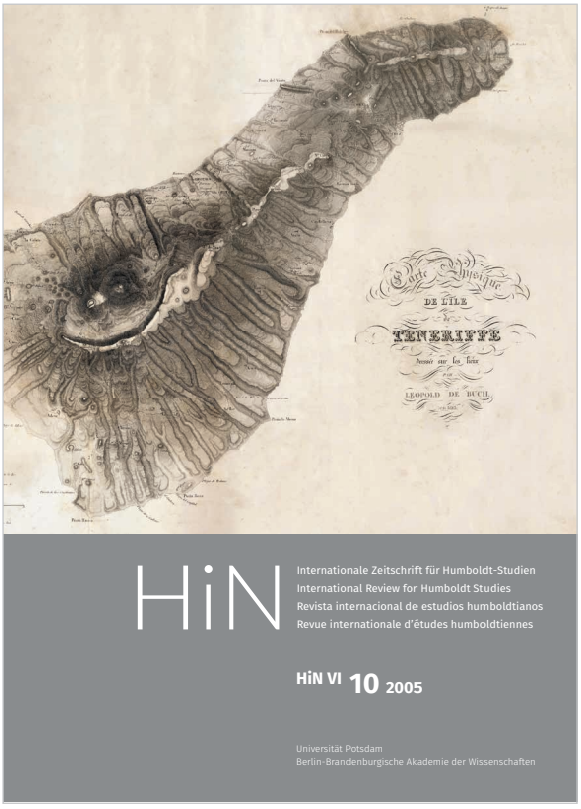
University Press Potsdam

Artikel publiziert in:

Ottmar Ette, Eberhard Knobloch (Hrsg..)

HiN : Alexander von Humboldt im Netz, VI (2005) 10

2019 – 82 p.
ISSN (print) 2568-3543
ISSN (online) 1617-5239
URN https://nbn-resolving.org/urn:nbn:de:kobv:517-opus-35175



Zitiervorschlag:
Doherr, Detlev: The Humboldt Digital Library: Exploring Innovative Structures, In: Ette,-Ottmar; Knobloch, Eberhard (Hrsg.). HiN : Alexander von Humboldt im Netz, VI (2005) 10, Potsdam, Universitätsverlag Potsdam, 2019, p. 30-34.
URN https://nbn-resolving.org/urn:nbn:de:kobv:517-opus-35215

The Humboldt Digital Library: Exploring Innovative Structures

Detlev Doherr

Summary

Alexander von Humboldt's maps, graphs and illustrations contain a great deal of detail, but in the available rare editions they are hardly visible to the naked eye. In many editions they have been reduced. In a digital library, they will become accessible in their entirety, and Internet technology will reproduce them in a form that overcomes the limitations of the original printing. The user will be able to enlarge the images and see details that might have been overlooked in the past. The Humboldt's digital library will adhere to the standards for digital libraries established by the Open Archives Initiative (OAI) and the tools EPRINTS and DSPACE to provide the Web services and determine the most effective way to establish dynamic linking and knowledge based searching of information within the archive.

Zusammenfassung

Alexander von Humboldts Karten, Bilder und Illustrationen enthalten eine Vielzahl von Details, die in den verfügbaren Ausgaben seiner Werke mit bloßem Auge kaum oder wegen Reduzierung der Auflösungen gar nicht sichtbar sind. In einer digitalen Bibliothek sind diese Informationen komplett zugänglich, ja die Internettechnologie lässt es zu, diese in einer Form zu reproduzieren, die die Grenzen des herkömmlichen Ausdruckens sprengen. Der Nutzer ist in der Lage, die Bilder zu vergrößern und Details zu erkennen, die möglicherweise sonst übersehen werden könnten. Die digitale Bibliothek über Humboldt entspricht den Standards, die die Open Archives Initiative (OAI) definiert hat. Als Plattform werden serverseitig die Programme EPRINTS und DSPACE eingesetzt, um einen geeigneten Internet-Dienst anzubieten und effektive Möglichkeiten zur dynamischen Verbindung von Informationen und wissensbasierter Informationsrecherche im gesamten Archiv zu entwickeln.

Resumen

Los mapas, cuadros e ilustraciones de Alexander von Humboldt contienen gran cantidad de detalles, los cuales, en las ediciones disponibles de sus obras, son apenas visibles, o por reducción de la resolución ya son invisibles. En una biblioteca digital esta información es completamente accesible, ya que la tecnología del internet permite reproducirlas de modo tal que va más allá de los límites de la impresión común. El usuario tiene la posibilidad de ampliar los cuadros y reconocer detalles que normalmente pasarían inadvertidos. La biblioteca digital sobre Humboldt cumple con los estándares definidos por la Iniciativa de Archivos Abiertos, OAI (Open Archives Initiative). Como plataforma del servidor se hace uso de los programas EPRINTS y DSPACE, permitiendo así brindar un adecuado servicio de Internet y desarrollar posibilidades efectivas de relacionar dinámicamente las informaciones con las búsquedas basadas en conocimientos de todo el archivo.

