



Separatum aus:

THEMENHEFT 12

*Elisabeth Lienert / Joachim Hamm
Albrecht Hausmann / Gabriel Viehhauser (Hrsg.)*

Digitale Mediävistik

Perspektiven der Digital Humanities für die Altgermanistik

Publiziert im November 2022.

Die BmE Themenhefte erscheinen online im BIS-Verlag der Carl von Ossietzky Universität Oldenburg unter der Creative Commons Lizenz [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/). Die ›Beiträge zur mediävistischen Erzählforschung‹ (BmE) werden herausgegeben von PD Dr. Anja Becker (München) und Prof. Dr. Albrecht Hausmann (Oldenburg). Die inhaltliche und editorische Verantwortung für das einzelne Themenheft liegt bei den jeweiligen Heftherausgebern.

<http://www.erzaehlforschung.de> – Kontakt: herausgeber@erzaehlforschung.de
ISSN 2568-9967

Zitiervorschlag für diesen Beitrag:

Glauch, Sonja: Welche Lebenserwartung haben digitale Editionen? In: Lienert, Elisabeth/Hamm, Joachim/Hausmann, Albrecht/Viehhauser, Gabriel (Hrsg.): Digitale Mediävistik. Perspektiven der Digital Humanities für die Altgermanistik, Oldenburg 2022 (BmE Themenheft 12), S. 65–75 (online).

Sonja Glauch

Welche Lebenserwartung haben digitale Editionen?

Abstract. Der Beitrag unterscheidet zunächst vier Hinsichten, in denen digitale Editionen altern und gegebenenfalls sterben können: 1. Datenträgertechnologien und Laufumgebungen, 2. Kodierungen, Verlinkungen, Metadatenformate, 3. softwaretechnische Aspekte der Präsentationsschicht, 4. Aspekte der Ästhetik und des Bedienungskomforts. Er wendet sich dann einer aktuellen Lösungsperspektive für das Langfristbewahrungsproblem zu, nämlich den im Aufbau befindlichen Forschungsdateninfrastrukturen, und erörtert die Frage, ob digitale Editionen Forschungsdaten sind und inwiefern ihre Dauerhaftigkeit im Rahmen von Initiativen wie Text+ gewährleistet werden könnte. Abschließend nennt der Beitrag vier Optionen, die konkret in den letzten Jahren für die Bewahrung von digitalen Editionen ins Spiel gebracht worden sind. Da es sich bei diesen Ansätzen zum größten Teil um Vorüberlegungen oder (teils gescheiterte) Prototypen handelt, ist zum jetzigen Zeitpunkt nicht entscheidbar, welcher dieser Wege konkret für jetzt entstehende komplexe, dynamische Editionen dauerhaft(er)en Bestand gewährleisten könnte.

Wie lang bleiben die digitalen Editionen funktionsfähig, an denen manche von uns arbeiten? Das ist eine Frage, die wir uns morgens und abends stellen. Nachhaltigkeit und Langzeitverfügbarkeit sind als (noch) ungelöste Probleme allseits erkannt und im Bewusstsein, nicht zuletzt, weil sie von den Drittmittelgebern eingefordert werden.

Ich möchte vorneweg betonen, dass ich an diese Frage aus der Perspektive von jemandem herangehe, die an einer konkreten einzelnen Edition arbeitet, und nicht aus der Perspektive von Informationswissenschaftlern und Datenmanagement-Spezialisten. Ich stelle sie also aus der Perspektive

›Was braucht das Fach, was brauchen die Digitalphilologen?‹, also von der Nachfrageseite, und bin nicht die, die das Angebot überblickt, das dieser Nachfrage gegenübersteht, weil die Angebotslandschaft für den einzelnen Philologen überhaupt nicht überschaubar ist.

