

Data Management: Documentation & Metadata

Sherry Lake, Senior Data Consultant

Bill Corey, Data Consultant

Jeremy Bartczak, Intellectual Access & Metadata Services Librarian

Documentation & Metadata

Agenda:

- Metadata: What is it? Why is it important?
- Capturing and managing metadata:
 - Best practices
 - Tools
- Hands-On Creation of Metadata
- Questions

Metadata in Everyday Life

Name	Time	Artist	Album	Genre	Rating	Plays
Skillet	4:48	AC/DC	Skillet - Single	Pop		
O Come All Ye Faithful	3:52	Amy Grant	iTunes Holiday Sa...	Holiday		
Angels We Have Heard On High	5:26	Artha Franklin	iTunes Holiday Sa...	Holiday		
Carol of the Bells / Jingle Bells	2:48	Barry Manilow	iTunes Holiday Sa...	Holiday		
Norway	3:50	Beach House	Norway - Single	Alternative		
Cloud Nine	3:28	Ben Howard	These Waters EP	Music		
A Few Horrend Words	3:42	Ben Sollee	Learning to Bend	Alternative		
I Am Not a Farmer	3:37	Bill Frisell	Distarmer	Jazz		
The Lazy Song	3:08	Bruno Mars	Doon-Wopps & Hood...	Pop		
Looked Out of Heaven	3:58	Bruno Mars	Uppercrust Jukebox	Pop		
Don't View	5:27	Cash McCombe	Cashcombs	Alternative		
Leap	3:50	The Cave Singers	Welcome Joy (Bon...	Singer/So...		
Dream a Dream	3:49	Charotte Church	iTunes Holiday Sa...	Holiday		
Jar of Hearts	4:07	Christina Perri	Jar of Hearts - Single	Pop		
Counting Back to You	3:40	Daughtry	Break the Spell (Si...	Rock		
The First Noel	4:32	David Archuleta	iTunes Holiday Sa...	Holiday		
Easy	3:52	Deer Tick	Born On Flag Day	Alternative		
Acorn Factory	4:04	The Dodos	Time To Die	Alternative		
In Love With a Girl	3:26	Gavin DeGraw	Gavin DeGraw	Rock		
A Drowsy Fall (And I Feel Like...	4:22	Glenvegas	iTunes Holiday Sa...	Holiday		
Beated Promise Ring	3:46	Icon & Wine	Around the Wall (B...	Alternative		
Melody	3:52	Kate Earl	Kate Earl	Pop		
The Nutcracker, Op. 71, Act 2: C...	2:14	Kirov Orchestra &...	iTunes Holiday Sa...	Holiday		
Baby, It's Cold Outside	3:24	Lady Antebellum	iTunes Holiday Sa...	Holiday		
Love Comes And Goes	3:11	Lee Fields & The E...	My World	R&B/Soul		
Happiest Times	3:27	Little Big Adventure	The Hateful Eye - EP	Alternative		
Auld Lang Syne	2:52	The Lonesome Tra...	iTunes Holiday Sa...	Holiday		
Above the Northern Lights	4:33	Martheim Stearn...	iTunes Holiday Sa...	Holiday		
She Will Be Loved	4:17	Maroon 5	Songs About Jane	Pop		
It Snowed	3:40	Meaghan Smith	iTunes Holiday Sa...	Holiday		
Aloette	2:21	Meilinda Carol	Girl Scouts Greete...	Children's...		
Gimme Sympathy	3:35	Metric	Fantasies	Alternative		
Twelve Days of Christmas	4:16	Mexicali Marimba...	iTunes Holiday Sa...	Holiday		
Passion	3:37	Moby	Wait for Me	Electronic		
I Will Wait	4:37	Mumford & Sons	Babel	Alternative		
The Cave	3:36	Mumford & Sons	Sigh No More	Alternative		
O Holy Night	4:22	Musik Souchnid	iTunes Holiday Sa...	Holiday		

Nutrition Facts	
Serving Size 4 OZ. SERVING (112g)	
Servings Per Container VARIED	
Amount Per Serving	
Calories 170	Calories from Fat 70
% Daily Value*	
Total Fat 8g	12%
Saturated Fat 3g	15%
Cholesterol 65mg	22%
Sodium 70mg	3%
Total Carbohydrate 0g	0%
Dietary Fiber 0g	0%
Sugars 0g	
Protein 23g	
Vitamin A 0%	Vitamin C 0%
Calcium 0%	Iron 15%
*Percent Daily Values are based on a 2,000 calorie diet.	

