

# Nachhaltige Konzeptionsmethoden für Digital Humanities Projekte am Beispiel der Goethe- PROPYLÄEN

**Kasper, Dominik**

dominik.kasper@adwmainz.de  
Akademie der Wissenschaften und der Literatur,  
Deutschland

**Grüntgens, Max**

max.gruentgens@adwmainz.de  
Akademie der Wissenschaften und der Literatur,  
Deutschland

Vor dem Hintergrund des Akademievorhabens *PROPYLÄEN. Forschungsplattform zu Goethes Biographica* werden zwei aktuelle Konzeptionsmethoden aus dem Bereich des Software-Engineering vorgestellt, die sich unserer Erfahrung nach als geeignete Grundlage für nachhaltige Konzeptionsprozesse in Digital Humanities Projekten erwiesen haben: *Domain Driven Design* sowie *Behaviour Driven Development*. Zunächst erfolgt ein Überblick über Aspekte digitaler Nachhaltigkeit bei der Beantragung geisteswissenschaftlicher Langzeitvorhaben im Akademienprogramm.

Die PROPYLÄEN sind ein 2015 im Akademienprogramm gestartetes Forschungsvorhaben der Klassik Stiftung Weimar, der Sächsischen Akademie der Wissenschaften zu Leipzig und der Akademie der Wissenschaften und der Literatur Mainz (<http://www.goethe-biographica.de>). Geplant ist bei einer Gesamtlaufzeit von 25 Jahren Fortführung und Abschluss von vier (Print-)Publikationsreihen: Die Editionen der Briefe von sowie an Goethe, seine Tagebücher und die Zeugnisse seiner Begegnungen und Gespräche. Die gleichzeitig entstehende Forschungsplattform wird sukzessive alle Biographica in einer technisch und inhaltlich offenen Infrastruktur bereitstellen sowie Recherchezugänge und Visualisierungen auf den Datenbestand anbieten.

Eine Voraussetzung für die Aufnahme eines Neuvorhabens in das Akademienprogramm ist ein ausgearbeitetes Digitalisierungskonzept, zu dem unter anderem eine Strategie zur Langzeitarchivierung und Langzeitverfügbarkeit der Forschungsergebnisse gehört (Vgl. Herrmann 2016: 9). Häufig konzentriert sich dieses auf eine enge Definition von Forschungsdaten, die unter diesem Begriff lediglich edierten Text versteht. Im Falle einer Forschungsplattform wie der

PROPYLÄEN sind nicht nur die in der Datenschicht – Faksimiles, Editionstext, kurz die Gesamtheit der digitalen Forschungsdaten – befindlichen Komponenten für eine nachhaltige und langfristige Bereitstellung vorzusehen (zu Archivierungsschicht und Präsentationsschicht vgl. Pempe 2012: 141). Daneben sollte auch die Auswahl geeigneter Technologien ein Teil der Antragsstrategie sein. Unter Technologie sei alles vom Datenformat über ein Content-Management-System bis zur virtuellen Forschungsumgebung verstanden. Angesichts des Verhältnisses zwischen langfristiger Projektförderung – die Akademienvorhaben werden 12 bis 25 Jahre gefördert – und der technischen Entwicklung ist die Vorstellung der einmaligen Einrichtung einer digitalen Arbeitsumgebung und Publikationsplattform am Anfang der Projektlaufzeit, welche daraufhin für 25 Jahre verwendet würde, nur mit begleitender Wartung und Weiterentwicklung denkbar. Wird das gewählte Datenformat in 30 Jahren noch maschinenlesbar sein? Absolute Sicherheit gibt es in diesem Fall nicht, aber allgemein gilt die Verwendung von (semistrukturierten) Rein-Textformaten – wie XML nach TEI – gegenüber proprietären Formaten als nachhaltiger.

Der *Mehrwert der elektronischen Fassung* entsteht aber vornehmlich durch neue Verbindungs-, Gruppier-, Sortier-, Filter- und Suchmöglichkeiten, die auf dynamischen Verarbeitungsmechanismen, also auf der Geschäftslogik der Anwendung, basieren. Diese nicht-statischen Elemente, die sich insbesondere in der Präsentationsschicht einer geisteswissenschaftlichen Webanwendung konkretisieren, müssen in ihrer Funktionalität genauso nachvollziehbar und reproduzierbar über das Projektende hinaus zur Verfügung stehen.

