

## Science Data Center für Literatur

### Ulrich, Mona

mona.ulrich@dla-marbach.de  
Deutsches Literaturarchiv Marbach

### Hess, Jan

jan.hess@dla-marbach.de  
Deutsches Literaturarchiv Marbach

### Kamzelak, Roland

roland.kamzelak@dla-marbach.de  
Deutsches Literaturarchiv Marbach

### Kramski, Heinz Werner

heinz.werner.kramski@dla-marbach.de  
Deutsches Literaturarchiv Marbach

### Jung, Kerstin

kerstin.jung@ims.uni-stuttgart.de  
Institut für maschinelle Sprachverarbeitung der  
Universität Stuttgart

### Kuhn, Jonas

jonas.kuhn@ims.uni-stuttgart.de  
Institut für maschinelle Sprachverarbeitung der  
Universität Stuttgart

### Schlesinger, Claus-Michael

claus-michael.schlesinger@ilw.uni-stuttgart.de  
Institut für Literaturwissenschaft / Digital Humanities der  
Universität Stuttgart

### Viehhauser, Gabriel

gabriel.viehhauser-mery@ilw.uni-stuttgart.de  
Institut für Literaturwissenschaft / Digital Humanities der  
Universität Stuttgart

### Schembera, Björn

schembera@hlrs.de  
Höchstleistungsrechenzentrum Stuttgart

### Bönisch, Thomas

boenisch@hlrs.de  
Höchstleistungsrechenzentrum Stuttgart

### Kaminski, Andreas

kaminski@hlrs.de  
Höchstleistungsrechenzentrum Stuttgart

Die Digitalisierung verändert die Bedingungen für die Produktion, Distribution und Rezeption und damit auch für die Erforschung von Literatur. In den Digital Humanities stehen dabei bislang insbesondere die neuen Möglichkeiten der digitalen Auswertung (Distant/ Scalable Reading) und die Digitalisierung vorhandener Druckbestände im Zentrum der Aufmerksamkeit. Die veränderten medialen Bedingungen führen jedoch nicht nur zu einer Übersetzung von gedruckten Texten in digitale Objekte, sondern bringen selbst produktiv neue literarische Formen und Gattungen hervor, für die computergestützte Elemente konstitutiv sind. Hierzu zählen etwa literarische Hypertexte, Blog-Formate, computergestützte kollektive und kollaborative Projekte, literarische Tweets und Twitter-Bots, Texte und Textgeneratoren, die auf computerlinguistische Methoden setzen, schließlich auch frühere Formen computergestützter Literaturproduktion wie der Poesieautomat von Hans Magnus Enzensberger oder die *Stochastischen Texte* von Theo Lutz. (Rettberg 2019, Suter 2012, Tomaszek 2011, Lutz 1959). Hinzu kommen im Bereich Literaturforschung und -archive zunehmend digitale Vor- und Nachlässe, die eine Vielzahl von unterschiedlichen Datenträgern und Datenformaten beinhalten.

Das jüngst ins Leben gerufene interdisziplinäre *Science Data Center für Literatur (SDC4Lit)* hat sich das Ziel gesetzt, die Anforderungen, die Digitale Literatur an ihre Archivierung, Erforschung und Vermittlung stellt, systematisch zu reflektieren und entsprechende Lösungen für einen nachhaltigen Datenlebenszyklus Literatur langfristig umzusetzen.

Für die Archivierung, Analyse und Vermittlung von Digitaler Literatur wird eine Forschungsplattform entwickelt. Da eine solche Plattform nur in der interdisziplinären Zusammenarbeit zu bewerkstelligen ist, sind im Projekt Partner mit unterschiedlichen Expertisen in den einzelnen Teilbereichen vereint, nämlich das Deutsche Literaturarchiv Marbach, das Höchstleistungsrechenzentrum Stuttgart, sowie das Institut für Maschinelle Sprachverarbeitung und die Abteilung Digital Humanities der Universität Stuttgart.

