hcs/article/view/21>

---

<sup>45</sup> Documentation is in fact currently being drafted within the general documentation, to help contributors: <https://ciris.huma-num.fr/manuel/doku.php?id=manuelciris:contribuer> ; consulted on 23 April 2021.