HELMHOLTZ METADATA COLLABORATION

Materials Data Science & Informatics IAS-9

Scientific metadata: Fundamentals of structured and standardized research data annotation

Annika Strupp // Silke Gerlich

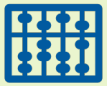
Institute for Materials Data Science and Informatics, Forschungszentrum Jülich //

www.helmholtz-metadata.de

Annika


Consultant web analytics


Master student
Digital Data Management


Data science training


Data Steward @ HMC
- Training & Outreach

Who we are

Annika

- Consultant web analytics
- Master student Digital Data Management
- Data science training
- Data Steward @ HMC
  - Training & Outreach

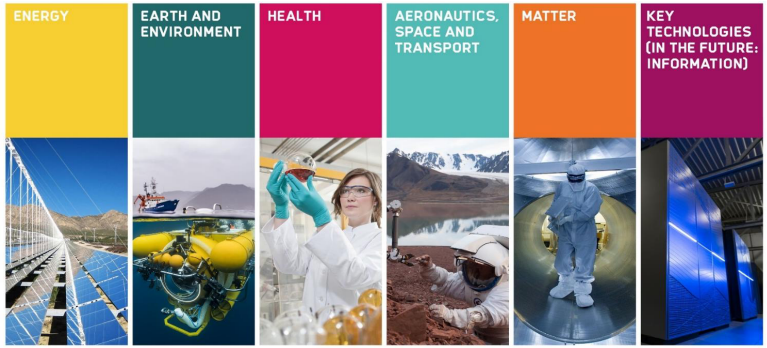
Silke

- PhD molecular plant sciences (University of Cologne)
- Coordinator for academic education
- Data science training
- Postdoc @ HMC
  - Ontology development
  - Training & outreach

What is HMC?

HMC | HELMHOLTZ METADATA COLLABORATION

- we support researchers & infrastructures to make HGF data FAIR
- we work across scientific boundaries to provide comprehensive and sustainable services, consulting, information and tools for metadata handling.
- we are located at 6 different locations in the Helmholtz Network



HMC @ JÜLICH Forschungszentrum

- team of data stewards, software developers and domain scientists from various disciplines
- located at IAS 9 Materials Data Science & Informatics
- collaborating with scientists, administration and infrastructure providers on various metadata projects
- always happy to discuss metadata:

  HMC@fz-juelich.de & Hub Information Webpage

Data & Metadata

- What is data?
- What is metadata and why is it important?
- How can metadata be classified?
- Where can we find metadata?

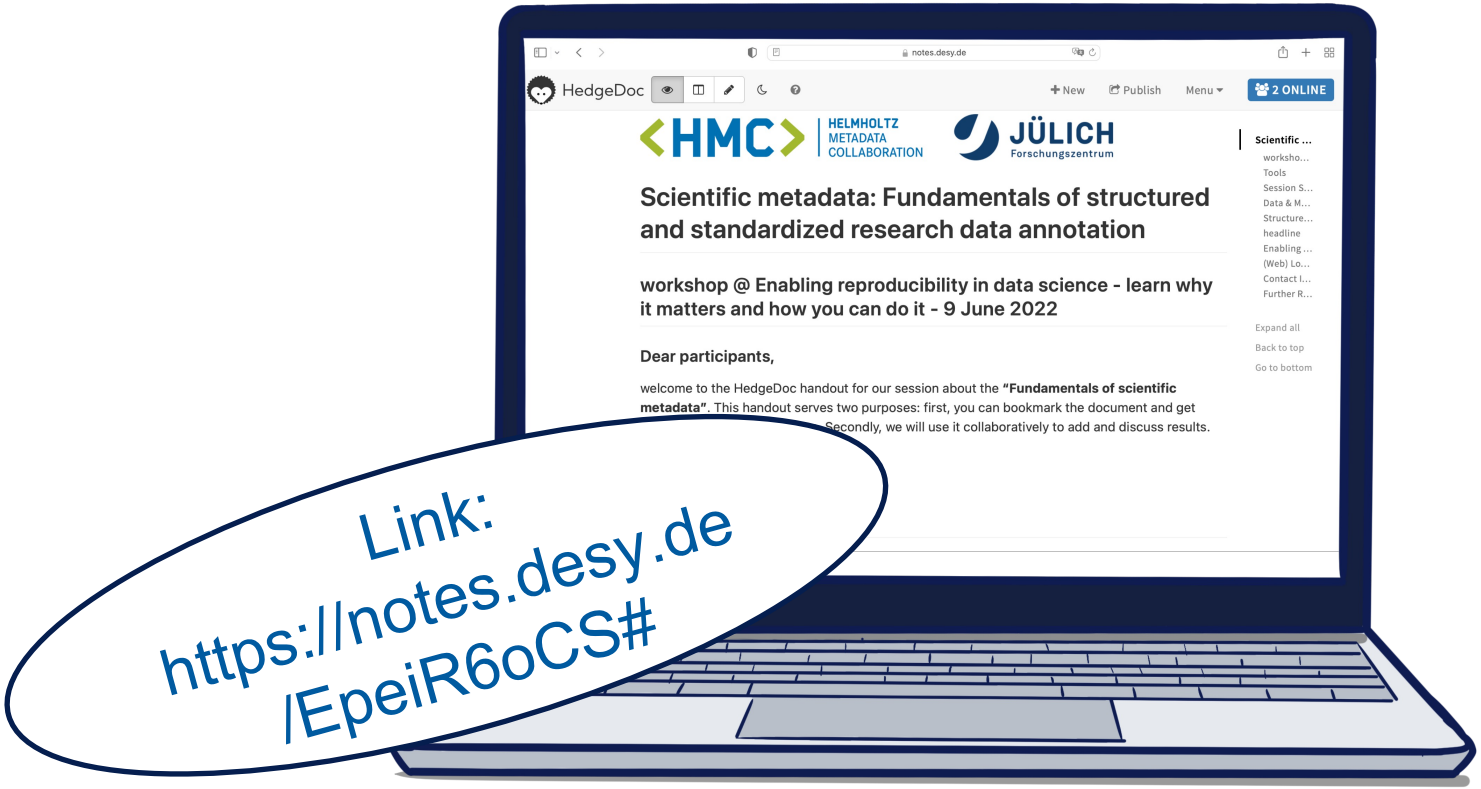
Structure & Schema

- What is structured metadata?
- What are the benefits of structured metadata?
- How do you record metadata in a structured way?
- What is a schema and how does it help to record metadata?

Enabling technology & standards

- What are the benefits of structured metadata?
- How are structured metadata applied in a linked world?
- What are metadata standards?
- How do I find appropriate standards for my research?

The Handout


notes.desy.de
HedgeDoc
New
Publish
Menu
2 ONLINE
HELMHOLTZ METADATA COLLABORATION
JÜLICH Forschungszentrum
Scientific metadata: Fundamentals of structured and standardized research data annotation
workshop @ Enabling reproducibility in data science - learn why it matters and how you can do it - 9 June 2022
Dear participants,
welcome to the HedgeDoc handout for our session about the "Fundamentals of scientific metadata". This handout serves two purposes: first, you can bookmark the document and get
Secondly, we will use it collaboratively to add and discuss results.
Scientific ...
worksho...
Tools
Session S...
Data & M...
Structure...
headline
Enabling ...
(Web) Lo...
Contact I...
Further R...
Expand all
Back to top
Go to bottom
Link:
https://notes.desy.de
/EpeiR6oCS#

Let‘s get to know each other!

Let’s type a small JSON metadata record about ourselves and the cities we live in 😀.

Copy the example below, paste it to text field ↗ **here** and fill in your values.

Example:

```
{
  "firstName": "value",
  "ORCID": "value",
  "researchField": "value",
  "currentPositon": "value",
  "favoriteCake": "value",
  "hobbies": ["value", "value"]
  "city": {
    "name": "value",
    "url": "value"
  }
}
```

Part 1: Data & Metadata

What are data?

What is data?

This is data


fPxsydzampgdVtbY3oc77GA53MD7yCthr4Vw3vb9LmACZ59cayrvSB/Gjy98Nx/wXmtX/qe8q5QX
o3lZmVdbCR2/G+/zRW5jvunBzRc4NR7eaGPCBPMn4+C42Prmcy1WlRuJvZWVuKR7cul7vbOut2A+
fbGN9x7YTPCK26z3N+ojVGQ8O/fx9Xe/FwZOnOu7F/aH7bpLzbMlDCb1/aSvEObaxNFwtf684EbJ
DQMC8tL8GhbEG3ztue7Ee48H18b6B/fIOswX235rT0+7WxNc+QoQkTbQmdfb5hO5KRagzD7mYFEL
2/8G4ADWUMCWGnMPBWSeDRgCVEY2at2gw5u/tQNbBKf7CCBvmvW99zf44Dne6w3xxdwXVAj4aCW8
885P+zJv2bgGoWLhzCPg/psNB7gCOw/CA3VYaXwzt/ldG9fMdcLAFraKpb2Vd8TKC7u9n8Z77wRd
a9wqrA8nxHbLE7Z2hPvwI2O9X77qGHd7Flv6xvIKaT3fWw4HXEyxhQ/kdoGBlzUHHDGTAYHD8MPz
YX0AAUzzyk/iance34t3Y3itL4IRcRFQYEFQLx9R6SFAgDc4wQNBOv7eDFSJiyxYImuEq+IxceDM
wGSu3/e9xKCbiyRuhv2tgbONCY9NbSZsx9iX2spv7GA0nktEfa6o18BPsXIup44Onw0VNJv419Pj
WithKPHGjDbxVlDHX1fks/dgMfTwueKF87LzoBCvJjKCUgkTFiVKZIUDtlRinvDlm29XvIWoVVkL
2XZ9ayYSsHb1Cu42HC2EDs/+WPaVueWPL7cPSpKmAJ7egmFgwPhEfAgtBTQiWHDfC2r47H41EDkD
SvtgdOwl5RfK3QkNzeXFFb4B/n4rvDAPXjIykN6eh+gDnkIoM2iN0+INn9fJZhHn8JfXWLDfG0R7
X8xlv5m4G9aNZwArb8KMvloW4QpkfQs+N955iZog1lw3V05MIy5ALZ+7YU4rGNAIykkvfefTxdcJ
TVgdq/u4IvgOixKhNejDxi4Daiqsrook1o6w8P4BsLDj/KlDRPCFUbjDe6QwuewWK6yhGqxyGldI
tfeHOEHM2IepBdYUrrcTIRthFOAllQAH+g1XeW6W7I18MHGkbgMJgTZdqRZgrHToE3xCLvEJrp8E
KiAbM+bxNtDALbgTdsYOyA/aLGHAw2CCCYO7CA33h5j6+GDQ2tjUifOJiFq+D2sjQlTwObYNe41Y
TXcj8ZkvYZwzsiuQznQNAghsrpa7twd7BkEwwNJUVgFQZa8A3wF94bOWzAM1jcN3zDOBFRPUG2nG
q+M5uCwwCwX8mpbM/7IfmyjNZfE/WOxXJCdANcTwcMoM7qIJIS/f+tyrCzcdG/gspWIwcc0bhIVi
9F4hJItFg3Hwie8gBuBUiZcSSmCqvGYS6MpmB+91sfZ4AJKSCwCl4FWyz3vimTcUBRuMCbIEvQjQ
SKwW6+sPqENQBD0n/tXhYvUC1Td8DlIBLUjf9xQF4Au7+io0j1CQNtvCf0MPhIJBGkMnRtY5440t
x8b/zAtIhtK/xKYmeBIjif/PC1GbWCpuzluVsVeaBaGGnUCt45wlD7Aa5CrcADCGOMZh/WBiU2cD
sMneQCL+v8JmYb6EzztCL+DnETkAyiHgj8+wUgQwZUZ/wvUoLKoglth1iFaJI8Bww4uDxCB7IeIQ
FzGMDYFn5dZ674yRYlmwanYTolCuQ44I/N3NCoSNdlZPes0VsOzECLhbaXesI8K22UjABbmBrknY
OyoTwEKLBGIfEPMaFHaaPUsPZ+tQLTEK3MKPE1x7YdYQkzyhr4tgChuTrGESwg+W3fgr34WdymoS
4SF84xuExyqUYtCw3a+CKZhEuIE6LALL+ADBYMIhD3TBBVPlYjUiXKMXY0bR+cF+aBOEAQtlcbBg
Y65BGuAfZRPF8dJFTIHorJkxSNaB6HLyYrODyWxlwcRrNt4k1/9YPn4teQd1cbceUEKmW2SrxAS2
v1c5NursUZV9EelRv/ws6HTlPvcmiHB9QMFqb+aWXq0K5qVx4OwfAWkVVPaHdIM7t00YFhoBZZg/
DvxIv8pGheT58JbQ3441EGnBRfwVAgxg880EJsV5PRVML7Ow2aZAWA7ESpqEf6gYTkBABa3nswnM

This is data


fPxsydzampgdVtbY3oc77GA53MD7yCthr4Vw3vb9LmACZ59cayrvSB/Gjy98Nx/wXmtX/qe8q5
o3lZmVdbCR2/G+/zRW5jvunBzRc4NR7eaGPCBPMn4+C42Prmcy1WlRuJvZWVuKR7cul7vbOu
fbGN9x7YTPCK26z3N+ojVGQ8O/fx9Xe/FwZOnOu7F/aH7bpLzbMlDCb1/aSvEObaxNFwtf
DQMC8tL8GhbEG3ztue7Ee48H18b6B/fIOswX235rT0+7WxNc+QoQkTbQmdfb5hO5KRaf
2/8G4ADWUMCWGnMPBWSeDRgCVEY2at2gw5u/tQNbBKf7CCBvmvW99zf44Dne6w3xxdwXV
885P+zJv2bgGoWLhzCPg/psNB7gCOw/CA3VYaXwzt/ldG9fMdcLAFraKpb2Vd8TC7u9
a9wqrA8nxHbLE7Z2hPvwI2O9X77qGHd7Flv6xvIKaT3fWw4HXEyxhQ/kdoGBl
YX0AAUzzyk/iance34t3Y3itL4IRcRFQYEFQLx9R6SFAgDc4wQNBOv7eDFS
wGSu3/e9xKCbiyRuhv2tgbONCY9NbSZsx9iX2spv7GA0nktEfa6o18BBs
WithKPHGjDbxVlDHX1fks/dgMfTwueKF87LzoBCvJjKCUgkTFiVKZI
2XZ9ayYSsHb1Cu42HC2EDs/+WPaVueWPL7cPSpKmAJ7egmFgwPhE
SvtgdOwl5RfK3QkNzeXFFb4B/n4rvDAPXjIykN6eh+gDnkIoM
X8xlv5m4G9aNZwArb8KMvloW4QpkfQs+N955iZog1lw3V05
TVgdq/u4IvgOixKhNejDxi4Daiqsrook1o6w8P4BsLDj
tfeHOEHM2IepBdYUrrcTIRthFOAllQAH+g1XeW6W7I
KiAbM+bxNtDALbgTdsYOyA/aLGHAw2CCCYO7CA33
TXcj8ZkvYZwzsiuQznQNAghsrpa7twd7BkEwwN
q+M5uCwwCwX8mpbM/7IfmyjNZfE/WOxXJCdAN
9F4hJItFg3Hwie8gBuBUiZcSSmCqvGYS6Mpm
SKwW6+sPqENQBD0n/tXhYvUC1Td8DlIBLUjf
x8b/zAtIhtK/xKYmeBIjif/PC1GbWCpuzluV
sMneQCL+v8JmYb6EzztCL+DnETkAyiHgj8+wU
FzGMDYFn5dZ674yRYlmwanYTolCuQ44I/N3NCoSNd
OyoTwEKLBGIfEPMaFHaaPUsPZ+tQLTEK3MKPE1x7YdYQkzyhr4tgChuTrGESwg+W3fgr34WdymoS
4SF84xuExyqUYtCw3a+CKZhEuIE6LALL+ADBYMIhD3TBBVPlYjUiXKMXY0bR+cF+aBOEAQtlcbBg
Y65BGuAfZRPF8dJFTIHorJkxSNaB6HLyYrODyWxlwcRrNt4k1/9YPn4teQd1cbceUEKmW2SrxAS2
v1c5NursUZV9EelRv/ws6HTlPvcmiHB9QMFqb+aWXq0K5qVx4OwfAWkVVPaHdIM7t00YFhoBZZg/
DvxIv8pGheT58JbQ3441EGnBRfwVAgxg880EJsV5PRVML7Ow2aZAWA7ESpqEf6gYTkBABa3nswnM

