

How to pick the low hanging fruits of Linked Data

Seth van Hooland
Ruben Verborgh

DCMI webinar, May 21st 2014





<https://www.flickr.com/photos/smithsonian/2584174182>



facet publishing

the publisher of choice for the information professions worldwide

- > E-books
- > Textbooks
- > Series

Subject Areas

Collection Management

- > Acquisitions & collection development
- > Cataloguing & classification
- > RDA: Resource, Description & Access
- > E-books & electronic resources
- > Preservation, digitisation & conservation

Library Management

- > Copyright & legal issues
- > Evaluation & measurement
- > Library planning & design
- > Management & leadership
- > Marketing & advocacy
- > Training & development

Library Systems & Technologies

- > Digital libraries
- > Intranet & website management
- > Mobile technology
- > The TECH SET
- > About The TECH SET
- > Tools & systems
- > Web 2.0 & beyond

Information Management

- > Information retrieval
- > Knowledge & information management
- > Information science

Professional Skills

- > Academic librarianship

[Home](#) | [Contact](#) | [Log In](#) | [View Basket \(0\)](#)

Cataloguing and classification



Linked Data for Libraries, Archives and Museums

How to clean, link and publish your metadata

Seth van Hooland and Ruben Verborgh

Price: £49.95

Price (to CILIP members): £39.96

ISBN: 978-1-85604-964-1

FORTHCOMING

Pre-order at Bookpoint by calling +44 (0)1235 827702, emailing facet@bookpoint.co.uk or faxing +44 (0)1235 827703

This highly practical handbook teaches you how to unlock the value of your existing metadata through cleaning, reconciliation, enrichment and linking and how to streamline the process of new metadata creation.

Libraries, archives and museums are facing up to the challenge of providing access to fast growing collections whilst managing cuts to budgets. Key to this is the creation, linking and publishing of good quality metadata as Linked Data that will allow their collections to be discovered, accessed and disseminated in a sustainable manner.

Metadata experts Seth van Hooland and Ruben Verborgh introduce the key concepts of metadata standards and Linked Data and how they can be practically applied to existing metadata, giving readers the tools and understanding to achieve maximum results with limited resources.

Readers will learn how to critically assess and use (semi-)automated methods of managing metadata through hands-on exercises within the book and on the accompanying website. Each chapter is built around a case study from institutions around the world, demonstrating how freely available tools are being successfully used in different

[Search Catalogue](#)



[Search Tips](#) | [Advanced Catalogue Search](#)

[Sign up to our eBulletin](#)

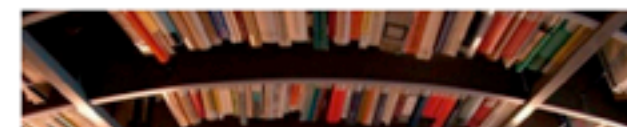
Catalogue 2.0

The future of the library catalogue

Browse a free sample chapter

Reflecting on the Future

of Academic and Public Libraries



Low hanging fruits

- Clean your metadata
- Reconcile with authoritative sources
- Enrich your metadata
- Publish your metadata



Putting LD in practice

- Clean your metadata :
 - UPenn Schoenberg Database of Manuscripts (Philadelphia)
- Reconcile with authoritative sources :
 - Powerhouse Museum (Sydney)
- Enrich your metadata :
 - British Library (London)
- Publish your added-value metadata :
 - Cooper Hewitt National Design Museum (New York)

Pablo Picasso

[Add More Info](#)

[Start New](#)

[Order](#)

[Options](#)

[Use It](#)

Pablo Picasso

picture:



[7]



[8]



[12]



[14]



[16]



[20]

comment: Pablo picasso [7,8,12,14]

fotos, biografia y cuadros de pablo picasso [16]

Voir toute l'actualité de Pablo Picasso [20]

admins: 729353303 [20]

a: Home Page [18]

More jpg wallpaper [18]

Search jpg wallpaper [18]

Download Pablo Picasso Seated Man jpg wallpaper free (50 KB) [18]

app id: 2490221586 [12,14,16]

210301922379857 [20]

alternate: <http://www.slideshare.net/rss/latest> [12,14,16]

<http://www.slideshare.net/api/oembed/1?format=xml&url=http://www.slideshare.net/jessicapavonramos/pablo-picasso-2581640> [14]

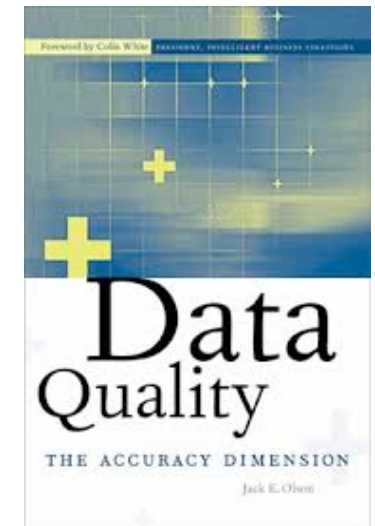
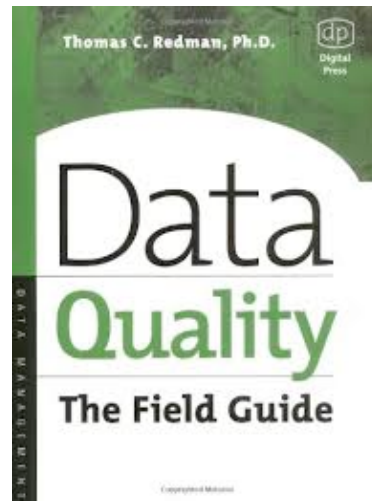
<http://www.slideshare.net/api/oembed/1?format=json&url=http://www.slideshare.net/jessicapavonramos/pablo-picasso-2581640> [14]

