

# Workflowautomatisierung in PubFlow



**Peer Brauer**  
**[peer.brauer@informatik.uni-kiel.de](mailto:peer.brauer@informatik.uni-kiel.de)**  
**Bremen, 18.04.2013**

# Agenda

1. Das Projekt
2. Workflows in PubFlow
3. Ausblick & Visionen

Abschnitt 1

# **Das Projekt**

COMPUTER SCIENCE

## Beyond the Data Deluge

Gordon Bell,<sup>1</sup> Tony Hey,<sup>1</sup> Alex Szalay<sup>2</sup>

Since at least Newton's laws of motion in the 17th century, scientists have recognized experimental and theoretical science as the basic research paradigms for understanding nature. In recent decades, computer simulations have become an essential third paradigm: a standard tool for scientists to explore domains that are inaccessible to theory and experiment, such as the evolution of the universe, car passenger crash testing, and predicting climate change. As simulations and experiments yield ever more data, a fourth paradigm is emerging, consisting of the techniques and technologies needed to perform data-intensive science (*1*). For example, new types of computer clusters are emerging that are optimized for data movement and analysis rather than computing, while in astronomy and other sciences, integrated data systems allow data analysis and storage on site instead of requiring download of large amounts of data.

Today, some areas of science are facing hundred- to thousandfold increases in data volumes from satellites, telescopes, high-

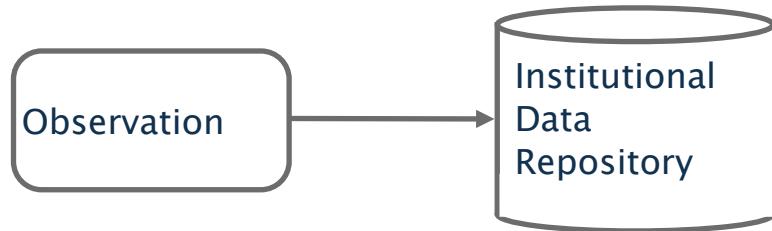
The demands of data-intensive science represent a challenge for diverse scientific communities.



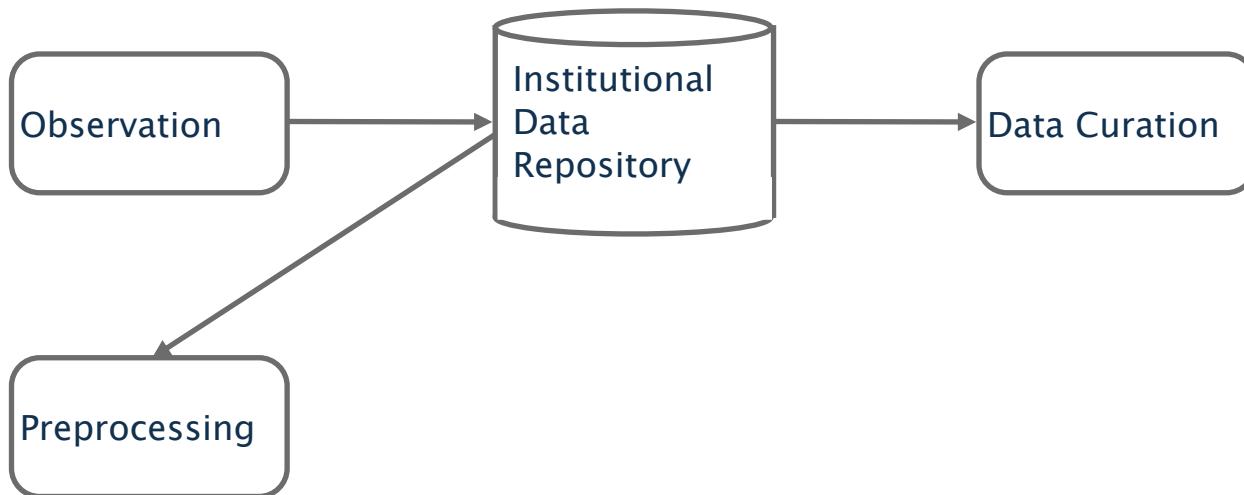
**Moon and Pleiades from the VO.** Astronomy has been one of the first disciplines to embrace data-intensive science with the Virtual Observatory (VO), enabling highly efficient access to data and analysis tools at a centralized site. The image shows the Pleiades star cluster from the Digitized Sky Survey combined with an image of the Moon centered within the *Wide World Telescope*.

