



Digital Libraries: Stimulating Open Innovation Through Crowdsourcing And Transparency

Jessica Colaço

E-mail: jessica[at]ihub.co.ke

&

Hilda Moraa

E-mail: hilda[at]ihub.co.ke

iHub - Nairobi's Premier Innovation Hub

Nairobi, Kenya

Meeting:

94 — *Creating a culture for innovation and change* — Management and Marketing with Academic and Research Libraries

Abstract:

With the rapid growth in technology and digital information, digital libraries are often seen as a replacement to archival paper repositories. Digital libraries encompass the storage of information in electronic form and put content onto the online world, making the information more open and accessible. Digital libraries come with the benefit of little to no time delay to access information. Users can also access new knowledge, share and learn new things, thus improving individuals' skill set positioned to innovate with the content.

In this paper, we discuss how digital libraries can stimulate open innovation through the power of crowdsourcing and transparency by sharing information and offering free access to recent literature. The first section provides insights on the need for researchers to make their work more visible on a digital platform to promote open access to other agents participating in the network. The second section highlights the relevance and benefits of digital libraries in stimulating open innovation through crowdsourcing, thereby promoting a culture of innovation creating change.

Keywords: digital libraries, open innovation, knowledge, crowdsourcing, change

Introduction

Many library institutions around the world and government organizations are engaging in projects incidental to digital libraries. A good example is the United States Congress library [1], which has been pushing this initiative from year 1998. Another organization that has tried their hand on this endeavor is Google [2], which came up with Google books library project in 2004. This project was conceived on the premise that it would be an overnight success based upon on its internet searching dominance however, this project faced hiccups as a result of petition from the publishing industry. Other initiatives include the California digital library, National Digital Library Program offering including books, articles, pamphlets, personal papers, legislative documents, prints, architectural drawings, photographs, maps, sheet music, sound recordings, and movies.

Digital libraries involve standard initiatives such as cataloging, Metadata, data definitions, data interchange, search and interface for the display of the data (source). Focusing on online access to data and analytical tools for research, education and digital government as examples of online content, is important to generate content, analyze and interact with complex data in these digital libraries.

However building an effective digital library is the most challenging thing as identified by most of the existing digital libraries such as the stuff of the National Digital Library Program at the Library of Congress (source). In some cases, there may be no technology solution to the challenge, but through sharing of ideas, new thinking need to emerge to help institutions such as

the Library of Congress formulate policy and get more skills from diverse users that have creative and innovative minds and can devise solutions to these challenges.

Crowdsourcing, a combination of the terms “Crowd” + “Outsourcing”, is the practice of companies, business owners or even individuals, making an open call to a broad community of people to help them solve a particular problem (source). In return drive down the cost of solving a problem quite substantially, in addition to being able to receive a huge amount of ideas and suggestions.

Crowd sourcing is a relatively new concept that came into being in 2006 (source). This can be attributed to the rise of a networked world through the web. Since the inception of the World Wide Web, technology has changed the way humankind approaches and solves problems. One of this has been through crowd sourcing where an individual will put up a challenge to the whole world.

Crowd sourcing is based on the concept of wisdom of the crowd. Wisdom of the crowd is a process of taking into account the collective opinion of a group of individuals rather than a single expert to explain a question. This phenomenon works best when you have a connected world through the World Wide Web (source).

Crowd sourcing has now evolved to include crowd funding where websites have been put up to request funding for projects. Excellent examples of crowd funding platforms are in Barack Obama’s campaign contribution and artistic projects. Other trends are crowd voting used in

reality television programmes to aggregate the society opinions and it has been stretched to encompass aspects of democratization. Since crowd sourcing has inputs from a cross section of society pundits have come to generalize crowd sourcing as having elements of democratization

Crowdsourcing and Digital Libraries

There is a deep relationship between the challenges of digital libraries and crowdsourcing. Through crowdsourcing, the digital libraries enlist a multitude of users to help solve a variety of problems transforming human computer interactions by outsourcing specific activities to a community through an open call.

Examples of crowdsourcing goals for libraries could be: getting users to mark the errors in their catalogues; rating the reliability of information/records; adding information to records; verifying name authority files; adding user created content to collections; creating e-books; correcting full text; transcribing handwritten records; and most especially describing items that they have not made accessible because they are not catalogued/described yet. A prime example of this is photographs. The normal procedure in a library is that a photograph is not digitized until it has been catalogued. If instead it is digitized first and users are given the chance to describe the content this would radically open up access to a lot of 'hidden' and difficult to describe photographic collections (source).

