TITLE: Massive Scale Librarianship
AUTHOR(s): R. David Lankes, IST 676
PUBLICATION TYPE: Presentation
DATE: 2006
VENUE: Plenary Presentation at the Charleston Conference, Charleston, SC.
ABSTRACT: Overview of digital reference training, and how it integrates with
digital reference management.
KEYWORDS: Massive Scale Librarianship, Participatory Librarianship
Massive Scale Librarianship

R. David Lankes
and
IST 676

http://drew.syr.edu/MSL
http://www.DavidLankes.org
Gigabyte per Mile
Gigabyte per Mile

- Soon Every Mile of Road will Generate a Gigabyte of data a day
  - Road Sensors, Real-Time Traffic Data, Weather Information, Toll Data, Car Black Boxes
• Soon Every Mile of Road will Generate a Gigabyte of data a day
  - Road Sensors, Real-Time Traffic Data, Weather Information, Toll Data, Car Black Boxes

• It is assumed that this will become a gigabyte an hour. As there are 3.5 million miles of highways in the U.S. that would be 3.3 petabytes of data per hour, or 28 exabytes per year.
Exabyte?!
| Byte     | I byte: a single character; |

http://www2.sims.berkeley.edu/research/projects/how-much-info/datapowers.html
## Exabyte?!

<table>
<thead>
<tr>
<th>Byte</th>
<th>1 byte: a single character;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Byte</td>
<td>1 byte: a single character;</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
</tbody>
</table>

http://www2.sims.berkeley.edu/research/projects/how-much-info/datapowers.html
<table>
<thead>
<tr>
<th>Byte</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 byte: a single character;</td>
<td></td>
</tr>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>2 Gigabytes: 20 meters of shelved books</td>
</tr>
</tbody>
</table>
### Exabyte?!

<table>
<thead>
<tr>
<th>Byte</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 byte</td>
<td>a single character;</td>
</tr>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>2 Gigabytes: 20 meters of shelved books</td>
</tr>
<tr>
<td>Terabyte</td>
<td>2 Terabytes: An academic research library</td>
</tr>
</tbody>
</table>

http://www2.sims.berkeley.edu/research/projects/how-much-info/datapowers.html
# Exabyte?!

<table>
<thead>
<tr>
<th>Byte</th>
<th>1 byte: a single character;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>2 Gigabytes: 20 meters of shelved books</td>
</tr>
<tr>
<td>Terabyte</td>
<td>2 Terabytes: An academic research library</td>
</tr>
<tr>
<td>Petabyte</td>
<td>2 Petabytes: All US academic research libraries;</td>
</tr>
</tbody>
</table>

http://www2.sims.berkeley.edu/research/projects/how-much-info/dapowers.html
## Exabyte?!

<table>
<thead>
<tr>
<th>Byte</th>
<th>1 byte: a single character;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>2 Gigabytes: 20 meters of shelved books</td>
</tr>
<tr>
<td>Terabyte</td>
<td>2 Terabytes: An academic research library</td>
</tr>
<tr>
<td>Petabyte</td>
<td>2 Petabytes: All US academic research libraries;</td>
</tr>
<tr>
<td>Exabyte</td>
<td>5 Exabytes: All words ever spoken by humans.</td>
</tr>
</tbody>
</table>

[http://www2.sims.berkeley.edu/research/projects/how-much-info/datapowers.html](http://www2.sims.berkeley.edu/research/projects/how-much-info/datapowers.html)
<table>
<thead>
<tr>
<th>Byte</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Byte</td>
<td>1 byte: a single character;</td>
</tr>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>2 Gigabytes: 20 meters of shelved books</td>
</tr>
<tr>
<td>Terabyte</td>
<td>2 Terabytes: An academic research library</td>
</tr>
<tr>
<td>Petabyte</td>
<td>2 Petabytes: All US academic research libraries;</td>
</tr>
<tr>
<td>Exabyte</td>
<td>5 Exabytes: All words ever spoken by humans.</td>
</tr>
<tr>
<td>Zettabyte</td>
<td></td>
</tr>
</tbody>
</table>
## Exabyte?!?

