DATA DOCUMENTATION INITIATIVE

DDI Lifecycle

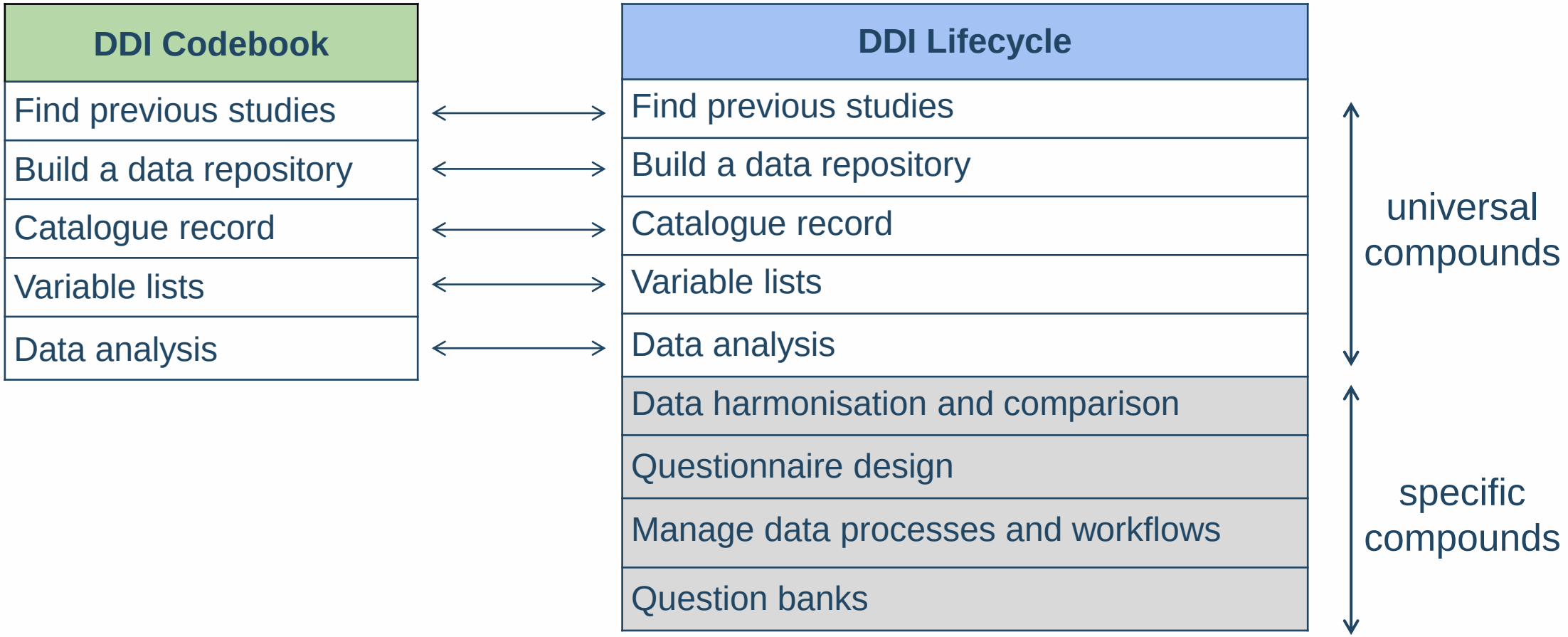
INTRODUCTION AND OVERVIEW

Hayley Mills, CLOSER, UCL

Contents

- Difference between DDI Codebook and DDI Lifecycle
- Features of DDI Lifecycle
- Activities DDI Lifecycle supports
- Summary

Difference between DDI Codebook and DDI Lifecycle



DATA DOCUMENTATION INITIATIVE

DDI Codebook

Unstructured un-standardised metadata

Lost metadata manager

Structured DDI-C metadata

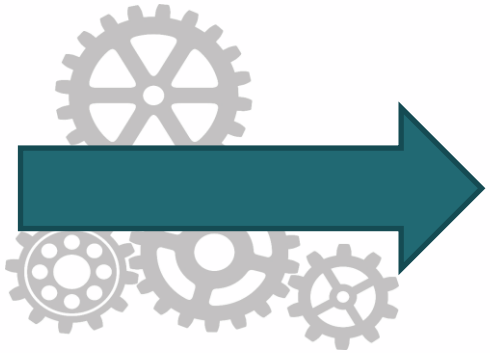
- Basic descriptive content for variables, files, source material, and study level information.
- Supports discovery, preservation, and the informed use of data.
- Documentation of a simple study data.

