Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Workflow-basierte Verarbeitung und Archivierung von Ozeanbeobachtungsdaten

Prof. Dr. Wilhelm (Willi) Hasselbring

Lehrstuhl Software Engineering
http://se.informatik.uni-kiel.de/

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Kompetenzverbund
Software Systems Engineering (KoSSE)
http://kosse-sh.de/

PubFlow

Kieler Marktplatz, 18.06.2013

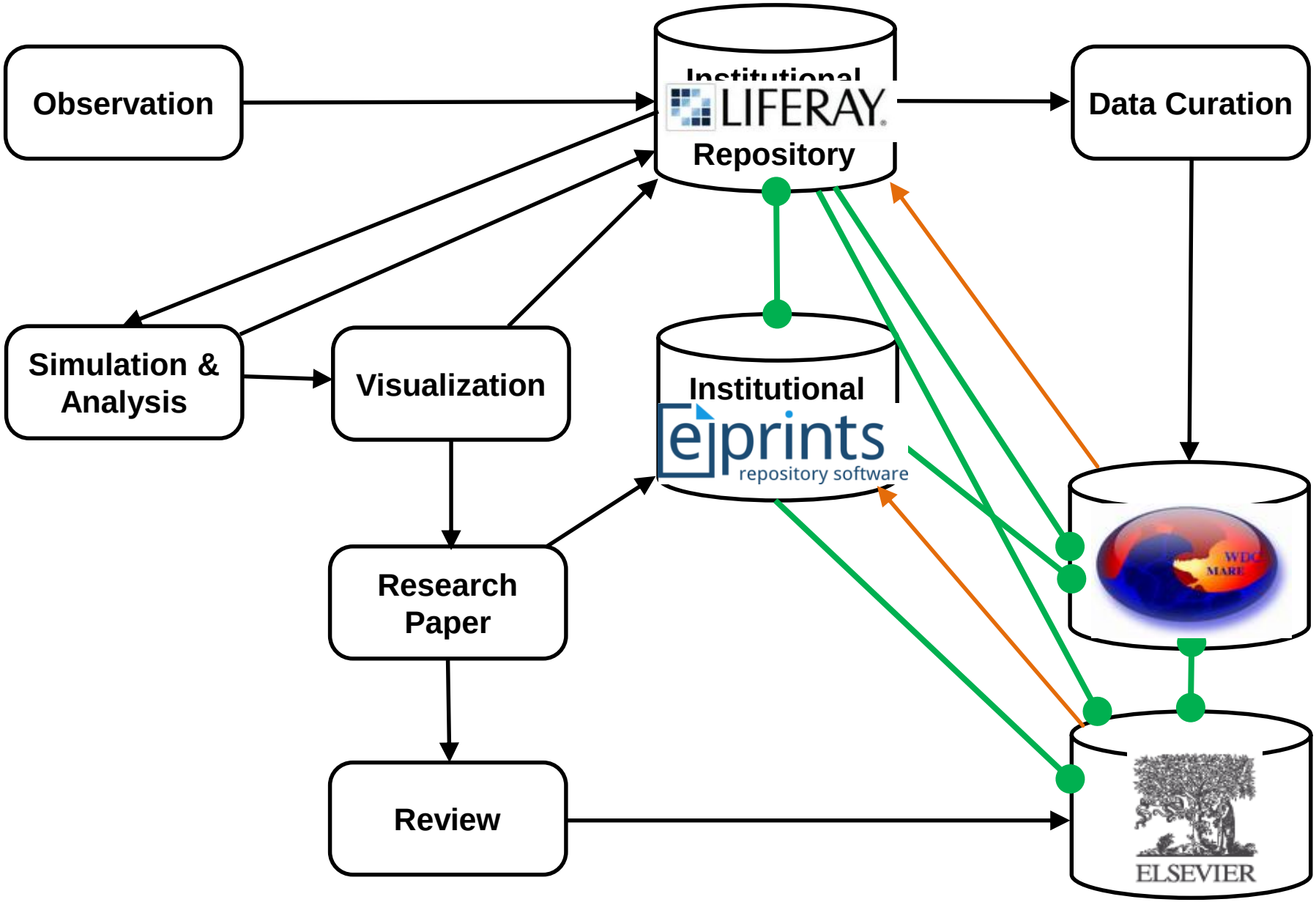
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Agenda

1. Publikations-Workflows für Forschungsdaten
2. CTD-Datenverarbeitung in PubFlow als Beispiel
3. Provenienzdaten-Erfassung
4. Zusammenfassung und Ausblick

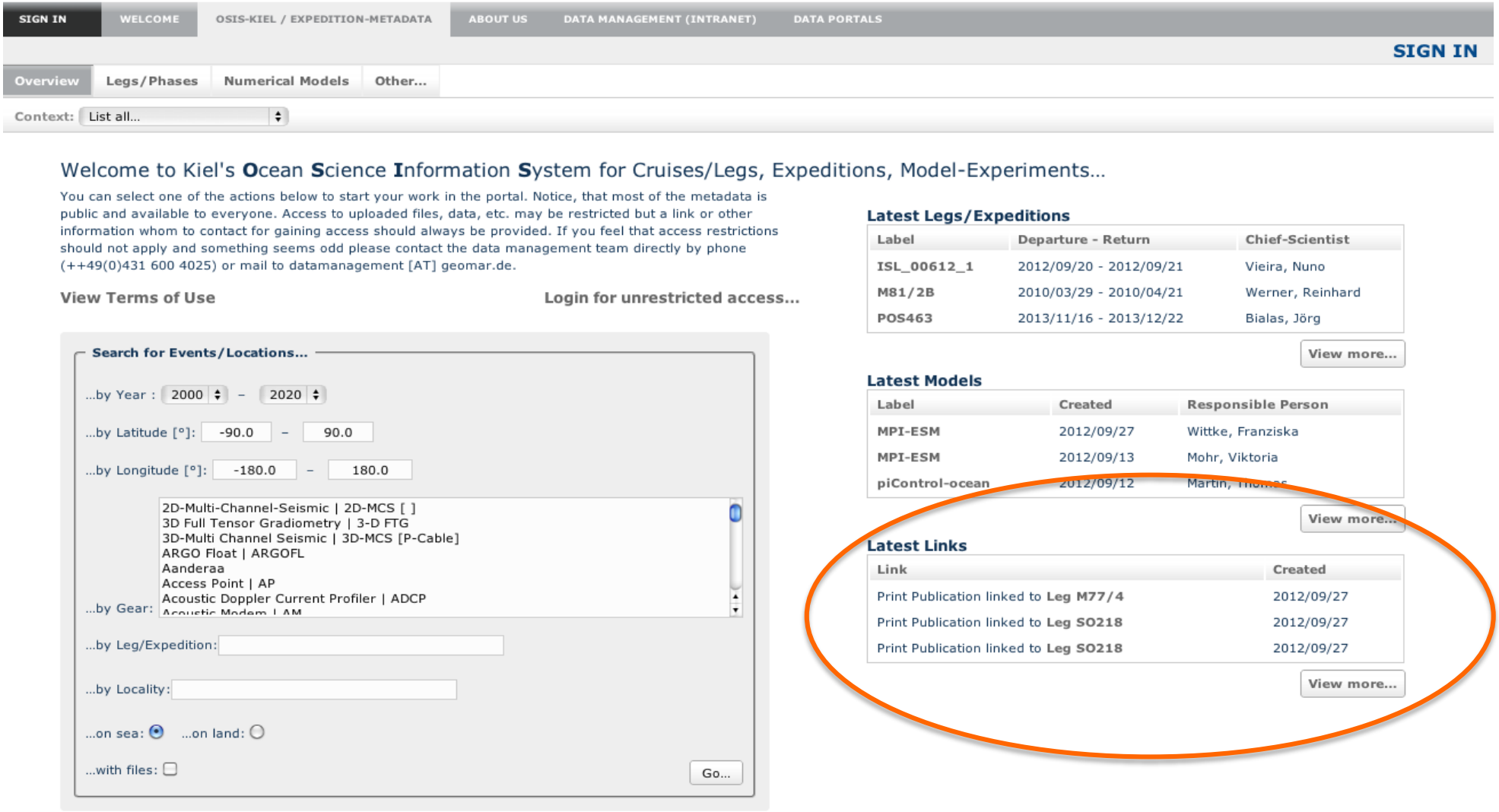
Data and Paper Flow (in Ocean Science)