Komplexer als die Datenschicht ist demnach die dauerhafte Erhaltung der Applikationslogik und Präsentationsschicht: Wie kann die webbasierte Verarbeitung von Anfragen und die Wiedergabe von Ergebnissen auch 10 Jahre nach Projektende noch funktionsfähig gehalten werden? Eine aktuell diskutierte Strategie ist die virtuelle Kapselung sämtlicher Softwarekomponenten einer Applikation, die künftig ein Emulieren aller Komponenten auf aktuellen Betriebssystemen ermöglicht (siehe dazu bspw. <http://recomputation.org/>). Vorhaben wie die PROPYLÄEN müssen dieser Herausforderung während der Projektlaufzeit begegnen, indem sie eine hohe Flexibilität sowie stete Aktualisierung und Anpassung der Technologien ihrer Infrastruktur in der Grundplanung vorsehen. Um diesen Prozess auf eine transparente Grundlage zu stellen, gilt es bei der informationstechnischen Gestaltung des Forschungsvorhabens eine möglichst *gegenstandsnahe* Abstraktion zu wählen und die Anforderungen an die Software auf *formalisierte* Weise zu dokumentieren. Folgende Annahmen und Schlüsse gehen dieser Überlegung voran:

- Es wird während der Projektlaufzeit intern wie extern neue Erkenntnisse zu Goethes Zeit und Wirken geben,

die unseren Blick auf den Forschungsgegenstand verändern. Dennoch ist es möglich, übergreifende Wissensobjekte zu identifizieren und in Form von informatischen *Entitäten* abzubilden.

- *Technologieunabhängige, anwendungsorientierte* und *allgemeinverständliche* Funktionsbeschreibungen der Forschungsplattform sind formulierbar. Neue Erkenntnisse im Bereich der Goethe-Forschung wie im Bereich der *usability* werden diese Anforderungen innerhalb der Projektlaufzeit höchstwahrscheinlich verändern.

In der *Konzeptionsphase* einer nachhaltigen digitalen Grundlage für den Daten-, Präsentations- und Applikationsteil der Forschungsplattform PROPYLÄEN sind daher zwei Fragen leitend, deren Beantwortung uns nicht nur für Langzeitvorhaben mit Digital-Humanities-Anteil zentral erscheint (vgl. zu Prozessphasen und zur Konzeption hier und folgend: Schrader 2016: Step 14–21, zur Konzeption Step 17):

- Wie lassen sich *Biographica* Goethes für technische und geisteswissenschaftliche Anforderungen nachhaltig modellieren bei gleichzeitiger Bewahrung der Flexibilität in der Anwendungsarchitektur für die Integration gegebenenfalls noch nicht bekannter digitaler Ressourcen?
- Wie lassen sich die funktionalen Anforderungen an eine digitale Rechercheumgebung aus der Fachcommunity nachhaltig formulieren und dokumentieren?

Zur Beantwortung können zwei Konzeptionsmethoden aus dem Bereich des Software-Engineering fruchtbar gemacht werden: *Domain Driven Design (DDD)* (siehe Evans 2003, Vernon 2013) und *Behavior Driven Development (BDD)* (siehe North 2016). Beiden ist gemein, dass sie unabhängig von Datenformaten, Programmiersprachen oder Präsentationstechnologien operieren.

*DDD* nimmt an, dass einem Konzeptions- bzw. Entwicklungsprozess dann die größten Erfolgschancen und daraus resultierend die beste Nachhaltigkeit zukommen, wenn die *virtuellen* Komponenten des informatischen Modells ausgehend von ihren *realen* Entsprechungen gebildet werden. Damit kommt der fachwissenschaftlichen Logik, die sich aus dem Wissen von Domänen-Experten (im Anwendungsfall Goethe-Philologen) und deren Niederlegung in publizierten Editionsbanden ableitet, eine zentrale Rolle für die Modellierung zu. Die Ausmodellierung der Wissensdomäne entsteht iterativ in einem konstanten kommunikativen Prozess aller Projektbeteiligten. *DDD* löst die gängige und nicht optimale Praxis einer ausschließlichen Bedarfsformulierung durch Geisteswissenschaftler und einer darauffolgenden Umsetzung durch Informatiker zugunsten eines gemeinsamen Modellierungsprozesses unter Verwendung einer für alle Beteiligten verständlichen (*ubiquitären*) Sprache auf. Dieser im Rahmen eines

