# "Lo-Fi to Hi-Fi" Crowd Cataloguing: Increasing E-Resource Records and Promoting Metadata Literacy within WiderNet

Deborah Maron, UNC Chapel Hill
Cliff Missen, WiderNet and UNC Chapel Hill
Katie McNeirney, UNC Chapel Hill
Elnora Tayag, UNC Chapel Hill

**Abstract**
Five billion people, as well as rural libraries, schools, and prisons across the globe, are not connected to the Internet. The WiderNet Project attempts to assist these underserved populations with its eGranary Digital Libraries, which make a wide variety of Web resources available offline. Access alone is not enough, however, as an individual's first exposure to Web resources is much like 'drinking from a firehose'. Additionally, most of the eGranary's 32 million resources remain uncatalogued and hence difficult to find. To address this issue, we are experimenting with 'crowd cataloguing' to create and cull cataloguing information for eGranary resources. An exploratory study is introduced in which we will draw on lessons learned from folksonomy as well as previous eGranary cataloguing projects. The purpose is to further develop a "Lo-Fi to Hi-Fi" model of eGranary metadata generation that emphasizes improving the metadata literacy of general users and library professionals.

## 1   Introduction

### 1.1   **Folksonomy to Cataloguing**

Crowdsourced metadata as folksonomy defines many popular social media spaces. It is also becoming more prevalent in LIS projects where staff time is scarce and resources need to be more organized and findable. By crowdsourcing metadata creation, a library or archive can benefit from two things: cultural congruence - i.e. relevance of terms within a particular population, say a geographic region like Kenya, or a "geek" online community like Stack Overflow[1]:- as well as popularity metadata, i.e. prevalent terms as determined by the masses. To quote Kroski: "No longer do the experts have a monopoly on this domain; in the new age users have been empowered to determine their own cataloging needs. Metadata is now in the realm of the Everyman."[2] This is not to say that traditional cataloguing/metadata procedure should be discarded: The dialectic between social media tagging practice and catalogue practice endures as libraries, archives and similar institutions apply the mechanisms used in social media to their own Web-based collections. Given this, is it possible to transcend tagging and have the user do actual cataloguing?

### 1.2   **Project Background**

Before delving into the cataloguing question posed, it is necessary to situate this problem within the experimental context. WiderNet, our study focus, provides 'snapshots' of the Internet in underserved areas around the world using eGranaries, hard disks with 32 million resources each. However millions of eGranary e-resources remain uncatalogued and hard to find.  The eGranary has a modestly effective full-text search tool, but few users have well-developed search skills. Server logs show that as many as 95% of the resources opened on an eGranary are found using the catalogue, indicating a strong user preference for using this location tool.

WiderNet's two staff cataloguers work part-time and the catalogued records grow slowly compared to the millions of documents added to the collection annually. To create more metadata, WiderNet cataloguers

---

have recruited volunteers to work on the catalogue and "portals"- eGranary sub-collections- to mixed but promising results. For instance, Zambian partners helped curate several very successful health care portals by advising catalogers, proving that domain experts and cataloging amateurs can play a critical role in identifying and classifying resources.