DATA DOCUMENTATION INITIATIVE

DDI Lifecycle - Broken down into different functions

Unstructured un-standardised metadata

Lost metadata manager



- Information is broken up into more discreet content.
- DDI-Lifecycle is modular and extensible.
- Metadata reuse.

DATA DOCUMENTATION INITIATIVE

DDI Lifecycle - Reuse

Structured DDI-L metadata

Metadata specified by DDI Lifecycle standard

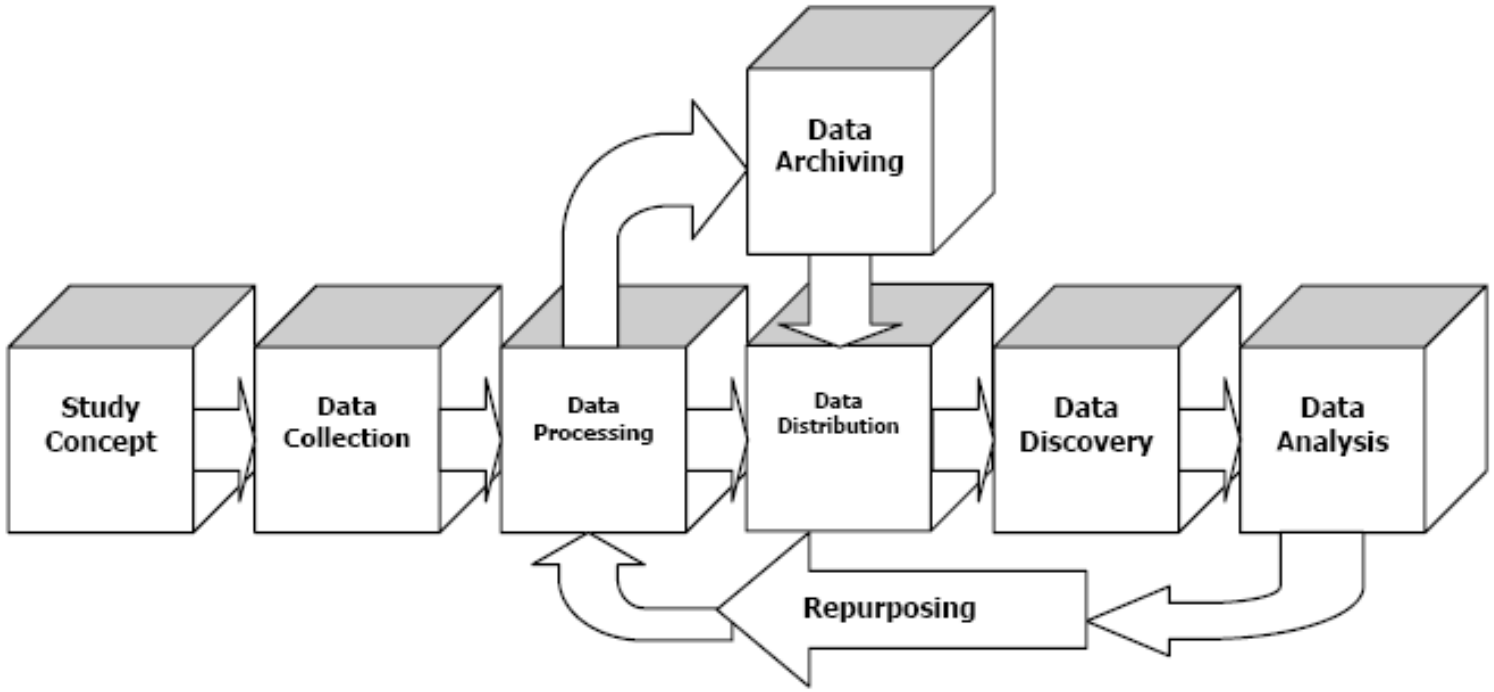
- Document and *manage* data across the entire data life cycle.
- Supports preservation, discovery, access and analysis in a much easier way.
- Documentation of repeated studies/surveys

