Preparing the Way: Creating Future Compatible Cataloguing Data in a Transitional Environment

Dean Seeman & Lisa Goddard Memorial University of Newfoundland

Faster, Smarter, Richer Conference Rome, Italy February 27th, 2014





Aspects of our Linked Data Future

Decentralization

Collaboration

Localization

Richness

Structure

1. Decentralization





Indexes



eJournals



Research Repositories



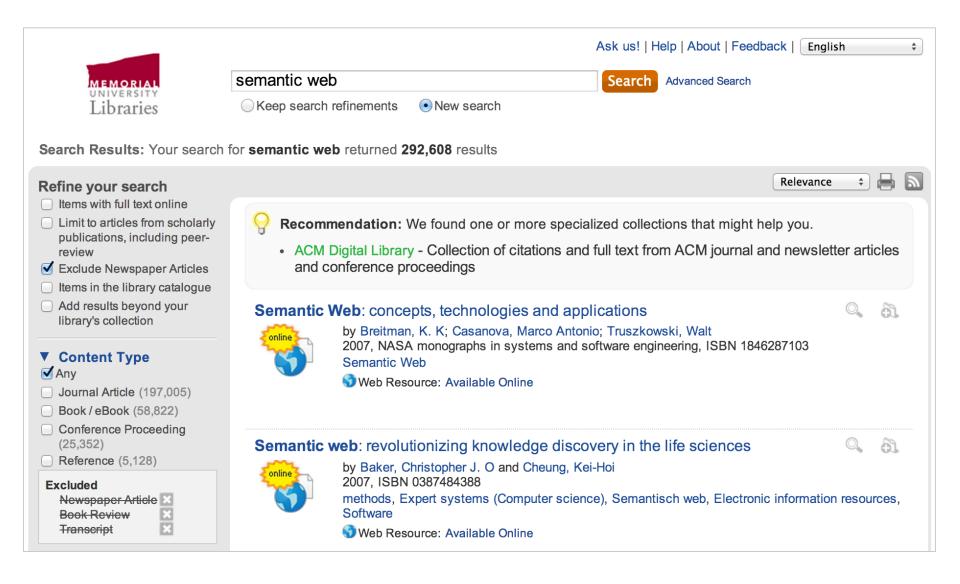
Digital Archives

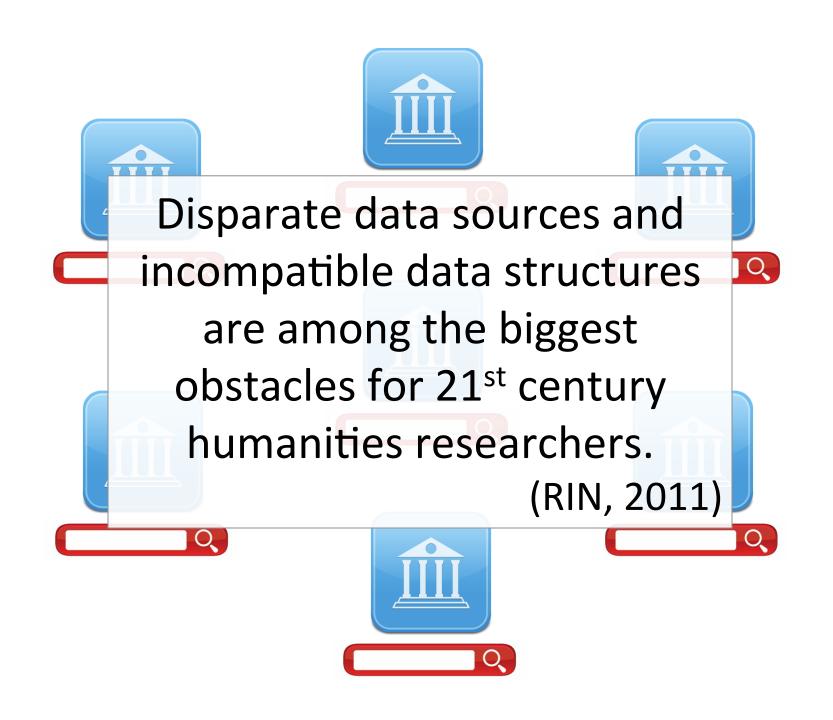


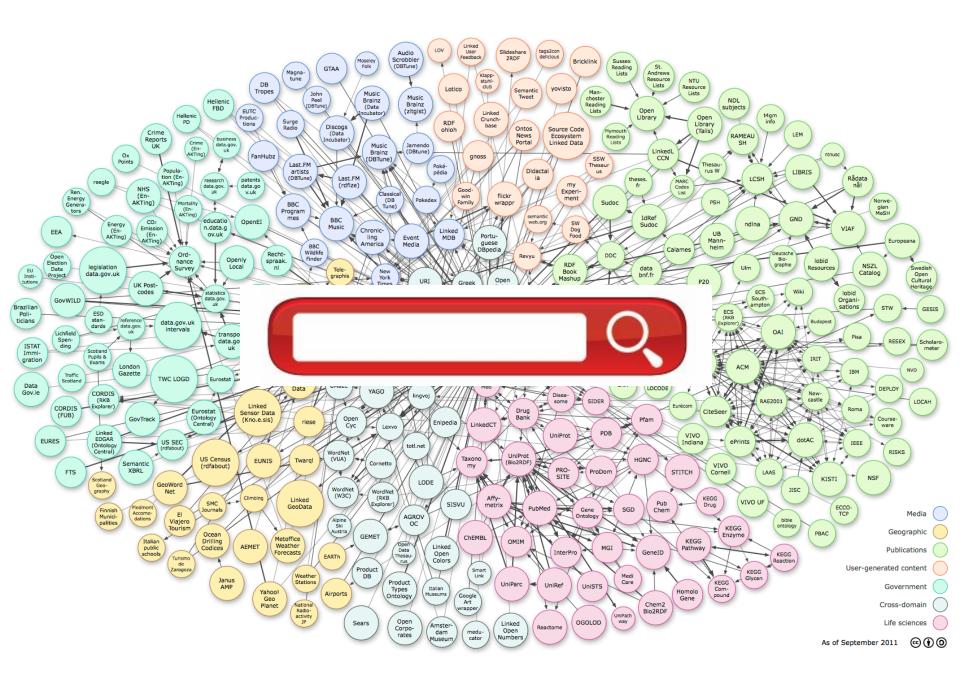
Data Sets



Extract, Transform, Load

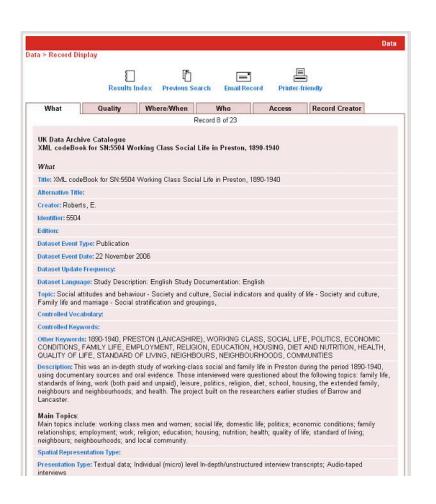






Decentralization

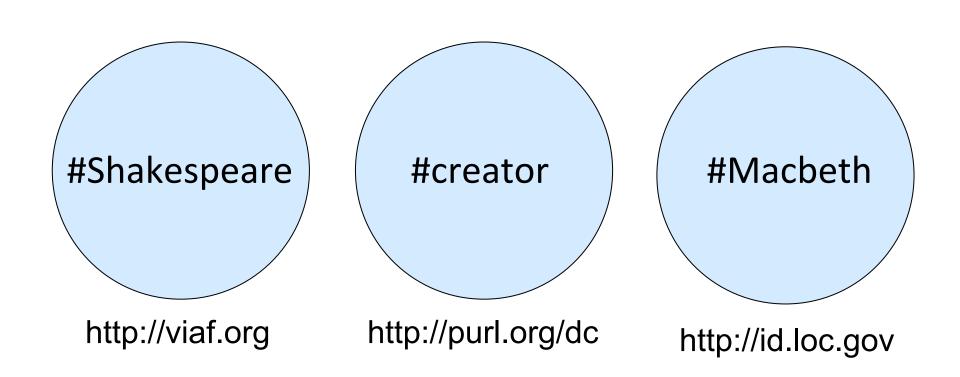
Statements not records.



Subject -> Predicate -> Object

Decentralization

Most data stored remotely.



2. Collaboration









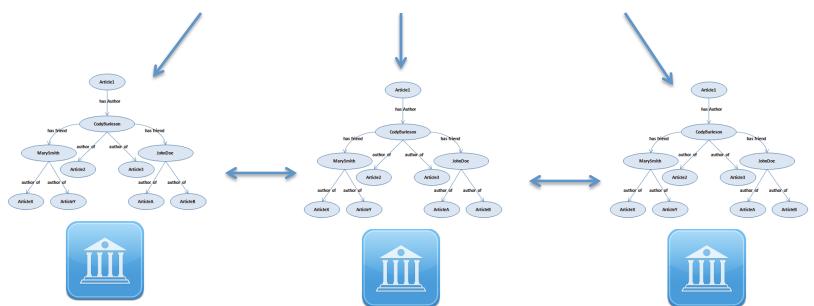




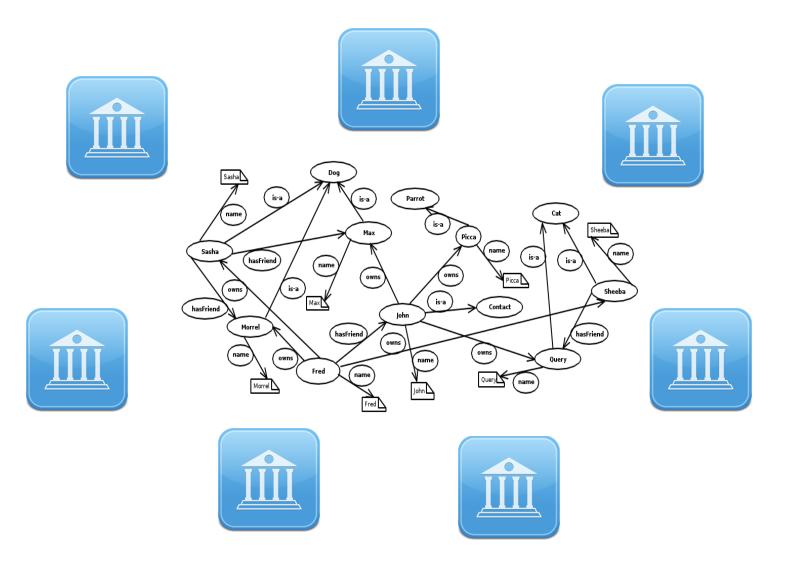


Collaboration

```
26132811
008
        920608s1992
                                     001 0 eng
