



# Handling Metadata

**Frans de Liagre Böhl**

**&**

**Ron Scholten**

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



**Universiteit Utrecht**

# Goal

Get a feeling for:

- What metadata is
- What purpose it has
- The variety of forms metadata can take
- First steps toward a structured approach towards metadata

Question: Who deals with metadata?

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Agenda

- Goal
- Introduction
- What do we want with Research Data – FAIR principles
- Relation FAIR principles and metadata
- What is metadata, its manifestations and why we bother
- Handling metadata – how would you position yourself?

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Introduction

- Datamanager at the UU is a function 'in development'
- Two datamanagers with different function types
  - Frans:
    - » Represents researchers in storage facility project
    - » Advisor for data-related issues of researchers
    - » Hands on support for organising storage, privacy impact assessments, juridical advice, etc.
  - Ron:
    - » Organises output files of measurements
    - » Transforms and redistributes data for further analysis
    - » Designing a structured way for describing the data

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# YOUth background

- Start 2015
- Cohort study in Utrecht and surroundings
- Among 6000 subjects (+parents)
  - Kids (as of fetus)
  - Teens (as of 9yrs)

- Partnership between



- YOUth has a broad central theme:

**Investigating the link between brain development, environmental influences and behaviour**

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Data collection

- ... used to follow research questions
- YOUth is different: data is collected, research questions are to follow  
(But have to fit within the central theme!)

- YOUth datatypes

MRI

Questionnaires

IQ

Eye-tracking

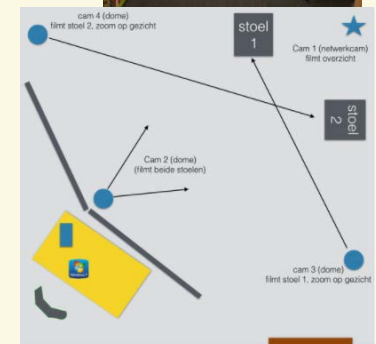
EEG

Computer Tasks

Bio

Parent Child  
Interaction

Echo



- So far 2,7TB collected, estimate 55TB in total

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# I-Lab for Yoda: what is it

Search by filename ▾ Search term...

Home / research-aldermen-and-notaries

## research-aldermen-and-notaries

Metadata Lock Actions ▾

Submit

Name	Modified date
<input type="checkbox"/> Aldermen and Notaries	2017-11-20 16:11:47
<input type="checkbox"/> yoda-metadata.xml	2017-11-02 16:29:12

Previous **1** Next



Metadata form - /research-aldermen-and-notaries

Close

Save

Submit



Delete all metadata

Descriptive

Title



Credit transactions by aldermen and notaries in the Low Countries

G  
S

## Yoda Open Data

You are viewing files in a web browser. For a better user experience we recommend that you open/map this location as a network drive using [https://i-lab.public.data.uu.nl/vault-aldermen-and-notaries/research-aldermen-and-notaries\[1511270521\]](https://i-lab.public.data.uu.nl/vault-aldermen-and-notaries/research-aldermen-and-notaries[1511270521]) as location.

### Index of /vault-aldermen-and-notaries/research-aldermen-and-notaries[1511270521]/ on nluu5p

Parent collection

Name	Size	Owner	Last modified
original/		rods	2017-11-21 14:22
License.txt	20K	rods	2017-11-21 14:22
yoda-metadata[1511270528].xml	10K	rods	2017-11-21 14:22

Location(s) covered

Leiden

Location(s) covered

Utrecht

Location(s) covered

Den Bosch

Period covered

Start Period

1500-01-01

End Period

1780-12-31

Tag

Early Modern History, Economic History, Republic

eeeting



# What do we want with research data?

## The FAIR principles

If you want the results of scientific research to have long term value, the needs to be stored as a comprehensible set which is:

- **Findable** – via a searchable characteristics, e.g. a unique identifier
  - **Accessible** – with descriptions of rights, contact persons
  - **Interoperable** – by info on used methodologies, taxonomies and datatypes
  - **Reusable** – by information on collection methods transformational algorithms, checking provenance, etc.
- FAIR data = well described data
  - Each principle implies a way to describe the research data

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# What is metadata?

