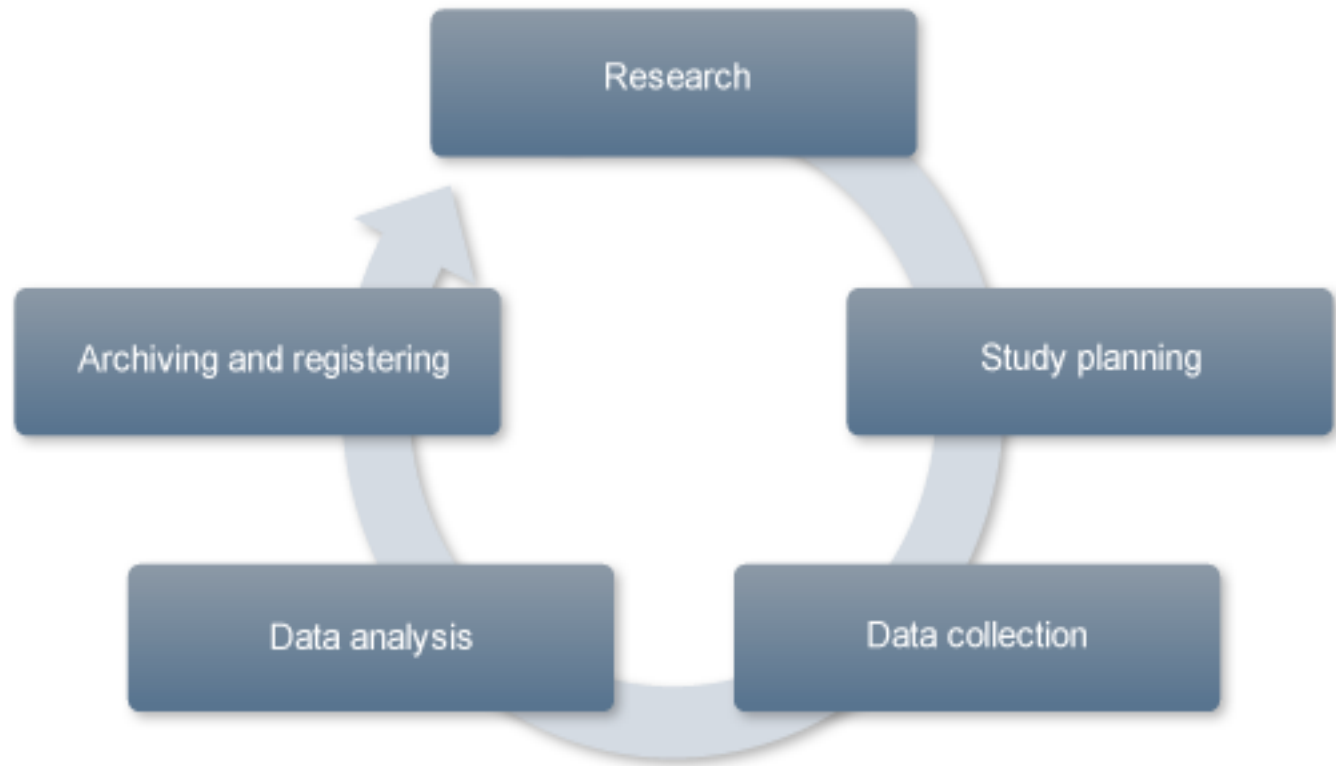
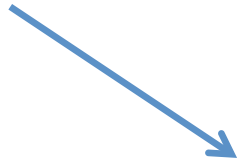


*DublinCore Conference 2014
Austin, Texas, USA*

Dublin Core Metadata for Research Data – Lessons Learned in a Real-World Scenario with datorium

Andias Wira Alam
andias-wira-alam@gesis.org

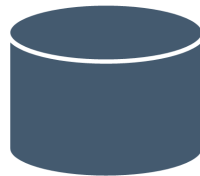
GESIS provides services...



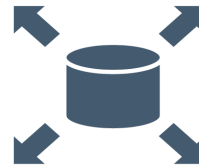
...a user-friendly tool for the autonomous documentation, upload and publication of research data.



