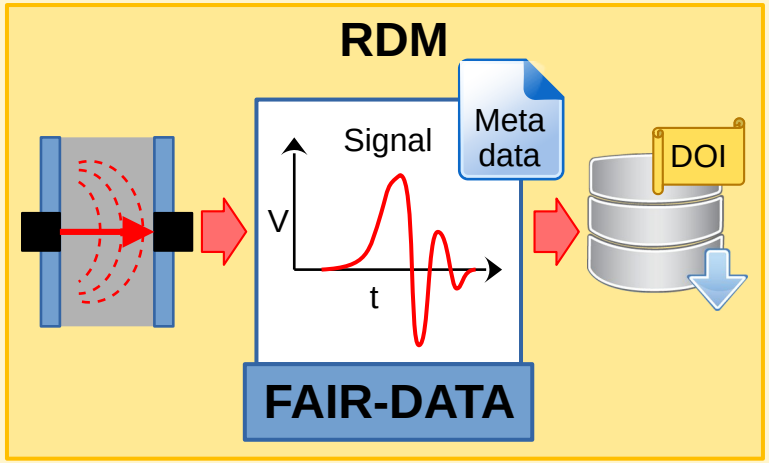


RDM - from Theory to Practice:

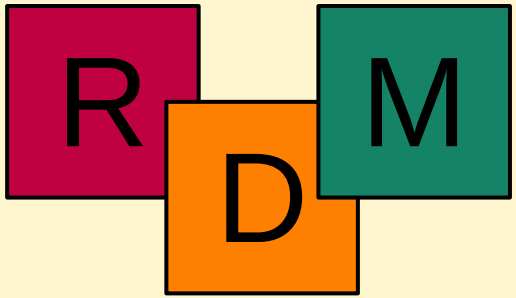
An Application Example based on Measurement Data gained by Ultrasonic Pulse Transmission Tests

Jakob Harden
Graz University of Technology

Table of Contents

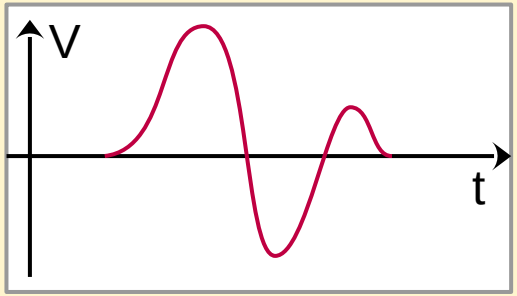
Introduction to RDM

- Data Lifecycle
- FAIR Data Principles
- Data Management Plans
- Metadata Standards
- Publish data



