

Web-Scale Cataloging and Metadata Management

Matt Goldner and Karen Calhoun
OCLC End User Services & OCLC Metadata Services
ALCTS Forum
Monday, January 14, 2008



Matt Goldner, slides 1-15

“Within the next five years ...



...there will no longer be a monolithic library Web site. Instead library data will be pushed out to many starting places on the Web and directly to users.”

Provocative Statement #6,
<http://www.taigaforum.org/docs/ProvocativeStatements.pdf>



**“They come and go and draw
from the well” *I Ching*, hexagram 48**



- The Library as a center of collections
- The Library as a center of experts and tools to guide users to appropriate resources

Three information market trends that have revolutionized information creation, access and user expectations



Evidence of convergence between .com/.org, between library and open web sites is everywhere we look today:

-Search services, such as google scholar, yahoo subscriptions and MS academic search seek to integrate traditional library content and services into mainstream search engines.

-At the same time, library service organizations and nonprofits are looking to integrate library collections into the open web through such programs as Gale's accessmylibrary, elsevier's scirus, and of course OCLC's open worldcat.

-A rash of mass-digitization projects through private/public partnerships, including google book search, the open content alliance, microsoft book project has begun moving legacy print collections to the network.

-There are also a host of supporting services for library users: citeulike, librarything, etc.

-Services are not far behind: we've seen the reference desk move online via questionpoint, wondir, yahoo answers, google answers and the like.

Within the library portal, change is also happening quickly – ERM, digital repositories,

Bottom line – The delivery system for library content and expertise increasingly consists of a network of public/private partnerships, rather than simply a library portal or physical building.

From a user's point of view, the separation between "them" (.com) and "us" (.org/.edu) is disappearing. This introduces a new level of complexity for us, around the formation of partnerships between and among private/public institutions. Supporting these partnerships requires new organizational capacity, new policy, and new infrastructure.

Market Trend #2: Changing nature of “content”



1. Well over 50 million Americans create and contribute web content. That's over 1/3 of US internet users, so it's becoming more pervasive than we might think. People are commenting on and cataloging everything: books, music, videos, podcasts, personal collections.
2. . This explosion of digital content is both evaluative in nature (opinions, ideas, ratings, etc.) and newly created content itself.
3. One of the reasons we're seeing this trend is b/c content is increasingly digital and downloadable – so users are willing and able to create, contribute and consume content in ways they simply weren't able to before.

Market Trend #3:
Changing user expectations

Customized information
Personally programmed devices
Collaborating at a distance

From "Future Worker 2015: Extreme Individualization,"
Gartner (March 2006)



Web 2.0, social networks, and folksonomy (tagging) have changed the way people find and share information online. The power of "collective wisdom", and the social aspects of web 2.0 make finding the right information easier (and more fun).

These changes in information delivery and the information landscape will of course shape consumer behavior. A recently-published Gartner Group report discusses the likely behavior of information consumers in 2015.

People will customize 90 percent of the information, tools, and educational and technological resources they use (0.8 probability).

18-year-olds in developed countries will spend 50 percent of their waking hours using personally programmed devices for communication, entertainment, education, news and time management requirements (0.7 probability).

By 2015, workers will spend more than 80 percent of their time working collaboratively, and not necessarily face-to-face (0.6 probability).

Clearly, this shift in behavior will require us to change the way the we deliver service . . . Not just content, but also other services.

The Search Engine Challenge



Q: Where do you typically begin your search for information on a particular topic?

Search Engines

84%



Library Portals

2%

From OCLC: *Perceptions of Libraries and Information Resources* (2005)

Then: the user built their
workflow
around the library

Now: the library must build its
service
around the user workflow

Then: resources were scarce and
attention
was abundant

Now: attention is scarce and
resources
are abundant

The River



How do we deliver our value - collections, services and community ...

- To the user
 - On the network
 - At the point of need?



Getting into the “flow”



User workflows, not products or services, should be the focal point

Users care about the “jobs” they need to complete, not about products

Solutions that users need will vary by task and change over time

From *Outside Innovation*, Patricia Seybold (2007)

<http://www.worldcat.org/oclc/70778382>

Services should not be the focal point of the library, because they are not the focal point of the library patron.

