# Grey Literature - A Digital Age Mosaic

Peter R. Young Chief Asian Division Library of Congress 14 December 2009



# Grey Literature - A Digital Age Mosaic

## 1. Grey Literature Challenges

Definitions - Characteristics Creation – Access - Use Search – Discovery - Delivery

## 2. Digital Grey Literature

Framing Digital Grey Literature Digital Technology Impact

### 3. Grey Web Opportunities Creation – Search – Delivery

BRARY OF CONGRESS

## 4. e-Grey

e-Science – e-Research Data-driven Science - Citizen Science Web 2.0 – Social Networking



# Grey Literature - A Digital Age Mosaic GL Quiz

### 1. True or False

- a) GL is easily defined
- b) GL access is problematic
- c) Only info professionals know about GL

### 2. The GL challenge is:

- a) GL lacks bibliographic control
- b) GL is ephemeral
- c) Demand for GL access is growing
- d) All of the above

### 3. The impact of the Web on GL is:

- a) No Web impact on GL
- b) The Web solves GL access problems
- c) The Web makes GL more of a problem

### 4. How much GL is there?

- a) More GL than published print content
- b) GL growing 10X every 5 years
- c) No one knows

BRARY OF CONGRESS



# 1.0 GL Challenges Definitions

- GL is "...produced on all levels of government, academics, business or industry in print and electronic formats, but is not controlled by commercial publishers." ICGL Luxembourg definition, 1997 - Expanded in New York, 2004
- 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." Wikipedia



## 1.1 GL Challenges Definitions

 GL 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." US Interagency Gray Lit Working Group 1995

### Organizations generating GL:

- Research institutes
- Academic institutions
- Federal agencies
- Private publishers
- Corporations
- Trade associations

BRARY OF CONGRESS

Think tanks



## 1.2 GL Challenges Characteristics



BRARY OF CONGRESS

 Research, Technical, & Economic Reports

> Pre-Prints - Fact Sheets -Discussion Papers

- Standards Patents Trade Literature - Lab Notebooks
- Working papers Newsletters

#### White papers - Business Documents

- Government Documents -Committee Reports - Statistics
- Technical Documentation -Conference Proceedings
- Symposia Bulletins PPTs Data
- Unpublished works Pamphlets -Market Surveys - Lecture notes
- Dissertations Theses Catalogs



# 1.3 GL Challenges Characteristics

- 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



## 1.4 GL Challenges Creation – Access - Use



BRARY OF CONGRESS

- Lack of commercial control of GL challenges researchers and information professionals
- 11 GreyNet International Conferences define & frame GL challenges
  - Who creates/produces GL?
  - Who uses GL and for what?
  - Why is GL important?
  - What is the size of GL corpus?
  - GL growth rate/demand estimates?
- GL documents & records science, technology, research & investigations
  - GL contains valuable and unique content
  - GL -> Digital GL -> Grey Web ->e-Grey

# 1.5 GL – Challenges Creation – Access - Use

#### GL value

- Estimating value of non-standard content
- GL bibliographic control
  - Format bifurcated universe
    - Monographic cataloging
    - Serial article indexing
    - Web metadata

#### GL dissemination & metadata

- Unpublished or semi-published or limited-distribution content
- Content that is not described is invisible and undiscoverable
- GL search & discovery
  - Semi-controlled works are not consistently cited
    - Value remains hidden
- GL & Grey data

BRARY OF CONGRESS

 Role of bibliographic control and libraries in an increasingly data dominant realm



# **1.6 GL – Challenges** Creation – Access - Use

### Published B&W Literature

### Authority based - credible

- Formal defined structure
- Cataloging rules
- Classification schemes
- Filtered peer reviewed
- Subject designations
- Metadata separate from content
- Organized collections
- Controlled access
- Organizationally intensive
- Collection-centric
- Formal publications
- Hierarchical arrangements
- Uniform, orderly, consistent, coherent