Christian-Albrechts-Universität zu Kiel
Technische Fakultät



Kiel Data Management Infrastructure – Ocean Science Information System

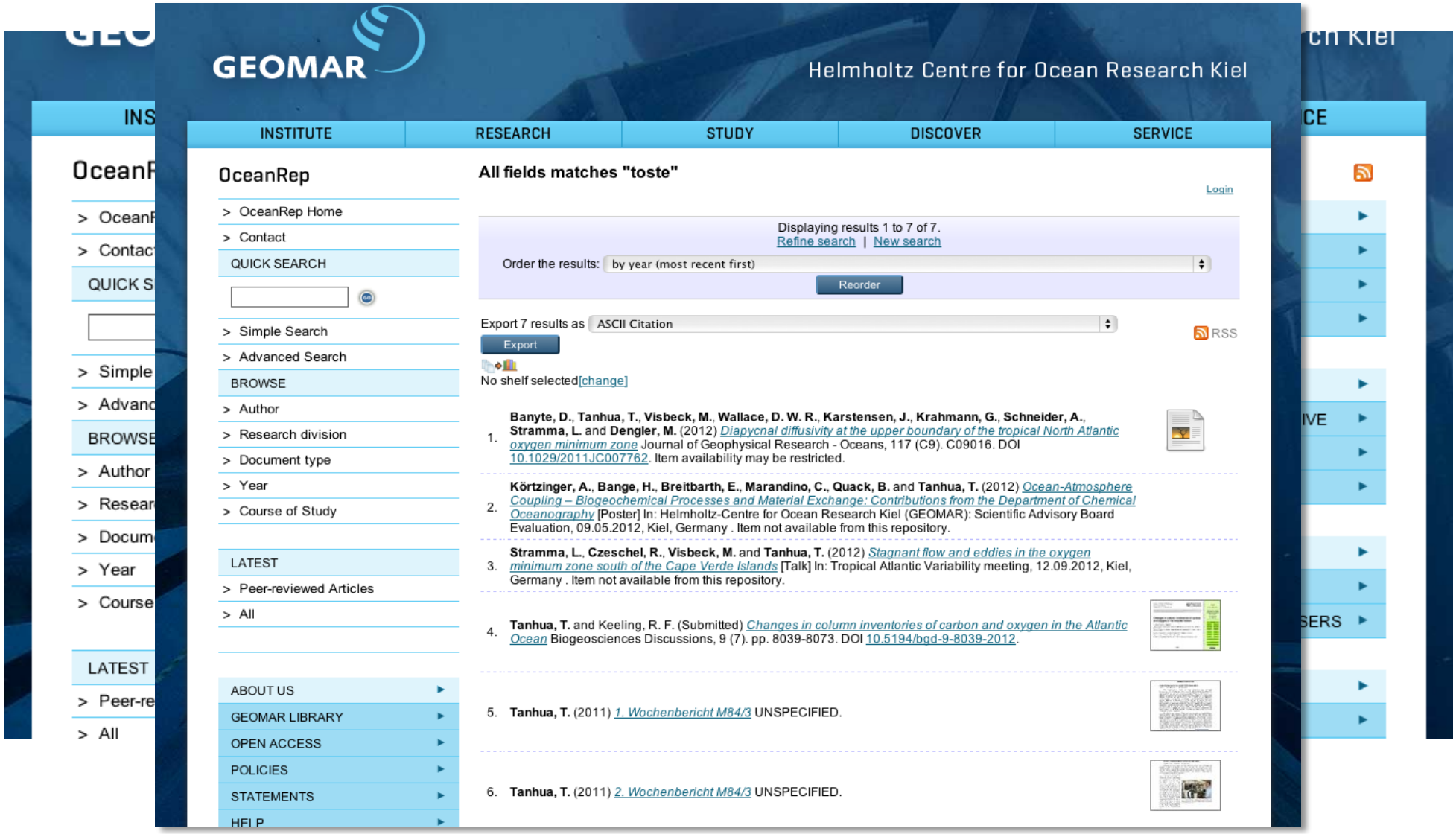
Christian-Albrechts-Universität zu Kiel
Technische Fakultät



This, and the following slides were provided by the Kiel Data Management Team.

