Development of text and image processing for digital libraries: the Bibliothèques Virtuelles Humanistes project and the digitization of Renaissance documents

Toshinori Uetani, Rémi Jimenes, Sandrine Breuil, Jorge Fins, Marie-Luce Demonet, Lauranne Bertrand

Succeed awards

The Succeed awards will recognise the successful implementation of a digitisation programme, especially those exploiting the latest technology and the output of research for the digitisation of historical text.

Nominee information

Institution name

Centre d'Études Supérieures de la Renaissance, University François-Rabelais, Tours, France

Head of institution

Philippe Vendrix

Address

59, rue Néricault-Destouches
BP 11328

City ***State/Province/Region***

Tours

Postal/ZIP Code ***Country***

37013 France

Phone

33 (0)2 47 36 77 61

URL

http://cesr.univ-tours.fr/

Contact information

Contact person

Marie-Luce Demonet

Institution name

Centre d'Études Supérieures de la Renaissance, University François-Rabelais, Tours

Address

59, rue Néricault-Destouches
BP 11328

City ***State/Province/Region***

Tours

Postal/ZIP Code ***Country***

37013 France

Phone

33 (0)2 47 36 77 65

Email

marie-luce.demonet@univ-tours.fr

Nomination summary

Title for the nomination

Development of text and image processing for digital libraries: the Bibliothèques Virtuelles Humanistes project and the digitization of Renaissance documents

Abstract

The BVH (Bibliothèques Virtuelles Humanistes: Virtual Humanistic Libraries) is a research program devoted to the digitization and electronic publication of original source documents from the Renaissance period. Since 2003, its website has published digital facsimiles, selected Early Modern imprints (1450-1650) mainly from regional collections, and transcriptions of French texts of the same period, encoded according to the XML-TEI standard. Particular attention is paid to achieving great accuracy in the bibliographical description as regards the true states of originals and the closest correspondence between two distinct corpora, facsimile and text, linked by several levels of metadata in the main catalogue. The BVH team works in close collaboration with researchers from the Computer Science Laboratory of Tours (LI-Tours) to develop new technologies in the fields of image processing and pattern recognition. Open source software for layout analysis and text transcriptions, AGORA and RETRO, enables us to perform automatic extraction of graphic components from digitized books, and thus to build up specialized databases of iconographic and typographical material. As a member of the TEI consortium, we actively contribute to the development of a specialised schema for the transcription of Renaissance documents. Each step of processing and every component developed at the BVH is also intended for use by the whole digital community, creating a model for the digital library of the future.

1- Introduction

The Bibliothèques Virtuelles Humanistes (BVH, or Virtual Humanistic Libraries) is a project run since 2002 by a research team in the Centre d'Études Supérieures de la Renaissance (CESR, or Center for Advanced Renaissance Studies) at the University of Tours, France. Founded in 1956, this institute is a multidisciplinary laboratory (UMR 7323) of the French National Centre for Scientific Research (CNRS), within the Department of Social Sciences and Humanities, recognised for its particular expertise in many aspects of Renaissance studies: literature, philosophy, history, history of the arts, musicology and book history. For example, catalogues of incunabula preserved in French public libraries – *Catalogues régionaux des incunables des bibliothèques publiques de France* – are currently published under the authority of the institute's emeritus professor.

Its goal is to develop a digital library (http://www.bvh.univ-tours.fr/) delivering two types of reliable digital representations, facsimile and electronic text, closely linked together, providing access to the original documents to a large public of researchers and non-specialists along with automatic tools for linguistic and historical research. Printed books and manuscripts of the XVth-XVIIth centuries, typical examples of intellectual production of the Renaissance period, are selected for digitization mainly from regional collections such as the public libraries and archives of Orléans, Tours, Poitiers, Blois, Châteaudun, Bourges, Châteauroux or Vendôme. Occasional one-off collaborations with more distant libraries in France or abroad are also carried out.

The digitization is performed at the institute by technical staff, using the institute's scanner specifically designed for the digitization of ancient books, or in partner libraries by their staff, or in a partner library *in situ* by contractors, in each case carefully following appropriate technical requirements, which are constantly kept up to date. After digitization, the BVH team processes all the image files, following an appropriate workflow.

Since 2009, the BVH website has included the Epistemon corpus of electronic editions of Renaissance texts in French, originally published online in 1998 at the University of Poitiers. In these editions, XVIth century French language texts are directly transcribed from their original editions and encoded with the XML-TEI standard.

