



**ANTIFASCHISTISCHES
SCHUTZ-INSTITUT**
für reflexive Gesellschaftsethik

KI-LAB-26

Nach der Architektur

&

Vom Archiv zum interaktiven Lernraum



Nach der Architektur

Wie Ki-Lab-26 mit Codex, GitHub und Zenodo zu einer freien Publikations- und Archivierungsinfrastruktur ausgebaut wurde

Abstract

Der vorliegende Beitrag schließt an das bereits veröffentlichte Projekt **KI-LAB-26** an, in dem ein publizistischer Grundsatzartikel zur Verflachung digitaler Archive mit einem methodischen Strukturpapier zur mehrschichtigen Ordnung heterogener Wissensbestände verbunden wurde. Im Zentrum der folgenden Ausarbeitung steht nicht mehr primär die innere Logik der Archivarchitektur selbst, sondern der Schritt von der Theorie zur öffentlichen, freien und wissenschaftlich belastbaren Infrastruktur. Dargestellt wird, warum das Projekt im Anschluss an das Methodenpapier mit Hilfe von Codex operativ aufbereitet, in ein versioniertes GitHub-Repository überführt und über Zenodo DOI-fähig archiviert wurde; welche begrifflichen und institutionellen Differenzen zwischen Website, Repository und Archivschicht bestehen; warum die Verfügbarkeit eines DOI für freie wissenschaftliche Arbeit von erheblicher Bedeutung ist; und wodurch sich Ki-Lab-26 seit diesem Ausbausschritt nicht mehr nur als Text, sondern als dokumentierte, versionierte und zitierfähige Arbeitsumgebung bestimmt. Der Aufsatz richtet sich ausdrücklich auch an Leserinnen und Leser, die weder den vorangegangenen Arbeitsprozess noch die technischen Einzelbegriffe bereits kennen.

1. Einleitung: Vom Methodenentwurf zur öffentlichen Infrastruktur

Mit dem bisherigen Stand von Ki-Lab-26 war ein entscheidender begrifflicher Schritt bereits vollzogen. Das Projekt hatte dargelegt, warum große digitale Archive unter Bedingungen begrenzter Kontextfenster nicht durch bloße Speicherung oder flache Suchbarkeit erschließbar werden, sondern eine mehrschichtige Ordnung benötigen: Roharchiv, Inventar, Segment, Dossier, Oberkarte, Root-Map, ergänzt durch Vektorräume, Relationen und kontrollierte Provenienz. Es war also gezeigt worden, was eine solche Architektur **in sich** leisten soll. Weniger deutlich war jedoch noch, welche **äußeren Publikations- und Archivierungsformen** notwendig sind, damit eine derartige Architektur nicht als isolierter PDF-Text stehenbleibt, sondern selbst jene institutionelle Form erhält, die ihrem Anspruch entspricht.

Genau hier setzt der vorliegende Anschlussartikel an. Er behandelt nicht mehr bloß die innere Struktur des Projekts, sondern dessen zweite, ebenso wichtige Frage: Wie wird aus einem Methodenpapier eine freie, wissenschaftlich anschlussfähige, versionierbare und zitierfähige Arbeitsinfrastruktur? Anders formuliert: Wie wird aus einer strukturellen Theorie digitaler Souveränität eine reale Publikationsform, die diese Souveränität auch praktisch trägt?

Die Antwort darauf besteht nicht in einem einzelnen Werkzeug. Sie besteht in einer **gestuften Publikationsarchitektur**. Für Ki-Lab-26 wurde diese Architektur im Anschluss an das Methodenpapier in drei Funktionsräume ausdifferenziert: erstens die institutionell verantwortete Website als Primärbühne, zweitens das versionierte öffentliche Repository als Arbeits- und Publikationscluster, drittens die externe Archivschicht mit DOI zur dauerhaften Referenzierbarkeit.

Dass hierfür Codex, GitHub und Zenodo verwendet wurden, ist nicht bloß technische Bequemlichkeit. Es ist Ausdruck einer bewussten Entscheidung, Text, Struktur, Versionsgeschichte und Zitation nicht voneinander zu trennen.

2. Ausgangslage: Warum das Methodenpapier allein nicht genügte

Das Methodenpapier Ki-Lab-26-Archive-Strukture-v1.1 formulierte eine anspruchsvolle Diagnose: Das eigentliche Problem digitaler Wissensarbeit liegt nicht nur in der Menge des Materials, sondern in der Gefahr des **Flattenings**, also jener Verflachung, bei der Herkunft, Zwischenstufen, Unsicherheit, Pfadabhängigkeit und strukturelle Mehrdeutigkeit in glatte und scheinbar neutrale Objekte überführt werden. Diese Diagnose hatte unmittelbare Konsequenzen für die Behandlung von Archiven, aber sie hatte ebenso Konsequenzen für die Form ihrer Veröffentlichung.