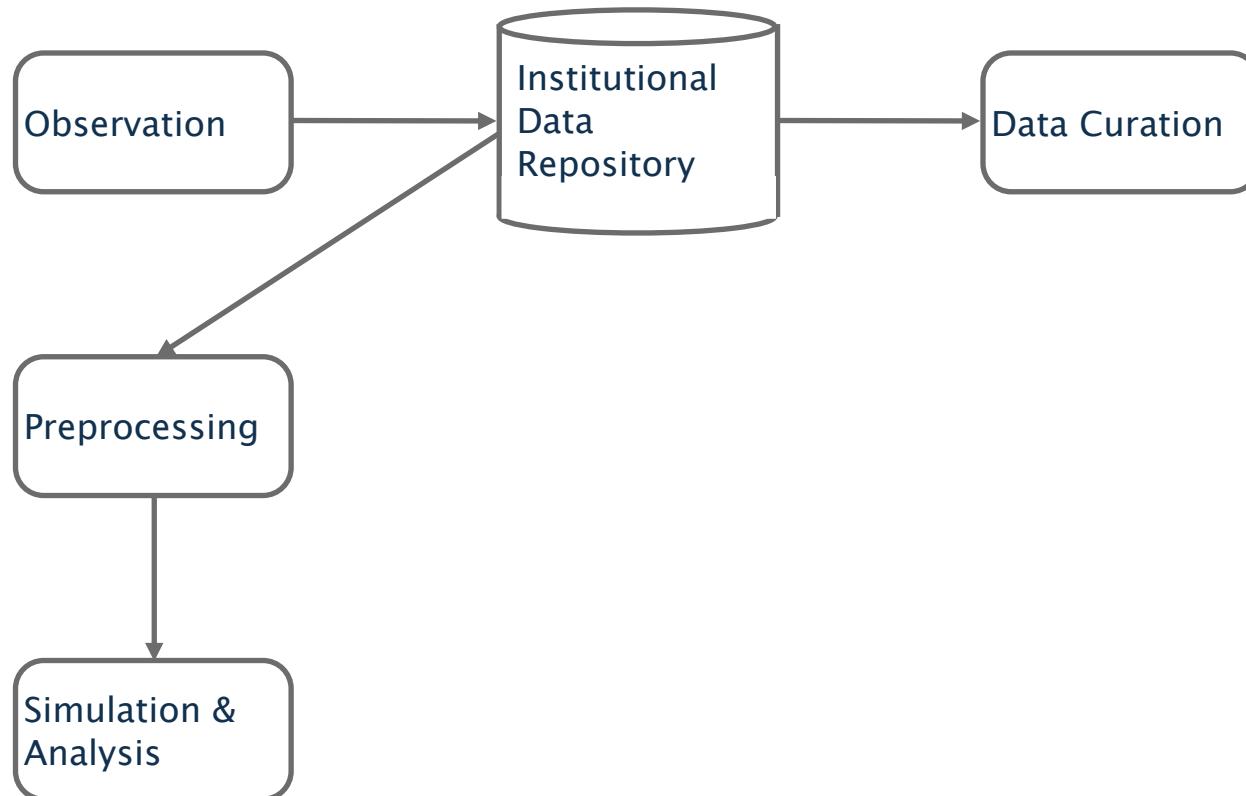
> Das Projekt

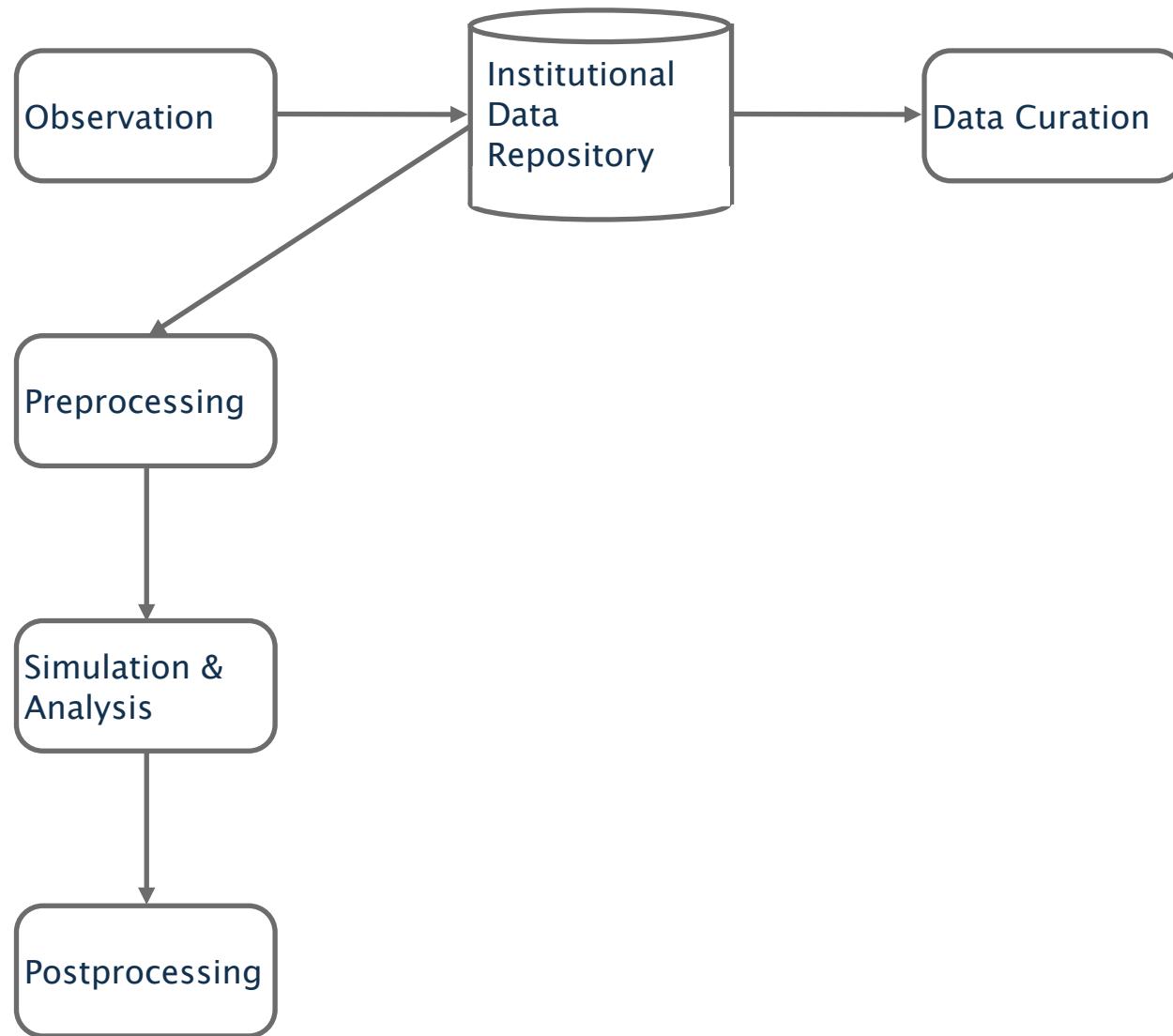
Observation

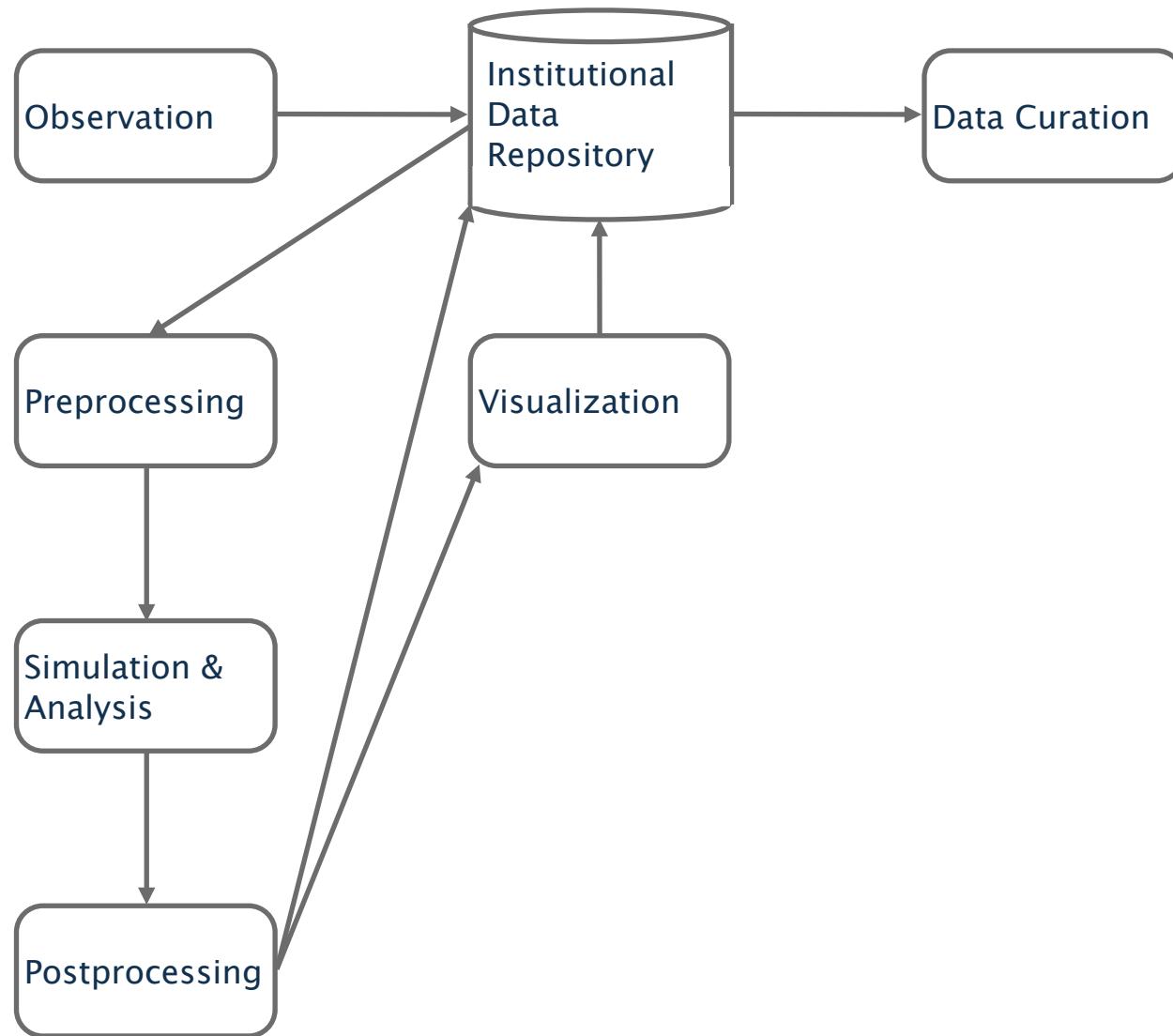


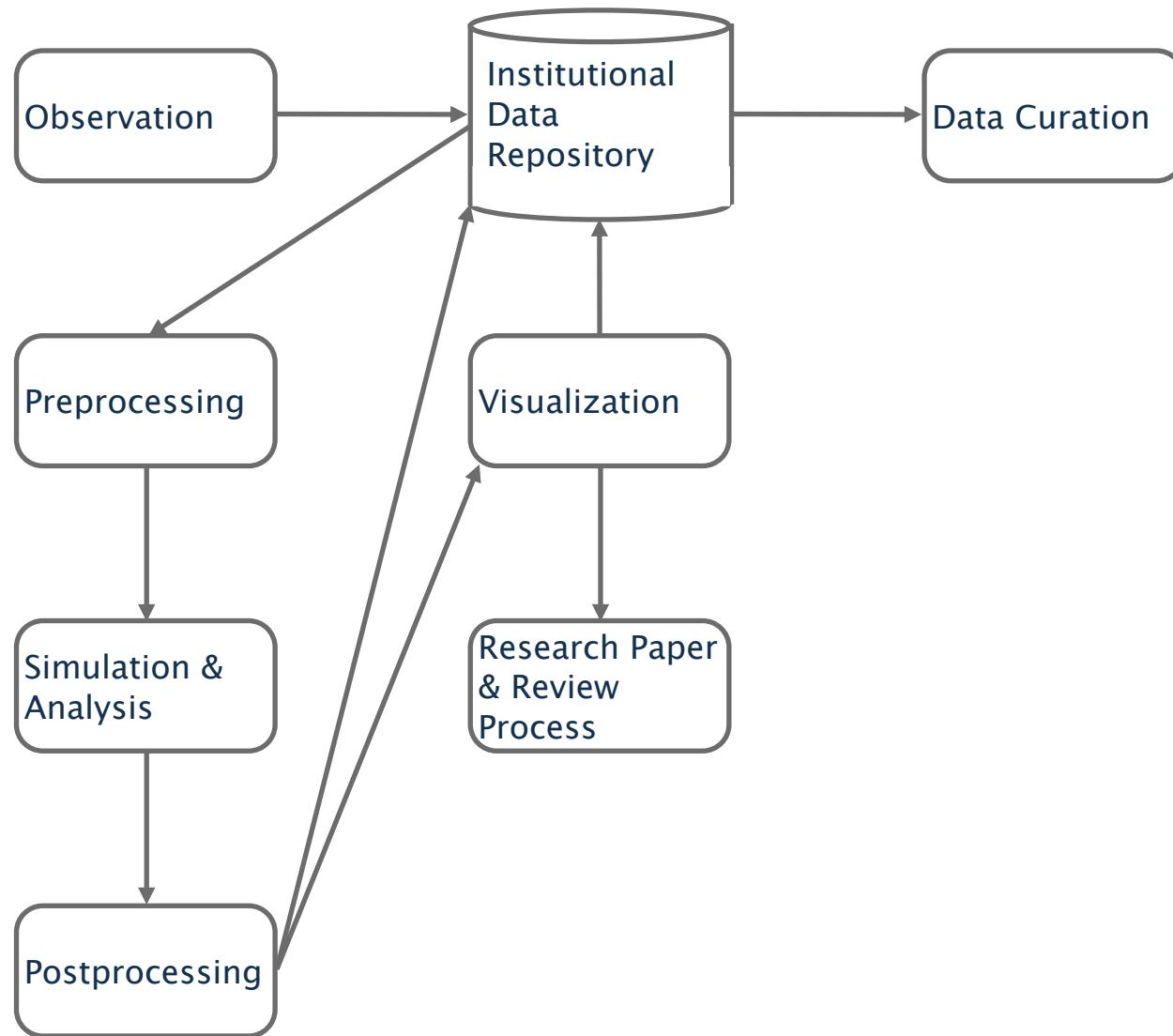


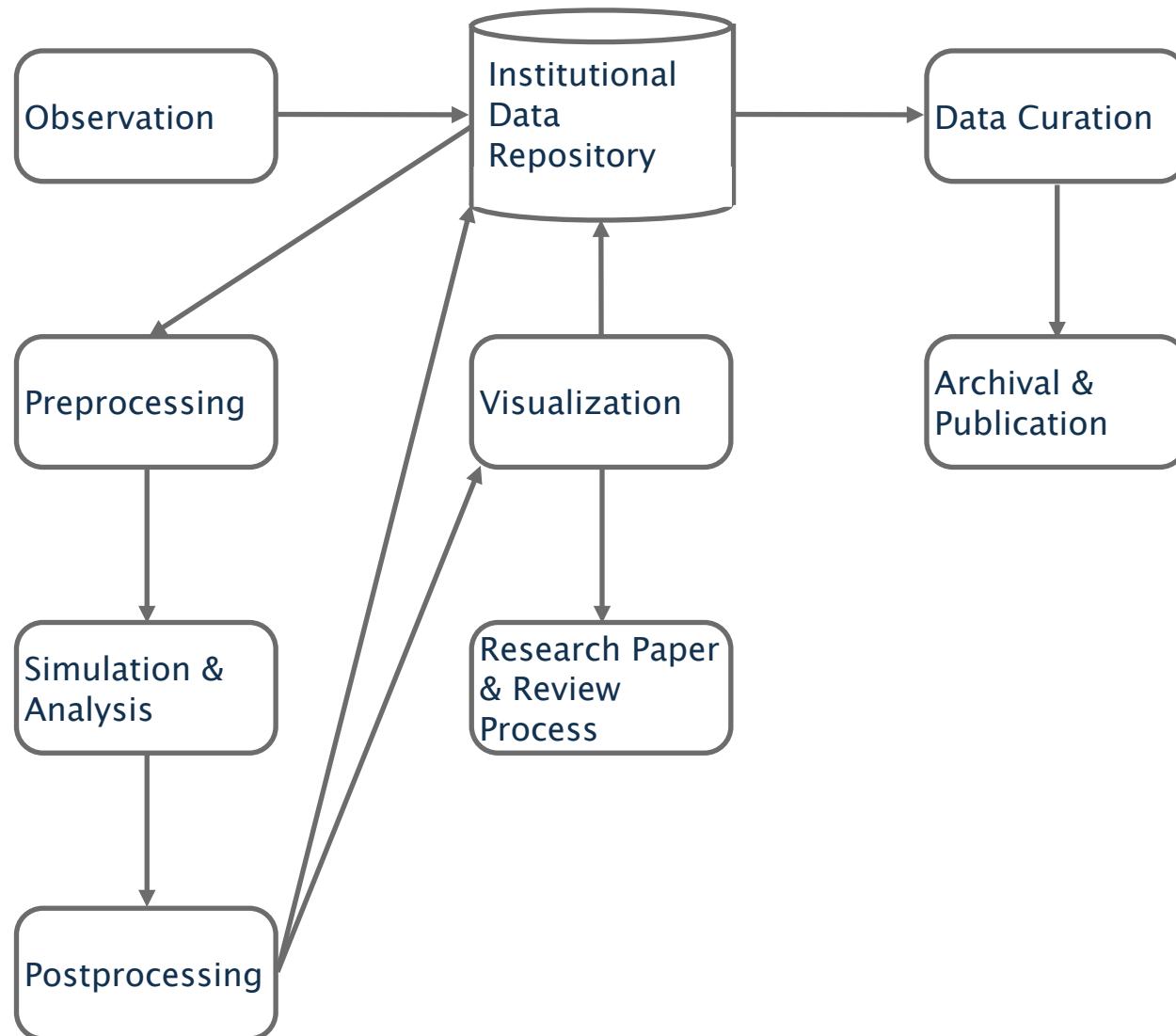


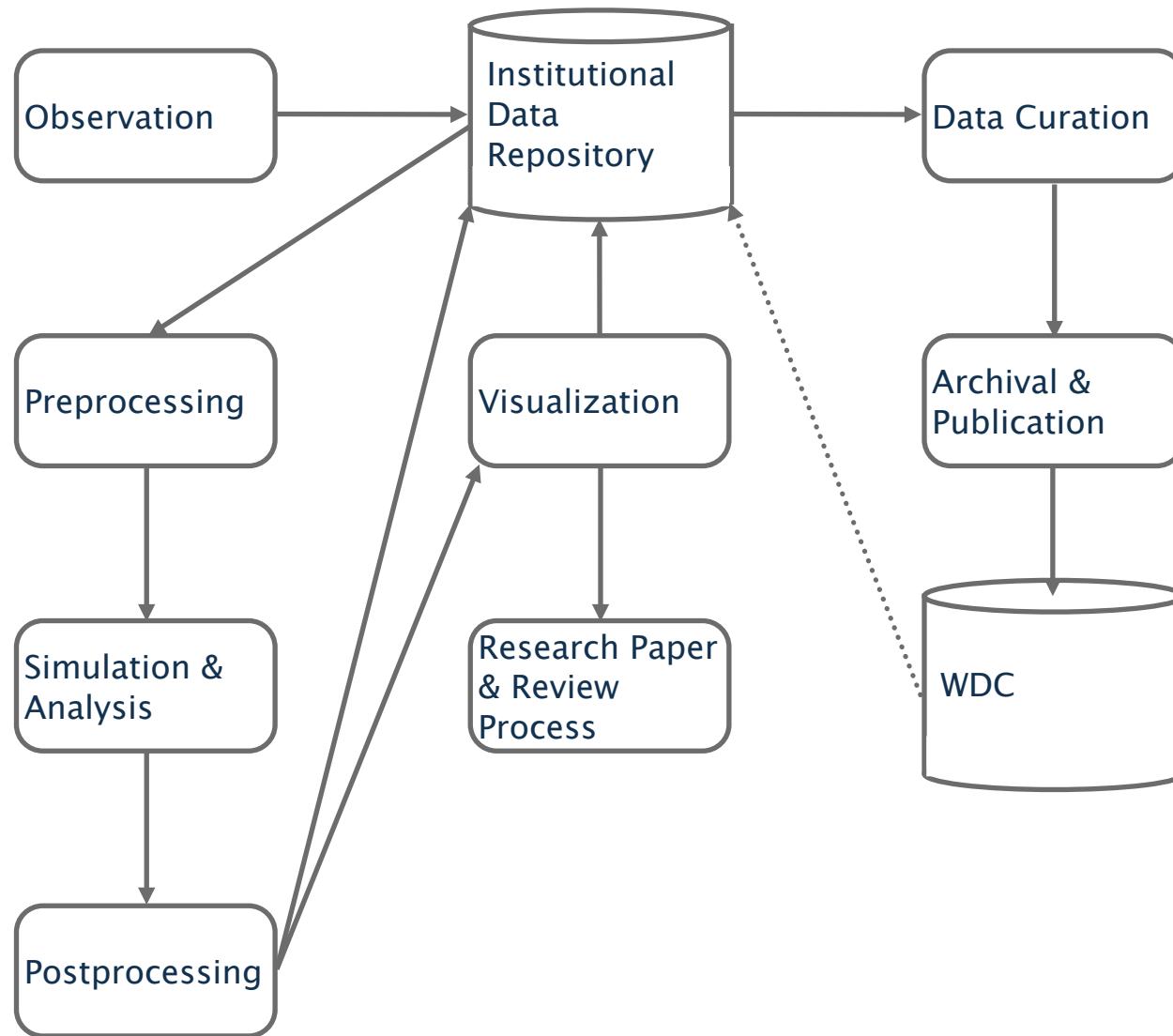


