

# How reproducible are your EPR results?

## Tools for fully reproducible data analysis

- Metadata acquisition
- Gap-less protocol of processing and analysis
- Reliable, high-quality software

Till Biskup

56<sup>th</sup> RSC ESR Meeting, Leeds, 2023  
30/03/2023



# Reproducible research

The FAIR Guiding Principles for scientific data management and stewardship



**F**indable   **A**ccessible   **I**nter-  
operable   **R**eusable



Wilkinson *et al.*, *Scientific Data* 3:160018, 2016Mons, *Nature* 578:491, 2020

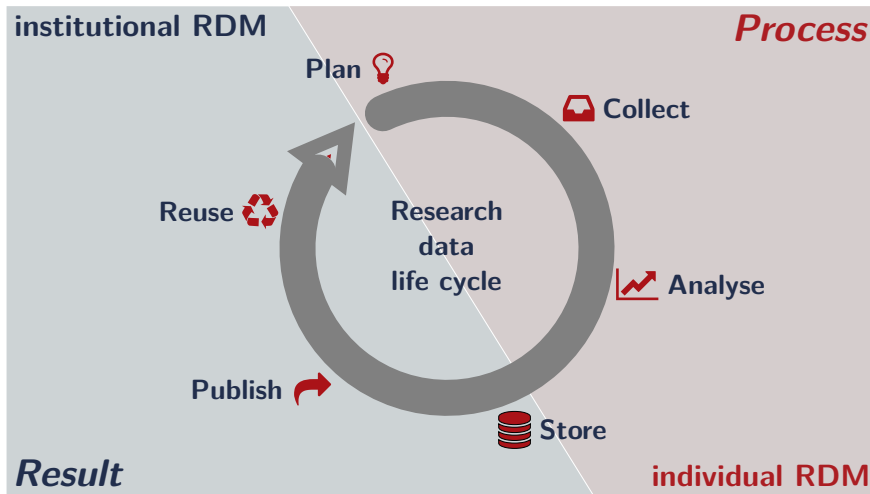
“ *If I have seen further  
it is by standing on y<sup>e</sup> shoulders of giants.*

– Sir Isaac Newton

- ❓ How can I make sure others can reproduce what I have done?
- ❓ For how long do I remember myself what I have done?
- ❓ What is the usual length of stay of people in a group?
  
- ❗ We need to document *in sufficient detail* what we have done.
- ❗ Documentation needs to be automated wherever possible.

# Reproducible research

The research data life cycle as an abstract model



### Collect

- ▶ recording metadata during data acquisition
- ▶ Who has done what with whom when how and why?
- ☞ machine-readable and *human-writable*
- ☞ recording cannot be (fully) automated

### Analyse

- ▶ gap-less protocol of each analysis step
- ▶ fully reproducible data processing and analysis
- ☞ recording *fully automatically*
- ☞ full reproducibility only with scientific workflow system

---

```
format:
  type: ASpecD recipe
  version: '0.2'
```

```
datasets:
  - /path/to/first/dataset
  - /path/to/second/dataset
```


```
tasks:
  - kind: processing
    type: BaselineCorrection
    properties:
      parameters:
        kind: polynomial
        order: 0
  - kind: singleplot
    type: SinglePlotter1D
    properties:
      filename:
        - first-dataset.pdf
        - second-dataset.pdf
```

---

---

```
system_info:
  python:
    version: "3.7.3 ..."
  packages:
    aspectd: 0.6.4
# ...
- kind: processing
  type: BaselineCorrection
  properties:
    parameters:
      kind: polynomial
      order: 0
      coefficients:
        - -0.04609818536259180
      fit_area:
        - 10
        - 10
      axis: 0
  apply_to:
    - /path/to/first/dataset
# ...
```

---

 J. Popp, T. Biskup: ASpecD: A modular framework for the analysis of spectroscopic data focussing on reproducibility and good scientific practice. *Chemistry Methods* 2:e202100097, 2022

- 🔑 Recipe-driven data analysis
  - No programming skills needed
- 🔑 Dataset as unit of data and metadata
  - Abstracts away from vendor file formats
- 🔑 Full reproducibility
  - History is a fully working recipe
- 🔑 Modular and extensible
  - Focus on the operation, not the infrastructure
- 🔑 Support for different spectroscopic methods
  - Python packages for dedicated methods available

