Quality assurance and metrics for FAIR data in engineering

Special Interest Group within NFDI4Ing

Iryna Mozgova, Michael Selzer, Daniela Hausen, Gerald Jagusch,
Christian Langenbach, Mario Moser, Sabine Schönau, Barbara Steude
Max Leo Wawer, Jürgen Windeck

March 03, 2022

SIG Quality assurance and metrics for FAIR data

Development and exchange of various products, content and processes for quality assurance of research data and processes in the engineering sciences

- Founders: Base Services Measure S-1 and Community Cluster 5 – Standardisation

- Duration: 36 months
- Meeting frequency: Quarterly

- Kick-Off: 11.06.2021
- 2nd meeting: 08.10.2021, focus: quality-assured publication of research data
- 3rd meeting: 14.01.2022, focus on data management plans

- Next meeting: 08.04.2022, 9:00-11:00 am, focus: research data management plans and research processes.

Who we are: SIG-Team: TA S-1 und CC-5

Iryna Mozgova

Max Leo Wawer

IPeG
Leibniz University Hannover

Gerald Jagusch

Jürgen Windeck

University and State Library
Darmstadt

Michael Selzer

IAM-MMS
Karlsruhe Institute of Technology

Mario Moser

Patrick Mund

WZL
RWTH Aachen University

Daniela Hausen

Sabine Schönau

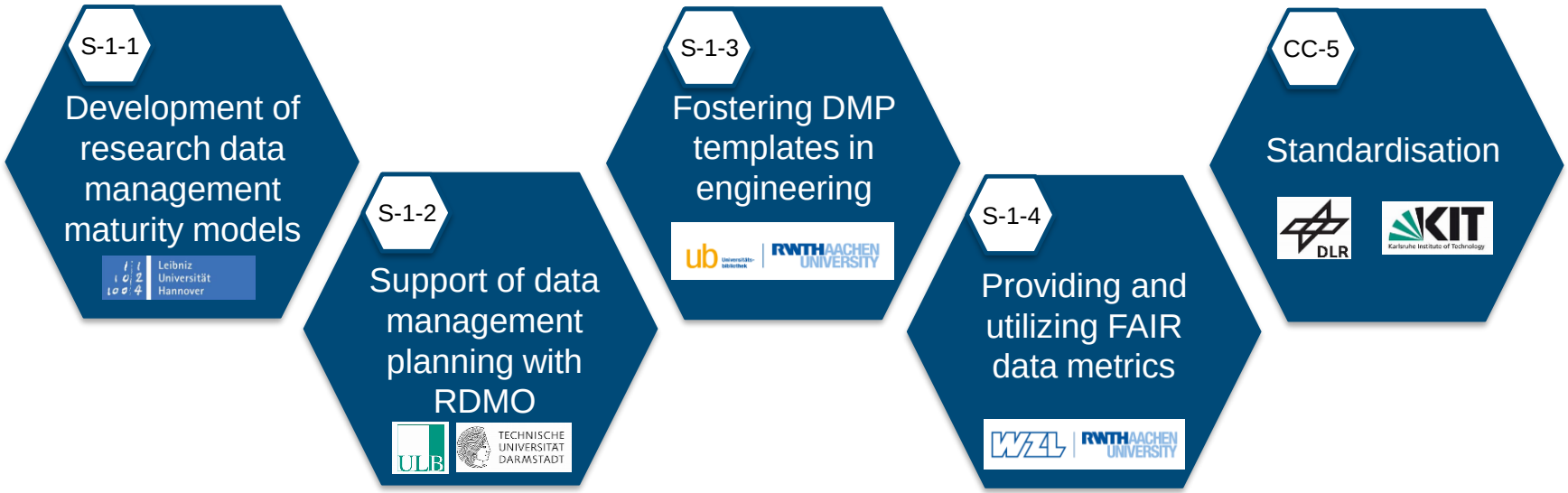
UB
RWTH Aachen University

Christian Langenbach

Barbara Steude

German Aerospace Center

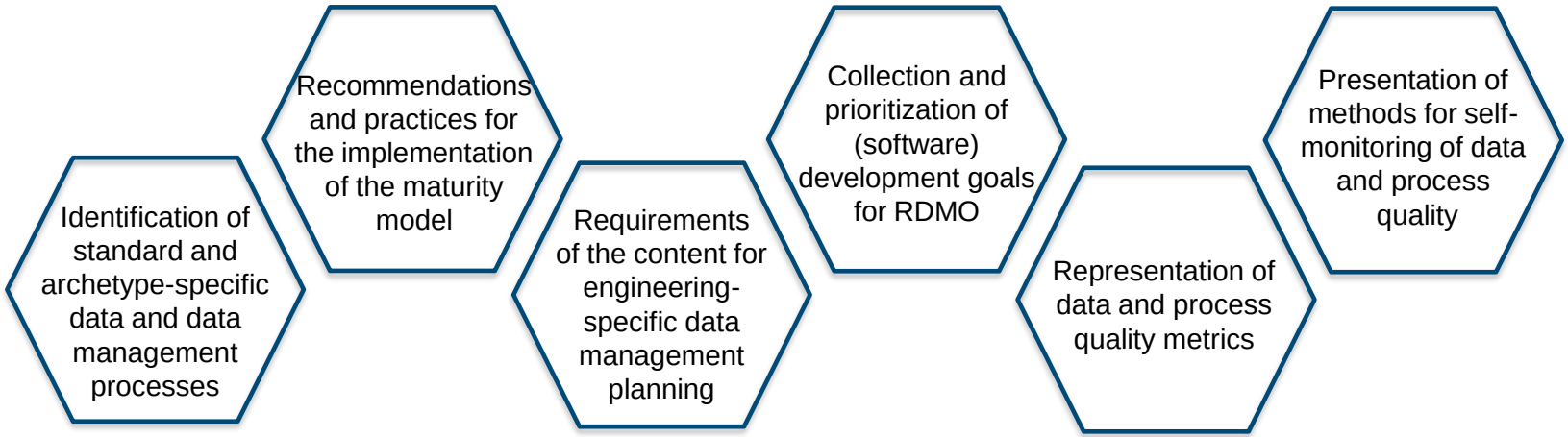
S-1: “Quality assurance in RDM processes and metrics for FAIR data” und CC-5: “Standardisation”



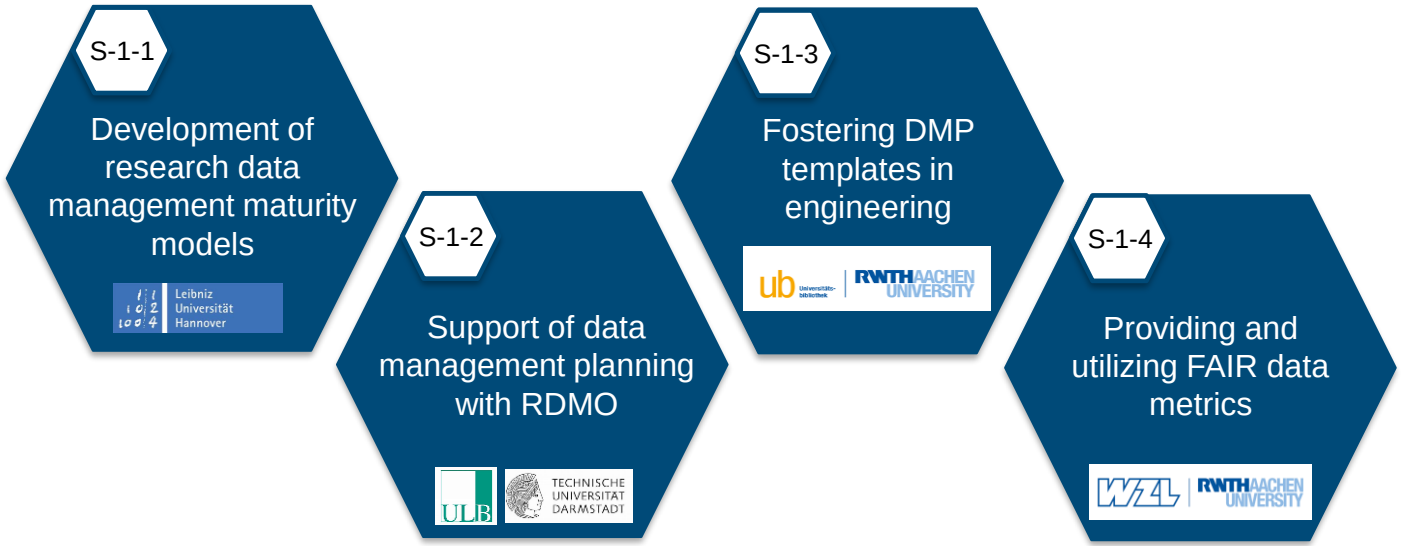
Quality assurance and metrics for FAIR data