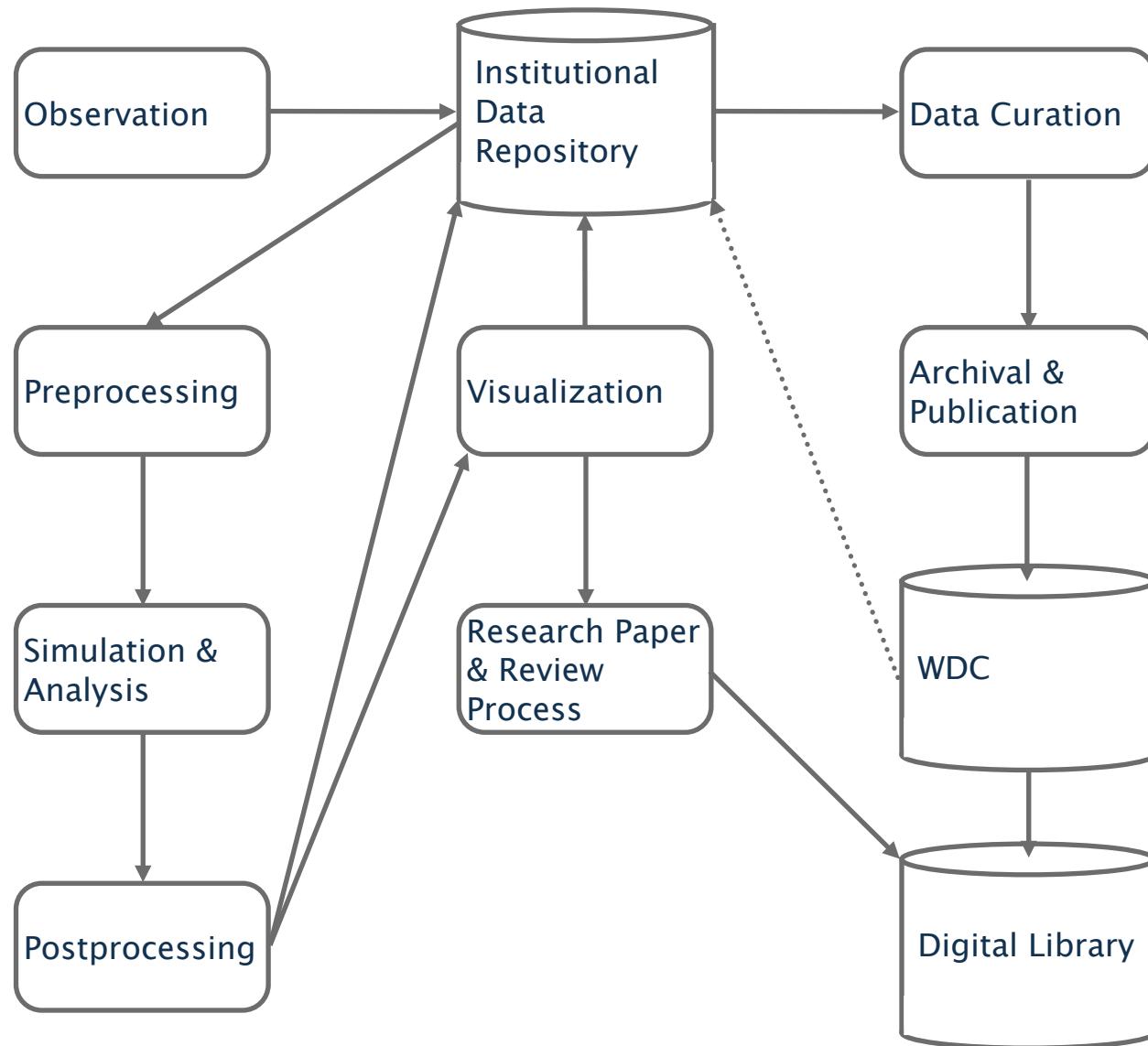


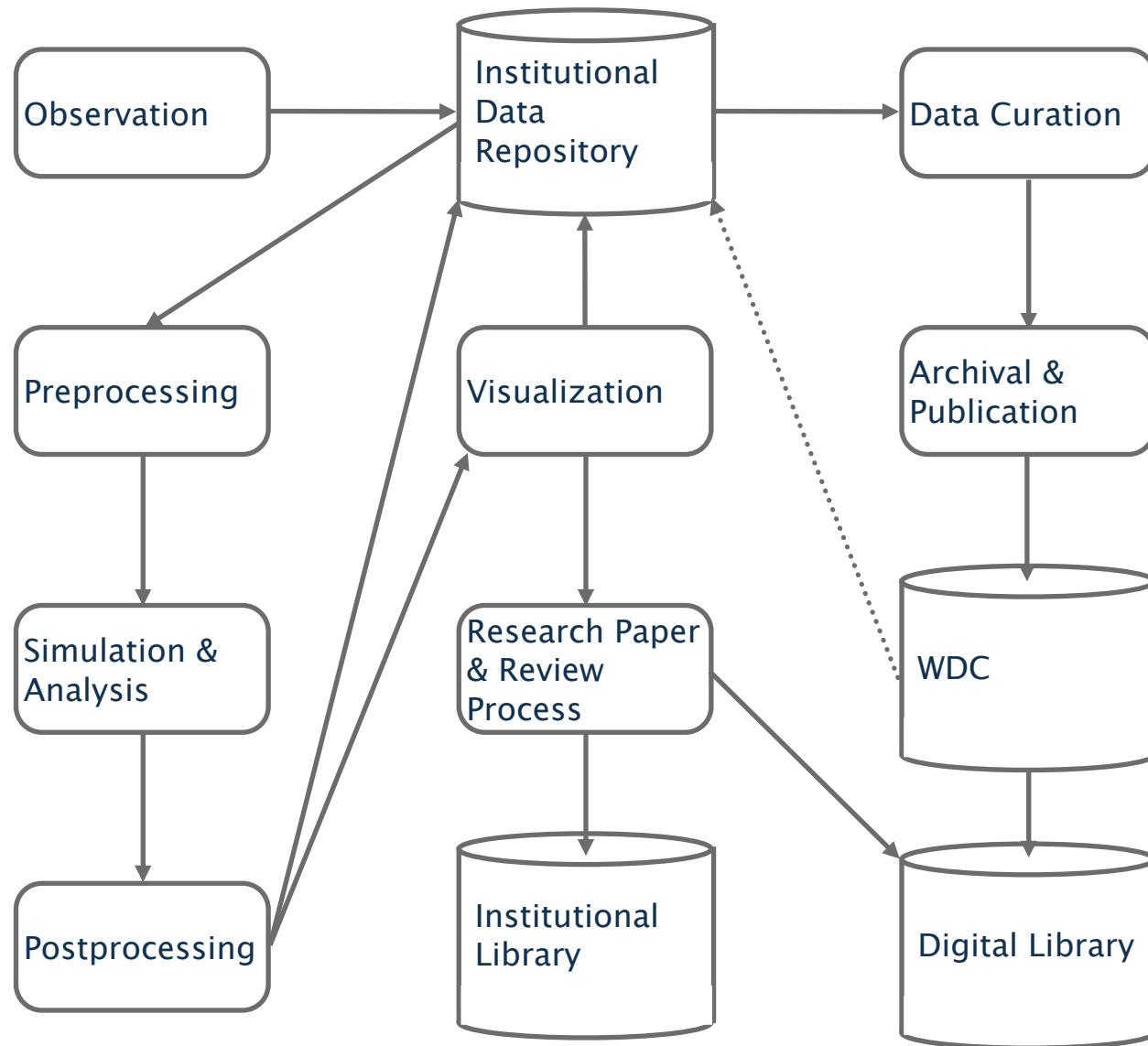


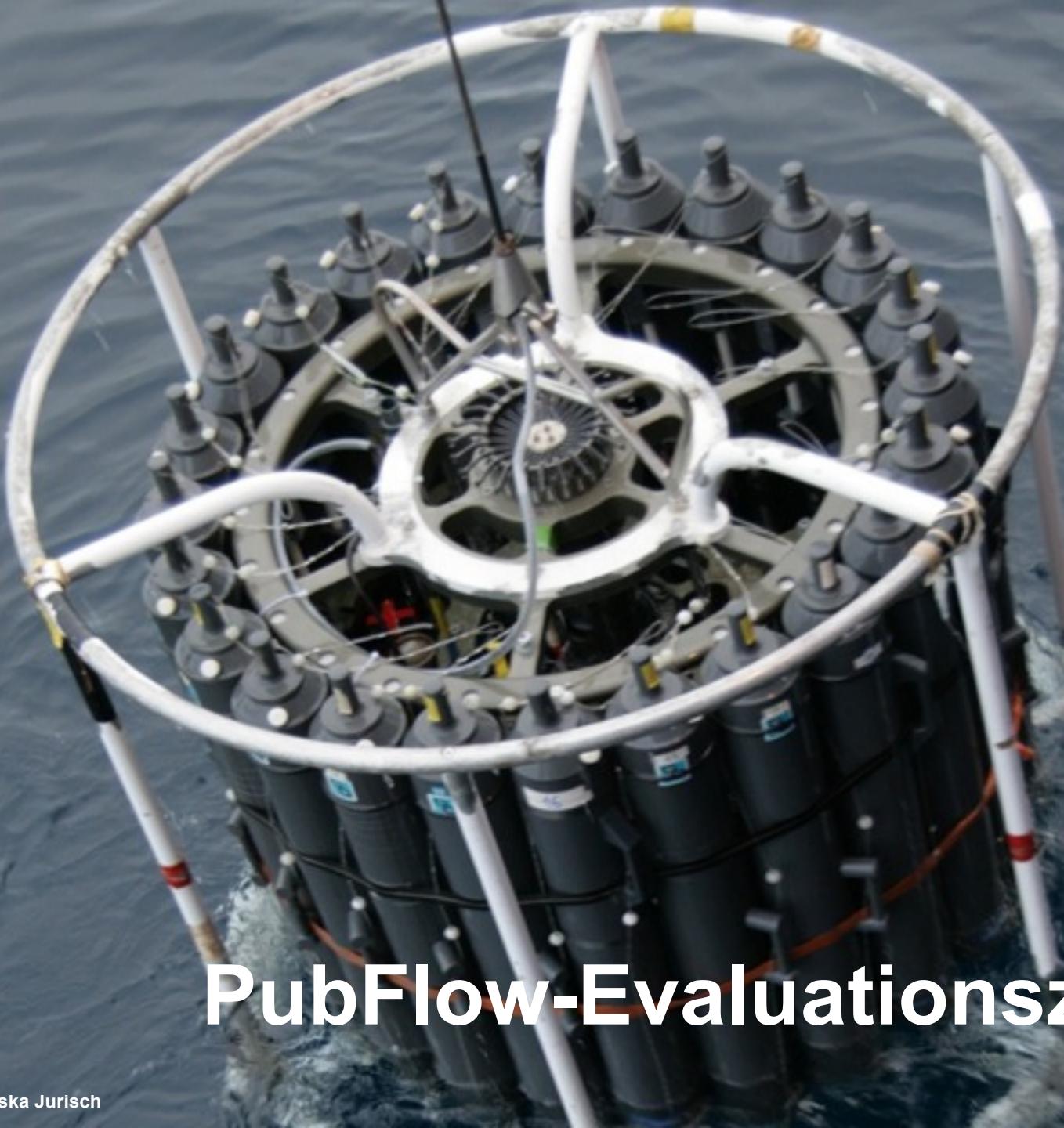












# PubFlow-Evaluationszenario

# Kernanforderungen

- ▶ Klare Trennung der Aufgaben
- ▶ Einfache Bedienbarkeit
- ▶ Integration in die tägliche Arbeit
- ▶ Hohe Konfigurierbarkeit

File ▾ Process ▾ Configuration

Available Processes  
No records found.

Deployed Processes  
Process name  
+ OCN

Refresh

run process

5

Leg Id	
PID	
Login	
Source	
Author	
Project	
Topology	
Status	
SavePath	
Reference	
Name of export file	
Comment	