<http://www.slideshare.net/api/oembed/1?format=xml&url=http://www.slideshare.net/guest1c5fb9/pablo-picasso-2491014> [16]

<http://www.slideshare.net/api/oembed/1?format=json&url=http://www.slideshare.net/guest1c5fb9/pablo-picasso-2491014> [16]

[show 46 more values](#)

<http://sig.ma/search?q=Pablo+Picasso&templateName=>



The ROI of Data Quality seminar on June 10th, read more on the event page.

READ MORE

Data Quality Association Belgium



DQA is a new organization in Belgium (started April 2012). Our goal is to bring together data quality professionals to learn, share experience and knowledge, network and get better at all aspects of their data quality work.

Interested? Subscribe to our newsletter to stay informed about our activities or drop us an e-mail at info@dqa.be if you have a specific question.

Activities



DQA will be organizing courses, seminars and a yearly congress about Data Quality.

Are you looking for information? A knowledge center will bring together all the presentations and vendor-information you need about Data Quality.

Membership is free. Just subscribe to our newsletter to stay posted.

Newsletter

Sign up for *Newsletter*
Sign up for the DQA newsletter now and stay informed of all our seminars and activities.

Email Address *

First Name

Last Name



The MIT Total Data Quality Management Program

Home

What is TDQM

TDQM Sponsors

TDQM Publications

MITIQ Program

IQ Conference

DQ Resources

News

Contact Us

MIT TDQM Program Highlight

A joint effort among members of the TDQM Program, MIT Information Quality Program, CITM at UC Berkeley, government agencies such as the U.S. Navy, and industry partners. Activities in the TDQM program has evolved into four key areas:

- [MIT Information Quality Program](#)
- [The International Conference on Information Quality](#)
- [The ACM Journal of Data and Information Quality](#)
- [Information Quality Certificate Program](#)

Corporate Householding Research

- [CHH White Paper](#) (Size 359KB)
- [MIT-CHH Introduction](#) (Size 35KB)
- [MIT-CHH Description](#) (Size 163KB)
- ["A Framework for Corporate Householding," Proceedings of the 7th International Conference on Information Quality, Nov. 2002, pp. 36-46. Madnick, S., Wang, R. & Zhang, W. \(Size 374KB\)](#)
- [MIT-CHH Password required area](#)

[Join the MITIQ/ICIOQ mailing list](#)

(Note: We do not distribute email addresses)

Search MIT TDQM Web Site

<http://www.dqa.be/>

<http://web.mit.edu/tdqm/>

Case-study

- Experiment with cleaning operations in a hands-on manner with the the Schoenberg Database of Manuscripts metadata
- Download the data from <http://book.freeyourmetadata.org/chapters/2/>

Google refine *A power tool for working with messy data.*

Create Project
Open Project
Import Project

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.

Get data from

This Computer

Web Addresses (URLs)

Clipboard

Google Data

Enter one or more web addresses (URLs) pointing to data to download:

Add Another URL Next »

Facet / Filter Undo / Redo 0

Refresh Reset All Remove All

Primary seller

ab
☐ case sensitive ☐ regular expression

Title

bel
☐ case sensitive ☐ regular expression

22 matching rows (206253 total)

Extensions: Named-entity recognition ▾ Freebase ▾ RDF ▾

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 10 next > last »

Catalogue ID	Lot or cat.	Price	Currency	Sold	Source	Current location	Author	Author (variant)	Title	Language	Material	Place
Handschriften, BÄcher, Autographen.	9	75					Bruni, Leonardo	Leonardo Bruni Aretino	Commentary on Sallust's Primo Bello Punico	Italian	V	Italy
No. 51: Handschriften; Bucher; Autographen _	2	90					Josephus, Flavius	Josephus Flavius	De Bello Judaico[De Judeorum Vetustate Sive Contra Apionem	Latin	P	Italy
No. 55: Bucher, Handschriften, Autographen ...	5	125					Josephus, Flavius	Josephus Flavius	De Bello Judaico[De Judeorum Vetustate Sive Contra Apionem	Latin	P	Italy
No. 103: Bucher, Autographen	3	3750					Bruni, Leonardo	Leonardo Bruni Aretino	Commentary on Sallust's Primo Bello Punico	Italian	V	Italy
No. 103: Bucher, Autographen	21	1250					Sallust	Crispus Sallust	De Conjuratone Catilinae, de Bello Jugurthino	Latin	V	Italy, Flore
No. 124: Bucher, Autographen	18	1750					Sallust	Crispus Sallust	De Conjuratone Catilinae, de Bello Jugurthino	Latin	V	Italy, Flore
Boston Pub	122					Boston Public Library, MS 1544		Jacopo Palladino Teramo	Consolatio Peccatorum Seu Processus Belial	German	P	Gern
						Cambridge, MA, Harvard University, Houghton Library, MS Typ 170		Camillo Bonadeo	Scelta Dell Piu Belle...	Italian	P	Italy, Plac
Livres precieux et _manuscrits avec miniatures composant la bibliotheque de M. E.M.B.	18								Book of Hours Which Belonged to Claude de Lorraine and Antoinette de Vendome the Ms. Includes 'Oraisons Speciales En Francais' By Pierre Dore At the End	Latin French	V	Fran

