

PsyCuraDat: Development of user-oriented curation criteria for psychological research data

Katarina Blask, Marie-Luise Müller, Marc Latz, Valentin Arnold, & Stephanie Kraffert
Leibniz Institute for Psychology

***‘Towards the sustainable use
of psychological research data’***

Structure

1. Project “PsyCuraDat”
 - Goals
 - Strategic Requirements
 - Work program and milestones
2. First empirical results
 - Evaluation of potential of existing standards for data re-use
 - Information processing aspects of data re-use
 - Formal specification of the standard
3. Discussion
 - Preliminary conclusion – summary
 - Outlook

1)The project „PsyCuraDat“ - Goals

Curation criteria

- Development of user-oriented curation criteria for psychological research data.

Re-usability

- Improving and facilitating re-usability of the research data.

Standard

- Establishing a coherent standard for sharing psychological research data.

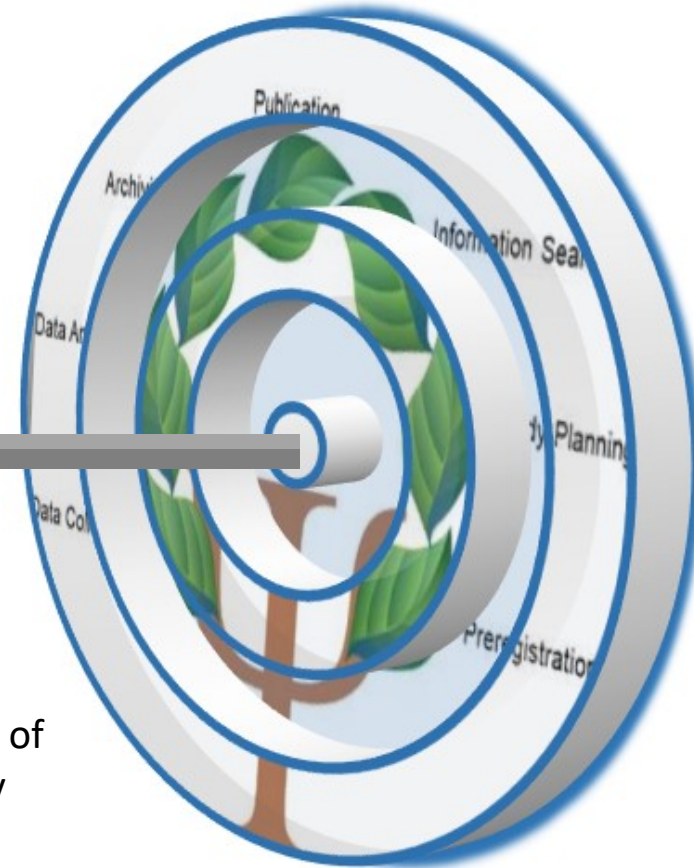
Data

- Increase sharing of research data by introducing the standard.

Funding: BMBF (Federal Ministry of Education and Research)

Duration: 3 years (2019-2022)

Project manager: Dr. Katarina Blask



Strategic Requirements

Including researchers, as data providers and data users, in the development of the standard

1. Simple implementation for data providers
2. Increased transparency in regards to re-using data by data users

Focus: Development of **user-oriented curation criteria**; **content** and **information architecture** are geared towards the **research process**

Benefits:

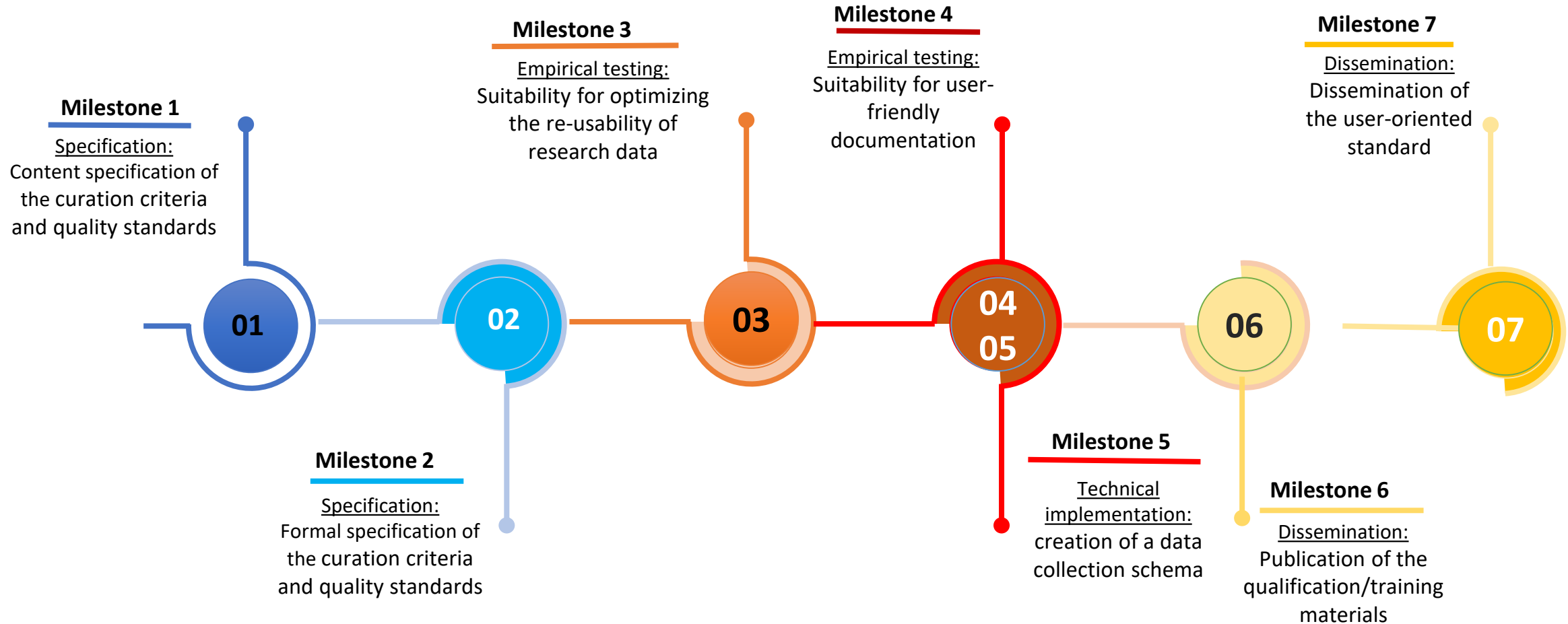
- Enhanced effectiveness and efficiency concerning the curation and re-use of data
- Increased satisfaction and quality within the field of psychology

Work program and milestones

Conceptualisation of user-oriented
criteria and standards

Testing of the user-oriented criteria
and standards

Establishment of the user-oriented
criteria and standards



2) First empirical results

Identification of existing curation criteria & quality standards

Milestone 1

Specification:

