

# The Value Proposition for Semantic Search: A User Panel

# Representation from 6 Communities

- Library – Diane Hillmann
- Data – George Thomas
- Health Informatics – Helga Rippen
- Intelligence – Leslie Mitchell
- Publishing – Joe Hilger
- An Enterprise – Farah Gheriss

# 5 Minute Lightning Talks Using Three Basic Questions

- What is the background or context of your community or project relevant to Semantic Search?
- What are the search needs of the users in your community or project that may be solved by semantic search, focusing on unique or high priority requirements?
- What challenges, problems or issues have you or members of the community encountered as you've tried to identify, evaluate or implement semantic search in this context?

# **Library Community**

**Diane Hillmann**

**Metadata Management Associates  
(MMA)**

# Library Community: Background Information

- The Library Community has worked within a data silo for over a century
  - Starting with LC catalog card distribution, through MARC and collaborative data creation in OCLC, libraries have shared the burden of metadata creation
  - MARC data has been largely unavailable outside library systems
- Early development of MARC in 1960's gave libraries a head start on machine readable data
  - Fifty years later, moving beyond MARC is a struggle: librarians are 'hard-wired' for MARC
  - Recent interest in the Semantic Web and Linked Open Data has broken open the silo some
- Resource Description and Access (RDA) has been in development for almost a decade, based on FRBR
  - The RDA Toolkit, a web service available via license, is the source of the guidance instruction ('cataloging rules') for RDA
  - The RDA Element Sets and Value Vocabularies are available for free at the Open Metadata Registry <http://rdvocab.info> (but still under review)

# Library Community:

## Search Needs Relevant to Semantic Search

- Traditionally, 'search' in libraries has been search of metadata records, not of resource content
  - Current ILS's based on MARC have not made good use of the extensive shared vocabularies in library data (topical, geographic, genre, name disambiguation)
  - Increasing understanding that library data needs to escape catalogs to be useful in the modern world
- Reliance on text strings in library data has made data transition scenarios very complicated
  - Data maintenance in current library systems still a very human, manual process
  - New understanding of needs for strong identifier systems has begun to result in some improvements for shared vocabularies
- Expansion of search into digital data has been slowed by legal issues around copyright & fair use (Google vs. Authors Guild vs. Hathi Trust)
  - Much of the innovation so far has been around image data, relying still on metadata
  - Rise of eBooks and the distribution troubles with libraries has created more friction

# Library Community: Challenges

- Innovation in libraries has been scattered, unclear where leadership will come from
  - LC and OCLC seem to be traveling paths with unclear implications for the community
  - Both seem to be eschewing FRBR (and thus RDA)
- FRBR supports browse explicitly via extensive relationships
  - Promises better search and browse by moving vocabularies to the web and providing access to legacy data and prospective data using those vocabularies
    - For LOD-LAM information, different metadata standards and practices the biggest impediment to better collaboration
- Search of digital content of interest outside of Google and other commercial spaces will depend on how effectively legacy data is mapped and repurposed
  - Ongoing conversations between libraries, archives and museums show great promise
  - Lack of resources for innovation and training of librarians in data management a major problem

# Key to Acronyms

- MARC: Formally MARC 21, stands for Machine Readable Cataloging
- FRBR: Functional Requirements for Bibliographic Records
- ILS: Integrated Library System
- LC: Library of Congress
- LOD-LAM: Linked Open Data-Libraries, Archives, Museums



# **The Data Community**

George Thomas

Department of Health & Human Services

Office of Chief Information Officer

# Data Community: Background Information

- open gov initiative, site
  - [hhs.gov/open](https://hhs.gov/open), [healthdata.gov](https://healthdata.gov)
- search
  - content and data, both currently using Solr
- linked data
  - hospitals, (doctors, drugs, genes, beneficiaries, claims)
  - ...

