Answering the call for more accountability:

Applying data profiling to museum metadata

Seth van Hooland, Seth Kaufman and Yves Bontemps

Outline

- Introduction to the issue of metadata quality
- Biography of a museum catalog
- Applying data profiling to museum metadata
- Pro-active approach: incorporating metadata quality tools into OpenCollection

Introduction to (meta)data quality

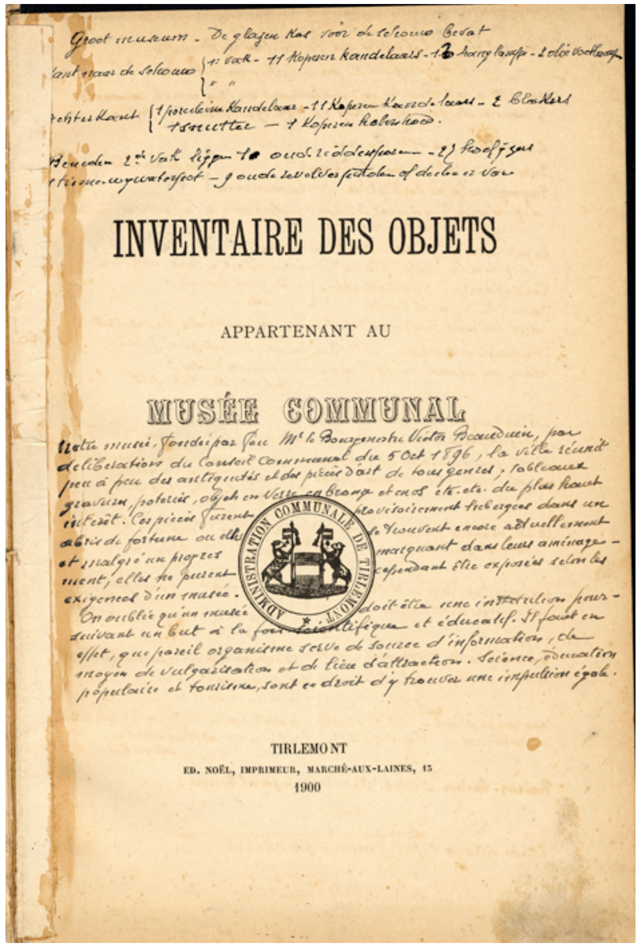
- Problematic definition
- Data quality approaches: Thomas Redman and Richard Wang
- Criticized by Isabelle Boydens who proposes a hermeneutic approach
- Key question: NOT “are the metadata accurate?” BUT “how are metadata progressively constructed?”

