

# Toward a RDM in Engineering Interest Group

Angelina Kraft<sup>(1)</sup>, Daniela Hausen<sup>(2)</sup>, Markus Stocker<sup>(1)</sup>

(1) TIB Leibniz Information Centre for Science and Technology

(2) RWTH Aachen University

tinyurl.com/**yaphjz6s**

# Origins

- German National Research Data Infrastructure initiative
- Consortium also in engineering, among other disciplines

# Origins

- German National Research Data Infrastructure initiative
- Consortium also in engineering, among other disciplines
- Many problems are not national and must be addressed globally
- For instance, standardization of terminology (vocabulary)
- Hence, the consortium started working on an RDA IG

# Why RDA

- RDA builds the social and technical bridges to enable open data sharing
- Strong focus on research data management
- Large, established, global organization
- Pragmatic and agile, with low entry barrier for new interest/working groups

# RDA disciplinary coverage

- Agriculture
- Linguistics
- Biomedical Sciences
- Chemistry
- Digital Humanities

# RDA disciplinary coverage

- Agriculture
- Linguistics
- Biomedical Sciences
- Chemistry
- Digital Humanities
- ... and soon Engineering?

# Challenges (also) in engineering

- High data heterogeneity
- Numerous proprietary data formats
- Lack of systematic acquisition/curation of data/software in infrastructures
- The key engineering “data” is often software
- Little conventions, e.g. in file naming and data/software documentation
- Close cooperation with industry
- Lack of interfaces between industry and academia
- Data locked up in silos with strong access constraints
- Little interoperability



# Proposed aims for the IG

- Mapping and understand the landscape
  - Engineering comprises many sub-disciplines, e.g. chemical, civil, electrical, mechanical
  - Heterogeneous practices, data, workflows, tools, services

# Proposed aims for the IG

- Mapping and understand the landscape
  - Engineering comprises many sub-disciplines, e.g. chemical, civil, electrical, mechanical
  - Heterogeneous practices, data, workflows, tools, services
- Assess the current application of
  - FAIR Data Principles in engineering
  - Open Data and Data Sharing in this discipline

# Proposed aims for the IG

- Mapping and understand the landscape
  - Engineering comprises many sub-disciplines, e.g. chemical, civil, electrical, mechanical
  - Heterogeneous practices, data, workflows, tools, services
- Assess the current application of
  - FAIR Data Principles in engineering
  - Open Data and Data Sharing in this discipline
- Identify and document engineering specific RDM concerns due to, e.g.
  - Close cooperation with industry
  - Strong focus on creating commercial advantage
  - Contract or mission oriented research

# Proposed aims for the IG

- Mapping and understand the landscape
  - Engineering comprises many sub-disciplines, e.g. chemical, civil, electrical, mechanical
  - Heterogeneous practices, data, workflows, tools, services
- Assess the current application of
  - FAIR Data Principles in engineering
  - Open Data and Data Sharing in this discipline
- Identify and document engineering specific RDM concerns due to, e.g.
  - Close cooperation with industry
  - Strong focus on creating commercial advantage
  - Contract or mission oriented research
- Collect and compare practices between academic and industrial research
  - Rather different practices in academia and industry
  - Industry data management typically follows strict protocols (e.g. Industry 4.0)

# Proposed aims for the IG

- Build on the collected information base and shared understanding
- Harmonize and standardize practices, workflows, (meta)data, software
- Create and publish best practices
- Catalyze the implementation of FAIR Data Principles
- Build awareness, e.g. for data/software publication/citation
- Analyse the relationships to other relevant disciplines, e.g. chemistry

# Current status

- First BoF meeting at RDA
- Draft Charter exists and can be shared with interested parties
- Review and revision of draft Charter
- Expand initial membership
- In particular also co-chairs

# Objectives today

- Assess the interest for the prospective IG in the wider RDA community
- Identify co-chairs for the prospective IG
- Discuss needs of engineering communities for RDM
- Obtain feedback and input on next steps
- Network among researchers, industrial members and related RDA groups

# RDM in Engineering IG: Draft Charter

Angelina Kraft<sup>(1)</sup>, Daniela Hausen<sup>(2)</sup>, Markus Stocker<sup>(1)</sup>