## **Metadata = Data describing data**

- But: lots of contextual aspects to 'data' transcending the pure data-side of things. The data has been collected:
  - In a research context (funders, PI's, contact persons)
  - With a goal (hypothesis)
  - With a methodology
  - In a semantic context
  - In a context with juridical obligations
  - In a process, with a history, a provenance

## **Metadata = Data describing data and its (relevant) context**

- What is relevant depends on why you collect metadata

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# FAIR-revisited: What do we want to do with metadata?

**Metadata informs users and systems on research data so actors know:**

- What the data means
- How the data can be re-used
- Under what conditions the data can be reused
- How the data should be maintained and for how long

**To function as such, metadata must be:**

- Created: describing datasets
- Maintained: properties can change over time; provenance
- Published: metadata must be readable by men as well as machine
- Exchanged: pushing metadata to third party data-catalogues

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Digital metadata – Two basic types

- **Structured – machine readable**
  - » In a separate file
  - » Encapsuled in the data-object

Examples:

- » xml- or csv file in a preconceived and explicit structure and semantics
- » (online) vocabularies for semantic standardization

- **Unstructured**

Examples:

- » a codebook in a separate word-file describing variables
- » A separate readme.txt describing methods, instruments etc.

**N.B. for a comprehensive description you'll probably need both structured and unstructured metadata.**

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Metadata – a structured approach for establishing a position

## Structured metadata

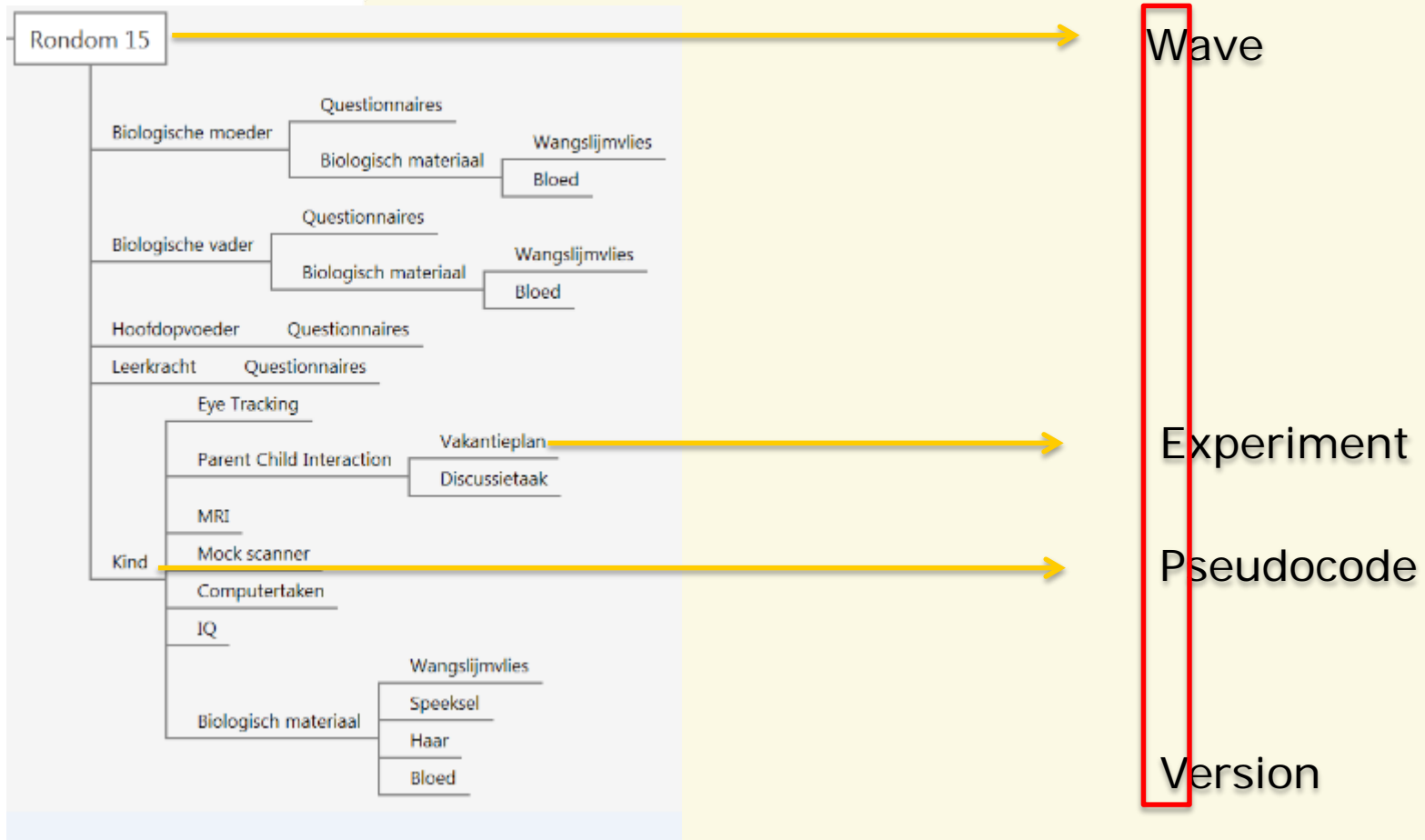
- What is the research discipline? -> standardized structure (DDI, DCC)
- What is the field(s) within the discipline? -> standardized structure and semantics: application profiles (NL LOM, ABCD) of standardized schemes
- FAIR: what structural metadata is needed to fulfil each of the four principles?

