

# The Incorporation of NID into Digital Libraries and Its Benefits

Eisenberger, Sonja

**Abstract:** *This paper addresses the topic of the ongoing digitization of previously analog knowledge collections, primarily libraries, i.e. mainly but not exclusively written materials in text format and additionally materials in image, video, and audio formats. A short introduction to digital books explains the extended ways of using them in contrast to their analog predecessors and their incorporation into various systems such as NID (Net Interactive Documents). In one section, the topic of digital libraries is reviewed. A few of the most important representatives are briefly introduced. The 3.2 chapter applies to the software, which can be used for digital libraries. In the main part, an introduction is given to NID (Net Interactive Documents) and its current applications as well as projects planned for the future. Finally, different features of NID are compared with a similar system - Hypothes.is, and the advantages and disadvantages are explained.*

**Index Terms:** *digital books, digital libraries, Greenstone, Hypothes.is, IILF, NID*

## 1. INTRODUCTION

As an after-effect of the immense technological progress that happened in the last twenty years, a large quantity of analog objects are being superseded by digital ones. This includes, but is not limited to, books, photos, and documents.

This process of digitization, from analog to digital objects, allows a user to use an object in a wider and far more versatile way than originally possible, or conceivable. For instance, books are no longer restricted by language barriers or even literacy. Most digital books can be directly translated with as little effort as a mouseclick. In the case of illiteracy, a read-aloud option can be found and used nearly everywhere nowadays. Another advantage of digital books is their availability in the digital realm. There is no hassle of driving to a library or a store, where you have to check if the desired book is even stocked. For digital books it is sufficient, most of the time, to just type the name into a search bar. This is just one of the many reasons why digital libraries are emerging all over the web and slowly succeeding their analog predecessors.

The simplest definition of a digital library is to see it as a container for digital books. A more detailed definition is discussed in later sections. But it is important to know this simple definition, to be able to exclude various www pages and all kinds of databases since these can be containers for various kinds of data and not, as discussed in this paper, digital books, documents, and libraries.

In summary, this paper will focus on digital books, digital libraries, and NID. In addition, examples of software used and a comparison between NID and Hypothes.is will also be presented in the following chapters.

## 2. DIGITAL BOOKS

A book or document can be simply seen as a medium for conveying information. This includes, but is not limited to, informative facts, educational material, discursive writings, and fiction. An e-book is its digital alternative. A medium in which information is organized and structured to be presented to the reader.[1]

Therefore the definition of an e-book in the Oxford dictionary is as follows: "a book that is displayed on a computer screen or on an electronic device that is held in the hand, instead of being printed on paper".[2]

The earliest e-books were manually typed by volunteers. This process has been almost entirely replaced by digitization or scanning followed by optical character recognition (OCR) to create text that can be edited and viewed on screen.[3] Another approach would be the pure digital creation of a book, without a previous version existing on paper. The term "born-digital" therefore came into existence.[3]

Consulting a book, including e-books, involves the following basic actions: Browsing, searching, extracting, comparing, and evaluating the relevance and quality of the presented information.[1]

Other than reference materials, most books in print form are meant to be read linearly. In

contrast, computer-mediated reading and writing involve opening several windows and skipping from one to another to pursue a line of thought. Some e-books, especially those created for this medium, can exploit this multi-layered nature. For example, additional windows can pop up with more information, movie and audio clips can be inserted into the text, or a reader can navigate an e-book to follow a topic or idea that spans multiple chapters. It may even be possible to make annotations or send a message to the author.[3] All these ideas and applications are incorporated into NID and they are described in chapter 4.

### 3. DIGITAL LIBRARIES

Digital libraries are defined as Internet sites dedicated to the creation and preservation of electronic collections of books and other materials without requiring end-users to purchase the materials they wish to view and read.[4]

Digital libraries include a broad range of materials, from books to depictions of three-dimensional artefacts. Content is either digitally generated or digitally converted from a multitude of analog sources.