## 2    The Model

"Crowd cataloguing"- a neologism reflecting that terms created and refined by both professionals and amateurs will directly inform the catalogue and new organizational hierarchy proper - will now be used by WiderNet in conjunction with a Lo-Fi/Hi-Fi metadata model[3]. The model name is taken from the different forms of prototyping- high fidelity and low fidelity- and relates to the hard (Hi) or soft (Lo) coded state of metadata at a particular tier[4].
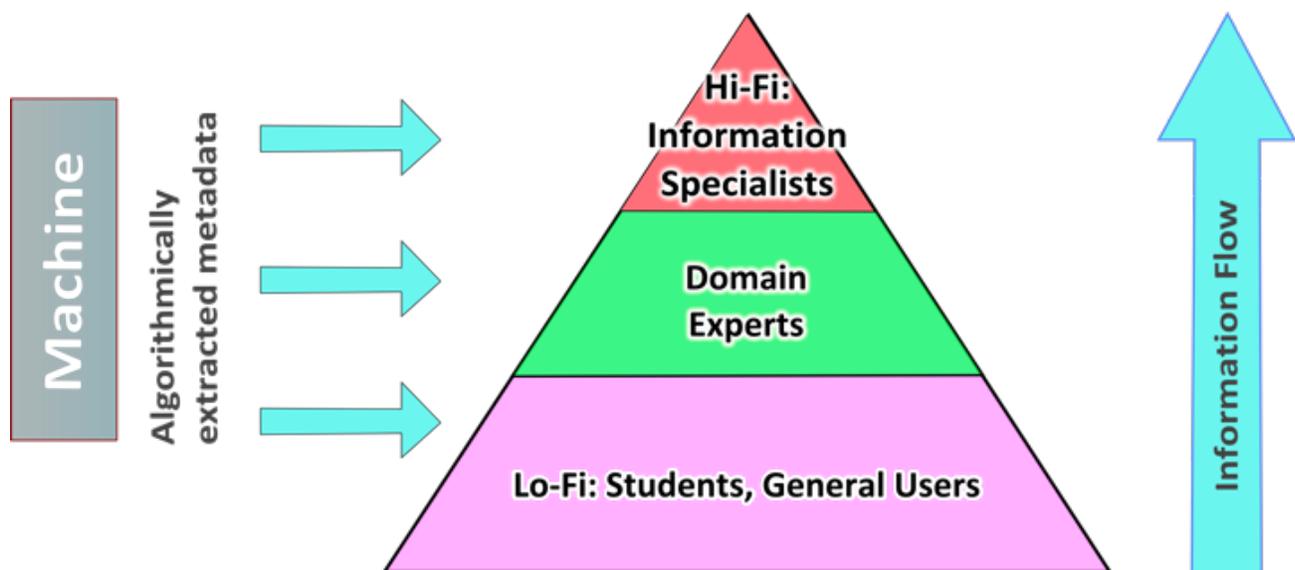
.



Figure 1. Lo-Fi to Hi-Fi Metadata Model

Our hierarchical reconceptualization employs tagging and cataloguing methods, but formalizes the process by expanding *every* user's ability to contribute catalog records within a system that provides accountability and fosters metadata literacy[5]. Metadata literacy, regarded here as the user's ability to apply relevant terms, is enhanced when one participates in such processes as tagging, and we argue that literacy can be further improved when one creates catalog records. Because many *good* records are needed and the top tier participants will not be able to do excessive cleanup, metadata literacy from the Lo-Fi to the Hi-Fi is paramount.

It is worth noting that organizations such as the Smithsonian and New York Public Library are undertaking metadata generation in the style of the Lo-Fi/Hi-Fi model, with some or all tiers present in their approaches.[678] We attempt to formalize these kinds of processes in the library, museum and archival community not by invention but by providing a cleaner metadata procedure with an emphasis on literacy.

---

[3] (Maron et al, 2014)
[4] ("Lo-Fi vs. Hi-Fi Prototyping," n.d.)
[5] (Mitchell & Greenberg, 2009)
[6] (Vershbow, 2013)
[7] (Short, 2014)
[8] (Britton, Level, & Gardner, 2013)

## 2.1    How does Lo-Fi/Hi-Fi cataloging work?

Amateurs (university students, retired librarians, subject matter enthusiasts) will comprise the lowest and largest tier of the metadata model (Lo-Fi), and be trained in, and assessed for, *metadata literacy*, so that they can apply culturally congruent information to resources determined to be the most relevant and in need of further metadata. Users at this tier locate new resources to highlight or manage resources culled from algorithmically extracted metadata. They create a new catalog record and indicate basic information (title, author, format, type), a summary, and assign LCSH *categories* to the best of their ability.  As they go up the metadata chain in the Lo-Fi to Hi-Fi model, these categories and other metadata will be cleaned up and augmented by Middle-Fi catalogers (domain experts). The Hi-Fi catalogers, who have significant training and expertise, will inspect and correct each record before it is submitted to the catalog.