DATA DOCUMENTATION INITIATIVE

Features of DDI Lifecycle

- Lifecycle expands on the idea of Codebook in terms of:
  - Content coverage and depth
  - Allows grouping and comparing related studies or series of studies
  - Metadata management over time
  - Reusable metadata
  - Manage questions and questionnaires
  - Support for the planning, capture, processing, storage, discovery and dissemination of data (allows for active management of metadata)

Data lifecycle model



DATA DOCUMENTATION INITIATIVE

Content coverage and depth

- **Conceptual objects**: concept, unit, unit type, universe, population, geographic structures, and representation
- **Methodological objects:** approaches to sample selection, data capture, weighting, quality control, and process management
- **Quantitative and qualitative data objects:** concept, universe, representation, usage, data type, record, record relationships, storage, access, and descriptive statistics
- **Data management:** ownership, access, rights management, restrictions, quality standards, organization, agent management, relationship between products, versioning, and provenance
- **Processing:** data capture, data processing, analysis, and data management

Metadata management over time

NACDA-ICPSR Portal

Name reascano7

Label CG1D SP CANT ANS SP TOO ILL

Statistics | Code Comparison | Correspondence Tree

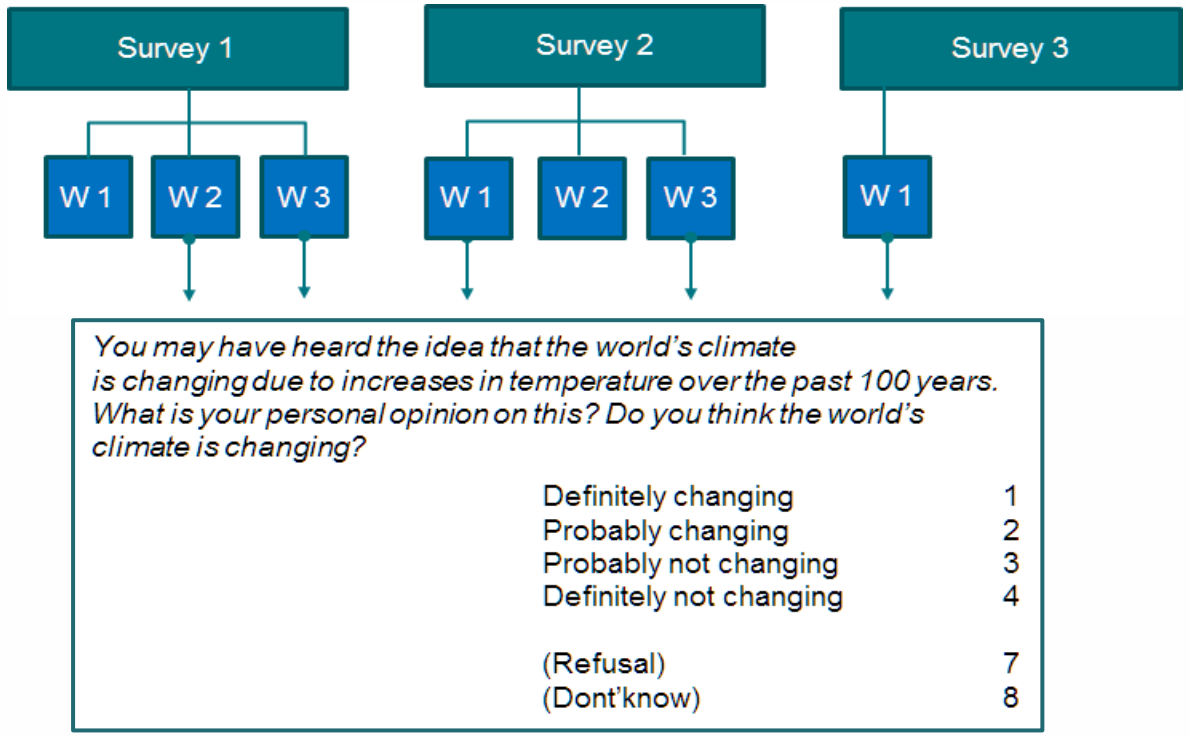
	National Health and Aging Trends Study (NHATS)							
	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8
	R1 SP cg1reascano7	R2 SP cg2reascano7	R3 SP cg3reascano7	R4 SP cg4reascano7	R5 SP cg5reascano7	R6 SP cg6reascano7	R7 SP cg7reascano7	R8 SP cg8reascano7
-9 Missing	-9	-9	-9	-9	-9	-9	-9	-9
-8 DK	-8	-8	-8	-8	-8	-8	-8	-8
-7 RF	-7	-7	-7	-7	-7	-7	-7	-7
-1 Inapplicable	-1	-1	-1	-1	-1	-1	-1	-1
1 Yes	1		1	1	1	1	1	1
1 SP HAS DEMENTIA/ALZHEIMER'S/ NOT ABLE		1						
2 No	2		2	2	2	2	2	2
2 SP IS UNABLE TO SPEAK		2						
3 SP IS UNABLE TO HEAR		3						
4 SP REFUSED		4						
5 PROXY REFUSED		5						
6 SP NOT PRESENT		6						
7 SP TOO ILL		7						
8 SP LANGUAGE BARRIER		8						
91 OTHER (SPECIFY)		91						