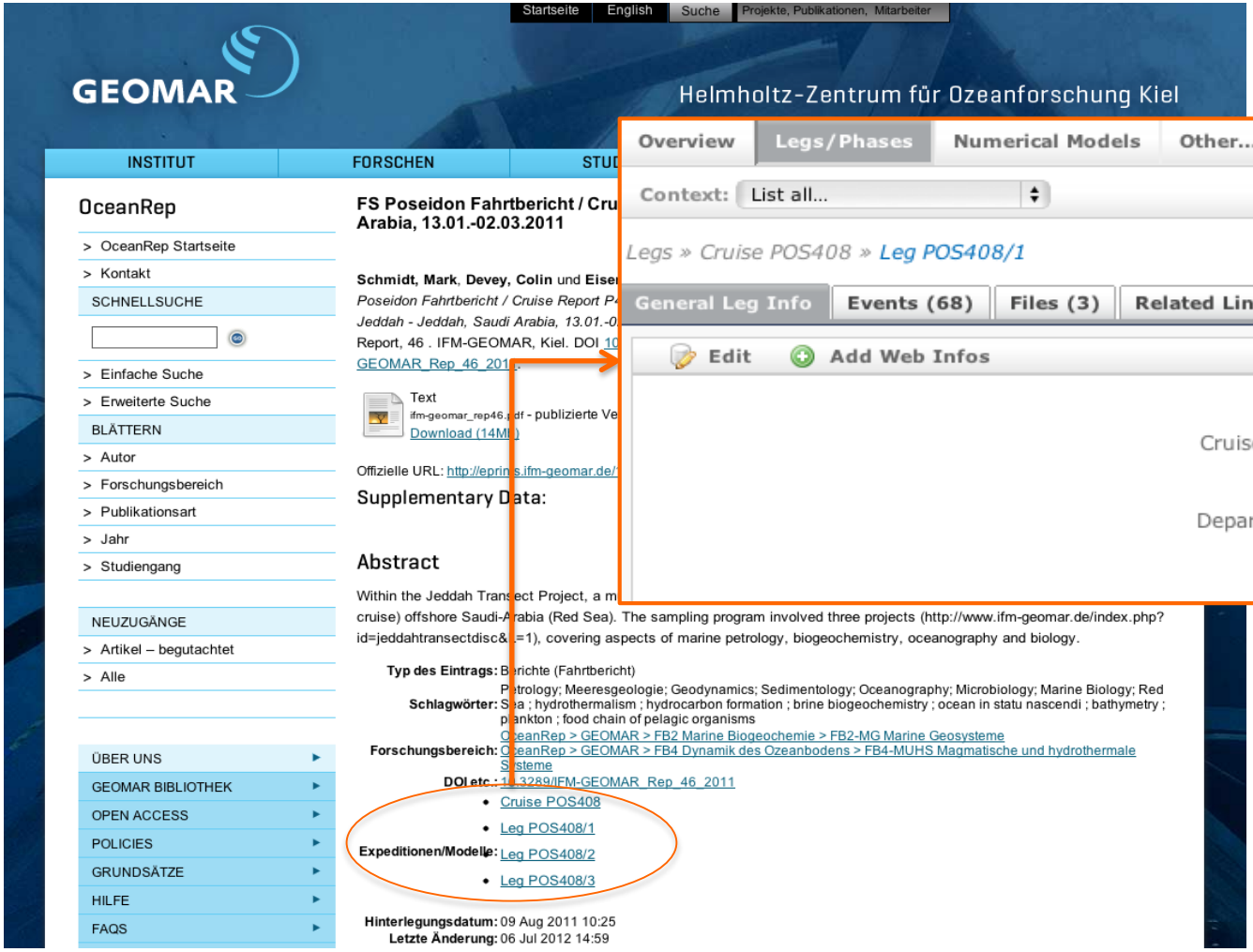
GEOMAR – OceanRep

Christian-Albrechts-Universität zu Kiel
Technische Fakultät



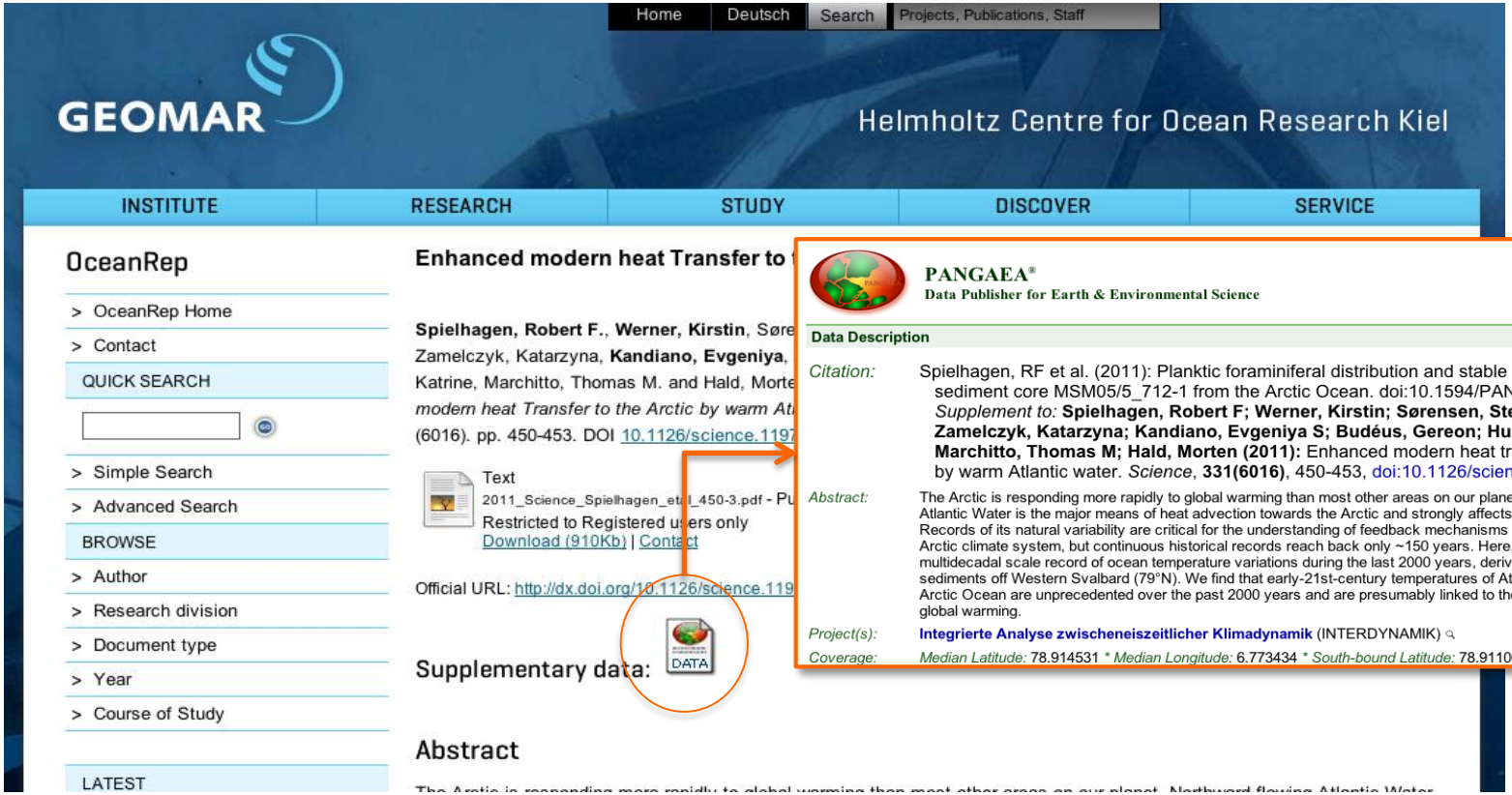
GEOMAR – OceanRep link to data

Christian-Albrechts-Universität zu Kiel
Technische Fakultät



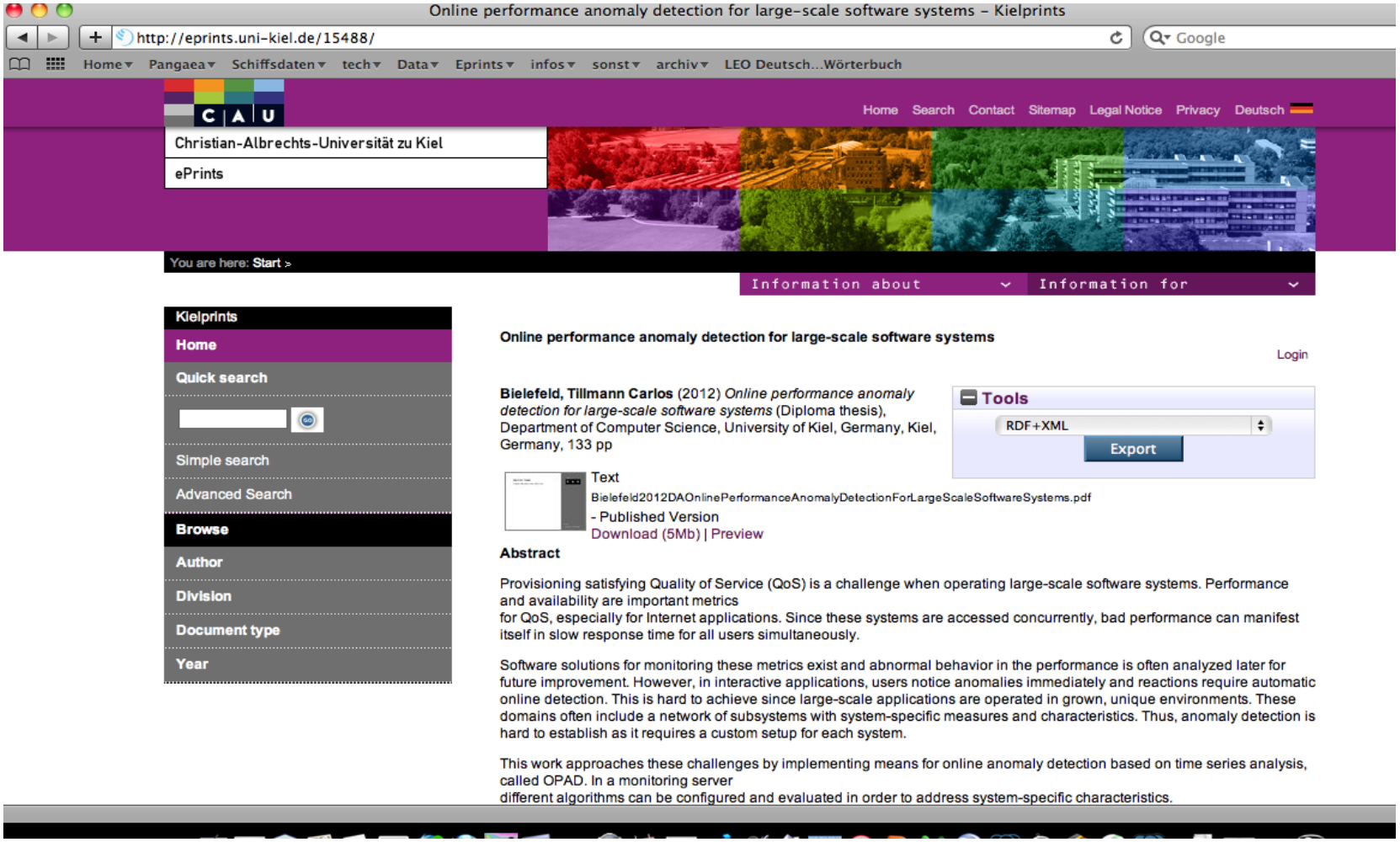
GEOMAR – OceanRep link to Pangaea

Christian-Albrechts-Universität zu Kiel
Technische Fakultät



Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Kielprints



Dr. Leslie Carr, scientific director of the Eprints project at the University of Southampton, gives a Kieler Woche Talk: Friday, June 28th ath 10:00 h.