Since its inception, the BVH program has worked to optimize these two distinct workflows and to automate each process, keeping in mind the double nature of the source documents as page-image and as text content and trying to represent image files and text files in an integrated way, preserving all available information. In this connexion, the BVH team is involved in several research projects in Digital Humanities, and also collaborates with computer science researchers as a data provider for several research projects concerned with the enhancement of Optical Character Recognition (Madonne, Navidomass, Impact). We are most closely linked with our colleagues from the computer science laboratory of the University of Tours (LI-Tours), in particular the RFAI team led by Jean-Yves Ramel, which specializes in image processing and pattern recognition, and the team led by Denis Maurel, which specializes in language processing.

The BVH program began its online publication of facsimiles in 2003, before many projects of mass digitization. It now offers 965 facsimiles and 47 encoded transcriptions, attracting a stream of more than 60,000 visitors a year. The BVH website includes many realizations, reflecting a decade of research, itself representing many aspects of the evolution of digital libraries; it also shows, in its procedures and newly developed applications, a promising outlook for the future.

2- Description

A. Before digitization: selection and description of books

The BVH is not a mass-digitization program such as Gallica or Google books, but a research "programme" endorsed by the CNRS. Supported partly by regional funding, its mission is to provide evidence about intellectual life during the Renaissance period at both regional and European levels. Texts are therefore selected according to explicit criteria:

- texts representative of Renaissance culture and of Humanism;
- documents representative of local history;
- important editions in the history of Classical or Medieval texts;
- rare editions (for instance, we consistently digitize *unica*);
- particular copies of specific interest (annotations, former ownership, binding).

→ See for instance, the file listing the books digitized for the campaign "Berry savant de la Renaissance" (Bourges, 2012): http://www.bvh.univ-tours.fr/actualites/12.07.13_numerisation_Berry_savant_2012.pdf

Our main task as a research project is to provide reliable data and to contribute qualified expertise about the books we reproduce. Before digitization, we compare different copies of one edition, and check their physical conditions. Each copy is collated page by page, in order to ensure the completeness of the volume: incomplete copies are excluded from our digitization process, except for copies of specific interest. In any case, the physical states of all copies digitized are described in the digitization record as precisely as possible. Before publication online, we control the quality of each digital image, with a book-in-the-hand comparison.

We compile detailed bibliographical records of the material state of the original document (binding, collation, signatures), following standard bibliographical recommendations (ISBD (a); AFNOR, Z 44-074). Since 2012, we have also indexed as exhaustively as possible the individuals involved in the preparation of a single edition (authors, translators, editors, dedicatees, preface writers, illustrators, etc.) and highlighted specific witnesses about the history of the copy (former owners' marks, manuscript annotations, etc.).

See for instance an edition gathering texts written by 35 authors: http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=875

B. Digitization process

Over the last ten years, the BVH has developed and evolved significantly. Regarding the digitization process itself, after some negative experiences (missed pages, poor image resolution), we now impose tight technical requirements on digitization contractors. First, we specify a comparatively slow rate of production and more detailed quality control (view by view). Paradoxically, this method has proved tremendously timesaving during post-production, it drastically reduces the number of post-production corrections. Secondly, being involved in image processing research, we require very high quality digital facsimiles. After some 300dpi grayscale photographs at the beginning of the project, we now take only color photographs, with a minimum resolution of 400dpi; or for some specific books or archives up to 600dpi. A graduation and color scale is added to each digitized book in order to ensure the reliability of the displayed resolution and colorimetry.

C. Several corpora, a single digital library

Each digitized book is downloadable in PDF format and also readable online. Online reading is facilitated by three types of browsing tools: page or folio numbers, structure in logical divisions (parts, chapters, etc.), and transcription of each division's title.

→ See for instance: http://www.bvh.univ-tours.fr/Consult/sommaire.asp?numtable=B372615206_5106&numfiche=358

Volumes combining several editions (bibliographical units) in a single binding are processed in a specific way: each edition gets its own bibliographical record, but browsing through the entire book from one publication to another is also facilitated by a link at the top of each record and of each consulted page. The composition of the whole collection is specified in its special descriptive note.

