
Von der Datenentstehung bis zur Publikation - das Kieler Datenmanagement-Portal

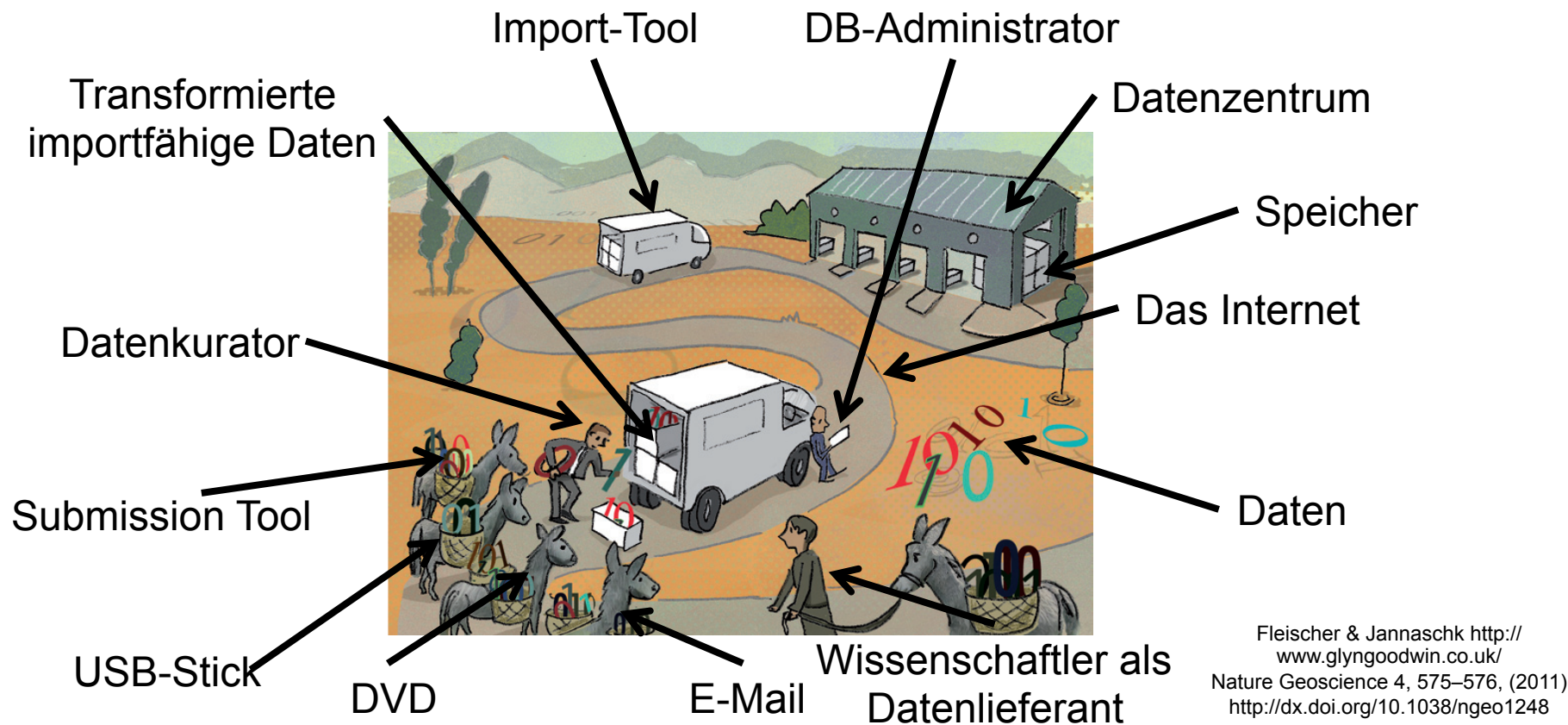
Dirk Fleischer
GEOMAR Helmholtz-Zentrum für Ozeanforschung Kiel

Programm

- Allgemeiner Stand
 - Entkoppelte Abschnitte im Lebenszyklus von Daten
 - VFU Kieler Datenmanagement Infrastruktur
- Provenienz
 - Was ist das - im Kontext Forschungsdaten?
 - Datenerfassung zum Zeitpunkt ihrer Entstehung
- Ausblick
 - Datenlandschaften
 - Int. Datenlandschaft – int. Forschungslandschaft

