

Nick Garabedian¹ (*Nikolay.Garabedian@kit.edu*),
 Iliia Bagov¹, Karlheinz Weber¹, Christian Greiner¹, Frederic Bock²,
 Benjamin Klusemann², Catriona Eschke³, Florian Wieland³, Martin Held⁴

¹ Karlsruhe Institute of Technology, Institute for Applied Materials, ² Helmholtz-Zentrum Hereon, Institute of Materials Mechanics,
³ Helmholtz-Zentrum Hereon, Institute of Metallic Biomaterials, ⁴ Helmholtz-Zentrum Hereon, Institute of Membrane Research

Motivation

- FAIR vocabularies are a prerequisite for FAIR data publication
- Tools for collaborative composition of such vocabularies are missing
- Vocabularies are an intuitive first step for any lab undergoing digitalization



OntoFAIRCook

- FAIR vocabularies are a stepping stone towards FAIR ontologies
- Vocabularies are converted semi-supervisedly to FAIR data ontologies
- Ontologies allow for scalable data-driven machine learning

VocPopuli

Vocabulary IDs

Global ID: set at birth and never changes
Local ID: a hash of the terms' Global IDs; changes when a term is added or deleted
Version ID: a hash of all terms' Local IDs; changes when a term is modified

Term IDs

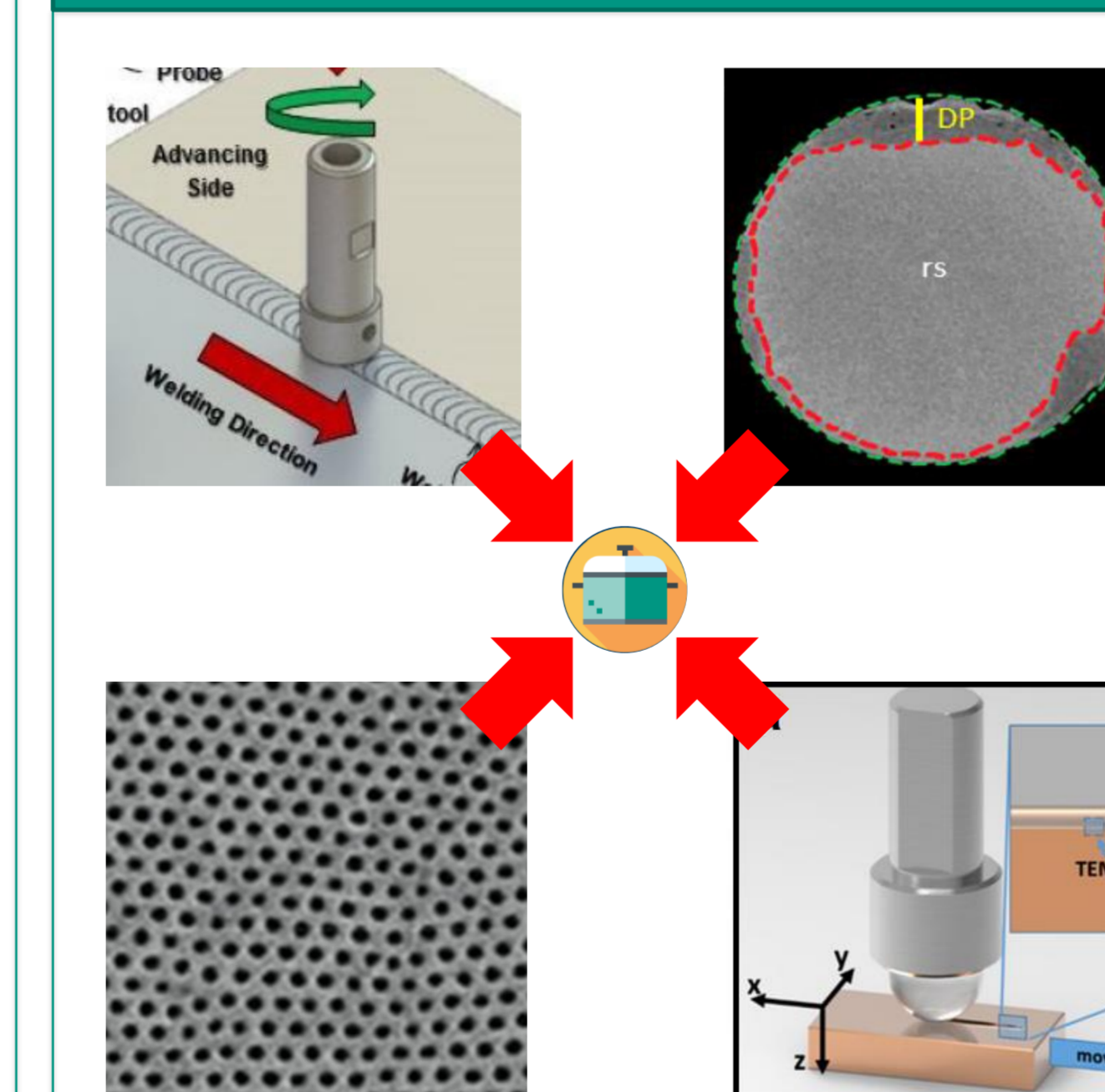
Global ID: set at birth and never changes; it represents a self-contained concept
Local ID: denotes the version of the term; changes for each version of the term while being discussed

- GitLab login is used to associate user contributions with term versions
- Each term receives a git branch for editing
- GitLab Issues are used to collect commentary on terms' definitions
- Terms are voted on and *approved* or *not approved*
- PROV Data Model (PROV-DM) and SKOS are used to track and structure the vocabularies

Electronic Lab Notebooks

- Terms used within ELNs are linked to the IDs in the related vocabulary and ontology for an unambiguous definition
- An export of the defined schemes can be used to automatically create a structured form in the ELNs
- The output from the developed tools will be exemplarily integrated into the ELNs Herbie and Kadi4Mat

Involvement Approach



- 4 labs with different research foci in materials science will test the solutions consecutively
- Each lab constructs a vocabulary
- → iterative improvements to software; 4 separate fresh tests
- Controlled vocabularies must be agreed upon by a critical number of scientists
- The size of deciding groups depends on the generality of the terms: general terms need wider agreement
- User-friendliness is a priority