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
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

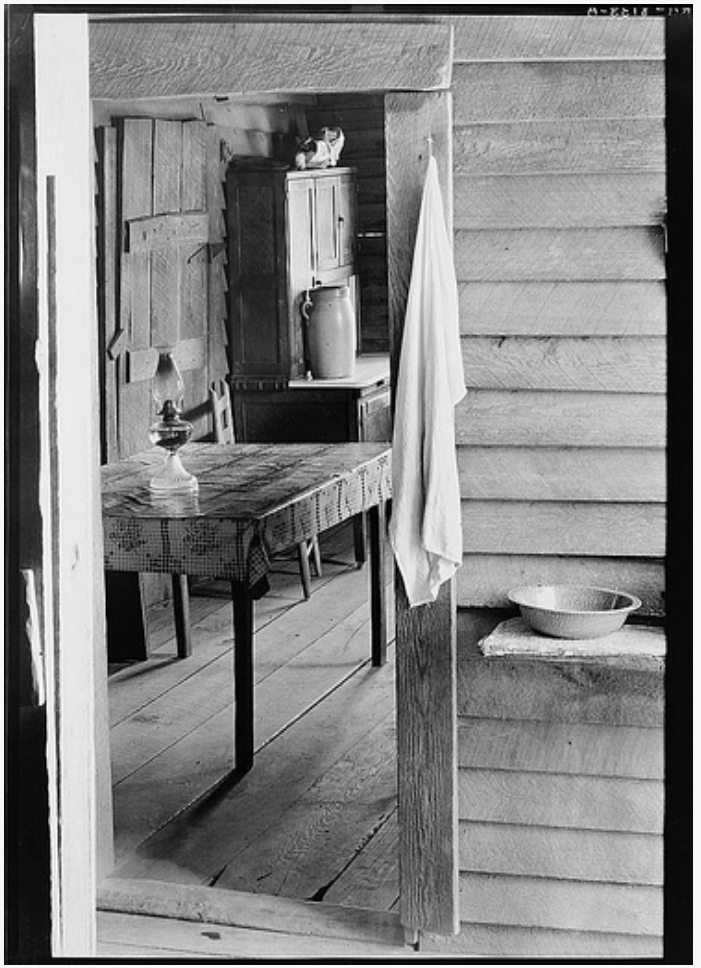
1. True or False
   a) GL is easily defined
   b) GL access is problematic
   c) Only info professionals know about GL
2. The GL challenge is:
   a) GL lacks bibliographic control
   b) GL is ephemeral
   c) Demand for GL access is growing
   d) All of the above
3. The impact of the Web on GL is:
   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

