

LSDMA AHM Darmstadt 2016

DLCL Neuroscience - Report André Giesler



DLCL Neuroscience Team and Tasks



- André Giesler
 - Myriam Czekala (part-time)
-
- Provenance Tracking
 - General data management support

Provenance Tracking in Neuroscience



Data ingest, e.g. scanning brain slices by microscope:
Contributors? Which hardware setup? By which procedure?



Data Transfers from temporary lab cache to final storage with preprocessing:
Preprocessing description? Reference to persistent identifiers?

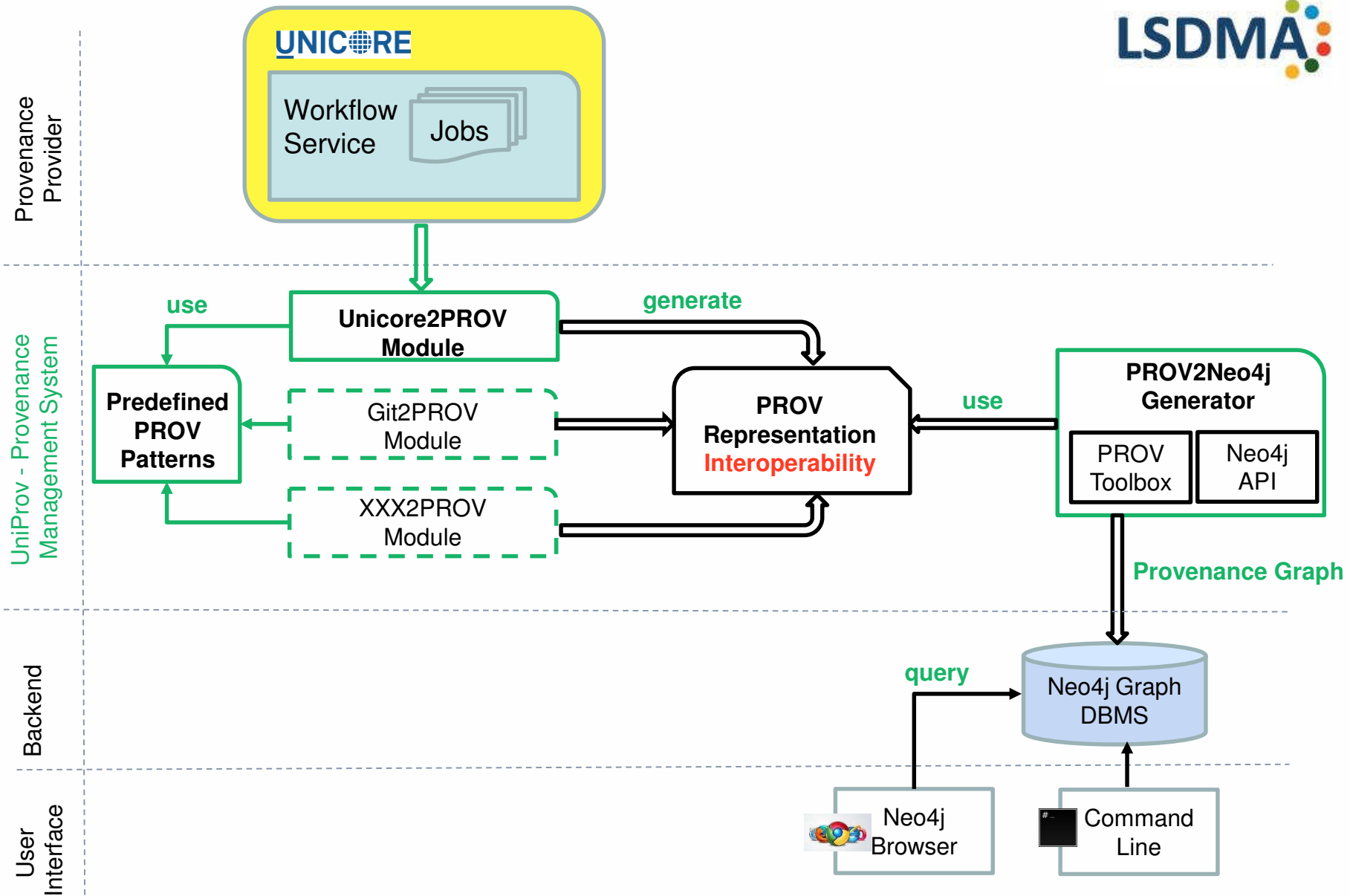
Query backtracking simulation XY submitted 3 years ago: **Which compiler, software versions, libs, target systems? Reproducing results.**