Allgemeiner Stand

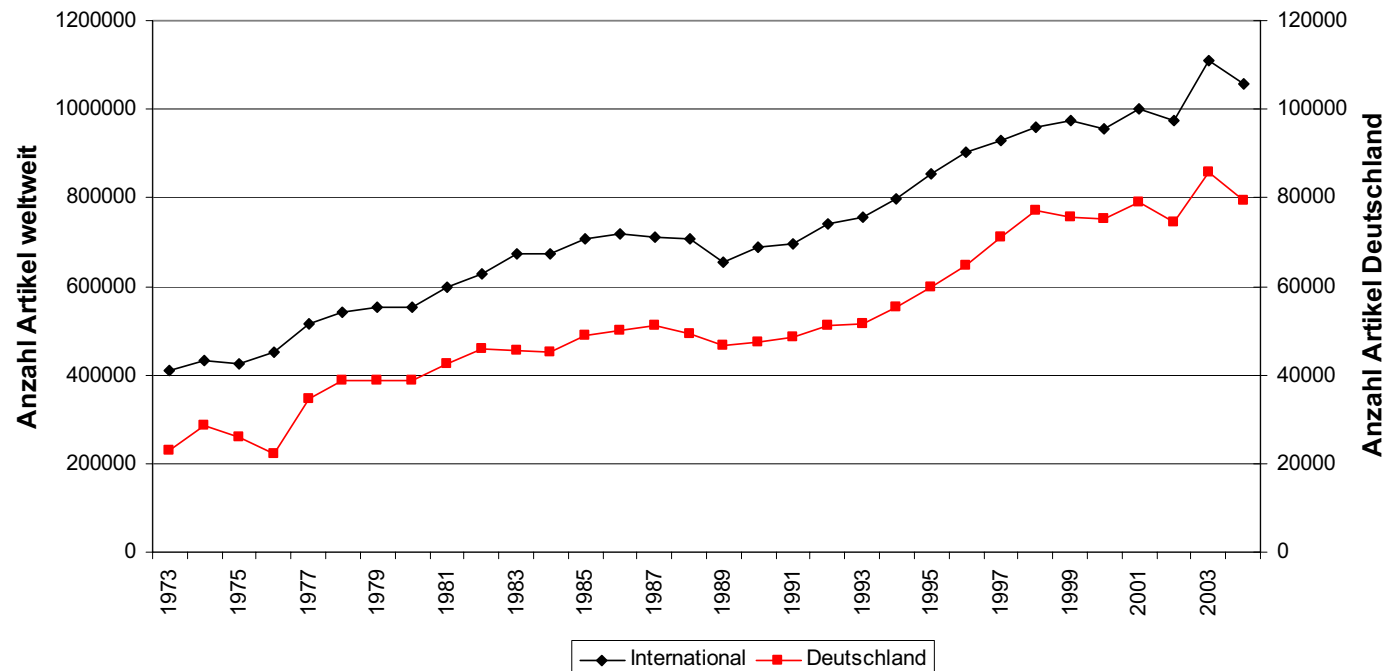
Entkoppelte Abschnitte im Lebenszyklus von Daten



Allgemeiner Stand

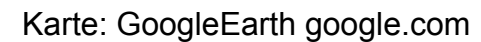
Entkoppelte Abschnitte im Lebenszyklus von Daten

Anzahl an naturwissenschaftlichen Artikeln im SCI



(Quelle: D. Tunger 2009 Forschungszentrum Jülich)

Entkoppelte Abschnitte im Lebenszyklus von Daten



Allgemeiner Stand

KDMI Rahmenbedingungen

Unpublizierte
Daten

KEINE
Langzeitarchivierung

Strukturierung

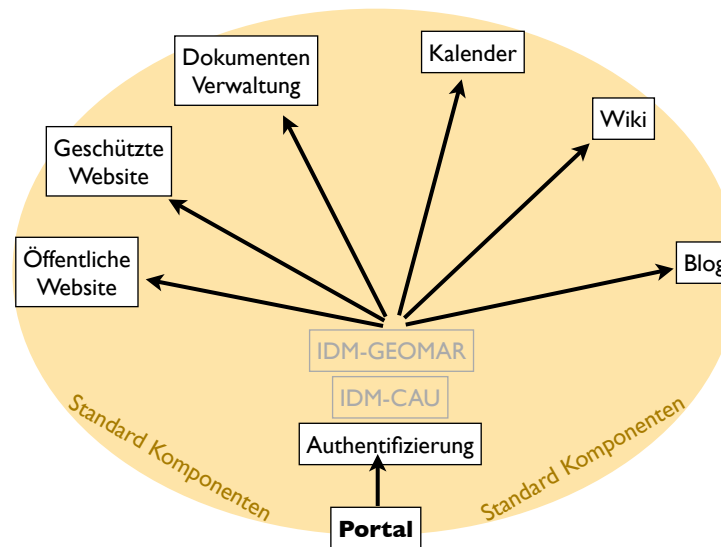
Beschleunigung des
Publikationsprozesses