Faceting

- One of the core functionalities of Refine, allowing you to discover quickly the true nature of your metadata
- What's the difference between *Primary Seller* and *Secondary Seller* ? Apply a text facet on both

Faceting

- New windows in the left side bar
- By default, ordered alphabetically but click on count
- Apply the same facet on Seller 2, so that we can compare the most popular values of both fields
- Experiment on other fields !
- Also check the outliers !



Facet Choices as Tab Separated Values

```
Sotheby's 26975
Evans 4263
Quaritch 4128
Thorpe, Thomas 3947
Puttick & Simpson 2846
Christie's 2477
Payne & Foss 1705
Rodd, Thomas 1538
Sotheby, Wilkinson & Hodge 1450
Maggs 1328
Kraus, H.P. 1233
De Bure 1121
Laud, William 982
Ashmole, Elias 979
Les Enluminures 915
Luchtmans 717
Saugrain & Leclerc 703
Hoepli 688
Techener 674
```


Clustering

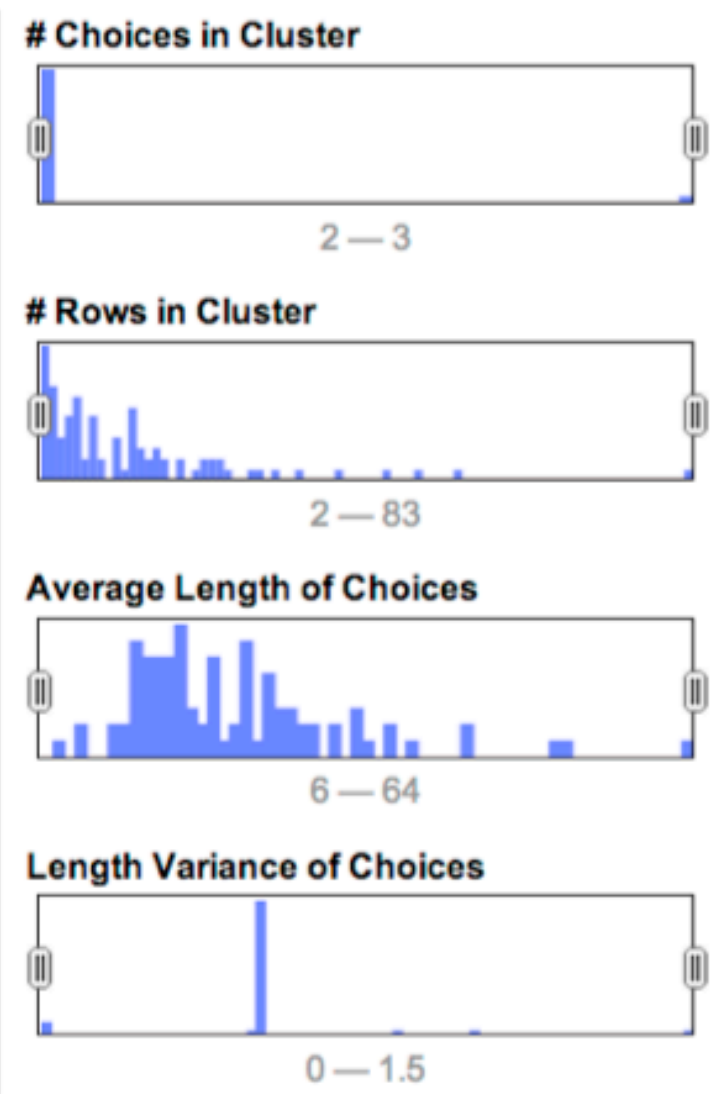
- Aggregate automatically different values regarding the same reality
- One of the best features of Refine
- Example : on the field Artist, apply Edit cells
> Cluster and edit
- New window pops up with clustering features and options

Cluster & Edit column "Artist"

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more ...](#)

