

GAMS, DERLA und Data Science

Experimentelle Datenanalyse der langzeitarchivierten Forschungsdaten

Selina Galka, Sebastian Stoff

Abstract: *Als digitale Edition am GAMS erfüllt DERLA (Digitale Erinnerungslandschaft) die nötigen Grundvoraussetzungen für die Anwendung aussagekräftiger digitaler Methoden. Dabei stellt DERLA nicht nur Werkzeuge zur Analyse der erfassten Erinnerungsorte bereit, sondern garantiert auch die Verwendung von nachhaltigen Forschungsdaten in den bereitgestellten Anwendungen. Über das GAMS können archivierte Forschungsdaten auch für eigene Untersuchungen weiterverwendet werden, wodurch DERLA einen wesentlichen Baustein zur nachhaltigen weiteren Erforschung österreichischer Zeitgeschichte mit Hilfe digitaler Methoden liefert und liefern kann.*

Keywords: *Digitale Edition; Digitale Langzeitarchivierung; Datenanalyse; FAIR-Prinzipien; Digitale Geisteswissenschaften; Open-Science; Forschungsdaten; RDM (Research Data Management)*

Einleitung

In den Geschichtswissenschaften gilt es als allgemein akzeptierte Tatsache, dass jede noch so methodisch gute Auswertung nur so gut ist wie das zugrundeliegende Quellenmaterial. Mit der Verwendung von digitalen bzw. digitalisierten Quellen bekommt diese Ansicht jedoch eine zusätzliche Bedeutung. Neben inhaltlichen (und natürlich klassisch-archivalischen) Kriterien gilt es nun auch die spezifischen Bedingungen der digitalen Welt zu berücksichtigen, wie beispielsweise Langlebigkeit oder Verständlichkeit verwendeter Forschungsdaten. Welche Aussage besitzt eine stichhaltige und methodisch gute historische Analyse, wenn das verwendete Quellenmaterial nicht mehr verfügbar oder gar nicht mehr verstehbar ist? Man denke beispielsweise an gesicherte Aufzeichnungen auf alten Datenträgern wie Disketten.

Nicht nur das bloße Auslesen der Daten ist eine Herausforderung, sondern auch deren sinnvolle Interpretation.¹

Wir behaupten, dass DERLA² als digitale Edition³ am GAMS⁴ die nötigen Voraussetzungen für die Anwendung aussagekräftiger digitaler Methoden⁵ erfüllt. DERLA stellt nicht nur Werkzeuge zur Analyse der erfassten Erinnerungsorte bereit, sondern garantiert auch die Verwendung von nachhaltigen Forschungsdaten in den bereitgestellten Anwendungen. Über die Schnittstellen des GAMS können die digitalen Inhalte auch für eigene Untersuchungen weiterverwendet werden.⁶

Der Beitrag gliedert sich in zwei Abschnitte. Der erste Teil versucht eine theoretische und grundlegende Annäherung an das Thema der digitalen Nachhaltigkeit aus der Perspektive von digitalen Editionen. Es werden wichtige Prinzipien, Konventionen und Referenzmodelle der digitalen Langzeitarchivierung vorgestellt, wie zum Beispiel die FAIR-Prinzipien und die Bedeutung des OAIS-Modells. Im zweiten

-
- 1 Vgl. Fotis Jannidis; Zahlen und Zeichen; in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities*, Stuttgart 2017, S. 63–67, DOI: 10.1007/978-3-476-05446-3_23 (abgerufen 8.9.2022).
 - 2 »DERLA ist ein interdisziplinäres Dokumentations- und Vermittlungsprojekt des Centrums für Jüdische Studien (CJS), des Zentrums für Informationsmodellierung (ZIM), dem Arbeitsbereich Geschichtsdidaktik der Karl-Franzens-Universität Graz und von _erinnern.at_ – Nationalsozialismus und Holocaust: Gedächtnis und Gegenwart. Neben der Dokumentation aller Erinnerungsorte und -zeichen für die Opfer sowie die Orte des Terrors des Nationalsozialismus in Österreich setzt es sich die Entwicklung digitaler Vermittlungsangebote zum Ziel.« – DERLA Digitale Erinnerungslandschaft Webauftritt und online Dokumentation – Siehe Lamprecht, Gerald et al., *Online-Dokumentation des DERLA Projektes*: <https://gams.uni-graz.at/archive/objects/context:derla/methods/sdef:Context/get?mode=about&locale=de>, CJS/ZIM Universität Graz (abgerufen 22.09.2022).
 - 3 Als Einführung zum Thema »Digitale Edition« empfehlen wir – Patrick Sahle; *Digitale Edition*; in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities*, Stuttgart 2017, 234–248, DOI: 10.1007/978-3-476-05446-3_23 (abgerufen 8.9.2022).
 - 4 »Das GAMS ist ein OAIS-konformes und zertifiziertes Asset Management System zur Verwaltung, Publikation und Langzeitarchivierung digitaler Ressourcen aus allen geisteswissenschaftlichen Fächern. Es bietet MitarbeiterInnen aus Forschung und Lehre, aber auch Studierenden in Projekten die Möglichkeit, diese Ressourcen zitabel und mit Metadaten versehen zu verwalten und zu veröffentlichen« – Siehe Steiner, Elisabeth; Stigler, Johannes; *Startseite des GAMS*, ZIM Zentrum für Informationsmodellierung – Universität Graz, <https://gams.uni-graz.at/> (abgerufen am 22.09.2022).
 - 5 Unter »digitalen Methoden« verstehen wir u.a. quantitative Analysen (wie statistische Auswertungen) aber auch Auswertungen basierend auf geografischen Informationssystemen, die im Wesentlichen computerbasiert sind. Vgl. Kapitel »Digitale Methoden« in: Fotis Jannidis/Hubertus Kohle/Malte Rehbein (Hg.), *Digital Humanities*, Stuttgart 2017, 214, DOI: 10.1007/978-3-476-05446-3_23 (abgerufen 8.9.2022);
 - 6 Vgl. Elisabeth Steiner/Johannes Stigler, *GAMS-Dokumentation*, Zentrum für Informationsmodellierung – Universität Graz, URL: <https://gams.uni-graz.at/o:gams.doku> (abgerufen 22.09.2022).

