PubFlow
a scientific data publication framework for marine science

Peer Brauer, Prof. Dr. Wilhelm Hasselbring
Software Engineering Group
Kiel University
Lucca, 23.09.2013

Christian-Albrechts-Universität zu Kiel

- *Software Engineering Group*
- Library of CAU
- Computing Center of CAU

GEOMAR

- Kiel Data Management Team
- Library of Geomar
- Data and Computing Center Geomar

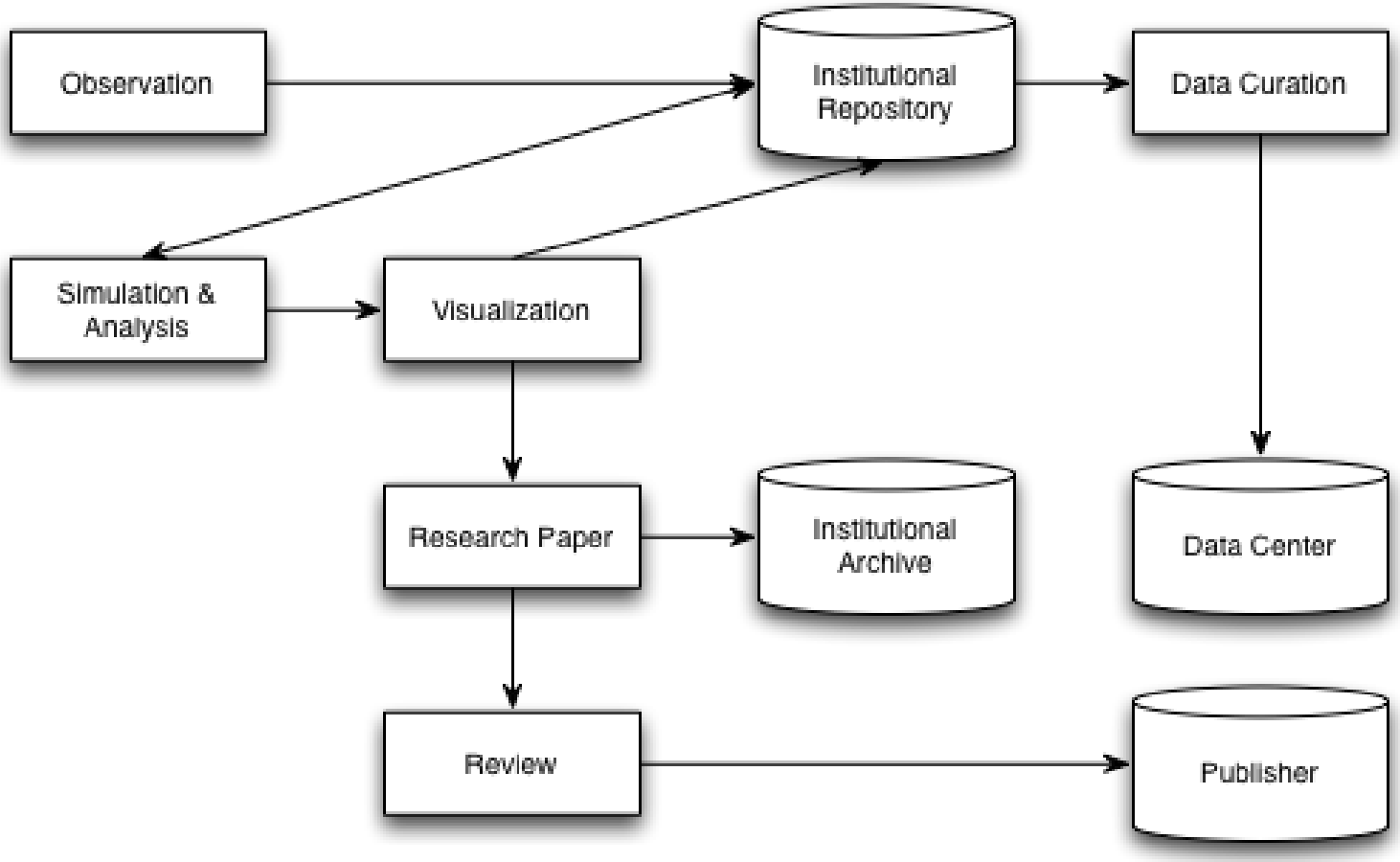
ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

