



eContentPlus

European Commission
Information Society and Media Directorate-General

ECP-2005-CULT-038097/Bernstein
<http://www.bernstein.oeaw.ac.at>

Deliverable no. 10, ref. 2.2, GIS data batch no. 1

Deliverable number	D2.2
Dissemination level	Public
Delivery date	30 July 2007
Status	Final
Author(s)	The Bernstein Consortium

This project is funded under the *eContentplus* programme (OJ L 79, 24.3.2005, p. 1), a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

1. Background

The goal of the Historical Geographical Information System of Project Bernstein (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 first step towards the GIS is the generation of a list of places occurring in the Bernstein databases matched to geographical coordinates (geo-list). This is the content of the deliverable D2.2 'GIS data batch no. 1' presented here. The webpage on GIS aspects on the Bernstein development site is <http://www.bernstein.oeaw.ac.at/twiki/pub/Main/PaperCartography/>.

2. Work accomplished

The work accomplished so far consists in a matched list of place names, coordinates and additional metadata. The list is explained hereafter and the process and issues generating the list described further down.

2.1 Data sources

The processed geographical data belong to the following sources:

- A. Piccard Findbuch watermarks and paper printed catalog (PF);
- B. Briquet printed catalog of watermarks and paper (BPC);
- C. Incunabula Short-Title Catalog, quasi complete (short) list of existent incunabula (ISTC, <http://www.bl.uk/catalogues/istc/>);
- D. Gesamtkatalog der Wiegendrucke, quasi complete (extended) list of existent incunabula (GW, <http://www.gesamtkatalogderwiegendrucke.de/>);
- E. International Paper Bibliography (IPB, <http://www.ddb.de/sammlungen/pdf/ibp.pdf>).

This represents 8.000 records, equivalent to approximate 3.000 individual localities.

2.2 Description of the geo-list

The core information provided by the geo-list are names of places as they will appear on the Bernstein maps and their respective geographical coordinates. Additional geographical information is given about variant names of the places and various levels of administrative units to which the places presently belong. The metadata identifies the source where the place names were found and records wherever an ambiguity surrounding the exact location which is meant by the source was found. Also given is information about wherever the place name given by the source refers to the place where the paper was produced (paper mill), where it was used (printing house, scriptorium), or where it is currently conserved (library, archive...).

The next two pages show the beginning of the geo-list. A digital version can be downloaded from http://www.bernstein.oeaw.ac.at/twiki/pub/Main/PaperCartography/BH-GIS_data_sample.xls. The fields read as follows:

