

ECP 2005 CULT 038097/Bernstein

Bernstein

Annual Report

September 2006 – August 2007

Bernstein – The Memory of Papers

<http://www.bernstein.oeaw.ac.at>

Deliverable number/name	<i>D-7.3 First annual report</i>
Dissemination level	<i>Public</i>
Delivery date	<i>27 September 2007</i>
Status	<i>Draft/Final</i>
Author(s)	<i>The Bernstein Consortium</i>



eContentplus

This project is funded under the *eContentplus* programme¹,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

¹ OJ L 79, 24.3.2005, p. 1.

1 Table of content

1 TABLE OF CONTENT 2

2 PROJECT OBJECTIVES 3

3 CONSORTIUM..... 3

4 PROJECT RESULTS/ACHIEVEMENTS..... 4

5 TARGET USERS & THEIR NEEDS 4

6 UNDERLYING CONTENT 6

7 SUMMARY OF ACTIVITIES..... 7

8 IMPACT & SUSTAINABILITY 9

9 FURTHER INFORMATION..... 10

2 Project Objectives

The present project revolves around paper. The significance of paper derives from being the ubiquitous physical support for information exchange until well into the present times. As such, the study of paper is used to identify undated paper documents or expertise documents of questioned authenticity. Paper studies have also a historical dimension, revealing aspects of technological evolutions, economical infrastructures, state policies, interwoven into human networks across countries. Knowledge of this kind is partly obtained from the physical characteristics of papers, a source of “hidden information” as opposed to the visible inscription of a paper object. The watermarks that are visible when holding a bank note against light are an example. They are also the most prominent characteristics examined by historians, reproduced and documented in catalogues. A great amount of paper data with a broad geographical and temporal spread is necessary in order to get this information.

The goal of project *Bernstein* is the creation of a European integrated digital environment about paper history and knowledge. The project will connect all presently existing European watermark databases and thus offer a comprehensive and significant information source about paper. The databases will be augmented by specialized image processing tools for measuring, authenticating and dating papers and by a plenitude of contextual data with bibliographical and geographical (GIS) contents. A substantial further project goal is the dissemination of the achieved results to the broad audience in the form of an easily installable software package.

3 Consortium

The consortium brings together all the major European actors in the field of digital historical paper expertise (hence the partner choices) coming from both humanities and computer sciences. The project consortium consists of 9 partners from 6 countries, among which the largest data collections about paper and watermarks are found. The *Bernstein* consortium includes:

1. Austrian Academy of Sciences, Vienna, Austria (OEAW, <http://www.oeaw.ac.at>), Commission for Paleography and Codicology of Medieval Manuscripts in Austria (<http://www.oeaw.ac.at/ksbm>), Commission for Scientific Visualization (<http://www.viskom.oeaw.ac.at>), Management, database of mediaeval watermarks, digital repertories, image processing tools,
2. Archives of the State of Baden-Württemberg, Stuttgart, Germany (LABW, <http://www.landesarchiv-bw.de>), huge collection of watermarks (Piccard, 95000),
3. Graz University of Technology, Institute for Information Systems and Computer Media, Austria (TUG, <http://www.iicm.edu/>), integration software (implementation), user interface,
4. Laboratory for Occidental Medieval Studies in Paris, France (LAMOP, <http://lamop.univ-paris1.fr>), historical GIS,
5. German National Library, Deutsches Buch- und Schriftmuseum, Leipzig, Germany (DNB, <http://www.d-nb.de/>), bibliography on paper, huge paper collection,
6. Dutch University Institute for Art History Florence, Italy (NIKI, <http://www.iuoart.org/>), paper database, art historical expertise,
7. Delft University of Technology, Information and Communication Theory Group, Netherlands (DUT, <http://tudelft.nl>), data mining software, intelligent image processing,
8. Koninklijke Bibliotheek, National Library of the Netherlands, The Hague, Netherlands

- (KB, <http://www.kb.nl>), watermarks in Dutch incunabulae (WILC) and
- University Liverpool, Great Britain (LU, <http://www.liv.ac.uk/>), integration architecture, bibliography integration.