Zunächst eine kleine Reflexion darüber, in welchen Hinsichten digitale Editionen altern und das Zeitliche segnen können:

1. Das Handgreiflichste sind zunächst einmal Datenträgertechnologien und Laufumgebungen, die außer Mode kommen. Das führt zum Exitus, falls es nicht durch eine Migration des Projekts aufgefangen werden kann. Ein Beispiel bieten Kiernan/Iacob (o. J.) mit einem Bericht zu den Zwangsmigrationen des ›[Electronic Beowulf](#)‹ ¹2000, ²2003, ³2011 (CD-ROM/DVD-ROM). Wir haben in der kurzen Geschichte der digitalen Edition schon mehrere Medienbrüche erlebt, so besonders von Editionen auf CD bzw. als lauffähige Programme für ein bestimmtes Betriebssystem hin zu Online-Editionen. Dieser Faktor dürfte mit der allgemeinen Umstellung auf browserbasierte Angebote für die Zukunft keine Rolle mehr spielen. Hier findet die Obsoleszenz stattdessen in einem anderen Bereich statt, der auf einer CD ausgeschlossen ist, nämlich:

2. Kodierungen, Verlinkungen, Metadatenformate: Das Auftauchen von falschen Sonderzeichen bzw. Ersatzzeichen für falsch kodierte Sonderzeichen dürfte ärgerlich, aber nicht letal sein; tödlicher ist dann schon ein 404-er, aus welchem Grund auch immer. Häufig müssen im Laufe der Jahre Daten und Ressourcen auf neue Server umgezogen werden; dabei werden oft interne Verzeichnisstrukturen umgestellt, was oft zu solchen Fehlern führt. Dergleichen hat nicht unbedingt das Ableben einer Edition zur Folge, sondern – je nachdem, wie zentral die fehlenden Daten sind – manchmal auch bloß eine etwas eingeschränkte Benutzbarkeit.

3. Diese Quelle von Vitalitätsverlusten lässt sich kaum trennen vom nächsten Punkt, nämlich technischen Aspekten der Präsentationsschicht. Mit zunehmend komplexer werdenden Editionen, die meist keine statischen Inhalte mehr präsentieren, wird auch die eingesetzte Software viel-

fältiger. Wir haben oft auf der Serverseite eine Datenbank laufen mit ihrer Software sowie eine Skriptsprache wie **PHP** (Hypertext Preprocessor) oder **Python** zur Generierung der Seiten, dazu weitere Skriptsprachen auf der Clientseite wie **JavaScript**, **XSLT** (Extensible Stylesheet Language Transformations) und dazu Stylesheet-, also Gestaltungselemente. Jedes dieser Elemente ist eventuell wiederum eingebettet in oder angebunden an ein Framework oder größere Komplettsysteme wie **Bootstrap**, **WordPress**, **Mediawiki**, **IIIF** (International Image Interoperability Framework) usw. Dazu kommt der Einsatz fertiger Module für schwierige Aufgaben wie einen Bildviewer, eine Textsynopse usw. Jeder dieser softwaretechnischen Bausteine stellt eine Abhängigkeit dar, jedes dieser Systeme kann theoretisch außer Betrieb gehen oder von seinen Entwicklern aufgegeben werden. Ein auf diese Weise erzwungener Umbau der Gesamtarchitektur kann natürlich brutal sein. Auch wenn man mit so etwas nicht unbedingt rechnen muss – auf die Dauer kann jeder der genannten Bausteine in dem Sinne veralten, dass er nicht mehr als *state of the art* gilt und eigentlich durch etwas anderes ersetzt werden müsste. Ein solcher Ersatz bedeutet meist einen kompletten Relaunch der Präsentation, aber auch unterhalb dieser Ebene und unsichtbar für die Benutzer müssen die ständigen Versionsanpassungen all dieser Bausteine und Pakete nachvollzogen werden, es sei denn, man hätte es geschafft, sein Gesamtpaket so zu schnüren, dass alle Programmier- und Anpassungsarbeit bei dem außer Haus gewarteten Framework liegen und man selbst für alle spezifischen Inhalte der Edition ein vollständig standardisiertes Datenformat gefunden hat, das in 50 Jahren immer noch gültig und von dem Framework oder einem Nachfolgerframework interpretierbar sein wird. Das wäre der ›heilige Gral‹, damit wäre auch die Frage der Langzeitlauffähigkeit gelöst. Anders gesagt: Das gibt es meines Wissens nicht oder nur in sehr günstigen institutionellen Konstellationen. Unterzieht man sein System den nötigen Aktualisierungen nicht, dann droht eben doch der Exitus, je nachdem wie zentral der Baustein und wie unglücklich im Nachhinein die ursprüngliche Ent-

scheidung für und wider eine Technik war. Sicher haben sich einige der verschwundenen digitalen Editionen auch aus solchen Gründen aus dem Netz verabschiedet.