Teil werden Aspekte der Datenhaltung und -analyse von DERLA als digitale Edition am GAMS vorgestellt. Es wird gezeigt, inwiefern DERLA die im ersten Abschnitt vorgestellten Prinzipien einhält und darüber einerseits selbst digitale Analyseverfahren für die Auswertung der Forschungsdaten anbietet und andererseits die nötige langzeitstabile Grundlage (in Form von qualitätsvollen Forschungsdaten) für erweiterte Analysen im Rahmen digitaler Methoden bereithält.

Digitale Editionen und Nachhaltigkeit

Noch nie war es so leicht, seine Texte und Daten an den Mensch zu bringen. In der Welt des »Digitalen« gestaltet es sich äußerst leicht, schnell und einfach digitale Inhalte zu erstellen und zu publizieren. Genauso schnell können sich digitale Inhalte aber auch verändern, plötzlich nicht mehr erreichbar sein oder sogar verschwinden. Im Forschungsbereich der Digitalen Geisteswissenschaften und im Hinblick auf digitale Editionen sind somit Langzeitarchivierung und digitale Nachhaltigkeit äußerst wichtige Aspekte.

Digitale Editionen erschließen geisteswissenschaftliche Objekte und stellen sie einem breit gefächerten Nutzer:innenkreis zur Verfügung, wobei sie einem »digitalen Paradigma« folgen, d.h. sie sind im Druck nicht reproduzierbar, ohne maßgebliche Funktionalitäten zu verlieren.⁷ Neben der Erarbeitung von Datenmodellen für die editionswissenschaftlichen Objekte, allfälliger Analysen und der Entwicklung eines Interfaces, spielt auch hier digitale Nachhaltigkeit eine wichtige Rolle.

Mit Langzeitarchivierung bzw. digitaler Nachhaltigkeit ist gemeint, dass digitale Ressourcen bewusst, dauerhaft und nachhaltig verwaltet werden, so dass sie unabhängig von den ständigen technischen Weiterentwicklungen und auch über Format-, Soft- und Hardwaregrenzen hinweg erhalten bleiben.⁸ Digitale Editionen nutzen neue Möglichkeiten, um Texte bzw. Daten aufzubereiten und sie darzustellen, was aber auch auf Kritik bei Editor:innen in der Fachcommunity stößt. Digitale Nachhaltigkeit ist somit nicht nur wichtig im Hinblick auf verantwortungsvollen Umgang mit digitalen Ressourcen, transparente Strukturen, die Erschließung der Ressourcen durch Annotationsstandards und Metadaten, sondern auch, um die Akzeptanz von digitalen Editionen innerhalb der Fachcommunity zu stärken.⁹

7 Vgl. Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung*, Norderstedt 2013, 148.

8 Vgl. Corinna Mayer, *Langzeitarchivierung*, Edlex, URL: <https://edlex.de/index.php?title=Langzeitarchivierung> (abgerufen 24.9.2022).

9 Vgl. Johannes Stigler, *Digitale Nachhaltigkeit*, in: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«, KON-DE Weißbuch, Graz 2021, PID: o:konde.6 (abgerufen 20.8.2022).

Die wichtigsten Pfeiler einer digitalen Nachhaltigkeitsstrategie sind nach Johannes Stigler¹⁰:

- offene, dokumentierte, maschinen- und menschenlesbare (Meta-)Datenformate
- offene Daten und Inhalte
- die Verwendung von offener und freier Software (open source)
- technische und institutionelle Infrastrukturen, die die digitalen Ressourcen unter der Prämisse der sich ständig ändernden technologischen Bedingungen kuratieren

Es existieren unterschiedliche Institutionen und Initiativen, die sich diesem Thema annehmen und Richtlinien entwerfen. So vernetzt das deutsche Kompetenznetzwerk *nestor* Forschungs- und Gedächtnisinstitutionen wie Museen, Bibliotheken und Archive, die sich mit der Langzeitverfügbarkeit von digitalen Ressourcen beschäftigen. Das Netzwerk fördert die Entwicklung von Standards und die Verbreitung dieser in den jeweiligen Communities. Verschiedene Arbeitsgruppen setzen sich u.a. mit dauerhafter Zitierbarkeit, Archivstandards oder Formaterkennung auseinander. Außerdem veranstaltet das Kompetenznetzwerk regelmäßig Workshops zum Thema der digitalen Nachhaltigkeit und bietet neben dem *nestor*-Handbuch, einer Enzyklopädie zur digitalen Langzeitarchivierung, weiteres Material sowie Ratgeber an.¹¹ Als Grundlage für eine disziplinen- und länderübergreifende Nachnutzung von Forschungsdaten dienen außerdem die FAIR-Prinzipien, welche 2016 veröffentlicht wurden.¹²

10 Vgl. Stigler, Digitale Nachhaltigkeit.

11 Vgl. Heike Neuroth et al. (Hg.), *nestor* Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Glückstadt 2016.

12 Vgl. FAIR-Prinzipien, URL: <https://www.go-fair.org/fair-principles/> (abgerufen 15.8.2022); vgl. Johannes Stigler, FAIR-Prinzipien, in: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«, KONDE Weißbuch, Graz 2021, PID: o:konde.7 (abgerufen 20.8.2022).

FAIR-Prinzipien

Die FAIR *Guiding Principles for scientific data management and stewardship* definieren, dass Forschungsdaten

- auffindbar (**f**indable)
- zugänglich (**a**ccessible)
- interoperabel (**i**nteroperable)
- und wiederverwendbar (**r**e-usable)

sein sollen.¹³

Näher ausgeführt besagen die FAIR-Prinzipien, dass für die Weiterverwendung von (Meta-)Daten diese zunächst auffindbar sein müssen.¹⁴ Daher sollten die in einer digitalen Ressource entstandenen (Meta-)Daten mit maschinenlesbaren Metadaten und persistenten Identifikatoren versehen sein, außerdem sollten sie auch den Identifikator der Daten anführen, die sie beschreiben. Als persistente Identifikatoren werden stabile Referenzen auf digitale Objekte bezeichnet, wobei hier unterschiedliche Systeme verwendet werden können – z.B. DOI (*Digital Object Identifier*), handle.net oder URNs (*Uniform Resource Names*).¹⁵ Für die Beschreibung der Metadaten gibt es standardisierte Datenformate wie z.B. das Schema *Dublin Core der Dublin Core Metadata Initiative*, welches ein kompaktes Vokabular zur Beschreibung von digitalen Ressourcen bereitstellt.¹⁶