## Unstructured metadata

- What should an anonymous user know to be able to work with the data without contacting the owner?

For both: how much is a researcher willing to fill out?

# YOUth timeline & red line in data-coding



Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# YOUth data life cycle and metadata

Universiteit Utrecht Log out r.h.h.scholten@uu.nl

Creating data **Yoda Portal** Research Statistics Intake Group Manager

Metadata form - /research-pilot-r3-pci Close

**Save** ✓ ✓ ✓ ✓ ✓

**Project information**

**Project ID**

**Project Title**

**Project Description**

**Name Principal Investigator(s)**  +

**Persistent Identifier PI (optional)**  +

**Link Project Website**

**(Sub) Discipline of research**  +

**Research type**

**Project approval information**

**Ethical Approval**

**Approval By**



on\_A00327\_verRaw

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017







# Positioning yourself w.r.t. metadata

Thesis 1

**Structural, machine readable, metadata is always insufficient for a comprehensive description of the data. Hence each dataset should have a form of unstructured metadata.**

Step 1: Take your position from agree <-> not agree

Step 2: Plenary discussion on why people choose their position

Step 3: Reposition yourself on the basis of the discussion

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Positioning yourself w.r.t. metadata

Thesis 2

**Time spent on selecting standardized metadata schemes for internal use is wasted. Best approach is to create your own scheme and standardize on interfaces (i.e. at the moment of metadata exchange).**

Step 1: Take your position from agree <-> not agree

Step 2: Plenary discussion on why people choose their position

Step 3: Reposition yourself on the basis of the discussion

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017



# Positioning yourself w.r.t. metadata

Thesis 3

**A dataset will be most optimally findable in a discipline specific repository with a (rich and discipline specific) metadatscheme.**

Step 1: Take your position from agree <-> not agree

Step 2: Plenary discussion on why people choose their position

Step 3: Reposition yourself on the basis of the discussion

Data Stewardship Networking Meeting  
Utrecht – Dec 1<sup>st</sup> 2017

