PubFlow
a scientific data publication framework for marine science

Peer Brauer, Prof. Dr. Wilhelm Hasselbring
Software Engineering Group
Kiel University
München, 23.10.2013

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

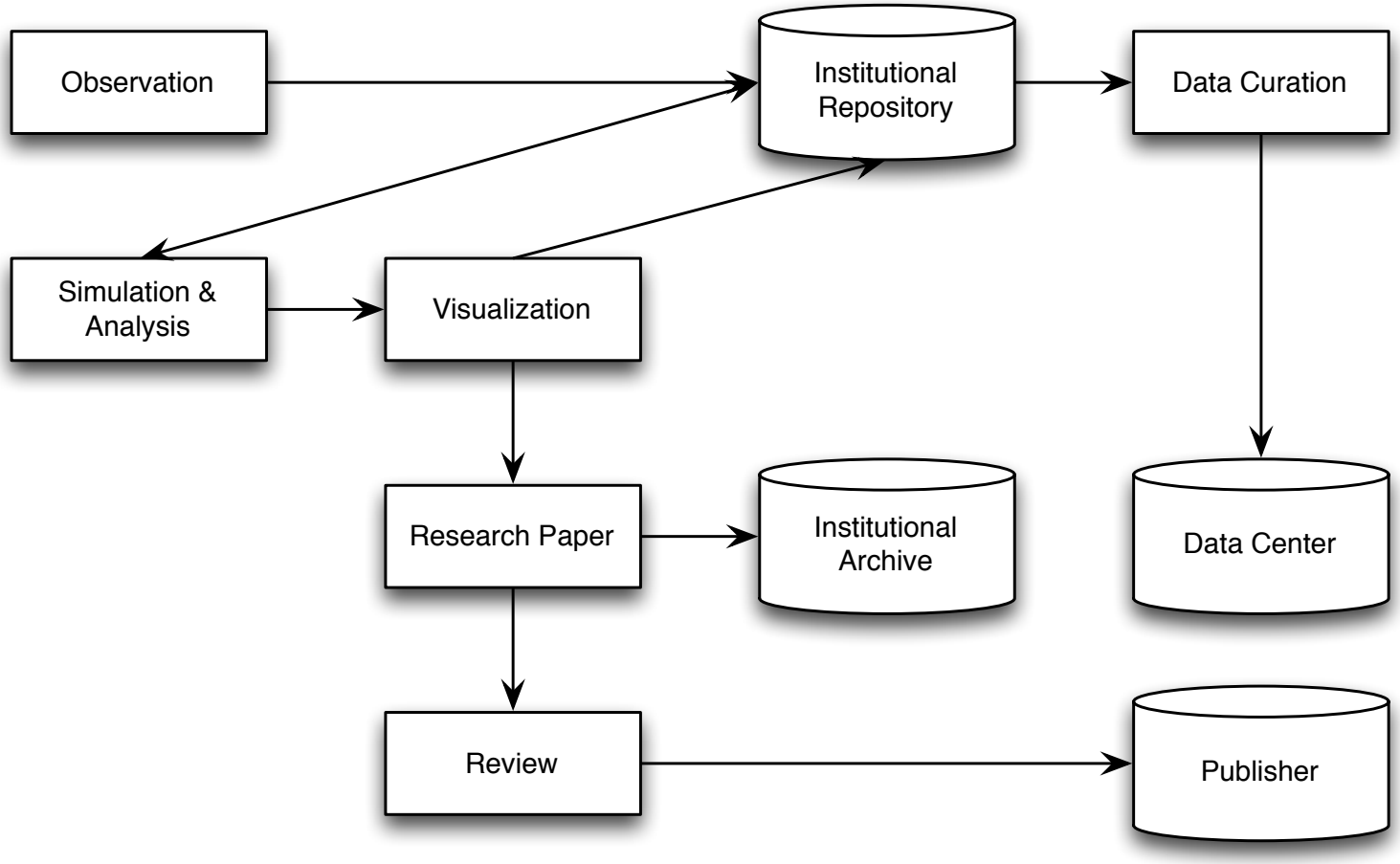
ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

