

Name of Proposed Interest Group:

Research Data Management in Engineering

Introduction (A brief articulation of what issues the IG will address, how this IG is aligned with the RDA mission, and how this IG would be a value-added contribution to the RDA community):

Research in Engineering comprises a vast span of sub-disciplines including for example chemical, civil, electrical, and mechanical engineering. Traditionally engineering disciplines belong to the applied sciences, which cooperate closely with industry and look to create commercial advantage from the research. Therefore open data and data sharing are rarely considered even when the research has been completed and the economic interests are secured.

This proposal for an Interest Group on “Research Data Management in Engineering” (IG RDM4Eng) will gather and compare industrial and institutional workflows, services and tools regarding ‘engineering research data’. An application for an initial BoF session will be submitted for the next RDA plenary in Gaborone, Botswana.

The Interest Group presents an opportunity to highlight emerging [FAIR \(Findable, Accessible, Interoperable, Re-usable\)](#) data approaches from scientific and industrial engineering disciplines and explores how data tools can be used ‘as a service’ to break up existing community specific ‘data silos’. The proposal’s background is based on projects and initiatives in the area of engineering sciences, and especially relates to the challenges posed by contract or mission oriented research which is performed together with industrial stakeholders.

User scenario(s) or use case(s) the IG wishes to address (what triggered the desire for this IG in the first place):

The engineering community is highly fragmented in terms of its RDM organization. Data and the descriptive documentation and software basis are crucial components of sustainable engineering research. Within the engineering science community at universities, initiatives such as the [American Association of Engineering Societies \(AAES\)](#) and [CESAER \(Conference of European Schools for Advanced Engineering Education and Research\)](#) with the [Task Force Open Science \(TFOS\)](#) have shown that the approach to RDM is often bottom-up. This is in contrast to other disciplines where RDM is strongly driven by professional associations (e.g. [DARIAH](#) in the humanities and [ELIXIR](#) in life sciences).

A major challenge is the heterogeneity of data generated by research groups, even if they investigate the same phenomenon. The data obtained in the primary data analysis are usually located on local data storage or on a backup storage by the doctoral students and scientists. Many scientists pursue their own strategy for file naming and documentation of their research data. Due to the different systems of the scientists, there are few systematic records within the discipline. Standardized metadata records are not yet widespread which

results in data that are difficult to retrieve and access. Data publication is far from being the norm. The FAIR principles are far from being properly implemented.

Within industry, on the other hand, particularly [Industry 4.0](#), data management comprises a strategic business approach that applies a consistent set of business processes in support of the collaborative creation, management, dissemination, and use of a product and/or a service. IT Systems supporting these industrial data management processes are known under different names:

- Engineering data management (EDM system)
- Product data management (PDM system)
- Product life cycle management (PLM system)
- Collaborative product development

Thanks to these management systems, the industrial workflows of product and service engineering are usually well documented. What is missing, however, are common interfaces and protocols for managing, accessing and re-using research data from industry and academics. Each data provider has its own service offer 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. The combination of these factors hinders interoperability and further uptake in FAIR data platforms (such as the proposed [European Open Science Cloud](#)) and a better data value chain around corporate information.

During the last year, several workshops and interviews in the CESAER context with participating engineers in the Netherlands, Ireland and Germany were performed, representing major engineering communities such as computational engineering, mechanical engineering, construction and thermodynamics.

Major outcomes are needs for

- an introduction to selected, existing methods and tools regarding RDM, which are adaptable to the needs of the specific sectors (also in combination with educational resources, e.g. persons who speak the 'language' and can demonstrate & adapt existing RDM tools)
- a harmonized software coding platform for engineers (having been told now several times that source code actually presents 'the most important' research data within many engineering communities)
- a basic data documentation guideline (or maybe even a standard), especially concerning contract and mission oriented research with industrial partners as well as journal guidelines
- better, coordinated access to HPC facilities.

We see the introduction of an IG in RDA as an opportunity to seek solutions in a broader international context, activating engineering scientists from all over the world. 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.

Objectives (A specific set of focus areas for discussion, including use cases that pointed to the need for the IG in the first place. Articulate how this group is different from other current activities inside or outside of RDA.):

The proposed “Research Data Management in Engineering” Interest Group (IG RDM4Eng) seeks to bring together scientific and industrial stakeholders from all relevant sectors. The IG RDM4Eng will provide its scientific and industrial members with the opportunity to discuss and improve the legal and technological challenges to the adoption of FAIR data and software management in Engineering, to exchange knowledge, opinions and experiences, and form or participate in existing Working Groups to address these challenges.