Über den Autor

Detlev Doherr

1953 in Göttingen geboren, 1983 Promotion an der Universität Göttingen, seit 1990 als Professor für Informatik und Umweltinformatik an der Hochschule Offenburg, 1998 Zertifizierung als „European Geologist“, seit 1993 wissenschaftlicher Leiter des Hochschulrechenzentrums, seit 2003 ‚Adjunct Research Associate' im Hall Center for the Humanities, University of Kansas.

The Humboldt Digital Library: Exploring Innovative Structures

Detlev Doherr

1. Introduction

A catalyst in the development of a Humboldt digital library has been the extraordinary adaptability and relevance of Humboldt's contributions to science. Humboldt, in contrast to his predecessors, was a „scientific traveler“, whose precision in reporting his observations and findings make him unique and still relevant.

His unorthodox combination of text, precise data and images reflects his search for totality and global interconnectedness. Because modes of publication available to him could not accommodate the ambitious scope of these aims, it remains for computer-based information technology to do justice to Humboldtian science, represented in Humboldt's twenty-nine volumes about the Americas with 30,000 pages and approximately 1500 images.

Using advanced information technologies we expect to provide multi-modal access to Humboldt's legacy on an Internet based knowledge network, integrating webservices like digital libraries with text and images, user driven search profiling, and linkage to evidentual data archives.

Humboldt's travel narratives suggest innovative features for our information system. Together with the rich content of his work and the exact measurements this reveals the obvious advantages of a Geographical Information System (GIS) to analyse the spatial connections between Humboldt's travels and observations. For example, Humboldt's observations on plant and animal distribution can be useful for detecting change in patterns during the last two hundred years.

2. Digital library

The text is the ultimate center of interest and point of orientation. Its presentation in the Internet is justified primarily because it is possible to present such a text in a way that brings to life the countless relationships that it suggests within and outside of Humboldt's works.

The procedures and operations of the Humboldt digital library will adhere to the standards for digital libraries established by the Open Archives Initiative (OAI) [http://www.openarchives.org] based on LINUX operating systems to check the searchability of texts, structured in pages and paragraphs from different volumes. We installed EPRINTS and DSPACE to provide the digital library on a Web server and determine the best way for dynamic linking and knowledge based searching of information within the archive. The XML documents were stored as document files in our digital library and were used for system checks and developments of a prototype for text search, linkage to interactive maps and further editing of text.

In an early stage of our work the important question arose whether the library should contain documents in XML form with related style sheets or PDF files, which could be searched directly by automatic search on Internet (for example, Google) or whether we need complex database structures in relational databases, which can be managed by SQL (Structured Query Language). These questions needed to be resolved, and as we worked with the texts we soon realized that we had in our possession a solution that gave satisfactory answer to basic questions, explained in the next chapter.

3. Project Developments and Testbed

In order to demonstrate the functionality of this information system, we developed a prototype that contains some basic functions for text search within all documents, tools to access volumes, chapters and paragraphs, and a search method for keywords or related synonyms. To avoid critical data transformation from XML and

SQL structures we provide the page-oriented scans for each paragraph or footnote. Those page numbers are given in parentheses and are linked with the image itself. Thus, the user may click on the marked page numbers in the digital text to see the scan of the original text or image.

The digital information system is based on modern client-server technology. This system contains the complex SQL structures of our database with Humboldt's texts and images, structured in paragraphs. This solution provides many more innovative features than the traditional digital library:

- Web services for information, not for documents
- Comparative analysis of paragraphs in different languages
- Identification of keywords and synonyms for multilingual approach
- Thematic search
- Chronology of travels
- Identification of names
- Links to archives from particular text paragraphs

For the presentation of geographic or thematic maps we need an interactive map server on the Internet, and this map server must contain digitalized information and provide a link to the text archive. We developed access to texts with points of interest on Humboldt's Orinoco map and tested navigation features. Our online prototype shows a simple Orinoco map which is georeferenced and has some marked points in thematic layers (See figure 1). Our experiment has been successful and shows the usefulness of navigation from texts to maps and their thematic layers.