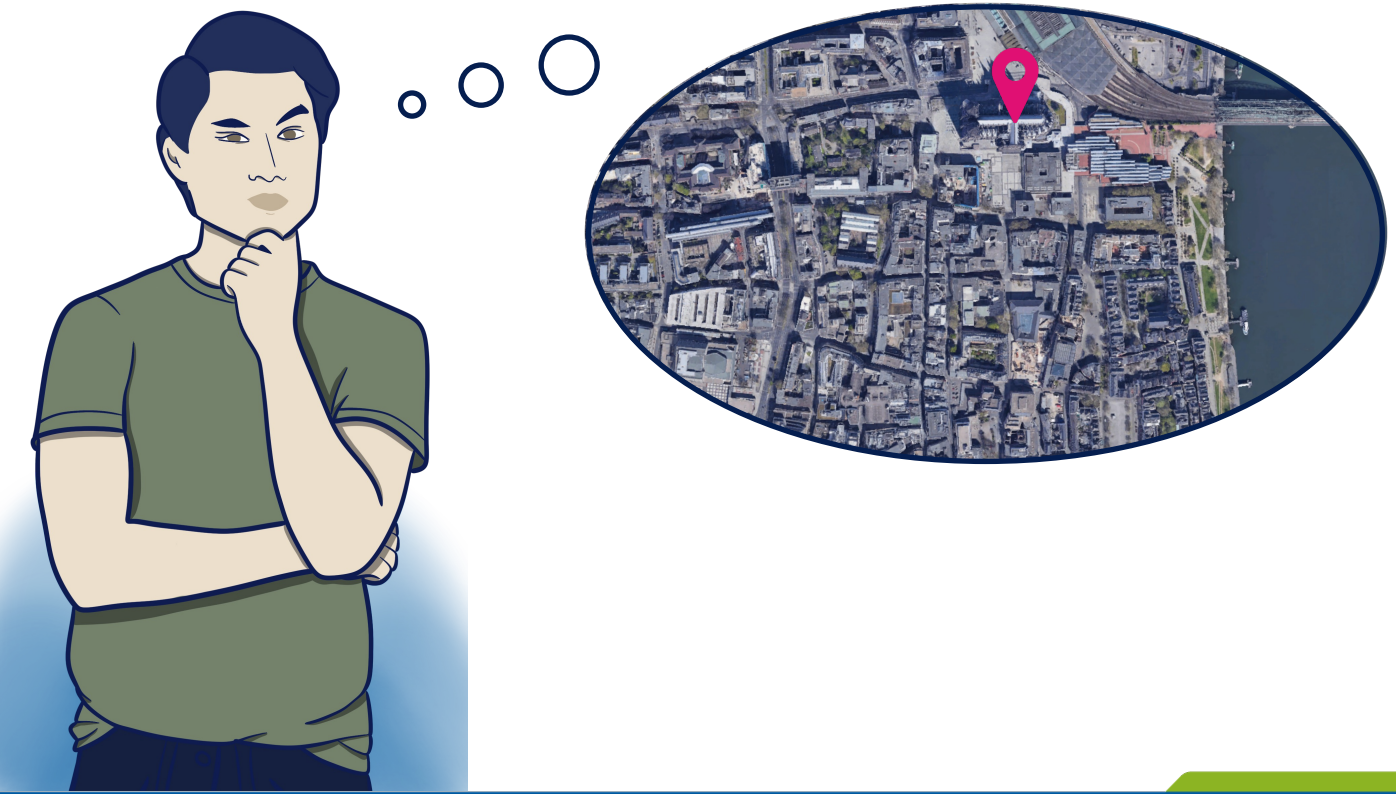
This is data

This is data

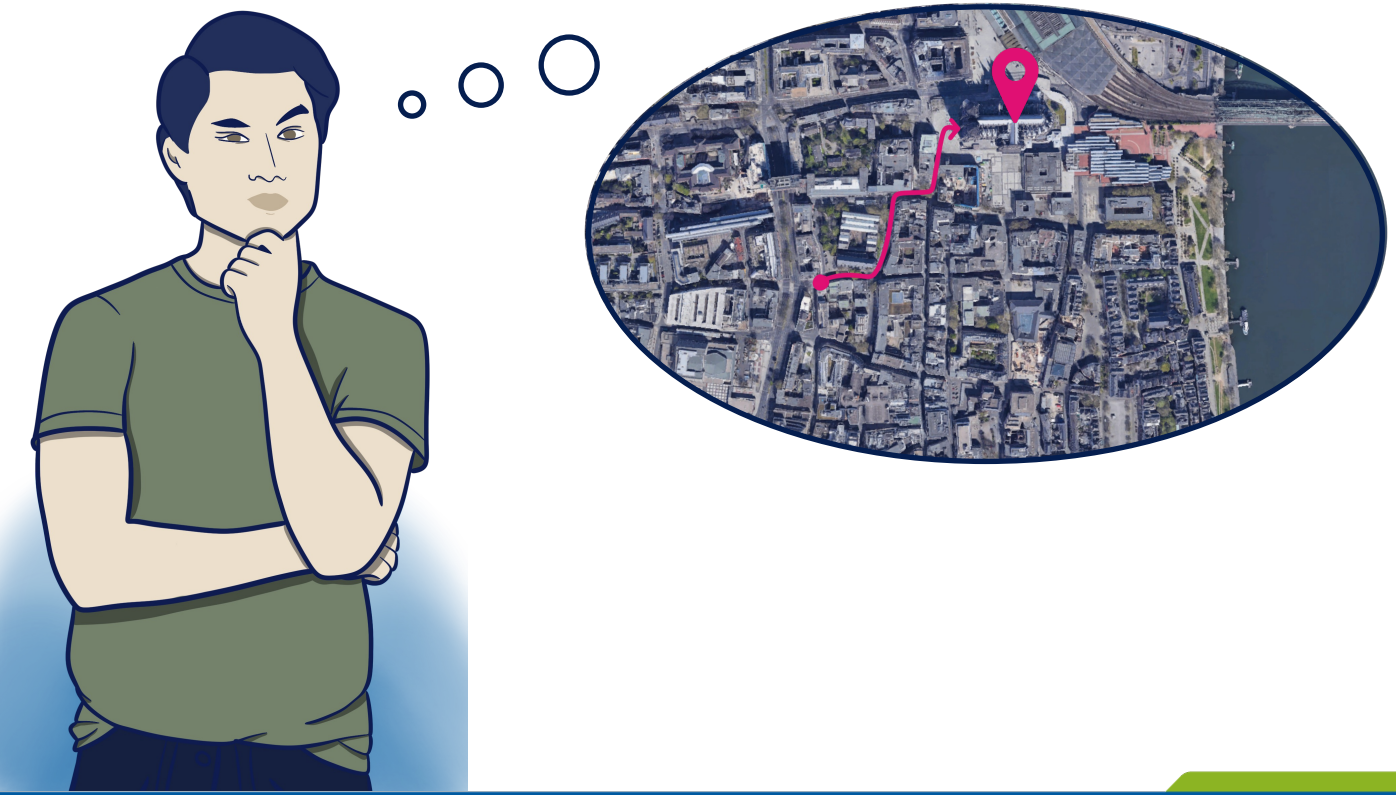
Information – the human-readable picture

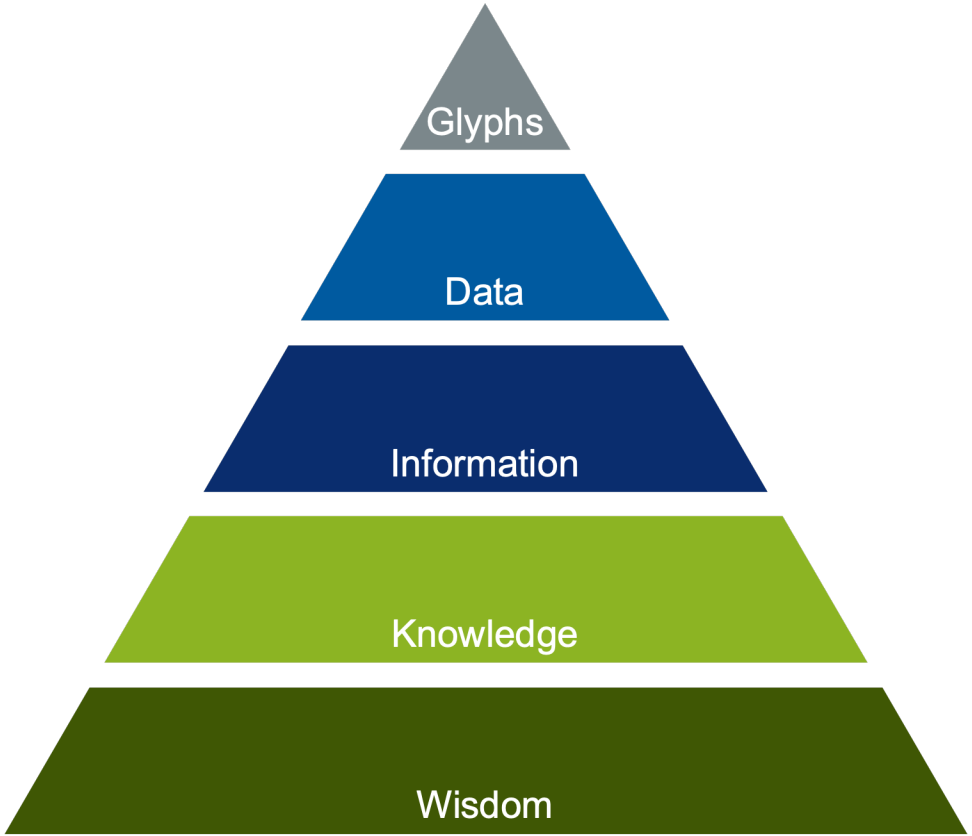
Information – the human-readable picture

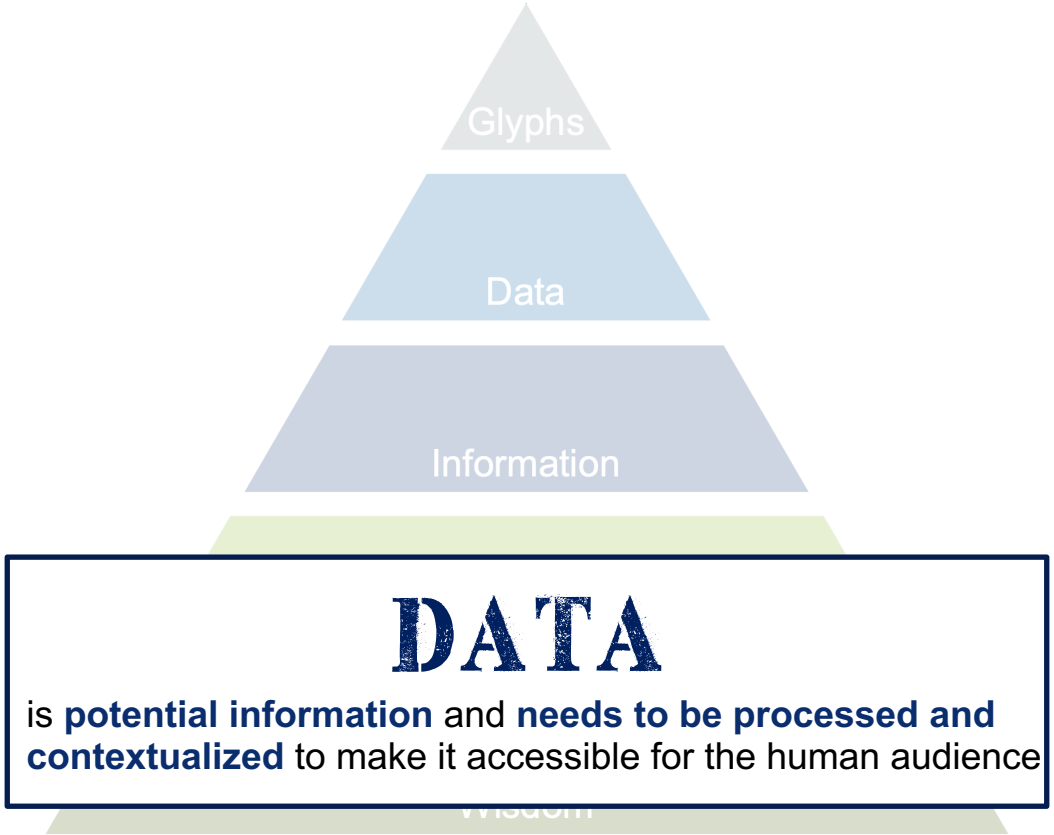
Knowledge – Where is the cathedral?



Wisdom – the bigger picture




Glyphs
Data
Information
Knowledge
Wisdom



DATA

is **potential information** and **needs to be processed and contextualized** to make it accessible for the human audience

>> Data are **REPRESENTATIONS** of **OBSERVATIONS**, **OBJECTS**, or other **ENTITIES** used as **EVIDENCE OF PHENOMENA** for the purposes of research or scholarship. <<

C.L. Borgman (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press

What is METADATA?

What is metadata? - Merriam Webster

metadata noun, plural in form but singular or plural in construction

Save Word

meta·da·ta | \ ˌme-tə-ˈdā-tə  , -ˈda- *also* -ˈdä- \

Definition of *metadata*

: data that provides information about other data

What is metadata? - Wikipedia

Definition [edit]

Metadata means "data about data". Although the "meta" prefix means "after" or "beyond", it is used to mean "about" in epistemology. Metadata is defined as the data providing information about one or more aspects of the data; it is used to summarize basic information about data which can make tracking and working with specific data easier.[12] Some examples include:

- Means of creation of the data
- Purpose of the data
- Time and date of creation
- Creator or author of the data
- Location on a computer network where the data was created
- Standards used
- File size
- Data quality
- Source of the data
- Process used to create the data

For example, a digital image may include metadata that describes the size of the image, its color depth, resolution, when it was created, the shutter speed, and other data.[13] A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document. Metadata within web pages can also contain descriptions of page content, as well as key words linked to the content.[14] These links are often called "Metatags", which were used as the primary factor in determining order for a web search until the late 1990s.[14] The reliance of metatags in web searches was decreased in the late 1990s because of "keyword stuffing".[14] Metatags were being largely misused to trick search engines into thinking some websites had more relevance in the search than they really did.[14]

Metadata can be stored and managed in a database, often called a metadata registry or metadata repository.[15] However, without context and a point of reference, it might be impossible to identify metadata just by looking at it.[16] For example: by itself, a database containing several numbers, all 13 digits long could be the results of calculations or a list of numbers to plug into an equation - without any other context, the numbers themselves can be perceived as the data. But if given the context that this database is a log of a book collection, those 13-digit numbers may now be identified as ISBNs - information that refers to the book, but is not itself the information within the book. The term "metadata" was coined in 1968 by Philip Bagley, in his book "Extension of Programming Language Concepts" where it is clear that he uses the term in the ISO 11179 "traditional" sense, which is "structural metadata" i.e. "data about the containers of data"; rather than the alternative sense "content about individual instances of data content" or metacontent, the type of data usually found in library catalogues.[17][18] Since then the fields of information management, information science, information technology, librarianship, and GIS have widely adopted the term. In these fields the word *metadata* is defined as "data about data".[19] While this is the generally accepted definition, various disciplines have adopted their own more specific explanation and uses of the term.