# Data Community: Search Needs Relevant to Semantic Search

- super rich snippets
  - [sindice](#), schema.org ++
- config Solr core
  - {keyword1} = foo:prop / bar:path :: datatype^^en
- open source
  - [linked media framework](#)

# Data Community: Challenges

- index multiple syntaxes?
  - html+rdfa, turtle
- security/privacy
  - data element secrecy, service requestor anonymity
- authn/authz
  - WebID + PPO (see SemTech [session](#) and [dev challenge](#))

# Health Informatics

**Helga Rippen**

**Westat**

# Health Informatics: Background Information

- Health informatics is broad (medical devices, electronic health records, bio-genomics, to consumer health tools)
- The optimal design and implementation of health information technology is unknown... it is a complex field and in its infancy.
- Health Informatics is looked upon as a tool to deliver cost-effective, quality, safe, efficient, health care.
- 30-70% of Electronic Health Record implementations fail.
- No standard implementation methodology, terms or measures related to success

Community: [www.healthITxChange.org](http://www.healthITxChange.org)

Reference: Rippen, et al Organizational Framework for Health Information Technology, Int J Med Inform, 2012  
[http://www.ijmijournal.com/article/S1386-5056\(12\)00031-7/abstract](http://www.ijmijournal.com/article/S1386-5056(12)00031-7/abstract)

# Health Informatics:

## Search Needs Relevant to Semantic Search

- How can one leverage data on the Internet (blogs and other social media, publications, website content, etc) to mitigate risk of implementation failure?
- Context is important (physician office versus hospital; resources; specialty; vendor).
- Can we learn what to do (or not to do) to successfully implement an Electronic Health Record?
- How do we transform information into knowledge?

# Health Informatics: Challenges

- Taxonomy/terms are not fully defined within the space.
- Bias may exist depending upon source (e.g., vendor) and expertise (use different terms)
- Difficult to pull out trends as there is a high “signal to noise”



# **The Intelligence Community**

**Leslie Mitchell**

**Information International Assoc.**

# Intelligence Community: Background Information

- **Multi-Agency/Inter-Agency**
  - 17 Intelligence Agencies
- **Multi-disciplined**
  - OSINT, SIGINT, MASINT, GEOINT, HUMINT, **INTER**  
INTELLIGENCE
- **Security**
  - Varied classifications of data and information



# Intelligence Community:

## Search Needs Relevant to Semantic Search

- **Volume**
  - Terabytes (1,000 gigabytes) & petabytes (1,000 terabytes) of information
  - More information than is humanly possible to analyze
- **Velocity**
  - Time sensitive data
  - Growing exponentially by the second
- **Variety**
  - Structured and unstructured data
  - Unclassified and Classified



# Intelligence Community: Challenges

- Lack of flexible Ontologies
- Restricted or limited access to data
- Training



# Acronyms:

- **Open Source Intelligence - OSINT**
- **Signals Intelligence - SIGINT**
- **Measures and Signals Intelligence - MASINT**
- **Geospatial Intelligence - GEOINT**
- **Human Intelligence - HUMINT**
- **Technical Intelligence - TECHINT**
- **COUNTER INTELLIGENCE**

# **The Publishing Community**

**Joe Hilger**

**Avalon Consulting**

# Publishing Community: Background Information

- Avalon Consulting, LLC has specialized in search since 2005. We have worked on search projects in a wide range of industries.
  - Public Sector
  - Financial Services
  - Publishing
  - Healthcare
  - Retail
- Avalon has recently worked with a wide range of publishers including:
  - Wolters Kluwer
  - McGraw-Hill
  - Simon Schuster
  - Lexis-Nexis and others
- The publishing projects we worked on span a wide range of uses.
  - Digital Products
  - Digital Asset Management Solutions
  - Large On-line Manuals

# Publishing Community: Search Needs Relevant to Semantic Search

- Books and Publications have hierarchical relationships that are best represented through search (long form search.)
  - Books and publications require sectionalized search.
- Publishers need new ways to monetize their content.
  - Search enables content re-use and paid features like custom publishing
- Publishers have vast amounts of content (many with more than 100 years of historical content.)
  - Faceted search and related content allow people to discover information in these vast repositories.