Content specification of the curation criteria and quality standards

01

ST
1.1

Identification of existing curation criteria & quality standards

Method

- Assessment of the suitability for the documentation of research processes

Selection criteria

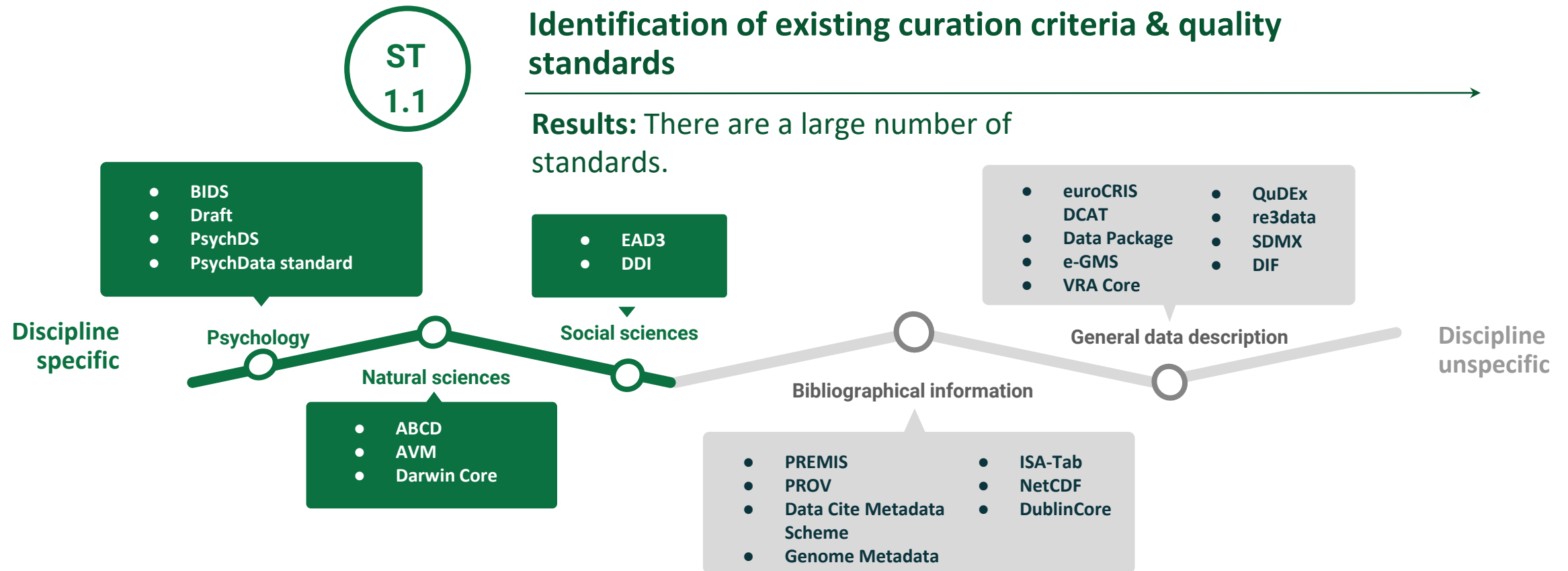
- Content relevance
- Discipline specificity
- Acceptance

Evaluation/Analysis

Is the standard suitable for psychology with regard to documentation of these central phases of the research process:

- Conceptualization
- Data collection
- Data analysis
- Publication
- Archiving

Identification of existing curation criteria & quality standards



Identification of existing curation criteria & quality standards



Identification of existing curation criteria & quality standards

Results: There is a large number of standards.

Existing standards **do not adequately match** the documentation requirements of the psychological research process.

→ Especially concerning **documentation** of data collection & -analysis process.

Evaluation of the re-use potential of existing metadata standards

Development and implementation of expert interviews for the evaluation of the re-use potentials of existing metadata standards

Milestone 1

Specification:

Content specification of the curation criteria and quality standards

01

ST
1.2

Method

- Expert interviews
- Online survey

Central question

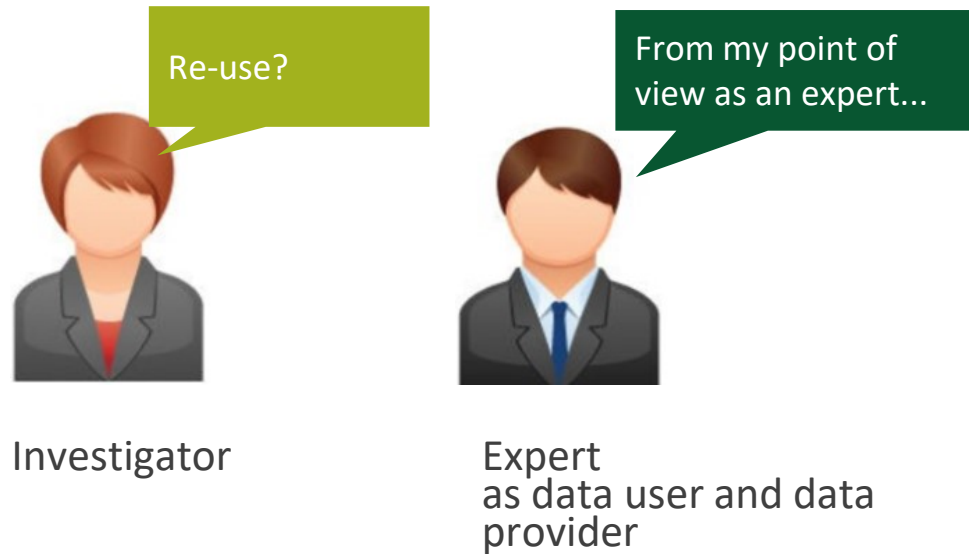
Which data and metadata are appropriate for which type of re-use?

Evaluation of the re-use potential of existing metadata standards

Expert interviews

Goal:

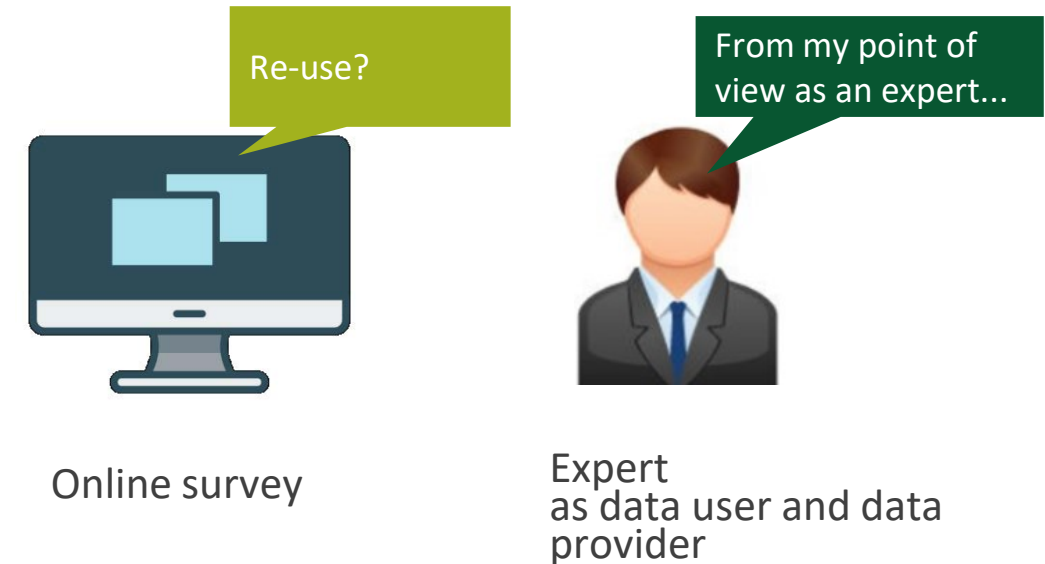
- Assessment and evaluation of the re-use potential of existing metadata standards from the experts' point of view



Online survey

Goal:

