

95. Deutscher Bibliothekartag

# Metadatenverwaltung für den Verteilten Dokumentenserver

Stefan Hohenadel

[[stefan.hohenadel@bsz-bw.de](mailto:stefan.hohenadel@bsz-bw.de)]

Bibliotheksservice-Zentrum Baden-Württemberg

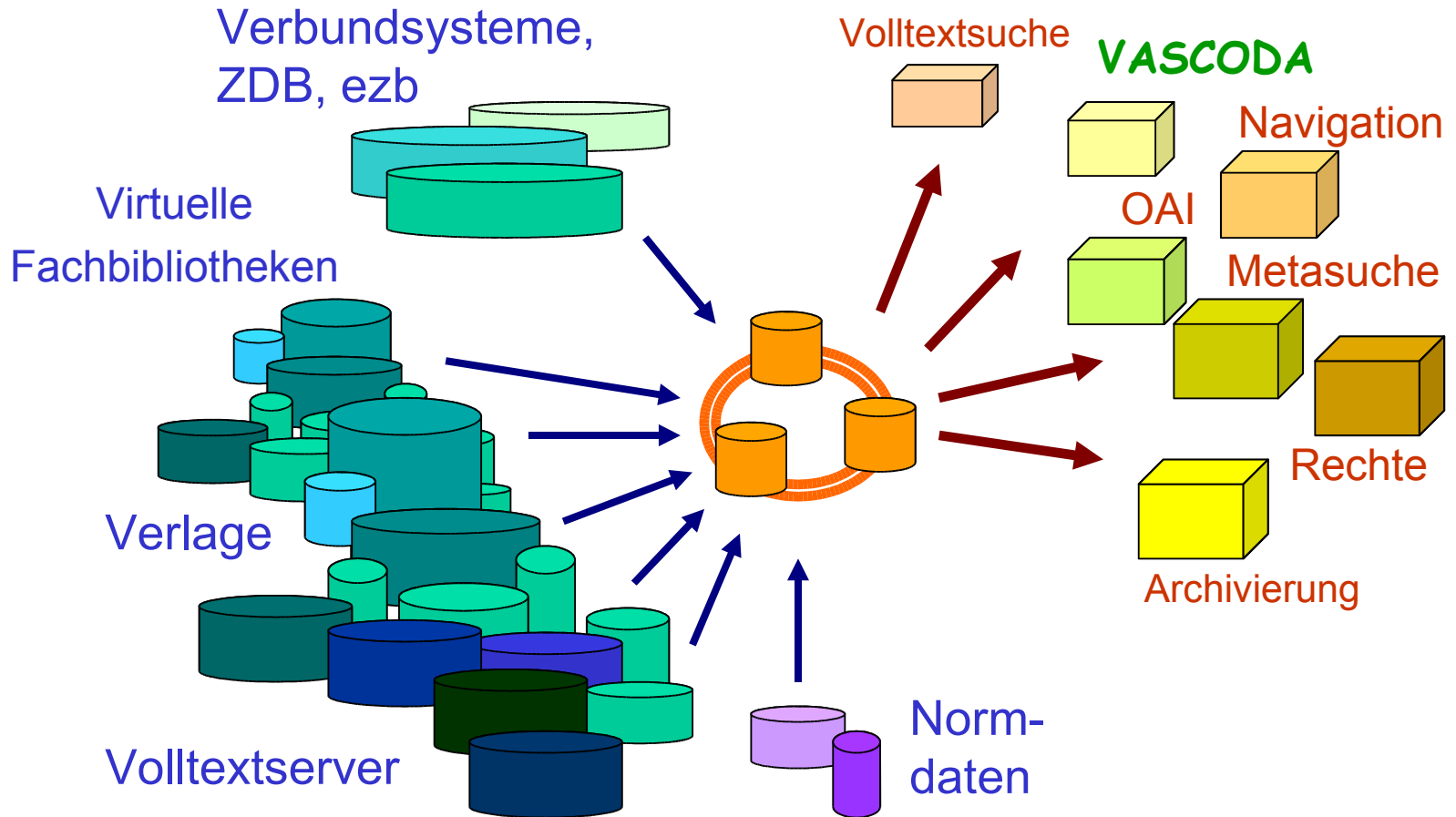
22. März 2006

---

## Digitale Dokumente

- **Zugriff** auf digitale Dokumente (Datenbanken, Artikel, Journale, Hochschulschriften etc.) ist für **Bildung** und **Wissenschaft** essentiell.
- Idealerweise: **einheitlicher, übergreifender und gut benutzbarer Zugang** zu den veröffentlichten **Dokumenten**.
- Wachsende **Probleme** durch unübersehbare **Vielfalt** von:
  - **Datenformaten** und -quellen
  - Erschließungsmethoden
  - **Lizenzen**
- Problem bei Lizenzen: wissenschaftliche Quellen sollen **permanent verfügbar** sein (sonst: Nachvollziehbarkeit des Forschungsprozesses gefährdet).

