HAL open science

Ancient written cultures and digital technology: Biblissima(+)

Anne-Marie Turcan-Verkerk

▶ To cite this version:

Anne-Marie Turcan-Verkerk. Ancient written cultures and digital technology: Biblissima(+). Bibliothek und wissenschaft, 2022, 55, pp.103-119. hal-04048320

HAL Id: hal-04048320

https://ephe.hal.science/hal-04048320v1

Submitted on 27 Mar 2023

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.

Distributed under a Creative Commons Attribution 4.0 International License

Anne-Marie Turcan-Verkerk, « Ancient written cultures and digital technology: Biblissima(+). Preparing tomorrow's research and providing access to the written heritage », *Bibliothek und Wissenschaft,* 55, 2022 : *Faszination (Buch-)Handschriften im Jahr 2022. Tradition und Zukunft ihrer Erschließung in Bibliothek und Wissenschaft*, cur. Claudia FABIAN, p. 103-119.

Anne-Marie Turcan-Verkerk

Ancient written cultures and digital technology: Biblissima(+).

Preparing tomorrow's research and providing access to the written heritage

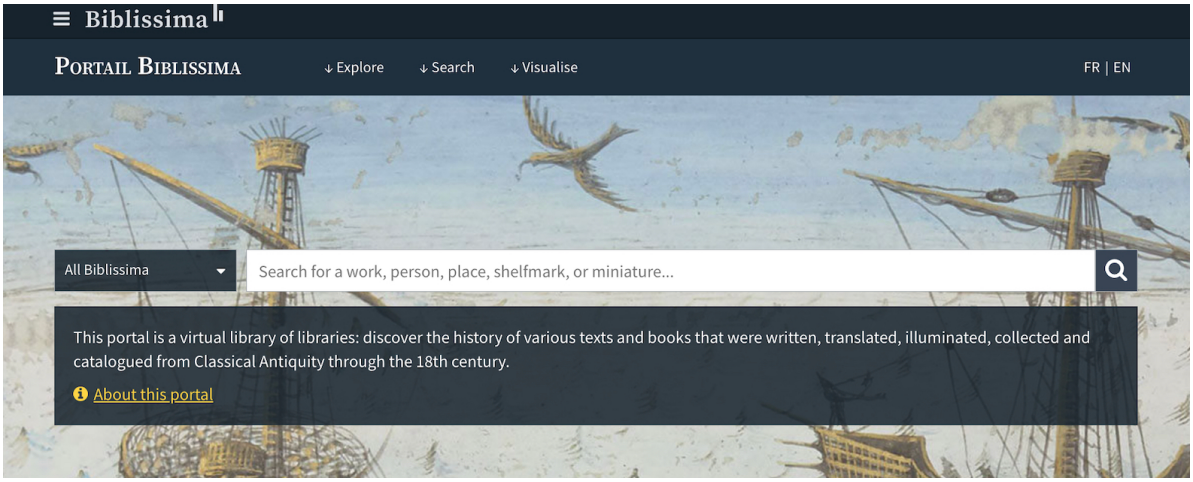
Biblissima, the Observatory of Written Heritage of the Middle Ages and the Renaissance, was created in 2012 thanks to a grant from the French government as part of the "Investissements d'avenir" programme; the project ended on 31 December 2021. This Observatory is a "data equipment" on the circulation of texts in the West from the 8th to the 18th century, mainly in Latin, Greek, Hebrew and Old French. Adapted to the needs of the Humanities and Social Sciences (HSS), their structures and their professions (research, training, conservation, mediation...), it is intended to accumulate the data necessary for the full exercise of these professions, and to make them interoperable in order to overcome the barriers caused by the heterogeneity of formats and vocabularies. Biblissima provides access to a number of tools for deploying the lifecycle of data, from creation to enrichment and reuse, while respecting and promoting the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. In this way, it contributes to the digital conversion of the humanities, both in France and internationally, while making ancient manuscripts and prints, their texts and iconography, accessible to all audiences. Biblissima is in fact the initial phase of a process that is intended to take on a quite different dimension. This first project laid the foundations for a national infrastructure with an international vocation, which began in 2021 and will continue until 2029: Biblissima+, Observatory of Ancient Written Cultures from Clay to Print. Biblissima(+) is first and foremost a research project. It aims at helping the research communities involved in the study of the transmission history of ancient texts to cross new frontiers. To do so, it is necessary to bring these communities closer together and to integrate their data, their methods, their practices and their tools (whatever their size, format and means) in order to build a coherent ecosystem that everyone can use with profit.[1]