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
  - Think tanks

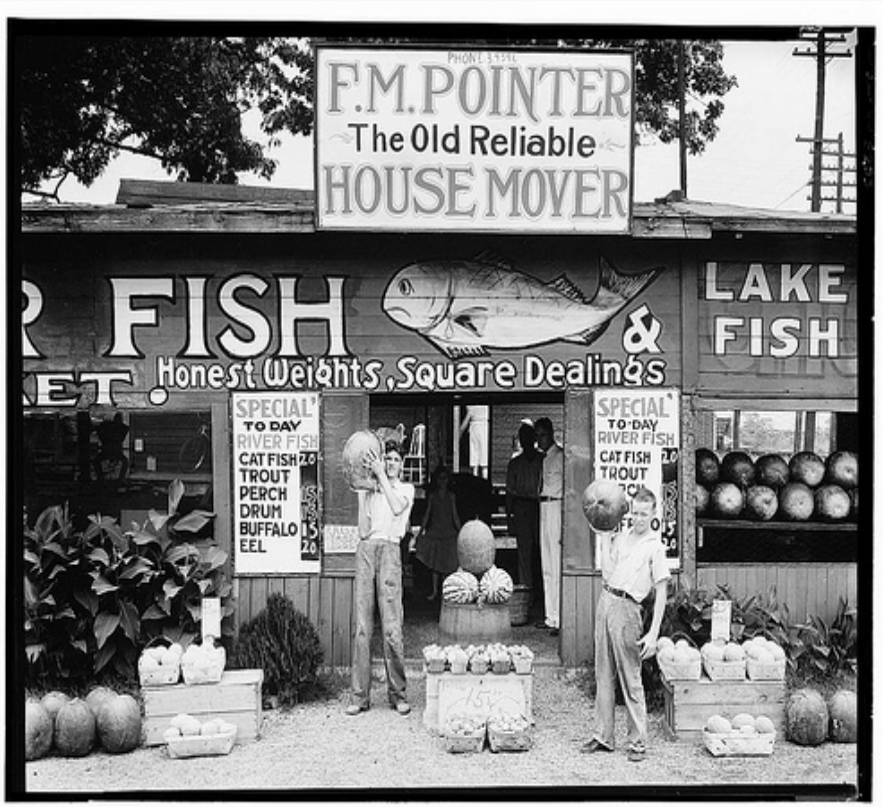


1.2 GL Challenges
Characteristics

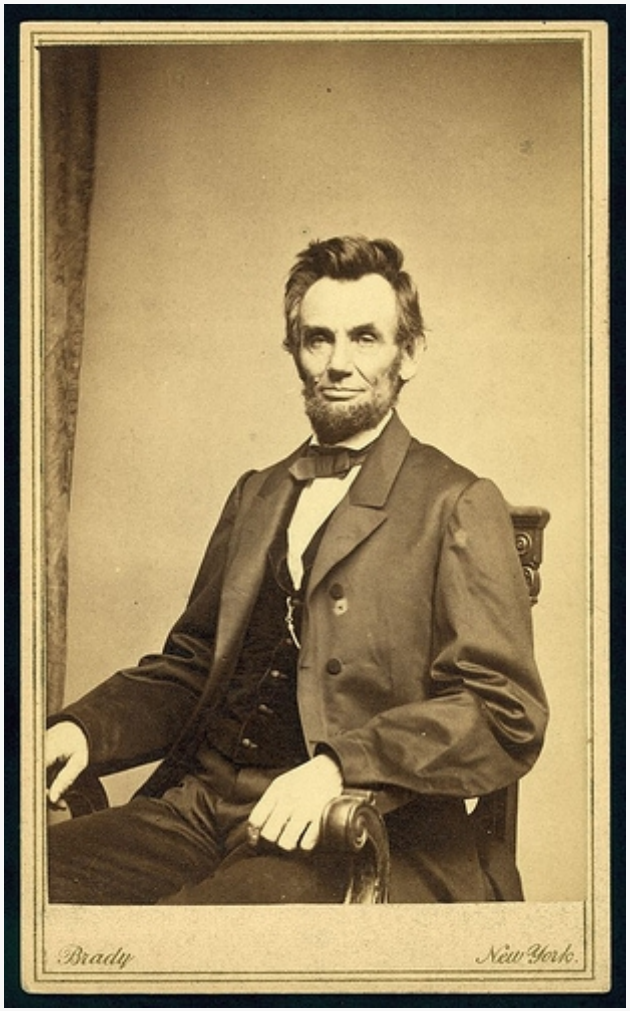
- Research, Technical, & Economic Reports
  - Pre-Prints - Fact Sheets - Discussion Papers
  - Standards - Patents – Trade Literature – Lab Notebooks
  - Working papers - Newsletters
- White papers - Business Documents
  - Government Documents - Committee Reports - Statistics
  - Technical Documentation - 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



- 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
  - 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
  - New GL Search/ Access models

2.0 Digital Grey Literature
Framing Digital GL

- 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

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

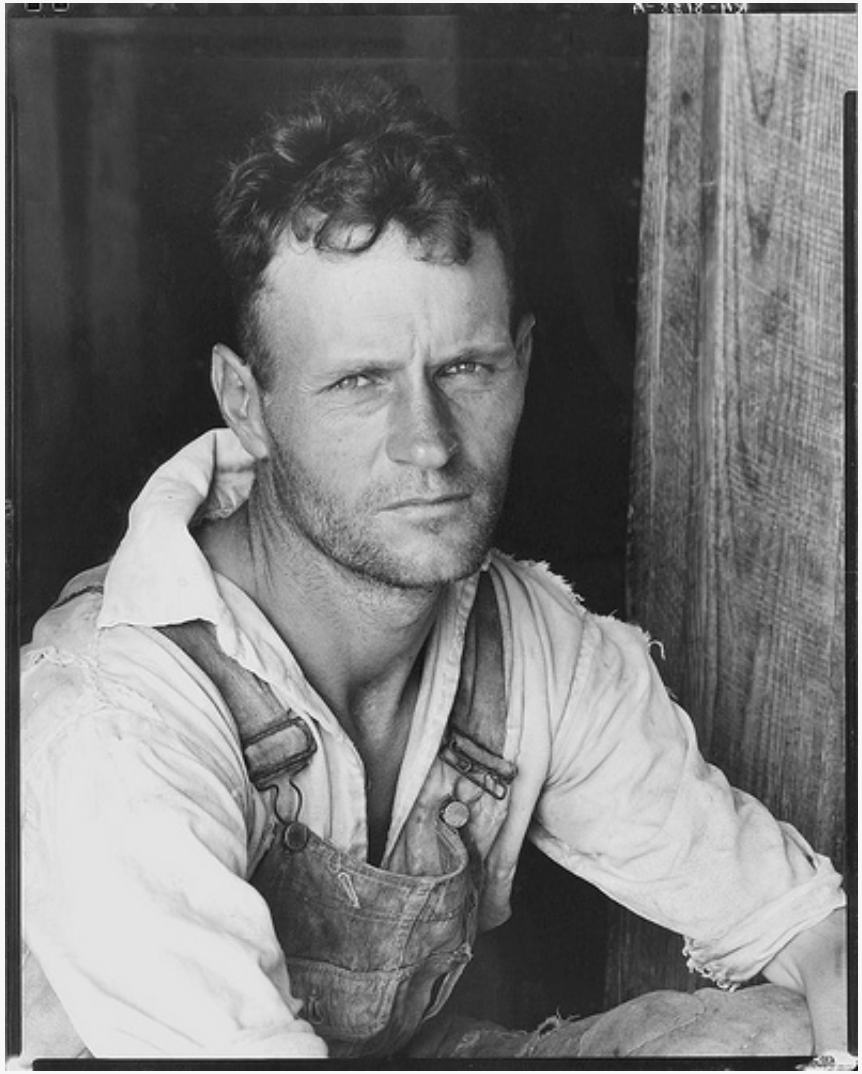
2.2 Digital GL
Digital Technology Impact

"… 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.



