

## *Name of Proposed Interest Group:*

# Research Data Management in Engineering

### *Introduction*

Engineering comprises a vast span of sub-disciplines including for example chemical, civil, electrical, and mechanical engineering. Traditionally, engineering disciplines apply scientific knowledge and empirical evidence to finding (innovative) solutions in areas that range from design, creating and building to production of software and innovative materials. Therefore, research in engineering disciplines is highly multi-disciplinary and also works in close collaboration with industry. Hence research data management practices within these sub-disciplines tend to be unaligned with e.g. open data initiatives/requirements and the implementation of the FAIR data principles. Publishing research data, code and workflows is rarely considered even when the research has been completed and the economic interests are secured. Engineers think and work for the future, but they usually do not think about embargos for their data. Sharing data and information often just takes place in a closed user group. The interpretation of “as open as possible” is in engineering more as closed as possible.

The aims of the Interest Group "Research Data Management in Engineering" (IG RDM4Eng) are changing the culture of handling data, creating awareness and bridging (sub-)communities and existing initiatives, which often exist only within a research group. Moreover, the IG will provide a platform for developing consensus on RDM best practices for engineering and to find convincing arguments for engineers in academia and industry to introduce or to improve RDM practices in their workflows.

### *User scenario(s) or use case(s) the IG wishes to address*

The IG RDM4Eng wants to address the engineer-specific challenges and aspects on RDM which are more often expected by research funding agencies and universities.

#### 1. Challenges

- Due to the **multi-disciplinary nature of the engineering research**, a close attention should be placed on the processes and workflows in order to define standards and best practices.
- Attention must be paid not only to the **research data per se**, but also to all **digital objects which help the generation and the (re-)use of the data**.
- **Breaking up data silos** is one of the major challenges.
- The **close collaboration with industry** can challenge researchers when confronting funder/journals/institutional requirements regarding RDM or simply when trying to follow the FAIR data principles.
- **No awareness and missing standards and skills for data, code and workflow publication as well as for FAIR and RDM workflow implementation** are controversial to the work of an engineer. Common to all engineers is that they work with standards like [ISO 8000](#) or [ISO 9000](#) every day.
- There is **no culture** or community efforts to **optimize or facilitate the sharing of data** with “external researchers” or other research communities (i.e. intellectual property rights are very relevant for researchers working in engineering).

- **Common interfaces and protocols** for managing, accessing and re-using research data from industry and academics are still **missing**. Each data provider has its own service offering and returns data in different (proprietary) formats with different licenses and costs. Additionally, commercial data providers are often constrained to particular business sectors in specific geographical areas and keep their data locked within isolated data sets.

## 2. Initial Use Cases to overcome the challenges

**Wind Energy goes FAIR.** This use case is supported by the researcher Nikola Vasiljevic from the Technical University of Denmark (DTU), who is also an RDA ambassador. It will develop a set of practical solutions for securing a long-term FAIRness and interpretability of (research) digital objects created in a typical lifecycle of wind energy projects. In particular, we will focus on improving the interoperability, reusability, and interpretability of digital objects generated in the [research project 'V52'](#) (internally funded by DTU Wind Energy), which produced extremely heterogeneous datasets (i.e. from meteorological data to [SCADA data](#)). The lack of a robust solution for preserving the knowledge that is generated in the interaction with the datasets is a common challenge in this kinds of research projects which produce large and complex datasets. Often there are no proper records of previously derived data sets and/or they are not systematically organized. This means that each time a researcher wants to re-use the data, the analysis starts from scratch. We will focus on finding and implementing the most suitable data formats, data structures/organization and naming conventions for recorded parameters in projects like 'V52'. Whenever possible, already existing conventions will be used. In case they do not exist (e.g. SCADA data), we will propose naming conventions, standardized data formats and structures, which will be discussed with the wind energy community. In this use case, relevant RDA recommendations will be revised and whenever possible adopted to try to maximize the re-usability of the data produced in the project 'V52'.

A second use case is the Collaborative Research Centre ([SFB](#) 1313 "Interface-driven multi-field processes in porous media," located at the University of Stuttgart and supported by the researcher Bernd Flemisch and research data manager Sibylle Hermann. **Porous media systems** are highly relevant in a wide range of applications, including ecological, technical and biological systems. Subsurface flow and transport processes, fuel cells, or vertebroplasty are just a few examples. Around 50 researchers from 18 different institutes of the University of Stuttgart and partners universities make up the team. A declared goal of the SFB is the integration of physical/mathematical models and the resulting computational models with high-resolution experiments.