future ocean
KIEL MARINE SCIENCES

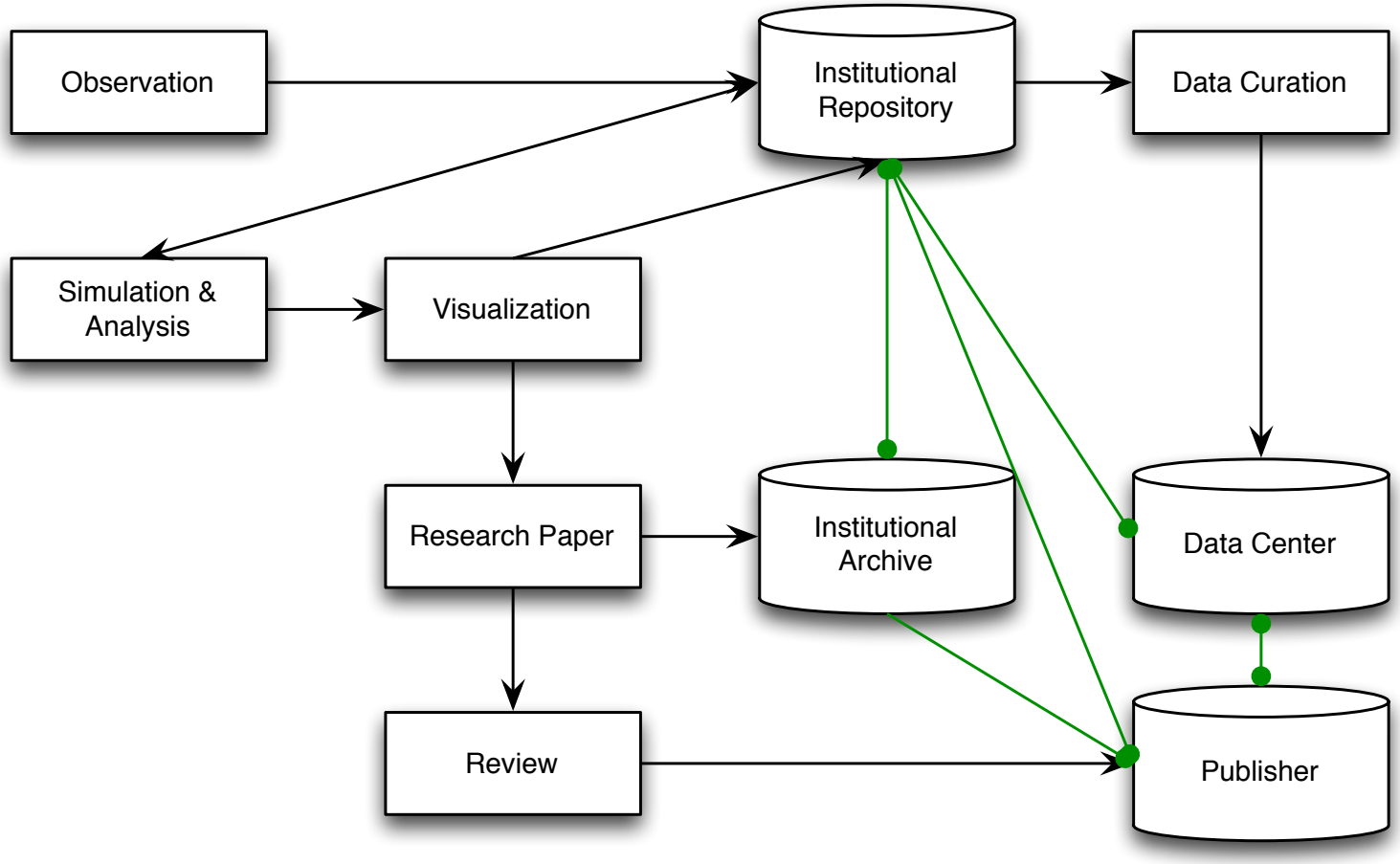
funded by DFG

PubFlow
www.pubflow.de

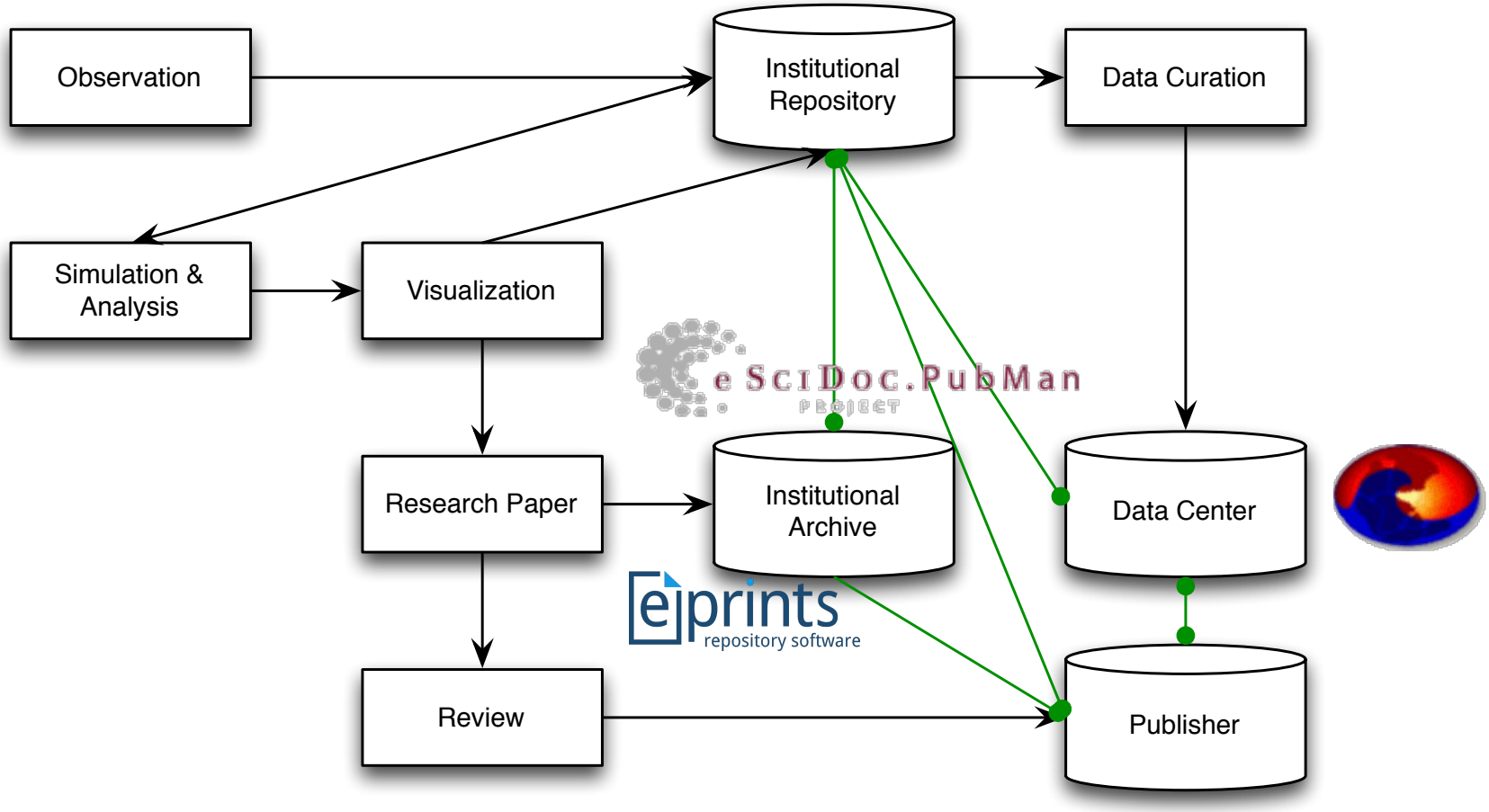
Motivation


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
eSciDoc.PubMan
Project
Research Paper
Institutional Archive
Data Center
eprints
repository software
Review
Publisher

Agenda

PubFlow Framework
Evaluation Scenario
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 in data management