DATA DOCUMENTATION INITIATIVE

https://harmonize.icpsr.umich.edu/

Reusable metadata

- Reuse of a question in different waves/rounds/sweeps of a survey and across surveys.



Manage questions and questionnaires

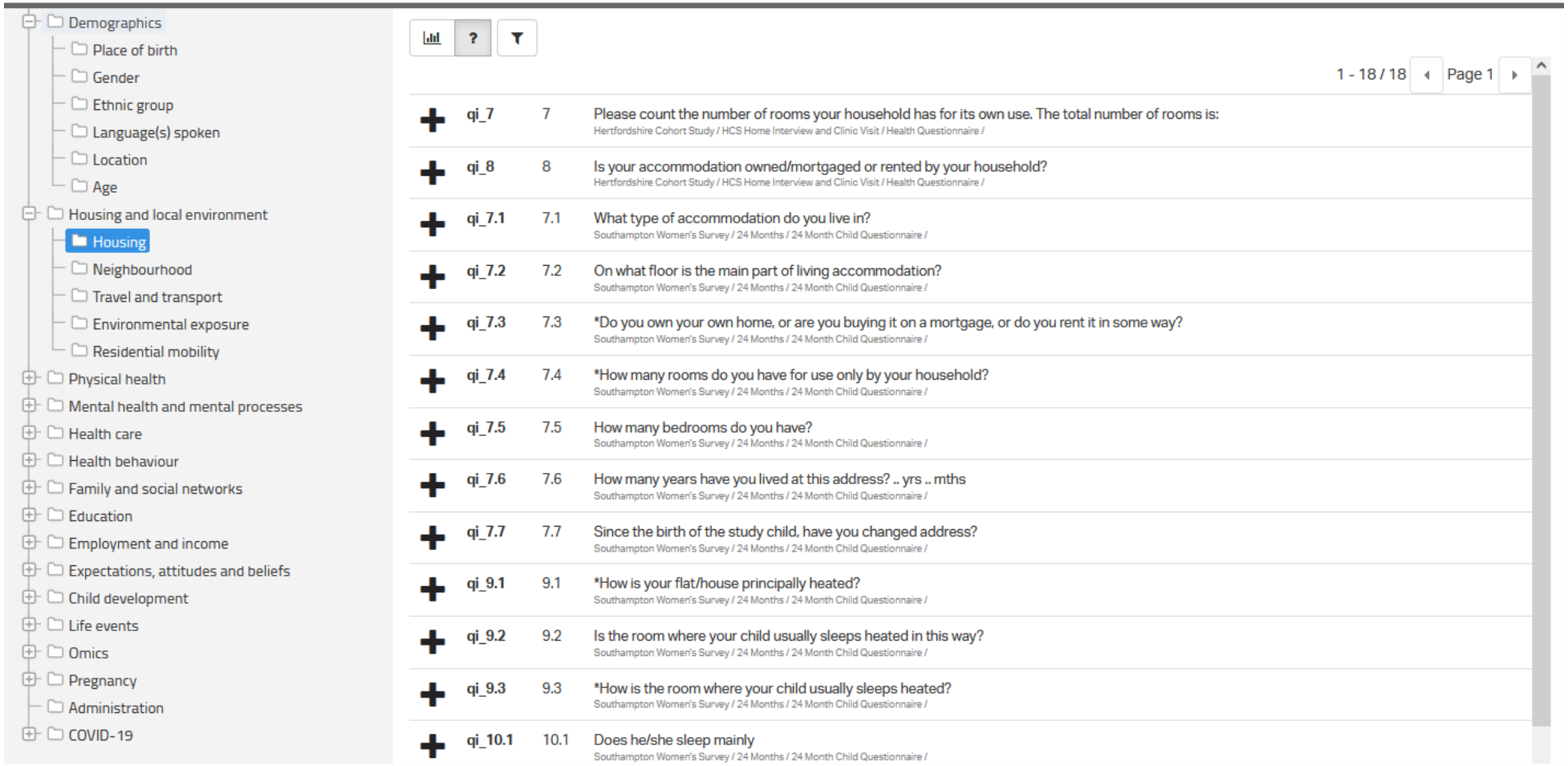
Topic	DDI Codebook	DDI Lifecycle
Question	qstn	QuestionItem, QuestionGrid, QuestionBlock
Non-question measures		MeasurementItem
Questionnaire		Instrument
Questionnaire routing		Sequence, IfThenElse, Loop, RepeatUntil, RepeatWhile
Questionnaire content usage		QuestionConstruct, MeasurementConstruct, ComputationItem, StatementItem