# Angestrebtes Szenario



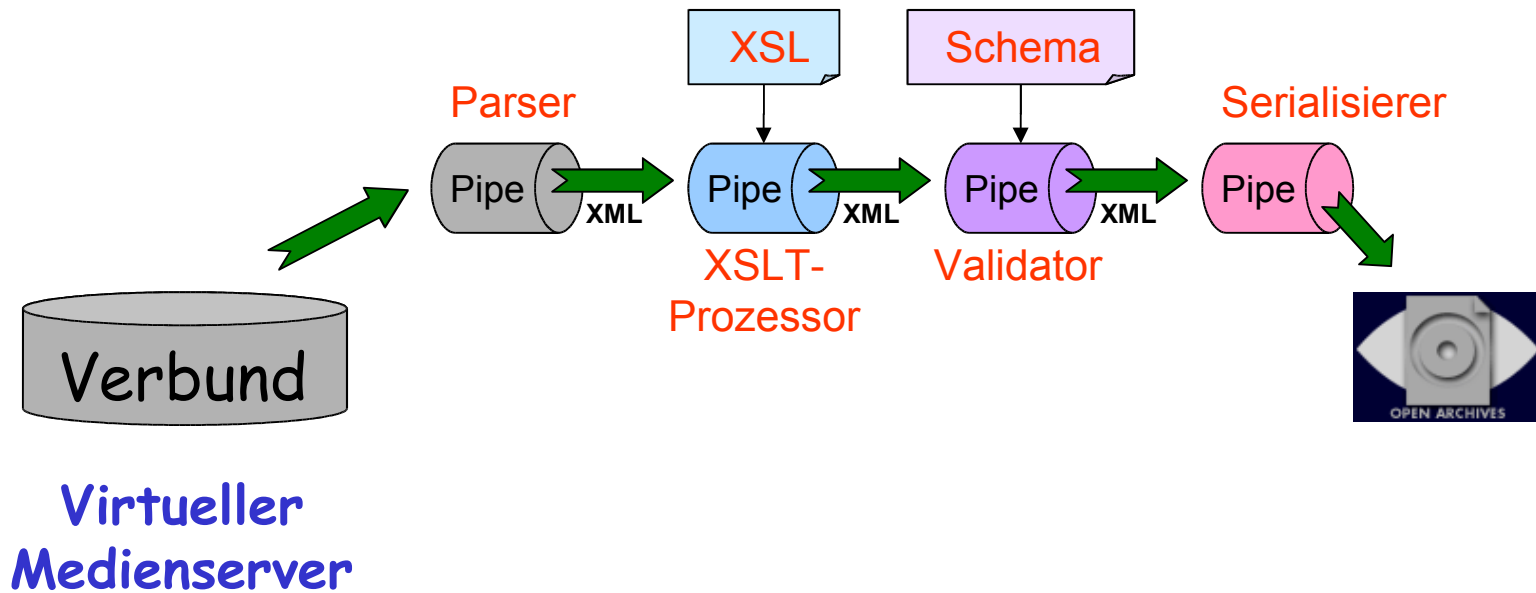
## Verteilter Dokumentenserver (VDS)

- AG Verbundsysteme: Alle **digitalen Dokumente** sollen in einem **verteilten Dokumentenserver** zugänglich sein.
- Ziel: Verteiltes **Depot** lizenzierter Publikationen und freier (digitaler) Dokumente. **Zugang soll strukturell einheitlich und komfortabel sein.**
- Umfasst **Speicherung, Erschließung, Verlinkung** und **Präsentation** von Dokumenten.
- VDS soll **kooperativ** von allen Verbänden betrieben werden. Ziel ist die bessere Nutzbarkeit von Hochschulservern
- Herkunft des Vorhabens: **Metadatenverwaltung** (BSZ), Verteilte **Rechteverwaltung und Authentifizierung** (UB Freiburg und UB Regensburg), **Volltextsuche** (vascode), **Prototyp** des Gesamtsystems (Kooperativer Bibliotheksverbund Berlin-Brandenburg).