Application example

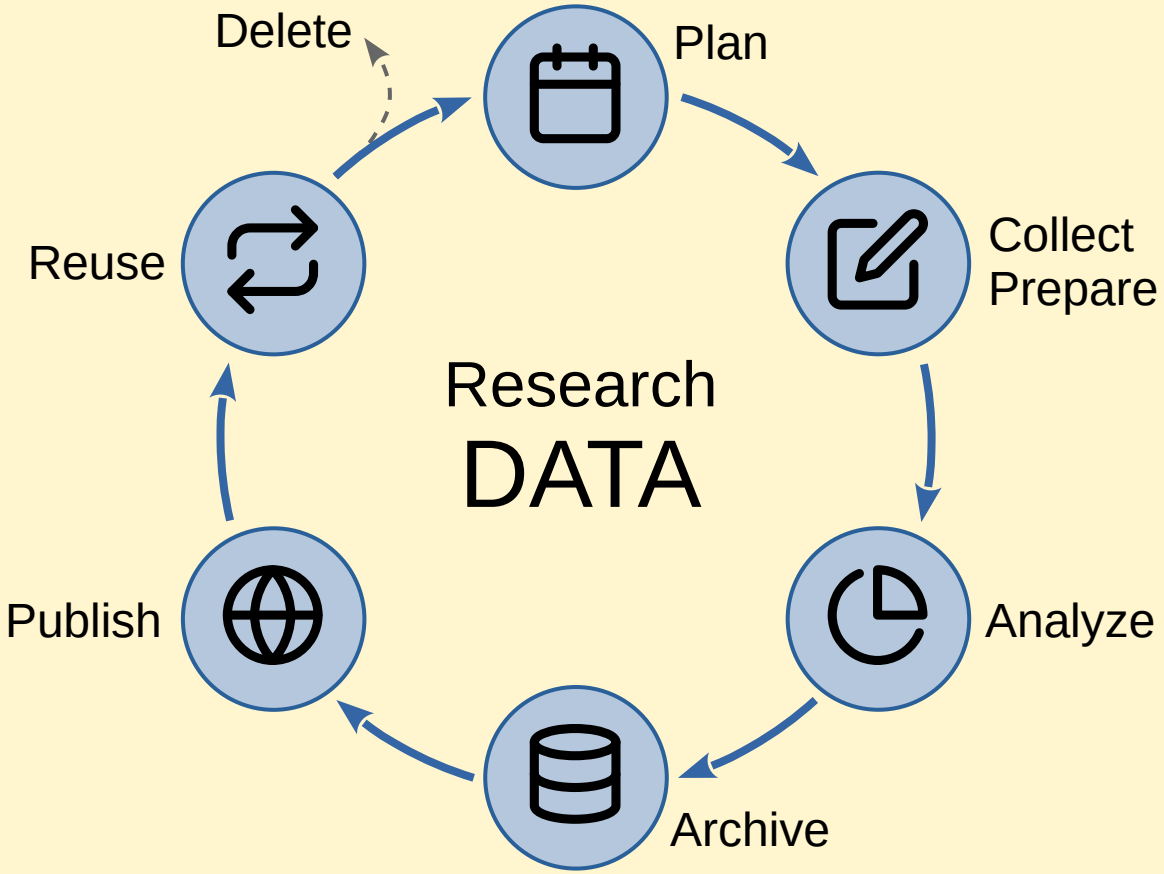
- Measurement data and metadata
- Build data set
- Reuse data set
- Show-case



Data Lifecycle

Data are „living objects“

- Livetime
- Resources
- Description



FAIR Data Principles

FAIR & CARE	… Example
• **F**indable	… available on repository, indexed data
• **A**ccessible	… free download, no paywall, long term storage
• **I**nteroperable	… common and open file formats
• **R**eusable	… well documented, metadata available
• **C**ollective benefit	… open access, publish data
• **A**uthority to control	… right to data erasure, anonymization of data
• **R**esponsibility	… consider impact on personal rights
• **E**thics	… rules to handle personal/patient data, good scientific practice

Data Management Plans

What is a DMP?

- How to deal with data
- Enable subsequent use
- Prevent loss of data
- Considerations:
  - Responsibilities
  - Human and financial resources
  - Metadata and data formats

Goal and benefit of DMP's

- More clarity in data handling
- Make data long-term
- Make data comprehensible and adaptable
- Make data handling more efficient

Metadata Standards

What is Metadata?

- Describe data
- Provided along with the data
- Required for subsequent use
- General standards
  - DataCite
  - Dublin Core
  - ...
- Discipline specific standards
  - Many standards available

Answer the 5 “W”-questions!

- **W**ho?
- **W**here?
- **W**hat?
- **W**hen?
- **W**hat for?

„Data without metadata is NOT research data!“

[Jakob Harden, 2023]

Publish Data

Why?

