Dr. Manolis Ulbricht
E-Learning Projekt „Digitale Erschließungs- und Editionsmethoden“ (2019–2021)

Fachbereiche Geschichts- und Kulturwissenschaften &
Philosophie und Geisteswissenschaften

Freie Universität Berlin

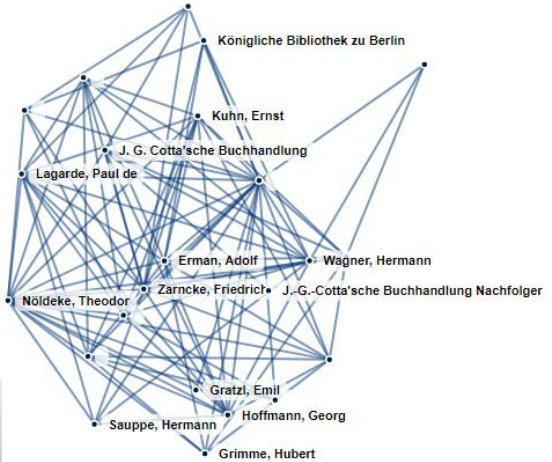
Grundkurs

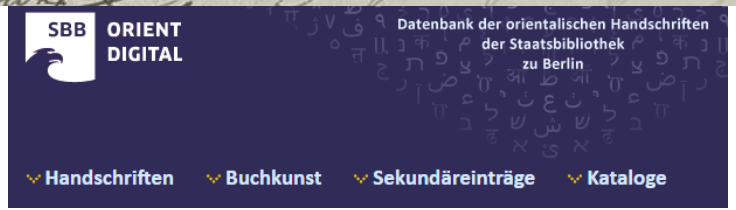
# Digitale Erschließungs- & Editionsmethoden: vom Handschriftendigitalisat zur Digitalen Edition

## Sitzung 2: Das XML-Universum

# Einführung: Digital Humanities und Digitale Editionen




# XML im Zentrum

XML bezeichnet ein Verfahren, um Texte auszuzeichnen und Informationen zu kodieren. Es ist entwickelt worden, um die Strukturen eines Dokumentes kenntlich zu machen und für eine Verarbeitung mittels des Computers vorzubereiten, indem Kodierungen (Auszeichnungen) in einen laufenden Text eingefügt werden.

- ein offener Standard
- einfache „Lingua franca“ für Inhalte und Strukturen
- Inhaltsbeschreibungssprache
  (keine Programmiersprache und keine Seitenbeschreibungssprache)

# Das XML-“Universum“

- XML – eXtensible Markup Language
  - Metasprache: textbasiert, strukturiert
- XML-Standards und XML-Vokabularien
  - aus XML abgeleitete Untersprachen, z.B. XHTML, TEI
- Hilfstechnologien
  - für den Zugriff auf XML-Dokumente (XPath), zur Datenabfrage (XQuery) oder zur Verknüpfung von XML-Ressourcen (XLink)