## Metadatenverwaltung und -konvertierung

- Ausgangssituation: Metadaten werden von den Inhabern zur Verfügung gestellt (**Dateien**).
- Ziel: Metadaten sollen für verschiedene Nutzungen **aufbereitet** werden (Onlinesuche/Open Access, Kataloganreicherung, etc.)
- Idee: Statt für jede Nutzung eigene Software zur Aufbereitung der Metadaten zu schreiben, schreibt man nur **Bausteine** – sog. „**Pipes**“ – die mehrfach verwendbar und frei kombinierbar sind.
- Diese **Pipes** werden zu **Pipelines** zusammengesetzt. Die Metadaten **passieren** die Pipeline. Jede Pipe wird einmal **durchlaufen** und führt ihre spezifische Funktion auf den Metadaten aus. So können diese flexibel verarbeitet werden.

## Anwendungsfall 1: OAI Zugriff auf Virtuellen Medienserver

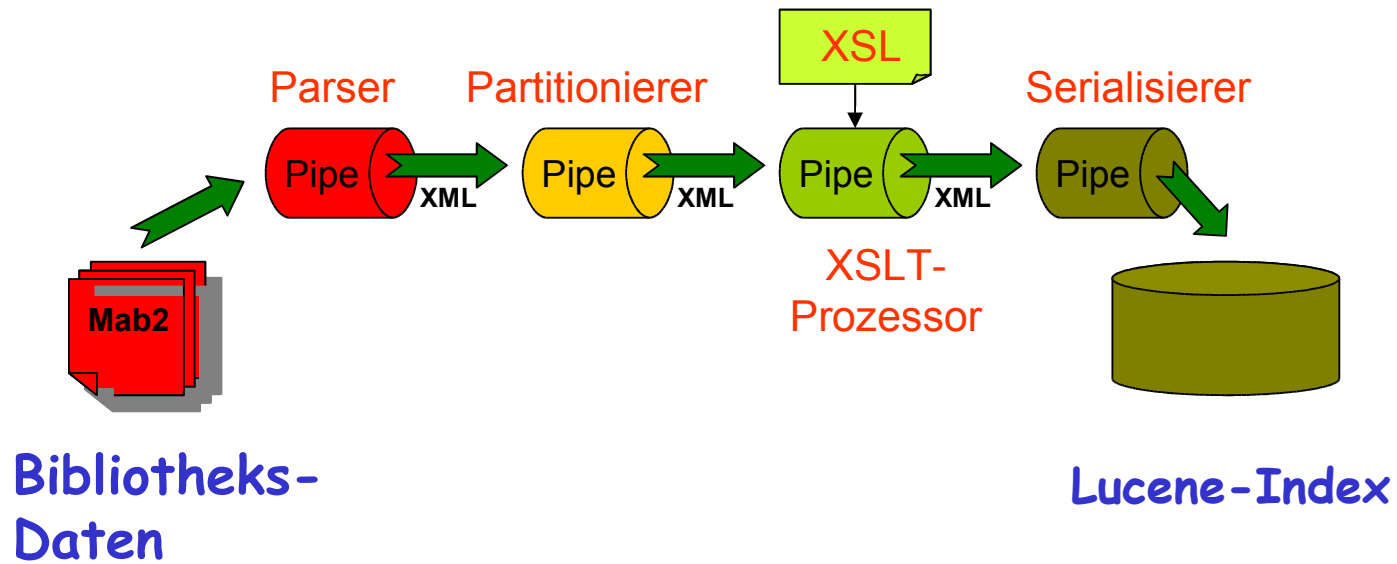


