Title
Sharing Metadata: Building Collections and Communities

Permalink
https://escholarship.org/uc/item/5qf9h2k1

Authors
Chandler, Robin L.
Westbrook, Bradley D.
Rundblad, Kevin

Publication Date
2009-05-09
Sharing Metadata: Building Collections and Communities

• Robin Chandler, UC San Diego Libraries
• Brad Westbrook, UC San Diego Libraries
• Kevin Rundblad, UCLA
• Presented at Annual Meeting of the Society of California Archivists, Riverside, CA, May 9, 2009

Libraries, archives, and museums cannot afford to think about collections only in the context of the local community. Sharing benefits our users, and we can no longer assume our users will come in through the "front door." Cultural institutions must get their content "into the flow" where working users will discover and reap the benefits of digital content delivery. The speakers in this session will describe the benefits of creating repurposable metadata discoverable in multiple environments; discuss some of the obstacles challenging institutions to create sharable metadata; compare two mechanisms for creating repurposable metadata; and share some success stories where leveraging metadata in social networking environments helped users discover the breadth and depth of an institution's collections.
Sharing Metadata: Building Collections, Community & Computing Power

Robin L. Chandler
Society of California Archivists
May 9, 2009
• **Major Theme:**
  • Metadata Standards are the key to Sharing Metadata
  • Repositories amass collection metadata in local databases *locked* away from the internet public in **Data Silos**
  • Standardized metadata is flexible, shareable and promotes interoperability

• **Minor Theme:**
  • Tear Down your Data Silos & Share Your Metadata
“Shareable metadata is metadata which can be understood and used outside of its local environment by aggregators to provide more advanced services”
Sara Shreeves *Moving Towards Shareable Metadata*
What’s Wrong with Data Silos?
Data SILOS limit exposure of our collections to users...
Data SILOS
Freeze
Innovation

• Our users are innovators

• “Innovation comes only from readily & seamlessly sharing information rather than from hoarding it” Tom Peters, author *In Search of Excellence*
Why Share Our Metadata?

• Sharing Benefits our users
  • Data must be where users are working -- getting “Into the Flow”
  • Federates distributed collections
  • Supports “one-stop” searching
  • Supports Web 2.0 culture and services

• Sharing Benefits our institutions
  • Increases exposure to collections
  • Broadens user base
  • Increases potential for collaboration & grant seeking opportunities
  • Enhances rather than limits your data
A History of Sharing:

• Knowledge
• Systems
• Standards
• Expertise

1877 – Standardization of size of catalog cards
1901 – LC catalog card distribution program
1968 – MARC
1971 – OCLC
1980 – RLIN
1987 – Z39.50
1994 – DCMI (Dublin Core Metadata Initiative)
1997 – MARC 21
1998 – EAD 1.0
2001 – OAI-PMH ; METS ; MODS
2004 - Web 2.0
2007 – OAI-ORE
Evolution of Sharing: Analog to Web 2.0

• Analog: Sharing the Creation of Authoritative Catalog Records
  • Librarian expert creates a catalog record and shares the record with other Librarian experts
  • Motives:
    • Encourages shared cataloging & conserving staff resources
    • Encourages cooperative collection development

• Web 2.0: Sharing the Production of Knowledge
  • Archivist expert creates an interoperable digital object with standardized metadata and shares the object with users
  • Motives:
    • Increase exposure of collections
    • Encourage metadata enhancement by distributed experts
  • In socially networked environment, expert users enhance digital object with additional metadata. This is joint authorship using the distributed expertise of human computing power.
Key attributes to shareable metadata:
• Human understandable outside of its local context
• Provides contextual information at the object level
• Consistent across a collection: uses same data fields and same controlled vocabulary across a collection
• Easily packaged in multiple views for multiple audiences
• Machines can process
• Uses Standards: descriptive, technical, administrative
What is a standard?

“a description of a measurable item or process” which is agreed upon and published by recognizable standards-issuing organization such as national and international associations.”

“Standards are used to assure that communication is clear. The use of a recognized industry standard assures everyone that processes and measurements are uniform.”

“Clear thinking about what we are doing with description is essential to creating useful tools and delivering them in a helpful and understandable format.”

