[This document contains the author's accepted manuscript. For the publisher's version, see the link in the header of this document.]

Abstract:
The Humboldt digital library (HDL) represents an innovative system to access the works and legacy of Alexander von Humboldt in a digital form on the Internet (www.avhumboldt.net). It contributes to the key question about how to present interconnected data in an appropriate form using information technologies. The HDL has been created as a dynamic digital library with the capability of connecting multilingual and multimedia data from diverse online archives. Humboldt's volumes have become available, but beyond that any relevant information related to the observations of Humboldt, even outside the works can become immediately accessible. This makes it possible to recognize natural changes and compare Humboldt's descriptions with recent situations. The technology we have developed addresses the issues of sustainability and makes it possible to detect changes in the environment since the time of Humboldt's observations.

Text of paper:

Humboldt digital library and interconnectedness

Detlev Doherr[1] and Frank Baron[2]

1. University of Applied Sciences Offenburg, Badstr, 24, 77652 Offenburg, Germany

2. German Department, University of Kansas, Lawrence, KS 66045-2127, USA

Detlev Doherr (Corresponding author)
Email: ddoherr@hs-offenburg.de

Frank Baron
Email: fbaron@ku.edu

1 Introduction

Now, after approximately 200 years, a comprehensive online access to the printed works of Alexander von Humboldt and his extraordinary exploration is within sight. Humboldt himself struggled to publish his works in a print version, in a form that could not do justice to the broad range of his discoveries, observations, and its interconnections (Baron and Doherr 2006).

The project of developing a digital library on the legacy of Alexander von Humboldt was initiated in the year 2000 and was organized in a partnership between the University of Kansas, USA, and the University of Applied Sciences Offenburg, Germany. The results of this research work have published under the name "Humboldt Digital Library" (HDL) on the Internet and contain a digital collection of Humboldt's works (Baron and Doherr 2006; Doherr and Brahaj 2009).

Starting with the platform for digital libraries called EPRINTS (see: http://www.eprints.org), we converted more than 29 volumes from the Humboldt publications via optical recognizers to text format. The illustrations, tables, and maps (around 3000) were scanned and preserved in image formats. All information has been made available on the Web as PDF files or scan images.

Not only questions of worldwide digital accessibility motivated us to develop an information system, which we called HDL, but it was Humboldt's profound grasp of a broad spectrum of knowledge and his insistence on precision, which placed him in a unique position to advance the natural sciences in the first half of the nineteenth century. He held the strong conviction that the distinct disciplines represented artificial divisions of knowledge, and the key to understanding nature was interconnectedness. As a result, he published a detailed narrative of his travels, constantly integrating his observations and data from various perspectives and disciplines. The need to make connections was characteristic of his publications.

So we considered the potential of the Humboldt Digital Library to present the works of Humboldt in an innovative way. We attempted to keep in mind the full range of his philosophical and scientific achievements, which invite information technology to provide access to his legacy.

2 Requirements of interconnectivity

In a digital library, the focus of the system is to provide digital documents representing the original works. In the HDL, the focus is to provide a research environment to navigate and do research on Humboldt's works. The documents of his writings are conserved in their original form, but the system should extend further by providing interconnectedness through different modes and features. It is not just the original documents per se that has concerned us, but also the way in which Humboldt represented knowledge. Humboldt saw all phenomena in the context of a historical development. For him, the historical perspective was an active dimension in the web of interconnectedness. Nature always existed for him in time and space.

The concept of interconnectedness, according to Susan Faye Cannon, is the key to understanding *Humboldtian science*. Cannon writes that Humboldt's primary goal was "the accurate, measured study of widespread but interconnected real phenomena in order to find a definite law and a dynamic cause" (Cannon 1978).

Using modern Internet technology, the HDL provides the key to understanding Humboldt's works more than before because of the interconnectedness of information objects within the HDL, which cannot be provided by other digital libraries (for example EPRINTS) today.

