



WissGrid AP3: LZA Dienste

Michael Lautenschlager (DKRZ)
WissGrid AP3 Review, Potsdam 28.01.10



Bundesministerium
für Bildung
und Forschung



Inhalt:

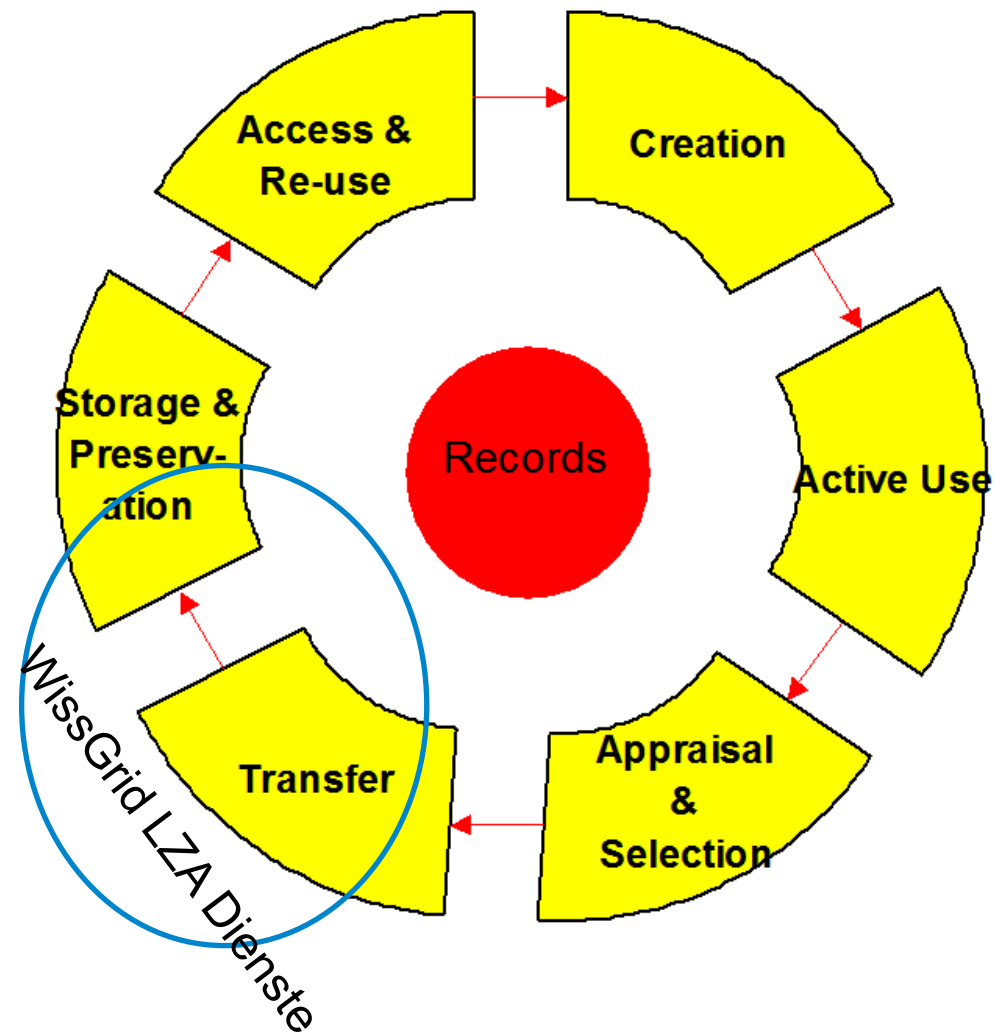
- Hintergrund und Motivation
- Aktuelle Umsetzung