--Michael Fox
2000 SAA Presentation
## Types of Standards

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>How descriptive data is expressed</td>
<td>DACS</td>
</tr>
<tr>
<td>Value</td>
<td>How descriptive content is controlled</td>
<td>LCNAF, LCSH, AAT</td>
</tr>
<tr>
<td>Structure</td>
<td>Specification of the elements of description, technical and administrative metadata and their relationship to one another</td>
<td>MARC, MODS, DC, VRA, MIX, PREMIS etc.</td>
</tr>
<tr>
<td>Communication</td>
<td>How description is formatted and transmitted</td>
<td>MARC, EAD, METS</td>
</tr>
</tbody>
</table>
A repository applying metadata standards will create effective metadata records for the

*discovery,
*sharing,

and,

*ongoing management

of its digital objects.
Data Silos Can Happen Because:

- We lack staff & skills to implement standards
- We perceive our Users as Local Only
- We recognize that storage is cheap
- We believe we can’t keep up with dynamic standards
• Data standards are stable....they are the products of thorough intellectual process and are maintained by groups of experts.
• Data standards are also dynamic – evolving in response to community needs.
• Institutions customizing standards face migration issues
• Leads to local instability as institutions must resolve problems without community solutions
ALA Tiaga Forum Provocative Statements Discussion, 2006

“Within the next five years.......”
• The continuing disaggregation of content from original containers causes a revolution in resource discovery
• There will no longer be a monolithic “library” Web site....instead, data will be pushed out to many starting places on the Web and directly to users
• All information discovery will begin at Google, including discovery of “archival” resources

- Since 2003, the number of scholars across disciplines who report starting their research at non-library discovery tools, has increased.
- Over 90% of librarians list the gateway/portal role as very important.
- There is a disconnect!
So, don’t let your data be shrouded in the fog of mystery............
Tear down SILOS!
Share your Metadata!
Promote Discovery & Use of your Data!

• Recommends making EAD finding aids accessible via online catalogs and internet

• Encourages inter-institutional collaboration for sharing metadata records
CLIR’s Cataloging Hidden Special Collections and Archives Program, 2009

“Applicants will be expected to base their proposals on technical solutions that already exist for swift and efficient entry of data which can be translated into standard records formats such as EAD and MARC”
Thank You!

rochandler@ucsd.edu

“Stop hugging your data”
Sir Tim Berners-Lee, 2009
Conveying Metadata Among Silos

Bradley D. Westbrook
Metadata Analysis and Specification Unit
UC San Diego Libraries
Overview

• Reframing the “silo” problem
• Challenges for metadata sharing
• Tools / methods for metadata sharing
• The problem of metadata synchronicity
Some facts about metadata silos

- Inevitable
- Ubiquitous
- Beneficial
What’s good about metadata silos?

• Metadata silos share core purposes
  – Finding
  – Identifying
  – Selecting
  – Acquiring

IFLA Study Group on the Functional Requirements for Bibliographic Records (1997)
So what’s the problem?

It is the conveyance, or transmission, of metadata. In a word, the “shareability” of metadata.

What is needed is better understanding of what is involved in sharing metadata, or conveying it from one silo to another, and development of techniques and processes that will promote metadata sharing in a manner both beneficial and cost-effective.
What’s so important about sharing?

• Efficiency

• Increased access

• New data products

• Professionalism
Challenges for metadata sharing

- Increasingly complicated data environment
- Transformation barriers
Increasingly complicated metadata environment

- Book catalogs
- AACR and card catalogs
- Multiple content standards for different kinds of objects and functions
- Multiple format standards for communicating metadata
- Intersecting but bounded knowledge consuming / producing communities
Transformation barriers

Inequalities among standards create barriers when conveying metadata from one environment to another.