                       nyu
010
        $a92021087
020
        $a1555701205
040
        $aDLC$cDLC$dAGL
049
        $aHNKA
050 00 $aZ678.9$b.C58 1992
082 00 $a025/.00285$220
100 1 $aCohn, John M
245 10 $aPlanning for automation :$ba how-to-do it manual for
        librarians /$cJohn M. Cohn, Ann L. Kelsey, Keith Michael
        Fiels
260
        $aNew York :$bNeal Schuman Publishers,$cc1992
300
        $avi, 116 p. ;$c28 cm
440
     0 $aHow-to-do it manuals for libraries :$uno. 25
        $aIncludes bibliographical references and index
504
     0 $aLibraries$xAutomation$xManagement$xHandbooks, manuals, etc
     0 $aLibrary planning$xHandbooks, manuals, etc
    10 $aKelsey, Ann L
     10 $aFiels, Keith Michael
```



Collaboration



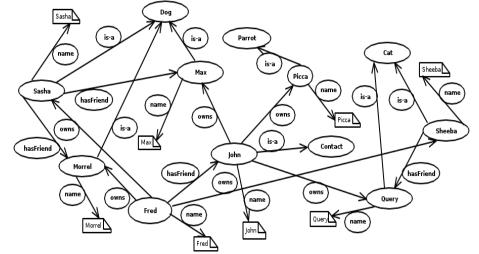
Collaboration













The New Hork Times



Enhancing Shared Records

NACO









Ontology Development

Library Related Ontologies

RDA Ontology

Dublin Core Ontology

Bibliographic Ontology (BIBO)

Citation (CITO)

Provenance Ontology (PROV-O)

