

Lernportfolio

Dokumentation meiner Lernreise im Teilmodul Datenformate im Studiengang Bibliotheksinformatik

Modul: Schnittstellen und Datenformate
Studiengang: Bibliotheksinformatik
Hochschule: Technische Hochschule Wildau
Student:in: Patryk Gadziomski
Matr.Nr.: 50226324
Dozent:in: Tracy Arndt
Datum: 28.01.2026
Online-Version: [LCS-learning-portfolio.github.io](https://github.com/patrykgadziomski/LCS-learning-portfolio)

Willkommen auf meinem Lernportfolio!

Dieses Portfolio dient der Dokumentation meiner Lernreise im Teilmodul *Datenformate* im Studiengang Bibliotheks-informatik an der Technischen Hochschule Wildau. Hier sammle ich Notizen und Erkenntnisse zu verschiedenen Themen aus dem Teilmodul.

Aber was ist ein Lernportfolio?

Das Lernportfolio bezeichnet eine Sammlung von Dokumenten (z. B. Arbeitsblätter, Protokolle, Essays) oder Materialien [...], die zu einem bestimmten Zeitpunkt unter Bezug auf ein inhaltlich umrissenes Lehrgebiet die Leistungen [...] dokumentieren und charakterisieren.[\[1\]](#)

Das Besondere an diesem Lernportfolio ist die Integration eines Knowledge Graphs auf der rechten Seite. (Aus Gründen der Anzeige muss der Knowledge Graph die Grenzen des Fensters einhalten. Leider wird es dadurch etwas voll bzw. unübersichtlich, aber hoffentlich hält es sich in Grenzen.) Der Graph visualisiert die Struktur und Beziehungen zwischen den verschiedenen Notizen, ähnlich wie in Obsidian. Keine Werbung! Aber eine sehr tolle Software ;)

[→ Obsidian](#)

Jeder Knoten repräsentiert eine Notiz, und die Verbindungen zeigen, wie die Themen miteinander zusammenhängen. Klicken Sie auf einen Knoten im Graph, um direkt zum entsprechenden Artikel zu springen!

Der Aufbau dieses Lernportfolios ist in zeitliche Abschnitte unterteilt, je nach den Terminen, an denen die Vorlesungen stattfanden. In diesen wird kurz zusammengefasst, was an diesen Tagen stattfand und was ich gelernt habe, in Bezug auf die Themen und weitere Erkenntnisse. Darauf folgen die zu bearbeitenden Pflichtaufgaben sowie ein abschließendes Fazit.

WICHTIG: Bei Pflichtaufgabe 2 sollte nur eine der drei Unteraufgaben bearbeitet werden. Aus Interesse am Thema wurden alle Unteraufgaben bearbeitet, jedoch wurde eine zur Bewertung für die Dozentin markiert! AUßERDEM befindet sich im Repository ebenfalls eine PDF-Version des Lernportfolios, ohne die Pflichtaufgaben, damit die Länge ersichtlicher wird. Das Dokument wurde anhand des [Orientierungsrahmens für die formale Gestaltung einer schriftlichen Beleg- oder Abschlussarbeit](#) der TH-Wildau angefertigt.

[→ PDF-Version des Lernportfolios ohne Pflichtaufgaben](#)

Bei der technischen Umsetzung dieses Lernportfolios wurden Claude (Anthropic) und Euria (Infomaniak) eingesetzt, um die grundlegenden Funktionalitäten zu generieren. Weitere Funktionen, Strukturierung und Datenkuratierung wurden durch den

Entwickler – mich :) – durchgeführt. Der Text in diesem Portfolio wurde ohne jegliche Hilfe von Sprachmodellen erstellt.

Folgende Fragen werden am Ende jeder Vorlesung als Zusammenfassung beantwortet und als zusammenfassendes Fazit in Form einer neuen Seite:

- Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?
- Was war Ihr größter Lerngewinn?
- Was war besonders überraschend? Was hat Sie irritiert?
- Wo sehen Sie für sich noch Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?

Aber ich will Sie nicht länger aufhalten. Stürzen Sie sich in die Welt der Daten und viel Spaß beim Erkunden meines Lernportfolios! :)

Vorlesung 1

Dies war der dritte Tag des Studiums der Bibliotheks-informatik an der TH Wildau und zugleich der erste Tag des Moduls Schnittstellen und Datenformate. Das Modul ist in zwei Teile gegliedert: Der erste Teil, der von Tracy Arndt unterrichtet wird, behandelt Datenformate.

Die Vorlesung begann mit den üblichen organisatorischen Informationen und der Vorstellung. Anschließend wurden die Ziele des Teilmoduls sowie die einzelnen Blöcke vorgestellt. Der erste Block (Tag 1 und 2) befasst sich mit Terminologie, Heterogenität, Interoperabilität und Organisation von Metadaten. Natürlich wurden auch die Prüfungsleistungen und die dazugehörigen Bewertungskriterien erläutert. Das Lernportfolio als Prüfungsleistung finde ich persönlich sehr gut. Bereits im Bachelorstudium hatte ich häufig diese Form der Prüfungsleistung. Das Schöne daran ist für mich die Freiheit, die den Studierenden eingeräumt wird, ihr Portfolio auf individuelle Weise zu gestalten. Außerdem bin ich der festen Überzeugung, dass man nur durch Reflexion des Gelernten wirklich erkennt, ob man die Inhalte eines Moduls oder Fachs verstanden hat. Manchmal führe ich selbst ein Lerntagebuch für Themen, die ich mir selbst aneigne, da ich mir dabei die Frage stellen muss: Was habe ich eigentlich gelernt?

Zunächst klärten wir, was Daten überhaupt sind, also die Terminologie. Das war für mich nichts Neues, da ich im Bachelor einen Schwerpunkt auf Daten gelegt hatte. Auch Metadaten waren mir bereits bekannt (Metadatenmanagement wurde im Bachelorstudium als Wahlmodul belegt. Damals habe ich mich ausgiebig mit dem Datenformat NeXus im Rahmen einer Hausarbeit beschäftigt). Bibliothekarische Metadaten waren mir jedoch nur teilweise vertraut. Ich hatte zwar PICA und MARC im Studium kennengelernt, aber weder BibFrame noch Linked-Data-Formate. Linked Data finde ich äußerst interessant und möchte es überall dort einsetzen, wo es möglich ist, um es wirklich zu verstehen und zu erfahren, wie es funktioniert. Wie sich zeigt, bekomme ich dazu auch die Gelegenheit. Was das Thema Heterogenität

angeht: Es war ein Thema, das mir erst in der Arbeitspraxis wirklich bewusst wurde. Im Studium wurde immer betont, dass man Standards einhalten muss, doch eine tatsächliche Rolle spielten sie erst, als man damit in Berührung kam.