Patrons care about the “jobs” they need to complete. They are happy to use tools and to buy and consume services that help them accomplish their scenarios, regardless of the provider.

The solutions that patrons need and value will vary by task and will change over time.

How do we get in the flow?



Aggregate services around users

- Discover resources, services, expertise
- Deliver needed items to users - wherever they are
- Share & Collaborate expertise, recommendations, experiences

Continue to focus on the core objectives

A New Kind of Library: Outreach and Engagement 

Build a vision of a new kind of library

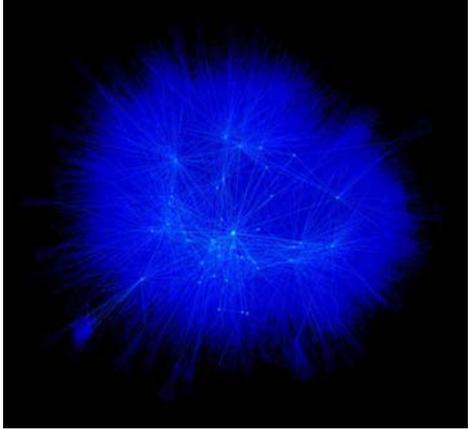
Be more involved with research, teaching, and learning materials and systems

Be more engaged with local communities

- The people
- Their data
- Their social networks

Make library collections and librarians more visible where users are

Move to next generation systems and services



An online social network

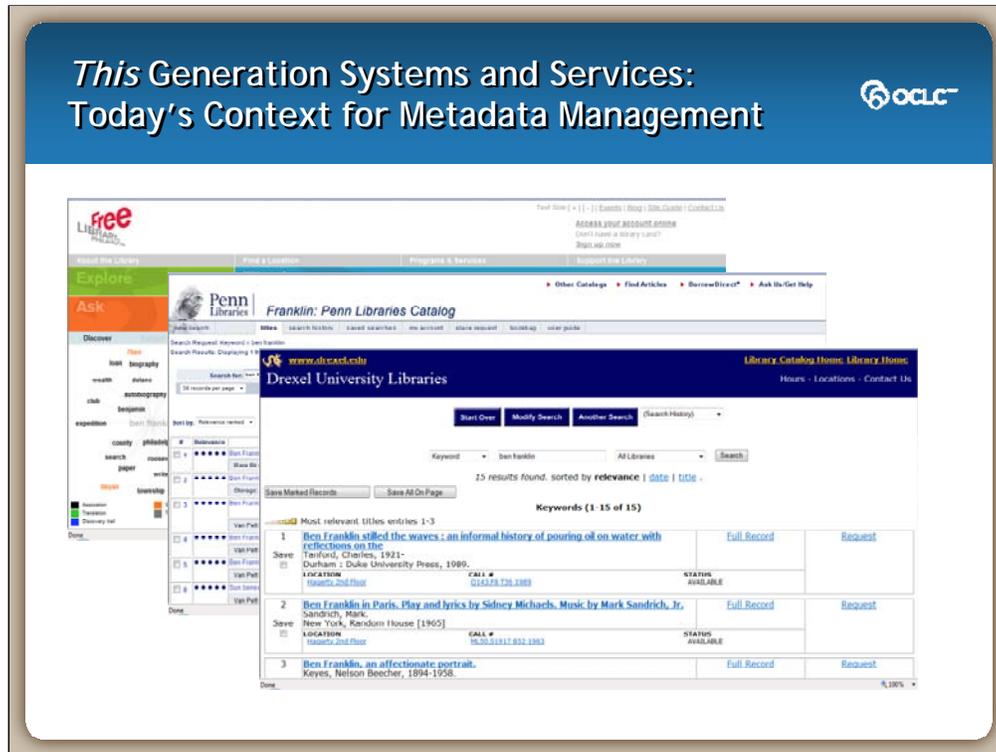
Karen Calhoun's notes for slides 15-32

Now that Matt has introduced the concepts of
A new kind of library characterized by outreach and engagement &
Next generation systems and services, in which users participate and
engage in online social networks,

Let us turn to our own expert community and how our present systems
help or hinder us from engaging with one another and doing work
together on behalf of end user communities

Yes, it can be quite illuminating to think of ourselves—that is collections
and technical services experts—as one of these online social
communities with shared purposes, using next generation systems and
services to accomplish our goals.