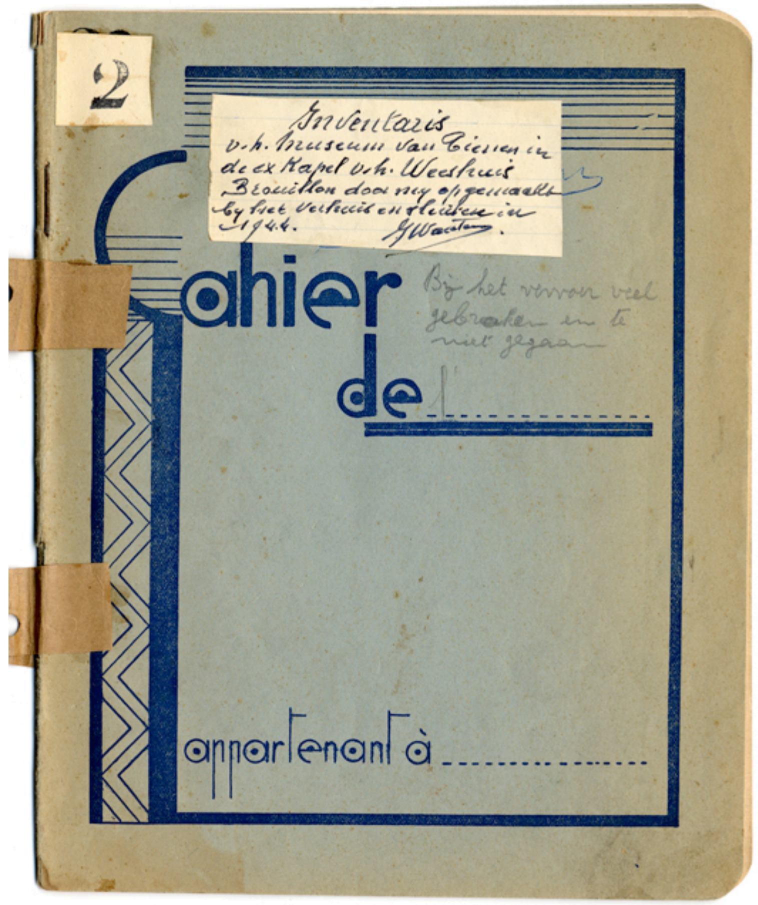
Metadata and change

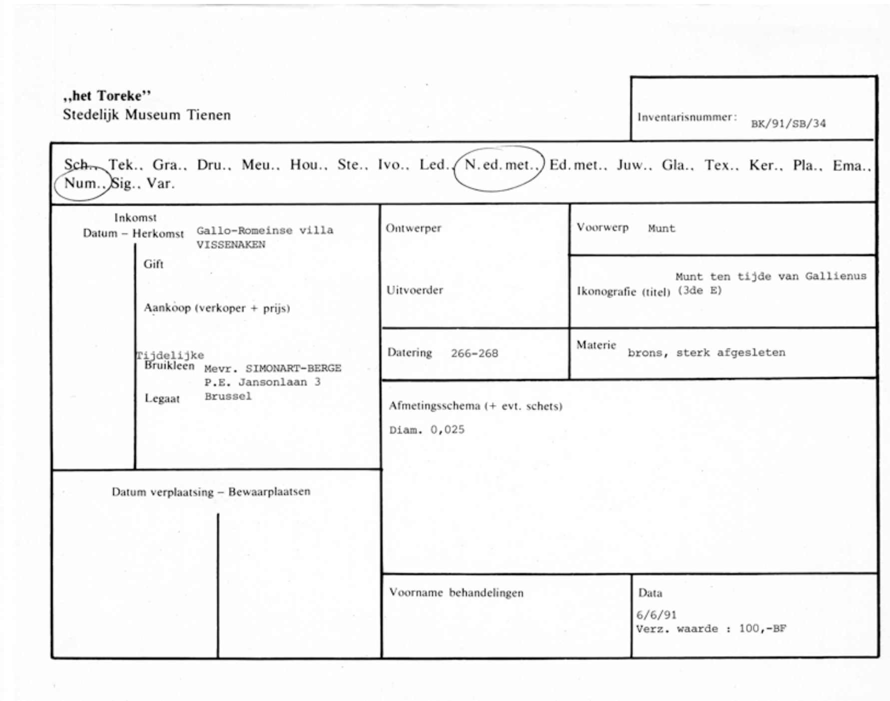
- “Meta” points out to a higher level
- But also to the idea of change, evolution (e.g. metamorphosis)
- Need of a framework to grasp changes and their impact upon metadata quality

