Documents as collaborative living objects in modern digital libraries

Bilal Zaka1, Hermann Maurer2

1COMSATS University Islamabad, Pakistan

zaka@comsats.edu.pk

2University of Technology, Graz, Austria

hmaurer@iicm.edu

Abstract

More digitally aware users of modern-day online libraries have a lot to offer to the data repositories in form of passive and active contributions. These contributions in form of browsing behavior, feedbacks, comments, reviews, annotations etc. become as valuable as the original contents with the passage of time as the original contents of the documents in the libraries. Digital libraries that allow such collaborative means of data acquisition and presentation make the otherwise static documents more dynamic. The multidimensional and rapidly increasing data sets of such online library platforms require nontraditional data-processing approaches to get maximum value out of the associated big data. In our paper we will discuss why an even more forward-looking online library platform is required. The paper covers a detailed comparison of existing commonly used library systems with an emphasis on advanced data analytics. The results of a comparative study point out gaps addressed by NID and set the direction for future digital library systems development. The paper further presents our efforts to develop such a platform, highlighting the basic and advanced features of the system.

*Keywords:* Digital Libraries; Net-Interactive Documents; Library Technologies; Library Analytics

1. Introduction

As mentioned above, modern digital libraries collect lots of data about how their books, documents or picture collections (we use the term document for any of those in the following) are used to learn to understand the behaviors and wishes of readers. However, usually digitally offered documents offer for the user just two advantages over printed material: They can be accessed around the clock, and they may contain interesting media content like pictures, clips, maybe even links to other material. Yet basically they can only be used “passively”, by reading//studying them.

In a way, this is strange: When one uses an E-Learning module, or a computer game, or a social network, or some blog, users can (are sometimes forced to) interact with the material and can communicate with other users or even the authors of the material.

Part of what will be discussed in this introduction has been mentioned in some ways already in [1] [2] [3] [4]. We are trying to summarize the most pertinent points in a concise manner here and express our belief that the ideas are particularly important for research papers and are independent of the subject involved.

To be concrete, the least one could expect from a digital document is that for every page of the document users can send some feedback, possibly anonymously, unless they want an answer to a question. This has been implemented already many years ago in e.g. [5].

Mentioning this idea to publishers or libraries, they are at first appalled: They are worried to get an avalanche of feedbacks that they could never handle: This is not just because of volume, but also because no person can be expert on all topics; and to always contact the authors responsible for a document would be much too cumbersome.

However, the answer to this dilemma is simple: One defines for each document, or a set of documents experts (typically the authors of the document), who will get and deal with the feedbacks, thus not creating any work for publishers or libraries. Yet libraries might also get possibly useful information, like that document authored by some persons get many more feedbacks than others.

However, there are two drawbacks to this: First, the same remark or question may be sent many times; secondly, if the remark is substantial, it should be available at the correct point. But even if the system would allow this, it would destroy the flow of the original presentation and the structure of the document, like number of pages, the table of contents and more.

For this reason, the two authors have discussed an alternative solution, called Net Interactive Documents (NID, for short) without such drawbacks. The NID library while facilitating the extensive user contributions in system, extends management services that helps in organization, analysis and presentation of data in appropriate form. A prototype has been developed, mainly by a team around the first author and has been used in quite a few applications by the team of the second author providing ideas for further useful features.

We will review the existing digital library platforms considering the heterogenous and incremental nature of digital information in section 2. In section 3 we will present a data analytics process model of NID library system in an effort to bridge the gaps identified in our review of prevailing library systems. In section 4 we will discuss the basic features of the NID library system. We will present a few extensions and applications from the user point of view in Section 5. Section 6 will deal with administrative aspects. Section 7 summarizes our findings. A list of references concludes the paper.

1. Prevailing digital library systems and big data analytics

Big data presents several challenges due to the large volume of data and greater complexity in terms of its variety and the expected speed of the system. In order to get the maximum value out of big data, context specific and non-traditional data processing and analytics techniques are required. During the past couple of decades libraries have gathered enormous amount of digital data both in structural, unstructured and semi structured format. To extend digital services, online publishing platforms and content management systems are extensively used by the institutional repositories and various other commercial and intellectual libraries. The information systems tailored to address the library requirements of information capture, curation, storage, search and provision of bibliometric services have been available for some time. There have been significant improvements in them to include aggregation services, better preservation and more user centric presentation approaches. However, implementation of big data analytics in libraries is still uncommon, we hardly see practical use of big data to support innovation in digital library services [6]. In our introductory section we talked about the need for user interactions within conventional digital libraries. In this section we will focus more on data analytics aspects of the large amounts of data that are managed by commonly used digital library systems. We mention a number of digital library software references in literature [7] [8] [9]. However, we will focus mainly on three of the more prevailing open-source software platforms that are being maintained by active communities.

* 1. DSpace

DSpace [10] is available as a free open-source full stack web application. It is used by over 3000 organizations worldwide. DSpace supports configurable workflows, metadata schemas, comprehensive search and display functionality. It is developed using stable Java based web technologies and maintained by active support groups. At the data layer, DSpace stores information in PostgreSQL relational data base store and offers local and cloud bit-store (e.g., Amazon S3) to hold and repository contents. The system allows storage and management of various data types such as Pdf, Word, Jpeg, Mpeg, Tiff files. Major strengths of DSpace includes its persistent existence in digital library software lineup, very extensive support of metadata creation (extensions of Dublin core based schema), import/export of data in a variety of formats such as Endnote, CSV, BibTex, TSV, RIS and compatibility with services such as OAI, arXiv, PubMed, CrossRef etc.

Use of Solr [11] search and indexing platform and support for the third-party web analytics tools like Google Analytics, Elasticsearch usage statistics and Solr Statistics also strengthens the system's bibliometric services.

The support for federated and stackable authentication and authorization methods is also a desirable feature that allows easy integration with organization's existing authentication models. DSpace also offers another very useful feature of data embargo scheme to restrict access to meta data and contents in repository.