Daten → XML Datensätze → Vorformatiertes XML → Valides XML → Ausgabe

## XML-Pipeline zur Metadatenkonversion

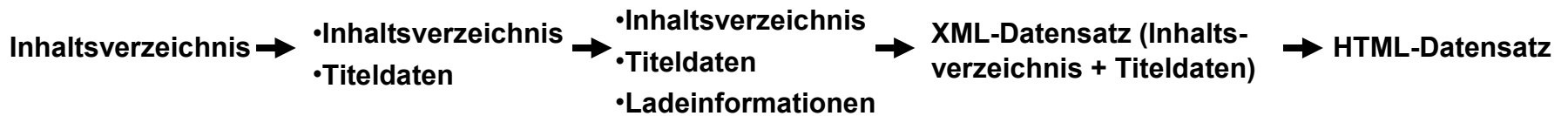
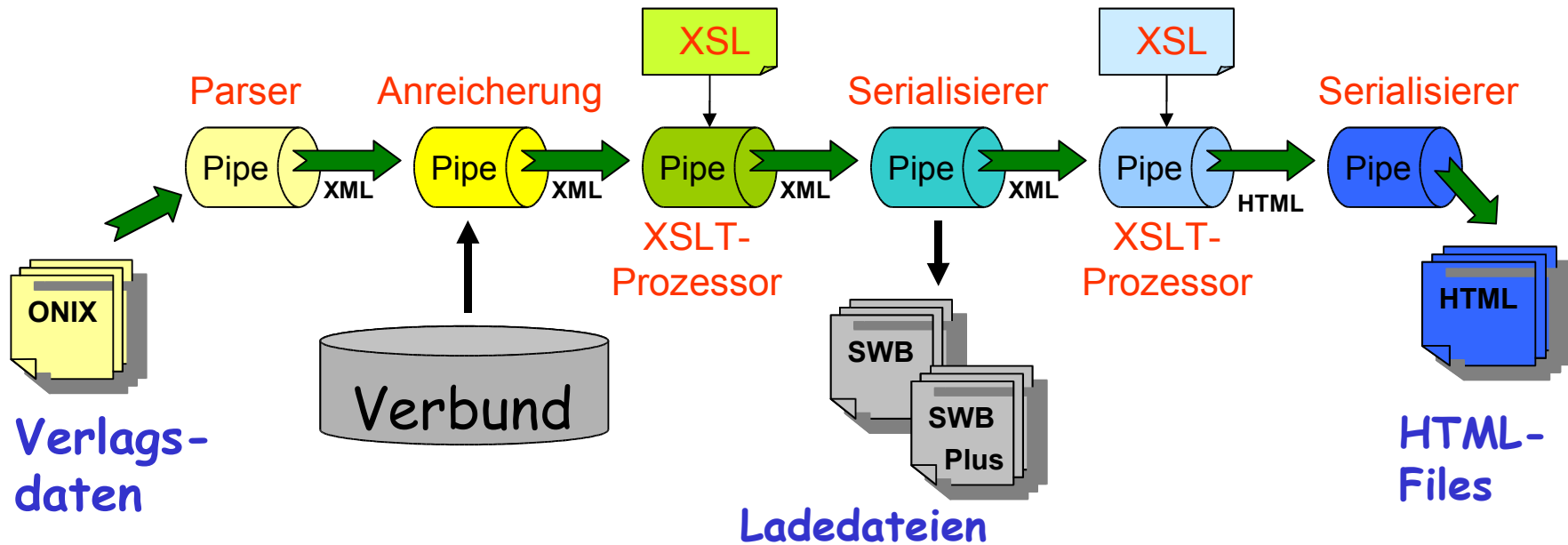
- Die Quelldaten werden zunächst in ein XML-basiertes Format gebracht. Die Ausgabe der ersten Pipe ist immer ein **XML-Datenstrom**.
- Auf diesem Datenstrom werden **Konversionen** durchgeführt, welche die Daten in das gewünschte Ausgabeformat transformieren.
- Mögliche **Hilfsschritte**: Aufspaltung eines großen Dokuments in viele kleinere / Validierung eines Zwischenformats / Zusammenfügen kleinerer Dokumente zu einem größeren / Serialisierung eines Zwischenproduktes / Abfrage einer Datenbank usw.
- Die letzte Pipe wandelt den Datenstrom in das gewünschte Endformat. Typischerweise wird der Datenstrom dabei **serialisiert**.