# Publishing Community: Challenges

- Metadata is critical for content re-use and faceted search. How does an organization manage metadata across a large content set with limited resources?
- Search needs to support content at the document and section level. How can a single search display documents and sections of documents?
- Related content increases engagement in on-line products. How can content relations be managed for a large dataset by organizations with limited resources?

# **For an Enterprise**

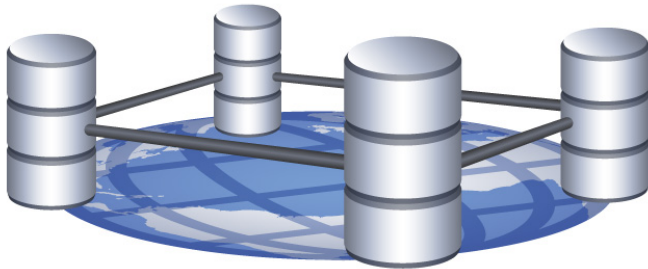
**Farah Gheriss**

**International Monetary Fund**

# Semantic search Challenge



## DATABASES



# Semantic search

- Challenge
  - Information is scattered in multiples repositories
  - Each repository has its own metadata scheme
  - Each repository uses different ways to describe the information stored within.
  - Organizations need to access information **quickly** regardless where the information is stored.

A lack of semantic search could yield to



# Semantic search Enterprise

- Auto-suggestion
- Type ahead
- Best bets
- Faceted navigation
- Related searches
- Social search
- Contextualization
- Expertise location

Wednesday, December 5, 2012



Staff Put New Laptops to the Test

Fund Adopts Institutional View on Capital Flows

Top 10 Most Visited Pages in November

Staff invited to IMFFA International Dress Showcase

Fund Adopts Institutional View on Capital Flows

Flexible and balanced view on capital flows to help give countries clear policy advice. [\[more\]](#)

MORE TOP STORIES RSS

Go to a Country

Select a Country

Go to a Department or Office

Select a Department or Office



Management Corner

Email any member of the Management Team

ManagementCorner@imf.org



Communications Toolkit

From Human Resources

JobLink/TalentLink

- New MBP Cards Coming in December
- 2012 Financial Certification
- VSP – Employee Briefings
- Leave Self Service
- Staff Survey – Read All About it

Ethics and Integrity

Go to HR Web

Noteworthy

- New F&D Focuses on Philanthropy Sowing Change
- HQ1 Renewal: Latest Designs Revealed
- Papers Underline Critical Role of Management of Natural Resources
- Book Launched on Oil Wealth in Central Africa

More Stories

See also the latest: **IMF digest Staff news**

Office of the Managing Director

- MD Speech: Asia and the Promise of Economic Cooperation
- FDMD Speech: Enabling Economic Transformation in the Middle East and North Africa
- DMD Shinohara Speech at Tax and Growth Seminar in

Notices

- Fundwide Toy Collection Drive for Children's National Medical Center- December 3-7
- Interior Window Washing
- Announcing Workplace Climate Survey on LGBT Issues
- HQ2 HVAC Maintenance Work

More Notices

RSS

- Daily Global Security News





All Sources

Advanced

Search:  Title Only  Exact Phrase  URL Only

Refine By

Reset All Refiners

Content Area

Any Content Area

Publications (11866)

HR Web (394)

News (390)

Blogs (75)

Services (34)

Department

Any Department

SPR (3316)

TGS (3314)

MCM (1366)

1-20 of 51,339 results

Sort by: Relevance



Suggest a Best Bet!

Best Bets

- ★ Annual Leave
- ★ Leave Self Service
- ★ My Leave Information

E-mail results!

My Leave Information

Provides access to online leave self-service tools

Date: 11/30/2012 Size: 93KB

<http://www-intranet.imf.org/departments/Hum ... yHR/myPersonalInformation/MyLeaveInformation>

E-mail it! ★ Suggest as a Best Bet!

