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:

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] 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