funded by DFG

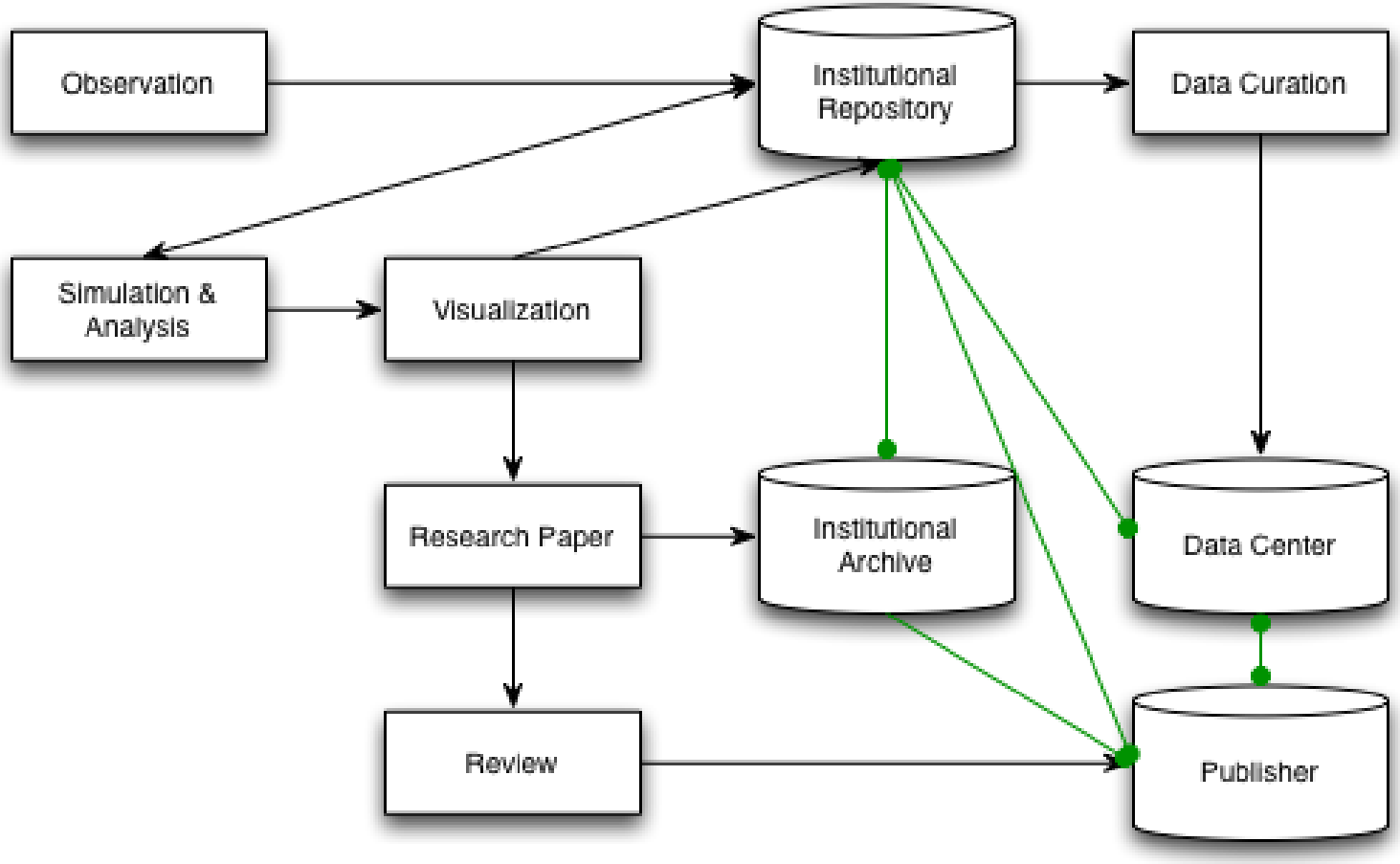
PubFlow

www.pubflow.de

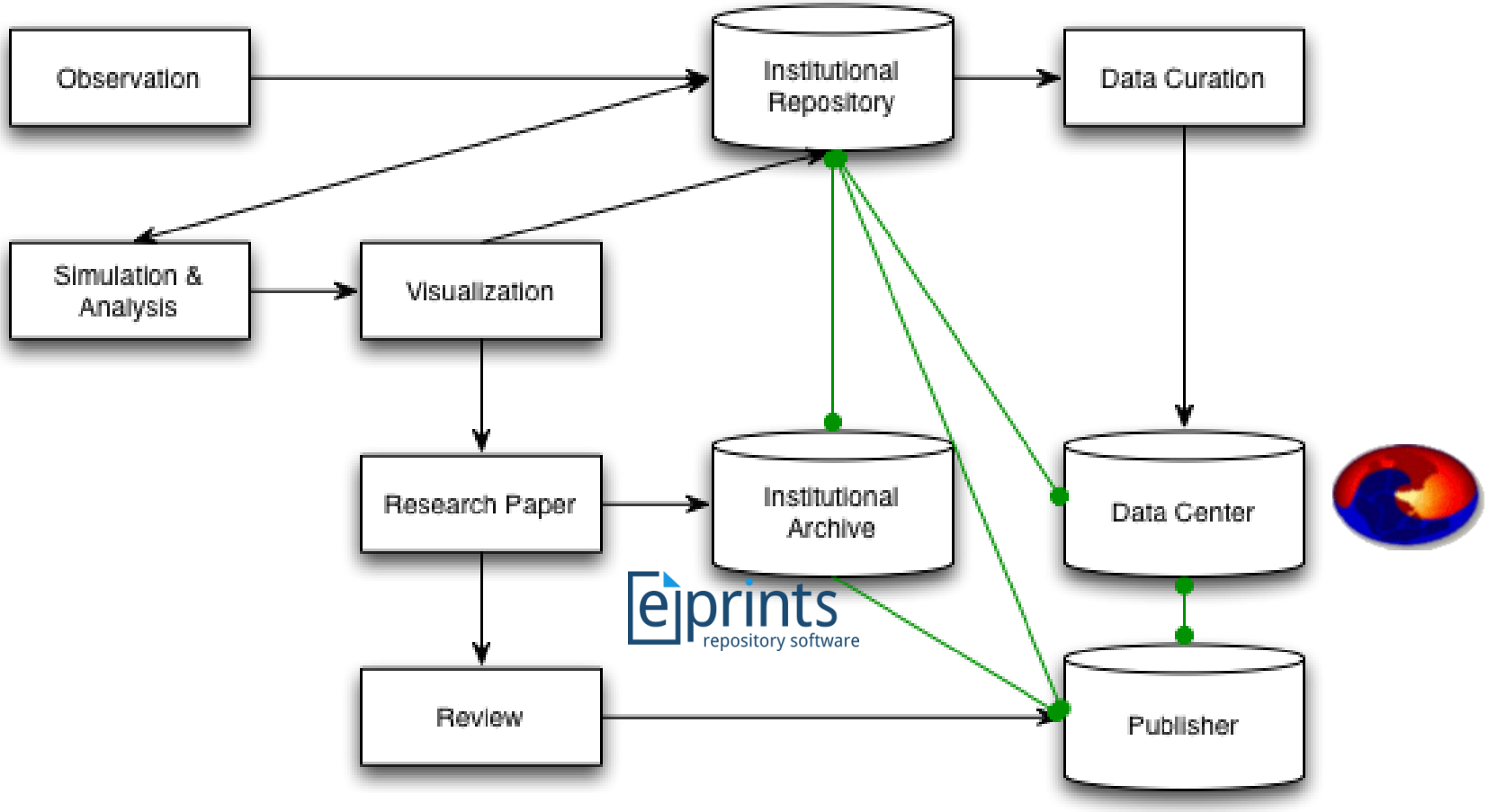
Motivation


Observation
Institutional Repository
Data Curation
Simulation & Analysis
Visualization
Research Paper
Institutional Archive
Data Center
Review
Publisher

Motivation


Observation
Institutional Repository
Data Curation
Simulation & Analysis
Visualization
Research Paper
Institutional Archive
Data Center
Review
Publisher

Motivation


Observation
Institutional Repository
Data Curation
Simulation & Analysis
Visualization
Research Paper
Institutional Archive
Data Center
Review
Publisher
eprints
repository software

- **Agenda**

