"Lo-Fi to Hi-Fi": A New Way of Conceptualizing Metadata in Underserved Areas with the eGranary Digital Library

Deborah Maron	Cliff Missen	Jane Greenberg
UNC Chapel Hill, USA	UNC Chapel Hill, USA	Metadata Research Center
maron@live.unc.edu	missenc@unc.edu	janeg@email.unc.edu

Abstract

Digital information can bridge age-old gaps in access to information in traditionally underserved areas of the world. However, for those unfamiliar with abundant e-resources, their early exposure to the digital world can be like "drinking from a fire hose." For these audiences, abundant metadata and findability, along with easy-to-use interfaces, are key to their early success and adoption. To hasten the creation of metadata and user interfaces, the authors are experimenting with "crowd cataloging." This report documents their experimental and intended work and Maron's Lo-Fi to Hi-Fi metadata pyramid model guiding a developing metadata initiative being pursued with the eGranary Digital Library, the technology used by Widernet in a global effort to ameliorate information poverty. The work in development, the Lo-Fi to Hi-Fi model, has principles adapted from technical design processes and tried and true methods within metadata creation such as crowdsourcing. It attempts to reconceptualize the metadata modeling paradigm and aligns with research that has shown that community-based librarians are better poised to identify culturally congruent resources, but many require significant training in metadata concepts and skills. The model has amateurs (mostly students) crowdsource "lo-fi" terms, which domain experts and information professionals can curate and cull in "hi-fi" to enhance findability of resources within the eGranary while simultaneously honing their own computer, information and metadata literacies. Though the focus here is on Africa, the findings and practices can be universalized to off-line collections around the globe.

Keywords: information literacy; computer literacy; WiderNet; eGranary; Africa; metadata literacy; crowdsource; crowd catalog; folksonomy; LIS education; industrial design; hi-fi prototype; lo-fi prototype

1. Introduction

Many citizens of first world nations have become accustomed to, and even routinely take for granted, accessing information immediately and easily through the Web and mobile technologies. With years of experience under their belts, they blithely operate search engines, barely recalling how they developed their search skills over hundreds of hours of Internet use. However, not all individuals have that luxury, with over five billion people lacking access to Internet resources. Information poverty, begotten of a lack of technology and knowledge of how to use it, is a pervasive problem that affects quality of life, as well as development of crucial 21st century skills like information and computer literacy, for children and adults around the globe.

Lack of the aforementioned literacies is an information and library science (ILS) issue that impacts many librarians and library workers in all aspects of work including metadata creation and management. More specifically "metadata literacy" is adversely affected by the conditions in environments in which workers lack sufficient computer or information literacy.

The focus of the work reported here is the rural, indigenous sub-Saharan library, where a dearth in literacy is exacerbated by a lack of connectivity to the Internet. Driving questions include the following: How should metadata for essential resources be developed in these and similarly afflicted regional libraries? Also, how can the knowledge and terminology of a particular society be leveraged within these regions to create metadata? We propose that concepts

from the informal technical design process employed by WiderNet over the last five years to develop topical portals for end users (in medicine, nursing, public health, rural agriculture, life skills, disability rights, etc.) be formally expanded and evaluated to provide a first-rate framework for conceptualizing and delivering metadata formation for people in underserved communities. This prototyping was further defined as "hi-fi" and "lo-fi" and extrapolated into a model by Maron in 2014.

We present this model in the context of WiderNet@UNC, an initiative to bring Internet content to places worldwide that are not connected, via massive hard disks of material that mirrors what the Internet has available. WiderNet's hard disks of material, called eGranary Digital Libraries, are currently used in over 800 locations and contain 35 million items each; while the majority of the documents can be searched using a built-in term search engine, only a fraction of the items are catalogued because the onus falls on the small team of paid and volunteer cataloguers to create records.

In evaluating logs from dozens of eGranary servers, it has been noted that users, generally unfamiliar with search engines, are more likely to use the limited catalog to locate resources. In most cases, 90-95% of the documents retrieved were listed in the catalog. Clearly, new users prefer well-cataloged resources.

This project report details the expanded concept of the Lo-Fi to Hi-Fi metadata pyramid, representing a method by which metadata can be crowdsourced and curated for resources by the very people that use and operate eGranaries within underserved areas of Africa in a tiered system; students and other general users in the respective communities identify folksonomic terms and useful resources as "suggestions" (lo-fi), which are then winnowed by domain specialists, approved, and finally become part of the canon of knowledge (hi-fi) in the hands of more expert catalogers. Hopefully this scheme would imbue metadata and other types of literacy in general users, scholars, practitioners and library professionals, and foster the creation of metadata in regions with eGranaries that critically need it so that more information can be found. As well, it is expected to reveal culturally congruent metadata that external agents can adopt and employ. The Lo-Fi to Hi-Fi metadata pyramid can also, if successful, be globally applied to other collections and digital libraries in communities facing similar obstacles; as such obstacles are fairly universal. Before delving into the method and model, it is imperative to go over definitions of terms and provide context for the problem.