3 System architecture of the HDL

The HDL integrates a graphical user interface, based on a Content Management System as a Web interface and a LINUX- based information system, developed in the programming languages PERL, PHP, and JavaScript. The relational database model is implemented in MySQL, containing the whole digital information. This model integrates the text data together with their references to documents, illustrations, maps, tables, page scans, and to translations. In addition, the system contains tables with metadata associated with the paragraphs like key word lists, synonyms of terms, interconnections defined by editors, user data (only accounted users) like profiles, bookmarks, and personal notes.

The text paragraphs are defined as the main information units, representing the key components of the system. As the original structure of the published volume, they contain the searchable text phrases, keywords, and terms that are used to establish the hyperlink interconnections between different text paragraphs, volumes, and documents within the HDL. They also allow a comparison of text phrases, the use of specific terms in different semantic contexts, or the definition of synonyms.

In addition, this structure opens an option to manage and provide not even the English version but other translations parallel to the given paragraph. An example is given in Fig. 1, showing the presentation of Humboldt's *Personal Narrative of Travel to the Equinoctial Regions of the New Continent during the Years 1799–1804* with the title of Chapter I, Part II in English and Spanish on the website of HDL.
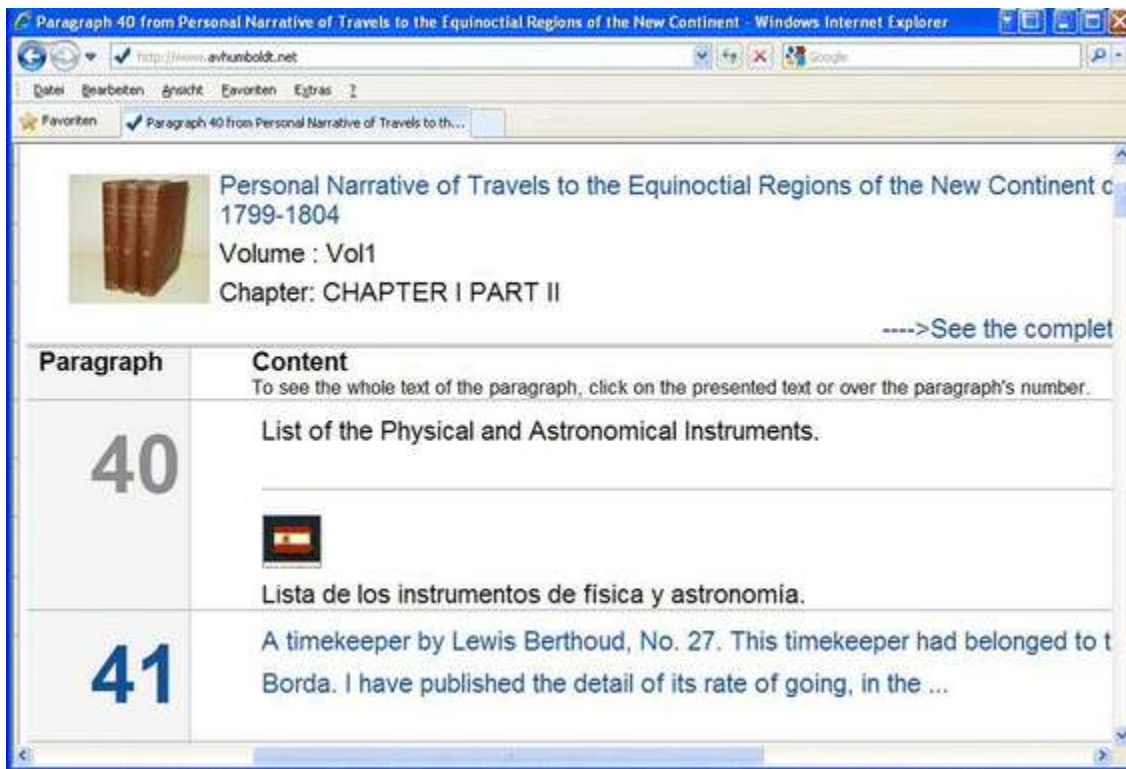

Fig. 1 Paragraphs (here titles) in different translations presented by the HDL

Other translations can be integrated, which would be indicated by a flag marker at the bottom of every single paragraph.