### Grey Literature

- Unstructured
- Messy & disorderly
- Informal stuff
- Unorganized
- Unpublished/published mix
- Confusing
- Unfiltered
- Lack of uniform standards
- Fragmented fugitive
- Unconnected
- Ephemeral
- Informal invisible
- Uncontrolled & unstructured
- Inaccessible hidden
- Undiscoverable invisible

## 1.7 GL – Challenges Search – Discovery - Delivery

- Alternate approaches:
  - Developmental Change:
    - Gradual, evolutionary, and developmental adaptation of traditional workflows, policies, practices
  - Revolutionary Change:
    - Radical and abrupt development of new institutional models and capabilities in response to disruptive technology
- Choice: apply standard library control structures or develop new approaches
  - Print control structures

BRARY OF CONGRESS

New GL Search/ Access models



# 2.0 Digital Grey Literature Framing Digital GL



RARY OF CONGRESS

- Digital networking technology is affecting the nature of scientific and research communications
- Digital networking technology and open access are changing the conduct of science and research
- Rapid Internet/Web development has affected distribution of and access to GL
  - Open source publishing
  - Web-based digital content

p.12

## 2.1 Digital GL Digital Technology Impact



"...it's hard for incumbent organizations in a disrupted industry to change to a new model....some of the forces preventing change are strongest in the best run organizations. The reason is that those organizations are large, complex structures, and to survive and prosper they must contain a sort of organizational immune system dedicated to preserving that structure...

"Most of the time the immune system is a good thing, a way of preserving what's good about an organization, and at the same time allowing healthy gradual change. But when an organization needs catastrophic gut-wrenching change to stay alive, the immune system becomes a liability."

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



## 2.2 Digital GL Digital Technology Impact

- "... the wider, more immediate and openly public, circulation of research and scholarship is trumping the traditional distinctions of the print literature. Much of this open research has been peer-reviewed, with the details of the journal publication clearly identified on the article, but it may not be in an officially published form.
- "On top of this, corresponding developments with "open data" are talking place, making another form of GL immediately and widely available for new forms of collaboration and reanalysis."
  - John Willinsky. Professor at the Department of Language and Literacy Education at UBC, and the Public Knowledge Project





## 2.3 Digital GL Digital Technology Impact

- "Governments and the courts are posting reams of material online. The Wikipedia and the blogosphere represent grass-root efforts to reshape the basis of participation in human knowledge.
- "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."
  - Blogger John Willinsky. Professor at the Department of Language and Literacy Education at UBC, and the Public Knowledge Project.

**BRARY OF CONGRESS** 



# 2.4 GL – Digital Challenges

- "The emergence of digital collections and the increasing amount of digital grey literature challenges the library's ability to consistently collect information using our traditional tests and tools."
- "Currently, the most familiar approach to improving access to digital grey materials is *linking* and not conscious collection development."
  - Heather Lehman and Janet Webster Describing Grey Literature Again: a survey of collection policies in Publishing Research Quarlerly, Spring 2005, pp. 65-72.





# 2.5 GL – Digital Challenges

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



BRARY OF CONGRESS



## 3.0 Grey Web Opportunities GL & 4<sup>th</sup> Generation Science

- 1st Generation
  - People to people
- 2nd Generation
  - People to people with machines
- 3rd Generation
  - Machines to people
- 4<sup>th</sup> Generation
  - Machines to machines



## 3.1 Grey Web Opportunities Social Networking Web 2.0

#### Wisdom of the crowd – Crowdsourcing

- Web 2.0 and GL
- Wikis
- Flickr
- Del.icio.us
- Tagging
- Publically negotiated knowledge
- Collaborative Projects Hyperlinked world
- Externallized knowledge and meaning
- Billions of links Culture of abundancy
- Web as a context of relationships
- Building a Web infrastructure of meaning
- Flexibility = Connectiveness
  - Thanks to David Weinberger, Harvard Berkman Center