DATA DOCUMENTATION INITIATIVE

Allows grouping and comparing related studies or series of studies

closer Discovery — Explore the content of UK longitudinal studies

UKRI Economic and Social Research Council



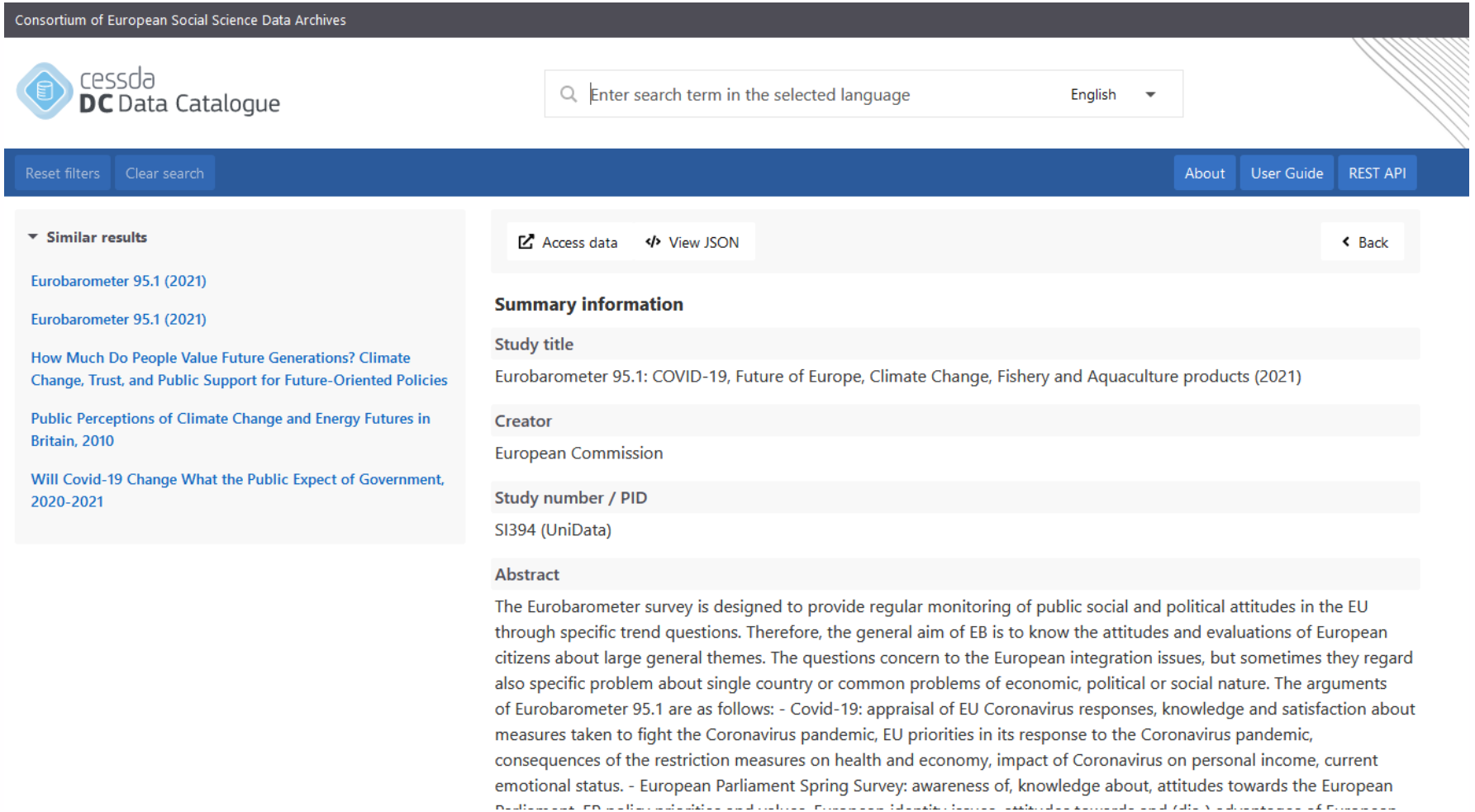
DDI DATA DOCUMENTATION INITIATIVE

https://discovery.closer.ac.uk/explore/

Activities DDI Lifecycle supports

- Data catalogue - Archival preservation of descriptive and production content
- Structured human-readable data dictionary
- Metadata driven statistical systems
- Descriptive documentation of the content, meaning, provenance, and access
- Discovery and exchange of data at the study, data file, variable, and question level
- Question banks

Example - Create a Data catalogue



DATA DOCUMENTATION INITIATIVE

https://datacatalogue.cessda.eu/

Example- Create a data dictionary

PDF basket download from Colectica portal

own89 - Ownership of house 1989

Type	Code
qi_12	Who owns it?
Study	National Survey of Health and Development
Sweep	1989 (Age 43)
Dataset	1989 Main Questionnaire Dataset
Variable Group	1989 Main Questionnaire Dataset
Variable Group	102 - Housing and local environment
Variable Group	10201 - Housing