This includes in particular contract research and its associated privacy and security concerns (such as the conditions included in non-disclosure agreements), and the important role of software and source code as ‘research data types’ within the engineering sector:

- 1) **Use case: coding base & project data:** Analyzing the use of software source code repositories such as [GitHub](#) and [GitLab](#) by the engineering sector and identifying common needs and best practices, such as a centralized GitLab/GitHub framework with similar best practices and standards for computational engineering and software management. In addition to software management, this use case will also identify workflows and services to facilitate, standardize and harmonize the transfer of engineering project results (e.g. [data.DURAARK.eu](#), an architectural engineering example, or [4TU.ResearchData](#)) into a broader FAIR knowledge base. Currently, outcomes of such projects, data results and accompanying repositories, are usually listed on discovery platforms such as [re3data.org](#). As a principle, these should be however also distributed in the emerging FAIR assessment platforms and tools such as [FAIRsharing.org](#).
- 2) **Use case: privacy and security in engineering data management:** Given the lack of legal national and international harmonisation of scientific and industrial data protection and access, different approaches, contracts (e.g. [non-disclosure agreements](#)) and protocols will have to be accessed and compared in order to improve the FAIRness of engineering research data. Scientists may have to be pointed to possibilities to get access to legal advice as well as appropriate contract templates for the planning and executing of FAIR best practices in research data management in the case of mission oriented research.

Based on the use cases as outlined above, the preliminary focus of the IG RDM4Eng will cover following areas:

- **Engineering Data and Code landscape**
 - defining a list of tools dealing with Engineering Data Management in academia and industry
 - defining and evaluating existing and developing engineering data platforms
 - disseminating the IG results within other relevant engineering organisations on a global, European and national scale

- **Privacy and Security in Engineering Data**

- sharing best practice on non-disclosure agreements with industrial stakeholders and differential privacy
- developing models for dynamic consent that protects industrial interest while enabling data sharing '*as open as possible, as closed as necessary*'
- providing a forum for discussing, explaining and responding to data regulation issues on a national and international level

As a result, the IG RDM4Eng will build a knowledge base in order to share technical practices, identify common data and service requirements, and facilitate search and analysis of existing FAIR data solutions for interoperability challenges that are shared among engineering research infrastructures, universities and companies. The IG will seek collaboration 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 sectors:

- Automotive sector
- Construction sector
- Computational engineering sector
- Mechanics
- Architectural engineering
- Chemical engineering
- and others

Participation (Address which communities will be involved, what skills or knowledge should they have, and how will you engage these communities. Also address how this group proposes to coordinate its activity with relevant related groups.):

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 contract 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 (Discuss what the IG intends to accomplish. Include examples of WG topics or supporting IG-level outputs that might lead to WGs later on.):

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

- ❖ Strengthen the connection between the industrial and academic sector

- ❖ Bring light to the issue of contract and mission oriented engineering research from global and national points of view
- ❖ Establish an exchange & information knowledge base for engineering data types and software products
- ❖ Display funder guidelines and best practices
- ❖ It is planned to solve particular problems like those identified in the CESAER context by spawning of RDA Working Groups

Mechanism (Describe how often your group will meet and how will you maintain momentum between Plenaries.):

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.

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 years.

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

- IG Chemistry Research Data
- IG Data Fabric
- IG Data Foundations and Terminology
- WG Data Type Registries
- WG Data Versioning
- IG Disciplinary Collaboration Framework IG
- IG Domain Repositories
- IG From Observational Data to Information
- IG Health Data
- WG Blockchain Applications in Health
- WG International Materials Resource Registries
- WG Metadata Standards Catalog
- IG RDA/CODATA Legal Interoperability
- IG RDA/CODATA Materials Data, Infrastructure & Interoperability
- IG RDA/NISO Privacy Implications of Research Data Sets
- IG Reproducibility
- 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 this 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.

Timeline (Describe draft milestones and goals for the first 12 months):

We are looking forward to having the IG established before the 12th RDA Plenary Meeting in November 2018 in Gaborone, Botswana, so that we may have a chance to gather more experts attend a possible BoF event by this time and have the first official meeting.

At P12, we will also contact fellow WG/IG as identified above to discuss common challenges and experiences concerning our main research and infrastructure related topics, as well as the involvement of the industry sector.

The first outcomes of this IG are planned to be presented in a timely fashion using the RDA platform and file repository structure, with a formal presentation and discussion at the latest after 12 months after the establishment of the IG.

List of initial members

Name	Affiliation	Country
All: Please feel free to enter your name or make suggestions for potential interested members; the IG will be open for everyone to participate		
Alastair Dunning	TU Delft	Netherlands
Rainer Stotzka	Karlsruhe Institute of Technology	Germany
Daniela Hausen	RWTH Aachen University	Germany
Angelina Kraft	Technische Informationsbibliothek (TIB) German National Library of Science and Technology	Germany

Markus Stocker	Technische Informationsbibliothek (TIB) German National Library of Science and Technology	Germany
----------------	---	---------