



What is metadata?



Metadata structure



Types of metadata



Conclusion

Metadata



Discovering Open Practices for Early Career Researchers
4 September 2014

Laurence Horton
Data Librarian, The London School of Economics and Political Science
L.Horton@lse.ac.uk

 This presentation is licensed under a Creative Commons Attribution 4.0 International license.





What is metadata?



Metadata structure



Conclusion



Types of metadata

Metadata



Discovering Open Practices for Early Career Researchers
4 September 2014

Laurence Horton
Data Librarian, The London School of Economics and Political Science
L.Horton@lse.ac.uk



This presentation is licensed under a Creative Commons Attribution 4.0 International license.



Metadata is data about data

Any object can have metadata. A datum (data) is data collection consisting of value, something digital

...it's what you need to have to create something

...it's the data, what, why, when, where and how (any) of data

If we can't read the story, the story is quickly forgotten

"Structured or semi-structured information which enables the creation, management, and use of records (i.e. data) through time and within and across domains"

Source: <http://www.structureddata.com>



Metadata can be structured

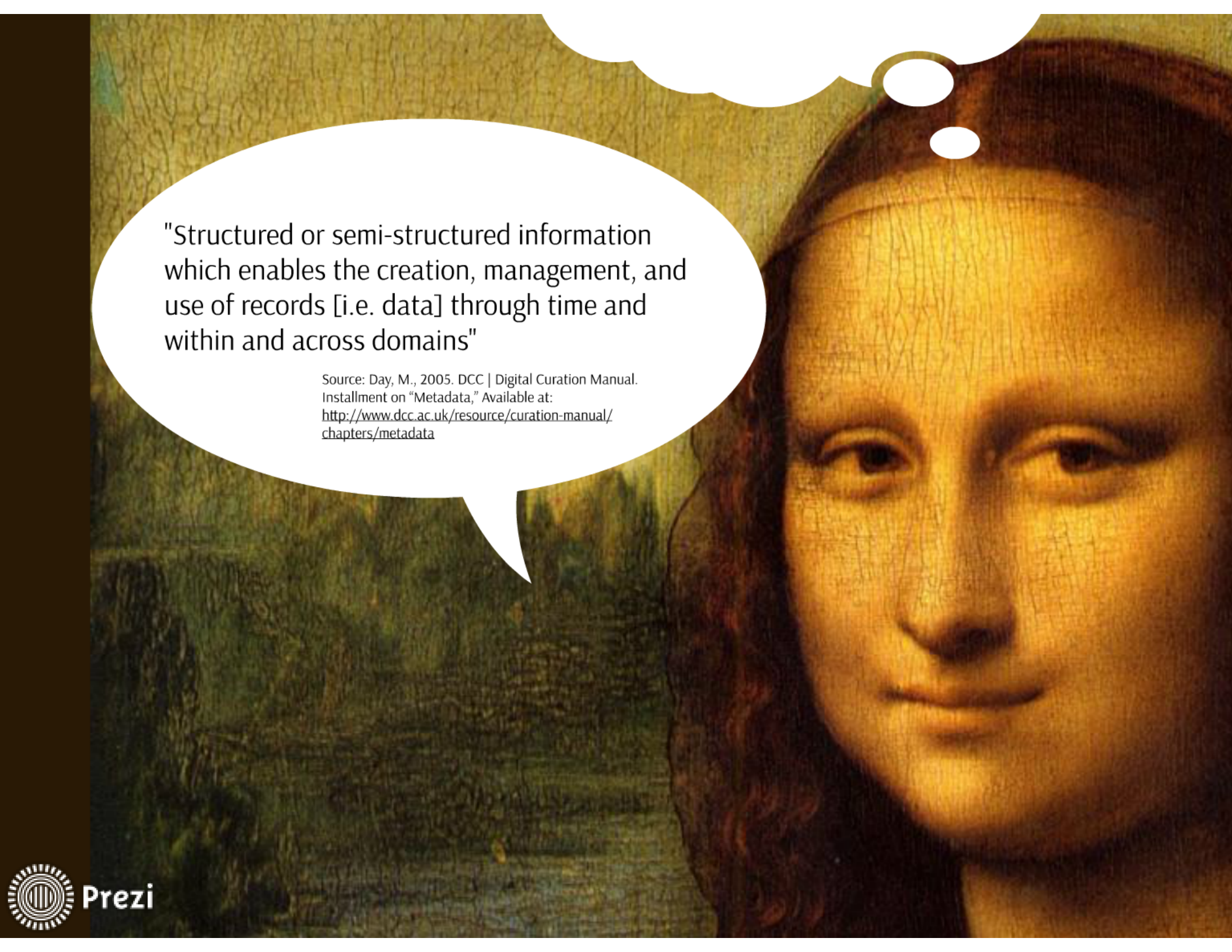
Metadata is structured which is organized by the owner for its use



Metadata



What is metadata?