Research-supported development and exchange of various products, content, and processes for quality assurance of engineering research data and processes

Discussion and validation of the tools and implementation of results

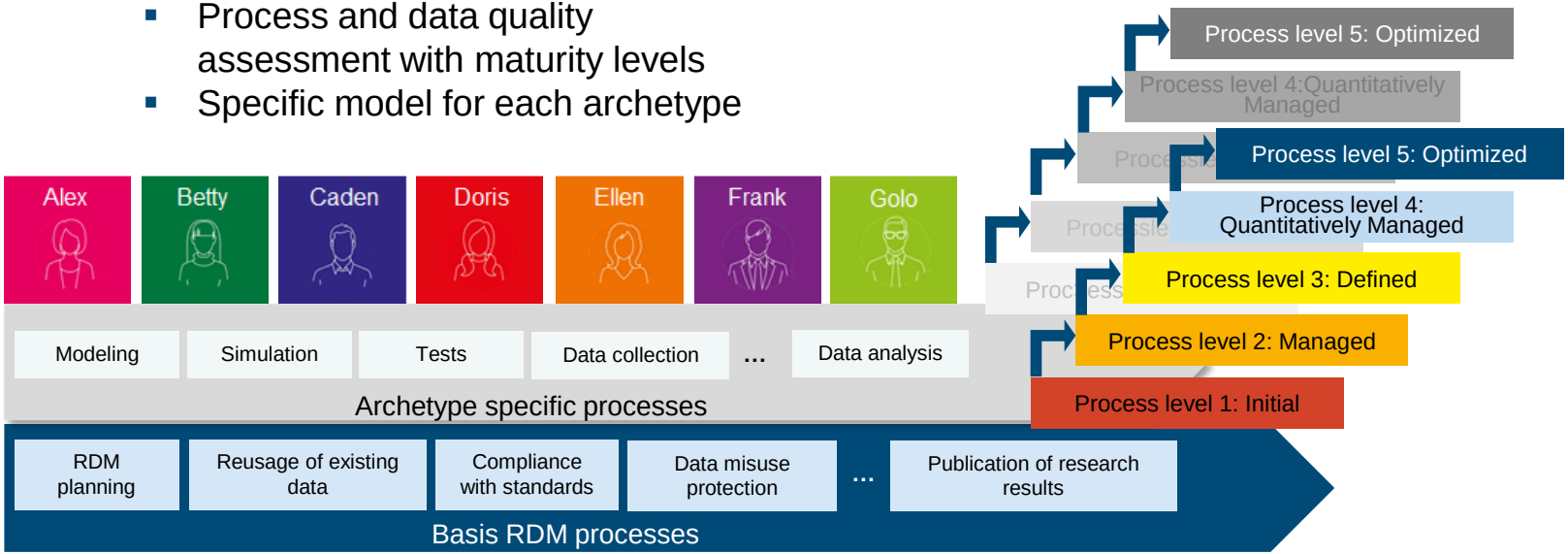


S-1: Quality assurance in RDM processes and metrics for FAIR data



S-1-1: Maturity model for research data management

- A framework for research data management processes and metrics for research data quality assurance
- Tool for self-monitoring and control
- Process and data quality assessment with maturity levels
- Specific model for each archetype