Denn ein Text, der Provenienz, Mehrschichtigkeit und Rückverfolgbarkeit verteidigt, gerät in einen performativen Widerspruch, wenn er selbst nur als unversionierte Einzelfassung zirkuliert. Eine PDF-Datei auf einer Website kann publizistisch sinnvoll sein; sie genügt aber nicht, wenn es darum geht, die eigene Entwicklungsgeschichte, unterschiedliche Dateitypen, ergänzende methodische Umfeldtexte, Lizenzlage, Zitationslogik und spätere Revisionsschritte transparent zu halten. Eine bloße Plattformveröffentlichung wäre hier also selbst zu schwach gewesen.

Hinzu trat ein zweiter Gesichtspunkt. Ki-Lab-26 sollte gerade nicht als geschlossener, statischer Text verstanden werden, sondern als strukturierter Arbeitszusammenhang. Der Haupttext ist zwar zentral, doch er steht in Beziehung zu methodischen Grundlagentexten wie der **Systemischen Integrationsarchäologie** und der **Algorithmischen Stratigraphie**, ebenso wie zu operativen KI-Methodiken wie **Prompthygiene**, **Master-Communication-Key** und **KI-Triangulierung**. Solange diese Beziehung nur implizit bleibt, ist das Projekt publizistisch sichtbar, aber strukturell noch unterschlossen. Es bedurfte daher einer Form, die Haupttext, Grundlagen, operative Methodik, Berichte, Metadaten und spätere Revisionsschritte als **zusammengehörigen, aber differenzierten Cluster** sichtbar macht.

3. Der operative Ausbau: Was mit Codex tatsächlich getan wurde

3.1 Codex nicht als Autorersatz, sondern als Strukturierungswerkzeug

Die Rolle von Codex bestand nicht darin, die inhaltliche Autorenschaft zu übernehmen. Seine Funktion lag vielmehr darin, eine gegebene Werklandschaft in eine saubere, nicht-destruktive und reproduzierbare Repository-Struktur zu überführen. Diese Unterscheidung ist wichtig. Wer Codex bloß als Textgenerator missversteht, unterschätzt seine eigentliche Stärke in diesem Zusammenhang: die präzise Ausführung wiederholbarer, strukturierender Arbeitsschritte bei gleichzeitiger Erhaltung von Quellen, Dateikontexten und Zielordnungen.

Der erste Schritt bestand in der Erzeugung **texttreuer Markdown-Fassungen** aus den vorhandenen PDF-Dateien. Dieser Schritt war nicht bloß eine Konvertierung. Er hatte eine doppelte Funktion. Erstens wurde dadurch eine offene, diffbare, versionsgeeignete Arbeitsfassung der Texte geschaffen. Zweitens konnte das Material sofort in eine methodisch sinnvolle Grundordnung überführt werden. Der Projektbestand wurde dabei in drei Hauptbereiche gegliedert: **Haupttext**, **Grundlagen** und **Operative Methodik**. Damit erhielt Ki-Lab-26 nicht nur eine Dateiablage, sondern eine erste publikationslogische Ordnung.

3.2 Klassifikation, Sidecars und Berichte

Zu diesen Markdown-Fassungen traten **Metadateien** und **Berichte**. Für jedes abgeleitete Dokument wurden begleitende Metainformationen erzeugt; dazu kamen Inventare und Reports, die den Zustand der Extraktion, die Klassifikation und eventuelle Unsicherheiten dokumentierten. Gerade darin zeigt sich die innere Kohärenz des Projekts: Die Architekturlogik des Methodenpapiers wurde auf die eigene Publikationsaufbereitung angewandt. Nicht nur das, was publiziert wird, sondern auch die **Art der Aufbereitung** erhielt selbst eine dokumentierte Form.

Anschließend erfolgte eine **Dateinamen-Normalisierung**. Dieser Schritt mag nebensächlich erscheinen, ist aber für professionelle öffentliche Arbeit unverzichtbar. Umlaute, Leerzeichen, Tippfehler, gemischte Groß- und Kleinschreibung sowie inkonsistente Pfadlogiken sind nicht bloß ästhetische Störungen. Sie erschweren Referenzierbarkeit, maschinische Verarbeitung, stabile Links und die saubere spätere Versionierung. Die Normalisierung war daher ein wichtiger Schritt, um aus einem ad hoc gewachsenen Dateibestand einen langfristig belastbaren öffentlichen Bestand zu machen.

3.3 README, LICENSE und CITATION

Im nächsten Schritt wurde das Repository nicht nur technisch, sondern **editorisch und rechtlich** geklärt. Ein `README.md` erläutert nun Struktur, Zweck und innere Ordnung des Repositories. Eine `LICENSE` fasst die repo-weite Lizenzierung der eigenen Text- und Dokumentationsmaterialien unter **CC BY-NC-SA 4.0**. Eine `CITATION.cff` wiederum legt fest, wie das Repository selbst und wie der Haupttext bevorzugt zitiert werden sollen.

Gerade diese drei Elemente markieren den Übergang von bloßem Dateibestand zu wissenschaftsnaher Infrastruktur. Ohne `README` bleibt ein Repository für Außenstehende strukturell stumm. Ohne Lizenz bleibt die Weiterverwendung unklar. Ohne Zitationsdatei bleibt offen, wie ein solcher Werkzusammenhang bibliographisch sauber anzusprechen ist. Mit diesen Dateien wird das Projekt also in einen Zustand überführt, in dem nicht nur die Inhalte, sondern auch ihre **öffentliche Benutzbarkeit** geregelt sind.

