

# Fostering the democratization of research data by using the Annotated Research Context (ARC) as practical implementation

Benedikt Venn<sup>1</sup>, Kevin Schneider<sup>1</sup>, Kevin Frey<sup>1</sup>, H. Lukas Weil<sup>1</sup>, Johannes Werner<sup>2</sup>, Fabian Wannemacher<sup>2</sup>, Thomas Zajac<sup>2</sup>, Dirk von Suchodoletz<sup>3</sup>, Björn Usadel<sup>4</sup>, Jens Krüger<sup>2</sup>, Christoph Garth<sup>5</sup>, Timo Mühlhaus<sup>1</sup>

<sup>1</sup>Computational Systems Biology, University of Kaiserslautern, Paul-Ehrlich-Straße 23, 67663 Kaiserslautern, Germany  
<sup>2</sup>High Performance and Cloud Computing Group (IT Center), University of Tübingen, Wächterstraße 76, 72074 Tübingen, Germany  
<sup>3</sup>eScience (IT Center), University of Freiburg, Hermann-Herder-Str. 10, 79104 Freiburg i. Br., Germany  
<sup>4</sup>IBG-4 Bioinformatics, Forschungszentrum Jülich, 52425 Jülich, Germany  
<sup>5</sup>Scientific Visualization Lab, University of Kaiserslautern, 67663 Kaiserslautern, Germany