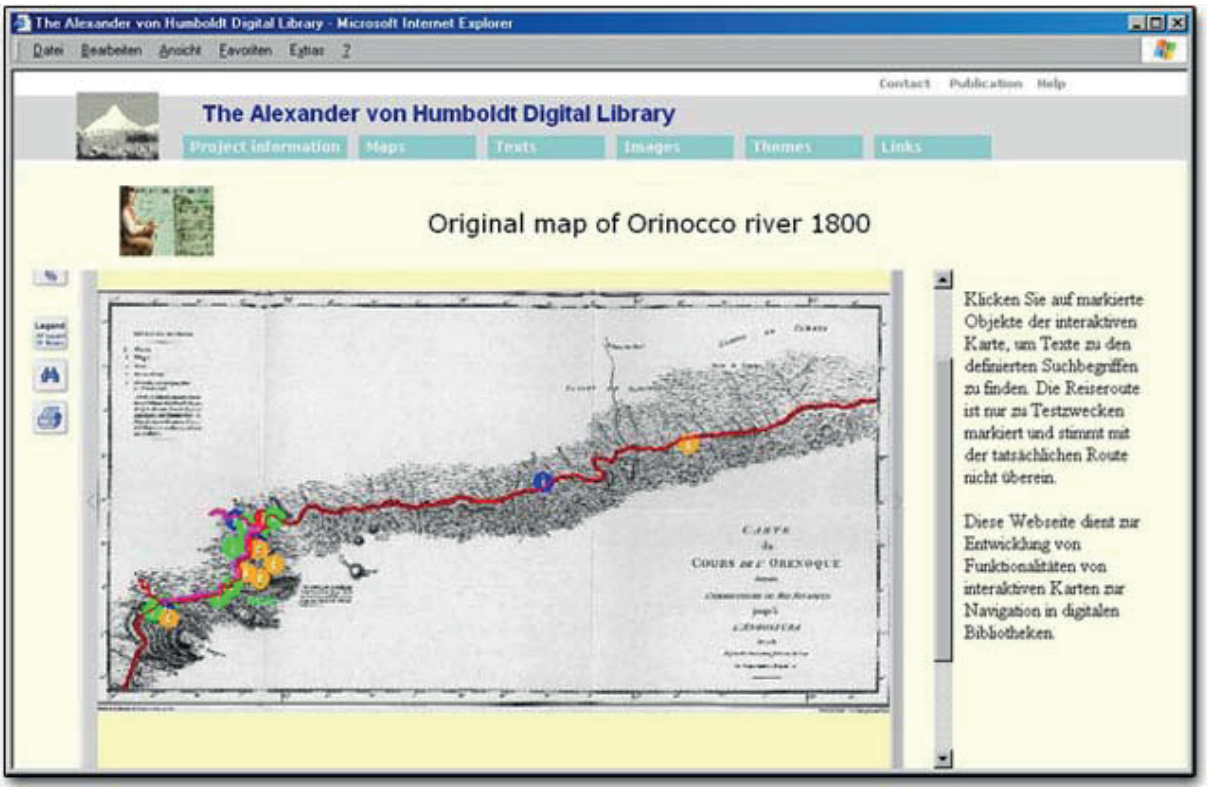


Fig. 1: Interactive map (Humboldt's travel map to the Orinoco) with interactive layer and points of interest, providing access to the original texts.

Furthermore, we also investigated the problem of accessing texts depending on a location on the map. We developed a search function to use a keyword, which is related to the entire point of interest in the interactive map. We defined objects on the map, which have attributes such as themes, dates, or notes. This problem revealed the usefulness of SQL structures for the correlation of objects inside the maps and the harvesting

of information from the digital archive. This is precisely the structure that we have developed successfully in our prototype.

Unfortunately, SQL structures involve a complex database design. So we defined media assets, which contain the formatted text, images, scans or maps. In addition, the digitized and formatless texts and notes are part of those assets. Together with unique identifiers for paragraphs, related keywords, themes, synonyms and other elements, it is possible to create dynamic links to other assets, no matter if maps, text or images. The system will provide access to those assets via Web browser. The linking can be predefined in expert shells, but also created automatically, whenever global information is necessary.

4. Testbed for data communication

The request for information from Internet is managed by the Humboldt portal, which offers a graphical user interface, depending on predefined user profiles. This structure offers search for text information on different paths. The first path is the search for text passages, keywords and themes. It leads directly to a text presentation of texts. Another option is to search by a specific date or location in Humboldt's travel entries. Using the spatial data engine, the user will have access to a variety of map layers. The user can use the maps for zooming in, check for locality names, search for topics in the legend to find points o interest. The result can be presented as a map or a text paragraph in the Web page.
The connection between the text and spatial data will be established by interactive maps, containing graphical objects with attributes. The focus will be on the identification of text attributes in the database and the definition of hyperlinks to access the archive.