3.4 Git-Initialisierung, Commit und öffentlicher Push

Erst danach erfolgte die eigentliche **Git-Initialisierung**. Das ist mehr als ein technischer Nebenschritt. Mit dem ersten Commit wurde der bis dahin lokal aufgebaute Bestand in einen expliziten Versionszustand überführt. Von da an war die Projektgeschichte nicht mehr nur stillschweigend im Dateiordner enthalten, sondern als nachvollziehbare Änderungsgeschichte strukturiert. Dass vor dem ersten öffentlichen Push noch einmal lokale Sicherungsreste und unerwünschte Zusatzordner entfernt wurden, gehört ebenfalls zu dieser Sorgfalt. Ein öffentliches Forschungsrepo ist nicht einfach ein Spiegel der lokalen Werkstatt, sondern eine redaktionell verantwortete Auswahl dessen, was tatsächlich in den wissenschaftlich-publicistischen Raum gehört.

Mit dem anschließenden Push auf GitHub erhielt Ki-Lab-26 seine erste öffentliche Repository-Gestalt. Dieser Schritt ist deshalb bedeutsam, weil GitHub nicht nur Speicherort, sondern auch eine Form der **Versionstransparenz** ist. Leserinnen und Leser können dort nicht nur die aktuelle Fassung sehen, sondern im Prinzip auch verstehen, dass es sich um einen sich entwickelnden, dokumentierten Zusammenhang handelt.

4. GitHub als zweiter Raum: Warum das Repository mehr ist als eine Ablage

Das GitHub-Repository erfüllt innerhalb der Gesamtarchitektur eine Funktion, die weder die Website noch Zenodo allein übernehmen könnten. Es ist der Raum, in dem **Arbeitsstruktur, Veröffentlichungslogik und Versionierbarkeit** zusammenfallen.

Die Website des AS-Instituts bleibt der primäre, institutionell gerahmte Publikationsort. Dort erscheint ein Text in einem eindeutigen publizistischen Kontext, versehen mit Einleitung, Einbettung, ästhetischer Rahmung und klarer Verantwortungszuweisung. GitHub dagegen übernimmt eine andere Aufgabe. Es macht das Projekt nicht nur lesbar, sondern **strukturell sichtbar**. Haupttext, Grundlagen, operative Methodik, Berichte, Lizenz, Zitationslogik und Arbeitsquellen liegen dort nicht bloß nebeneinander, sondern in einer nachvollziehbaren Repo-Struktur. Dadurch wird die Publikation in eine offene, arbeitbare Umgebung überführt.

Gerade für freie wissenschaftliche Arbeit ist dieser Unterschied wesentlich. Ein PDF auf einer Website ist zitierbar, aber nur begrenzt revisierbar. Ein GitHub-Repository kann dagegen eine Werklandschaft versioniert, dokumentiert und später differenziert fortschreiben, ohne dass jede kleine Änderung sofort das gesamte Publikationssystem destabilisiert. Für Ki-Lab-26 bedeutet das: Das Projekt ist nicht länger nur eine Schrift über digitale Souveränität, sondern selbst eine **praktizierte Form digitaler Souveränität**.

5. Zenodo als dritte Schicht: Warum Archivierung und DOI mehr leisten als bloße Sichtbarkeit

5.1 Der Unterschied zwischen Erreichbarkeit und Referenzierbarkeit

Mit GitHub war das Projekt öffentlich. Mit Zenodo wurde es darüber hinaus **archiviert und DOI-fähig**. Dieser Unterschied ist für wissenschaftliche Arbeit fundamental. Eine GitHub-URL macht einen Bestand zugänglich. Ein DOI macht ihn **dauerhaft referenzierbar**.

Das ist keine bloß bibliographische Formalie. URLs können sich ändern, Repositories können umbenannt, verschoben oder gelöscht werden. Ein DOI hingegen fungiert als persistenter Identifikator, der auf Dauer gerade dazu dienen soll, eine bestimmte Ressource eindeutig und dauerhaft ansprechbar zu halten. In einer Wissenslandschaft, die von schnellen Plattformwechseln und flüchtigen Linkstrukturen geprägt ist, ist das von erheblicher Bedeutung. Die Verfügbarkeit eines DOI ist daher nicht nur ein Komfortmerkmal, sondern ein Schritt aus der bloßen Netzsichtbarkeit in die Sphäre **belastbarer wissenschaftlicher Referenz**.

5.2 Warum ein Folge-Release nötig wurde

Im konkreten Fall musste hierfür eine kleine, aber instruktive Operation erfolgen. Das Repository war auf GitHub bereits veröffentlicht und mit einem ersten Release versehen, bevor Zenodo aktiviert wurde. Zenodo archiviert jedoch zuverlässig jene Releases, die nach aktivierter Integration erscheinen. Deshalb wurde ein kleiner Folge-Release erzeugt, der nicht eine neue inhaltliche Hauptversion markierte, sondern die **Archivierungssynchronisation** auslöste. Gerade diese Episode ist aufschlussreich, weil sie zeigt, dass Publikationsinfrastruktur nicht aus bloßer Theorie besteht. Sie verlangt genaue Kenntnis von Zuständigkeiten, Reihenfolgen und Systemgrenzen.

