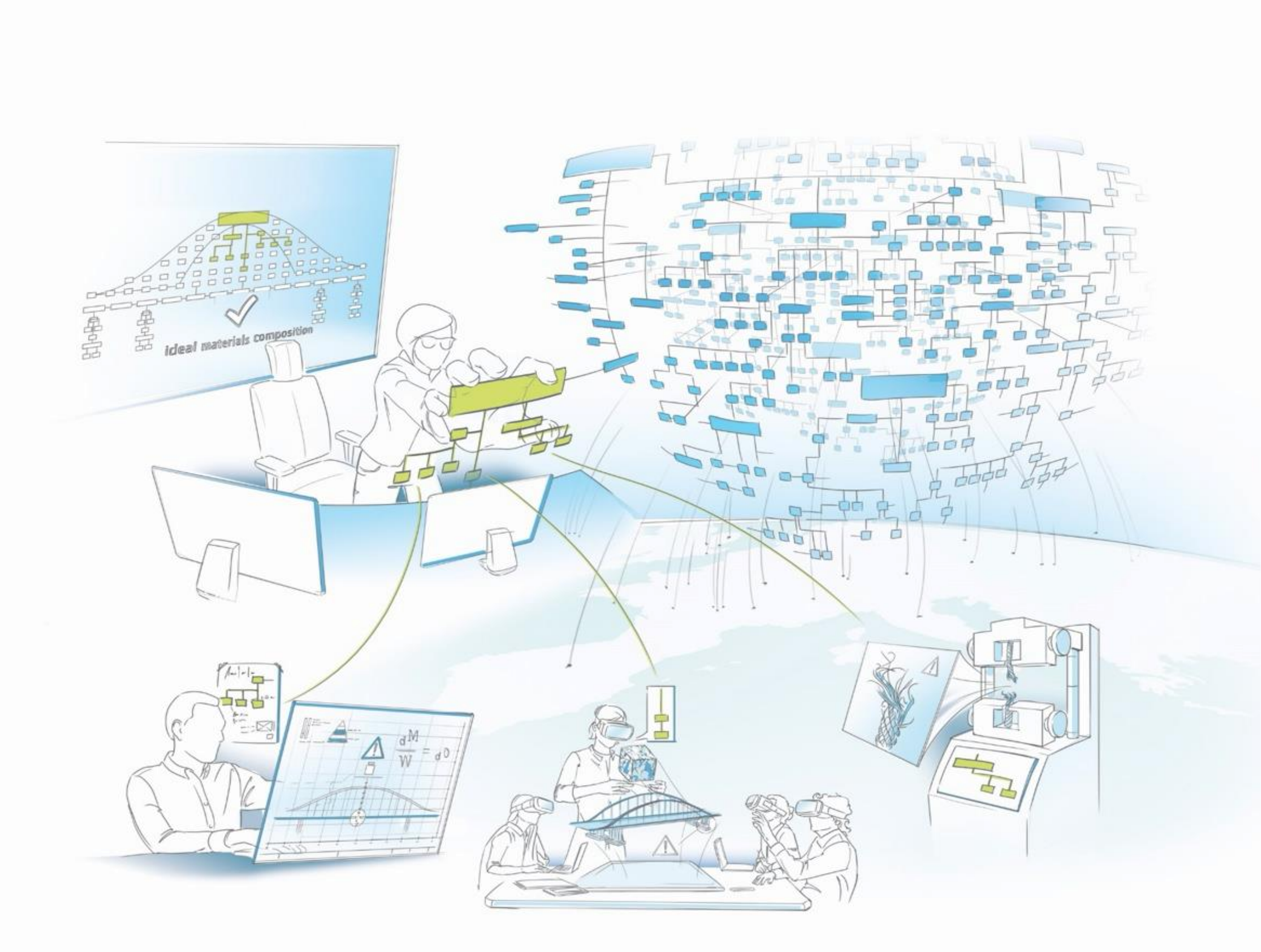
Creating and running automated workflows for material science simulations