Es hat schon einen Grund, warum professionelle Webanwendungen, etwa der Banken oder der Presse- und Medienhäuser, kaum je länger als drei Jahre existieren, bevor wieder ein kompletter Relaunch, eine komplette Neuprogrammierung nötig wird. Das Tempo der Entwicklungen in der Webprogrammierung ist enorm. Bei den Neugestaltungen solcher Profi-Webpräsentationen spielt freilich ein letzter Faktor der Obsoleszenz stark mit herein:

4. Aspekte der Ästhetik der Präsentationsschicht, des Bedienungskomforts, der Nutzeransprüche, denen nicht mehr entsprochen wird. Layout, Typographie, Bedienung von Suchfeldern und Eingabemasken, Look-and-Feel, das sind weiche Kriterien. Den aktuellen *state of the art* nicht mehr liefern zu können, führt nicht dazu, dass ein elektronisches Angebot nicht mehr funktioniert. Ob es sich in der Benutzungsfrequenz niederschlägt, wäre interessant zu wissen, aber vermutlich besteht in Wissenschaftskreisen eine recht hohe Toleranz gegenüber altem Webdesign, Unübersichtlichkeit und umständlicher Funktionalität, sofern die Inhalte nützlich sind. Für manche Zwecke haben statische Seiten der 90er vielleicht sogar gewisse Vorteile. Jedenfalls hat sich die Optik von Textdarbietung im Web selbst innerhalb der letzten zehn Jahre noch einmal deutlich geändert; das Tempo, in dem Editionen veralten, nimmt in dieser Dimension also eher zu als ab. Ich möchte betonen, dass es dabei nicht allein um Layout und Optik geht, sondern auch um Werkzeuge wie Suchfunktionen, Filter und Visualisierungen. In diesem Bereich überschneiden sich Aspekte reiner Bequemlichkeit mit Aspekten der Funktion, die sehr wohl zum Kern einer Edition beitragen und ihre Leistungsfähigkeit bestimmen können. In anderen Beiträgen in diesem Band werden schon so viele mögliche künftige Anbindungen von Editionen an Tools angesprochen, dass es evident sein dürfte,

dass auch auf dieser Ebene fertige Editionspräsentationen veralten, wenn sie solchen künftigen Standards nicht mehr entsprechen.

Der Rückblick zeigt so viele damals nicht antizipierbare Entwicklungen, dass wir uns davor hüten sollten, Prognosen über die nächsten zehn oder zwanzig Jahre abzugeben. Natürlich waren die Pioniere immer besonders betroffen, weil sie nicht wissen konnten, was sich in der Folge zum Standard entwickeln würde; was lässt uns glauben, dass dies in den nächsten zwanzig Jahren anders sein wird?

Das Problem als solches ist natürlich erkannt. So adressieren die neueren Bemühungen um eine dauerhafte Bewahrung von Forschungsdaten und um Repositorien für Forschungsdaten genau diese Frage. Das neue Konsortium [Text+](#) im Rahmen der großen Initiative zum Aufbau einer [nationalen Forschungsdateninfrastruktur](#) (NFDI) beschreibt seine Zuständigkeit damit, »text- und sprachbasierte Forschungsdaten langfristig [zu] erhalten und ihre breite Nutzung in der Wissenschaft [zu] ermöglichen« (www.text-plus.org). Text+ läuft seit Herbst 2021 für fünf Jahre und will die Wissenschaft nicht nur beraten und bei der Weiterentwicklung von Standards und Normdaten unterstützen, sondern offenbar auch Dienste und Routinen installieren, mit denen Daten gehostet und archiviert werden können. Editionen sind eine von drei Domänen, also Zuständigkeitsbereichen, von Text+.