Humboldt published numerous tables with valuable observations, measurements, and environmental data. Following the concept of the paragraph-oriented data sets, the HDL contains the table data from Humboldt's documents as text information inside the database as HMTL-formatted text.

We decoded the Chimborazo tables with this feature, integrating the numerous measurements and observations in the system and made them searchable and comparable with other data. As a result, the title of each column gives the attribute names inside the database. The altitudes are the identifiers of the data sets, containing the textual information of the Chimborazo tables (see Fig. 2). As a consequence, the HDL can present the data now as a table output on the Web with the text in different translations in an easily searchable form.

Data tables containing Humboldt's measurements and observations related to the altitude of the Chimborazo. The column titles are used as attribute names of the data table

The database includes data sets of different translations of every each information cell.

Each information cell contains text data, which are searchable and comparable with other text in the HDL no matter of language, volume, or table

Fig. 2 The table shows the internal database structure of the HDL, decoding the image data of the Chimborazo image

4 System functionality

The HDL provides a new dimension to Humboldt's works by means of interconnection: connecting texts, paragraphs, translations, and other forms of data, such as images and graphic illustrations. The HDL can recreate the context of a particular text, making it possible to view images, geo-referenced and interactive maps, information about plants and animals, as well as scientific data relevant to Humboldt's observations.

A potential user can access to the online HDL system free of charge and without any authentication or login.

In order to locate the required paragraph, there are two different available browsing methods. The first logic of browsing is document and book-oriented, offering documents in PDF format, or presenting a list of books and chapters within Humboldt's works as text. An image reference will be displayed together with the text, whenever images are linked. Each paragraph has a hyperlink containing an automatic link to the scanned image of the page to which it belongs.

The other option to use the HDL is given by the information retrieval module (IRM) (Doherr and Brahaj 2009). This approach offers the advantage of a most efficient search for any word, term, key word, synonym, or thematic topics. An example of the use of IRM would be a search of all table data within the HDL, where Humboldt described measurements about "Snow." The result of such a search inside the HDL is presented in Fig. 3, which shows a table together with the citation of the volume.

Search results *snow*

Back to search

Document	Result
	To see the whole text of the paragraph, click on the presented text or over the paragraph's number.
Aspects of Nature Chapter:131 Paragraph:70	My statements respecting the mean height of the Snow-line in the Himalaya. (Asie Centrale, tom. lii. P. 326.)

	Paris feet.	English feet.
Northern declivity	15,600	16,626
Southern declivity	12,180	12,981
Difference	3,420	3,645

Fig. 3 Result of a search of table data for the keyword "snow"

A click on the paragraph number on the left side bar will open the paragraph, which contains the presented table.

Every online access will cause a set of statistical data, which is used to improve the search capabilities of the Humboldt Digital Library (Doherr and Brahaj 2009). The HDL is calculating weight maps for the paragraphs, depending on the user's profiles and search activities, which leads to an efficient use of the IRM and highlighted paragraphs, whenever they are calculated as important.

5 Searchable graphical data

Humboldt's maps, graphs, and illustrations are valuable, but they are available mostly only in rare editions. Texts are often hardly visible to the naked eye because of the reduction of the image resolution (Doherr 2006). But, in the HDL the images are linked to the paragraphs to which they belong to. Further efforts are taken to improve the presentation of the images and their context to make the image data searchable.

As a prototype of integrating graphical data to the digital library, we used the image of the Chimborazo (Fig. 2) to demonstrate the options of online presentations and data interconnections.