The creation of digital libraries has been accompanied by vast digitization efforts to turn the abundance of traditional scholarly and cultural resources held in libraries, archives, and museums into a digital format.[5]

#### 3.1. Digital Libraries Examples

##### 3.1.1. World digital library

The World Digital Library, established in 2009, was a collaborative project of the U.S. Library of Congress, with support from UNESCO and contributions from libraries, archives, museums, educational institutions, and international organizations from all over the world. The goal of WDL was to preserve and share access to some of the world's most important cultural treasures and significant historical documents to enable discovery, scholarship, and use.[6]

##### 3.1.2. Project Gutenberg

Michael Hart created Project Gutenberg in July 1971 to offer free electronic access to public domain literary works. A forerunner in many areas, Project Gutenberg was the first supplier of information services on the Internet and the oldest digital library. With 55 languages and 40 mirror sites across the globe, thousands of books are downloaded each day. Project Gutenberg encourages digitization in "text format," enabling a book to be copied, indexed, searched, analyzed, and compared with other books.[7]

##### 3.1.3. Universal digital library

Universal Digital Library (UDL), also known as Million Book Project, is an online information database developed by Carnegie Mellon University School of Computer Science and other university libraries. A number of other universities and government institutions from China, India, the United States, and Europe are also partners in this project.[8]

##### 3.1.4. The Bavarian State Library & The Munich Digitization Center

The Munich Digitization Center (MDZ) has been making the extensive collections of the Bavarian State Library (BSB) available on the Internet since 1997. The MDZ is the BSB's central innovation and production unit for the development, testing and realization of new products and processes related to the diverse and extensive topic of digital libraries. With around 2.7 million digitized titles from the collections of the Bavarian State Library and the provincial state libraries in Augsburg, Bamberg, Regensburg and Passau, the Digital Collections are now the biggest service of its kind in Germany, with over 355 million digitized pages.[9]

### 3.2. Software Examples Used by Digital Libraries

#### 3.2.1. International Image Interoperability Framework format (IIIF) [10]

IIIF is a technique to universally distribute images, audio, and video data from servers to various Web contexts where they can be viewed and used in a variety of ways.

Modern Web browsers can show file types like .jpg and .mp4 up to specified sizes, but they are limited in most other aspects. The IIIF specifications follow common Web standards that outline how each browser functions to allow using more advanced features outside merely viewing images and audio/visual files. This entails providing deep zoom, comparison, structure (for an object like a book, the structure is defined as the page order), and annotation for images. It involves being able to give complicated structures for audio/visual materials, including many film reels that together make a single movie, along with other elements like captions, transcriptions/translations, annotations, and more.

These things are made to function consistently via IIIF. This makes it possible for materials to be portable across viewers, to connect and unite them beyond institutional borders, and for many other things.

### 3.2.2. Greenstone [11]

Software called Greenstone is used to create and share digital library collections. It offers a method for compiling data and releasing it online or on portable storage devices like DVDs and USB flash drives.

The University of Waikato's New Zealand Digital Library Project created and circulated Greenstone, which was built in collaboration with UNESCO and the Human Info NGO. It is a multilingual, open-source software distributed in accordance with the GNU General Public License.

The Greenstone program intends to give users the tools they need to create their own digital libraries, especially at academic institutions, libraries, and other public service organizations. In the domains of education, research, as well as cultural heritage around the globe, digital libraries are fundamentally altering how information can be accessed and shared, especially in developing nations.

Creators hope that this software will promote the efficient use of digital libraries to spread knowledge and make it accessible to all.

## 4. NET INTERACTIVE DOCUMENTS – NID [12]

NID is an innovative system, which allows networking of materials and networking of users with users, experts, and providers.

The term networking materials, as it is understood in this context, is the ability to access additional, referenced, information at any time. This can be a passage in the same or another book, or a contribution to the World Wide Web, regardless of whether it is a text, an image or a multimedia clip. It can be one's own contribution or even be a piece of special software offering interesting material. Some of these additions and explanations could even be made specifically to the user's specifications.