- A. *Source id nr.* – Number of the record in the source from where the place name was found.
- B. *Source* – Short name of the source from where the place name was found.
- C. *Information type* – Records if the place name is that where the paper was produced, where it was used, or where it is currently conserved.
- D. *Accepted from GW* – Records if coordinates could be attributed to the place name found in source GW.
- E. *Accepted from PF* – Records if coordinates could be attributed to the place name found in source Piccard.
- F. *Accepted from BPC* – Records if coordinates could be attributed to the place name found in source Briquet.
- G. *Acceptance status* – Records if the person processing the record has found the place name to be ambiguous (ex.: several places with the same name) or couldn't locate it geographically.
- H. *Name in the source* – Name of the place as given in the source.
- I. *Normalized name* – When several names are available for the same location, one name is chosen to be displayed on the Bernstein maps.
- J. *Latitude* – Decimal latitude coordinate of the place.
- K. *Longitude* – Decimal longitude coordinate of the place.
- L. *Country* – Country in which the place is located at the present time.
- M. *Admin region 1* – 'First order' administrative unit (such as 'Land' in Germany) in which the place is located at the present time.
- N. *Admin region 2* – 'Second order' administrative unit (such as 'county' in England) in which the place is located at the present time.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	Bernstein Geographical Data on Watermarks and Paper features (Sample)														
2	Source														
3	Accepted from:														
4	id nr.	Source	Place type	GW	PF	BPC	status	Name in the source	Normalized name	Lat	Lon	Country			Admin region 2
5	1	GW	incunabula location	y				's-Gravenhage	's-Gravenhage	51,52	5,92	NL	Zuid-Holland		
6	953	GW	edition place	y				Den Haag	's-Gravenhage	51,52	5,92	NL	Zuid-Holland		
7	2260	GW	incunabula location	y				's-Heerenberg	s-Heerenberg	51,88	6,25	NL	Gelderland		
8	2	GW	incunabula location	y				's-Hertogenbosch	's-Hertogenbosch	51,70	5,32	NL	Noord-Brabant		
9	4578	Piccard Findbuch	place of paper use		n			Aach	ambiguous (4)						
10	3	GW	incunabula location	y				Aachen	Aachen	50,46	6,06	D	Nordrhein-Westfalen	Köln	
11	4580	Piccard Findbuch	place of paper use	y				Aalen	Aalen	48,83	10,10	D	Baden-Württemberg	Stuttgart	
12	4	GW	incunabula location	y				Aalst	Aalst	50,56	4,02	B	Oost-Vlaanderen		
13	6	GW	incunabula location	y		y		Aarau	Aarau	47,38	8,05	CH	Aargau	Aargau	
14	4582	Piccard Findbuch	place of paper use	y		y		Aardenburg	Aardenburg	51,17	3,27	NL	Zeeland	Zeeland	
15	7	GW	incunabula location	y				Aarhus	Aarhus	56,15	10,22	DK	Arhus Amt		
16	4583	Piccard Findbuch	place of paper use	y				Abbach	Abbach	48,93	12,05	D	Bayern	Oberpfalz	
17	3104	Briquet	place of paper use		n			Abbaye de Bonne-Espérance	Vellerelle-les-Brayeux	50,38	4,13	B	Hainaut		
18	4584	Piccard Findbuch	place of paper use	y				Abbaye de Lyrn	Abbaye de Saint-Jacques	50,38	5,34	B	Liège		
19	3105	Briquet	place of paper use		n			Abbeville	Abbeville	50,06	1,50	F	Picardie	Somme	
20	9	GW	incunabula location	y				Abdinghof							
21	3107	Briquet	place of paper use		n			Aberdeen	Aberdeen	57,13	-2,10	UK			
22	10	GW	incunabula location	y				Aberystwyth	Aberystwyth	52,42	-4,07	UK			
23	11	GW	incunabula location	y				Ablate	Abbiategrasso	45,40	8,90	I	Lombardia	Milano	
24	4585	Piccard Findbuch	place of paper use	y				Åbo	Turku	60,45	22,28	FI	Lansi-Suomen		
25	2522	GW	incunabula location	y				Åbo							
26	4586	Piccard Findbuch	place of paper use		n			Åbo	ambiguous (2)						
27	3108	Briquet	place of paper use		n			Abstberg	not identified						
28	3109	Briquet	place of paper use		y			Acey	Vitreux	47,14	5,41	F	Franche-Comté	Jura	
29	12	GW	incunabula location	y				Achel	Achel	51,15	5,29	B	Limburg		
30	4587	Piccard Findbuch	place of paper use	y				Achern	Achern	48,63	8,07	D	Baden-Württemberg	Freiburg	
31	13	GW	incunabula location	y				Acireale	Acireale	37,62	15,17	I	Sicilia	Catania	
32	4588	Piccard Findbuch	place of paper use		n			Acken	Acquapendente	42,73	11,87	I	Lazio	Viterbo	
33	14	GW	incunabula location	y				Acquapendente	Acquapendente	in area	in area	AU			
34	15	GW	incunabula location	y				Adelaide	Adelaide						
35	4589	Piccard Findbuch	place of paper use		n			Adelberg	Adelmannsfelden	48,57	10,01	D	Baden-Württemberg	Stuttgart	
36	4590	Piccard Findbuch	place of paper use	y				Adelmannsfelden	Adelmannsfelden						
37	4591	Piccard Findbuch	place of paper use		n			Adelsbach	Adelsbach						
38	4592	Piccard Findbuch	place of paper use		y			Adelsheim	Adelsheim	49,40	9,39	D	Baden-Württemberg	Karlsruhe	