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Agenda

1. Publikations-Workflows für Forschungsdaten
2. **CTD-Datenverarbeitung in PubFlow als Beispiel**
3. Provenienzdaten-Erfassung
4. Zusammenfassung und Ausblick

PubFlow Project Consortium

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Funded:

Software Engineering Group, University Kiel

Associated:

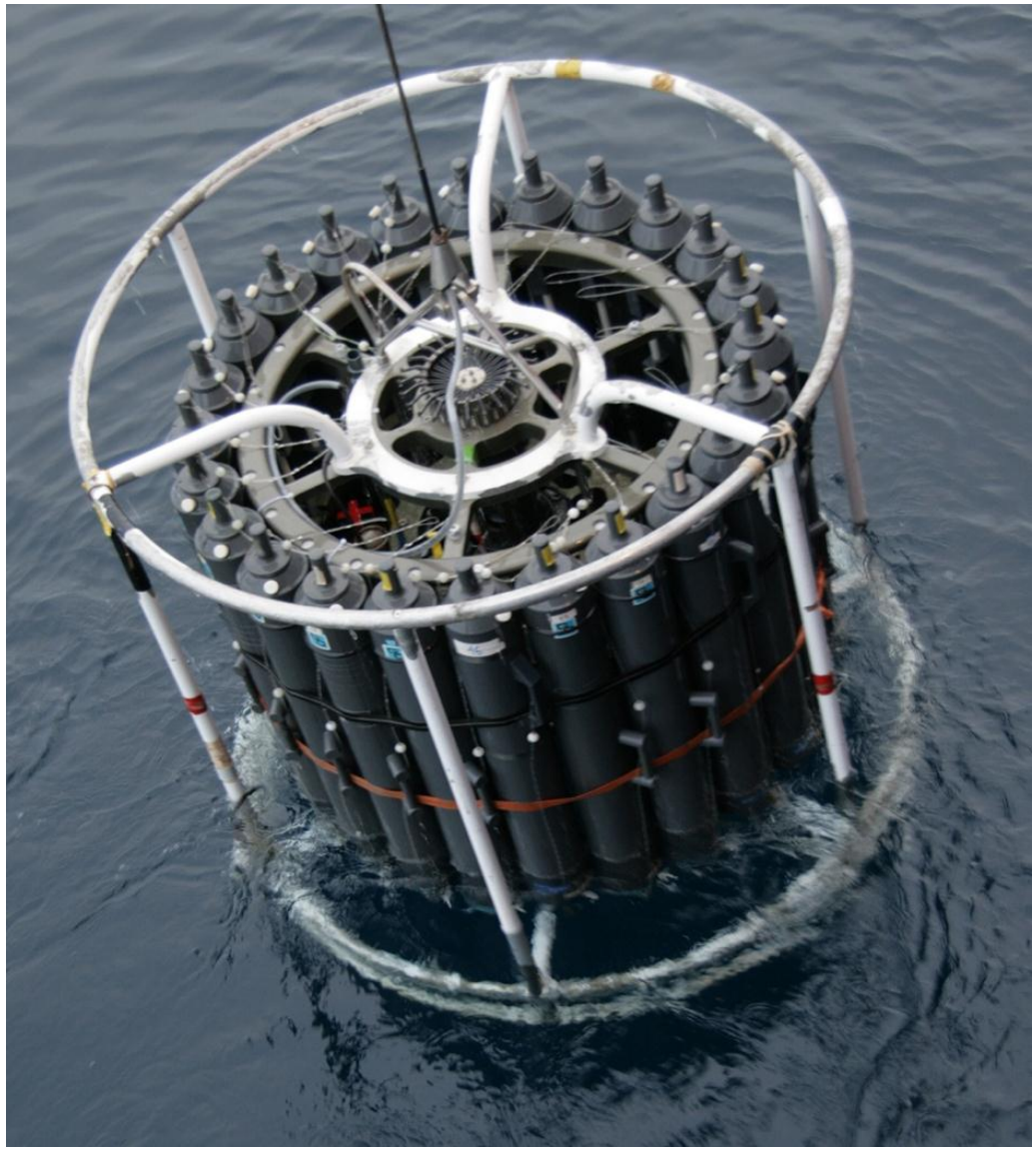
- Excellence cluster “Future Ocean”
- Data and computing center of GEOMAR
- **Library** of GEOMAR
- Computing center of University Kiel
- **Library** of University Kiel
- ZBW
  German National **Library** of Economics -
  Leibniz Information Centre for Economics

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

The initial PubFlow case study

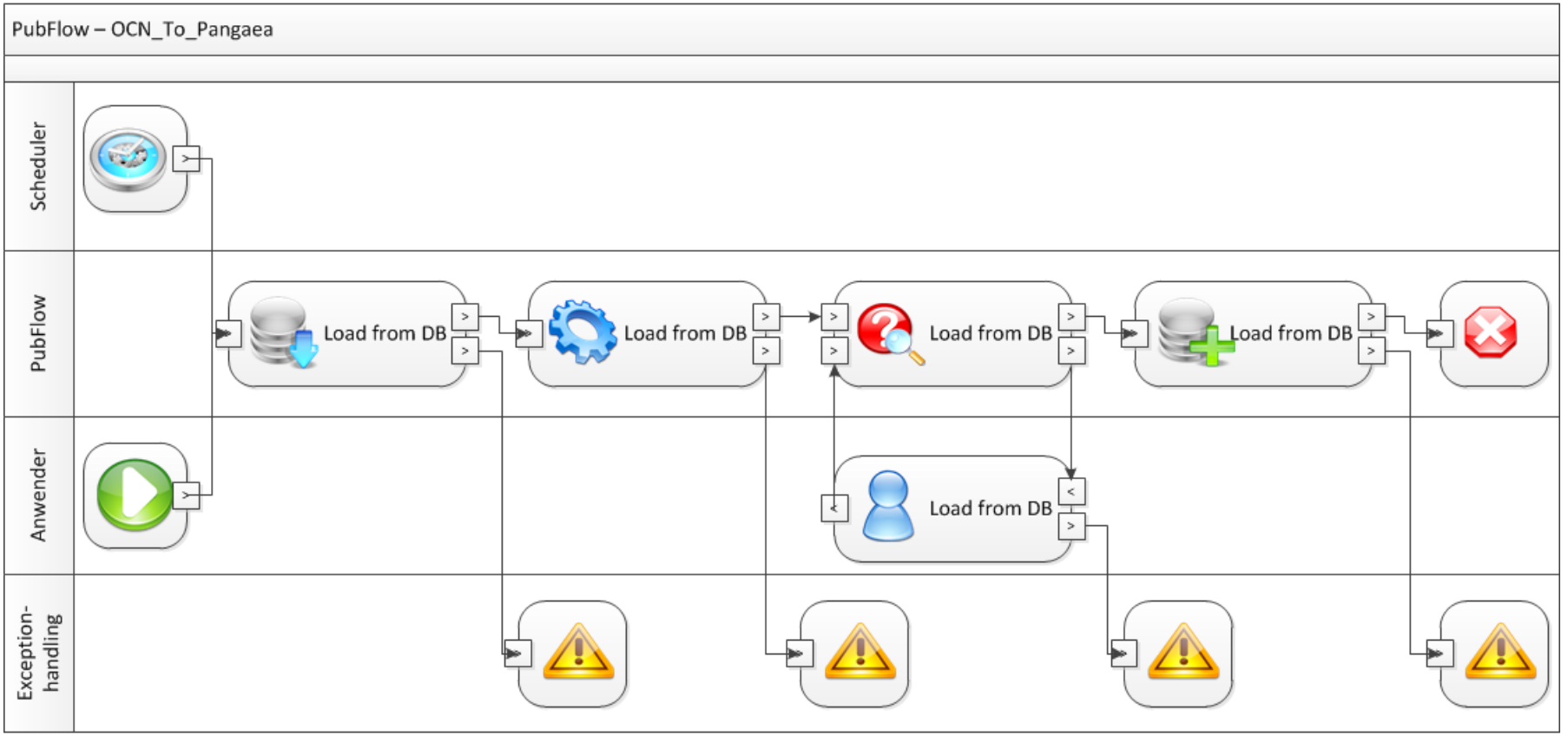
CTD Rosette System

(Conductivity / Temperature / Depth/pressure sensor).