Die born-digital Bestände des Deutschen Literaturarchivs bestehen zum einen aus digitalen Nachlässen und zum anderen aus archivierten netzliterarischen Werken. Der umfangreichste digitale Nachlass am Deutschen Literaturarchiv ist von Friedrich Kittler und umfasst 1,5 Millionen Dateien. Zur deutschsprachigen Netzliteratur können weitaus weniger Objekte gezählt werden. Netzliteratur ist durch Verlinkungen und Multimedialität geprägt. Das erschwert die Definition von Objektgrenzen und führt zu nichtlinearen Objektstrukturen, die in der Rezeption nichtlineare Handlungen ermöglichen

Zum einem scheinen sich diese Texte also zur Anwendung computergestützter und computerlinguistischer Methoden besonders anzubieten, da sie genuin in elektronischer Form vorliegen. Zum anderen bringt gerade diese Form für ihre Archivierung und Bereitstellung eine Reihe von besonderen Anforderungen mit sich.

Digitale Nachlässe sind aufgrund großer Mengen an Daten ohne computergestützte Methoden kaum erschließbar und zugänglich zu machen. Um auf diese wachsende Herausforderung in Archiven und Bibliotheken einzugehen, soll der Einsatz von Methoden der Digital Humanities für die inhaltliche Erschließung textbasierter born-digital Bestände erprobt werden. Wenn digitale Nachlässe bereits obsoletere Dateiformate enthalten, sind diese nicht ohne vorherige Formatmigration für aktuelle computergestützte Analysen zugänglich.

Auch literarische Webseiten sind von den hochfrequenten Erneuerungszyklen digitaler Technik betroffen. Weiterentwicklungen der Betriebssysteme, der Browser, des HTML-Standards und gängiger Webtechnologien können zu fehlerhafter Darstellung oder fehlenden Funktionen einer Webseite führen. Um ein Werk der Netzliteratur dokumentieren zu können, sind daher neue Formen der Modellierung von Texten, die über eine bloß lineare Form hinausgehen, gefragt.

Diese und weitere Bestände sollen mit modernen digitalen Methoden erschlossen, erforscht und vermittelt werden können. Im Zentrum stehen daher der Aufbau verteilter langzeitverfügbarer Repositories für Digitale Literatur inklusive Forschungsdaten und die Entwicklung der SDC4Lit-Forschungsplattform. Die Repositories werden vom Projekt und seinen Kooperationspartnern regelmäßig erweitert und bilden den zentralen Speicher für das Harvesting von Netzliteratur und weiteren Formen elektronischer Literatur im künftigen Betrieb des SDC. Die Forschungsplattform bietet die Möglichkeit zum computergestützten Arbeiten mit den Beständen der Repositories.

Bereits entwickelte oder in der Entwicklung befindliche Ansätze zur Archivierung und Bereitstellung von WARC-Archiven (Lin et al. 2017), Textkorpora (Fischer et al. 2019) und Analysefunktionen (Hinrichs et al. 2010) sowie strukturierte Reflexionen eigener Strategien (Kramski, von Bülow 2011) weisen auf eine modulare und integrierte Lösung bei der Bereitstellung von Daten und Services. Die entsprechend geplante modulare Architektur der bereitgestellten Services ermöglicht eine nachhaltige Integration von Repositories und Analysemethoden sowie die Möglichkeit zur späteren bedarfsorientierten Einbindung von Korpora und Analysewerkzeugen.

Für die Entwicklung des Repositories und der Forschungsplattform ist der Kontakt zu an der Herstellung, Verbreitung, Erforschung und Vermittlung von elektronischer Literatur beteiligten Communities ein entscheidendes Element. Diese Beteiligung wird über einen mehrköpfigen Beirat und Outreach-Maßnahmen wie Workshops, Seminare und die Arbeit mit Fokusgruppen

erreicht. Eine wichtige Aufgabe des Projekts ist in diesem Zusammenhang die Modellierung von Formen digitaler Literatur, die zunächst beispielorientiert im Umgang mit einem bereits vorhandenen Corpus digitaler Literatur erfolgt.<sup>1</sup> Daraus entstehen sowohl technische als auch gattungspoetologische Herausforderungen, etwa bei der Begriffsbildung (digitale vs. elektronische Literatur), bei der medienbezogenen Abgrenzung von digitaler und nicht-digitaler und post-digitaler Literatur, und schließlich in Bezug auf gattungspoetologische und literaturgeschichtliche Fragen zur elektronischen Literatur seit den 1950er Jahren mit einem Fokus auf den deutschsprachigen Raum und mit Blick auf internationale Entwicklungen in Literatur und Literaturforschung. (Block, 2004; Gould, 2012; Rettberg, 2019; Seça, 2015)