Kinds of Inequalities

• Specific to general
• Differences in data format/values
• One to many (splitting)
• Many to one (aggregation)
• Structural differences
• Missing data
• Superfluous data
Specific to general

Creator

<table>
<thead>
<tr>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustrator</td>
</tr>
<tr>
<td>Sculptor</td>
</tr>
<tr>
<td>Etc.</td>
</tr>
</tbody>
</table>

Subject

650 0 Northern elephant seal | x Geographical distribution |
| z Southern California Bight (Calif. and Mexico) |

9 May 2009  SCA Annual Meeting
Format / value differences

Syntax

**source**
06-11-1958
mm-dd-yyyy

**target**
1958-11-06
yyyy-mm-dd

Vocabulary

file types:

**source**
Image
Text
Audio
Video

**target**
still image
text
sound recording
moving image

9 May 2009
SCA Annual Meeting
One to many

The value of a single source element is divided into multiple elements in the target schema:

source (local database)

<table>
<thead>
<tr>
<th>Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Submarine valleys; Ocean bottom; Marine Sediments; Scientific Expeditions—Mexico</td>
</tr>
</tbody>
</table>

processing: Multiple values are separated by a semicolon in the database, these should be separated and each value put into a separate <mods:genre> element.

target (dc)

<dc:subject>Submarine valleys</dc:subject>
<dc:subject>Ocean bottom</dc:subject>
<dc:subject>Marine Sediments</dc:subject>
<dc:subject>Scientific Expeditions -- Mexico</dc:subject>
Many to one

The values of multiple source elements are combined into a single target element.

**source (database)**
- Repositories/repositoryName
- Repositories/address1
- Repositories/address1
- Repositories/address3
- Repositories/city
- Repositories/region
- Repositories/mailCode
- Repositories/country
- Repositories/url

**target (mods)**
- mods/note[@displayLabel="Digital object made available by"]

Processing:
- Concatenate repositoryName, address1, address2, address3, city region mailCode, country (url)

<note displayLabel="Digital object made available by">Mandeville Special Collections Library, University of California, San Diego, La Jolla, CA 92093-0175 (http://orpheus.ucsd.edu/speccoll/)</note>
Structural difference

Hierarchical

<mods xsi:schemaLocation="http://www.loc.gov/mods/v3 http://w
<relatedItem type="constituent">
  <titleInfo>
    <title>Report to the Commissioner of the Federal Public Housing Authority</title>
  </titleInfo>
  <identifier type="ARK">
    http://ark.cdlib.org/ark:/13030/hb50000652/FID28
  </identifier>
</relatedItem>
<relatedItem type="constituent">
  <titleInfo>
    <title>Minutes from Housing Authority meeting</title>
  </titleInfo>
  <identifier type="ARK">
    http://ark.cdlib.org/ark:/13030/hb50000652/FID15
  </identifier>
</relatedItem>
</mods>

<dc xsi:schemaLocation="http://www.cdlib.org/schemas/xmldata http://ark.cdlib.o:
<relation>Minutes from Housing Authority meeting</relation> 1
<relation>Report to the Commissioner of the Federal Public Housing Authority 2
<identifier>http://ark.cdlib.org/ark:/13030/hb50000652/FID28</identifier> 1
<identifier>http://ark.cdlib.org/ark:/13030/hb50000652/FID15</identifier> 2
</dc>

9 May 2009  SCA Annual Meeting  14
Missing data

• Collection Description

<table>
<thead>
<tr>
<th>Item Titles</th>
<th>Theodore Roosevelt on Film</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On a horse</td>
<td>The Library of Congress</td>
</tr>
<tr>
<td>2. Czar Nicholas</td>
<td>American Memory Home</td>
</tr>
<tr>
<td>3. Peace envoys</td>
<td>Browse Collections</td>
</tr>
</tbody>
</table>

Features:
- Timeline
  - 1889-1919
  - 1690-1921
- Essays
  - T.R. on Film
  - Sound Recordings of T.R.'s Voice

Browse Collection by:
- Title
- Subject

Theodore Roosevelt

Scenes of T.R. and his sons Quentin and Archie, 1917-1918
About this image

Overview

Theodore Roosevelt was the first U.S. president to have his career and life documented on film, and is known for his involvement in preserving the American heritage.

• Technical Metadata
to support different system functions, like preservation
Replacing missing data

• Take from collection level metadata or context
• Draw on non-explicit metadata (e.g., organizational practices, workflow documentation, cataloging guidelines)
• Extract from files (mainly for technical metadata)
• Infer from existing metadata (e.g., if a file has an mp3 extension, assign a mime type of audio/mpeg)
Superfluous local metadata

Relevant in the original context of an item but not in its new environment.