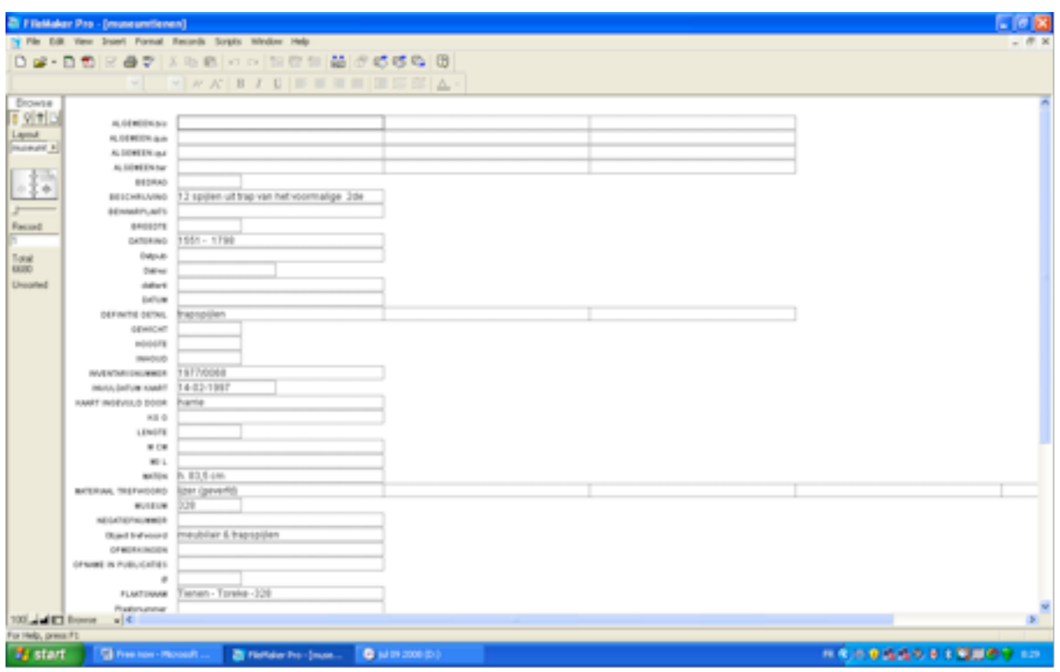
“Temporalités étagées” (Braudel/Boydens)

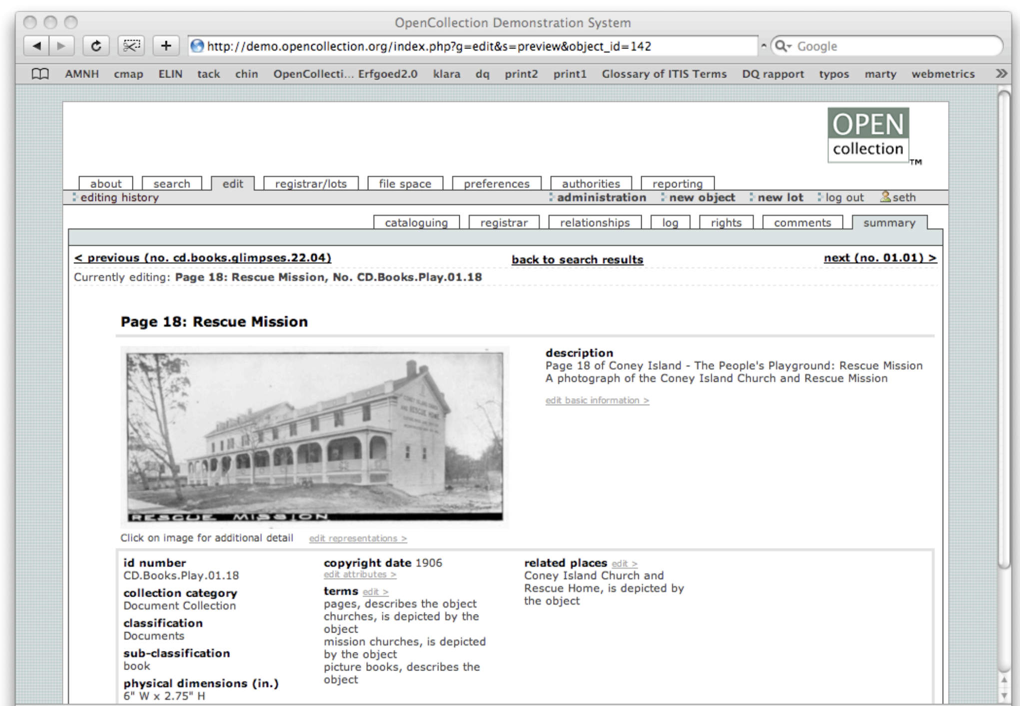
- Provides a framework to grasp MDQ
- Temps longs: long term evolutions that are reflected in digitization policies
- Temps intermédiaire: evolution in standards and technologies
- Temps courts: evolution of the metadata themselves