To prevent havoc and endorse metadata literacy, every user, from the amateur to the domain expert to the librarian, will be held accountable at each tier. Mistakes and changes will continually be logged so that quality control can be ensured.  Catalogers will be able to communicate horizontally with their colleagues to share advice, and information about edits made to a catalog record will flow vertically so that trainees can see how their worked fared in the hands of higher-level editors.
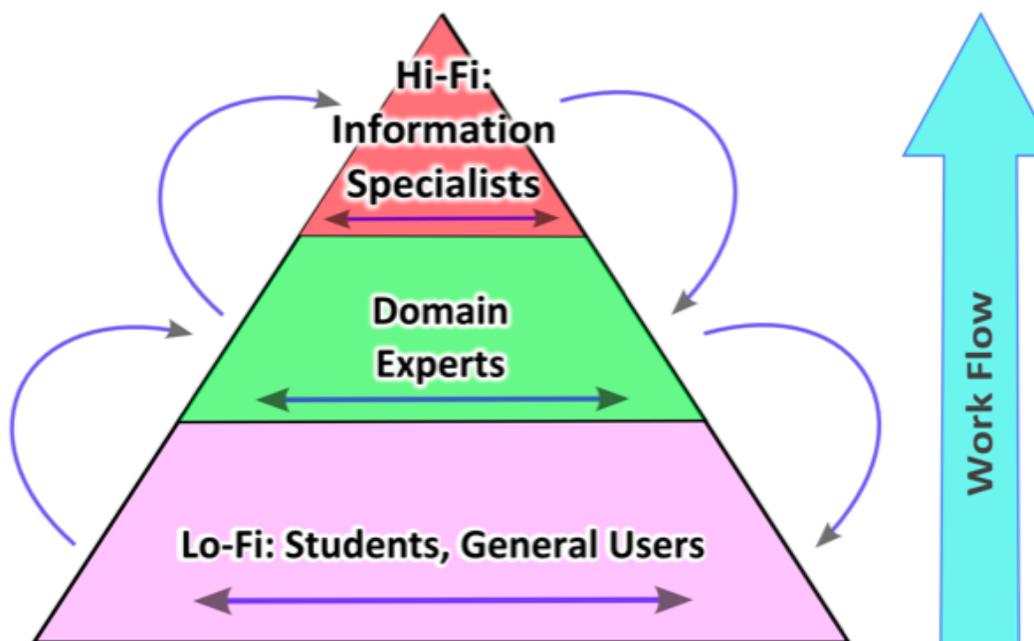


Figure 2. Feedback Flow within Model

## 2.2    The Role of Tagging

Tagging of eGranary resources can take place at any point: tags can be gathered from the public at the original Web site, added in the process of cataloging, or assigned from within the collection (either on-line or off-line).  Affinity groups can be invited to tag specific resources.

The authors propose utilizing tagging functionality on all levels so users can contribute terms to provide other users an alternative wayfaring search experience. Later, these tags can become categories, resolved to LC subject headings, and possibly added to the 'canon' of e-Granary subject headings/category options. This will advance the organizational structure of the catalog and hierarchy. Such terms can also possibly be reconceptualized as a thesaurus, an ontology, or other sort of categorization. A good example of ontology might be "terms describing African medical conditions and ailments, curated by people in Africa".

# 3    Planned Exploratory Study

## 3.1    **Research Design**
Longitudinal, to evaluate subjects' cataloguing expertise/progress over a several month period, e.g. a semester in the case of LIS student volunteers, and experimental, with pretest/posttest control group design to test a subject's metadata literacy (defined below) pre- and post-treatment in order to evaluate how well subjects fare with treatment.

Questions and Hypotheses
To what extent can the Lo-Fi/Hi-Fi model be effective as a framework for crowd cataloguing in terms of creating a higher volume of quality records?