This free annotation feature leads to a few basic issues that need to be addressed. The first required and implemented element is a hierarchy of rights. For example, any user (A) can add any annotation for himself, called a private annotation; or for a group (G) of users, also called a group annotation, who have agreed to see A's annotations. Annotations for a wider public are controlled by an editor with special rights, who can share them either only with user A or with the whole group G, or with the public.

A further important aspect is that these annotations are not part of the material itself, but additions to it.

The material itself cannot be edited after it has been uploaded as a NID document. Annotations can also be edited, deleted, or made available to user groups other than the one originally defined, by authorized persons.

Through these annotations, a document in a NID collection can acquire increasing amounts of available information over time. Therefore it is an interesting task to automate or semi-automate this linking between books and other material of interest. In this context, semi-automated means that the system suggests relevant links, but a user decides whether they are useful or not.

At least as important as the aforementioned networking of materials is the networking of users with other users, specialists, and providers. It can be considered a particularly valuable feature of NID that a user (1) can send feedback on any page to the person responsible for the page/document and (2) expect a response. Since any user can start a discussion on any page, NID offers on all sections of the website the possibility that a user contacts the responsible expert, either anonymously or by name, with a request for feedback or comment, or help.

Another quite useful feature of multimedia materials managed using NID is the ability to conduct full-text searches. This is not restricted to real texts, but also covers texts that are part of images. The reason for that is that NID is based on the IIIF standard.[10]

Especially important is to be able to apply the full-text search not only to one book but to multiple books simultaneously. With the majority of full-text search engines, you have to run the full-text search on each item individually. In NID, on the other hand, an entire "category" of documents can be searched with a single request. Besides full-text search, another possible and useful way to search a NID library or category is by author, title, or area.

With NID, multiple pages from different NID or IIIF compliant servers can be displayed on the screen at the same time. This allows a user to view descriptions in different documents simultaneously.

NID software is designed using state-of-the-art programming tools, platform-independent frameworks and components. This means that the software can be used on all major server operating systems. This also includes Windows-based server platforms (also client OS) and Unix/Linux.

## 5. HYPOTHES.IS A SHORT INTRODUCTION

The following section focuses mainly on the Hypothes.is web app, open-source software that

anyone can use, it just needs to be installed via a plugin. Hypothes.is lets the user create annotations, not only in documents but anywhere on the web. For this, each user is required to create an account.

Creating annotations via Hypothes.is is quite simple. One only needs to mark a desired word or sentence segment and then choose between the highlighting or annotation functions. When using annotation functions, one can freely choose what kind of annotation, i.e. plain text, links or images, one wants to incorporate. Additionally, tags can be created for this annotation. Tags are single words that should make it easier to search for the annotation. The annotation can be published either privately, within a group, or publicly. To publish annotations within a group, one must either have created the group or have been invited to the group.

#### 6. NID VS HYPOTHES.IS

There are some similarities between NID and Hypothes.is, however, they perform differently in individual categories and offer various applications.

Hypothes.is is universal, when the software is installed for the browser, one can annotate on any website. In NID, annotations are only possible in NID documents. However, NID has the advantage, as mentioned in chapter 4. that it is not needed to be installed and does not require registration to use it.

Additionally, these two systems differ in the annotation verification process. In NID documents, it is possible, if desired, to use a feature that forwards annotations to a respective expert of the document before they are published. The annotations are then reviewed and if they do not meet requirements determined by the expert, they are not published. Through extensive testing of the Hypothes.is software, it was concluded that there is still no automatic verification of created annotations. That is if a user posts an annotation that doesn't meet the experts' requirements, that is useless, inappropriate, or offensive, it cannot be verified in advance by an authorized person. Such an annotation can only be reported after posting with the report icon to be deleted by the admin, who may not necessarily be an expert in the field. This scenario does not arise when using NID.

Another important aspect is the annotation capabilities. With Hypothes.is, it is possible to annotate the entire web, but this only applies to text paragraphs. One can't mark up images or text in images. With NID, one can annotate anywhere in the specialized NID documents, be it an image or text. This wide range of annotation options is

made possible by the IIIF format in NID documents.