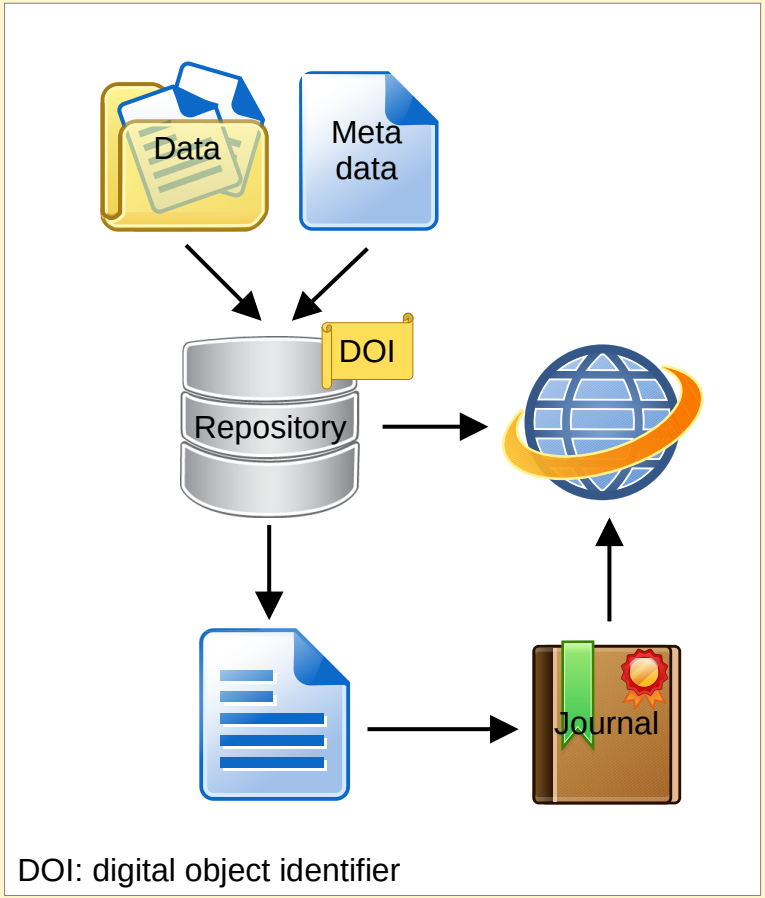
- Make your work more visible
- Increase impact of your research
- Collective benefit for the scientific community

Where?

- Repositories
  e.g. TU Graz, Zenodo, ResearchGate, ...

How?

- Record on repository (DOI)
- Data paper (data journal)
  e.g. Science Data Journal (Nature Publishing Group)