I. The first Biblissima and its achievements

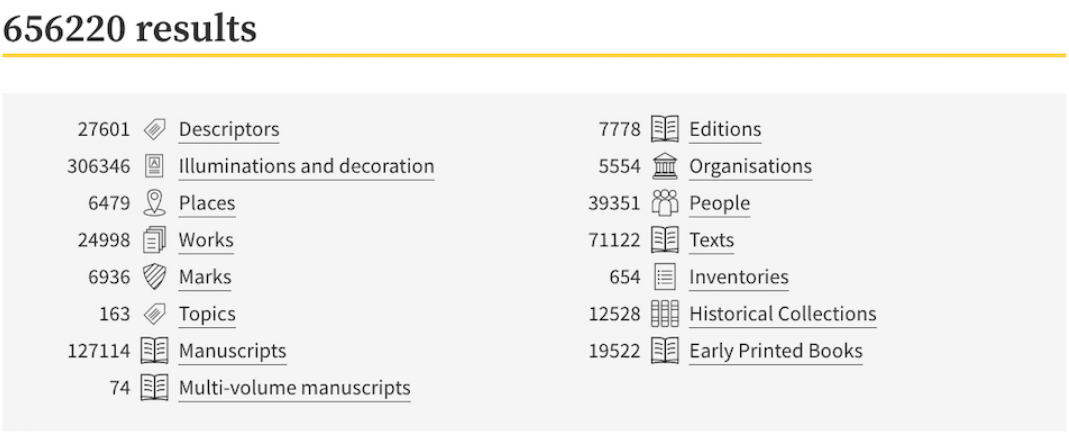
With equal urgency, Biblissima had a major and a minor. The major one was data acquisition, as the aim was to gather a massive corpus of data on the transmission of ancient texts in the West from the 8th to the 18th century, to allow new research in this field, new questions, unexpected results. Around 70% of the funding was devoted to the acquisition, collection and online publication of data. The teams worked as much on the digital preservation and dissemination of legacy data as on the production of new data, based on the exploration of poorly known documentary material and the use of new technologies. It was particularly strategic to bring together the major online catalogues of manuscripts and incunabula, such as the one of the Bibliothèque nationale de France (BnF Archives et manuscrits) or the catalogues of incunabula of French libraries, and the databases developed by researchers on the same collections, but with a

1 Biblissima was funded by the French government as part of the "Programme d'investissements d'avenir", with the reference ANR-11-EQPX-0007. Biblissima+ is funded by the French government as part of the "Programme d'investissements d'avenir", with the reference ANR-21-ESRE-0005. The project is coordinated by Anne-Marie Turcan-Verkerk (Ecole Pratique des Hautes Études - PSL) on behalf of the Campus Condorcet, a new university campus dedicated to the humanities and social sciences that has been set up at the north of Paris, in Aubervilliers. The Campus Condorcet brings together higher education institutions and research centres, several of which work together within Biblissima partnership; they have grouped their documentary and heritage collections within a single library, the Grand Equipement Documentaire (GED).

different point of view (in particular the databases and catalogues developed by the Institut de recherche et d'histoire des textes, IRHT, in the field of library history, on Greek manuscripts, Old French manuscripts, illuminated manuscripts...). It was also crucial to encourage research laboratories or scientific projects of limited duration to work closely with libraries and archives, and to safeguard the results of their work, make them interoperable and increase their impact.[2] The minor task, in financial terms but not in terms of impact, has been the creation, development and enrichment of a digital infrastructure making all these data interoperable, and which gives the project its meaning: the Biblissima portal and the various related sites, a single entry point into interoperable resources.



Today, the objectives of Biblissima have been achieved on the whole: some projects could not be completed, but others, which were not foreseen, have been developed, so that the critical mass of data expected by the project has been reached, sometimes exceeded, archived, and progressively made interoperable by the portal, which has more than 650,000 pages to date.

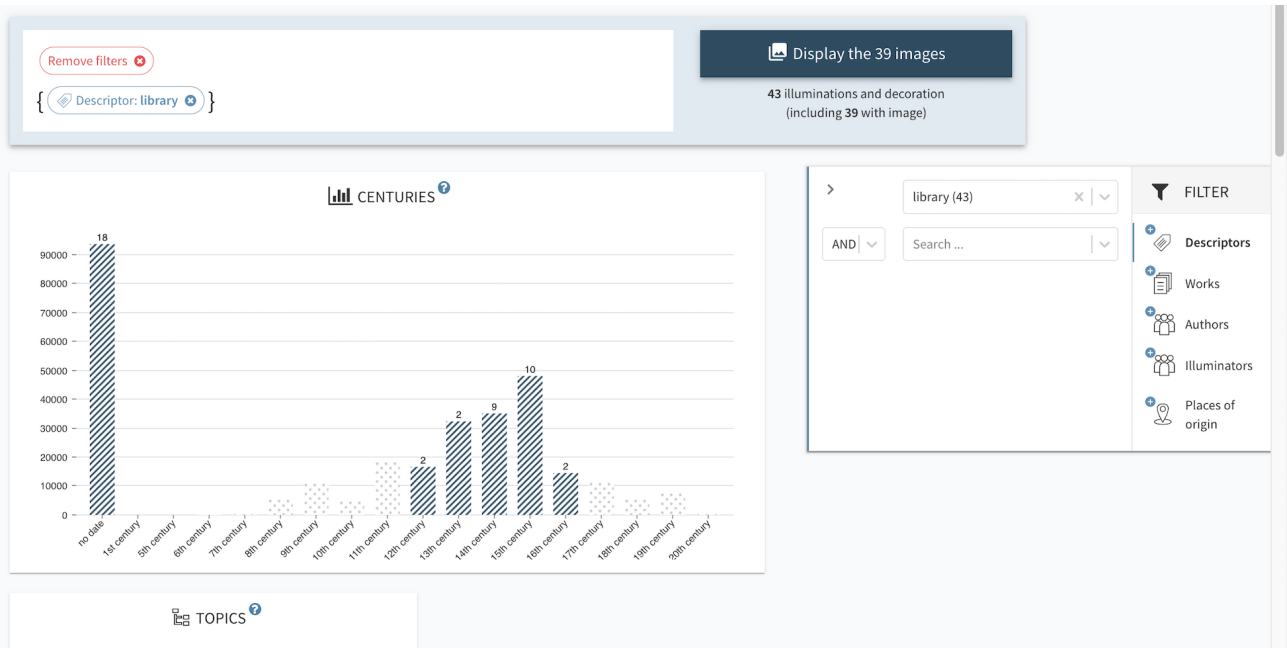


I.1 Progressive enrichment of the Biblissima portal (2013-2021)