Features

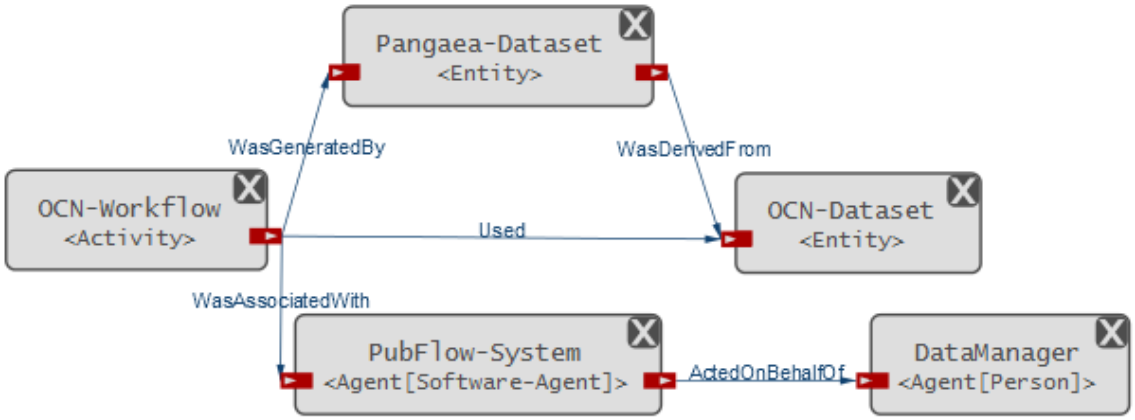
Build upon proven workflow technology

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

Features

Provenance Awareness

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



Features

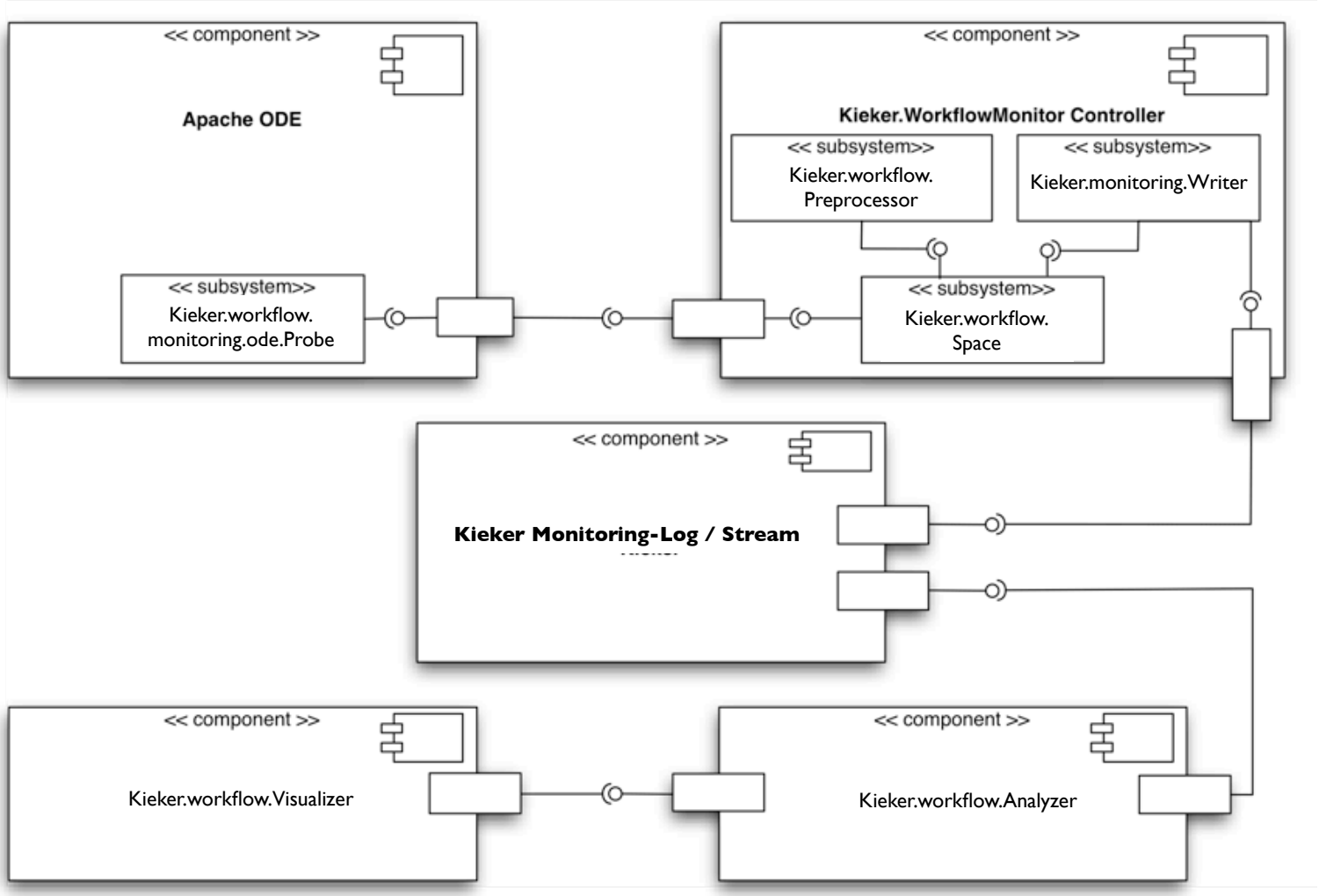
Provenance Awareness

W3C Working Group Note

31	Policy Reasoning Framework	Framework / API	PROV-O
32	Informed Rural Passenger Information Infrastructure	Application	PROV-O
33	PubFlow Provenance Archive	Application Framework / API	PROV-O, PROV-XML
34	PROV Python library	Framework / API	PROV-N, PROV-JSON
35	csv2rdf4lod-automation	Application	PROV-O
36	recoprov	Application	PROV-O, PROV-N
37	DataFAQs	Application	PROV-O
38	provx2o	Application	PROV-O, PROV-XML
39	Hedgehog	Application	PROV-XML
40	**QuerioCity research prototype**	Application Framework / API Service	PROV-O
41	Tinga Provenance Service	Service	PROV-O, PROV-JSON
42	Human Computation ontology	Vocabulary Extension	PROV-O
43	tavernaprov	Vocabulary Extension	PROV-O
44	The Open Provenance Model for Workflows (OPMW)	Vocabulary Extension	PROV-O