Laut dem FAIR-Prinzip der Zugänglichkeit sollten die (Meta-)Daten so aufbewahrt werden, dass sie über ein anerkanntes, standardisiertes Kommunikationsprotokoll (z.B. OAI-PMH), welches offen und frei zugänglich ist, aufgerufen werden können. Das Kommunikationsprotokoll sollte es außerdem erlauben, Authentifizierungs- und Autorisierungsprozeduren aufzusetzen. Hinzu kommt, dass die Metadaten der digitalen Ressource auch dann noch zugänglich sein sollten, wenn die digitale Ressource selbst gar nicht mehr existiert.¹⁷

Die innerhalb einer digitalen Edition erstellten und publizierten Inhalte bzw. Daten sollten ausgetauscht, interpretiert und mit anderen Daten kombiniert werden können – dies wird durch die Verwendung von XML-basierten Datenformaten

13 Vgl. Mark D. Wilkinson et al., *The FAIR Guiding Principles for scientific data management and stewardship*, in: *Scientific Data* 3 (2016), 1–9.

14 Vgl. FAIR-Prinzipien.

15 Vgl. Roman Bleier/Helmut W. Klug, *Persistent Identifier*, in: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«, KONDE Weißbuch, Graz 2021, PID: o:konde.12 (abgerufen 20.8.2022).

16 Vgl. *Dublin Core Metadata Initiative*, URL: <https://dublincore.org> (abgerufen 24.9.2022)

17 Vgl. FAIR-Prinzipien.

für die Primärquelle und die Verwendung von Metadaten, die auf kontrollierten Vokabularen, Ontologien oder Thesauri etc. basieren und ebenfalls den FAIR-Prinzipien folgen, ermöglicht.¹⁸ Für digitale Editionen haben sich mittlerweile einige derartige Datenformate etabliert, wie z. B. das Schema der *Text Encoding Initiative* (TEI) für Textdaten oder die Richtlinien der *Music Encoding Initiative* (MEI) für Musiknotation. Im Bereich der Normdaten spielen z. B. GeoNames, GND, Getty und VIAF eine große Rolle.¹⁹

Damit die Forschungsdaten auch von Dritten weiterverwendet werden können, sollten die Metadaten von digitalen Editionen akkurate und relevante Informationen und Attribute enthalten.²⁰ So sollten im TEI-Dokument einer digitalen Edition auch die Editionsrichtlinien, die angewandt wurden, vorhanden und kodiert sein, und durch die Zurverfügungstellung eines Schemas können die Daten validiert und leichter ausgetauscht werden. Auch sollte mit standardisierten Lizenzmodellen, wie z. B. *Creative Commons*, beschrieben werden, unter welchen Bedingungen die Inhalte zugänglich sind.²¹ *Creative Commons* bezeichnet eine Organisation, die Tools und verschiedene Lizenzen zur Verfügung stellt, um Personen dabei behilflich zu sein, Nutzungsbedingungen für ihre Werke zu formulieren.²²

Technische Infrastrukturen

Bei der Erstellung von digitalen Editionen kommt auch der technischen Infrastruktur eine bedeutende Rolle zu, welche die Befolgung der FAIR-Prinzipien erlaubt und unterstützt. Die entstandenen und modellierten Daten sollten idealerweise in einer Infrastruktur abgelegt werden, welche die digitalen Ressourcen archiviert, verwaltet und eine nachhaltige Nutzung dieser gewährleistet²³; außerdem sollte auch eine weitere Kuratation der Daten nach Ende der Projektlaufzeit erfolgen. Mittlerweile

18 Vgl. Stigler, FAIR-Prinzipien.

19 Vgl. TEI (Text Encoding Initiative), URL: <https://tei-c.org> (abgerufen 24.9.2022); GeoNames, URL: <https://www.geonames.org/> (abgerufen 24.9.2022); Gemeinsame Normdatei, URL: <https://www.dnb.de/gnd> (abgerufen 24.9.2022); OCLC: VIAF, URL: <https://www.oclc.org/de/viaf.html> (abgerufen 24.9.2022); Getty Vocabularies (Getty Research Institute), URL: <https://www.getty.edu/research/tools/vocabularies/> (abgerufen 24.9.2022).

20 Vgl. FAIR-Prinzipien.

21 Vgl. Stigler, FAIR-Prinzipien.

22 Vgl. Creative Commons, URL: <https://creativecommons.org> (abgerufen 20.8.2022); vgl. Helmut W. Klug, Creative Commons, in: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«, KONDE Weißbuch, Graz 2021, PID: o:konde.45 (abgerufen 20.8.2022).

23 Vgl. Stigler, Digitale Nachhaltigkeit.

gibt es einige Open Source-Initiativen, die Software für derartige technische Infrastrukturen entwickeln, wie z.B. FEDORA (*Flexible Extensible Digital Object Repository Architecture*).²⁴

Als Referenzmodell für digitale Archive mit Nachhaltigkeitsstrategie dient das OAIIS-Referenzmodell (*Open Archival Information System*), welches abstrakt die Zusammenarbeit von Menschen und Maschinen beschreibt und sich als Ziel setzt, Ressourcen möglichst nachhaltig zu verwalten.²⁵ Das Modell wurde vom *Consultative Committee for Space Data Systems* entwickelt und sieht als zentrale Bestandteile »Information Packages«, konzeptuelle Container, vor: das SIP (*Submission Information Package*), AIP (*Archival Information Package*) und DIP (*Dissemination Information Package*).

Ein Produzent der Information übergibt das SIP an das System, dort wird es in ein AIP umgewandelt (z.B. Anreicherung mit weiteren Informationen, Hinzufügen eines persistenten Identifikators, eventuell Änderung des Dateiformats, Qualitätskontrolle) und bei der Anfrage durch Benutzer:innen wird daraus ein DIP generiert, welches die gespeicherten Informationen ausgibt.²⁶

Langzeitrepositorien können unterschiedliche Zertifizierungen erwerben, um als *trustworthy digital repository* zu gelten und um nachzuweisen, dass die digitalen Ressourcen nachhaltig langzeitarchiviert werden, wie z.B. im Bereich der Digitalen Geisteswissenschaften häufig das *CoreTrustSeal*.²⁷ Voraussetzung für eine derartige Zertifizierung ist oft die Verwendung des OAIIS-Referenzmodells.