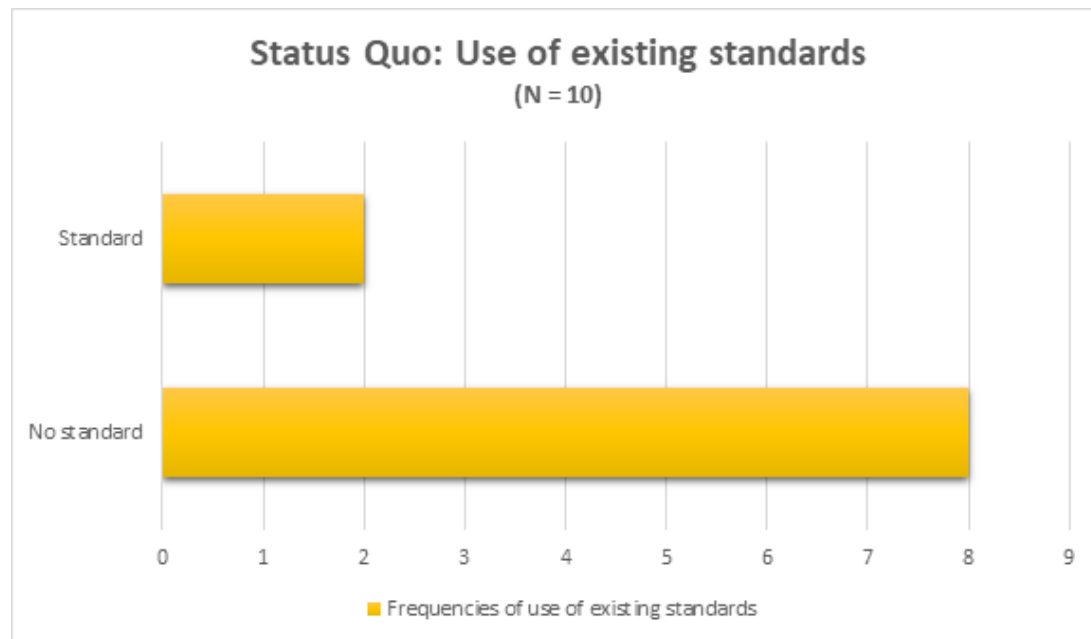
- Validation of the actual-target value deviation derived from the expert interviews
- Gaining further insights into the needs of researchers



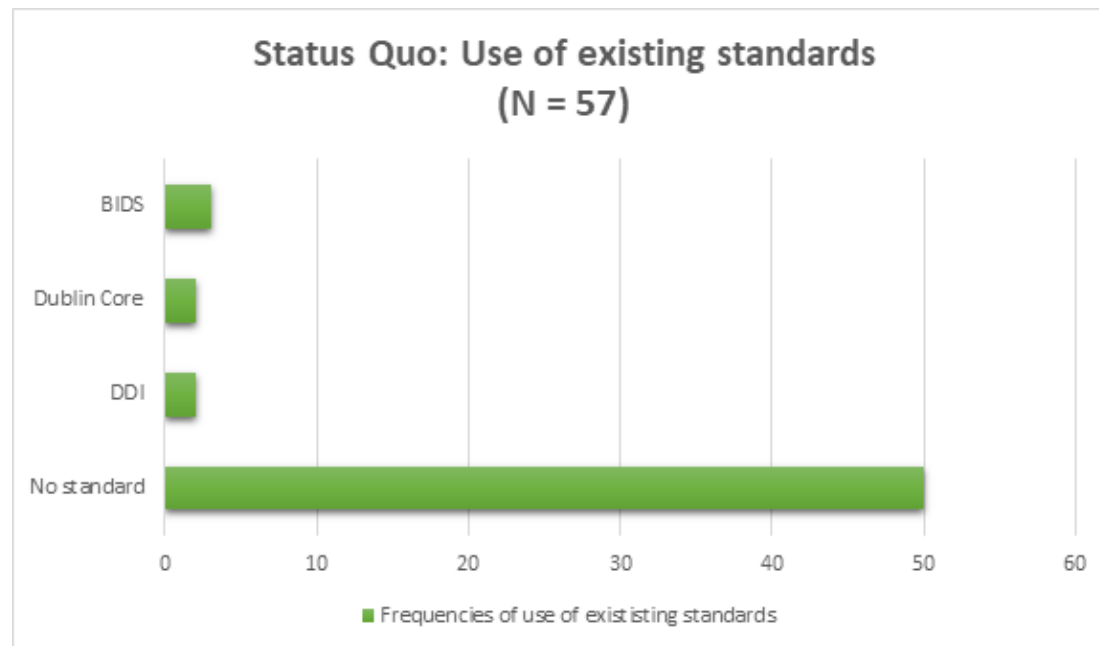
First results from ST 1.2

Use of existing standards

Expert interviews



Online survey

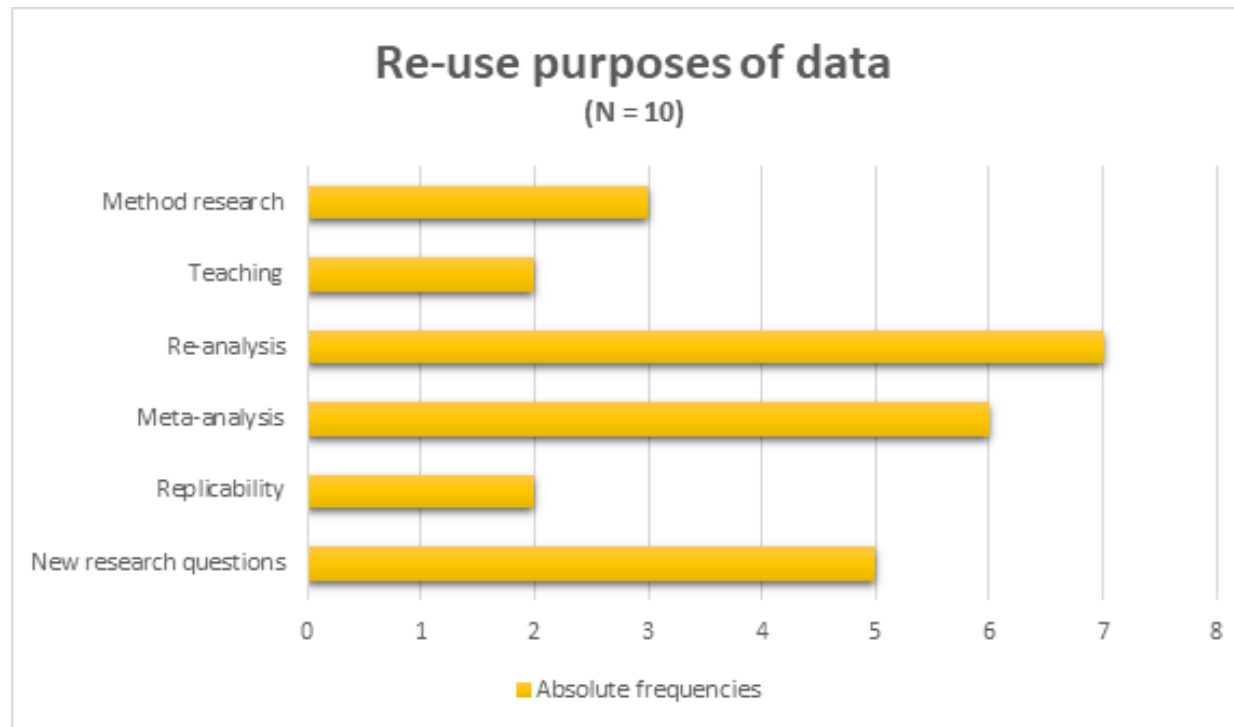


Existing standards for the documentation of psychological research data are rarely known and used.