Anschließend folgte eine Gruppenarbeit, in der wir uns mit verschiedenen Begriffen auseinandersetzen sollten. Dazu wurden uns Quellen zur Verfügung gestellt. Meine Gruppe und ich entwarfen daraufhin eine Visualisierung. Die Ergebnisse wurden anschließend präsentiert. Es wurde lebhaft darüber diskutiert, was genau ein Datenmodell ist und ob bzw. wie überall Standards auftreten können. Es war interessant zu sehen, dass selbst Menschen, die täglich mit dieser Thematik arbeiten, Schwierigkeiten haben, alles präzise zu beschreiben. Das war für mich zum Teil ein Aha- und ein Wow-Moment. Früher dachte ich immer, dass man, sobald man in einem Bereich arbeitet, diesen vollständig überblickt und versteht. Mittlerweile weiß ich es besser und dass es eine ständige Lernreise ist. In der Gruppe haben wir viel diskutiert und die Abbildung komplett neu gestaltet.

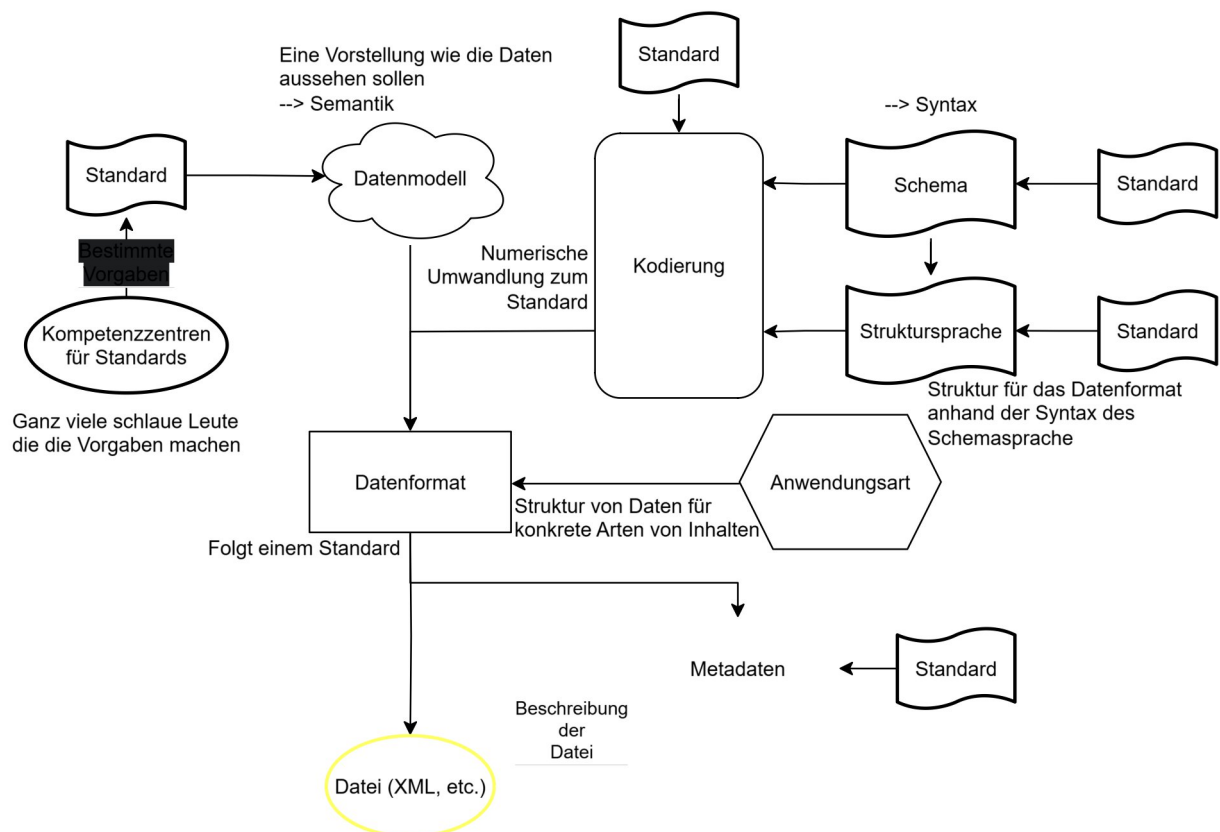


Abbildung 1: Gruppenarbeitsergebnis im Teilmodul Datenformate, Visualisierung zur Beschreibung verschiedener Datenbegriffe und ihrer Zusammenhänge

Spannend fand ich die Interoperabilität von Daten. Damit habe ich bisher nur wenige Berührungspunkte gehabt, was ich gerne ändern möchte. Ich fand die Konvertierung zwischen Formaten schon immer faszinierend. Ein Skript zu schreiben, das ein Format in ein anderes konvertiert, hat etwas beruhigendes. Dabei merke ich, dass ich statt vorhandener Software lieber eigene Skripte schreibe, was Vor- und Nachteile hat. Als auf der Folie Kenntnis von Metadaten-Standards [...] auftauchte,

erinnerte ich mich an ein Gespräch mit einer Bekannten über die Datenspeicherung physikalischer Vorgänge (Flugfahrttechnik) und ihr Problem, dass sie nicht wusste, wie sie solche Daten speichern und anschließend optimal ins System weitergeben sollte. Hier fehlte das Wissen über geeignete Datenformate, insbesondere über spezialisierte Formate für physikalische Vorgänge oder Experimente. Was man nicht kennt, kann man eben nicht nutzen.