(1) TIB Leibniz Information Centre for Science and Technology

(2) RWTH Aachen University



[tinyurl.com/yb4ec8vb](https://tinyurl.com/yb4ec8vb)

# RDA IG: The Charter

- Introduction
- Use/r cases/scenarios the IG wishes to address
- Objectives
- Participation
- Outcomes
- Mechanism
- Timeline
- Initial members

# Use/r cases/scenarios

- What triggered the desire for this IG?
- High fragmentation in terms of RDM organization
- Unaddressed international challenges
- Gaps between industrial and academic research
- Need to activate the engineering community
- Seek solutions in broader international context
- Leverage in engaging industry

# Objectives

- Connect scientific and industrial stakeholders from all relevant sectors
- Provide members with the opportunity to
  - Discuss and improve the legal and technological challenges to FAIR data adoption
  - As well as software management in engineering
  - To exchange knowledge and experiences
  - To form or participate in groups that address these challenges
- Address discipline specific concerns, e.g.
  - Contract research and its associated privacy and security concerns
  - The role of software and source code as 'research data types'
  - FAIR (meta)data in engineering disciplines

# Participation

- Open to all RDA members from all countries and scientific disciplines
- In particular
  - Scientists involved in contract research
  - Industrial representatives in engineering
  - Industrial software engineering experts
  - Policy-makers for non-disclosure agreements and legal experts
  - Data stewards and related research data experts
  - HPC and distributed computing experts

# Outcomes

- Knowledge base and exchange platform for
  - Engineering metadata and data types
  - Metadata and data standards
  - Existing vocabularies
  - Software management
  - Others TBD
- Shared understanding for differences and commonalities of contract and mission oriented engineering research from global and national points of view
- Stronger connection between industry and academia
- Others TBD

# Mechanism

- Open discussion using the RDA platform, document store, and wiki
- Regular (monthly) conference calls
- Outputs based on consensus of the participating group members
- Group and joint sessions at RDA Plenaries
- Spawn RDA WGs that address scope-limited concerns
- Interaction with relevant RDA working and interest groups
- Among 19 identified groups, for instance
  - IG RDA/CODATA Materials Data, Infrastructure & Interoperability
  - IG RDA/NISO Privacy Implications of Research Data Sets
- Co-chair election every two years

# Timeline

- Further grow initial members and identify co-chairs
- Finalize and submit the Charter for review
- Obtain RDA endorsement for (i.e., establish) the IG for P13
- If endorsed IG kickoff meeting at P13
- Otherwise second BoF at P13



# Initial members

Name	Affiliation	Country
Marta Teperek	TU Delft	Netherlands
Susanna-Assunta Sansone	Uni of Oxford, Dep of Engineering Science (and RDA FAIRsharing WG)	UK
Alastair Dunning	TU Delft	Netherlands
Daniela Hausen	RWTH Aachen University	Germany
Angelina Kraft	TIB Leibniz Information Centre for Science and Technology	Germany
Markus Stocker	TIB Leibniz Information Centre for Science and Technology	Germany
Gerald Jagusch	ULB Darmstadt	Germany
Nanette Reißler-Pipka	Karlsruhe Institute of Technology (KIT)	Germany