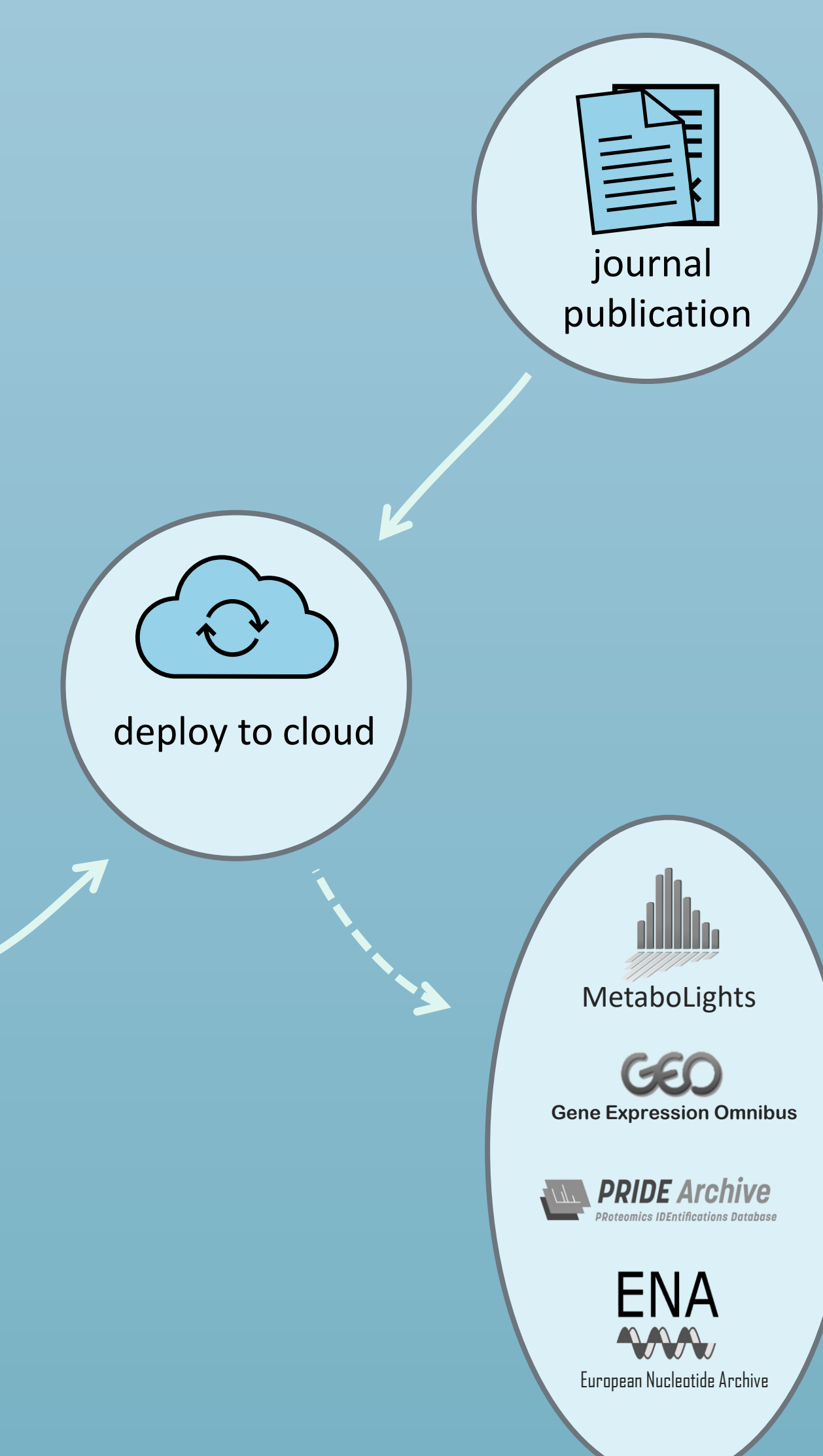
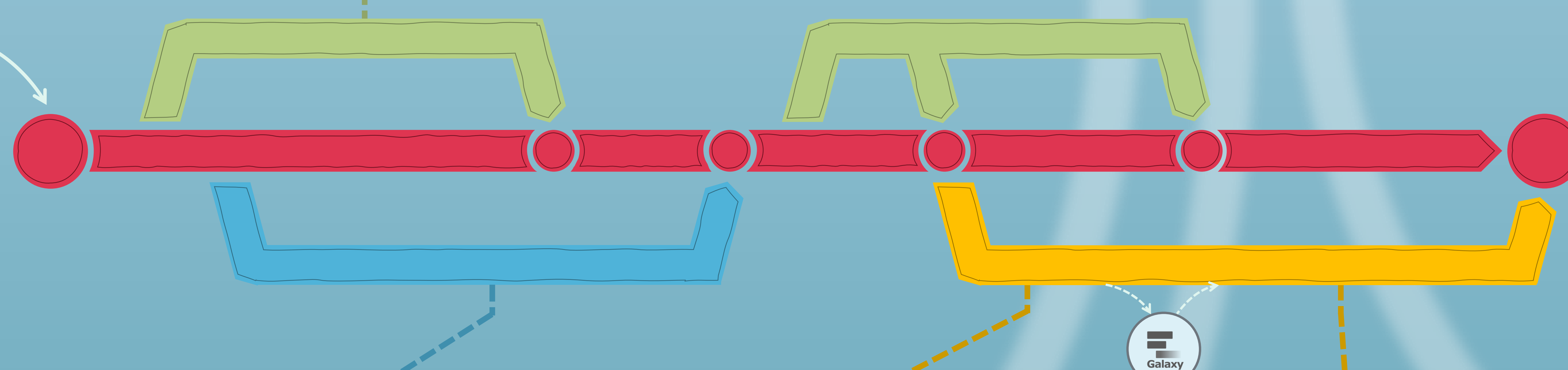