Neben digitalen Objekten und entsprechenden Metadaten wird auch ein Repository der anfallenden Forschungsdaten nachvollziehbar und nachhaltig gespeichert. Zu den Forschungsdaten zählen erstens die bei der Arbeit des SDC anfallenden Forschungsdaten, insbesondere solche, die für das Anbieten von Diensten auf der Plattform notwendig sind, etwa mittels Machine Learning errechnete Datenmodelle für an das Corpus angepasste computerlinguistische Analysewerkzeuge (Eigennamenerkennung, Parser, Topic Models etc.). Zweitens soll das Repository die Möglichkeit bieten, die von Nutzer\*innen der Forschungsplattform generierten Forschungsdaten strukturiert zu speichern und für die weitere Forschung zur Verfügung zu stellen, etwa Annotationen oder ergänzte Metadaten zu einzelnen Objekten oder zu Objektklassen.

Die Sammlung, Bereitstellung, Erforschung und Vermittlung von Literatur im medialen Wandel ist eine Aufgabe, die Forschung und Archive gleichermaßen betrifft. SDC4Lit verfolgt deshalb das Ziel, diese Aufgabe und die entsprechenden Unteraufgaben interdisziplinär zu bearbeiten.

## Fußnoten

1. Deutsches Literaturarchiv Marbach: Literatur im Netz, <http://literatur-im-netz.dla-marbach.de/>, Zugriff 20.9.2019.

## Bibliographie

**Block, Friedrich W.** (2004): *p0es1s. Ästhetik digitaler Poesie = The aesthetics of digital poetry*. Erscheint anlässlich der Ausstellung "p0es1s. Digitale Poesie" im Kulturforum Potsdamer Platz, Berlin, 13. Februar bis 4. April 2004. Ostfildern: Hatje Cantz.

**Gould, Amanda Starling** (2012): „A Bibliographic Overview of Electronic Literature“. In: *Electronic Literature Directory* o.V.

**Hinrichs, Erhard W., Marie Hinrichs und Thomas Zastrow** (2010): *WebLicht: Web-Based LRT Services*

for German, Proceedings of the ACL 2010 System Demonstrations, S. 25–29.

**Kramski, Heinz Werner, Ulrich von Bülow** (2011): „Es füllt sich der Speicher mit köstlicher Habe“ – Erfahrungen mit digitalen Archivmaterialien im Deutschen Literaturarchiv Marbach, in: Caroline Y. Robertson von Trotha. Robert Hauser (Hg.), Neues Erbe : Aspekte, Perspektiven und Konsequenzen der digitalen Überlieferung, Karlsruhe: KIT Scientific Publishing, S. 141-162.

**Rettberg, Scott** (2019): *Electronic literature*. Cambridge, UK: Polity Press.

**Seiça, Álvaro** (2015): *Um Feixe Luminoso: Uma Leitura da Coleção de Literatura Eletrônica Portuguesa*. Florianópolis: Universidade Federal de Santa Catarina.

**Suter, Beat** (2012): *Von Theo Lutz zur Netzliteratur. Die Entwicklung der deutschsprachigen elektronischen Literatur*, <[https://netzliteratur.net/suter/Geschichte\\_der\\_deutschsprachigen\\_Netzliteratur.pdf](https://netzliteratur.net/suter/Geschichte_der_deutschsprachigen_Netzliteratur.pdf)>, Zugriff am 31.12.2019.

**Tomaszek, Patricia** (2011): German Net Literature: In the Exile of Invisibility, < <http://elmcip.net/critical-writing/german-net-literature-exile-invisibility> >, Zugriff am 19.9.2019.