The background of the slide is a close-up of the Mona Lisa painting, showing her face and upper torso. A white thought bubble is positioned above her head, and a larger white speech bubble is on the left side of the image.

"Structured or semi-structured information which enables the creation, management, and use of records [i.e. data] through time and within and across domains"

Source: Day, M., 2005. DCC | Digital Curation Manual. Installment on "Metadata," Available at: <http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata>

The background of the entire slide is a reproduction of the Mona Lisa painting. The woman's face is the central focus, with her enigmatic smile and the hazy, atmospheric landscape behind her. Several white thought bubbles and a speech bubble are overlaid on the image, containing text about metadata.

Metadata is data about data

Any object can have metadata. A datum, data, a data collection; something physical, something digital.

...it's what you need to know in order to understand something

...it's the who, what, why, where, when and how story of data.

If we can't read the story, the story is quickly forgotten.

structured information
creation, management, and
[data] through time and
remains"

M., 2005. DCC | Digital Curation Manual.
"Metadata," Available at:
<http://www.dcc.ac.uk/resource/curation-manual/metadata>

Metadata can be structured

Predefined or standardized fields to be populated by the researcher or archivist

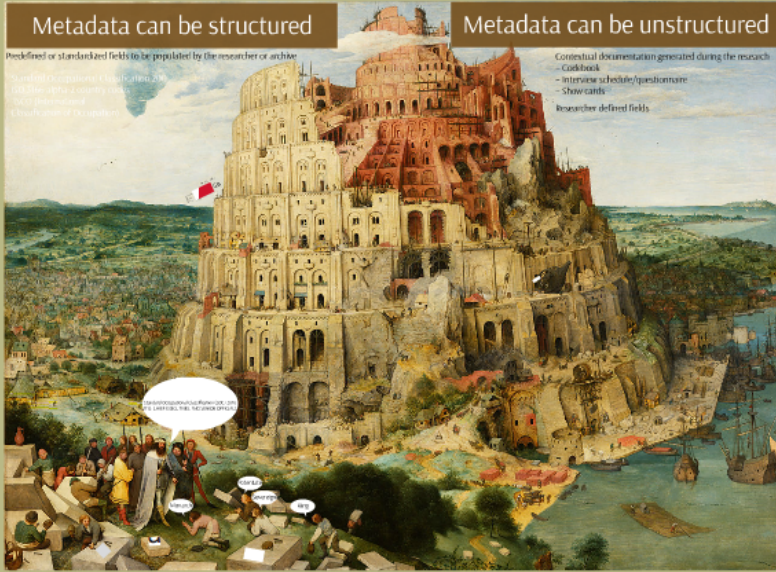
- Archival description (MARC, Dublin Core)
- Metadata standards (RDF, XML)
- Metadata schemas (Dublin Core, MARC)