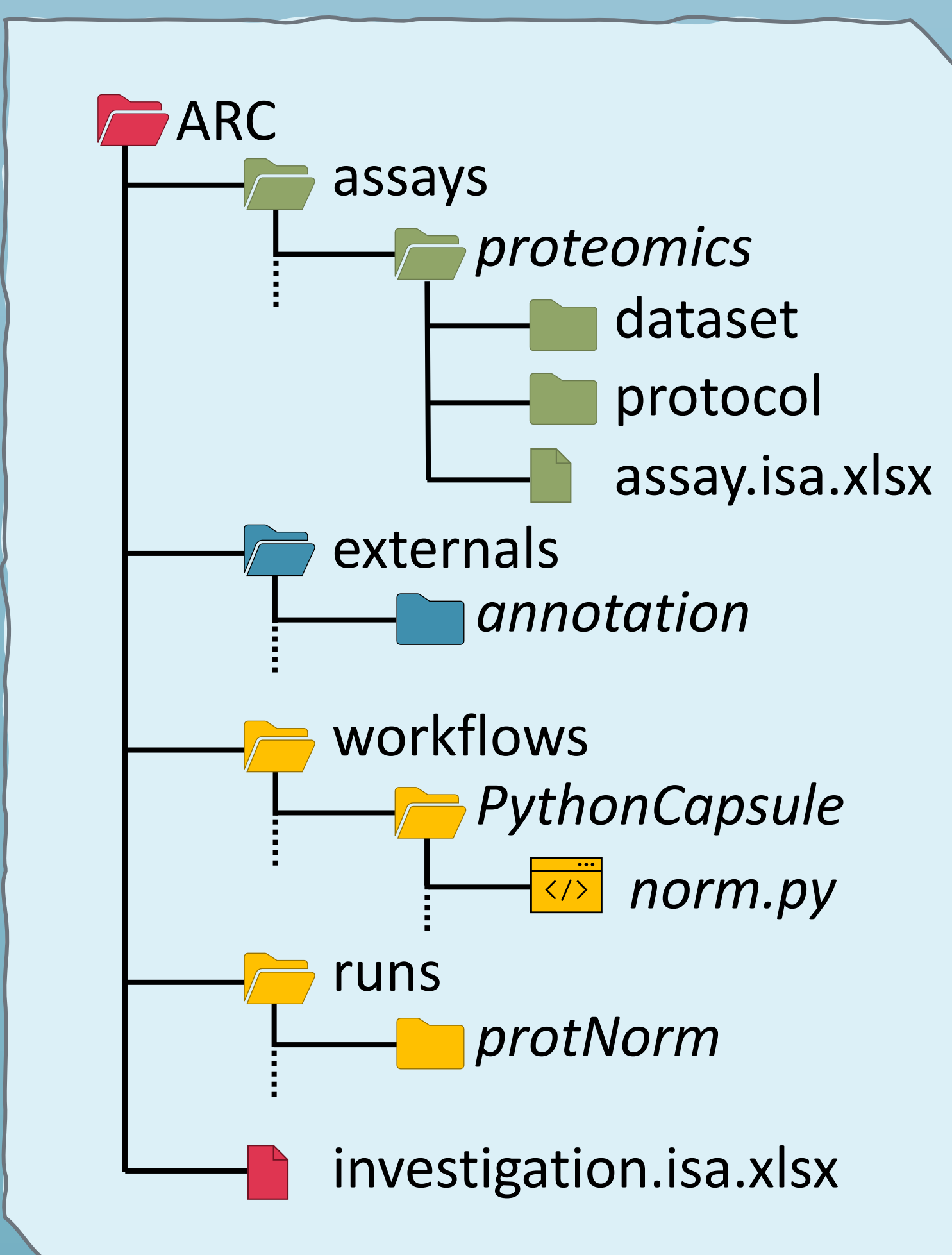
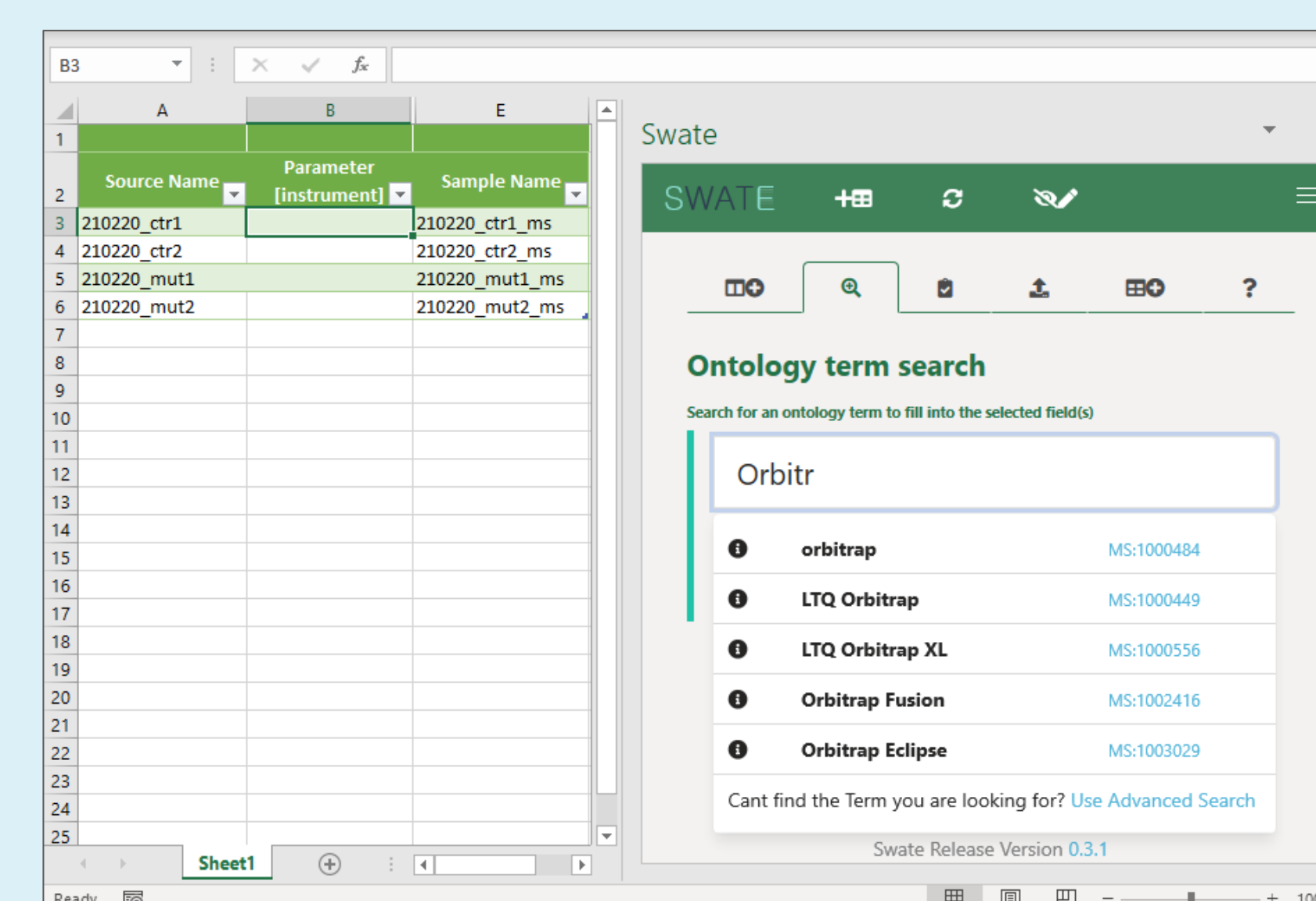
## Annotated Research Context (ARC)