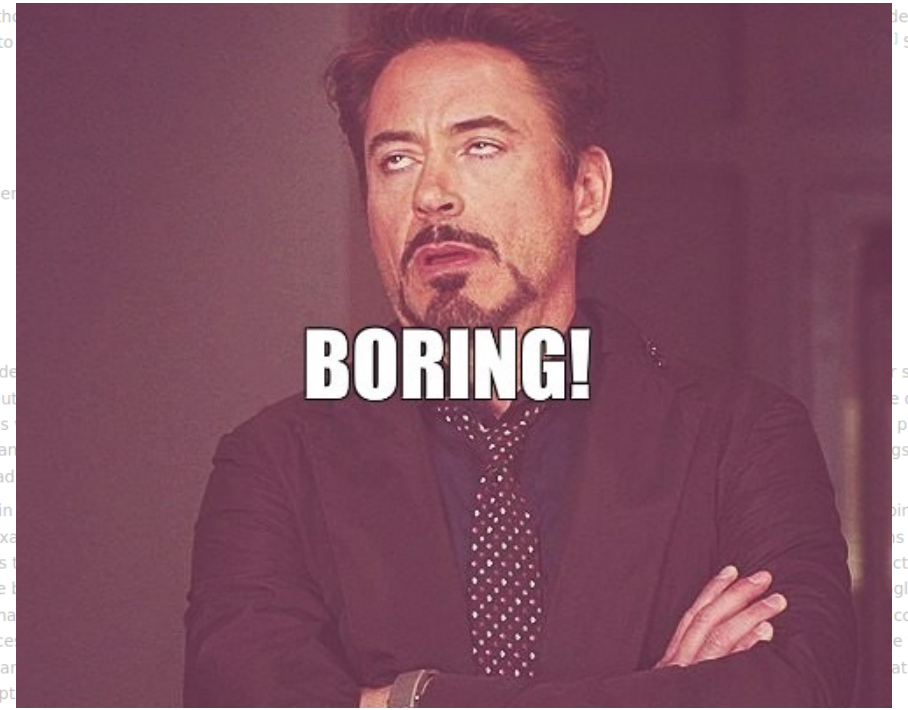
Slate reported in 2013 that the United States government's interpretation of "metadata" could be broad, and might include message content such as the subject lines of emails.[20]

What is metadata? - Wikipedia

Definition [edit]

Metadata means "data about data". Alth… defined as the data providing information about one or more aspects of the data; it is used to … Some examples include:

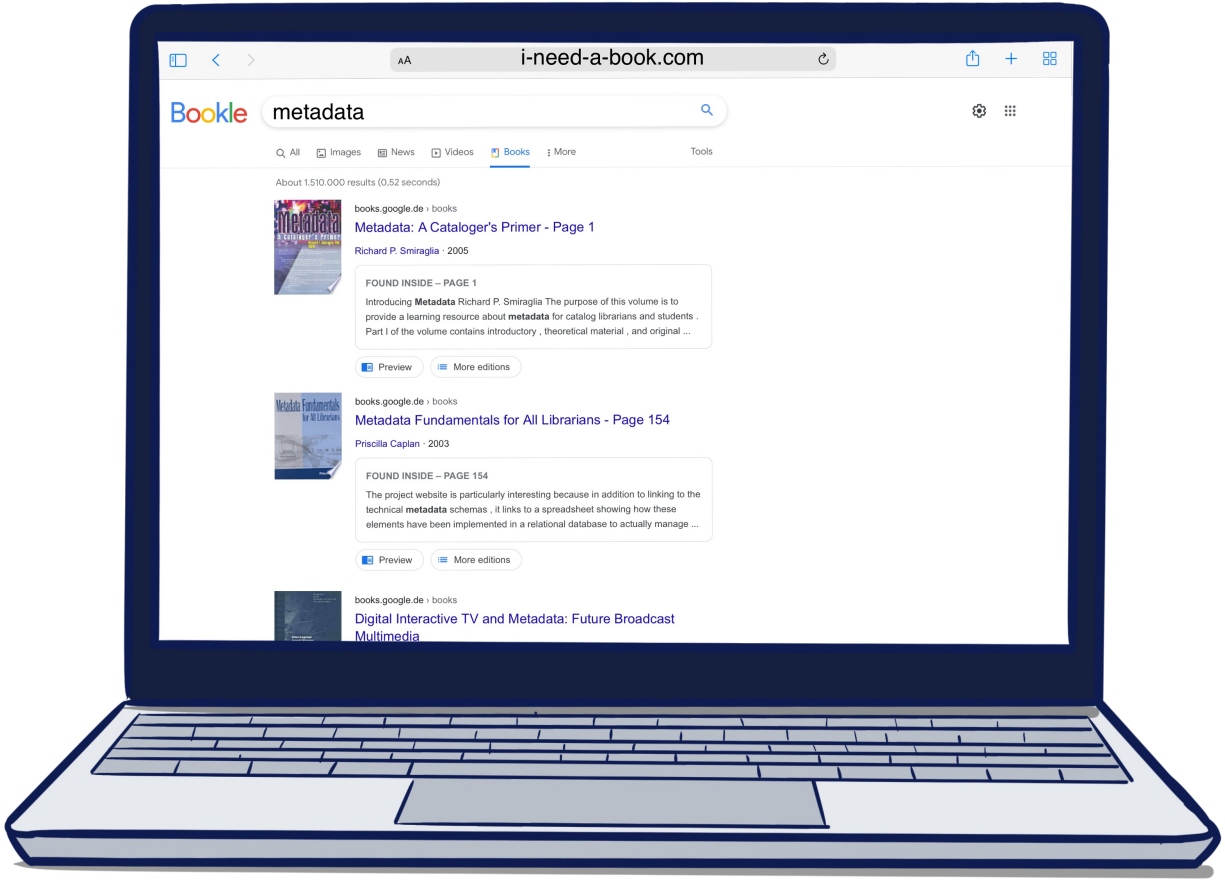
- Means of creation of the data
- Purpose of the data
- Time and date of creation
- Creator or author of the data
- Location on a computer network wher…
- Standards used
- File size
- Data quality
- Source of the data
- Process used to create the data

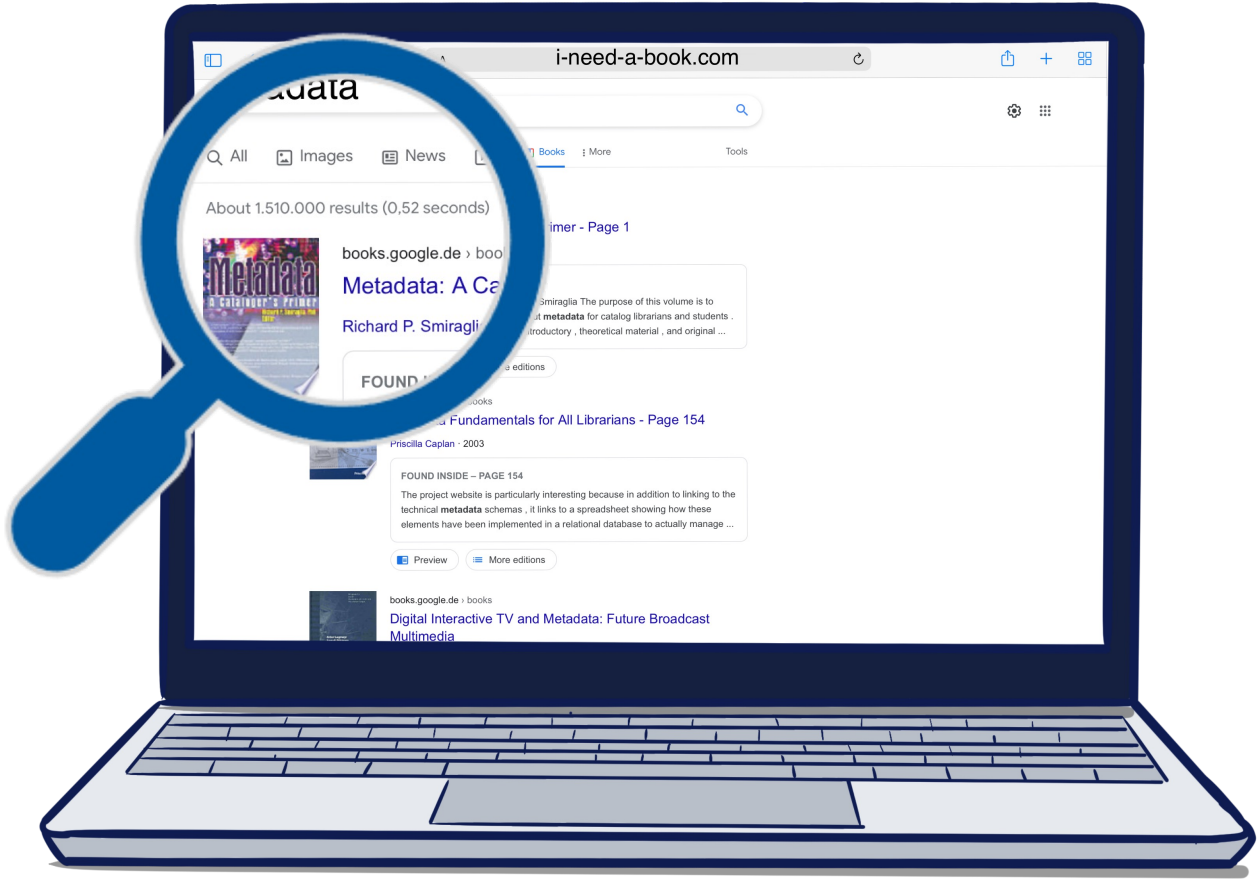


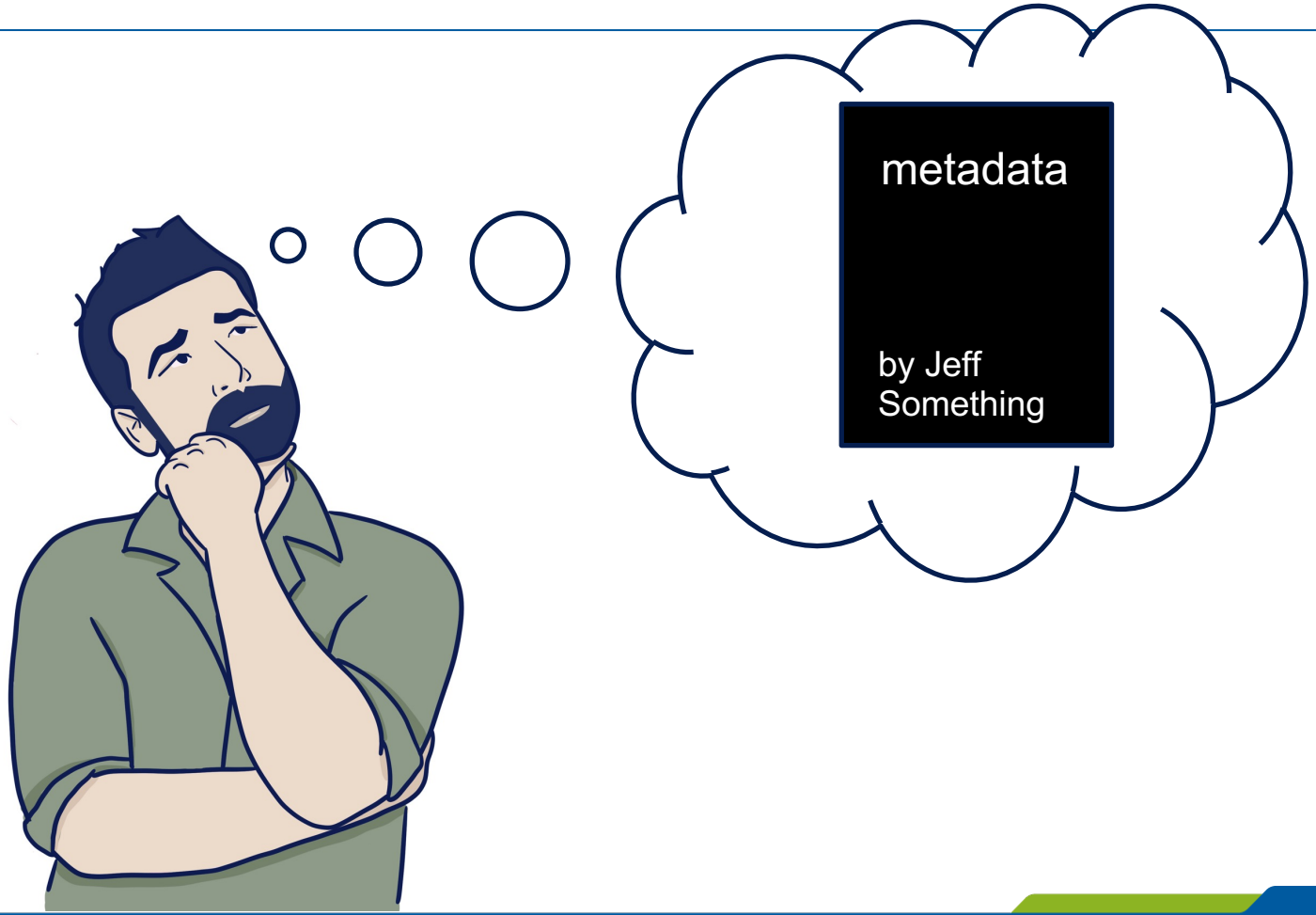
For example, a digital image may include … speed, and other data.[13] A text document's metadata may contain information about … document. Metadata within web pages can also contain descriptions of page content, as … primary factor in determining order for a web search until the late 1990s.[14] The reliar… gs were being largely misused to trick search engines into thinking some websites had …

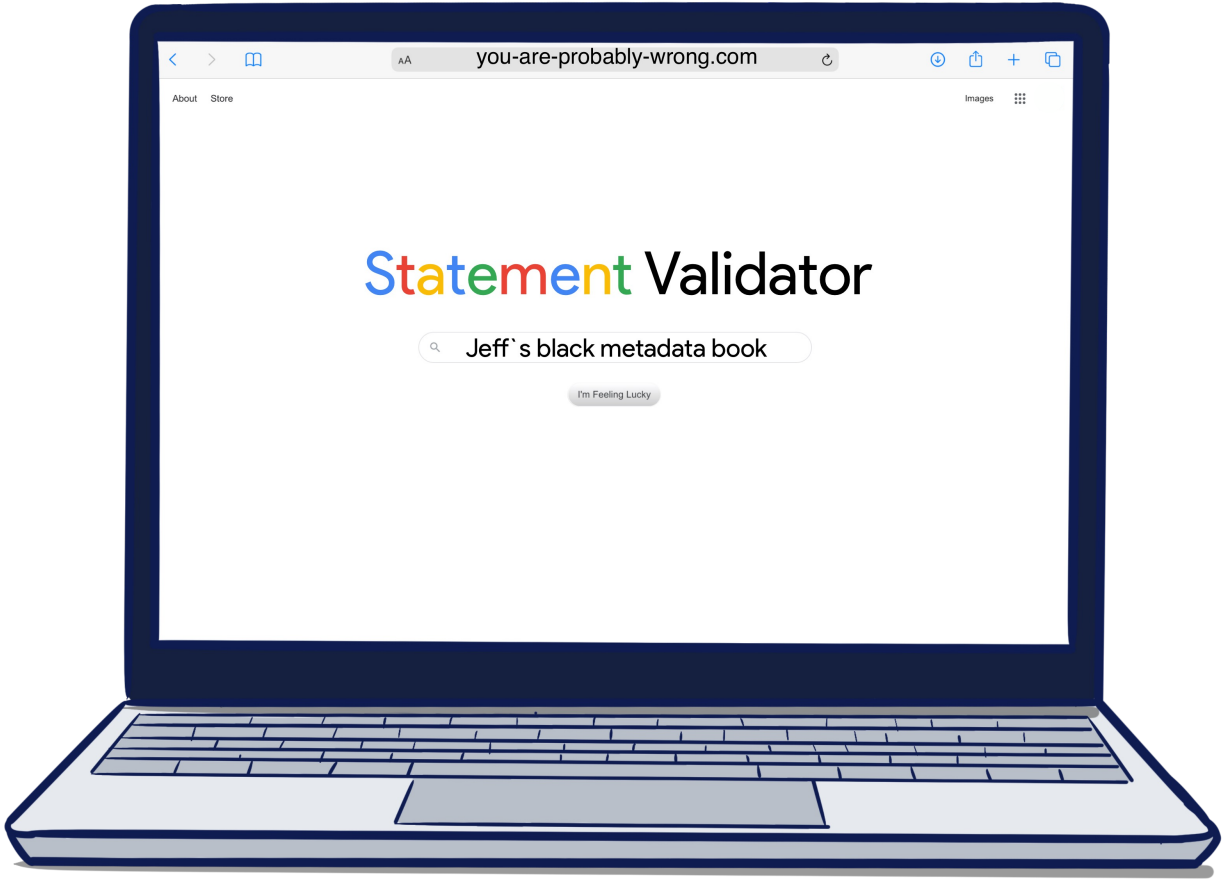
Metadata can be stored and managed in … int of reference, it might be impossible to identify metadata just by looking at it.[16] For exa… s or a list of numbers to plug into an equation - without any other context, the numbers … ction, those 13-digit numbers may now be identified as ISBNs - information that refers to the … gley, in his book "Extension of Programming Language Concepts" where it is clear tha… containers of data"; rather than the alternative sense "content about individual instance… e fields of information management, information science, information technology, librarian… ata".[19] While this is the generally accepted definition, various disciplines have adopt…