Langzeitarchivierung im Life Cycle

Langzeitarchivierung (LZA):
Sicherung der Nachnutzbarkeit

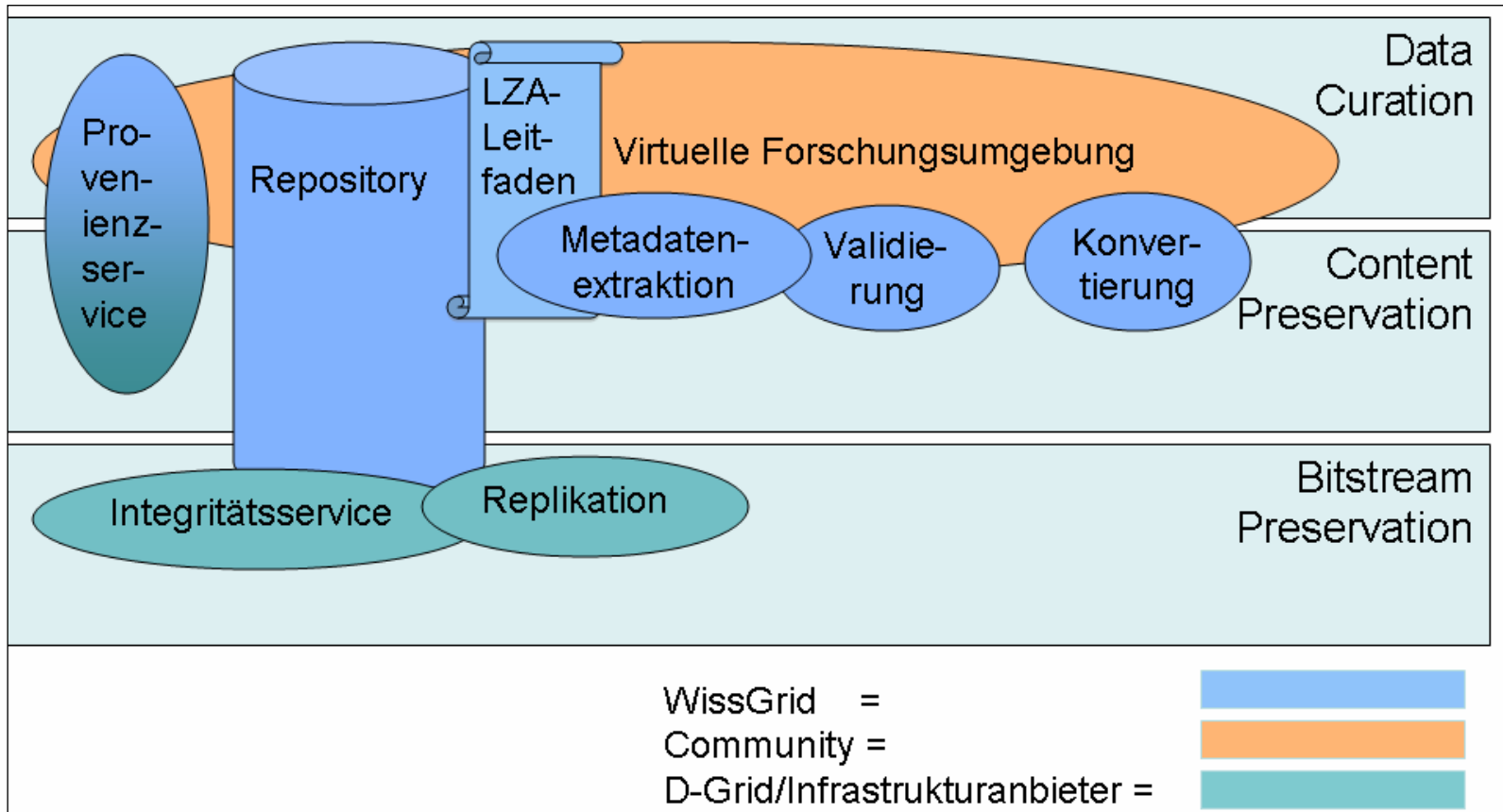
- Bitstream Preservation:
Datenentitäten bleiben erhalten
- Content Preservation:
Datenentitäten bleiben lesbar
- Data Curation: Datenentitäten
bleiben interpretierbar

Datenentitäten:
Digitale Medien und
wissenschaftliche Primärdaten





LZA-Dienste, LZA-Ebenen und Akteure



WissGrid Dienste-Fokus: Content Preservation



Schwerpunkt

Ein zentraler Aspekt in Content Preservation:
Anpassung vorhandener Datenformate an sich verändernde
SW-Werkzeuge

Strategie:

A) Konvertieren: Veraltete Speicherformate in aktuelle
Formate konvertieren (z.B. digitale Medien)

B) Konservieren: SW-Werkzeuge zum Lesen der
Speicherformate auf aktuelle Rechnerarchitekturen
migrieren (z.B. hochvolumige Daten aus
(Klima-)Modellierung und Erdbeobachtung)

WissGrid Fokus für LZA-Dienste auf A):

Transformation definierter Formate kann am ehesten
automatisiert und als Grid-Dienst implementiert werden.