INVENTAIRE DES OBJETS
APPARTENANT AU
MUSÉE COMMUNAL
TIRLEMONT
ED. NOËL, IMPRIMEUR, MARCHÉ-AUX-LAINES, 15
1900


2
Cahier de
appartenant à


„het Toreke"
Stedelijk Museum Tienen
Inventarisnummer: BK/91/SB/34
Sch., Tek., Gra., Dru., Meu., Hou., Ste., Ivo., Led., N.ed.met., Ed.met., Juw., Gla., Tex., Ker., Pla., Ema., Num., Sig., Var.
Inkomst
Datum – Herkomst Gallo-Romeinse villa VISSENAKEN
Gift
Aankoop (verkoper + prijs)
Tijdelijke Bruikleen Mevr. SIMONART-BERGE P.E. Jansonlaan 3 Brussel
Legaat
Ontwerper
Uitvoerder
Voorwerp Munt
Ikonografie (titel) Munt ten tijde van Gallienus (3de E)
Datering 266-268
Materie brons, sterk afgesleten
Afmetingsschema (+ evt. schets)
Diam. 0,025
Datum verplaatsing – Bewaarplaatsen
Voorname behandelingen
Data
6/6/91
Verz. waarde : 100,-BF




OpenCollection Demonstration System
http://demo.opencollection.org/index.php?g=edit&s=preview&object_id=142
OPEN collection
about
search
edit
registrar/lots
file space
preferences
authorities
reporting
cataloguing
registrar
relationships
log
rights
comments
summary
< previous (no. cd.books.glimpses.22.04)
back to search results
next (no. 01.01) >
Currently editing: Page 18: Rescue Mission, No. CD.Books.Play.01.18
Page 18: Rescue Mission
description
Page 18 of Coney Island - The People's Playground: Rescue Mission
A photograph of the Coney Island Church and Rescue Mission
Click on image for additional detail
id number
CD.Books.Play.01.18
collection category
Document Collection
classification
Documents
sub-classification
book
physical dimensions (in.)
6" W x 2.75" H
copyright date 1906
terms
pages, describes the object
churches, is depicted by the object
mission churches, is depicted by the object
picture books, describes the object
related places
Coney Island Church and Rescue Home, is depicted by the object

Biography of a catalog:

- Temps longs: evolution from French to Dutch as an indexing language
- Temps intermédiaire: evolution from inventory in a notebook > catalog cards > FileMaker Pro > web-based applications (OpenCollection)
- Temps courts: evolution of the metadata themselves

Changes in metadata from 1900 to 2008

- ``20. Un mortier moyen modele en bronze portant l'inscription Petrus Vanden Gheyn me fecit 1546. Le pilon manque"

```
<?xml version="1.0"?>
<metadata
  xmlns="http://example.org/myapp/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://example.org/myapp/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <dc:title>Vijzel in gegoten brons.</dc:title>
  <dc:description>Vijzel in gegoten brons, met insnoeringen,
middenste band versiert. Omschrift: PETRUS VANDEN GHEIN ME
FECIT MCCCCC XLVI</dc:description>
  <dc:subject>mortier, vijzel</dc:subject>
  <dc:publisher>Stadsmuseum Tienen</dc:publisher>
  <dc:identifier>http://www.tienen.be/</dc:identifier>
</metadata>
```

From theory to practice...

- Conclusion: change is a fundamental notion when working on metadata quality
- => Incorporate tools within collection management software that help metadata practitioners to monitor change
- Answers the call for objective information needed to develop a long-term metadata policy

Data profiling: “chique et pas cher”

- First step to MDQ
- Simple and straight-forward tool that allows a basic understanding of the formal characteristics of a metadata set

Profile execution: africa-museum-analysis 6/3/08 9:35 AM

Context: objectnumber

Analysis: objectnumber.patternanalyzer(Pattern A...