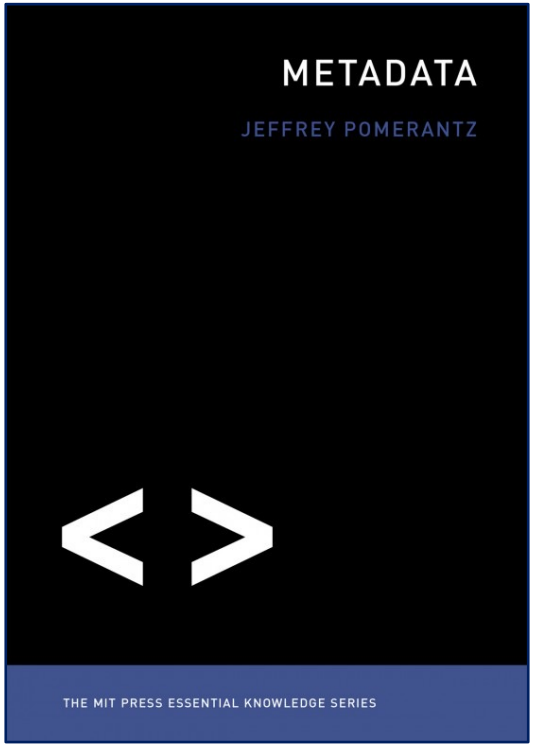
Slate reported in 2013 that the United States government's interpretation of "metadata" could be broad, and might include message content such as the subject lines of emails.[20]


i-need-a-book.com
Bookle
metadata
All
Images
News
Videos
Books
More
Tools
About 1,510,000 results (0,52 seconds)
books.google.de › books
Metadata: A Cataloger's Primer - Page 1
Richard P. Smiraglia · 2005
FOUND INSIDE – PAGE 1
Introducing Metadata Richard P. Smiraglia The purpose of this volume is to provide a learning resource about metadata for catalog librarians and students . Part I of the volume contains introductory , theoretical material , and original ...
Preview
More editions
books.google.de › books
Metadata Fundamentals for All Librarians - Page 154
Priscilla Caplan · 2003
FOUND INSIDE – PAGE 154
The project website is particularly interesting because in addition to linking to the technical metadata schemas , it links to a spreadsheet showing how these elements have been implemented in a relational database to actually manage ...
Preview
More editions
books.google.de › books
Digital Interactive TV and Metadata: Future Broadcast Multimedia


i-need-a-book.com
All
Images
News
Books
More
Tools
About 1.510.000 results (0,52 seconds)
books.google.de
Metadata: A Ca
Richard P. Smiragli
Metadata
A Cataloger's Primer
imer - Page 1
Smiraglia The purpose of this volume is to
metadata for catalog librarians and students .
troductory , theoretical material , and original ...
editions
Fundamentals for All Librarians - Page 154
Priscilla Caplan · 2003
FOUND INSIDE – PAGE 154
The project website is particularly interesting because in addition to linking to the technical metadata schemas , it links to a spreadsheet showing how these elements have been implemented in a relational database to actually manage ...
Preview
More editions
books.google.de › books
Digital Interactive TV and Metadata: Future Broadcast Multimedia


metadata
by Jeff Something


you-are-probably-wrong.com
About
Store
Images
Statement Validator
Jeff`s black metadata book
I'm Feeling Lucky


METADATA
JEFFREY POMERANTZ
THE MIT PRESS ESSENTIAL KNOWLEDGE SERIES

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.
pages cm.—(The MIT Press essential knowledge series)
Includes bibliographical references and index.
ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organization. I. Title.
Z666.7.P66 2015
025.3—dc23

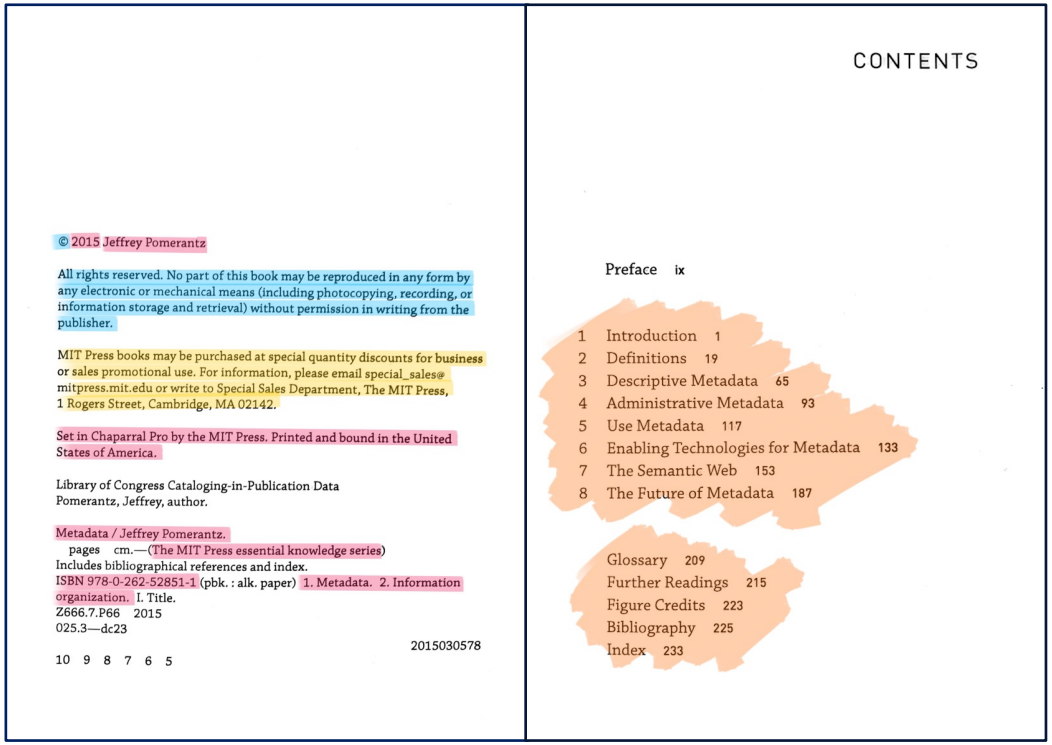
2015030578

10 9 8 7 6 5

CONTENTS

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier



MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 1 Rogers Street, Cambridge, MA 02142.

Set in Chaparral Pro by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data
Pomerantz, Jeffrey, author.

Metadata / Jeffrey Pomerantz.
pages cm.—(The MIT Press essential knowledge series)
Includes bibliographical references and index.
ISBN 978-0-262-52851-1 (pbk. : alk. paper) 1. Metadata. 2. Information organization. I. Title.
Z666.7.P66 2015
025.3—dc23

2015030578

10 9 8 7 6 5

CONTENTS

Administration

- Distribution
- Responsibility
- Conditions

Legal terms

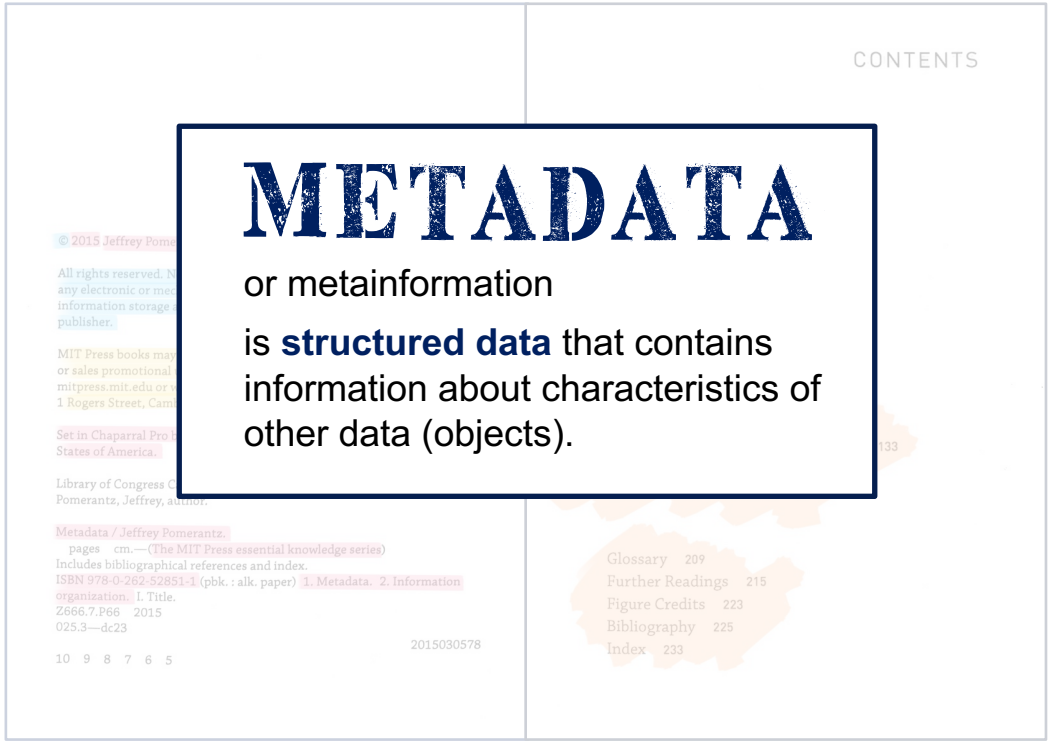
- Copyright issues
- Terms of distribution

Structure

- Content
- Chapters
- Pages

Description

- Publication year
- Author
- Title
- Publisher / Series
- Keywords
- Persistent Identifier



METADATA

or metainformation

is **structured data** that contains information about characteristics of other data (objects).

Administration

- Distribution
- Responsibility
- Conditions

Legal terms

- Copyright issues
- Terms of distribution

Structure

- Content
- Chapters
- Pages

Metadata can be:

descriptive

Information about the intellectual content

e.g. title author, date of publication, subject, description, unique identifier

administrative

Information to support management of a resource

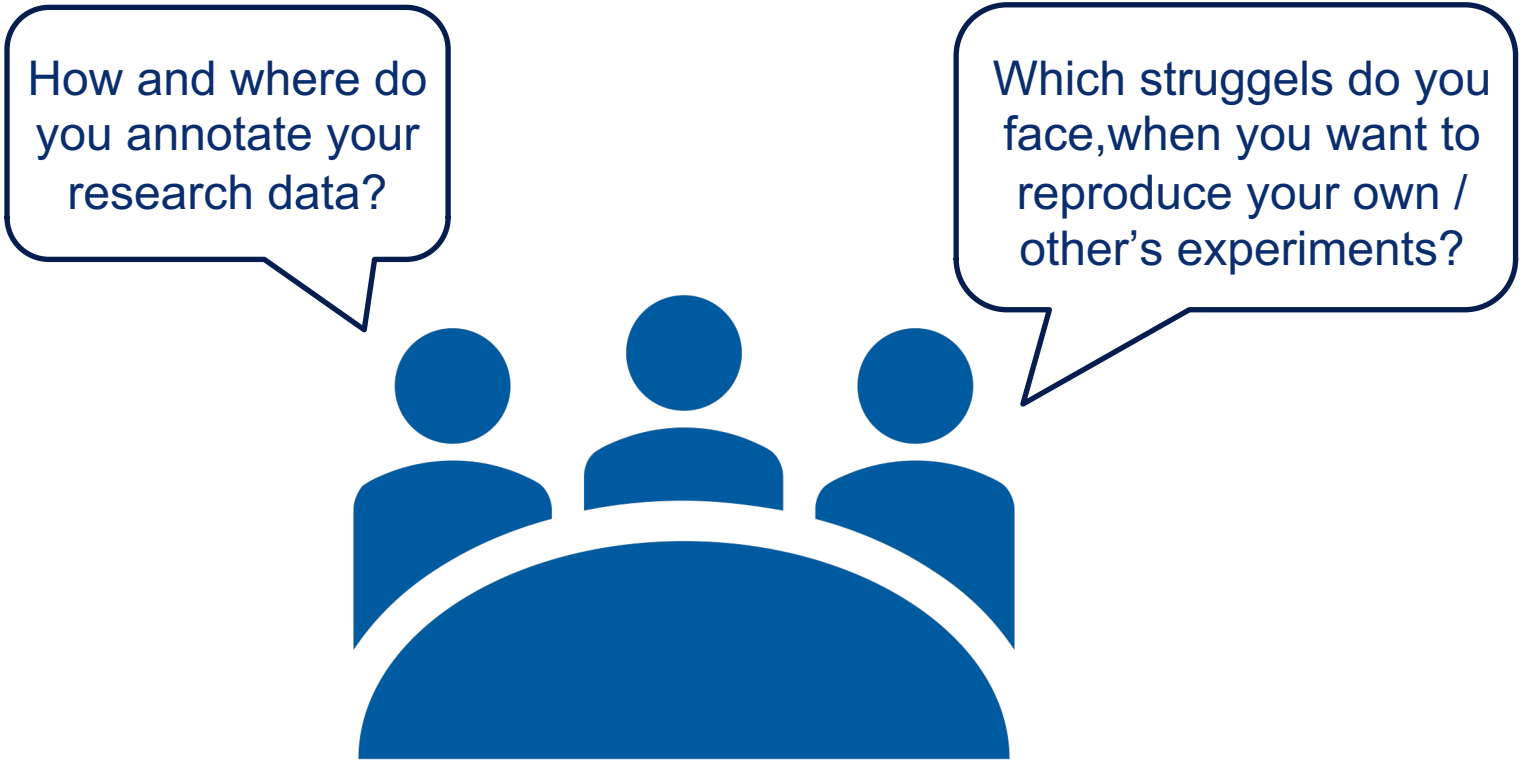
e.g. technical information on the file's creation and format, version, copyright information, licence

structural

relationships between components of a data object

e.g. chapters in a book, files in a data set

Discussion: Your experiences with research data



Discussion: Your experiences with research data

Data Annotation	Reproducibility
hand written notes -> lack of findability / readibility	time passes -> harder to find the notes / make sense
different spellings / terminologies	no description of other's data -> can't reproduce data from others
different "languages" between fields of research -> hard to join the languages	missing information in publications
ambiguity in data annotation -> solution: controlled vocabulary	inventing data

Metadata annotation in the scientific context


You should start your project with **repeating** your collaborator's results

You should start your project with **repeating** your collaborator's results

The Publication

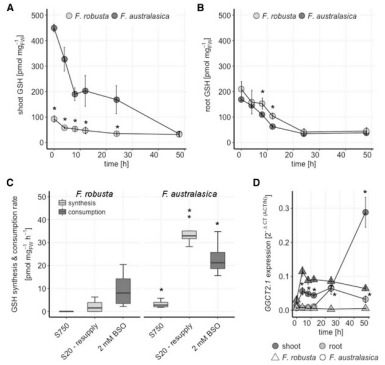


Figure 4. GSH turnover in *F. robusta* and *F. australasica*. GSH concentrations in shoots (A) and roots (B) of 20-d-old seedlings of *F. robusta* (C_3) and *F. australasica* (C_3) in a time course of 48 h after transfer to medium supplemented with 2 mM BSO. Data are presented as means and SE, $n = 4$. C, GSH synthesis was analyzed in 20-d-old seedlings exposed to low sulfate (20 µM sulfate; S20) or adequate sulfate (750 µM sulfate; S750) for 6 d by resupply with 0.2 mM [^{35}S]sulfate solution for 4 h. GSH consumption rate is calculated from A at 4 h after treatment with 2 mM BSO. Data are shown as box plot (25%–75%); the line represents median, and the whiskers represent 1.5 IQR, $n = 4$. D Transcript levels of *GGCT2.1* in shoots and roots of 20-d-old seedlings in a time course of 48 h after transfer to medium supplemented with 2 mM BSO. Data are presented as means and SEM, $n = 4$. Asterisks represent significant differences between *F. robusta* and *F. australasica* at $P < 0.05$ (Student's *t* test).

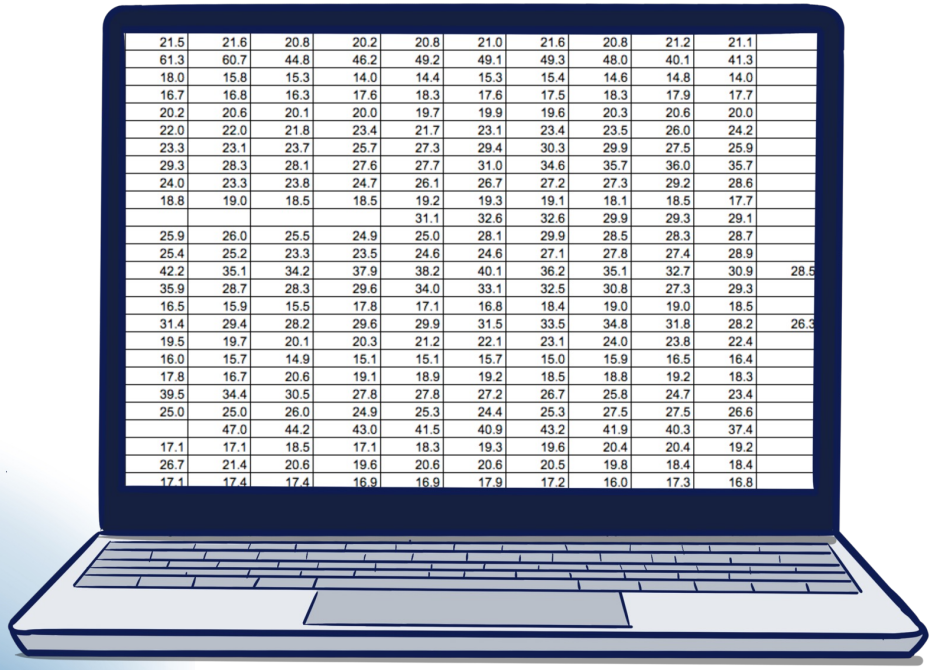
to higher GSH synthesis are therefore likely to be involved in the adjustment of S supply and GSH homeostasis of C_4 plants.