Method key collision Keying Function fingerprint **91 clusters found**

Cluster Size	Row Count	Values in Cluster	Merge?	New Cell Value
3	19	<ul style="list-style-type: none">Robert Boyvin (12 rows)Robert Boyvin? (6 rows)Boyvin, Robert (1 rows)	<input type="checkbox"/>	Robert Boyvin
3	6	<ul style="list-style-type: none">Stefano di Tommaso Lunetti? (4 rows)Lunetti, Tommaso di Stefano (1 rows)Stefano di Tommaso Lunetti (1 rows)	<input type="checkbox"/>	Stefano di Tommaso Lunetti
3	23	<ul style="list-style-type: none">Master of Edward IV (21 rows)Master of Edward IV? (1 rows)Master of Edward Iv (1 rows)	<input type="checkbox"/>	Master of Edward IV
3	14	<ul style="list-style-type: none">Girolamo da Cremona (12 rows)Girolamo, da Cremona (1 rows)[Girolamo da Cremona] (1 rows)	<input type="checkbox"/>	Girolamo da Cremona
2	5	<ul style="list-style-type: none">Memling, Hans, school (4 rows)Memling, Hans, school? (1 rows)	<input type="checkbox"/>	Memling, Hans, school
2	4	<ul style="list-style-type: none">Mercati, Venturino (2 rows)Venturino Mercati (2 rows)	<input type="checkbox"/>	Mercati, Venturino

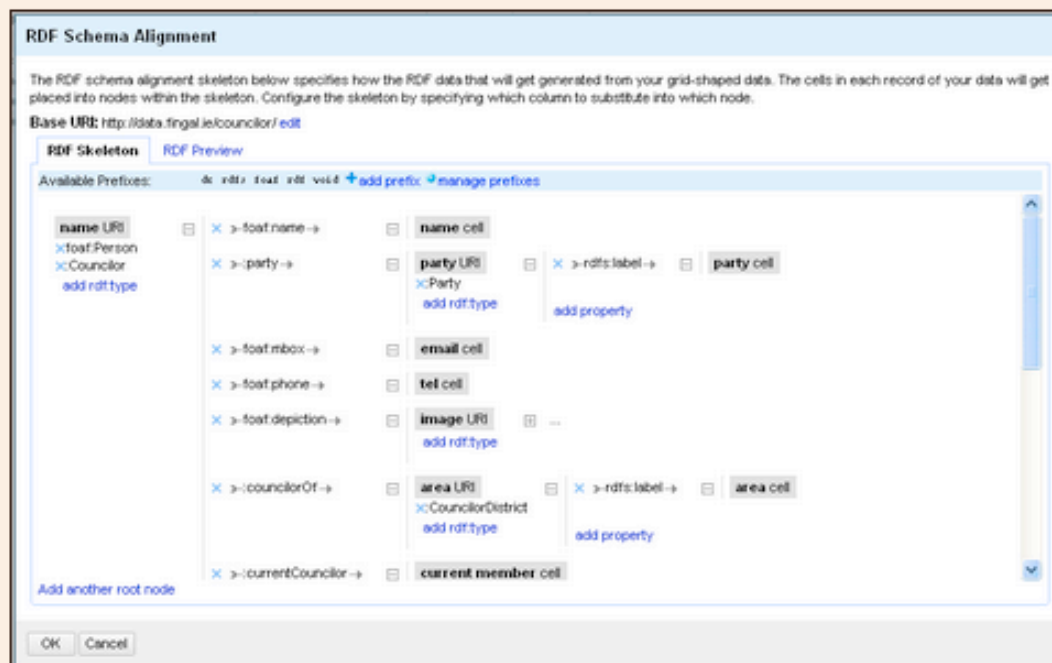


Select All Deselect All Merge Selected & Re-Cluster Merge Selected & Close Close

Putting LD in practice

- Clean your metadata :
 - UPenn Schoenberg Database of Manuscripts
- Reconcile with authoritative sources :
 - Powerhouse Museum
- Enrich your metadata :
 - British Library
- Publish your added-value metadata :
 - Cooper Hewitt National Design Museum

RDF Refine

[Home](#)[Showcases](#)[Docs](#)[Support & Devel](#)[Download](#)

RDF Refine - a Google Refine extension for exporting RDF

Main features

Reconcile & interlink

- Reconcile against SPARQL endpoints, RDF dumps
- Search the Web for related RDF datasets

Export RDF

- GUI for defining the shape of the RDF graph
- Use your own vocabulary or import existing ones
- Autocomplete for property and class names

[Download extension 0.8.0](#)[\(Test/ALPHA!\) Download ORefine extension 0.9.0](#)

Getting started...

- 1 Make sure "extensions" folder exists in your Google Refine workspace
- 2 Download the extension
- 3 Extract the downloaded zip file to the "extensions" folder
- 4 Restart Google Refine

Visit [the installation guide](#) for more detailed instructions. For tutorials-like material see our [showcases](#). For detailed and technical documentation see the [documentation section](#).

News

2014-03-06: a test version compatible with Open Refine 2.6 beta is released!

2012-07-27: version 0.8.0 released
Support for reconciling using [Apache Stanbol](#) is added. This feature has been initially implemented for [LMF](#) and partially funded by the Republic of Austria within the COMET project [Salzburg NewMediaLab](#).