MODS/MADS

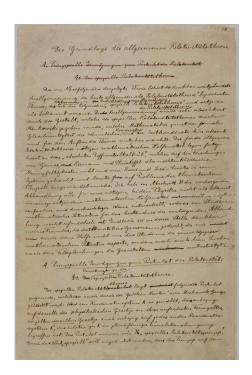
FRBR

Holding Ontology

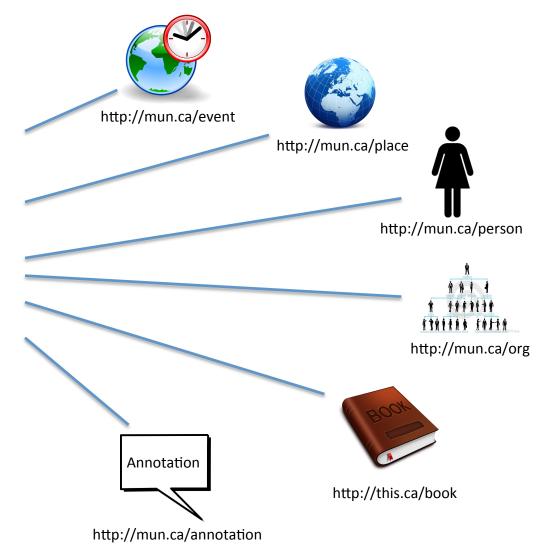
3. Localization

Freday Oct. 30. 1936 "[I]t is their accumulated special collections that increasingly define the uniqueness and character of individual research libraries." things that have happened in my life which are a bit out of the

Expose Entity URIs



http://mun.ca/doc123



4. Richness

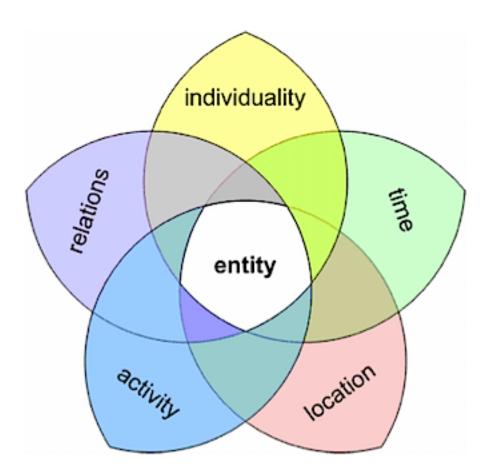
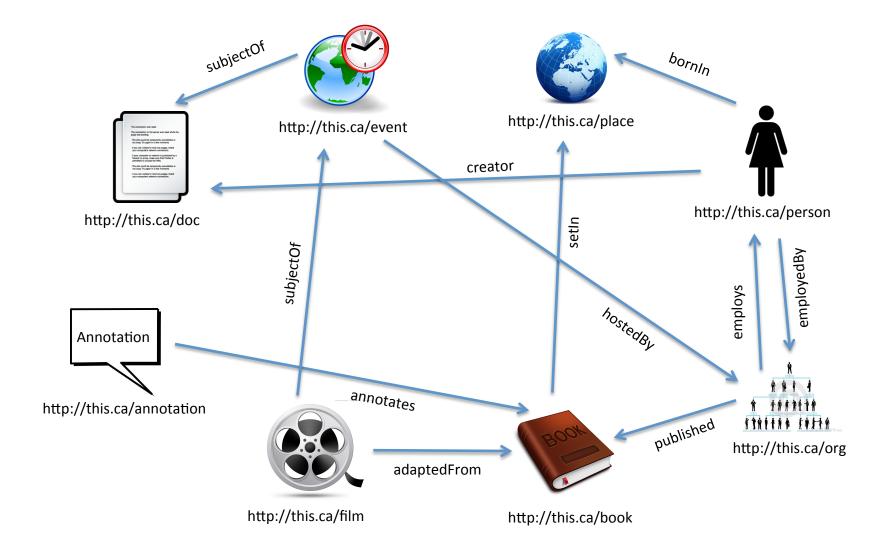


Fig. 1. Five Fundamental Categories for Context Information

Define Relationships



5. Structure

Collaboration Decentralization Richness Localization Structure

Structured Data

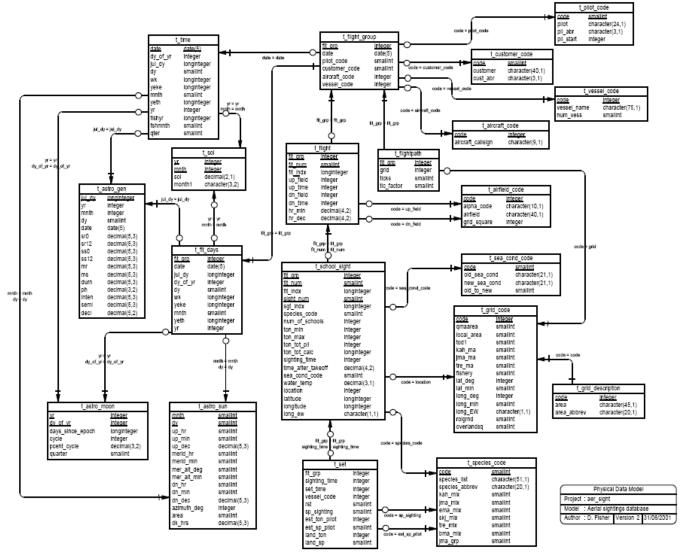
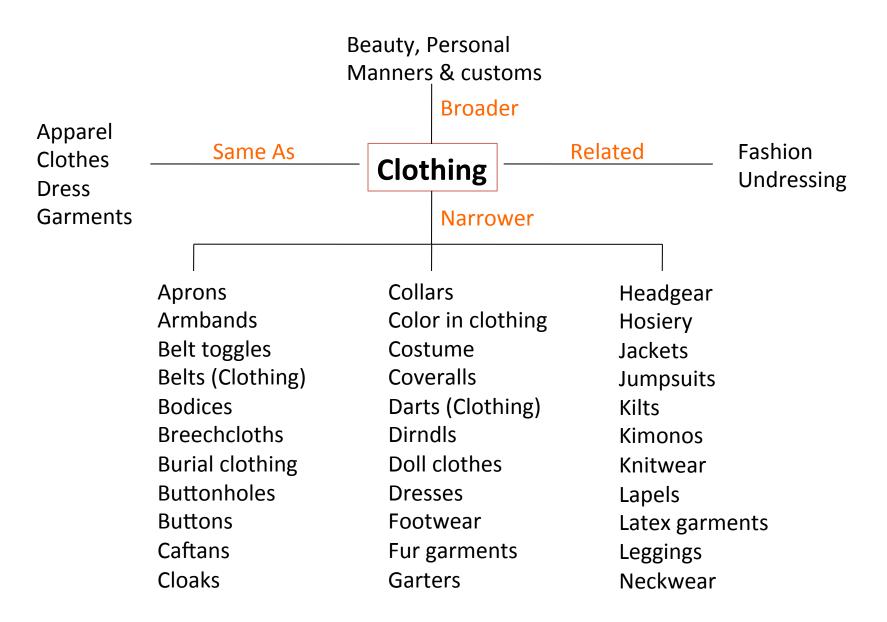


