

# entityXML

## Ein Speicher- und Austauschformat (für die GND)

Uwe Sikora



Der Hintergrund

# Warum gibt's entityXML?

# Die Ausgangssituation: Eine Menge Fragen

- Konzeption einer GND-Agentur für das NFDI Text+ Konsortium
- “Wie können wir Daten mit der GND austauschen?”
- “Wie dokumentieren wir unsere Recherche und die daraus entstandenen Daten als Forschungsdaten?”
- “Wie können wir Daten erschließen, die von der GND nicht unterstützt werden?”
- “Wie können wir Eingabeverfahren beschleunigen, um schneller an GND-URLs zu kommen?”

Das Werkzeug

# Was ist entityXML?

## ... ein Werkzeugkasten

Markupformat auf XML Basis, das als **Speicher-** und **Austauschformat** den GND-Ansprüchen genügt und flexibel auf Bedürfnisse von Forschungsprojekten reagiert.

- Konzeptstudie auf dem Weg zum BETA Status
- natürlich Open Source
- oXygen Framework

entityXML **GitLab**: <https://gitlab.gwdg.de/entities/entityxml>

entityXML **Handbuch**: <https://entities.pages.gwdg.de/entityxml/#>

## entityXML als Speicherformat

- unabhängig von einer komplexen Speicherlösung (z.B. Erschließungssysteme, Datenbanken etc.) in **XML-Dateien**
- Daten bleiben als **Forschungsdaten** transparent und lassen sich in Forschungsdatenmanagementkonzepte integrieren.
- Anreicherung mit projektspezifischen Daten durch **Custom Namespaces**
- Dokumentation von Bearbeitungsständen
- **Entkopplung** nicht-kompatibler Zeitpläne (GND-Einspeisung vs. Forschung)

## ... und als Austauschformat

- Implementierung von Klassen und Properties der **GND-Ontologie in XML**
- **Validierung** auf Grundlage des entityXML Schemas (grundständige Qualitätssicherung)
- **Konversion** von entityXML Daten in unterschiedliche Formate und XML-Dialekte (Integrierbarkeit)
- **Steuerung des Workflows** einer GND-Agentur

Das Design

# Wie sieht entityXML “unter der Haube” aus?

# Die drei Grundpfeiler

- Handbuch, Kapitel “Ressourcen”:  
<https://entities.pages.gwdg.de/entityxml/#resources/>
1. **Validierbarkeit** durch ein dokumentiertes maschinenlesbares Datenschema (Relax NG & Schematron)
  2. **Anwenderorientierte Eingabe** durch den oXygen Autor-Modus
  3. **Integrierbarkeit** in andere System durch Transformationen (XSLT & XQuery)

# Vielen Dank für die Aufmerksamkeit

Gibt's *Fragen, Anregungen,  
Freude, Lob, Kritik, Skepsis,  
Bedenken?*

...

immer her damit!



<http://ottoazubiblog.de/wp-content/uploads/2012/01/Lernberg.jpg>