The first phase of data processing and integration in the portal focused on the history of early manuscript and printed collections. The portal has also become a leading working tool in the field

2 All Biblissima projects, except the most recent ones, are described in the "Biblissima booklet" prepared for the study day on 15 March 2018: Biblissima. Innover pour redécouvrir le patrimoine écrit. Paris 2018, 200 pages (by Anne-Marie Turcan-Verkerk with Eduard Frunzeanu and Régis Robineau, and the collaboration of Kévin Bois), downloadable: https://projet.biblissima.fr/fr/evenements/journee-biblissima-iiif-innover-redecouvrir-patrimoine-ecrit-2018.

of manuscript iconography, with the interoperability in 2020 of the two major French iconographic databases, Mandragore (Bibliothèque nationale de France - BnF) and Initiale (Institut de recherche et histoire des textes - IRHT), thanks to the development of an iconographic thesaurus with more than 27,000 entries reconciling the vocabularies of the two databases, giving thematic access to their 230,000 images (and more than 306,000 records). In 2021, an innovative interface for searching and visualising iconography was created, with the creation of co-occurrence diagrams of iconographic descriptors.



In 2021, the portal also opened up to full-text bibliography, thanks to a test carried out with the national Persée platform (persee.fr), which provides free and open access to complete collections of scholarly publications (journals, books, conference proceedings, serial publications, primary sources, etc.) and to a range of research and exploitation tools. Given the nature and history of these manuscript collections, the records of the catalogue of latin classical manuscripts of the Vatican Library[3] were made interoperable with the other resources of the portal, in particular the Pinakes[4] and Bibale[5] databases, and also with the records of the Heidelberg manuscripts[6] and the digitisations of the Vatican Library.

The current version of the Biblissima portal integrates datasets from 20 sources, some of which were not originally planned, both in France[7] and abroad[8], among them Europeana Regia, Virtual library on the great royal collections of manuscripts in Europe during the Middle Ages and the

3 AAVV, *Les manuscrits classiques latins de la Bibliothèque Vaticane*, I-III,2, Paris, 1975-2010 (Documents, Études et Répertoires publiés par l'IRHT, XXI,1-3/2) (https://www.persee.fr/collection/dirht)
4 The Pinakes database (IRHT) lists all known Greek texts up to the 16th century and the catalogued manuscripts that transmit them (except for papyri).
5 Bibale (IRHT) is a database on the transmission of medieval texts in the West and the study of ancient and modern collections of ancient manuscripts and printed books and their owners.
6 Heidelberger historische Bestände. Manuscripts from the Historical Collections of the University Library of Heidelberg (handschriften-digital.uni-hd.de).
7 For example, the digitised manuscripts and prints available on Mazarinum, the digital library of the Mazarine Library in Paris (mazarinum.bibliotheque-mazarine.fr).
8 For example, the Medieval manuscripts from the Wellcome Collection Library (Wellcome Library, London) (wellcomelibrary.org) or the Bremen manuscripts can be found in the portal.

Renaissance (BnF, BSB, BHUV, HAB, KBR). In addition to the large databases, digital libraries and catalogues that are at the basis of the project, other resources have gradually been added, either because Biblissima has contributed to their funding, or because those responsible for them have seen the Biblissima portal as a means of ensuring that their data has a certain influence, a form of sustainability, and above all, an enrichment thanks to the scientific connections made possible by the portal.[9]

I.2 IIIF Collections and the development of IIIF in France

A new research platform, which was not originally planned but the need for which has been felt over the years, was born in 2018: IIIF Collections of Manuscripts and Rare Books[10]. IIIF Collections provides simple and unified access to approximately 85,000 early books, both manuscripts and prints, whose scans are interoperable through the International Image Interoperability Framework (IIIF) protocol. At present, 27 digital libraries from Belgium, Canada, France, Germany, the Netherlands, Spain, Switzerland, the United Kingdom and the United States can be searched together on any term, in particular provenance data and author names.

The latest data integration efforts in the Biblissima portal are part of a process of bringing together the Biblissima portal itself and the IIIF Collections platform. We realise that the libraries that join IIIF Collections think of integrating Biblissima, which means Biblissima is increasingly expected to be exhaustive. Therefore, in 2021, we have started to include the data from IIIF Collections into the Biblissima portal. This will enable users to access all the interoperable data offered by the project in its various platforms. For all those who appreciate the ergonomic and very simple interface of IIIF Collections and who are mainly looking for digitisations of manuscripts and early prints, the platform is maintained for itself and continues to play its role as a research tool in the interoperable digitised collections that have adopted the IIIF protocol. Biblissima thus offers two complementary working tools.

Finally, a service to accompany HSS communities in the adoption of the IIIF image interoperability protocol, IIIF 360, was created by Biblissima in 2018, with the top-level research infrastructure Huma-Num[11] and the support of the Campus Condorcet. To date, Régis Robineau, who is also a member of the Technical Review Committee of the IIIF community and one of the five IIIF ambassadors in the world, has responded to some forty requests, mainly from the world of higher education and research (laboratories, projects, university libraries).[12]

I.3 The Biblissima authority files

In 2018 another platform was created, data.biblissima.fr. Data.biblissima is a platform for managing and sharing the authority files which are the backbone of both the Biblissima portal and IIIF Collections. There is indeed a close link between the quality of the aggregation and interoperability work and that of the authority files: it is not enough to assign an identifier to each entity, the data must be aligned, deduplicated or, on the contrary, merged together, and disambiguated, a particularly delicate task in all disciplines of high erudition, in particular for texts

---

9 For full reference see: "Version du portail en cours" (https://portail.biblissima.fr/a-propos).

10 (https://iiif.biblissima.fr/collections/).