2012-01-18: version 0.7.5 released

Case study

- Experiment with reconciliation operations in a hands-on manner with the metadata of the Powerhouse museum and the LCSH
- Download the data from <http://book.freeyourmetadata.org/chapters/3/>
- Focus on the Categories field, populated with the Powerhouse museum Object Names Thesaurus (PONT), a locally created vocabulary

Extensions: Named-entity recognition ▾ Freebase ▾ RDF ▾

Edit RDF Skeleton...

Reset RDF Skeleton...

Add reconciliation service ▶

Based on SPARQL endpoint...

Based on RDF file...

Based on a Sindice site search...

Based on a Apache Stanbol EntityHub...

Add SPARQL-based reconciliation service

Name: LCSH - PH museum test

A human readable name

Endpoint details

Endpoint URL:

Graph URI:

Leave empty to use the default graph

Type:

This determines the syntax that will be used for search

Label properties

Select properties that are used to label resources in the endpoint. These properties will be used to match resources:

☐ rdfs:label

☒ skos:prefLabel

☐ dcterms:title

☐ dc:title

☐ foaf:name

☐ Other...

OK

Cancel

Facet / Filter

[Undo](#) / [Redo](#) 3

[Refresh](#)

[Reset All](#)

[Remove All](#)

[x](#) Record ID

99

☐ case sensitive

☐ regular expression

3442 matching records (100170 total)

Show as: [rows](#) [records](#)

Show: [5](#) [10](#) [25](#) [50](#) records

v All	v Record ID	v Categories
☆ 🔍 80.	6991	Specimens
☆ 🔍		Botanical specimens
☆ 🔍 99.	6799	Hand fans
☆ 🔍		Personal Effects
☆ 🔍		Clothing and Dress
☆ 🔍 222.	369939	Specimens
☆ 🔍		Wool specimens
☆ 🔍		Animal Samples and Products
☆ 🔍		Textile Technology
☆ 🔍 223.	369981	Wool specimens
☆ 🔍		Animal Samples and Products
☆ 🔍 226.	369945	Specimens
☆ 🔍		Wool specimens
☆ 🔍		Animal Samples and Products
☆ 🔍		Textile Technology
☆ 🔍 229.	371995	Wool specimens
☆ 🔍		Animal Samples and Products
☆ 🔍 230.	371994	
☆ 🔍 284.	5599	Castings
☆ 🔍		Sculpture
☆ 🔍 314.	5499	Sculptures
☆ 🔍		Writing equipment
☆ 🔍		Sculpture
☆ 🔍 373.	5299	Castings
☆ 🔍		Draughts sets
☆ 🔍		Sculpture

Facet / Filter

Undo / Redo 5

Refresh

Reset All

Remove All

Categories: judgment [change](#) [invert](#) [reset](#)

3 choices Sort by: **name** count

(blank) 5664

matched 24198

none 174786

[exclude](#)

[Facet by choice counts](#)

Categories: best candidate's score [change](#) [reset](#)



0.07 — 1.00

☒ Numeric 71574 ☐ Non-numeric 0 ☐ Blank 0 ☒ Error 67872

Categories [change](#)

4995 choices Sort by: **name** count

[Cluster](#)

[Photographs](#) 10027

[Numismatics](#) 8505

[Clothing and Dress](#) 7636

[Ceramics](#) 7478

[Glass plate negatives](#) 7122

[Pictorials](#) 5143

[Documents](#) 4681

[Photographic prints](#) 3870

[Specimens](#) 2608

[Botanical specimens](#) 2589

[Domestic Equipment-Home](#) 2359

[Timber Samples](#) 2322

91881 matching records (100170 total)

Show as: **rows** records

Show: **5** 10 25 50 records

All

Categories

★ 1.

[Meteorites](#)

[Choose new match](#)

★

[Models](#)

☒ ☒ [Models \(Patents\)](#) (0.375)

☒ ☒ [Models \(Persons\)--Vocational guidance](#) (0.162)

☒ ☒ [Models and modelmaking--Radio control systems](#) (0.133)

☒ ☒ [Create new topic](#)

★

[Models](#)

☒ ☒ [Models \(Patents\)](#) (0.375)

☒ ☒ [Models \(Persons\)--Vocational guidance](#) (0.162)

☒ ☒ [Models and modelmaking--Radio control systems](#) (0.133)

☒ ☒ [Create new topic](#)

★

[Mineral Samples-Geological](#)

☒ ☒ [Create new topic](#)

★ 3.

[Botanical specimens](#)

[Choose new match](#)

★

[Timber Samples](#)

☒ ☒ [Create new topic](#)

★ 4.

[Wool specimens](#)

☒ ☒ [Create new topic](#)

★

[Specimens](#)

☒ ☒ [Specimens](#) (1)

☒ ☒ [Specimens](#) (1)

☒ ☒ [Voucher specimens \(Biological specimens\)](#) (0.225)

☒ ☒ [Create new topic](#)

★

[Animal Samples and Products](#)

☒ ☒ [Create new topic](#)

★

[Textile Technology](#)

☒ ☒ [Create new topic](#)

★ 5.