Workflow and management metadata, e.g.
who digitized an item and other work actions
accession and stack location information
notes about the physical condition of a source item
terms used to drive local access mechanisms
Tools / methods sharing metadata

• Basic software programs
  – Excel
  – XML editors
  – XSLT
  – Altova Map Force

• Collection management / metadata authoring systems
  – UC San Diego Libraries’ DAMS
  – Archivists’ Toolkit
Excel

- Good for visual scanning of data for patterns or disruptions of patterns
- Easy to add data
- Supports a variety of data manipulation
  - Find and Replace
  - Concatenation
  - Date formatting
  - A lot more...
Archivists' Toolkit export metadata map

<table>
<thead>
<tr>
<th>Source</th>
<th>Processing</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Names</td>
<td>personalDirectOrder variable</td>
<td>IF personalDirectOrder = True 0</td>
</tr>
<tr>
<td>Names</td>
<td>personalDirectOrder</td>
<td>IF personalDirectOrder = NULL 1</td>
</tr>
<tr>
<td>Names</td>
<td>nameType</td>
<td>IF nameType = family 3</td>
</tr>
<tr>
<td>n/a</td>
<td>n/a</td>
<td>constant</td>
</tr>
<tr>
<td>Names</td>
<td>personalPrimaryName</td>
<td>IF nameType = personal and personalRestOfName = NULL</td>
</tr>
<tr>
<td>Names</td>
<td>personalRestOfName</td>
<td>IF nameType = personal and personalDirectOrder = True</td>
</tr>
<tr>
<td>Names</td>
<td>familyName</td>
<td>IF nameType = family</td>
</tr>
</tbody>
</table>

9 May 2009  SCA Annual Meeting
XML Editors

• For working with XML encoded data
  – Can import excel spreadsheets into xml format
  – Can be used to process xml data with an xslt style sheet
  – Contextual find and replace
  – Record / model validation
XSLT

Extensible Stylesheet Language Transformations
- For transformation of XML encoded data
- High technical requirement to create, modify and troubleshoot
- Can be reused & shared
- Supports more advanced data transformations
UC San Diego Libraries’ DAMS
Multiple metadata views via XSLT
METS view
RDF graph view
Archivists’ Toolkit
AT DO exports
The problem of metadata synchronicity
Linked resources

Rendered locally, with additional local metadata, as:

Creator: Duchamp, Marcel, 1887-1968
Title: Nude descending a staircase, 1912
Dim.: 58” x 35”
Type: Oil painting
Topics: Cubism
Notes: Part of 1913 Armory show in NYC
Thank You!

Bradley D. Westbrook
bdwestbrook@ucsd.edu
Metadata in a
Metadata in a Crowd
Metadata in a...
Metadata in a crowd
Metadata in a crowd shared knowledge
Hi, I’m Kevin
Hi, I’m Kevin Rundblad
What is Metadata in a Crowd?

INTRO: Real 2.0 vs. Simulated 2.0

MAIN: Simon Says “Tag This Photo”
       Human Computation Models
Thinking in 2.0 – Engaging the User
2.0 is a Culture, Not
2.0 is a Culture, Not Tech
If the 2.0 idea is not about the user …it is artificial 2.0
Real 2.0 comes directly from users
Traditional development will struggle with 2.0 development, since it is focused on the process, not the user.
Technology only expresses 2.0 ideas
2.0 is about Creators / Participants
Individuals post valuable info
Many consume it
Expert

Seeks Objective Articulation
Structured and Precise
Narrowly Defined Expertise
Social

Very Wide Knowledge
Fragments of Expertise
Tends to be Locally Derived
Subjective Perspective
Emotional Metadata
Social knowledge: Many Perspectives
Is the Crowd a Viable Source of Metadata?
Aren’t Users Lazy?
Um, Yes and Yes
But they do like to play games…
...spending many hours earning points that don’t mean anything
…motivation is key for participation
...and participation often means completing an intelligent task

(ok, mindless too)
Human Computation
Human Computation

Finding tasks that humans can do better than computers
Human Computation

Finding tasks that humans can do better than computers

And creating motivation to perform them
Human Computation and
Human Computation and P2P Model
Human Computation and P2P Model
Human Computation and P2P Model
Human Computation and P2P Model

One Video File requested

BitTorrent P2P Architecture
Human Computation and P2P Model

One Video File requested

Many systems each delivering parts of file, until entire file is complete
Socially formed knowledge
is like the P2P distribution architecture
Social Knowledge - Human Computation

Content formed from many individuals

Many individuals contributing parts of the content
But “…unlike computer processors, humans require some incentive to become part of a collective computation”.