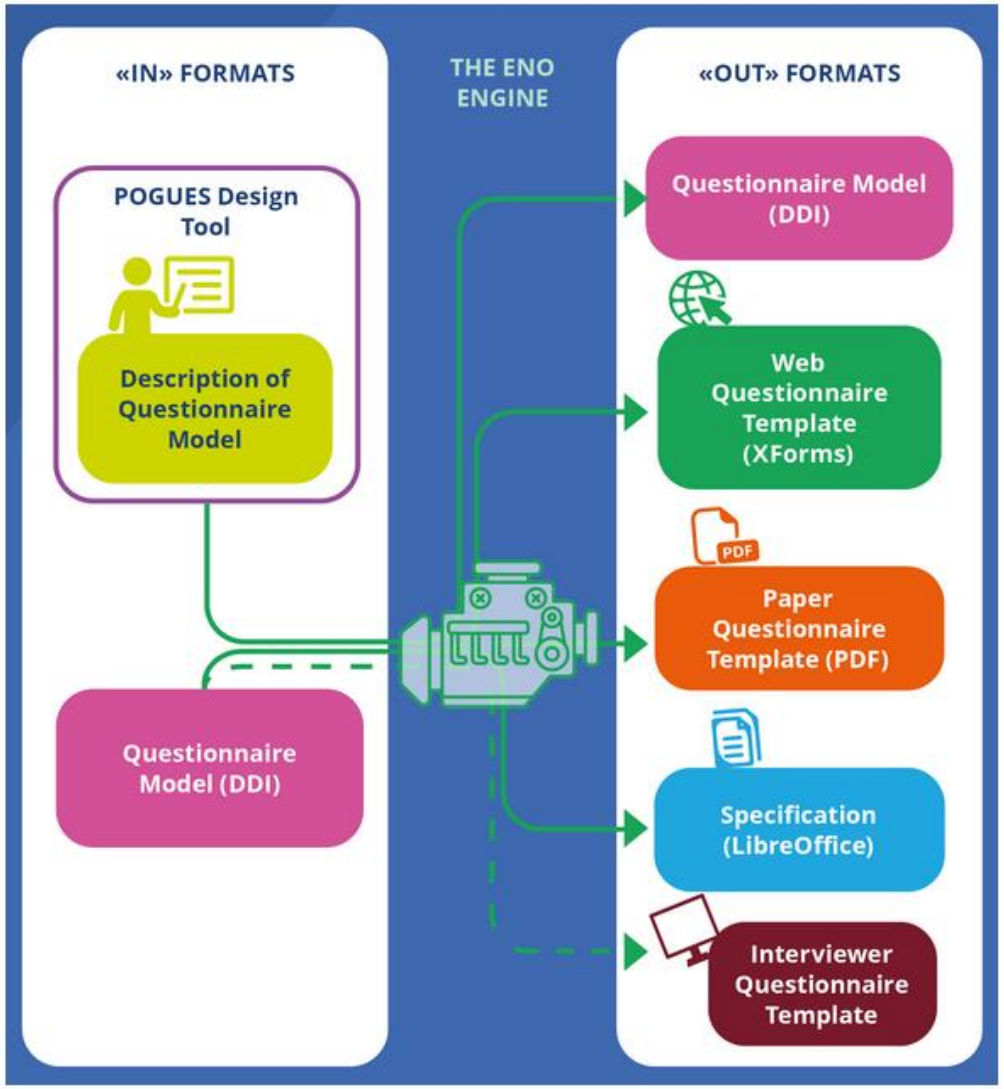
			Frequency
Valid	0	Owns it or is buying it	2,687
	1	Renting it from the Council	294
	2	Renting it from a relative	21
	3	Renting it from a private landlord	104
	4	Renting it from a housing association	14
	5	Other, specify	137
	9	Unknown	5