This Generation Systems and Services: Today's Context for Metadata Management



I would like to talk with you today about our present systems and services for sharing the work of metadata creation and maintenance and how those systems help or hinder us.

This is the Aquabrowser interface of the Free Library of Philadelphia. I don't mean to speak about the interface itself. I want us to think collectively about the metadata that underpins this interface and in particular how that metadata is created, captured, and in particular, maintained over time.

We manage metadata largely in silos today. Although the Free Library staff might have captured a record from a system like OCLC or a set of LC records or from a vendor initially, once it is in the local system the metadata tends to be managed, enriched, and maintained locally. [click twice]

Across town, the staff at the U of Penn and Drexel do pretty much the same, and on and on around the world's libraries.

This is where we are today and our collective systems, taken as a whole, mandates redundant effort, adds cost, and decreases the responsiveness we can provide to end user communities.

What If? Community-Based Metadata Management at Web-Scale



Complementary global, group and local metadata management? What would that look like? How would it work?

How would we aggregate such services around users?

With thanks to “Network Level Catalog Discussion” by Steve Shadle (University of Washington), PCC Policy Committee, November 1, 2007

- Community-based OCLC Enhance program
- Network-level catalog maintenance tools
- Network-level authority control processing

So I would like to challenge what IS today and talk together with you about how we might do better as a community of experts, and what the systems and services to support community based metadata management on the network would look like.

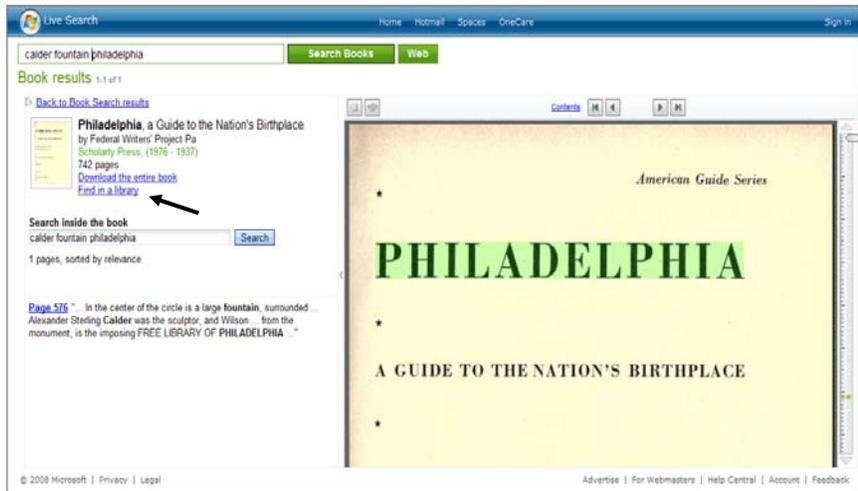
How could local, regional or national, and global efforts complement one another more effectively?

How could we assure that we are building community-based efforts around end user workflows?

Is there a way to engage users themselves in this work?

I'd like to pause here and thank colleagues at the University of Washington, particularly Steve Shadle, for challenging the cooperative cataloging community and OCLC to rethink how we collectively accomplish catalog management. A couple of Steve's suggestions conclude this slide. I think I can safely say that without the forward thinking of the University of Washington staff, we would not be having this program today.

What's Possible *Today?* (1a) Find Globally, Get Locally (or Regionally)



I hope you will agree, it should be as easy to borrow a book on the Web as it is to buy one there. To accomplish easy and convenient connections for users, and to effectively expose our rich library collections where users' eyes are, community and Web based metadata management hold the key.

Metadata created and managed at local, group and global levels is already being used today to connect users who find things in a variety of places on the Web with libraries that can supply the items to them.