<table>
<thead>
<tr>
<th>Byte</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Byte</td>
<td>1 byte: a single character;</td>
</tr>
<tr>
<td>Kilobyte</td>
<td>2 Kilobytes: A typewritten page;</td>
</tr>
<tr>
<td>Megabyte</td>
<td>2 Megabytes: A high resolution photograph;</td>
</tr>
<tr>
<td>Gigabyte</td>
<td>2 Gigabytes: 20 meters of shelved books</td>
</tr>
<tr>
<td>Terabyte</td>
<td>2 Terabytes: An academic research library</td>
</tr>
<tr>
<td>Petabyte</td>
<td>2 Petabytes: All US academic research libraries;</td>
</tr>
<tr>
<td>Exabyte</td>
<td>5 Exabytes: All words ever spoken by humans.</td>
</tr>
</tbody>
</table>

http://www2.sims.berkeley.edu/research/projects/how-much-info/datapowers.html
Paper Experiment

-Gladwell Tipping Point
Paper Experiment

- Imagine I Give you a Large Piece of Paper (.01” thick)

-Gladwell Tipping Point
Paper Experiment

• Imagine I Give you a Large Piece of Paper (.01” thick)
• You Fold It

-Gladwell Tipping Point
Paper Experiment

- Imagine I Give you a Large Piece of Paper (.01” thick)
- You Fold It
- And Fold It Again...50 Times

-Gladwell Tipping Point
Imagine I Give you a Large Piece of Paper (.01” thick)
You Fold It
And Fold It Again...50 Times
How Tall Would It Be?

-Gladwell Tipping Point
• Imagine I Give you a Large Piece of Paper (.01” thick)

• You Fold It

• And Fold It Again...50 Times

• How Tall Would It Be?

• Approximately the Distance Between the Earth and the Sun

-Gladwell Tipping Point
Predictable Change

• Unlike the Web, Massive Scale Computing is “Foreseeable”

• Moore’s LAW
  - Storage is Exceeding Moore’s Law
Issues

- **Privacy**
  - Total Information Awareness (DARPA)

- **Ethics of Involvement**
  - Unintended Censorship

- **Commercialization and Scarcity**
  - Radar Data

- **Unintended Consequences**
  - AOL Search Logs

- **Security**
  - eVoting and Diebold

- **Preservation: Migration and Emulation**
  - The Abandoned Grad Student at DRI

- **Culture and Control**
  - Google goes to China

- **Cognition and Perceptual Scarcity**
  - A weekday edition of the New York Times contains more information than the average person was likely to come across in a lifetime in seventeenth-century England. More new information has been produced in the last 30 years than in the previous 5,000. (Wurman)

- **Scalability and Sustainability**
  - Ask.com operations VP Dayne Sampson estimates that the five leading search companies together have some 2 million servers, each shedding 300 watts of heat annually, a total of 600 megawatts. These are linked to hard drives that dissipate perhaps another gigawatt. Fifty percent again as much power is required to cool this searing heat, for a total of 2.4 gigawatts. With a third of the incoming power already lost to the grid's inefficiencies, and half of what's left lost to power supplies, transformers, and converters, the total of electricity consumed by major search engines in 2006 approaches 5 gigawatts...almost enough to power the Las Vegas metropolitan area – with all its hotels, casinos, restaurants, and convention centers – on the hottest day of the year.
Options?
Options?

• Catalog it All
  - Tried It
Options?

- Catalog it All
- Tried It
- Ignore It
  - Commercialization of Data Stewardship
  - Dissolution of the Library as Institution?
    Continued Reduction to a Niche Player
Options?

• Catalog it All
  - Tried It

• Ignore It
  - Commercialization of Data Stewardship
  - Dissolution of the Library as Institution?
    Continued Reduction to a Niche Player

• Limit the Library
  - Selection versus Intellectual Freedom
Options?

