



# DemocraSci & Renku

Collaborative research to foster interdisciplinarity

# The current state of the project

- Streamlined the implementation of the code
- Integration of the project into Renku
  - Specially tricky due to the number of files and their total size
  - Renku is still under development
  - So far WP1, preprocessing of the data
- A productive collaboration thanks to frequent meetings

# The current state of the project

- Streamlined the implementation of the code
- Integration of the project into Renku
  - Specially tricky due to the number of files and their total size
  - Renku is still under development
  - So far WP1, preprocessing of the data
- A productive collaboration thanks to frequent meetings

We aim at improving the collaboration and carry out interdisciplinary research by making use of **Renku**

## 5 Questions we need to ask

1. How did I compute this results?
2. How does new data change this result?
3. How did you compute your result? Can I use your data to reproduce it? With your code? On your infrastructure?
4. Has anyone ever used an <XYZ-algorithm> on this data? How?
5. Who is using my data? And my algorithms? Why are they not citing me?

# 5 Questions we need to ask

1. How did I compute this results?

2. How can I ensure the results are reproducible?  
If you can answer these questions confidently:

• You can collaborate easily

3. How can I ensure the results are reproducible?  
• Your team is efficient

• You participate in Open Science/Open Data

4. How can I ensure the results are reproducible?  
• You are properly acknowledged (and you acknowledge others!)

• You can be held accountable

5. Will my results be cited?  
• Your results are **trustworthy**

citing me?

# Goals of Renku

1. Provide the means to create **reproducible** (data) science

2. Facilitate the **sharing** and **reuse** of research artifacts

3. Foster a **collaborative environment** for interactive prototyping

4. Enable the **discovery** of relevant data and methods

5. Allow **federated access** across institutions giving each the freedom to impose its own access controls over resources

# Goals of Renku

1. Provide the means to create **reproducible** (data) science

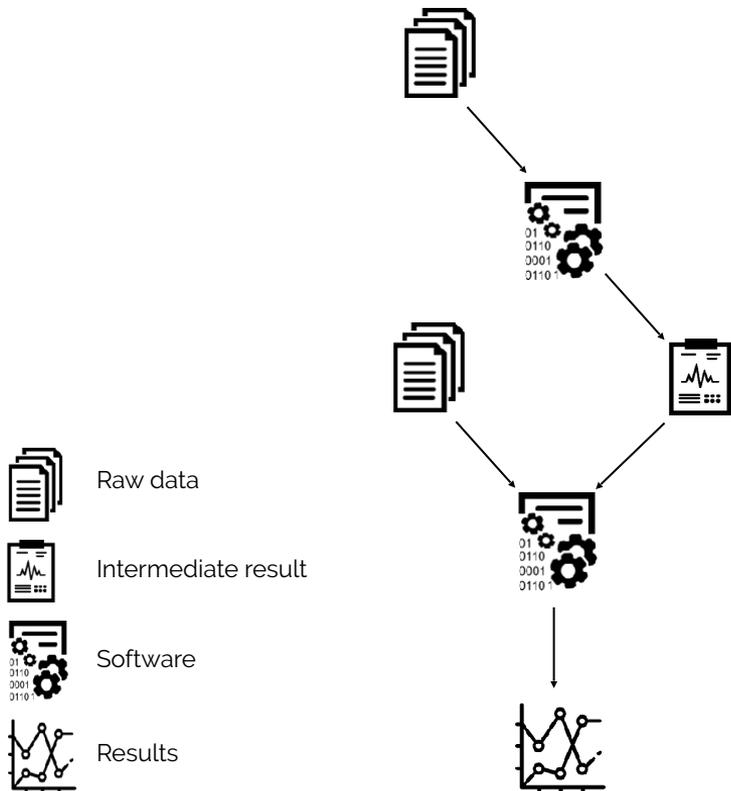
2. Facilitate the **sharing** and **reuse** of research artifacts

3. Foster a **collaborative environment** for interactive prototyping

4. Enable the **discovery** of relevant data and methods

5. Allow **federated access** across institutions giving each the freedom to impose its own access controls over resources