Figure 1: Entity Relationship Diagram (ERD) for the aer_sight database

Semantic Structure



Machine-Actionable Data

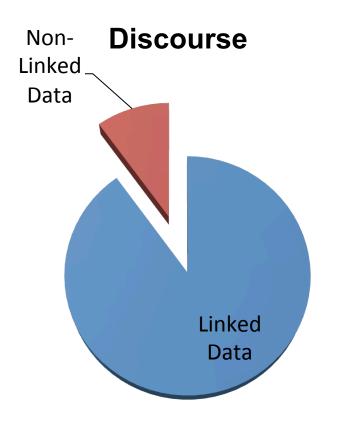


mpedson Michael Peter Edson

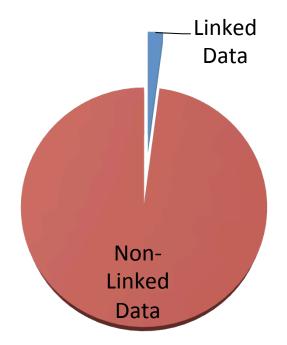
"Our top two users are computers." Says @udcmrk about the importance of API's and machine readable data.

24 Jan

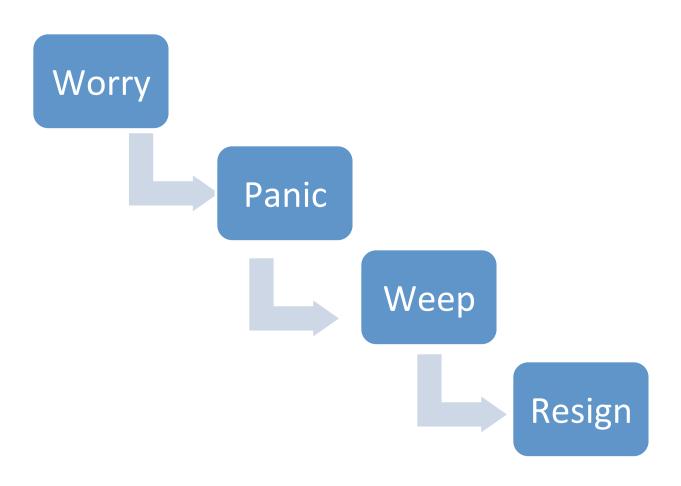




Cataloguing Practice



4 Cataloguer Tasks in Relation to Linked Data



Trust Standard Development?





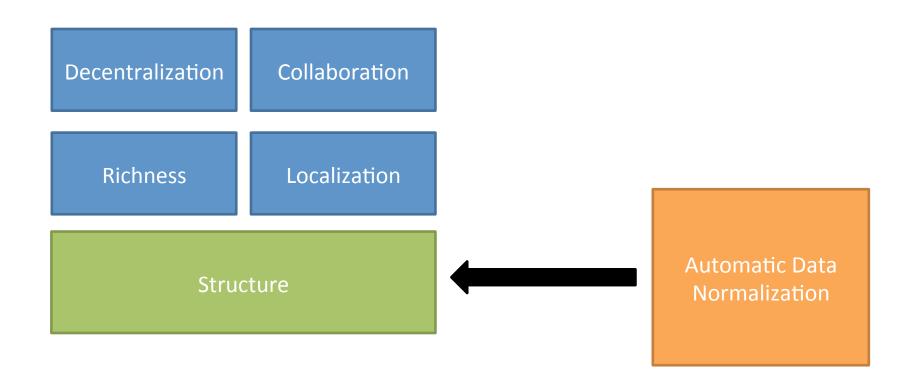


Collaboration Decentralization Richness Localization Structure

The greatest consumer of our data is going to be the machine.

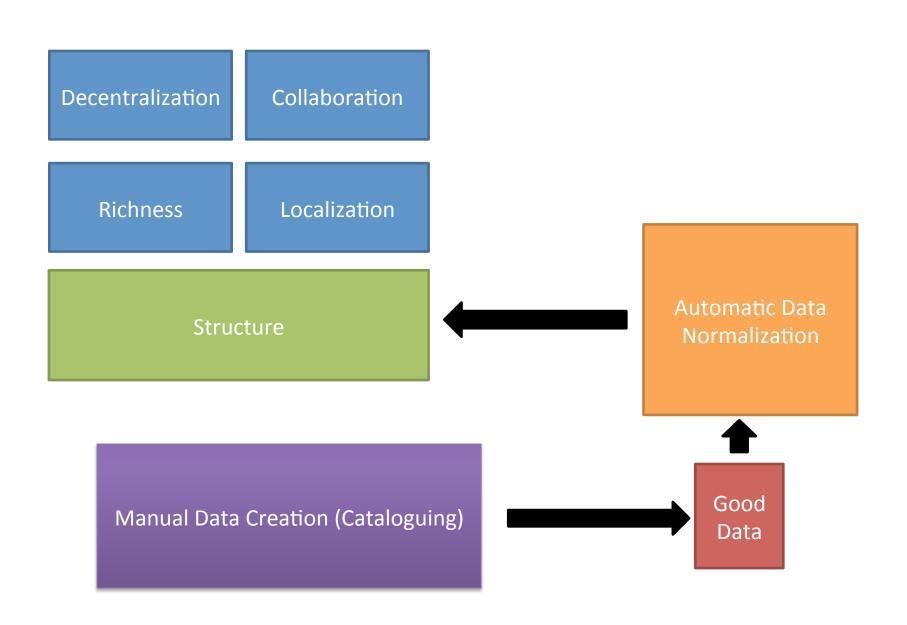
We have to make our data machine understandable.



