

Vernetztes Wissen

Hypertext und Semantic Web

Im Folgenden werde ich der Frage nachgehen: Was bleibt von der Utopie der Vernetzung nach Jahrzehnten des Online-Stellens, der Zugänglichmachung im Netz, der Verknüpfung von Sammlungen? Sind neue Konzepte denkbar, die über die Zugänglichkeit durch Veröffentlichung und Vielfalt durch Vernetzung hinausgehen? Hierbei ist darauf zu achten welche etablierten Konzepte der Tech-Giganten dem wiederum im Weg stehen. Der Kontext ist ein weiteres Mal der Wandel von digitaler Kommunikation mit Beginn des World Wide Webs, der grafischen Variante des Internets. Denn Mitte der 1990er-Jahre wurde das Internet semantisch, Verbindungen wurden zur zentralen Information, so wie die Wörter in der Semantik eines Satzes: Der Sinn der Daten trat in den Vordergrund. Das passierte nicht nur durch die Einführung von grafischer Navigation über Hyperlinks – seitdem können wir durch Informationen »surfen« –, sondern in der Konzeption von semantischen Verknüpfungen, das Netz wurde damit zu einem logischen Netz.

Hypertext war das Prinzip des Internets seit den 1990er-Jahren. Texte wurden zu tief vernetzten Medien. Auf kommerziellen Social Media-Plattformen wird dieses Prinzip komplett zerstört. In den Texten gibt es keine Verknüpfungen mehr.

Was war so besonders an Hypertext, und warum ist das Vernetzungsprinzip, das daraus entstand, das semantische Web, so wichtig für Kulturerbeinstitutionen? Ein Bild, das sich mit Hypertext verbindet, ist der Garten: »The Garden is an old metaphor associated with hypertext. [...] The Garden of Forking Paths from the mid-20th century. The concept of the Wiki Gardener from the

1990s. Mark Bernstein's 1998 essay *Hypertext Gardens*.¹ Man stelle sich zum Beispiel ein Wiki vor. Das ist eine einfache Art und Weise, verknüpfte Texte zu veröffentlichen, quasi ein Speicher für so funktionierende Texte; das bekannteste Wiki der Welt ist die Wikipedia als eine gemeinschaftlich betriebene vernetzte Enzyklopädie. Am Anfang ist so ein Wiki leer, dann legen wir erste Einträge an, die aufeinander verweisen, über hervorgehobene Wörter – also die »Hyperlinks« im Web, die über Markup Languages wie HTML (Hypertext Markup Language) – definiert sind, das Ganze wächst zum Hypertext heran. Dann fangen wir an, diesen Prozess zu reflektieren: Welche Bereiche unseres Speichers fehlen noch, in welche Richtungen soll es wachsen und wuchern, welche Bereiche sind unnötiges Wachstum, das nicht zu unseren Interessen und Zwecken passt? So fangen wir an, dieses Wachstum zu steuern, wie das eines kleinen Gartens. Ist so ein flexibler Wissensgarten als System vorstellbar, um Sammlungen zu beschreiben? Und was brauchen wir dafür? Und wie ließen sich die kritischen Perspektiven auf Museen und Kunstgeschichte hier auch einpflegen?

Die Grundlagen dafür, Menschen und Maschinen im Internet zu vernetzen, in einer Art und Weise, die erlaubt, Information zu transportieren, zu lesen, zu schreiben und zu verarbeiten, wurden Ende der 1990er-Jahre maßgeblich weiterentwickelt. Das oben beschriebene Hypertext-Prinzip war noch eher vergleichbar mit anderen Methoden digitale Texte über Telekommunikationsnetze zu verbreiten, etwa Mailboxen oder Bildschirmtext der 1980er, nur dass die Navigation flüssiger wurde durch direkte Verbindungen in Texten, die Links. Das Format dafür heißt Hypertext Markup Language (HTML). Wenn dieses Buch in HTML geschrieben wäre, könnte ich direkt ein paar Wörter anklickbar machen, die auf ein anderes Dokument verweisen, das weiterführende Informationen enthält. Ende der 1990er-Jahre wurde eine weitere Markup Language definiert, Extended Markup Language (XML), die dann erlaubte, mehr als nur verknüpfte Texte zu beschreiben oder zu gestalten. Jetzt ging es darum, Einheiten im Text genauer zu definieren, nicht nur Verknüpfungen oder Überschriften hervorzuheben, sondern Informationen genau zu erfassen.