Vor allem die Tatsache, dass es so viele Standards gibt und man sich auf keinen einigen kann (aus verschiedenen Gründen), fand ich sehr spannend. Im Modul Künstliche Intelligenz arbeite ich gerade daran, einen Linked-Data-Datensatz aufzubauen (dazu später mehr) und hätte beinahe einen eigenen Standard entwickelt, weil ich dachte: Ah, wieso nicht? Dann kann ich es genau an meine Wünsche anpassen! Am Ende habe ich mich jedoch für bestehende Standards entschieden, da sie alles enthielten, was ich benötigte, ich musste sie mir nur genauer ansehen.

In meinem Alltag kann ich durchaus behaupten, dass ich viel mit Daten zu tun habe. Das sollte eigentlich jeder behaupten können, da wir von Daten umgeben sind. Im beruflichen Kontext arbeite ich mit ISBD und MARC, da ich in meiner Bibliothek die Katalogisierung mit Koha durchführe. In meiner anderen Tätigkeit bin ich mit Webentwicklung beschäftigt, weshalb ich auch mit verschiedenen Datenformaten umgehen muss. Das alles war mir bereits aus dem Bachelorstudium bekannt. Dieser erste Tag war für mich somit eine nette Auffrischung, Neues habe ich nicht viel gelernt. Mein Hunger auf Linked Data ist jedoch gewachsen :)

Zusammenfassend kann man sagen:

- **Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?**

Inhaltlich habe ich mich mit den Grundlagen wie Terminologie, Heterogenität, Interoperabilität und Organisation von Daten beschäftigt, dabei habe ich mein Wissen zu Metadaten, PICA und MARC aufgefrischt und erstmals BibFrame sowie Linked Data kennengelernt. Persönlich hat mich überrascht, wie schwer es selbst erfahrenen Fachleuten fällt, Begriffe wie Datenmodell präzise zu beschreiben. Das hat mir gezeigt, dass Lernen nie wirklich fertig ist und dass es okay ist, nicht alles sofort zu verstehen. Ansonsten habe ich bestätigt bekommen, dass Reflexion der Schlüssel zum Verständnis ist. Genau deshalb führe ich auch weiterhin mein Lerntagebuch, weil es mich zwingt, mir selbst zu fragen: Was habe ich eigentlich gelernt?

- **Was war Ihr größter Lerngewinn?**

Mein größter Lerngewinn liegt in der Erkenntnis, dass Standards nicht nur theoretisch, sondern praktisch relevant sind und dass es sinnvoll ist, bestehende Standards zu nutzen, statt eigene zu entwickeln, selbst wenn sie zunächst nicht perfekt erscheinen.

- **Was war besonders überraschend? Was hat Sie irritiert?**

Überraschend fand ich, dass auch erfahrene Menschen Schwierigkeiten haben, grundlegende Begriffe wie Datenmodell präzise zu definieren. Das hat mich aus meiner Annahme gerissen, dass Expertise = vollständiges Verständnis bedeutet. Irritierend fand ich die Tatsache, dass es so viele Standards gibt und keine Einigung besteht.

- **Wo sehen Sie für sich noch Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?**

Ich möchte mich intensiver mit Interoperabilität und Linked Data beschäftigen, insbesondere mit praktischen Anwendungen und Konvertierungsprozessen. Ich plane, Linked Data im Modul Künstliche Intelligenz praktisch anzuwenden und dabei bestehende Standards zu nutzen, statt eigene zu entwickeln. Zudem möchte ich meine Skripting-Fähigkeiten weiter ausbauen, um Formate selbst zu konvertieren.

Vorlesung 2

An diesem Tag sind wir tiefer in die Materie der Metadaten eingetaucht und haben mit der Geschichte von PICA und MARC begonnen. Persönlich fand ich das Thema nicht besonders spannend, aber es war dennoch interessant zu sehen, wann diese Formate entstanden sind und wie sich die Anforderungen an Metadaten im Laufe der Zeit gewandelt haben. Ich merke, dass mein Interesse eher der Verwendung von Daten- und Metadatenformaten gilt, nicht ihrer Geschichte.

Das zentrale Thema des Tages war die Metadatenorganisation, ein hochspannendes, aber auch komplexes Feld. Allein die Tatsache, dass es Austauschformate braucht, weil es so viele unterschiedliche Metadatenformate gibt, macht das Ganze kompliziert. Ich habe mich gefragt, warum es nicht ein universelles Daten- oder Metadatenformat für alles geben kann. Doch das ist klar: Man kann nicht alles abbilden. Selbst wenn man ein solches Format ständig erweitern würde, würde es so viele Felder enthalten, dass es aus Speichergründen nicht mehr praktikabel wäre. Es geht vielmehr um aufgabenspezifische Nutzung. Die Vorstellung eines Omni-Formats ist zwar schön, aber unmöglich..

Was man hierbei wunderbar sieht: Metadaten gibt es überall. Egal, in welchem Bereich man arbeitet, man kommt immer mit ihnen in Berührung, auch wenn man es nicht bewusst wahrnimmt. Für mich wurde noch einmal deutlich, wo überall Metadaten vorkommen: in jedem Datenfluss, in jeder Datei, in jeder Kommunikation. Ich habe mir meine alten Projekte im Bachelor noch einmal angesehen und bemerkt, wie oft Metadaten darin vorkamen und wie wenig ich damals darauf eingegangen bin. Jetzt hätte ich gerne die Zeit, alles noch einmal durchzugehen und zu überarbeiten.

Das FRBR-Modell war mir aus dem Bachelor bekannt, aber ich wusste nicht, dass es mehrere Varianten gibt oder weiterentwickelt wurde. Das FRBR-LRM hat mich überrascht. Da es ein sehr theoretisches Modell ist, entsteht für mich wieder das gleiche Problem: Es lässt sich nicht alles abbilden, und man kann sich ewig darüber streiten. Beim Thema ETL musste ich zugeben, dass mir das bereits bekannt war. Im Bachelor hatte ich viele Module aus dem Bereich Data Science belegt.

