Grey Literature – A Digital Age Mosaic
Keynote Address at GL11
14 December 2009
Peter R. Young[1]
Chief, Asian Division
Library of Congress
Washington, D.C.

Over much of the past decade, I have worked to understanding the impact of digital networking and Web 2.0 technologies on national and research libraries. This concentration was the focus of my work at the U.S. National Agricultural Library and guides my efforts at the Library of Congress. It is instructive to consider grey literature from the perspective of digital science data and eScience, and to consider the Eleventh International Grey Literature Conference themes[2] in relation to digital age challenges.

1. Grey Literature Challenges

GL consists of content "...produced on all levels of government, academics, business or industry in print and electronic formats, but is *not controlled by commercial publishers.*"[3] GL consists of a "...body of materials that *cannot be found easily* through conventional channels such as publishers, but which is frequently original and usually recent."[4] GL in general consists of "...foreign or domestic open source material that usually is *available through specialized channels* and may not enter normal channels or systems of publication, distribution, bibliographic control, or acquisition by booksellers or subscription agents."[5]

With the following characteristics, it is instructive to consider GL challenges:

- GL falls between open and classified/sensitive literature
- GL varies in quality
- GL intended to serve a limited audience
- GL issued in limited quantities
- Most GL is non-proprietary
- Hard-copy GL repositories are scarce & incomplete
- GL lacks outreach/marketing
- GL falls between "black & white"
- "Grey" implies incompleteness
- GL reflects gradations
- GL definitions are fluid

---

[1] This presentation reflects the ideas of the author and does not represent official policies of the Library of Congress or the Federal government.
[2] Impact of GL on Net Citizens, Corporate governance of GL, Uses and applications of subject based GL, GL repositories revisited and Open access to Grey resources
[3] *ICGL Luxembourg definition, 1997 - Expanded in New York, 2004.*
[4] *Wikipedia*
[5] *US Interagency Gray Literature Working Group 1995*

GL challenges form into the following questions:

1. What is needed to make GL more accessible and valuable?
2. Do we need to depend on information professionals to place this large body of valuable content under bibliographic control to enable search engines to provide better access?
3. Are GL content and metadata standards sufficient and universally applied so that GL could be accessible if sufficient resources were available?
4. What would it take for the corpus of current and retrospective GL to be accessible via the Web?
5. How would we decide what GL was going to be more valuable to future generations of scientists and researchers, especially if we do not have sufficient resources to assure access to all GL content?

2. Digital Grey Literature

Digital networking technology is affecting the nature of scientific and research communications. These same forces of change are affecting the nature of grey literature content. In addition, digital networking technology and the trend toward open access are changing the conduct of science and research through the rapid development of the Internet/Web. These same forces are having a profound impact on the distribution of and access to GL content.

> "What has been made public by being published is no longer a black and white issue. There is still plenty of room for judgments and distinctions to be made about the quality, type, and nature of this knowledge. *This growing openness around what is known assists in the very assessment and verification. I, for one, do not see grey skies ahead, but something brighter."*[6]
> "Since there has been an increase in publication and dissemination of materials from the producers of grey literature and other materials, establishing the means to work directly with users rather than relying on the traditional means of evaluating and collecting becomes necessary." [7]

It is difficult to see how the library and information community could successfully attract sufficient resources to address the growing body of GL content and to comprehensively work to provide consistent and uniform access to an ever-increasing tsunami of print and digital content. In fact, digital machines are ideally suited for performing certain functions exceptionally well, but these functions are not identical to the things that humans are capable of replicating at anything near machine speed or accuracy.

---

[6] Blogger John Willinsky. Professor at the Department of Language and Literacy Education at UBC, and the Public Knowledge Project.

[7] Heather Lehman and Janet Webster *Describing Grey Literature Again: a survey of collection policies* in Publishing Research Quarlerly, Spring 2005, pp. 65-72.

The combination of demand-driven science, together with the development of new digital storage and Web exchange tools are creating a new dynamic model for conducting data-intensive research in new ways. These trends are creating a lot of excitement among an array of scientific domains from micro-biology to astrophysics, and from ecological to oceanographic studies, where the creation of new knowledge employs new digital tools and Web methods. These new dynamics require that the more traditional library and information science communities join in creating the emerging Data Web through which links and object relationships between and among content and users are established in a evolutionary and iterative way.

It is not difficult to see that the Web is escalating the GL challenge. A deluge of Web-based grey digital content is sufficient to drown any plan for placing these materials into a controlled and accessible environment through our traditional print-based bibliographic control methods. David Weinberger speaks on the "Future of Knowledge" in his *Everything is Miscellaneous: the Power of the New Digital Disorder (2007)*. His thoughts on the relationship between data, information, knowledge, and wisdom are both enlightening and challenging. From an epistemological standpoint David's ideas are interesting. But, from a librarian's standpoint, these concepts are revolutionary. He writes of a third order of knowledge (a new paradigm) as one where things that one can do easily with digital technology are those things that the real world makes really hard. In this third order, users (not owners) of information organize that information according to their interests.

Flickr offers an example of Web 2.0 tools that offer opportunities for individuals and communities to help organize Web content. The Library of Congress's project with Flickr to allow users to help identify assets and in so doing to assist curators in describing and tagging individual photographs with descriptive terms is but one example of engagement of users in helping to organize digital content. Del.ici.ous is another example, that offers tagging in the form of bookmarks to Web sits of interest. It offers personalization in tagging, but it also provides a social networking community where a user can see what everyone else using the public site has tagged. These tags are social links that grow and evolve through the addition of more and more users, illustrating the "Wisdom of Crowds"that WikiPedia reflects.

Another example of citizen science is drawn from astronomy. The Galaxy Zoo site contains a quarter of a million galaxies which have been imaged with a camera attached to a robotic telescope the Sloan Digital Sky Survey. More than 150,000 people have taken part in Galaxy Zoo to help classify images of galaxies.

"Today, scientific publishers are production companies, specializing in services like editorial, copyediting, and, in some cases, sales and marketing... in ten to twenty years, scientific publishers will be technology companies.... they'll be technology-driven companies in a similar way to, say,

Google or Apple. That is, their foundation will be technological innovation, and most key decision-makers will be people with deep technological expertise. Those publishers that don't become technology driven will die off."

"Scientific publishers should be terrified that some of the world's best scientists, people at or near their research peak, people whose time is at a premium, are spending hundreds of hours each year creating original research content for their blogs, content that in many cases would be difficult or impossible to publish in a conventional journal. What we're seeing here is a spectacular expansion in the range of the blog medium. By comparison, the journals are standing still."[8]

The global library and information services community does not need to be standing still. We have the responsibility to address the growing need to both organize and to provide access to grey content, especially as digital data increases in the future. The astonishing growth of digital content requires that we elicit new methods and new models for assuring that the value of grey literature content is available to those scientists and researchers addressing the grand challenges of the future.

---

[8] Michael Nielsen blog.