Description

Log Run

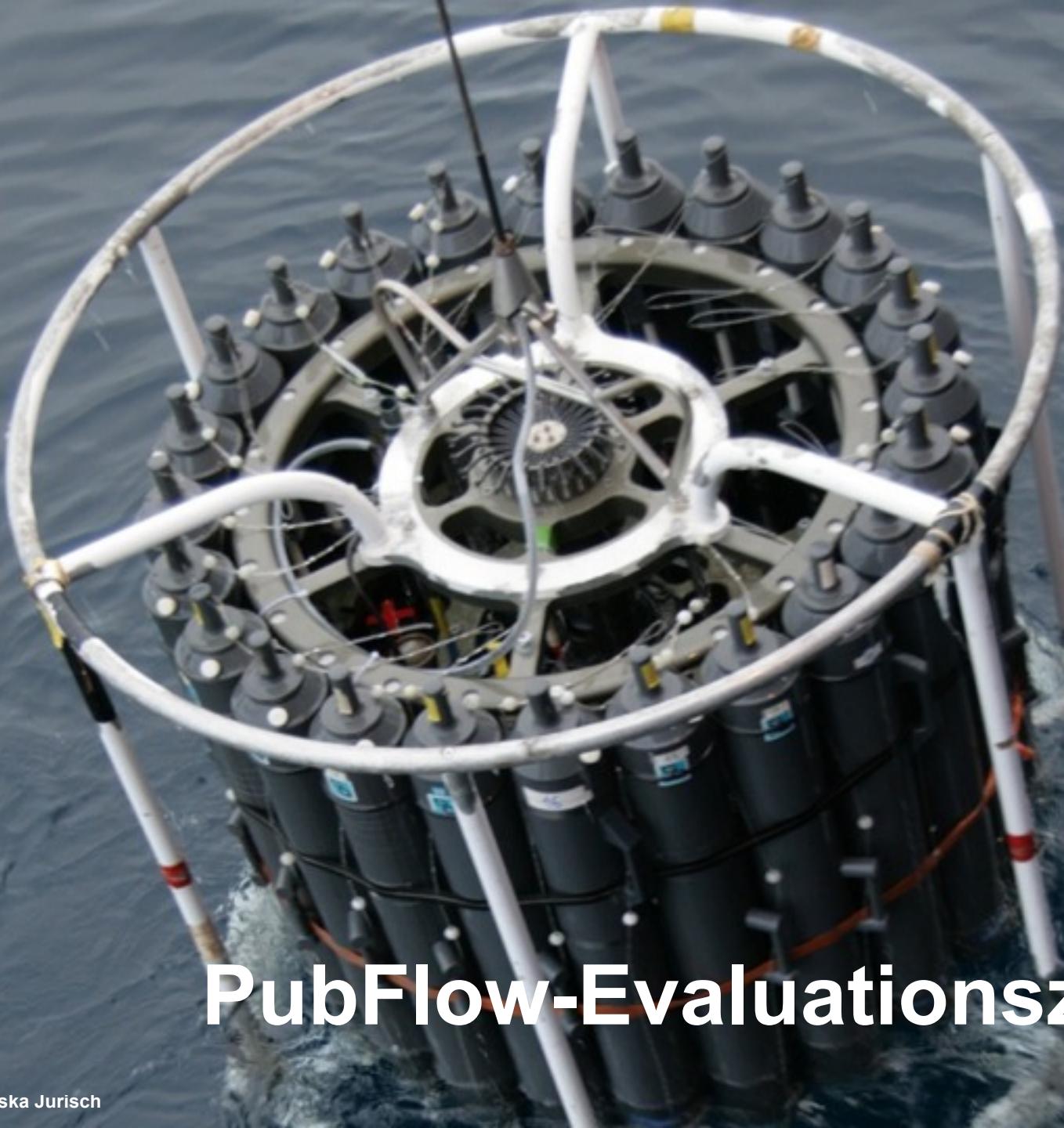
Do it! close

The screenshot shows the PubFlow Version 1.0 application. On the left, there's a sidebar with 'Available Processes' (empty) and 'Deployed Processes' (one entry: 'OCN'). A 'Refresh' button is also present. The main area features a 'run process' dialog box. This dialog has a title bar 'run process' with a close button, a numeric field '5', and a table with 12 rows. The table columns are 'Label' and 'Value'. The labels are: Leg Id, PID, Login, Source, Author, Project, Topology, Status, SavePath, Reference, Name of export file, and Comment. Below the table are two buttons: 'Do it!' and 'close'. To the right of the dialog, there's a large empty area with a 'Description' label and 'Log'/'Run' buttons.