1. PubFlow Framework
2. Evaluation Scenario
3. Conclusion

What is PubFlow about?

- Creating a scientific workflow environment for data publication
- Introducing role based working models to the domain of data management
- Increasing the degree of automation data management

Features

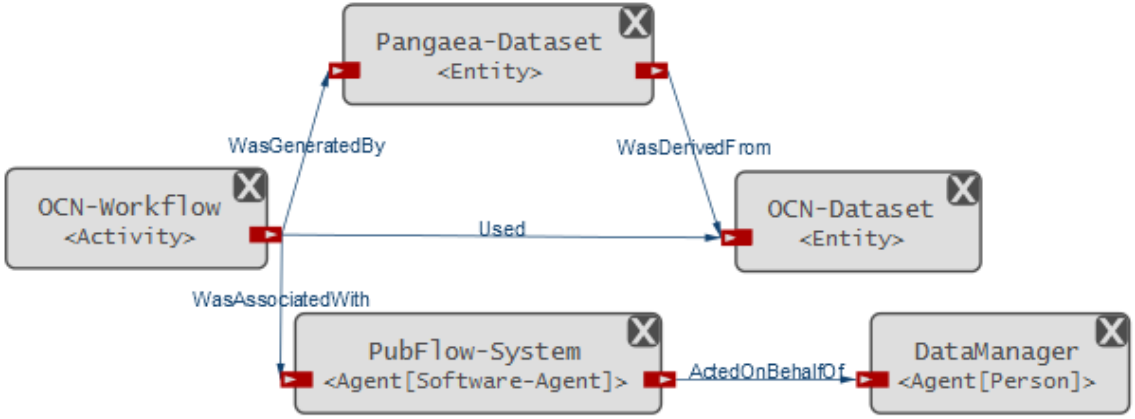
Build upon proven workflow technology

- Build in support for BPMN, BPEL
- Extensible by other workflow engines
- Designed for high throughput

Features

Provenance Awareness

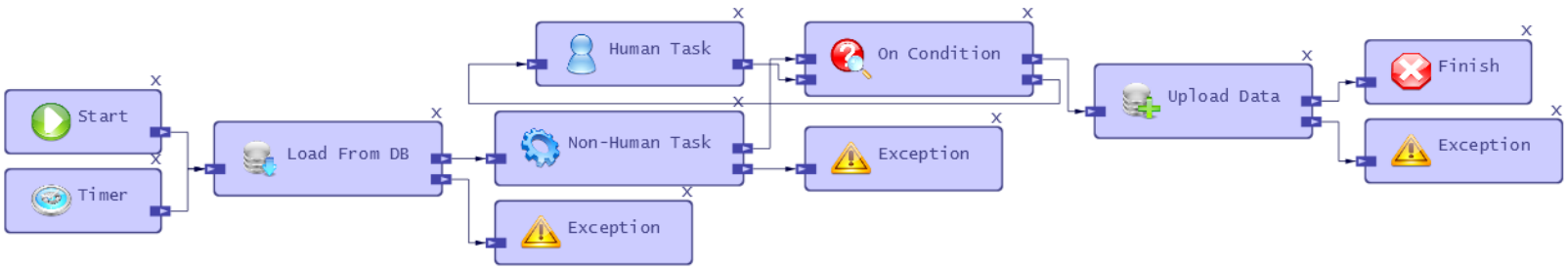
- Automatically capturing of provenance data
- Integrated W3C Prov-O compliant provenance archive
- Workflow based provenance browser