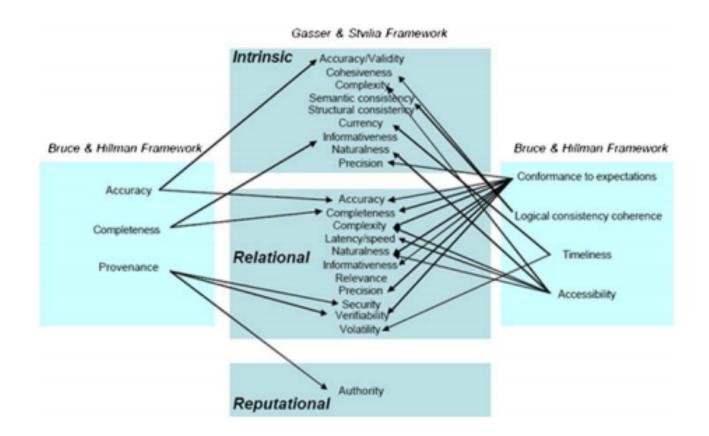


"In computer terms, we have a data normalization problem."

Ross Singer



What is Good Data?



Ochoa, X., & Duval, E. (2009). Automatic Evaluation of Metadata Quality in Digital Repositories, 10(2), 67–91. doi: 10.1007/s00799-009-0054-4

A Few Markers of Good Data for Data Normalization

Discrete

Each element asserts a single thing

Semantically Unambiguous

Data should be clear in its meaning and minimize multiple interpretations

Consistent

Predictable values

This kind of data ...

Helps us in our current environment

Helps the process of data normalization

Helps the future ... even if it isn't Linked Data



Looking at the future ...

... what can cataloguers practically do to plug into it?

O C U S

Focus On Authorities & Controlled Access Points

Authorities

Contain mostly differentiated values

```
    ➡ Hilton, John, 1804-1878
    ➡ Hilton, John Anthony, 1941-
    ➡ Hilton, John Buxton
    ➡ Hilton, John (John L.)
    ➡ Hilton, John W. (John William), 1943-
```

Better for machine processing

Authorities



Controlled Access Points (MARC 1xx, 6xx, 7xx)

=100 1\\$aStoppard, Tom,\$eauthor.

Automatically Normalized / Translated into URIs

Tom Stoppard

Author of Work

Parade's End

http://viaf.org/viaf/ 101362857/rdf.xml http://rdvocab.info/roles/authorWork

http://www.worldcat.org/oclc/827974267

Controlled Access Points

```
F035 \\$a(SIrSI) a4 120534
=040 \\$aCaNfSM$beng$cCaNfSM$erda
                                                                                 Better to have this
=041 \\$geng
=043 \\$an-cn-nf
                                                                        compacted in one statement
=098 \\$aL.L
=100 1\$aKavanagh, Ed,$d1954-$einstrumentalist,$eproducer.
=245_10$aOn_strings of light :$bChristmas melodies performed on Celtic harp /$cEd Kavanagh.
=264 \1$a[Mount Pearl, Nfld.] :$bEd Kavanagh,$c2009.
=264 \4$cphonogram copyright 2009.
=300 \\$a1 audio disc :$bdigital ;$c12 cm
=336 \\$aperformed music$2rdacontent
=337 \\$aaudio$2rdamedia
=338 \\$aaudio disc$2rdacarrier
=500 \\$aProgram notes printed on container.
=500 \\$aCompact disc.
=505 0\$aO little town of Bethlehem (2:34) -- What child is this? (4:40) -- The cherry tree carol (1:47) -- I wonder as I wander (3:57) -- Angels we have
heard on high ; Il est n{acute}e ; Venite adoremus dominum (3:42) -- Bring a torch, Jeannette, Isabella (2:35) -- Snow angels / Ed Kavanagh (2:54) -- The
Huron carol ('Twas in the moon) (2:59) -- God rest you merry gentlemen (1:30) -- In the bleak midwinter (2:36) -- Entre le boeuf (3:07) -- It came upon the
midnight clear (2:34) -- O come, o come Emmanuel (2:21) -- Silent night (2:59).
=508 \\$aProduced by Ed Kavanagh
511 \\$aEd Kavanagh, harp 🔀eather Wright, violin, viola ; Rick Hollett, piano accordion, flute ; Christina Smith, cello ; Gerry Strong, tin whistle, wooden
lute, Dave Panting, bass.
=518 \\$aRecorded at MMaP Research Centre (Memorial University) and Recordtime Productions, $t. John's, NL.
=650 \0$aCeltic harp music.
=650 \0$aChristmas music$zNewfoundland and Labrador.
=651 \0$aNewfoundland and Labrador$vSongs and music.
                                                                              As opposed to spread
=700 1\$aWright, Heather,$einstrumentalist.
=700 1\$aHollett, Rick,$einstrumentalist.
                                                                             throughout the record
=700 1\$aSmith, Christina,$einstrumentalist.
=700 1\$aStrong, Gerry,$einstrumentalist.
=700 1\$aPanting, Dave,$einstrumentalist.
=740 02$aO little town of Bethlehem.
=740 02$aWhat child is this?
=740 42$aThe cherry tree carol.
=740 02$al wonder as I wander.
```

=740 02\$aAngels we have heard on high

=100 1\\$aStoppard, Tom,\$eauthor.