Louis Von Ahn

Source: invited talk at CMU on Human Computation
Models of **Human Computation**

**AKA** Human Intelligence Tasks (HITs)
(Coined by Amazon)
Human Intelligence Tasks – 3 Types

- Socially motivated
- Economically motivated
- Tacit (user may be unaware of task itself they perform)
Human Intelligence Tasks (social)
(or social knowledge systems)

Tagging (Flickr) and Comments (Amazon)
Ratings (Yelp)
Problem Solving (StackOverflow)
Tagging @ flickr
Tagging @ flickr

Wider Discovery

Public Tagging
Ratings @ Yelp

Japanese auto Los Angeles

1. Fuji Auto
Category: Auto Repair
Neighborhood: Pasadena

The best, most honest auto repair place I have ever done business with. Competent repair folks and they work quickly and the prices are not as high as the dealerships. They only work on Japanese cars.

2. Ed Little Auto Service
Category: Auto Repair
Neighborhood: Culver City

love ed little auto service love bob love yelp (for letting me find ed little) it was the fourth time in two months that I had to visit auto mechanics because of my right headlight. thank GOD!
Ratings @ Yelp

Efficient Discovery
Ratings @ Yelp

Efficient Discovery

“Hearsay” becomes Reliable
Ratings @ Yelp

Efficient Discovery

“Hearsay” becomes Reliable

Marketing becoming less credible
Ratings @ Yelp

Efficient Discovery

“Hearsay” becomes Reliable

Marketing becoming less credible

Many perspectives creates trustworthiness
Vinh Loi Tofu
Categories: Vegetarian, Vietnamese
Neighborhood: Reseda

It's good!

I've been walking around for months thinking that Vietnamese food was off the menu for me; I stopped eating meat. I think this place just put it right back on the menu!

I think it's really scary how close they come to mimicking meat at the same time not! You'll just have to try it to maybe understand.

While at the place, the owner was quite friendly and warm; same with other customers patronizing the place. I asked a couple of the customers if they were vegan or vegetarian, they both said no. They just mentioned that the food was good.

Another cool thing to mention is that the own mentioned that he would be willing to teach people how to cook anything he served.

Enough said, go try this place if you like Vietnamese food.
Problem Solving @ stackoverflow

Recent Questions

- **Dynamic search result when typing**
  - Votes: 0
  - Answers: 3
  - Views: 51
  - Tags: asp.net, search
  - Time: 2m ago
  - Author: Community

- **Changing flash files on screen by clicking a button**
  - Votes: 0
  - Answers: 1
  - Views: 124
  - Tags: flash, change
  - Time: 2m ago
  - Author: Community

- **Logging in to meetup.com with Curl**
  - Votes: 0
  - Answers: 0
  - Views: 1
  - Tags: curl, php
  - Time: 2m ago
  - Author: Jonathan Lyon

- **Threading for distance vector which does not drop packets.**
  - Votes: 0
  - Answers: 0
  - Views: 1
  - Tags: java, threading, distance, vector, algorithm
  - Time: 2m ago
  - Author: sunny

- **Reading text with Java Scanner next(Pattern pattern)**
  - Votes: 0
  - Answers: 0
  - Views: 2
  - Tags: java, scanner, next
  - Time: 2m ago
  - Author: burntsugar

- **Regular Expressions, overhyped??? [closed]**
  - Votes: 2
  - Answers: 13
  - Views: 180
  - Tags: discussion, regular-expression, regex, subjective, argumentative
  - Time: 3m ago
  - Author: rooskie
I'm a member on several forums that have a subforum dedicated to programming questions.