Features

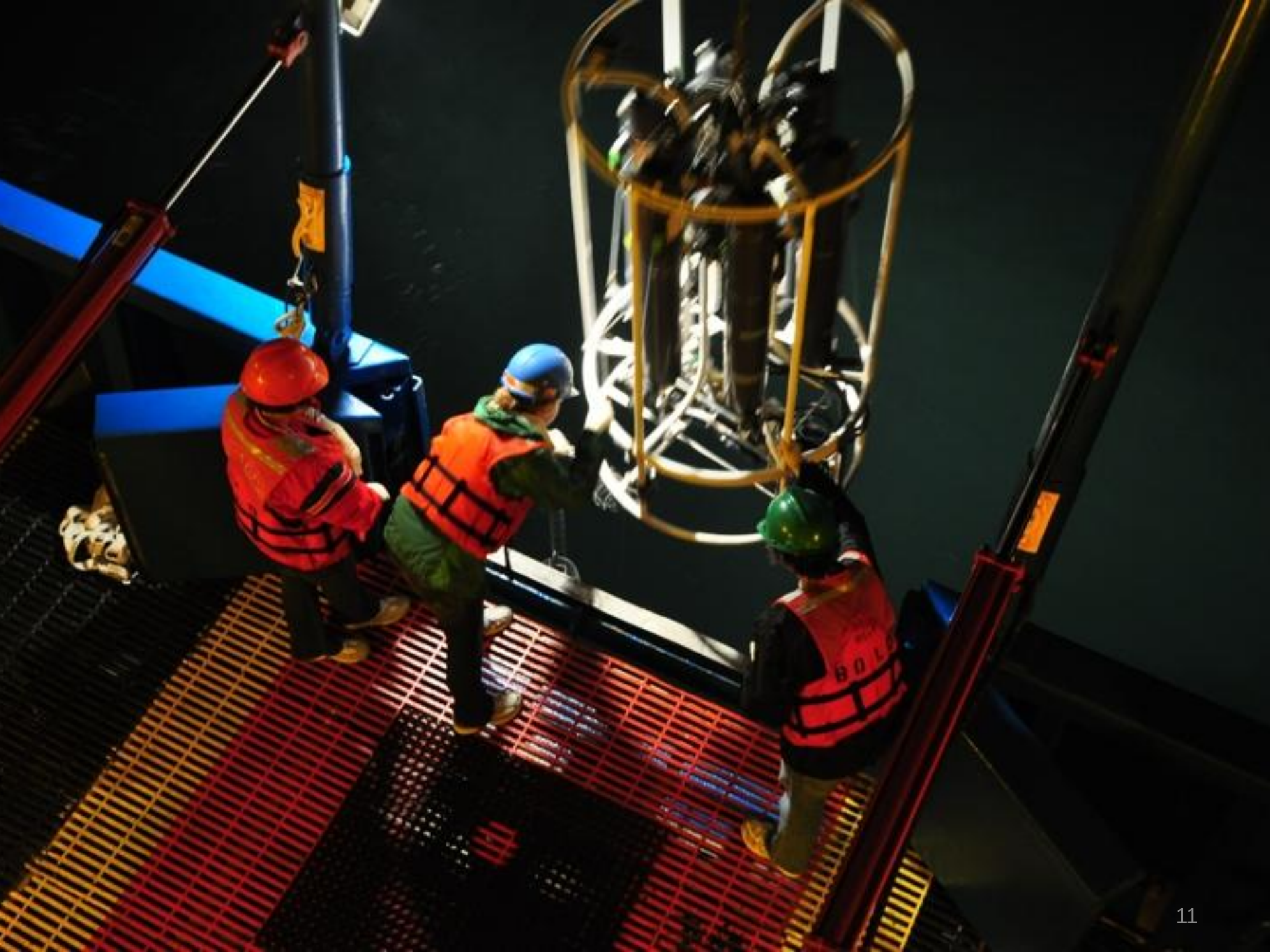
Graphical Workflow Editor

- Supports graphical DSLs
- Data managers can easily define own workflows
- Workflows can be transformed to selected target execution environment

