



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 be confident that these results are correct?
If you can answer these questions confidently:

3. How can I be confident that my results are reproducible?
• You can collaborate easily

4. How can I be confident that my results are reliable?
• Your team is efficient

5. How can I be confident that my results are valid?
• You participate in Open Science/Open Data

6. How can I be confident that my results are trustworthy?
• You are properly acknowledged (and you acknowledge others!)

7. How can I be confident that my results are useful?
• You can be held accountable

8. How can I be confident that my results are impactful?
• Your results are **trustworthy**

9. How can I be confident that my results are relevant?
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

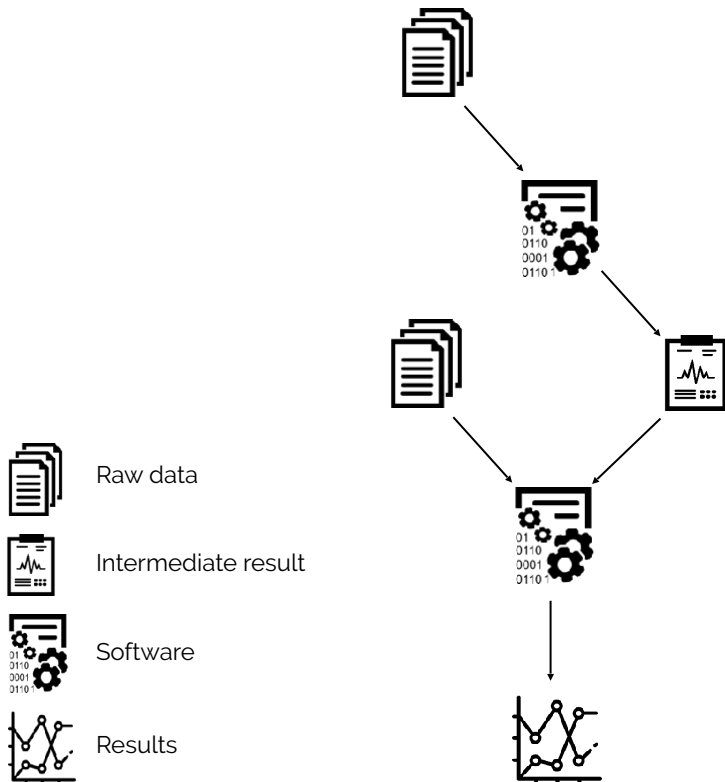
4.