Metadata can be unstructured

Contextual documentation generated during the research

- Checkbook
- Interview schedule/questionnaire
- Showcards

Researcher defined fields



Metadata structure

Metadata can be structured

Predefined or standardized fields to be populated by the researcher or archive

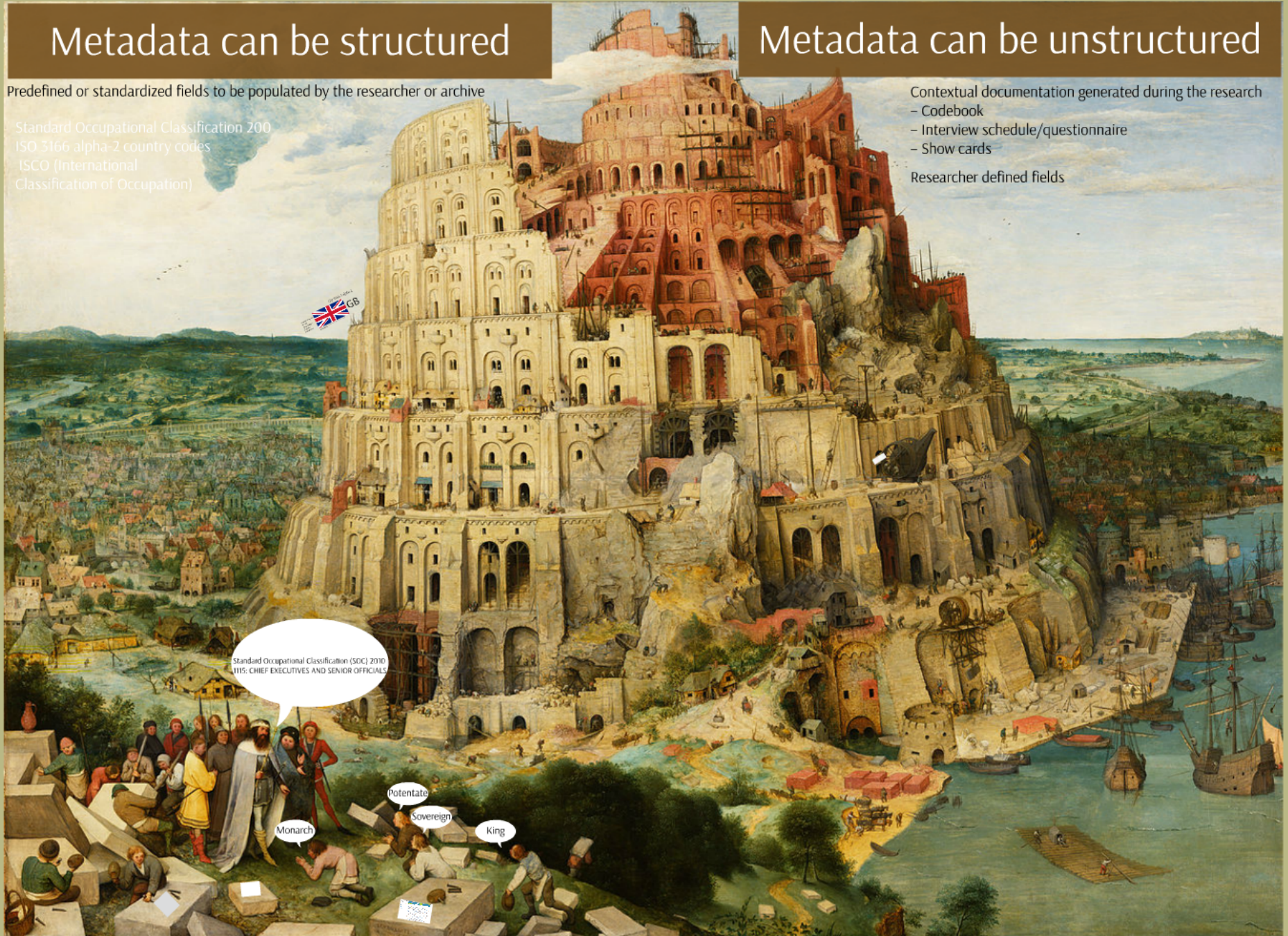
Standard Occupational Classification 200
ISO 3166 alpha-2 country codes
ISCO (International
Classification of Occupation)

Metadata can be unstructured

Contextual documentation generated during the research

- Codebook
- Interview schedule/questionnaire
- Show cards

Researcher defined fields



"Pieter Bruegel the Elder - The Tower of Babel (Vienna) - Google Art Project - edited" by Pieter Bruegel the Elder (1526/1530-1569) - Levels adjusted from File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project.jpg, originally from Google Art Project.. Licensed under Public domain via Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_\(Vienna\)_-_Google_Art_Project_-_edited.jpg#mediaviewer/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_\(Vienna\)_-_Google_Art_Project_-_edited.jpg](http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project_-_edited.jpg#mediaviewer/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-_Google_Art_Project_-_edited.jpg)



Prezi

Metadata can be structured

Predefined or standardized fields to be populated by the researcher or archive

Standard Occupational Classification 200

ISO 3166 alpha-2 country codes

ISCO (International
Classification of Occupation)



ISO 3166-1 alpha-2



GB

United Kingdom
Britain
Great Britain
England
British Isles

"Flag of the United Kingdom". Via Wikipedia - http://en.wikipedia.org/wiki/File:Flag_of_the_United_Kingdom.svg#mediaviewer/File:Flag_of_the_United_Kingdom.svg



Standard Occupational Classification (SOC) 2010
1115: CHIEF EXECUTIVES AND SENIOR OFFICIALS

Monarch

Potentate

Sovereign

King

Pieter Bruegel the Elder - The Tower of Babel (Vienna) - Google Art Project - edited" by Pieter Brueghel the Elder (1526/1530-1569) - Levels adjusted under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_-_Google_Art_Project_-_edited.jpg





Monarch

Potentate

Sovereign

der - The Tower of Babel (Vienna) - Google Art Project - edited" by Pieter Bruegel the
Prezi Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_



Benefits of DDI

Last Updated: Mon, 2009-12-07 18:42 — edwardsm

The DDI facilitates:

- **Interoperability.** Codebooks marked up using the DDI specification can be exchanged and transported seamlessly, and applications can be written to work with these homogeneous documents.
- **Richer content.** The DDI was designed to encourage the use of a comprehensive set of elements to describe social science datasets as completely and as thoroughly as possible, thereby providing the potential data analyst with broader knowledge about a given collection.
- **Single document - multiple purposes.** A DDI codebook contains all of the information necessary to produce several different types of output, including, for example, a traditional social science codebook, a bibliographic record, or SAS/SPSS/Stata data definition statements. Thus, the document may be repurposed for different needs and applications. Changes made to the core document will be passed along to any output generated.
- **On-line subsetting and analysis.** Because the DDI markup extends down to the variable level and provides a standard uniform structure and content for variables, DDI documents are easily imported into on-line analysis systems, rendering datasets more readily usable for a wider audience.
- **Precision in searching.** Since each of the elements in a DDI-compliant codebook is tagged in a specific way, field-specific searches across documents and studies are enabled. For example, a library of DDI codebooks could be searched to identify datasets covering protest demonstrations during the 1960s in specific states or countries.

- Benefits of DDI
- DDI projects

Source: DDI Alliance (2009) "Benefits of DDI" <http://www.ddialliance.org/ddi-at-work/benefits>

TT15: CHIEF EXECUTIVES AND SENIOR OFFICIALS



Bruegel the Elder - The Tower of Babel (Vienna) - Google Art Project - edited" by Pieter Bruegel the Elder (1563-1638) - Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_-_Google_Art_Project_-_edited.jpg

What is a DOI?



A DOI is a Digital Object Identifier. That is a **Digital Identifier** of an **Object**, not an **Identifier** of a **Digital Object**.

10.1234/456-mydoc-456584893489 is an example of the DOI format. It consists of a **Prefix** and a **Suffix**. The prefix always begins '10' indicating this is a DOI name; the second part of the prefix is typically four digits, and is allocated to the DOI assigner. The suffix is created by the assigner, and can be any length and incorporate other numbering schemes (including an existing identifier) if desired. Existing data can also be linked in the DOI record.

A DOI can be assigned to any **Object**. An object can be anything you want to manage: a piece of online content in a particular version, or a related object, such as an author or a party to an agreement; or a physical asset, a DVD or an item of equipment, for example.

DOIs can be assigned at different levels of granularity to suit differing needs, e.g. a collection, a book, a table within the book, the related source data for the table etc. When assigning DOIs, the registrant links metadata with an appropriate level of detail to identify a separate entity within the DOI system.

DOIs in use

- First applications launched in 2000
- Currently being assigned by well over 5000 naming authorities e.g. publishers, science data centres, movie studios etc
- Around 100 million DOI names assigned
- Over 1.5 billion DOI resolutions per year





Elder - The Tower of Babel (Vienna) - Google Art Project - edited" by Pieter Brueghel the
in Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the
ect - edited.png](http://commons.wikimedia.org/wiki/File:Pieter_Bruegel_the_Elder_-_The_Tower_of_Babel_(Vienna)_-edited.png)

Search



English

ORCID

Connecting Research
and Researchers

FOR RESEARCHERS

FOR ORGANIZATIONS

ABOUT

HELP

SIGN IN

WHAT IS ORCID?

THE ORCID TEAM

THE ORCID COMMUNITY

MEMBERSHIP

NEWS

EVENTS

OUR MISSION

OUR PRINCIPLES

OUR GOVERNANCE

OUR POLICIES

- [PRIVACY POLICY](#)

As researchers and scholars, you face the ongoing challenge of distinguishing your research activities from those of others with similar names. You need to be able to easily and uniquely attach your identity to research objects such as datasets, equipment, articles, media stories, citations, experiments, patents, and notebooks. As you collaborate across disciplines, institutions and borders, you must interact with an increasing number and diversity of research information systems. Entering data over and over again can be time-consuming, and often frustrating.

ORCID is an open, non-profit, community-driven effort to create and maintain a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. ORCID is unique in its ability to reach across disciplines, research sectors and national boundaries. It is a hub that connects researchers and research through the embedding of ORCID identifiers in key workflows, such as research profile maintenance, manuscript submissions, grant applications, and patent applications.

ORCID provides two core functions: (1) a registry to obtain a unique identifier and manage a record of activities, and (2) APIs that support system-to-system communication and authentication. ORCID makes its code available under an open source license, and will post an annual public data file under a CC0 waiver for free download.

The ORCID Registry is available free of charge to individuals, who may obtain an ORCID identifier, manage their record of activities, and search for others in the Registry. Organizations may become members to link their records to ORCID identifiers, to update ORCID records, to receive updates from ORCID, and to register their employees and students for ORCID identifiers.