Valid	Invalid	Minimum	Maximum
3262	2100	0	9

DATA DOCUMENTATION INITIATIVE

Example - Create a metadata driven statistical system

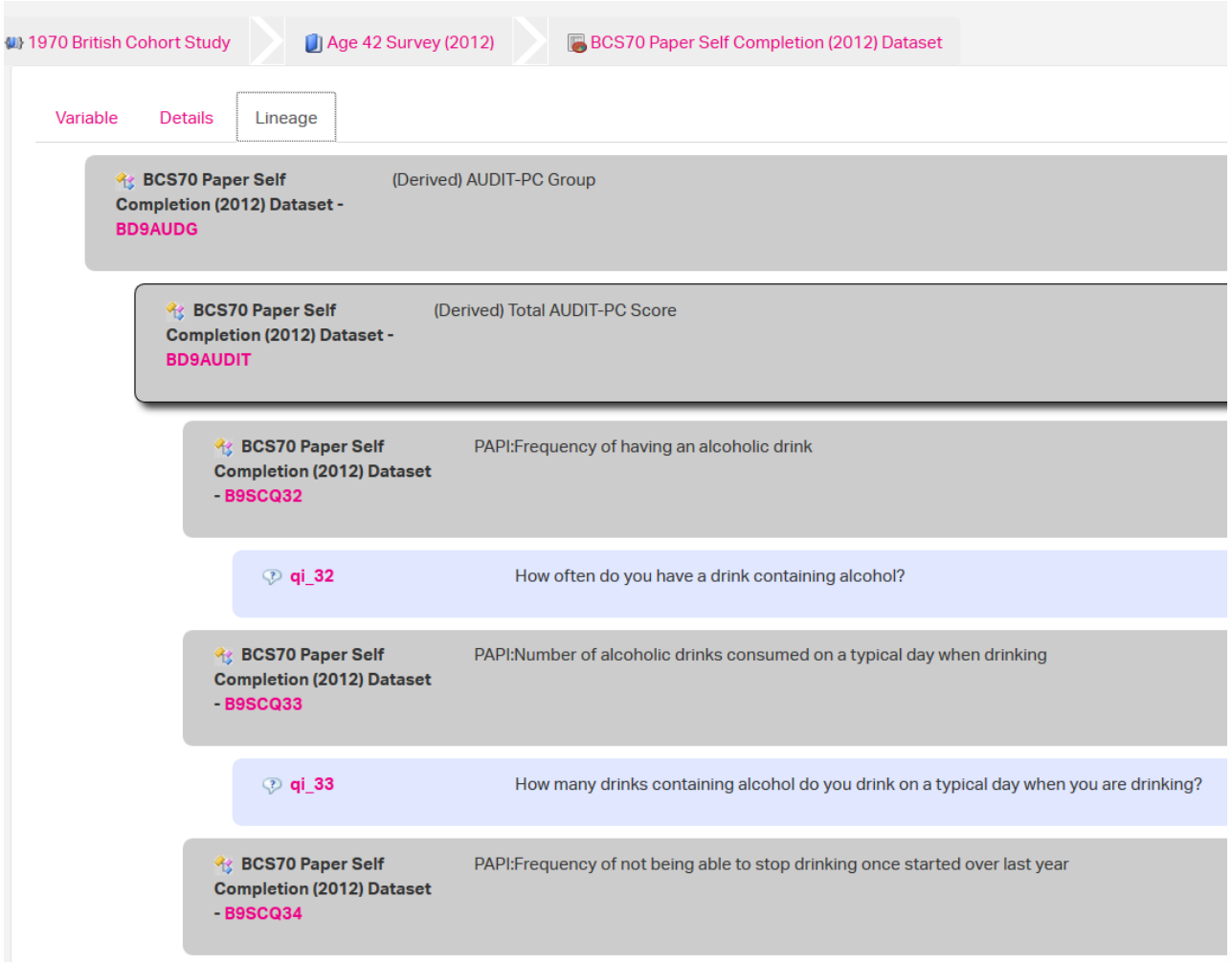
INSEE - Architecture of the Eno Engine



https://www.insee.fr/en/information/5014703?sommaire=5014796

Example - Provide descriptive documentation of the content, meaning, provenance, and access

CLOSER Discovery lineage



DATA DOCUMENTATION INITIATIVE

Summary

- Supports the whole data lifecycle for the planning, capture, processing, storage, discovery and dissemination of data
- More suitable for:
  - multi-series/study
  - projects with surveys
  - larger projects
- Allows active management of metadata
  - Create products and processes for efficiency, consistency, quality, transparency and discovery

DATA DOCUMENTATION INITIATIVE

Image Attributions

Lego bricks	“LEGOS workers fired from job" by woodleywonderworks is licensed under CC BY 2.0