Dann kam das unglaublich spannende Thema: Linked (Open) Data. Im Studium hatte ich bereits damit zu tun, da sich mein Professor damit beschäftigt, aber ich konnte mir darunter nie etwas vorstellen. Für mich war es eine rein theoretische Vorstellung: ein Web, in dem Daten als Knoten definiert sind, die Verbindungen zu anderen Knoten und somit zu anderen Daten herstellen. Theoretisch verstanden, aber praktisch? Keine Ahnung. Bis jetzt. Jetzt habe ich es verstanden: Es ist ein Knowledge Graph! Ein Knowledge Graph entsteht durch das Verbinden von Daten (Nodes) und das ist Linked Data. Die Idee, dass das Web im Web-3.0-Modell komplett aus aufeinander verweisenden Daten besteht, finde ich unglaublich schön. Allein die Möglichkeit, semantische Anfragen zu stellen, um genau die gesuchten Daten zu finden, ist für die Suche und Suchmaschinen enorm wichtig. Leider wurde diese Entwicklung durch das Aufkommen großer Sprachmodelle gebremst, oder zumindest in den Hintergrund gedrängt. Ich stelle mir ein perfektes Modell vielmehr als eine Verbindung von beidem vor: Das Semantic Web könnte eine unglaubliche Datenquelle sein, auf die KI-Agenten zugreifen könnten. Ich habe dazu einmal eine Visualisierung gesehen, finde sie aber nicht mehr. Also habe ich versucht, sie mit ChatGPTs DALL-E zu generieren.

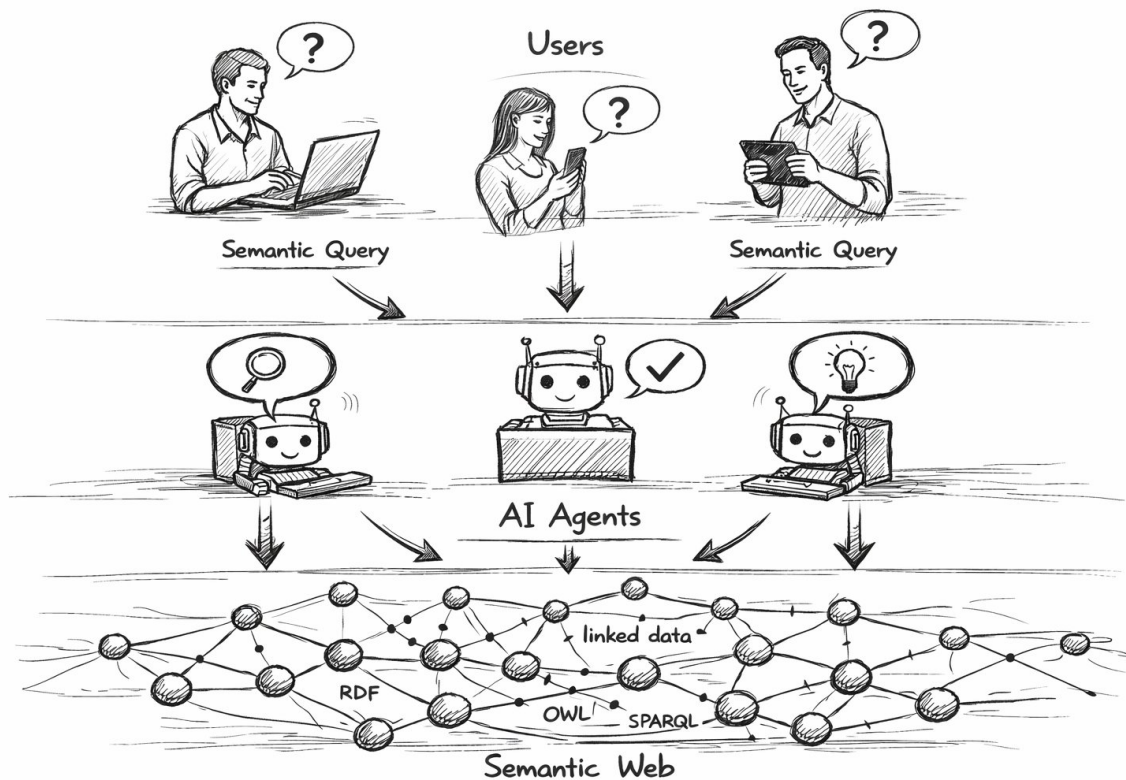


Abbildung 2: Vorstellung eines Web-Models, bei welchem Semantic Web und KI-Agenten harmonisch koexistieren - erstellt mit ChatGPTs DALL-E

Jetzt: Warum glaube ich, dass ich Linked Data bereits seit Jahren nutze?

Ich benutze seit Jahren Obsidian, um meine Notizen zu verwalten. Die Software erlaubt es, Verbindungen zwischen Notizen herzustellen und damit war mir das Prinzip, Informationen zu verknüpfen, bereits sehr vertraut. Aber jetzt habe ich es vollständig verstanden: Die einzelnen Knoten in meinem Kopf zum Thema Linked Data haben sich endlich verbunden :) Ich habe meine Notizen die ganze Zeit wie Linked Data behandelt, ohne es zu wissen. Seitdem nutze ich Standards wie Schema.org, Dublin Core und Friend of a Friend, um meine Daten noch besser zu strukturieren und zu verknüpfen. Jetzt kann ich meine ganzen Notizen wirklich als Linked Data bezeichnen. Da sie aber zum Teil privat sind, nicht als Linked Open Data. Ich habe aber tatsächlich schon darüber nachgedacht meine ganze Aufschriebe und Lerntagebücher, genau wie dieses per GitHub Repository als Linked Open Data zu veröffentlichen. Die Frage bleibt nur ob sich das lohnt, und ob man das nicht evtl. anders tun sollte. Ich bin für Vorschläge offen :) Im Modul Künstliche Intelligenz arbeite ich mit einer Kommilitonin an einem Projekt zum Thema Retrieval Augmented Generation (RAG). Dafür erstelle ich, wie bereits erwähnt, einen Linked-Data-Datensatz. Ziel ist es, am Ende ein RAG-System zu haben, mit dem man im Obsidian-Vault suchen oder Fragen stellen kann. Dabei soll das System nicht nur die Ähnlichkeit zwischen Anfrage und Dokumenten in der Vektor-Datenbank berücksichtigen, sondern auch die Dokumente, die mit diesen verknüpft sind. Sehr interessant war auch die Erkenntnis, dass Linked Data besser für das Training oder die Nutzung von KI-Modellen wie LLMs geeignet ist. Das passt perfekt zu unserem Projekt. Wenn es funktioniert, werde ich es auch für meinen privaten Vault nutzen und kann dann quasi mit mir selbst reden, da er viele Informationen aus meinem Leben enthält. (Der Datenschutz muss natürlich beachtet werden. Ich möchte nicht, dass meine Daten in irgendein LLM gelangen.)