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

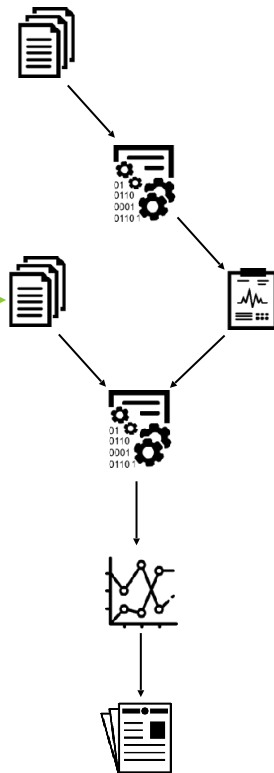
Lineage of a simple analysis



- 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

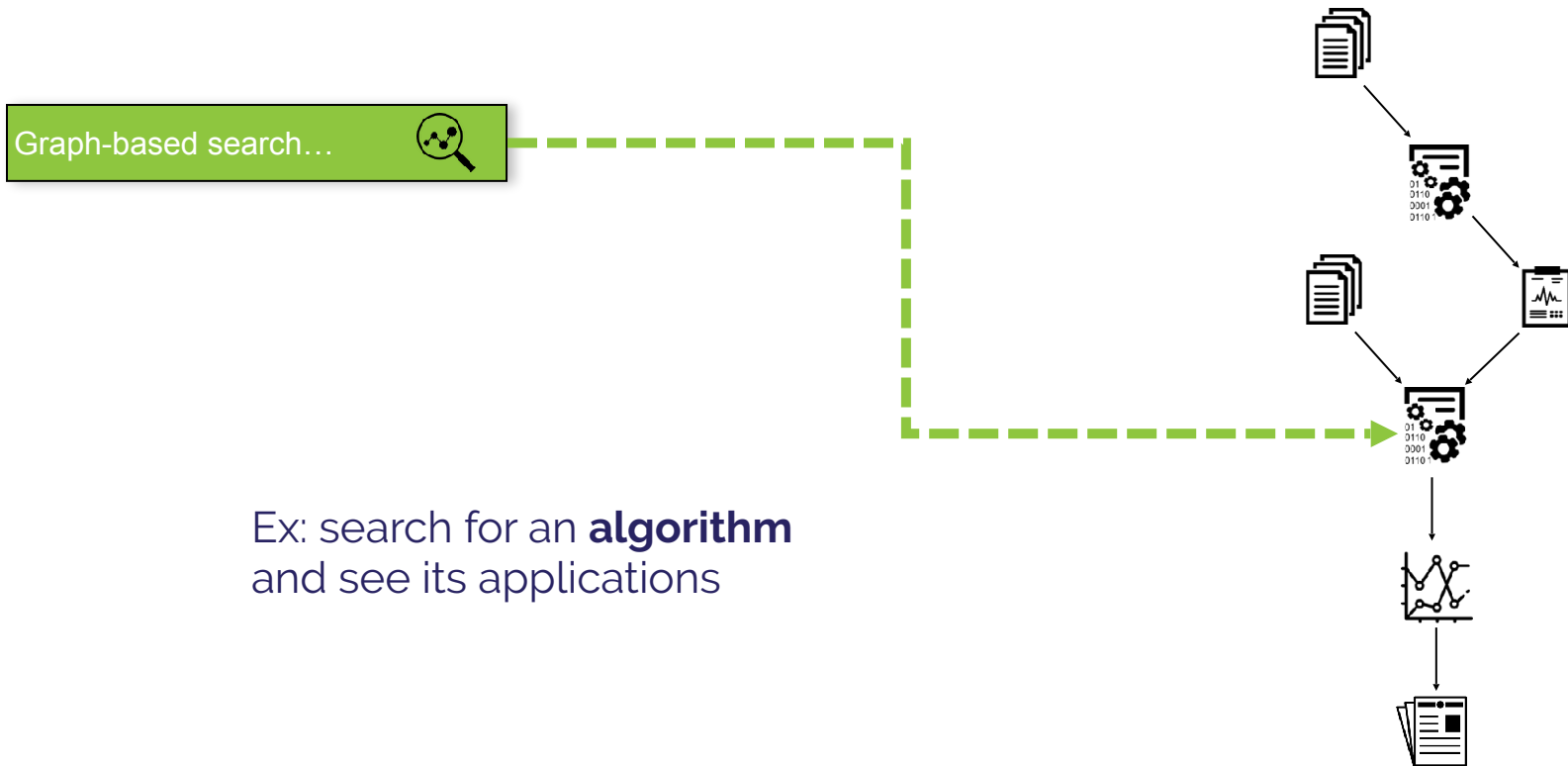
Discover and understand the analysis process

Graph-based search...



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

Discover and understand the analysis process

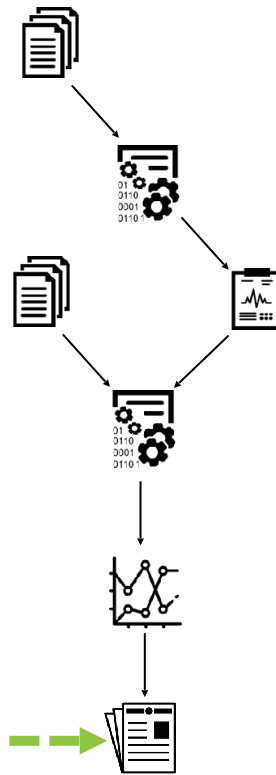


Discover and understand the analysis process

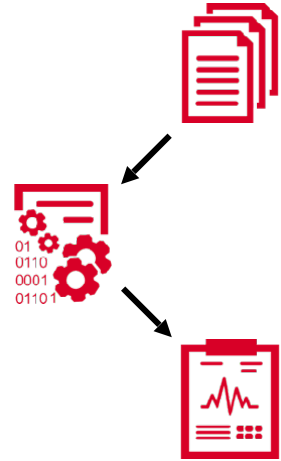
Graph-based search...