LZA-Dienste, die sich aus WissGrid Architektur ergeben:

- **Repository:** SW-System, das Daten mit Metadaten im Grid verwaltet
- **Formatkonvertierung:** Konvertierung in neue technische Form für dasselbe intellektuelle Objekt
- **Formatcharakterisierung:** Auslesen technischer Metadaten aus Dateien
- **Formatvalidierung:** Prüft Übereinstimmung von Objekten mit Definition bzw. Charakterisierung
- **Provenienz-Service:** Informationen um Entstehung und Modifikation von Daten sollen in LZA-Repositoryn integriert werden, um Nachvollziehbarkeit sicherzustellen
- **WissGrid Dienste Framework (WDF):** Stellt LZA-Dienste im Grid zur Verfügung



LZA-Dienste im OAIS-Ingest

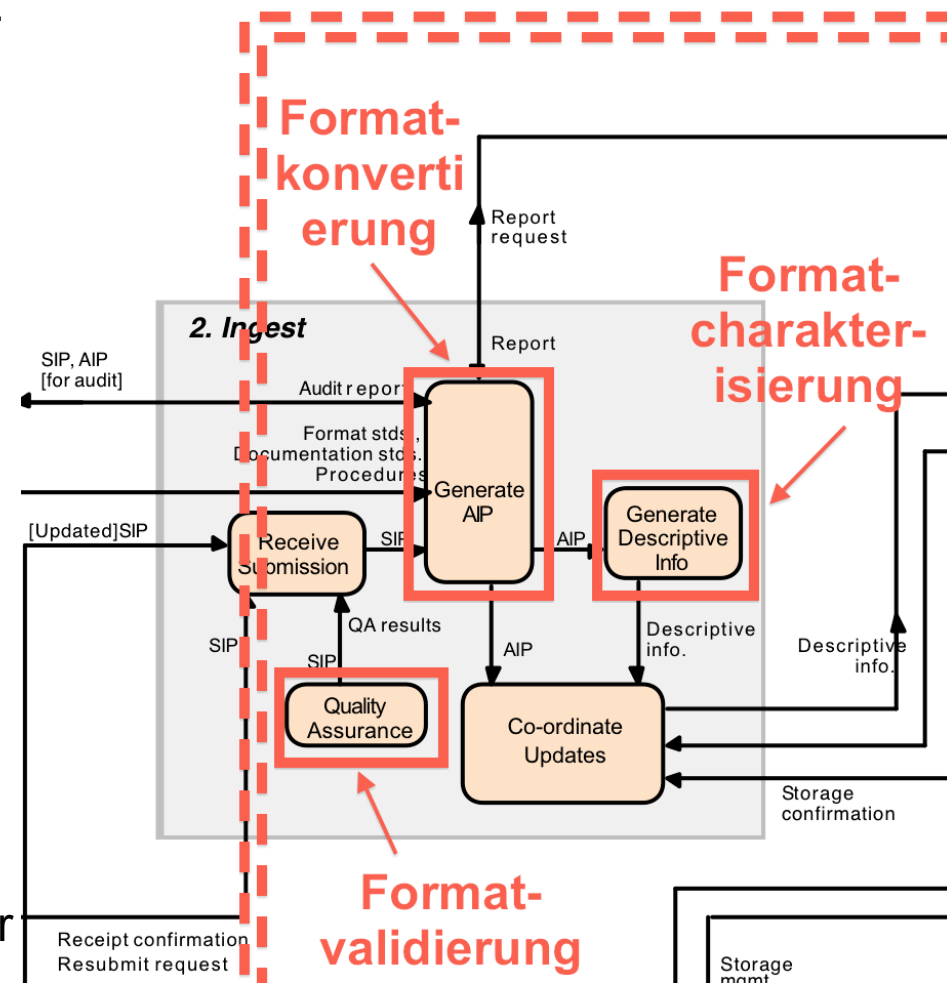
LZA-Dienste im OAIS-Ingest als Interface zum Forschungsdatenarchiv:

Formatvalidierung: Prüfung auf Formatfehler beim Einfüllen in Forschungsdatenarchiv (OAIS „quality assurance“)

Formatcharakterisierung: bietet Provenienzinformation. Übergang zu Data Curation und inhaltlichen Metadaten fließend (OAIS „general descriptive information“)

Formatkonvertierung: unterschieden nach „migration on ingest“, migration on access“ und „migration on obsolence“ (verschiedene OAIS Referenzen, hier „generate AIP“)

Provenienz-Service: Speicherung der Prozessinformation, enthalten in OAIS Informationspaketen.