Zusammenfassend kann man sagen:

- **Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?**

Inhaltlich habe ich mich mit der Metadatenorganisation auseinandergesetzt, insbesondere mit der Notwendigkeit von Austauschformaten und der Begrenztheit universeller Formate. Ich habe das FRBR-LRM vertieft und Linked Data endlich verstanden, nicht nur theoretisch, sondern auch praktisch im Kontext von Knowledge Graphs und Obsidian. Persönlich habe ich erkannt, dass ich bereits seit Jahren mit Linked Data arbeite, ohne es zu wissen. Ansonsten habe ich gesehen, wie Metadaten überall vorkommen, auch in meinen eigenen Arbeiten und ich will sie nun bewusster nutzen. Ich habe auch erkannt, dass KI und Semantic Web gut zusammenpassen und will das in meinem Projekt nutzen.

- **Was war Ihr größter Lerngewinn?**

Mein größter Lerngewinn ist die Erkenntnis, dass ich Linked Data bereits in meiner täglichen Praxis anwende und dass ich es jetzt verstanden habe.

- **Was war besonders überraschend? Was hat Sie irritiert?**

Überraschend fand ich, dass ich Linked Data bereits seit Jahren nutze und dass es nicht nur Theorie ist, sondern praktisch anwendbar. Auch das FRBR-LRM hat mich überrascht, weil es weiterentwickelt wurde. Irritiert hat mich die Tatsache, dass große Sprachmodelle die Entwicklung des Semantic Web gebremst haben.

- **Wo sehen Sie für sich noch Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?**

Ich möchte tiefer in Linked Data einsteigen, insbesondere in Praxisanwendungen, Standards (Schema.org, FOAF, Dublin Core) und Integration mit KI-Systemen wie RAG. Ich will außerdem mein RAG-Projekt mit Linked Data erfolgreich abschließen und es dann für meinen privaten Obsidian-Vault nutzen, um mit mir selbst zu reden.

Vorlesung 3

Das Thema *Linked Data* und *Semantic Web* wurde an diesem Tag weiter vertieft und man kann durchaus sagen: Meine Lernkurve in diesem Bereich war exponentiell!

An diesem Tag habe ich endlich verstanden, wie man *Ontologien* aufbaut und *RDF-Standards* verwendet. Zwar hatte ich damit bereits experimentiert, aber erst durch die Übung in der Vorlesung wurde alles viel klarer. Bisher dachte ich immer, Ontologien seien komplexe, textuelle Bäume, was so gesehen stimmt, aber auch wieder nicht.

PS: Vor dem Masterstudium habe ich mir Jobs im Bereich Ontologien angesehen und dachte, sie klingen zwar interessant, aber ich würde es bestimmt nicht hinbekommen, weil mir das Wissen fehlte. Jetzt weiß ich: Mir hat das Wissen gefehlt, aber die Praxis war da. Ich wusste nur nicht, dass das, was ich in Obsidian erstellt habe, mehr oder weniger Ontologien waren.

Ontologien lassen sich als das darstellen, was ich sowieso fast jeden Tag mache, oder zumindest für mein privates Leben oft tue: Ich erstelle Beschreibungen von Objekten und deren Beziehungen zueinander, in Form eines Knowledge Graphs. Das war wie eine Erleuchtung für mich. Anhand dieser Ontologien kann man Objekte beschreiben und dafür gibt es verschiedene Standards. Das war ein großer Lerneffekt und ein noch größerer Aha-Moment. Vor allem durch das Üben mit *WebVOWL* wurde mir klar, wie Ontologien visuell aussehen können.

Zwei Übungen haben mir unglaublich viel gebracht: [Eigene Ontologie](#) und [BIM25 Ontologie](#).

Neben diesem Thema wurde SPARQL kurz angesprochen, das wird aber am nächsten Tag genauer behandelt. Das RDF-Schema als Tripel kannte ich bereits, es war somit leicht verständlich, und ich konnte schnell in die Übungen einsteigen. Was ich nicht kannte, waren die Standards wie FOAF, sowie die RDF-Serialisierung. Doch alles war sehr verständlich.

Für mein Lernportfolio habe ich mich für *JSON-LD* als Datenformat für Linked Data entschieden. Der Grund: die einfache Lesbarkeit, da es auf der Syntax von JSON basiert und ein modernes Format ist. Am Anfang finden Definitionen im Context statt, unter anderem auch von meinem eigenen Standard, da ich für das Portfolio teilweise sehr spezifische Beziehungen brauchte, oder es einfach ausprobieren wollte, wie es mit einem eigenen funktioniert. Ich hätte natürlich Standards wie Schema.org verwenden können, aber der Lerneffekt war so größer. Der weitere Aufbau des Datensatzes erfolgte wie bei einer normalen JSON-Datei, mit der Ausnahme, dass die Keys die Syntax *standard:relation* verwenden. Die Erstellung dieses Datensatzes sah ich als sehr wichtige Übung an um mit einem echten Linked-Data-Format zu arbeiten und vor allem, um die Visualisierung als Knowledge Graph zu üben.

Sehr schöne Ressourcen, die genannt wurden, möchte ich hier noch einmal festhalten:

- [RDF Grapher](#)
- [Visualize RDF graph linked data as a connected diagram](#)
- [Online RDF converter and validator to JSON-LD, Microdata, Turtle, TriG, RDF-star or any other serialization format](#)
- [IDLab Turtle Validator](#)

Und jetzt noch einmal zum Thema Jobsuche: Jetzt, da ich endlich verstanden habe, was Linked Data und Ontologien sind, habe ich mir erneut Jobs in diesem Bereich angesehen. Hier eine kleine Sammlung, damit ich später noch weiß, was in diesem Bereich möglich ist. (Die Links könnten mittlerweile ungültig sein, da die Stellenanzeigen abgelaufen sein könnten.)