Stellen wir uns vor, ich möchte eine Statue in einem Museum genau im Netz beschreiben. Ich könnte einen Hypertext schreiben und darin erwähnen,

1 Mike Caulfield, »The Garden and the Stream: A Technopastoral«, *Hapgood* (blog), 17. Oktober 2015, <https://hapgood.us/2015/10/17/the-garden-and-the-stream-a-technopastoral/>.

wer sie gemacht hat, und auf einen anderen Text verlinken, der die Biografie der Künstlerin bereithält. Ist die Statue aus Marmor, könnten wir noch auf eine Information zu diesem Material verweisen usw. In XML könnte ich eindeutig definieren, um was für Informationen es hier geht und wie sie zusammenhängen. Dass die Künstlerin die Urheberin der Statue ist, wird dann nicht nur als Satz erfasst, sondern als Daten, die das Verhältnis von einer Informationseinheit zu einer anderen beschreiben. Diese Art und Weise Daten aufeinander zu beziehen, wird semantisch genannt, weil sie auf einer Bezeichnungslogik (Semantik) mit definierten Vokabularen basiert. Das sind die konzeptuellen Bausteine der Markup Language, und wie sie sich aufeinander beziehen lassen. Die Fachbegriffe dafür sind Resource Description Framework (RDF), Schema und Ontologie.

Die Kunstwissenschaft kennt solche Unterschiede der Wissensrepräsentation, wie sie hier zwischen HTML und XML beschrieben wurden, zum Beispiel als den Unterschied von Werkverzeichnis und Werkbeschreibung. Nur, und jetzt wird es spannend, gibt es ein Ungleichgewicht in deren Informationsgehalt und -struktur. So sind die strukturierten Informationen in einem Werkverzeichnis meist recht kurz gefasst, die grundlegendsten Informationen zu einem Werk werden erfasst, wie das Entstehungsdatum, Urheber*in, Material usw. und detailliertere Informationen kommen im Text der Werkbeschreibung unter. Im Web können wir diesen Unterschied aufbrechen, nicht nur können die Texte Hypertext werden, also mit Referenzen verwoben werden, sondern die in der Werkbeschreibung erwähnten Beziehungen und Eigenschaften können eindeutig als solche definiert werden. Man stelle sich vor, ich möchte beschreiben, dass ein Werk von der Berliner Künstlerin Anna Eichens gemalt wurde, und möchte erfassen, dass sie Schülerin von Theude Grönland war. Dann wäre im Sinne des Semantic Web das Ziel, dass sowohl die Verhältnisse zwischen den Personen, ihre Identitäten, also auch die Verhältnisse zum Werk als digitale Informationen strukturiert sind. Das erlaubt Eindeutigkeit, Vergleiche, Statistik, Auffindbarkeit, Rechnen. Zum Beispiel würde auf Wikidata das XML – also der Teil den Maschinen lesen, für Menschen gibt es ein Interface – zum Geburtsort von Eichens so aussehen:

```
<wdt:P19 rdf:resource="http://www.wikidata.org/entity/Q64"/>
```

Der direkte Verweis auf eine »Wahrheit« im Wikidata-System wird mit »wdt:« (Wikidata truth) gemacht, hier haben wir die Eigenschaft P19 (P für property), die in Wikidata die Angabe des Geburtsorts bedeutet, und den Verweis auf

die Einheit (entity) mit ihrer Nummer Q64, wohinter sich Berlin verbirgt. Damit ist auch eindeutig, dass Anna Eichens nicht zum Beispiel in der Kleinstadt Berlin im US-Bundesstaat Connecticut geboren ist, denn das wäre Q821199.

Warum ist es für Kunstgeschichte und Museologie wichtig zu verstehen, was das »Semantic Web« ist? Das Semantic Web ermöglicht uns, bestimmte Dokumentationsformate mit denen wir arbeiten – von der Inventarkarte bis zum Katalog – in offenere, flexiblere Varianten zu übersetzen, die nicht nur von Menschen, sondern auch von Computerprogrammen gelesen werden können. Das ist zum Beispiel notwendig, um so etwas wie eine Social Media-Plattform zu betreiben, aber auch für künstliche Intelligenz. Denn hinter einem Beitrag auf den sozialen Medien versteckt sich eigentlich eine Art Gerüst, die die verschiedenen Bestandteile, etwa das Bild, die Autor*innenschaft, das Datum, den Text, Geo-Daten, Kommentare, und so weiter zusammenhält. Und künstliche Intelligenz, die mehr kann als nur wie ChatGPT, Grok und andere Chat-Bots eine*n Büro-Assistent*in zu simulieren – etwa die Kriegs- und Verwaltungsmaschinen, die der KI-Konzern Palantir anbietet –, basiert auf Datenordnungen, das heißt auf spezifischen konzeptuellen Datensystemen (Ontologien). Für soziale Medien und für die neuen Datenmaschinen braucht es strukturiertes Wissen. Das Semantic Web ist also eine Schlüsseltechnologie für Social Media und künstliche Intelligenz. Diese beiden Phänomene verstellen aber auch den Blick darauf, was es außerdem ist.