CC image by USDAGov on Flickr

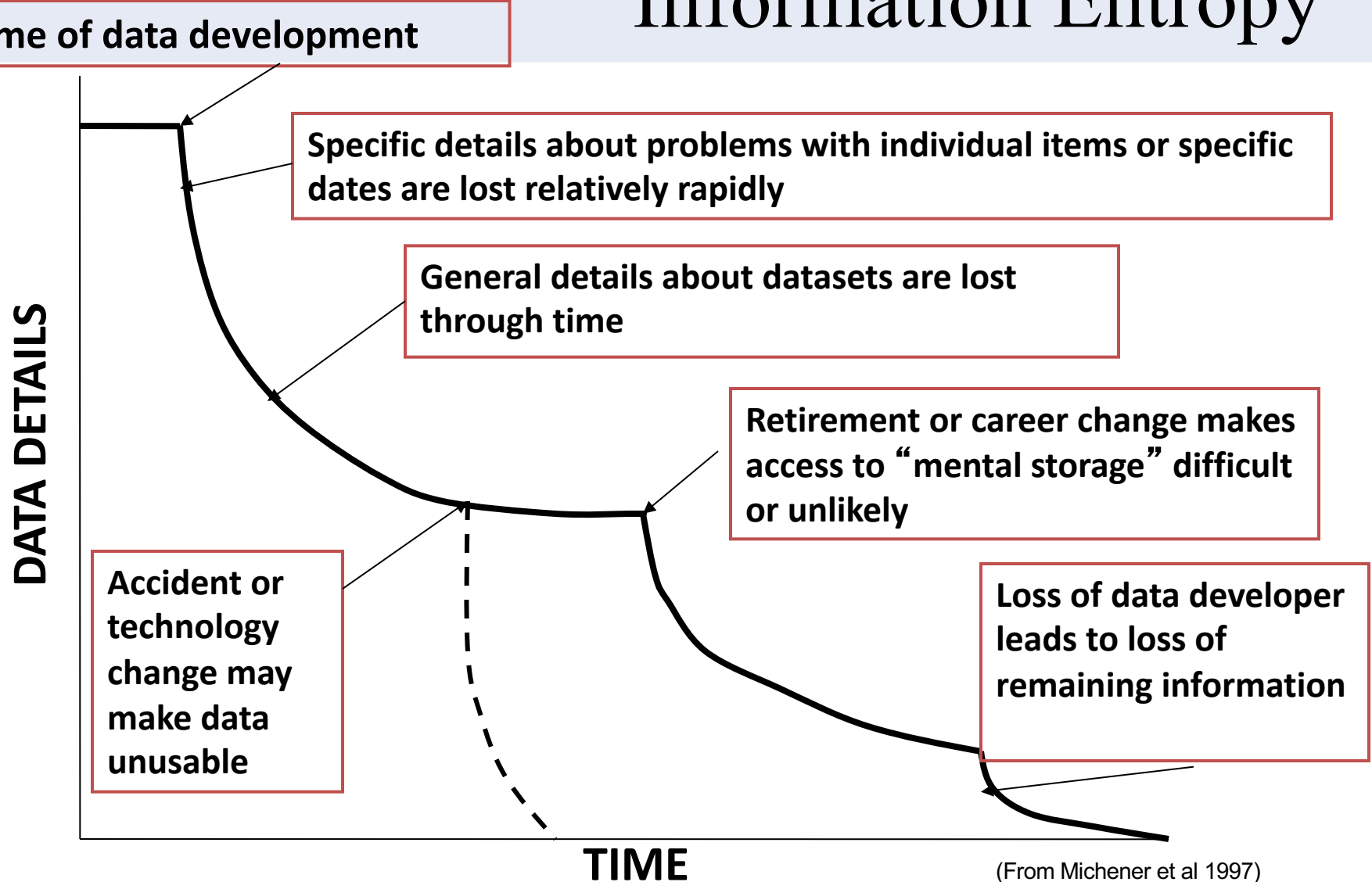
Author(s)	Boullosa, Carmen.
Title(s)	They're cows, we're pigs / by Carmen Boullosa
Place	New York : Grove Press, 1997.
Physical Descr	viii, 180 p ; 22 cm.
Subject(s)	Pirates Caribbean Area Fiction.
Format	Fiction