- **Wissenschaftliche:r Mitarbeiter:in (w/d/m) Schwerpunkt Wissensgraph** - [Link](#) - Dabei geht es um den Aufbau und Pflege eines Wissensgraphen sowie der Datenmodellierung und Ontologieintegration. Klingt mittlerweile nicht mehr so abstrakt für mich :)
- **Wissenschaftliche:r Mitarbeiter:in (w/d/m) Schwerpunkt ETL-Prozesse & Ontologie-Entwicklung** - [Link](#) - Hier lese ich viel über Transformationsmodellen mit Datenformate, die mir leider noch nicht ganz bekannt vorkommen, aber auch der Erarbeitung von Ontologien!
- **Knowledge Engineer (m/w/d)** - [Link](#) - Diese Stelle, auch wenn Worte wie Ontologien, Metadaten und Wissensgraphen gefallen sind, ist viel

wirtschaftlicher und vielvältiger aufgestellt. Hier muss man auch Data Science Erfahrung haben

- **Senior Developer (f/m/d) for Knowledge Graph Research** - [Link](#) - Eine sehr interessante Stelle, mit einer riesigen akademischen Verantwortung. Nichts für mich (noch nicht).
- **Knowledge and Ontology Engineer - Insurance (m/f/d)*** - [Link](#) - Wieder eine sehr wirtschaftliche Stelle, die auch wieder sehr im Bereich von Data Science liegt.

Was mir bei der Recherche aufgefallen ist: Viele Stellen sind sehr verwandt mit denen von Data Engineer oder Data Scientist. Mir ist klar, dass man Ontologien und Knowledge Graphs auch in diesen Bereichen einsetzt, aber man muss stark darauf achten, zwischen bibliothekarischen und wirtschaftlichen Stellen zu unterscheiden. Ich fand es schön, die Stellenanzeigen jetzt lesen zu können, ohne das damalige Gefühl der Überforderung. Jetzt habe ich das Wissen aus dem Teilmodul und muss es nur noch in der Praxis umsetzen und meine Fähigkeiten in diesem Bereich weiter ausbauen.

Zusammenfassend kann man sagen:

- **Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?**

Inhaltlich habe ich verstanden, wie Ontologien aufgebaut und RDF-Standards verwendet werden, insbesondere mit WebVOWL und JSON-LD. Ich habe auch SPARQL und RDF-Serialisierungen kennengelernt und konnte mein Wissen in Übungen anwenden. Persönlich habe ich erkannt, dass ich bereits praktische Erfahrung mit Ontologien habe, durch meine Arbeit mit Obsidian. Das hat mein Selbstvertrauen gestärkt. Ansonsten habe ich Ressourcen gesammelt (RDF Grapher, Turtle Validator etc.) und Jobs im Bereich Ontologien und Knowledge Graphs recherchiert und dabei festgestellt, dass ich jetzt die Stellenanzeigen ohne Angst lesen kannst.

- **Was war Ihr größter Lerngewinn?**

Mein größter Lerngewinn ist die Erkenntnis, dass ich bereits Ontologien erstellt und dass ich sie jetzt verstanden habe. Außerdem war die Erstellung meines JSON-LD-Datensatzes eine praktische Übung, die mir gezeigt hat, wie man mit Linked Data arbeitet und wie man es visualisiert.

- **Was war besonders überraschend? Was hat Sie irritiert?**

Ich war überrascht dass ich bereits Ontologien erstellt habe, ohne es zu wissen. Auch die Vielfalt der Jobs im Bereich Knowledge Graphs und Ontologien hat mich überrascht. Irritierend fand ich, dass viele Stellen sehr nah an Data Science liegen und dass man zwischen bibliothekarischen und wirtschaftlichen Stellen unterscheiden muss.

- **Wo sehen Sie für sich noch Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?**

Ich möchte tiefer in SPARQL einsteigen und weitere RDF-Serialisierungen kennenlernen. Ich will auch mehr mit WebVOWL und anderen Visualisierungstools arbeiten und weitere Standards wie SKOS oder OWL erforschen. Ich willst mein JSON-LD-Projekt weiter ausbauen (mein Portfolio fertigstellen) und es in meinem RAG-System nutzen. Ich will auch mehr praktische Erfahrung mit Ontologien sammeln und vielleicht sogar eine eigene Ontologie für meinen Vault erstellen. Langfristig: Ich will mich evtl. auf Jobs im Bereich Linked Data/Knowledge Graphs/Ontologien bewerben, aber erst wird der Master beendet.

Vorlesung 4

An diesem Tag wurde *SPARQL* vorgestellt und ich muss zugeben: Es war der Tag, an dem ich am wenigsten mitgenommen habe. Nicht, weil das Thema uninteressant wäre, im Gegenteil: Ich finde es unglaublich, was man mit SPARQL alles abfragen kann. Aber ich bin einfach nicht für *Online-Vorlesungen* gemacht. In einer Präsenzvorlesung höre ich zu, achte aufmerksam hin und stelle gelegentlich Fragen, da kann ich mich komplett darauf einstellen. Bei einer Online-Vorlesung funktioniert das nicht. Meine Konzentration schweift ab. Deshalb fiel es mir schwer, der Vorlesung zu folgen, aber ein paar Dinge habe ich doch mitgenommen.

Mit SPARQL hatte ich bisher nur einen Berührungspunkt: Im Bachelor-Modul *Metadatenmanagement* habe ich es kennengelernt, aber ich muss gestehen: Bis heute habe ich es nicht ganz durchblickt. Mir ist grundsätzlich klar, wie SPARQL funktioniert, aber die Syntax ist mein größtes Problem. Theoretisch müsste ich mir alle Codes für Entitäten merken, wenn keine Liste zur Hand ist. Das ist ungewohnt und für mich eine Herausforderung. Wenn ich Python-Code oder eine SQL-Abfrage schreibe, muss ich normalerweise nichts nachschlagen, außer den Namen der Spalten. Ob man das vergleichen kann, weiß ich nicht. Wegen meiner mangelnden Online-Konzentration und der ungewohnten Syntax hatte ich Schwierigkeiten, den Übungsaufgaben zu folgen. Was mir jedoch leichtfiel, waren Funktionen wie Filtern oder Gruppieren, da diese SQL stark ähneln. Ich habe mir vorgenommen, das nachzuholen, denn ich sehe riesiges Potenzial in SPARQL, vor allem für *Information Retrieval* und auch für meine privaten Projekte. Besonders spannend finde ich, dass man geografische Daten abfragen kann und dass die Ergebnisse visuell aufbereitet werden können und das nicht nur als Balkendiagramme, sondern auch Karten und Timelines. Im Modul *Management and Leadership*, in dem wir ein Projekt mit der *Apple Vision Pro* durchführen, wollen wir SPARQL nutzen. Meine Teilgruppe soll untersuchen, wie man Medien in Bibliotheken oder Kultureinrichtungen während der Nutzung der Apple Vision Pro anreichern kann. Da wir für jedes Objekt spezifische Informationen abfragen müssen, eignet sich SPARQL perfekt dafür.