The exchange, description and administration of data records is undertaken via the Data Repository of the University of Stuttgart ([DaRUS](#)) as the central link, via which data records can also be published directly. The automation of the data processes is regarded as essential. Automation initially involves the acquisition of all relevant input and output data that describe the respective experiment or simulation. In a further step, these data are automatically described with the metadata schema [EngMeta](#) and stored together on DaRUS. For the automated upload APIs of the underlying system [Dataverse](#) are used. The uploaded data sets are to be enriched with visualizations that allow a quick preview of the data. In order to ensure the reproducibility of simulation data, so-called software stacks are to be

stored in form of containers that contain the complete operating system as well as all packages and executables required for the execution of the simulation.

The emphasis in this use case is set on cooperation between infrastructure and researchers. The development of interfaces between institute and university infrastructure can be used as an example for the development of further interfaces. In addition, the interaction between experiment and simulation is a widespread phenomenon in the engineering sciences, as it is of utmost importance for the validation of computational models. So far, there are hardly any efforts in integrating systematically experimental and simulation data for model validation. Therefore, findings from the IG are to be incorporated into this use case, tested and evaluated and can be presented as best practices.

### Objectives

Possible objectives the IG RDM4Eng will work on:

- **Bring together scientific and industrial stakeholders** from all relevant sectors to discuss their legal and technological challenges around RDM practices.
- Provide a **forum for exchanging knowledge, options and experiences** on a national and international level.
- Bring light to the **issue of contract- and mission-oriented engineering research** from global and national points of view.
- Disseminate the IG **results within other relevant engineering organisations** on a global, European and national scale.
- Evaluate **RDA outputs and recommendations** and consider them with the **focus on engineering**.
- **Collaborate** with those **RDA groups** that have affinity to the objectives mentioned above, as well as with **external organisations** (such as [AAES](#), [CESAER](#), [NIST](#)), past and ongoing **engineering projects** ([Big Data Europe](#), [BOOST 4.0](#), [DURAARK](#)), and **industrial stakeholders** from different engineering disciplines like the automotive, construction, energy or architectural sector.

We also expect that an IG will provide a stronger leverage when it comes to engaging the industrial sector. Industry members are usually not present at university-based or scientific-community-based workshops, but RDA provides a framework which is nowadays widely recognized.

### Participation

This IG will be open to all RDA members from all countries and scientific disciplines. Particularly, but not exclusively, the IG will welcome members from the following backgrounds:

- Scientists involved in contracted research, to share their experience in dealing with RDM questions and non-disclosure agreements
- Industrial representatives from major and minor companies representing engineering science and the industry (particularly industry 4.0) sector
- Practitioners of software engineering for the industry sector
- Policy-makers for non-disclosure agreements & legal experts
- Data Stewards and related research data experts

- HPC and distributed computing experts

### Outcomes

Major/Preliminary outcomes of the IG RDM4Eng will include the following:

- Identify, collect, evaluate and compare industrial and institutional workflows, services, tools and best practices to establish an **exchange & information knowledge base** for engineering disciplines.
- Develop a **best data management practices guide for engineering researchers** based on the results drawn from the experiences of the use cases.
- Develop **models for dynamic consent** that protects industrial and institutional interest while enabling data sharing '*as open as possible, as closed as necessary*'.
- Define and evaluate existing and developing **engineering data platforms or repositories**.
- **Map RDM specific workflows, tools and topics** e.g., a first approach of a [mechanical engineer-specific DMP template](#) is given by RWTH Aachen University and Technical University of Darmstadt; another metadata scheme [EngMeta](#) for engineering science is developed by the University of Stuttgart.

### Mechanism

Outputs and recommendations will be produced based on consensus of the participating RDA group members. All topics will be openly discussed via the RDA communication platform providing a CMS, document store, and Wiki.

At the RDA plenaries the IG will organize group sessions and will interact with other RDA groups, e.g. by the organization of joint sessions. In between plenaries regular virtual conferences will guarantee the continuity of activities and encourage the continuous exchange of information. A summary of the virtual meetings will be provided for all initial and group members on the IG website. Initial and group members will be invited via mail to all face-to-face and virtual meetings and afterwards to the summary. As a first and ongoing step, the initial members are to be used as ambassadors to invite known engineers to participate in the IG RDM4Eng and to look for solutions in an international context.

**In order to draw the attention of engineers to RDA** and, in particular to the IG RDM4Eng, we will actively participate at engineer-specific conferences and seminars like the annual [VDI international conference](#) or the [WESC](#) by organizing RDM workshops, presentation, posters and/ or bringing information material. The first positive experiences were collected using this approach in the context of the [german national research data infrastructure](#). At the WGP (Wissenschaftliche Gesellschaft für Produktionstechnik) annual conference 2018, a poster and information material were used to draw attention to the initiative of [NFDI4Ing](#). This method is also useful to present the use cases and to show best practices. The goal is to reach engineer in their familiar environment which lowers the barrier to formulate their wishes and suggestions about RDM. In our view, it would be a fruitful combination for a use case if engineers from academia and industry as well as an infrastructure partner take care of the continuation of a use case. We see engineers are multipliers for industry partners.