Group: Histogram

name	value
AA.9999.99.99	14555.0
AA.9999.99.9999	11862.0
AA.9.9.99999	10213.0
AA.9999.99.999	8123.0
AA.9999.99.9	7328.0
AA.9999.9.99	2467.0
AA.9999.9.999	2451.0
AA.9.9.9999	2186.0
AA.9999.9.9999	1719.0

Export results | Show examples | Export examples

objectid	objectnumber	object...	date_of...	date...	dat...	title	medium
122523	EO.1946.36.11	... 1				sceptre	
63332	EO.1948.14.14	... 1				objet sculptÈ	
63144	EO.1948.14.15	... 1				objet sculptÈ	
69367	EO.1948.14.16	... 1				lance d'ostentation	
63133	EO.1948.14.17	... 1				lance d'ostentation	
64279	EO.1948.14.35	... 1				canne sculptÈe	bois
64280	EO.1948.14.36	... 1				canne sculptÈe	bois
64518	EO.1948.14.46	... 1				couteau de jet	
62777	EO.1948.14.78	... 1				lance	
62778	EO.1948.14.79	... 1				corne de búuf	
62779	EO.1948.14.80	... 1					
128698	EO.1948.14.82	... 2					
128698	EO.1948.14.82	... 2					
98215	EO.1948.27.10	... 1				statuette	
98216	EO.1948.27.11	... 1				statuette	
98217	EO.1948.27.12	... 1				statuette	
98218	EO.1948.27.13	... 1				statuette	

Examples of analyses:

Pattern	Number of occurences	Example
(empty)	65011	
9999-9999	1564	1891-1914
9999	1105	1909
99-99/9999	574	09-10/1992
99/9999	374	01/1994
99-9999	346	08-1950
99/99/9999	312	04/08/1963
AAA 9999	90	Mai 1938
AAAAAAA-AAAA 9999	84	Janviers-mars 1999
99-99 9999	61	01-02 1993

Title	Number of occurences
(empty)	5623
statuette	2043
panier	1800
bracelet	1792
collier	1376
masque	1324
groupe	1250
couteau	1073
sifflet en bois	1012

Taking MDQ a step further

- Data profiling only analyzes the current state of metadata
- But a pro-active approach also analyzes user needs and perceptions of metadata
- By understanding how metadata is *actually* used one can better deploy resources for data enhancement and creation
- Can also provide hints about problems with metadata quality