ORCID records hold non-sensitive information such as name, email, organization and research activities. ORCID understands the fundamental need for individuals to control how their data are shared, and provides tools to manage data privacy. We take steps to protect your information, consistent with the principles set forth in our Privacy Policy, which are intended to comply with the Safe Harbor Principles issued by the U.S. Department of Commerce.

Source: ORCID (2014) "What is ORCID?" <https://orcid.org/content/initiative>



Small white text box with illegible content, possibly a watermark or a small note.

Unstructured metadata at work....



John Roberts @john_roberts

25 Mar 12

Few serious attempts at implementation of **#digital_preservation** - esp with long-term aspirations **#fp2012**



Annette Strauch @annettestr

5h

@asist_org @nytimes interesting article about the right to forget in **#digitalpreservation** and **#selectionprocesses**



Rick Cockram @le_re

11 Nov 11

"Sharing knowledge: open access and **preservation** in Europe" _ ec.europa.eu/research/scien_____ **#dp** **#openaccess** **#EC** **#ERA**



APARSEN @APARSENproject

12 Jun

13Jun:**#APARSEN** Webinar: Digital Preservation Cost Models and Sustainability 3pmCEST _bit.ly/10LIFh5 **#digpres** **#costmodels**

Source: Jessica Trixa (2014) "Metadata and documentation" CEESDA Training, licensed under a Creative Commons Attribution 3.0 Unported License.



Question 8: Does the research you are currently working on have a data management plan?
* If you are working on more than one project or theme, answer for the one that takes most of your time* Required response. Select one:

dmp

4
0
4
5
0
1

- 0 I don't know what a data management plan is
- 1 I don't know if my project has a data management plan
- 2 My project does not have a data management plan
- 3 A data management plan was written as part of the bid but has not been updated
- 4 My project has a data management plan that covers the duration of the research
- 5 My project has a data management plan that extends beyond completion of the research

Object level

Collection level

Digital Communication Enhancement (DICE) project.

24 variables

167 respondents

13-24 March 2012

The JISC funded DICE project ran from November 2011 until July 2012 with the intention of raising the awareness of library staff and researchers to the increasing importance attached to the preservation of research material and data, not just publications.

Staff and PG students at LSE

"Creación de Adán (Miguel Ángel)" by Michelangelo Buonarroti - See below.. Licensed under Public domain via Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Creaci%C3%B3n_de_Ad%C3%A1n_\(Miguel_%C3%81ngel\).jpg#mediaviewer/File:Creaci%C3%B3n_de_Ad%C3%A1n_\(Miguel_%C3%81ngel\).jpg](http://commons.wikimedia.org/wiki/File:Creaci%C3%B3n_de_Ad%C3%A1n_(Miguel_%C3%81ngel).jpg#mediaviewer/File:Creaci%C3%B3n_de_Ad%C3%A1n_(Miguel_%C3%81ngel).jpg)

Question 8: Does the research you are currently working on have a data management plan?
* if you are working on more than one project or theme, answer for the one that takes most of your time * Required response. Select one.

