RePlay-DH

Realization of a **pla**tform and accommodating services for research data management in the research communit**y** **D**igital **H**umanities

www.ub.uni-stuttgart.de/replay

Markus Gärtner[1*], Uli Hahn[2], Sibylle Hermann[3*]**

markus.gaertner@ims.uni-stuttgart.de, uli.hahn@uni-ulm.de, sibylle.hermann@ub.uni-stuttgart.de

[1]Institute for Natural Language Processing [2]Communication and Information Centre [3]University Library [*]University of Stuttgart [**]Ulm University

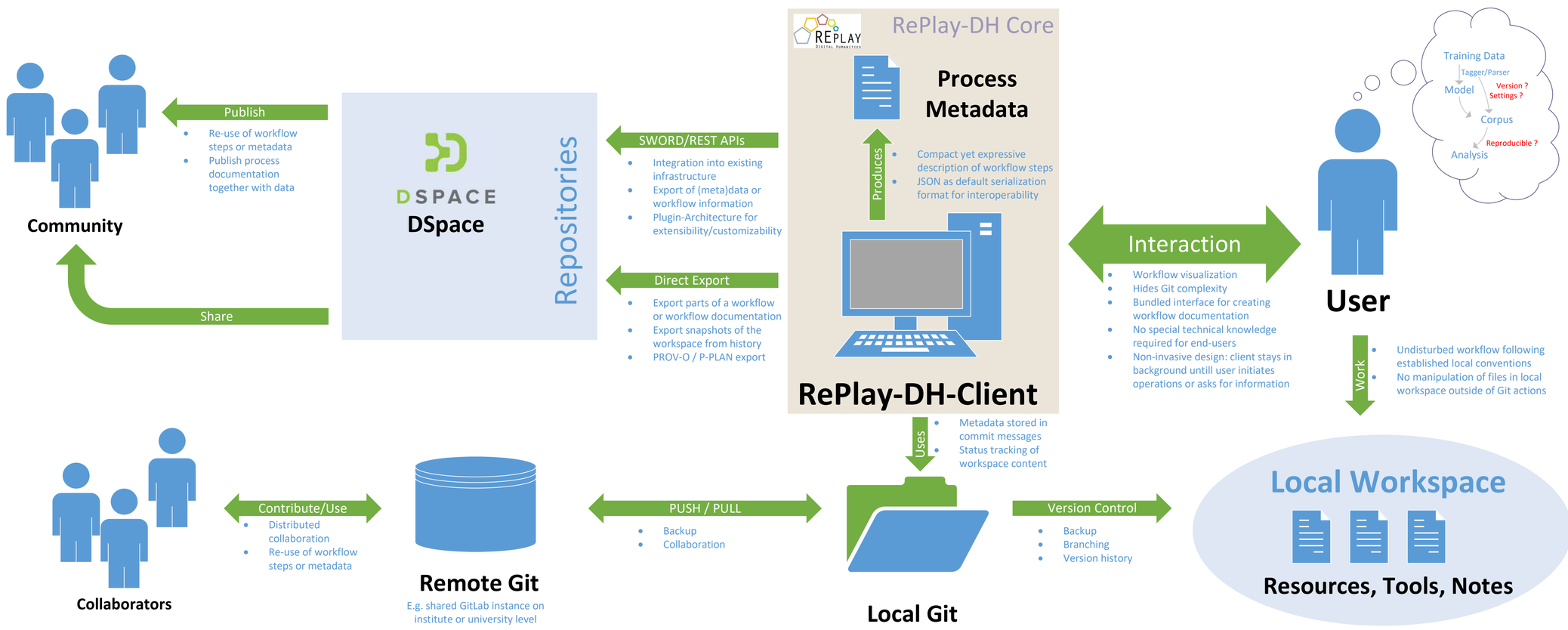
Preserving Workflow Reproducibility: The RePlay-DH Client as a Tool for Process Documentation

Motivation

- Documentation of complex research processes is often **lacking**.
- If done at all, it usually is performed **after** the process.
- Sustainable process documentation requires lots of additional **effort**.
- Existing version control solutions or workflow management systems are typically **not suitable** for processes in the fields of **CL** and **DH**.

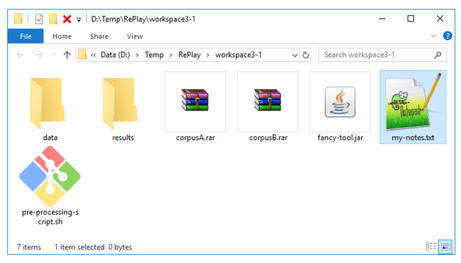
Goals

- Assist in creating documentation already during an active research workflow.
- Provide a simple metadata schema for workflow documentation.
- Minimize effort required from researchers for clean process documentation.
- **Idea:** Build on Git as foundation for workflow tracking, but hide the complexity by channeling all the documentation work through a single graphical application.