[Specimens](#)

☒ ☒ [Specimens](#) (1)

☒ ☒ [Specimens](#) (1)

Meteorites

From [Library of Congress Subject Headings](#)


[Details](#)[Visualization](#)[Suggest Terminology](#)

Meteorites

URI(s)

- > <http://id.loc.gov/authorities/subjects/sh85084315>
- > <info:lc/authorities/sh85084315>
- > <http://id.loc.gov/authorities/sh85084315#concept>

Instance Of

- > [MADS/RDF Topic](#)
- > [MADS/RDF Authority](#)
- > [SKOS Concept](#) 

Scheme Membership(s)

- > [Library of Congress Subject Headings](#)

Collection Membership(s)

- > [LCSH Collection - Authorized Headings](#)
- > [LCSH Collection - General Collection](#)
- > [LCSH Collection - May Subdivide Geographically](#)

Broader Terms

- >  [Meteors](#)

Narrower Terms

- >  [Achondrites](#)

100170 records									
Show as: rows records					Show: 5 10 25 50 records				
▼ All					▼ Categories				
☆	🗨				Models				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Models (Patents) (0.375)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Models (Persons)—Vocational guidance (0.162)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Models and modelmaking—Radio control systems (0.133)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic				
☆	🗨				Models				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Models (Patents) (0.375)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Models (Persons)—Vocational guidance (0.162)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Models and modelmaking—Radio control systems (0.133)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic				
☆	🗨				Mineral Samples-Geological				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic				
☆	🗨	2.			Botanical specimens				
					Choose new match				
☆	🗨				Botanical specimens				
					Choose new match				
☆	🗨	3.			Botanical specimens				
					Choose new match				
☆	🗨				Timber Samples				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic				
☆	🗨	4.			Wool specimens				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic				
☆	🗨				Specimens				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Specimens (1)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Specimens (1)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Voucher specimens (Biological specimens) (0.225)				
					<input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic				

Specimens

From [Library of Congress Subject Headings](#)

Details

Visualization


Suggest Terminology

Specimens

URI(s)

- > <http://id.loc.gov/authorities/subjects/sh87006764>
- > <info:lc/authorities/sh87006764>
- > <http://id.loc.gov/authorities/sh87006764#concept>

Instance Of

- > [MADS/RDF Topic](#)
- > [MADS/RDF Authority](#)
- > [SKOS Concept](#) 

Scheme Membership(s)

- > [Library of Congress Subject Headings](#)

Collection Membership(s)

- > [LCSH Collection - General Collection](#)

Specimens

From [Library of Congress Subject Headings](#)

Details

Visualization

Suggest Terminology


Specimens

Use as a form subdivision under types of publications, printed matter, etc., for actual specimens of the material.

URI(s)

- > <http://id.loc.gov/authorities/subjects/sh99001749>
- > <info:lc/authorities/sh99001749>
- > <http://id.loc.gov/authorities/sh99001749#concept>

Instance Of

- > [MADS/RDF GenreForm](#)
- > [MADS/RDF Authority](#)
- > [SKOS Concept](#) 

Scheme Membership(s)

- > [Library of Congress Subject Headings](#)

Collection Membership(s)

- > [LCSH Collection - Subdivisions](#)
- > [LCSH Collection - GenreForm Subdivisions](#)
- > [LCSH Collection - General Collection](#)
- > [Pattern Heading Collection - H1095](#)

Putting LD in practice

- Clean your metadata :
 - UPenn Schoenberg Database of Manuscripts
- Reconcile with authoritative sources :
 - Powerhouse Museum
- Enrich your metadata :
 - British Library
- Publish your added-value metadata :
 - Cooper Hewitt National Design Museum

What is NER ?

- Consider the sentence « On 25 September 2006, we visited Washington to see the White House »
- First step => identification
 - 25 September 2006
 - Washington
 - White House
- Second step => disambiguation

What is NER ?

- Each entity is associated with a meaning :
 - http://dbpedia.org/resource/White_House
 - http://dbpedia.org/page/Washington,_D.C
- NE extraction workflow consists of analyzing input content for detecting named entities, assigning them a type weighted by a confidence score and by providing a list of URIs for disambiguation



Named-Entity Recognition extension for Google Refine / OpenRefine

🕒 90 commits

🔗 1 branch

🏷️ 5 releases

👤 3 contributors



🔗 branch: master ▾

Refine-NER-Extension / +

Add WikiMeta service. ...



RubenVerborgh authored 4 days ago

latest commit 1879e4e48a 📄

📁 module	implement the "reconstruct" method to allow for history-apply / reapply	3 months ago
📁 src	Add WikiMeta service.	4 days ago
📄 .classpath	Add base class for NER commands with JSON output.	a year ago
📄 .gitignore	Initialize service manager with default list if no previous configura...	a year ago
📄 .project	Initialize project with menu placeholders.	a year ago
📄 LICENSE.md	Add license.	a year ago
📄 README.md	Update paper title.	6 months ago
📄 build.xml	Added missing line to copy DefaultServices.json to classes dir	a year ago

<> Code

🔔 Issues

2

🔗 Pull Requests

1

📶 Pulse

📊 Graphs

🔗 Network

HTTPS clone URL

https://github.com 📄

You can clone with [HTTPS](#) or [Subversion](#). ?