Humboldt suggested that it would be possible to show the features of the tropics and the gradual changes up to the snow-capped mountains. Humboldt foresaw an integration of esthetics, botany, and geography. Joining the resources of various disciplines pointed the way to breakthroughs in scientific exploration.

The same vision of unity is embodied in Humboldt's famous graphic image of the Chimborazo (see Fig. 4), which shows plants at several altitude levels together with related physical measurements described in the right and left tables beside the cross-section of the mountain.
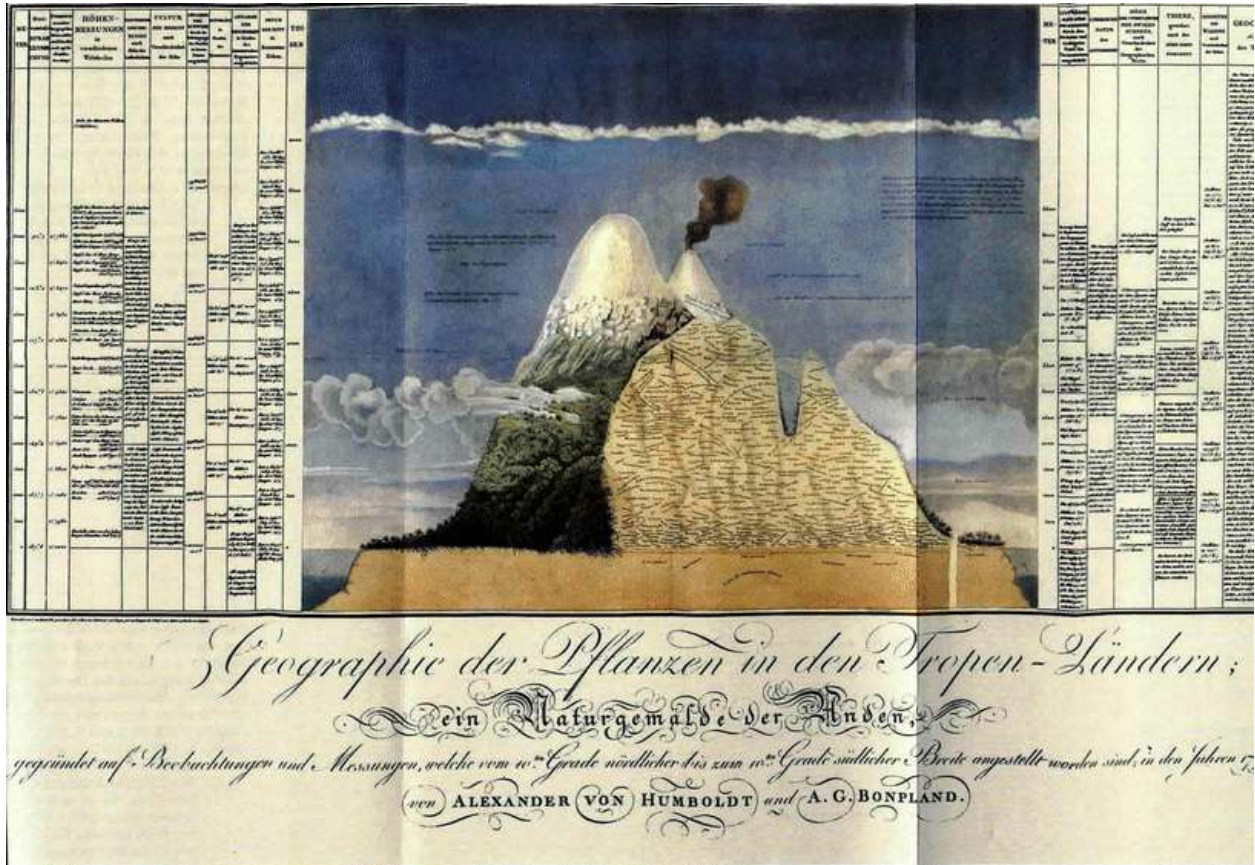