## **3.2 Grey Web Opportunities** Library of Congress P&P Flickr Pilot Project





### 3.3 Grey Web Opportunities Library of Congress P&P Flickr Pilot Project



LIBRARY OF CONGRESS

### **3.4 Grey Web Opportunities** Library of Congress P&P Flickr Pilot Project



LIBRARY OF CONGRESS

## 3.5 Grey Web Opportunities Del.icio.us

🚰 Delicious - Microsoft Internet Explorer	_ 8 ×
File Edit View Favorites Tools Help	
🛛 🚱 Back 🔹 🕑 - 💌 😰 🏠 🔎 Search 🤺 Favorites 🤣 😥 + چ 🖆 - 🛄 🎇	
Address 🗃 http://delicious.com/	e »
delicious social bookmarking	
The tastiest bookmarks on the web. Save your own or see what's fresh now!	
HIDE INTR	0 🕄
Search the biggest collection of bookmarks in the universe	
	_
Fresh Bookmarks Popular Bookmarks Explore Tags	
The freshest bookmarks that are flying like hotcakes on Delicious and beyond. See more recent bookmarks  YouTube - Beyonce ft Lady Gaga - Video Phone (remix) (Official Video) HQ	_
SAVE	
via yourube.com videos	
76 Related Tweets	
Not a Hoax: Ken Ober, MTV's 'Remote Control' Host, Dies at 52 - Yahoo! TV Blog save via tu-yahoo.com	
N 31 Delated Tweets	-
Cone	
🟄 Start 🚯 Novell GroupWise - Mailbox 🗁 Grey Literature 14 Dec 2 🔞 Microsoft PowerPoint - [ 🦓 Delicious - Microsoft I 🧷 🦉 🔏 Desktop 🎽 « 🐉	👗 12:48 PM