Sarath Menon

Max-Planck-Institute for Sustainable Materials

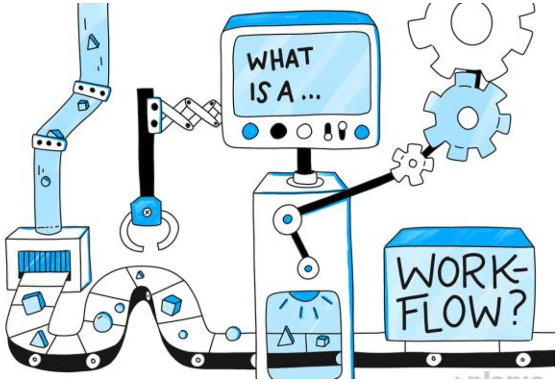
NATIONAL RESEARCH DATA INFRASTRUCTURE FOR MATERIALS SCIENCE & ENGINEERING

NFDI MatWerk

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524

Introduction to FAIR workflows

- Complex multi-step methods
- Used for data collection, data preparation, analytics, predictive modelling, and simulation
- Computational workflows are enablers of automated data processing.



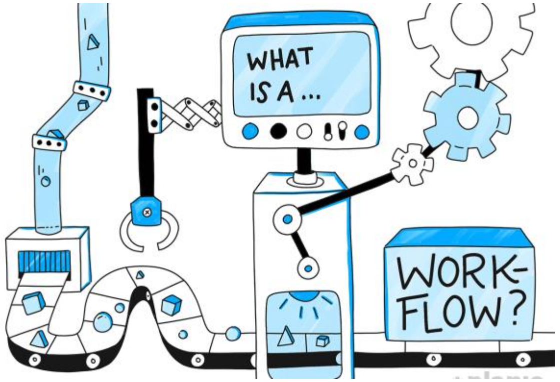
https://plan.io/blog/what-is-a-workflow/

FAIR workflows = FAIR data + FAIR software + Unique requirements

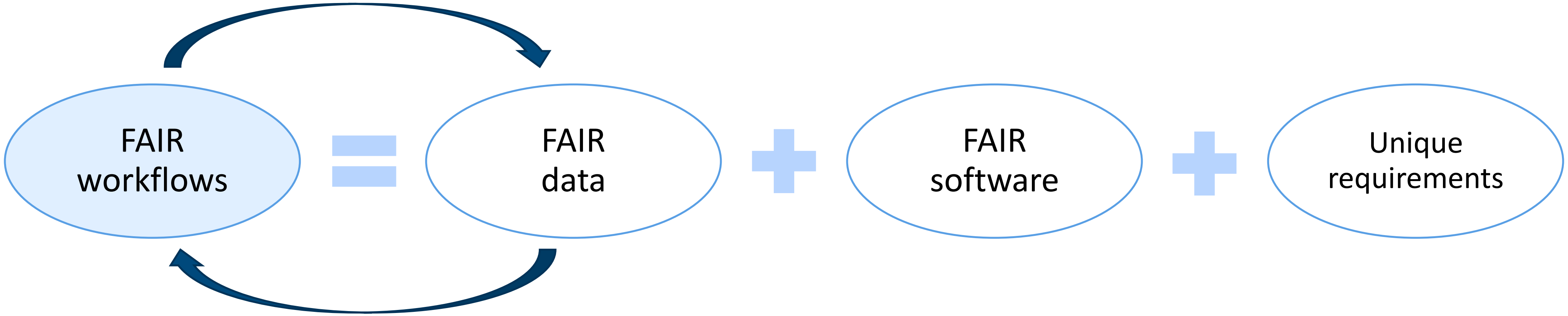
Goble *et al*; FAIR Computational Workflows. *Data Intelligence* 2020; 2 (1-2): 108–121. doi: https://doi.org/10.1162/dint_a_00033

Introduction to FAIR workflows

- Complex multi-step methods
- Used for data collection, data preparation, analytics, predictive modelling, and simulation
- Computational workflows are enablers of automated data processing.