It is an observation that to almost every single question about finding data in strings, the most common answer is "use regular expressions". And if you're lucky, a link to a tutorial.

For many situations that people will generally recommend regular expression. It's pure overkill.

Many things can be done with substring() and split() functions, yet people always seem to put regular expressions on a throne. Why is this?

Does this have something to do with the history of regular expressions versus these OO-language functions? Are the people who recommend regular expressions old farts?

Why in the name of the Matrix should I use a regular expression and test it, to see if a string contains a questionmark, if I can just do string1.contains("?");

While in some cases, such as form validation, regular expressions are useful, I find that in most cases I don't need them.

So how often do you use regular expressions, not because it's useful, but because it's what you're used to using.

EDIT: If you're gonna vote to close this, at least have the decency to make a comment as to why!
Motivation to Participate

Problem Solving @ stackoverflow

Badges

Ask questions, provide answers, vote for the things you find helpful — and Stack Overflow will bestow badges upon you. Here's a list of all the badges, along with a count of how many users have earned each one so far:

- Autobiographer × 11388: Completed all user profile fields
- Beta × 3771: Actively participated in the Stack Overflow private beta
- Citizen Patrol × 2591: First flagged post
- Civic Duty × 1638: Voted 300 times
- Cleanup × 1927: First clean edit
- Commentator × 8474: Left 10 comments
- Critic × 10493: First down vote
- Disciplined × 538: Deleted troll post with 3 or more upvotes
- Editor × 20795: First edit
- Enlightened × 3203: First answer was accepted with at least 10 up votes
- Famous Question × 43: Asked a question with 10,000 views
- Favorite Question × 309: Question favored by 25 users

Legend

- Gold Badge
  - Gold Badges are rare. You'll have to not only participate but be skilled and knowledgeable to earn these. They're something of an accomplishment!
- Silver Badge
  - Silver badges are earned for longer term goals. Silver badges are uncommon, but definitely attainable if you're interested enough.
- Bronze Badge
  - Bronze badges are earned for basic use of the Stack Overflow site; they are relatively easy to earn.
Human Intelligence Tasks (economic)

Mechanical Turk (Amazon) - small tasks $0 - $10
TextEagle (Nathan Eagle) – tasks for mobile phone
Small Tasks @ Mechanical Turk

### Total Earnings

<table>
<thead>
<tr>
<th>Rewards You Have Earned</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approved HITs</td>
<td>$0.20</td>
</tr>
<tr>
<td>Bonuses</td>
<td>$0.00</td>
</tr>
<tr>
<td>Total Earnings</td>
<td>$0.20</td>
</tr>
</tbody>
</table>

### Your HIT Status

<table>
<thead>
<tr>
<th>Date</th>
<th>Submitted</th>
<th>Approved</th>
<th>Rejected</th>
<th>Pending</th>
<th>Earnings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep 20, 2007</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>$0.20</td>
</tr>
</tbody>
</table>

### HIT Totals

<table>
<thead>
<tr>
<th>HITs You Have Accepted</th>
<th>Value</th>
<th>Rate</th>
<th>HITs You Have Submitted</th>
<th>Value</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>HITs Submitted</td>
<td>13</td>
<td>43.3%</td>
<td>... Submitted</td>
<td>13</td>
<td>66.7%</td>
</tr>
<tr>
<td>Approved</td>
<td>2</td>
<td></td>
<td>... Rejected</td>
<td>1</td>
<td>33.3%</td>
</tr>
<tr>
<td>Rejected</td>
<td>1</td>
<td></td>
<td>... Pending</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