11 Huma-Num is a French national Research Infrastructure devoted to Social Sciences and Humanities. Huma-Num offers a variety of platforms and tools for the processing, conservation, dissemination and long-term preservation of digital research data.

12 Recently, Biblissima organised an IIIF day on 24 March 2021 (online, due to the health crisis). The event gathered 10 speakers and nearly 200 participants (materials and video recordings on: https://projet.biblissima.fr/en/courses-and-events/rendez-vous-iiif360-2021). On 24 June 2021, Biblissima co-organised with the BnF the regional session "IIIF France" within the framework of the international IIIF conference (22-24 June 2021; materials and video recordings at: https://projet.biblissima.fr/en/courses-and-events/2021-iiif-annual-conference-regional-meeting-france).

in ancient languages. Authority files are only published on data.biblissima when they are fully developed and reusable. So far, more than 300,000 items have been published on the platform: there are persons and organisations, geographical names, and about 200,000 shelfmarks of manuscripts and early printed books. The authority files for textual works and iconographic descriptors will soon be published.

With regards to shelfmarks, work is being carried out systematically for France. For foreign institutions, the integration of shelfmarks into the authority file is carried out as the data arrives in the partners' resources and, much more massively, as the resources increase in the IIIF Collections platform. Gradually, a large authority file of ancient books in the world is being built up, which should be extended at least to archives, papyri and ostraka, and probably one day to other text-bearing objects identifiable by their inventory numbers. For manuscripts, the international consortium ISMI (International Standard Manuscript Identifier), with cooperation in Germany by the Bayerische Staatsbibliothek – BSB, in Switzerland by E-Codices and the University of Fribourg, and in France by the IRHT and Biblissima, has a higher ambition thanks to the federation of conservation institutions on a worldwide scale.[13] Biblissima+ will fund the ISMI consultation and query interface, using data.biblissima as a starting point and as a storage/management platform for the authority file, in order not to create new duplication.

Biblissima has chosen the Wikibase model for its authority files, because it is the one that France has chosen for its National Entity File (Fichier national d'entités - FNE) project led by the Bibliographic Agency for Higher Education (ABES) and the BnF. The data.biblissima platform, which federates data produced in an uncoordinated manner by the scientific partners, is therefore prepared to become a supplier of the FNE. Biblissima+ will finance the development of authority files and federate them, in order to become a national service for the reconciliation highly erudite data. Data.biblissima will be able to supply the FNE, but also international authority files such as ISNI or Geonames, with verified data, stably aligned with the major authority files (such as BnF, Library of Congress or Deutsche Nationalbibliothek [DNB] for names of persons and organisations and works [rather than VIAF, which is proving unstable]).

Biblissima, by aggregating foreign resources, is already aiming at least at bilingualism of its authority files. Through its international network, thanks to the extension of partner projects to international collaborations (see below), and thanks to IIIF Collections, Biblissima+ will give life to a large pool of data to be made interoperable, from which it will be necessary to extract multilingual descriptors, align them, produce enriched thesauri, redistribute them and make them collaborative, or even participatory; the data.biblissima platform is also planned to be opened up to participatory science.

I.4 The Biblissima tools

In order to broaden its audience and contribute to their education, Biblissima provides innovative tools for the description of bindings, for editing in XML-TEI (Text Encoding Initiative), and for the analysis of ancient languages. These tools attract a large part of the flow of Internet users to the site.

I.4.1 Bindings

The study of bindings was particularly advanced at the Reserve des livres rares et précieux of the Bibliothèque nationale de France. The BnF has created, in connection with its bindings database,

13 François Bougard, Matthieu Cassin, William Duba, Claudia Fabian, Christoph Fluëler, Anne-Marie Turcan-Verkerk: International Standard Manuscript Identifier (ISMI): pour un registre électronique des identifiants des livres manuscrits. In: Digitalia 15/1 (2020), pp. 45-52.

reliures.bnf, an XML-TEI encoding scheme for ancient bindings, accompanied by its manual and its dictionary of tags[14].

I.4.2 Editing chain and data display

The University of Caen, thanks to the collaborative work of the Pôle Document Numérique of the Maison de la Recherche en Sciences Humaines, the Centre de Recherches Archéologiques et Historiques Anciennes et Médiévales (CRAHAM), the Presses Universitaires de Caen and the Métopes research infrastructure, has become a leader in the digital scholarly editing of ancient sources as well as in online university publishing. An important collaboration between the IRHT, the CRAHAM and especially the Pôle Document Numérique gave birth to an environment of critical edition in TEI of ancient inventories of books, thanks to which the collections Thecae (*Thesaurus Catalogorum Electronicus*)[15], for completed critical editions, and Thecae-Lab, for editions in progress, were created. This programme has contributed to the creation of a complete chain of generic tools, based on a global system combining an XML-EAD cataloguing environment, an XML-TEI source editing framework and a thesaurus creation framework, initially developed for a pilot corpus, the sources for the history of Mont-Saint-Michel.