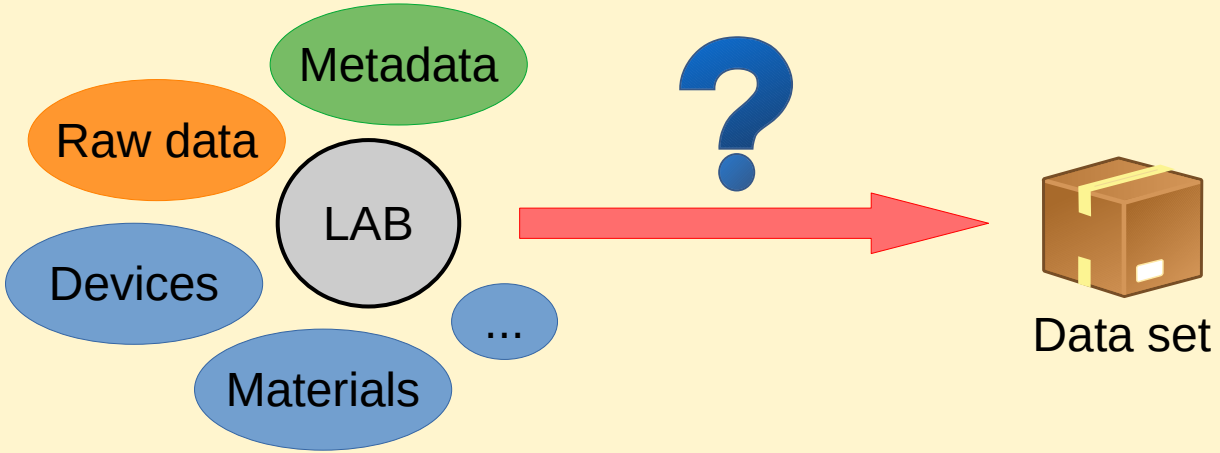


DOI: digital object identifier

Laboratory test data

Ultrasonic pulse transmission tests on cement pastes at early stages

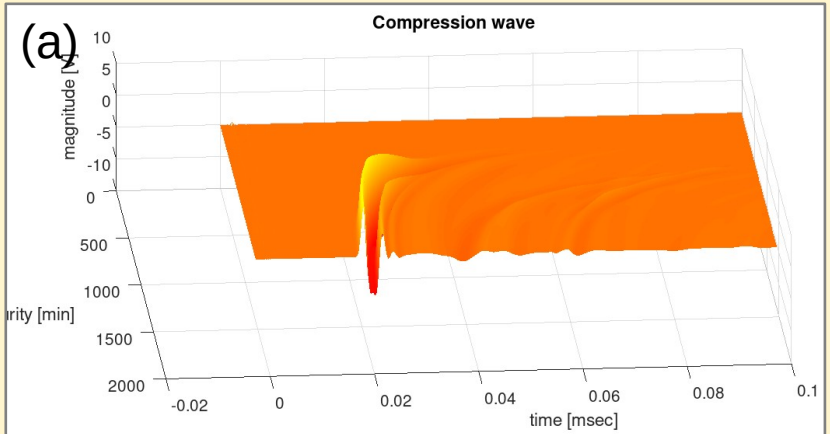
- **Metadata:** test series information (author, name, start-date, end-date, …)
- **Metadata:** test information (author, name, description, …)
- **Metadata:** supplementary information (devices, materials, mix recipe, specimen, …)
- **Raw data:** different sources (device output, lab notes)
- **Raw data:** different file formats (plain text files, spreadsheet)

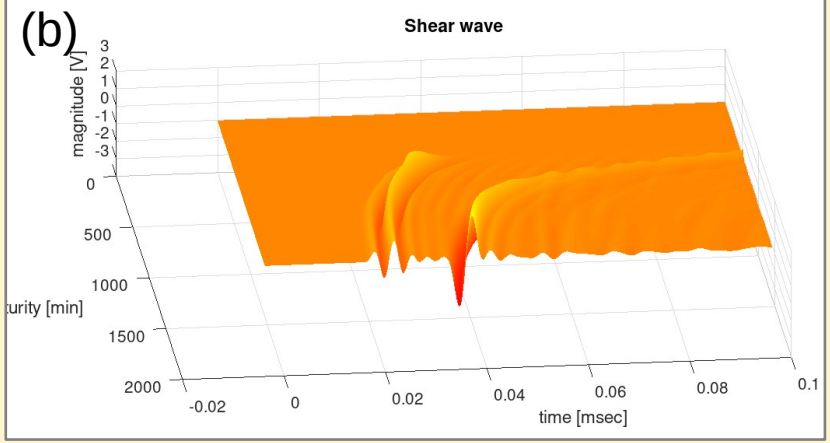


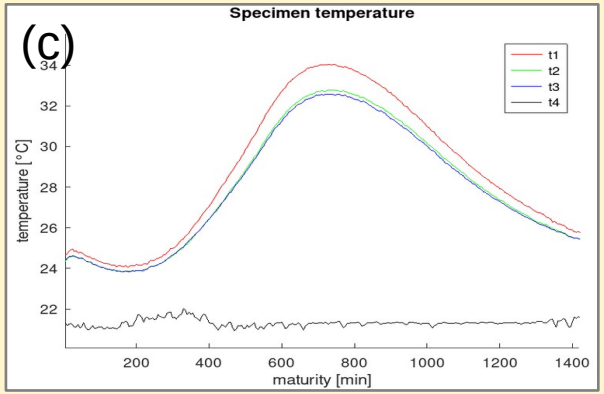
Laboratory test data

Raw data

- US test I, compression waves, 290 text files (a)
- US test II, shear waves, 290 text files (b)
- Specimen temperature, text file (c)
- Specimen size (lab notes)
- Fresh paste density (lab notes)
- Solid specimen density (lab notes)
- Environment temperature (lab notes)