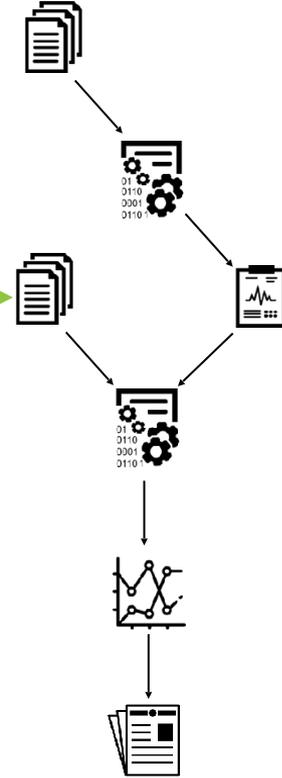
# Lineage of a simple analysis



- Inputs and outputs of analysis steps are recorded into a **knowledge graph** *as the work is being done*
- Steps can be **repeated** or integrated into more complex **workflows**
- **Provenance** of all data products is always accessible via simple tools
- **Version control** is built-in for data, code, and workflows

# Discover and understand the analysis process

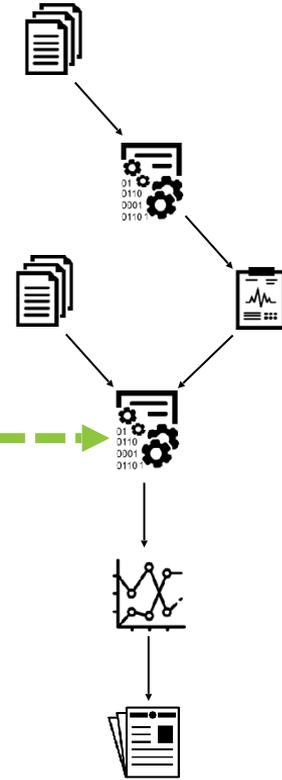
Graph-based search...



Ex: search for **data** and explore the tools to efficiently generate results

# Discover and understand the analysis process

Graph-based search...



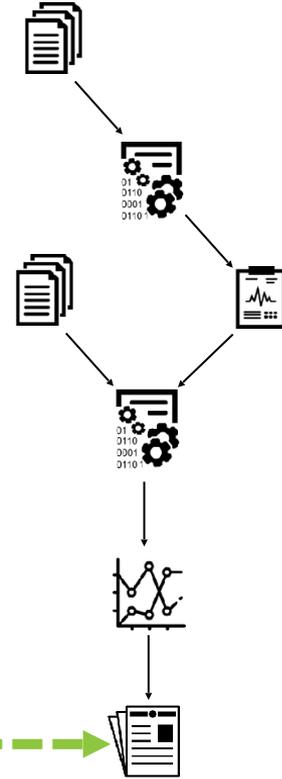
Ex: search for an **algorithm**  
and see its applications

# Discover and understand the analysis process

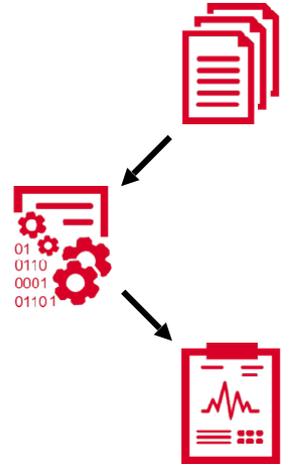
Graph-based search...



Ex: search for a **publication**, obtain a full view of how the results were obtained