Documentation and
Upload



Storage and
Review

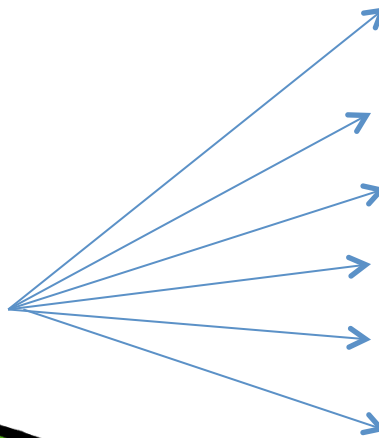


Publication



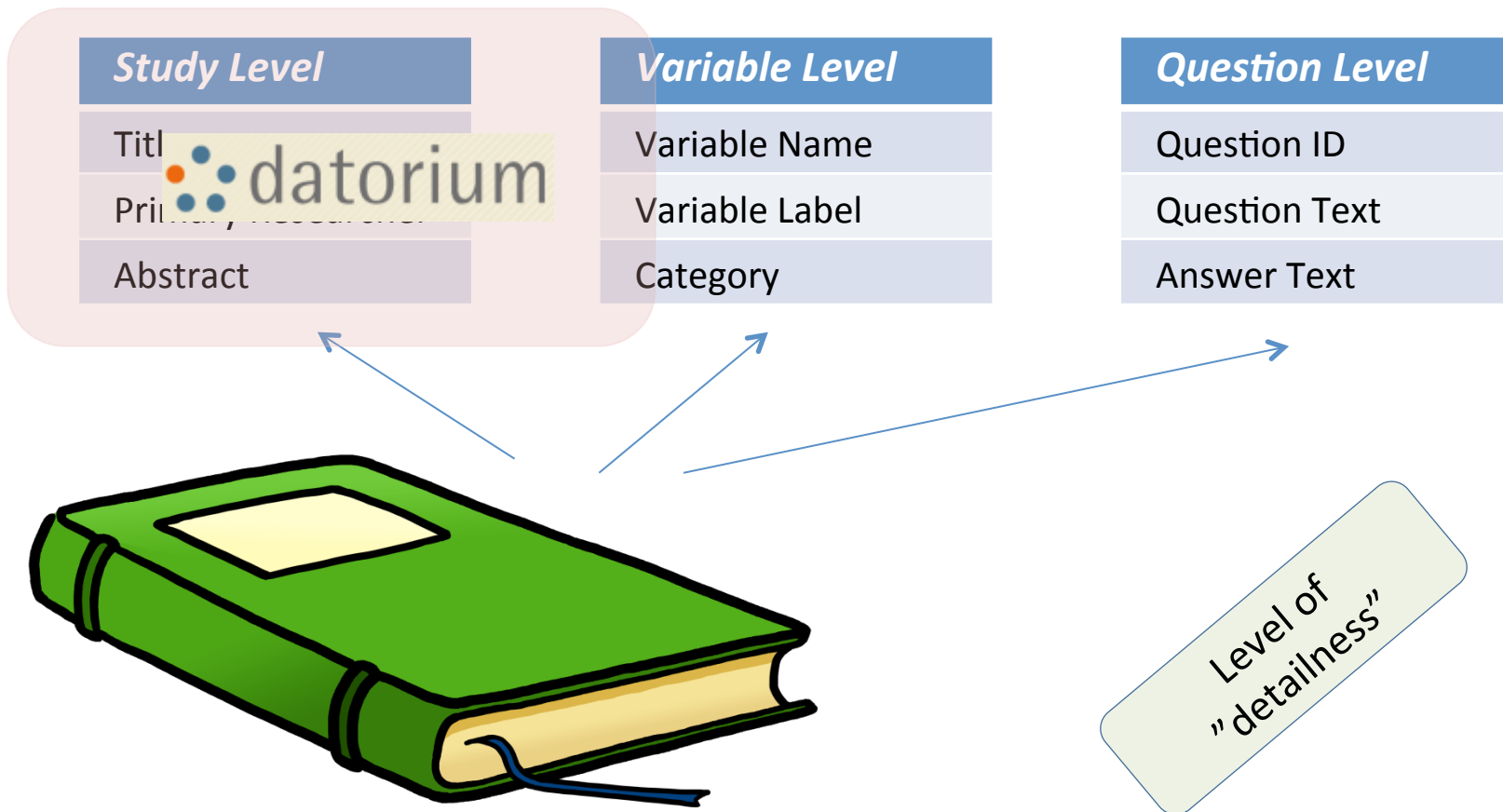
Search and
Download

Describing Document



<i>Metadata</i>
Title
Author
Abstract
ISBN
Publisher
...

Describing (Research) Data



Ideal Vision

Any piece of information in the research data should be well-documented and described.