Nachweis

Für den
Wissenschaftsstandort
Kiel

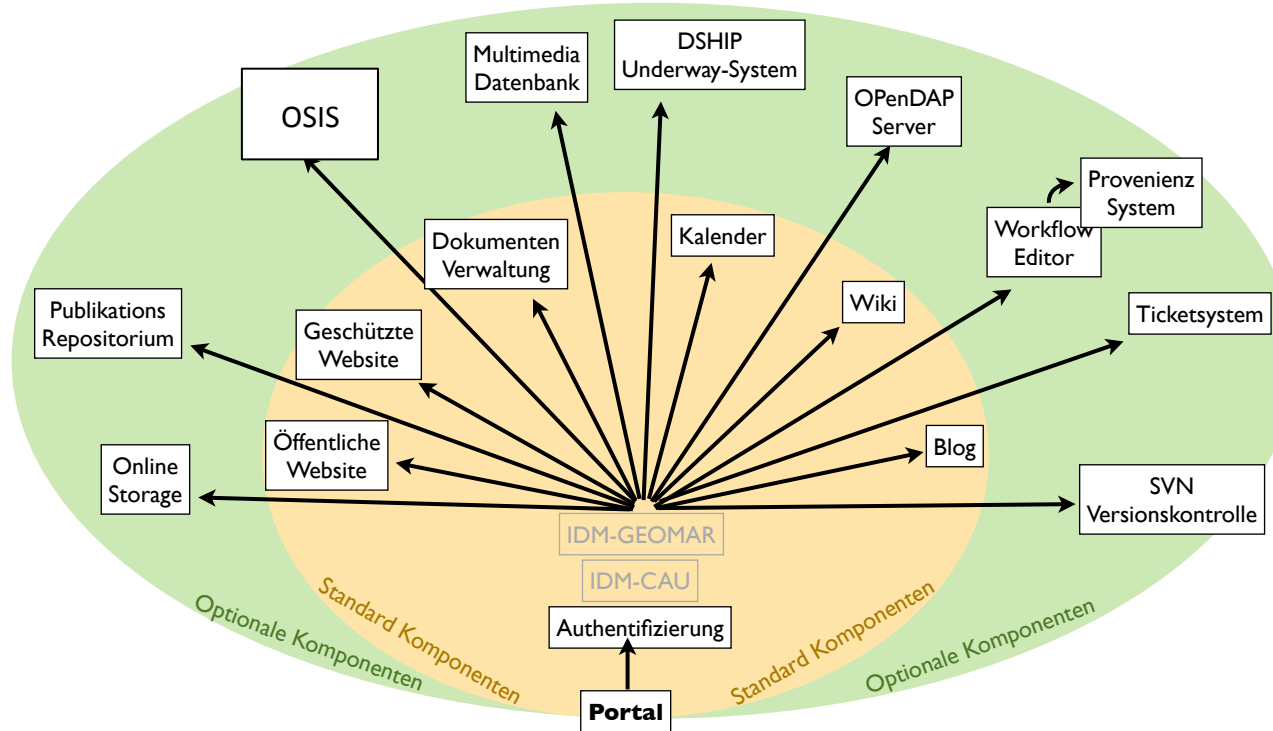
Allgemeiner Stand

VFU Kieler Datenmanagement Infrastruktur



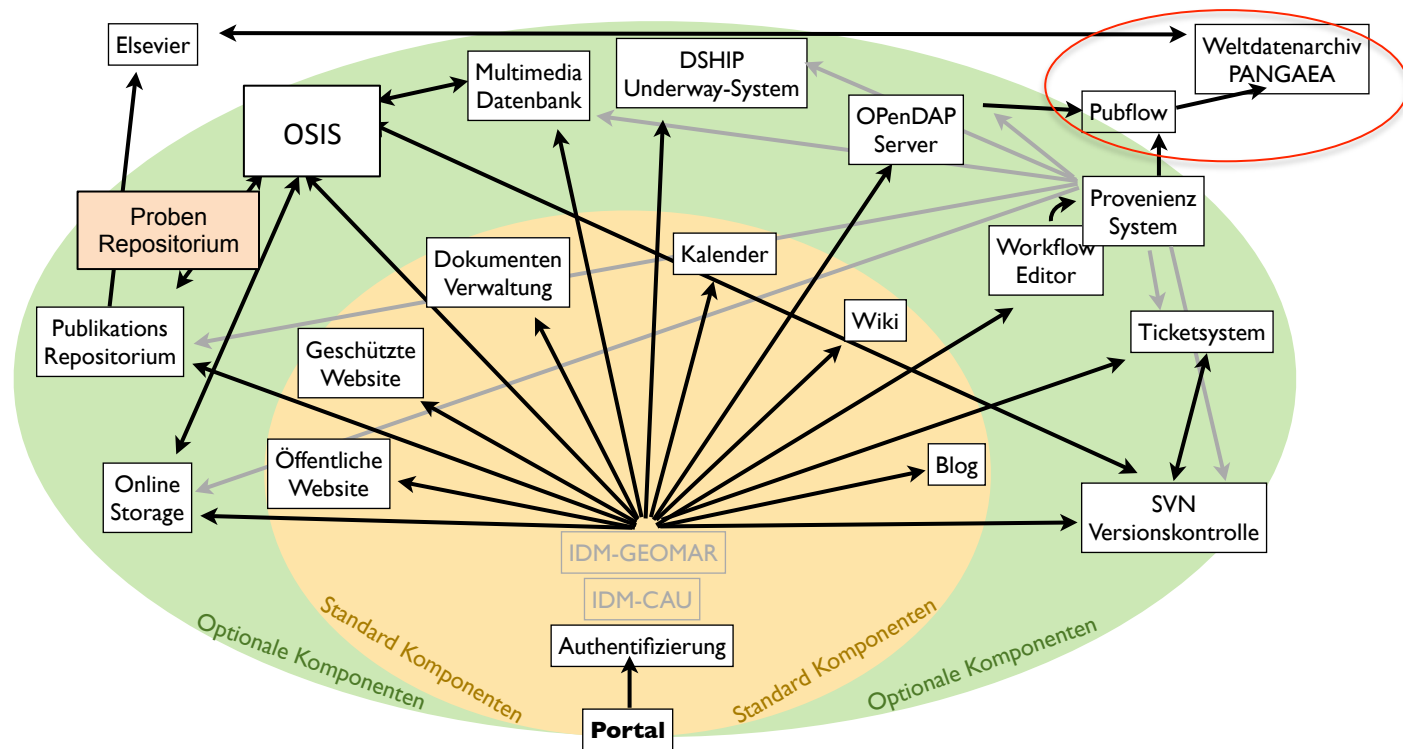
Allgemeiner Stand

VFU Kieler Datenmanagement Infrastruktur



Allgemeiner Stand

VFU Kieler Datenmanagement Infrastruktur



Provenienz

Was ist das – im Kontext Forschungsdaten?