First results from ST 1.2

Re-use purposes of data

Expert interviews

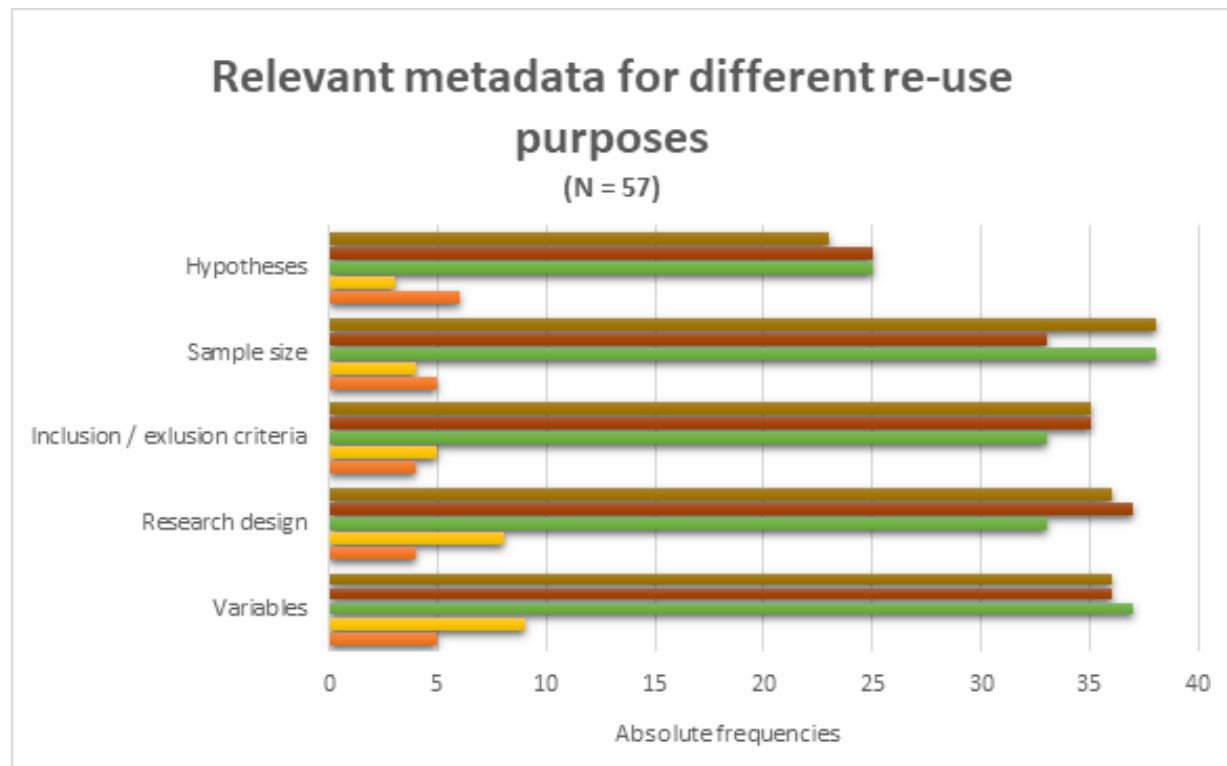


The main focus for researchers when re-using data is on re-analysis, meta-analysis and answering new research questions.

First results from ST 1.2

Relevant metadata for different re-use purposes

Online survey



Metadata are relevant to:

- Meta-analysis
- Systematic reviews
- Re-analysis
- (And less: illustrations & others)
- others

Evaluation of the re-use potential of existing metadata standards

Milestone 1

Specification:

Content specification of the curation criteria and quality standards

01

ST
1.3

Actual/target comparison and derivation of the content specification

Method

- Merging the results of ST 1.1 & 1.2

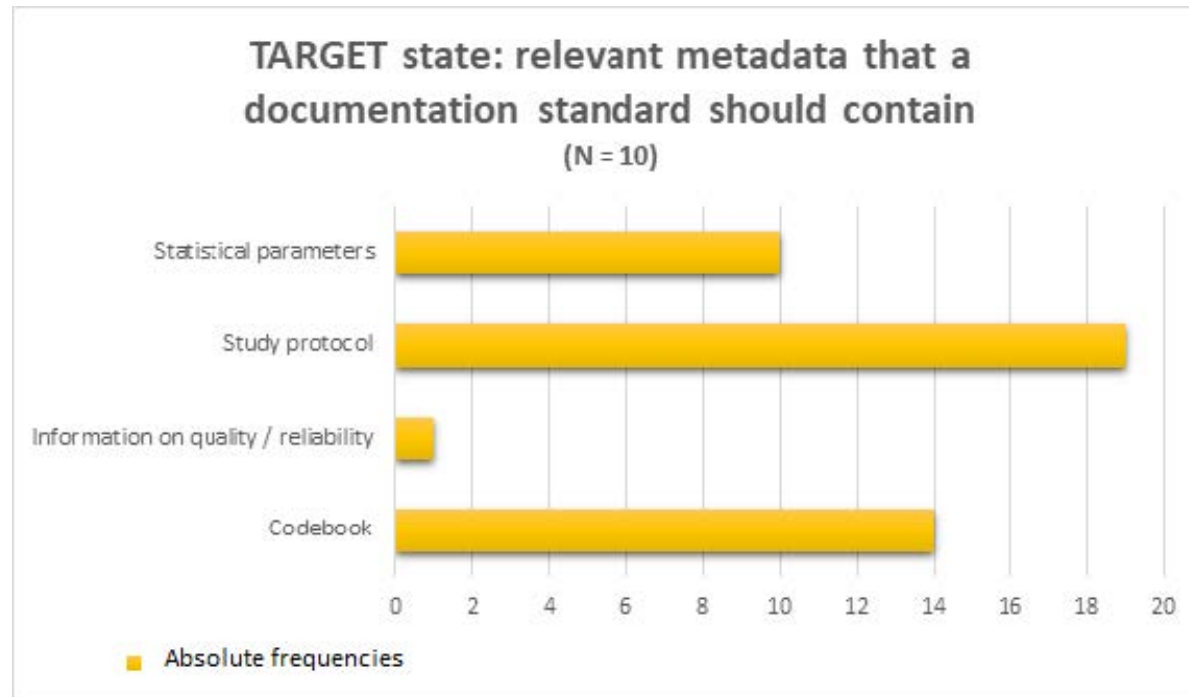
Central question

To what extent are theoretical requirements for a standard corresponding with researchers daily business?