### ASpecD



<https://docs.aspecd.de/>

J. Popp, T. Biskup. *Chem. Meth.* **2**:e202100097, 2022

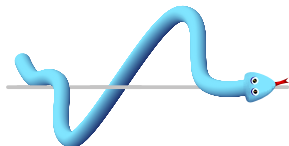
### cwepr



<https://docs.cwepr.de/>

M. Schröder, T. Biskup. *J. Magn. Reson.* **335**:107140, 2022

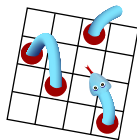
### trEPR



<https://docs.trepr.de/>

J. Popp, M. Schröder, T. Biskup. *repr* (2022).  
doi:10.5281/zenodo.4897112

### FitPy



<https://docs.fitpy.de/>

T. Biskup. *fitpy* (2022). doi:10.5281/zenodo.5920380



Following best practices in software development, e.g.:

- ▶ test-driven development
    - high test coverage, better reliability
  - ▶ clean code
    - readable, expressive, self-documenting
  - ▶ fully documented
    - <https://docs.aspecd.de/>
  - ▶ version control system
    - <https://github.com/tillbiskup/aspecd/>
  - ▶ open source
    - BSD license: everybody is allowed use and modify it
- 👉 Everybody who can program Python could take over the project.

---

```
common Infile - v. 0.1.0
```

```
GENERAL
```

```
Date start: 2020-04-04
Time start: 11:05:00
Date end: 2020-04-04
Time end: 15:50:00
Operator: John Doe
Purpose: Kill time
```

```
SAMPLE
```

```
Name: Random sample 1
Description: Nicked from bench neighbour
```


```
COMMENT
```

```
To be or not to be...
```

---

- Who has done
- what
- with whom
- when
- how
- and why?

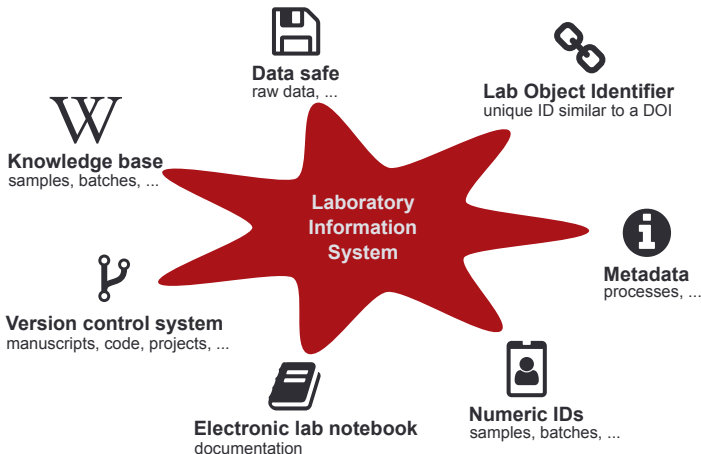
 Many more blocks/parameters for specific methods


 B. Paulus, T. Biskup: Towards more reproducible and FAIRer research data: documenting provenance during data acquisition using the Infile format. *Digital Discovery* 2:234–244, 2023.

- 🔑 Plain text file, human-writable
  - No external dependencies
- 🔑 Resides next to the data
  - Independent of vendor formats; data and metadata always together
- 🔑 Machine-actionable metadata
  - Analysis routines can make sense of your data.
- 🔑 Sufficient detail
  - Never forget an important detail
- 🔑 Modular and extensible
  - Easy to adapt to specific needs

# Reproducible research – the larger picture

Making use of the tools available in the digital era



 T. Biskup: LabInform: A modular laboratory information system built from open source components. *ChemRxiv* 2022, DOI:10.26434/chemrxiv-2022-vz360

# Acknowledgements

People helping to shape the ideas presented and make them become true



## Implementations

- ▶ Bernd Paulus
- ▶ Jara Popp
- ▶ Mirjam Schröder

## Crucial ideas

- ▶ Thomas Berthold

## Encouragement

- ▶ Erik Schleicher
- ▶ Stefan Weber

## Discussions

- ▶ Carolin Odebrecht
- ▶ Kathrin Schmitt

**You for your kind attention!**