Features

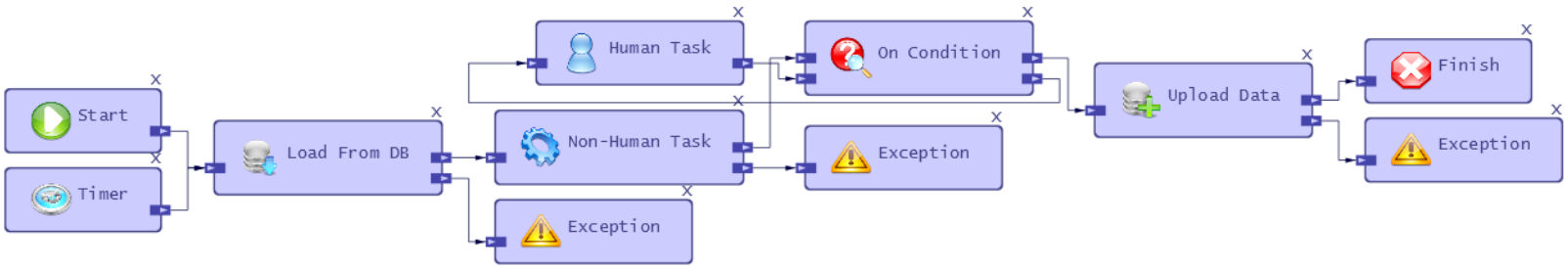
Provenance Awareness

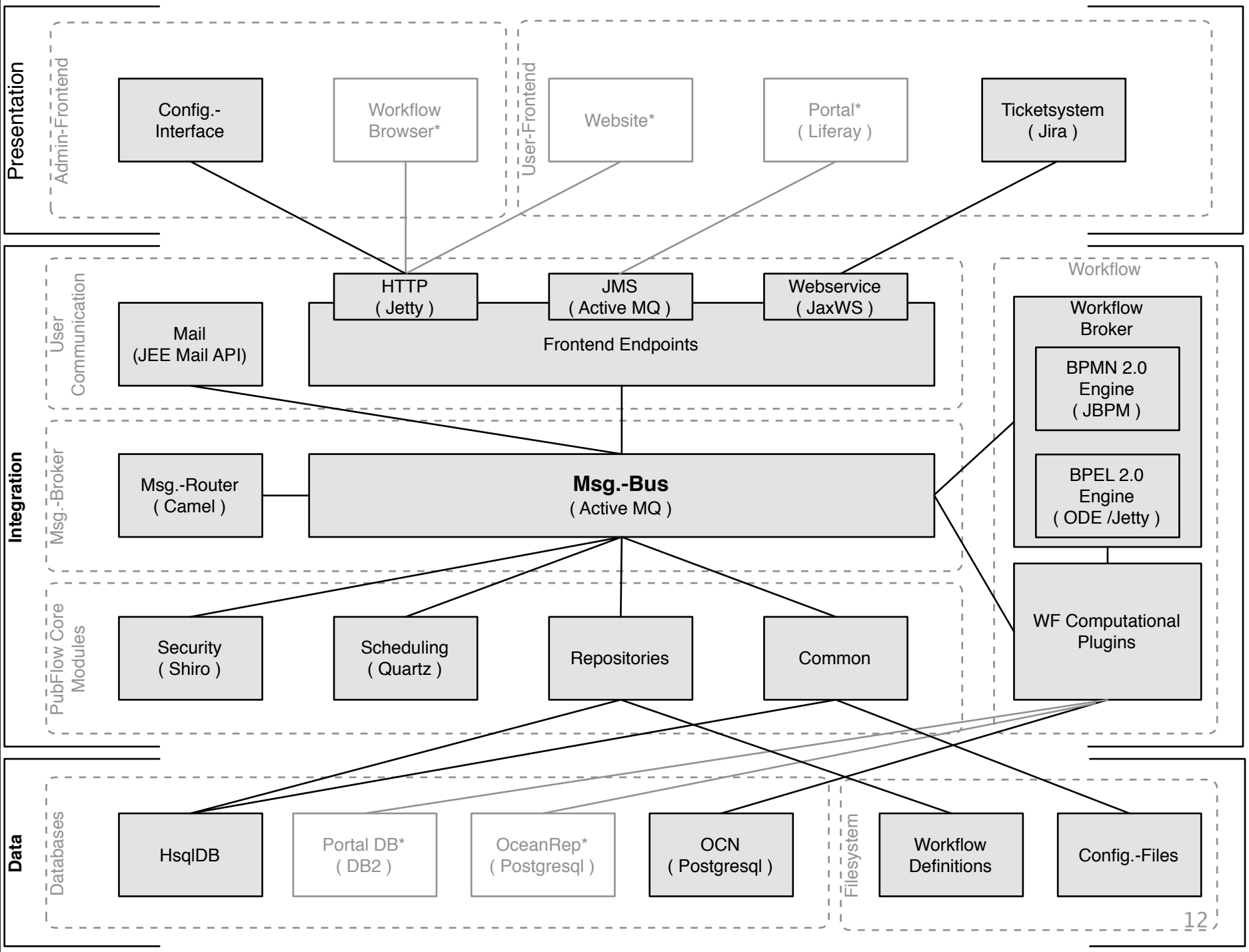


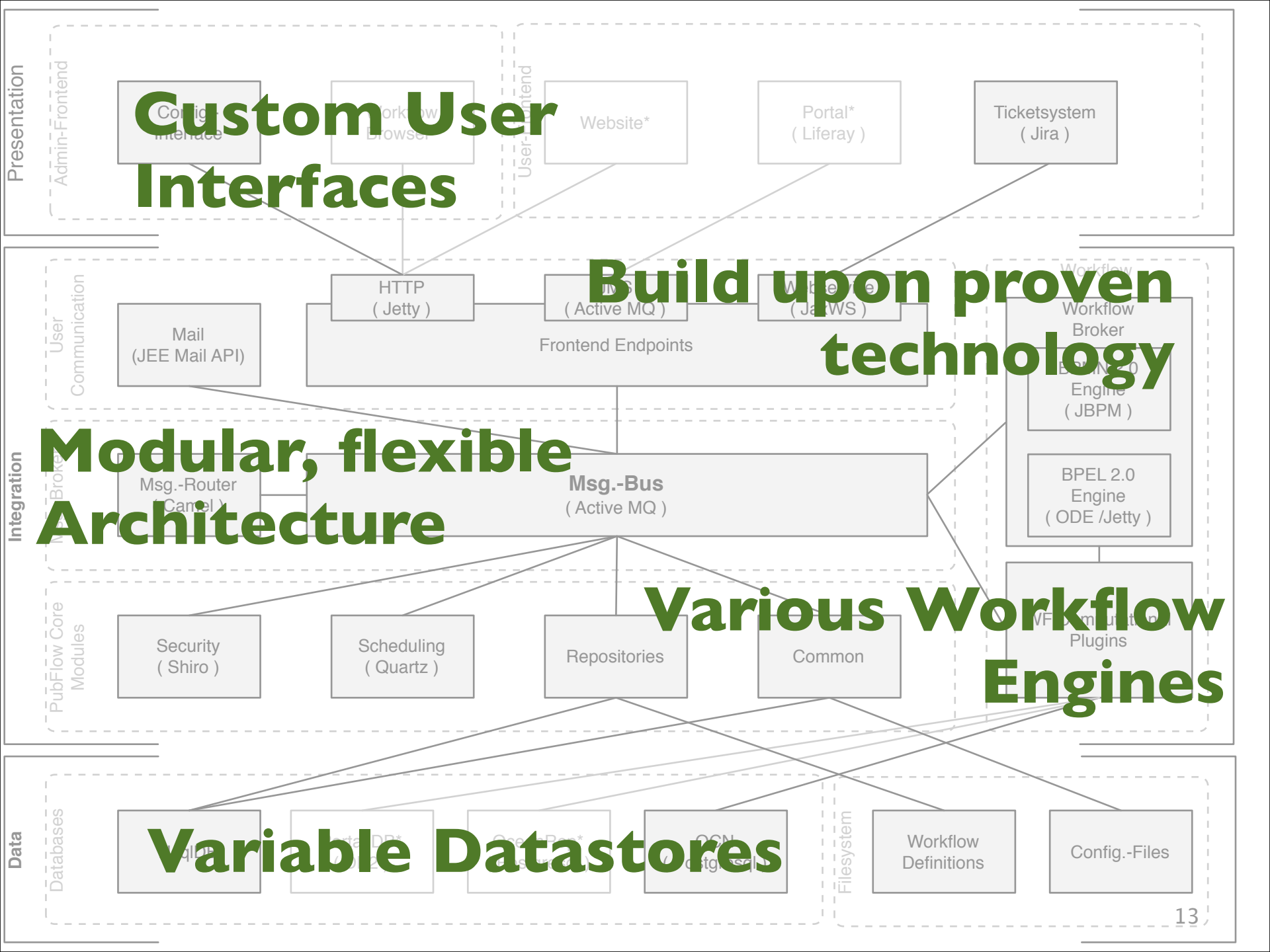
Features

Graphical Workflow Editor

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

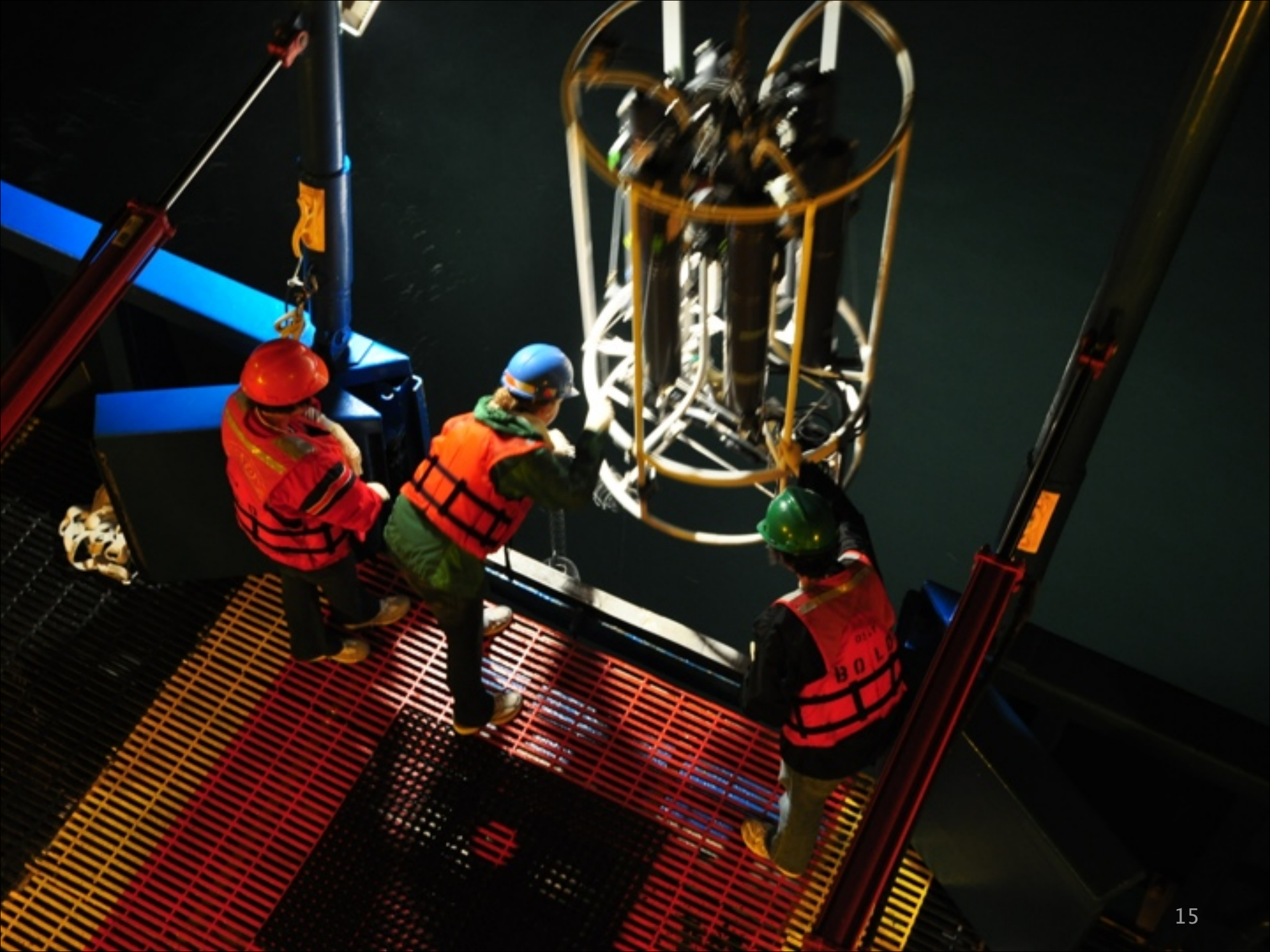



Presentation
Admin-Frontend
Config.-Interface
Workflow Browser*
User-Frontend
Website*
Portal* (Liferay)
Ticketsystem (Jira)
Integration
User Communication
Mail (JEE Mail API)
HTTP (Jetty)
JMS (Active MQ)
Webservice (JaxWS)
Frontend Endpoints
Workflow
Workflow Broker
BPMN 2.0 Engine (JBPM)
BPEL 2.0 Engine (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