Partitioning of S in Shoots and Roots of *Flaveria* Species

To test the significance of the root for S metabolism in the context of the evolution of C_4 photosynthesis, the five species were grown under full nutrient and low S conditions. Total S_i sulfate and low M_r thiols were determined in shoots and roots (Supplemental Fig. S7). Whereas total S and sulfate did not show any clear patterns relative to photosynthetic type, Cys, and GSH at full nutrition. To better understand the partitioning of S in the different species, the relative portions of total S in sulfate, Cys, and GSH were calculated (Fig. 5). In the shoots of fully nourished *Flaveria* species, the fraction of total S occupied by inorganic sulfate was relatively stable at 50%–70%. However, in the roots, the fraction of inorganic sulfate was higher in the C_4 species. Exposure to S deficiency reduced the sulfate pool in the shoots and roots of *F. robusta*, *F. linearis*, *F. anomala*, and *F. palmeri* to 3.5%–16%. The C_3 species *F. australasica* suffered little loss of relative sulfate pool in shoots, but showed a strong decrease in roots. The increase in GSH fractions of total S in shoots and roots

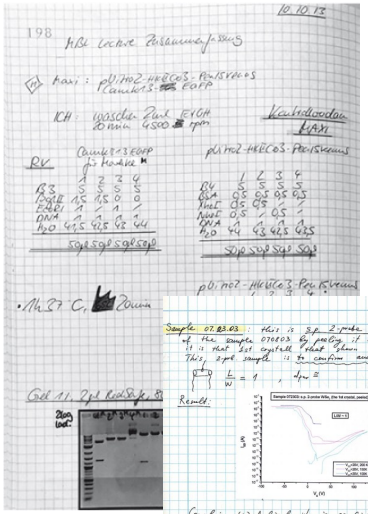
You should start your project with **repeating** your collaborator's results

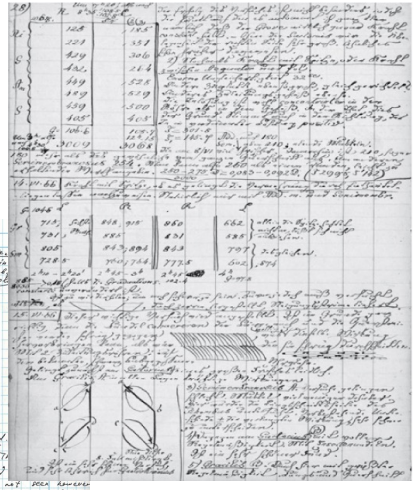
The Data



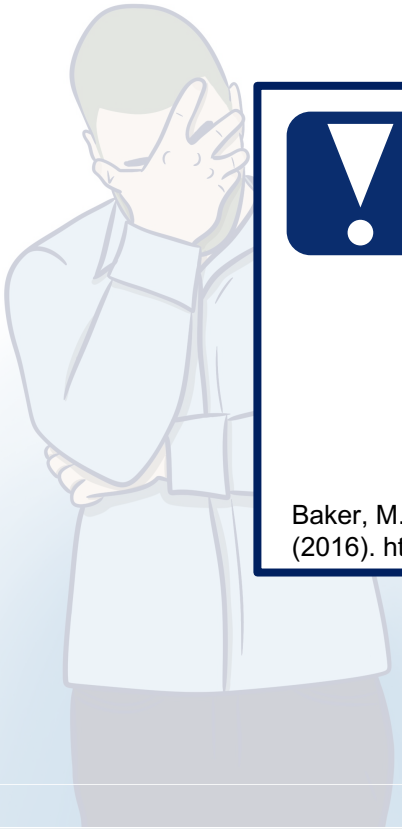
21.5	21.6	20.8	20.2	20.8	21.0	21.6	20.8	21.2	21.1	
61.3	60.7	44.8	46.2	49.2	49.1	49.3	48.0	40.1	41.3	
18.0	15.8	15.3	14.0	14.4	15.3	15.4	14.6	14.8	14.0	
16.7	16.8	16.3	17.6	18.3	17.6	17.5	18.3	17.9	17.7	
20.2	20.6	20.1	20.0	19.7	19.9	19.6	20.3	20.6	20.0	
22.0	22.0	21.8	23.4	21.7	23.1	23.4	23.5	26.0	24.2	
23.3	23.1	23.7	25.7	27.3	29.4	30.3	29.9	27.5	25.9	
29.3	28.3	28.1	27.6	27.7	31.0	34.6	35.7	36.0	35.7	
24.0	23.3	23.8	24.7	26.1	26.7	27.2	27.3	29.2	28.6	
18.8	19.0	18.5	18.5	19.2	19.3	19.1	18.1	18.5	17.7	
				31.1	32.6	32.6	29.9	29.3	29.1	
25.9	26.0	25.5	24.9	25.0	28.1	29.9	28.5	28.3	28.7	
25.4	25.2	23.3	23.5	24.6	24.6	27.1	27.8	27.4	28.9	
42.2	35.1	34.2	37.9	38.2	40.1	36.2	35.1	32.7	30.9	28.5
35.9	28.7	28.3	29.6	34.0	33.1	32.5	30.8	27.3	29.3	
16.5	15.9	15.5	17.8	17.1	16.8	18.4	19.0	19.0	18.5	
31.4	29.4	28.2	29.6	29.9	31.5	33.5	34.8	31.8	28.2	26.3
19.5	19.7	20.1	20.3	21.2	22.1	23.1	24.0	23.8	22.4	
16.0	15.7	14.9	15.1	15.1	15.7	15.0	15.9	16.5	16.4	
17.8	16.7	20.6	19.1	18.9	19.2	18.5	18.8	19.2	18.3	
39.5	34.4	30.5	27.8	27.8	27.2	26.7	25.8	24.7	23.4	
25.0	25.0	26.0	24.9	25.3	24.4	25.3	27.5	27.5	26.6	
	47.0	44.2	43.0	41.5	40.9	43.2	41.9	40.3	37.4	
17.1	17.1	18.5	17.1	18.3	19.3	19.6	20.4	20.4	19.2	
26.7	21.4	20.6	19.6	20.6	20.6	20.5	19.8	18.4	18.4	
17.1	17.4	17.4	16.9	16.9	17.9	17.2	16.0	17.3	16.8	

The Documentation





The Documentation



»More than 70 % of researchers have tried and failed to reproduce another scientist's experiments.

More than half have failed to reproduce their own experiments.«

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533,** 452 – 454 (2016). https://doi.org/10.1038/533452a

Worst practice – no documentation

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

someRandomFileName.csv

Slightly better – lab notes

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	
12	0.16	0.5534802	-0.2035575	-0.4056435	
13	0.17	0.5527935	-0.1961019	-0.4098618	
14	0.2	0.558189	-0.1908045	-0.4121181	
15	0.21	0.5764356	-0.1865862	-0.4162383	
16	0.22	0.589581	-0.18639	-0.4258521	
17	0.25	0.6049827	-0.1941399	-0.4243806	
18	0.26	0.619992	-0.206991	-0.4192794	
19	0.27	0.6320583	-0.2191554	-0.4092732	
20	0.3	0.6392196	-0.2279844	-0.3975993	
21	0.31	0.6465771	-0.2317122	-0.3908304	
22	0.32	0.6583491	-0.2291616	-0.3950487	
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

2022-02-28

Gotham City, New Jersey, USA

Flight of the bat

weather: more clouds than sun, 11°C, 74% humidity, 1023 mbar, SSW, 17 km/h

recording device strapped to upper arm

Slightly better – lab notes

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	
12	0.16	0.5534802	-0.2035575	-0.4056435	
13	0.17	0.5527935	-0.1961019	-0.4098618	
14	0.2	0.558189	-0.1908045	-0.4121181	
15	0.21	0.5764356	-0.1865862	-0.4162383	
16	0.22	0.589581	-0.18639	-0.4258521	
17	0.25	0.6049827	-0.1941399	-0.4243806	
18	0.26	0.619992	-0.206991	-0.4192794	
19	0.27	0.6320583	-0.2191554	-0.4092732	
20	0.3	0.6392196	-0.2279844	-0.3975993	
21	0.31	0.6465771	-0.2317122	-0.3908304	
22	0.32	0.6583491	-0.2291616	-0.3950487	
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

2022-02-28

Gotham City, New Jersey, USA
Flight of the bat

weather: more clouds than sun, 11°C, 74% humidity, 1023 mbar, SSW, 17 km/h

recording device strapped to upper arm

Slightly better – lab notes

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	
12	0.16	0.5534802	-0.2035575	-0.4056435	
13	0.17	0.5527935	-0.1961019	-0.4098618	
14	0.2	0.558189	-0.1908045	-0.4121181	
15	0.21	0.5764356	-0.1865862	-0.4162383	
16	0.22	0.589581	-0.18639	-0.4258521	
17	0.25	0.6049827	-0.1941399	-0.4243806	
18	0.26	0.619992	-0.206991	-0.4192794	
19	0.27	0.6320583	-0.2191554	-0.4092732	
20	0.3	0.6392196	-0.2279844	-0.3975993	
21	0.31	0.6465771	-0.2317122	-0.3908304	
22	0.32	0.6583491	-0.2291616	-0.3950487	
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

LAB NOTES IV
LAB NOTES III
LAB NOTES II
LAB NOTES I

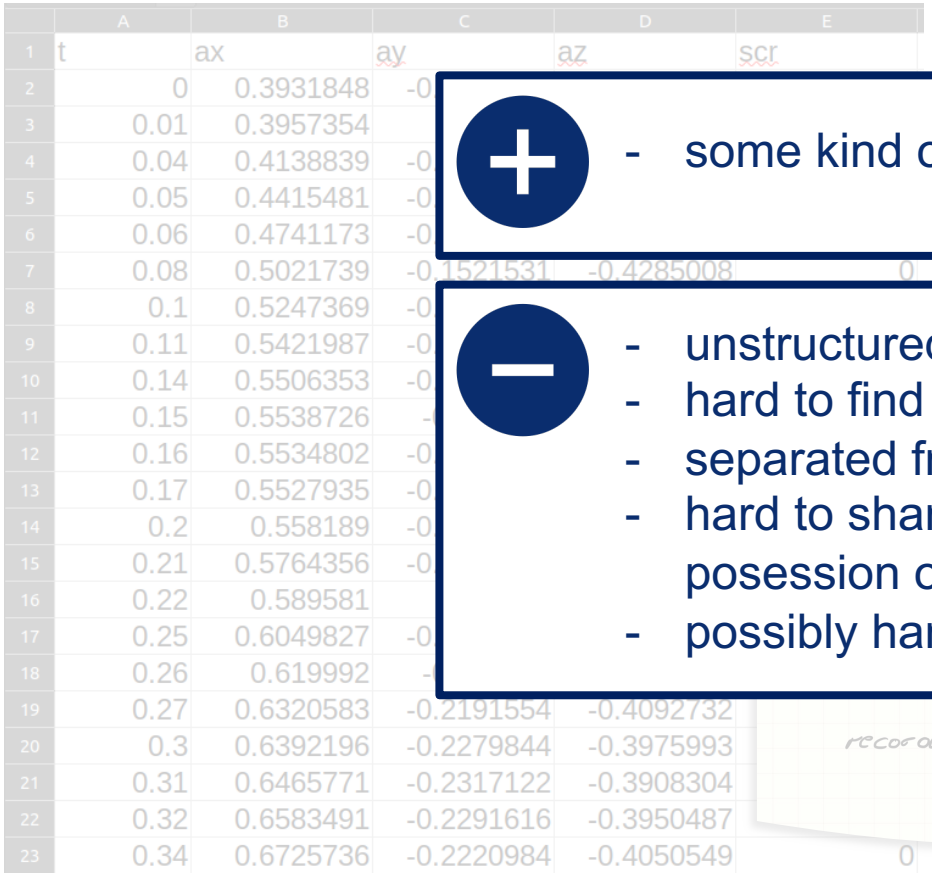
2022-02-28

Gotham City, New Jersey, USA
Flight of the bat

weather: more clouds than sun, 11°C, 74% humidity, 1023 mbar, SSW, 17 km/h

recording device strapped to upper arm

Slightly better – lab notes



	t	ax	ay	az	scr
2	0	0.3931848	-0		
3	0.01	0.3957354			
4	0.04	0.4138839	-0		
5	0.05	0.4415481	-0		
6	0.06	0.4741173	-0		
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0		
9	0.11	0.5421987	-0		
10	0.14	0.5506353	-0		
11	0.15	0.5538726	-		
12	0.16	0.5534802	-0		
13	0.17	0.5527935	-0		
14	0.2	0.558189	-0		
15	0.21	0.5764356	-0		
16	0.22	0.589581			
17	0.25	0.6049827	-0		
18	0.26	0.619992	-		
19	0.27	0.6320583	-0.2191554	-0.4092732	
20	0.3	0.6392196	-0.2279844	-0.3975993	
21	0.31	0.6465771	-0.2317122	-0.3908304	
22	0.32	0.6583491	-0.2291616	-0.3950487	
23	0.34	0.6725736	-0.2220984	-0.4050549	0

\+
- some kind of documentation

−
- unstructured
- hard to find
- separated from data
- hard to share / only in the posession of the experimentator
- possibly hard to read