DataONE Education Module: Metadata. DataONE. Retrieved Nov 12, 2012. From http://www.dataone.org/sites/all/documents/L07_Metadata.pptx

Metadata

- What is it?
 - Structured information that describes a resource
- Why is it important?
 - Enables a resource or data to be easily discovered
 - Good metadata will help others understand and use your data

Information Entropy



Metadata Answers...

- **Who** created the data?
- **Who** maintains it?
- **When** were the data collected? **When** were they published?
- **Where** was it collected (geographic location)?
- **What** is the content of the data? The structure?
- **Why** were the data created?
- **How** were they produced/analyzed?

Metadata in Research

Project Documentation	Dataset Documentation
<ul style="list-style-type: none">• Context of data collection• Data collection methods• Structure, organization of data files• Data sources used• Data validation, quality assurance• Transformations of data from the raw data through analysis• Information on confidentiality, access and use conditions	<ul style="list-style-type: none">• Variable names and descriptions• Explanation of codes and schemas used• Algorithms used to transform data• File format and software (including version) used

Working with Data

- When you ***provide*** data to someone else, what types of information would you want to include with the data?



- When you ***receive*** a dataset from an external source, what types of details do you want to know about the data?

Critical Roles of Metadata

- **Data Discovery**
 - To be able to identify important data sets
- **Data Retrieval**
 - To know how and where to access data
- **Data Use**
 - To know enough details about how the data were collected and stored
- **Data Archiving**
 - Data can grow more valuable with time, but only if the critical information required to retrieve and interpret the data remains available

Metadata Formats

- Documentation for understanding & re-use
 - Readme File
 - Data Dictionary
 - Codebook
- Structured documentation in XML format for use in programs
 - DDI
 - FGDC
 - EML

Unstructured Documentation

- Data Dictionary
<http://people.virginia.edu/~sah/bse1/DataDefinitions.pdf>
- ReadMe File
http://libra.lib.virginia.edu/dataset_readme_template
- Dryad Example (lab notebook)
http://wiki.datadryad.org/wg/dryad/images/3/3d/DryadLab_example_readme.pdf
- Google Forms

Structured XML

Standard Schemes (XML)