▶ PubFlow Version 1.0

Abschnitt 2

# **Workflows in PubFlow**

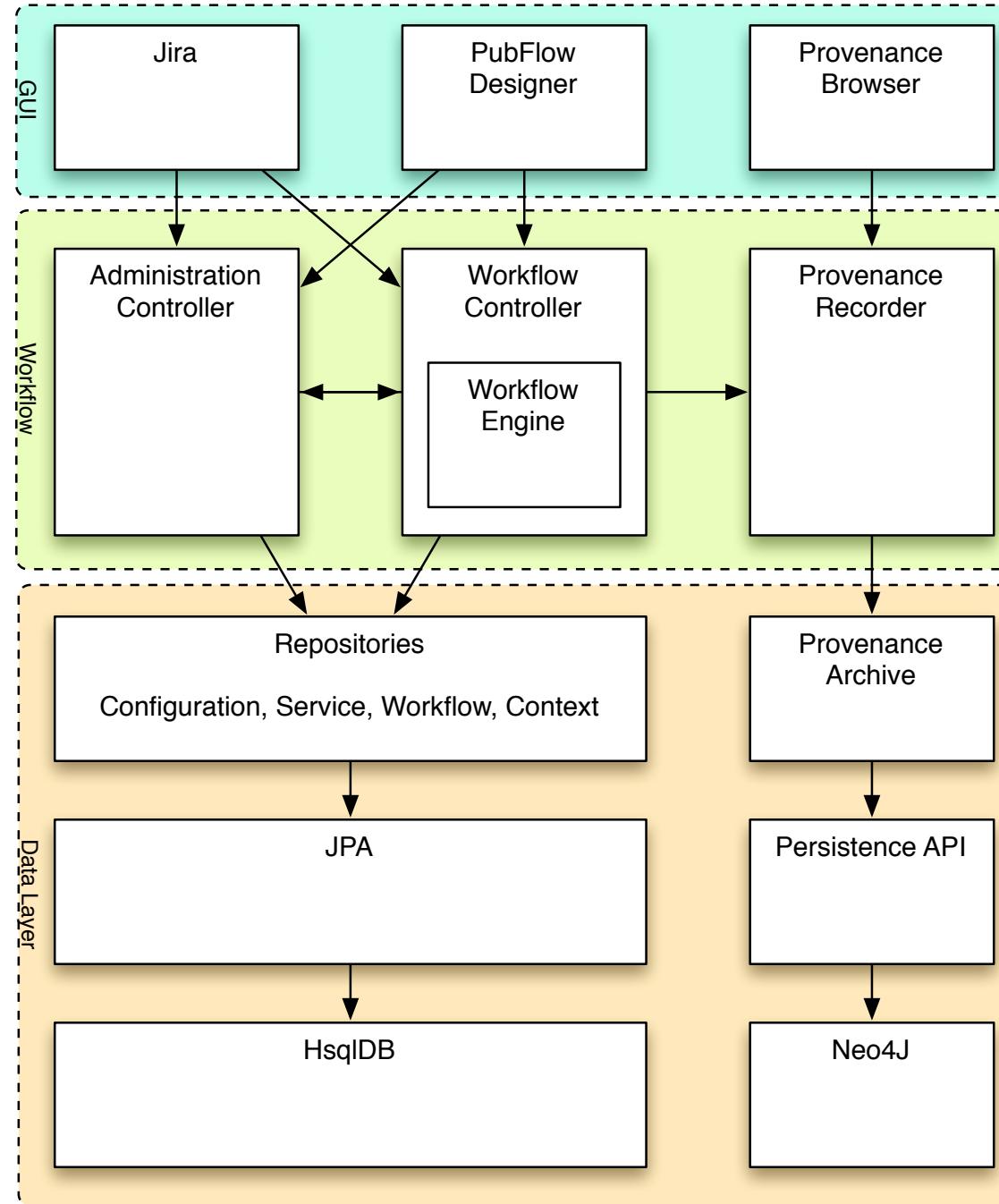


# PubFlow-Evaluationszenario

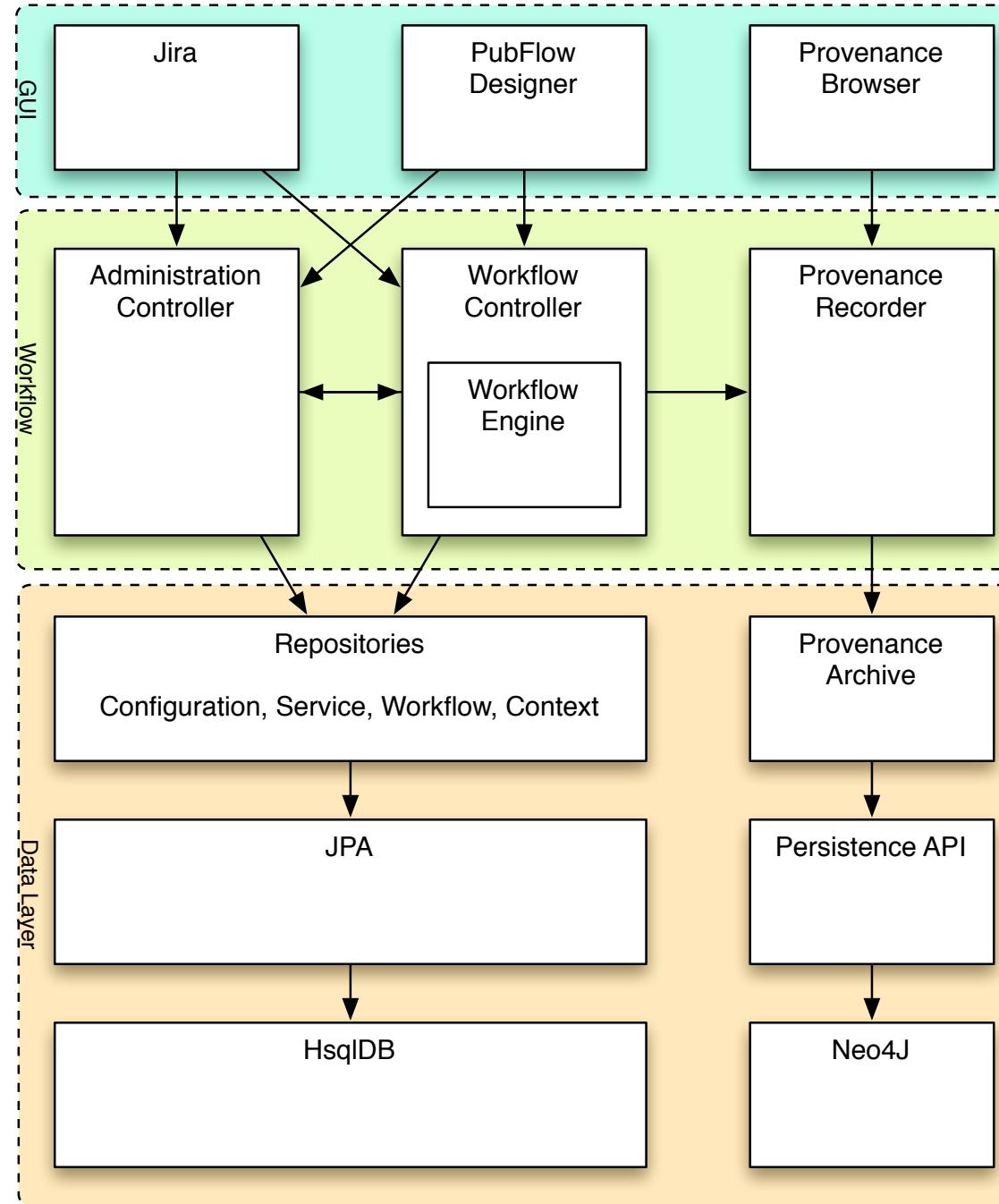


# Flow-Evaluationszenario

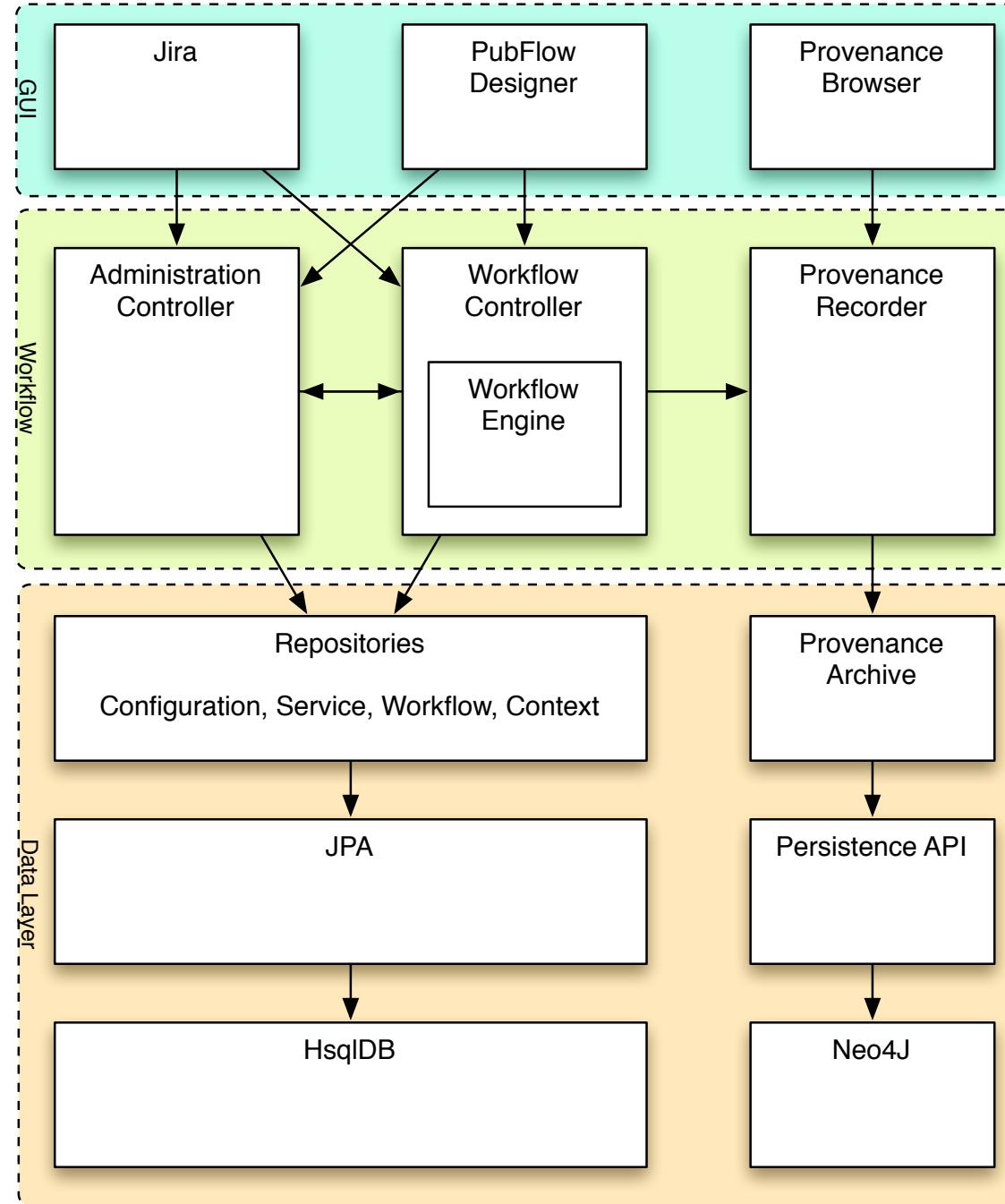
## > Workflows in PubFlow



## > Workflows in PubFlow



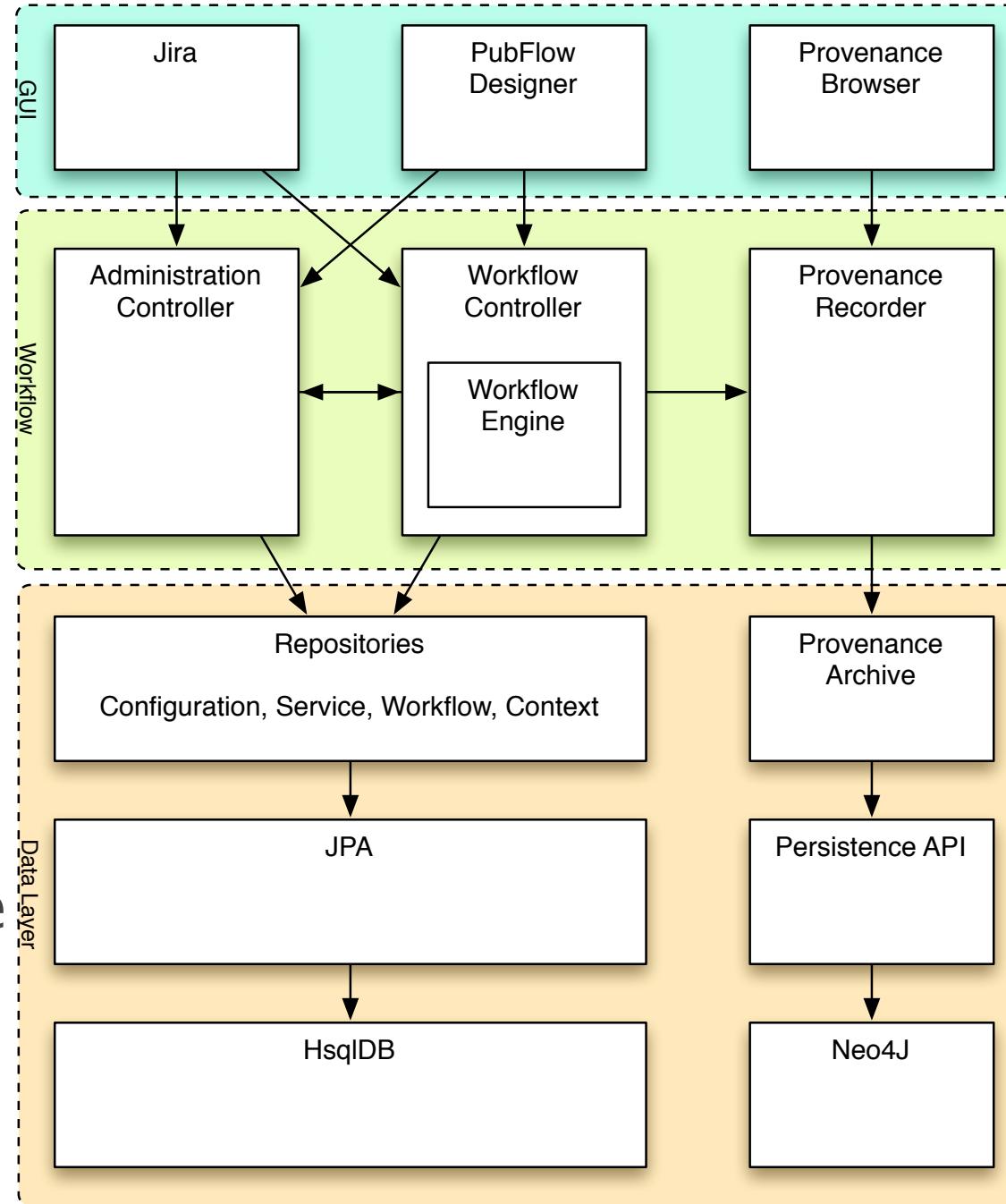
**GWT**  
Jira



**BPEL**  
Trans-  
formation  
Framework  
**BPMN 2.0**  
Jetty  
**Apache**  
**ODE**

**GWT**

Jira

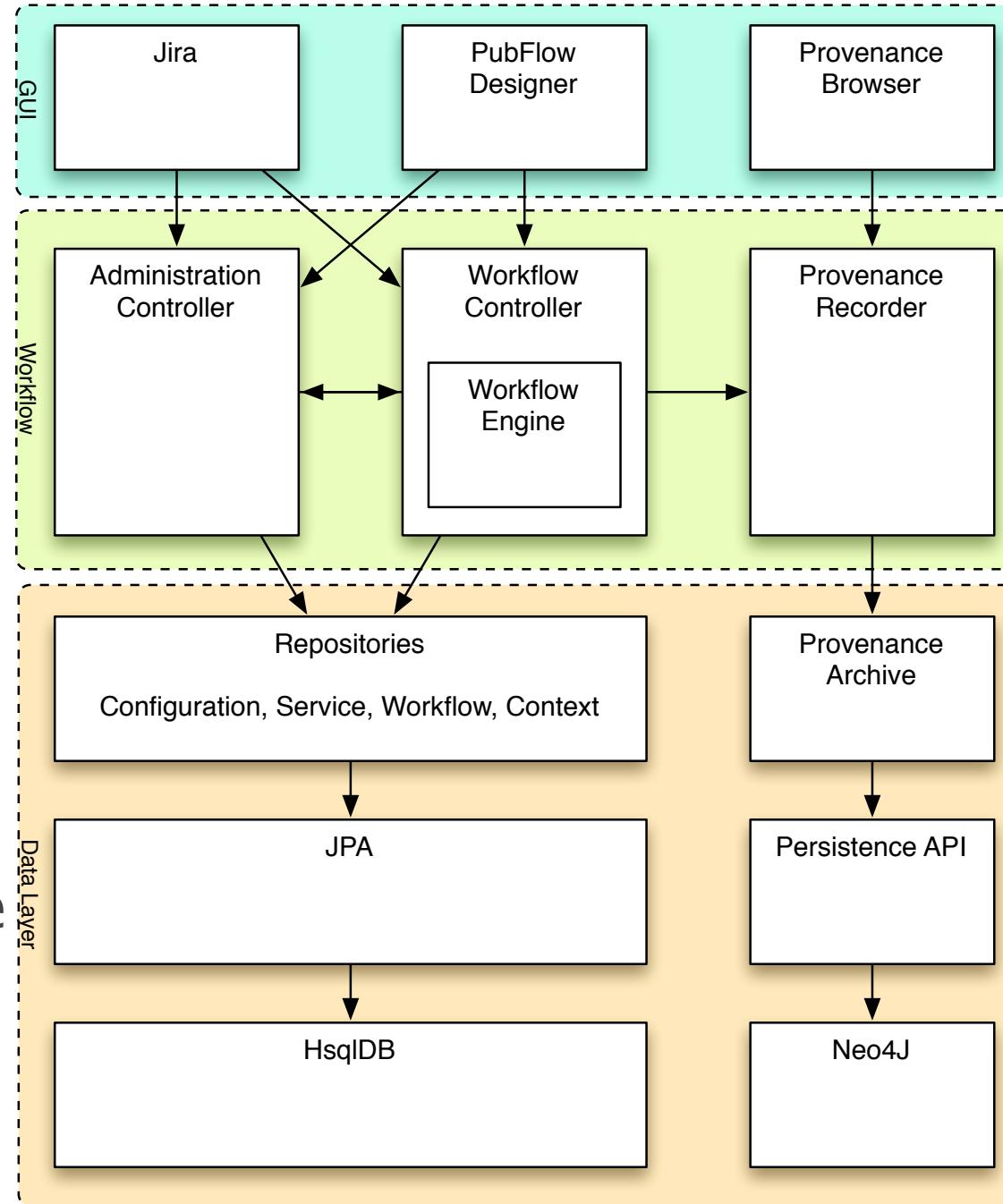


**BPEL**  
**Trans-**  
**formation**  
**Framework**  
**BPMN 2.0**  
**Jetty**  
**Apache**  
**ODE**

Apache Shiro  
 Hibernate  
**Quartz**  
 Hsqldb

**GWT**

Jira



**BPEL**  
Trans-  
formation  
Framework  
**BPMN 2.0**  
Jetty  
**Apache**  
**ODE**

Apache Shiro  
Hibernate  
**Quartz**  
HsqlDB

# Verschiedene Sichten

- ▶ PubFlow: Workflow Sicht für Daten-Manager
- ▶ Jira: Aufgaben als Tickets

## Create Issue

Configure Fields

Quick Search

Administer Project



PubFlow

Summary

Issues

Popular Issues

Labels

Project \* PubFlow Workflow Processing

Issue Type \* OCN\_IssueType

Source\_OCN 

Source-custom field for OCN

Login\_OCN 

Login-custom field for OCN

Status\_OCN 

Status-custom field for OCN

File name\_OCN 

File name-custom field for OCN

Leg comment\_OCN 

Leg comment-custom field for OCN

PID\_OCN 

PID-custom field for OCN

Topology\_OCN 

Topology-custom field for OCN

Author\_OCN 

Author-custom field for OCN

Leg ID\_OCN 

Leg ID-custom field for OCN

Reference\_OCN 

Reference-custom field for OCN

Project\_OCN 

Project-custom field for OCN

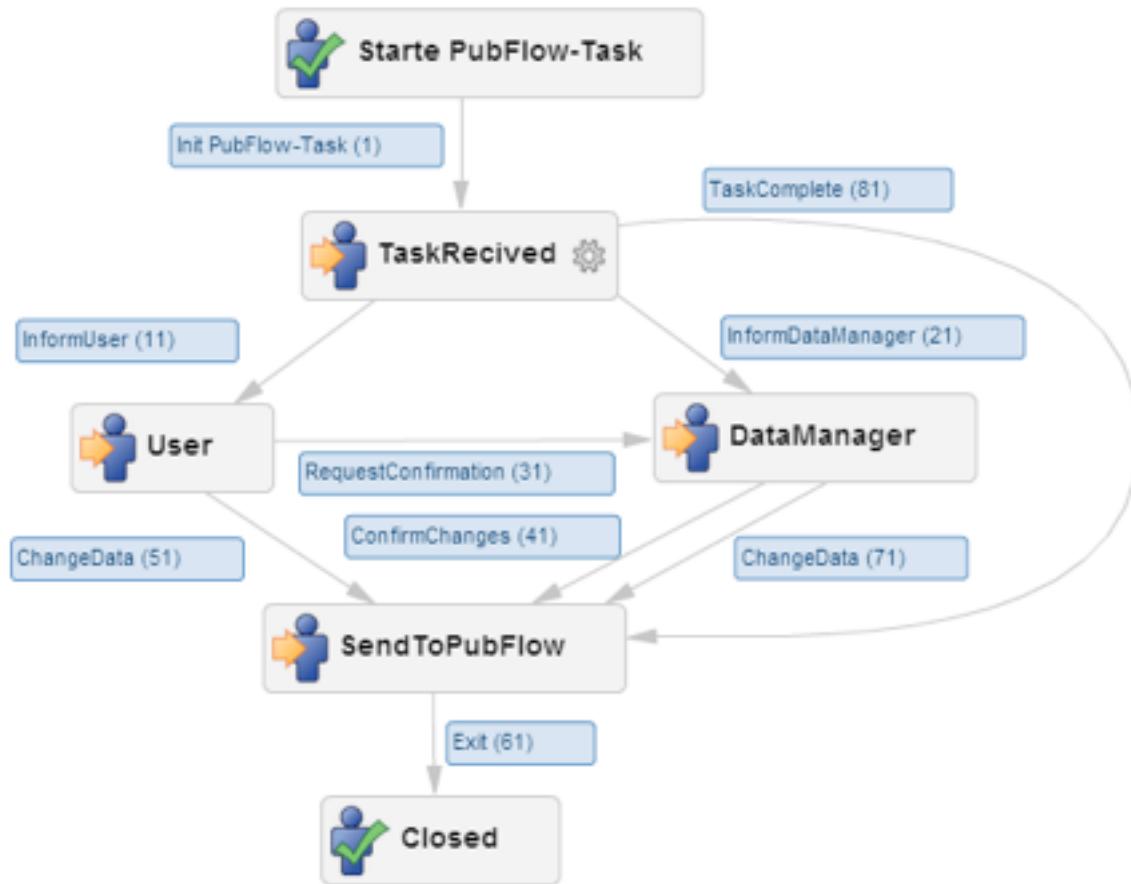
Zielpfad\_OCN 

Zielpfad-custom field for OCN

Reporter \* admin

Start typing to get a list of possible matches.

Summary \*  Create another



## ► Der PubFlow Jira Workflow

 **PubFlow Jira Listener** Disable

[Uninstall](#)