I.4.3 Helping people understand Latin and Greek: Collatinus and Eulexis

In order to open up the content of ancient books to a wide range of audiences, it is necessary to help them understand ancient languages, and first of all Latin and Greek. To this end, a tool created by Yves Ouvrard for Latin, Collatinus, has been strongly developed, and Eulexis has been created for Greek[16]. Over a period of ten years, Collatinus, a resident and web-based Latin translation tool, has been upgraded with the addition of syllable quantities, a scansion module and a probabilistic tagger. The range of searchable dictionaries has been considerably expanded (from one, the digitised french dictionary of Félix Gaffiot, to 16). The lexicon has grown from 11,000 lemmas to 85,000 with the systematic scanning of digital dictionaries. In 2021, Dutch and Basque were added to the translation languages; Vietnamese is being validated. Versions 11.3 and 12beta are being finalised. Eulexis was created by Philippe Verkerk in 2014 as a web version and in 2017 as a resident version. The resident version paved the way for translations into several languages. In 2020, Gérard Gréco's Bailly was added to the three original dictionaries (Bailly's abridged for French, LSJ for English and Pape for German). In the consultation of the dictionaries, the ability to use wildcards or regular expressions has been added for the resident version. The consultation of Collatinus and Eulexis is increasing and has seen a leap since 2020, which can be explained by the pandemic but also by their referencing on Lexilogos.

***

From now on, even if exhaustiveness is impossible and far from being achieved, the data equipment provided by the project exists, it constitutes a base, and it has built around it a community closely associating the world of libraries and archives and that of research. Most of the growth must now come from the dynamics of the portal itself, capable of constantly aggregating new content from its partners but also from outside its initial partnership, following what the IIIF Collections platform is already doing. For this reason, in the new Biblissima+ project (2021-2029), efforts will focus on the mechanisms for increasing and updating the portal's data, on the tools for enriching the corpus thus constituted, and on the ergonomics of the query and work interfaces. Thereby, research provides itself an ambitious working tool thoughtfully designed from its current

14 (https://outils.biblissima.fr/en/tei-schemas-bookbindings/)
15 (https://www.unicaen.fr/services/puc/sources/thecae//accueil)
16 Ph. Verkerk, Y. Ouvrard: Collatinus & Eulexis : Latin & Greek Dictionaries in the Digital Ages. In: Digital Classics III: Re-thinking Text Analysis (Heidelberg mai 2017). Classic@ Journal, 20, 2019, Harvard University, online : https://classics-at.chs.harvard.edu/collatinus-eulexis-latin-greek-dictionaries-in-the-digital-ages/

needs, and which foster an ecosystem effect that will also create future working methods well adapted to the field.

II. Towards Biblissima+, an Observatory of ancient written cultures

Up to now, Biblissima has worked mainly on French collections, because the type of funding provided by the French government required it, and because Biblissima was a prototype at national level. However, the history of the transmission of texts shows how essential it is to take into account the whole range of collections existing in the world in order to reconstruct the ancient phases of the history of written cultures. Today, new research practices and new technical possibilities have been added to this scientific requirement: no one works only with his or her national community, research can only be international, and it is at this level that collaborations and the design of new tools must take place. Solidly built on the foundation of its data cluster, on the initiatives and partnerships of the previous project, the Biblissima+ infrastructure will therefore be deployed on a much wider perimeter to take into account the needs of research and the institutions that support or use it, which have emerged in the course of the project's development. Biblissima+ will therefore be an Observatory of ancient written cultures, focused as before on written heritage, with a primary interest in ancient texts and their transmission, but

- with no time limit,
- without limit of language,
- with no limit on the type of documentation (ancient artefacts of all kinds bearing text, ancient and modern bibliography, historical or scientific archives, sound documents, computer simulations and vision, etc.)
- with the whole range of sciences interacting with the more traditional disciplines in the field of humanities, which today include STEM, cognitive sciences, artificial intelligence, resulting in a broad redefinition of the disciplines of scholarship,
- with particular attention to its international outreach.

The result will be a digital infrastructure devoted to the history of the transmission of texts produced from Antiquity to the Renaissance in the East and the West, whatever the medium and language. The focus will be on the collections of documents and heritage objects held by libraries, archives and museums. Biblissima+ will develop its portal by providing easy access to original written documentation, full-text bibliography and related research archives; the project will also provide access to a working environment offering tool chains for enriching, sharing and reusing corpora. Its aim is to thoroughly renew research on ancient written cultures in France and throughout the world.

II.1 Biblissima+ and the expansion of its offer

The existence of a portal such as Biblissima, which offers the possibility of searching through a variety of resources in a single interface, leads to a legitimate desire for completeness. The informed user does not understand that there are missing data known to exist elsewhere and expected to be found within the portal. The non-expert user misses some of the information and does not necessarily know that it exists, let alone where to find it. In order to constitute a national research infrastructure, such a portal must therefore go further than the first Biblissima: it must broaden the spectrum of resources, and it must project itself into the future, i.e. it must have a mode of operation that allows the aggregation of new resources and the automation of their updating.

The role of the data.biblissima authority files

To achieve this, the team in charge of the Biblissima+ portal[17], coordinated by Emmanuelle Morlock (HiSoMA laboratory, Lyon), is putting in place the logic and mechanisms for expanding the pool of data aggregated by the portal via its authority files: document descriptions, bibliographic data, scientific archives, and open access publications. Biblissima+ will thus be able to interface as well as possible with the main programmes and tools that France has set up, by showing how they can be linked together: in particular the Catalogue collectif de France (CCFr), the national Open Archive HAL, the digitisation programme for scientific archives and journals Collex Persée, Istex, a set of open access resources acquired for the academic world, OPERAS, "the European Research Infrastructure for the development of open scholarly communication in the social sciences and humanities", in connection with the Grand Equipement Documentaire on the Campus Condorcet and the new EPHE-PSL Institute for Linguistic Heritage and Diversity, ILARA, in connection with the data of the Archives nationales and FranceArchives, the national portal of the archives of France.
Today, the full-text bibliography is missing from the Biblissima working environment. However, France has created an infrastructure for digitising and making available its HSS journals and collections, Persée, whose resources are particularly relevant to the written heritage of French libraries and archives, research in art history and archaeology, philology and, more generally, the erudition and history of written cultures. The journals and collections put online by Persée are accessible free of charge to any French or foreign Internet user. Biblissima+ and the Persée platform, a founding partner of the programme, will therefore share their authority files, in particular those of shelfmarks, names of historical places, names of persons and institutions, in order to systematically link the data of the portal to the full text of the books and articles that document them, and vice versa. In Persée, the bibliographic search will be able to use the Biblissima+ portal to access primary sources and the latest research on them, in order to put the bibliography into perspective and to go beyond it. It is possible to imagine, at a later stage, research of a different nature, for example to reconstitute the documentation of a particular scholar, or to follow the developments in the history of ideas and criticism concerning a particular concept, a particular person, a particular text and so on, especially on the basis of reviews. This resource is very complementary to the BnF's digital library Gallica, where there is also a large number of digitised collections and journals.