Bestehende Infrastrukturen in Österreich

In Österreich gibt es bereits einige wenige Infrastrukturen, die die digitale Langzeitarchivierung und/oder Dissemination von Daten, die im Digital-Humanities-Kontext entstehen, erlauben. Dazu gehören z.B. das Service ARCHE (*A Resource Centre for the Humanities*), entwickelt am *Austrian Centre for Digital Humanities and Cultural Heritage*, welches Langzeitarchivierung und Dissemination von Daten aus dem gesamten Bereich der Geisteswissenschaften anbietet; DHPlus, eine Plattform zur Langzeitarchivierung von Digital-Humanities-Vorhaben vorzugsweise für Projekte der Universität Salzburg, und das *Geisteswissenschaftliche Asset Management System*

24 Vgl. Fedora, URL: <https://duraspace.org/fedora/> (abgerufen 24.9.2022).

25 Vgl. ISO 14721:2012. Space data and information transfer systems — Open archival information system (OAIIS) — Reference model, URL: <https://www.iso.org/standard/57284.html> (abgerufen 18.9.2022); vgl. Mayer, Langzeitarchivierung.

26 Vgl. Elisabeth Steiner, OAIIS RM, in: Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt »Kompetenznetzwerk Digitale Edition«, KONDE Weißbuch, Graz 2021, PID: o:konde.11 (abgerufen 20.8.2022).

27 Vgl. CoreTrustSeal, URL: <https://www.coretrustseal.org/> (abgerufen am 24.9.2022).

(GAMS), welches am »Institut Zentrum für Informationsmodellierung« entwickelt wurde und im DERLA-Projekt eine wesentliche Rolle spielt, wie in Kapitel zwei näher erläutert wird.²⁸

GAMS basiert auf der Open-Source-Lösung FEDORA, entspricht den Empfehlungen des OAIS-Referenzmodells, orientiert sich an den FAIR-Prinzipien (Vergabe von persistenten Identifikatoren, Anreicherung mit Metadaten, standardisierte Kommunikationsprotokolle können verwendet werden, Angabe von Lizenzmodellen, Verwendung von standardisierten Datenformaten) und ist seit 2019 mit dem *CoreTrustSeal* als vertrauenswürdiges digitales Repositorium zertifiziert.

DERLA als Digitale Edition am GAMS

DERLA und OAIS

Gemäß dem OAIS-Referenzmodell archiviert das GAMS digitale Objekte, bei denen es sich um die »zentrale Archivierungseinheit« handeln soll und die natürliche bzw. reale Dinge wie zum Beispiel Kunstwerke digital repräsentieren. Über diese Objekte ist es einem OAIS-konformen System möglich, standardisierte Verfahren der Digitalen Langzeitarchivierung zur Verfügung zu stellen. Digitale Objekte beinhalten laut OAIS einerseits eine inhaltliche Beschreibung des real repräsentierten Dinges (beispielsweise Bilddateien, eine textuelle Erfassung des Motivs des Kunstwerks oder die Nennung des Künstlers) und andererseits begleitende Metainformation, wie zum Beispiel zur Autorenschaft des Digitalen Objektes selbst.²⁹

DERLA versteht dem fachwissenschaftlichen Verständnis folgend jeden einzelnen Erinnerungsort, als selbständige, durch OAIS-konforme digitale Objekte repräsentierbare archivalische Einheit. Diese als digitale Objekte repräsentierten Erinnerungsorte am GAMS beinhalten folgend alle im Projekt erschlossenen Forschungsdaten (Beschreibung des Ortes, Position), erweiterte Metadaten (wer hat den Ort er-

28 Vgl. Martina Trognitz/Matej Ďurčo, One Schema to Rule them All. The Inner Workings of the Digital Archive ARCHE, in: Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 71 (2018), 217–231; vgl. Johannes Stigler/Elisabeth Steiner, GAMS – Eine Infrastruktur zur Langzeitarchivierung und Publikation geisteswissenschaftlicher Forschungsdaten, in: Mitteilungen der Vereinigung österreichischer Bibliothekarinnen und Bibliothekare 71 (2018), 207–216.

29 Vgl. Stigler, Digitale Nachhaltigkeit; vgl. Steiner, OAIS RM; vgl. Steiner, Elisabeth; Stigler, Johannes, GAMS-Dokumentation, Zentrum für Informationsmodellierung – Universität Graz, Abschnitt »Basic Concepts« <https://gams.uni-graz.at/o:gams.doku> (abgerufen 22.09.2022).

DERLA, FAIR und Open Science

Bei Aufruf eines Erinnerungsortes wie zum Beispiel der »Grabanlage für Karl Spindler« über den Webbrowser (unter <https://gams.uni-graz.at/o:derla.sty21>) wird nicht eine »bloße« Internetseite angesteuert, sondern das GAMS liefert eine Web-Ansicht für das zugrundeliegende digitale Objekt. Diese Weboberfläche wird gemäß den Mechanismen des GAMS dynamisch aus den eigentlichen Forschungsdaten des digitalen Repositoriums erzeugt.³⁴ Daher stehen der Webpräsenz DERLAs dieselben qualitätssichernden Maßnahmen zur Verfügung wie allen anderen digitalen Objekten am GAMS. Folgend ist die digitale Edition DERLA FAIR-konform und bemüht sich darüber hinaus, generelle Gütekriterien moderner Wissenschaften, wie Open Science, einzuhalten.³⁵

Über die Verwendung von sogenannten permanenten Identifikatoren in den Adressen der Digitalen Objekte von DERLA wird eine dauerhafte und offene Verfügbarkeit der Forschungsdaten und Webpräsentation der Digitalen Edition garantiert.³⁶ (**Open Access** bzw. FAIR-accessible) Alle Webseiten DERLAs (aber auch Forschungsdaten) sind über standardisierte, REST-konforme URLs ansprechbar und dauerhaft zitierfähig. Beispielsweise lässt sich der Erinnerungsort (bzw. das Digitale Objekt, das einen Erinnerungsort repräsentiert) mit dem permanenten Identifikator `o:derla.sty21` über die URL <https://gams.uni-graz.at/o:derla.sty21> über einen Webbrowser aufrufen. »Enthaltene« Forschungsdaten lassen sich hierarchisch adressieren. So gelangt man zu den eigentlichen Inhaltsdaten des Erinnerungsortes über https://gams.uni-graz.at/o:derla.sty21/TEI_SOURCE bzw. zu begleitenden Metadaten in Dublin Core über <https://gams.uni-graz.at/o:derla.sty21/DC>. Auf der Weboberfläche DERLAs wird Nutzer:innen der Zugang zu diesen standardisiert adressierbaren Forschungsdaten über textuelle oder symbolische Hyperlinks erleichtert bzw. aktiv angeboten.