🖥️ Clone in Desktop

☁️ Download ZIP

<https://github.com/RubenVerborgh/Refine-NER-Extension>

Adding extra services

- You need to request an API key to make use of the services of Alchemy and Zemanta :
 - <http://www.alchemyapi.com/api/register.html>
 - <http://developer.zemanta.com/member/register/>
- Click the Named-entity recognition toolbar button and choose Configure API keys
- Add the keys you received and click Update

Case-study

- Experiment with reconciliation operations in a hands-on manner with the metadata of the British Library (CSV conversion from an RDF file available through Europeana)
- Download the data from <http://book.freeyourmetadata.org/chapters/4/>
- We're only interested in the description field
=> Choose View > Collapse other columns

Facet / Filter

[Undo / Redo 3](#)

[Refresh](#)

[Reset All](#)

[Remove All](#)

[x DBpedia Spotlight: judgment](#)

[change](#)

2 choices Sort by: [name](#) [count](#)

(blank) 274

matched 177

[Facet by choice counts](#)

[x AlchemyAPI: judgment](#)

[change](#)

3 choices Sort by: [name](#) [count](#)

(blank) 416

(unreconciled) 18

matched 17

[Facet by choice counts](#)

[x Zemanta: judgment](#)

[change](#)

2 choices Sort by: [name](#) [count](#)

(blank) 105

matched 346

[Facet by choice counts](#)

100 records

Extensions:

[Named-entity recognition](#)

[Freebase](#)

[RDF](#)

Show as: [rows](#) [records](#)

Show: [5](#) [10](#) [25](#) [50](#) records

[« first](#) [« previous](#) [1 - 25](#) [next »](#) [last »](#)

v All	v description	v AlchemyAPI	v Zemanta	v DBpedia Spotlig
☆ 🗨 1.	Sir David (1758-1825) dancers and musicians hookahs portraits sports and pastimes water-colour Delhi style South Asia – History 954	Sir David	Outline of South Asian history Choose new match	
☆ 🗨			Delhi Choose new match	
☆ 🗨			David Choose new match	
☆ 🗨			Watercolor painting Choose new match	
☆ 🗨			Hookah Choose new match	
☆ 🗨 2.	Gouache painting with gold of village life in Kashmir	Kashmir	Kashmir Choose new match	Gouache Choose new match
☆ 🗨 3.	Watercolour depicting a Durga Puja		Durga Puja Choose new match	Durga Puja Choose new match
☆ 🗨 4.	India Tamil Nadu Vellore clerks occupations opaque water-colour Tanjore/Vellore style South Asia – History 954		Outline of South Asian history Choose new match	Tamil Nadu Choose new match
☆ 🗨			Tamil Nadu Choose new match	
☆ 🗨			Vellore Choose new match	
☆ 🗨			Watercolor painting Choose new match	
☆ 🗨				Vellore Choose new match
☆ 🗨				Tanjore Choose new match
☆ 🗨				Vellore Choose new match

Putting LD in practice

- Clean your metadata :
 - UPenn Schoenberg Database of Manuscripts
- Reconcile with authoritative sources :
 - Powerhouse Museum
- Enrich your metadata :
 - British Library
- Publish your added-value metadata :
 - Cooper Hewitt National Design Museum

Introducing REST

You *don't* need an API – your *website* is the API.

resources

representations

self-describing messages

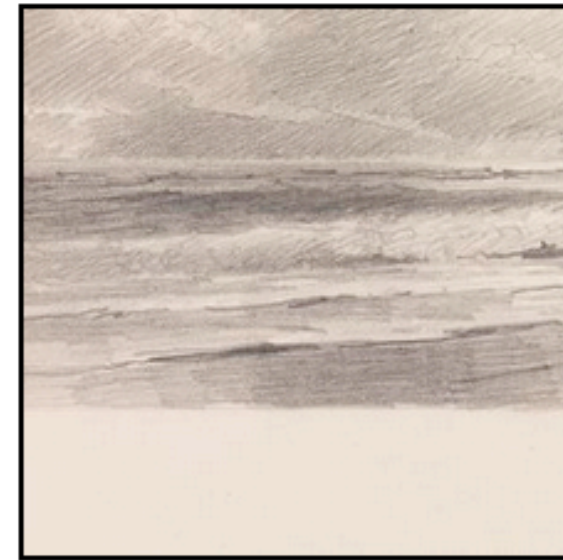
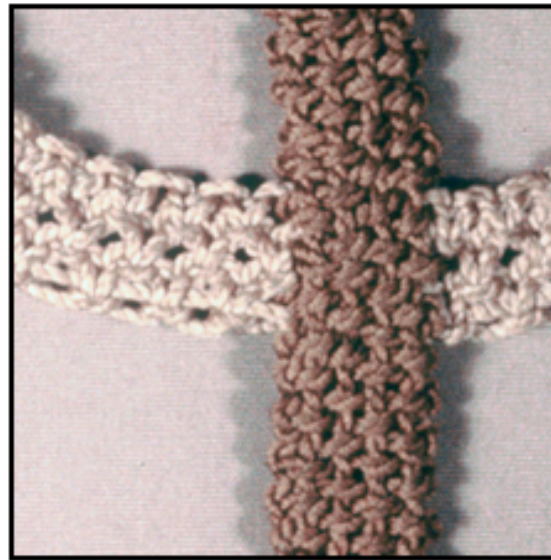
hypermedia

REST – REpresentational State Transfer architectural style



Welcome to the beta release of our new collection database!