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.

3.0 Grey Web Opportunities

GL & 4th Generation Science

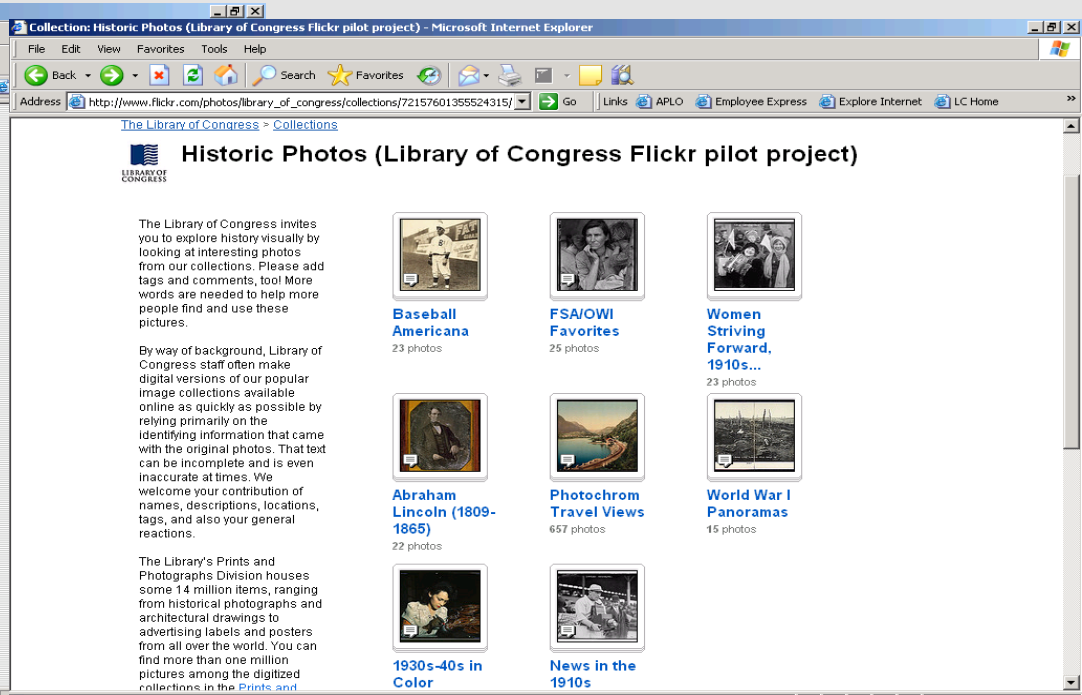
- 1st Generation
  - People to people
- 2nd Generation
  - People to people with machines
- 3rd Generation
  - Machines to people
- 4th Generation
  - Machines to machines

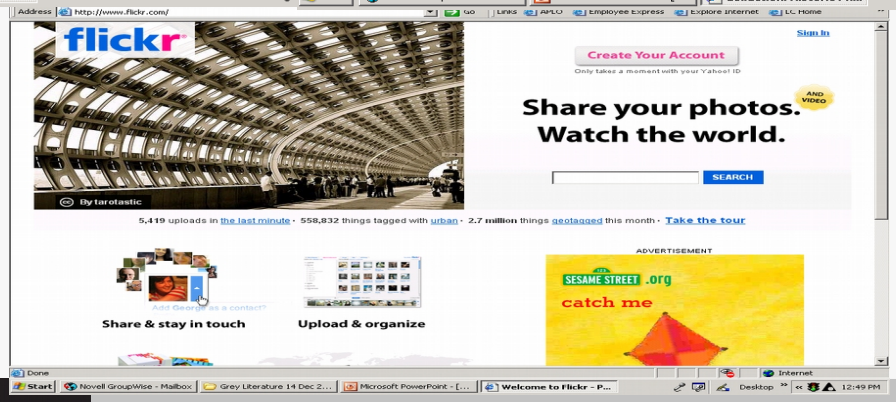
3.1 Grey Web Opportunities
Social Networking Web 2.0