## Anwendungsfall 2: Projekt BAM



Katalogdaten → Mab XML Datensätze → Datensätze in verarbeitungsfähiger Größe → Indexierbares Format

# Anwendungsfall 3: Catalogue Enrichment



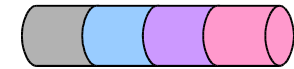
## Pipeline-Ansatz und Metadaten-Registry

- Jeder Teilschritt entspricht einer **Verarbeitungseinheit** (Pipe) in einer Metadaten-Verarbeitungspipeline. Die Verarbeitungseinheiten können durch die Anwendung **beliebig** zu neuen Pipelines **zusammengesetzt** werden.
- Die meisten Teilschritte kommen in vielen Anwendungsfällen vor. Daher sind diese Verarbeitungseinheiten **wieder verwendbar**. Die Software kann neu konfiguriert oder ergänzt werden.
- Jedem **Anwendungsfall** entspricht eine Pipeline.
- **Pipeline-Beschreibungen im XML-Format** können zusammen mit Formatbeschreibungen für Metadaten in einer **Metadaten-Registry** vorgehalten werden.
- So kann **Wissen** über Anwendungsfälle **gespeichert, distribuiert** und **wieder verwendet** werden.

- In der Metadaten-Registry werden alle **Format-, Daten- und Pipeline-Beschreibungen** vorgehalten.
- Der Nutzer kann eine **Pipeline-Beschreibung** durch Anfrage nach Quell- und Zielformat anfordern.
- Alle zur Konstruktion der Pipeline nötigen **Spezifikationen** werden zurückgeliefert.
- Die Pipeline wird **lokal konstruiert** und kann sofort **genutzt** werden.

## Metadaten-Registry

SWB → OAI



Mab2 → BAM



C.E. (ONIX)