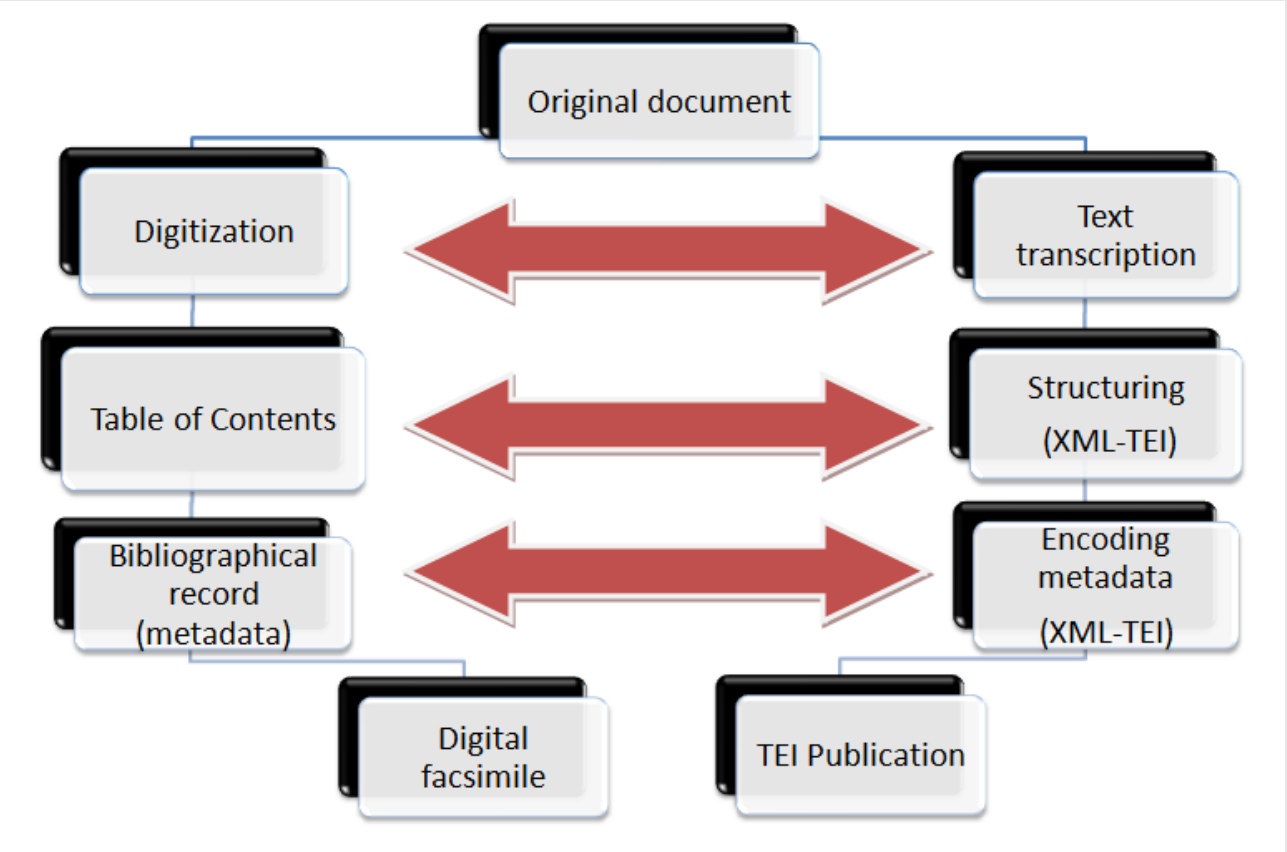
→ See for instance: www.bvh.univ-tours.fr/Consult/index.asp?numfiche=948

Some of our documents are available in both facsimile and electronic edition, encoded in XML-

TEI. In such cases, the digitized page and its transcription can be displayed in parallel.

→ See: http://goo.gl/6jgV6d

At present, the BVH program uses two different workflows for facsimiles and for transcriptions. In its main catalogue, metadata are established for each corpus independently: the accuracy of the results must be preserved. Our main concern is to ensure a close correspondence between these two workflows and to guarantee a reliable link between digital text, digital images and original document. In the long run, we plan to merge the two different workflows using a single XML schema.