dmp

4
0
4
5
0
1

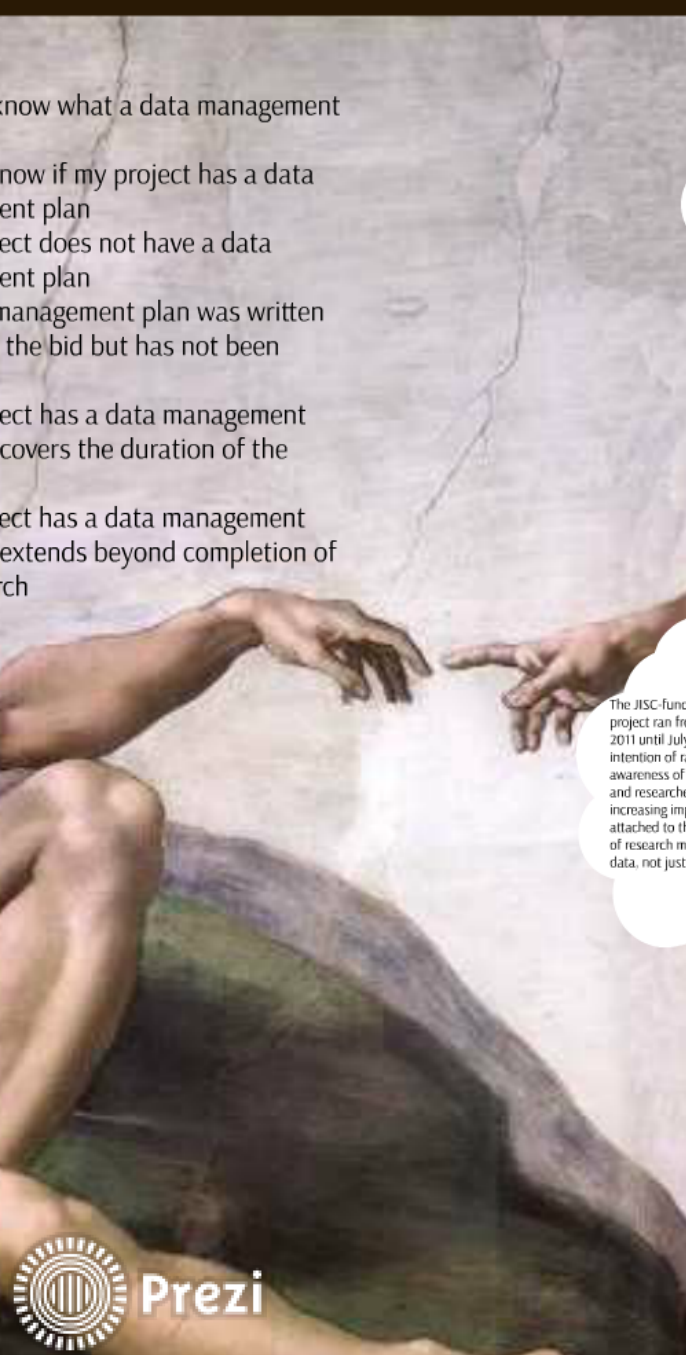
- 0 I don't know what a data management plan is
- 1 I don't know if my project has a data management plan
- 2 My project does not have a data management plan
- 3 A data management plan was written as part of the bid but has not been updated
- 4 My project has a data management plan that covers the duration of the research
- 5 My project has a data management plan that extends beyond completion of the research

Question 8: Does the research you are currently working on have a data management plan?
* if you are working on more than one project or theme, answer for the one that takes most of your time * Required response. Select one.

dmp

4
0
4
5
0
1

- 0 I don't know what a data management plan is
- 1 I don't know if my project has a data management plan
- 2 My project does not have a data management plan
- 3 A data management plan was written as part of the bid but has not been updated
- 4 My project has a data management plan that covers the duration of the research
- 5 My project has a data management plan that extends beyond completion of the research



Collection level

know what a data management
ent plan
ect does not have a data
ent plan
management plan was written
the bid but has not been
ect has a data management
covers the duration of the
ect has a data management
extends beyond completion of
rch

Digital Communication
Enhancement (DICE) project.

24 variables

167 respondents

13-24 March
2012

Staff and PG students at LSE

The JISC-funded DICE
project ran from November
2011 until July 2012 with the
intention of raising the
awareness of library staff
and researchers to the
increasing importance
attached to the preservation
of research material and
data, not just publications.

Collection level

Digital Communication Enhancement (DICE) project.

24 variables

167 respondents

13-24 March 2012

Staff and PG students at LSE

The JISC-funded DICE project ran from November 2011 until July 2012 with the intention of raising the awareness of library staff and researchers to the increasing importance attached to the preservation of research material and data, not just publications.

Collection level

Digital Communication Enhancement (DICE) project.