- EDV - Informatik – Scientific Workflow

- Workflow-Engine mit BEPL oder BPMN Workflowbeschreibung
- Ausgangslage ist ein vorliegendes digitales Datenset
- Endprodukt ist ein verändertes digitales Datenset
- BEPL/BPMN Beschreibung + Engine Laufzeit = Provenienz

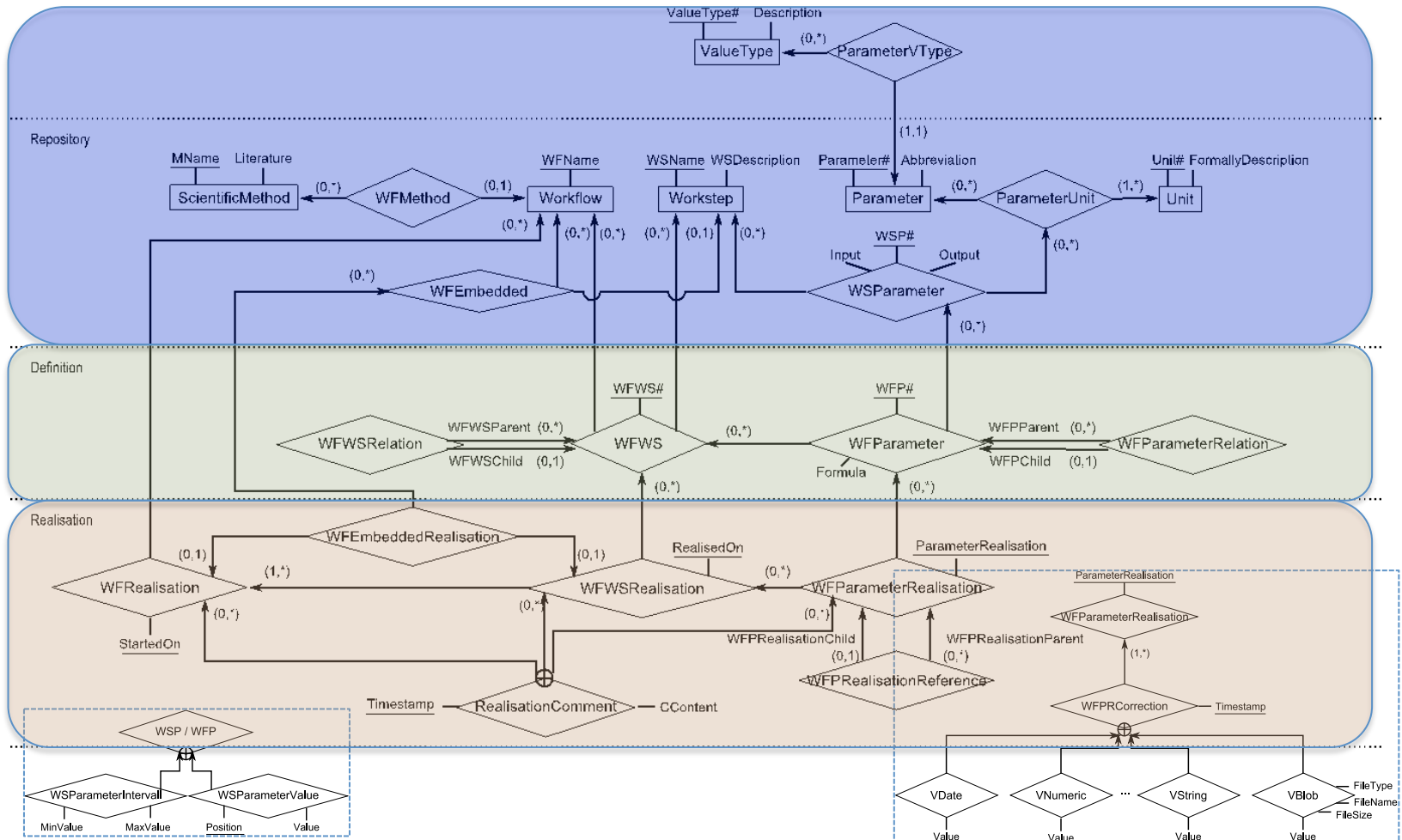
- Beobachtungsdaten

- Beruhend auf realen Objekten (Proben)
- Daten werden gemessen/erzeugt
 - Wie wird das gemacht – Methode
 - Wie wurde die Methode ausgeführt