Fig. 4 Cross-section of the Chimborazo Mountain, containing details of plant geography, physical measurements, and observations in tables, related to altitude. (Essai sur la Geographie des Plantes—A. von Humboldt and A. Bonpland, Paris, 1918)

To include the graphic information of the image, we redesigned the internal part of the cross-section of Chimborazo with the plant names (see Fig. 5) and included areas, which contain hyperlinks to our system (Baron and Doherr 2006). Those hyperlinks point to related text paragraphs or plant illustrations.
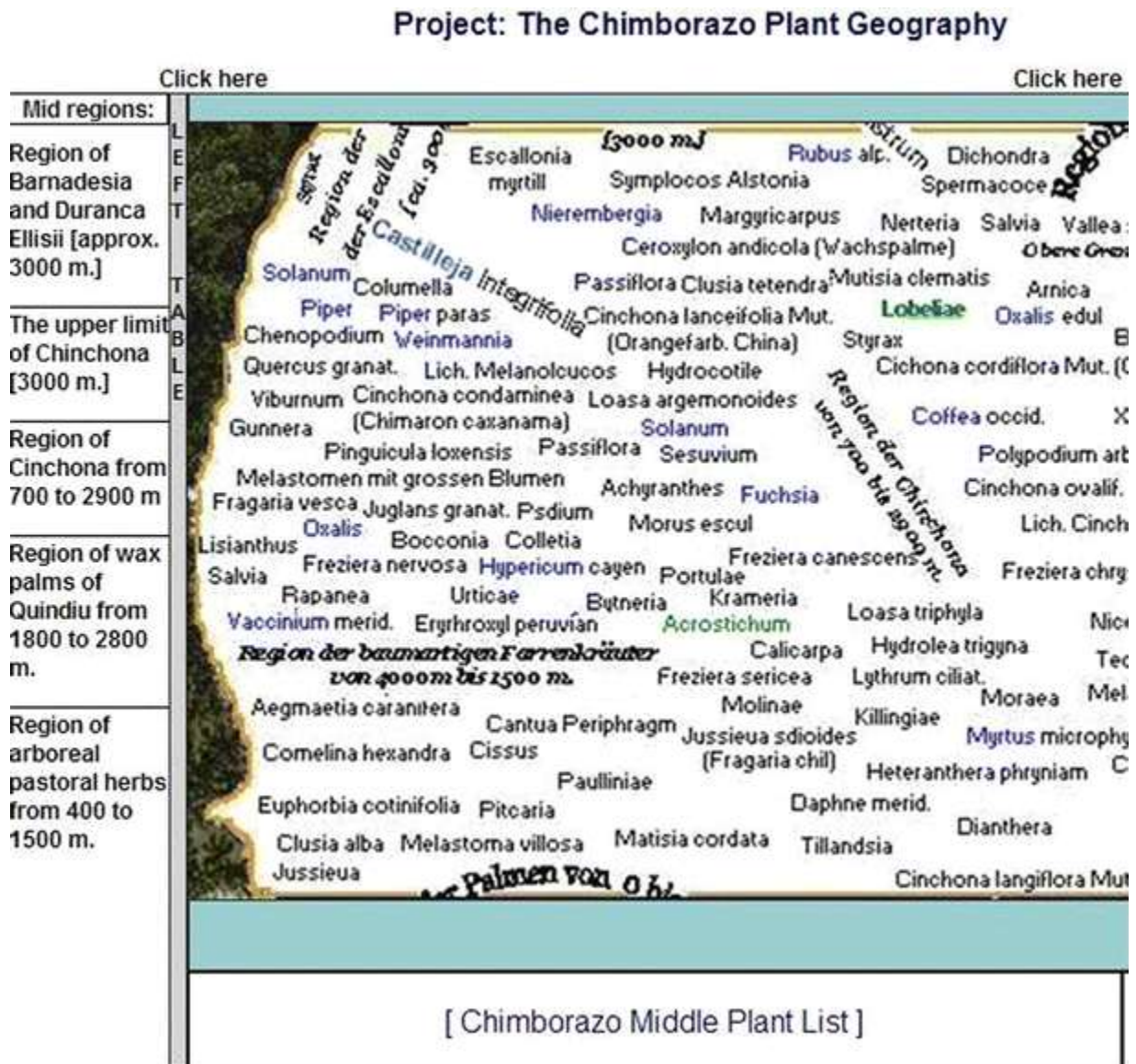