# 3.6 Grey Web Opportunities

🗿 Wikipedia - Microsoft Internet Explorer		_ B ×	🎒 Gray literature - Wikiper	dia, the free encyclopedia - Microsoft Inti	ernet Explorer
File Edit View Favorites Tools Help			File Edit View Favorit	es Taols Help	
🛛 🚱 Back 🔹 🕥 - 💌 😰 🏠 🔎 Search 🤺 Favorites 🤣 🔗 😓 🖬 - 📙 🎎			🛛 😋 Back 🔹 🕥 👻 💌	😰 🏠 🔎 Search 🤺 Favorites	🚱 🍰 🗟 - 🤤 🛍
Address 🕘 http://www.wkipedia.org/	🝘 Employee Express 🛛 Explore Internet 🏾 🍪 LC Home	»	Address 🕘 http://en.wkipe	dia.org/wiki/Grey_literature	💌 🛃 Go 🔤 Links 🝘 APLO 🍘 Employee Express 🔮 Explore Internet 🍘 LC Home 🛛 🛪
WIKIPEDIA English 日本語 770 Fires Forycobedia ジリー百姓事 1990 000 + 414/54	ł	<u>_</u>	WIKIPEDIA	article discussion edit th	Try Beta & Log in / create account  as page nistory  Wikipedia. Ad-free forever.  S0.6M USD  Donate Nov
De triete Enzyklopadie Be toos wekate Français De triete Kongologie die liker Sez soor wat is Bellano Lerrici Copadia liker Sez soor wat	Expanina inclopeda Mare observative Polski Mora encyklykleodia 661000+ havet costni 04 maretei n		In the Prec Encyclepedial     navigation     Main page     Contents     E Contents     Courtent svents     Courtent vents     Courtent vents	Gray literature From Vilipedia, the free encyclopedia (Maerical from Grey Merature) Grey Inerature (or gray Merature) to refer to a body of materials that ca and usually recent" in the words of M The U.S. Interagency Gray Literature 'foreign or domestic open source ma systems of publication, distribution, discontinued.	is a torm used variably by the intelligence community, librarians, and medical and research professionals next be found easily through conventional channels such as publishers, "but which is frequently original IC. Debachard <sup>11</sup> . Working Group (GLWG): "Gray Information Functional Plan," <b>18 January 1995</b> , defines gray Iterature as terial that usually is available through specialized channels and may not enter normal channels or bibliographic control, or acquisition by booksellers or subscription agents." In early 2000, the IGLWG was defines are iferature as "information produced on all levels of agreement, academics, business and
			<ul> <li>Recent changes</li> </ul>	industry in electronic and print formal	ts not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing
search • suchen • rechercher • szukaj • 檢索 • ricerca • zoeken • buscar • b			Contact Wikipedia	body." (Luxembourg, 1997 - Expande	ed in New York, 2004).
搜索 • søk • haku • cerca • ⊓owyx • keresés • hledání • ara • căutare • serĉu • i bladat • dav • tim klám	Citizen science - wikipedia, the free ency	Topedia - Microsoft Internet Explorer			echnical reports from government agencies or scientific research groups, working papers from research or preprints. The term grey literature is offen, but not exclusively, used for scientific research
grey literature English	File Edit View Favorites Tools Help	Search 🤺 Favorites 🚱 🔗 🔌 🖬 - 📴 🐔		<b>A</b>	rey literature poses difficulties for librarians and other information professionals for several reasons. bibliographic control, meaning that basic information such as author, publication date or publishing body
	Address Address Address Address Address Address	science 🔽 🍮 Go 🛛 Links 🖨 APLO 🔗 Emp	olovee Express 🙆 Explore	Internet 🙆 LC Home »	ly, non-professional layouts and formats and low print runs of grey literature make the organized collection
	Main page				pared to more traditional published media such as journals and books. <sup>[2]</sup>
	Contents     From Wikipedi	a, the free encyclopedia		-	Is generally draw a distinction between ephemera and grey literature, however, there are certain overlaps ainly share common frustrations such as hibliographic control issues
e) Done	= Featured content Citizen scien	ce is a term used for projects or ongoing program of scientific work in which	individual volunteers or i	networks of volunteers,	ant share common nasharino daen da sharadabile comfortistico.
🦺 Start 🔞 Novell Group Wise - Malbox 🔯 Grey Liberature 14 Dec 2 🔯 Microsoft PowerPoint - [ 🖗 Wikipedia - Mi	<ul> <li>Current events</li> <li>Random article</li> <li>many of whon</li> </ul>	n may have no specific scientific training, perform or manage research-related	l tasks such as observa	tion, measurement or	💽 Microsoft PowerPoint - [ 🕼 Gray literature - Wiki 🧷 🖉 🔏 Desktop 🎽 « 🐉 📐 12:48 PM
	computation.				
	The use of citi	zen-science networks often allows scientists to accomplish research objecti	ves more feasibly than v	would otherwise be possible.	
	In addition, the	ese projects aim to promote public engagement with the research, as well as	with science in genera	I. Some programs provide	
	Go Search materials spe	cifically for use by primary or secondary school students. As such, citizen so	cience is one approach	to informal science	
	interaction education.				
	About Wikipedia     Contents	i [hide]			
	Community portal     1 Examples				
	Contact Wikipedia     2 Other Uses	of the Term			
	Donate to Wikipedia     A External link	s			
	= Help	-			
	toolbox Examples	;		[edit]	
	What links here     Related changes     Upload file     Special pages     Printable version     Permanent link     Cite this page     this citizen sc	nning citizen science project currently active is probably the Audubon Sociel wn examples of citizen science programs include World Water Monitoring D crts run by the Comell Laboratory of Omithology <sup>121</sup> , such as Ebird, NestWatar y Zoo project. Another example of an effective citizen science project in the l lework (CoCGRAHS), run by the Colorado Climate Center at the Colorado St ience project is used for weather forecasting and monitoring, severe weather	ty's Christmas Bird Cou lay <sup>[1]</sup> , NASA's Stardustr h, Project FeederWatch United States is the Cor ate University in Fort Co alerts, and climate stuc	nt, which started in 1900. @home and Clickworkers, a n, and Celebrate Urban Birds mmunity Collaborative Rain, lins, Colorado. Data from lies.	
	Francais     Citizen science	e networks are extensively involved in phenology, the observation of cyclic ev	vents of nature, in order	to investigate how global	
	warming affect considered cit	is plant and animal life in different geographic areas. <sup>191</sup> Distributed computing izen science, even though the primary task of computation is performed by v	ventures such as SETI olunteers' computers.	@home may also be	
	The increasin citizen scienc the San Franc	g prevalence and use of consumer electronic devices that can record media, e data collection about the condition of public spaces, such as public parks isco Neighborhood Parks Council.	such as mobile phones, as seen in the ParkSca	, has allowed for easier n & website developed by	
Other Uses of the Term famil					
	e			🔮 Internet	
	🐮 Start 🚳 Novell GroupWise - Mailbox 🦉 C	itizen science - Wiki i Microsoft PowerPoint - [	E 🖉 1	🖌 Desktop 🎽 « 🐯 📐 7:55 AM	