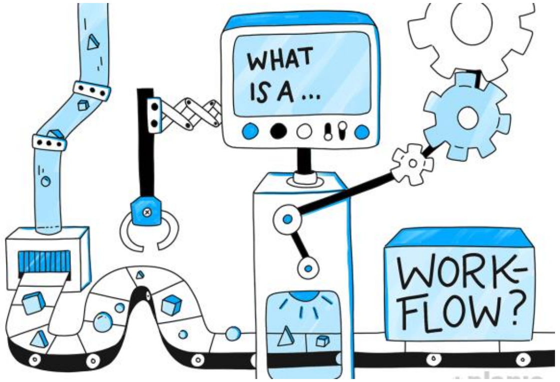
https://plan.io/blog/what-is-a-workflow/



Goble *et al*; FAIR Computational Workflows. *Data Intelligence* 2020; 2 (1-2): 108–121. doi: https://doi.org/10.1162/dint_a_00033

Introduction to FAIR workflows

- Complex multi-step methods
- Used for data collection, data preparation, analytics, predictive modelling, and simulation
- Computational workflows are enablers of automated data processing.



https://plan.io/blog/what-is-a-workflow/

FAIR workflows = FAIR data + FAIR software + Unique requirements

FAIR**R**
Reuse, Rerun,
Replicate,
Repurpose, Recycle,
Reproduce

Goble *et al*; FAIR Computational Workflows. *Data Intelligence* 2020; 2 (1-2): 108–121. doi: https://doi.org/10.1162/dint_a_00033

Published does not mean reproducible

[HTML] A general-purpose machine learning framework for predicting properties of inorganic materials

L Ward, A Agrawal, A Choudhary, C Wolverton

npj Computational Materials, 2016 • nature.com

Abstract

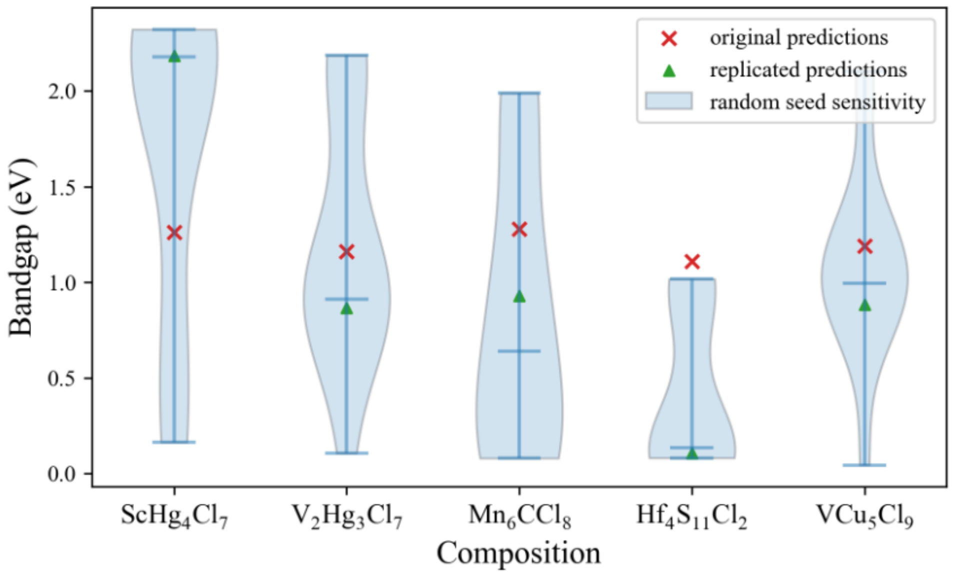
A very active area of materials research is to devise methods that use machine learning to automatically extract predictive models from existing materials data. While prior examples have demonstrated successful models for some applications, many more applications exist where machine learning can make a strong impact. To enable faster development of machine-learning-based models for such applications, we have created a framework capable of being applied to a broad range of materials data. Our method works by using a