Persönlich habe ich mir vorgenommen, das Buch *Learning SPARQL: Querying and Updating with SPARQL* von O'Reilly durchzuarbeiten, um am Ende sagen zu können: *Ich habe SPARQL wirklich verstanden*. Doch aktuell habe ich kein Projekt, in dem ich persönlich SPARQL einsetzen muss und auch nicht die Zeit, das Buch komplett zu lesen. Deshalb wird sich das Thema SPARQL für mich leider noch etwas nach hinten verschieben. Ein weiteres Thema, das mir an diesem Tag wieder in den Sinn kam, war *Wikidata.org*, welches mir bereits bekannt war. Ich habe in der Vergangenheit an *Wikipedia-Workshops* teilgenommen und fand die Initiative immer spannend. Bei Wikipedia selbst bin ich jedoch nie ganz warm geworden. Wikidata hingegen finde ich ein unglaubliches Werk: Man kann dort über alle möglichen Dinge Informationen abfragen und SPARQL ist dabei unverzichtbar. Ein weiterer Grund, SPARQL zu lernen!

Zusammenfassend kann man sagen:

- **Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?**

Inhaltlich habe ich SPARQL wiederholt und versucht zu verstehen. Ich habe auch erkannt, dass Filtern und Gruppieren in SPARQL ähnlich wie in SQL funktioniert und dass es im Projekt mit der Apple Vision Pro eine zentrale Rolle spielen wird. Persönlich habe ich ehrlich eingeschätzt, dass ich nicht für Online-Vorlesungen gemacht bist und dass mir die Syntax von SPARQL schwerfällt. Ansonsten habe ich Wikidata.org wiederentdeckt und erkannt, dass SPARQL dort unverzichtbar ist.

- **Was war Ihr größter Lerngewinn?**

Mein größter Lerngewinn ist die Erkenntnis, dass SPARQL für meine Projekte relevant ist, insbesondere im Kontext der Apple Vision Pro und der Anreicherung von Medieninhalten.

- **Was war besonders überraschend? Was hat Sie irritiert?**

Überraschend fand ich dass ich SPARQL bereits im Bachelor kennengelernt habe und dass es jetzt wieder auftaucht. Auch die Möglichkeit, geografische Daten abzufragen und zu visualisieren, hat mich überrascht und begeistert. Irritierend fand ich, dass man die Codes für die Entitäten evtl. immer nachschlagen muss.

- **Wo sehen Sie für sich noch Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?**

Ich möchte SPARQL vertiefen, insbesondere die Syntax und praktische Anwendung (z. B. mit Wikidata). Ich will auch mehr über die Anwendung von SPARQL in KI- oder AR-Projekten lernen. Ich habe vor das Buch *Learning SPARQL* lesen, sobald ich Zeit habe.

Vorlesung 5

Unverzichtbar sind Werkzeuge, mit denen sich Daten ansehen, bearbeiten, analysieren und verarbeiten lassen. Dabei wurden verschiedene Softwarelösungen vorgestellt und zwei davon haben wir in einem kleinen Workshop selbst ausprobiert.

Am letzten Tag des Teilmoduls fand ein Workshop zu *OpenRefine* und *Metafacture* statt. Bei *OpenRefine* wurden verschiedene Funktionen anhand der Dokumentation gezeigt und direkt ausprobiert. Bei *Metafacture* wurden die Funktionen anhand eines Tutorials erklärt. Persönlich nutze ich solche Software in meiner Praxis nicht, denn wenn ich Operationen wie Filtern oder Clustering großer Datenmengen durchführe, mache ich das meistens in *Python* mit Bibliotheken, die gut mit großen Datenmengen umgehen können. Da ich das oft tue, bin ich mit Python bei der Datenanalyse auch schneller als mit einer GUI-basierten Software. Ich kann aber nachvollziehen, dass eine grafische Oberfläche für Personen, die nicht regelmäßig programmieren, von Vorteil ist. Sehr praktisch fand ich den Abgleich von Daten mit externen Datensätzen, zum Beispiel mit dem Datensatz der DNB.

Bei *Metafacture* war mein Eindruck ähnlich: Das Einlesen eines Datensatzes, etwa in JSON, und das Exportieren in ein anderes Format fand ich sehr praktisch. Ansonsten konnte ich für meine alltägliche Arbeit keinen großen Nutzen daraus ziehen, das kann sich aber noch ändern.

Am Ende folgte eine kleine Diskussion, die genau in diese Richtung ging: Einerseits hat man das Potenzial der Software gesehen, aber da sie nur im Nischenbereich von sehr wenigen Personen eingesetzt wird, ist fraglich, ob man sie überhaupt lernen sollte, vor allem, weil man sich erst mit ihr auseinandersetzen müsste, da sie nicht ganz unkomplex ist. Ich glaube, die in der Gruppe, die täglich programmieren, sind eher gewohnt, solche Operationen mit Python und Pandas durchzuführen, solange der Speicher es erlaubt. Die anderen hingegen haben das Problem, dass sie in ihrer alltäglichen Arbeit einfach keine Verwendung für solche Datenwerkzeuge finden, weil sie nicht täglich mit Daten arbeiten.

Zusammenfassend kann man sagen:

- **Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?**

Inhaltlich habe ich OpenRefine und Metafacture kennengelernt, insbesondere deren Funktionen zur Datenbereinigung, -konvertierung und -abgleich (z.B. mit DNB). Ich habe erkannt, dass beide Tools für bestimmte Aufgaben nützlich sind, vor allem für Nutzer:innen ohne Programmierkenntnisse. Persönlich aber, arbeite ich lieber mit Python, und bin dadurch schneller und flexibler. Ansonsten habe ich die Nischenrolle dieser Tools erkannt und dass sie nicht für alle relevant sind. Die Diskussion hat mir gezeigt, dass die Wahl des Tools vom Kontext und der täglichen Praxis abhängt.