Die im Zuge des DERLA Forschungsprojektes erarbeiteten und archivierten OA-IS-konformen digitalen Objekte werden vom GAMS – wie zuvor erwähnt – mit dauerhafter und offener Verfügbarkeit ausgestattet. Jedem archivierten Erinnerungsort stehen dieselben Basisfunktionalitäten zur Verfügung wie jedem anderen digitalen Objekt am GAMS auch. Somit sind auch die URLs DERLAs langzeitstabil und nach den FAIR-Kriterien zitierfähig (**Open Citation**). Nebst der Bereitstellung der eigentlichen Funktionalität verfügt das GAMS über dauerhaftes Personal, welches mit der Wartung und Erhaltung der Infrastruktur betraut ist. Das GAMS garantiert zumindest die Langzeitverfügbarkeit der Metadaten der archivierten digitalen Objekte.

34 Vgl. Steiner/Stigler, GAMS-Dokumentation.

35 Vgl. UNESCO Recommendation on Open Science, URL: <https://www.unesco.org/en/natural-sciences/open-science> (abgerufen 23.09.2022).

36 Vgl. Bleier/Klug, Persistent Identifier.

Open Data bzw. das FAIR-Kriterium der Interoperabilität fordert die Erfassung der Daten eines Forschungsprojektes gemäß internationaler Datenstandards ein. Wie bereits erwähnt, sind die Forschungsdaten DERLAs in XML unter Verwendung verschiedener Standards (wie TEI) über die standardisierten Schnittstellen des GAMS erreichbar. All diese Daten stehen folgend für weitere Nachnutzung, wie zum Beispiel eigene oder erweiterte Forschungsbemühungen, zur Verfügung.

Neben den inhaltlichen Forschungsdaten zu den einzelnen Erinnerungsorten beinhaltet DERLA als digitale Edition auch didaktische Begleittexte (**Open Educational Resources**), die ebenso als eigenständige Forschungsdaten am GAMS archiviert sind. Somit sind über DERLA nicht nur direkte Daten und Metadaten zu den Erinnerungsorten (wie zum Beispiel Koordinaten des Erinnerungszeichens oder zur Autorenschaft des Digitalen Objektes) erreichbar, sondern auch begleitendes didaktisches Material wiederum als eigene digitale Objekte abgelegt. Diese didaktischen Materialien werden von DERLA als gleichrangige Forschungsdaten verstanden. Ein beispielhaftes didaktisches Objekt DERLAs findet sich hier: <https://gams.uni-graz.at/o:derla.fixsty346>, unter https://gams.uni-graz.at/o:derla.fixsty346/TEI_SOURCE lassen sich erneut gemäß der Absprachen des GAMS die zugrundeliegenden Forschungsdaten des Digitalen Objektes ansprechen.

Auf der Weboberfläche DERLAs befindet sich eine Beschreibung der Lizenzbedingungen des Materials der digitalen Edition und es ist somit FAIR-reusable. Dazu gehören neben Datenmaterial auch eigenentwickelte Programme und Projektabläufe. Ebenso beinhalten die erarbeiteten Inhaltsdaten eine Anführung der zugrundeliegenden Lizenz, im Falle von TEI-XML Dokumente im TEI-Header. Aus den einzelnen Forschungsdaten können weiter Verantwortlichkeiten und Rahmenbedingungen von DERLA als Forschungsprojekt ausgelesen werden, wodurch Bedingungen der Weiterverwendung auch erfragt werden könnten. Dies dient dazu, den Forderungen von **Open License** gerecht zu werden.

Open Research schreibt die transparente Beschreibung von Methodologie, Arbeitsschritten und Forschungswerkzeugen eines digitalen Projektes vor. Folgend bietet DERLA auf der Weboberfläche unter »Projekt« eine transparente Darlegung der Projektabläufe dar. Neben den eigentlichen Forschungszielen werden unter anderem auch Projekthintergründe und das Projektteam genannt. Auch werden erarbeitete und verwendete Datenmodelle, kontrollierte Vokabularien und Thesauri mitsamt der dahinterliegenden Projektabläufe dokumentiert. Die Beschreibungen der Datenmodelle sind ebenso im Rahmen der üblichen GAMS-Abläufe dauerhaft zitierfähig und werden innerhalb der inhaltlichen Forschungsdaten verwendet, um auf maschinelle wie auch natürlichsprachliche Erläuterungen der Datenmodelle zu verweisen. Auch Quellcodes zur Transformation bzw. Weiterverarbeitung nötiger XSLT-Stylesheets (wie zum Beispiel zur Generierung der Weboberfläche aus den TEI-XML basierten Forschungsdaten) sind über das GAMS transparent und stabil erreichbar. So lässt sich nachvollziehen, dass dem Erinnerungsort mit dem perma-

nenten Identifikator »o:derla.sty45«, erreichbar unter <https://gams.uni-graz.at/o:derla.sty45>, das XSLT-stylesheet <https://gams.uni-graz.at/archive/objects/o:derla.sty45/datastreams/STYLESHEET/content> für die Erzeugung der Weboberfläche zugewiesen ist.