Projektes von allen Beteiligten zu entwickelnden Sprache liegen eine Reihe von Komponenten für ein Modell der Wissensdomäne (*domain model*) zugrunde, das in seiner Gesamtheit alle Eigenschaften, Beziehungen und „Geschäftsprozessen“ der zukünftigen Anwendung abbilden kann. Die wichtigsten Komponenten für die Fachdomäne „PROPYLÄEN“ sind:

- Objekte mit eigener Identität (*entities*): Briefe, Faksimiles, Personen, Orte, Werke
- Objekte, die sich über die Gesamtheit ihrer Eigenschaften definieren (*value objects*): Geokoordinaten (Länge-Breite), Datierungen (Anfangs-Enddatum), Korrespondenzvorgang (Sender-Empfänger) etc.
- Die Beziehungen dieser Objekte untereinander (*associations*): Person # erwähnt in: Brief
- Zusammenfassungen von *entities*, *value objects* und *associations* zu Aggregaten (*aggregates*); der Zugriff auf einzelne Komponenten eines Aggregates darf ausschließlich über die Entität an der Aggregat-Wurzel (*aggregate root*) erfolgen. Aggregate gewährleisten logische Konsistenz und funktionale Integrität einer Anwendungs- bzw. Wissensdomäne: Brief (Brieftext, Kommentar, Faksimile, Datierung, im Brief erwähnte Orte und Personen)
- Dienste zwischen Objekten der Wissensdomäne (*services*): Sortiere nach Datum etc.

Die Fachdomäne kann grafisch dargestellt werden. Abbildung 1 dient der Illustration des DDD-Konzepts am Beispiel eines Teils der PROPYLÄEN-Domäne.

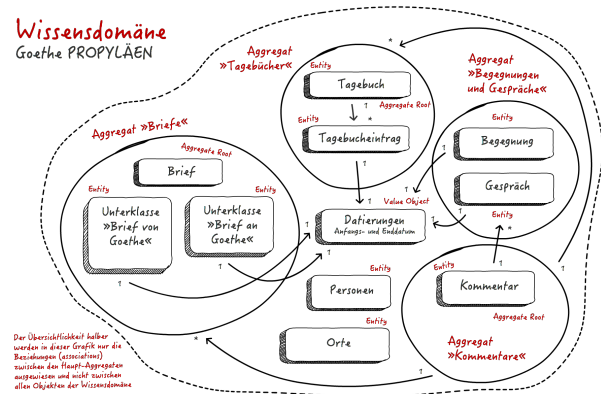


Abbildung 1: Wissensdomäne Goethe PROPYLÄEN

Dan Norths Idee des *BDD* ist, Anforderungen an die geplante Applikation in Form von *user stories* und verhaltensorientierten Testszenarien im Vorfeld der Realisierung unter Zuhilfenahme einer ubiquitären Sprache zu beschreiben. Durch die Verwendung von Schlüsselwörtern in formalisierter Notation lässt sich eine einheitliche und leicht verständliche Dokumentation erstellen. Aufgrund starker Nutzerorientierung wird *BDD* im Qualitätsmanagement von Softwareanwendungen

zu den *Akzeptanztests* gerechnet (neben Unit-Tests, Funktionstests, Performance-Tests oder manuellen Tests).

Abbildung 2 zeigt ein BDD-Testszenario in den Goethe-PROPYLÄEN, das auf Gherkin-Notation basiert (Siehe <https://github.com/cucumber/cucumber/wiki/Gherkin>, Stand 26.08.2016). Die Schlüsselwörter *Funktionalität*, *Beschreibung* und *Details* enthalten übergeordnete Informationen, während im Bereich der *Szenarios* mit Hilfe der Operatoren *Szenario* bzw. *Gegeben*, *Wenn*, *Dann*, *Und* einzelne Anforderungen formuliert werden. Die Ausführungen können dabei allgemein gehalten werden und lassen sich im späteren Verlauf auffächern.