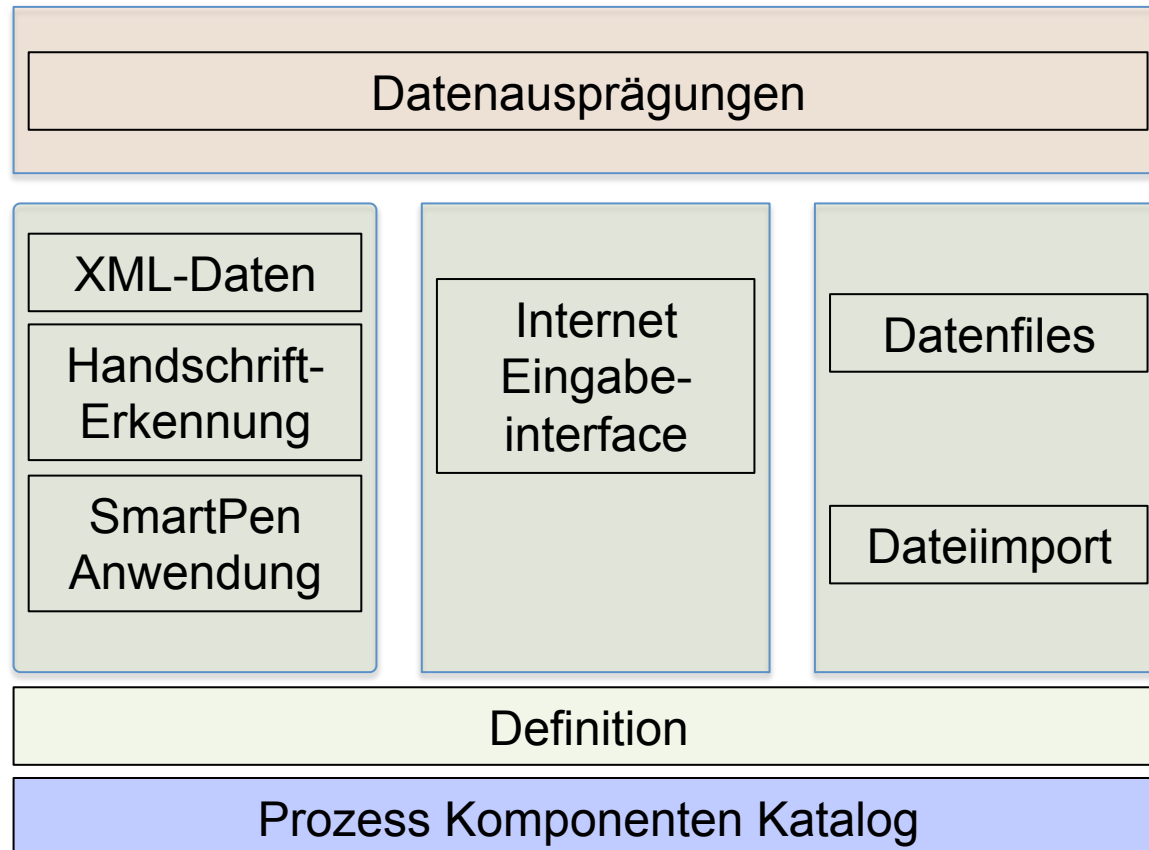
- Datenspeicherung als Humanprozess

- Proben werden sequenziell bearbeitet
- Menschen arbeiten sequenziell

Provenienz Datensystem

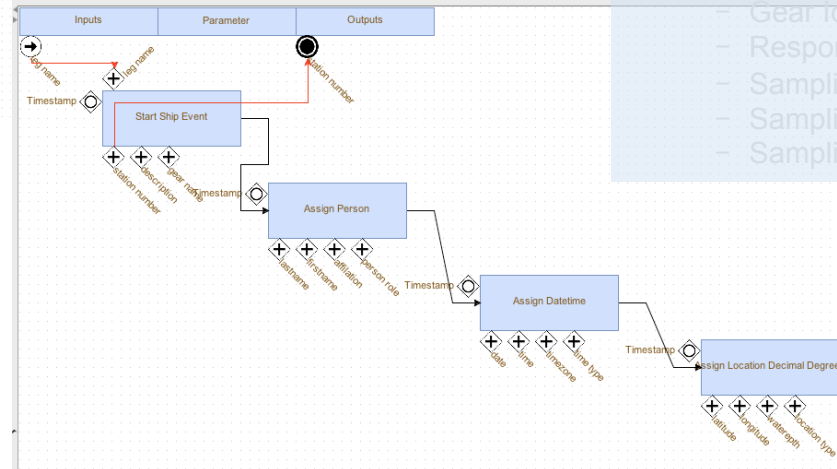
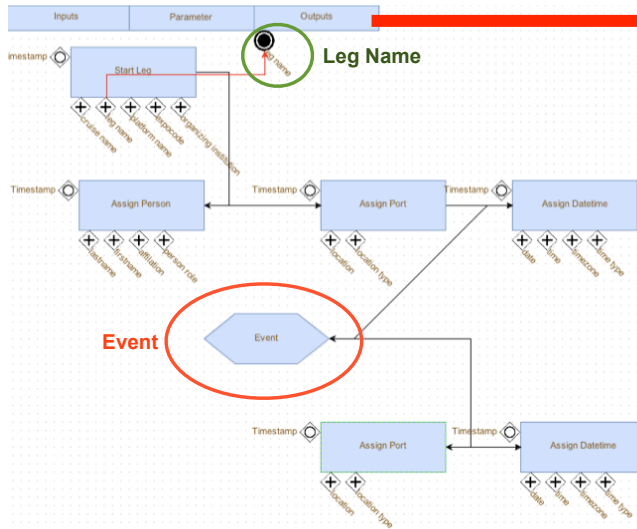