This experience could serve as a prototype for others of the same kind, for example by making data from data.biblissima available to the HAL open archive, or Huma-Num and the projects with which it is associated, in particular the European TRIPLE project and his platform, "an innovative multilingual and multicultural discovery solution for the social sciences and humanities (SSH)"[18], but also with the aim of collaborating with online publishing houses. Similarly, any project funded or co-financed by Biblissima+, or simply wishing to make its data available on the portal and interoperable with others, will have to integrate the Biblissima+ identifiers and contribute to the alignment of the data. The enrichment of the portal is thus directly linked to the development of data.biblissima and its role as a platform for data reconciliation in areas where information is particularly lacking and complex, as mentioned above. The project for an international identifier for manuscripts, ISMI, is one of the most important elements (see above).

Biblissima+ partner projects and the gradual expansion of the ecosystem

From 2013 to 2018, Biblissima's calls for partner projects have encouraged collaboration between conservation institutions and research teams. Funding for digitisation, cataloguing and the development of tools was only granted if the operation was based on a scientific project, shared between a research team and the conservation institution carrying out the project. Following an evaluation by Biblissima's International Scientific Advisory Board, funding was granted to

17 Régis Robineau (head), Eduard Frunzeanu and Kévin Bois: https://projet.biblissima.fr/fr/organisation/equipe-biblissima

18 (https://project.gotriple.eu/).

Biblissima's partners, which guaranteed the proper execution of the project. These partner projects have breathed new life into libraries and archives and have promoted a scientific policy of digitisation of ancient documents. Scientifically, they were the unexpected and therefore the inventive part of the project. They have also created a network of libraries, archives and research teams that have benefited from the help of Biblissima[19]. It is also through these partner projects that training operations, the Biblissima summer schools, were set up, thanks to the leverage of the funding: any project that received at least €40,000 had to organise an international summer school by its own means; 120 French and foreign students and professionals were trained as part of these courses. Many people working in digital humanities projects, in heritage collections and in higher education have participated in these summer schools, which have played a decisive role in demonstrating to them the fruitfulness and therefore the need to combine the disciplines of scholarship and digital practice.

For Biblissima+, it remains vital to provide funding for innovation and training, be it for inventive scientific projects producing new data, new resources to be made interoperable via the portal, or new tools and services for research. This is all the more vital as the portal is opening up much more widely to archival services, to new linguistic domains, to new communities, in particular epigraphists, archaeologists, numismatists and sigillographers. Biblissima+ is not a funding agency, but a project that seeks to enrich itself in an interesting and coherent way. Through its calls for projects, coordinated by François Bougard (IRHT, Aubervilliers - Paris - Orléans), it therefore promotes a rich co-construction between the founding teams of Biblissima+ and new partners, while maintaining a strong link with libraries, archives and museums. This process guarantees a gentle, gradual expansion and constant attention to the emergence of new ideas and tools.

In 2022-2023, Biblissima+ will launch a call for projects in three categories, in order to compare what is comparable and to meet a wide range of needs. The first will be for projects of significant size. The second category will encourage risk-taking and innovation, with 'small' exploratory funding. It is hoped that this category will provide seed funding for projects that can be scaled up the following year. It may also be a question of unblocking a project with additional funding. Finally, Biblissima will encourage projects by young researchers, the aim of which will be to find innovative usage for the portal, to expand the Biblissima+ network, to make it more dynamic and to enhance its value, while at the same time investing in the future by helping young researchers. The evaluations will be carried out by the International Scientific Advisory Board. In all cases, the association with a heritage institution will be a decisive element of the application, as the FAIRisation of the data and the process of integration into the Biblissima+ portal. For the Biblissima+ team, two things are indeed very important: the possibilities of interconnection with the data of the portal (via authority files), the ease of retrieval of the metadata (by making the sources available on an accessible data repository, the use of open licences and the documentation of the structuring of data). The writing of a data management plan (DMP) for each project will therefore be crucial.

All projects will be able to involve foreign teams, but only 20% of selected projects can be led by an international team. The fundamental objective of these projects is to foster innovation and open up new horizons in every senses of the word (scientific, geographical, social...).

II.2 Digital innovations and ancient sources: preparing for the future

Just as Biblissima has federated research projects and data collection, Biblissima+, on this basis, at the crossroads of the disciplines of erudition and other sciences, will play the same federating role for tools and encourage the development of services for research, in order to become a real

19 (https://projet.biblissima.fr/en/community/other-partnerships)

ecosystem favouring the emergence of new knowledge. To achieve this, it must bring together very diverse communities, avoid sterile competition and costly redundancies, encourage the emergence of new ideas and create genuine cohesion. Biblissima+ brings together, throughout France, 17 research teams working on texts, from antiquity to digital publishing, including a micro-enterprise and the Service Interministériel des Archives de France, which had never worked together in a concerted manner, but 8 of which were already involved in Biblissima. The project must therefore federate these teams and their supervisory institutions at several levels that are both relevant for them and for the project.