S-1-2: RDMO

Research Data Management Organiser

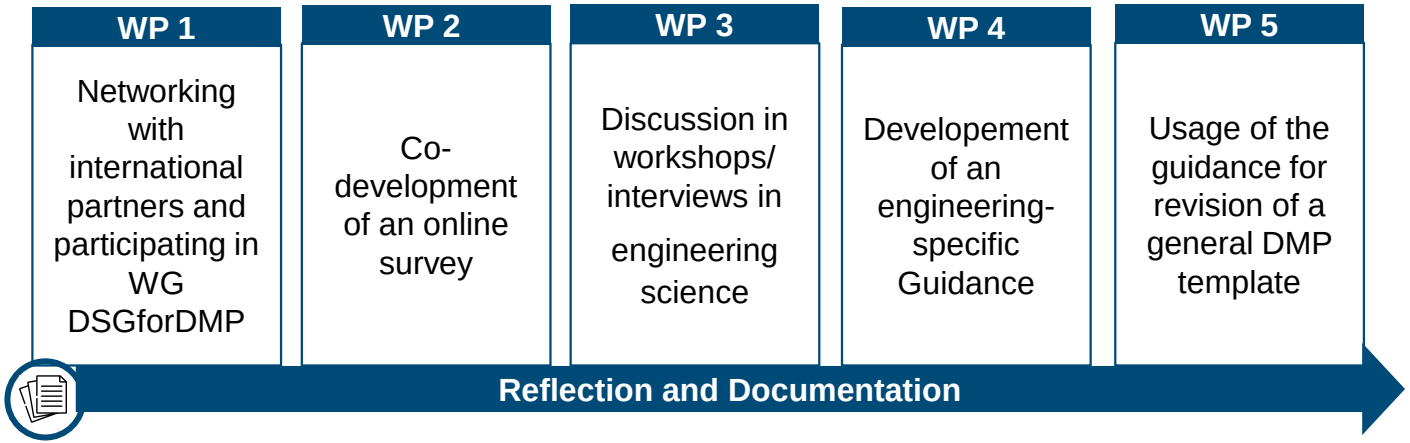
- Interactive web application for RDM planning
- DMP export for grant applications via adjustable templates
- 100% Open Source Software
  - DFG-funded project (2015 - 2020)
  - Established RDMO user community