Usability vs. Completeness



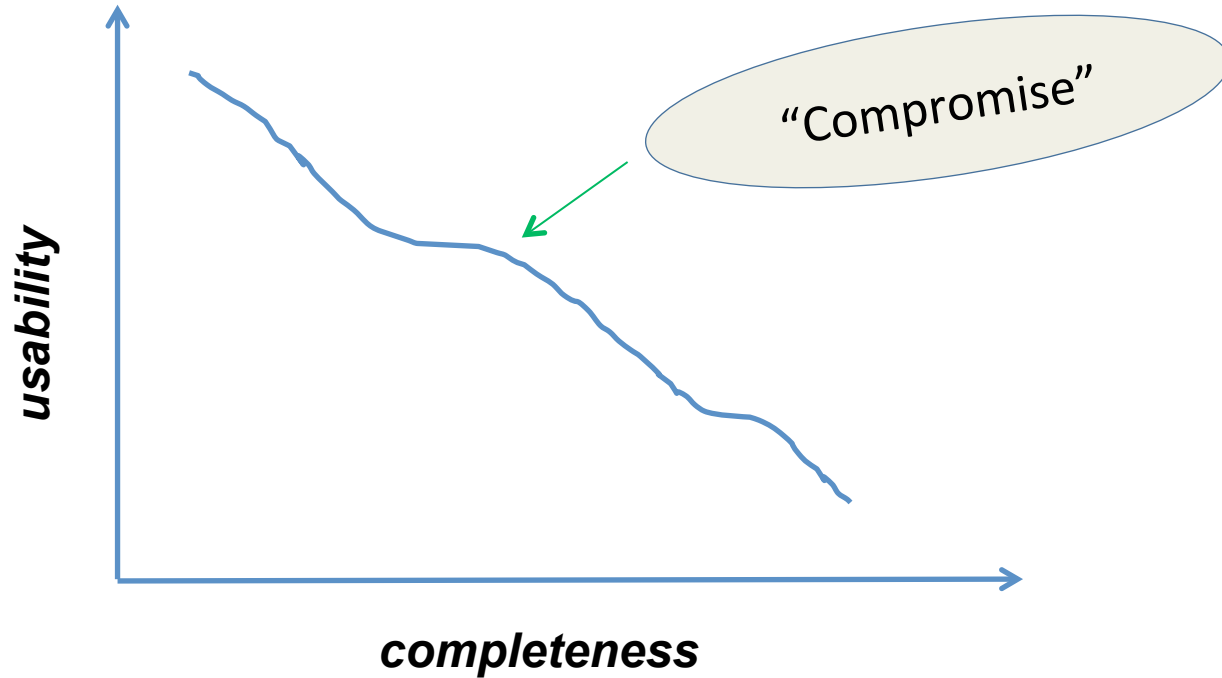
I am simple

VS.



I want details

Usability vs. Completeness



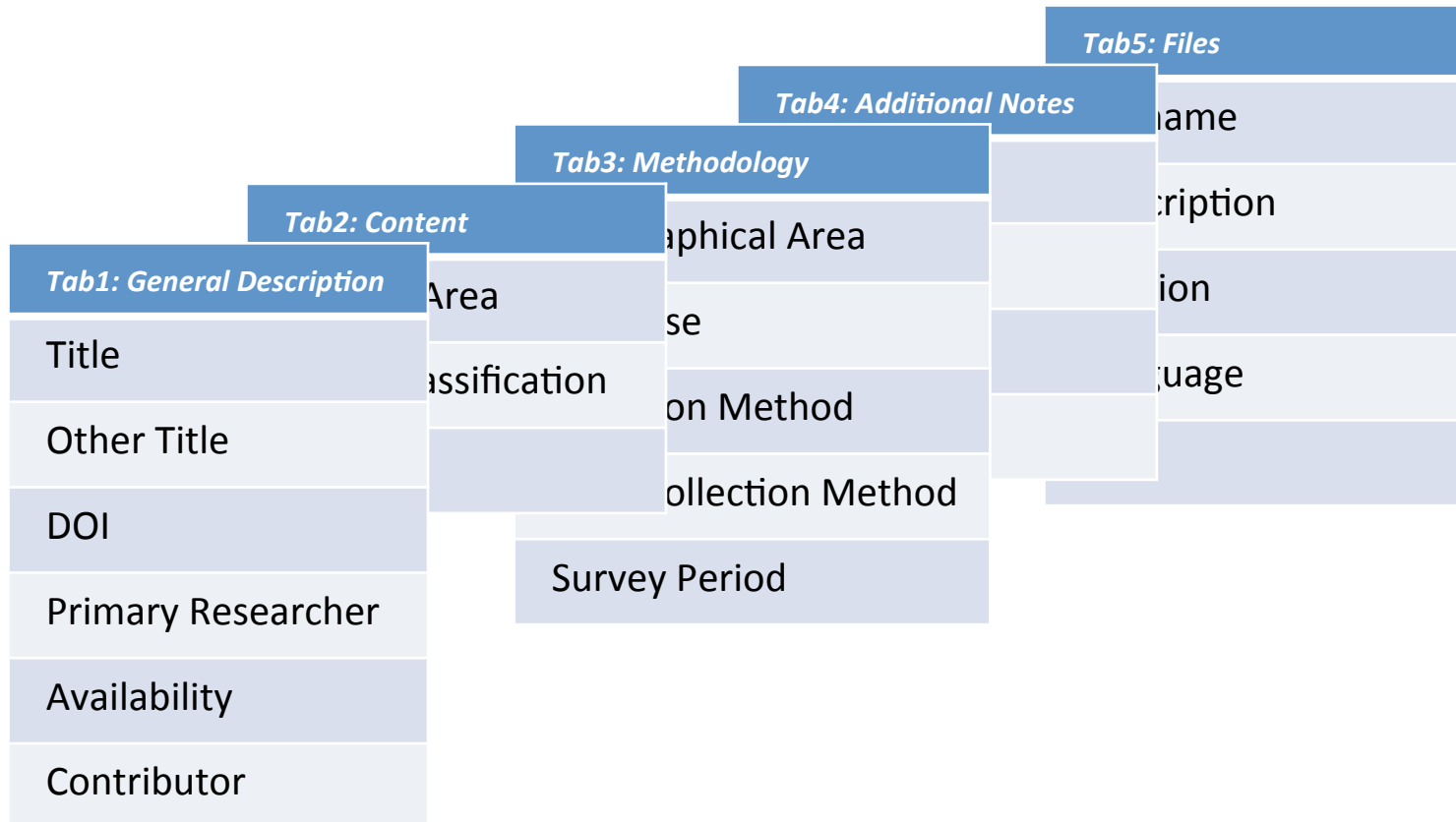
Why Dublin Core?