Even better – Readme-style metadata

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

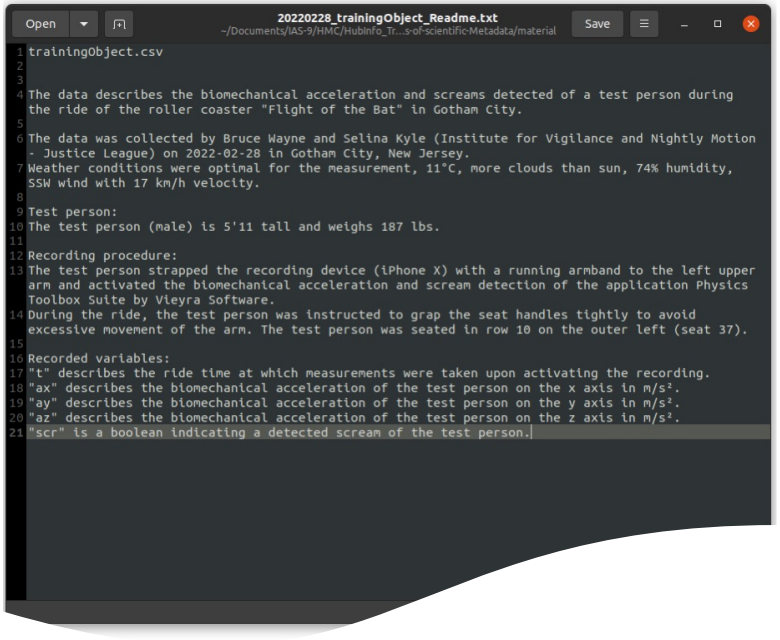
txt 20220228_recordingData_Readme.txt

Even better – Readme-style metadata

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

txt 20220228_recordingData_Readme.txt

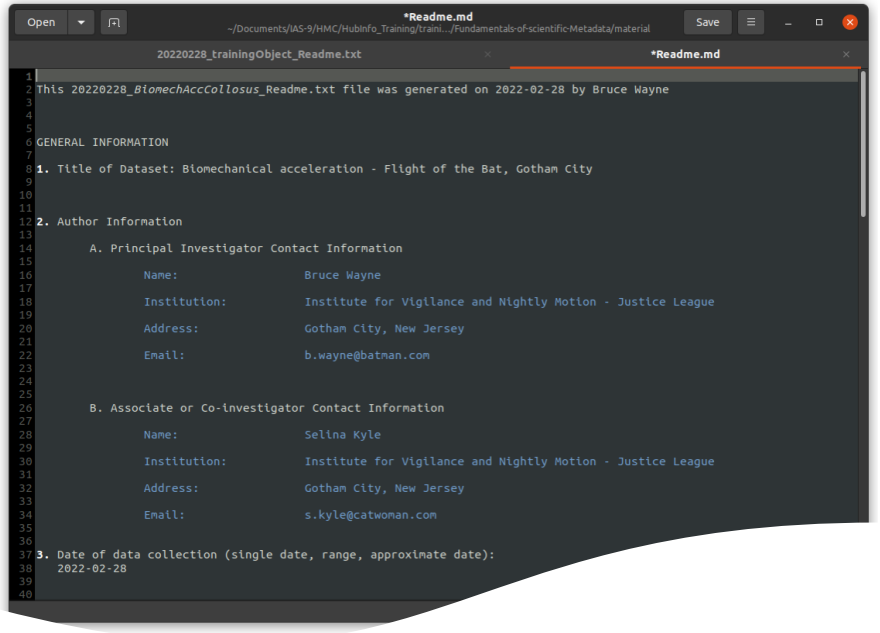


Even better – Readme-style metadata

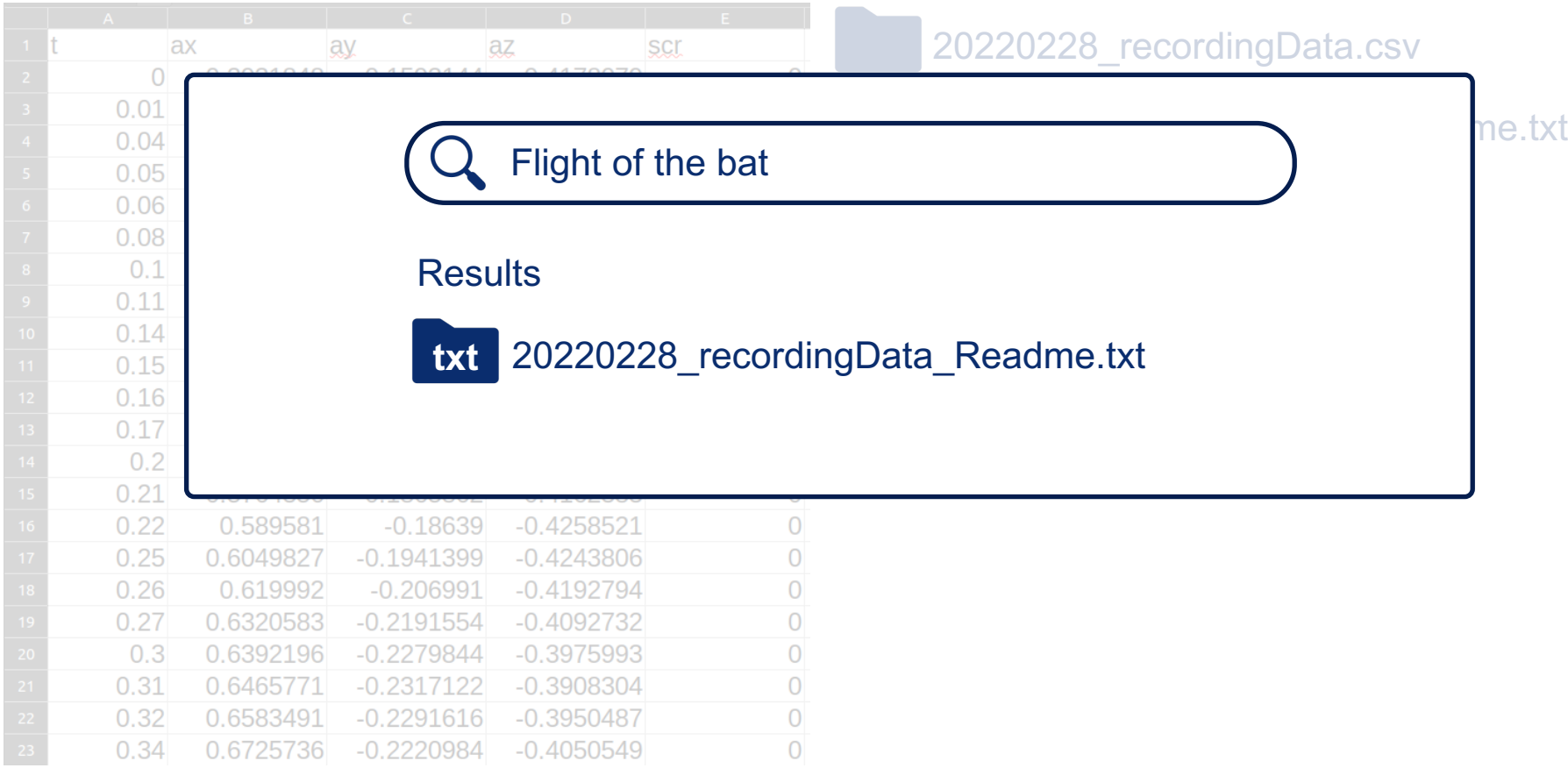
	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1	0.5247369	-0.1669662	-0.420849	0
9	0.11	0.5421987	-0.1813869	-0.4160421	0
10	0.14	0.5506353	-0.1947285	-0.4094694	0
11	0.15	0.5538726	-0.203067	-0.4057416	0
12	0.16	0.5534802	-0.2035575	-0.4056435	0
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0.1908045	-0.4121181	0
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

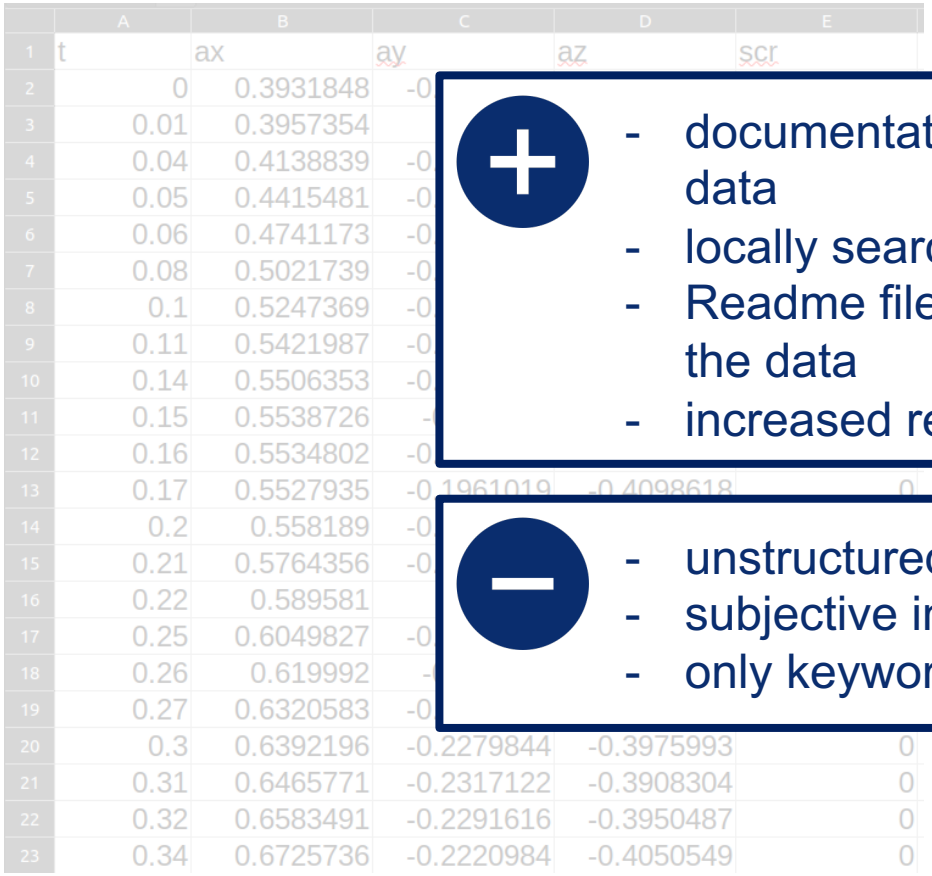
txt 20220228_recordingData_Readme.txt



Even better – Readme-style metadata



Even better – Readme-style metadata



	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0		
3	0.01	0.3957354			
4	0.04	0.4138839	-0		
5	0.05	0.4415481	-0		
6	0.06	0.4741173	-0		
7	0.08	0.5021739	-0		
8	0.1	0.5247369	-0		
9	0.11	0.5421987	-0		
10	0.14	0.5506353	-0		
11	0.15	0.5538726	-		
12	0.16	0.5534802	-0		
13	0.17	0.5527935	-0.1961019	-0.4098618	0
14	0.2	0.558189	-0		
15	0.21	0.5764356	-0		
16	0.22	0.589581			
17	0.25	0.6049827	-0		
18	0.26	0.619992	-		
19	0.27	0.6320583	-0		
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

gData_Readme.txt

\+

- documentation linked to the data
- locally searchable
- Readme file can be shared with the data
- increased readability

−

- unstructured
- subjective information
- only keyword search possible

Even better – Readme-style metadata

	A	B	C	D	E
1	t	ax	ay	az	scr
2	0	0.3931848	-0.1593144	-0.4178079	0
3	0.01	0.3957354	-0.15696	-0.4242825	0
4	0.04	0.4138839	-0.1547037	-0.429678	0
5	0.05	0.4415481	-0.1512702	-0.4325229	0
6	0.06	0.4741173	-0.1488177	-0.434583	0
7	0.08	0.5021739	-0.1521531	-0.4285008	0
8	0.1				
9	0.11				
10	0.14				
11	0.15				
12	0.16				
13	0.17				
14	0.2				
15	0.21	0.5764356	-0.1865862	-0.4162383	0
16	0.22	0.589581	-0.18639	-0.4258521	0
17	0.25	0.6049827	-0.1941399	-0.4243806	0
18	0.26	0.619992	-0.206991	-0.4192794	0
19	0.27	0.6320583	-0.2191554	-0.4092732	0
20	0.3	0.6392196	-0.2279844	-0.3975993	0
21	0.31	0.6465771	-0.2317122	-0.3908304	0
22	0.32	0.6583491	-0.2291616	-0.3950487	0
23	0.34	0.6725736	-0.2220984	-0.4050549	0

20220228_recordingData.csv

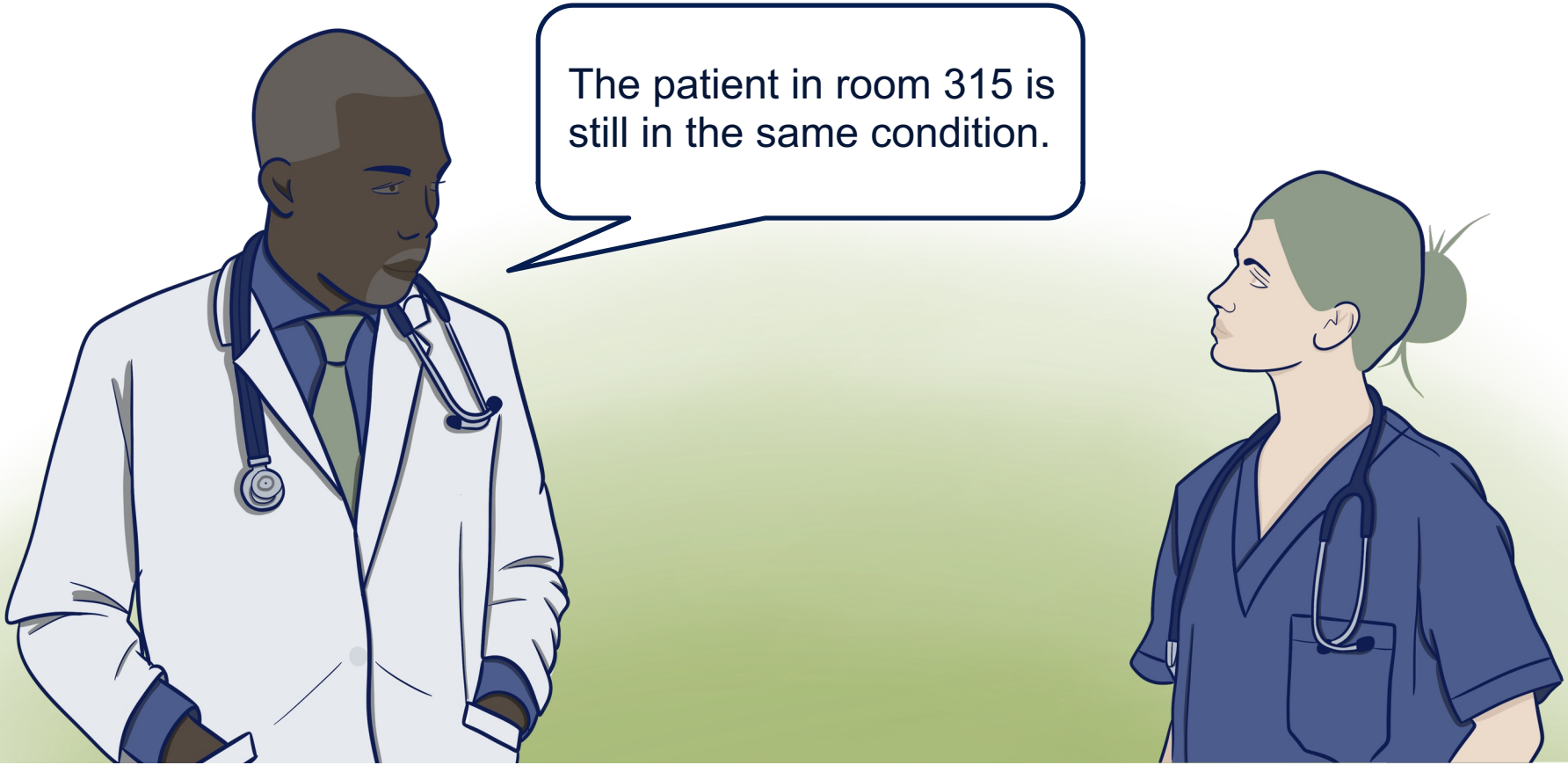
txt 20220228_recordingData_Readme.txt

https://ordo.open.ac.uk/articles/dataset/Template_for_a_README_file_for_data_uploads/13332743/1

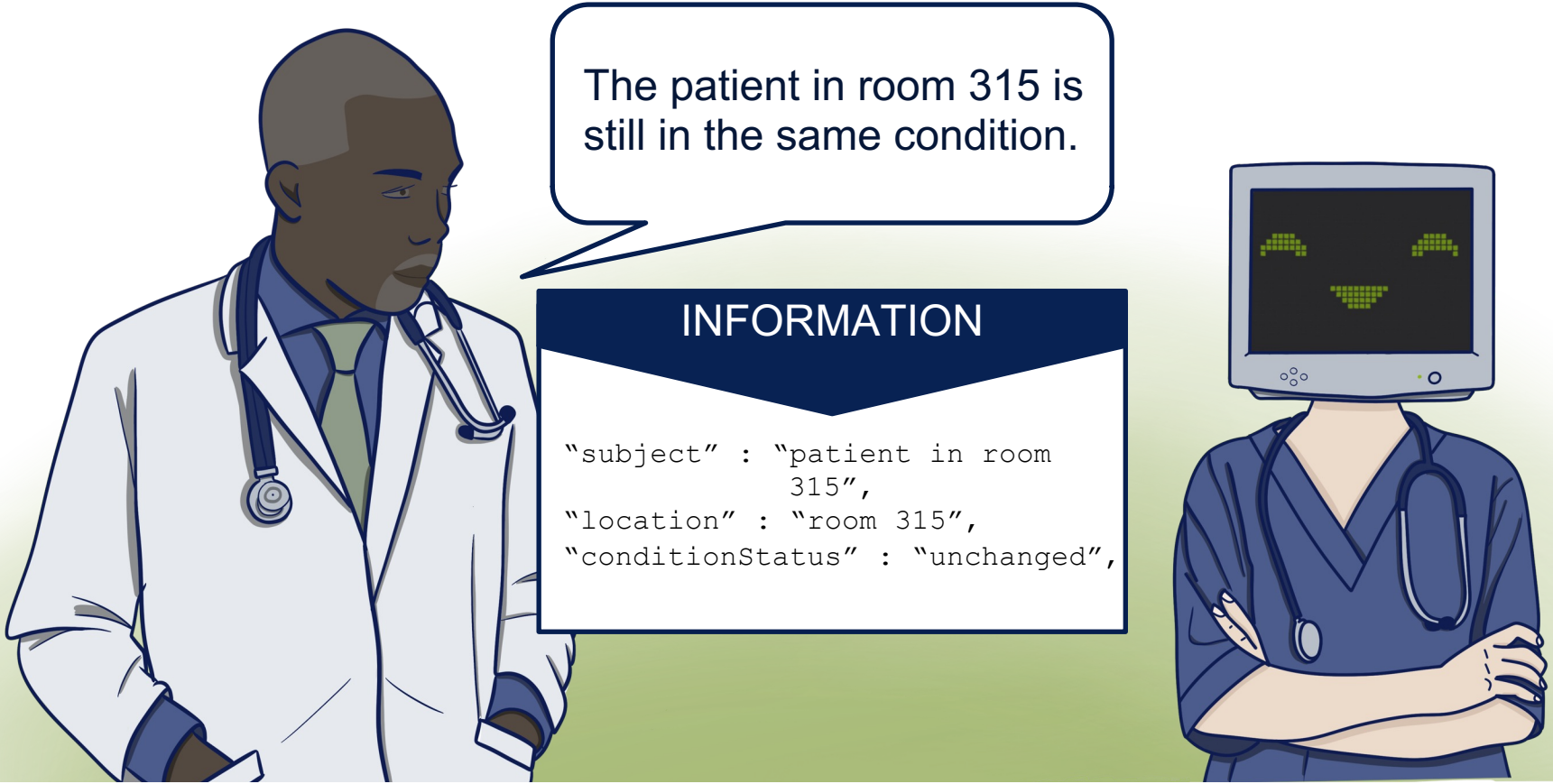
Link in Handout!