# XML Auszeichnungssprachen

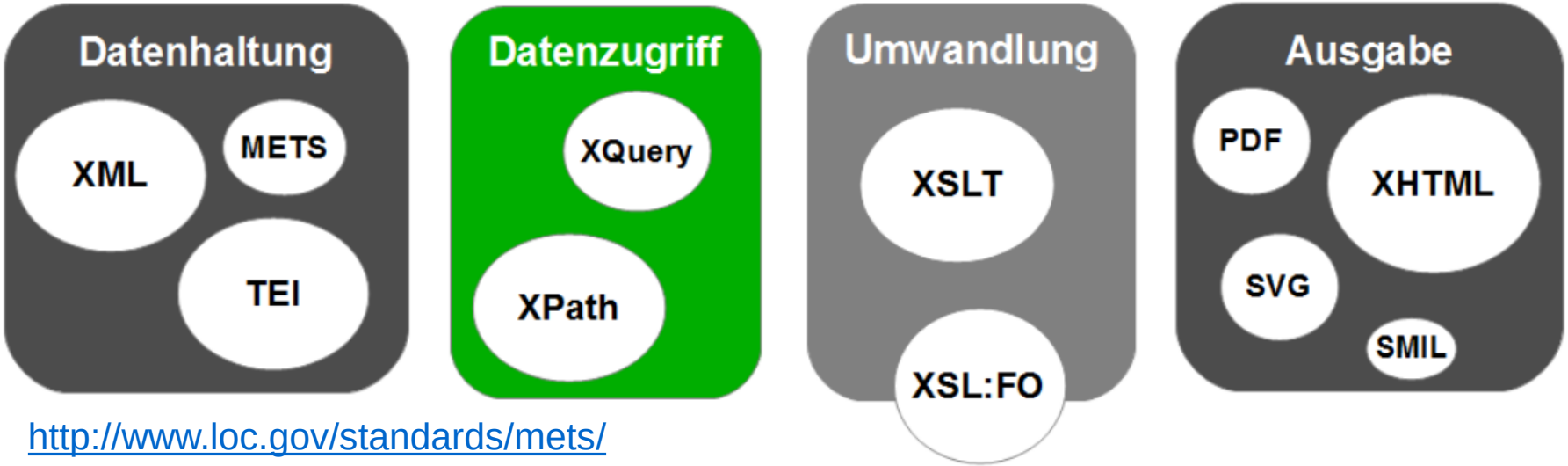
XML als flexible Auszeichnungssprache

Konkrete Auszeichnungssprachen, die durch Schemata kodifiziert sind und dann zur Auszeichnung von Texten / Beschreibung von Daten eingesetzt werden, sind Anwendungen von XML.

Auf XML basieren:

- TEI (Text Encoding Initiative) für Kodierung von Texten
- XHTML als XML-konforme Variante der Sprache für Webseiten
- DocBook für die Erstellung von Dokumentationen
- SVG (Scalable Vector Graphics) für die Beschreibung von Vektorgrafiken

# X-Technologien: Workflow




http://www.loc.gov/standards/mets/

Vgl. Ulrike Henny, https://www.i-d-e.de/wp-content/uploads/2014/07/X-Technologien.pdf

# Wiederholung: XML Grundlagen

- Verwendung der gleichen Zeichen (Text oder beschreibende Daten) für die Kodierungen
- Anfang und Ende des Markups: <Auszeichnung>Text</Auszeichnung>
- Element: Einheit aus Tag und allem, was zwischen den Tags steht
- Elemente können mit Attributen um weitere Informationen ergänzt werden

  <text sprache="deutsch">, ein Element kann viele verschiedenen Attribute enthalten, die durch Leerzeichen getrennt werden

# XML Grundlagen

Schachtelung: XML-Elemente können andere Elemente enthalten, aber: Alle Elemente müssen vollständig in einem anderen Element enthalten sein. Diese Schachtelung führt dazu, dass

- die Struktur der Kodierung strikt hierarchisch ist.
- alles in einem obersten Element enthalten ist, das man Wurzelelement (root) nennt.
- man ein XML-Dokument als wohlgeformt bezeichnet, wenn alle Tags geschlossen sind und sich nicht „überlappen“.

Ein XML-Dokument ist valide, wenn es korrekt strukturiert ist und nur die in der Struktur erlaubten Tags und Attribute verwendet.

# XML Grundlagen

Neben dem Wurzelelement gibt es noch die sog. Präambel (Prolog), der vor dem Starttag des Wurzelelementes steht (enthält mindestens XML-Deklaration und weitere optionale Angaben, z.B. welchem Regelwerk (Schema) das Dokument entsprechen soll).

```
<?xml version="1.0" encoding="UTF-8"?>
```

Parser sind Programme, die XML-Daten einlesen und prüfen.

# XML Schemata

- Valide wird ein XML-Dokument bezeichnet, wenn es den Regeln eines Schemas (z.B. DTD [Document Type Definition]) entspricht.
  XML-Schemata werden über Namensräume (namespaces) einem Dokument zugewiesen.

- Namensräume werden im Wurzelelement eines XML-Dokumentes deklariert und mit Hilfe von Attributen abgekürzt. XML-Namensräume werden benutzt, um in einem einzelnen Dokument mehrere XML-Sprachen zu mischen.
  - Beispiel:

    xmlns:tei="http://www.tei-c.org/ns/1.0" erlaubt es, im Dokument Elementnamen mit dem Präfix **tei:** zu versehen. Ein solches Präfix legt fest, dass das Element aus dem Namensraum der TEI stammt.