4 Project Results/Achievements

The **integrated workspace** is the backbone of the project that provides the digital environment necessary for the integration of resources. Specifically this is an internet application that allows the access to all resources. It was a milestone of Bernstein when the first version of the integrated workspace was presented in Fabriano (Italy) in July 2007. The first proposal for the design of the integrated workspace (mock-ups of user interface) is available online under <http://bernstein.iicm.tugraz.at/bernstein/>. This proposal will be the basis for the discussion about the final functionality and design of the integrated workspace. The main structure of the workspace consists of the navigation panel on top of the page and the search overview section on the top left of the workspace. These main parts are always present guaranteeing simple navigation through the different sections of the page. The navigation panel consists of the seven main parts of the site: Start page (“About Bernstein”), “Advanced Search”, “Specific Search”, “Component Search”, “Browse Motif”, “Bibliographic Search” and “Help Section”. Above the panel there are three language change buttons for English, French and German symbolised by flags.

An essential step for the integration of the four watermark databases (Piccard-Online, WILC, WZMA, NIKI – ordered according to the number of records in each) is the adoption of a first version of **watermark standard**. The major work to do was the comparison of Piccard-Online (<http://www.piccard-online.de>) and WILC (<http://watermark.kb.nl>). The result is a trilingual (English, French, German) thesaurus of watermark descriptive terms. The standard includes a classification scheme for the hierarchically organized watermark types and a nomenclature giving a name to each type. The future potential of this watermark standard is obvious for it is useful and extensible to other collections of watermarks beyond this project’s lifetime. Watermarks of the same type can now be described trilingually by the same names and data interoperability and search across all databases is now possible. As a by-product a concordance between the motives represented in Piccard-Online and those corresponding in Briquet’s watermark collection was produced (“Briquet” is a yet not digitally available watermark collection widely used for paper studies). This concordance could help to complete the future linking of a digitized Briquet to the other data-bases within the portal.

The goal of the **Bernstein Historical Geographical Information System** (BH-GIS) is the representation across space and time of the distribution of watermarks and other paper features stored in the Bernstein databases. Additionally, Bernstein should be able to provide bibliographical data incorporated in the maps. The focus on this task is in the second project year. The first step towards the BH-GIS is the generation of a list of places occurring in the Bernstein databases matched to geographical coordinates. The work accomplished so far consists in a matched list of place names, coordinates and additional metadata. The processed geographical data belong to Piccard-Online, Briquet printed catalogue of watermarks and paper, Incunabula Short-Title Catalogue, Gesamtkatalog der Wiegendrucke (GW, <http://www.gesamtkatalogderwiegendrucke.de/>) and the International Paper Bibliography of partner DNB. At the time of this report this represents 16.000 records, equivalent to approximate 5.000 individual localities. Contacts with The Consortium of European Research Libraries (CERL, <http://www.cerl.org/>) were established in order to investigate cooperation possibilities. CERL maintains a database of historical variants of place names. Discussions

are underway on providing CERL with the place coordinates in exchange for variants of place names.

The **cross-link between the “Watermarks in Incunabula printed in the Low Countries” (WILC) and the “Incunabula Short-Title Catalogue” (ISTC, <http://www.bl.uk/catalogues/istc/>)** has been implemented and tested. WILC is the consortium’s database on papers of Dutch early printed books and ISTC is the database of the British Library developed in collaboration with LU, providing information on the content and physical aspects of virtually all known incunabula editions in the world (29.224 editions). The linking consists in Digital Object Identifier (DOI) hyperlinks representing Incunabula of the Low Countries numbers, a metadata information already present in WILC and ISTC (in Marc21 format), which status as standard for incunabula studies will ensure persistency of interoperability.

5 Target Users & their Needs

The project is characterized by a diversified source of demand for the functionalities we offer. Hereafter we describe who the user communities are, what are their specific needs and what is the critical level expected to satisfy these requirements.

1. **Cultural demand / Historians** – The original initiative for the project came from historians who wish to study the culture of Europe at the onset of the Renaissance by mean of paper studies. These aspects to all of which our project brings solutions are of foremost importance to the successful work of historians: identification of date and place of origin of papers based on objective measurements; a statistical and historical cartography of paper features in order to map over space and time paper technology and trade evolutions; and finally the capability to think about paper history in the broader terms of European cultures and changes in societies.

2. **Economic demand / Curators and Industry** – For art dealers the ability to correctly authenticate, date and locate paper documents is the basis itself on which their business relies. The same is true for curators of public collections (libraries, archives, museums), who need to know the value of objects they possess or wish to acquire or sell.