Open Source fordert die transparente öffentliche Bereitstellung des im Projekt entwickelten Quellcodes ein. Über die Schnittstellen des GAMS lässt sich nicht nur das Programm zur Erzeugung der Weboberfläche einsehen, sondern auch Schnittstellen zur Bereitstellung interner Logik, wie beispielsweise zur Erzeugung geeigneter Datenbankformate oder der zum Digitalen Objekt anliegenden Metadaten aus dem Hauptdatenstrom. So lässt sich zum oben genannten Objekt mit dem PID »o:derla.sty45« auch das Programm zur Erzeugung der Dublin Core Metadaten unter https://gams.uni-graz.at/archive/objects/o:derla.sty45/datastreams/DC_MAPPING/content adressieren, welches dieses Ergebnis produziert: <https://gams.uni-graz.at/archive/objects/o:derla.sty45/datastreams/DC/content>.³⁷

DERLA und Datenanalyse

DERLA bietet als digitale Edition verschiedene Werkzeuge für die Analyse der Forschungsdaten an. Als sichtbarstes Hauptelement kann dabei die »Karte der Erinnerung« genannt werden, die über verschiedene Filtermöglichkeiten eine räumliche, kategoriale und zeitliche Analyse über alle Erinnerungsorte hinweg ermöglichen soll. Die kategoriale Analyse folgt dabei dem auf der Weboberfläche dargelegten Datenmodell. Es können beispielsweise mit Hilfe der Applikation alle Erinnerungsorte in der Steiermark des Typs Gedenktafel angezeigt werden. Eine komplexere Abfrage könnte lauten: *»Zeige alle Erinnerungsorte, deren Erinnerungszeichen eine Straße ist und deren Denkmalsetzung zwischen 1970 und 2000 erfolgte.«*

Die Applikation greift auf die Forschungsdaten abgelegt als digitale Objekte am GAMS zu und füllt damit die auf der Programmiersprache Javascript basierende Kartenanwendung. Somit sichert das GAMS die Aussagekraft der Analyse über den

37 Vgl. What is Open Science? Introduction, FOSTER, <https://www.fosteropenscience.eu/content/what-open-science-introduction> (abgerufen 23.09.2022); vgl. Benedikt Fecher/Sascha Friesike, Open Science: One Term, Five Schools of Thought, in: Könke Bartling/Sascha Friesike (Hg.), *Opening Science, The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*, 2014, 17–49; vgl. Michelle Sidler, *Open Science and the Three Cultures: Expanding Open Science to all Domains of Knowledge Creation*, in: Könke Bartling/Sascha Friesike (Hg.), *Opening Science, The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*, 2014, 81–89; vgl. auch Beschreibung von DERLA und Open Science unter »Projekte/Open Science und DERLA«, Gerald Lamprecht et al., Startseite des DERLA Projektes: <http://gams.uni-graz.at/context:derla>, CJS/ZIM Universität Graz (abgerufen 22.09.2022).

Zugriff auf die eigentlichen, archivierten Forschungsdaten des DERLA Projektes ab. Selbiges trifft auch auf die diversen bereitgestellten Suchen der digitalen Edition zu. Volltextsuche und auch die »Erweiterte Suche« (basierend auf den Opferkategorien DERLAs) verwenden ebenso die langzeitstabilen Daten am GAMS.³⁸

Unter »Orte der Erinnerung/Datenanalyse« findet sich eine Anbindung DERLAs an verschiedene, etablierte Analysewerkzeuge der Digitalen Geisteswissenschaften. »Räumlich-zeitliche Analyse« führt dabei zu einer in die Weboberfläche integrierten Version des Dariah Geo Browser,³⁹ der eine umfangreichere zeitliche Zählung als die »Karte der Erinnerung« erlaubt. Nicht jede Forschungsanwendung unterstützt eine technische Integration »on-the-fly« – das bedeutet, dass die Forschungsdaten durch eine Forscher:in teilautomatisiert übertragen werden müssen. DERLA stellt in solchen Fällen für die gängigsten Werkzeuge der Digitalen Geisteswissenschaften eine zumindest Leitfaden basierte Hilfestellung an.

Unter »*Inhaltlicher Analyse*« steht eine Anbindung an das Forschungswerkzeug Palladio zur Verfügung. Palladio erlaubt nebst anderen Funktionalitäten eine umfangreiche Facettierung der Erinnerungsorte und eine darauf basierende quantitative Auswertung. Beispielsweise können alle Erinnerungsorte gemäß der Art des Erinnerungszeichens gezählt und sortiert werden.⁴⁰

»*Quantitative Analyse*« hält eine Integration von rawgraphs.io⁴¹ bereit und erlaubt die Erzeugung von diversen Graphen zur Veranschaulichung quantitativer Zusammenhänge. Beispielsweise kann eine Treemap zur Darstellung der Relation der Anzahl der Erinnerungsorte pro Bundesland oder pro Kategorie erzeugt werden.

38 Vgl. Gerald Lamprecht et al., Karte der Erinnerung des DERLA Projektes: <https://gams.uni-graz.at/context:derlageo/sdef:Context/get?mode=map,CJS/ZIM> Universität Graz (abgerufen 22.09.2022).

39 Dariah Geo-browser, DARIAH-DE, URL: <https://geobrowser.de.dariah.eu/> (abgerufen 23.09.2022)

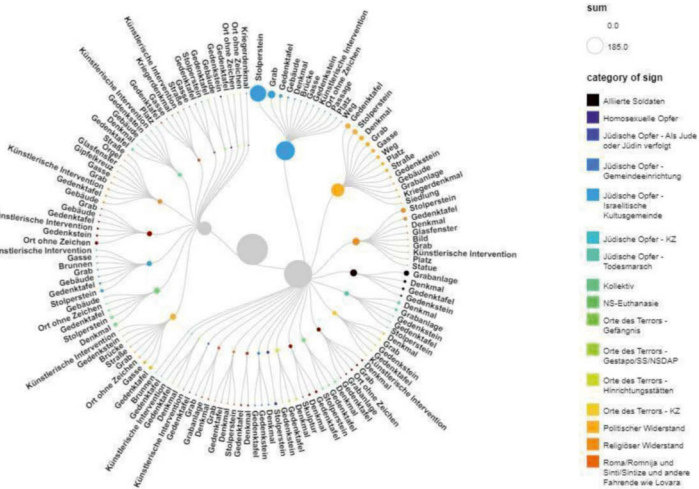
40 Startseite des Palladio-Tools, <https://hdlab.stanford.edu/palladio/> (abgerufen 29.09.2022)

41 Rawgraphs Startseite, DensityDesign Research Lab/Calibro/Inmagik, <https://www.rawgraphs.io/> (abgerufen 23.09.2022).

Abb. 1: Treemap – erzeugt via rawgraphs.io. Relation der vergebenen Opferkategorien pro Erinnerungszeichen über alle Bundesländer hinweg. Ein großer Teil der Erinnerungsorte gedenkt an jüdische Opfer oder an Widerstand. (Quelle: Eigenproduktion via rawgraphs.io).