I invite you at this point to think about metadata not only as a means to describe some object, but as the PRIMARY means of making CONNECTIONS between where users find things to how they can get hold of them.

Here is an example of network level metadata management in action today. I used to live in Philadelphia and I love the Calder fountain at Logan Square [explain example]

What's Possible *Today?* (1b) Find Globally, Get Locally (or Regionally)



The screenshot shows the WorldCat.org interface. At the top, there's a search bar with the query "in Philadelphia.+a+Guide+to+the+Nations+Birthplac". Below the search bar, the title "Philadelphia, a guide to the nation's birthplace." is displayed, along with the author "by [Eudora Wither's Project \(Pa.\)](#)" and the publisher "Philadelphia: William Penn Association of Philadelphia, 1937." The page includes sections for "Get It", "Save It", "Add to It", and "Share It". A "Find in a Library" link is highlighted with a black arrow. Below this, there's a "Enter Location Information" field with "43017" entered, and a "Go" button. The results show "Displaying libraries 1-10 out of 178 for all 6 editions" and "43017 Dublin, OH". A list of libraries is shown, with "1. [Ohio Hist Society](#) Columbus, OH 43211 United States" listed. The page also features a "Buy This Item" section on the right.

Worldcat.org / Find in a Library

The Find in a Library link leads to the record for this guidebook in WorldCat.org, which is freely available to everyone in the world with an Internet connection. Through the Open WorldCat program, WorldCat.org enables library metadata, created by our expert community, to function as a kind of switching system to help users easily navigate from wherever they find things – like on a Microsoft or Google page or on Yahoo – to local library collections.

Thank heaven for library metadata! Without the last 30 years efforts to share our records and holdings, this functionality would not be possible.

There are many features I could talk about on this page, but I'd like to draw your attention to the metadata management part that drives a search started on the Open Web to an individual library's holdings.

For the Philadelphia guidebook I found, I could have looked at the digitized copy on the MS Live Search Academic site, but if I want a print copy I can locate one by typing in my zipcode. In this case I learn that I can pick up a copy at the Ohio Historical Society.

So in a way we have already come a long way toward realizing a vision of managing the metadata we collectively have to support user workflows at the network level.

What's Possible *Today?* (2) Local Interface with Group and Global Reach

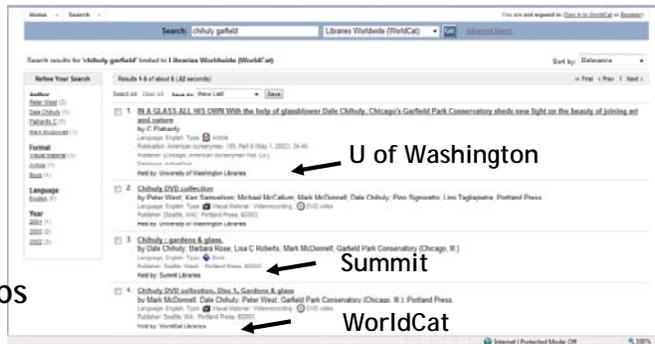


In this case:

Local = my
library in
WorldCat

Group = my
library's groups
in WorldCat

Global = all of
WorldCat



I showed you an example of a search that started globally and ended up at the local level. Here is an example of network level metadata management in the service of a user workflow that starts at the local level and ends up at a group or global level.

I did a search on the U Wa WorldCat Local catalog for the Dale Chihuly glass exhibit at the Garfield Park Conservatory in Chicago. I find holdings at the U W level but also items of interest to me in the regional consortium, Summit, and in WorldCat. If I click on the WorldCat item I am taken again to the screen where I can learn what other collections hold the item.

I should also mention another metadata management feature that is essential for supporting this kind of user easy user navigation. It is the FRBRization of WorldCat manifestations or edition records into work clusters, so the user doesn't need to understand or be hindered by the presence of many editions scattered across the holdings of many libraries.