DC metadata as a fundamental metadata infrastructure for scientific publications is therefore slowly evolving into an infrastructure for research data as well.

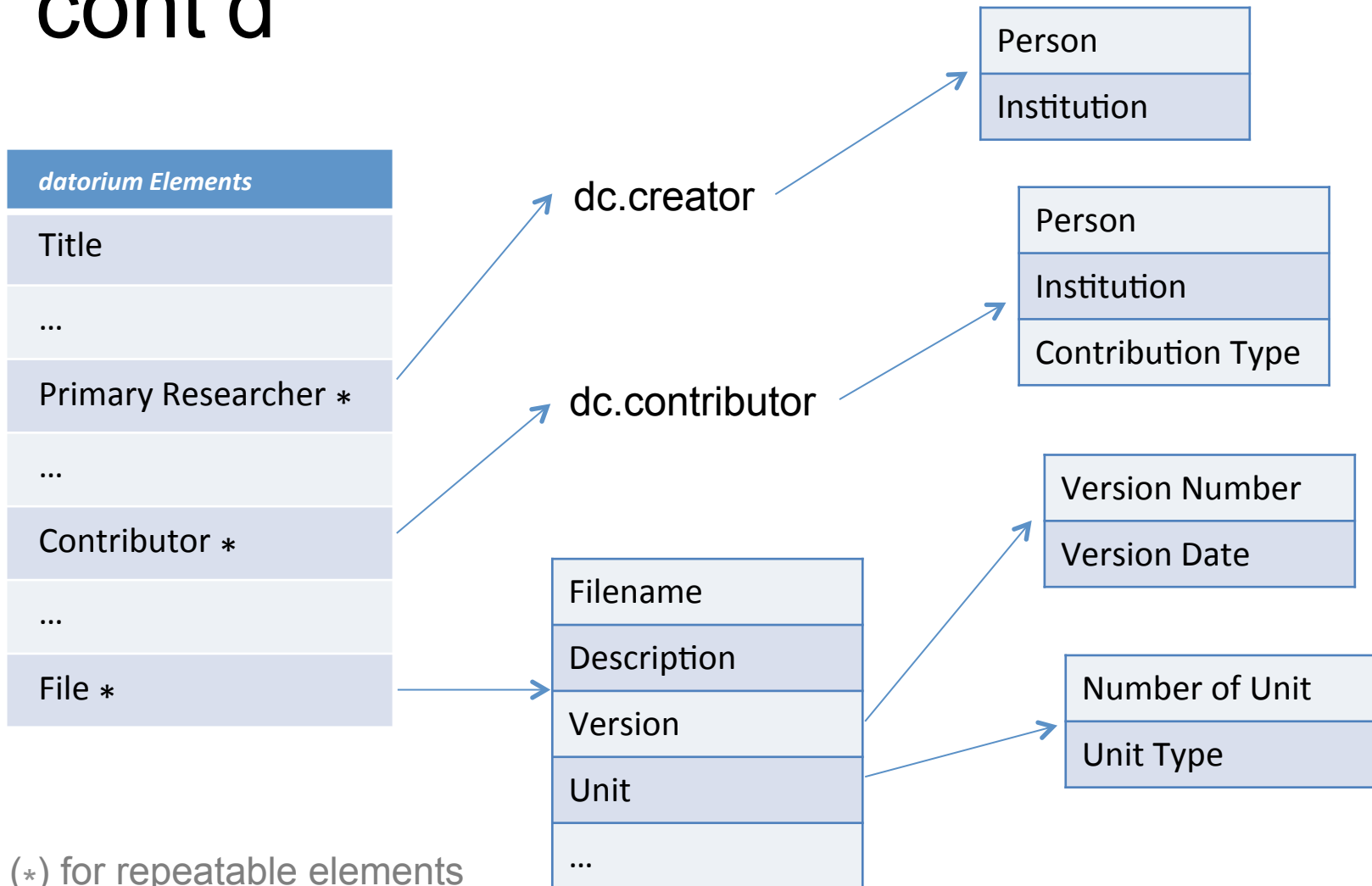
Dublin Core for Users?