Datenakquise



Provenienz

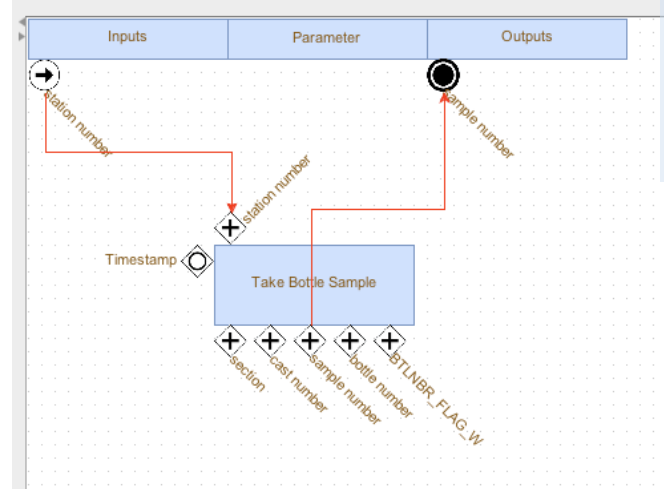
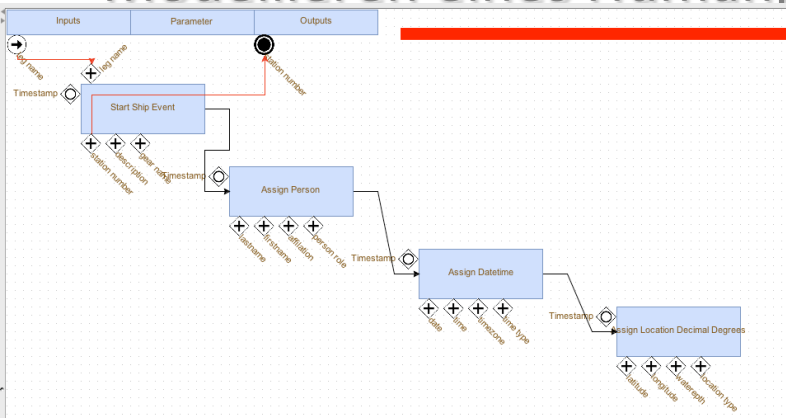
Modellieren eines Humanprozesses am Bsp. Forschungsschiffreise



- Cruise Name
 - Leg Name
 - Schiffsname (GND)
 - ExpoCode
 - Org. Institut
 - Fahrtleiter (GND)
 - Abfahrtsorten
 - Abfahrtszeit
 - Leg Name
 - Stationnummer (IGSN)
 - Description
 - Gear Identifier
 - Responsible Person (GND)
 - Sampling Timestamp
 - Sampling Location (Lat/Lon)
 - Sampling Altitude
- Fahrtleiter**

Provenienz

Modellieren eines Humanprozesses am Bsp. Forschungsschiffreise



- Cruise Name
 - Leg Name
 - Schiffsname (GND)
 - ExpoCode
 - Org. Institut
 - Fahrleiter (GND)
 - Abfahrthafen
 - Abfahrtszeit