What's Possible *Today?* (3) Local Interface with Group and Global Reach

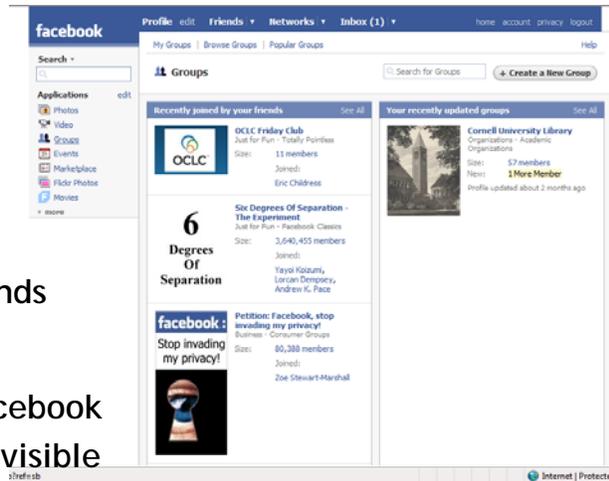


In this case:

Local = my
Facebook page

Group = the
groups my friends
and I belong to

Global = all Facebook
pages that are visible
to me



So I've described the network level metadata management that already exists that is built on library bib and holdings data. Here is another metadata management system that I'll bet you've never considered as such.

Yes, metadata management runs Facebook, just like it runs library sites, but the metadata comes largely from Facebook users. The site organizes that metadata into services for me, for my Facebook friends, for the groups that my friends belong to, and at the global level.

This is a screenshot of all the groups that my Facebook friends and I belong to.

If you would like, we might spend some time in the discussion teasing out the differences in where metadata comes from and how it is managed in Facebook versus library systems. Something to think about.

Choices



1. Continue distributed metadata management
2. Build community-based metadata management at Web scale
 - More effectively share library community expertise
 - Engage users in a social space
3. A blend of both

So where might we, as a community of experts, decide to go from here?

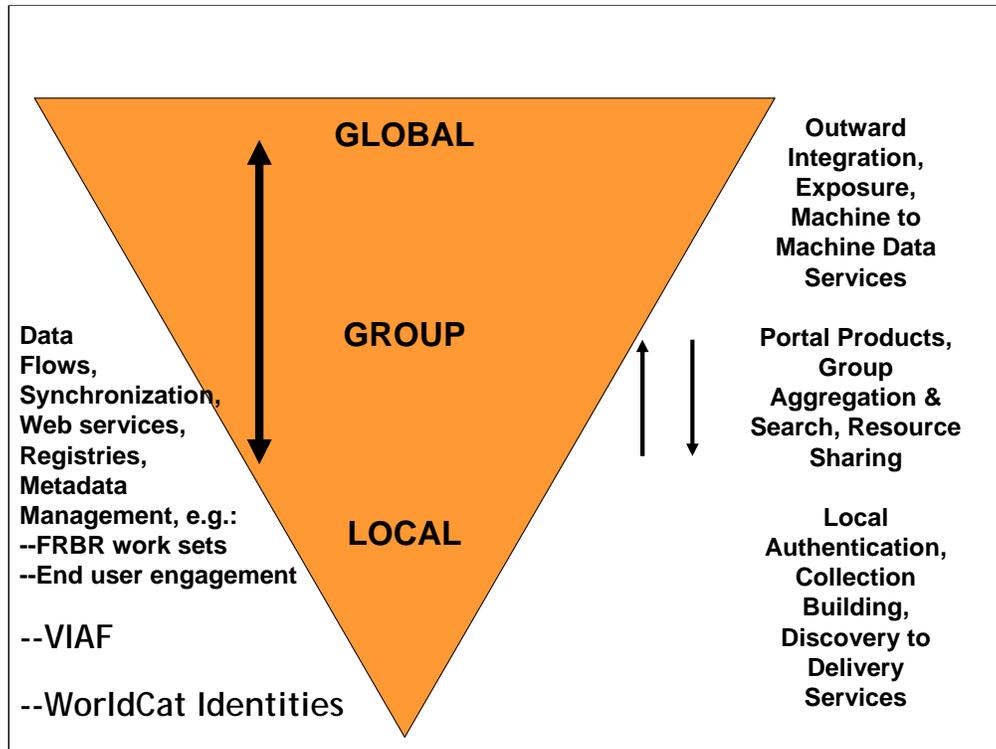
Going back to the point I made at the beginning, right now metadata management (corrections and enrichments to records) is mainly done in silos, within our local systems.