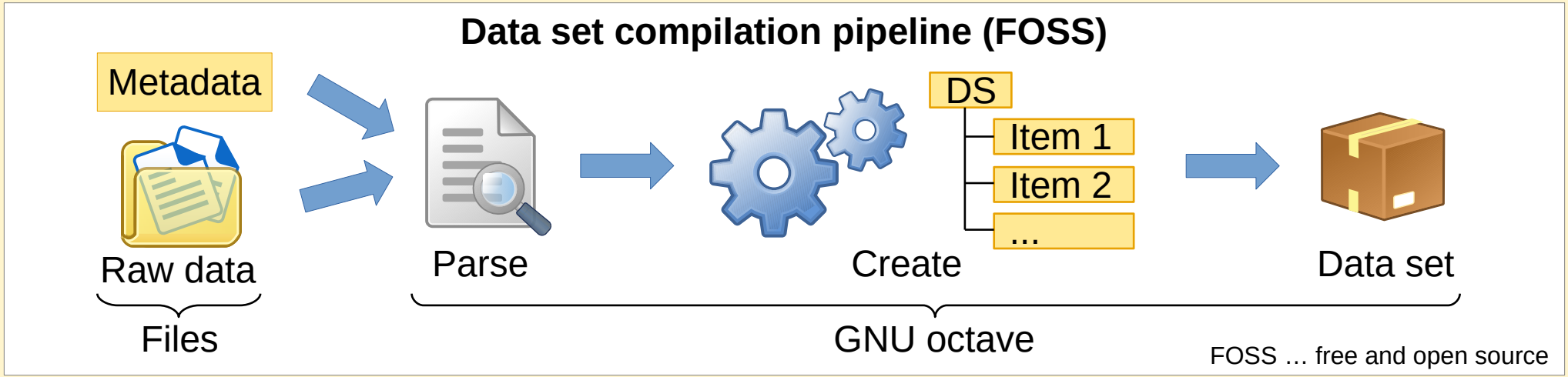




Data set compilation

From raw data to data set

- Device data and lab notes (text files, spreadsheets)
- Metadata (test series, devices, materials, mixture recipe, specimen, location, license...)
- Parse files (GNU octave file parser)
- Create data structure (GNU octave import script)
- Save **data set** (GNU octave binary file)



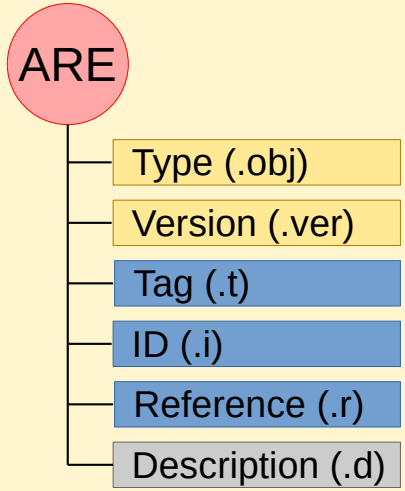
Hierarchical data structure

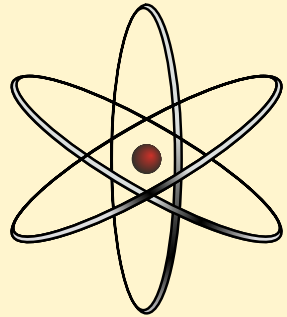
Hierarchical data structure

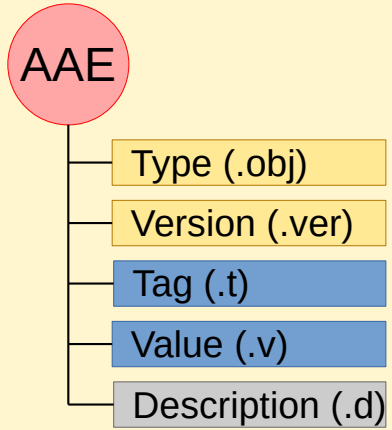
- C-like structure type
- Consists of substructures
- Made of simple base elements (atomic)

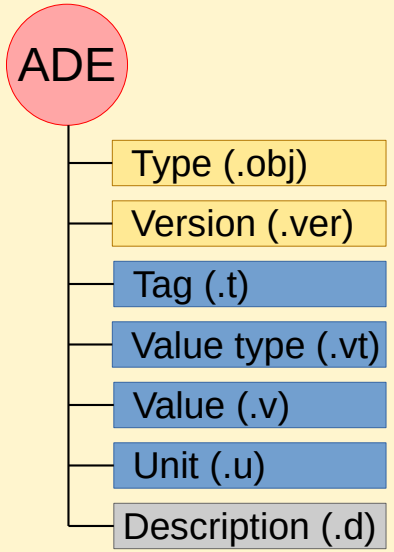
Atomic structure elements

- Used to store data, attributes, references
- Fields: type, version, tag, value, ...
- Atomic data element (ADE)
- Atomic reference element (ARE)
- Atomic attribute element (AAE)