- Catalog it All
  - Tried It
- Ignore It
  - Commercialization of Data Stewardship
  - Dissolution of the Library as Institution?
    Continued Reduction to a Niche Player

- Limit the Library
  - Selection versus Intellectual Freedom
- Embrace It
  - The Ethical Responsibility of LIS Education
Participatory Librarianship

- Knowledge is Created through Conversation
  - Conversation Theory
  - Scalable
    - Intra-individual, Inter-personal, Organizational
- Artifacts are Simply By-Products

http://iis.syr.edu/
Mission of the Participatory Library

• To facilitate the knowledge creation of its community through conversation
Not Really New
Not Really New

• Socratic Method
Not Really New

• Socratic Method
Not Really New

- Socratic Method
- Book Groups
Not Really New

- Socratic Method
- Book Groups
- Programs
  - Speaker Series
  - Story Time
Not Really New

- Socratic Method
- Book Groups
- Programs
  - Speaker Series
  - Story Time
- Reference
Not Really New

- Socratic Method
- Book Groups
- Programs
  - Speaker Series
  - Story Time
- Reference
- Collection Development
Not Really New

- Socratic Method
- Book Groups
- Programs
  - Speaker Series
  - Story Time
- Reference
- Collection Development
- Preservation
The Hidden Item Focus

- LIS Theory Development
- Evaluation
- Difficulty Incorporating All Library Functions
- CORC, LII
- Bibliofundamentalism
  - An Extreme that Sees the Role of the Library as a Haven of Order and Quality
  - Ultimate Focus on Collection
  - NOT SCALABLE!
Did it Rain June 18th in Yellowstone?
Did it Rain June 18th in Yellowstone?
Yes
Did it Rain June 18th in Yellowstone?
Yes
How Sure Are You?
Did it Rain June 18th in Yellowstone?
Yes
How Sure Are You?
Did it Rain June 18th in Yellowstone?

Yes

How Sure Are You?
Did it Rain June 18th in Yellowstone?

Yes

How Sure Are You?
Did it Rain June 18th in Yellowstone?
Yes
How Sure Are You?
Did it Rain June 18th in Yellowstone?

Yes

How Sure Are You?
Did it Rain June 18th in Yellowstone?
Yes
How Sure Are You?
Did it Rain June 18th in Yellowstone?
Yes
How Sure Are You?
Did it Rain June 18th in Yellowstone?
Yes
How Sure Are You?
Participatory Library System

Catalog → "Amazon-Like" Catalog → True Catalog → Participatory Library

User Data

Digital Catalog → Federated Search

Databases

External Feeds and Sites

Institutional Repository → Digital Repository → Community Repository

Digitization and Preservation

Reference → Knowledge Base

Virtual Reference

Hosted Community Services (Wiki’s, blogs, etc.)
Example

Library
Example
Library

Library

Lib.Local
PEOPLE, Reference Transactions, blog posts, wikis, events, annotations, books, articles, digital images....

Lib.Network
PEOPLE, Reference Transactions, blog posts, wikis, events, annotations, books, articles, digital images....

Web Resources
Commerce, videos, search engines, general web available information

Institutions
Governments, Humanities

Massive Data
Research results, sensor data, imagery, interaction data
C O N V E R S A T I O N S

Lib.Local
PEOPLE, Reference Transations, blog posts, wikis, events, annotations, books, articles, digital images....

Lib.Network
PEOPLE, Reference Transations, blog posts, wikis, events, annotations, books, articles, digital images....

Web Resources
Commerce, videos, search engines, general web available information

Institutions
Governments, Humanities

Massive Data
Research results, sensor data, imagery, interaction data
Engage!

- To Ignore the Implications of Massive Scale Computing is Dangerous.
  - It Abdicates Serious Decisions and Consequences to Others

- Massive Scale Librarianship is an Opportunity to Not Only Enhance the Mission of the Library, but Proactively Position Librarians at the Forefront of the Information Field...Where they Belong