We can already do some nifty things with the metadata we already have at local, group and global levels to expose our collections better on the Web and make it easier and more convenient for users to find and get to them.

I see three choices going forward [read slide]

Under bullet 2, I am introducing a new concept – what if our expert community was to welcome the assistance of end users for community-based metadata management on the network?

I think the third bullet is the most likely scenario because it will take time to build the tools we need to make significant progress changing our current siloed metadata management practices.



This is a concept diagram of how local, group, and global metadata might work together. On the right are the roles that metadata might play at the three levels.

On the left are some tools to support the connections that metadata can make. If we are serious about creating a community-based metadata management system on the network, more tools will be needed.

I'd like to conclude by talking about some new tools that could play a part in a truly global metadata network that plays well on the Web and is centered around the user experience.

Viaf and WorldCat Identities.

Think Globally, Act Locally: VIAF (Virtual International Authority File)



VIAF memorandum of understanding, November 2007

- Bibliothèque nationale de France, Deutsche Nationalbibliothek, Library of Congress

Developed and hosted by OCLC Research

Goal is to include authoritative names in a freely available global service

Virtually combines multiple name authority files

Promotes greater re-use of bibliographic data from libraries serving English-, French-, and German-speaking communities

Talk about slide

In my recent travels to speak with librarians outside the US about their goals for the future, it has come home to me how much the present North American systems, with few exceptions, are built upon an English-language bias. Sure, we have descriptions of lots of non-English titles, but the metadata itself—the notes, the heading forms and so on—show a strong bias for English speaking audiences.

To make the World in WorldCat real, we need to think about what it means to serve communities for whom English is not the first language.

The VIAF is a first step in that direction. One day, it might enable us to display and link the heading forms in the language most helpful for the user.

Same Person, Different Headings in North America and in Germany



Virtual International Authority File Hosted by OCLC Research

Names

2 headings found for *teresa of avila*

	Name	Authority File	Sample Title	Links
1	Teresa, Jde, Jesu†1515-1502	DNB	sämtliche Schriften der hl. Theresia von Jesu	LC
2	Teresa, Tof Avila, Saint †1515-1502	LC	libro de la vida	DNB

Here is an example of a person, Saint Teresa of Avila, whose preferred heading is different in the LC vs DNB authority files.

WorldCat Identities: Getting to WorldCat Identities from WorldCat.org



Home Search You are not signed in (Sign In to WorldCat or Register)

Search for items: au:James Thurber Search Advanced Search

The beast in me, and other animals : a new collection of pieces and drawings about human beings and less alarming creatures
by James Thurber

Type: Book
Language: English
Publisher: Harmondsworth : Penguin, 1976.
OCLC: 59144269
Related Subjects: American wit and humor
Citations: Cite this Item Export to EndNote Export to RefWorks

Get It
Search my library
Buy This Item

Save It
Add This Page to Favorites
Save to: [New List]
Save

Add to It
Review this Item
Edit Public Notes

Share It
Link to this Page
BOOKMARK

Services from OCLC Library for this item: OCLC Partner Access 1.0 Check OCLC Library Connect to the catalog of the OCLC Library Request item Order Item from CISTI OCLC FirstSearch

Libraries **Details** Subjects Reviews

About the Author(s)
James Thurber

Item Details

Buy This Item
Search for
The beast in me, and other animals : a new collection of pieces and drawings about human beings and less alarming creatures

Now WorldCat Identities. It has just gone into production in November. It is a product that was initially developed in the OCLC Office of Research.

The metadata that underpins WorldCat Identities is from both national authority files and the WorldCat bib records that come from all of you.

WCat Identities takes advantage of the millions of hours of effort that the library community has contributed to shared cataloging and presents all that data as a new END USER facing service.