Das Semantic Web ermöglicht flexible digitale Wissensordnungen. HTML fungierte noch als eine Grammatik für das Netz, wie Kim H. Veltman schreibt, ein Kulturhistoriker, der viel über semantische Netze nachgedacht hat: »From an historical point of view these new markup languages in their attempt to deal with the structure of language are effectively re-inventing the principles of grammar in electronic form.«² RDF hingegen ist die Dialektik im Netz, so beschreibt Veltman die Konzepte des semantischen Webs:

»The vision outlined by Tim Berners Lee goes far beyond this. [...] His global reasoning web or semantic web requires that every unit of knowledge will contain summary information describing itself in terms of its logical truth-value, its quality and so on. To achieve this the W3 Consortium is building on earlier attempts to create a Protocol for Internet Content Selection (PICS)

2 Kim H. Veltman, »Conceptual Navigation in Multimedia Knowledge Spaces«, in TKE-1999, *5th International Congress and Workshops on Terminology and Knowledge Engineering*, TermNet, (Wien, 1999), 2.

and is creating a Resource Description Format (RDF), which will enable all information to have a self-describing component in machine-readable form. From an historical point of view, RDF is re-inventing the principles of dialectic.«³

Die Markup Languages erlaubten, Text als digitales, vernetztes Wissen aufzubauen. RDF legt nun noch eine Ebene darauf, die es erlaubt, Bedeutungen in einer digitalen Semantik zu schreiben.⁴ Die Daten werden nämlich als Subjekt-Prädikat-Objekt geschrieben, als Daten-Triples, zum Beispiel für ein Gemälde: »Werk« – »hat Urheber« – »Name der Künstlerin«. Die Unterscheidung zwischen Inventar und Historisierung zu durchbrechen, würde zum Beispiel beinhalten, Informationen zu Werkgenese und Sozialgeschichte in Daten zu überführen, statt einem Text »Künstlerin X schuf dieses Werk in ihren jungen Jahren am Hof von Y während der Aufstände der XY« müssten wir die einzelnen Einheiten in diesem Satz als Daten verstehen und seine Semantik als digitale Zusammenhänge, als Semantik der Informationen. Wenn wir Texte als Netze der Information nutzen, verändern wir die Schemata digitaler Sammlungen.

Die Informationswissenschaftlerin und Philosophin Ingetraut Dahlberg hat Veltman den Tipp gegeben, wie konzeptuelles Denken zur Arbeit mit Klassifikationen gehört:

»The analytical tables of facts (e.g. concepts and classes of concepts) must also contain all the possible concepts (in the form of facets, just as Ranganathan had built his faceted Colon Classification system) by which synthetic patterns of true statements on referents can be constructed. Somewhere in your paper this ought to be said. And here our good Kant with his analytical and synthetic judgments could also be brought into the picture!«⁵

Veltman führt die Möglichkeiten, so etwas mit Computern zu machen, bis auf die Rechenmaschinen von Leibniz zurück und bemängelt, dass damalige digi-

3 Ebd.

4 Kim H. Veltman, »Towards a Semantic Web for Culture«, *Journal of Digital Information* 4, Nr. 4 (2004): 4.

5 Ebd., 71.

tale Kataloge immer noch nicht die konzeptuelle Kombination von Kategorien erlauben.⁶

Eine ganze Architektur an Technologien wirkt zusammen, um vernetztes Wissen zu ermöglichen. Erst wenn wir diese durchgehen, können wir verstehen, was die weiteren Unterschiede zu anderen Wissensspeichern sind, wie etwa Büchern.