SHOW MORE

Reproducibility in Computational Materials Science: Lessons from'A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials'

D Persaud, L Ward, J Hattrick-Simpers

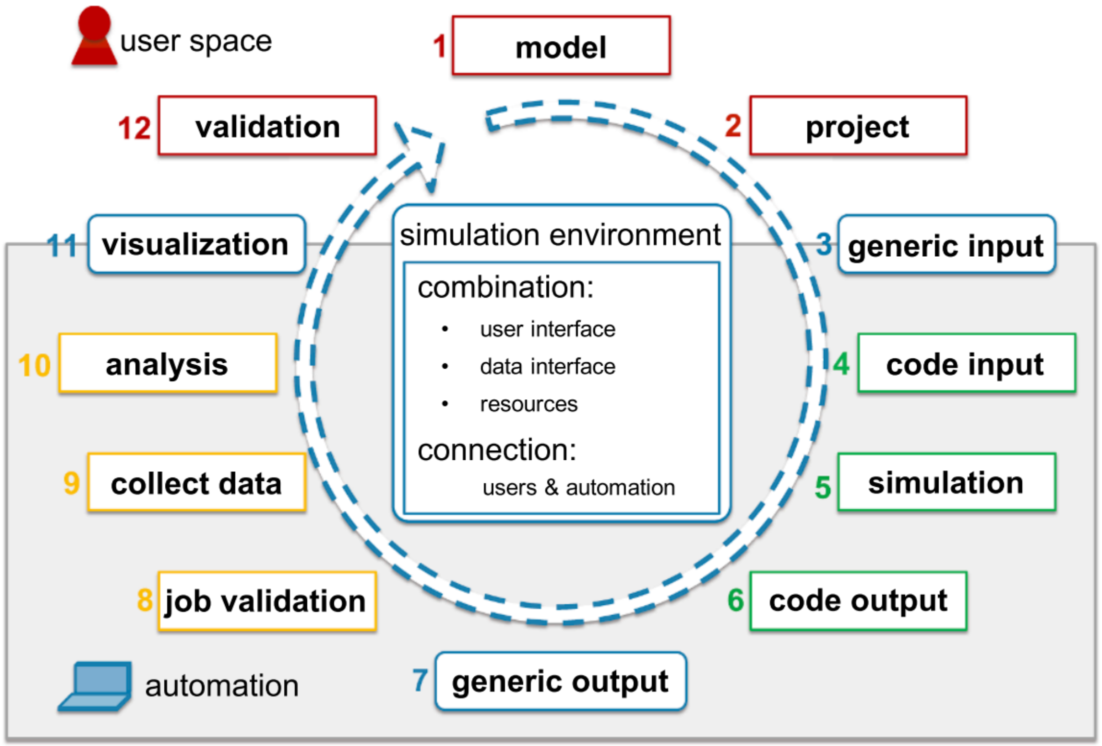
arXiv preprint arXiv:2310.07044, 2023 • arxiv.org



Major challenges:

- Sequential code organization: which steps in what order?
- Reporting computational dependencies and versioning

Computational workflows in MatWerk



- **Ease of use:** Separate user and code space, generic interface, code agnostic
- **Scalable:** Prototype in notebooks, run on HPC resources
- **Shareable:** Mybinder, workflow templates
- **Extensible:** Integrate codes easily, examples