Users need platforms which allow them to publish their data in an easy way, making the data visible and citable.

Well-documented and described?



cont'd



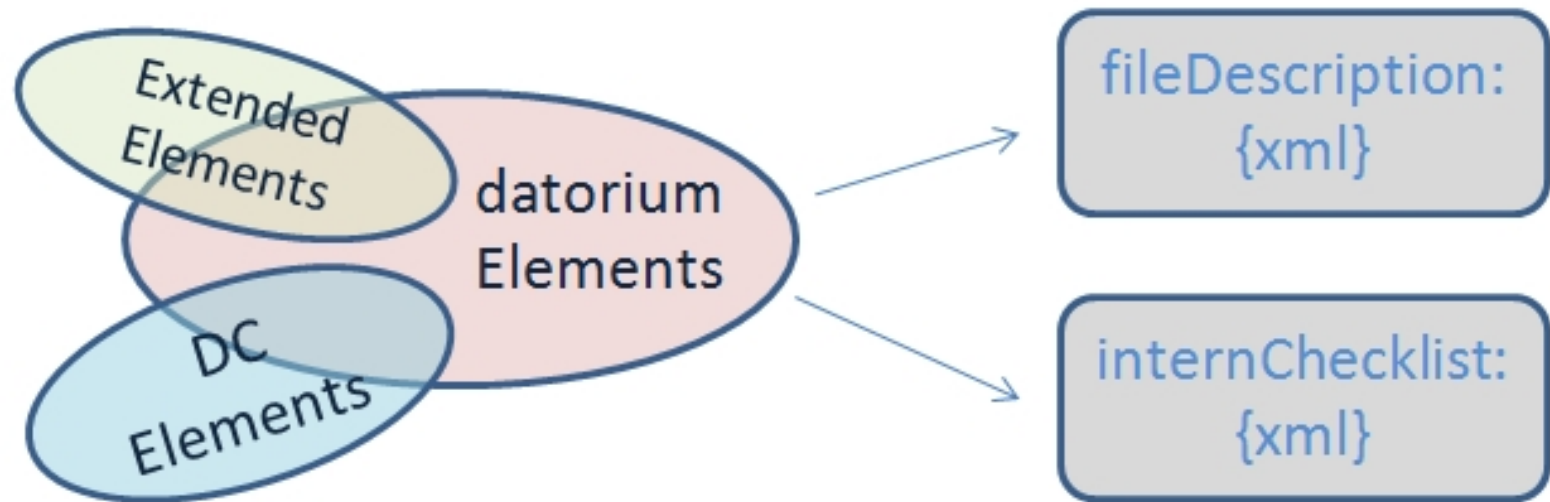
(*) for repeatable elements

but, DublinCore ...

is simple and flat.

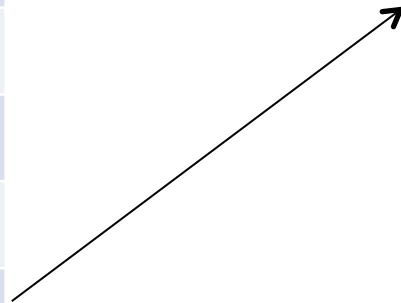
How do we cover complex description?