5.3 Was jetzt vorliegt: Concept DOI und Version DOI

Im Ergebnis liegen nun zwei DOI-Ebenen vor. Zum einen ein **versionsübergreifender DOI** für das Repository als fortlaufenden Gesamtzusammenhang, zum anderen ein **versionsspezifischer DOI** für den aktuell archivierten Release. Diese Unterscheidung ist zentral. Wer auf das Repository als strukturierten Publikationscluster insgesamt verweist, sollte den versionsübergreifenden DOI verwenden. Wer auf einen exakt archivierten Stand Bezug nimmt, verwendet den versionsspezifischen DOI des Releases.

Zugleich bleibt eine wichtige begriffliche Präzisierung bestehen: Der derzeitige Zenodo-Datensatz ist ein **repository- und releasezentrierter Datensatz**, kein gesonderter, papierzentrierter Einzeldatensatz für das Methodenpapier selbst. Genau deshalb wurde in der `CITATION.cff` des Repositories darauf verzichtet, den Release-DOI fälschlich als Paper-DOI auszugeben. Die bevorzugte Zitation des Haupttextes bleibt papierzentriert; der DOI dagegen archiviert derzeit die Repository-Schicht. Diese Trennung ist kein Mangel, sondern Ausdruck begrifflicher Disziplin.

6. Warum die DOI-Verfügbarkeit für freie wissenschaftliche Arbeit wichtig ist

Die Bedeutung eines DOI erschöpft sich nicht in der Zitierform. Für freie, institutsunabhängige oder nicht unmittelbar universitär eingebundene Forschung ist ein DOI in mehrfacher Hinsicht relevant.

Erstens schafft er **Seriosität ohne institutionelle Unterordnung**. Forschung, die außerhalb klassischer Verlags- und Hochschulinfrastrukturen entsteht, leidet häufig nicht an methodischem Mangel, sondern an der Schwierigkeit, ihre Ergebnisse in eine Form zu bringen, die von Dritten dauerhaft referenziert und ernst genommen werden kann. Ein DOI ersetzt keine wissenschaftliche Qualität. Er macht Qualität aber **auffindbar, stabil ansprechbar und bibliographisch anschlussfähig**.

Zweitens schafft er **zeitliche Stabilität**. Gerade weil Ki-Lab-26 ausdrücklich als versioniertes Projekt gedacht ist, bedarf es einer Instanz, die bestimmte Zustände fixiert, ohne die Entwicklungsfähigkeit zu unterbinden. Genau das leistet die Kombination aus GitHub und Zenodo: GitHub erlaubt Entwicklung, Zenodo fixiert veröffentlichte Zustände.

Drittens schafft er **freie Sichtbarkeit jenseits proprietärer Plattformlogiken**. Ein PDF in einem Messenger-Kanal oder auf einer Plattform ohne stabile Archivierung ist zwar publik, aber prekär. Ein DOI-gebundener Datensatz hingegen bleibt Teil einer wissenschaftsnahen, offenen Referenzstruktur. Für das AS-Institut bedeutet das, dass freie wissenschaftliche Arbeit nicht länger auf die Binnenlogik kurzfristiger Plattformöffentlichkeiten reduziert bleibt, sondern in eine Form übergeht, die sowohl offen als auch strukturell belastbar ist.

7. Was Ki-Lab-26 jetzt ausmacht

Mit diesen Schritten hat sich Ki-Lab-26 in seiner Gestalt verändert. Es ist jetzt nicht mehr nur ein Methodenpapier, das eine bestimmte Architektur beschreibt. Es ist auch nicht nur ein Projektcluster aus verwandten Texten. Es ist vielmehr zu einer **publikationsfähigen, versionierten und archivierten Arbeitsinfrastruktur** geworden.

Diese Infrastruktur besitzt mindestens fünf klar unterscheidbare Eigenschaften.

Erstens ist sie **provenienzbewusst**. Die Herkunft der Dateien, ihre Bearbeitungsschritte, ihre Klassifikation und ihre Versionierung werden nicht verdeckt, sondern dokumentiert.

Zweitens ist sie **mehrschichtig**. Website, Repository und DOI-Archivierung erfüllen unterschiedliche Funktionen und sind gerade deshalb zusammen stärker als jede Ebene für sich.

Drittens ist sie **lizenzrechtlich geklärt**. Durch die explizite Wahl von CC BY-NC-SA 4.0 wird die freie nicht-kommerzielle Weitergabe ermöglicht, ohne die Textarbeit in begriffslose Beliebigkeit zu entlassen.

Viertens ist sie **zitationsfähig**. Mit `CITATION.cff`, Release-Struktur und DOI ist festgelegt, wie auf das Repository und wie auf den Haupttext verwiesen werden soll.