II.2.1 Specialised regional centres

First of all, there is a double reality: a disciplinary reality (we have communities of epigraphists, literary scholars, musicologists, philologists and text historians, linguists, developers, etc.) and a regional reality, with laboratories that have developed an internationally recognised specialisation. Biblissima+ aims to encourage the development of these regional centres and at the same time to give them their place in a large-scale construction in which each one contributes its own skills to the service of the whole.

On the Condorcet Campus, the project's coordinator, the Biblissima+ portal team is at the heart of the project. It acts as a link between all the communities and the heritage institutions. It ensures the convergence of authority files and their reuse by the research teams, making them available to all for future interactions. It ensures the interoperability of digital resources and the relevance of data retrieval and updating systems. It develops and maintains the central Biblissima+ sites by adapting them to the needs of users and ensuring that the FAIR principles are respected throughout the processing chain. It ensures the acculturation of the communities and their support, in particular for the adoption of the IIIF protocol.

In Aubervilliers, Paris and Pierrefitte, the actors of the Condorcet Campus, the CNRS, the Archives nationales and PSL University (Ecole Pratique des Hautes Etudes [EPHE], Ecole nationale des Chartes [ENC], Ecole normale supérieure [ENS]) together form a "Swiss Army knife".

- FranceArchives is a single entry point offering federated access to millions of archival documents from 75 partner archives and, in the long term, from all French archives.
- With the ENC's Digital Projects Mission, the Centre Jean Mabillon (CJM) is working on OCR of ancient texts, their editing in TEI, corpus data mining, textometry and stylometry, lemmatisation, and the development of reference systems.
- The Centre de recherche sur la conservation (CRC, at the Muséum national d'histoire naturelle), contributes its knowledge of manuscript materials and the physico-chemical data from its various projects.
- Classicists and medievalists at EHESS (École des Hautes Études en Sciences Sociales) have developed several databases (Images, Thema, Quodlibase) and work environments (ALGreM, Savoirs) that bring together TEI, GIS, and bibliographic resources, feeding the new Didomena data warehouse and the DATU thesaurus of EHESS.
- Within the PSL University, the Scripta-PSL programme. Histoire et pratiques de l'écrit (History and practices of writing) brings together the fundamental sciences of writing (palaeography, codicology, epigraphy, history of the book, etc.) with other HSS (linguistics, history, anthropology, etc.) and the digital and computational humanities, leveraging artificial intelligence (AI) for the automatic recognition of writing and page layout in all graphic systems.
- In the AOROC laboratory (Archéologie et philologie d'Orient et d'Occident), these researchers join numismatists, epigraphists, archaeologists specialised in GIS and 3D reconstruction.

- Within the SAPRAT laboratory (Savoirs et pratiques du Moyen Âge à l'époque contemporaine), Multipal, a transversal palaeography tool, and e-Signa, a portal for medieval symbols, are being developed to pool data (coats of arms, crests, mottos, signatures) from databases and corpus of medieval sources listing these signs and their supports: seals, monuments, stained glass, manuscripts, etc.
- The Grand Equipement Documentaire (GED) is at the crossroads of all the communities involved, on the Condorcet Campus, but also at the national level through its role in the Collex-Persée project (see above) and its links with the national and international networks of libraries and archives.
- The IRHT is now located on the Condorcet Campus and is very involved in machine learning for the automated recognition of forms and characters and in electronic publishing. It works with TEKLIA on machine learning, deep learning and automatic natural language processing (NLP) technologies for the automatic recognition of handwritten and printed Latin script.

In Caen, the CRAHAM laboratory and the Pôle Document numérique (PDN) of the Maison de la Recherche en Sciences Humaines (MRSH) are highlighting their experience in publishing ancient sources, which has the specificity of covering the entire publishing chain, from the researcher's file to publication and distribution by a publishing house, the Presses Universitaires de Caen (PUC), in close collaboration with the METOPES research infrastructure (the 'Caen chain' used by the other partners).

In Lyon and Avignon, with the CIHAM (Histoire, archéologie, littérature des mondes chrétiens et musulmans médiévaux) and HiSoMA (Histoire et sources des mondes antiques) laboratories, classicists, late classicists and medievalists work in multilingual fields around the Mediterranean and closely associate textual science and archaeology, with a strong interest in epigraphy and a great involvement in digital humanities, the TEI consortium and the EpiDoc standard.

In Poitiers, since 1974, the CESCM (Centre d'études supérieures de civilisation médiévale) has published the Corpus des inscriptions de la France médiévale (CIFM), now in TEI (TITULUS project) and extended to the Latin Kingdoms of Jerusalem.

In Tours, with the Centre d'études supérieures de la Renaissance (CESR), the two major fields are the study of the written and musical heritage of the Renaissance. The corpus and data that make up the Bibliothèques virtuelles humanistes (BVH) feed Gallica, Europeana, EDIT16 and are aligned with the main international repositories and catalogues (VIAF, USTC, ISTC). They are also the subject of research in the field of linguistics, OCR and provenance studies. The Ricercar team is introducing the music data processing chain into Biblissima+.

The teams are represented on the Biblissima+ Executive Committee, which is actually building the infrastructure.