Schema Overview



cont'd

DATORIUM Elements
...
...
...
...
File



```

<?xml version="1.0" encoding="UTF-8"?>
<file id="1">
  <fileName>mySPSS_File01.spss</fileName>
  <fileDesc>All survey datasets</fileDesc>
  <fileMD5>719a47df73860ff179de21f42fd1fa150</fileMD5>
  <fileLang>en_US</fileLang>
  <fileVersion>
    <fileVersionNumber>1.0</fileVersionNumber>
    <fileVersionDate>2014-10-09</fileVersionDate>
  </fileVersion>
  <fileSoftware id="1">IBM SPSS 22</fileSoftware>
  <fileSoftware id="2">GNU R 3.1.1</fileSoftware>
  <fileResource>
    <fileResourceType></fileResourceType>
    <fileResourceTypeGeneral>Dataset</fileResource>
  </fileResource>
</file>

```


Labels	DC Elements	Extended Elements	Type
Title	dc.title	-	text
Other Title *	-	dbk.othertitle dbk.othertitle.text dbk.othertitle.type	text text text
DOI	dc.identifier.uri	-	URI
Primary Researcher *	dc.creator	dbk.primaryresearcher.person dbk.primaryresearcher.institution	text text
Publisher	dc.publisher	-	text
Publication Year	-	dbk.publicationyear	date: YYYY
Availability	-	dbk.availability	text
Embargo	-	dbk.embargo.availability	text
Embargo (until)	-	dbk.embargo.end	date: YYYYMMDD
Contributor *	dc.contributor	dbk.contributor.person dbk.contributor.institution dbk.contributor.type	text text text
Subject Area *	dc.subject.other	-	text
Topic Classification *	dc.subject.classification	-	text
Abstract	dc.description	-	text
Geographical Area *	dc.coverage.spatial	-	text
Universe *	-	dbk.universe	text
Selection Method	-	dbk.selectionmethod	text
Data Collection Method *	-	dbk.datacollectionmethod	text
Survey Period *	-	dbk.surveyperiod dbk.surveyperiod.start dbk.surveyperiod.end	date date
Rights *	dc.rights	-	text
Notes *	-	dbk.notes dbk.notes.text dbk.notes.type	text text text

(*) for repeatable elements

cont'd

Labels	DC Elements	Extended Elements	Type
Source *	-	dbk.source	text
Publications *	-	dbk.publication dbk.publication.text dbk.publication.id	text text
File *	-	dbk.file dbk.file.filename dbk.file.filedescription dbk.file.version dbk.file.versionNumber dbk.file.versionDate dbk.file.resource dbk.file.resourceType dbk.file.resourceTypeGeneral dbk.file.language dbk.file.numberofvariables dbk.file.unit dbk.file.unitNumberOf dbk.file.unitType dbk.file.software dbk.file.alternateld dbk.file.relatedld dbk.file.relatedldIdentifier dbk.file.relatedldType	text text text text date: YYYYMMDD text text text text text int text int text text text text text
Date Issued (for sorting purpose)	dc.issued	-	date: YYYYMMDD
Checklist (for Curators only)	-	intern.cheklist	text

(*) for repeatable elements

cont'd

`dbk.other.title.type` – Type of other title can be selected from the controlled vocabularies provided by DBK (Zenk-Möltgen et al. 2012), such as “project title” or “original title”.

`dbk.availability` – It consists of three controlled vocabularies: “free access”, “restricted access”, and “embargo”.

`dbk.contributor.type` – Type of contributor is based on the category scheme of the ContributorType from DataCite.

`dc.subject.other` – Subject Area has been chosen from the disciplines in SSOAR - Social Science Open Access Repository.

`dc.subject.classification` – Topic Classification is based on DBK, consists of overall 38 terms, such as “Economic Systems” or “Social Policy”.

`dc.coverage.spatial` – Geographical Area consists of places, such as countries, cities, or provinces / states, based on ISO-3166.

cont'd

`dbk.datacollectionmethod` – Data Collection Method consists of the controlled vocabularies provided by DDI (unreleased beta version, March 2013), such as “Email interview”, “recording”, or “Telephone interview: CATI”.

`dbk.notes.type` – It is based on `DescriptionType` provided by DataCite, such as “Abstract” or “TableOfContents”.

`dbk.file.language` – Languages provided by ISO-639.

`dbk.file.unitType` – Unit Type is based on “Analysis Unit” provided by DDI, such as “Family”, “Individual”, or “Organization”.

`dbk.file.relatedIdType` – Type of the related identifier is based on the `RelationType` provided by DataCite, such as “IsCitedBy”, “IsDocumentedBy”, or “IsPartOf”.