# 4.0 e-Grey eScience – Data-Driven Science

 Computationally intensive science using distributed network environments using immense data sets that require grid computing; the term sometimes includes technologies that enable distributed collaboration, such as the Access Grid.

### Data-driven science

BRARY OF CONGRESS

There is a rapidly growing set of applications, referred to as data driven applications, in which analysis of large amounts of data drives the next steps taken by the scientist, e.g., running new simulations, doing additional measurements, extending the analysis to larger data collections.



## 4.2 e-Grey Data-Intensive Resarch

Data-intensive Research Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies.





## 4.3 e-Grey Citizen Science

### Citizen Science

**Citizen science** is a term used for projects or ongoing program of scientific work in which individual volunteers or networks of volunteers, many of whom may have no specific scientific training, perform or manage research-related tasks such as observation, measurement or computation.



## 4.4 e-Grey Web Opportunities Citizen Science





## 4.5 e-Grey World Internet/Web Usage

**360,985,492** Internet users in 2000 1,733,993,741 Internet users in 2009 4.6 fold increase 25% of world population

IBRARY OF CONGRESS



## **4.6 e-Grey** World Internet/Web Domains





## **4.7 e-Grey** World Internet/Web Content

The world produces between 1 and 2 exabytes of unique information per year, which is roughly 250 megabytes for every man, woman, and child on earth. An exabyte is a billion gigabytes, or 1018 bytes. Printed documents of all kinds comprise only .003% of the total. Magnetic storage is by far the largest medium for storing information and is the most rapidly growing, with shipped hard drive capacity doubling every year. Magnetic storage is rapidly becoming the universal medium for information storage.



## 4.8 eGrey A Sea Change in Computing

#### **Massive Data Sets**

Federation, Integration & Collaboration There will be more scientific data generated in the next five years than in the history of humankind

Evolution of Many-core & Multicore

Parallelism everywhere

What will you do with 100 times more computing power?

The power of the Client + Cloud

Access Anywhere, Any Time

Distributed, loosely-coupled, applications at scale across all devices will be the norm

#### Thanks to: Lee Dirks, Microsoft



## **4.9 eGrey** The Exploding Digital Universe



- The digital universe will grow 10-fold in five years
  - From ~160-170 exabytes in 2006 to >1,600 exabytes in 2011
- Information created surpassed available storage in 2007, will be 2X five years
  - Unstructured information accounts for >90% of the digital universe
- Consumers account for ~70% of information created, yet enterprises have "responsibility" for ~85%
  - Preservation "intense" information will grow 9-fold in 5 years

Thanks to: John Gantz, Chief Research Officer, IDC Eleventh International Grey Literature Conference - 14 December 2009 p.3





## 4.10 eGrey Michael Nielsen

"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."



## 4.11 eGrey Michael Nielsen

"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."



# Grey Literature - A Digital Age Mosaic

Peter R. Young Chief, Asian Division Library of Congress 14 December 2009