Fünftens ist sie **weiterentwicklungsfähig**. Die jetzige Gestalt ist kein Endpunkt, sondern eine methodisch kontrollierte Schwelle. Neue Versionen, zusätzliche Texte, spätere separate Paper-Datensätze oder weiter ausgearbeitete technische Komponenten können anschließen, ohne dass die Grundlogik des Projekts verloren ginge.

8. Bedeutung für das AS-Institut: Freie Wissenschaft als institutionell disziplinierte Offenheit

Für das AS-Institut liegt die Tragweite dieser Schritte nicht nur in einem einzelnen Projekt. Sie markieren vielmehr eine Form, in der freie wissenschaftliche Arbeit institutionell diszipliniert werden kann, ohne in klassische Abhängigkeit zurückzufallen.

Das AS-Institut ist weder Großverlag noch Universität im traditionellen Sinn. Gerade deshalb ist die Frage, wie Texte, Methoden und Langzeitprojekte öffentlich werden, keine Nebenfrage, sondern eine Strukturfrage. Ohne eigene Publikations- und Archivierungsarchitektur droht freie Wissenschaft entweder in informellen Halböffentlichkeiten zu versickern oder sich doch wieder an fremde Infrastrukturen binden zu müssen, deren Regeln, Filter und Prioritäten von außen gesetzt werden.

Mit der Aufstellung von Ki-Lab-26 als Website-Text, GitHub-Cluster und Zenodo-Archiv entsteht ein anderes Modell. Wissenschaftliche Arbeit bleibt hier offen, aber sie bleibt nicht formlos. Sie bleibt zugänglich, aber nicht strukturlos. Sie bleibt frei, aber nicht ungebunden. Genau dies ist für freie Institute von großer Bedeutung: Nicht die improvisierte Sichtbarkeit um jeden Preis, sondern eine **kontrollierte öffentliche Nachvollziehbarkeit**, die der Sache selbst gerecht wird.

In diesem Sinn ist Ki-Lab-26 nun auch ein Demonstrationsfall dafür, wie kleinere oder budgetär begrenzte Akteure eine ernstzunehmende, anschlussfähige und dauerhaft referenzierbare Forschungsumgebung aufbauen können, ohne auf die Infrastruktur großer Plattformmonopole vollständig angewiesen zu sein. Die technische Schwelle hierfür ist real, aber sie ist nicht unüberwindbar. Entscheidend ist weniger rohe Kapitalmacht als eine methodisch reflektierte Kombination aus Strukturdisziplin, offener Infrastruktur und publizistischer Sorgfalt.

9. Grenzen, Präzisierungen und nächste Schritte

Der erreichte Zustand ist belastbar, aber nicht abschließend. Drei Präzisierungen sind wichtig.

Erstens ist der derzeitige DOI noch **repository-zentriert**, nicht papierzentriert. Für die Zukunft kann es sinnvoll sein, zusätzlich einen eigenen, manuellen Zenodo-Datensatz nur für das Hauptpapier anzulegen. Damit ließe sich die bibliographische Lage weiter schärfen.

Zweitens bleibt die **Metadatenpflege** eine reale Aufgabe. Titel, Beschreibung, Keywords, Zitationshinweise und institutionelle Rahmung müssen nicht nur einmal gesetzt, sondern über weitere Versionen hinweg kohärent gehalten werden.

Drittens ist die eigentliche inhaltliche Arbeit mit dem Projekt keineswegs abgeschlossen. Die nun geschaffene Infrastruktur erzeugt vielmehr die Voraussetzung dafür, dass v1.2 und folgende Stufen nicht wieder in informelle Dateigeschichte zurückfallen, sondern auf einem bereits versionierten, archivierten und öffentlich lesbaren Fundament aufsetzen können.

Gerade deshalb ist es wichtig, den erreichten Zustand weder zu unterschätzen noch zu mystifizieren. Es handelt sich nicht um die Vollendung des Projekts, wohl aber um dessen erste **institutionell belastbare Außenform**.

10. Schluss: Von der Theorie der Souveränität zu ihrer Praxis

Im ersten großen Aufsatz zu Ki-Lab-26 wurde gezeigt, dass geistige Souveränität im digitalen Raum nur dort möglich bleibt, wo Archive nicht flach, sondern geschichtet, provenancebewusst und reflexiv organisiert werden. Der nun vollzogene zweite Schritt hat dieser These eine praktische Entsprechung gegeben. Denn dieselben Prinzipien, die im Methodenpapier für die Ordnung von Wissensbeständen formuliert wurden, sind nun auf die eigene Publikationsweise angewandt worden.

Das Projekt hat damit eine wichtige Schwelle überschritten. Es ist nicht länger nur ein Text **über** digitale Souveränität, sondern selbst eine kleine, operative Form digitaler Souveränität. Codex diente der strukturellen Aufbereitung, GitHub der offenen, versionierten Sichtbarkeit, Zenodo der dauerhaften archivischen Referenz. Zusammen ergeben diese Ebenen eine Infrastruktur, in der freie wissenschaftliche Arbeit nicht nur behauptet, sondern **praktisch vollzogen** wird.