Der Semantic Web-Stack, also der »Stapel« von technischen Funktionen, beinhaltet aufeinander aufbauende Ebenen. Auf den untersten Ebenen finden wir die grundlegenden Möglichkeiten ein adressierbares digitales Textdokument mit einer internen Struktur zu schreiben, also das bisher erwähnte XML und RDF sowie eine eindeutige Adresse (URI). In den darüber gelagerten Ebenen geht es um die Konzepte, die hier angewandt werden und ihre Struktur bzw. Logik. Noch mal am Beispiel einer Künstlerin und ihres Werks. Da sie eine Künstlerin ist, ist sie auch ein Mensch, ein Kind von Eltern, lebt auf der Erde, Urheberin von Werken (oder Konzepten) usw. Diese Informationen sind logisch miteinander verbunden.

Die konzeptuellen Schemata, werden dann – eine weitere Ebene oben drauf, wenn wir beim beschriebenen Stapel von HTML und RDF bleiben – als Ontologien konzipiert, eine für die Archivwissenschaft und das Sammlungsmanagement genauso wichtige Technik wie auch für die digitalen Enzyklopädien im Netz. Der Semantic Web-Stack erlaubt es, sie genau so auch im Netz verbunden zu repräsentieren. Und er sieht auch weitere Funktionen vor, wie zum Beispiel technologisches Vertrauen durch digitale Signaturen.

Wenn wir heute, fast 20 Jahre nach Einführung dieser Technologie, Informationen unstrukturiert oder schlecht strukturiert (etwa auf Social Media-Plattformen) im Netz veröffentlichen, begrenzen wir ihren Informationsgehalt. Das heißt dann nicht nur, dass sie nicht automatisiert verarbeitet werden können, sondern auch, dass sie schwerer zu finden oder gar zu verknüpfen sind.

Veltman, der die Entwicklung dieser Systeme von der Warte der Kunst- und Wissenschaftsgeschichte aus beobachtete, bemerkte rückblickend zynisch, dass die Systeme darauf basierten, den Ort für Bedeutung immer weiter nach oben zu verschieben:

6 Ebd., 22f. Die SPARQL-Endpunkte, die programmierte Abfragen zum Beispiel auf Wikidata erlauben, können wir heute also eine Suchtechnik auf Basis synthetischer Konzeptarbeit verstehen.

»Even more fortunately, no one was so rude as to ask what had happened to the elusive concept of meaning, which the 1997 diagram had linked neatly with RDF. Nor did anyone ask why the dimension of meaning keeps being pushed up the layers of the cake to a level that is not yet working.«⁷

Er beschreibt, wie hier von einem unumstößlichen und universellen Weltbild ausgegangen wird: »AI and semantic web communities create data structures that assume a single world-view.«⁸ Eine Ontologie beschreibt zum Beispiel, das eine Sammlung in einem Museum liegt und dass sich in dieser Sammlung Werke befinden und dass diese Urheber*innen haben und eine Provenienz.

Museen sind leichter zu vernetzen und digital zu visualisieren als andere Sammlungen von Dingen, denn Museen gleichen sich in ihrer Funktionsweise. Das liegt an geteilten digitalen Ontologien, der Menge an Typisierungen, die auf Sammlungsgegenstände angewandt werden. Ein Zitat aus dem Jahresbericht 2002 der Staatlichen Museen Berlin fasst das Vorgehen zusammen:

»Für beide Aktivitäten – Errichtung und Pflege großer digitaler Objektdatenbanken bzw. von Internet-Portalen – stellt sich erneut und mit noch größerer Dringlichkeit die Notwendigkeit, gemeinsame Verfahrensweisen, Standards und Informationspools zu etablieren und zu pflegen.«⁹

Als konkrete »Gemeinschafts-Tools« nennt der Bericht: »Austausch- und Datenformate, damit elektronische Daten (oft von ganz unterschiedlich strukturierten Datenbanken stammend) in einer einheitlichen Präsentationsweise dargestellt werden können;« sowie »Regelwerke, damit die Erfassung und Haltung der Daten von vornherein gleichförmiger geschehen kann,« und »Thesauri und Wortgutbestände – denn ein gemeinsam benutztes Vokabular hilft, gleiche Objekte in unterschiedlichen Museen auch gleich zu bezeichnen,« die auch verstanden werden als eine »Orientierungshilfe für die Benutzer, sich mit den Suchanfragen an einem Vokabular zu orientieren, in dem auch die Erfassung der Objekte erfolgte.«¹⁰ Die teilweise seit mehreren Jahrzehnten in Digitalisierung begriffenen Sammlungen um die Jahrtausendwende dann ans Netz zu

7 Ebd. S. 5.

8 Ebd. S. 7.

9 Staatliche Museen zu Berlin Preußischer Kulturbesitz, »Jahresbericht 2002 der Staatlichen Museen Preußischer Kulturbesitz Berlin«, 77.