In this example, I've searched for the American humorist James Thurber's title The Beast In Me. I want to learn more about Thurber, so I click his name under the Details tab of the bib record display.

James Thurber's Identity



WorldCat Identities BETA

Search Identities
Name: Go

Jump To: Overview | Publication Timeline | Works About | Works By | Audience Level | Related Names | Useful Links | Fast Headings

Thurber, James 1894-1961

Overview

Works: 593 works in 1,183 publications in 30 languages and 63,651 library holdings

Genres: Humorous fiction Fantasy fiction Humorous stories, American American wit and humor Comedies Humorous stories Photographs

Subject Headings: Humorists, American--20th century Cartoonists--United States Authors, American--20th century

Roles: Bibliophile

Classifications: ps2539194, 017.5

Publication Timeline

1900 2010

Key

- Publications about James Thurber
- Publications by James Thurber
- Publications by James Thurber, published posthumously

Most widely held works about James Thurber

Alternative Names

00'3, 127 D

Languages

- English (1,122)
- German (11)
- No Linguistic content (2)
- French (16)
- Japanese (11)
- Spanish (11)
- Swedish (9)
- Czech (3)
- Dutch (7)
- Hebrew (3)
- Italian (5)
- Chinese (3)
- Russian (4)

And I'm taken to a pretty good proxy for a MySpace page for this author ... who could never be in MySpace because he's been dead for nearly 50 years.

Mining and bringing together metadata from bib and authority data, WCat Identities clusters information about this author's works, genres, subject headings and so on.

A graph of when this author flourished, plus cover art and much more are presented to the user.

Most widely held works ABOUT Thurber are also included...

Thurber (2)



Most widely held works about James Thurber

- **James Thurber** by Robert Eustis Morabber
- **Thurber : a biography** by Burton Bernstein
- **The clocks of Columbus: the literary career of James Thurber** by Charles S Holmes
- **The art of James Thurber** by Richard C Tobias
- **Thurber: a collection of critical essays** by Charles S Holmes

more ▼

Most widely held works by James Thurber

The Thurber carnival by James Thurber

51 editions published between 1931 and 1999 in 5 languages and held by 3,528 libraries worldwide
Includes the complete book "My life and hard times." Also selected cartoons, humorous essays, and stories from Thurber's earlier books.

Many moons by James Thurber

38 editions published between 1943 and 2003 in 8 languages and held by 3,427 libraries worldwide
Though many try, only the court jester is able to fulfil Princess Lenore's wish for the moon.

Lanterns & lances by James Thurber

12 editions published between 1961 and 1992 in 3 languages and held by 2,327 libraries worldwide

Writings and drawings by James Thurber

19 editions published between 1944 and 2002 in 3 languages and held by 2,288 libraries worldwide
Contains the greatest man in the world, interview with a lemming, and other humorous pieces read by Henry Morgan, interspersed with commentary by Thurber on his writing and drawing and with commentary by friends and associates on Thurber and his work.

The years with Ross by James Thurber

10 editions published between 1957 and 2001 in English and held by 2,157 libraries worldwide

My life and hard times by James Thurber

48 editions published between 1933 and 2000 in 3 languages and held by 1,872 libraries worldwide

Selected letters of James Thurber by James Thurber

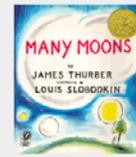
14 editions published between 1953 and 2003 in English and held by 1,789 libraries worldwide

more ▼

Chinese (1)
Russian (1)
Ukrainian (4)
Portuguese (2)
Arabic (2)
Polish (2)
Persian (2)
Greek, Modern (2)
Danish (2)
Norwegian (2)
Telugu (1)
Kiswahili (1)
Korean (1)
Afrikaans (1)

more ▼

Covers



As well as the most widely held works by this author in libraries. These are all FRBRized clusters, by the way.