> *Hypothesis: More records will be created for eGranary resources with the addition of crowd cataloguing within the model's framework.*

How effective and precise can trained amateurs (Lo-Fi) be as cataloguers, and do they improve over time as measured by the correctness of LCSH submitted to the top-level participants?

> *Hypothesis: Subjects who have undergone training will gain metadata literacy in the context of WiderNet cataloguing; that is, subjects will perform better (with a larger percentage of terms correct) after the 4-hour training and vertical/horizontal feedback gotten through the pyramid model. Additionally, over time cataloguers will improve (get more terms accepted into the canon and have fewer mistakes overall).*

Further research questions will be pursued within future studies and include: cultural congruence within and between participant groups of different SES/geographic locations; efficacy comparisons between participant groups (e.g. retired librarians and students); trustworthiness of terms as perceived by top-tier cataloguers; deanonymization of participants and the effect on quality and accountability; and the effect the presence of different types of incentives (grades and rankings, among others) has on a participant's performance.

## 3.2    **Setting**
Training would occur in school labs and during class time. Records would be created and submitted online from multiple locations, including homes and classrooms.

## 3.3    **Population, Subjects and Sampling**
Our sample will be comprised of one class from each of three African i-Schools and one class from each of three American i-Schools. The rationale here is that this would be a manageable sample size to start out with and the researchers already have access to institutions in these regions. Further, analyses can be run later to determine cultural differences (and cultural congruence) by comparing American and African populations' choice of LCSH terms, as well as item descriptions, within respective cataloguing records. The Lo-Fi tier will be comprised of LIS students who are new to metadata (n=120); in Middle-Fi, domain experts at American universities and at African institutions (n=60); in Hi-Fi, metadata and cataloguing experts in the US and Africa (n=30)

## 3.4    **Treatments**
A treatment is comprised of a 4 hour interactive training webinar or recorded video explaining LCSH and how to catalog specifically for WiderNet, in addition to the vertical and horizontal feedback loops (with peers and experts) in the Lo-Fi/Hi-Fi metadata model.

## 3.5    **Methods and Tools**
Semi-structured interviews and surveys of the volunteers' experiences and opinions at all tiers, descriptive statistics (user logs to see which headings and resources had most hits, and if they increased following subject heading addition)

## 4    Conclusion

Folksonomic practice in libraries has leveraged the power of the crowd in making the user wayfaring experience more efficient in digital environments. WiderNet seeks to learn lessons from other organizations doing these things, and augment the process by implementing crowd participation for catalog record generation. We expect that our findings and practices, in terms of crowd cataloguing and the Lo-Fi/Hi-Fi model, can be extrapolated to other organizations seeking to scale cataloguing efforts, regardless of metadata standards employed.

## References

Britton, C. J., Level, A. V., & Gardner, M. A. (2013). Crowdsourcing: divide the work and share the success. Library Hi Tech News, 30(4), 1–5. doi:10.1108/LHTN-03-2013-0017

Eden, B., & Steele, T. (2009). The new cooperative cataloging. Library Hi Tech, 27(1), 68–77. doi:10.1108/07378830910942928

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

Maron, D., Missen, C. & Greenberg, J. (2014, in press). Lo-fi to Hi-fi: A new way of conceptualizing metadata in  underserved areas using the eGranary Digital Library. International Conference On Dublin Core And Metadata Applications.

Mitchell, E., & Greenberg, J. (2009). Metadata literacy [electronic resource] : an analysis of metadata awareness in college students. University of North Carolina at Chapel Hill, Chapel Hill, N.C.

Short, J. (2014). Take Ten to Tag! Smithsonian Gardens Public Tagging Initiative. Technical Services Quarterly, 31(4), 319–331. doi:10.1080/07317131.2014.943005

Stack Overflow. (n.d.). Retrieved October 11, 2014.

Vershbow, B. (2013). NYPL Labs: Hacking the Library. Journal of Library Administration, 53(1), 79–96. doi:10.1080/01930826.2013.756701

## Table of Figures