		A	B	C	D	E	F	G	H	I	J	K	L	M	N			
Bernstein Geographical Data on Watermarks and Paper features (Sample)																		
Accepted from:																		
id nr.	Source	Place type	GW			PF			BPC			status			Acceptance			
			GW	PF	BPC	status	GW	PF	BPC	status	GW	PF	BPC	status	Name in the source	Normalized name	Lat	Lon
38	16 GW	incunabula location	y						Admont	Admont	47,34	14,27	AT	Steiermark				
39	17 GW	incunabula location	y					ambiguous (3)	Adria	Adria								
40	3111 Briquet	place of paper use						ambiguous (3)	Adria	Adria								
41	4593 Piccard Findbuch	place of paper use	y				n	ambiguous (3)	Aelst	Aelst	50,46	4,52	B	Vlaams-Brabant	Vlaams-Brabant			
42	3112 Briquet	place of paper use						not identified	Aerdenbourg	Aerdenbourg	51,17	3,27	NL	Zeeland	Zeeland			
43	3113 Briquet	place of paper use						not identified	Aeren	Aeren								
44	3114 Briquet	place of paper use						not identified	Aeren	Aeren	46,40	8,15	CH	Valais				
45	4594 Piccard Findbuch	place of paper use					n	not identified	Afflitz	Afflitz								
46	18 GW	incunabula location	y					ambiguous (3)	Affligem	Affligem	50,50	4,20	B	Vlaams-Brabant				
47	19 GW	incunabula location	y					ambiguous (3)	Affligem	Affligem	50,50	4,20	B	Vlaams-Brabant				
48	20 GW	incunabula location	y				y		Agen	Agen	44,22	0,63	F	Aquitaine	Lot-et-Garonne			
49	21 GW	incunabula location	y						Agira	Agira	37,65	14,52	I	Sicilia	Enna			
50	22 GW	incunabula location	y						Agnone	Agnone	41,80	14,37	I	Molise	Isernia			
51	2772 GW	incunabula location	y						Agram	Agram	45,80	16,00	CR	Grad Zagreb				
52	23 GW	incunabula location	y						Agrigento	Agrigento	37,32	13,57	I	Sicilia	Agrigento			
53	4596 Piccard Findbuch	place of paper use					n	not identified	Ahanau	Ahanau								
54	4597 Piccard Findbuch	place of paper use					n	ambiguous (6)	Ahausen	Ahausen								
55	4598 Piccard Findbuch	place of paper use					n	not identified	Ahirch	Ahirch								
56	4599 Piccard Findbuch	place of paper use					n	ambiguous (3)	Ahlen	Ahlen								
57	4600 Piccard Findbuch	place of paper use	y						Ahus	Ahus	55,92	14,28	SV	Kristianstads Lan	Linderödsåsen			
58	4601 Piccard Findbuch	place of paper use					n	ambiguous (31)	Aich	Aich								
59	4602 Piccard Findbuch	place of paper use					n	ambiguous (2)	Aichach	Aichach								
60	4603 Piccard Findbuch	place of paper use	y						Aichbach	Aichbach	48,10	15,27	AT	NiederAT	NiederAT			
61	4604 Piccard Findbuch	place of paper use					n	ambiguous (3)	Aichelberg	Aichelberg								
62	3117 Briquet	place of paper use					n	ambiguous (2)	Aichstätten	Aichstätten								
63	25 GW	incunabula location	y						Aidona	Aidona	48,27	9,38	D					
64	3118 Briquet	place of paper use							Aidone	Aidone	37,42	14,45	I	Sicilia	Enna			
65	3119 Briquet	place of paper use					y		L'Aigle	L'Aigle	48,75	0,63	F	Basse-Normandie	Orne			
66	3120 Briquet	place of paper use					y		Aigueperse	Aigueperse	46,02	3,20	F	Auvergne	Puy-De-Dôme			
67	3121 Briquet	place of paper use					y		Airasca	Airasca	44,92	7,48	I	Piemonte	Cuneo			
68	26 GW	incunabula location	y				n	not identified	Aisey-Le-Duc	Aisey-Le-Duc								
69	29 GW	incunabula location	y				y		Alud	Alud	46,30	23,72	RO	Alba				
70	3123 Briquet	place of paper use					y		Aix-en-Provence	Aix-en-Provence	43,32	5,26	F	Provence-Alpes-Côte [Bouches du Rhône			
71	30 GW	incunabula location	y				y		Aachen	Aachen	50,46	6,06	D	Nordrhein-Westfalen	Köln			
									Ajaccio	Ajaccio	41,55	8,44	F	Corse	Corse-du-Sud			

2.3 Generation process of the geo-list

The generation of the geo-list delivered here was not straightforward: it implies several steps in a lengthy process, as detailed hereafter.