- https://rdmorganiser.github.io/

S-1-3: DMP templates for the engineering science

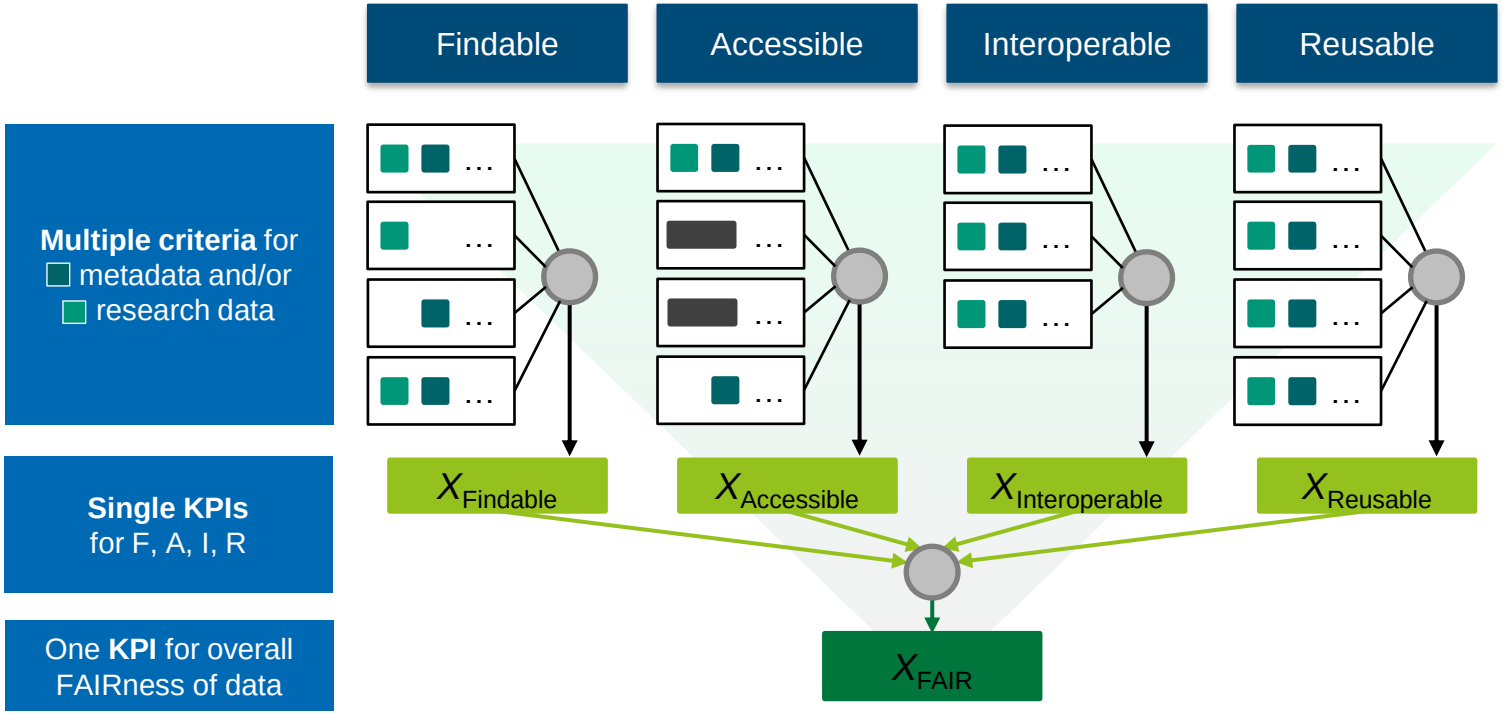
- Aims:
  - Development of engineering-specific guidance for DMP templates
  - Creation of engineering-specific templates

- Work packages (WP):