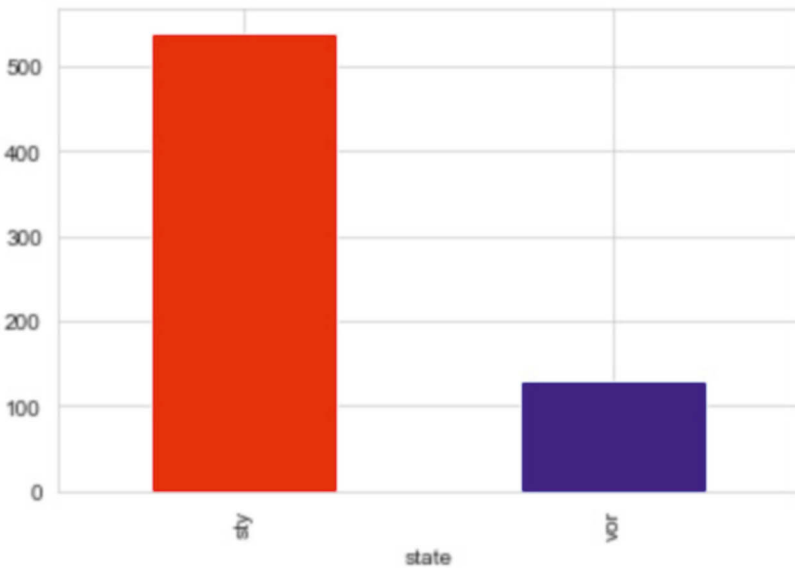
Abb. 2: Dendrogramm – Relation der Erinnerungsorte pro Bundesland, Kategorie der Opfergruppe («Alliierte Soldaten«, «Jüdische Opfer«), und Art des Erinnerungszeichens («Gasse«, «Stolperstein« usw.). Ein großer Teil der steirischen Erinnerungsorte sind Stolpersteine in Gedenken an jüdische Opfer. (Quelle: Eigenproduktion via rawgraphs.io).



Neben den direkt an der Weboberfläche integrierten Anwendungen, besteht aufgrund der offenen Verfügbarkeit der Forschungsdaten am GAMS auch die Möglichkeit der spezifischen Analyse über gängige Techniken der Data Science, wie zum Beispiel via Python und begleitenden Werkzeugen. Folgend soll eine beispielhafte Auswertung demonstriert werden.

Es soll die Verteilung der Erinnerungsorte im Bezug auf Kategorie der zugewiesenen Opfergruppen («Politischer Widerstand«, «Alliierte Soldaten« usw.), Art des Erinnerungszeichens («Gasse«, «Stolperstein«), der Bundesländer («Steiermark«, «Vorarlberg«) und Datum der Denkmalsetzung untersucht werden.⁴² Es wird die Behauptung aufgestellt, dass die Opfer-Kategorien und die Art der Zeichen gleichmäßig auf die Erinnerungsorte und Bundesländer verteilt seien. Über ein eigenes Python-Programm werden die Forschungsdaten von DERLA für die Steiermark und Vorarlberg abgegriffen und für die folgende quantitative Analyse umgewandelt.

Abb. 3: Anzahl der Erinnerungsorte pro Bundesland (Quelle: Eigenproduktion via python pandas, seaborn, matplotlib).

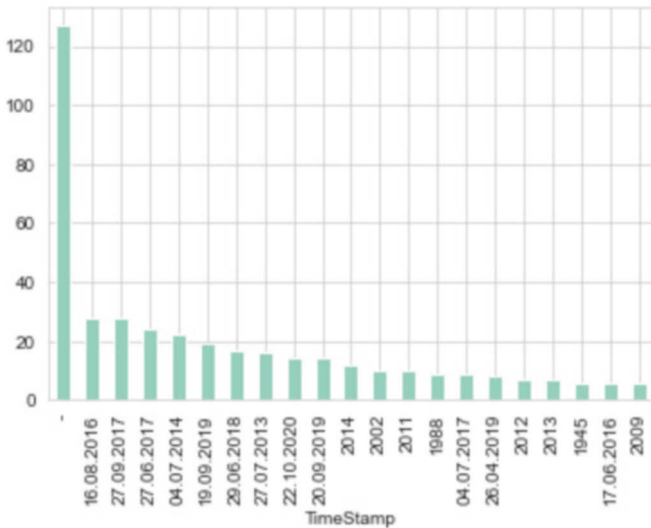


42 Eine detaillierte Beschreibung des Datemodells DERLAs findet sich unter: Gerald Lamprecht et al.; CJS/ZIM Universität Graz, https://gams.uni-graz.at/archive/objects/o:derla.ontology/methods/sdef:Ontology/get?locale=de&mode=ont_overview (abgerufen 29.09.2022)

Wie zu erwarten ergibt die Zählung der Erinnerungsorte eine höhere Anzahl für das Bundesland Steiermark.