Christian-Albrechts-Universität zu Kiel
Technische Fakultät

CTD Workflow



Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Agenda

1. Publikations-Workflows für Forschungsdaten
2. CTD-Datenverarbeitung in PubFlow als Beispiel
3. **Provenienzdaten-Erfassung**
4. Zusammenfassung und Ausblick

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Collecting provenance data

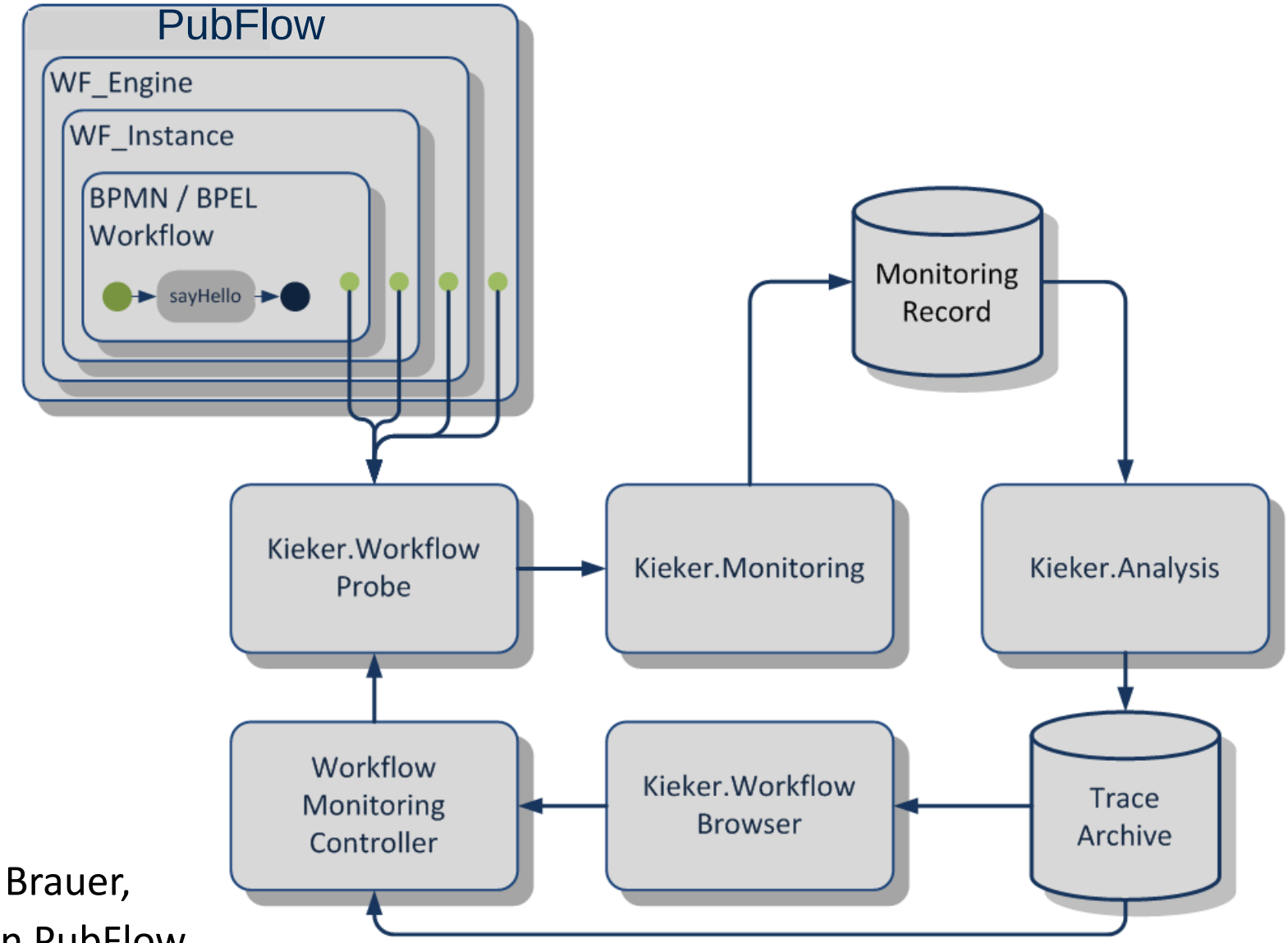
Provenance information describes the origins and the history of (research) data in its life cycle, and the process by which it arrived:

- A description of all modifications and changes
- A list of all methods applied
- A list of all systems and persons involved
- Answers to
  - When,
  - where,
  - how,
  - why.
- Given the documented history of an object, the object attains an **authority** that allows scholars to appreciate its importance, whereas, in the absence of such history, the object may be treated with some skepticism.
- Usually collected and used in databases, data warehouses, workflow systems, scientific data processing, Web services, etc.