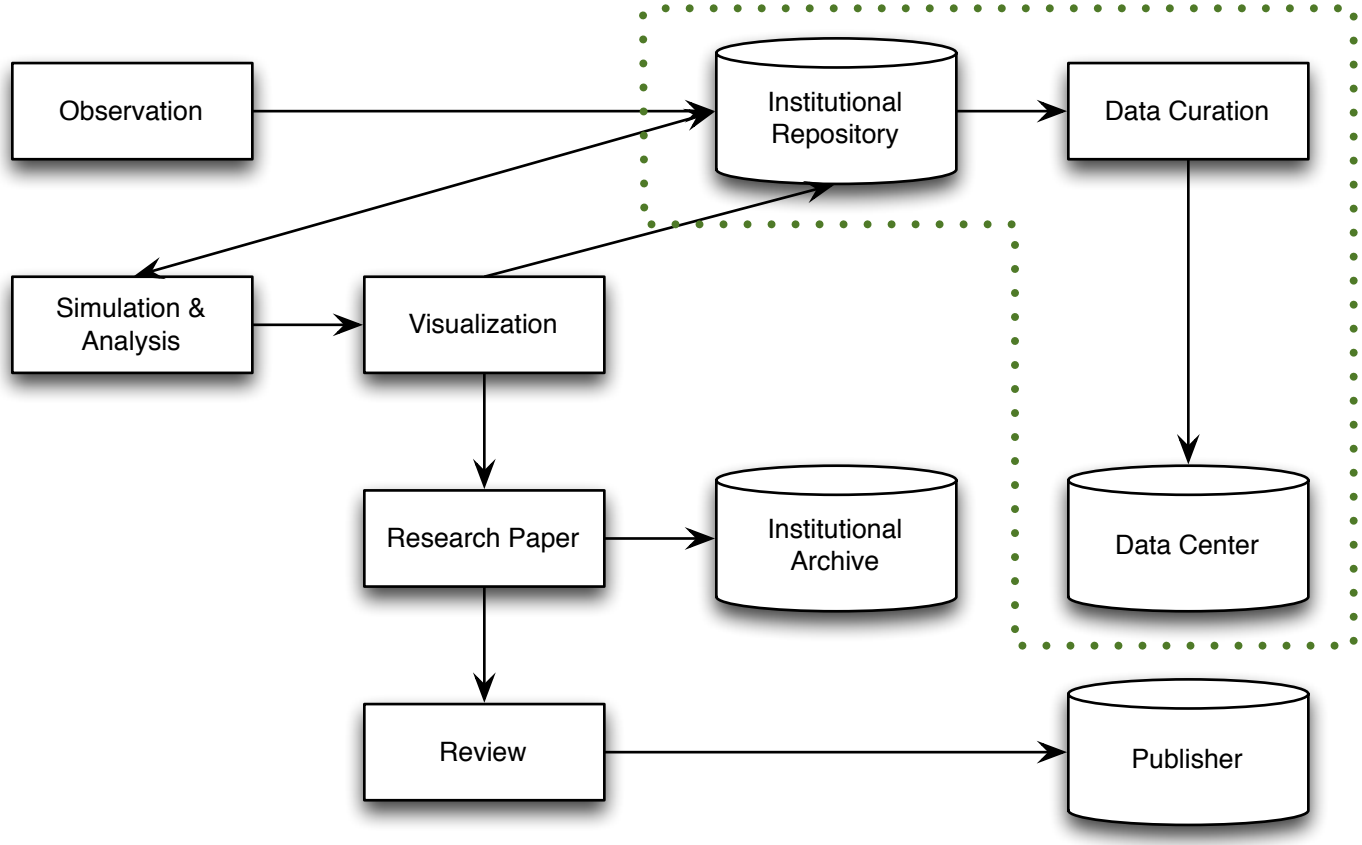

Custom User Interfaces
Build upon proven technology
Modular, flexible Architecture
Various Workflow Engines
Variable Datastores
Presentation
Admin-Frontend
User-Frontend
Website*
Portal*
(Liferay)
Ticketsystem
(Jira)
Integration
User Communication
Mail
(JEE Mail API)
HTTP
(Jetty)
(Active MQ)
Frontend Endpoints
Workflow Broker
Engine
(JBPM)
BPEL 2.0
Engine
(ODE /Jetty)
Msg.-Router
(Camel)
Msg.-Bus
(Active MQ)
PubFlow Core Modules
Security
(Shiro)
Scheduling
(Quartz)
Repositories
Common
Plugins
Data
Databases
Filesystem
Workflow
Definitions
Config.-Files