1.1. Definitions of the terms

Information literacy is defined as "the set of skills needed to find, retrieve, analyze, and use information." Those who are information literate "have learned how to learn" and find information for virtually any task ("Introduction to Information Literacy," n.d.). Computer literacy is a term being continually redefined, but Childers writes that "a person is either computer literate or not based on how proficient they are at some basic computer tasks" ("Computer Literacy," n.d.). Finally, metadata literacy is a term coined by Erik Mitchell and concerns a person's ability to cultivate adequate metadata for digital objects (Mitchell, 2010).

1.2. Computer and information literacy in areas of sub-Saharan rural Africa

For members of communities around the world, computers are critical in terms of cultivating skills necessary to be an active, participatory member of the information age. In fact, the computer as a beacon of hope and its ability to revolutionize and improve many facets of an African citizen's life was recognized in the early 1990's by Oduaran and others, but a lack of computer literacy persists even today (Oduaran, 1991; du Plessis & Webb, 2012). This paucity of computer literacy begets information illiteracy, a problem pervading not just the general, indigenous rural populace in sub-Saharan Africa but the population of teachers and information workers as well (Jager & Nassimbeni, 2007) . This problem manifests itself in not only libraries but also in the issues for which information professionals are to provide information, such as the AIDS epidemic and prevention. Compounding a lack of computer and information literacy in

certain African libraries is a lack of metadata literacy, a skill not possessed even by many American library professionals (Park, Tosaka, Maszaros, & Lu, 2010).

2. WiderNet@UNC: Bringing information and literacies to the masses

2.1. Overview of WiderNet

WiderNet@UNC is a research program at the University of North Carolina, Chapel Hill that focuses on low-cost, high-impact uses of ICT and training modalities for under resourced communities worldwide. Its sister organization, the WiderNet Project, is a non-profit service program founded in 2001 that aims to bring educational digital content to places worldwide that lack adequate Internet connectivity. Using massive hard disks of material that mirror thousands of World Wide Web sites, WiderNet's eGranary Digital Libraries are currently used in over 800 locations and contain 35 million items each.

2.2. Metadata Principles and Challenges

Utilizing Dublin Core and Library of Congress (LoC) standards, WiderNet cataloguers have developed a protocol for adding metadata, highlighting resources, and creating user-centric collections, for e-Granaries. However, only a fraction of the items are findable through the catalogue because the onus falls on the small team of student and volunteer cataloguers to create records. Many more resources could be found and privileged if there was more metadata available, and if users and library workers in Africa were contributing to the process.

WiderNet has worked with partners in developing countries to create custom user-centric "portals" from catalogued records. For example, in 2008 they launched collaboration with the medical college at the University of Zambia and the School of Public Health at the University of Alabama to create a portal for teaching health sciences in Zambia. Over 1.5 million documents were garnered from the inputs of dozens of educators and practitioners around the world (lo-fi) and then WiderNet librarians cataloged over 2,000 items that had been highlighted by the expert advisors. Then, in consultation with their Zambian counterparts, they mapped 600 cataloged items to the Zambia national medical curriculum. Students and instructors were quick to adopt this curated collection and eventually insisted on it being installed in dozens of other institutions where they practiced and taught.

In another example, they worked with the United States International Council on Disabilities and over 100 advocacy groups around the world to create a dozen portals around disability rights and resources for persons with disability. Over 2.5 million new resources were added to the eGranary library and mostly librarians in the U.S. and Europe catalogued 4,000 items.

3. Hi-Fi/Lo-Fi Prototyping: Can the principle be adapted to metadata?

A prototype is defined by Merriam-Webster as an "original or first model of something from which other forms are copied or developed", or a "first or early example that is used as a model for what comes later" ("prototype", n.d.). It is proposed that methods involving high and low fidelity prototyping (hereafter called "hi" and "lo" fi) be used as a model for creating and curating metadata for resources in eGranaries. Egger describes prototypes thus:

> Low-fidelity (lo-fi) prototyping is characterized by a quick and easy translation of high-level design concepts into tangible and testable artifacts. Lo-fi is also know as low-tech, as the means required for such an implementation consist, most of the time, of a mixture of paper, cardboard, post-it notes, acetone sheets etc. A clear advantage of lo-fi prototyping is its extremely low cost and the fact that non-programmers can actively be part of the idea-crystallization process.