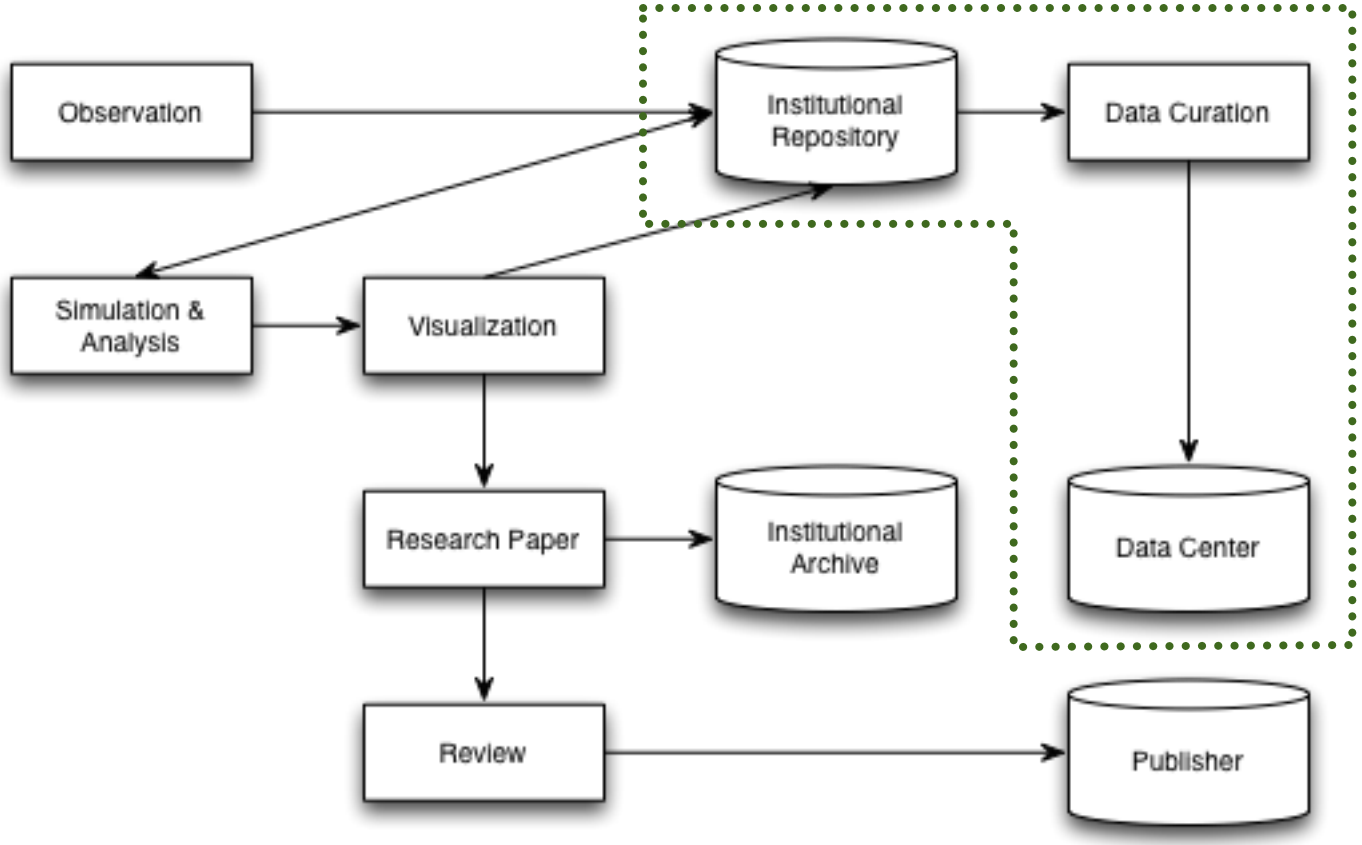


Section 2

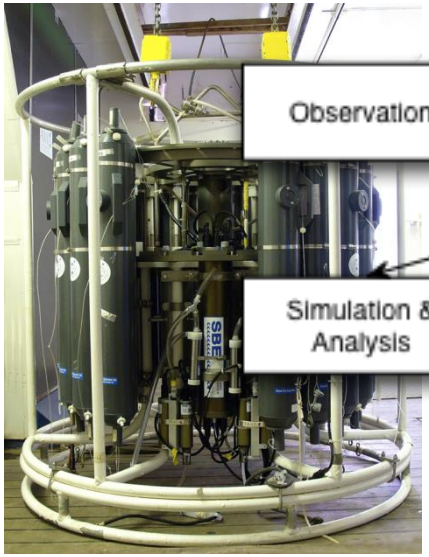
The Evaluation Scenario

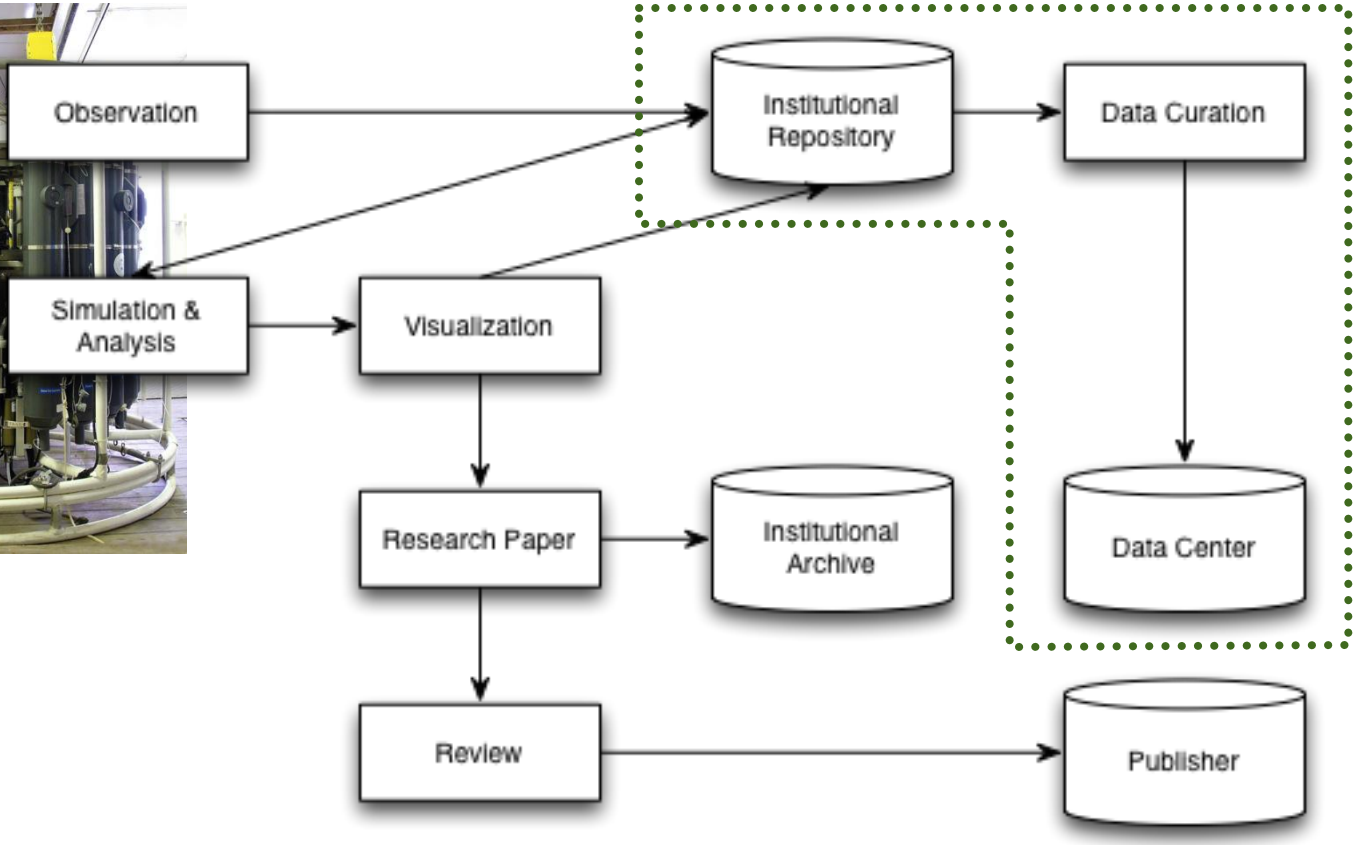


Evaluation Scenario

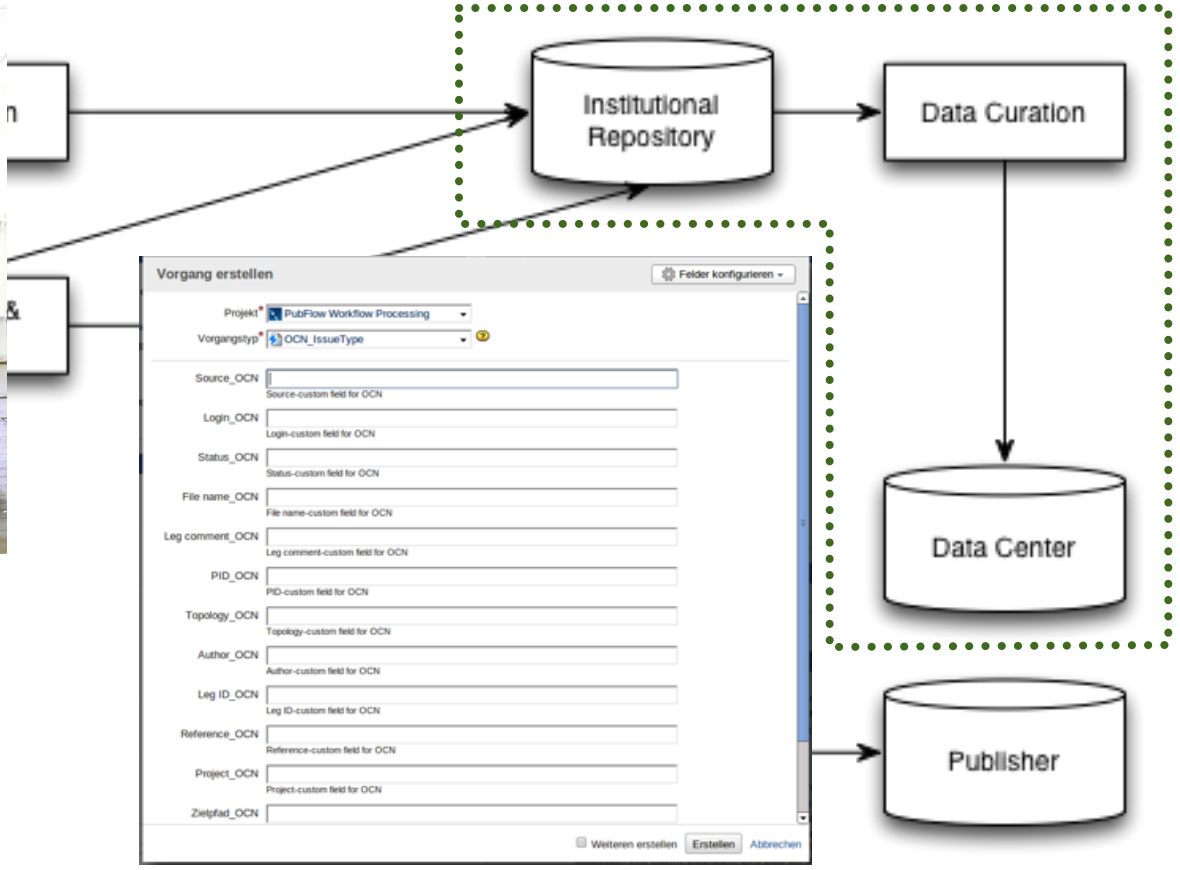


Evaluation Scenario

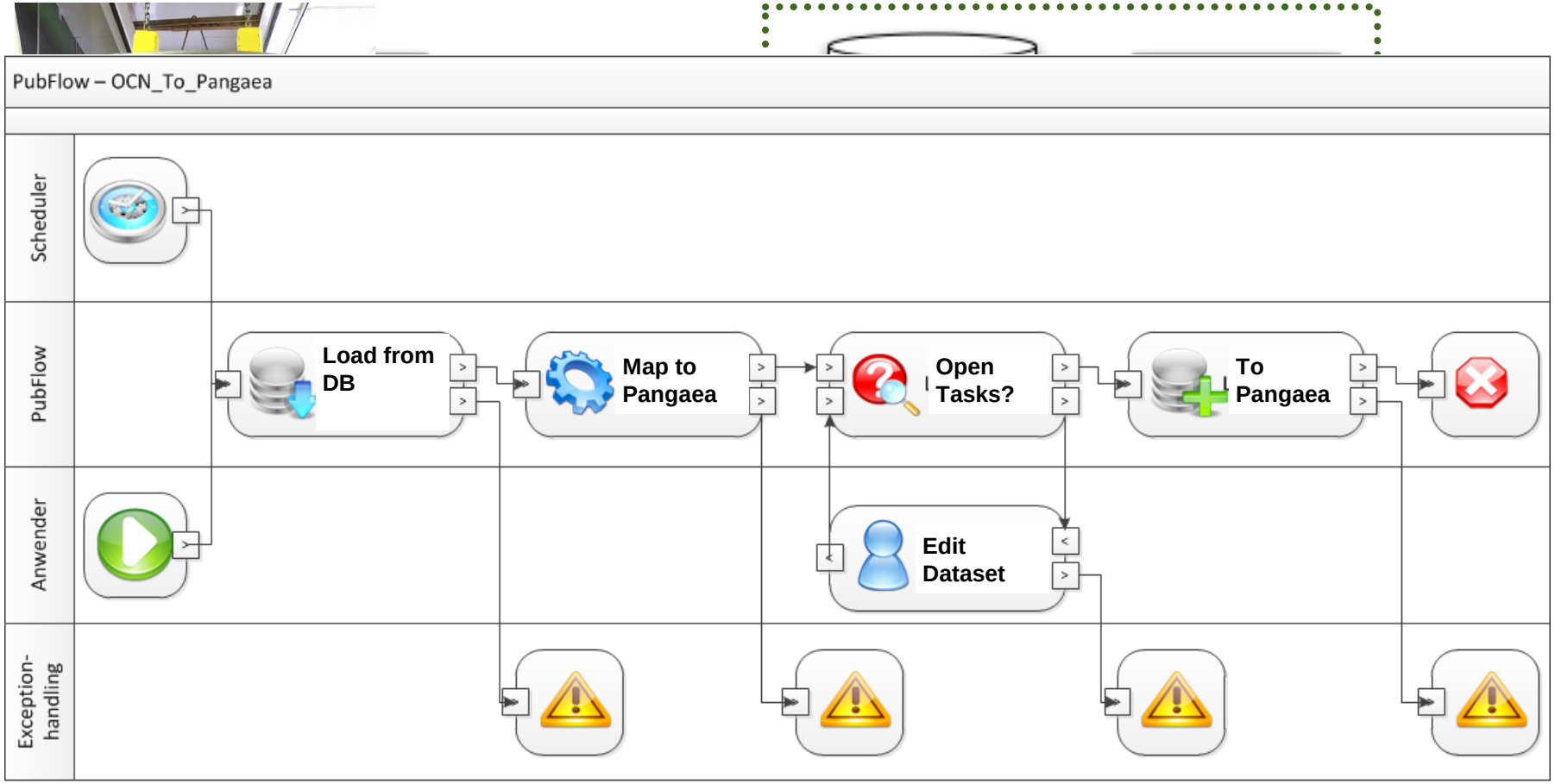




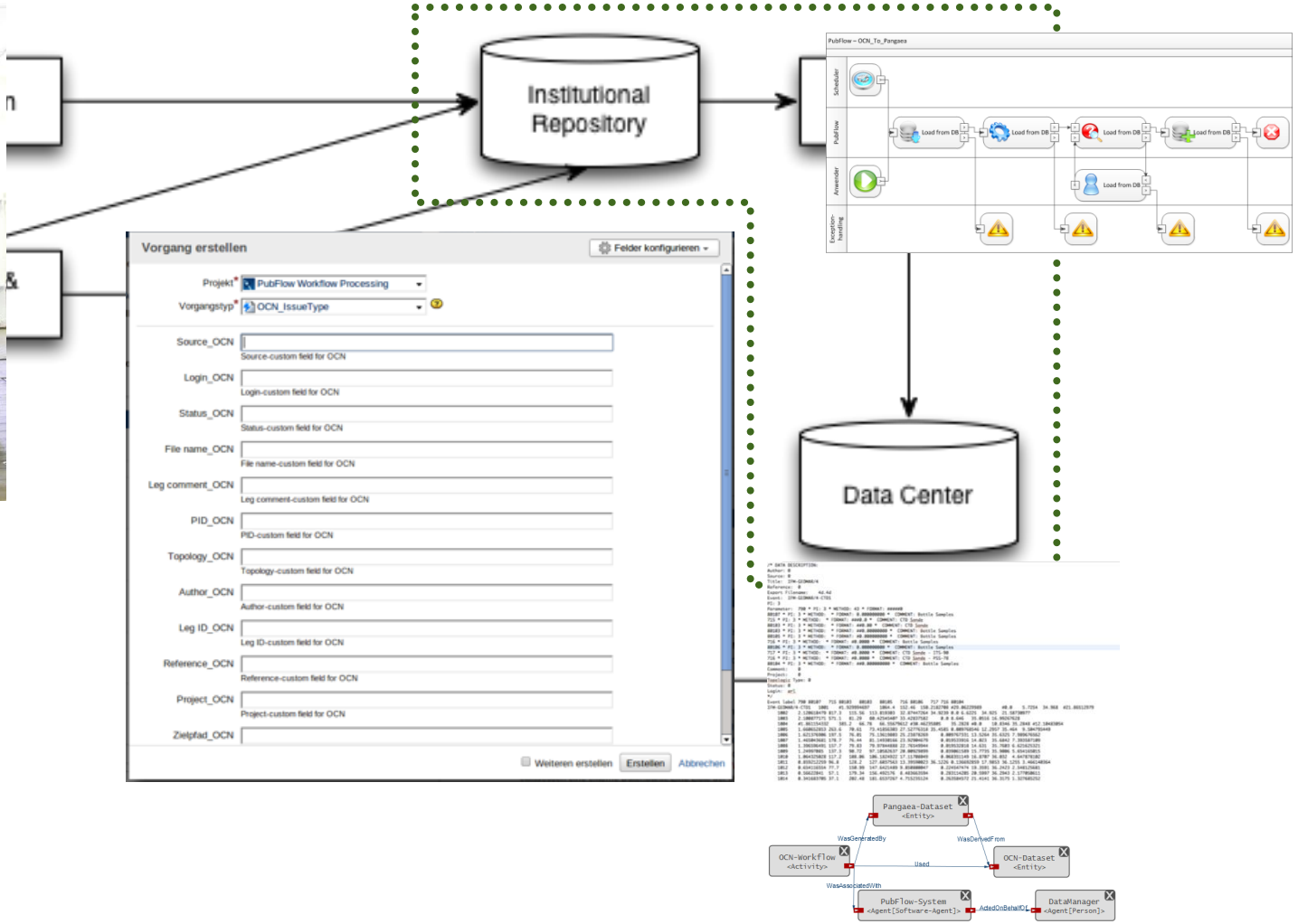