Section 2

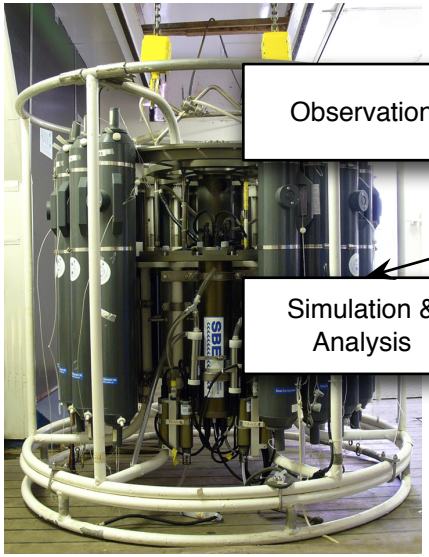
The Evaluation Scenario

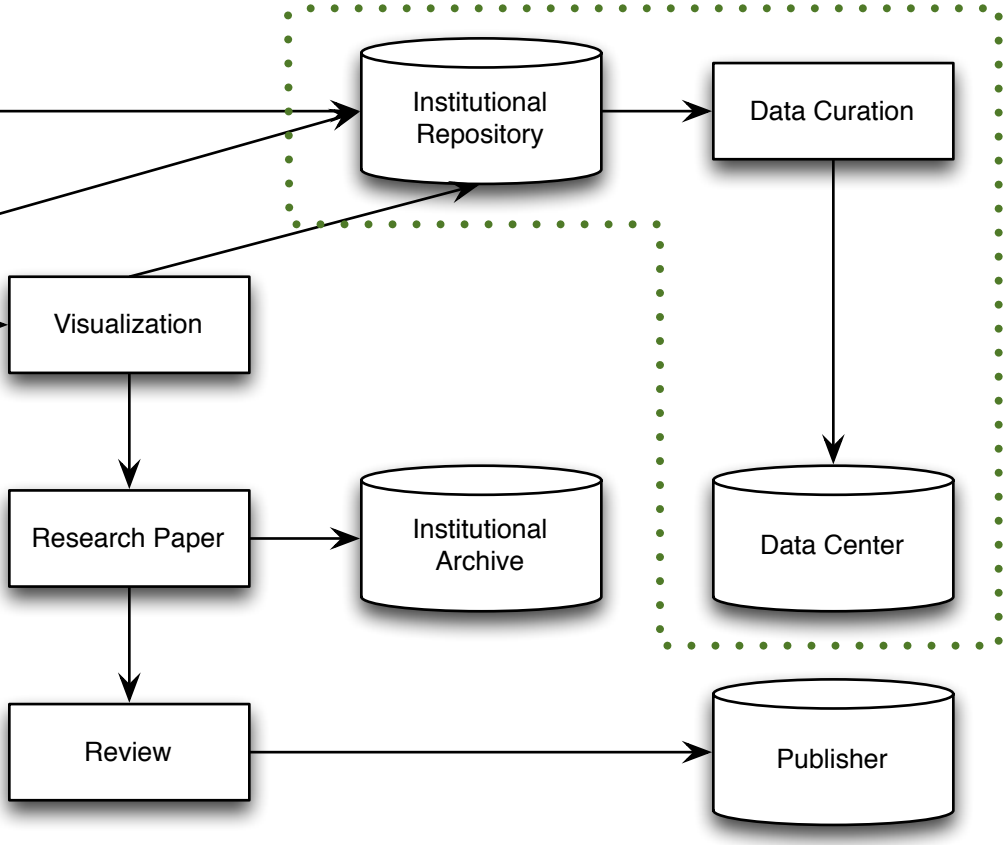


Evaluation Scenario

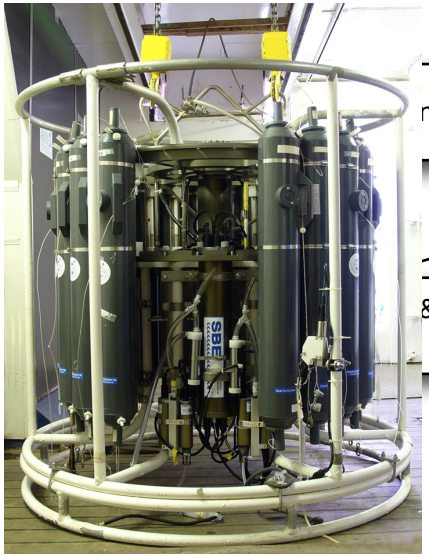


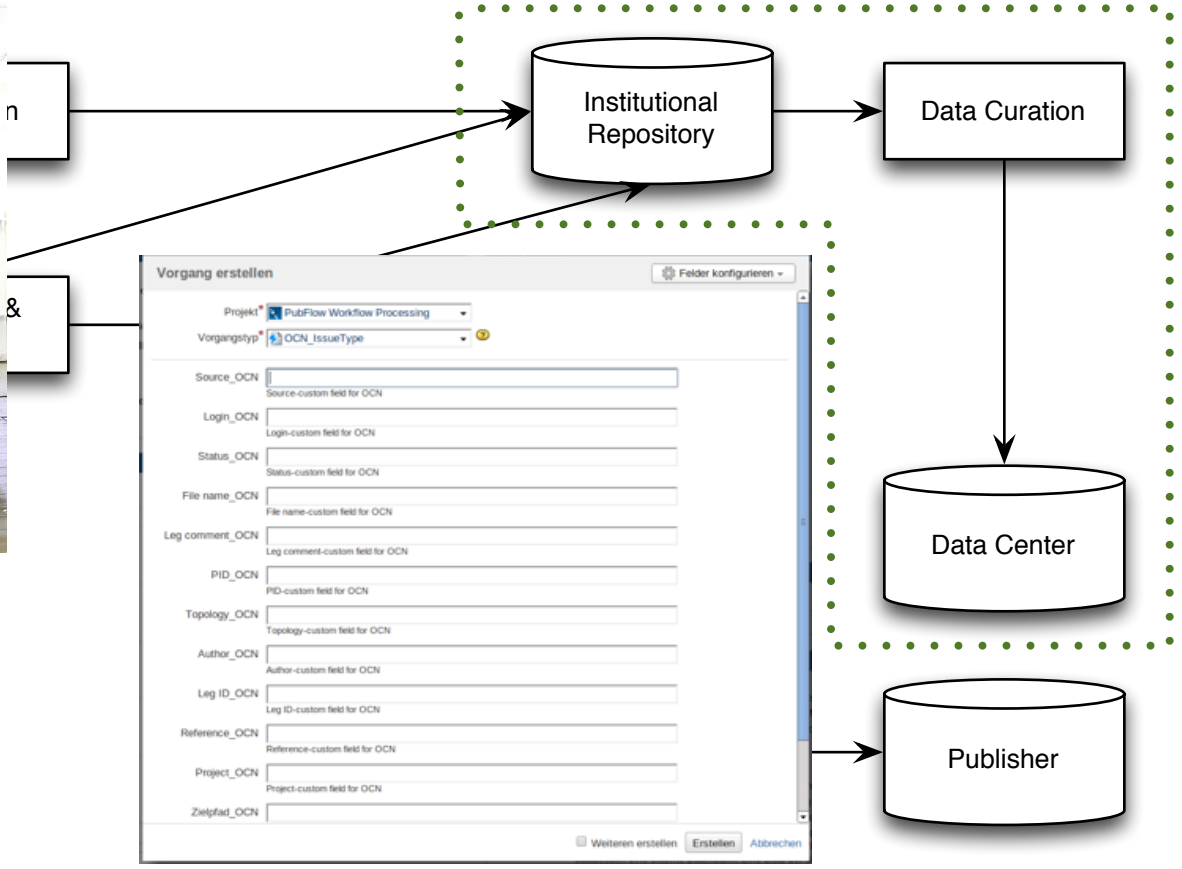
Evaluation Scenario