Related Searches

- Home Leave
- Leave Without Pay
- Annual Leave
- Compensatory Leave
- Leave Self Service

Continue search in...

- telis catalog
- eLibrary
- Google
- Google Scholar

Send Feedback or Question



**Countries**

All

**Key Output Areas**

- All
- Bilateral Surveillance
- Facilities Specific to Low Income Countries
- General Research
- Generally Available Facilities
- Multilateral Surveillance
- Regional Surveillance
- Standards and Codes and Financial Sector Assessment
- Support and Governance
- Technical Assistance

**Topics**

- All
- External Sector
- Fiscal Sector
- Monetary and Financial Sector
- Real Sector
- Social and Demographic Sector

[Please provide feedback](#)

**Departments**

All

**Search KE Documents**

**Recent KE Documents**

Results by Modified Date | View by Relevance | [Alert Me](#) | [RSS](#)

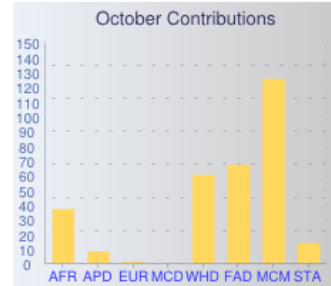
- Chair - Staff Report for the 2012 Article IV Consultation**  
Source: Institutional Repository | Department: African Department | Date: 12/4/2012 | Series ID: SM/12/292  
[Show Keywords](#)
- BTO\_MWT\_Price Statistics Course at CEF\_Graf\_November 18 - 29, 2012.docx**  
Source: EDMS | Department: STARE | Date: 12/4/2012 | Document #: 5040117
- Bangladesh - 2012 - Briefing Paper for Ms. Rebecca Millar - January 7-16.docx**  
Source: EDMS | Department: FADR1 | Date: 12/4/2012 | Document #: 5040359
- Moldova 2012 BRF LTX Sokoler (Second Mission) Forecasting and Policy Analysis TA JC.docx**  
Source: EDMS | Department: MCMTA | Date: 12/4/2012 | Document #: 5040322
- The Acting Chair's Summing Up - Albania 2012 Article IV Consultation - Executive Board Meeting 12/111 - December 3, 2012**  
Source: Institutional Repository | Department: Secretary's Department | Date: 12/4/2012 | Series ID: SUR/12/103  
[Show Keywords](#)
- BRP\_ERETES Steering Committee in Luxembourg (December 7, 2012) and Meeting with the INSEE-France (December 5-6, 2012).docx**  
Source: EDMS | Department: STARE | Date: 12/4/2012 | Document #: 5040161
- IMF Executive Board Concludes 2012 Article IV Consultation with Nepal**  
Source: Institutional Repository | Department: External Relations Department | Date: 12/3/2012 | Series ID: PIN/12/138  
[Show Keywords](#)
- Kosovo TEC Mission RPT on Regulations Under the Central Bank Law (2010) - PDF 2012 11 CO**  
Source: EDMS | Department: MCMCO | Date: 12/3/2012 | Document #: 5038917
- Mauritania - 2013 - Briefing Paper on Customs Administration - January 13-24.docx**  
Source: EDMS | Department: FADR1 | Date: 12/3/2012 | Document #: 5039453
- Belgium: Back-to-office Report on the European Commission Workshop on Fiscal Decentralization and Budgetary Discipline (November 27, 2012).msg**  
Source: EDMS | Department: FADTP | Date: 12/3/2012 | Document #: 5038841

**About KE Documents**



KE Documents is your gateway to the core country documents of the Fund. \*KE Document responds directly to the need for better information sharing. Memo from IMF Managing Director, June 24, 2009.....More

**Departments Monthly Contributions to**



[More KE](#)

**Help**

- [Tips and Tricks](#)
- [Frequently Asked Questions](#)
- [How to Get Started](#)

**Related Resources**

- [Crisis Group](#)
- [Toolkit for Crisis Management](#)
- [Financial Architecture](#)