However, the system is yet to introduce automated means of transforming the unstructured information available in its bit-store to machine useable structured models. This type of processing provides the means of inferring useful results for the consumption of its users. The system also lacks user interactivity and collaborative data enrichment functions.

* 1. Omeka

Omeka [12] is another web-based library platform available as free and open source software. It is developed using PHP Zend framework and supports MySQL relational database along with local file store as well as cloud storage such as Amazon S3. The digital library platform supports various multimedia content types for documents, images, audio and video. The system is capable of creation and sharing of rich content collection adhering to Dublin Core standards. Omeka is also maintained by an active community of developers and similar to DSpace and is also offered as cloud service. Omeka offers metadata sharing services based on Atom, DCMES-XML, JSON, RSS2 and RDF feeds. The metadata harvesting through OAI-PMH [13] data is also possible in repositories based on Omeka. Unicode (UTF-8) supports, and multilingual content support is also present similar to DSpace. The major strength of Omeka is its extensible and plugin based architecture that allows the addition to the core features of the digital library platform. The developer community contributed many useful system extensions that allows user contributions in form of annotations (hypothes.is[[1]](#footnote-1)) on text and image data. Advanced data analytics on library contents is possible using the Ngram and Text analysis Plugins. The advanced search functionality in Omeka is also available through plugin extensions.

Omeka offers some really interesting, advanced analytics and user data acquisition features. However, it is important to note that according to the independent software marketplace, analysis portals like SaaSHub[[2]](#footnote-2) and whatCMS[[3]](#footnote-3) , the Omeka market share and popularity among institutional repositories still falls behind a rather modest and consistent digital library platform like DSpace.

* 1. Islandora

We selected Islandora [14] as the third prevalent digital library platform in our review being an active open source digital repository ecosystem. It is an extensible and modular platform where a number of standalone systems are glued together to build multimedia digital content repositories. The Islandora platform is capable of managing complex data curation workflows and display schemes using the strengths of a very comprehensive and widely used content management system Drupal. Islamdora provides cross platform responsive, adaptive shareable presentation features through use of International Image Interoperability Framework (IIIF) [15]. Islandora makes use of different microservices for content transformation for better content preservation and use. The system offers creation, serialization, export, and exposition of metadata in customizable format using OAI-PMH and SparQL endpoints. This platform draws its strengths from widely adopted best practices and open standards. It uses notable technologies including Drupal, Apache Solr, Cantaloupe (an IIIF compliant Image server), Mirador (IIIF Compliant media viewer), Matomo (web analytics), ImageMagick (processing of images), FFMpeg (processing of video files), Tesseract (Optical Character Recognition Engine), FITS (File Information Toolset) and Fedora (Repository system architecture). The availability of data transformation and information management components in the technologies used signifies the capabilities of system towards advanced data analytics. The project road map lists the strategic objectives of adapting to modern dev-ops, detailed documentations, and process automation. However, there is no mention of innovations planned in user end services in the system that make use of data analysis.

Once popular digital library platforms such as Eprints[16], GreenStone Digital Library (GSDL)[17] are no longer updated frequently and are losing their market presence. These platforms still offer standard repository features with metadata harvesting, indexing and linking services and are in service being legacy data stores and being easy to maintain.

It is also pertinent to mention some of the proprietary digital repositories with massive contents collections and a large user base. These online digital libraries include ACM Digital Library[[4]](#footnote-4), IEEE Xplore[[5]](#footnote-5), ScienceDirect[[6]](#footnote-6), SpringerLINK[[7]](#footnote-7) and Scribd[[8]](#footnote-8). These systems mainly focus on subscription-based business models where authentication, authorization and content embargo schemes are given importance at system level. They also have user centric services related to profiling, and recommender systems in order to deliver personalized services. The proprietary digital libraries also make use of advance bibliometric services that suggests the applications of data analytics at certain levels. However, considering the digital rights management implementations in these libraries, collaborative and user contributed data features are not common.

Our study of open digital library platforms and proprietary library portals indicates that there is a need for introduction of user specific data contribution features both active and passive. Besides the system-based collaborations, user community collaboration makes digital repositories true living entities.

Data transformation features are also missing from many of the commonly used library platforms. It is important that unstructured or raw data is processed to produce meaning full information sets that can be better consumed by the system and provides users with better insights about repository data.

There is still no widespread awareness and use of advance analytics tools available in some of the open digital libraries; this is mainly because the effective implementation of big data analytics is sometimes hindered by the lack of technical competencies of library administrators [6]. It is important that a more seamless and fluid approach is used towards the inclusion of data analytics services in digital library platforms.

1. Bridging the gap with Net Interactive Document Library framework

From our earlier described review of existing digital library platforms and use experiences of major online repositories, it was deduced that the collaborative services and big data analytics are two key areas that can significantly improve user services in online information repositories.

We propose a digital library model where more system interaction opportunities are given to library users for a more engaging experience. The user and system-based collaborations can help in keeping the information base of libraries more updated. Learning being intrinsically a social process and the main objective of any library is greatly complemented when users get a chance to share information with others.

Furthermore, in our proposed library model, to make a better and broader use of advanced big data analytics we propose integration of data transformation, processing and inference features in the process flows of a library platform.

A system developed considering enhanced process flows provides the insights, help in decision making and delivers adaptive information. Insight is meaningful information generated when machine understandable and structured information is combined with a system available knowledge model. Thus, the first step in leveraging big data analytics is to have consistent data pipelines transforming the raw information into a standardized form. The data insights lead to the business intelligence for the system to help in decision making processes. With the help of informed decisions, adaptive information delivery is possible.

The model given in figure 1 presents the core system data flow with mapping of data creation, acquisition, transformation, processing and presentation services at various stages.