... needed are co-chairs, max four and at least two regions

# Research Data Management in Engineering disciplines



# Where is TU Delft?

## (Delft University of Technology)



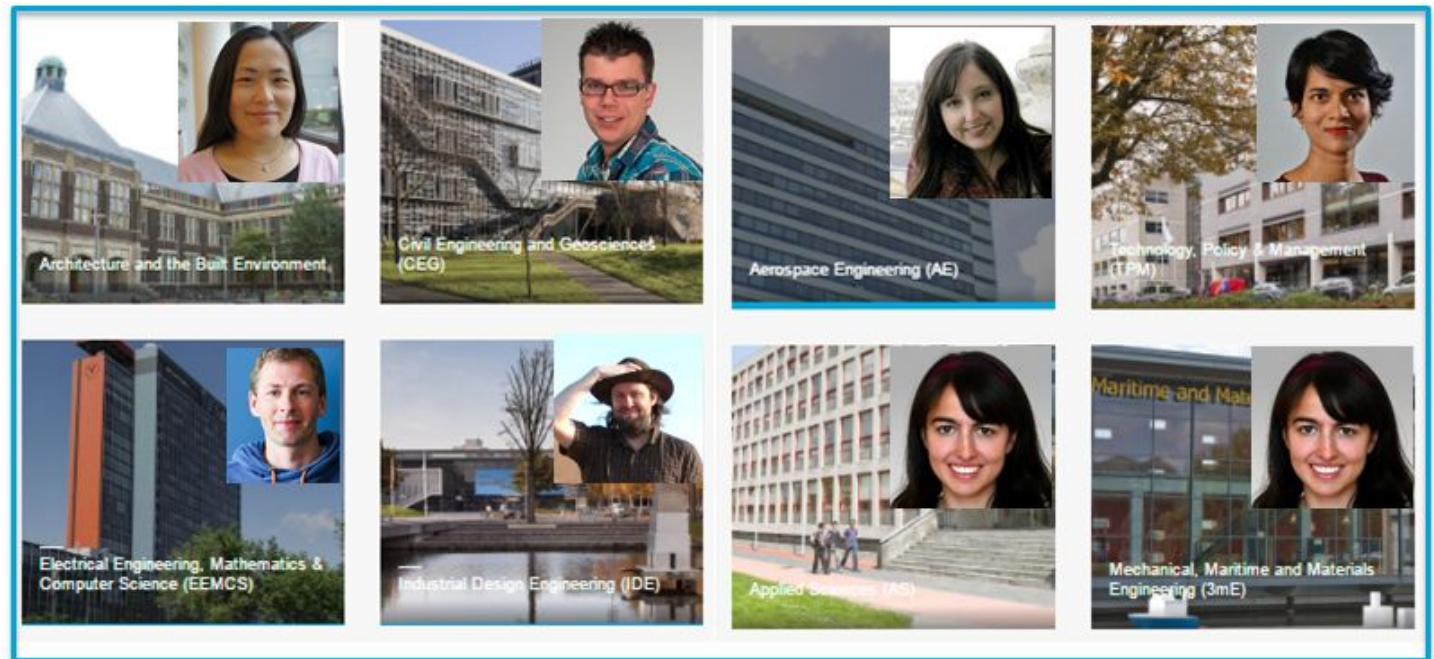


Good data management is a necessary prerequisite to FAIR Data and Open Science



# Research is central: importance of disciplinary support

Subject-specific Data Steward at every Faculty



<https://www.tudelft.nl/en/library/current-topics/research-data-management/research-data-management/data-stewardship/contact/>

But what does it mean to manage data properly in Engineering disciplines?



# But what does it mean to manage data properly in Engineering disciplines?

- There are not many standards for Engineering disciplines
- Problems with:
  - Day to day data management
  - Collaboration with different research groups
  - Collaboration with industry
  - Interdisciplinary collaboration
  - Data sharing
  - Funders' policy compliance (data management plans)
- See <https://rdm.engineering/> for more

# Three-ish initiatives

- Domain data protocols
- Cross-disciplinary collaboration in coastal engineering
- Common taxonomy and metadata in wind energy by the FP7 IRPWind Project

(Also use of NetCDF as a format within engineering -  
<https://zenodo.org/record/1465950>)

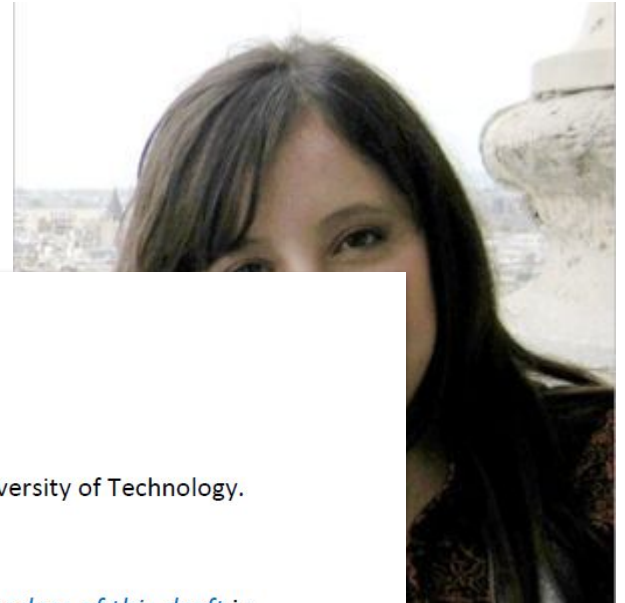


# Domain data protocols



- Instead of writing lengthy data management plans, researchers could “endorse” protocols from their research domains