Lego excavator	"Lego Excavator (8853)" by Stephen Edmonds is licensed under CC BY SA 2.0
Lego sorted	“Technic bits“ by Windell Oskay is licensed under CC BY 2.0
Lego race car	"Lego Rally Shock & Roll Racer (8840)" by Stephen Edmonds is licensed under CC BY SA 2.0
Lego bulldozer	“Lego Motorized Bulldozer (8275)" by Stephen Edmonds is licensed under CC BY SA 2.0
Lego control centre	"Lego 8094 Control Centre (4)" by Stephen Edmonds is licensed under CC BY SA 2.0

Credits: DDI Train the Trainer Workshop 2018

Alina Danciu
Guillaume Duffes
Adrian Dușa
Lauren Eickhorst
Dan Gillman
Arofan Gregory
Taras Günther
Lea Sztuk Haahr
Sanda Ionescu
Jon Johnson
Chifundo Kanjala
Kaia Kulla
Amber Leahey
Alexandre Mairot
Johan Fihn Marberg
Hayley Mills
Olof Olofsson
Hilde Orten
Anja Perry
Dan Smith
Wendy Thomas
Joachim Wackerow
Knut Wenzig

Wendy Thomas chair of the Technical committee

DATA DOCUMENTATION INITIATIVE

Thank you

QUESTIONS?