10 Ebd.

bringen, bedeutete, sie über geteilte Standardisierungen mit anderen Sammlungen zu vernetzen, wie im Kapitel »Das vernetzte Museum« am Beispiel der Plattform Europeana, die einen gemeinsamen europäischen Museumskatalog im Netz anbietet, untersucht wurde.

Die digitalen Sammlungen sind über die Logik des semantischen Webs direkt mit den Museen als physikalischen Räumen verbunden, diese Bezüge machen die Ebene des Vertrauens oberhalb der Ontologien aus:

»A web site that refers to Leonardo da Vinci's Mona Lisa in the Louvre should provide a direct link to the Louvre's digital version of that painting. By so doing one solves in large part the existence and identity dimensions of the logic layer. Creators of websites will of course be at liberty to refer to other copies and versions, legitimate or pirate, but in the absence of a direct link their claims will be idle. This direct electronic link will effectively function in much the same way that a pedigree does in the world of professional breeders. Without a pedigree, without a proof of a direct link, claims of identity will be unclear and thus also at a lower level of trust.«¹¹

Mit der Erzeugung, Verwendung und Weiterentwicklung, ja auch Kritik dieser digitalen Ontologien beschäftigen sich die digitale Museologie und Kunstgeschichte intensiv. Standards wie LIDO (Lightweight Information Describing Objects) und das CIDOC Conceptual Reference Model (CIDOC CRM), aber auch digitale Fassungen ikonographischer Bildbeschreibung (ICONCLASS), bilden die Basis einer digitalen Sammlungspraxis und werden fortwährend weiterentwickelt. Ein Beispiel wäre hier CIDOC CRMInf (CRM Inference), ein Konzept um Meinungen in Daten abzubilden. Frühere Ansätze waren hier das Konzept von Annotationen im semantischen Web, als eine spezielle Form von Meta-Daten:

»First, there is the identity and (self-)description of ›objects,‹ which can be subdivided into a) objective facts and b) subjective interpretations of those facts. Second, there are pointers to that knowledge in the form of metadata and ontologies (classification systems, thesauri). Third, there are tools for annotating (e.g. W3's Annotea), which are about adding comments and potentially new knowledge to the existing corpus.«¹²

11 Veltman, »Towards a Semantic Web for Culture«, 40.

12 Ebd., 41.

Wenn wir uns an die Kritik von Veltman erinnern, dass die Bedeutungen, die in logischer Beschreibung entstehen, immer weiter aus der grundlegenden Technik in neue Ebenen verschoben werden, können wir mit CRMInf hier ein Auslagern in Nebensysteme beobachten. Bestimmte, als zu subjektiv verstandene, Bedeutungen sollen nun als Meinungen ausgegliedert werden. Veltman hat dieses Problem schon 2006 beschrieben:

»The universal is static and can readily be reduced to mechanistic metaphors. The particular is growing (or decreasing) and is more readily amenable to organic metaphors. To ensure the progress of science, it thus made sense to separate the organic, changing, (subjective) aspects of subjects from the mechanical, unchanging, (objective) elements of objects.«¹³

Trotzdem sind die Ansätze, Bedeutungen und Beschreibungen spezifischer zu fassen, sehr interessant, da sie teilweise versuchen, Fragestellungen aus der kritischen Museologie aufzugreifen und zum Beispiel andere Stimmen als die der Autoritäten innerhalb der Museen sichtbar zu machen. Veltmans Plädoyer war, neue Wissenssysteme zu entwickeln und mit den älteren zu verknüpfen, um den Wandel von Interpretationen sichtbar zu machen:

»Creating our own modern classification is not enough to understand how, and more significantly why, earlier cultures had very different ways of organising knowledge. Thus we need bridging and mapping devices that allow us to move dynamically through different languages, different levels of vocabularies, different chronologies (in the sense of time systems), different cartographical methods and policies [...]. Such dynamic lists of knowledge will allow us to trace changes of interpretation over time, have new insights and help us to discover new patterns in knowledge.«¹⁴

Linked Open Data

Mitte der Nullerjahre begann sich die Formulierung der Linked Open Data (LOD) zu etablieren, als neue Konzeption des semantischen Web. Suggestiert das semantische Web eine Kontinuität von Entitäten oder Datenpunkten, die global verbunden sind, so macht der Begriff LOD vielleicht deutlicher, dass es

13 Ebd. S. 25.

14 Ebd., 28.