- DDI– Data Document Initiative

<http://www.ddialliance.org/>

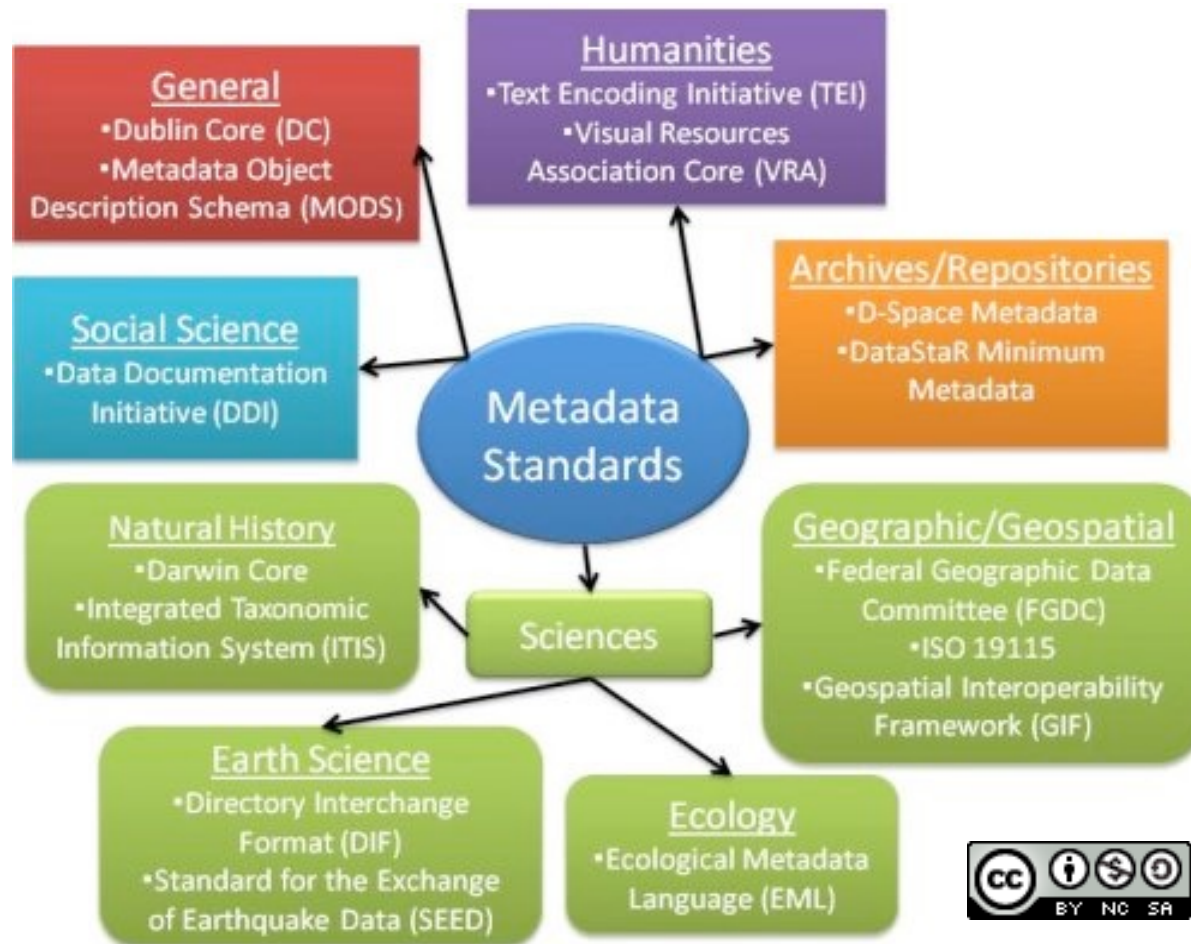
- FGDC– Geospatial Metadata Standard

<http://www.fgdc.gov/metadata/geospatial-metadata-standards>

- EML– Ecological Metadata Language

<http://knb.ecoinformatics.org/software/eml/>

Metadata Standards



Structured Metadata Tools

Tools

- Colectica add-on for Excel (DDI)
- Nesstar (DDI)
- Metavist (FGDC)
- Morpho (EML)

<http://dmconsult.library.virginia.edu/metadata-workshop/>

Colectica for Excel

- Excel Addin (DDI)
- Describes data files, variables, and code listings (metadata saved in the excel file)
- Import SPSS (.sav) & Stata (.dta) files into Excel, along with metadata
- Code books can also be customized and generated by the tool with various outputs

<http://www.colectica.com/software/colecticaforexcel>

Nesstar Publisher

- DDI Metadata Editor
- Creates codebooks

<http://www.nesstar.com/software/publisher.html>

Metavist

- Metadata editor for FGDC
- Includes fields for the Biological Data Profile

<http://metavist.djames.net/>

Morpho Data Management Software

- Creates EML metadata
- Create a catalog of data & metadata upon which to query, edit and view data collection
- Easy-to-use, cross-platform application for local and network access

<http://knb.ecoinformatics.org/morphoportal.jsp>

QUESTIONS?

Sherry Lake, Senior Data Consultant
shlake@virginia.edu

Bill Corey, Data Consultant
wtc2h@virginia.edu

Jeremy Bartczak, Intellectual Access & Metadata Services Librarian.
Health Sciences Library
jbartczak@virginia.edu

Data Management Consulting Group
<http://dmconsult.library.virginia.edu>