# Domain data protocols



## Core Requirements for Astronomical Imaging Data

Version September the 20<sup>th</sup>, 2018

Creator: Dr. Heather Andrews, Data Steward at Faculty of Aerospace Engineering at Delft University of Technology.

*Text to be filled in by the researcher* is depicted in red, while *extra information for readers of this draft* is described in italic blue font. This documents is for one type of data used by astronomers. An astronomical project would in most cases include more than imaging data only.

Andrews  
Faculty of  
Engineering

Core Requirement	Guiding questions
1. Data description and collection or re-use of existing data	Data Type: Astronomical imaging data from ground-based and/or space-based telescopes.  Data Format: Flexible Image Transport System format

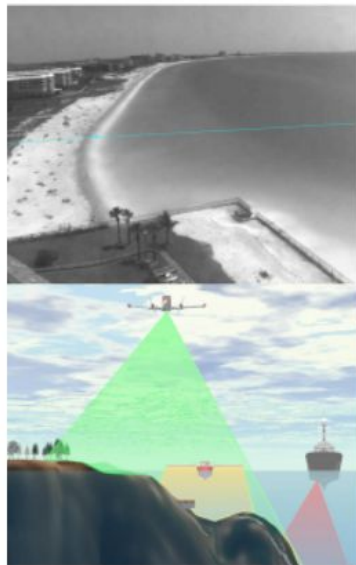
# Open Earth - Cross-disciplinary collaboration in coastal engineering

## Procedure

The data collection procedure consists of four phases.

1. Extract. Collecting data and storing the measured data into files.
2. Transform. Enriching gathered data with metadata and storing in a standard file format.
3. Load. Storing the files in a database.
4. Provide. Giving access to the database.