No Screenshots Available	Version: 1.0.0.0 Developer: CAU Plugin key: de.pubflow.jira.listener.PubflowListener	
<input type="checkbox"/> 2 of 2 modules enabled		
eventPublisher		
<b>EventListener</b> Class that processes the incoming JIRA issue events.		

## ► Das PubFlow Jira Plugin

# pubflow-graph-visualization-framework

A visualization framework for graphs written in JavaScript

[Project Home](#) [Downloads](#) [Wiki](#) [Issues](#) [Source](#) [Administer](#)

[Summary](#) [People](#)

Tip: Discuss and then document [each teammate's project duties](#).

## Project Information

Recommend this on Google

Starred by 0 users  
[Project feeds](#)

Code license  
[Apache License 2.0](#)

Labels  
Academic, JavaScript,  
HTML5, Visualization

Members  
peer.brauer  
2 committers  
1 contributor

Your role  
Owner

# pubflow-provenance-archive



An archive to store the provenance information collected by the PubFlow research data publication framework.

[Project Home](#) [Wiki](#) [Source](#) [Administer](#)

[Summary](#) [People](#)

Tip: Discuss and then document [each teammate's project duties](#).

## Project Information

Recommend this on Google

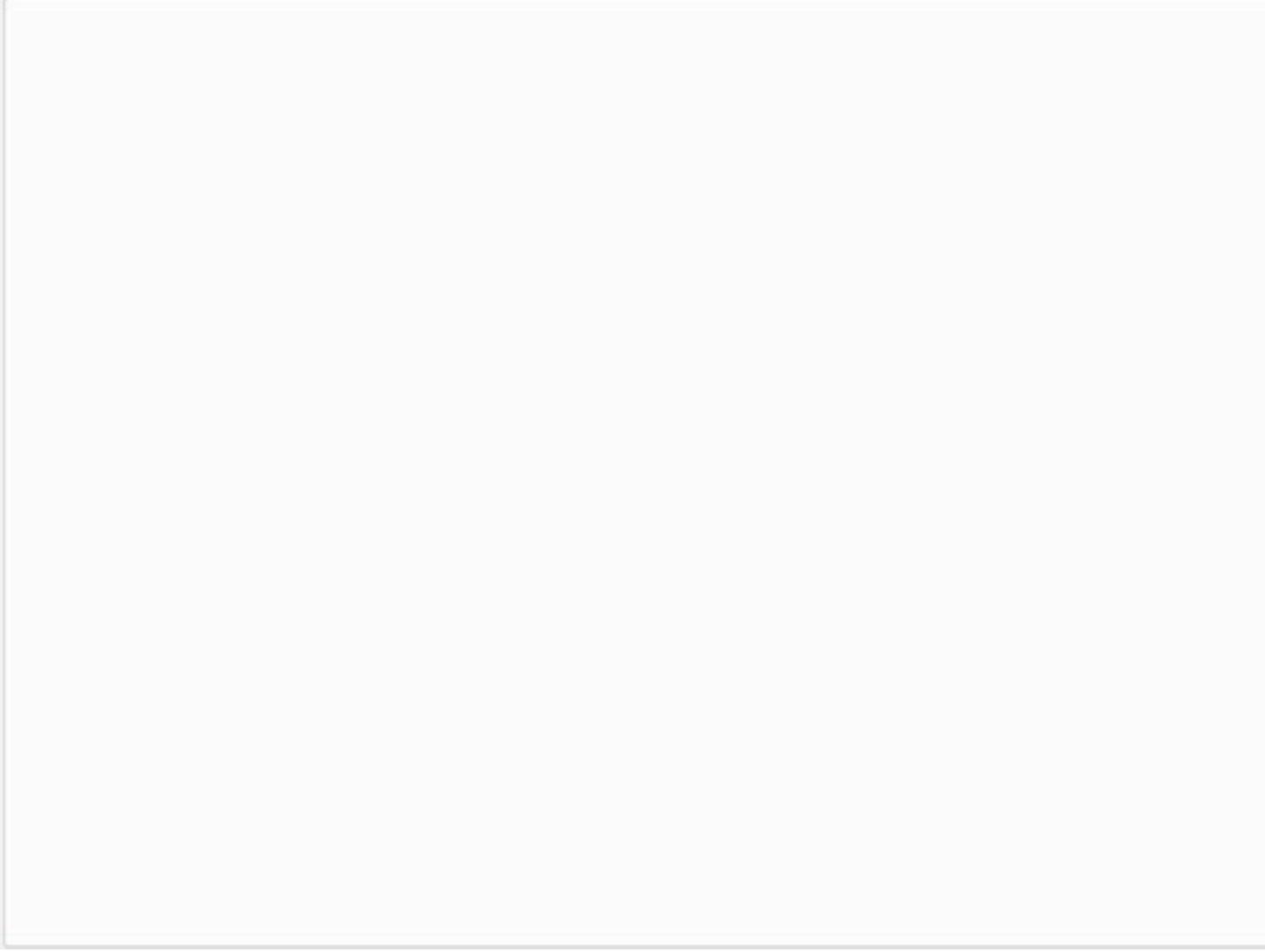
Starred by 0 users  
[Project feeds](#)

Code license  
[Apache License 2.0](#)

Labels  
Academic,  
Datamanagement,  
Provenance

The PubFlow Provenance Archive is a subproject of the PubFlow research data publication framework ([www.pubflow.de](http://www.pubflow.de)). PubFlow is a framework, which helps datamangers to transfer data from an institutional repository to a public available archive. During the transfer PubFlow provenance information for the data transferred, describing what happened to the dataset. This provenance information record is then stored in this archive.