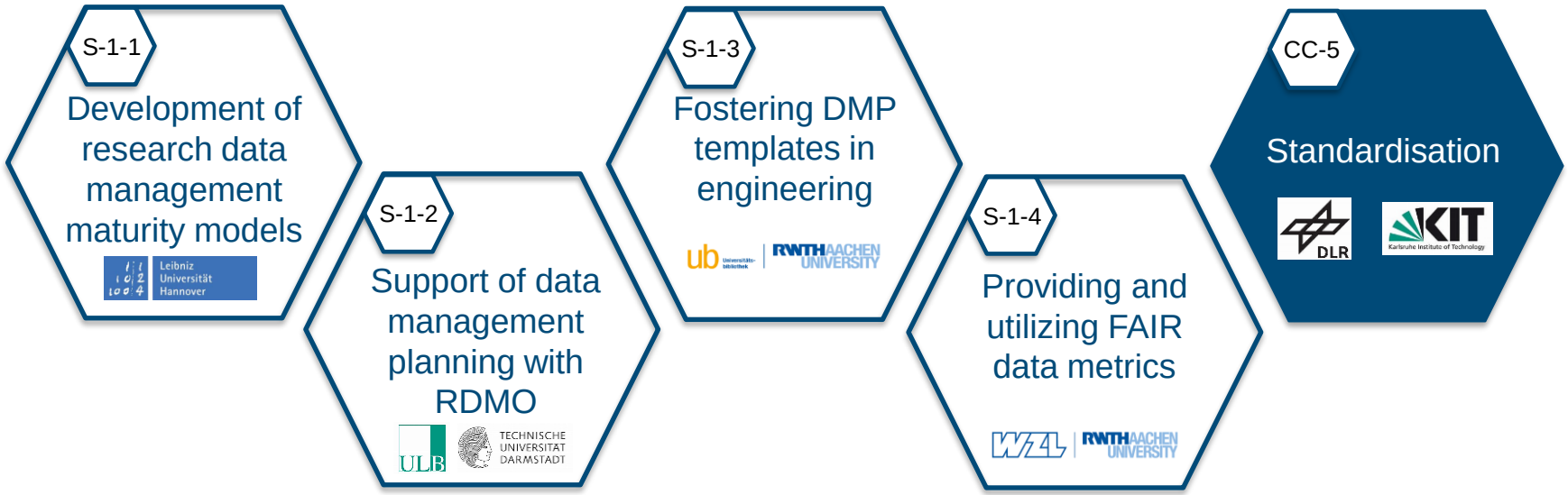


S-1-4: FAIR data metrics

Structure of the Key-Performance-Indicator system for FAIR research data



S-1: “Quality assurance in RDM processes and metrics for FAIR data” und CC-5: “Standardisation”



RFC concept of Community Cluster 5

Goals/Vision

What should be achieved

Establishment of (quasi-)standards for research data management

Based on the IETF (Internet Standards) concept, request for comments (RFC) for research data management will be built for NFDI(4Ing)

- Uniform platform for publishing (quasi-)standards (guidelines) for research data management
- Extracted from best-practice examples for broad reuse across NFDI initiatives and subject communities

Definition of RFC

Once an RFC is classified, it is published as an NFDI draft and discussed and tested by research and technical groups and individuals. It may eventually go through the following stages of development:

- Proposed standard: Stable, well understood, and generally considered useful
- Draft standard: Stable enough to use implementations of the standard in FDM applications
- NFDI standard: technically mature, widely implemented, and of great benefit to FDM in academia
- RFCs are numbered sequentially and are reviewed and published by a panel within the NFDI

Schematic structure of the RFC concept

Discussion on FDM issues in SIGs/AGs

Result is processed RFC-conform

Subject/discipline-specific panel reviews

RFC is published under a consecutive number

Communities have access to cross-disciplinary FDM solutions and implement them according to their individual needs

What has been implemented so far

Current status

- Two RFCs for creating RFCs are in the drafting phase
  - Structure of the publication path with the structure of the review process
  - Style guide for the RFCs

- Three initial possible community-driven RFCs on the topics of
  - Data Management Plans
  - Metadata (and ontology) on engineering-relevant data
  - Descriptive elements of an automated workflow

The results will be discussed with various partners from the NFDI and aligned with the members of the Task Area "Community Cluster"