This is our stuff, we have lots of it



[show me more things!](#)

This database represents approximately 75% of the Smithsonian Cooper-Hewitt, National Design Museum's diverse collection spanning thirty centuries of historic and contemporary design, including [thoughts on typography](#), [electronic intelligence](#), [a cat wishing it could play the guitar](#), [George "The Iceman" Gervin](#), [a Chinese grocery](#) and [icebergs](#).

A beta release is [an early release and means that active development is ongoing](#) — object research is being added, bugs are being fixed, and erroneous terms are being revised. In spite of the eccentricities of raw data, you can [begin exploring the collection](#) and discovering unexpected connections among objects and designers.

<https://collection.cooperhewitt.org>

A URL uniquely identifies a conceptual resource

Don't.

**http: //example.org/
collection/
showObject.aspx**

What is this?

Can I bookmark this?

Can I share this?

Do.

**http://example.org/
objects/18353113/**

What is this?

Can I bookmark this?

Can I share this?

Each resource can have multiple representations

Don't.

<http://example.org/objects/18353113/> gives HTML

<http://api.example.org/getObjectJson.php?id=18353113> gives JSON

Can I bookmark this?

Can I share this?

Do.

<http://example.org/objects/18353113/>

gives HTML.

gives JSON.

gives RDF.

Can I bookmark this?

Can I share this?

Use self-descriptive messages

Don't.

/objects?filter=toy

/?page=2

Can I bookmark this?

Can I share this?

Do.

/objects?filter=toy

**/objects?
filter=toy&page=2**

Can I bookmark this?

Can I share this?

Use hypermedia in all your representations

Don't.

```
{  
  "title": "Spun Chair",  
  "producer": {  
    "id": 1804  
  }  
}
```

Can I act on this?

Do.

```
{  
  "title": "Spun Chair",  
  "producer": {  
    "url": "/producers/  
1804"  
  }  
}
```

Can I act on this?

What happens if you don't ?

See DPLA and Europeana

What people need to do:

<http://dp.la/item/ecdafcf9b06be6efed042e40b3923e57>

What machines need to do:

Request an API key.

Receive an e-mail with this key.

Find the right URL template for the “API call”.

Fill out details in the template to construct the URL.

Open this URL.

Drawing, "Page from a sketchbook showing St. Luke.", ca. 1590

type: [Drawing](#)

See [all objects](#).

Vertical rectangle. The Apostle stands before a niche, thinking about what he will write next. The Bull crouches behind him. The left lower part of the coat shows the correction of a first design. Verso: "B," horizontally. The bull runs towards left. The putto flies beside him, holding his head. Caption: "A and scribbling/ a fare col vise fra le corne e con lemane sula fronte che/ il toro vede lume corendo." Notes vertically, on top: "adi 18 dicembre bighest et fatken van/ 2 trelle tetrinken/ La di 14 di gennaio comencai a/ bere la bottade 4 borili e 1/2."

Participants

Department

[Drawings, Prints, and Graphic Design](#)

Images



<http://dataplatfom.freeyourmetadata.org/>

GET http://datapatform.freemetadata.org/objects/18109141

Edit and Retry

200 OK 4.29 kB 201 ms

[View Request](#) [View Response](#)

HEADERS

Access-Control-Allow-Origin: *
Connection: keep-alive
Content-Encoding: gzip
Content-Type: application/json
Date: Thu, 15 May 2014 19:12:33 GMT
Etag: "136774574"
Server: nginx/1.1.19
Transfer-Encoding: chunked
Vary: Accept-Encoding
X-Powered-By: Express

BODY

[view raw](#)

```
{
  "accession_number": "1901-39-126",
  "title": "Drawing, \"Page from a sketchbook showing St. Luke.\", ca. 1590",
  "url": "http://datapatform.freemetadata.org/objects/18109141",
  "department_id": "35347493",
  "period_id": null,
  "media_id": "35365609",
  "type_id": "35237093",
  "date": "ca. 1590",
  "year_start": null,
  "year_end": null,
  "year_acquired": "1901",
```

<http://hurl.it>

Give humans and machines the same API: the Web

It's all you need now and in the future.

**Technologies will change,
so identify your concepts,
not the technology used to retrieve them.**

**Use the Web's links and form
to navigate between concepts.**

Get in touch !

- Handbook will be available from 19th of June - a review copy anyone ?
- Follow @freemetadata, @RubenVerborgh and @sethvanhooland
- EU and US promo tour - contact us if you want to collaborate or co-organize a workshop