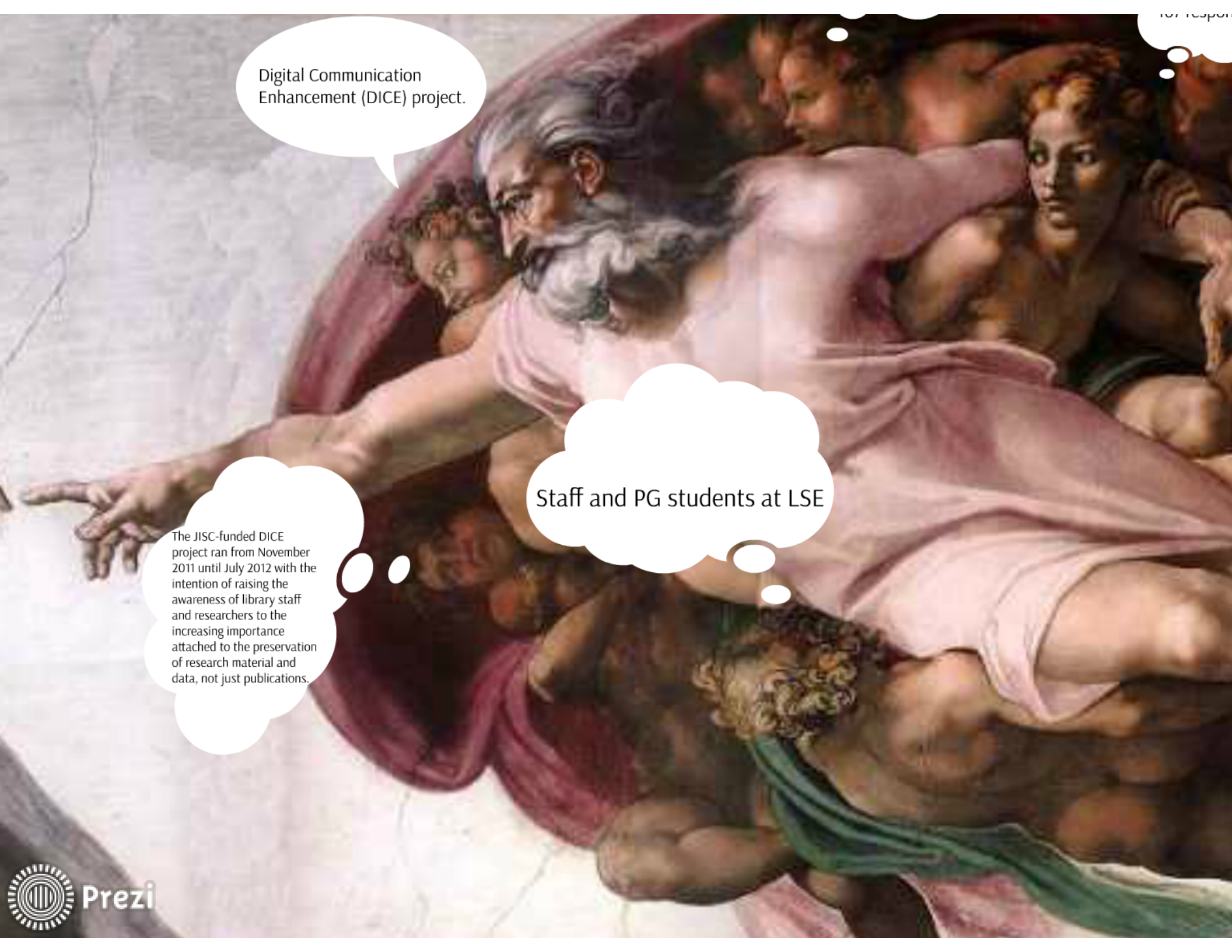
24 variables

167 respondents

13-24 March 2012

Staff and PG students at LSE

The JISC-funded DICE project ran from November 2011 until July 2012 with the intention of raising the awareness of library staff and researchers to the increasing importance attached to the preservation of research material and data, not just publications.