- **Was war Ihr größter Lerngewinn?**

Mein größter Lerngewinn ist die Erkenntnis, dass Werkzeuge wie OpenRefine und Metafacture nicht *überflüssig* sind, sondern für bestimmte Nutzer:innen und Anwendungsfälle sehr wertvoll.

- **Was war besonders überraschend? Was hat Sie irritiert?**

Überraschend fand ich, dass Metafacture so gut für Formatkonvertierungen geeignet ist und dass OpenRefine so gut mit externen Datensätzen (wie DNB) verbinden kann. Irritierend fand ich an diesem Teil der Vorlesung nichts.

- **Wo sehen Sie für sich noch Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?**

Ich will meine Python-Skills weiter ausbauen, besonders in Bezug auf Datenbereinigung und -konvertierung.

Fazit

Abschließend kann ich durch aus behaupten, dass das Teilmodul *Datenformate*, auch wenn es nur im ersten Semester stattfindet, sich durch das ganze Semester zieht. Wieso? Weil man als Bibliotheks-informatiker jeden Tag mit verschiedenen Datenformaten in Berührung kommt. Sei es in Form in *ISBD* und *MARC*, oder *XML* und *JSON*. Vor allem in dem jetzigen Zeitalter ist es ein Wunder, wenn man damit nicht in Kontakt kommt. Allein im Studium merke ich das sehr. Wir haben noch andere Module im Studium, und in fast allen kommt das Thema nebei vor, bzw. wird vorausgesetzt. In meinem KI-Projekt, wo ich ein Linked-Data Datensatz erstelle für ein RAG-System, kommt Linked Data im YAML Format vor. Im Apple Vision Pro Projekt beschäftigen wir uns mit der Anreicherung von Medien durch Daten von Schnittstellen und verwenden hierbei auch SPARQL. Selbst bei diesem Lernportfolio kommen Datenformate und Metadaten vor. Ich muss zugeben, dass ein Lernportfolio in dieser Form nicht notwendig war, aber auf diese Weise konnte ich das JSON-LD Datenformat besser kennenlernen und mir allgemein ein besseres Bild von Linked Data machen. Und bei Internetprogrammierung ist es nur eine Frage der Zeit, bis wir uns auch mit Datenformaten beschäftigen. In meinem privaten Leben beschäftige ich mich fast täglich mit Datenformaten und Linked Data, wenn ich meine Notizen von der Hochschule erstelle oder bearbeite, oder auch von anderen Dingen aus meinem Leben. Also ja, ich habe was gelernt, aber was und wie, auf das gehe ich nochmal in den folgenden Fragen ein:

Was haben Sie gelernt (inhaltlich, persönlich, sonstiges)?

Gelernt habe ich nicht nur, welche Datenformate existieren und wie sie aussehen, sondern auch die Unterscheidung zwischen Schemasprachen und Strukturierungssprachen. Wichtige Begriffe wurden dabei geklärt und auch wenn das meiste für mich nichts neues war, da ich es bereits im Bachelorstudium hatte, so war

es eine sehr gute Auffrischung der ganzen Thematik. Vor allem die theoretische Tiefe hat mir viel gebracht, da dies bei mir früher nur angeschnitten wurde.

Was war Ihr größter Lerngewinn?

Mein größter Lerngewinn war definitiv das Thema Linked Data. Nicht nur wurde mir endlich klar wie die Praxis bei Linked Data aussieht, sondern auch, wie man Ontologien erstellt. Wichtig für mich ist auch die persönliche Anwendung in meinem Obsidian Vault, da ich es nun, da ich das Wissen habe, auf das nächste Level bringen kann. Linked Data ist für mich ein Thema, welches sich bis Ende vom Studium durchziehen wird und ich kann mir sehr gut vorstellen in dieser Richtung eine Masterarbeit anzufertigen. Wie genau diese aber aussehen wird, wird sich noch zeigen.

Was war besonders überraschend? Was hat Sie irritiert?

Meine größte Überraschung war tatsächlich, wie viel sich über die ganzen theoretischen Dinge wie Datenmodelle, diskutieren lässt. Ich dachte, mehr oder weniger, dass man das beschließt und fertig ist. Es ist so definiert, also akzeptieren wir es alle. Aber nein! Auch Menschen die Tag für Tag in diesem Bereich arbeiten, diskutieren immer noch über verschiedene Definitionen und Strukturen (Siehe Vorlesung 1). Das hat mir gezeigt, dass das alles in Bewegung ist. Das finde ich toll! Ich hatte nie wirklich einen Einblick in die Praxis bei Datenformaten und Linked Data bekommen, und jetzt wo ich es hatte, kann ich viel besser behaupten, dass es mir gefällt.

Wo sehen Sie für sich noch weiteren Bedarf an Weiterbildung zum Thema Datenformate? Was nehmen Sie sich für die nächste Zeit vor?

Was die Weiterbildung angeht, so sehe ich bei mir persönlich vor allem zwei Bereiche, die ich weiter vertiefen will: Einerseits natürlich Linked Data (Ich glaube das wurde aus dem gesamten Lernportfolio ziemlich deutlich). Ich finde diese Thematik äußerst interessant und will mein Wissen und mein Können in diesem Bereich weiter vertiefen und verfeinern. Das beinhaltet nicht nur theoretisches Wissen, in dem ich Bücher und Artikel darüber lesen will, sondern auch die praktische Übung durch die Einbindung von Linked Data in meine Hochschulprojekte als auch meine Privatprojekte. Andererseits wie bereit bei Vorlesung 4 erwähnt will ich mich unbedingt in SPARQL weiterbilden. Ich sehe sehr viel Potenzial darin, auch wenn ich es mir noch nicht 100%ig vorstellen kann, einfach weil mir das SPARQL Wissen fehlt. Die Syntax ist für mich schwer verständlich. Durch die Verwendung von SPARQL in einem Projekt will ich das ändern und auch mich privat darin weiterbilden, sobald die Zeit verfügbar ist.