Nun erhebt sich freilich die Frage, ob mit ›Forschungsdaten‹ im Kontext von digitalen Editionen eigentlich die digitalen Editionen selbst gemeint sind. Diese Sichtweise fände ich nicht selbstverständlich, weil wir Editionen als Publikationen und damit als Resultate unserer Forschungsarbeit (nicht als anfallende Begleit- und Rohdaten) verstehen und weil wir den Begriff ›Daten‹ dafür etwas befremdlich finden, während viele Definitionen von Forschungsdaten das Instrumentelle hervorheben: Daten, die im Forschungsprozess anfallen und für die Forschung (nach-)genutzt werden. Andererseits sind Definitionen des Begriffs ›Forschungsdaten‹ aber meist so allgemein, dass sie Editionen sehr wohl erfassen sollten – diese sind

digitale Repräsentationen von Texten, deren zentraler Zweck darin besteht, für weitere Forschung zur Verfügung zu stehen, somit sind sie Daten aus der und für die Forschung, also Forschungsdaten. Die Frage mag gesucht klingen, aber sie entscheidet schließlich darüber, in welcher Hinsicht sich etwa eine nationale Forschungsdateninfrastruktur zuständig fühlt für eine digitale Edition.

Wenn man hier etwas weiter die momentane Lage sondiert, dann zeigt sich, dass die Frage doch nicht ganz müßig ist. Ein zentraler Baustein einer Infrastruktur, die Nachhaltigkeit herstellen und garantieren soll, sind Repositorien: Orte, an denen Daten abgelegt, aufbewahrt und zugänglich gehalten werden. Jedoch, sieht man sich einmal in einem großen Verzeichnis von solchen Repositorien um, re3data.org, einem am [KIT](http://www.kit.edu) (Karlsruher Institut für Technologie) angesiedelten »global registry of research data repositories« (www.re3data.org/about), dann stellt man fest, dass dort Projekte als Repositorien eingetragen sind, die man ebenso als Editionsprojekte begreifen kann; so etwa eine Handschriftenpräsentation des [Codex Sinaiticus](http://www.codex-sinaiticus.org) mit Digitalisaten und Transkription: www.re3data.org/repository/r3d100010560.

Das lässt erahnen, dass Editionsprojekte sich selbst als Archiv oder Repositorium von Originaldokumenten verstehen können. Ich kann nicht ermes sen, ob das nur fragwürdige Grenzfälle einer Eintragung in diese Repositorienliste sind, es scheint mir aber sichtbar machen zu können, dass es zwischen digitalen Daten und den Strukturen, in denen die Daten aufbewahrt werden, keine so klaren Unterscheidungen gibt wie in der analogen Welt zwischen dem Buch und der Bibliothek. Ein Buch kann sich zwar ›Archiv‹ betiteln, aber niemand wird wirklich die Dienstleistungen eines Archivs von ihm erwarten. Eine digitale Edition leistet dagegen immer in gewissem Grad sowohl das, was ein Buch als auch was eine Bibliothek leistet, und folglich besteht eine Grauzone in der Zuordnung dessen, was Editionsprojekte sind und tun. Wenn nun eine digitale Edition wie die des [Codex Sinaiticus](http://www.codex-sinaiticus.org) sich selbst als ein Forschungsdatenrepositorium begreift,

was genau wären (in technischer Hinsicht) dann die dort deponierten Forschungsdaten? Und welche Infrastruktur wäre zuständig für den Erhalt des Repositoriums, also der Edition als ganzer?

Das mag jetzt als sehr künstliche Zuspitzung erscheinen. Sicher spielen auch Verständigungsschwierigkeiten eine Rolle, weil die Datenmanagement-Leute ihr Tun mit Begriffen beschreiben, die für uns Mediävisten sehr ungewohnt sind: Da sind wir Stakeholder, da werden Daten und Software kuratiert etc. Es macht aber auf das Problem aufmerksam, das ich hinsichtlich der Dauerhaftigkeit von digitalen Editionen als das zentrale wahrnehme: Editionen bestehen aus Inhalten wie Texten und Bildern sowie einer Darbietungsoberfläche und Zugangsstruktur. Das ist in der Tat eine sehr ähnliche funktionale Zweiteilung wie bei einem Datenrepositorium, denn das muss genauso eine Zugangsstruktur und eine Benutzeroberfläche haben, damit die Daten, die in ihm deponiert sind, nutzbar sind. Der Unterschied zum Repositorium besteht darin, dass weite Teile der Darbietungsoberfläche und der Zugangsstrukturen integrale und wesentliche Bestandteile der Edition bilden. Die Funktionsschicht ist mit den Daten aufs engste verzahnt, und ohne ihre Funktionsschicht sind die Daten weitgehend nutzlos. Folglich bietet die Idee, die Daten allein (also z. B. die XML/TEI-Dateien) in einem Repositorium abzulegen, nicht wirklich Abhilfe gegen das schnelle Altern von digitalen Editionen; da sage ich nichts Überraschendes.