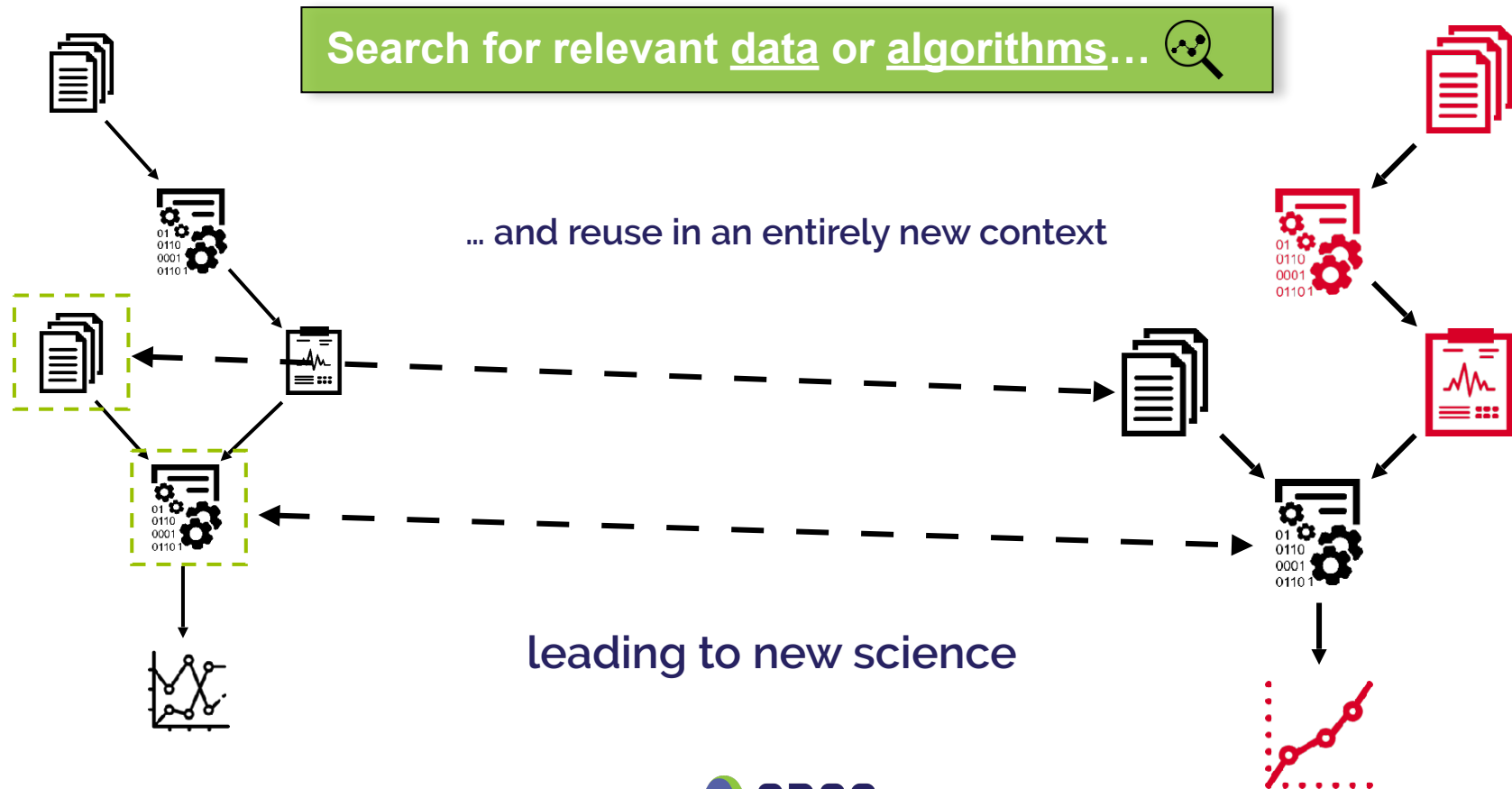
Ex: search for a **publication**, obtain a full view of how the results were obtained



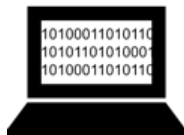
Reuse and repeat



Reuse and repeat



Create locally, share easily (and globally)



```
Requirement already satisfied: htai5lib<1.0b1, >=1.0b2, <=1.0b3, <=1.0b4, <=1.0b5, <=1.0b6, <=1.0b7, <=1.0b8, >=0.99999999 in /opt/conda/lib/python3.6/site-packages (from blaach->nbconvert->ipython-weather-ch==0.1.0)
Requirement already satisfied: parson<0.1.1 in /opt/conda/lib/python3.6/site-packages (from jedi==0.10.0->ipython==4.0.0->ipykernel->ipython-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->ipython-weather-ch==0.1.0)
Requirement already satisfied: webencodings in /opt/conda/lib/python3.6/site-packages (from htai5lib<1.0b1, >=1.0b2, <=1.0b3, <=1.0b4, <=1.0b5, <=1.0b6, <=1.0b7, <=1.0b8, >=0.99999999->blaach->nbconvert->ipython-weather-ch==0.1.0)
Installing collected packages: seaborn, paty, statmodels, weather-ch
Running setup.py install for seaborn: finished with status 'done'
Running setup.py install for statmodels: finished with status 'done'
Successfully installed paty-0.5.0 seaborn-0.8.1 statmodels-0.8.0 weather-ch
Running intermediate container: 796ad8676d5
---- 397786c9587
Step 7/7 : USER root
---- Running in 33ae022ee88
Running intermediate container: 33ae022ee88
Successfully built 2e1032dced9
Successfully tagged gitlab_renga:build-5801/rok-weather-ch/review-master-ghvov12344583c71a7b0363328eb395fa0120b21ef
root@P-master:~/weather-ch# docker-compose -f docker-compose.yml up --build
on branch master
All files were generated from the latest inputs
root@P-master:~/weather-ch# docker-compose -f docker-compose.yml up --build
git/
  .gitignore
  .ipynb_checkpoints
  .renga/
  .renga.lock
  Dockerfile
  README.ad
  data/
  requirements.txt
  src/
root@P-master:~/weather-ch# docker-compose -f docker-compose.yml up --build
conda_no_SMA.txt
meta_data.yml
standardized.csv
root@P-master:~/weather-ch# docker-compose -f docker-compose.yml up --build
renga log data/zh/
renga log data/zh/standardized.csv
+ 60cf190 cdata/zh/standardized.csv
+ 60cf190 renga/workflow/931136c0204668361c7f6ff5ffc_pythn_cx1
+ 5a5b3d6 cdata/zh/hung_no_SMA.txt
root@P-master:~/weather-ch# docker-compose -f docker-compose.yml up --build
```