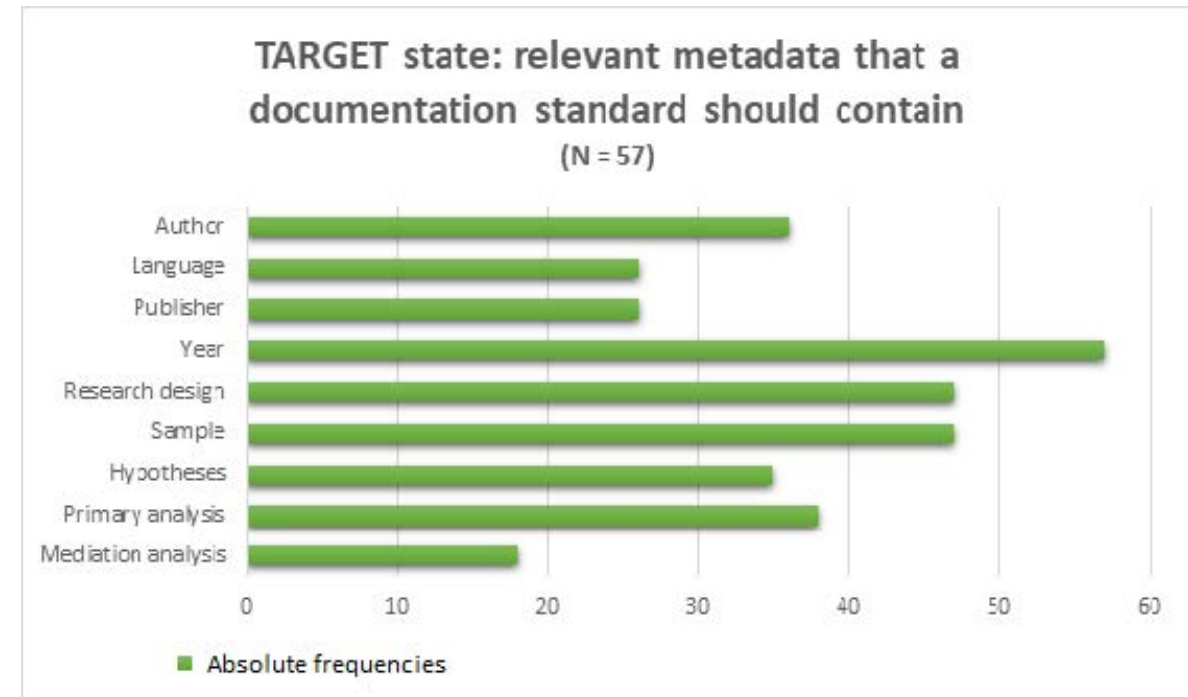
Results from ST 1.2

A prototypical documentation standard from the researchers' point of view

Expert interviews

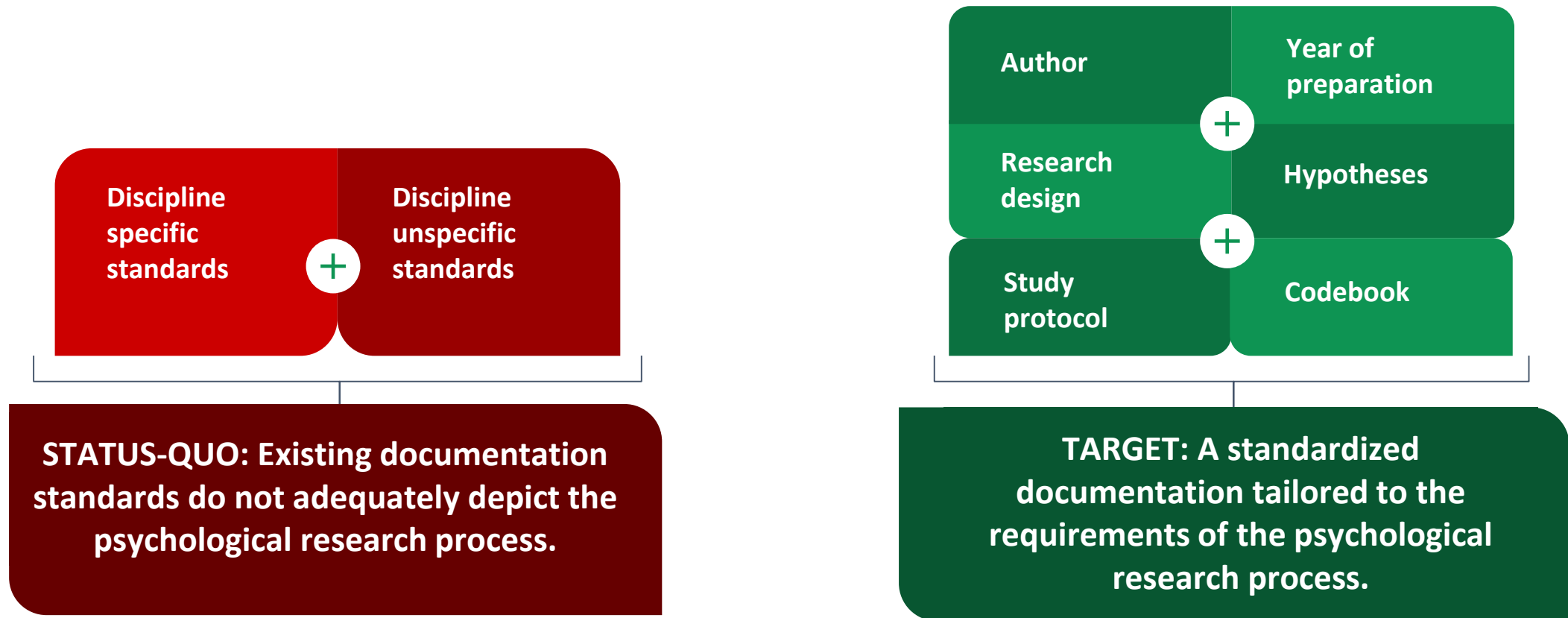


Online survey

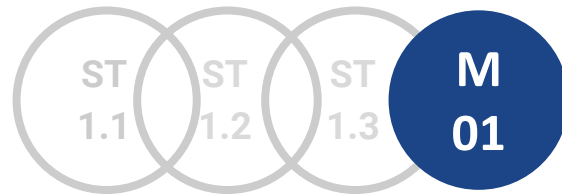


ST 1.3 Actual/target comparison

A prototypical documentation standard the researchers' point of view



Evaluation of the re-use potential of existing standards

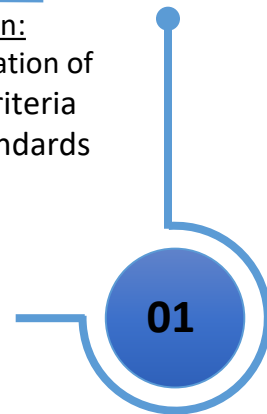


Contextual specification of the user-oriented curation criteria and quality standards

Milestone 1

Specification:

Content specification of the curation criteria and quality standards

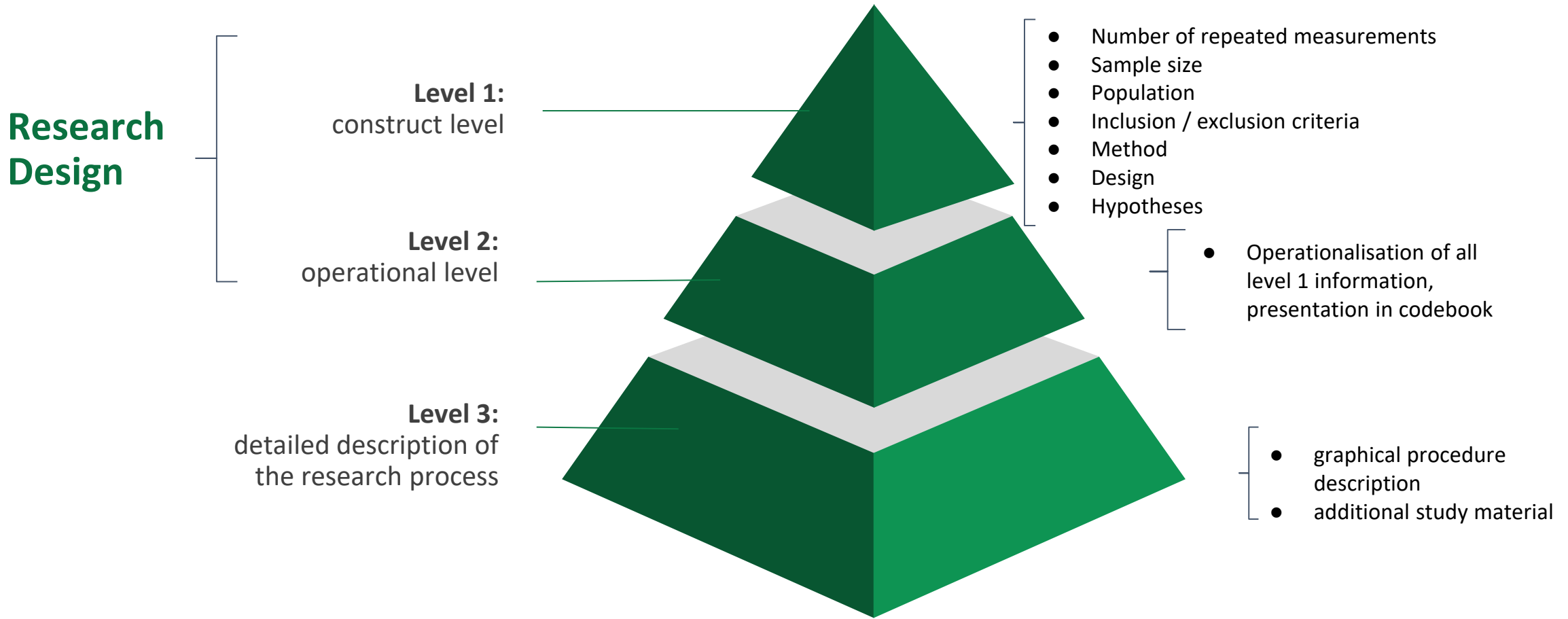


Important results

- Clear consensus on the documentation of psych. research data is missing; existing standards largely insufficient to document the psych. research process
-> Acceptance and re-use by psychologists impeded
- Relevant re-use purposes: re-analysis, meta-analysis, answering new research questions
- Standard should contain: research design, hypotheses, codebook, administrative metadata, study protocol

Milestone 01:

Content specification of the user-oriented curation criteria and quality standards



ST
1.4

Preparation of a user study to investigate information processing while re-using data

02

Milestone 2

Specification:

Formal specification of the curation criteria and quality standards

Method

- **Online user study incl. cognitive interviews**
- **Participants:** Researcher from the field of psychology ($N = 6$), randomised selection
- **Design: 2 (documentation: integrated vs. excluded) X 2 (order: correct vs. incorrect) mixed-factorial design**

Central question

- Which **metadata** are most important for optimal **re-usability** of psychological research data and in what **form** should it be presented?
- Which **information architecture** should a data set be based on?

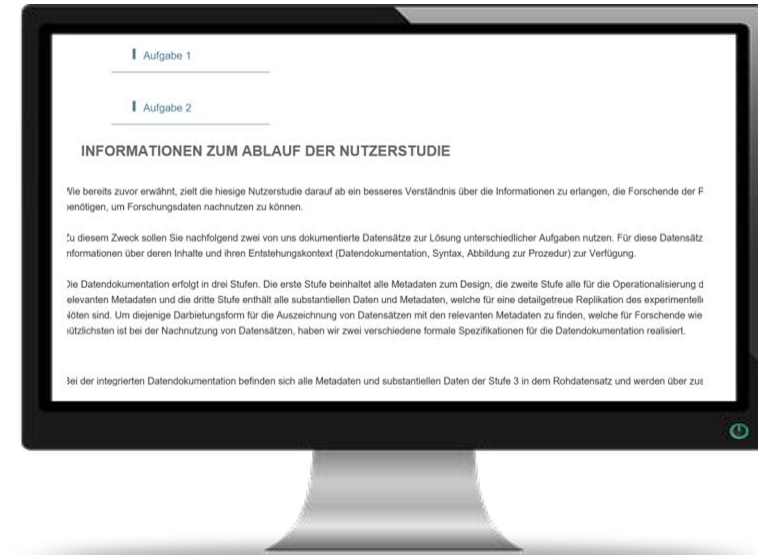
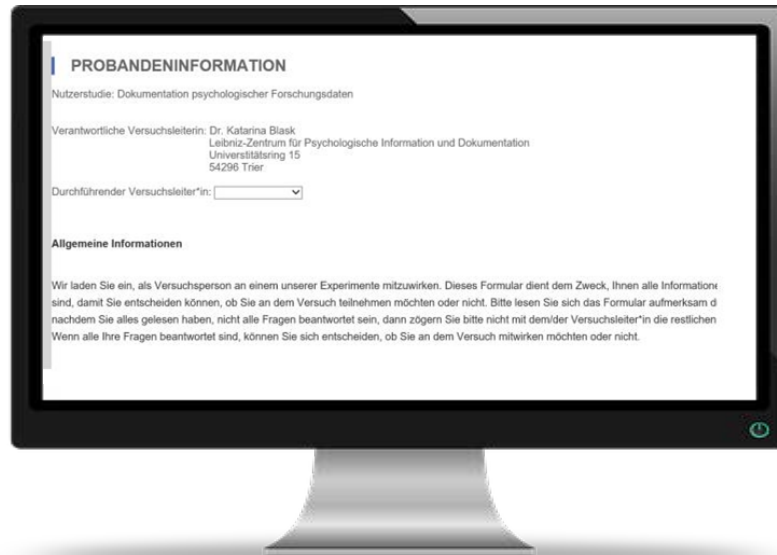
ST
1.5

User study - Implementation and analysis

02

Consent form

Instructions



Milestone 2

Specification:
Formal specification
of the curation
criteria and quality
standards

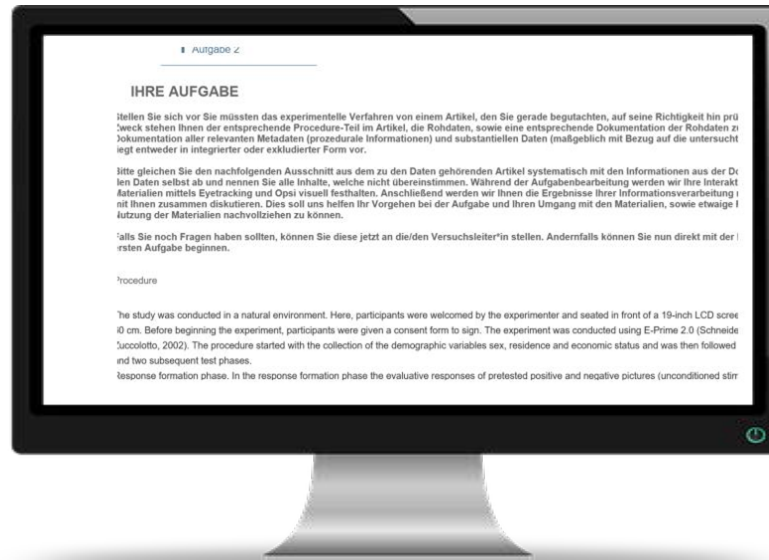
02

Milestone 2

Specification:
Formal specification
of the curation
criteria and quality
standards

Task

Cognitive interviews



Investigator

Have you
noticed...



Participant

I have
noticed...