Crowdsourcing provides a paradigm in which the components of digital libraries become integrated; that is users can play an active role in producing content as well as consuming it

hence providing collective knowledge, strengths of different kinds of communities with linked data, automated analysis and openness.

Current digital libraries implementation, require designing tools that facilitate the enhancement of cataloging or descriptive information by incorporating the contributions of users, hence taking advantage of the distributed expertise. This can be enhanced by the potential of crowdsourcing through collaborative tools, that could allow the crowd of users enhance the cataloging of the content, add comments, adding value to the resource. The vision for digital libraries should include fluid, easy access to a wide variety of materials. This is often in conflict with the duties of libraries and archives entrusted with care and management of materials that may be subject to privacy rights or other needs for security.

The more academics and researchers publicize their insights and useful materials for the community, it could get thousands of volunteers with internet connection to assist build local capacity, local content and improve their skills set. A shift in thinking is required by librarians and researchers to fully embrace the potential of crowdsourcing by giving other users the freedom to interact, share, collect and add value to their insights or data into their collection hence create relevance in society.

Researchers need to focus on global reach as their vision, rather than just local reach about crowdsourcing their data. This can be enhanced through creating a centralized pool of local volunteers who can interact, learn and create more content to their digital repositories. Focusing

on a global reach, digital libraries, allow users to access new knowledge in multi-modal format anywhere, anytime breaking down the barriers on institutional walls, culture and distance.

Building a multi-modal infrastructure of sharing information, not only breaks the institutional walls but also supports further the educational processes, adding value to the digital libraries which are by definition complex systems. Interoperability- the ability of the digital libraries and the traditional libraries to work together, is a key step in ensuring the digital collection grows in a way that allows users to navigate through different sources within an integrated single environment which can be analyzed from organizational, semantic, and technical levels.

On the other hand, the role of open access has been crucial to broaden the function of Digital Libraries within the research community by making their research more transparent facilitating more innovations from their insights.

Crowdsourcing has not been attempted to a larger scale by both digital and traditional libraries to date but could prove to be the most useful tool in the near future due to the increase in digital local content and open data for them to be successful. If the public can act as volunteers and given a high level of trust, control and responsibility then they will actively participate in contributing to creating a portal of more meaningful content leading to a natural way of building capacity through access to a pool of experts with a mix of skill sets.

There is need to harness digital volunteers to transcribe, create, enhance and correct text, images and archives. Crowdsourcing offers a cloud-based workforce that is available while improving

time to analyze content to the users. Most important there needs to have proper management and organization of the content deposited in the digital libraries for it to be of value.

Benefits Of Implementing Crowdsourcing In Digital Libraries

Saving Huge Costs

By allowing the public to share and analyze the digital information it saves huge costs by the librarians or institutions achieving their goals much faster through additional capacity unlike if they would have done it on their own.

Building a wider virtual network

Libraries can create more user groups and virtual users who share and interact with the online content on a frequent basis. These users can also learn, engage and transfer knowledge if actively involved with the digital libraries hence utilizing the knowledge, expertise and interest of the community.

Improving the quality of data

Through open access to the online content, the librarians- research institutions, companies, schools can be able to add value to their data, example through comments, tags, ratings, reviews making the data more discoverable in different ways for a more diverse audience.

First-hand communication

The librarians get first-hand responses and insights from the users and digital volunteers that helps them to verify their assumptions by listening to the crowd. This also demonstrates the

relevance of the digital libraries in the community by the active public involvement and responses.

Strengthening and building trust

The librarians strengthen and build loyalty of the users to the digital library. Users feel free to share information and their contributions without thinking of making money hence encouraging a sense of public ownership and responsibility towards cultural heritage collections.

How digital libraries can promote the culture of open innovation

It's for sure that through crowdsourcing in the digital libraries, a culture of open innovation is created through discovery of new knowledge that can be used to develop new innovations; growth of digital collection, creating a pool of data that can be outsourced beyond the local boundaries to make decisions that lead to positive impact.

Digital libraries have played a role of acting as an intermediary that brings users and information together on a single platform. Users can share information at different levels and create meaningful data sets out of it.

Through data mining- extracting the knowledge to meet the unmet needs of the users is important (source). Hence there is need for the librarians to go an extra mile in understanding unsupervised learning techniques such as clustering, and composite term discovery techniques, which are useful innovative skills while extracting the knowledge.