Another interesting point is the interactivity between users as well as the validity of comments. In Hypothes.is, users can reply to public comments of any person. There is no certainty that these responses are correct because there is no verification of the responses. On the other hand, in NID, there is no way to reply to annotations. Instead, there is a possibility to start a discussion feature. However, a user can ask a question directed to the expert of the document using annotations, in order to receive a professional and direct answer. With this feature, a certain level of professionalism and validity can be guaranteed. These are two factors, which are very important to the operators of NID.

These were only some differences between the two annotation programs, which were pointed out here. Both offer many more features, but discussing all of them requires a separate paper.

#### 7. CONCLUSION

In summary, it can be said that the digital advance will be unstoppable and will accompany us in every area, be it technology, information, or life itself. Therefore, it is important to use a system for digital libraries that is clearly designed, easy to use and can maintain a certain minimum level of quality.

A well-thought-out interaction of the software in systems like NID thus creates a hitherto unique element of high-quality knowledge transfer, not only between digital materials but also between different users and corresponding experts. In addition, the accumulated knowledge can be distributed in a simple way among all interested users. In the age where all the knowledge and ignorance of mankind is just a mouse click away and should be filtered accordingly, this is an important milestone. High-quality collections of scientific material bring a variety of benefits that previous analog collections could not offer. Therefore, it is important to develop and expand unique projects such as NID in order to make the accumulated knowledge available to mankind according to the current needs of users.

#### REFERENCES

- [1] L. Bennett and M. Landoni, "E-books in academic libraries," *Electron. Libr.*, vol. 23, no. 1, pp. 9–16, 2005, DOI: 10.1108/02640470510582709.
- [2] "e-book noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner's Dictionary at OxfordLearnersDictionaries.com." <https://www.oxfordlearnersdictionaries.com/definition/english/e-book?q=Ebook> (accessed Jun. 22, 2022).
- [3] C. Armstrong, "Books in a virtual world: The evolution

of the e-book and its lexicon," *J. Librariansh. Inf. Sci.*, vol. 40, no. 3, pp. 193–206, 2008, DOI: 10.1177/0961000608092554.

- [4] J.-A. Cerdón-García, J. Alonso-Arévalo, R. Gómez-Díaz, and D. Linder, "4 - Open access eBooks," in *Social Reading*, J.-A. Cerdón-García, J. Alonso-Arévalo, R. Gómez-Díaz, and D. Linder, Eds. Chandos Publishing, 2013, pp. 121–141.
- [5] I. Xie and K. K. Matusiak, *Discover Digital Libraries: Theory and Practice*. 2016.
- [6] "About this Collection | World Digital Library | Digital Collections | Library of Congress," *Libr. Congr. Washington, D.C. 20540 USA*.
- [7] "The Project Gutenberg eBook of Project Gutenberg (1971-2008), by Marie Lebert." <https://www.gutenberg.org/cache/epub/27045/pg27045.html> (accessed Feb. 13, 2022).
- [8] "ULIB." <http://ulib.isri.cmu.edu/ULIBAboutUs.htm> (accessed Feb. 06, 2022).
- [9] "Munich Digitization Center (MDZ) - Homepage." <https://www.digitale-sammlungen.de/en> (accessed Feb. 13, 2022).
- [10] "Home — IIIF | International Image Interoperability Framework." <https://iiif.io/> (accessed Jun. 26, 2022).
- [11] "Welcome :: Greenstone Digital Library Software." <https://www.greenstone.org/> (accessed Jun. 27, 2022).
- [12] S. Eisenberger, H. Maurer, and B. Zaka, "The power of NID."
- [13] "Home : Hypothes.is." <https://web.hypothes.is/> (accessed Jun. 27, 2022).

**Sonja Eisenberger** is currently completing her master's degree in computer science with a major in data science and machine learning at Graz University of Technology. She is currently working on the NID project, on which she is also writing her thesis. Her previous scientific work, especially on the topic of NID, can be viewed at Austria Forum. E-Mail: [sonja.eisenberger@hotmail.com](mailto:sonja.eisenberger@hotmail.com)