Fahrleiter

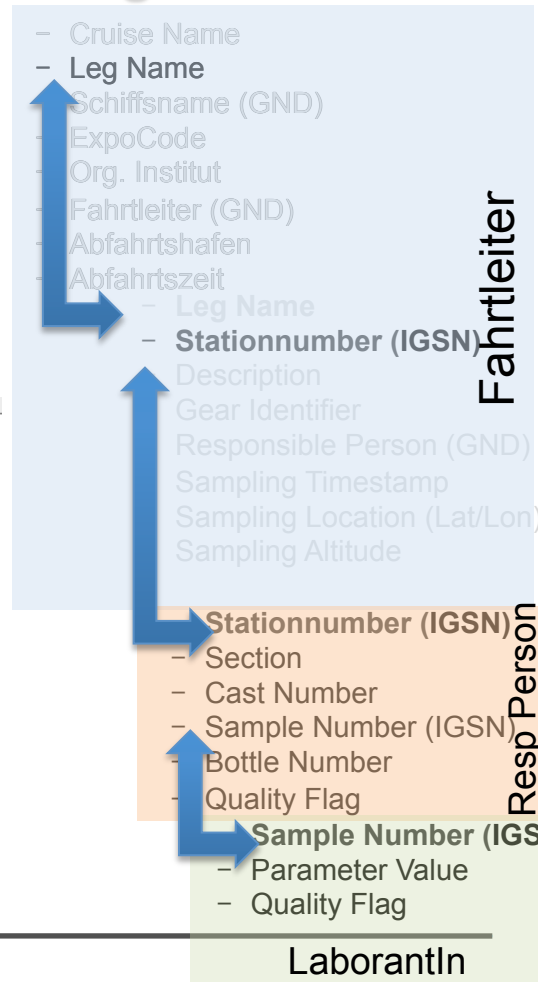
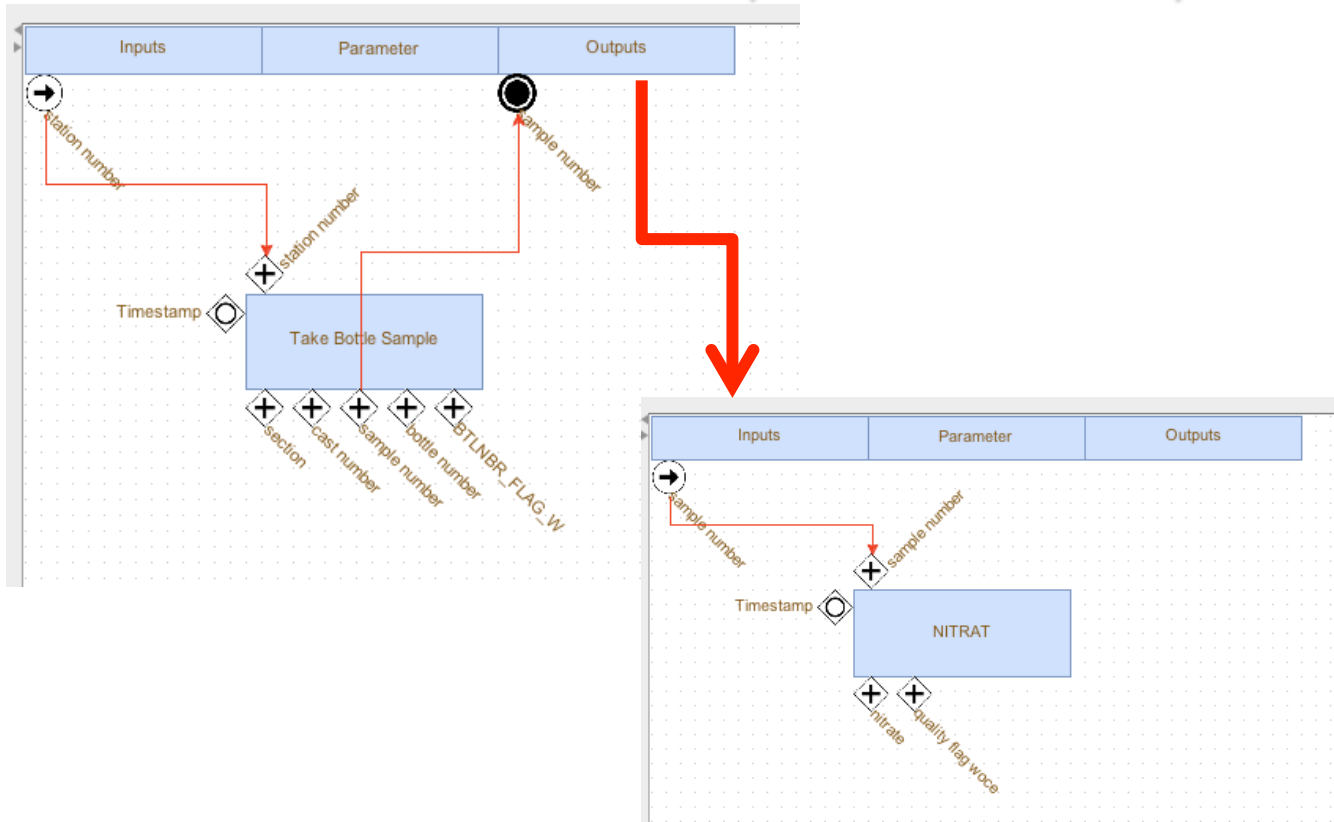
- Leg Name
 - Stationnummer (IGSN)
 - Description
 - Gear Identifier
 - Responsible Person (GND)
 - Sampling Timestamp
 - Sampling Location (Lat/Lon)
 - Sampling Altitude

- Stationnummer (IGSN)
 - Section
 - Cast Number
 - Sample Number (IGSN)
 - Bottle Number
 - Quality Flag

Resp Person

Provenienz

Modellieren eines Humanprozesses am Bsp. Forschungsschiffreise



Datenakquise

Am Ort der Datenentstehung - Handschriften Erkennung



Datenakquise

SmartPen-Anwendung

Beispiel Anwendung für Thor Hansteen ein Feldtagebuch

Position	Grad	Minute	Dezimal	Ort:
Breite	-39	23	570	Vittarica
Länge	-71	49	475	Land:
Datum:	07.03.2013			chile

Text

Go with Christian into the field to search for the second location. Problem with original road planning closed gate to private property go to alternative the Altitude: 1380 compass direction 244 Start: 19:24 UTC Stop: 20:11 UTC

Bild

Vittarica Vulk.

S

N

Gas (Proben) genommen
Sensor kalibriert:
Logger kopiert/ausgetauscht

Probennummern:

Beispiel Anwendung für Thor Hansteen ein Feldtagebuch

Position	Grad	Minute	Dezimal	Ort:
Breite	-39	23	570	Vittarica
Länge	-71	49	475	Land:
Datum:	07.03.2013			chile

Text

Go with Christian into the field to search for the second location. Problem with original road planning closed gate to private property go to alternative the Altitude: 1380 compass direction 244 Start: 19:24 UTC Stop: 20:11 UTC

Bild

Vittarica Vulk.

S

N

Gas (Proben) genommen
Sensor kalibriert:
Logger kopiert/ausgetauscht

Probennummern:

Beispiel Anwendung für Thor Hansteen ein Feldtagebuch

Position	Grad	Minute	Dezimal	Ort:
Breite	-39	23	570	Vittarica
Länge	-71	49	475	Land:
Datum:	07.03.2013			chile

Text

Go with Christian into the field to search for the second location. Problem with original road planning closed gate to private property go to alternative the Altitude: 1380 compass direction 244 Start: 19:24 UTC Stop: 20:11 UTC

Bild

Vittarica Vulk.