Digital libraries create database repositories for information retrieval, which can be openly accessed through the process of combining internal and external ideas, as well as internal and external paths to market to advance the development of new technologies. This can be nurtured through an environment where users can share information, ideas and comments hence stimulating a culture of openness that can lead to capacity by new ways of doing things and meeting the changing and ever-increasing users expectations.

Challenges of Crowdsourcing

Despite the powerful approach of crowdsourcing that digital libraries can implement, it also has its own fallbacks if not well executed and managed by the people involved. The acceptance and openness of such applications has made it easy to contribute poor quality content. However, various solutions have been proposed for the Web-based domain, to assist with monitoring and filtering poor quality content, but these methods fall short when applied to ubiquitous crowdsourcing, where the task of collecting information has to be performed continuously and in real-time, by an always changing crowd.

Crowd

Most crowd outsourcing projects have large masses of people, which can be hard to manage and control.

Volunteers

Having a volunteering network that crowdsources, finds many of its participating non-profits struggle because they haven't written a compelling challenge for their community. That's when

the network's community manager steps in to resolve these issues so that participants can maximize their exposure.

Costly

It takes time to crowdsource effectively. Indeed, although an organization might outsource innovation, it cannot outsource the labor necessary to be successful. If there is a lack of structure, you can expect to invest even more time. "When you don't have systems and structures, and you try to do ambitious things, it takes time," said Brian Reich, co-founder of WeCanEndThis.

Quality of data

Today, the technology for digital conversion is, at best, emergent and often forces a library to choose between risking damage to precious originals or producing the highest quality reproductions. There are few established standards or best practices and a shortage of tools for the objective measurement of reproduction quality. There is a need for more automated support for capturing in explicit data structures the navigational and organizational clues implicit in printed works through page numbers, tables of contents, and indices (source).

Standardization

What types of protocols and what degree of standardization on types of digital libraries will achieve a balance between feasibility of widespread implementation and coherence of access e.g, through distributed search ? How can distributed digital libraries best safeguard the rights associated with content (including rights of privacy and conditions imposed as well as copyright) while still providing the broadest possible access?

Recommendations

Make sure you have the right crowd.

While the digital libraries initiatives want to have a diverse crowd it needs to be knowledgeable and/or passionate about the topics. More to crowd management, digital library projects should focus on inviting relatively small interested and engaged members of the public to participate in the projects and adding resources to the virtual collection hence easier to control the inflows and outflows of the data.

Specific and well-structured tasks for volunteers.

People don't respond well to vague or overly wide tasks. Digital libraries need to break down the tasks into small chunks that the volunteers involved can perform in a short time, leading to a better quality of data deposited in the digital repositories.

Rewards Appropriate for the Task.

Digital libraries need to understand the reward mechanism. Some tasks are inherently boring and people will expect financial compensation (Mechanical Turk), others are more fun/interesting and people will do it for their reputation.

Crowdsourcing needs structure and rules.

While the crowd craves freedom, people need to be told the rules of engagement and how to participate. These rules have to be clear, empowering the crowd and directive in their end result. While enforcing the rules in place, community management becomes easier. Rules can be broken, the spirit of a contest can be thwarted, and unforeseen behaviors can necessitate action.

If the larger community's interests are to be encouraged and sustained, then community-centric behavior needs to be enforced. That means the organization will have to use some of its community management resources to enforce and even build new rules.

Motivation

Towards a richer crowdsourcing in digital libraries they have to be scaffolding users into expertise and above all making users understand the motivation of adding their data via wisdom of crowds where the librarians empower and consult with users. Most important, the librarians need to consider certain factors while sharing and creating virtual collections: quality assurance, cataloging and organizing, extracting meta data and conceptualizing the data into valuable content.

Conclusion

Crowdsourcing has not been attempted on any significant scale by libraries to date, but could prove to be the most useful tool a library can have in the future (source). If the facts known about crowdsourcing and the benefits outlined in this article are applied any crowdsourcing project such as a library is likely to be successful. If the public is given a high level of trust and responsibility they will respond with loyalty and commitment hence a culture of innovation can be created through a symbiotic pipeline in the massive digitization of information, storage and dissemination. Crowdsourcing has the potential to revolutionize information collection and processing systems by enabling in-depth, large-scale, and cost-effective information gathering, more accurate techniques for information extraction from data.