II.2.2 Thematic structuring: the 7 Biblissima+ clusters

All these teams design very sophisticated working environments, developed by the communities according to their own needs, which are sometimes only suitable for a small number of users. How do you make it easy for everyone to access? A working environment that is too rich, overflowing with possibilities, confuses the user more than it helps. Lists of tools are multiplying, but they do not seem to be reaching many people. And yet, we need to facilitate access to the whole range of tools, while helping highly innovative teams to design new working tools that will revolutionise tomorrow's science.

In order to encourage all the teams, but also to prevent the tools from developing in a totally anarchic, uncoordinated way, without finally finding new users, Biblissima+ proposes to develop

and organise the innovations created by the teams thematically according to the life cycle of the data and the progress of research, and to design a chain of tools going from the discovery of the source to the in-depth analysis of the text it transmits. This organisation defines major areas of expertise (e.g. 3D, HTR, TEI publishing, etc.), which bring together the players in the field. Current research carried out by the Biblissima+ partners throughout France has led to the emergence of 7 clusters, with their own questions, tools and training courses: these are the 7 clusters which structure Biblissima+:

1. the acquisition of interoperable image corpora (2D, 3D, hyperspectral imaging), supported by the IIIF 360 service (cluster 1);

2. the searchability of physico-chemical analysis data of ancient artefact materials (cluster 2);

3. Artificial intelligence for the automatic reading of ancient writings and pattern recognition (cluster 3);

4. a platform for the analysis and learning of various ancient graphic systems of any kind (cluster 4);

5. a unique cluster dedicated to scholarly editing in TEI (cluster 5), with two sub-groups, one dedicated to epigraphic sources from antiquity to modern times (cluster 5a), the other to the way of editing different textual typologies (cluster 5b);

6. the retrieval, editing in MEI and integration of ancient musical sources in the portal (cluster 6);

7. the most revolutionary tools for automated text analysis (cluster 7).

These clusters bring together several teams, with their own projects, their own networks, which are all funded, and which have an interest in pooling their skills and reflecting together on their own practices. As Biblissima did for research projects, Biblissima+ finances developments at two levels: the main building blocks are funded from the outset by the project and over the long term, and additional building blocks are federated over the course of the annual calls for projects, for unpredictable developments.

The clusters have a decentralised governance. While the digital infrastructure is coordinated by the Biblissima+ portal team, based on the Condorcet Campus, the clusters have their own life, while being linked to the whole and to each other by various means: Each cluster has a coordinator, who regularly exchanges with Marie-Agnès Lucas-Avenel (CRAHAM, Caen), the Biblissima+ cluster coordinator, and who reports on his activities at least once a year to the Biblissima+ executive committee. Each cluster has its own page on the Biblissima+ website[20]; this page gives access to the cluster's data and softwares. Each cluster meets once a year for several days of intensive training and discussion to take stock, share results and problems, and set a common roadmap. The annual Biblissima+ days for appraisal and forward planning enable the clusters to compare their results, to work on the links between them, to discuss the chaining of tools and to determine the roadmap for the following year. They take place just before the Scientific Advisory Board, so that its members can join in and, in the process, better assess the relevance of the partner projects proposed to Biblissima+[21]. In the long term, the Biblissima+ website will reflect in its

20 (https://projet.biblissima.fr/en/project/clusters-biblissima).

21 The first Biblissima+ Days took place on 28 and 29 April 2022: https://projet.biblissima.fr/en/news/journees-biblissima-2022-digital-innovations-and-ancient-sources-how-prepare-future. The presentations will soon be available on the YouTube channel of Biblissima+ and the Digital Forge of the University of Caen.

presentation and ergonomics the developments in the chaining of tools, in order to provide all users with an enriched working tool.

In these groups, ideas and solutions will be discussed and broad guidelines will be determined. In return, Biblissima+ will retrieve the referenced metadata and the annotations and deliverables that researchers are willing to share, to enrich the portal and its authority files with validated data from the funded projects. This consensus building will create a reciprocal cross-fertilisation but also a simplification of the landscape over the years, and will allow the emergence of new ideas, which will make the clusters evolve. The Biblissima+ project thus carries in itself the principles of its evolution.

***

Biblissima+ hopes to take up several challenges:

Scientific, with the aim of accelerating the evolution of HSS communities working on the written memory of humanity to enable them to access information that no one dominates today. At the end of the project, a paradigm shift is expected in the discovery of texts from all cultures, their contextualisation, their understanding, their availability, their integration into the scientific discourse and into the collective memory.

Structural and sociological: Biblissima+ brings together communities of literary scholars, philosophers, historians, archaeologists, art historians, epigraphists, numismatists, and curators of libraries, archives and museums who deal with textual documents, regardless of cultural area or period. The infrastructure brings them together for the first time in a coherent way, allowing them to overcome their differences, to dialogue, to exchange their skills, and to increase them significantly, in an ecosystem approach. It invests in ergonomics and facilitation to help all the communities change their uses and make significant progress in mastering the most innovative digital tools.

Technical, by accompanying these high erudition communities in digital and computational conversion, in particular in the appropriation of artificial intelligence, but also in the fruitful dialogue with physics and chemistry, thus contributing to the redefinition of the humanities of the 21st century. The infrastructure is also helping some STEM (science, technology, engineering and mathematics) communities to take up the challenges offered by the humanities, with their complex data, their culture of exception, and their fuzzy notions.

Finally, the computational turn in the sciences of texts does not only concern research, conservation and training, but every citizen, of every age, of every environment, of every culture, whose heritage is ancient documents, and whose memory and identity are nourished by the written word. The infrastructure develops the most sophisticated technologies of fundamental research to facilitate the availability of the sources and results of the science of texts to all[22].

22 All websites mentioned in this text were last visited on 31 July 2022.