Evaluation Scenario



Evaluation Scenario



Evaluation Scenario

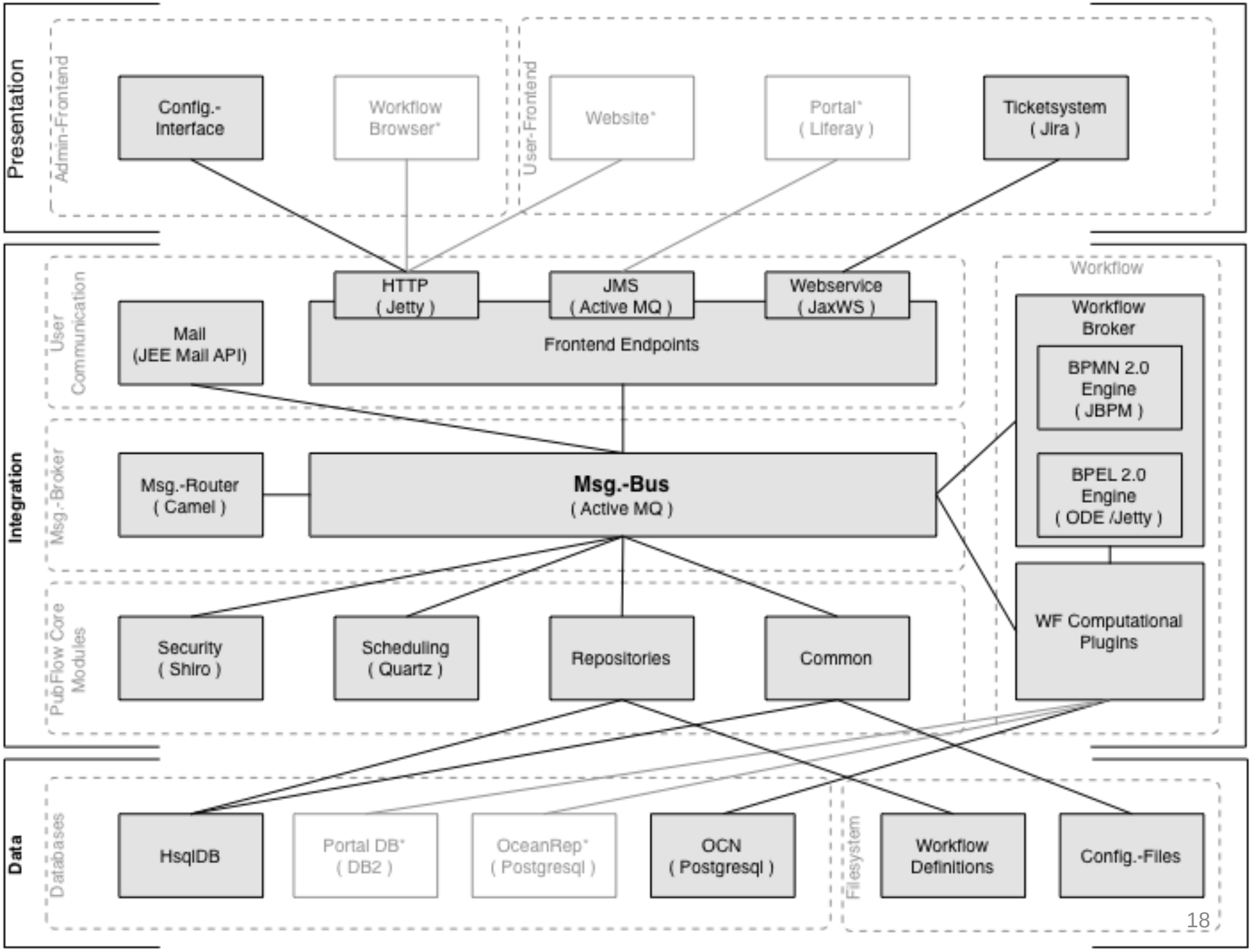


Conclusion

- PubFlow increases the degree of automation in the data publication process
- Is build upon proven workflow technology
- Brings the division of work to data management
- Collects provenance information

PubFlow

www.PubFlow.de


Presentation
Admin-Frontend
Config.-Interface
Workflow Browser*
User-Frontend
Website*
Portal* (Liferay)
Ticketsystem (Jira)
Integration
User Communication
Mail (JEE Mail API)
HTTP (Jetty)
JMS (Active MQ)
Webservice (JaxWS)
Frontend Endpoints
Workflow
Workflow Broker
BPMN 2.0 Engine (JBPM)
BPEL 2.0 Engine (ODE /Jetty)
Msg.-Broker
Msg.-Router (Camel)
Msg.-Bus (Active MQ)
PubFlow Core Modules
Security (Shiro)
Scheduling (Quartz)
Repositories
Common
WF Computational Plugins
Data
Databases
HsqlDB
Portal DB* (DB2)
OceanRep* (Postgresql)
OCN (Postgresql)
Filesystem
Workflow Definitions
Config.-Files