14

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

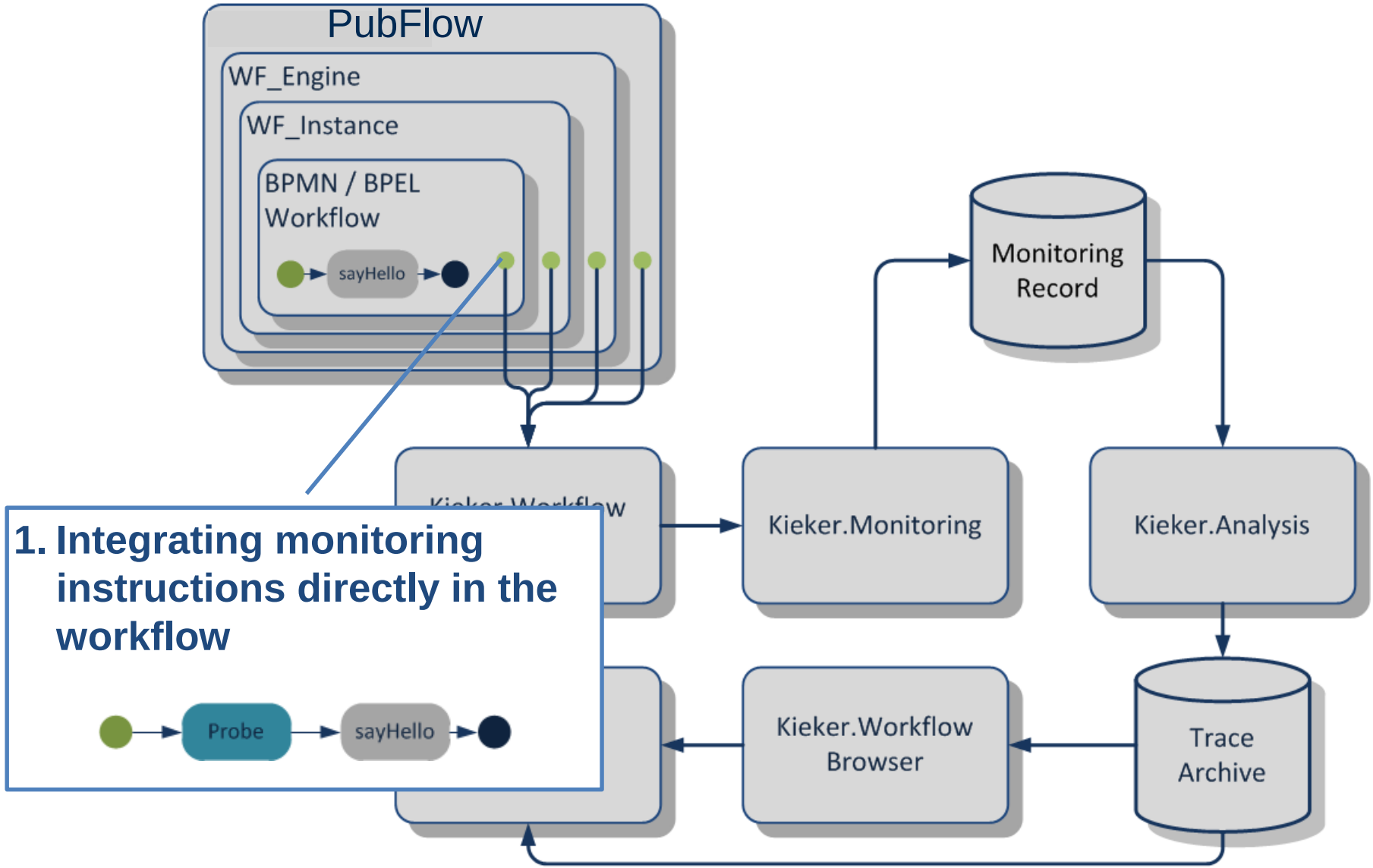
Provenienzdaten-Erfassung in PubFlow



Quelle: Peer Brauer, Mitarbeiter in PubFlow

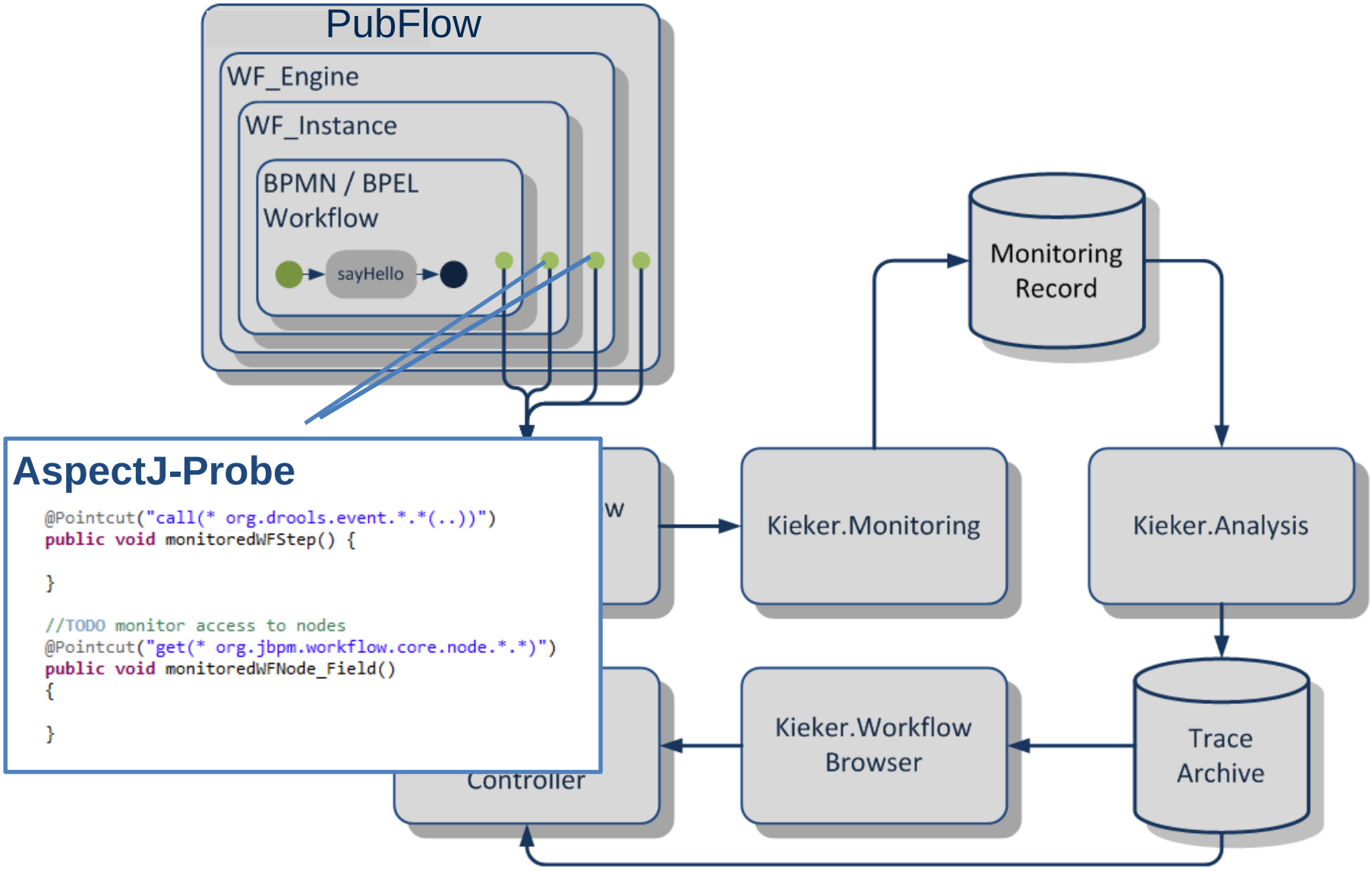
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Provenienzdaten-Erfassung in PubFlow



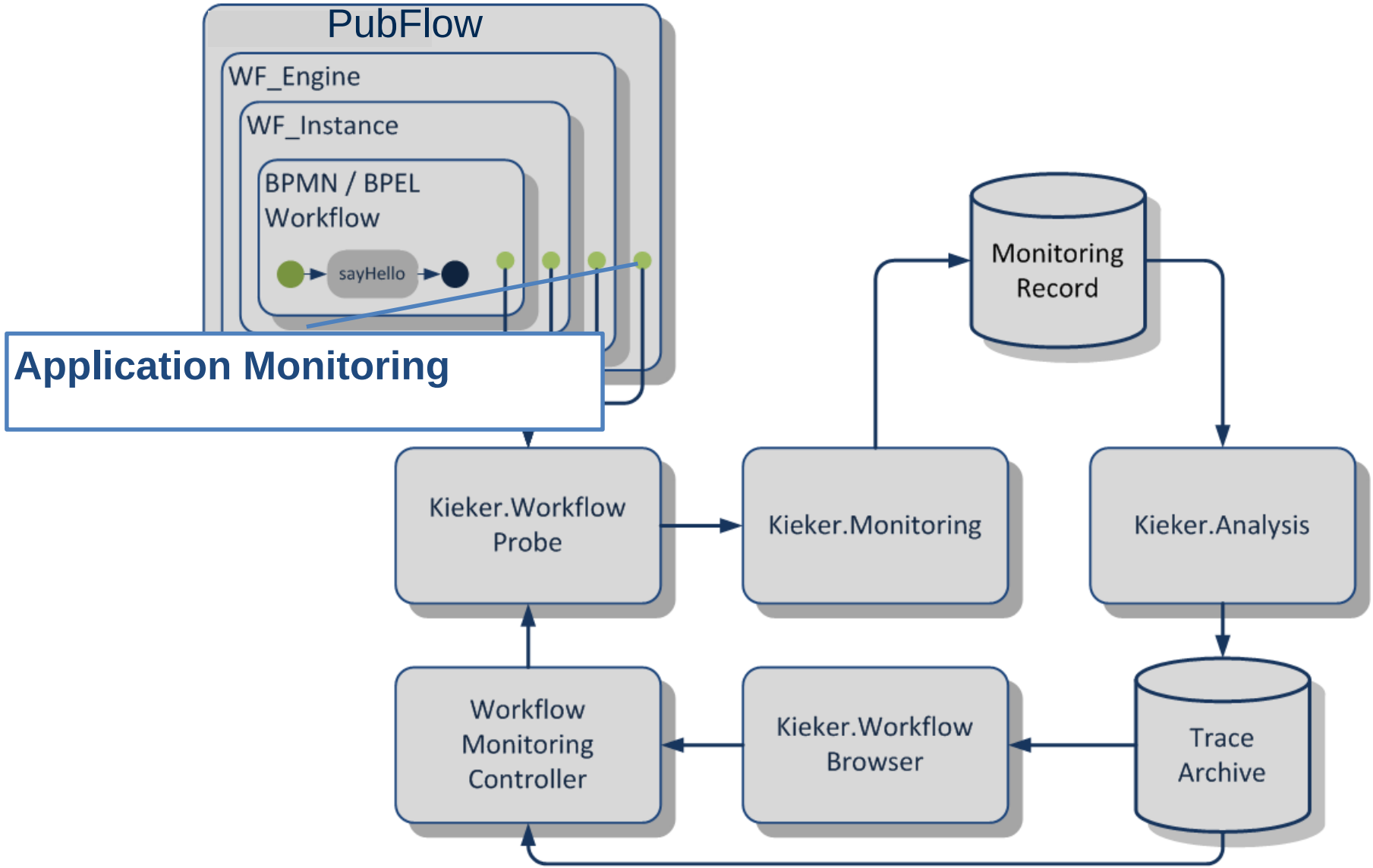
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Provenienzdaten-Erfassung in PubFlow