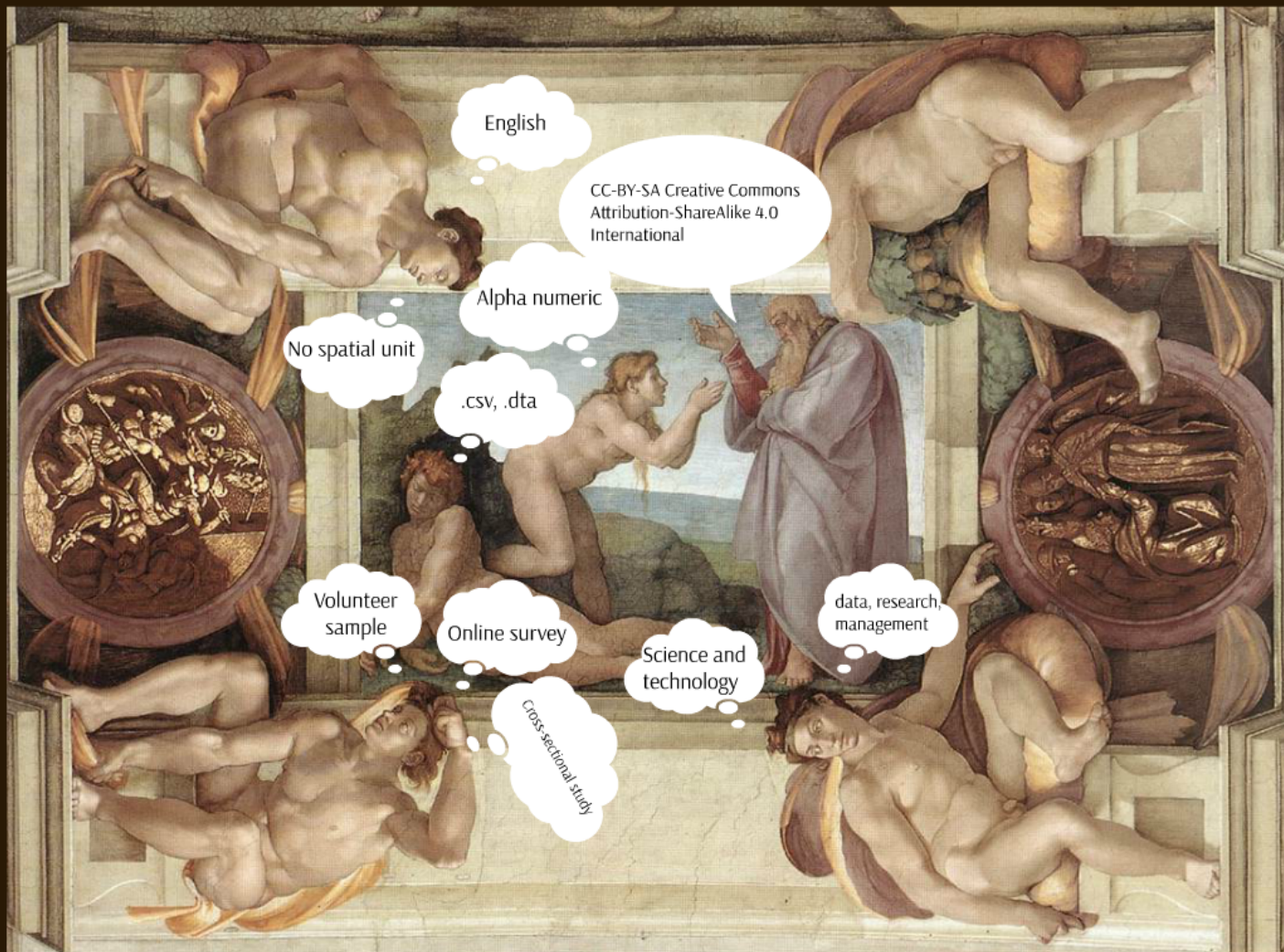
Digital Communication Enhancement (DICE) project.

The JISC-funded DICE project ran from November 2011 until July 2012 with the intention of raising the awareness of library staff and researchers to the increasing importance attached to the preservation of research material and data, not just publications.

Staff and PG students at LSE



"Creación de Adán (Miguel Ángel)" by Michelangelo Buonarroti - See below. Licensed under Public domain via Wikimedia Commons - [http://commons.wikimedia.org/wiki/File:Creaci%C3%B3n_de_Ad%C3%A1n_\(Miguel_3C3%A1ngel\).jpg#mediaviewer/File:Creaci%C3%B3n_de_Ad%C3%A1n_\(Miguel_3C3%A1ngel\).jpg](http://commons.wikimedia.org/wiki/File:Creaci%C3%B3n_de_Ad%C3%A1n_(Miguel_%C3%A1ngel).jpg#mediaviewer/File:Creaci%C3%B3n_de_Ad%C3%A1n_(Miguel_3C3%A1ngel).jpg)



English

CC-BY-SA Creative Commons Attribution-ShareAlike 4.0 International

Alpha numeric

No spatial unit

.csv, .dta

Volunteer sample

Online survey

Science and technology

data, research, management

Cross-sectional study

"Michelangelo, Creation of Eve 00" by Michelangelo Buonarroti - Web Gallery of Art. Image Info about artwork. Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Michelangelo_Creation_of_Eve_00.jpg#mediaviewer/File:Michelangelo_Creation_of_Eve_00.jpg



What is metadata?



Metadata structure



Conclusion



Types of metadata

Metadata



Discovering Open Practices for Early Career Researchers
4 September 2014

Laurence Horton
Data Librarian, The London School of Economics and Political Science
L.Horton@lse.ac.uk



This presentation is licensed under a Creative Commons Attribution 4.0 International license.

atured

the research



"Gustav Klimt, 016" by Gustav Klimt - <http://www.bethedore.nl/en/swim/ingen/bethedore/jugendstil/and-avener-successor/gustav-klimt>. Licensed under Public domain via Wikimedia Commons - http://commons.wikimedia.org/wiki/File:Gustav_Klimt_-_016.jpg#mediaviewer/File:Gustav_Klimt_-_016.jpg

Conclusion

What information do I need to understand, evaluate, and build upon data without having to ask for additional information from the author?

**METADATA IS A
LOVE NOTE
TO THE FUTURE**

"Metadata is a love note to the future" by Playing Futures: Applied Nomadology <https://www.flickr.com/photos/centralasian/8071729256/>
available under CC-BY 2.0 license <https://creativecommons.org/licenses/by/2.0/>

Are there existing standards I can use to collect or structure that information?