Wenn man die neue Lage von Ki-Lab-26 in einer knappen Formel fassen wollte, könnte man sagen: Aus einem Methodenpapier ist ein frei zugänglicher, versionierter und dauerhaft referenzierbarer Publikationszusammenhang geworden. Gerade darin liegt sein gegenwärtiger Wert. Denn die Frage, ob Wissen bloß erzeugt oder auch in einer tragfähigen Form veröffentlicht, geordnet, weiterentwickelt und dauerhaft ansprechbar gemacht wird, entscheidet zunehmend darüber, ob freie Forschung bloß sichtbar oder tatsächlich wirksam sein kann.

Dieser Unterschied ist nicht administrativ, sondern kulturell. Und genau deshalb ist der Schritt von der Architektur zur Infrastruktur nicht bloß ein technischer Nachtrag, sondern ein integraler Teil des Projekts selbst.

Vom Archiv zum interaktiven Lernraum

Produktarchitektur, Stand der Technik und aktueller Demonstratorstatus eines Archive-to-Learning-Compilers

Abstract

Der vorliegende Beitrag beschreibt eine Produktlinie, die nicht bei der bloßen Digitalisierung von Dokumenten stehen bleibt, sondern heterogene Wissensbestände in interaktive, statische und epistemisch kontrollierbare Lernräume überführt. Ausgangspunkt ist die Beobachtung, dass große persönliche oder institutionelle Archive zwar textuell reich, didaktisch und operativ jedoch oft unterschlossen bleiben. Im gegenwärtigen Stand der Technik dominieren hierfür Retrieval-Augmented-Generation-Ansätze, die externe Wissensquellen in die Antwortgenerierung von Sprachmodellen einbeziehen; weiterentwickelte Varianten wie GraphRAG ergänzen dies um Wissensgraphen, Community-Hierarchien und strukturierte Zusammenfassungen. Der hier beschriebene Ansatz verschiebt den Fokus jedoch von der antwortzentrierten Retrieval-Optimierung zu einer quellengebundenen, statisch publizierbaren und editorisch überprüfbareren Lernarchitektur. RAG im allgemeinen Sinn dient dazu, autoritative Wissensbestände außerhalb der Trainingsdaten zur Answerzeugung heranzuziehen; GraphRAG erweitert dieses Paradigma um strukturierte, hierarchische Graphrepräsentationen. Der hier vorliegende Prototyp knüpft an diese Entwicklung an, übersetzt sie aber in einen inspectable learning space, also in einen Raum, in dem Quelle, maschinelle Ableitung und redaktionelle Bearbeitung sichtbar getrennt bleiben.

Die konzeptionelle Herkunft dieses Produkts liegt im Schnittpunkt von systemischer Integrationsarchäologie, reflexiver Gesellschaftsethik und einer expliziten Archivtheorie für fragmentierte Wissensräume. Die systemische Integrationsarchäologie setzt genau dort an, wo Biografien, Institutionen und digitale Speicherformen von Lücken, Tabus, Unsichtbarkeiten und algorithmisch erzeugter Marginalisierung geprägt sind, und formuliert dafür einen methodischen Zugriff aus Spurenlesen, Mustererkennung, Kontextualisierung, Integration und Validierung. Parallel dazu ist im Werkzusammenhang des AS-Instituts ausdrücklich vorgesehen, komplexe Werkstattfassungen nicht nur als PDF, sondern perspektivisch auch in barrierefreier HTML-Form, mit Rückmeldungsoffenheit und fortlaufender Weiterentwicklung bereitzustellen.

1. Einleitung: Das Grundproblem großer Archive

Die Gegenwart ist von einer eigentümlichen Asymmetrie geprägt: Noch nie standen so viele Texte, Notizen, PDFs, Protokolle, Chats, Skizzen und Fragmente zur Verfügung, und doch bleibt ihre tatsächliche Erschließung meist unterkomplex. Archive wachsen, ohne intelligenter zu werden. Dokumente liegen vor, aber ihre Begriffe sind nicht kartiert, ihre Nachbarschaften nicht sichtbar, ihre didaktischen Wege nicht gebaut. Genau dieser Befund bildet den Ausgangspunkt des hier beschriebenen Produkts: Nicht das einzelne Dokument steht im Zentrum, sondern die Transformation eines Korpus in einen begehbaren Wissensraum. Diese Denkfigur ist mit dem Ki-Lab-26-Zusammenhang und der systemischen Integrationsarchäologie eng verwandt: Es geht nicht nur um Wiederfinden, sondern um kontrollierte Sichtbarmachung, um strukturierte Verdichtung und um die methodische Wahrung epistemischer Unterschiede.

2. Stand der Technik: Von RAG zu strukturierter Wissenserschließung

Der gegenwärtige technische Mainstream im Bereich domänenspezifischer Wissenssysteme wird weitgehend von Retrieval-Augmented Generation getragen. RAG ergänzt Sprachmodelle um eine externe Wissensbasis, damit Antworten nicht allein aus Trainingsdaten, sondern aus aktuellen oder eigens bereitgestellten Quellen generiert werden. In der neueren Entwicklung verschiebt sich der Schwerpunkt dabei zunehmend von naiver semantischer Suche auf stärker strukturierte Verfahren. GraphRAG steht exemplarisch für diesen Übergang: Dort werden aus unstrukturiertem Text Wissensgraphen, Community-Hierarchien und Zusammenfassungen erzeugt, um Retrieval und reasoningfähige Zugriffspfade zu verbessern.