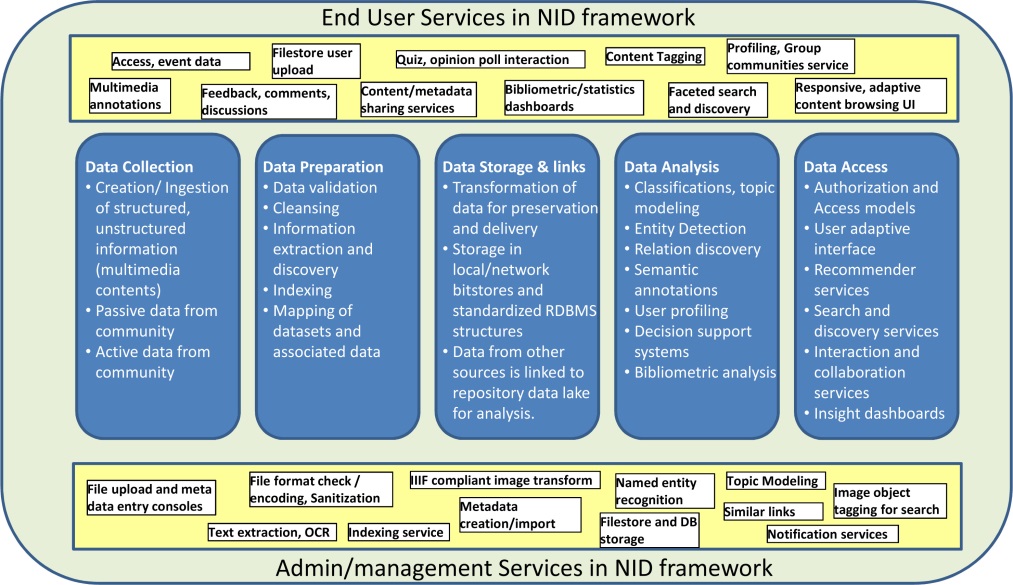


Fig. 1. Big data analytics process flow in NID framework

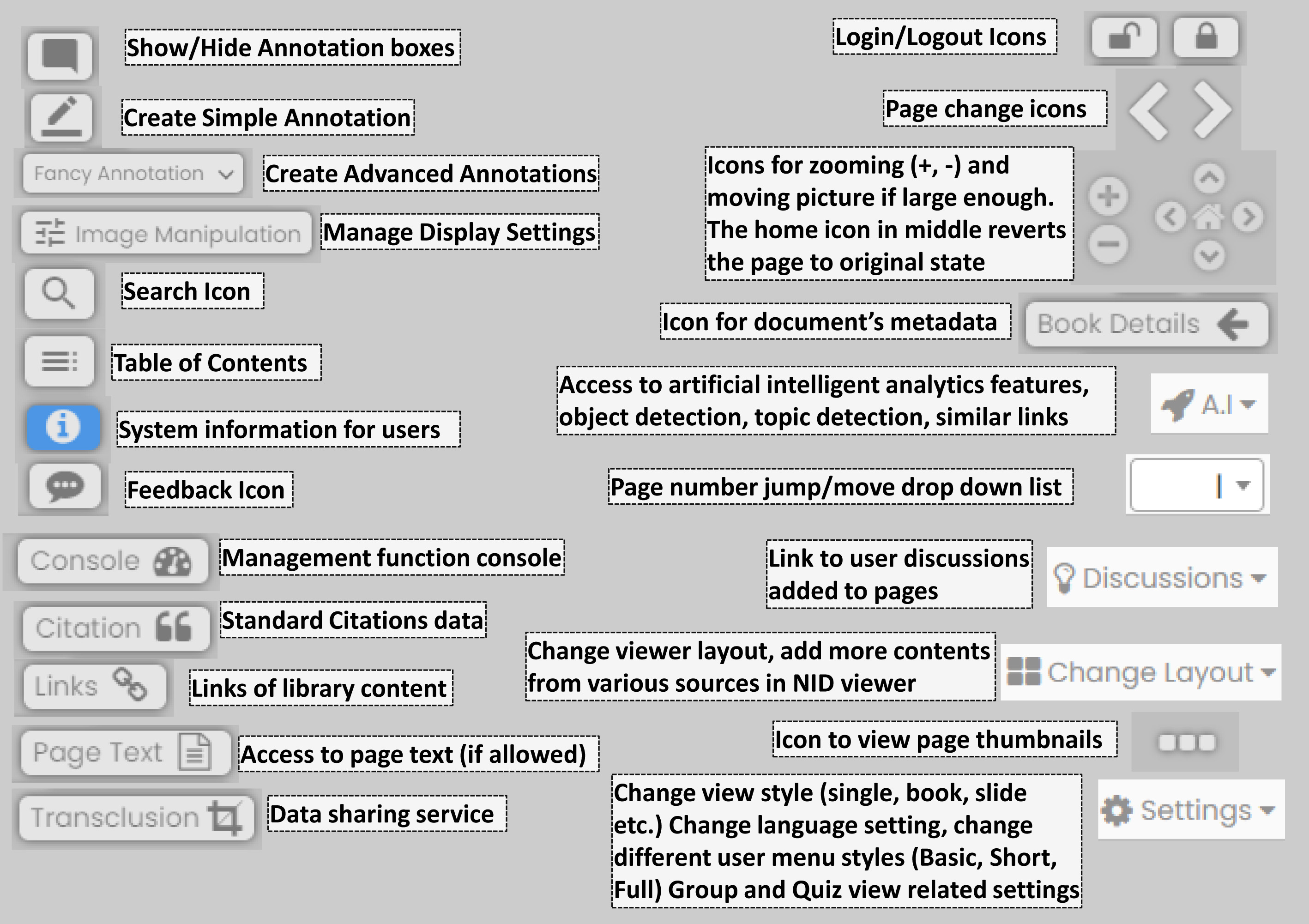
We introduced a number of user information creation services such as annotation, feedback, discussion board, file stores, tagging etc. Besides such user generated active or explicit collaborative data, the system also has means to gather implicit data through information use monitoring. This data supplements the existing metadata and original contents available in the library store in an ongoing manner. The frame-work also provides transformation services for easy application on library data to transform raw information into a more standardized, structures and system usable format. These services include system integrated Optical Character Recognition (OCR) engine, multilingual indexing, validation and sanitization of stored textual contents. The system also uses image processing tools to preserve the visual bitstreams in a storage efficient and more compatible format.