- 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



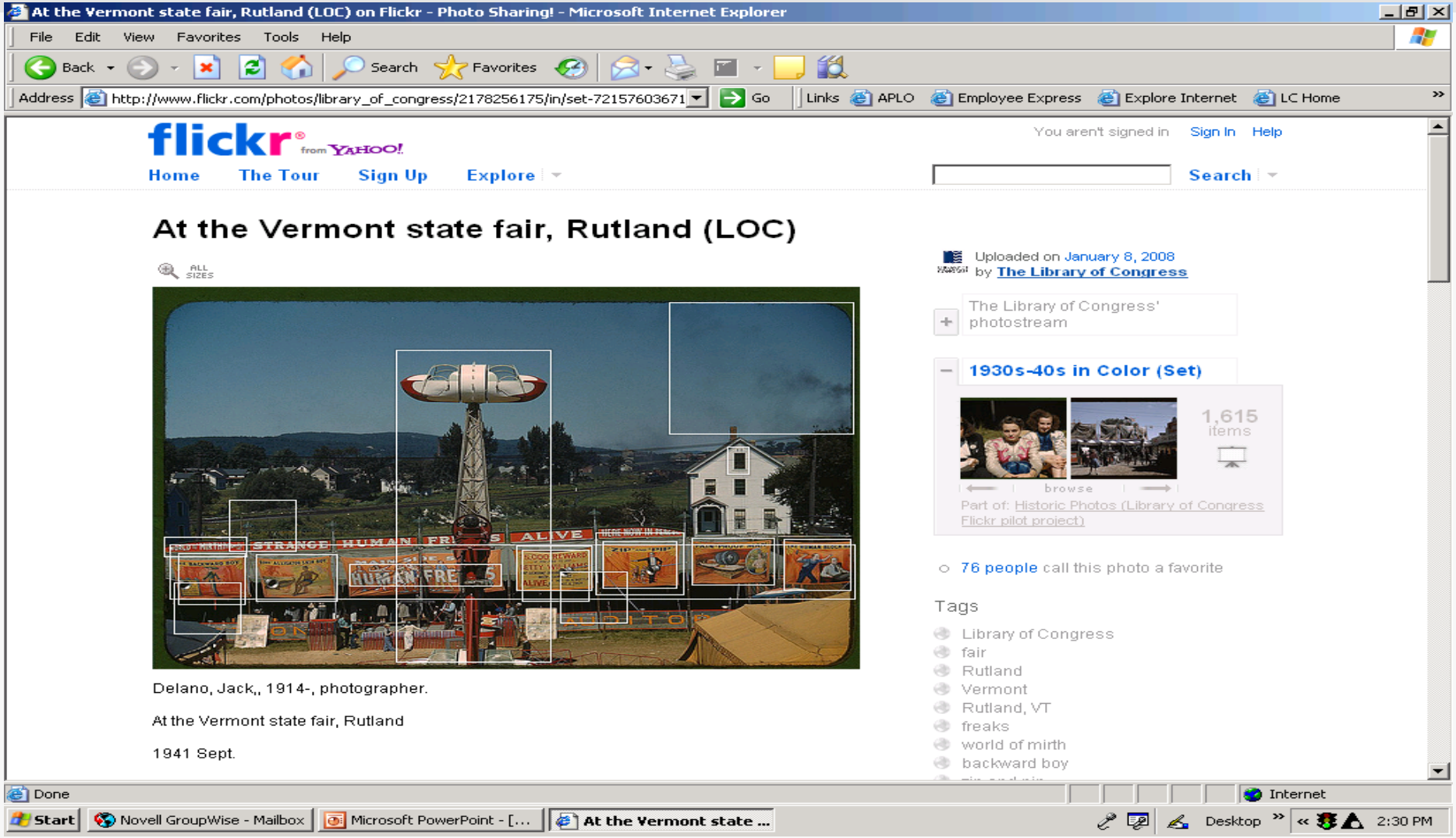


3.3 Grey Web Opportunities

Library of Congress P&P Flickr Pilot Project

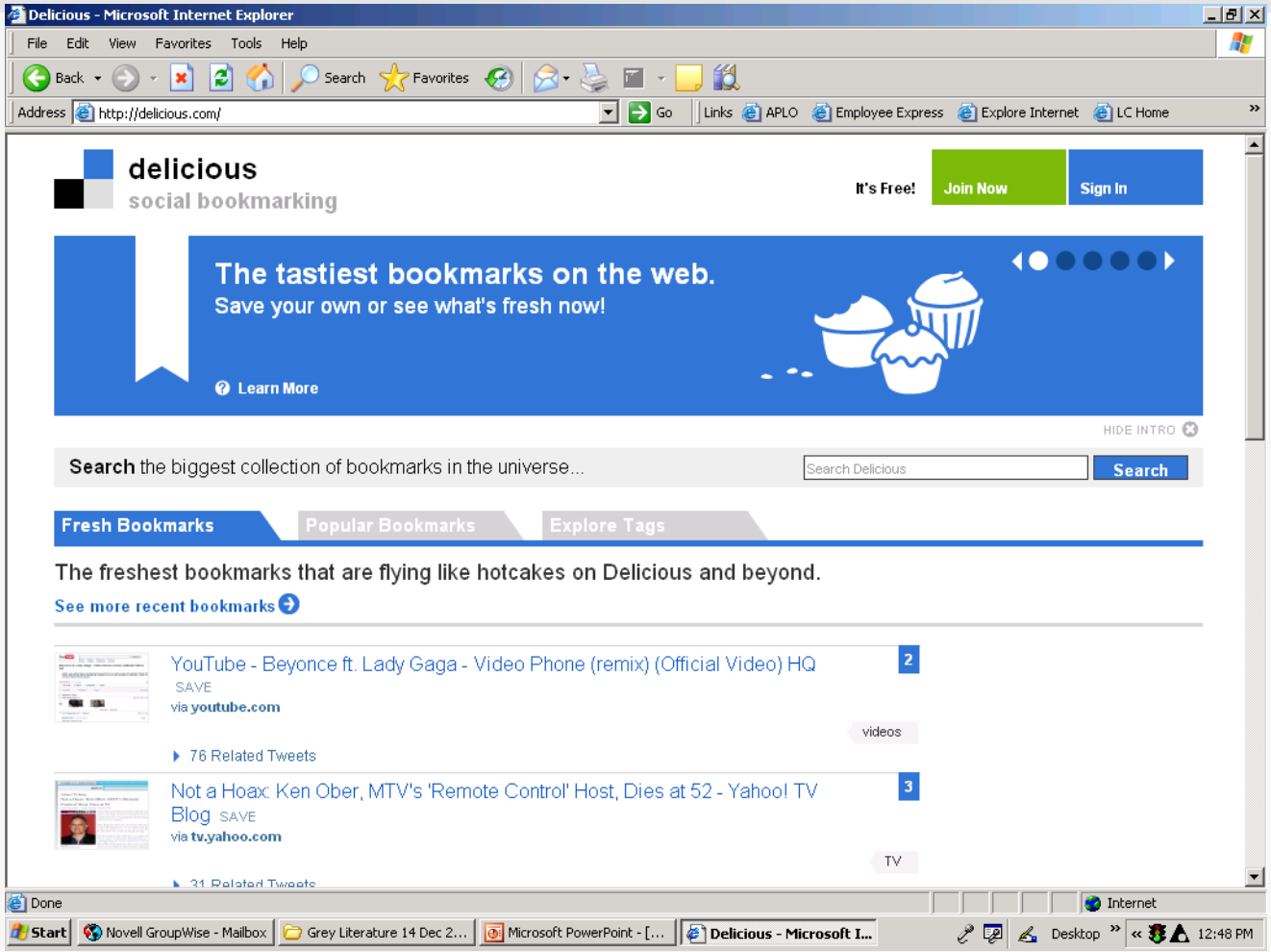
3.4 Grey Web Opportunities

Library of Congress P&P Flickr Pilot Project