02

Milestone 2

Specification:
Formal specification
of the curation
criteria and quality
standards

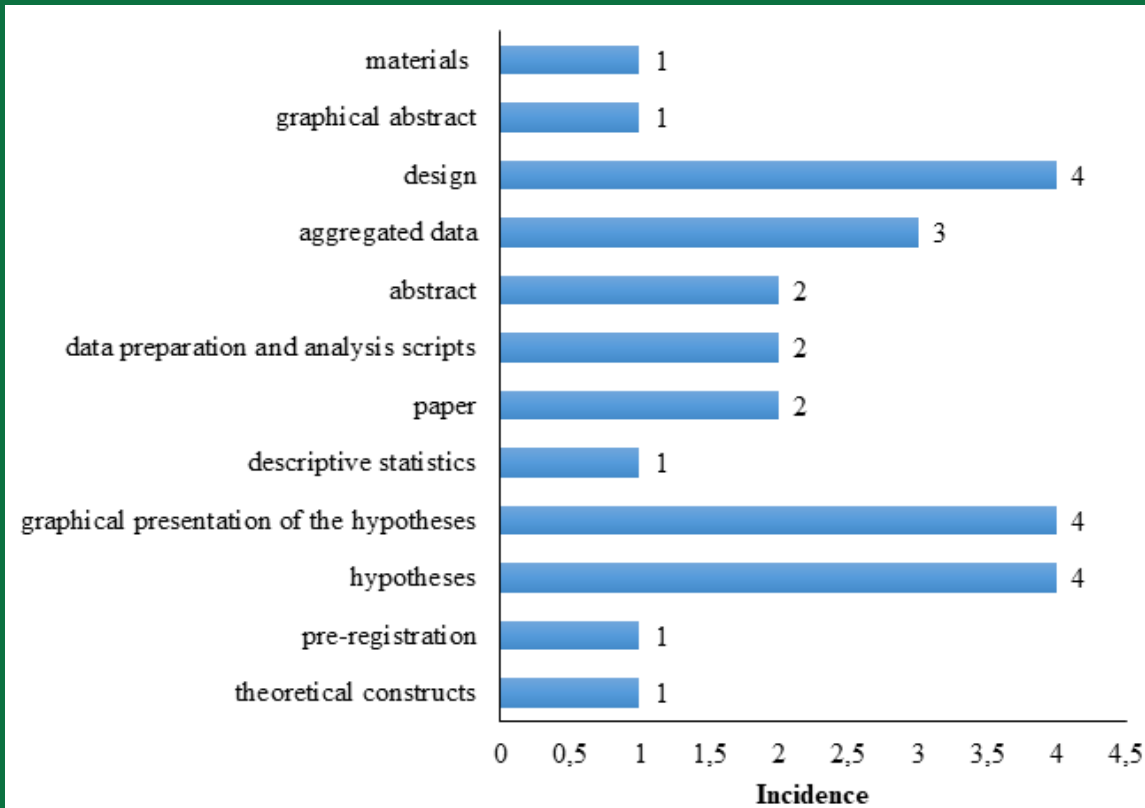
Results of the user study

- **Formal structure**
 - Majority saw potential for improvement regarding the **formal structure (5 out of 6)**
 - **5 out of 6 participants** found the **graphical presentation of the procedure** as well as the **codebook** helpful/useful
- **Order of variables in dataset**
 - Majority (**4 out of 6**) did not **notice any influence on the re-usability**
 - **Two participants** thought the **correct order** is influential

02

Milestone 2

Specification:
Formal specification
of the curation
criteria and quality
standards

Results: Requested information

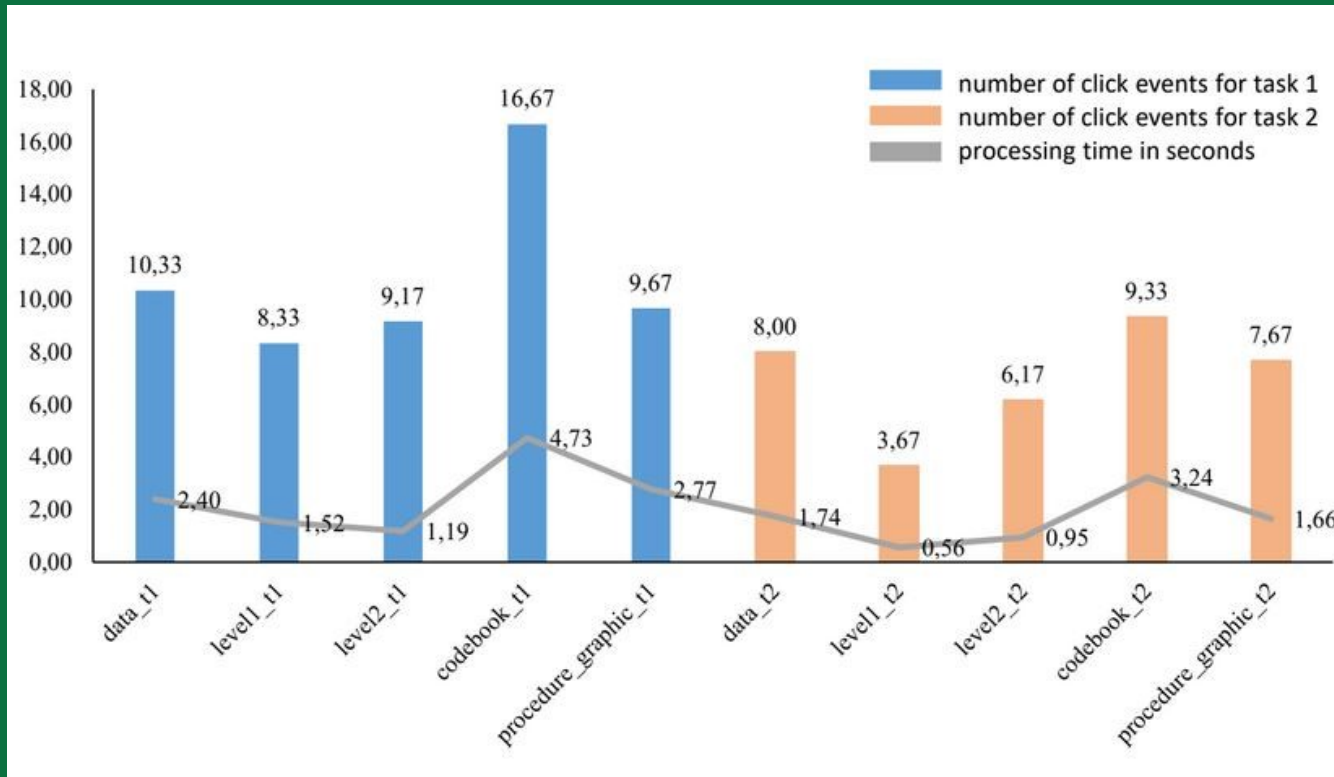
- Information about the **design, hypotheses and a graphical representation of the hypotheses** was mentioned the most (**4 out of 6**)
- 50% of participants (**3 out of 6**) requested a provision of **aggregated data**

02

Milestone 2

Specification:
Formal specification
of the curation
criteria and quality
standards

Results: Interaction with data & documentation



Click events while
completing the task:
Most frequent access to
**codebook > data >
procedure visualization**

02

Milestone 2

Specification:
Formal specification
of the curation
criteria and quality
standards

Discussion

- There is potential for improvement regarding the **formal** and **content structure**
 - Content description / differentiation of documentation levels was perceived as insufficient
- **Requested information:** participants preferred a similar structure to APAs **JARS** (Appelbaum et al., 2018)
- Clear description necessary on a **study (e.g., design, hypotheses, procedure)** and **data level (e.g., codebook)**

Evaluation of the re-use potential of existing standards

Preparation of a user study to refine the standard

ST
1.6

02

Milestone 2

Specification:

Formal specification
of the curation
criteria and quality
standards

Method

- **Online user study incl. cognitive interviews**
- **Participants:** Spokespersons of NOSI ($N = 7$), randomised selection
- **Adaptive design**