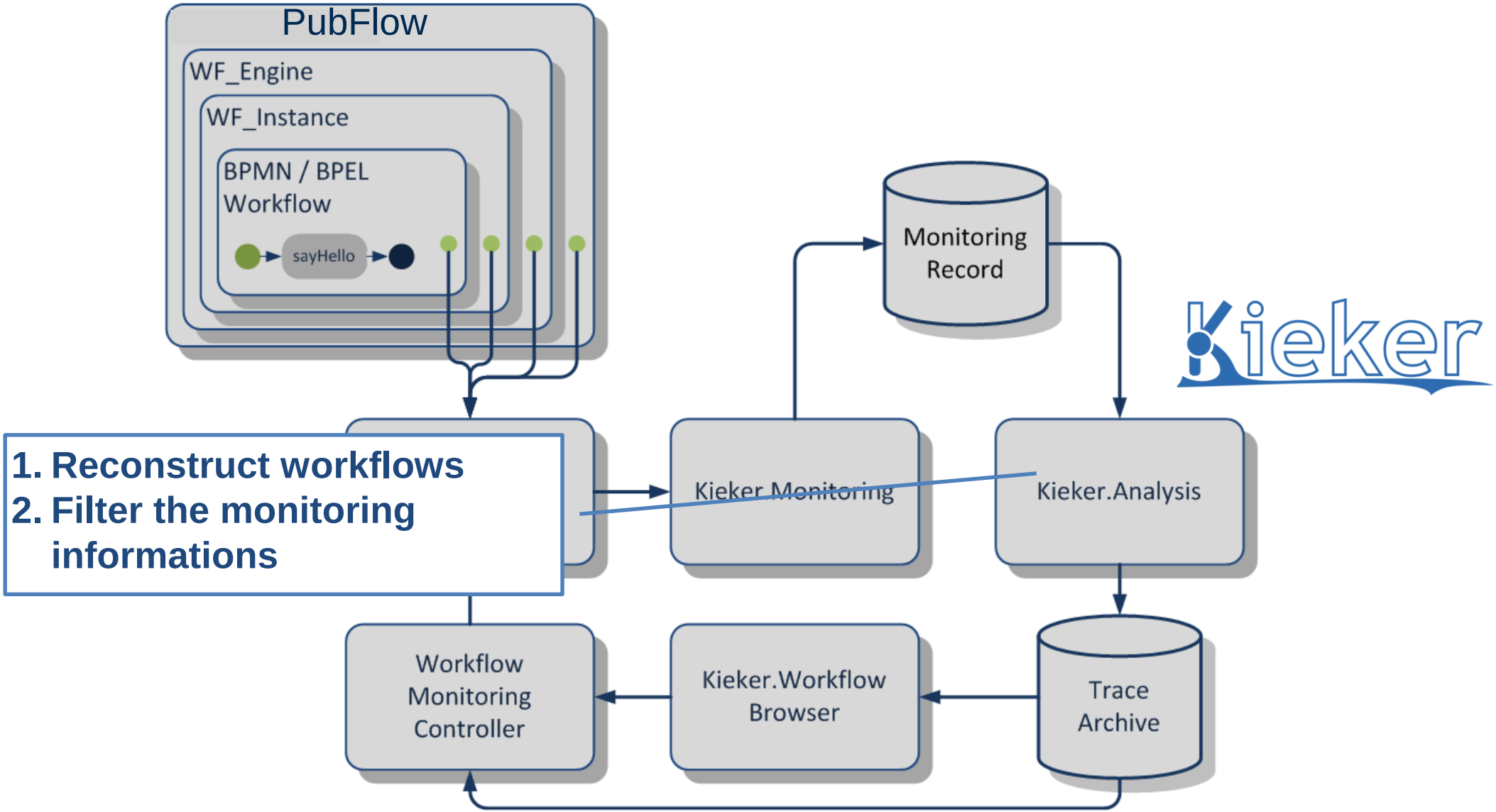
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Provenienzdaten-Erfassung in PubFlow



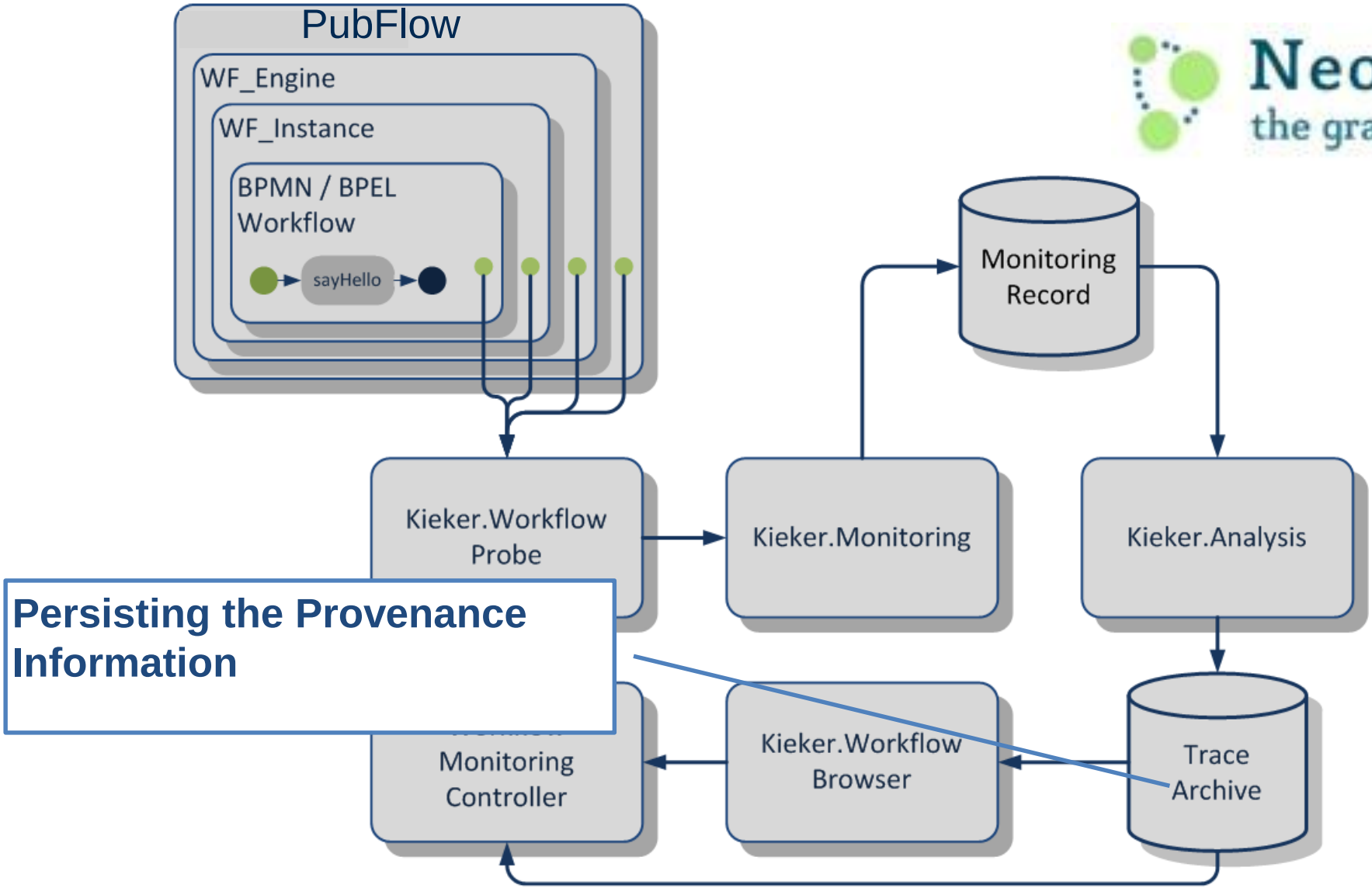
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Provenienzdaten-Erfassung in PubFlow