### HITs Available to You

<table>
<thead>
<tr>
<th>Title</th>
<th>Requester</th>
<th>Reward</th>
<th># of HITs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create short, interesting video</td>
<td>Ask.com</td>
<td>$10.00</td>
<td>1</td>
</tr>
<tr>
<td>Write a 350-500 word article on property for sale in Larnaca</td>
<td><a href="http://www.ContentSpooling.net">www.ContentSpooling.net</a></td>
<td>$4.00</td>
<td>1</td>
</tr>
<tr>
<td>Write a 350-500 word article on webbing recipes</td>
<td><a href="http://www.ContentSpooling.net">www.ContentSpooling.net</a></td>
<td>$4.00</td>
<td>1</td>
</tr>
<tr>
<td>Write a 350-500 word article on see keyword queue</td>
<td><a href="http://www.ContentSpooling.net">www.ContentSpooling.net</a></td>
<td>$4.00</td>
<td>1</td>
</tr>
<tr>
<td>Call in your best family travel story!</td>
<td>VoiceByVoice</td>
<td>$2.00</td>
<td>1</td>
</tr>
<tr>
<td>Add numbers to the wikipedia of statistics</td>
<td>Eric J. Silverberg</td>
<td>$2.00</td>
<td>1</td>
</tr>
<tr>
<td>Looking for great success stories ...</td>
<td>VoiceByVoice</td>
<td>$2.00</td>
<td>1</td>
</tr>
<tr>
<td>Have you ever worked for Gold's Gym?</td>
<td>David</td>
<td>$1.00</td>
<td>1</td>
</tr>
<tr>
<td>Have you ever worked for Bally Total Fitness?</td>
<td>David</td>
<td>$1.00</td>
<td>1</td>
</tr>
<tr>
<td>Have you ever worked for 24 Hour Fitness?</td>
<td>David</td>
<td>$1.00</td>
<td>1</td>
</tr>
</tbody>
</table>
Small Tasks @ Mechanical Turk

Small Tasks

Time vs. Economics
(does not add up)
Small Tasks @ TextEagle

Mobile Human Computation in Africa

TextEagle is a…”system enabling the 3 billion mobile phone subscribers living in the developing world to earn small amounts of money by completing short, SMS-based tasks.”

Nathan Eagle
Research Scientist, MIT

Source: http://web.media.mit.edu/~nathan/
Translation/Transcription Services

Question: Translate the phrase "Address Book" into Girama.

Question: Transcribe the following audio clip from a New York hospital.

Also Citizen Journalism
Human Intelligence Tasks (tacit)
*User may not be aware of how tasks are utilized.*

reCAPTCHA (based on CAPTCHA)
ESP Game (now Google Image Labeler)
CAPTCHA:

“Are you human?”

(CAPTCHA: Developed by Louis Von Ahn)
CAPTCHA:
(Completely Automated Public Turing Test To Tell Computers and Humans Apart)

Security based on human perception
Turing Test administered by AI

(CAPTCHA: Developed by Louis Von Ahn)
reCAPTCHA

Security + Failed OCR = Opportunity

(CAPTCHA: Developed by Louis Von Ahn)
reCAPTCHA

Currently being used to help scan books for Internet Archive

(graphic - http://recaptcha.net/learnmore.html)
reCAPTCHA

Bots getting better at deciphering the CAPTCHAs

(graphic - http://recaptcha.net/learnmore.html)
reCAPTCHA

Bots getting better at deciphering the CAPTCHAs

Creates feedback loop - Good thing, since it means OCR gets more precise at the same time

(graphic - http://recaptcha.net/learnmore.html)
ESP Game (now Google Image Labeler)

2 individuals match = high probability of reliable result
With Human Computation…

...computing becomes the coordinating force between many individuals and intelligent or perceptual tasks.
Human Computation Models

Social
Tagging (Flickr) and Comments (Amazon)
Ratings (Yelp)
Problem Solving (StackOverflow)

Economic
Mechanical Turk (Amazon) - small tasks $0 - $10
TextEagle (Nathan Eagle) – tasks for mobile phone

Tacit
reCAPTCHA (based on CAPTCHA)
ESP Game (now Google Image Labeler)
How can you create a platform for capturing social knowledge?
Shared Knowledge Production

Public platforms – ex. Flickr

Develop platform – ex. Simul8 Model
UCLA Library Simul8 Group Model

Listen to users – experience paradigm
See user mode, lifestyle, aesthetic
Engage in their mode – playful and experimental dev
Fast prototyping with high level programming
Student developers and designers
Creative Commons License
One Question:

Kevin Rundblad
kevinrundblad@library.ucla.edu
One Question: Are You Human?

Kevin Rundblad
kevinrundblad@library.ucla.edu