1. *Extract place names* – Various methods have been used to extract the data, from a simple download of the appropriated field in the case of digital databases (IPB) to typing in the computer list the information found in a printed source (BPC).
2. *Identify locations* – Using various digital and printed maps and atlases the place names given in the sources had to be identified. For each record several sources were checked each time to minimize the risk of excluding homonym places.
3. *Disambiguation* – In many cases the same place name is given to distinct location ('Paris' in France and 'Paris' in Texas, USA). Following rules have been followed to decide to which location to attribute the place name (if possible):
 - a. *Additional data in the source* – Sometimes the source provides clues to the location of the place name, such as mentioning a nearby city or river.
 - b. *Geographical likelihood* – A bigger city is more likely to be meant than a smaller city.
 - c. *Historical likelihood* – Paris, Texas, USA is not a likely location of a place name of a book printed during the time of the Renaissance.
4. *Find coordinates* – Once it was established which location the source means, its geographical coordinates had to be found using the Internet and printed resources. Giving the various coordinates systems used by the reference materials, some had to be converted into the decimal system used in this geo-list. The main sources of information are (1) for coordinates and administrative units the Complete Files of Geographic Names for Geopolitical Areas (ISO/IEC 10646) from the GEOnet Names Server of the National Geospatial-Intelligence Agency, USA (<http://earth-info.nga.mil/gns/html/namefiles.htm>) and (2) for place name variants the Global Gazetteer Version 2.1 (Falling Rain Genomics, <http://www.fallingrain.com/world/>).
5. *Attribute administrative units* – Providing the administrative unit to which a place belongs allows doing geographical statistics by clustering point-locations into areas. The overview thus gained helps the historian in his investigative work of the past. This is the reason why along with the coordinates' information on the administrative units was provided.

Because administrative units are dynamical historical entities – created, abolished, their boundaries changed – it was decided to use the administrative units as they are at the present time.

2.4 Amount of data the geo-list

The geo-list contains 3.000 place names ordered alphabetically. 250 (A-Brest) have been processed, i.e. name-normalized, matched with coordinates, attributed to administrative units, place type identified and given identification references.

There are two reasons why no more data was processed: work on the GIS functionality is scheduled for the second year of the Bernstein project and much time was spend in extracting the place names from the sources.

The remaining work (Brest-Z) is estimated at one month for one person.

3. Collaborations initiated

1. *ECAI* – Through participation in an Electronic Cultural Atlas Initiative (ECAI) conference, contact has been established with an international community active in the area of historical GIS research and development. The core of this community is the so-called 'clearing house' (<http://ecaimaps.berkeley.edu/clearinghouse/>), a software that interconnects various projects with geographical content so that datasets from a project can be used by other projects. The Bernstein GIS developing team intends to take advantage of this opportunity and contributes with its work to the geographical information environment managed by ECAI.

Noteworthy, some of the participants at the ECAI meeting were projects funded by the European Commission, such as Query and context based visualization of time-spatial cultural dynamics (QVIZ, <http://qviz.eu>), Great Britain Historical Geographical Information System (GBHGIS, <http://www.gbhgis.org>), Server for digital historical maps at the Institute of European History Mainz (<http://www.ieg-maps.uni-mainz.de>).

2. *CERL* – The Consortium of European Research Libraries (CERL, <http://www.cerl.org>) maintains a database of historical variants of place names, which is a useful extension to the Bernstein GIS in so far as (1) it allows users a wider range of search criteria and (2) it provides a greater flexibility of integrating the Bernstein digital

workspace with other databases and Internet services. Discussions are underway on providing CERL with the place coordinates in exchange for variants of place names.

3. *BKG* – In order to solve the issue of digital boundary coordinates for administrative units it is intended to contact the German Federal Agency for Cartography and Geodesy (BKG, <http://www.bkg.bund.de>), which has work experience in spatial data projects of the European Commission. The contact was suggested by the Bernstein Project Officer, who we thank for providing us with the name of the persons to write to in this matter.

4. Future steps

The next steps to complete the geographical data stock for the Bernstein GIS, are:

1. Complete the matching of place names in the geo-list.
2. Incorporate the place names as given by three Bernstein databases (NIKI, WILC, WZMA, c. 200 place names).
3. Acquire boundary coordinates for the administrative units in the geo-list.