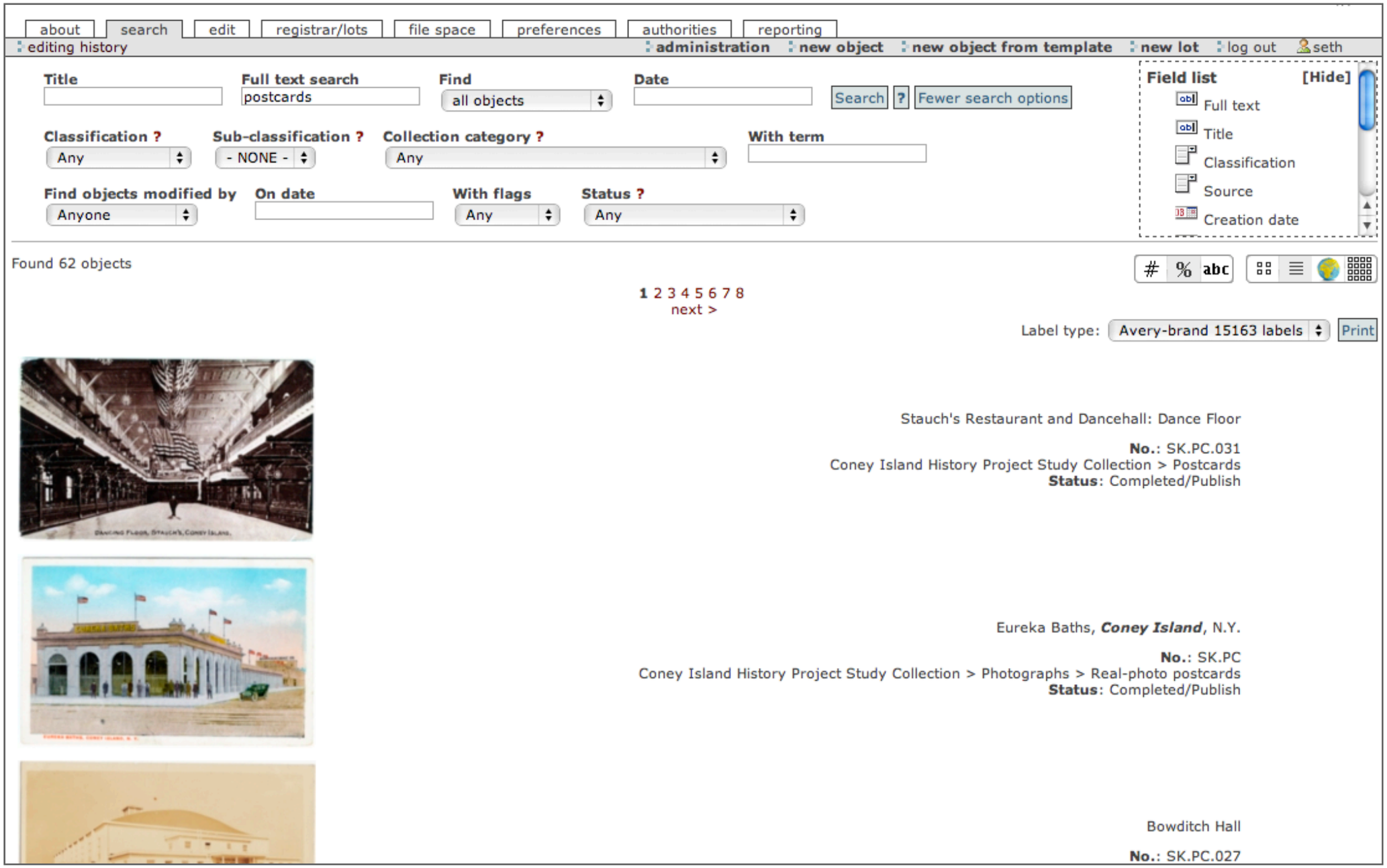
Passive user behavior analysis

- Classic approaches:
  - Surveys – users lie
  - Query logs – interpretation can be difficult; often inaccurately convey user intentions
- Needed: a transparent and objective means of ascertaining what type of metadata users are interested in

Dynamic search interfaces

- Analyze *how* users search by providing a mechanism for users to construct their own custom search forms
  - discover what types of data they are interested in searching on
  - discover what types of data they *actually* search on
  - detect possible quality issues by looking at data elements that are removed from forms after use

Dynamic search interfaces



Dynamic metadata views

- Analyze demand for specific metadata by providing a mechanism for user creation of custom data displays
  - discover what kinds of data they are interested in using
  - discover what types of data they *actually* use
  - discover what data elements are included in displays and subsequently removed

Dynamic metadata views

about | search | edit | registrar/lots | file space | preferences | authorities | reporting

editing history | administration | new object | new object from template | new lot | log out | seth

cataloguing | registrar | relationships | log | rights | comments | summary

back to search results — next (no. sk.pc) >

Currently editing: **Stauch's Restaurant and Dancehall: Dance Floor, No. SK.PC.031**

Stauch's Restaurant and Dancehall: Dance Floor

Click on image for additional detail
edit representations >

description
Stauch's Restaurant and Dancehall: Dance Floor
edit basic information >

Field list [Show]
- Full text
- Title
- Classification
- Source
- Creation date

id number
SK.PC.031

collection category
Coney Island History Project Study Collection

classification
Postcards

postmark date June 05 2007 at 9:00 to 17:00

postmark date June 05 2007 at 9:00 to June 05 2008 at 17:00

date created 2006

postmark date July 24 2008 at 20:46

postmark date June 05 to June 15 2007
edit attributes >

related places edit >
Stauch's Restaurant and Dance Hall, was creation location of the object

current location edit >
cabinet a

Implementation within OpenCollection

- An open-source collections management application for museums and archives
- Handles a wide-variety of media and descriptive, technical and administrative metadata
- In use at 25+ institutions in the North America, South America and Europe for a variety of collections: fine art, moving images, anthropology, photography, oral history, natural history, etc.

Implementation within OpenCollection

- Profiling tools are integrated with reporting engine
- Dynamic interfaces are integrated with existing OC cataloguer's search and browse facility
- Currently metadata profiling and dynamic user interface tools are in prototype phase
- Plan to rollout to selected OC sites in Fall 2008

Danke!

Questions or comments?

Contact:

Seth van Hooland
(svhoolan@ulb.ac.be)

Seth Kaufman
(seth@opencollection.org)

For more information on OpenCollection:
http://www.opencollection.org