> At the other extreme, high-fidelity (hi-fi) prototypes are characterised by a high-tech representation of the design concepts, resulting in partial to complete functionality. High-tech, however, implies higher costs, both temporal and financial, and necessitates good

programming skills to implement the prototype. The main advantage of hi-fi, high-tech prototyping is that users can truly interact with the system, as opposed to the sometimes awkward facilitator-driven simulations found in lo-fi prototyping. Obviously, there is a continuum from low to high-fidelity prototyping that usually stretches out from early to late design.("Lo-Fi vs. Hi-Fi Prototyping," n.d.)

Lo-fi prototyping, which Egger explains is "cheap, fast and accessible to non-programmers" aids participants of all levels of computer and information literacy to assist in idea and product formation, and is therefore proposed as the first step of the pyramid process, outlined in section 4.

4. The Lo-Fi to Hi-Fi Metadata Model: Crowd-Cataloguing the eGranary

This section introduces the model that is being developed, a study that is a result of meetings and the exchange of ideas at UNC Chapel Hill.

4.1. Tier 1: Lo-Fi

(lo-fi): Crowdsourcing a folksonomy

Example of Participants of Tier 1 populace: Mitchell's metadata literacy study focused on the ability of college students to create and curate metadata (Mitchell, 2010). It therefore is proposed here that college students form the majority of the lower tier of the model, the "lo-fi" stage of the process. Additionally, a machine algorithm will automatically extract metadata (indicating anything from whether something is, say, a book or web site only, to other technical details) and will feed it into this tier (or higher tiers, if a resource already contains adequate metadata to go straight to tier 2 or 3). African and international university students (graduate and undergraduate) familiar with a particular domain, e.g. hydrology, and possessing some degree of and or aptitude for metadata literacy, will create metadata. Here, terms and relationships can be drafted, thrown out and drafted again in iterative, rapid succession, in either an analog or digital environment. An eGranary resource page might have for instance a pop-up that allows one to easily tag it with descriptive terms. Alternatively, there could be paper-only environment in which students, some of whom might be more comfortable with lower-tech, are collaboratively brainstorming terms on post-its, which are later added digitally to the system. Creating terms in this manner prevents what Egger calls "tunnel vision," when people get caught up in the design of the product or resort to processes most comfortable to them, instead of focusing on what best benefits end users ("Lo-Fi vs. Hi-Fi Prototyping," n.d.). Further, people at this level are imbued with metadata, information and computer literacy through their efforts. Items in the "lo-fi" tier are not hardcoded into the canon, but Tier 1 products are passed to Tier 2 upon completion.

4.2. Tier 2: Middle-Fi

(middle-fi): Refining the terms and their relationships (synonyms, broader, narrower, if applicable). The participants are regional and international domain experts (e.g. hydrology professors/researchers, practicing hydrologists); here, the participants are fewer than in Tier 1.

4.3. Tier 3: High-Fi

(hi-fi): The smallest tier. Information specialists (African and international) approve and refine terms and relationships and add them to the canon of knowledge (hard coded as a nearly final product) in the form of a vocabulary, ontology or descriptive metadata applied to records. Domain experts and other information specialists can review this almost-final product though changes to the canon are harder to make. It is expected that at this level terms are more or less definitive and reflect what are used in a particular culture and discipline. Such high-level activities also enhance the indigenous library worker's multiple literacies, so the benefits of this process are multitudinous.
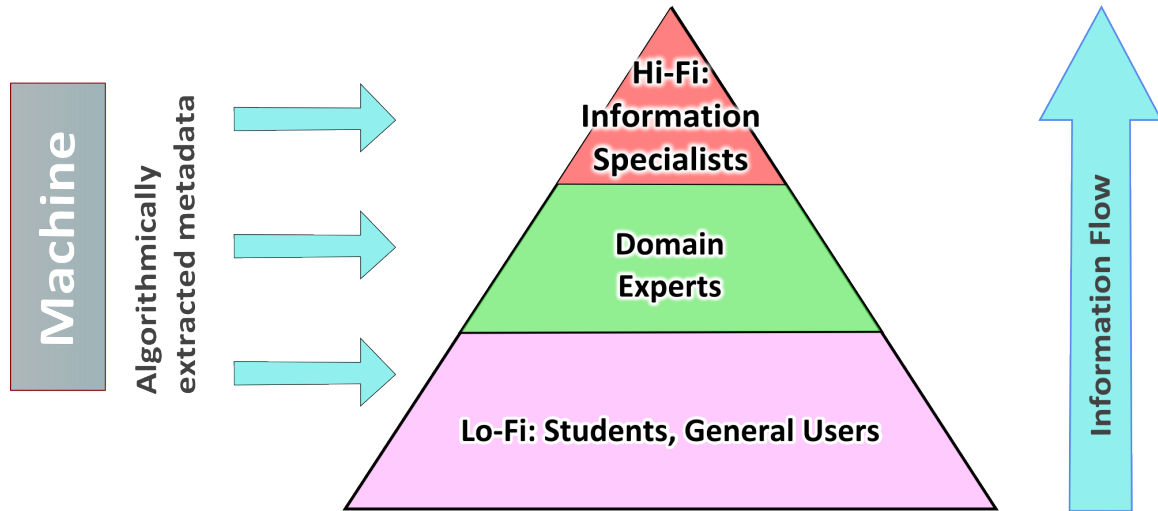
FIG. 1. Illustration of the pyramid model