Part 2: Structure & Schema

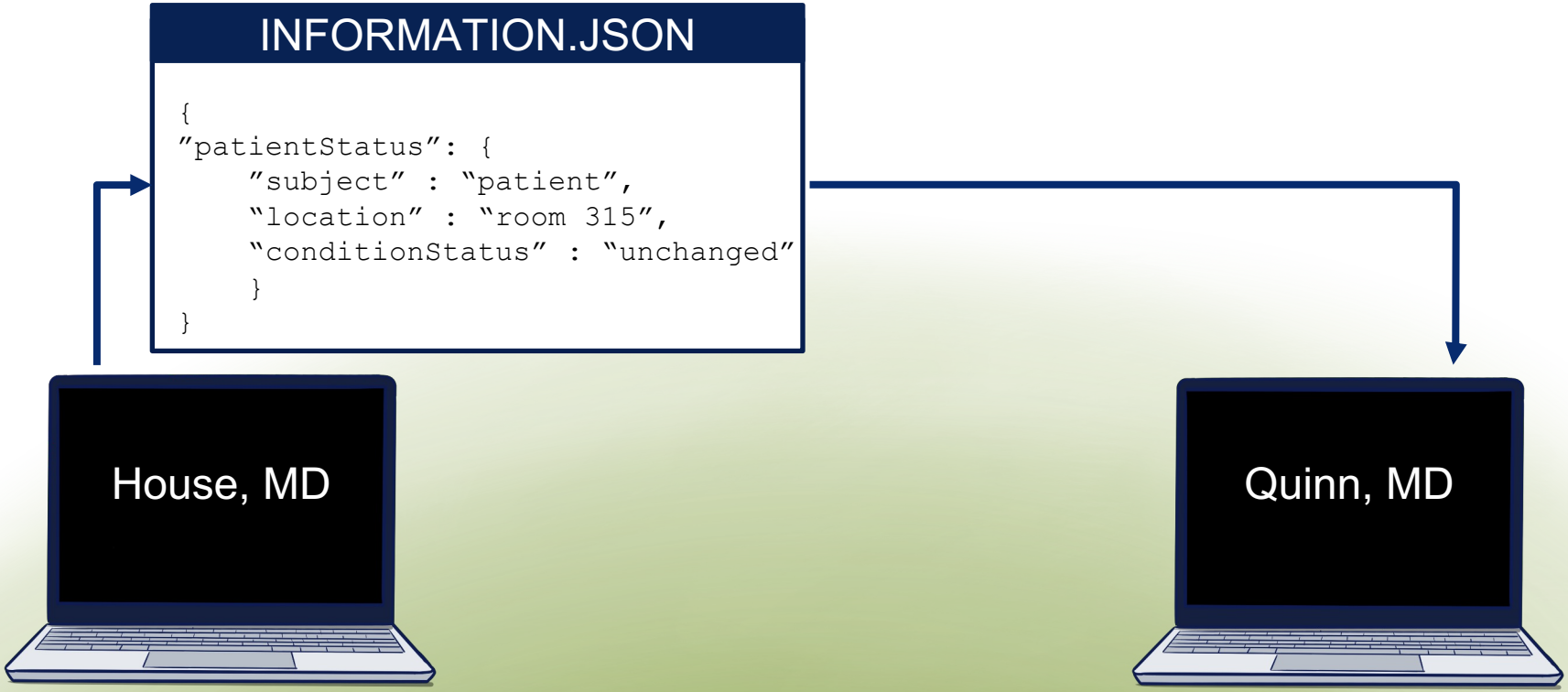
Transferring information – Natural Language



Transferring information – Data Transfer



Transferring information – Data Transfer



HELMHOLTZ METADATA COLLABORATION

Structured metadata & markup

www.helmholtz-metadata.de

What is markup?

Markup is not part of the text or content but tells something about it ...

Making markup work

To make markup work, the writer and the interpreter of the marked up content have to **agree on the interpretation of the markup symbols**. [1]

[1] Cynthia Zender (2005). Markup 101: Markup Basics. SAS Institute. https://www.lexjansen.com/pharmasug/2005/Tutorials/tu12.pdf
Interrobang punctuation mark: https://www.merriam-webster.com/dictionary/interrobang

Some types of markup

Punctuational markup
!.?

Presentational markup.

```
**bold**
```

Descriptive or declarative markup

```
<h1>The most important headline per page</h1>
```

Referential markup

```
<a href="url">link text displayed to reader on screen</a>
```

[1] James H. Coombs et al. (November 1987). Markup Systems and the Future of Scholarly Text Processing. Communications of the ACM 30.
http://xml.coverpages.org/coombs.html#Note1

Computer markup vs manual markup

Marking up a manuscript or page proof is usually a manual process.

In computer files, markup includes formatting instructions and additional information to the natural text so that software can format the text or a printer can print the document. [1]

```
**make this text bold**
```

[1] Cynthia Zender (2005). Markup 101: Markup Basics. SAS Institute. https://www.lexjansen.com/pharmasug/2005/Tutorials/tu12.pdf

Types of markup

Punctuational markup
!.?

Presentational markup.

```
**bold**
```

Descriptive or declarative markup

```
<h1>The most important headline per page</h1>
```

Referential markup

```
<a href="url">link text displayed to reader on screen</a>
```

[1] James H. Coombs et al. (November 1987). Markup Systems and the Future of Scholarly Text Processing. Communications of the ACM 30.
http://xml.coverpages.org/coombs.html#Note1

Benefit of rigorous markup

(Meta)data exchange formats need to be read and processed by humans and computers.
Descriptive & referential markup makes natural text more accessible for computer analysis. [1]

[1] Charles F. Goldfarb (1990). The SGML Handbook. Clarendon Press. https://books.google.com/books?id=RilvKya0EnwC

<language>SGML</language>

SGML (Standard Generalized Markup Language) was one of the first industry standards for electronic publishing – a meta-language for generalized, descriptive markup languages – first accepted as an ISO standard in 1986.

Image: HTML source code, https://unsplash.com/photos/Wyd_PkCa1BY

<language>XML and HTML</language>

Both, HTML (1989) and XML (1998) are based on SGML. HTML (HyperText Markup Language) is the standard markup language for web pages. In contrast, the main purpose of XML (eXtensible Markup Language) is the transfer and storage of arbitrary data on the World Wide Web.

[1] https://www.iso.org/standard/16387.html

Image: HTML source code, https://unsplash.com/photos/WYd_PkCa1BY

<example>XML</example>

XML is software- and hardware-independent. It is considered human-readable and allows for hierarchical (tree-like) structures. Data elements are wrapped in start and end “tags”. [1]

```
<example>
    <title>This is the example title</title>
    <description>A simple XML example</description>
    <wordCount>1</wordCount>
</example>
```

[1] “XML Tutorial”. © 1999-2022. Refsnes Data, W3Schools. https://www.w3schools.com/xml/

{"format":"JSON"}

JSON (JavaScript Object Notation) is not a markup language. It is a lightweight, human-readable, hierarchical format to store and transport data. JSON syntax is inspired by JavaScript object notation. [1] Like XML, JSON is software- and hardware-independent.

```
{
    "key":"value",
    "aString":"string",
    "anInteger":5,
    "aBoolean":true,
    "anArray": ["item1", "item2", "item3"]
}
```

[1] https://www.ecma-international.org/publications-and-standards/standards/ecma-404/

{"format":"JSON"}

- curly braces hold objects
  (collections, dictionaries of key/value pairs)
- square brackets hold arrays
  (ordered lists of values)
- keys must be of data type string (in quotes)
- values must be of data type string,
  number, boolean, array or object
- elements are separated by commas
- no comments supported
  (for interoperability)

```
{
    "key":"value",
    "aString":"string",
    "anInteger":5,
    "aFloat":0.5,
    "aBoolean":true,
    "anArray": ["item1", "item2", "item3"],
    "anObject": {
        "key1":"value1",
        "key2":"value2",
        "key3":"value3"
    }
}
```

[1] also see www.json.org

XML and JSON side-by-side

```xml
<example>
	<superhero>Wonder Woman</superhero>
	<publisher>DC Comics</publisher>
	<identities>
		<identity>Princess Diana</identity>
		<identity>Diana Prince</identity>
	</identities>
	<pet>
		<name>Jumpa</name>
		<species>kangaroo</species>
	</pet>
</example>
```

```json
{
  "superhero": "Wonder Woman",
  "publisher": "DC Comics",
  "identities": [
    "Princess Diana",
    "Diana Prince"
  ],
  "pet": {
    "name": "Jumpa",
    "species": "kangaroo"
  }
}
```

XML repeated child elements & JSON list

```xml
<example>
	<superhero>Wonder Woman</superhero>
	<publisher>DC Comics</publisher>
	<identities>
		<identity>Princess Diana</identity>
		<identity>Diana Prince</identity>
	</identities>
	<pet>
		<name>Jumpa</name>
		<species>kangaroo</species>
	</pet>
</example>
```

```json
{
  "superhero": "Wonder Woman",
  "publisher": "DC Comics",
  "identities": [
    "Princess Diana",
    "Diana Prince"
  ],
  "pet": {
    "name": "Jumpa",
    "species": "kangaroo"
  }
}
```

XML nested elements & JSON nested object

```xml
<example>
	<superhero>Wonder Woman</superhero>
	<publisher>DC Comics</publisher>
	<identities>
		<identity>Princess Diana</identity>
		<identity>Diana Prince</identity>
	</identities>
	<pet>
		<name>Jumpa</name>
		<species>kangaroo</species>
	</pet>
</example>
```

```json
{
  "superhero": "Wonder Woman",
  "publisher": "DC Comics",
  "identities": [
    "Princess Diana",
    "Diana Prince"
  ],
  "pet": {
    "name": "Jumpa",
    "species": "kangaroo"
  }
}
```

XML, JSON (& YAML)

```xml
<example>
        <superhero>Wonder Woman</superhero>
        <publisher>DC Comics</publisher>
        <identities>
                <identity>Princess Diana</identity>
                <identity>Diana Prince</identity>
        </identities>
        <pet>
                <name>Jumpa</name>
                <species>kangaroo</species>
        </pet>
</example>
```

```json
{
  "superhero": "Wonder Woman",
  "publisher": "DC Comics",
  "identities": [
    "Princess Diana",
    "Diana Prince"
  ],
  "pet": {
    "name": "Jumpa",
    "species": "kangaroo"
  }
}
```

```yaml
---
superhero: Wonder Woman
publisher: DC Comics
identities:
- Princess Diana
- Diana Prince
pet:
  name: Jumpa
  species: kangaroo
```

[1] if you are interested in YAML, also see https://yaml.org/

HELMHOLTZ METADATA COLLABORATION

The web is not the internet

World-Wide Web

„During some sessions in the CERN cafeteria,
Tim and I try to find a catching name for the system. […]
Tim proposes "World-Wide Web". I like this very much,
except that it is difficult to pronounce in French…“

(Robert Cailliau, 1995)

Quote: http://www.netvalley.com/archives/mirrors/robert_cailliau_speech.htm

Science created the World Wide Web

- CERN research centre in Geneva, Switzerland
- researchers Tim Berners-Lee and Robert Cailliau
- joint proposal for „World-Wide Web“
- developed to “meet the demand for automated information-sharing between scientists in universities and institutes around the world”. [1], [2]

1989

[1] http://info.cern.ch/hypertext/WWW/Proposal.html

[2] https://arxiv.org/pdf/1701.04765.pdf

1960s terms hypertext, hypermedia coined by Ted Nelson

1970s

Transmission Control Protocol/Internet Protocol (TCP/IP) invented by Vint Cerf and Robert (Bob) Elliot Kahn [1]

1980s Mail Transfer Protocol (MTP, SMTP); Suzanne Sluizer, Jon Postel [2]

[1] https://www.darpa.mil/attachments/ARPANET_final.pdf, https://doc.lagout.org/network/The%20Illustrated%20Network.pdf

[2] https://www.cnet.com/tech/tech-industry/end-of-the-road-for-smtp/

HTML (HyperText Markup Language with „hyperlinks“)

HTTP

(HyperText Transfer Protocol; conventions for client-server communication on the Web)

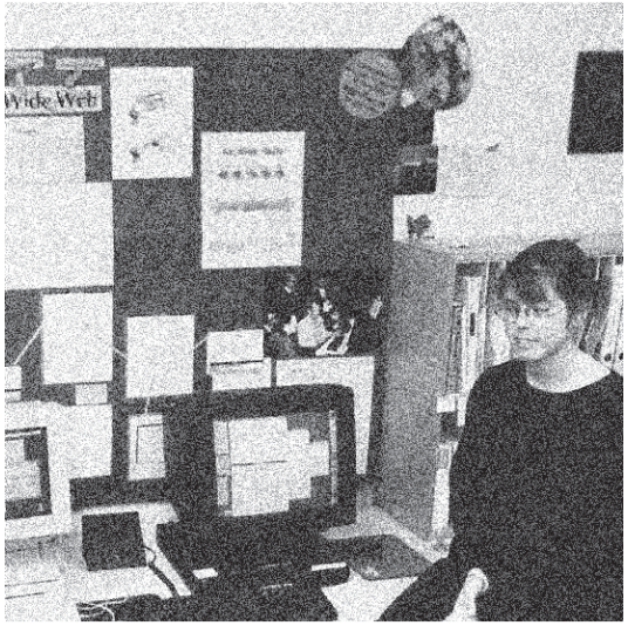
URI (Uniform Resource Identifier)

HTTP GET request in client-server communication

1. Client software (often a „web browser“) establishes connection.
2. Client sends GET request for resource URI and waits for an answer.
3. Server software processes the request, sends back represenation of resource (data and metadata).

The web browser

First operating-system-independent web browser – "Line-Mode browser" – was **written by undergraduate CERN intern Nicola Pellow** in 1990. [1]



[1] https://nowebwithoutwomen.com/images/Nicola_Pellow.pdf

Image: screenshot PDF https://nowebwithoutwomen.com/images/Nicola_Pellow.pdf, snapshot taken APR 2022

Early web servers & repositories

- **early 1990s** arXiv preprint repository switches from email to HTTP access for manuscript transmission. [1]
- **1992** Deutsches Elektronen-Synchrotron DESY in Hamburg connects a web server to the WWW.

Web repositories store and publish (scholarly) digital objects – like paper publications and research data – and their metadata records. They aim to **improve the persistent findability and accessibility of research output on the Web.** [2]

Photo: Kindly provided by Paul Ginsparg

Über welches Protokoll wurde Anfang der 1990er auf das heutige arXiv.org zugegriffen?

[1] https://ar5iv.labs.arxiv.org/html/1709.07020

[2] https://depositonce.tu-berlin.de/handle/11303/5330

Image: screenshot of slide 24; Pascal Becker (2022). „Digitale Repositorien“. Potsdam University of Applied Sciences

Registries of repositories

Repositories are indexed for findability in registry services.

www.re3data.org

v2.sherpa.ac.uk/opendoar

risources.dfg.de

Image: screenshot of re3data.org search box, snapshot taken APR 2022

Metadata schemas

Metadata schemas

Metadata schemas express **expectations in the structure of metadata records.**

Benefits of schemas

A metadata schema is – basically – a set of **conventions or contraints.** [1]

Schemas are expressed in formal languages like XML, JSON or else
so that **(meta)data can be parsed and validated automatically according to the schema.** [2]

[1] https://www.merriam-webster.com/dictionary/schema, https://www.merriam-webster.com/dictionary/protocol
[2] https://gitlab.hzdr.de/hmc/hmc/cct-7-semantics/hmc-glossary-initialization/-/blob/master/terms/schema.yaml (HMC CCT7, not yet ratified)
Image: Child plays with wooden shape sorter toy, https://unsplash.com/photos/ehaO7XywMGM

Writing schemas

XML Schemas (.xsd) are written in XML and used to describe & syntactically validate the structure of XML documents or (meta)data records. [1]

The JSON Schema vocabulary is used to describe & syntactically validate the structure of JSON (meta)data records. [2]

[1] “XML Schema Tutorial”. © 1999-2022. Refsnes Data, W3Schools. https: //www.w3schools.com/xml/schema_intro.asp
[2] “Understanding JSON Schema. The basics”, © Copyright 2013-2016 Michael Droettboom, Space Telescope Science Institute; Last updated on Feb 07, 2022. https://json-schema.org/understanding-json-schema/basics.html

"$schema" vocabulary example

- JSON Schema version in $schema
- list of required properties
- one required property
- one optional property
- data type constraints
- descriptions for the human reader

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "description": "In real life you would add a meaningful description here.",
  "type": "object",
  "required": [
    "superhero"
  ],
  "properties": {
    "superhero": {
      "description": "A mandatory string property.",
      "type": "string"
    },
    "power": {
      "description": "An optional numeric property.",
      "type": "integer"
    }
  }
}
```

"$schema" vocabulary example

- list of required properties
- one required property
- one optional property
- data type constraints
- descriptions for the human reader

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "description": "In real life you would add a meaningful description here.",
  "type": "object",
  "required": [
    "superhero"
  ],
  "properties": {
    "superhero": {
      "description": "A mandatory string property.",
      "type": "string"
    },
    "power": {
      "description": "An optional numeric property.",
      "type": "integer"
    }
  }
}
```

"$schema"

- list of required properties
- one required property
- one optional property
- data type constraints
- descriptions for the human reader

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "description": "In real life you would add a meaningful description here.",
  "type": "object",
  "required": [
    "superhero"
  ],
  "properties": {
    "superhero": {
      "description": "A mandatory string property.",
      "type": "string"
    },
    "power": {
      "description": "An optional numeric property.",
      "type": "integer"
    }
  }
}
```

[1] www.json.org

"$schema"

- list of required properties
- one required property
- one optional property
- data type constraints
- descriptions for the human reader

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "description": "In real life you would add a meaningful description here.",
  "type": "object",
  "required": [
    "superhero"
  ],
  "properties": {
    "superhero": {
      "description": "A mandatory string property.",
      "type": "string"
    },
    "power": {
      "description": "An optional numeric property.",
      "type": "integer"
    }
  }
}
```

[1] www.json.org

"$schema"

- list of required properties
- one required property
- one optional property
- data type constraints
- descriptions for the human reader

A JSON instance is syntactically valid, if it conforms to the definition described by the JSON schema.