3. **Societal demand / Forensic experts** – Paper documents such as passports are the most widely used identification documents for individuals across the world. The ability to quickly and surely identify fakes can prevent crimes and law infringements. Paper expertise by police agencies is done however not at the place where the paper documents are presented (borders, police checks, banks ...), but upstream in the chain of criminal expertise, by experts in laboratories. They are mostly relying on their visual experience and less on machines. There is clear need for image processing software for paper analysis and networked databases to authenticate the documents.

4. **Industrial demand / Papermakers** – Working with curators, paper manufacturers (artisans and big companies) provide specialized papers reproducing ancient characteristics for the need of restoration of damaged cultural goods. From an art-historical point of view it is important that their products match closely the old models and that confusions should be avoided about where and when a particular type of paper did occur so as not to produce chimerical restored objects. Therefore paper measurement and knowledge of paper history & cartography is critical to papermakers.

6 Underlying Content

Our resources consist of content (images, metadata and contextual resources) and content processing software (for image processing measurements, data mining, statistics and cartography).

1. **Images** – The images are reproductions of the physical structure of papers through several techniques: radiography, backlight, rubbing and tracing. Several features are made visible: the sieve of the paper mould (watermarks, chain and laid lines), parts of the wooden frame and the paper pulp distribution. The measurements of the variation of each of these elements make a unique identifier for each paper sheet and mould and more generally for the paper mill and the know-how of a region or time period. Thus the identification of the origins of papers can be established and by comparison of watermarks (which were peculiar to individual papermakers and replaced every 2 to 4 years), paper documents before the 17th century (in the age of manually produced papers) can be dated with a precision of ± 9 month (by intersection of watermark date ranges from the several paper batches that usually compose a manuscript, book or newspaper).

Quantity: In total we have 120.500 images distributed among four on-line databases: LABW, Germany (95.000), KB, Netherlands (16.000), OEAW, Austria (8.000) and NIKI, Italy (1.500). This covers the quasi totality of digital primary resources on paper studies actually available in the world (2005).

Quality: Our collections represent the reference material for historical studies on paper and watermark expertise. LABW provides the entire “Piccard” repertory of watermark tracings, a monumental work that in print version spans over 25 volumes and covers the watermarks used in the whole Occidental Europe, from the 13th to the 19th century. Although other watermark repertories exist, none is equal in importance to Piccard and no other is digitized or susceptible to be in the near future. KB bring its own impressive collection which records by electron radiography and rubbing paper types of all incunabula printed in the Low Countries (the first books produced after the invention of print with movable types by Gutenberg). OEAW’s (KSBM) collection is yet of another area, recording by X-rays Austrian manuscripts of the late Middle Ages. NIKI’s special contribution to the project is in outstanding reproductions of art drawings and prints of such key figures of European culture as Leonardo da Vinci and Rembrandt.

2. **Metadata** – The images are backed in the databases by metadata providing information first of all about the measured characteristics of the paper and classification of the watermarks. Next there is information about the date and place of production of the reproduced papers and about the documents for which they are used (for example for books, the title, author, publisher name and date).

3. **Contextual resources** – Besides content pertaining strictly to papers, we provide contextual data that helps advance the study of papers. This is provided by DNB in form of the most recent and complete bibliography on paper studies ever published. The paper data & bibliography integration of the last two volumes of the work will be part of our projects achievements. We mention also another contextual resource, that is not owned by our partners, but for which we will provide interoperability: the Incunabula Short-Title Catalogue (ISTC) of the British Library, the complete on-line catalogue of all incunabula printed in Europe (29.244 editions).

4. *Content processing software* – The image processing software for the expertise infrastructure is provided by OEAW and DUT. These are: [1] pre-processing software to improve image quality for human and machine examination (BlueNile for image filtering in the frequency domain) or remove ink traces from backlight reproductions and [2] measurement tools for chain and laid lines density measurements (AD751). [3] Textual data-mining will be provided by DUT and LU as part of the integrated software. [4] LAMOP contributes with its tools for historical cartography (developed in MATLAB), aided for the statistical aspects by the upgrading of the databases to be undertaken by their owners. LAMOP is pioneering historical cartography for paper studies.

7 Summary of Activities

This chapter presents the principal activities of the Bernstein Consortium for the first year of the project. For a detailed view please consult the half year progress reports (deliverables no. 6 D7.1 and nr. 12 D7.2, confidential documents). Also mentioned here are the next steps scheduled for year two for each activity.