Fig. 5 The redesigned graphical presentation of the Chimborazo image contains plant names in different colors, which represent the hyperlink status inside the HDL

In addition, we have created a virtual representation of Humboldt's travels on the basis of Google Earth Technology. The travel covers basic information of Humboldt's journey in the New World. Included are important places, images, and paintings of artists, texts, descriptions, maps, and lines marking Humboldt's approximate path. This tool allows the user to see the journey chronologically, something which is not always easy when following his written works.

Google Earth helps users appreciate Humboldt's writings and accomplishments in a visual fashion and go further by linking the displayed information with texts of the digital library (see Fig. 6). So users can visually follow the travel routes and click on marked locations on the overlay map to jump into the HDL's database to read Humboldt's observations.
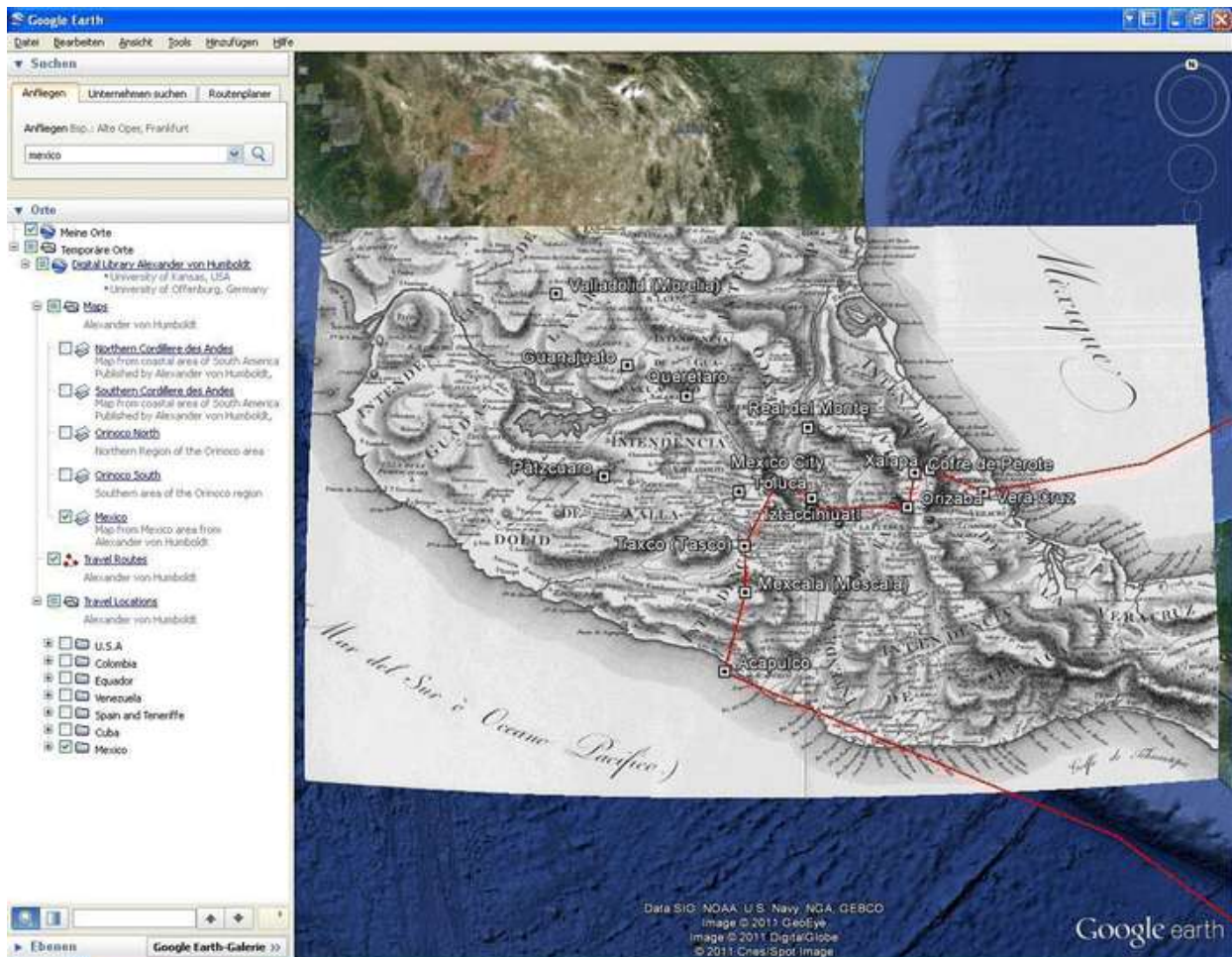