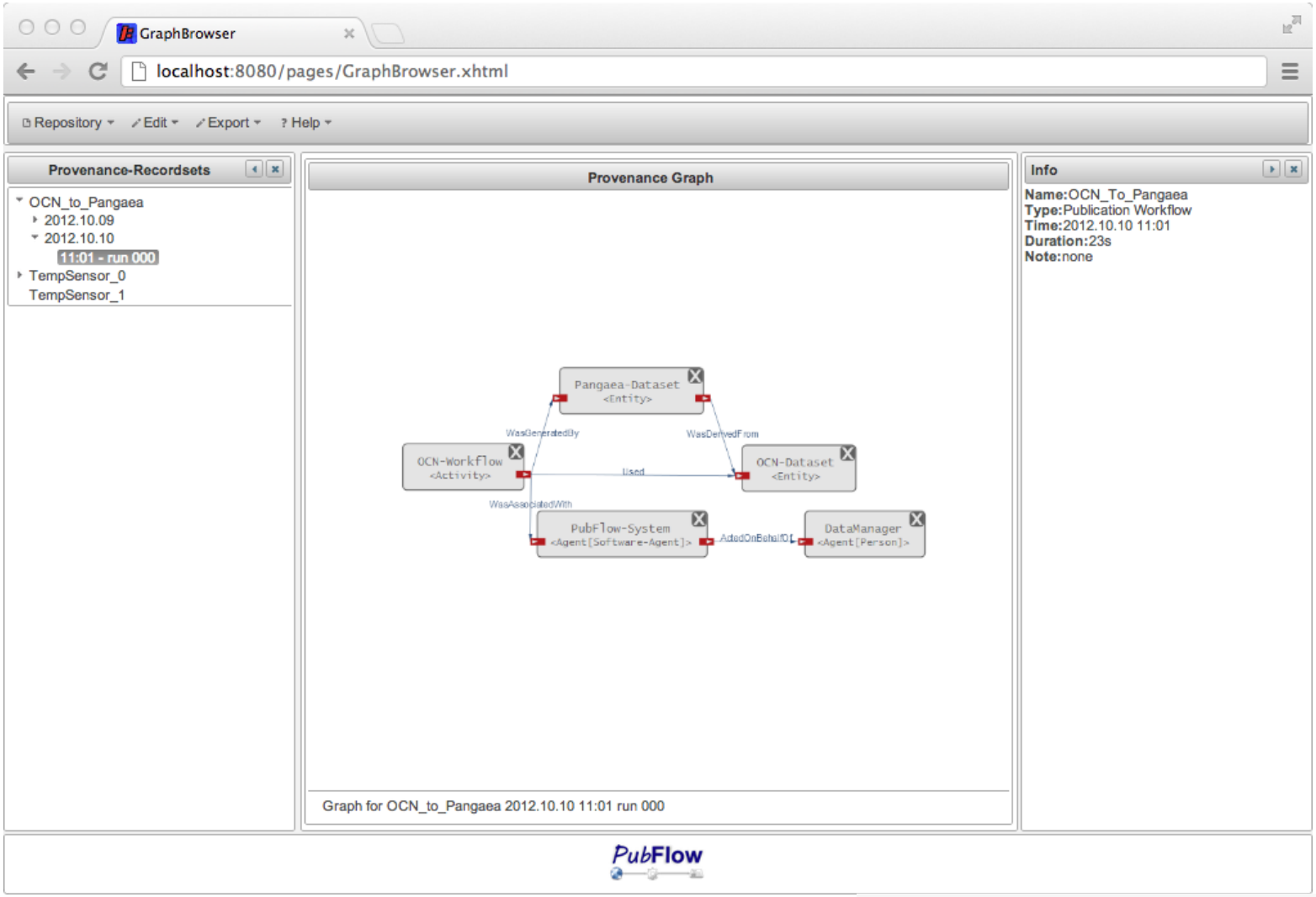
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Provenienzdaten-Erfassung in PubFlow



PubFlow Provenienz-Daten Viewer

Christian-Albrechts-Universität zu Kiel
Technische Fakultät



JavaScript InfoVis Toolkit
Create Interactive Data Visualizations for the Web

Christian-Albrechts-Universität zu Kiel
Technische Fakultät

Zusammenfassung und Ausblick

- Workflow-basierte Verarbeitung und Archivierung von Ozeanbeobachtungsdaten
  - Insbesondere zur Publikation von Forschungsdaten
  - Kombination von Workflow-Management und Aufgaben-Management
- Interdisziplinäre Zusammenarbeit in PubFlow u.a.
  - mit dem Kieler Datenmanagement Team,
  - mit der ZBW und den Bibliotheken der CAU und des GEOMAR,
  - mit den Rechen- und Datenzentren der CAU und des GEOMAR,
  - mit MaNIDA, dem Portal Deutsche Meeresforschung (Helmholtz-Initiative mit AWI, BSH, GEOMAR, HZG, Marum; CAU assoziiert) und
  - mit der Universität Southampton (UK).
- Fachbeirat mit Mitgliedern aus AWI, Pangaea/Marum, GEOMAR, DKRZ, DLR, IKMB
- Ausblick u.a.
  - Direkte Integration in das KDMT-Portal
  - Weiterer Anwendungsfall im Bereich CVOO (Cap Verde Ocean Observatory) angedacht
  - Grundlegende Arbeiten im Bereich der Monitoring-basierten Provenienzdaten-Erfassung.

C | A | U
Christian-Albrechts-Universität zu Kiel
Technische Fakultät

References

1. Peer Brauer, Wilhelm Hasselbring: “PubFlow: provenance-aware workflows for research data publication.” In: 5th USENIX Workshop on the Theory and Practice of Provenance (TaPP '13), April 2-3, 2013, Lombard .
2. Peer Brauer, Wilhelm Hasselbring: “Capturing provenance information with a workflow monitoring extension for the Kieker framework.” In: Proceedings of the 3rd International Workshop on Semantic Web in Provenance Management, May 2012, Heraklion, Greece
3. Matthias Rohr, André van Hoorn, Jasminka Matevska, Nils Sommer, Lena Stoever, Simon Giesecke, and Wilhelm Hasselbring: Kieker: “Continuous monitoring and on demand visualization of Java software behavior.” In Claus Pahl, editor, Proceedings of the IASTED International Conference on Software Engineering 2008 (SE’08), pages 80–85, February 2008.
4. André van Hoorn, Matthias Rohr, Wilhelm Hasselbring, Jan Waller, Jens Ehlers, Sören Frey, and Dennis Kieselhorst: “Continuous monitoring of software services: Design and application of the Kieker framework.” Technical Report TR-0921, Department of Computer Science, University of Kiel, Germany, November 2009.
5. André van Hoorn, Jan Waller, and Wilhelm Hasselbring: “Kieker: A framework for application performance monitoring and dynamic software analysis.” In Proceedings of the 3rd ACM/SPEC International Conference on Performance Engineering (ICPE 2012), pages 247–248. ACM, April 2012.