`dc.issued` – Date Issued has been generated by the system at the time of publication.

Implementation

- DSpace 1.8.x as basis platform for the implementation
- JavaScript (jQuery) for front-end manipulation
- da|ra API for DOI registration

Front End Implementation (Input Types)

File Description

DATORIUM Elements

Title

Subject Area

Abstract

Geographical Area

Survey Period

File

File: No file chosen
Please enter the full path of the file on your computer corresponding to your item. If you click "Browse...", a new window will allow you to select the file from your computer.

File Description:
e.g., Questionnaire, data set, Methodological Report, etc.

Version Number: +
Number of the dataset version

Version Date:
Date of the version of the dataset

Resource Type: +
Description of file type

Resource Type General:
Please select the general resource type of the research data by choosing a category from the drop down menu

Language: +
Language(s) of the research dataset

Number of Variables:
Number of variables in the dataset

Number of Units:
Number of units in the dataset

Unit Type:
Please select an unit type by choosing a category from the drop down menu


Software: +
Software that is needed to use the dataset

Alternate ID: +
Additional other identifiers of the research dataset

Related ID: +
Persistent identifiers that are related to the dataset

Type of Relation:
Type of relation of resources der Ressourcen

Editor – Tab1

General Description	Content	Methodology	Additional Notes	Files	Verify	License
Describe Item						
Title: *	<input type="text" value="Political Survey"/>					
	<small>Main title of the data set</small>					
Other Title:	<input type="text"/>					<input type="button" value="+"/>
	<small>Additional titles of the data set</small>					
Other Title Type:	<input type="text" value=""/>					
	<small>e.g. project title, original title, subtitle ↗</small>					
Principle Investigator: <i>and / or *</i>	<input type="text"/>					<input type="button" value="+"/>
	<small>Responsible author of the research data (last name, first name)</small>					
Institution:	<input type="text"/>					
	<small>Institution of the Principle Investigator; e.g. GESIS.</small>					
	<input type="checkbox"/> Einstein, Albert; University of Relativity					
	<input type="button" value="Delete"/>					
Publisher:	<input type="text" value="GESIS - Leibniz-Institut für Sozialwissenschaften"/>					
	<small>Institution that publishes the research data</small>					
Publication Year: *	<input type="text" value="2014"/>					
	<small>Year of publication of the dataset</small>					
Availability: 	<input type="text" value="Free Access"/>					
	<small>The class of availability that the research data has</small>					

Editor – Tab2

The screenshot shows the 'Content' tab of the GESIS editor. The main interface has a navigation bar with tabs: 'General Description', 'Content' (active), 'Methodology', 'Additional Notes', 'Files', 'Verify', and 'License'. Below the navigation bar is a 'Describe Item' button. The main content area is divided into sections: 'Subject Area:', 'Topic Classification:', and 'Abstract:'. A dialog box titled 'Please select...' is open, showing a tree view of subject categories. The dialog box has a 'show all' checkbox and 'Cancel' and 'Done' buttons at the bottom.

Subject Selection Dialog:

- Please select...
- Select subjects: show all
- Social Sciences
- Interdisciplinary and Applied Fields of the Social Sciences
- Humanities
 - Philosophy, Ethics, Religion
 - Science of Literature, Linguistics
 - History
 - Other Fields of Humanities
- Jurisprudence and Administrative Science
- Science
 - No classification applied

Editor – Tab3

General Description ▶
Content ▶
Methodology ▶
Additional Notes ▶
Files ▶
Verify ▶
License ▶

Describe Item

Geographical Area: i +

- German Democratic Republic (1949-1990) / DDDE
- Germany / DE
- German Reich (1871-1945) / DXDE
- Historical G...

Universe:

Selection Method:

May

2014

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
					3	
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Data Collection Mode: +

Description of the... (n selection)

Description of the... (click in the input field and choose a... specification)

Survey Period: i +

Year
Year+Month
Remove

Start (from)

End (to)

Editor – Tab4

General Description ▶
Content ▶
Methodology ▶
Additional Notes ▶
Files ▶
Verify ▶
License

Upload File(s)

File: Choose File No file chosen

Please enter the full path of the file on your computer corresponding to your item. If you click "Browse...", a new window will allow you to select the file from your computer.

File Description:

e.g., Questionnaire, data set, Methodological Report, etc.