# Reuse and repeat



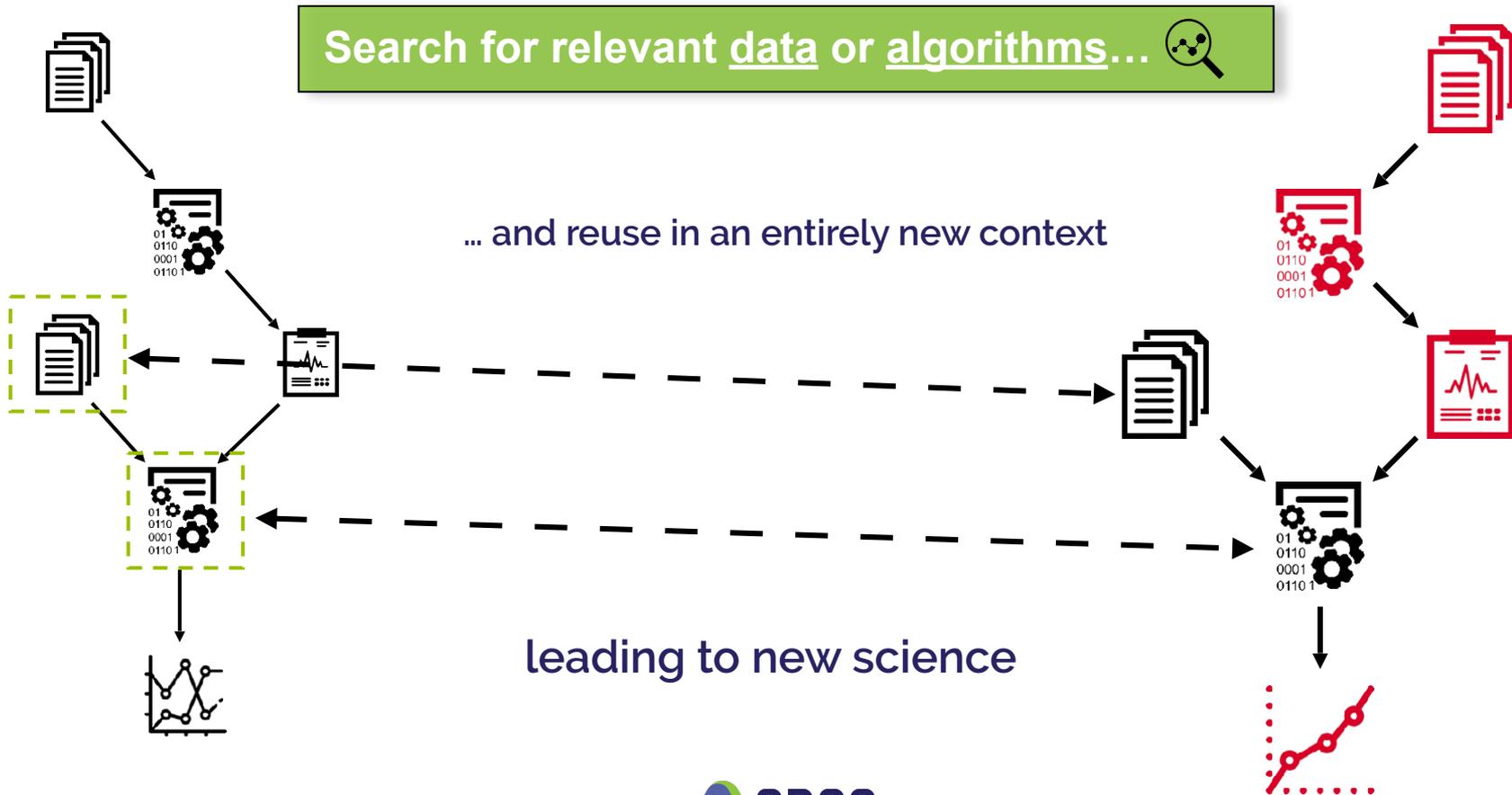
# Reuse and repeat

Search for relevant data or algorithms...

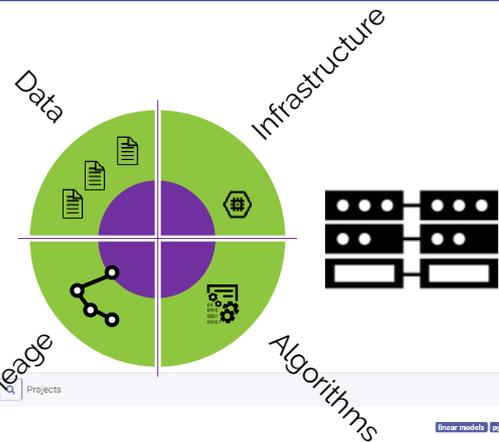
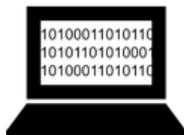