D. Text transcriptions

Because of inherent spelling variations, French texts of the Renaissance period require very careful attention to establish a reliable textual resource. In our transcriptions, we first try to establish a representation as close as possible to the original imprint or manuscript applying strict rules to render written forms, punctuation, Ramist letters and brevigraphs. This first state is then regularized with a semi-automatic script. Once processed, the text is encoded with TEI P5 in such a way as to support dynamic display of the Epistemon corpus, either as an original transcription or in regularized form. An additional modernized level can also be developed in the same way. Handwritten interventions, restitutions, and corrections of inaccurate words are equally highlighted. Our TEI files also include elaborate metadata, name mark-up and rich structural details.

→ See the toggle display module: http://goo.gl/69GEv8

E. From image to text: image processing and OCR

No Optical Character Recognition software currently available is satisfactory for the printed works from the XVth-XVIIth centuries. Degradation due to aging, stains or repeated use, can all disrupt OCR systems. Moreover, non-standard fonts remain beyond the capabilities of traditional OCR systems as yet. Although our text is currently transcribed manually, we are also closely involved in image processing research. For more than ten years, the BVH team has collaborated with the Computer Science Laboratory of Tours (LI-Tours) in the development of alternative techniques to traditional OCR, focusing on the three following steps of a full processing chain: (i) layout analysis, (ii) specific content extraction and indexing, and (iii) text transcription.

This collaboration has led to the development of AGORA, software capable of analyzing the page layout of historical books. AGORA generates XML-Alto files localizing each element composing the page and can also extract pictures of characters. Another application, called RETRO, developed as a means of exploiting the output from AGORA, performs clustering of extracted patterns. Once the clustering is done, a user (or a computer) can assign a label and other metadata to each cluster. These labels are then automatically propagated to each pattern instance, thus achieving the indexing and transcribing of the whole of a book. In this way, if 90% of patterns are detected as redundant, i.e. only one character in ten needs to be identified by the user in order to transcribe the whole book. Currently under development, these two applications are open source and can be freely downloaded.

→ More details about this project at the following URL: https://sites.google.com/site/paradiitproject/

3- Impact

A. Preserving and promoting cultural heritage

The impact of the BVH program as regards the preservation and promotion of cultural heritage is self-evident.

Digital facsimiles make ancient books which are otherwise very difficult to access freely available for all people. In our digitization campaigns, we therefore try to select the rarest and the most precious books kept in local libraries, in order to enhance their visibility to a wide public. This policy has enabled us to digitize some real "treasures", such as a priceless copy of *Le Roman de Tristan* (Paris, 1496) printed on vellum and splendidly decorated with paintings and illuminations, preserved in Châteauroux.

→ See: http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=635

Our specialist knowledge of the details of each copy can contribute to an enhanced awareness of local intellectual history and society. On the one hand, we can facilitate the sharing of such knowledge with a broad public interested in Renaissance studies or local history. On the other hand, even though a digitized surrogate cannot replace the original document, it contributes to its preservation by diminishing the frequency of its consultation.

B. Improving research on text transcriptions using new technologies

Thanks to our sophisticated transcription processes and the use of XSLT transformations, our contents are downloadable in several different formats, including PDF, HTML and XML-TEI files. But our objective is not simply to display our transcriptions but also to support linguistic analysis by software such as PhiloLogic (developed by ARTFL project, Chicago) or TXM (textometry tool developed by ICAR, ENS-Lyon). It is important for us to keep an eye on emerging digital technology with the clear goal to offer dynamic display and to improve knowledge. Our latest realisation in collaboration with a language processing laboratory is the ReNom website. This provides navigation tools usable by the general public to explore geographical places and historical or fictional characters in the works of Renaissance authors such as Rabelais and Ronsard. This project offers a different perspective on name tagging, combining scientific objectives with a strong public appeal. The technology used is automatic cascade transformation, which recognizes, normalizes and marks up named entities, also possibly associating them with a cartographic visualisation.

→ See: http://renom.univ-tours.fr/en

C. Iconographic researches

The BVH program also provides some answers to specific requests concerning art history and book history. The AGORA layout analysis software can easily extract graphical contents such as illustrations, initial letters, portraits or ornamental strips, associating each element extracted with its specific metadata: location in the page (ALTO coordinates), dimensions in pixels and real dimensions in millimeters (calculated automatically from the pixel coordinates and displayed resolution). We have been performing automatic extraction of illustrations and ornaments from our digital facsimiles since 2006.

This process can lead to several outputs. We are now able to include the original illustrations into a TEI transcription.