- » The ARC aims at capturing the entire research cycle and defines a framework for organization, sharing, versioning, reuse, and evolution of research projects.
- » It respects FAIR and linked open principles.
- » The usage of common spreadsheet software (e.g. Excel) with low system requirements minimizes the friction to the user.
- » It provides full metadata annotation based on the ISA model.
- » The modular structure facilitates a granular reuse and development of raw data or processing workflows.
- » Guided creation and metadata annotation leading to an interface between research communities and public repositories.



## Assays

- » All raw data must be stored as immutable data within assays.
- » Metadata annotation is realized by user friendly isa.xlsx sheets.
- » Excel plugin SWATE assists in assay file annotation.



## Metadata annotation toolbox

- » The data annotation is based on ISA model hierarchy and implemented as isa.xlsx files.
  - » investigation -> study -> assay
- » ArcCommander generates folder structure and registers global ARC metadata.
- » SWATE ensures proper assay annotation.

## Externals

- » Auxiliary, experiment independent resources
- » Sequence annotation mapping files that cannot be referenced due to missing release cycles.
  - » e.g. FastA or gff

## Workflows

- » Application code, scripts, executables
- » Workflows include their own containerized running environment to ensure independent executability and reproducibility.
- » Workflow reuse can be realized for example by export to and import of Galaxy resources.

## Runs

- » Any result derived from workflows applied to assays/runs/externals are stored as run result.
- » A Common Workflow Language file links workflows to their respective input data to ensure easy usage and reproducibility.