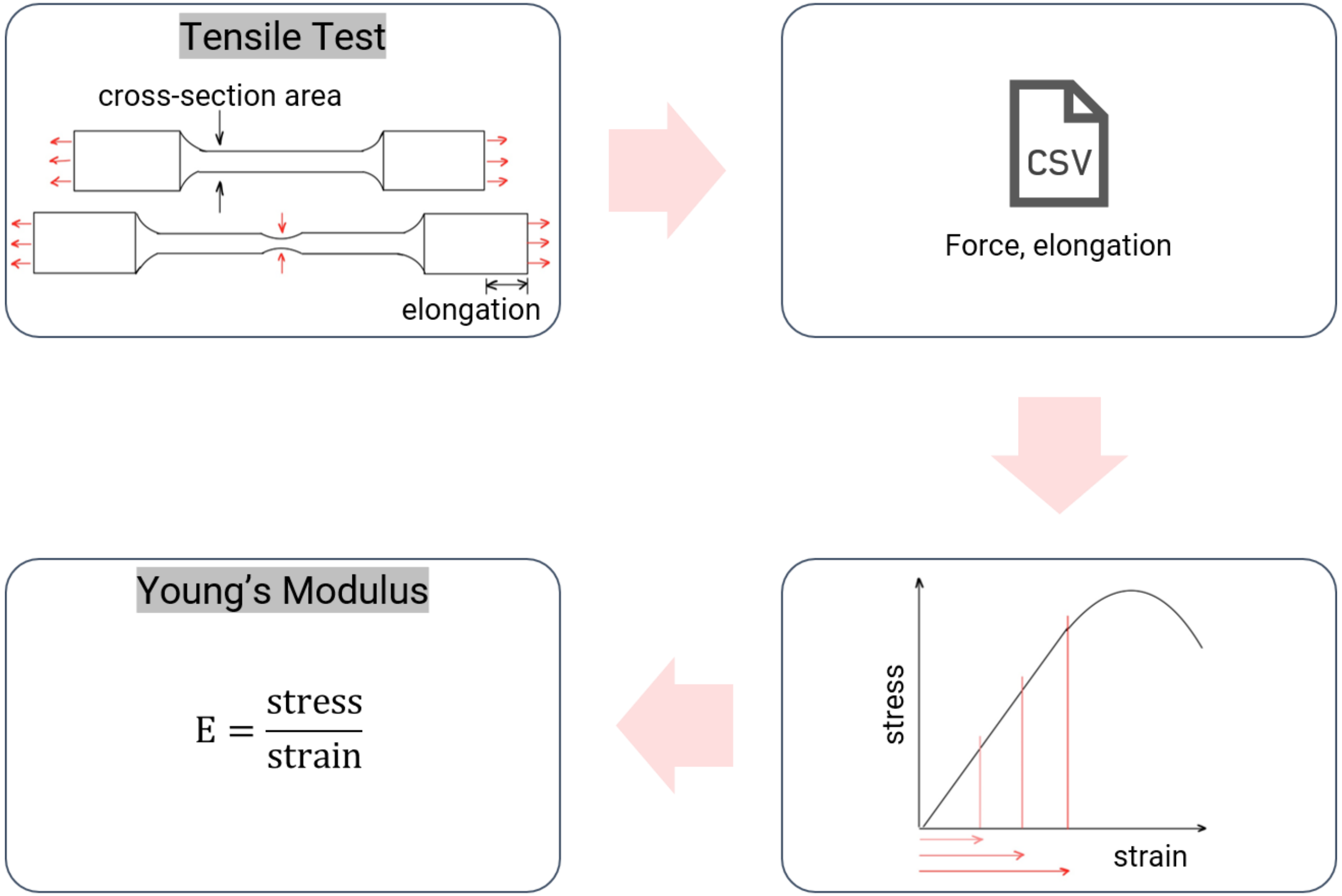
Janssen *et al*, Comput. Mater. Sci. 163 (2019)
https://doi.org/10.1016/j.commatsci.2018.07.043

Now: Tutorial

Calculation of Young's modulus

from experimental data

https://LMY.DE/WORKFLOW



Funded by

Deutsche Forschungsgemeinschaft

German Research Foundation

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524