Like any capable data lake, a digital library should also offer efficient means of data storage both at local filesystem level, but can also leverage latest cloud storage technologies. There must be standard file sharing and data consuming end points to export or import information from external sources. In the NID platform the IIIF [15] content sharing API as transclusions feature is made available for end users. Besides direct content sharing the system is proposed to make use of standard citation services and metadata digestion and syndication using advance link data schemas [18] and conventional metadata harvesting and exposure based on the OAI-PMH standard [13]. The analysis stage proposed in NID framework applies different algorithms to generate actionable insights on defined needs of a digital information repository. These needs include detection of similar items in the repository at document and page level, identification and tagging of notable entities in library objects i.e. topics or image objects. The platform finally presents the information available in a library system to its users under a layered access control regime. The library contents infused with user data is made available together with enhanced search and discovery options.

1. Introduction to the NID system

A NID server contains a large set of documents. When uploading documents (or also later) a number of parameters can be set. These metadata parameters are built around the Dublin Core metadata element set [19]. Typical ones that govern the content presentation are the interface language for the document, or the menu options offered to users (like just a few very simple ones if one expects naive users for the document, or larger sets for advanced users.) These and other parameters can also be set by users with a special “Settings” button.

No matter what interface is offered, there is always the possibility to send for every page feedback that goes only to the person responsible for the particular document. Such a feedback form is activated by clicking at the Feedback icon (in line eight from left side as shown in Fig.2.)

  
Fig. 2. Various icons and options available to users of NID library viewer

Feedbacks can be sent on every page, and are sent anonymously. If persons sending feedbacks want an answer they have to explicitly state their E-Mail address.

Before looking in more detail at some of the most important possibilities users have, once they have chosen a particular document, we discuss what a typical NID server looks like. A good example is the server [20]. It allows to find relevant books by either entering (part of) the title, or an author’s name, or pertinent keywords, or by looking into one of the at the time of writing 18 subcategories, many with subcategories. Note that in addition there is a row of 15 books rotating above the categories shown. How often a book cover is shown can be controlled by a parameter when uploading a book.

Let us concentrate on the category “Zeitschriften” (newsletters) by a click. Choosing from the first category “options Magazine” we get the last nine issues of the magazine published by IIASA, an internationally funded research organization located near Vienna.

Notice in passing that for each magazine shown then number of accesses to the journal obtained so far is shown. The first magazine shown “A systems look at hunger” [21] has been accessed 4.800 times.

By clicking through the journal in the obvious way (using the big > and < buttons) one finds many pages where parts of them are surrounded by a rectangle. Those are not part of the journal, but were added later by users to insert additional information that we call “annotations”). Annotations can be (arbitrary long) pieces of text, links, pictures, clips, etc. You will notice that some rectangles are blue, others, red. This is an ingenious trick: red means that it includes a reference to a contribution on some server where the contribution has added a link to the page of the journal “A systems look at hunger”. Thus, anyone on any webserver linking from a relevant contribution to one dealing with a similar topic in an IIASA journal (and notifying IIASA of this) gets as “thank you” a link back! Thus, similar contribution on servers all over the world can profit by linking to IIASA journal: This is clearly responsible for the large number of accesses now obtained, and may be a something to be copied by many other servers.

Note that annotation can be made by someone anonymously. When loading up a book, this may be allowed or not. The owner of the document is informed when an annotation happens, can allow immediate visibility or only after checking it, and can change or erase the annotation, or only make it available to a certain group of persons. We will return to the issue of groups a bit later, but first we will address the most important issue: How can one create annotations.

To create an annotation, you have to click at the pencil icon called “Create simple annotations” in Fig. 2. If this is not visible, you are not allowed to add annotations. Probably you have not logged in and are not allowed to make anonymous annotations. In this case, use the login icon (the first one at right side in Fig.2) to register (only the first time) and then log in. Once you have clicked at the annotation icon, you can draw a rectangle around some area of the screen. In doing so, the form shown in Fig.3 is displayed.