Zugleich zeigt der Stand der Technik, dass statische Generierung keineswegs eine veraltete Publikationsform ist, sondern im Gegenteil ein robustes Prinzip für wiederverwendbare, build-time erzeugte HTML-Ausgaben darstellt. Bei Static Site Generation wird HTML beim Build erzeugt und anschließend für Anfragen wiederverwendet; dies begünstigt Stabilität, Cachebarkeit und eine saubere Trennung von Erzeugung und Auslieferung. Für wissensbasierte Lernräume ist das deshalb relevant, weil hier nicht nur Interaktion, sondern auch Revisionsfähigkeit, Sichtbarkeit von Zuständen und publizierbare Beständigkeit zählen.

Der hier beschriebene Archive-to-Learning-Compiler steht somit in einem state-of-the-art-orientierten, aber eigenständig akzentuierten Feld. Er übernimmt aus dem aktuellen Entwicklungsstand die Einsicht, dass externe Wissensbestände strukturiert aufbereitet werden müssen; er verschiebt den Schwerpunkt jedoch von der bloßen Fragebeantwortung auf die **Publikation eines inspizierbaren Lernraums**. Nicht das Modell soll nur „besser antworten“, sondern der Korpus selbst soll in eine sichtbare, kontrollierbare Wissensinfrastruktur überführt werden.

3. Produktdefinition: Archive-to-Learning als eigenständige Klasse

Das Produkt lässt sich präzise als **universelles Authoring- und Publishing-System zur Transformation heterogener Wissensbestände in interaktive, multimodulare Lernräume** definieren. Sein Ziel ist es, aus Ordern, Uploads und Dokumentbeständen nicht bloß Suchindizes oder Chatkontexte zu erzeugen, sondern eine mehrschichtige Oberfläche, in der Knoten, Quellstellen, Slotstrukturen, Querverweise und spätere Medienausbauten geordnet erscheinen. Damit wird aus einem Archiv ein Lernraum erster Ordnung. Die Produktidee folgt keinem plattformökonomischen Primat der Sichtbarkeit, sondern einer epistemischen Logik von Provenienz, Ebenentrennung und Reviewbarkeit. Diese Ausrichtung steht in Kontinuität mit der am AS-Institut formulierten Kombination aus systemischer Analyse, Meta-Reflexion, evidenzbasierter Unterscheidung und ethischer Prozessbegleitung.

Im Unterschied zu vielen gegenwärtigen Wissenssystemen ist damit auch die Zielgestalt anders bestimmt. Das Produkt erzeugt nicht primär einen Assistenten, sondern einen **Raum**. Dieser Raum soll lesbar, durchquerbar und überprüfbar sein. Quelle, maschinelle Vorstrukturierung und redaktionelle Bearbeitung werden nicht ineinander verschmolzen, sondern sichtbar auseinandergehalten. Gerade hierin liegt sein wissenschaftlicher Mehrwert: Der Nutzer erhält keine epistemisch geglättete Oberfläche, sondern eine explizite Architektur der Differenz.

4. Implementationsarchitektur v0.1

Der bisher realisierte Prototyp besteht aus vier funktionalen Bausteinen. Baustein A implementiert eine vollständige Ingest-Strecke mit Scan, Klassifikation, Extraktion, konservativer Normalisierung und Berichtswesen. Baustein B realisiert eine Node Engine, die aus normalisierten Artefakten konservative Kandidaten, validierte Knoten und erste Relationen erzeugt. Baustein C ergänzt standardisierte Slotstrukturen für Text, Referenz, Medien, Prompt und Redaktion. Baustein D führt diese Ebenen in einen statischen HTML-Publisher über. Nach dem dokumentierten Projektstand umfasst der Demonstrator 106 validierte Knoten, 2756 Slots, eine Startseite, 106 Knotenseiten sowie die zugehörigen Index- und Reportartefakte.

Die technische und methodische Logik dieses Schritts besteht darin, dass aus heterogenen Rohdokumenten eine **statische, aber dennoch interaktive Wissensoberfläche** entsteht. Interaktiv ist sie nicht im Sinn eines sozialen Netzwerks, sondern als navigierbarer, typisierter und quellengebundener Raum. Dies entspricht sowohl der Zielrichtung der systemischen Integrationsarchäologie als auch dem expliziten Projektinteresse an PDF-/HTML-Doppelformen, dialogischer Werkstattlogik und maschinell unterstützter, aber nicht delegierter Strukturarbeit.