Version Number: +

Number of the dataset version

Version Date:

Date of the version of the dataset

Resource Type: +

Description of file type

Resource Type General:

Please select the general resource type of the research data by choosing a category from the drop down menu [↗](#)

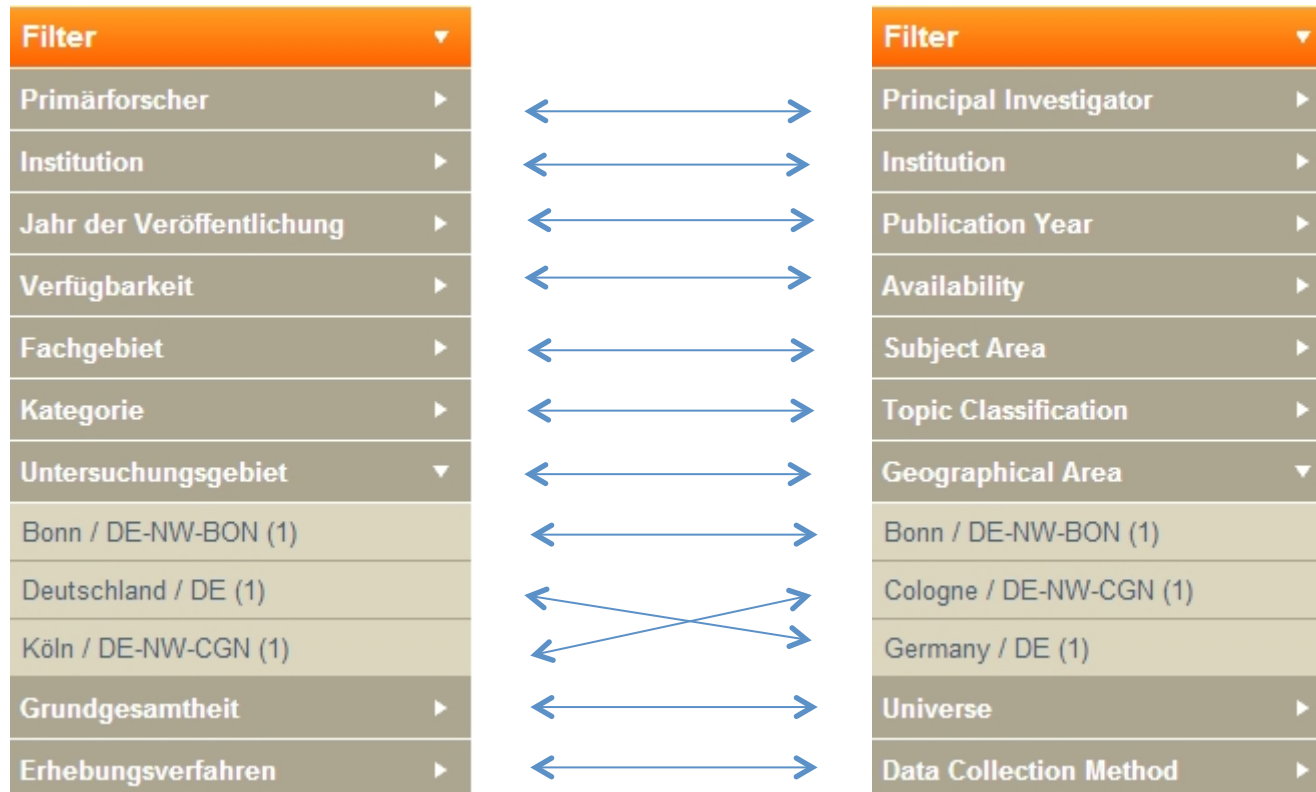
Language: i +

Language(s) of the research dataset [↗](#)

Number of Variables:

Number of variables in the dataset

Multi-language for facets (controlled vocabularies)



Curious? Please give
it a try! 😊



<https://datorium.gesis.org/>

Conclusion

- Dublin Core fits our purpose since it is a simple and appropriate schema for documenting research data.
- However, to meet all requirements, some extensions are needed.
- The front-end based manipulation compromises the limitation of the schema.

Acknowledgement

This project is funded by GESIS – Leibniz Institute for the Social Sciences. We thank our project members, *Monika Linne*, *Wolfgang Zenk-Möltgen*, *Oliver Hopt*, *Natascha Schumann*, *Stefan Müller*, *Reiner Mauer*, and *Martin Friedrichs*, for their useful inputs on requirements analysis and for their feedback and fruitful discussions during the implementation. Special thanks go to *Sigit Nugraha* who helped extensively with the technical implementations and *Astrid Recker* for proof reading this paper. Last but not least, we also thank the *DSpace developers* for their effort to continually develop this great repository system as an open source project. datorium can be accessed online at <https://datorium.gesis.org/>.

Thank You for Your Attention!

*I am looking forward to your remarks,
questions, critics, feedback.*