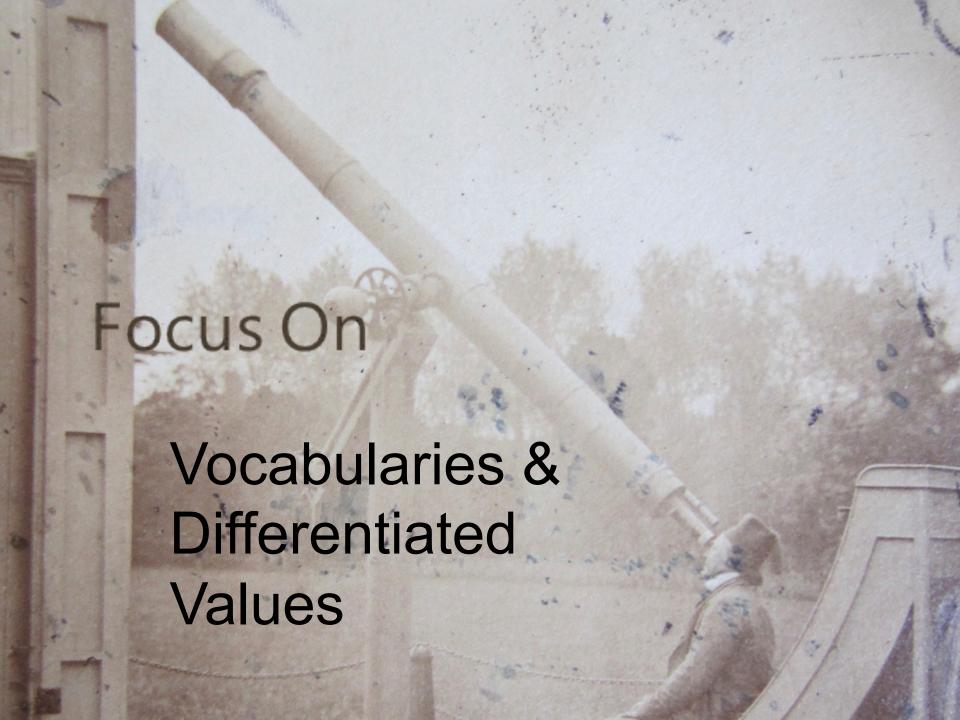
AUTHORIZED

=100 1\\$aStopard, Tom,\$eauthor.

UNAUTHORIZED



But in our Current
Cataloguing
Environment, It
May Be the Best
We Can Do



Vocabularies

Provide Consistent Values for Normalization

3.2.1.3 Recording Media Type

Record the media type using one or more of the terms listed in table 3.1.

Alternative

If the resource being described consists of more than one media type, record only:

a) the media type that applies to the predominant part of the resource (if there is a predominant part)

or

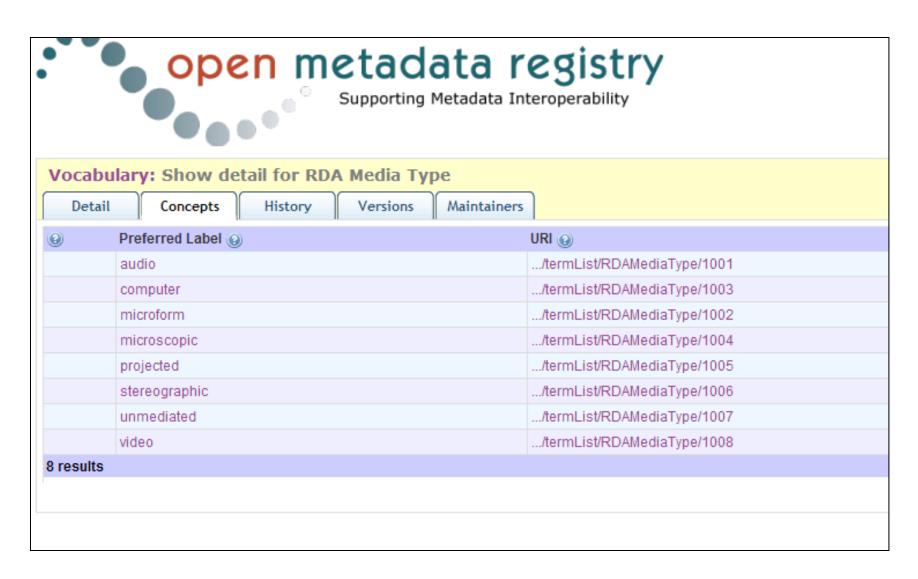
b) the media types that apply to the most substantial parts of the resource (including the predominant part, if there is one).

Use one or more of the terms listed in table 3.1, as appropriate.

TABLE 3.1

audio	Media used to store recorded sound, designed for use with a playback device such as a turntable, audiocassette player, CD player, or MP3 player. Includes media used to store digitally encoded as well as analog sound.
computer	Media used to store electronic files, designed for use with a computer. Includes media that are accessed remotely through file servers as well as direct-access media such as computer tapes and discs.
microform	Media used to store reduced-size images not readable to the human eye, designed for use with a device such as a microfilm or microfiche reader. Includes both transparent and opaque micrographic media.
microscopic	Media used to store minute objects, designed for use with a device such as a microscope to reveal details invisible to the naked eye.
projected	Media used to store moving or still images, designed for use with a projection device such as a motion picture film projector, slide projector, or overhead projector. Includes media designed to project both two-dimensional and three-dimensional images.
stereographic	Media used to store pairs of still images, designed for use with a device such as a stereoscope or stereograph viewer to give the effect of three dimensions.
unmediated	Media used to store content designed to be perceived directly through one or more of the human senses without the aid of an intermediating device. Includes media containing visual and/or tactile content produced using processes such as printing, engraving, lithography, etc., embossing, texturing, etc., or by means of handwriting, drawing, painting, etc. Also includes media used to convey three-dimensional forms such as sculptures, models, etc.
video	Media used to store moving or still images, designed for use with a playback device such as a videocassette player or DVD player. Includes media used to store digitally encoded as well as analog images.

Already Equipped with URIs



Differentiable Values

Example: Exercise the option at RDA 2.8.2.3 for place of publication

Supply the name of the larger jurisdiction (state, province, etc., and/or country) as particles information was taken from a source outside the resource itself (see 2.2.4 RDA).