TABLE 1: Summary of the pyramid model

Tier level	Type of terms	Creator population	Metadata Creator	Fidelity
1	Folksonomy, technical metadata	large	Student, machine	low
2	Folksonomy/ontology/vocabulary/technical metadata	medium	Domain expert, machine	med.
3	Ontology/vocabulary/other descriptive metadata/technical metadata	small	Information specialist, machine	high

Employing rudimentary examples of Lo-Fi/Hi-Fi metadata creation, WiderNet@UNC has demonstrated promising ideas for scaling up the creation of culturally-congruent metadata and user-centric portals through crowd-cataloging and tiered expertise. The authors will continue to explore these concepts as they expand metadata knowledge and use in target populations.

5. Conclusions

Metadata developments have progressed at a tremendous pace, particularly in technology-rich first world nations. The attention to metadata has been basic in developing countries, given more substantial priorities, such as implementing networking capacities. As technologies and opportunities such as the eGranary Digital Library are implemented, the need to address metadata issues has become increasingly apparent. This paper reported on steps taken to address metadata challenges and advance current practices. The Lo-Fi to Hi-Fi metadata pyramid model, taking its cues from other fields like design, is guiding a developing metadata initiative being pursued with the eGranary Digita Library and helping the initiative to understand how to expedite the creating of good quality metadata, making resources more findable and usable.

Next steps including testing the model in information and technology-poor areas of Africa by assessing the needs and available manpower to source the effort through a series of methods including surveys and experiments. We hope to discover through our research how to best implement the pyramid model, thereby eliminating much of the information, computer and

metadata illiteracy plaguing certain areas while bolstering eGranary resource findability. If successful, the effort can be duplicated in other countries, such as Bangladesh and India, and environments such as prisons, with eGranaries.

References

Computer Literacy: Necessity or Buzzword? (n.d.). Retrieved May 16, 2014, from http://www.ala.org/lita/ital/22/3/childers

Du Plessis, A., & Webb, P. (2012). Teachers' Perceptions about their Own and their Schools' Readiness for Computer Implementation: A South African Case Study. Turkish Online Journal of Educational Technology - TOJET, 11(3), 312–325.

Global access to aging information and the gerontology healthy ageing portal. Lisa E Skemp, Ji Woon Ko, Cliff

Missen, Diane Peterson. The University of Iowa College of Nursing, Iowa City, IA, USA.

Introduction to Information Literacy. (n.d.). Retrieved May 16, 2014, from http://www.ala.org/acrl/issues/infolit/overview/intro

Jager, K. de, & Nassimbeni, M. (2007). Information Literacy in Practice: engaging public library workers in rural South Africa. IFLA Journal, 33(4), 313–322. doi:10.1177/0340035207086057

Journal of Gerontological Nursing 01/2011; 37(1):14-9.

Lo-Fi vs. Hi-Fi Prototyping: how real does the real thing have to be? (n.d.). Telono. Retrieved from http://www.telono.com/en/articles/lo-fi-vs-hi-fi-prototyping-how-real-does-the-real-thing-have-to-be/

Mitchell, E. T. (2010). Metadata literacy: An analysis of metadata awareness in college students (Ph.D.). The University of North Carolina at Chapel Hill, United States -- North Carolina. Retrieved from http://search.proquest.com.libproxy.lib.unc.edu/docview/304160575/abstract?accountid=14244

Oduaran, A. (1991). The Computer Revolution and Adult Education. Growth Prospects in Africa.

Park, J., Tosaka, Y., Maszaros, S., & Lu, C. (2010). From Metadata Creation to Metadata Quality Control: Continuing Education Needs Among Cataloging and Metadata Professionals. Journal of Education for Library & Information Science, 51(3), 158–176.

prototype. (n.d.). Merriam-Webster. Retrieved May 17, 2014, from http://www.merriam-webster.com/dictionary/prototype.

Zambia (n.d.). Sparkman Center for Global Health brings eGranaries to Universities in Zambia. Retrieved September 18, 2014, from http://www.widernet.org/node/891 and http://www.sparkmancenter.org/egranary.