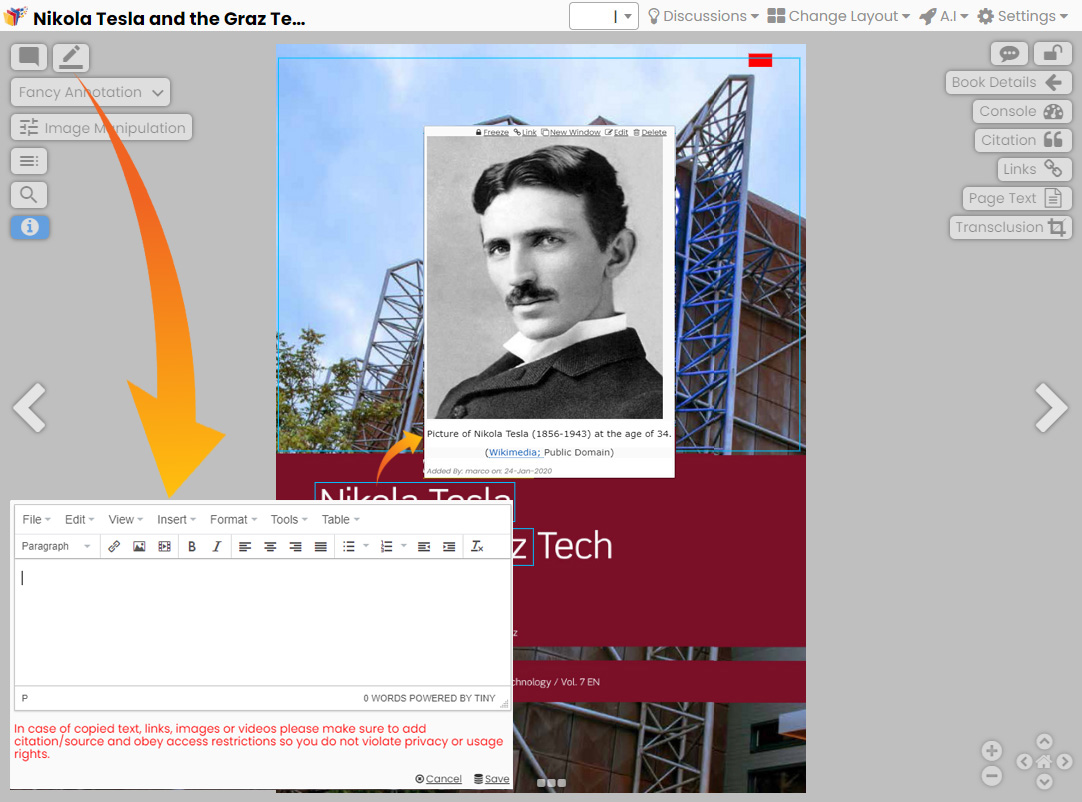


Fig. 3. Annotation addition process, showing the data entry form and annotation displayed in NID viewer

Now you can just type in any text (possibly making it bold or preceding it with a blank line to increase readability) and click at save, if you only want to add some textual annotation. If you do so, make sure that you are not plagiarizing by copying from a source without proper citation.

If you want to add a link, click at the annotation icon also shown in Fig.2 and 3, fill in the URL of the link and in the second line a few words with an explanation to tell the users what they will get when they follow the link. If you want to add a picture or clip, click at the icon next to the link icon.

You will get the form shown in Fig.4. Note that in this is case you are not supposed to upload a picture, but the source has to be the address of a picture on some other website. Such address can usually be obtained by a right mouse-click at the picture and using the URL shown behind “Copy graphic address” (German: Grafik Adresse kopieren).

The figure number and caption should be typed below the illustration in 8 pt and left justified [***Note:*** One-line captions of length less than column width (or full typesetting width or oblong) are centered]. For more guidelines and information to help you submit high quality artwork please visit: <http://www.elsevier.com/artworkinstructions> Artwork has no text along the side of it in the main body of the text. However, if two images fit next to each other, these may be placed next to each other to save space. For example, see Fig. 1.

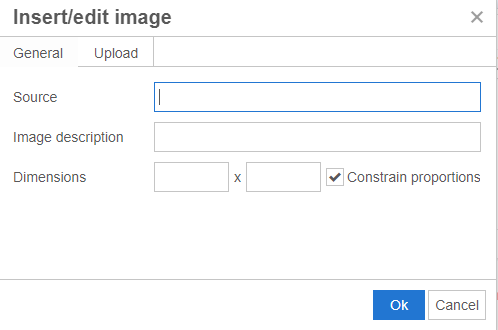


Fig. 4. How to add pictures and clips as annotations.

You may also add a brief description of the picture. Concerning dimensions, we suggest you use 560 as maximum in the first field (the second field will be adjusted by the system).

To be polite, add a line indicating the source you got the picture from. Doing it this way means that whenever the page with annotation is used, the picture is fetched from the external server. Such links are as legal as any other links. Jurisdiction is that any link (ordinary one or link to a graphic address) is permitted, if the organization from which you get the picture is not damaged by what you are doing. This means that links to any text or to any picture are OK, if there is a direct way to get to it: But if you “have to wade through” advertisements to get to it, it is hurting the organization offering the picture, since by directly going to it, it loses advertisement revenue (!!!).

In the form in Fig.4 you can also use the upload feature, but in this case you must have the right to use the picture: because you photographed it (in a place where this was allowed i.e. not of a piece of arts in a museum, or of persons who did not give you permission, etc.). But pictures in the public domain are OK, or pictures under Creative Common license by mentioning the correct spot and license.)

An annotation can be a combination of pieces so text, clips and links in any order. Because of this, annotations offer the option to “Freeze” them (to drag them around to a place where you can better read them), or to open them in a new window. This is particularly recommended for video clips, and before starting them, switch to full screen mode!

1. More features of NID from the users’ point of view

It has been explained how books can be located: Either by searching, or going to the appropriate category or subcategory and looking for what is of interest.

Of course, you also must be able to locate items in a book. NID uses Apache’s Solr open source search and indexing platform [11] as basis for searching. A search can be initiated on any page of any book using the search icon (magnifying glass) shown in Fig.2.

Figure 5 show a search menu where user can perform advance search by using boolean operators, wildcard searches and setting search scope by setting various filters.

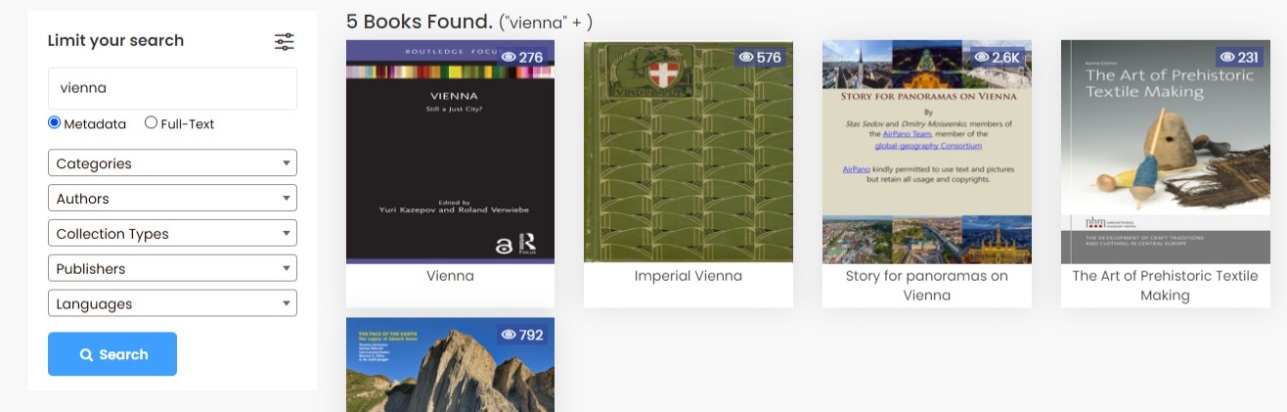


Fig. 5. Search on meta data and all library documents with options to limit the search output.