Another mechanism to invite new members to join the IG is through workshops and meetings organized by the national RDA nodes, e.g. the next [Danish RDA node meeting](#).

We will invite the participating engineers and infrastructure partner to take part in the RDA-EU ambassador program. The initial co-chairs will accompany the group's creation and establish the activities. It is intended to conduct a co-chair election every two to three years.

The proposed IG has identified overlap with regard to contents with the following RDA groups:

- IG Active Data Management Plans
- IG Chemistry Research Data
- IG Preservation Tools, Techniques and Policies
- Disciplinary Collaboration Framework IG
- WG International Materials Resource Registries
- WG Metadata Standards Catalog
- Preserving Scientific Annotation WG
- IG RDA/CODATA Legal Interoperability
- IG RDA/CODATA Materials Data, Infrastructure & Interoperability
- IG RDA/NISO Privacy Implications of Research Data Sets
- WG Research Data Collections
- IG Software Source Code
- IG Vocabulary Services

While especially the IG RDA/CODATA Materials Data, Infrastructure & Interoperability as well as the IG RDA/NISO Privacy Implications of Research Data Sets have conceptual similarities with the IG RDM4Eng, to our knowledge, none of the above IG focus on an inclusion of both, industrial and scientific stakeholders from the engineering sector and bringing them together both on an European and on an international scale. Co-chairs and/or active participants from IGs and WGs that address a common topic to the IG RDM4Eng will be invited as guests for a lecture and discussion during a session, or a joint meeting will be organized. Whenever possible we will exchange knowledge and adopt it to the engineer-specific area.

### **Timeline**

We had a BoF session at the 12th RDA Plenary in Botswana and a first IG session at the 13th RDA Plenary in Philadelphia. We are looking forward to having the IG established at the 14th RDA Plenary in Helsinki in October 2019. The first outcomes of this IG are planned to be presented in a timely fashion using the RDA platform and file repository structure.

### **List of initial members**

<b>Name</b>	<b>Affiliation</b>	<b>Country</b>	<b>Chair</b>
Marta Teperek	TU Delft	Netherlands	
Susanna-Assunta Sansone	University of Oxford, Dep of Engineering Science (and RDA FAIRsharing WG)	United Kingdom	
Alastair Dunning	TU Delft	Netherlands	
Daniela Hausen	RWTH Aachen University	Germany	Co-Chair
Angelina Kraft	Technische Informationsbibliothek (TIB) German National	Germany	

	Library of Science and Technology		
Markus Stocker	Technische Informationsbibliothek (TIB) German National Library of Science and Technology	Germany	
Gerald Jagusch	ULB Darmstadt	Germany	
Nanette Rißler-Pipka	Karlsruhe Institute of Technology (KIT)	Germany	
David Wallom	University of Oxford	United Kingdom	
Kyong-Ha Lee	Korea Institute of Science and Technology Information	South Korea	
Jonathan Petters	Virginia Tech	USA	Co-chair
Nikola Vasiljević (RDA Ambassador for Engineering and Technology / Wind Energy)	Technical University of Denmark	Denmark	Co-chair
Paula Martinez Lavanchy	TU Delft	Netherlands	Co-Chair
Gretchen Greene	National Institute of Standards and Technology (NIST)	USA	
Joao Rocha da Silva	INESC TEC	Portugal	
Madeleine de Smaele	TU Delft	Netherlands	
Shaun de Witt	UK.Aea	United Kingdom	
Gretchen Sneff	Temple university	USA	
David Fearon	John Hopkins University	USA	
Phoebe Ayers	MIT	USA	
Alex Ossono	UC3M	Spain	
Laura Palumbo	Rutgers University	USA	
Megan Force	Clarivate Analytics	USA	
Leighton Christiansen	US.DOT	USA	
Gerhard Goldbeck	Goldbeck Consulting	United Kingdom	
Ralph Müller-Pfefferkorn	TU Dresden	Germany	
Berta Chang	North Carolina State University	USA	
Alexis Lewis	US National Foundation	USA	
Rainer Stotzka	KIT	Germany	
Thomas Jejkal	KIT	Germany	
Swerker Holmgren	Uppsala University	Sweden	
Andrew White	Rensselaer Polytech Institute	USA	

Sibylle Hermann	University of Stuttgart	Germany	
Bernd Flemisch	University of Stuttgart	Germany	