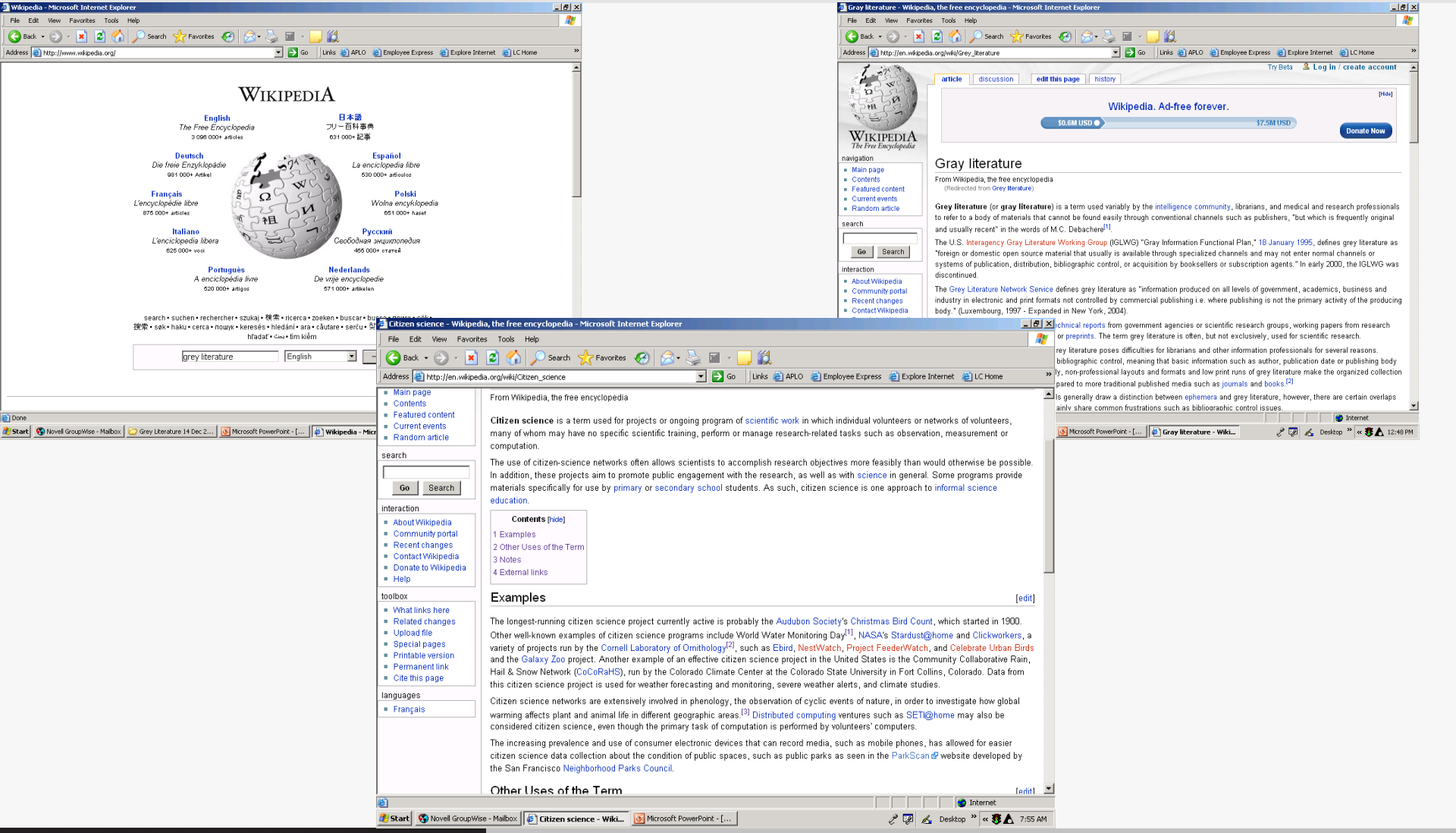


3.5 Grey Web Opportunities

Del.icio.us



3.6 Grey Web Opportunities



4.0 e-Grey
eScience – Data-Driven Science

- eScience
  - 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
  - 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