Workflow-centric

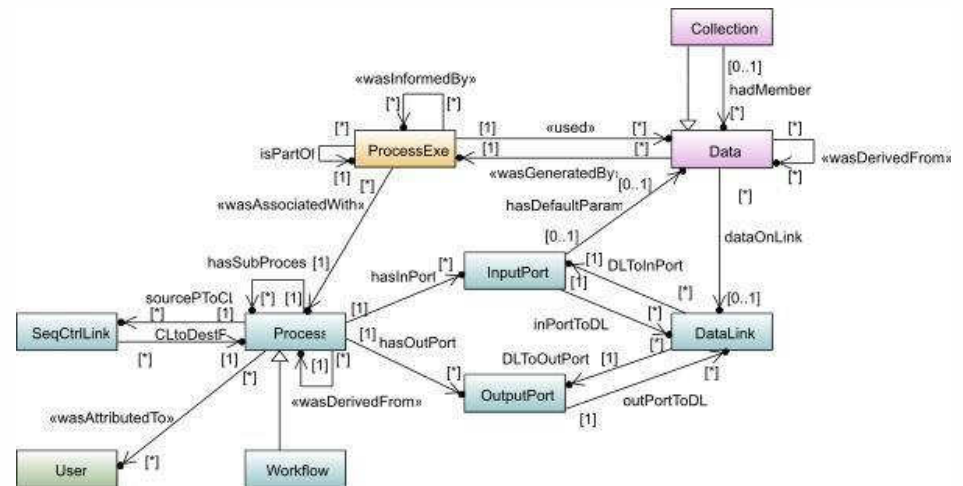
Workflow based image processing: **Which program A used in workflow B? Find all invocations that used parameter XY=100 ? Adding specific annotations...**

Provenance scenarios

- Designing a modular architecture
 - Tracking provenance information from various sources
 - Concentrating first on UNICORE based scientific workflows
- Applying the emerging **W3C-PROV** standard and PROV extensions for mapping workflow models (ProvONE)
 - Prospective provenance (static workflow model)
 - Retrospective provenance (executions)
- Mapping UNICORE job and workflow structures to PROV/ProvONE
- Store data in suitable backend (**Neo4j Graph Db**)
 - Efficient querying with Cypher
 - Visualize provenance graphs in Neo4j web interface
 - APIs and drivers for many common programming languages



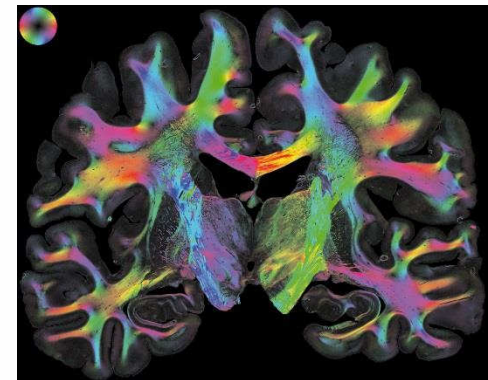
- Paper submitted:
 - A. Giesler, M. Czekala, B. Hagemeyer, R. Grunzke. UniProv: A flexible Provenance Tracking System for UNICORE, Theory and Practice of Provenance 2016 (TaPP'16), McLean



- In general: How to capture the big data life cycle?
- Data ingest, pre- and post-processing, data transfers, script-based workflows
- Working group with neuroscientists
 - Identifying provenance graphs outside of UNICORE workflows
 - Finding tools supporting provenance tracking (Git2PROV, Git-Annex, ...)
 - Mapping to W3C-PROV standard
- HBP Integration
 - SP5: Provenance – Evaluation of UniProv in HBP Provenance Management System (both PROV based)
 - SP7: Integrating PLI workflow in the HBP infrastructure

■ 3D-PLI Workflow

- Technique to map nerve fibers and fiber tracts in postmortem brains at the micrometer scale
- Using UNICORE workflows to automate the image-processing chain
- S. Köhnen, A. Huynh, G. Tabbi, A. Lührs, A Giesler, B. Hagemeyer, K. Amunts, Th. Lippert, M. Axer "Automation and Parallelization of a 3D Polarized Light Imaging Workflow" at OHBM 2016 Geneva
<http://www.humanbrainmapping.org/i4a/pages/index.cfm?pageID=3662>
(submitted)



Thank you for your attention!