Hierarchical data structure

Content	Level	Type	Path
Dataset	0	Structure root	ds
Test series metadata	1	Substructure	ds.meta_ser
Data set metadata	1	Substructure	ds.meta_set
Author information	1	Substructure	ds.aut
Location information	1	Substructure	ds.loc
Device information	1	Substructure	ds.dev
Material information	1	Substructure	ds.mat
Specimen information	1	Substructure	ds.spm
...	1	Substructure	ds. ...
Test collection	1	Substructure	ds.tst
Solid specimen density I	2	Substructure	ds.tst.s02
US test I	2	Substructure	ds.tst.s06
...	2	Substructure	ds.tst. ...

General information

Tables

Metadata

Measurement data, References to tables, Device settings

Hierarchical data structure

Content	Level	Type	Path
Author info	**1**	**Substructure**	**ds.aut**
Object type	2	Structure field	ds.aut.obj
Object version	2	Structure field	ds.aut.ver
Author ID	2-3	ADE	ds.aut.d01
Author name	2-3	AAE	ds.aut.a01
...	...	...	ds.aut. ...

Example: author information

ds.tst.s06.r01.r

{“aut”}

“struct_test_utt“

[1, 0]

ds.tst.s06.r01.i

1

Content	Level	Type	Path
US test I	**2**	**Substructure**	**ds.tst**
Object type	3	Structure field	ds.tst.s06.obj
Object version	3	Structure field	ds.tst.s06.ver
Related author(s)	3-4	ARE	ds.tst.s06.r01
Signal time array	3-4	ADE	ds.tst.s06.d12
Signal magn. array	3-4	ADE	ds.tst.s06.d13
...	...	...	ds.tst.s06. ...

Example: ultrasonic pulse transmission test, specimen I

ds.tst.s06.d12.v

1	0.1231
2	0.1232
3	0.1233
4	0.1235

ds.tst.s06.d12.u

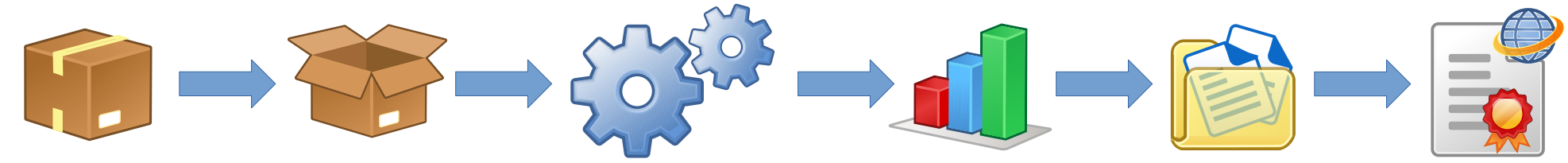
“sec”

Reuse data set

From data set to publication

- Access data (GNU octave structure handling)
- Analyse data (GNU octave analysis script)
- Plot data (GNU octave plot functions)
- Export data (GNU octave TeX export script)
- Publish data (LaTeX, TeX compiler)

Data analysis and publishing pipeline (FOSS)



FOSS … free and open source

Reuse data set

Access data set and plot data set content

Load dataset, access structure elements

```
> ds = load('/path/to/file.oct').dataset;
> us1_t_v = ds.tst.s06.d12.v;
> us1_t_u = ds.tst.s06.d12.u;
> us1_m_v = ds.tst.s06.d13.v;
> us1_m_u = ds.tst.s06.d13.u;
```