NID offers four unusual extensions for searching. First, it is possible to not just search in one book, but in all books of a (sub)category with a single query; second, it allows to not only searching in one document, but also in annotations of documents (just choose the option when clicking at the search icon): this was quite a technical challenges, since annotations are added or deleted dynamically, so the usual pre-processing for fast full text searching cannot be used; third, the full-text search is not just applied to text, but also to text that is part of pictures (of course this does only work if the text is horizontal and clearly recognizable); fourth, stunning (but still limited) is the possibility to not just search for textual terms but for objects. An example will explain this best: If you search for “person” in [21] Fig.6 shows a part of the list obtained.

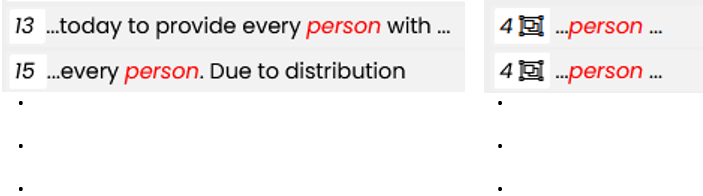


Fig. 6. Search options within documents; Object search results vs text search results.

The system has found on page 13 and page 15 the word person, but on e.g. page 4 (indicated by the icon before the word person) it has located two persons as such.

Have a look at them. Is it not impressive that the software identifies a person, even if seen from the back, or almost hidden by a dress, or just showing parts of a face from whatever angle, etc. This is based on work in Artificial Intelligence. We have used open source data. The feature is available in NID at this point only for some 80 types of objects. The object detection limit is mainly due to the use of a pre-trained knowledge model [24] for library’s object detection and annotation component. In the future our research groups plan to extend this object classification to more problem specific domains e.g., medical images, architectural designs etc.

NID also allows to start a discussion on any page. If a discussion on a page has already started, the word Discussion is lit up as seen on page 5 of [22]. The very top part of the discussion is shown in Fig. 7.

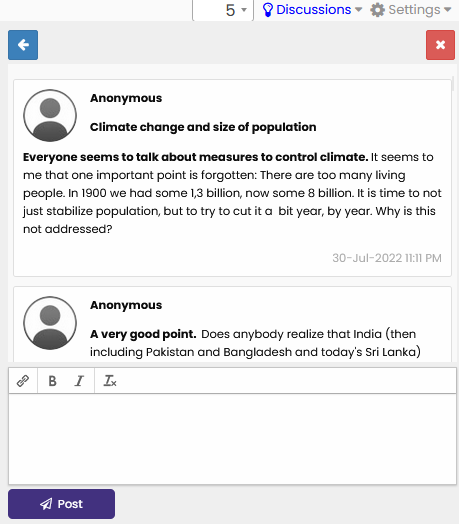


Fig. 7. First entries of a discussion.

The discussion at issue is actually quite long, so one can scroll down a lot, or even enlarge the window using the little arrow in the left upper corner.

NID also provides two types of quizzes are available, one just for self-checking, to find out if one has understood previous material; the other type to actually evaluate the knowledge of users, possibly influencing grades or such.

A particular important aspect is that groups of users can be defined. Different groups can have access to different documents. Even the same document might look rather different for different groups, since they might see completely different annotations.

This leads to important different applications of which we will mention a few.

Example 1: Consider a complicated mathematical formula.

* + It may be shown as such to university students in mathematics.
  + But it must be shown with a short explanation to senior students in high school.
  + And it may be shown only with extensive explanations, including maybe links leading to still easier to understand material for mathematically less experienced persons.

Example 2: Consider an application of a car company. A picture of a new car is shown.

* + For the general population, annotations will explain some great features.
  + For sales persons, annotations might explain advantages over rival products.
  + For engineers, annotations will give more detail on technicalities.
  + For repair crews, annotations will give hints how to handle potential problems, those annotations getting longer as more incidents happen.

Example 3: A person (here Thomas Alva Edison) may be shown to different groups in very different moods, as demonstrated on page 29 of [4].

Example 4: NID is an ideal tool to encourage young students to get involved in reading, eLearning and information technology, see the invited talk [4]. This is just one of the number of examples NID has been used for E-Learning.

A nice feature of NID is the table of contents icon in Fig.1. If it is shown (next to the search icon) on a page a click shows the table of contents on that page, not requiring to look for it at the beginning or end of the document, where it is usually located.

When applying NID it turns out that some users are not much familiar with internet and IT interfaces, others are. To satisfy both the “naïve” users, the “average user” and the “expert” one can use the Settings option (see Fig.2) to select the basic, the short or the full menu, overriding the default option defined when uploading the book.

There are many other important features of NID, some will be dealt with in the following sections and are just mentioned here for completeness’s sake:

* + Annotations can come in many forms, colors, shapes, texture of frame etc.
  + Annotations cannot just be added to text but anywhere, like in pictures or drawings. One interesting case was the complex architectural drawing of a living room. When moving the cursor to the elaborately designed window a video starts playing, showing what users might see in the future when they look out of the window of the finished house
  + Parts of a page can be defined (“transclusions”) and can be used in many ways without copying parts of the page (coordinates are just associated with the part of interest).
  + NID allows the incorporation of arbitrary high-resolution pictures.
  + If one wants to quote a part of a page properly, citation in various formats is supported.
  + Access to underlying page text if copy rights of objects allow text content copying.
  + If one has a map with many annotations highlighting special features on can use a “link” feature to the annotation on the map: this highlights the correct location.
  + Very exciting is the possibility to allow to place one or more pages also from different servers on the same screen and work with them independently (available via the “change layout option” in the full menu) allowing to compare e.g., a reproduction of a classical picture on a sever at Stanford with one in Bavaria, using the zoom option to examine the kind of brush-stroke used. This often allows to determine if the copies were done by the same person.