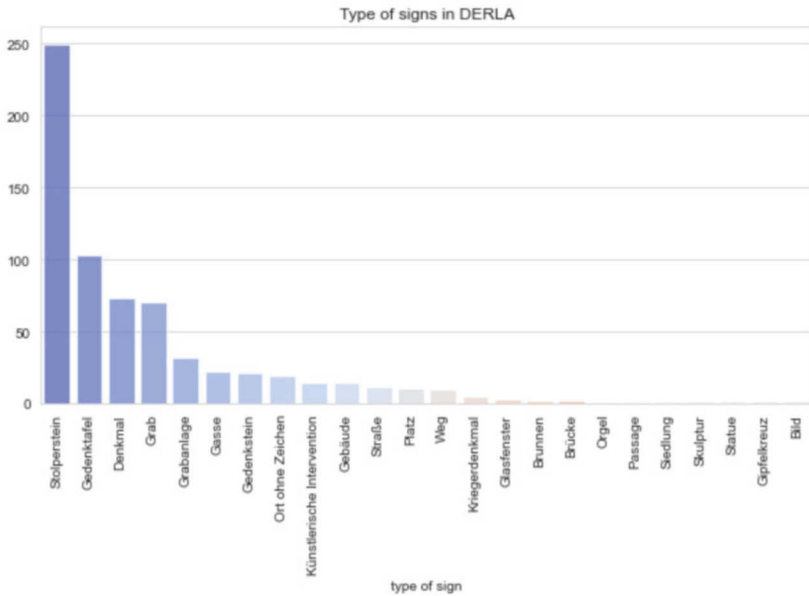
Die Gruppierung nach Datum der Denkmalsetzung zeigt eine gewisse Häufung von einigen Werten an. So haben über zwanzig Erinnerungszeichen als Datum der Denkmalsetzung den 16.08.2016 oder den 27.09.2017.

Abb. 4: Gruppierung nach Datum der Denkmalsetzung. (Quelle: Eigenproduktion via python pandas, seaborn, matplotlib).



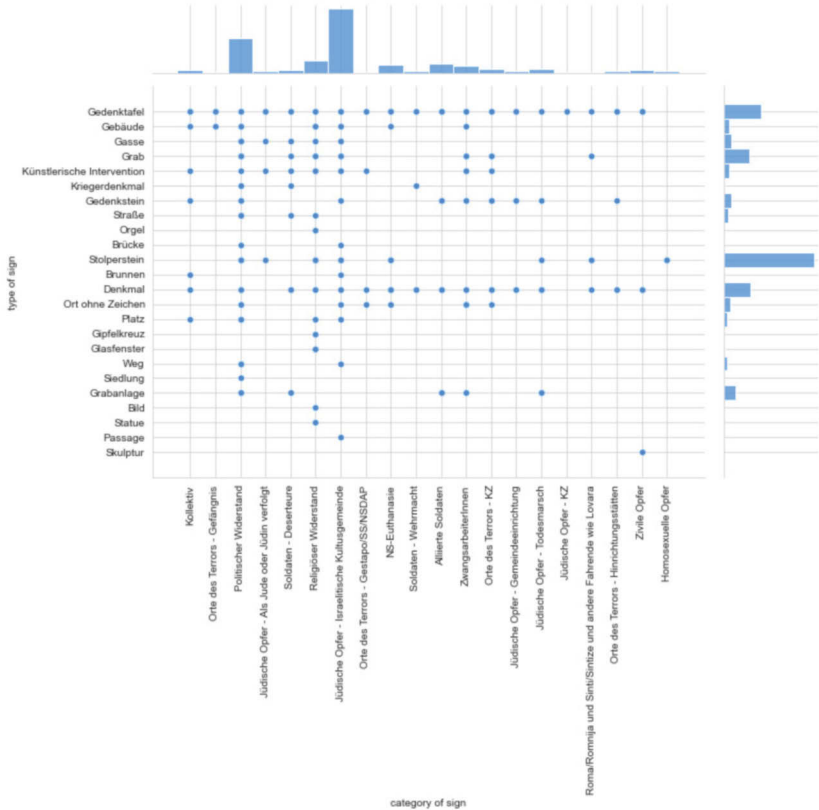
Die Verteilung der Art des Erinnerungszeichens zeigt eine deutliche Häufung von Stolpersteinen. Die Mehrheit der Erinnerungszeichen wird von nur wenigen Werten bestimmt.

Abb. 5: Art des Erinnerungszeichens der Erinnerungsorte (Quelle: Eigenproduktion via python pandas, seaborn, matplotlib).



Auf einem sogenannten »joint-plot« kann der Zusammenhang zwischen den vergebenen primären Kategorien der Opfergruppen und der zugewiesenen Art des Erinnerungszeichens ausgelesen werden. Die Balkendiagramme an den Rändern zeigen die (relative) Gesamtzahl der Erinnerungsorte mit vergebenem kategorischem Wert an, wie zum Beispiel Anzahl der Orte mit der Zuweisung »Stolperstein«.

Abb. 6: Jointplot – Zusammenhang zwischen Kategorien der gedachten primären Opfergruppe und der Art des Erinnerungszeichens. (Quelle: Eigenproduktion via python pandas, seaborn, matplotlib).



Der Graph zeigt eine größere Ungleichverteilung in Bezug auf die Art des Zeichens und der »primären« Zuweisung der Opfer-Kategorien an.⁴³ Man kann beispielsweise gut erkennen, dass, obwohl es sich bei vielen Erinnerungsorten um Stolpersteine handelt, zu diesen Orten nur eine moderate Anzahl an unterschiedlichen primären Opfer-Kategorien zugewiesen wurden.

43 Im Zuge der beispielhaften Auswertung wurden Mehrfachzuweisungen der Opfer-Kategorien pro Erinnerungsort ignoriert und nur die erste = primäre Zuweisung berücksichtigt.

Conclusio

Sämtliche Daten aus DERLA werden als digitale Objekte im GAMS angelegt und erfüllen somit wesentliche Prinzipien einer digitalen Nachhaltigkeitsstrategie. Diese umfasst die Nutzung von standardisierten Datenformaten (TEI-XML, GML-XML), die Anreicherung der Daten mit umfangreichen Metadaten, die Archivierung unter einem *Persistent Identifier* (wodurch sämtliche Webseiten und Forschungsdaten dauerhaft zitierfähig sind) und die Bereitstellung unterschiedlichster Disseminationsmethoden. Die so aufbereiteten und archivierten Daten sichern damit die Aussagekraft von folgenden Auswertungen (hinsichtlich Qualität der verwendeten Forschungsdaten) in unterschiedlichen Analyseszenarien ab. Einerseits besteht die Möglichkeit, die direkt mit der Digitalen Edition ausgelieferten Analysewerkzeuge (wie zum Beispiel die »Karte der Erinnerung«) zu nutzen, andererseits stellen die OAIS-konformen Schnittstellen des GAMS das nachhaltige Datenmaterial DERLAS auch für weiterführende eigene Untersuchungen zur Verfügung.