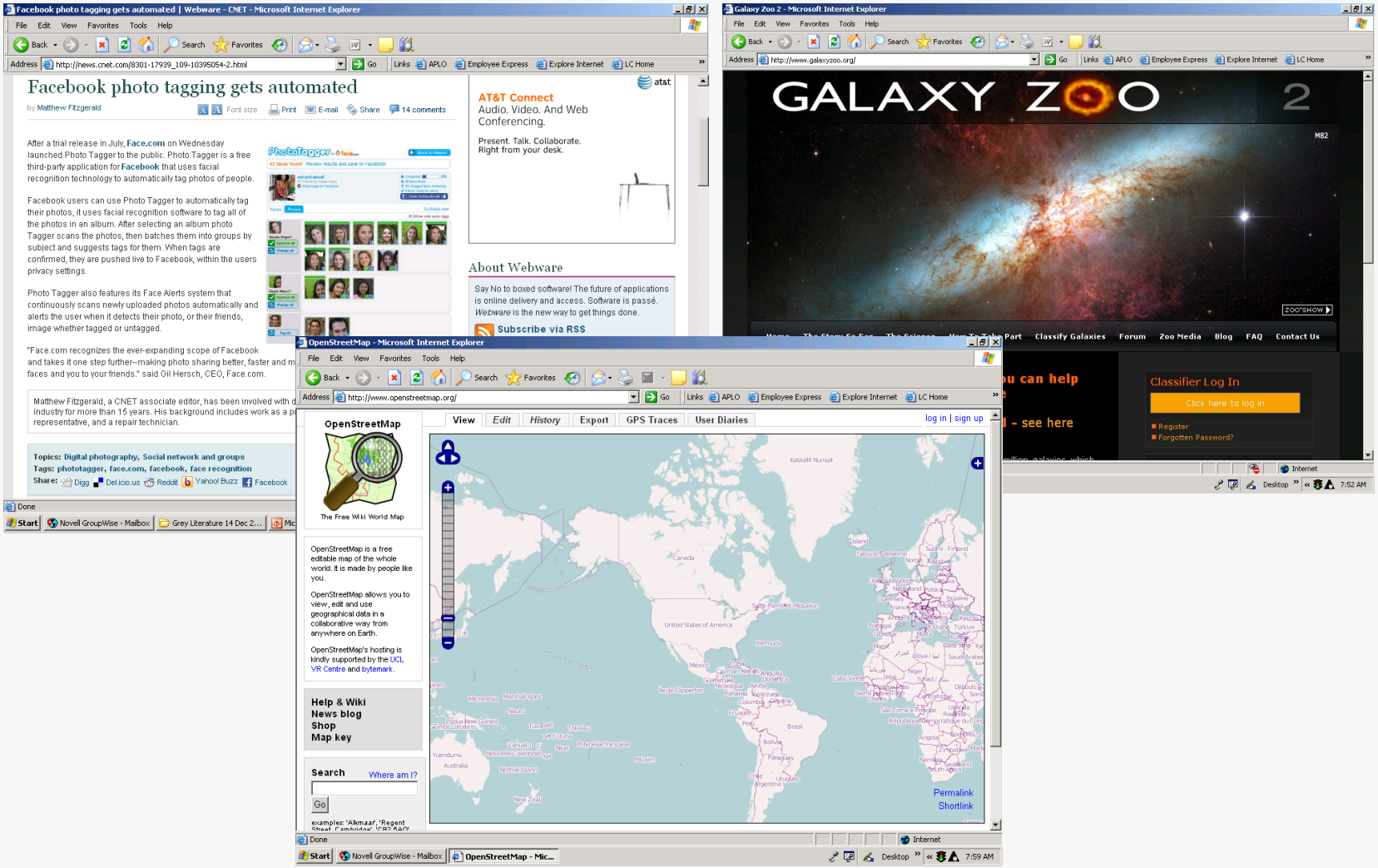
  Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, 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.