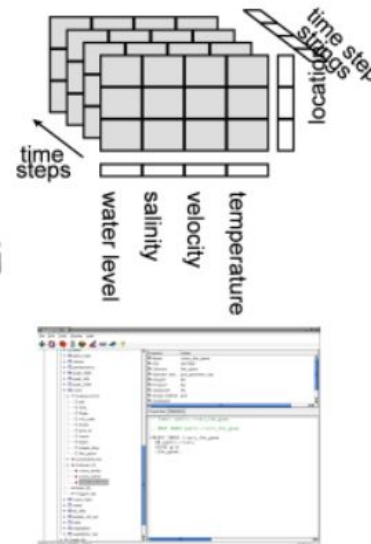
### Extract



### Transform



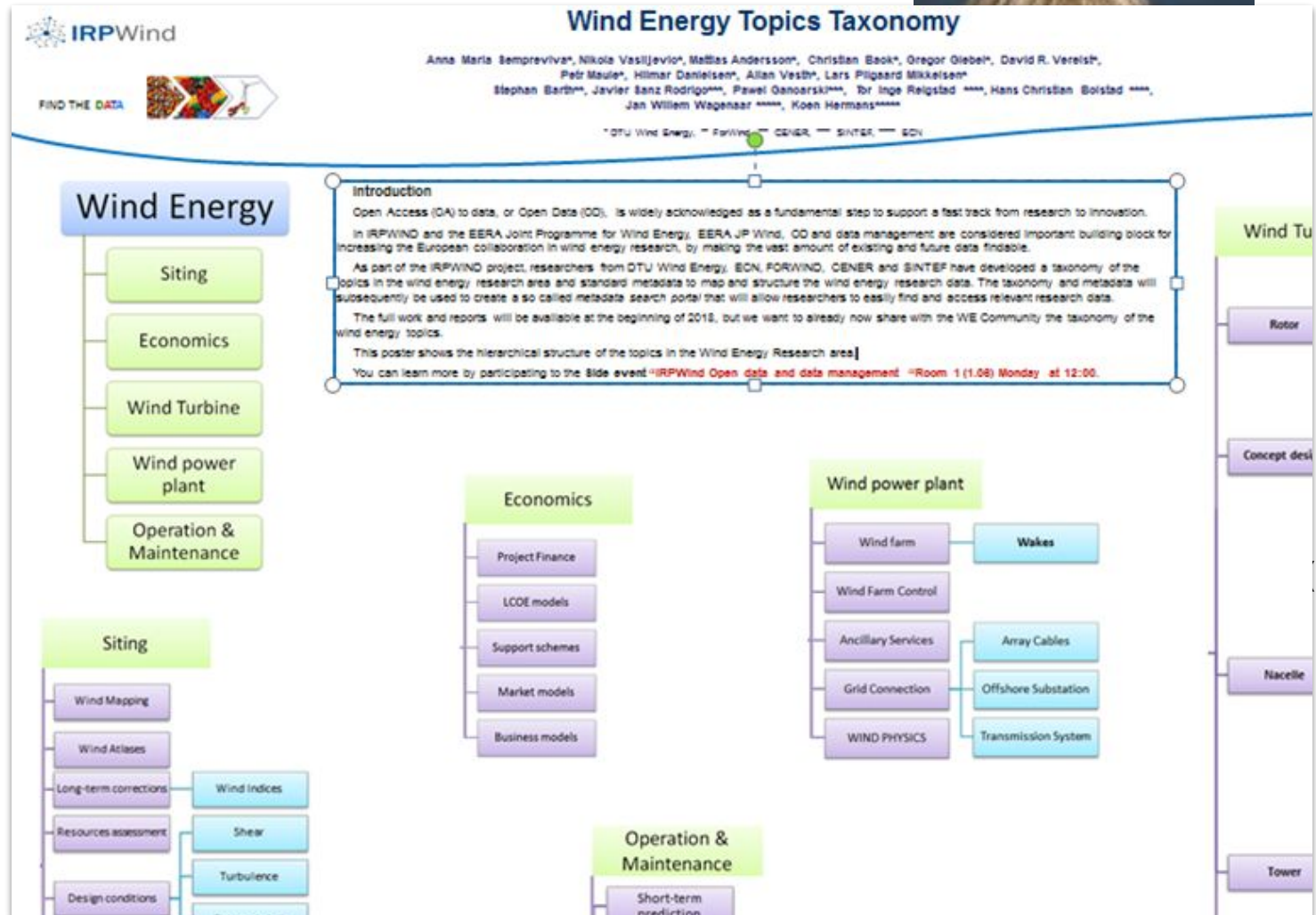
### Load



### Provide



# Common taxonomy and metadata in wind energy by the FP7 IRPWind Project



# Remaining questions

- How to support community efforts?
- How to allow discipline level consultations and adoption?
- How to support these efforts and how to provide adequate resources?
- Collaboration with FAIRsharing?
  - How to describe the standards?
  - How to make standards for data management findable in FAIRsharing?



# Contact



# Discussion

- Which problems with data (management, sharing, storing, etc.) in engineering can be specified?
- Which of them can be addressed interdisciplinary and which of them are specific for sub-disciplines (like constructing engineering, material science, mechanical engineering, etc.)?
- Possible differences of workflows, interests, problems?
- Necessary steps to be done during the data life cycle might be rather specific (context of metadata, provenance, analysis, preprocessing, etc.)?
- Relevant disciplines and concrete use cases?

# Dialog with Engineers present at BoF session

- Which disciplines are present in the room?
- Can you describe the specific character of data problems giving an example of your discipline?
- Example: Material Science (KIT): researchers come with probes of material to use the microscope (for taking the images (2D, 3D, etc.) at home): problem: they come without metadata for the probes (we know nothing about the origin, storing, handling of the material - factors influencing the results)  
Solution: general, standardized method to describe (to enrich) probes with metadata



# Next steps and conclusion

- Is a RDM for engineering an important addition to RDA IG/disciplines?
- If yes,
- Finalize the Charter
- Include BoF feedback and possible post-meeting commentary
- Name co-chairs and extend initial membership
- Submit the Charter
- Plan for P13 (Philadelphia)
- Ideally endorsement by P13
- Should be feasible if we submit by the end of the year