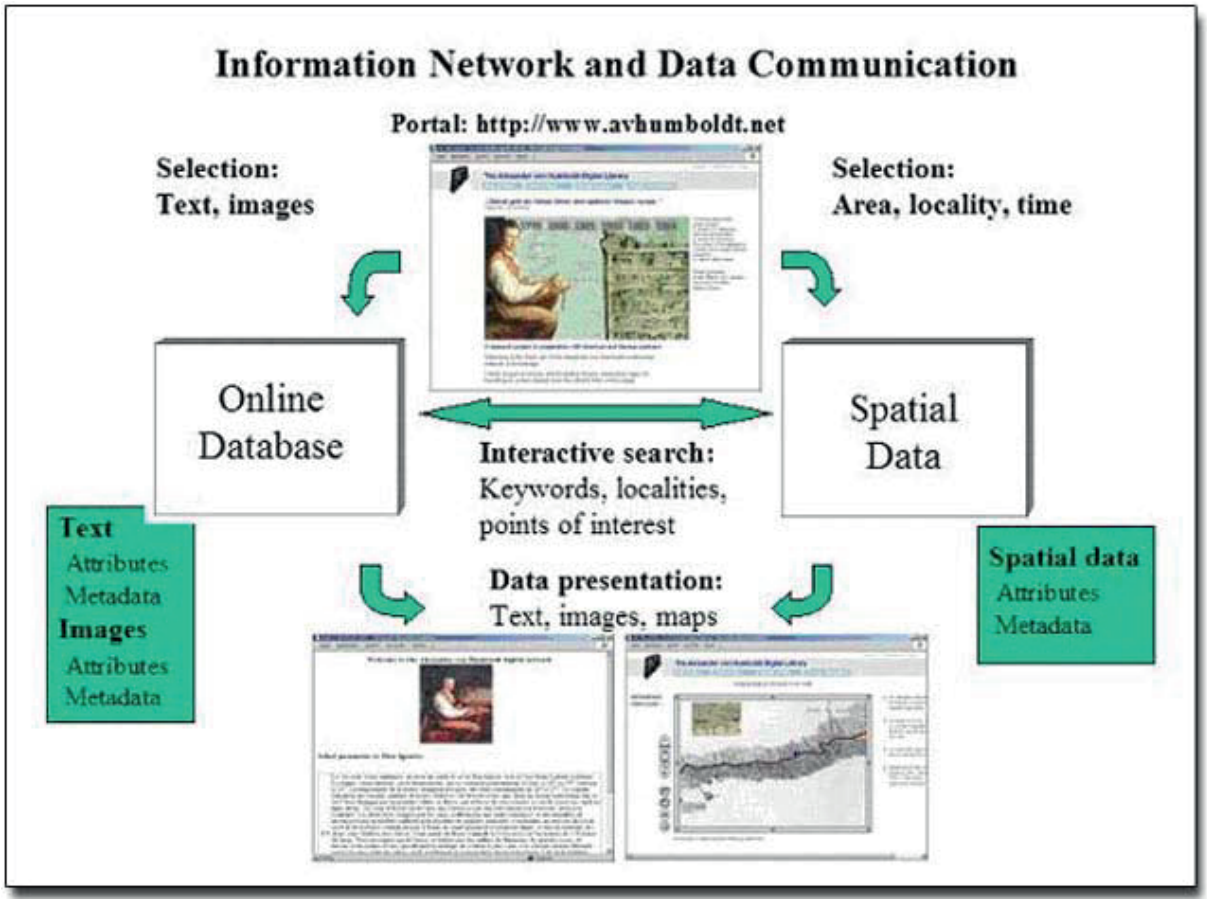


Fig. 2: Testbed for an information network and data communication within the archive

To be able to compare Humboldt's insights with what is known today, it is necessary to allow the user to navigate outside the system to other relevant data banks and return. Thus, the Humboldt digital library needs to be more than a collection of static images with hyperlinks. Only a dynamic system can accommodate navigation among independent data banks and can accomplish the diverse tasks we have outlined.

The consequences of the concept are dealing with search functions in the three dimensional scale, containing predefined keywords as well as a network of information, which is established on the basis of definitions of synonyms to existing keywords no matter of languages, spellings, or kind of texts.

5. Conclusion

Our testbed is running on tools from the Open Archives Initiative, opening the scope to a global network of knowledge. First steps are made to develop the structures of a digital information system, integrating text archives in several languages, images, georeferenced maps, and notes together with other information about original volumes from Humboldt and additional sources. We envision a Humboldt's digital library that allows the visitor to reconstruct Humboldt's path, encounters, and discoveries simultaneously from different perspectives and more precisely than ever before.

6. References

Open Archives Initiatives (OAI): http://www.openarchives.org

D. Doherr (2004): The Humboldt Digital Library: Exploring innovative Structures. http://web.gc.cuny.edu/dept/bildn/humboldtconference/Abstracts.htm