Es gibt nun verschiedene Ansätze, dem Problem zu begegnen:

- generische Viewer für XML/TEI, womit die Wartung und Weiterentwicklung des Viewers in institutioneller Hand liegen könnte und dem einzelnen Projekt erspart bliebe. Das ist auch von Seiten der Infrastrukturanbieter erwogen worden, wurde aber nicht als erfolgversprechend betrachtet (vgl. Buddenbohm [u. a.] 2016): »[...] stellte sich aber heraus, dass mittelfristig wohl nur eine Anwendungskonservierung eine breitere Akzeptanz bei den Nutzerinnen und Nutzern erlangen kann, wenn es um die Langzeitverfügbarkeit einer digitalen Edition als Webanwendung geht.«

- ›Anwendungskonservierung‹: Wenn ich es richtig verstehe, dann handelt es sich dabei um die Emulation eines gesamten Systems, wie es zum Zeitpunkt der Übergabe an das Datenzentrum beschaffen ist, als virtueller Server. Das klingt genial, und ein solches Angebot scheint auch bereits im Portfolio des Göttinger Humanities Data Centre vorhanden. Es gibt auch schon zwei solche einbalsamierten Projekte auf der Göttinger Seite aufzurufen (humanities-data-centre.de/?page_id=688). Leider klappt aber der Zugriff nicht; man landet in einer Dauerschleife von Verbindungsfehlern. Ganz so einfach ist der Umgang mit den lebenden Leichen oder das Channelling in die Cloud also nicht. Selbst wenn es technisch funktionieren sollte, wäre das doch nur eine unveränderliche Archivierung, abgekoppelt vom Internet in dem Sinne, dass Links und Webservices nicht mehr funktionieren. Das ist nicht das, was man mit Nachhaltigkeit meint, und es verlängert die Lebenserwartung einer Edition nicht.
- Ein anderer Ansatz ist die Entwicklung einer generalisierten Arbeits- und Publikationsplattform für Editionen, was eine Homogenisierung von Editionsprojekten *ab ovo* bedeutet. Zu nennen ist hier das schweizerische Projekt ›Nationale Infrastruktur für Editionen‹ (NIE): »Die Plattform soll den spezifischen Bedürfnissen umfangreicher und komplexer Editionsprojekte gerecht werden und insbesondere die elektronische Publikation und die langfristige Verfügbarkeit von Forschungsdaten und -ergebnissen in einem zentralen Bereich der nationalen geisteswissenschaftlichen Forschung gewährleisten« (fee.unibas.ch/de/nie-ine/). 2016 bis 2019 wurden 5 Mio. Franken dafür bewilligt, aber die Homepage der NIE ist im April 2021 zum letzten Mal gesichtet worden (web.archive.org/web/20210423073819/https://www.nie-ine.ch/). Das Unterfangen dürfte also gescheitert sein. Lebendig ist dagegen GAMS, das ›Geisteswissenschaftliche Asset Management System‹ der Uni Graz, wo im Augenblick in der Größenordnung von fast 100 Einzelprojekten gehostet werden (gams.uni-graz.at/context:gams.projekte?locale=de).

Die dort gewählte Herangehensweise bringt es mit sich, dass Editionsprojekte bereits im Benehmen mit der künftigen Publikationsumgebung konzipiert und entwickelt werden müssen. Dies dürfte auch bedeuten, dass dieses Verfahren nicht skalierbar ist: Es können also nicht beliebig viele Projekte betreut werden. Wenn Digitalisierung (im Sinne eines gesamt-kulturellen Prozesses) aber nicht eine verbesserte Skalierbarkeit zeitigt, darf man sie kontraproduktiv nennen.