→ See for instance: http://www.bvh.univ-tours.fr:8080/xtf/view?docId=tei/B410186201_I65/B410186201_I65_tei.xml;chunk.id=n10.14;toc.depth=1;toc.id=n10;brand=default

We also build up several specific databases devoted to iconographic studies. The first one is a database of the portraits extracted from our digital facsimiles.

→ http://www.bvh.univ-tours.fr/img_portrait.asp

The second one is a database of 3186 illustrations, correctly indexed with Iconclass, a multilingual thematic thesaurus.

→ http://www.bvh.univ-tours.fr/Iconclass_browse.asp

This extraction of images makes it possible to develop a new way of browsing facsimiles:

through the main consultation interface (chemin de fer), the user can access a specific page using these illustrations. From this page, the user can open a high-definition image by clicking on the picture or return to the digital facsimile (using the icon ); if the picture is indexed, the user can also read the Iconclass record (using the icon ).

→ See for instance: http://www.bvh.univ-tours.fr/Consult/imgcherche.asp?numtable=B360446201_INC6_1&mode=3&numfiche=635&position=6&ecran=0

D. A database for typographic studies

The database of ornamental letters we have been building up since 2006 will soon be replaced by a global database devoted to Renaissance typographic ornaments called BaTyR (Base de Typographie de la Renaissance, or Renaissance Typography Database). This is currently available as a beta-version, being still under development.

→ Browse the beta website: http://www.bvh.univ-tours.fr/batyr/beta/formulaire_bois.php

Historians wishing to identify the origin of some anonymous printed book, or trying to investigate the evolution of typography during the Renaissance can search BaTyR. It deals with a considerable challenge: the distinction between an ornament and its different appearances, in one or several books. It thus provides the opportunity of investigating the history of each ornament, by offering an inventory of its usages by different printers. BaTyR currently consists of 7000 distinct ornaments linked to more than 13000 items; the extraction and indexation processes are still in progress. The database can be searched through a global search form, or using specific indexes. For instance, an index of printers and booksellers provides access to a "publisher record", listing all the devices etc. used by a specified printer.

→ See, for instance, Jean de Tournes, printer in Lyon: http://www.bvh.univ-tours.fr/batyr/beta/notice_libraire.php?Libraire=5

BaTyR is a good example of the way in which technical innovations can encourage the development of digital humanities research. Our typographic expertise also allowed us to identify some anonymously printed books we had already digitized. We give here a few examples of the items we have assigned to printers according to a typographic analysis:

→ Pope Clemens VII, three indulgences printed in Bourges, ca. 1534:
http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=867
http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=868
http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=869

→ Euclides, *Elementorum geometricorum*, Basel, 1546 : http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=558

→ Le Saunyer, *Sommaire et brefve interpretation*, Paris, 1551 : http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=1013

→ G. Ruscelli, *Epistres des Princes*, Paris, 1572 : http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=517

→ J. de Coras, *In titulum Codicis Justiniani*, Lyon, 1550 : http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=163

→ J. de Rély, *Remonstrances faictes au Roy* , Paris, 1560 : http://www.bvh.univ-tours.fr/Consult/index.asp?numfiche=140

4- Extensibility

With its ten years of experience, the BVH program has acquired enough expertise in digitization and text/image processing to share it with other institutions or research programs. We frequently transfer our technical specifications model to local institutions for their digitization projects.

As we aim to solve problems shared by many digitization projects (for example, the best way of handling texts collected within a single binding), we hope some of our reflections and technical attempts will be useful for other digital libraries. Our aim for a greater accuracy of data and metadata about original source documents and our efforts to develop a global architecture ensuring close integration of text transcription and digital facsimile may be both a technical model and also a conceptual challenge for the digital libraries of the future.

We also wish to share some specialist tools developed in Tours. The two open source image processing applications AGORA and RETRO are freely downloadable, and many digitization programs could use them to develop their own iconographic databases.

A. Automated scripting

We are currently completing the development of eXtensible Stylesheet language (XSLT) scripts that generate a table containing foliation/paging and signature lists, on the basis of a book's detailed collation as recorded in its bibliographic description. This provides a framework for online browsing, and can be enriched (with structural information or title transcriptions...) to form a true table of contents. The same technology is used to control the accuracy of our collation formulas, checking the correspondence between foliation description and signatures list. Such a tool could be very useful for other digital libraries, and we would be glad to share it with institutional partners.

