RENGA - 連歌

Data Science Collaboration Platform

Rok Roškar - Swiss Data Science Center (ETH + EPFL)
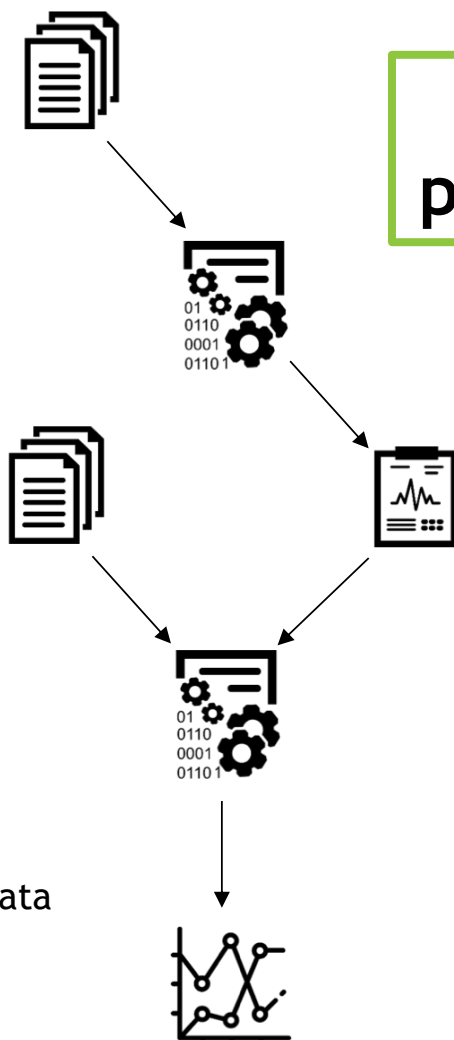
The Goals of Renga

<u>Renga</u>:

**the Swiss Data Science Center
collaborative data science platform**

- Provide (data) scientists means to create *reproducible* data science

- Facilitate the *sharing* and *reuse* of research artifacts

- Enable *rapid discovery* of relevant data and methods

- Foster a *collaborative environment* for rapid prototyping of novel approaches

- Allow federated access across institutions, giving each the freedom to impose its own access controls over resources
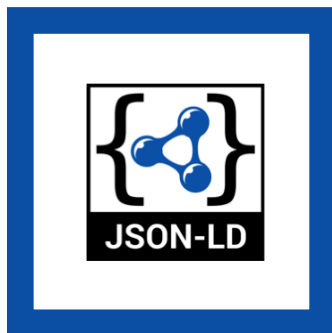
SDSC

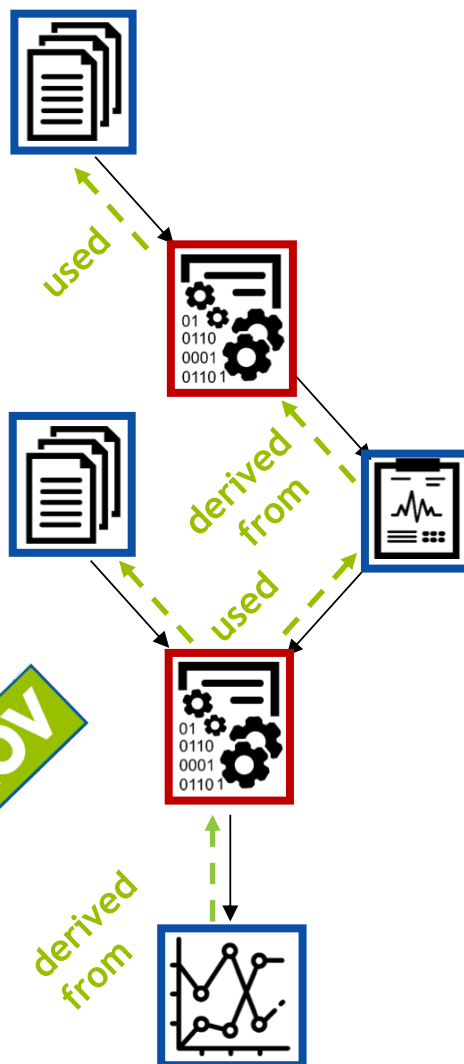**Capturing and recording data provenance is the core of Renga**

1. Inputs and outputs of analysis steps are recorded into a knowledge graph

2. Steps can be repeated or integrated into more complex workflows

3. Provenance of all data products always accessible via simple tools

4. Version control built-in for data, code, and workflows

Raw data

Processed data

Software

Results

SDSC

1. Tracking research provenance

- JSON-LD for better interpretability and interoperatibility

- Extended Dublincore ontology

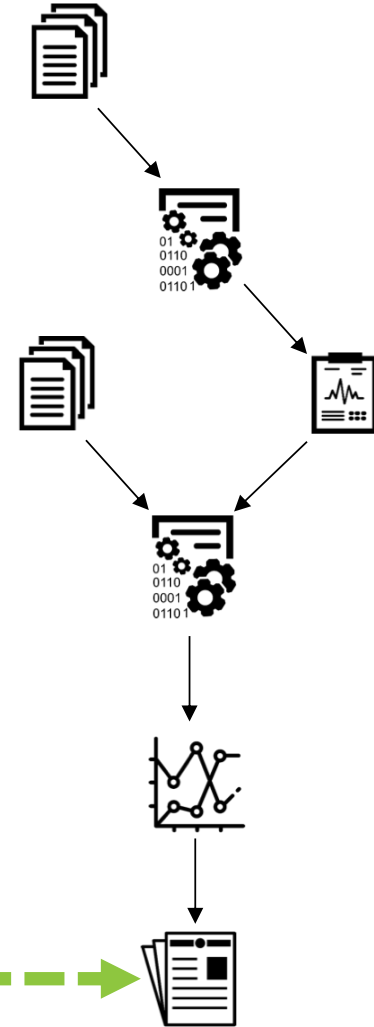- Provenance graphs based on PROV-O W3C recommendation