Central question

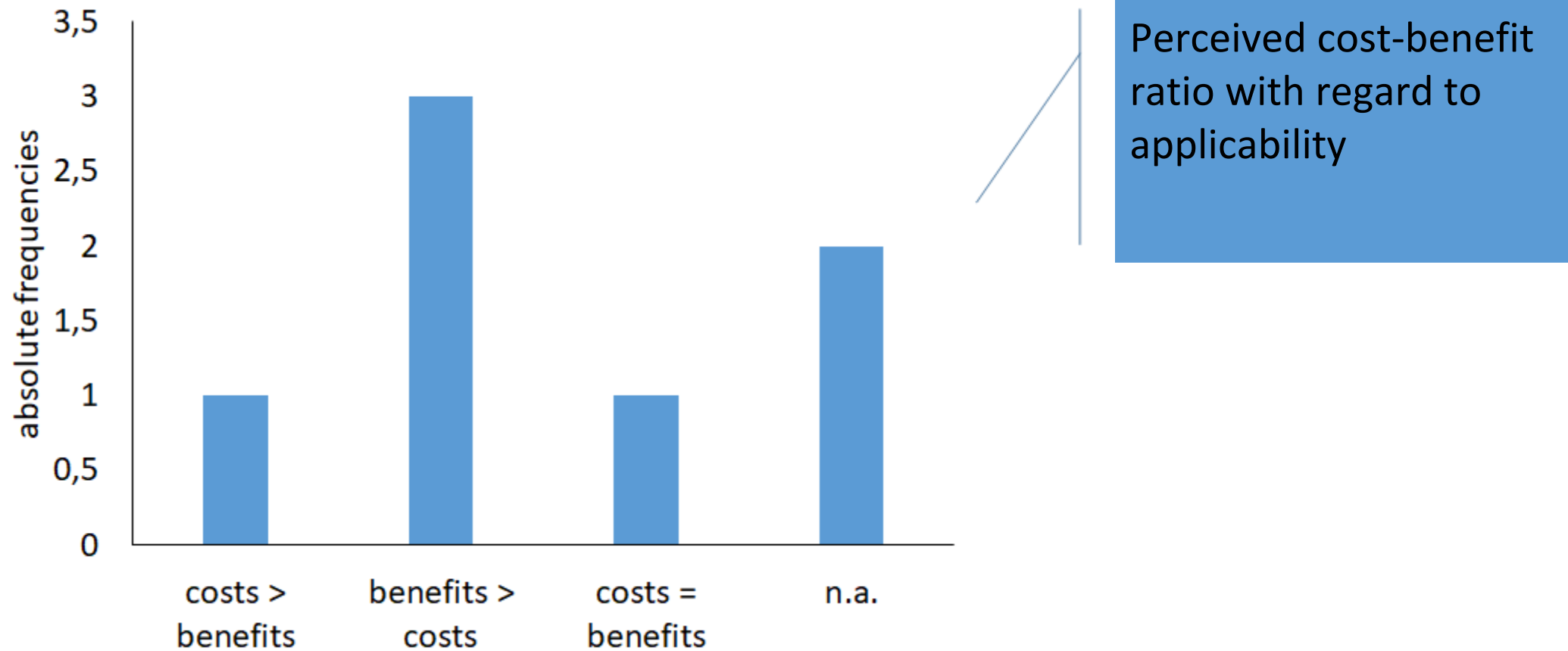
- Are the three levels of documentations experienced as useful/helpful?
- What else is missing for an optimal re-usability of data that has been processed according to this standard?

ST 1.7 Findings from the user study to refine our standard

- Overall positive evaluation regarding structure (i.e., three levels of documentation)
- Fit to everyday work: 100%



ST 1.7 Findings from the user study to refine the standard



ST 1.7 Findings from the user study to refine our standard

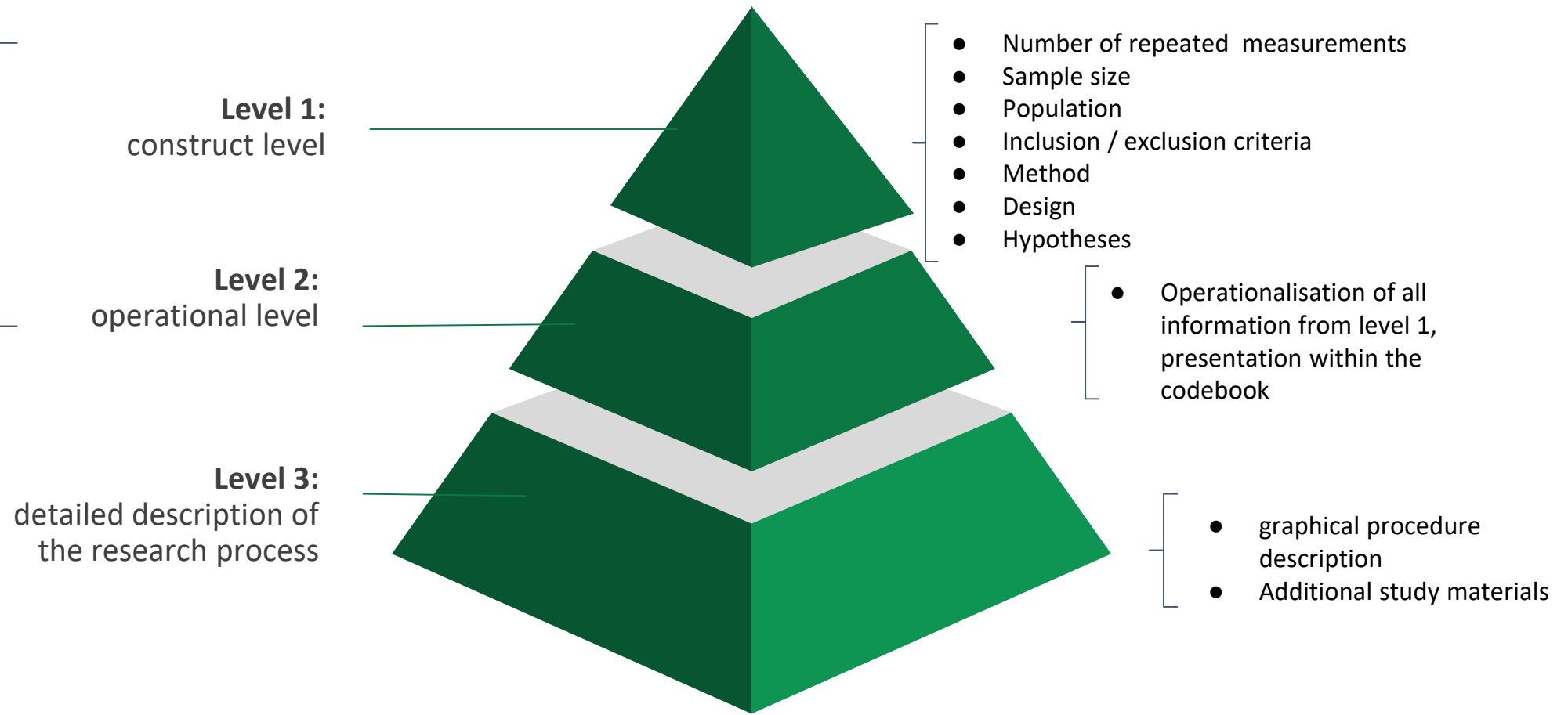
Potential for improvement / possible extension:

- User manual / readme
- Provision of study materials
- stronger integration of documentation levels (e.g., a hierarchical structure)

3) Discussion

What have we achieved...

Research Design



What is still missing...

1. Greater integration of levels (hierarchical structure)
2. More precise definition of formal specification
3. Exact guidelines for implementation

Outlook

Empirical review of the standards in regards to:

1. Improvement of the re-usability
2. User-friendly documentation

Thank you for your attention!