1. Data and data integration

Databases – The core of the project is the integration of several databases. The first necessary step undertaken was the harmonization of the terminology used to describe the data (paper characteristics) across the databases. Next, the hierarchical classification schemes of watermark motifs had to be addressed (“bull’s head with horns” and “bull’s head without horns” are both watermark motifs catalogued as part of the class “bull’s head”) and ensure that they are compatible (one database might consider “bull’s head” as an independent, top of the hierarchy, class, while in another it might be subsumed in “animals”). The terminology was also translated and is at present trilingual (English, French, German). The harmonization of the data and data structure took a lot of discussions between historians, the computer people being progressively drawn in.

This activity area will continue with work on the harmonisation of data units and formats, and filling the gaps of data types that exist in only some of the databases to be integrated (the height and width of a watermark for example).

Integrated workspace – Data and tools integration work concerned the integrated workspace, a Web interface that through the underlying software the project is developing will allow users to explore the data in the interconnected databases and apply on it a number of data processing tools such as image processing and cartography tools. Discussions between the development teams and data owners lead to the exhaustive listing of the use cases to be supported by the workspace and the adoption of a software-architecture for the workspace. On that basis a preliminary version was produced and submitted for review to the partners. Concomitantly an alternative scheme, called “component model”, for searching the databases and doing historical research on their content was proposed, discussed, elaborated, implemented into software and tested – decision was taken to make part of the data searchable with this method. The next integration steps will be the implementation of the various requirements as they have emerged in the first year into working software.

Repertories – There exist printed repertories of watermark descriptions that can enhance the coverage of the data offered by the project and the quality of expertise, like document dating,

of the tools under development. Such repertoires were digitized during the first project year, a work – of limited time spending it must be underlined – that will continue.

2. Data processing tools

Image processing – As the core data of the project are images reproducing paper features (backlight images of a papers for example), it is important that the users is provided with image processing capabilities to be able to measure those features in the images. On top of the existent tools provided by the partners, new ones were developed (chain lines measurement). Additional new products served the partners for preparing the data in the databases for integration (ex.: automatic segmentation of hand drawings into watermark, chain and laid lines). Next, work will be done on implementing user feed-back and finalizing the products.

Bibliography – The International Bibliography of Paper History allows access to more than 20.000 bibliographical records – now partially accessible through the Bernstein workspace (<http://dnb.cheshire3.org>). More to come in year two of the project!

Cartography – Since part of the paper documents described in the Bernstein databases are dated and localized and some others can be identified on the basis of the first, their spatio-temporal distribution can be visualized in form of maps. The first stage in this direction was geoparsing – extraction of place names from the databases, their identification in space and attribution of coordinates. Further work will concern the cartography software integrated to the workspace.

Incunabula link – A link was established between the WILC database that contains paper descriptions of incunabula, books of the first 50 years of printing in Europe, and the Incunabula Short Title Catalogue database of the British Library containing incunabula content, physical aspects and location information. Users of both resources can now easily gain more knowledge on the subject. Work under way is adding metadata to the other Bernstein databases to allow them too to be linked to the ISTC.

3. Dissemination

General public – Apart from reports in national and local newspapers about the project, a poster and flyer, its activities were brought to the general public mainly through a travelling exhibition held in four cities, in Germany, Austria and Italy. The exhibition focuses on watermarks, their history and usefulness, and comes along with a catalogue. Workshops on paper reproduction and digital tools for paper studies were held in Austria, England and Egypt. We expect to continue offering workshops and hold exhibitions in other countries, as well as prepare a technical and scientific book on the project and paper studies and present the exhibition through offering it its own website.

Scientific media – Project participants have contributed scientific papers to specialized journals in Humanities and Sciences, as well as participated in congresses. See list at <http://www.bernstein.oeaw.ac.at/twiki/bin/view/Main/ProductsDissemination>.

Contacts – The project has enabled the establishment of numerous contacts between institutions and scientists. Most notable are those with newly emerged paper database owners (two of them in Italy who became hosts to the travelling exhibition and the annual general meeting of the Consortium) or organisations seeking digital interlinking with the Bernstein integrated workspace (for the bibliographical references of The European Library, for the Historical Geographical Information System of the Electronic Cultural Atlas Initiative, or for the place names of the Consortium of European Research Libraries). These and new contacts

will be developed in the future for interconnecting Bernstein with other digital networks in new projects.

4. Management

Main aspects of the project management were meetings between partners and with contacts outside the project. This was partly supported by electronic communication means: a public project webpage, a development website open to the partners and a mailing list.