    Beispiele s. unter: https://de.wikipedia.org/wiki/Liste_der_XML-Namensr%C3%A4ume

# „Nichts ist so beständig wie der Wandel.“

Heraklit von Ephesus
(um 520 v. Chr.; † um 460 n. Chr.)

Fortsetzung folgt …

Freie Universität Berlin

Fachbereiche Geschichts- und Kulturwissenschaften &
Philosophie und Geisteswissenschaften

**Die vorliegenden Folien sind Ergebnis des E-Learning Projektes „Digitale Erschließungs- und Editionsmethoden in den Philologien des christlichen Orients: vom Handschriftendigitalisat zur Digitalen Edition (</xml>)“ (2019–2021)**

Eingeworben aus Fördermitteln des E-Learning Förderprogramms der Freien Universität Berlin mit einem Finanzvolumen von insgesamt 41.440 € (davon eingebrachter Eigenanteil 17.940 €).

Durchgeführt in Kooperation mit den Fachbereichen Geschichts- und Kulturwissenschaften (GeschKult) sowie Philologie und Geisteswissenschaften (PhilGeist) der Freien Universität Berlin.

Finanziell und personell unterstützt vom Center für Digitale Systeme (CeDiS) im Rahmen der Projektmaßnahme „Exploring Data Literacy“ (2019–2020), Bestandteil des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projektes „Learning Environments Online“ (LEON).

Mitfinanziert von der Professur Semitistik (Seminar für Semitistik und Arabistik) sowie vom Fachbereich Philologie und Geisteswissenschaften (PhilGeist) der Freien Universität Berlin.

**Projektleiter Dr. Manolis Ulbricht**

Wissenschaftlicher Mitarbeiter (2016–2019) an der Professur Byzantinistik/ Institut für Griechische und Lateinische Philologie (FU Berlin),
sowie Feodor Lynen-Forschungsstipendiat (2019–2021) der Alexander von Humboldt-Stiftung.

**Dr. Manolis Ulbricht**
**E-Learning Projekt „Digitale Erschließungs- und Editionsmethoden“ (2019–2021)**

Freie Universität Berlin

**Fachbereiche Geschichts- und Kulturwissenschaften &**
**Philosophie und Geisteswissenschaften**

## Projektbeschreibung und -ziele

Zentrales Ziel des E-Learning-Projekts „Digitale Erschließungs- und Editionsmethoden in den Philologien des Christlichen Orients: vom Handschriftendigitalisat zur Digitalen Edition (</xml>)“ war die Entwicklung von Blended-Learning-Konzepten sowie digitalen Lehr- und Lernmaterialien für einen Kurs, der über zwei Semester Anwendungsmöglichkeiten der Digital Humanities im Bereich der Philologien aus den Kultur- und Geisteswissenschaften aufzeigt und vermittelt. Zu den erarbeiteten Unterrichtsmaterialen zählen Foliensätze für den Grund- und Aufbaukurs, die Kenntnisse zu XML und TEI für die Erschließung und Auszeichnung textueller Daten, deren Verarbeitung (z. B. in XML-Datenbanken) und zu Abfragesprachen (XPath) vermitteln.

Weitere Lehrmaterialien wie Lehrvideos oder Selbsttests befähigen die Studierenden, sich eigenständig mit Fragen zu beschäftigen, die auf der Schnittstelle zwischen klassischer philologischer Arbeit (Studium von Manuskripten, Auswertung von Handschriftendigitalisaten, Print-Editionen etc.) und zukunftsorientierter digitaler Implementierung (Digitale Editionen, Erstellung von Datenbanken etc.) liegen.

Die Lehrmaterialien wurden erarbeitet von den Dozenten Dr. Ute Pietruschka, Dr. Marco Büchler und Daniel Haas M.A. Unterstützt wurden sie von den studentischen Hilfskräften Noah Witte-Winnett, Alexandros Boukevalos und Noël van den Heuvel sowie dem studentischen Mitarbeiter Sandro Morgenstern. Besonderer Dank gilt Dr. Victoria Mummelthei (Seminar für Semitistik und Arabistik, Fachbereich Geschichts- und Kulturwissenschaften/ FU Berlin) für die stete Unterstützung.

Erarbeitung der Folien und didaktische Konzeption: Dr. Ute Pietruschka

Endredaktion: Sandro Morgenstern

Stand: 09.08.2021



CeDiS Learning Environments Online