- Eine weiterte Option läge in einer stärkeren Formalisierung und Abstraktion, indem nämlich das »funktionale[] Zusammenspiel aller Komponenten von Text, Struktur, Layout, Schnittstelle und Metadaten« standardisiert beschrieben wird, und zwar im XML-Universum (Stäcker 2020) oder mit Linked Open Data. Das »Desiderat eines abstrakten Modells für Strukturen und Inhalte digitaler Editionen« hat auch Patrick Sahle bei dem Workshop »Nachhaltigkeit digitaler Editionen« 2018 formuliert (Dängeli 2019).

Ich habe bei dem letztgenannten Ansatz den Eindruck, *this is the way*. Aber ich kann nicht abschätzen, wie aufwendig es bei einer komplexen Edition eigentlich wäre, nachträglich formalisiert und maschineninterpretierbar zu beschreiben, wie die Gesamtheit der algorithmischen Prozesse beschaffen sind, die aus den Textrohdaten eine Benutzererfahrung generieren, und mit welchen Werkzeugen man eine solche Modellierung und Beschreibung überhaupt angehen würde, zumal bestehende Editionen in ihrer Funktionsschicht viel Wildwuchs zeigen dürften, weil es eben bislang keine Standardisierung für diese Prozesse gibt. Mir ist auch nicht klar, wie man sich die Software vorstellen soll, auf der das Paket dann quasi »läuft«, das aus allen Beschreibungsdaten eine dynamische Benutzeroberfläche macht.

Solange also keiner dieser Ansätze wirklich umsetzbar ist, wird es wohl noch eine Weile in punkto Lebenserwartung dabei bleiben, was schon seit den 90ern gilt: Eine Edition lebt so lange, wie sich jemand aktiv um die Softwarebausteine der Funktionsschicht kümmert, plus schätzungsweise fünf bis fünfzehn Jahre.

Literaturverzeichnis

Primärliteratur

Electronic Beowulf. Student Edition, hrsg. von Kevin Kiernan. 3. Aufl. London 2011 (DVD-ROM).

Electronic Beowulf 4.0, hrsg. von Kevin Kiernan. 2015 ([online](#)).

Sekundärliteratur

Buddenbohm, Stefan/Engelhardt, Claudia/Wuttke, Ulrike: Angebotsgenese für ein geisteswissenschaftliches Forschungsdatenzentrum, in: Zeitschrift für digitale Geisteswissenschaften 1 (2016) ([online](#)).

Dängeli, Peter: Die Nachhaltigkeitsproblematik digitaler Editionen – Workshopbericht ([online](#)).

Kiernan, Kevin/Iacob, Emil: Going Online (online erreichbar über: <https://web.archive.org/web/20181004083649/http://ebeowulf.uky.edu:80/>)

Stäcker, Thomas: >A digital edition is not visible< – some thoughts on the nature and persistence of digital editions, in: Zeitschrift für digitale Geisteswissenschaften 2020. text/html Format ([online](#)).

Online-Ressourcen

Bootstrap: <https://getbootstrap.com/>.

Codex Sinaiticus: <https://codexsinaiticus.org/de/>.

Electronic Beowulf: <http://ebeowulf.uky.edu:80/>.

GAMS (Geisteswissenschaftliche Asset Management System): <https://gams.uni-graz.at/>.

HDC (Humanities Data Centre): <https://humanities-data-centre.de/>.

IIIF (International Image Interoperability Framework): <https://iiif.io/>.

Mediawiki: <https://www.mediawiki.org/wiki/MediaWiki>.

NFDI (Nationale Forschungsdateninfrastruktur): <https://www.nfdi.de/>.

PHP (Hypertext Preprocessor): <https://www.php.net/>.

Python: <https://www.python.org/>.

re3data: <https://www.re3data.org/repository/r3d100010560>.

Text+: <https://www.text-plus.org/>.

WordPress: <https://wordpress.com/de/>.

Anschrift der Autorin:

Prof. Dr. Sonja Glauch
Friedrich-Alexander-Universität
Institut für Germanistik
Bismarckstraße 1
91054 Erlangen
E-Mail: sonja.glauch@fau.de