→ collation scripts: www.bvh.univ-tours.fr/form/collation.html

B. Encoding practices

The transcription of Renaissance texts and the structuring of their contents in XML-TEI format are not a simple task. As a member of the TEI consortium, the BVH team frequently reports some of the problems we encounter to other members of consortium, and contributes actively to debates about this technical specification. With a constant view to standardization of our practices, our goal is to customize the TEI consortium Recommendations to the specificities of Renaissance documents, and we have produced a handbook devoted to the encoding of Renaissance texts (first version in 2008; third version available since 2012, July 25). This handbook is currently only in French, but an English version is expected.

→ PDF: http://www.bvh.univ-tours.fr/XML-TEI/manuelTEIrenaissance3_2012.pdf
→ HTML: http://www.bvh.univ-tours.fr/XML-TEI/ManuelWeb/index.html

This on-line handbook provides access to our specific ODD ('One Document Does it all') a standardized documentation of all the elements and attributes used in our Renaissance-TEI customization.

→ Direct access to ODD: http://www.bvh.univ-tours.fr/XML-TEI/odd/odd.html

Our team frequently organizes and contributes to outreach activities concerning TEI encoding such as training workshops for PhD students, teachers, librarians, archivists or editors.

C. Image processing

In the long run, we hope that our research into enhanced OCR will benefit computer science and that we will make a contribution to the development of efficient software for the processing of ancient print materials. This remains a major challenge for the whole researcher community.

As of now, our image-extraction workflow is quite efficient. Current versions of AGORA and RETRO software are open source tools freely downloadable.

5- Other considerations

The Virtual Humanistic Libraries is a scientific project of the CESR (an institute depending both on the CNRS and the University of Tours). Its scientific design has been built up in collaboration with the IRHT (Institut de Recherche et d'Histoire des Textes, CNRS). Since this project is labelled by the CNRS as a scientific "program", its sustainability depends on the long-term political will of the French ministry of Research.

The BVH program has worked actively for the development of "digital humanities" and especially in the domain of "cultural heritage preservation" with its expertise both in text processing and in book history: for instance, the organisation in 2004 of the "Semaine du document numérique" (Digital Document Week) in La Rochelle and many participations to other scientific meetings.

1. Cultural heritage Network

Since its creation the BVH has collaborated with the National Library of France (BnF, or Bibliothèque nationale de France) Gallica project. In 2006, the CESR was appointed an associated pole ("pôle associé") of the BnF, specializing in digitization of imprints of the Renaissance period as a consequence of the BVH project. At the European level, the BVH is a member of the Europeana network. Its metadata is thus available on Gallica and Europeana via OAI-PMH exchanges.

Working with cultural institutions such as rare book collections of libraries and archives during our digitization activities, we naturally share our expertise in digitization and metadata organization with them, including, for example, VIAF for authority files, MARC-XML, Dublin Core, etc. to dialogue with different communities; reference catalogues such as USTC, ISTC, etc. for database management. Our membership in the European network (Europeana) has given us experience of the different visions characterising cultural heritage institutions and research projects.

Original copies of most BVH digitizations are kept in partner libraries and archives in the Region Centre of France. By publishing these regional materials online, the BVH website has assumed, in a sense, the role of digital aggregator of cultural heritage in this region of France for the Renaissance period. This regional function of the BVH can be extended by programs such as the project ReNom (renom.univ-tours.fr), a touristic application of facsimile and textual resources of authors like Rabelais or Ronsard with geolocation applications. Our network is also expanding at the international level with new major projects such as MONLOE ("Montaigne at Work"): http://www.bvh.univ-tours.fr/Montaigne_en.asp

In 2012, we have become a member of the Biblissima Equipex (Equipment of Excellence): http://www.biblissima-condorcet.fr/en.These projects are carried out in close collaboration between research teams and cultural institutions.

2. Dissemination policy

The original materials digitized by the BVH project, mainly old printed books of 1540-1650, are all in the public domain. As regards their digitization and diffusion online, however, an agreement is always signed between the CESR and the partner institution.

The BVH-Epistemon database contains unpublished complete texts or original transcripts which are distributed for free for private use, reading and research. These texts were transcribed by students or

researchers associated with academic institutions, or by providers under contract. The intellectual property and reproduction rights of organisms that have funded a step of the BVH workflow and copyright of the libraries that have authorized the reproduction of the original are protected. As for articles and searchable information in Epistemon, any reproduction for commercial purpose is subject to authorization. The contents of Epistemon are currently licensed for distribution under CC-BY-NC-ND creative commons license and will soon be available under CC-BY-NC-SA.