1. NID management functions

The back-office portal of NID library system has a variety of tools available to library administrators, editors, and privileged users for the overall management of the system. The system provides a reporting dashboard for a holistic view of what is available at the system. The library management functions are available on both the main library website along with exclusive administrative functions on a back-office portal.

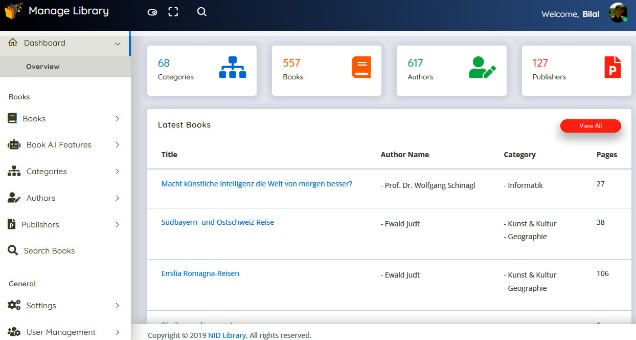


Fig. 8. Backoffice portal and main site management dashboard view of the NID library

The library management back-office portal has following cascaded function menus.



Fig. 9. Backoffice NID management portal features menu links.

The library administrators can perform data curation with easy form-based upload of variety of documents along with metadata. Data transformation services are activated automatically to convert the added documents into a standard image-based format. The added repository data is made available to the direct users of system and other peer library platforms using the IIIF standard. The Book menu also gives access to verify the API based annotation addition to library contents. There are several advanced data processing steps available through Artificial Intelligence (A.I) features menu. The system makes use of a built-in neural network based Optical Character Recognition (Tesseract OCR[[9]](#footnote-9)) engine to detect text in over 100 languages. The other AI services in the system includes Natural Language Processing (NLP) based services to detect named entities from the unstructured text of library documents, e.g. table of content entries, topics and automated tagging and linking of similar documents. The object detection and search index supplementing service make use of You-Only-Look-Once (YOLO) algorithm [23] with COCO dataset [24] to recognize and tag 80 object types in images of NID library system. The data organization and system management features include structured content category management, profile handling of authors and publishers. The system configuration settings related to the messaging, content types, digital rights management and role-based user access and authorization management are also part of administrative portal.