Quality assurance and metrics for FAIR data in engineering

Special Interest Group within NFDI4Ing

SIG QA Kick-Off – intermediate result of group work

Typical reasearch activities

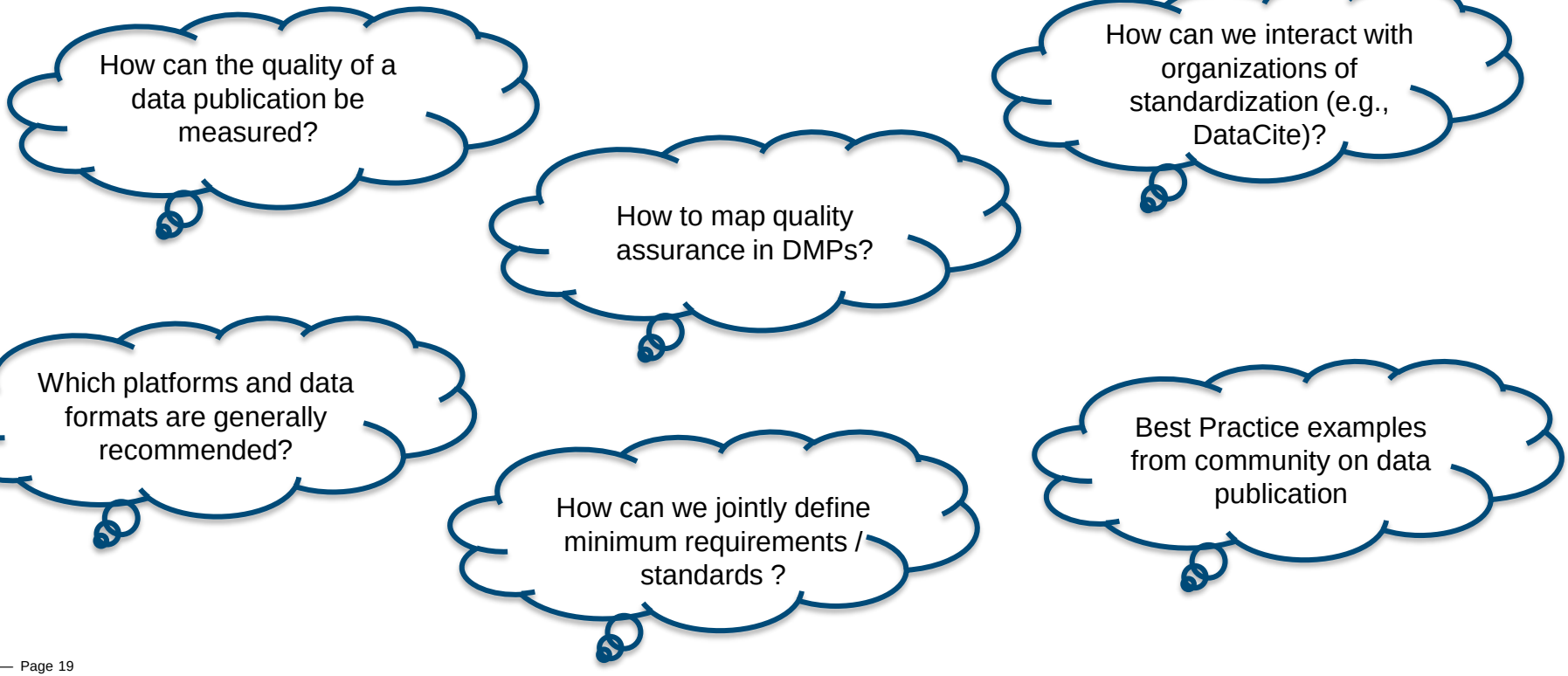
Documenting datatypes

Tools for data processing

Systems connected to a Data Management Planning Tool

Support of quality assurance measures

SIG QA – Questions for consideration



SIG QA 2. Meeting
Quality assurance and metrics for FAIR data

Discussed topics

- Vocabulary and Metadata schemas
- Ontologies for Provenance Tracking
- Data formats
- Licences
- Machine readable (meta) data
- Quality ensurance

SIG QA 3. Meeting
Data Management Plans in engineering science

- Short introduction to the content of DMP and the tool RDMO

https://www.openaire.eu/blogs/why-is-this-a-good-data-management-plan

RDMO

- Active participation in the online survey about engineering-specific aspects in a DMP
  - Results of the online survey presented at the next SIG meeting

- Active discussion in small groups
  - Which elements are important in a DMP/ DMP template?

SIG QA 3. Meeting -> Data Management Plans

Intermediate results of the group work

Understandablity clear, Relevance high

- Licences & terms of use
- Responsibilities & resource planning
- Versioning & Code
- Structuring of information, naming conventions & metadata
- Rights and role management
- Data & file format

Understandability unclear, Relevance high

- Data provenance & data sources
- Data processing
- Legal and ethical framework
- Overview of standards & standardisation
- Data sharing & data flow
- ELN

Save the Date

- Next meeting of SIG Quality assurance and metrics for FAIR data:
  - 08. April 2022, 8 a.m. UTC
- Homepage:
  - https://nfdi4ing.de/community-hub/special-interest-groups-sig/qa-metrics/
- Contact: contact@nfdi4ing.de

Thank you for your interest and cooperation!