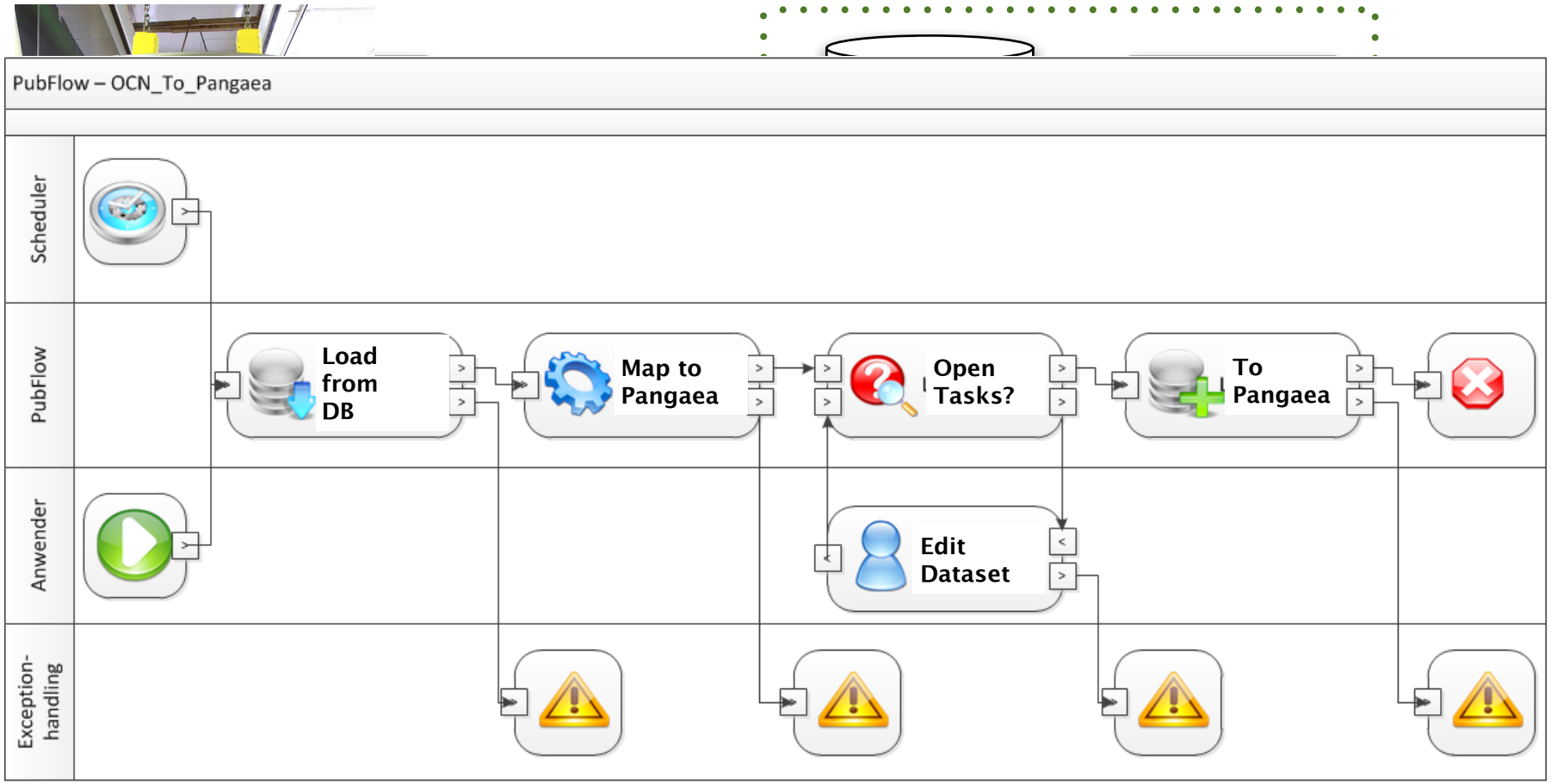


Evaluation Scenario

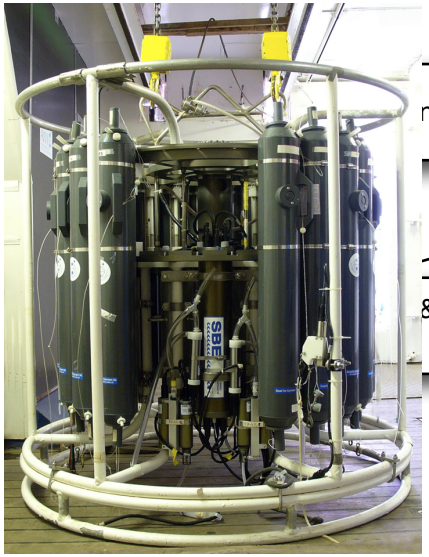


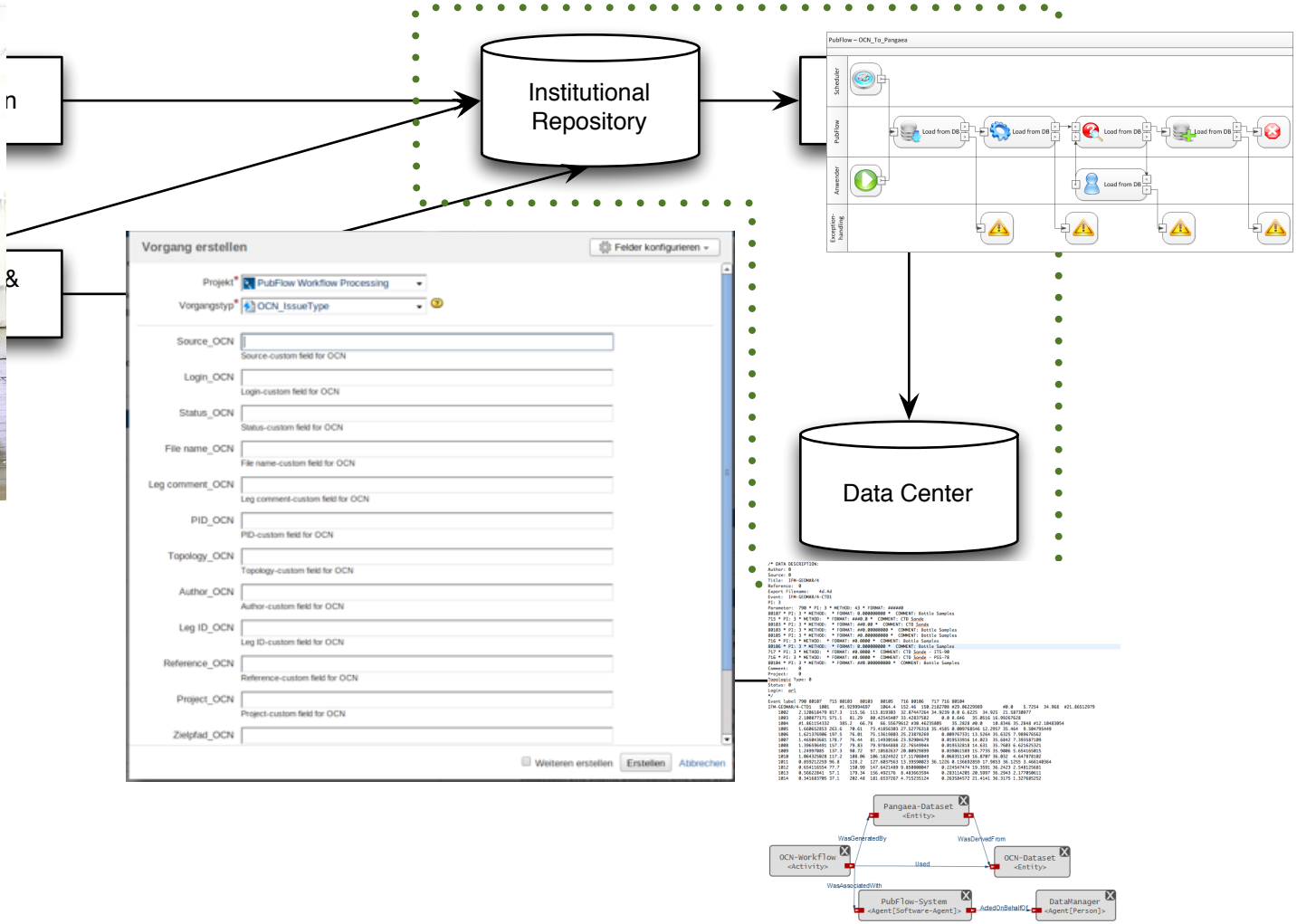


Evaluation Scenario



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