The archive is build upon the W3C ProvO standard and neo4j.



## ► PubFlow Graph-Framework



## PROV Implementation Report

W3C Working Draft 12 March 2013

### 2. List of Implementations

The following lists the reported implementations, the type of implementation, supported PROV encodings and the URL of the implementation.

Implementation Type:

- Application
- Framework/API
- Service
- Vocabulary
- Constraints Validator

Table 1: List of implementations reported to the PROV Working Group.

#	Name	Type	PROV Encodings
1	<a href="#">StatJR eBook system</a>	Application	PROV-O, PROV-JSON
2	<a href="#">PoN</a>	Application	PROV-N
3	<a href="#">WebLab-PROV</a>	Application	PROV-O, PROV-N
4	<a href="#">Taverna</a>	Application	PROV-O
...			
33	<a href="#">PubFlow Provenance Archive</a>	Application Framework / API	PROV-O, PROV-XML
34	<a href="#">PROV Python library</a>	Framework / API	PROV-N, PROV-JSON
35	<a href="#">... additional entries ...</a>	Application	PROV-O

► PubFlow Provenance Archive

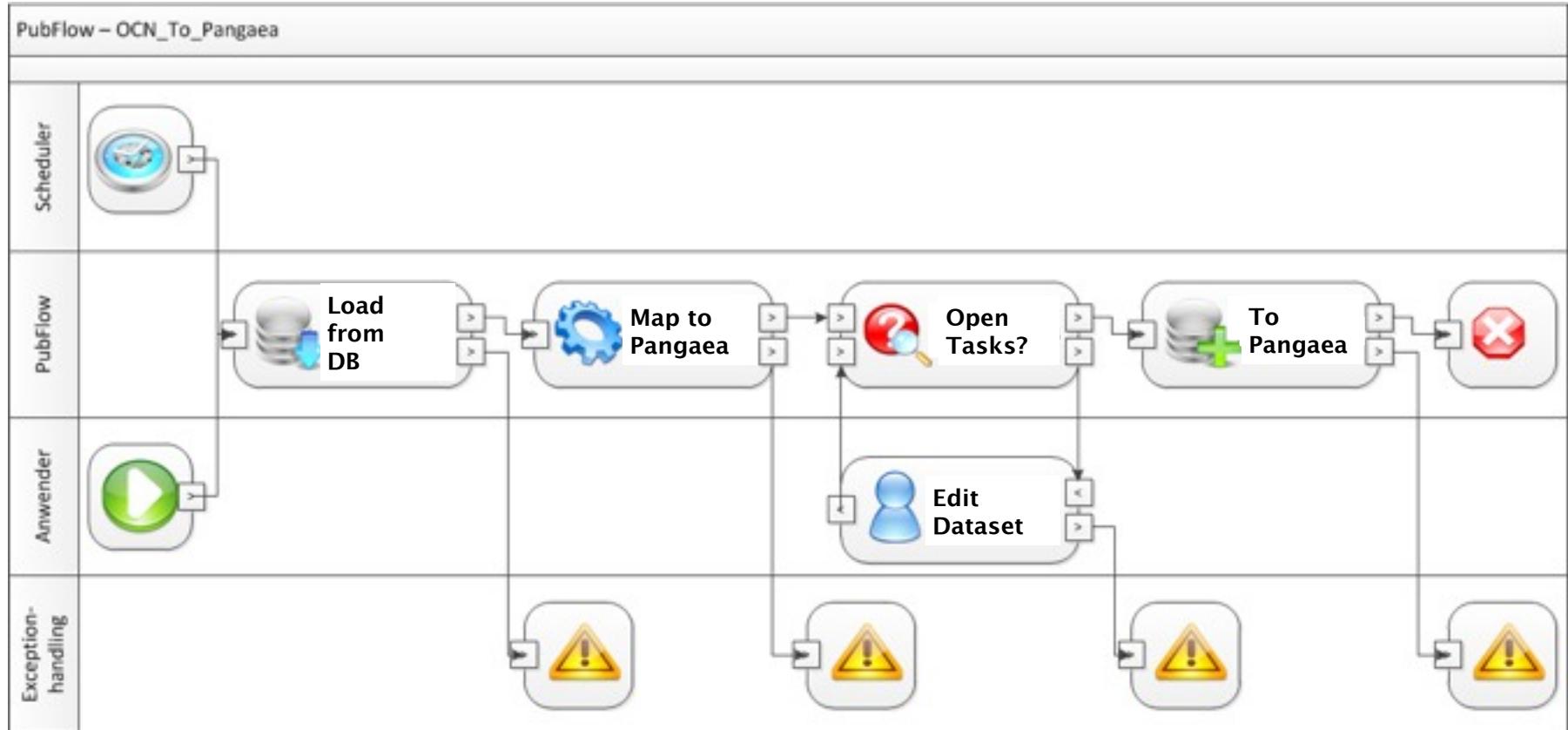
Abschnitt 3

# **Ausblick & Visionen**

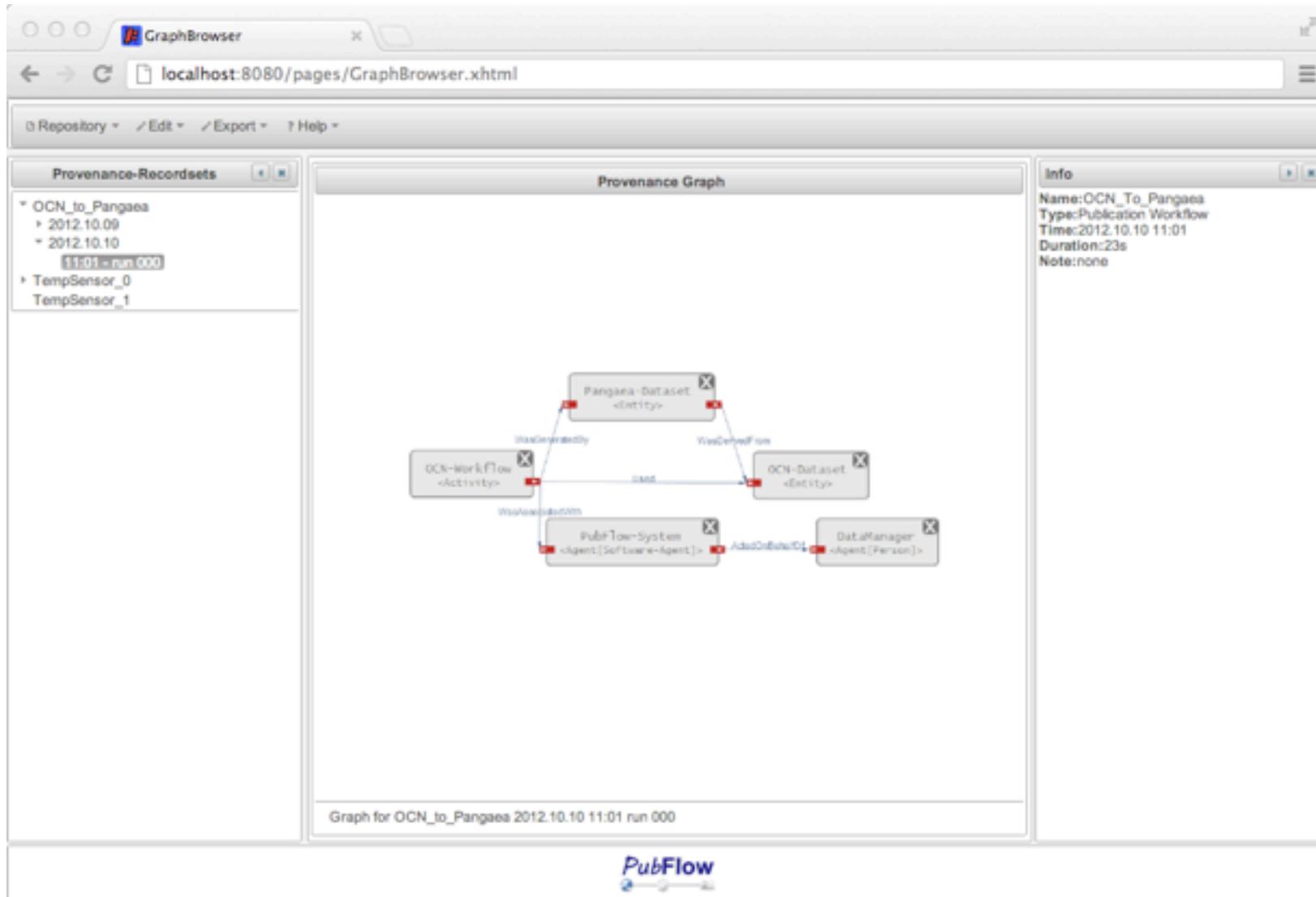
The screenshot shows a digital library interface with the following elements:

- Top Navigation:** Folders, My Documents, Recent Documents.
- Breadcrumbs:** Folders > SampleData.
- Actions Bar:** Add Subfolder, Permissions.
- Folder List:** A table showing 0 results under the "Folder" tab.
- Document List:** A table showing 1 result under the "Documents" tab. The document is "SampleFile.rtf" (modified 10/29/12 2:10 PM) by Peer Christoph Brauer. It has 0.2K size and 0 comments. Actions include Provenance, PubFlow, and Actions.
- Search:** Search input field and "Search this Folder" button.
- Callout Box:** A green callout box highlights the "PubFlow" button in the document actions area.
- Bottom Bar:** Imprint.

► Integration von PubFlow in das Portal



- ▶ Entwicklung von domainenspezifischen DSLs für PubFlow



► Fertigstellung des Provenance-Browsers

**Vielen Dank  
für Ihre  
Aufmerksamkeit**