GNU octave CLI

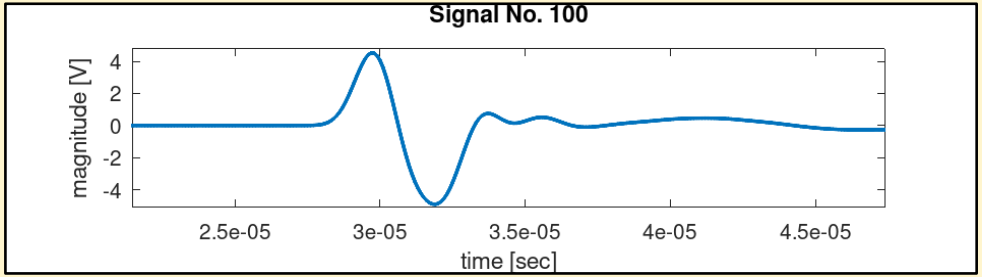
```
→ struct
→ „sec"
→ [0.1; 0.2; 0.3; 0.4; …]
→ [[0.1; 0.2; …], [0.3; 0.4; …]; …]
→ „V"
```

Output

Plot signal

```
> figure();
> plot(us1_t_v, us1_m_v(:, 100));
> xlabel(sprintf(„time [%s]", us1_t_u));
> ylabel(sprintf(„magnitude [%s]", us1_v_u));
> title(„Signal No. 100");
```

GNU octave CLI



Output

Reuse data set

Export and publish data set content

Export variable to TeX file

```
> s = sprintf(„\\def\\myunit{%s}", ds.us1_t_u);
> fid = fopen(„./path/to/myexport.tex");
> fprintf(fid, „%s\n", s);
> fclose(fid);
```

GNU octave CLI

```
→ s = ‘\def\myunit{sec}‘


→ ./path/to/myexport.tex
```

Output

Integrate TeX file in LaTeX code

```
\document{standalone}
\input{myexport}
\begin{document}
   The signal sample time unit is \myunit.
\end{document}
```

LaTeX document

The signal sample time unit is sec.