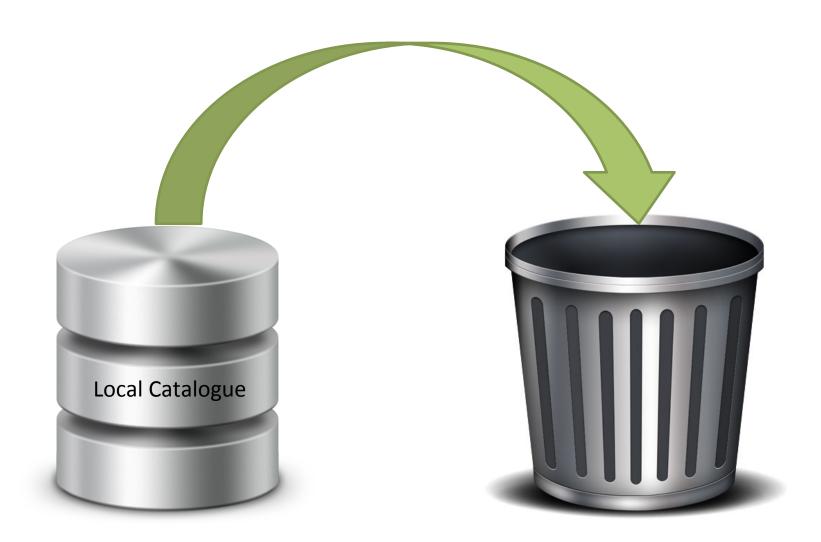
EXAMPLE

Dublin [Ireland]

Dublin [Ohio]

... make the implicit explicit





Future Value

Unique Assertions Ubiquitous Assertions

Local & Unique Resources **Local Authorities** Localization Local Aspects of **Smart Fields** Ubiquitous and Values Authorities Controlled access points Vocabularies Differentiable values Richness

Some practices DO NOT prepare us for the future.

Trapping Data in Free Text Fields Spread Throughout the Record

=035 \\\$a(Sirsi) a4 120534

```
=040 \\$aCaNfSM$beng$cCaNfSM$erda
=041 \\$geng
=043 \\$an-cn-nf
=098 \\$aL.L
=100 1\$aKavanagh, Ed,$d1954-$einstrumentalist,$eproducer.
-245 10$aOn strings of light :$bChristmas melodies performed on Celtic harp /$cEd Kavanagh.
=264 \1$a[Mount Pearl, Nfld.]:$bEd Kavanagh,$c2009.
=264 \4$cphonogram copyright 2009.
=300 \\$a1 audio disc :$bdigital ;$c12 cm
=336 \\$aperformed music$2rdacontent
=337 \\$aaudio$2rdamedia
=338 \\\$aaudio disc\$2rdacarrier
=500 \\\$aProgram notes printed on container.
=500 \\$aCompact disc.
505 0\$aO little town of Bethlehem (2:34) -- What child is this? (4:40) -- The cherry tree carol (1:47) -- I wonder as I wander (3:57) -- Angels we h
heard on high; Il est n{acute}e; Venite adoremus dominum (3:42) -- Bring a torch, Jeannette, Isabella (2:35) -- Snow angels / Ed Kavanagh (2:54)
Huron carol ('Twas in the moon) (2:59) -- God rest you merry gentlemen (1:30) -- In the bleak midwinter (2:36) -- Entre le poeut (3:07) -- It came up
midnight clear (2:34) -- O come, o come Emmanuel (2:21) -- Silent night (2:59).
=508 \\$aProduced by Ed Kavanagh.
511 \\$aEd Kavanagh, harp: Heather Wright, violin, viola ; Rick Hollett, piano accordion, flute ; Christina Smith, cello ; Gerry Strong, tin whistle, w
iute, Dave Panting, bass.
```

=518 \\\$aRecorded at MMaP Research Centre (Memorial University) and Recordtime Productions, St. John's, NL.

Addicted to Keystrokes



Addicted to Keystrokes

Principle of Common Usage or Practice (RDA 0.4.3.7)

"illustrations" instead of "ill."

"pages" instead of "p."

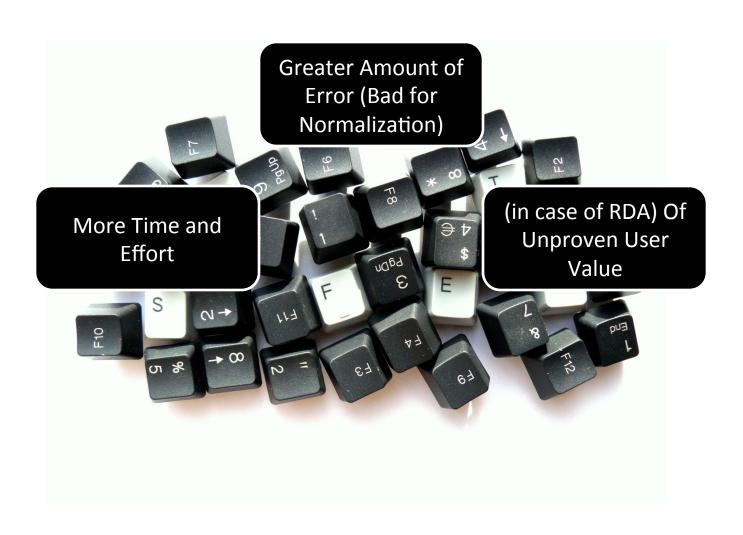
"publisher not identified" instead of "s.n."

Addicted to Keystrokes

Principle of Representation (RDA 0.4.3.4)

Increased transcription of text

Problems with Keystrokes



"I think a good rule of thumb is that if you're having to type data, and you haven't been transported back in time, then you're doing it wrong."

Richard Cotton, Bad Data Handbook

Punctuation

=100 1\\$aStevens, James,\$d1892-1971,\$eauthor.

We Need To Focus on Values, Not Display

Punctuation

LC-PCC PS 1.7.1

```
=300 \\$axii, 285 pages :\$billustrations (black and white) ;\$c24 cm =336 \\\$atext\$2rdacontent =337 \\$aunmediated\$2rdamedia =338 \\\$avolume\$2rdacarrier =504 \\$alncludes bibliographical references and index.
```

No 490, no period to end 300

```
490 present, 300 field ends in a "."
```

```
=300 \\$aviii, 275 pages ;$c25 cm.

=336 \\$atext$2rdacontent

=337 \\$aunmediated$2rdamedia

=338 \\$avolume$2rdacarrier

=490 1\$aAAR academy series
```

Local Practice

Genre/Form terms in Topical Heading fields (MARC 650)

=245 04\$aThe King and I

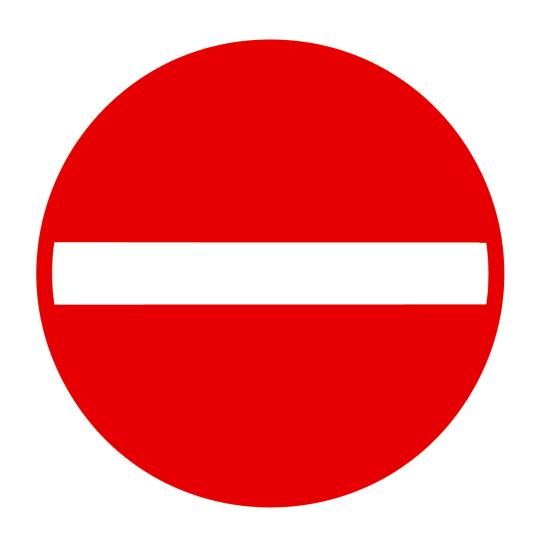
```
=600 10$aLandon, Margaret,$d1903-1993$vFilm adaptations.
```

- =600 10\$aLeonowens, Anna Harriette,\$d1831-1915\$vDrama.
- =600 00\$aMongkut,\$cKing of Siam,\$d1804-1868\$vDrama.
- =650 \0\$aFeature films.
- =650 \0\$aMusical films.