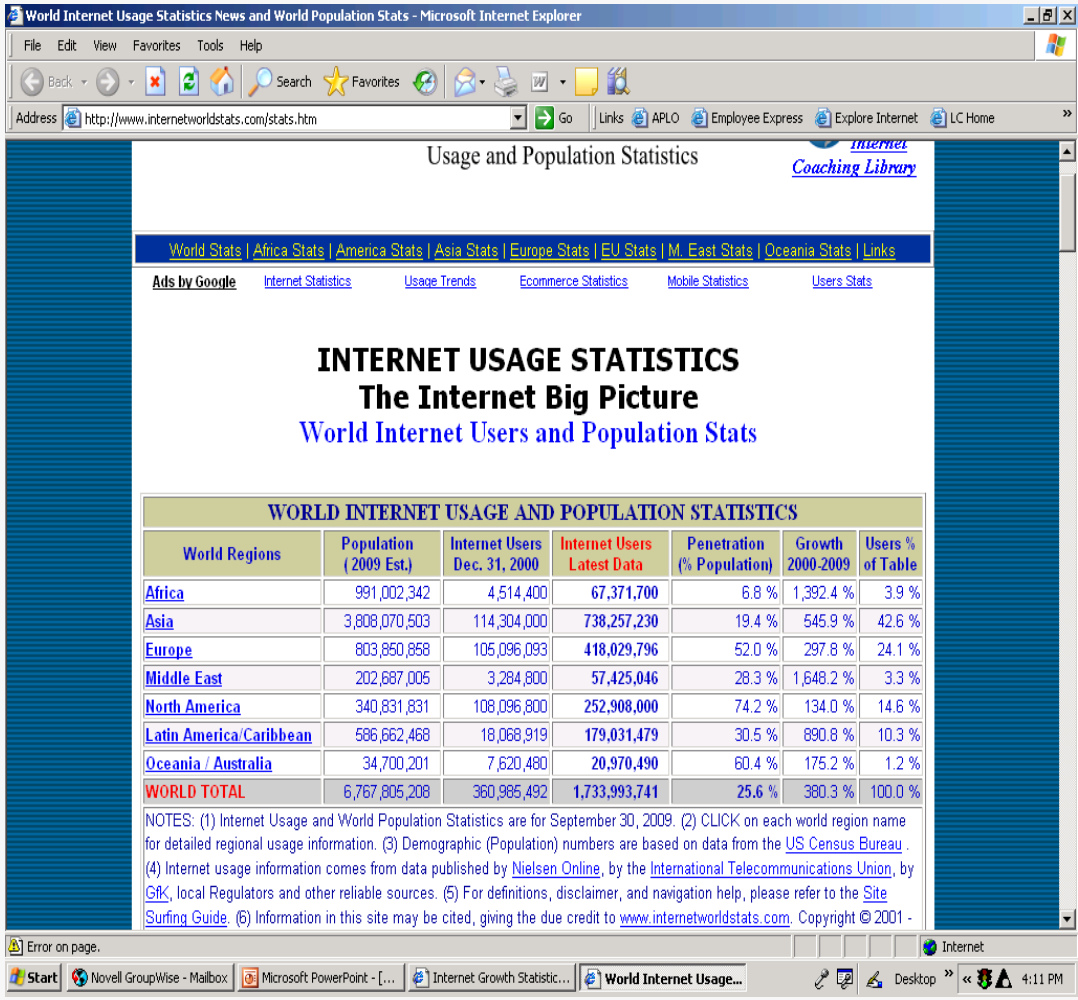
4.4 e-Grey Web Opportunities

Citizen Science



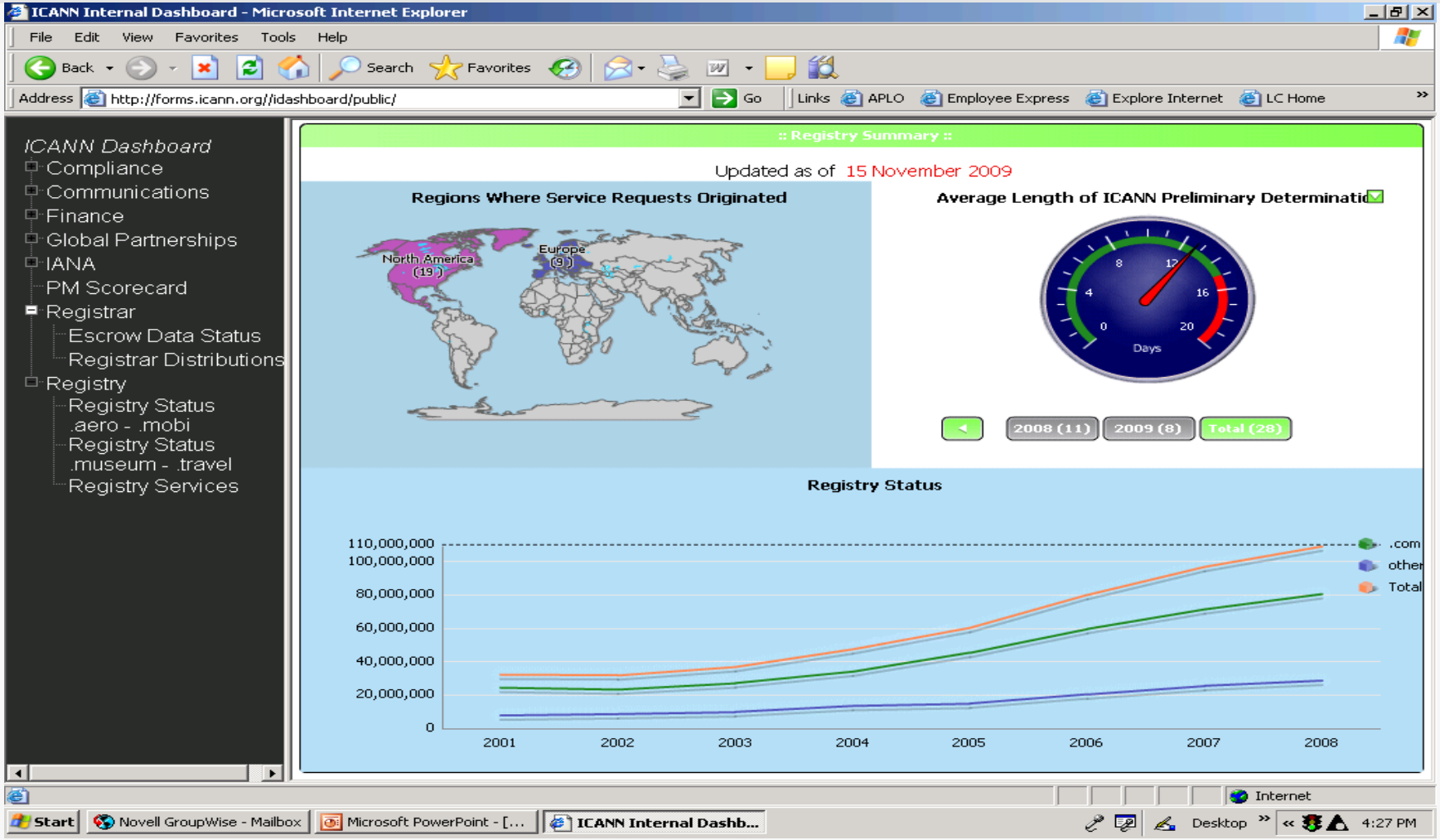
4.5 e-Grey
World Internet/Web Usage

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



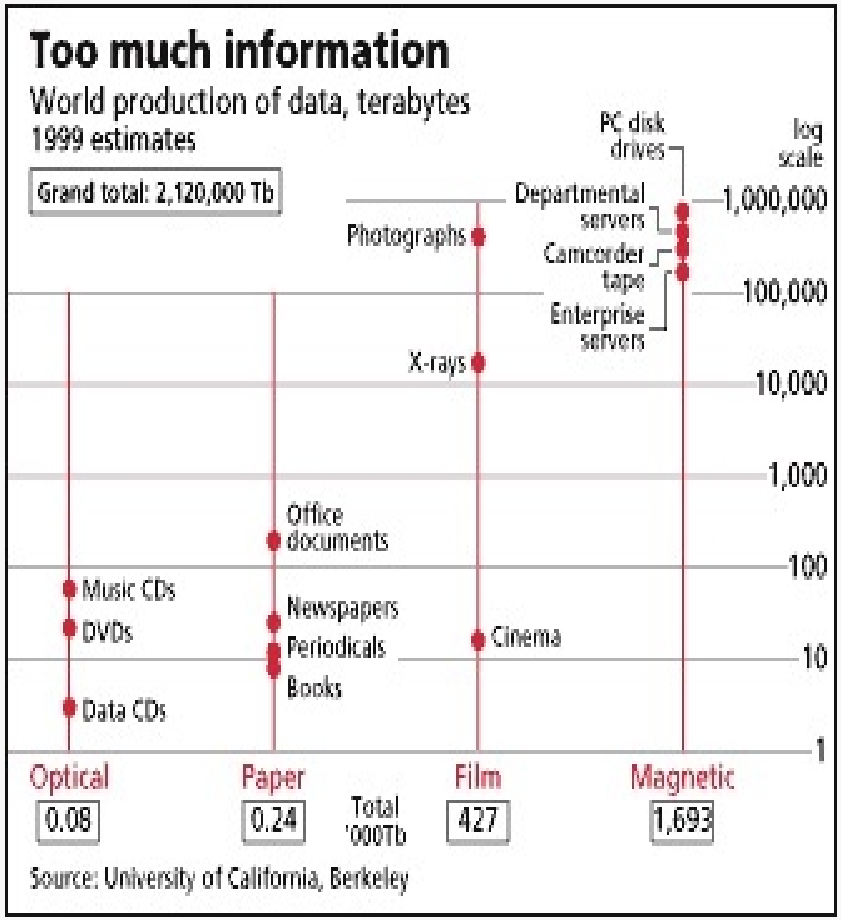