Thurber (3)



Related Names

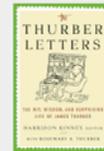
- White, E. B. (Elwyn Brooks) 1899-1985 [+]
- Nugent, Elliot 1899-1980 [+]
- Ustinov, Peter. Prefessor [+]
- Learning in Fobus, Inc. [+]
- Morgan, Henry 1915- Prefessor [+]
- Ross, Harold Wallace 1893-1951 [+]
- Sirovkin, William 1909-1961 [+]
- Faulkner, William 1897-1962 [+]
- Simont, Marc. Illustrator [+]
- Lasker, Ring 1885-1933 [+]

Useful Links

- Library of Congress Authority File (English)
- Deutsche Nationalbibliothek Authority File (German)
- Wikipedia James Thurber

Associated Subjects

American literature [American wit and humor](#) [American wit and humor](#),
Pictorial [audiobooks](#) [Authors](#) [Authors, American](#) [Bibliography](#) [Biography](#) [Caricatures and cartoons](#) [Cartoonists](#) [City and town life](#) [comedy](#) [Comedy films](#) [Comic, The](#) [Crime](#) [Criticism, interpretation, etc.](#) [Detective and mystery stories](#),
American Dogs [Drama](#) [Drama](#) [Drama-Bibliography](#) [Fables, American](#) [Feature films](#) [Fiction](#) [Films](#) [Humor](#) [Humorists, American](#) [Humorous stories](#),



Links to collaborators of biographers, links to authority records, and a tag cloud based on subject headings are also here.

If I click on any of the titles or on the cover art, I'm am taken to ...

Thurber (4)



The screenshot displays the WorldCat Beta interface for the book 'The Thurber carnival' by James Thurber. The page includes a search bar at the top, a navigation menu, and a detailed record for the book. The record lists the following information:

- Title:** The Thurber carnival
- Author:** by James Thurber
- Type:** Book
- Language:** English
- Publisher:** New York : Modern Library, 1994.
- Editions:** 42 Editions
- ISBN:** 0679600892 9780679600893
- OCLC:** 29632006
- Related Subjects:** United States -- Social life and customs -- 20th century -- Humor. Humorous stories, American. Fables, Americas.
- Citations:** Cite this item | Export to EndNote | Export to RefWorks
- Additional info:** Contributor biographical information | Publisher description

Below the record, there are sections for 'Get it' (with links to 'Find in other WorldCat libraries' and 'Buy from Amazon.com'), 'Save It' (with 'Add This Page to Favorites' and a 'Save' button), 'Add to It' (with 'Review this item' and 'Edit Public Notes'), and 'Share It' (with 'Link to this Page' and social media icons). At the bottom, there are tabs for 'Libraries', 'Details', 'Subjects', 'Editions', and 'Reviews', along with a 'Buy This Item' button.

The bib record, from whence I can navigate to other editions and/or to my local library to get the book.

Again, as we look at these systems, I encourage you to look beyond the interface to consider metadata's role not just in describing things or people, but for making connections between things at a variety of levels in service of user discovery to delivery workflows.

I mention WorldCat Identities not just as a product but to provoke our thinking about the potential for community-based metadata management of author names in direct service to end users.

For example, what if we thought about WorldCat Identities pages the way we think about Wikipedia pages?

Another idea: what if we thought about WorldCat Identities in terms of the subsets we can produce and manage for a nation's or discipline's most important writers?

Collaboratively Rethinking Metadata Management



What opportunities exist beyond transferring today's catalog maintenance practices to the network level?

How can we take advantage of the Web as a platform?

How can we share catalogers' expertise and efforts more effectively?

How can we engage end users in metadata management?

How can our approach be global in scope but respectful of linguistic and cultural differences?

What are the next steps? Who should be involved?

I conclude with these questions for us to ponder or discuss later in this session or when we return to our libraries or offices.

The U Wash staff and others have made excellent suggestions for making today's expert-based metadata management systems work better for users and most cost effectively for libraries.

[read first bullet]

Thank You!



Matt Goldner

Executive Director, End User Service

OCLC, Inc.

goldnerm@oclc.org

Karen Calhoun

VP, WorldCat and Metadata Services

OCLC, Inc.

calhounk@oclc.org

My colleagues and I at OCLC look forward to hearing your ideas and suggestions for the role that OCLC might play.

Thank you.