S

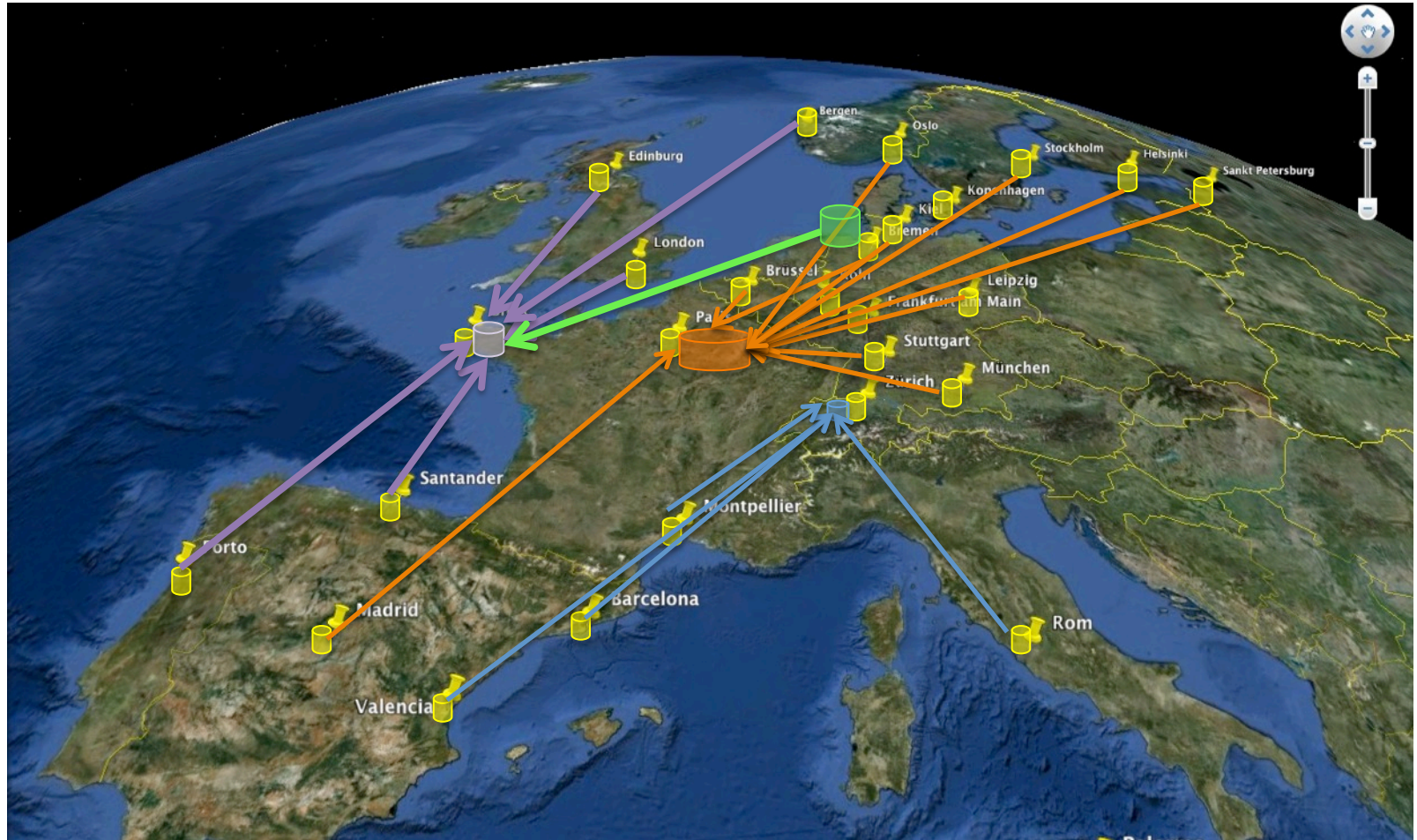
N

Gas (Proben) genommen
Sensor kalibriert:
Logger kopiert/ausgetauscht

Probennummern:

XML-Daten für Datenbanken

Datenlandschaft der Zukunft



Datenlandschaft der Zukunft



Fazit

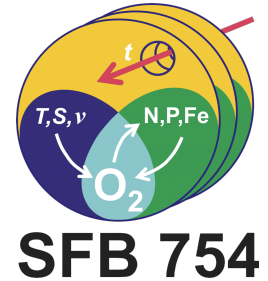
Datenmanagement in jedem Haus

- **Aktive Datenakquise und -management**
 - Daten zum Zeitpunkt der Entstehung
 - Vorstrukturierung zur weiteren Verwendung
 - Datenmanager gleichberechtigte Partner in Projekten
 - Teilprojekte um Datenmanagementmethoden innerhalb des Projektes zu erarbeiten
 - Projektplanung ebenfalls auf Basis der Datengrundlage
- **Archivierung und Verteilung**
 - Datenpublikationen stärker technisch unterstützen
 - Weltdatenarchive nutzen
 - Kooperationen mit wiss. Verlagen
 - Weltweiter Metadaten-Austausch bereits etabliert
- **Provenienz als multidimensionales Qualitätsmerkmal**
 - Einzelne Quality Flags sind subjektiv
 - Dokumentation der Entstehung von Daten ist ein Qualitätsmerkmal

Danke für Ihre Aufmerksamkeit



Fragen?
Anmerkungen?
Hinweise?



Danke an meine Kollegen

- Andreas Czerniak
- Hela Mehrrens
- Carsten Schirnack
- Pina Springer