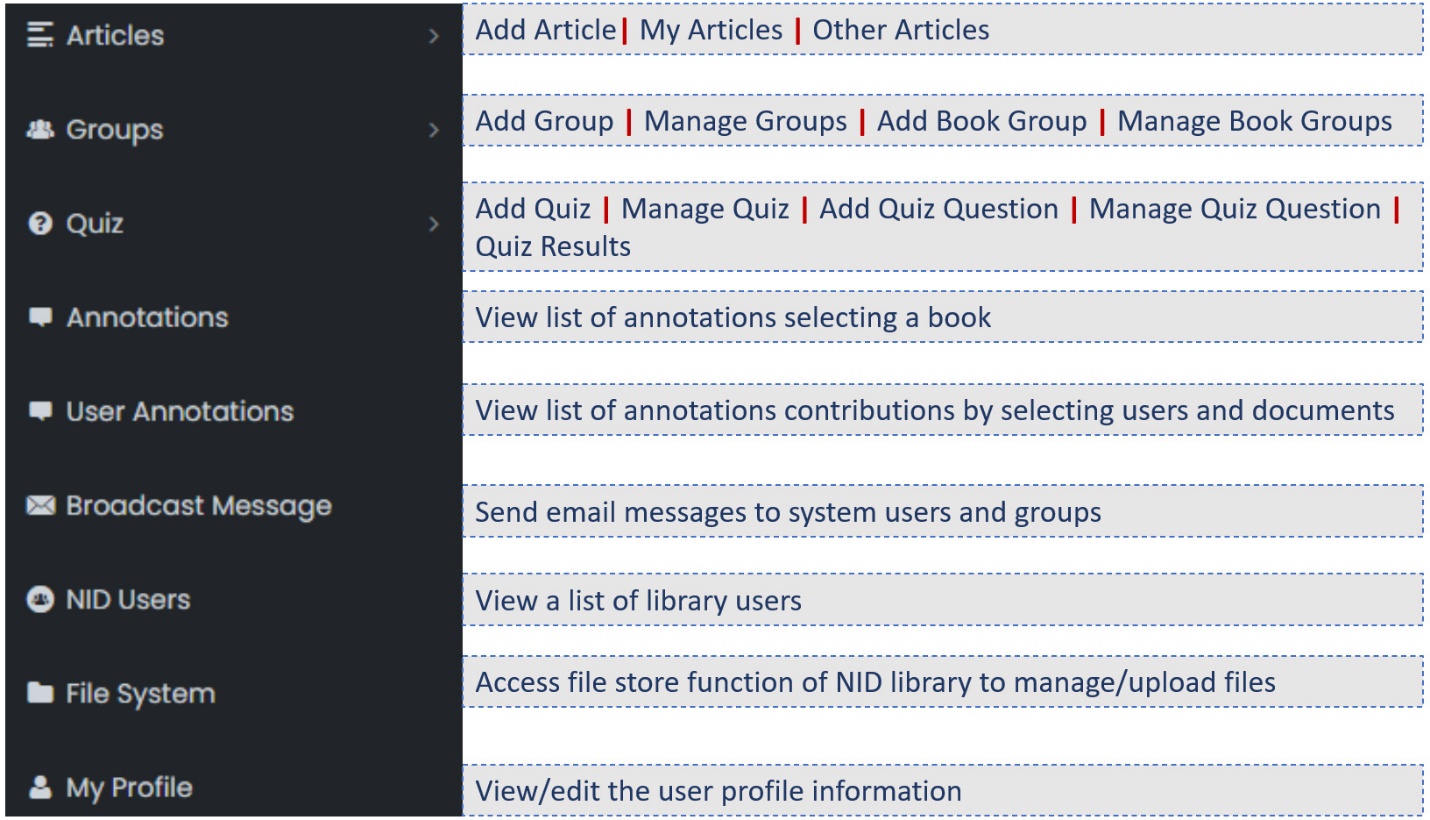


Fig. 10. Additional system services available to privileged users in main library website.

Besides the exclusive management portal, the system and data management features are also made available for the privileged users in the main library frontend portal. The console section of library users has a number of features to manage user blog entries as articles of the system, access to user group functions and assign access rights of marked contents to selected user groups. The user with rights can create quiz, opinion polls, view annotation listings, broadcast messages to users and groups, use system file store functions etc. Figure 10 lists some of the management features available on the main library site.

1. Conclusions

Our work shows that controlled integration of user data into digital libraries turns the library contents into true living entities. Multidimensional and growing data of such digital libraries has the characteristics of a data lake that demands enhancements beyond intuitive information processing and access methods. The second major aspect of our work suggests that we can leverage the use of big data analytics if the related technologies are seamlessly integrated with general workflows of a library system. The practical integration of analytics tools with standard library processes will ensure broader use of enabling technologies by general library staff and users. The Net Interactive digital library framework we proposed improves the traditional data lake architecture of digital libraries by complementing it with data warehousing characteristics. The analytics performed on library and user contributed data provides insights for more coherent information access rather than the typical information delivery.

References

[1] Web Sites Need Much More User-Driven Interactivity (Mauer, H., Zaka, B.), IPSI Transactions on Internet Research, vol 18, no. 2, (June 2022), 63-68

[2] Digital Libraries, IPSI Transactions on Internet Research, vol 18, no. 2 (June 2022), 1-4 (J)

[3] Large webistes have to become more interactive: <https://www.nid-library.com/Home/ViewBook/716> (Keynote 2022)

[4] Digital Libraries of the Future: https://nid.kinderphilosophie.at/Home/ViewBook/315 (Invited talk)

[5] Austria-Fourm: https://austria-forum.org

[6] Anna, N.E.V. and Mannan, E.F. (2020), "Big data adoption in academic libraries: a literature review", Library Hi Tech News, Vol. 37 No. 4, pp. 1-5. https://doi.org/10.1108/LHTN-11-2019-0079

[7] Ravikumar, S & Hazarika, Hirak. (2020). Comparative Study between Omeka, DSpace and E-Print: A Special Reference with Plugin Features. INFLIBNET Centre, Gandhinagar ISBN: 9789381232095 http://ir.inflibnet.ac.in/handle/1944/2330

[8] Verma, L., & Kumar, N. (2018). Comparative Analysis of Open Source Digital Library Softwares: A Case Study. DESIDOC Journal of Library & Information Technology, 38(5), 361-368. https://doi.org/10.14429/djlit.38.5.12425

[9] Mathieu Andro, Emmanuelle Asselin, Marc Maisonneuve, Digital libraries: Comparison of 10 software, Library Collections, Acquisitions, and Technical Services, Volume 36, Issues 3–4, 2012, Pages 79-83, ISSN 1464-9055, https://doi.org/10.1016/j.lcats.2012.05.002.

[10] DSpace an open source repository software, https://dspace.lyrasis.org/

[11] Solr, An open-source enterprise-search platform, https://solr.apache.org/

[12] Omeka, An open source web-publishing platform, https://omeka.org/

[13] OAI-PMH, Open Archives Initiative Protocol for Metadata Harvesting, https://www.openarchives.org/pmh/

[14] Islandora, An open-source software digital repository system based on Fedora Commons, Drupal and a host of additional applications, https://www.islandora.ca/

[15] IIIF, International Image Interoperability Framework Open standards for digital objects delivery, https://iiif.io/

[16] EPrints, The first open-source repository software, https://www.eprints.org/

[17] GSDL, Greenstone Digital Library software, https://www.greenstone.org/

[18] G. Loh, "Linked data and IIIF: Integrating taxonomy management with image annotation," 2017 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC), 2017, pp. 50-55, doi: 10.23919/PNC.2017.8203521.

[19] Dublin Core™ Metadata Element Set, Version 1.1: Reference Description, https://www.dublincore.org/specifications/dublin-core/dces/

[20] NetInteractive Documents Server, https://www.nid-libray.com

[21] A systems look at hunger, https://www.nid-library.com/Home/BookDetail/127

[22] NID library discussion example: https://www.nid-library.com/Home/ViewBook/483/5/view

[23] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[24] COCO, A large-scale object detection, segmentation, and captioning public dataset, https://cocodataset.org/

1. <https://web.hypothes.is/> [↑](#footnote-ref-1)
2. <https://www.saashub.com/compare-dspace-vs-omeka> [↑](#footnote-ref-2)
3. <https://whatcms.org/compare/Omeka-vs-DSpace> [↑](#footnote-ref-3)
4. <https://dl.acm.org/> [↑](#footnote-ref-4)
5. <https://ieeexplore.ieee.org/> [↑](#footnote-ref-5)
6. <https://www.sciencedirect.com/> [↑](#footnote-ref-6)
7. <https://link.springer.com/> [↑](#footnote-ref-7)
8. <https://www.scribd.com/> [↑](#footnote-ref-8)
9. <https://en.wikipedia.org/wiki/Tesseract_(software)> [↑](#footnote-ref-9)