4.6 e-Grey

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.



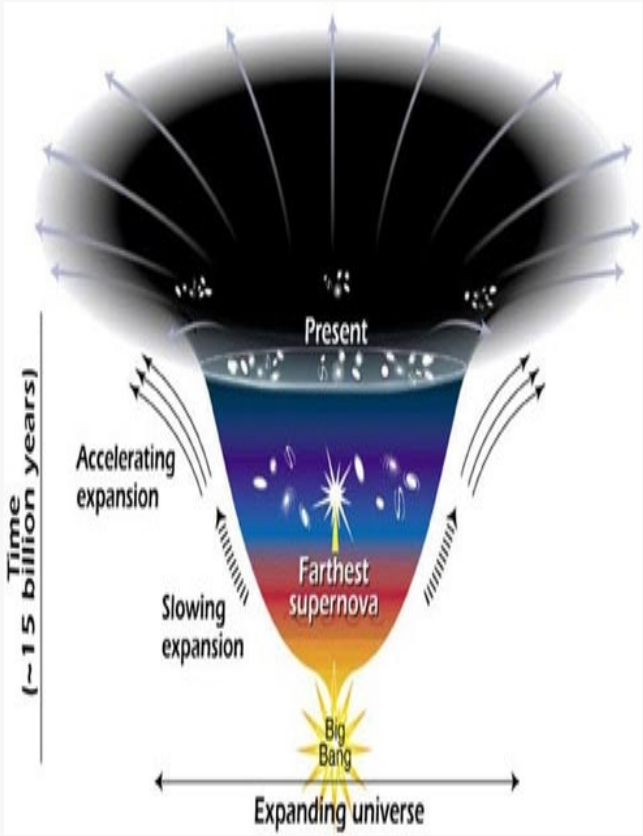
4.8 eGrey
A Sea Change in Computing

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

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