... and reuse in an entirely new context

leading to new science



# Create locally, share easily (and globally)



```
Requirement already satisfied: htai51lib<1.0b1, >=1.0b2, <=1.0b3, <=1.0b4, <=1.0b5, <=1.0b6, <=1.0b7, <=1.0b8, >=0.9999999pre in /opt/conda/lib/python3.6/site-packages (from sbatch-mountain=>ipyter-weather-ch==0.1.0)
Requirement already satisfied: parso<0.1.1, in /opt/conda/lib/python3.6/site-packages (from jedi<=0.10.0->ipython<4.0.0->ipykernel=>ipyter-weather-ch==0.1.0)
Requirement already satisfied: ptyprocess<=0.5 in /opt/conda/lib/python3.6/site-packages (from pexpect; sys_platform == 'win32'-->ipython<4.0.0->ipykernel=>ipyter-weather-ch==0.1.0)
Requirement already satisfied: webencodings in /opt/conda/lib/python3.6/site-packages (from htai51lib<1.0b1, >=1.0b2, <=1.0b3, <=1.0b4, <=1.0b5, <=1.0b6, <=1.0b7, <=1.0b8, >=0.9999999pre->sbatch-mountain=>ipyter-weather-ch==0.1.0)
Installing collected packages: seaborn, patsy, statsmodels, weather-ch
Running setup.py install for seaborn: finished with status 'done'
Running setup.py develop for weather-ch
Successfully installed patsy-0.5.0 seaborn-0.8.1 statsmodels-0.8.3 weather-ch
Removing intermediate container 790ad8076us
---- 39778609507
---- Running in 33ae022ee88
Removing intermediate container 33ae022ee88
---- 2e6105dced9
Successfully built 2e6105dced9
Successfully tagged ranga:bullid5001/rok.weather-ch/review-master-ghenv1:2546598c71e7b936328e639f6f6812b021ef
rok@P-master: ~$ docker exec -it ranga-sb001 bash
root@pre-sentation(weather-ch):# ranga status
On branch master
Your files were generated from the latest commits
rok@P-master: ~$ ranga doc
~/project/pre-sentation(weather-ch) ranga log
git/      .gitignore      .ipynb_checkpoints/  ranga/      .ranga.lock      Dockerfile      README.md
data/    requirements.txt  src/
rok@P-master: ~$ ranga doc
~/project/pre-sentation(weather-ch) ranga log data/zh/
konzo_no_SMA.txt  metadata.yml  standardize2.csv
rok@P-master: ~$ ranga doc
~/project/pre-sentation(weather-ch) ranga log data/zh/standardize2.csv
* 0f1c1f91 csta/zh/standardize2.csv
* 0f1c1f91 ranga/workflow/931115acc2054668b361c79f6fffffc_python_cw1
* 5e6105d csta/zh/hwop_no_SMA.txt
```



Search RENGA Projects

## weather-ch

An investigation into weather trends in Zürich, Switzerland.  
Updated 1 hour ago.

Overview Kus Files Settings

**open** Analyze data  
Perform analysis of the data to understand the weather trends. Updated 5 hours ago.

**complete** Preprocess data  
Convert values to deviation from monthly mean. Updated 1 hour ago.

**complete** Data reader  
Data reader. Updated 1 hour ago.

**complete** Implement code to read the data. Updated 1 hour ago.

### Preprocess data

Convert values to deviation from monthly mean  
Updated 1 hour ago.  
For the pre-processing results, see this notebook

**Rok Roikar** Updated 1 hour ago

```
import pandas as pd
import numpy as np
import scipy
import weather_ch

import matplotlib as mpl
from matplotlib import cm
import matplotlib.pyplot as plt
import seaborn as sns
from IPython.display import display, HTML
sns.set()

df = weather_ch.read_standardized('.../data/zh/standardized.csv')
df.head()
```

Local / Command-line interface

Cloud / web front-end



# Platform components

- **renku-ui**: react.js + bootstrap
- **renku-python**: python CLI and API client
- **renku-notebooks**: customized JupyterHub
- **renku-gateway**: API Gateway
- **renku-graph**: Knowledge Graph (JSON-LD + SPARQL + GraphQL)
- Repository management + CI (vanilla GitLab)
- Authentication (Keycloak)
- Runtime (Docker + Kubernetes)
- Workflow execution (CWL+???)



docker



# Demo of WP1 Data preprocessing & Project documentation



# Collaborative research for WP1

- Data scientists: development of new methods, and improvement of existing ones
- Domain experts: assessment of
  - Current results
  - Usability of the code
  - Additional preprocessing tasks needed
  - New functionalities and visualization tools
  - ...
- All these tracked through issues, in the project Wiki, etc.

# WP2 in Renku: even more interesting

- WP2, focused on the topic modeling and semantic analysis of the corpus, will offer a wider and more interesting range of possibilities:
  - Naming and discovery of topics
  - Easily testing other methods (dynamic LDA, etc.)
  - Results that will allow a more comprehensive understanding of the corpus
  - Exploring the use of embeddings
  - Semantic analysis of speeches and political positions
  - ...

# Conclusions

- The project is already advanced enough to enable the collaboration between domain experts and data scientist
  - Renku is the best platform to carry this out
  - All this together with an exhaustive documentation
- Also WP2 is advanced enough to be integrated into Renku
  - More possibilities for collaboration
- We have already taken some steps on the development of the knowledge graph
- Now the use of Renku may boost the enhancement of the code



Thank you so much!

**ETH** zürich