XSL

XSL

XSL

XSL

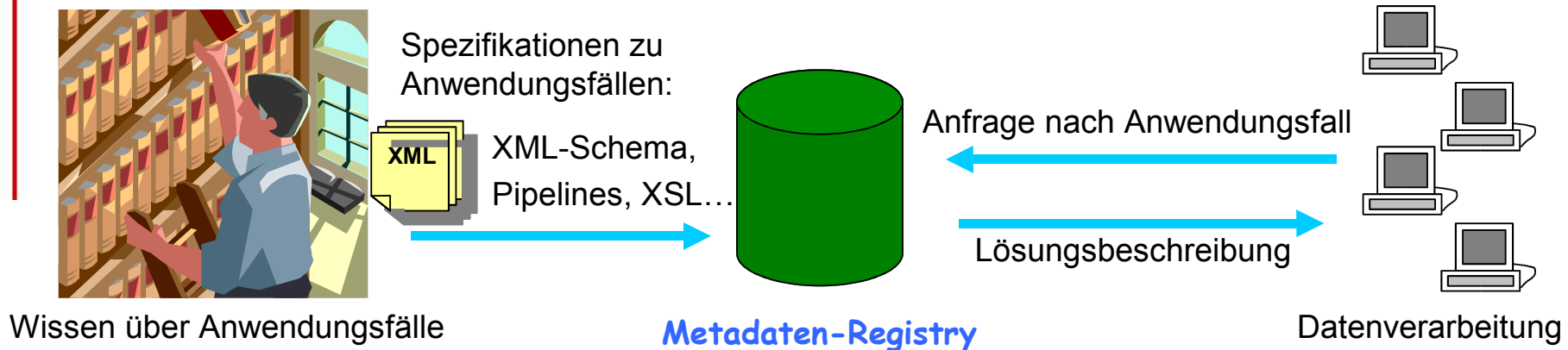
S

S

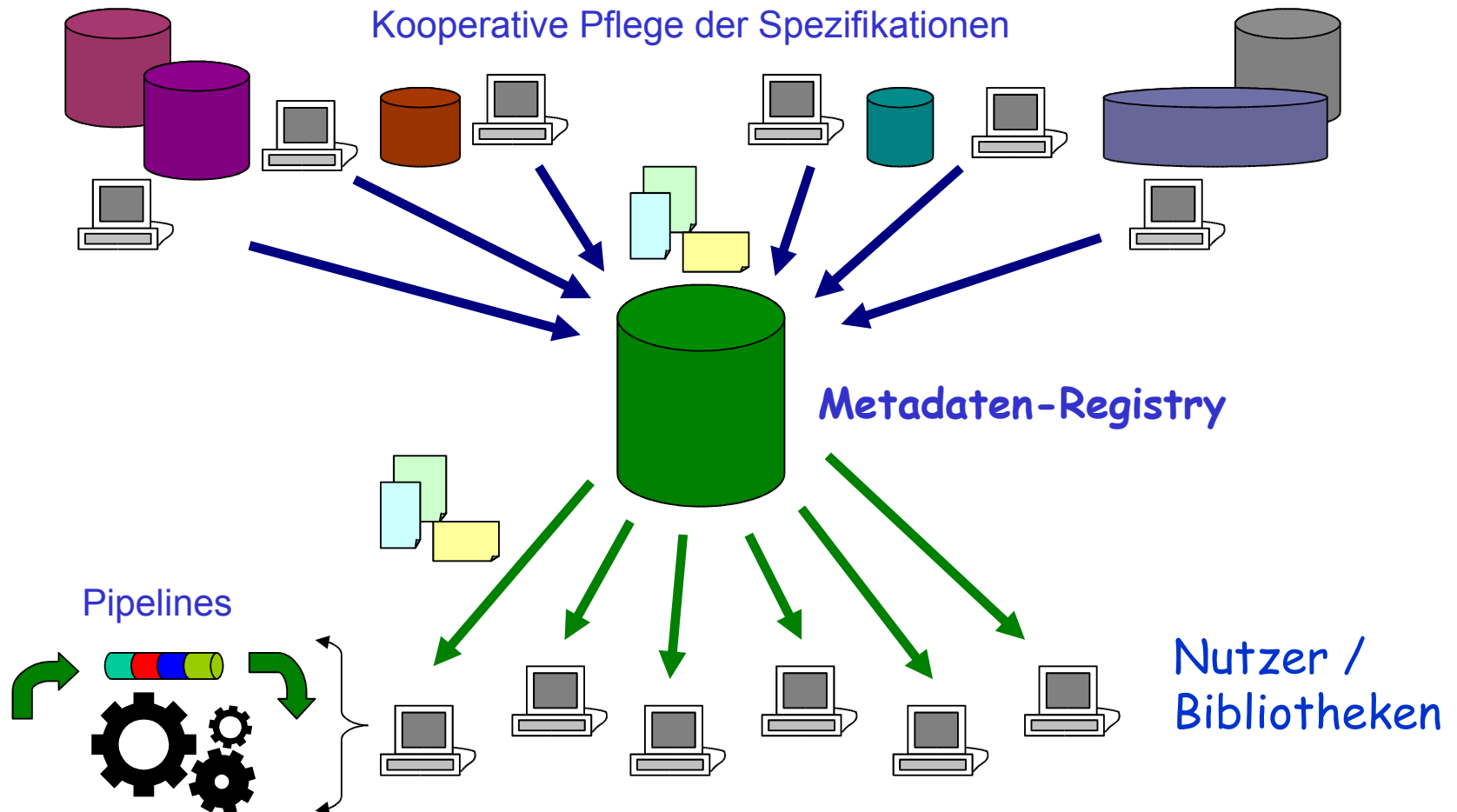
XML  
Schema

# Metadaten-Registry: Kooperative Nutzung

- Die Metadaten-Registry, in welcher die Pipeline-Beschreibungen vorgehalten werden, stellt ein **kooperativ gepflegtes, ständig wachsendes Reservoir** an objektiviertem Wissen über Metadatenverwaltung dar.
- Diese Registry kann mit wenig technischem Aufwand durch alle kooperierenden Institutionen genutzt und erweitert werden, so dass sie zur **nachhaltigen Grundlage** für **Mehrwert durch technisches Wissen** wird.



# Metadaten-Registry: Kooperative Administration



Vielen Dank für Ihre Aufmerksamkeit.

Fragen:

Bibliotheksservice-Zentrum Baden-Württemberg

[Stefan.Hohenadel@bsz-bw.de](mailto:Stefan.Hohenadel@bsz-bw.de)