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "description": "In real life you would add a meaningful description here.",
  "type": "object",
  "required": [
    "superhero"
  ],
  "properties": {
    "superhero": {
      "description": "A mandatory string property.",
      "type": "string"
    },
    "power": {
      "description": "An optional numeric property.",
      "type": "integer"
    }
  }
}
```

[1] www.json.org

Validate JSON record

A JSON instance is syntactically valid, if it conforms to the definition described by the JSON schema.

```
{
  "superhero": "String Hero"
}
```

```
{
  "superhero": 5
}
```

"$schema"

```
{
  "superhero": "String Hero"
}
```

A JSON instance is syntactically valid, if it conforms to the definition described by the JSON schema.

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "description": "In real life you would add a meaningful description here.",
  "type": "object",
  "required": [
    "superhero"
  ],
  "properties": {
    "superhero": {
      "description": "A mandatory string property.",
      "type": "string"
    },
    "power": {
      "description": "An optional numeric property.",
      "type": "integer"
    }
  }
}
```

[1] www.json.org

Challenges in schema development

The most challenging part of schema development
can be to have everyone **agree on the same expectations.**

HELMHOLTZ METADATA COLLABORATION

Minimal metadata standards

www.helmholtz-metadata.de

Schema or standard?

A well established metadata schema can become a standard.

The Dublin Core

Researchers, librarians and web technologists drafted the Dublin Core – a set of library-card-catalog-like metadata elements for the web – in 1995 at a meeting in Dublin, Ohio (USA). [1]

Creator
Contributor
Publisher
Title
Date
Language
Format
Subject
Description
Identifier
Relation
Source
Type
Coverage
Rights

[1] https://www.dublincore.org/resources/metadata-basics/
[2] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-3
[3] https://www.dublincore.org/about/
[4] https://www.iso.org/standard/71339.html

The Dublin Core

Dublin Core and its extensions are widely used and referenced today. The Dublin Core Metadata Initiative (DCMI) states to work openly, with a paid-membership model. [3] The 15 Dublin Core metadata elements have been formally standardized for cross-domain resource description as e. g. **ISO 15836-1:2017.** [4]

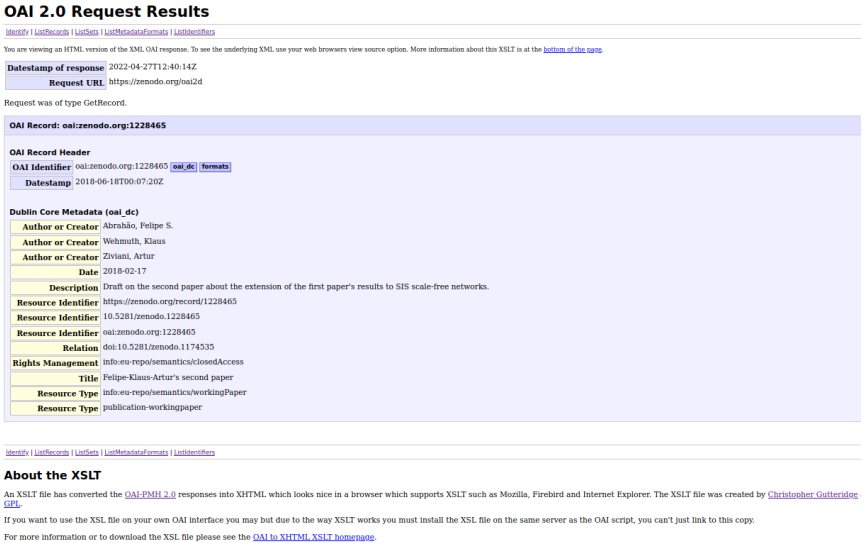
Creator
Contributor
Publisher
Title
Date
Language
Format
Subject
Description
Identifier
Relation
Source
Type
Coverage
Rights

[1] https://www.dublincore.org/resources/metadata-basics/
[2] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-3
[3] https://www.dublincore.org/about/
[4] https://www.iso.org/standard/71339.html

Dublin Core in the wild

Many scholarly repositories expose a standardized application programming interface (API) for the harvesting of **Dublin Core metadata as specified in the Open Archives Initiative Protocol for Metadata Harvesting.** [1]

Try it yourself and check **oai_dc XML records from Zenodo OAI-PMH endpoint** `https://zenodo.org/oai2d`



OAI 2.0 Request Results

Identify | ListRecords | ListSets | ListMetadataFormats | ListIdentifiers

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option. More information about this XSLT is at the bottom of the page.

Datestamp of response	2022-04-27T12:40:14Z
Request URL	https://zenodo.org/oai2d

Request was of type GetRecord.

OAI Record: oai:zenodo.org:1228465

OAI Record Header

OAI Identifier	oai:zenodo.org:1228465 oai_dc formats
Datestamp	2018-06-18T00:07:20Z

Dublin Core Metadata (oai_dc)

Author or Creator	Abrahão, Felipe S.
Author or Creator	Wehmuth, Klaus
Author or Creator	Ziviani, Artur
Date	2018-02-17
Description	Draft on the second paper about the extension of the first paper's results to SIS scale-free networks.
Resource Identifier	https://zenodo.org/record/1228465
Resource Identifier	10.5281/zenodo.1228465
Resource Identifier	oai:zenodo.org:1228465
Relation	doi:10.5281/zenodo.1174535
Rights Management	info:eu-repo/semantics/closedAccess
Title	Felipe-Klaus-Artur's second paper
Resource Type	info:eu-repo/semantics/workingPaper
Resource Type	publication-workingpaper

Identify | ListRecords | ListSets | ListMetadataFormats | ListIdentifiers

About the XSLT

An XSLT file has converted the OAI-PMH 2.0 responses into XHTML which looks nice in a browser which supports XSLT such as Mozilla, Firebird and Internet Explorer. The XSLT file was created by Christopher Gutteridge - GPL.

If you want to use the XSL file on your own OAI interface you may but due to the way XSLT works you must install the XSL file on the same server as the OAI script, you can't just link to this copy.

For more information or to download the XSL file please see the OAI to XHTML XSLT homepage.

[1] http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm#dublincore
[2] https://developers.zenodo.org/#oai-pmh

Minimal FAIR interoperability: RO-Crate

The RO-Crate (Research Object Crate) specifies a method of aggregating and describing research data with associated metadata.

[1] https://doi.org/10.3233/DS-210053

Minimal FAIR interoperability: RO-Crate

RO-Crates can be stored, transferred or published in multiple ways, e. g. downloadable ZIP files. [1]

[1] https://doi.org/10.3233/DS-210053

The RO-Crate Metadata File

RO-Crates describe data with metadata to aid in discovery, re-use and long term management of data.

The core of RO-Crate is the RO-Crate Metadata File `ro-crate-metadata.json`.

This file must be present in the root directory of e. g. the archived Zip file. It contains structured JSON-LD metadata about the dataset.[1]


```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [

 {
    "@type": "CreativeWork",
    "@id": "ro-crate-metadata.json",
    "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
    "about": {"@id": "./"}
 },
 {
    "@id": "./",
    "identifier": "https://doi.org/10.4225/59/59672c09f4a4b",
    "@type": "Dataset",
    "datePublished": "2017",
    "name": "Data files associated with ...",
    "description": "Description ...",
    "license": {"@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/"}
 },
 {
  "@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
  "@type": "CreativeWork",
  "description": "This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0
Australia License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/au/ or
send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.",
  "identifier": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
  "name": "Attribution-NonCommercial-ShareAlike 3.0 Australia (CC BY-NC-SA 3.0 AU)"
 }
 ]
}
}
```


[2]

[1] https://www.researchobject.org/ro-crate/1.1/introduction.html

[2]https://www.researchobject.org/ro-crate/1.1/root-data-entity.html#minimal-example-of-ro-crate

JSON-LD: Context for JSON keys

@context [1]

[1] https://www.youtube.com/watch?v=vioCbTo3C-4

The RO-Crate Metadata File

The JSON-LD @graph array describes data entities and contextual entities, cross-referenced using @id.

RO-Crate relies heavily on Schema.org vocabulary.

```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [

 {
    "@type": "CreativeWork",
    "@id": "ro-crate-metadata.json",
    "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
    "about": {"@id": "./"}
 },
 {
    "@id": "./",
    "identifier": "https://doi.org/10.4225/59/59672c09f4a4b",
    "@type": "Dataset",
    "datePublished": "2017",
    "name": "Data files associated with ...",
    "description": "Description ...",
    "license": {"@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/"}
 },
 {
  "@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
  "@type": "CreativeWork",
  "description": "This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Australia License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/au/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.",
  "identifier": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
  "name": "Attribution-NonCommercial-ShareAlike 3.0 Australia (CC BY-NC-SA 3.0 AU)"
 }
 ]
}
}
```

[1]

[1]https://www.researchobject.org/ro-crate/1.1/root-data-entity.html#minimal-example-of-ro-crate

The RO-Crate Metadata File

The JSON-LD @graph array describes data entities and contextual entities, cross-referenced using @id.

RO-Crate relies heavily on Schema.org vocabulary.


```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [

 {
    "@type": "CreativeWork",
    "@id": "ro-crate-metadata.json",
    "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
    "about": {"@id": "./"}
 },
 {
    "@id": "./",
    "identifier": "https://doi.org/10.4225/59/59672c09f4a4b",
    "@type": "Dataset",
    "datePublished": "2017",
    "name": "Data files associated with ...",
    "description": "Description ...",
    "license": {"@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/"}
 },
 {
  "@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
  "@type": "CreativeWork",
  "description": "This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0
Australia License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/au/ or
send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.",
  "identifier": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
  "name": "Attribution-NonCommercial-ShareAlike 3.0 Australia (CC BY-NC-SA 3.0 AU)"
 }
 ]
}
}
```


[1]

[1] https://www.researchobject.org/ro-crate/1.1/root-data-entity.html#minimal-example-of-ro-crate

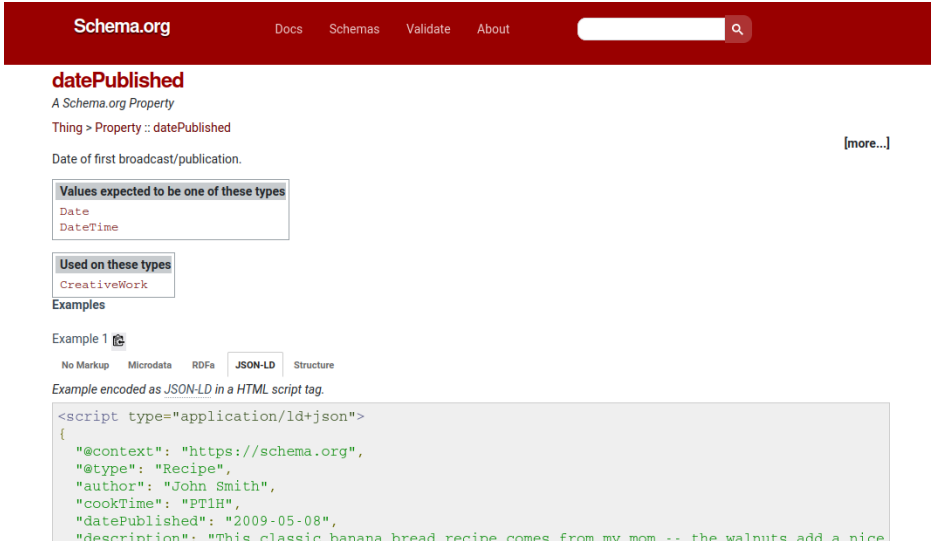
The RO-Crate Metadata File

The JSON-LD @graph array describes data entities and contextual entities. **RO-Crate relies heavily on Schema.org vocabulary.**

```
"key": "value"
```

```
"datePublished": "2017"
```

```
"https://schema.org/datePublished": "2017"
```



[1] Image source: https://schema.org/datePublished, snapshot taken 8 JUN 2022

Domain specific metadata terminologies & standards

Now, let‘s try and look up some **domain specific metadata templates** ...

TASK 2: Domain specific metadata terminologies & standards

1. Open one of these metadata standard registries in your preferred browser:
   - FAIRsharing.org
   - RDA Metadata Directory
   - RDA Metadata Standards Catalog
   - RDA Metadata Directory
   - DCC List of Metadata Standards
2. Search for a metadata schema, standard or vocabulary **relevant to your research domain**.
3. Inspect the **information provided**.
4. Take notes to **discuss your findings** with the group. Did you get any 404 (not found) responses clicking on links? Do you want to try a Google search in addition? Share some of your notes below.

Notes:

Recap

- ✓ data & metadata
- ✓ types of metadata
- ✓ unstructured & structured metadata records
- ✓ (web) locations
- ✓ metadata and the web
- ✓ finding metadata standards

HedgeDoc Handout

Further Reading

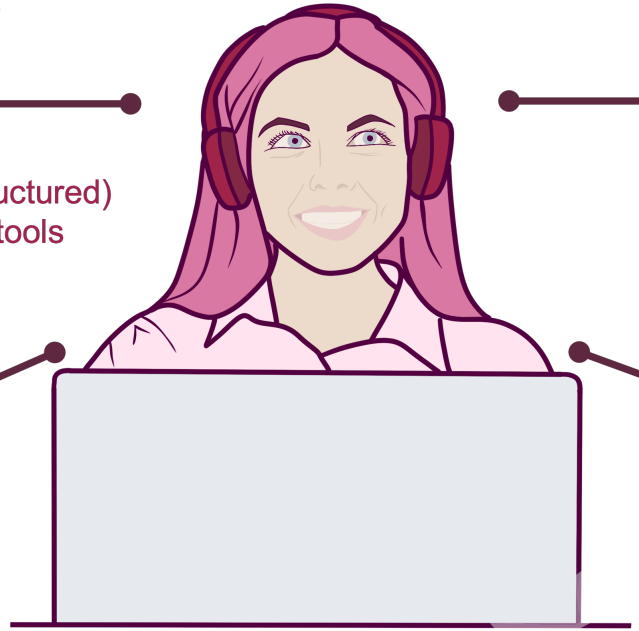
1. Pomerantz, J. (2015). Metadata. Cambridge, MA: MIT Press. ↩
2. Zhang, A. B., Gourley, D. (2008). "Metadata strategy" in Creating Digital Collections. Sawston, UK: Woodhead Publishing. ↩ ↩ ↩
3. Chadwick, I. (2016). "Research Data Management: guide to writing "readme" type metadata." The Open University. https://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/RDM-Guidelines-for-creating-readme-style-metadata.pdf ↩

Fundamentals of scientific metadata: why context matters

HMC HELMHOLTZ METADATA COLLABORATION

Training Data
Practice metadata annotation with exciting data sets.

Hands-on
Learn to write Markup, JSON & XML



Lectures
Learn about data, (structured) metadata, standards, tools and the internet work.

Lightning Talk
Get inspired by invited experts

Online Training
Join the class from wherever you are

Networking
Meet young professionals who share your interests

NEXT TRAINING

Incubator Summer Academy
From Zero to Hero

12. – 23. September 2022

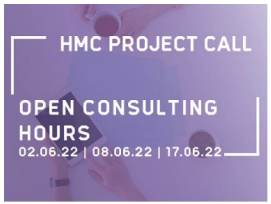
Annual HMC Project call

HMC is looking for projects that:

- address practical challenges in (meta)data generation, curation and enrichment
- have a core idea that adds new scientific knowledge to the respective field
- are promising to be generalizable & integratable into HMC in the long term

NUMBERS

- volume: 400 kEUR (200 + 200kEUR)
- duration: 2 years
- deadline: 06.07.2022
- funded proposals in 2021 ~ 26%



further information:
https://www.helmholtz-metadaten.de

Thank you!

www.helmholtz-metadaten.de
HMC@fz-juelich.de
@helmholtz_hmc

a.strupp@fz-juelich.de
s.gerlich@fz-juelich.de