3. Research Network

The BVH work with national and European infrastructures at several levels in developing tools and data sets.
On the national level, the BVH are a member of the CAHIER consortium (*Corpus d'Auteurs pour les Humanités: Informatisation, Édition, Recherche* – Authors' Corpora for the Humanities: Digitisation, Edition, Research), created in 2011, for the TGIR Huma-Num (French part of the European DARIAH infrastructure: http://www.dariah.eu/). The BVH project helps to coordinate and to aggregate author corpora for research (http://www.cahier.paris-sorbonne.fr). Many literary projects, particularly in France, do not use TEI encoding, and scholarly corpuses seem to be specific to each project. Corpora of Rabelais and Montaigne are part of the "CAHIER" Consortium.
As a research data set provider, our repository has been harvested by Isidore, the portal for scientific research (CNRS), since its launch in 2011.

The BVH program contributes to the European DARIAH infrastructure in providing digital materials for tool development. The BVH-Epistemon database also provides textual resources to many projects in linguistics for their research works and projects. For example, the development of a modernization tool by the Forell team at the University of Poitiers, with a set of rules and specific dictionaries, benefited from one of the two Google awards that the University of Tours obtained in December 2010 for "Full-text retrieval and indexation for Early Modern French documents". Epistemon is already accessible using TXM (a textometry tool developed by ICAR, ENS-Lyon), using PhiloLogic (developed by ARTFL, Chicago), and using Analog (developed by Forell, University of Poitiers) among many others.
Our use of XML following the recommendations of the TEI provides the result of standardization that we have stabilized after years of experimentation with the TEI consortium in which we are an active member and contributor since 2007.
BVH material and expertise – both images and texts – contribute to the development of specialized and generic open source software in collaboration with our computer science partners (in Paris, Rouen, La Rochelle and of course in Tours by the RFAI team, our historical partner); some of them participated also in the IMPACT project (Improving access to text; http://www.impact-project.eu/).

We have been a member of the CenterNet of ADHO (the Alliance of the Digital Humanities Organizations) since 2009 (http://digitalhumanities.org/centernet/), and may soon become an active member of the francophone association for the DH currently under construction. We continue to contribute actively to the landscape of the digital humanities by attending international meetings.

4. Degree courses in Digital humanities

Created in 2004-2005 school year, the Masters degree "Written Heritage: History and practice of publishing" ("Patrimoine écrit: histoire et pratique de l'édition") directly benefited from the first "Digital Document Week" held in La Rochelle in June 2004.
The Master degree was renamed "Heritage writing and digital editing" in 2007 (PEEN, Patrimoine écrit et édition numérique) and closely associated with research in book history conducted by the BVH team: its main focus is professional training and contemporary issues related to the digital

humanities. Students are provided with an initiation to relevant economic and legal realities, particularly intellectual property issues, and can choose from a set of options: digital edition, musical publishing, book production and digital humanities. Training for this last option is provided by the BVH team and its professional network and skills. Each year, a TEI workshop is organized and a "Day of digital professions" brings together key actors in digital publishing (see the program for this year:
http://cesr.univ-tours.fr/actualites/journee-des-professionnels-les-metiers-du-numerique-375710.kjsp?RH=CESR_FR)

In conclusion, if other projects provide data with friendly display and functionalities, our project offers also reusable sets of transcriptions, facsimiles and new tools for further analysis, supported by a strong collaborative network, with an eye firmly fixed on the future.

Visit our virtual library at www.bvh.univ-tours.fr

CESR BVH
Marie-Luce Demonet (Head)
Lauranne Bertrand, Sandrine Breuil, Jorge Fins, Rémi Jimenes, Toshinori Uetani,
Anne-Laure Allain, Christine Bénévent, Elise Gauthier, Anne Guérineau, Myriam Olivier,
Marie Olivron, Laetitia Bontemps, Charlotte Van der Werf

LI RFAI
Jean-Yves Ramel (Head)
Frédéric Rayar
Pascal Bourquin
Nicolas Ragot

With thanks to all our collaborators, past and present:
http://www.bvh.univ-tours.fr/presentation.asp#presentation_equipe

With thanks to Lou Burnard for help in translating this document.