"About" Musical Films OR
"Is a" Musical Film?

- =651 \0\$aThailand\$xSocial life and customs\$vDrama.
- =650 \0\$aFilms for the hearing impaired.
- =650 \4\$aMusical films.

Semantic Ambiguity



Focus on ...

Structure

Statements

Differentiable Values

Vocabularies

Authorities

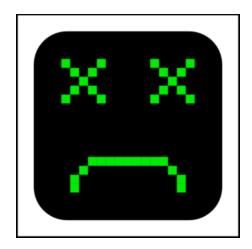
Controlled Access Points

Local Data

Good Data

... and start to focus less on

- Punctuation
- Free Text Fields
- The Data we Can't Easily Normalize
- The Data Machines Can't Use





Thank you.

References

Allemang, Dean and James A. Hendler. 2011. Semantic Web for the Working: Modeling in RDF, RDFS and OWL, 2nd ed. Amsterdam; Boston: Elsevier/Morgan Kaufmann

American Research Libraries (ARL). March 2009. Special Collections in ARL Libraries: A Discussion Report from the ARL Working Group on Special Collections. Washington, DC:ARL. http://www.arl.org/storage/documents/publications/scwg-report-mar09.pdf

Bowen, Jennifer B. 2010. "Moving Library Metadata Toward Linked Data: Opportunities Provided by the eXtensible Catalog." International Conference on Dublin Core and Metadata Applications 0 (0) (September 20): 44–59.

Cole, Timothy W., Myung-Ja Han, William Fletcher Weathers, and Eric Joyner. 2013. "Library Marc Records Into Linked Open Data: Challenges and Opportunities." Journal of Library Metadata 13 (2-3): 163–96. doi:10.1080/19386389.2013.826074.

Coyle, Karen. 2011. "Library Linked Data, Part I: Introduction to the Semantic Web (Mar. 8, 2011)" March 8. http://www.kcoyle.net/presentations/asisti.pdf.

Hilliker, Robert J., Melanie Wacker, and Amy L. Nurnberger. 2013. "Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons." Journal of Library Metadata 13 (2-3): 80–94. doi:10.1080/19386389.2013.826036.

Kincy, Chamya P., and Michael A. Wood. 2012. "Rethinking Access with RDA (Resource Description and Access)." Journal of Electronic Resources in Medical Libraries 9 (1): 13–34. doi:10.1080/15424065.2012.651573.

Lahanas, Stephen. March 9, 2009. "Understanding The Semantic Value Proposition." Semanticweb.com, http://semanticweb.com/understanding-the-semantic-value-proposition b11492?red=su

Lawson, Mark. August 9, 2005. "Berners-Lee on the read/write web", BBC News. http://news.bbc.co.uk/2/hi/technology/4132752.stm

Online Computer Library Centre (OCLC). 2011. "Perceptions of Libraries, 2010: Context and Community". Dublin,OH:OCLC. http://www.oclc.org/content/dam/oclc/reports/2010perceptions/2010perceptions all.pdf

Research Information Network. April 2011. Reinventing Research? Information Practices in the Humanities. http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities

Schreur, Philip Evan. 2012. "The Academy Unbound: Linked Data as Revolution." Library Resources & Technical Services 56 (4) (October 1): 227.

Singer, Ross. 2009. "Linked Library Data Now!" Journal of Electronic Resources Librarianship 21 (2): 114–126. doi:10.1080/19411260903035809.

Tillett, Barbara B. 2011. "Keeping Libraries Relevant in the Semantic Web with Resource Description and Access (RDA)." Serials: The Journal for the Serials Community 24 (3) (November 1): 266–272. doi:10.1629/24266.

W3C Library Linked Data Incubator Group (W3C). 2011. Library Linked Data Incubator Group Final Report. October 5. http://www.w3.org/2005/Incubator/lld/XGR-Ild-20111025/.

Welsh, Anne, and Sue Batley. 2013. Practical Cataloging. AACR, RDA and MARC21. American Library Association.

Zeng, Marcia Lei, Karen F. Gracy, and Laurence Skirvin. 2013. "Navigating the Intersection of Library Bibliographic Data and Linked Music Information Sources: A Study of the Identification of Useful Metadata Elements for Interlinking." Journal of Library Metadata 13 (2-3): 254–278. doi: 10.1080/19386389.2013.827513.

Image Credits

```
http://gregbrainos.files.wordpress.com/2012/10/denmark-outlet.jpg
http://framehawk.com/wp-content/uploads/2012/10/Focus.jpeg
http://www.stencilease.com/gif/CC0094.jpg
RDA Toolkit
http://access.rdatoolkit.org/rdachp3 rda3-2029.html
RDA Metadata Registry
http://metadataregistry.org/concept/list/vocabulary_id/37.html
RDA Toolkit
http://access.rdatoolkit.org/rdachp2 rda2-6453.html
http://www.codeforest.net/wp-content/uploads/2010/09/Database 1.png
http://corrupteddevelopment.com/wp-content/uploads/2012/06/trash-bin-icon-psd.jpg
http://3dwritingservices.files.wordpress.com/2012/10/scattered-keyboard-buttons2.jpg
http://www.proprofs.com/quiz-school/user_upload/ckeditor/do%20not%20enter(1).jpg
http://www.clker.com/cliparts/U/9/E/y/L/Z/sad-computer-md.png
http://commons.wikimedia.org/wiki/
File:Saint John the Baptist Preaching to the Masses in the Wilderness oil on oak panel by Piet
er Brueghel the Younger.jpg
```