Formatcharakterisierung und –validierung

- Funktionaler Ablauf:
 - Identifikation des Datenformats
 - Identifikation des richtigen Formatmoduls
 - Formatmodul parst Daten
 - Anwendung der Validierungskriterien
 - Serialisierung der gewonnenen Metadaten in Zielformat
- Technische Umsetzung: JHOVE2 wurde entwickelt für Standard-Medienformate, Erweiterung auf NetCDF in Klima-Community im Rahmen von C3-Grid (Details AWI, Bremerhaven)

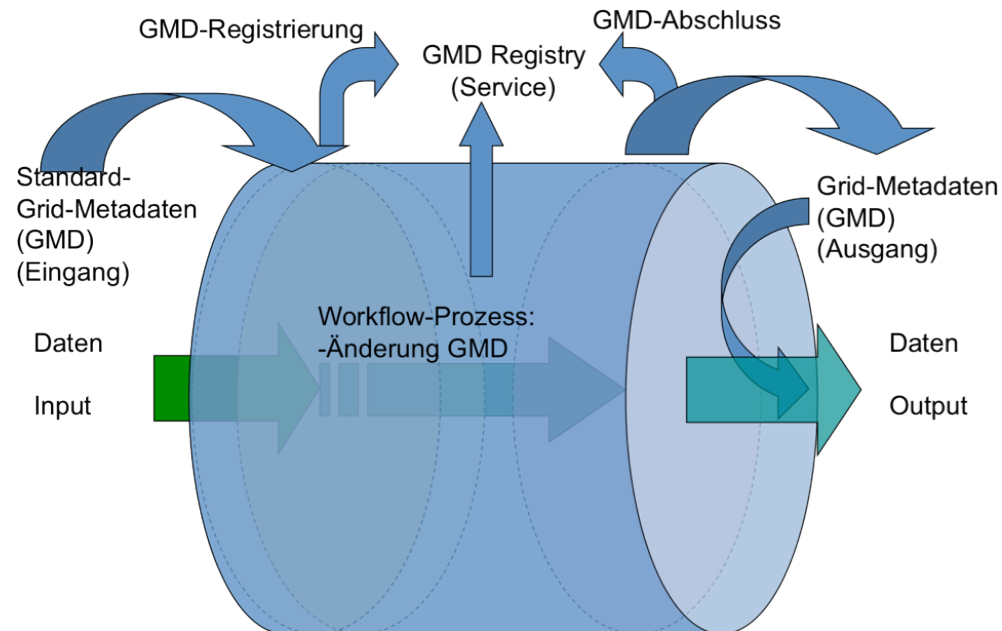


Formatkonvertierung

- Funktionaler Ablauf:
 - Identifikation des Datenformats (z.B. JHOVE)
 - Evaluation der Abhängigkeiten und Entscheidung zur Aggregation in LZA
 - Identifikation des richtigen Konvertermoduls
 - Ausführung des Konvertermoduls und Validierung des Zielformats
 - Generierung von Provenienzmetadaten
- Technische Umsetzung:
 - CriB und kopal Migration Manager sind Frameworks zur Formatkonvertierung (Prototypen), eigentliche Konvertierung durch externe Dienste
 - Klima-Community:
 - Aggregation von Einzelfiles in Container-Format (tar)
 - CMOR-2: Konvertierung von Modelloutput in NetCDF4
 - CDOs (Climate Data Operators): Konvertierung in NetCDF, GRIB, ASCII
 - WDCC des DKRZ: Konvertierung NetCDF oder GRIB in ASCII

Metadaten der Prozessierung werden in XML-Container (z.B. OGF-Usage-Record) gespeichert und mit XSLT-Mapping auf Metadaten in LZA Repositorien der Forschungsdatenarchive abgebildet (z.B. PREMIS-Standard für preservation metadata, CERA-2 im WDCC/DKRZ, ISO19115 im C3-Grid)).

- Entwurf:



C3-Grid: Metadaten werden in den einzelnen Datenbearbeitungsschritten der Workflows angepasst



WissGrid Dienste Framework (WDF)

- Ablauf einer LZA Aufgabe:
 1. Submit: Annahme der LZA-Aufgabe
(z.B. Konvertiere JPEG Daten, die älter als zwei Jahre sind, in JPEG 2000)
 2. Extract: Filterung der gefragten Daten aus Repository
 3. Identify: Auswahl einer verfügbaren CE (computing element), auf dem der gefragte Dienst installiert ist
 4. Stage-In: Transfer der Daten auf das CE
 5. Stage-Out: Annahme der konvertierten Daten, nach Ablauf des Jobs
 6. Re-Ingest: Rückführung der Daten in das Repository und Aktualisierung der Metadaten