Local / Command-line interface



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

Preprocess data
Convert values to deviation from monthly mean
Rok Roilkar updated 1 hour ago.
For the pre-processing results, see this notebook

```
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()
```

Cloud / web front-end



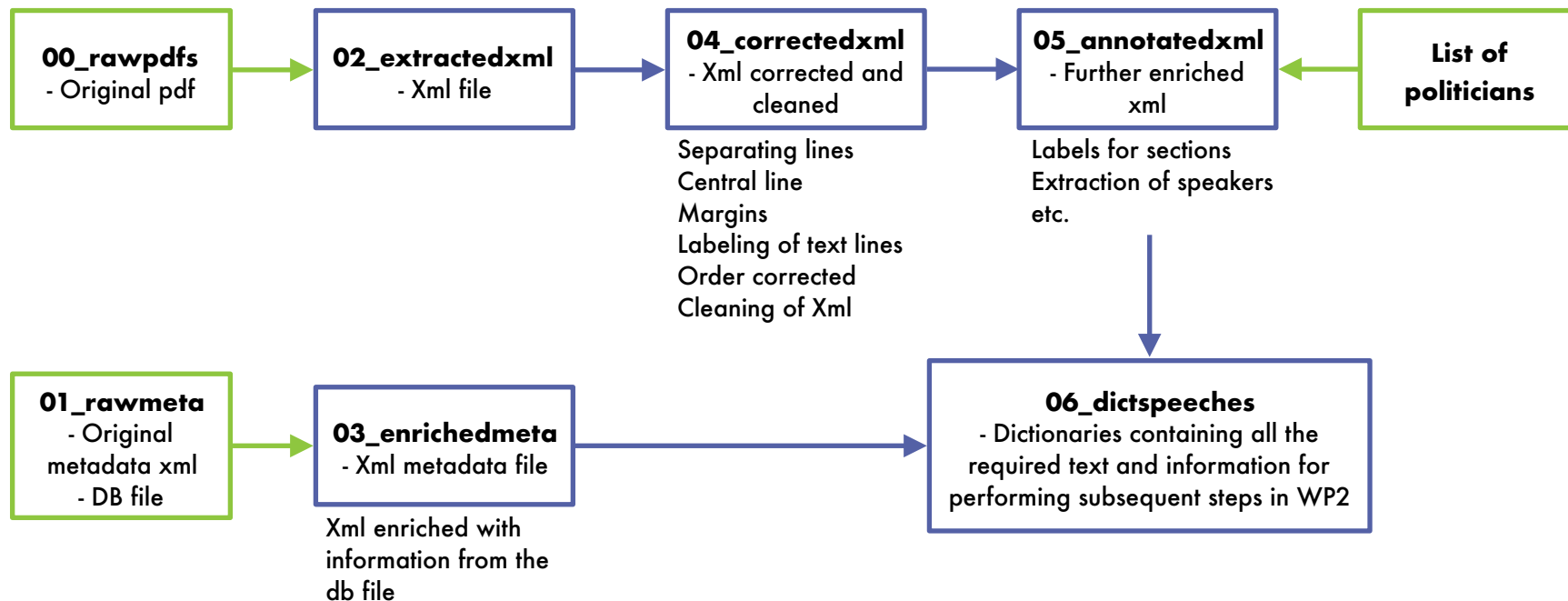
Platform components

- **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+???)



Demo of WP1 Data preprocessing & Project documentation

Workflow for WP1



All are .tar.gz files

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!