- CWL for representing all computational steps

- Capture individual steps from user input

- Tools for constructing workflows from basic pieces

- Rely on container technologies to ensure reproducibility

Graph-based search…

Search for a publication, obtain a full view of the relevant *data lineage*
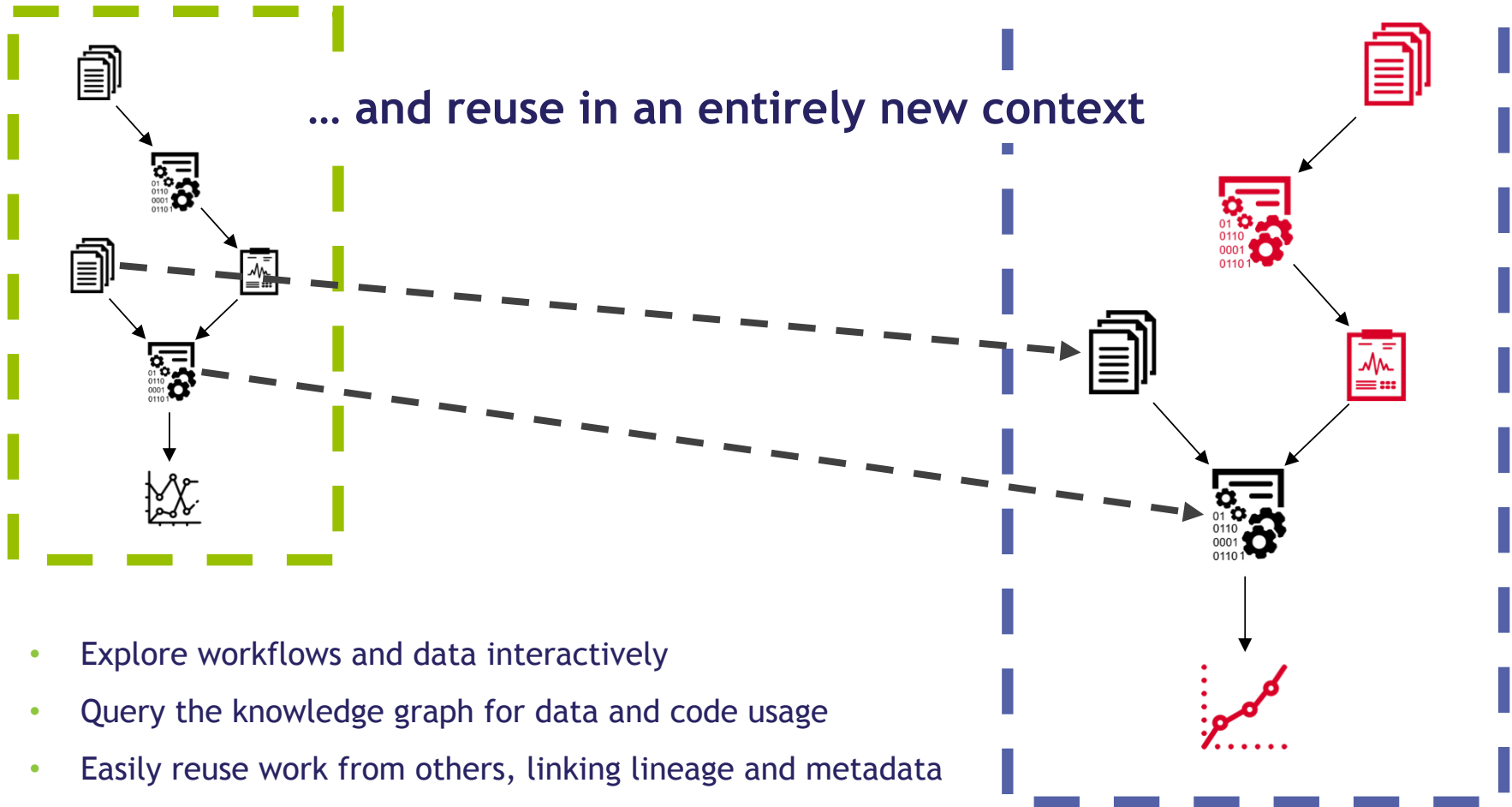
SDSC

3. Reusability

Search for relevant data or algorithms…

… and reuse in an entirely new context

- Explore workflows and data interactively
- Query the knowledge graph for data and code usage
- Easily reuse work from others, linking lineage and metadata
- Identify popular datasets and algorithms across the platform; audit

Current status

- Under <u>very active</u> development
- First released in September 2017, but currently undergoing complete redesign
- Expected new release end of April 2018

Find us on

http://github.com/SwissDataScienceCenter/renga

SDSC