8 Impact & Sustainability

1. **Impact** – The project is expected to have a considerable impact on paper studies, its “target market”, by facilitating and bringing innovation in the field and broadening and synergising the market.

Facilitating role. Obviously the creation of digital resources (the paper databases, image processing software and geographical and bibliographical contextual resources) will facilitate a work until now done “manually”, improving speed and breath of research (in excess of 100.000 records accessible with one query).

Innovation role. By integrating content (the paper databases) with content processing tools (the image processing, cartography and bibliography software) existing data can be used in ways not possible before (for example the ability of measuring paper features provided by image processing and the statistical information generated by paper databases make the dating of paper documents possible).

Broadening and synergizing role. The individual data resources, software and know-how having converged into the project were initially intended for specific user communities (databases for historians, image processing for experts, integration-capacity at partner TUG for applications outside the Humanities). The project has to enable each one to broaden its reach and become valuable for new users. Conversely the polyvalence of the projects products has the potential to create synergies between user communities.

2. **Sustainability** – The lasting and evolving of the projects achievements is based on the credit of the consortium partners and spin-in & spin-off effects.

Credit. The partners that will be the physical holders of the projects digital products are in their respective countries institutions of national importance (national libraries, academies...). This is one way how the consortium is ensuring the continuity of its work.

Spin-in. The project has already created an effect of “spin-in” by which new parties (such as database holders) are interested in joining the effort. While some collaboration takes part in the framework of the present project, others need a substantial allocation of resources and thus will make the object of new projects (for example integrating 31 non-Bernstein paper databases created or identified during the project). This is only possible if the project’s products are maintained alive.

Spin-out. In the context of the project “spin-out” activities refer to contacts between the primary products users (paper historians and experts) and users from untargeted fields. For example the linking of a watermark database (WILC of the Koninklijke Bibliotheek) and an incunabula database (ISTC of the British Library) is a by-product seen as the first step in integrating Bernstein with other European cultural assets such as The European Library

(<http://www.theeuropeanlibrary.org>). Again, these development opportunities are also a guarantee for the lasting of the Bernstein endeavour.

Migration. There will come a time after the end of the project when the software used to provide the services developed now will be obsolete. To maintain these services migration of data and software replication will be necessary. To ensure that the project partners store the data in non-proprietary, unencrypted, well documented and human readable formats (XML for the KB) or use database software that can export data to such format (MySQL and similar software for the other partners). As for the software the solution adopted was to make the source code publicly available under Open Source licences.

9 Further Information

In addition to the information provided in this report the reader might find of interest to consult the following resources:

- The Bernstein website (<http://www.bernstein.oeaw.ac.at>) which is the Internet “face” of the project and gives access to its “body”, the integrated workspace.
- The TWiki development platform (<http://www.bernstein.oeaw.ac.at/twiki>) serving as a common blackboard and document repository for the projects partners.
- A visit to the itinerant exhibition (order of the catalog and check locations at <http://www.bernstein.oeaw.ac.at/twiki/bin/view/Main/ProjectExhibitions>) where one can learn more about watermarks

Abbreviations

BH-GIS	Bernstein Historical Geographical Information System
BL	The British Library, London, United Kingdom
CERL	The Consortium of European Research Libraries
DNB	Deutsche Nationalbibliothek, Leipzig, Germany
DUT	Delft University of Technology, Delft, Netherlands
GIS	Geographical Information System
GW	Gesamtkatalog der Wiegendrucke (State Library, Berlin)
KB	Koninklijke Bibliotheek, The Hague, Netherlands
KSBM	Commission for Paleography and Codicology of Medieval Manuscripts, OEAW
IPB	International Paper Bibliography
ISTC	Incunabula Short Title Catalogue (by BL)
LABW	Archives of the State of Baden-Württemberg, Stuttgart, Germany
LAMOP	Laboratory for Occidental Medieval Studies in Paris, Paris, France
LU	Liverpool University, Liverpool, United Kingdom
NIKI	Dutch University Institute for Art History Florence, Florence, Italy
OEAW	Austrian Academy of Sciences, Vienna, Austria
TUG	Technical University Graz, Graz, Austria
VISKOM	Commission for Scientific Visualization, OEAW
WILC	Watermarks in Incunabula printed in the Low Countries (by KB)
WZMA	Watermarks of the Middle Ages (by OEAW (KSBM))