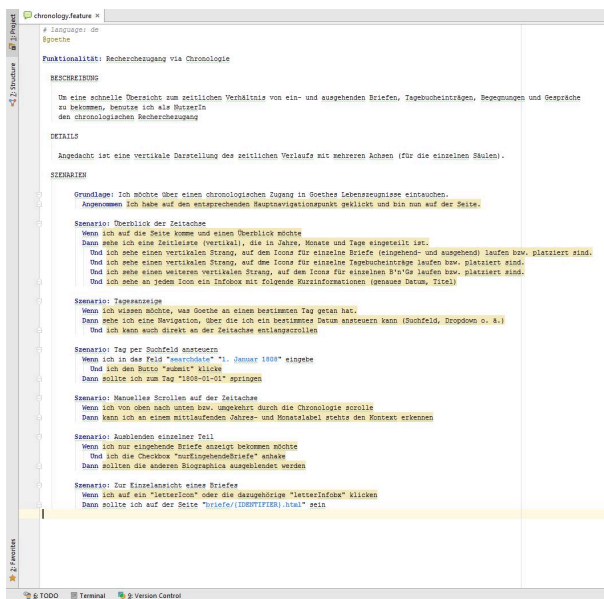


Abbildung 2: BDD-Testszenario in Gherkin-Notation

Ein Vorteil ist, dass Szenarios durch den Einsatz einer verständlichen, formalisierten Sprache nicht nur gemeinsam zwischen Geisteswissenschaftlern und Informatikern während des Entwicklungsprozess formuliert werden, sondern dass bei syntaktisch korrekter Formulierung diese als Grundlage für automatisierte Testverfahren der Präsentationsschicht einer Webanwendung herangezogen werden können (bspw. mit dem BDD Testing-Framework behat; siehe <http://docs.behat.org> (Stand 26.08.2016)). BDD ermöglicht somit eine zeit-ökonomische Überprüfung aller Bereiche und Funktionalitäten der Präsentationsschicht bei gleichzeitiger Dokumentation der Funktionalitäten in Reintext und natürlicher Sprache. Änderungen in der technischen Infrastruktur einer Plattform (Softwarewechsel der Datenbankschicht, Einführung neuer ProgrammROUTINEN) können mit deutlich gesteigerter Verlässlichkeit im Hinblick auf eine gleichbleibende, dauerhafte Funktion der Webanwendung durchgeführt werden.

Die gemeinsame Modellierung einer Wissensdomäne mittels DDD und BDD führt notwendigerweise

zu intensiver Auseinandersetzung mit allen geistes- wie informationstechnischen Aspekten des Projektes. Dadurch wird projektintern eine kommunikative Ebene entstehen, die eine gemeinsame und unmißverständliche Sprache verwendet und perspektivisch eine verbesserte Tiefenschärfe in Bezug auf funktionale Aspekte der zu realisierenden Forschungsanwendung ermöglicht.

DDD und BDD helfen als Konzeptionsmethoden, alle gemeinsam getroffenen Entscheidungen über die gesamte Projektlaufzeit transparent und nachvollziehbar zu dokumentieren. Da es sich gleichzeitig im Bezug auf die BDD-Akzeptanztests um „ausführbares Wissen“ handelt, wird eine dauerhafte und gleichbleibende Funktionalität der Präsentationsschicht geisteswissenschaftlicher Webanwendungen gewährleistet. Dies trägt erheblich zu gesteigerter Nachhaltigkeit digitaler Komponenten eines geisteswissenschaftlichen Forschungsprojektes bei.

## Bibliographie

**Evans, Eric** (2003): *Domain-Driven Design*. Tackling Complexity in the Heart of Software. Boston et al.

**Herrmann, Dieter** (2016): „E-Humanties im Akademienprogramm“, in: Union der Deutschen Akademien der Wissenschaften (Hrsg.): *Die Wissenschaftsakademien – Wissensspeicher für die Zukunft. Forschungsprojekte im Akademienprogramm*. Berlin / Mainz 9–11.

**North, Dan** (2016): *Introducing BDD*. <https://dannorth.net/introducing-bdd/> [letzter Zugriff 26. August 2016].

**Pempe, Wolfgang** (2012): „Geisteswissenschaften“, in: Neuroth, Heike et al. (eds.): *Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme*. Göttingen 137–159.

**Schrade, Torsten** (2016): *Nachhaltige Online-Applikationen in den Geisteswissenschaften – Modellierung und Implementierung*. Vortrag vom 11. April 2016, Hochschule Mainz, <http://metacontext.github.io/nachhaltige-online-apps/> [letzter Zugriff 26. August 2016].

**Vernon, Vaugh** (2013): *Implementing Domain-Driven Design*. Boston et al.