Fig. 6 Google Earth Presentation of Humboldt's travels on an overlay map of Mexico and locations with link to the text descriptions in the HDL

The HDL system also supports the communication between the specialists to share expert's knowledge, by allowing specific observations to be related to a paragraph in several contexts at the same time. Visitors interested in the works of Humboldt can create their own profile together with a personal user ID. The login opens the option of bookmarking and adding personal and public notes on each paragraph to the HDL.

6 Conclusions

As far as we are concerned, an effective Internet presentation has numerous preconditions. The Website must represent an advance over the printed and Internet resources now available (Baron and Doherr 2006). Moreover, because this subject matter is of interest not only to scholars but also to the general public, a primary consideration must be user-friendly access to texts and related materials.

The HDL preserves the author's original intention, retains an awareness of all relevant works, and still adheres to the requirements of a scholarly edition. The system reaches beyond the range of traditional digital libraries (which are limited to the function of word searches) by supplying dynamic links to sources, maps, images, graphs, and relevant texts. The system provides new forms of interaction and synthesis between humanistic texts and scientific observation.

Using modern Internet technologies, our project opens options for doing justice to his legacy. The first consideration is that Humboldt's original publications about his travels are now rare books; only a handful of libraries in the world have complete sets. Humboldt invested a fortune in creating expensive publications (folio volumes, frequently with plates colored by hand). Efforts to make works such as his travel narrative available to a wider public have resulted in the abridgement of the original texts, and many works still remain outside the scope of research.

By providing different representations of knowledge, by offering the chance of connecting different facts, and by bringing in new tools that allow interactions with other researchers, we provide a modern virtual research environment for the works of Alexander von Humboldt.

We developed the HDL as a prototype of a virtual research environment for Humboldt being aware of the fact that we did not capture all his volumes. But we expect to grow and expand international partnerships and cooperation and that will continue to increase the digital data on our platform.

References

Baron F, Doherr D (2006) Exploring the Americas in a Humboldt digital library: problems and solutions. *Geogr Rev* 96(3):439–451

Cannon SF (1978) Science in culture. Science History Publications, New York

Doherr D (2006) The Humboldt digital library, exploring innovative structures. ''Humboldt im Netz'', ISSN 1617-5239

Doherr D, Brahaj A (2009) Information management beyond digital libraries: Alexander von Humboldt in the Web. PIK, 32, pp 161–
166, ISSN 0930-5157, München

von Humboldt A (1918) Personal narrative of travel to the equinoctial regions of the New Continent during the years 1799–1804 (trans: Helen Maria Williams), London 1818