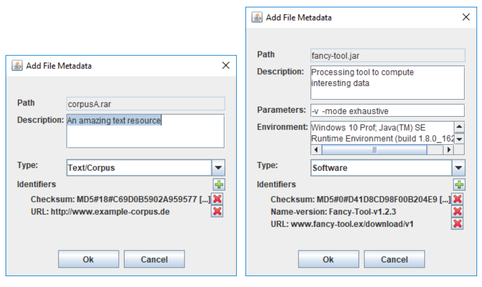


Documentation Process

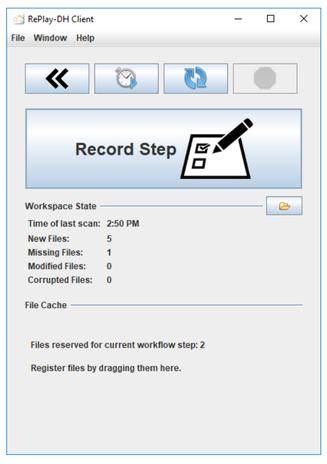
Assign a local workspace to be monitored by the RePlay-DH Client for changes:

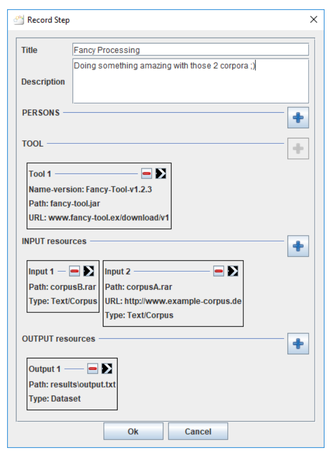


Keep familiar workflow patterns, no need to change how research workflows are performed. At any time files can be described early on and cached by the client:

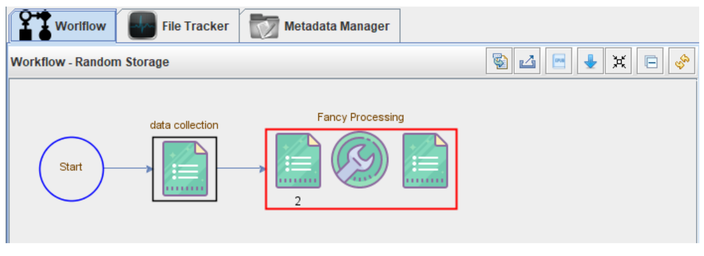


Compact client window provides overview of the current workspace state and allows to record completed workflow steps:





The resulting workflow graph is visualized and can be interacted with to navigate the version history of the workspace or to export (meta)data:



Integration

Independence: No external infrastructure or additional third-party software other than Java required for the basic client. Workflow documentation and local object metadata management in a simple schema following Dublin Core [1] available.

Extensibility: Plugin-architecture to incorporate the client into existing institutional infrastructure such as repositories for metadata or publishing.

Publishing: Interfaces with repository software DSpace [2] (`http://www.dspace.org`) for initiating the publication process of data with a persistent identifier (DOI).

Availability

Client available as open-source on GitHub:

`https://github.com/RePlay-DH`

[1] Andy Powell, Mikael Nilsson, Ambjörn Naeve, and Pete Johnston. Dublin core metadata initiative - abstract model, 2005. White Paper.

[2] MacKenzie Smith. Dspace: An institutional repository from the mit libraries and hewlett packard laboratories. In Maristella Agosti and Costantino Thanos, editors, *Research and Advanced Technology for Digital Libraries*, volume 2458 of *Lecture Notes in Computer Science*, pages 543–549. Springer Berlin Heidelberg, 2002.

University of Stuttgart
Germany

University Stuttgart:
University Library (UB),
Institute for Natural Language Processing (IMS),
Clarin-D Center, SFB-732 Project INF

ulm university universität ulm

Ulm University:
Communication and Information Centre (kiz)
Institute for Organisation and Management of Information Systems (OMI)

This work was funded by:
Ministry for Science, Research and the Arts in Baden-Württemberg (MWK)
German Research Foundation (DFG) via the SFB 732, project INF

Baden-Württemberg
MINISTERIUM FÜR WISSENSCHAFT, FORSCHUNG UND KUNST

DFG