Output

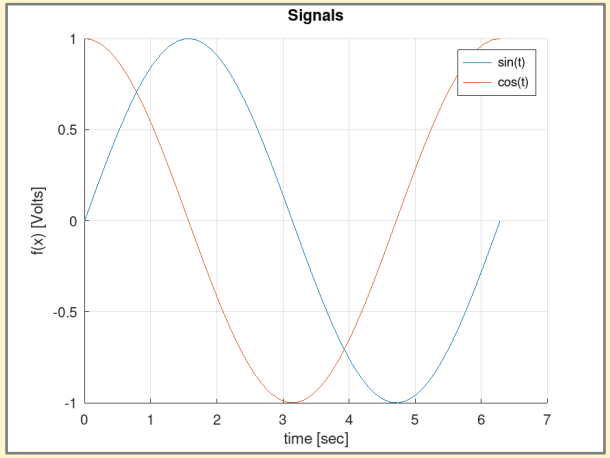
Show-case

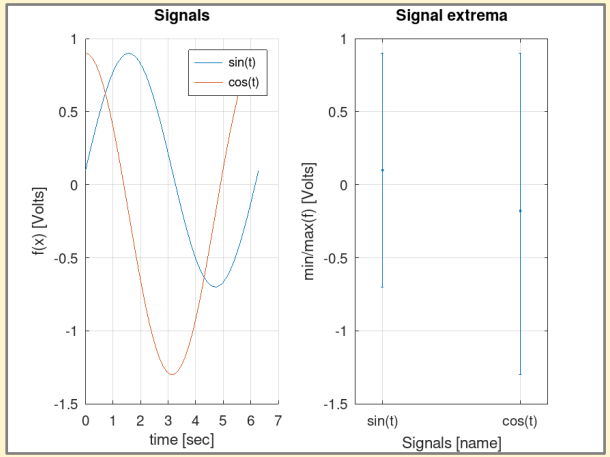
RDM workflow

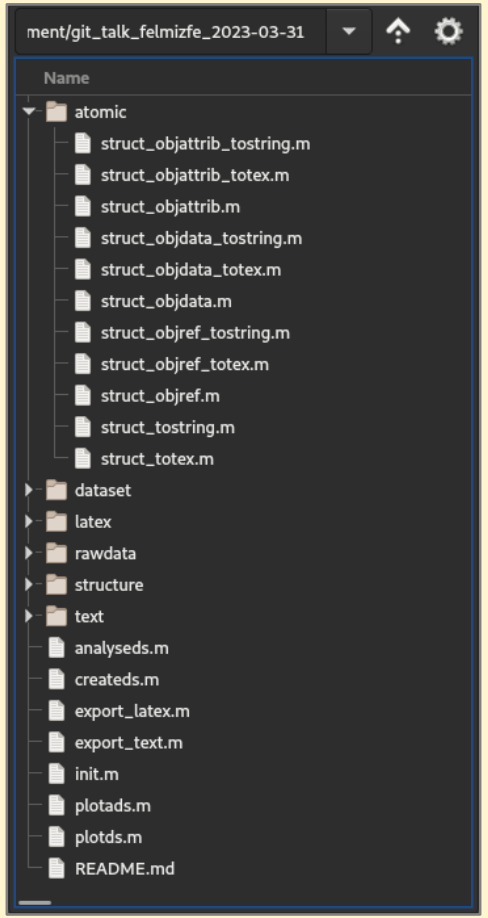
- Compile data set from raw data
- Analyse data
- Plot data
- Export data to plain text file
- Export data to TeX file
- Publish data with LaTeX

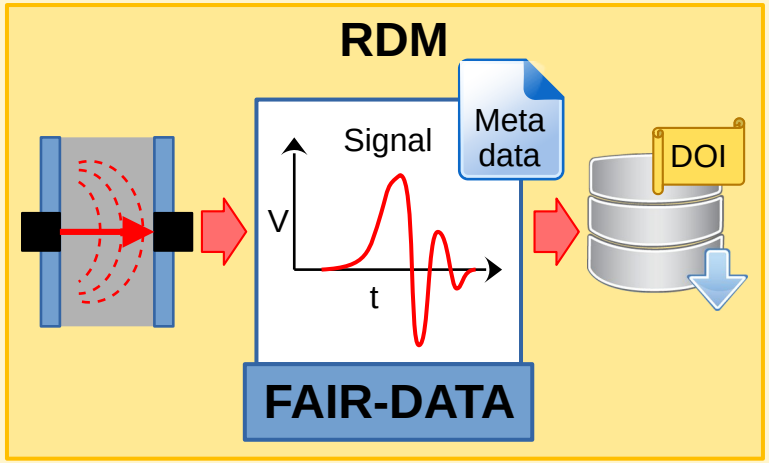
Code available

- DOI: 10.3217/58jhw-1b343
  https://repository.tugraz.at/









Thank you for your attention!

Literature and Links

RDM, FAIR & CARE principles, data lifecycle

- TU Wien: https://rdm.univie.ac.at/research-data-management/
- TU Graz: https://www.tugraz.at/sites/rdm/home

Data management plans

- TU Graz: https://www.tugraz.at/sites/rdm/dmps/data-management-plans

Metadata

- Metadata Standards Directory: https://rd-alliance.github.io/metadata-directory/standards/
- fairsharing.org: https://fairsharing.org/
- Datacite: https://datacite.org/
- Dublin Core: https://www.dublincore.org/

Repositories

- TU Graz: https://repository.tugraz.at/
- Zenodo: https://zenodo.org/
- ResearchGate: https://www.researchgate.net/

Jakob Harden
Graz University of Technology

Literature and Links

Data journals

- Forschungsdaten.org: https://www.forschungsdaten.org/index.php/Data_Journals (list of journals)
- Science Data: https://www.nature.com/sdata/ (Nature Publishing Group)

Hierarchical data formats

- hdfgroup.org: https://hdfgroup.org/solutions/hdf5 (HDF5)
- ASDC: https://asdc.larc.nasa.gov/documents/tools/hdf.pdf (HDF5)
- GNU octave: https://docs.octave.org/v4.0.3/Simple-File-I_002fO.html#Simple-File-I_002fO
- Matlab: https://de.mathworks.com/help/matlab/import_export/supported-file-formats-for-import-and-export.html
- Unidata: https://www.unidata.ucar.edu/software/netcdf/ (netCDF)

Jakob Harden
Graz University of Technology

Document Information

Jakob Harden
Graz University of Technology