5. Der aktuelle Demonstrator: analytische Beschreibung der Website

Die öffentlich zugängliche Demo unter https://as-institut.de/demo_as_institut/index.html dokumentiert den gegenwärtigen Projektstand in verdichteter Form. Bereits die Startseite bezeichnet die Ausgabe ausdrücklich als „generated static learning space“ und benennt als Grundprinzip die Trennung zwischen validierten Knoten, autogenerierten Slot-Entwürfen und quellengebundenen Exzerpten. Zugleich weist sie 106 validierte Knoten, sieben Knotenklassen, eine Standard-Sichtbarkeit auf Privatsniveau sowie einen Review-Vorbehalt für Publikation aus. Damit wird nicht nur ein Korpus dargestellt, sondern eine explizite epistemische Schutzfunktion der Oberfläche sichtbar gemacht. (as-institut.de)

Die Startseite enthält „Central Entry Points“, also jene Knoten, die momentan die reichste Kombination aus Slot-Inhalt und stabiler Provenienz aufweisen. Darunter erscheinen etwa „Antifaschistische Schutz-Institut“, „Ki-Lab-26-Archive-Strukture-v1.1“, „Flattening“, „Cluster- und Dossier-Ebene“ oder „Funktion der Taxonomie“. Bereits an dieser Auswahl wird deutlich, dass der Demonstrator sich nicht auf reine Dokumentdarstellung beschränkt, sondern einen Übergang von Werk, Methode und Metastruktur vollzieht. Besonders aufschlussreich ist hier, dass manche Knoten bereits belastbare Kurzbestimmungen tragen, während andere ausdrücklich als noch nicht stabil definiert ausgewiesen werden. Auch dies ist kein bloßes UI-Detail, sondern Ausdruck einer kontrollierten, nicht-simulierten Wissensordnung. (as-institut.de)

Unter „Browse The Corpus-Derived Node Space“ werden die erzeugten Knoten nach Typen gegliedert. Die derzeitige Verteilung beträgt 79 Concepts, 13 Methods, 5 Patterns, 4 Visuals, 3 Institutions, 1 Path und 1 Work. Diese Verteilung zeigt zweierlei zugleich: erstens, dass die Korpusverdichtung tatsächlich eine systematische Typisierung erzeugt; zweitens, dass der gegenwärtige Demonstrator stark begriffs- und methodenlastig ist. Für einen frühen Prototyp ist das plausibel, weil zunächst das begriffliche Skelett des Raums sichtbar werden musste; didaktisch bedeutet es jedoch auch, dass spätere Lernpfade und kuratierte Einstiege erforderlich sein werden. (as-institut.de)

Eine weitere aufschlussreiche Beobachtung ist die zugleich vorhandene Stärke und Unfertigkeit des Systems. Die Startseite vermerkt, dass in `project.yaml` noch keine projektspezifische Beschreibung vorliegt. Damit ist der Raum bereits formal funktionsfähig, aber noch nicht vollständig kuratiert. Ebenso finden sich unter den Zentralpunkten mehrere Knoten ohne stabile Kurzdefinition, während einzelne andere bereits relativ präzise formuliert sind. Das System zeigt also seine eigene Entwicklungsphase offen an, statt eine scheinbare Vollendung zu simulieren. Dies ist wissenschaftlich und methodisch überzeugend, auch wenn es publizistisch noch nachgeschärft werden muss. (as-institut.de)

6. Bisheriger Nutzen und Stand der Technik

Der bisherige Nutzen des Demonstrators liegt auf vier Ebenen.

Erstens wird ein zuvor verstreuter Dokumentenbestand in einen **navigierbaren semantischen Raum** überführt. Das allein ist bereits mehr als eine technische Komfortfunktion, denn es verändert den epistemischen Status des Korpus: Aus Sammlung wird Infrastruktur.

Zweitens wird eine **sichtbare Ebenentrennung** erreicht. Der Demonstrator hält validierte Knoten, autogenerierte Slots und quellengebundene Exzerpte sichtbar auseinander. Damit adressiert er ein Kernproblem vieler KI-gestützter Wissenssysteme, nämlich die stillschweigende Vermischung von Quelle, Inferenz und Darstellung. Genau diese Trennung wird auf der Website selbst explizit benannt.

Drittens entsteht ein **statisch publizierbarer Lernraum**, der nicht von einer permanent laufenden Interaktionsinfrastruktur abhängig ist. Im Lichte aktueller state-of-the-art-Entwicklungen ist das bemerkenswert, weil es graph- und retrievalnahe Strukturierung mit einer build-time publizierten Oberfläche verbindet. Technologisch steht das Produkt damit zwischen strukturiertem Corpus Engineering und statischer Wissenspublikation; funktional ist es näher an einer quellengebundenen Wissensarchitektur als an einem bloßen Chat-Frontend.

Viertens erzeugt der Prototyp bereits jetzt einen **praktischen Erkenntnisgewinn über das eigene Material**. Die Website macht sichtbar, welche Knoten stark sind, welche noch keine stabile Kurzdefinition tragen, welche Typklassen dominieren und wo Kurationsbedarf besteht. Mit anderen Worten: Das System erschließt nicht nur das Archiv, sondern diagnostiziert zugleich seinen eigenen Verdichtungszustand.

Im engeren Sinn des „Stands der Technik“ lässt sich der Demonstrator deshalb als **state-of-the-art-orientierter, quelleninspektiver Lernraumgenerator** beschreiben. Er ist nicht deshalb fortgeschritten, weil er global die komplexeste Retrieval-Infrastruktur hätte, sondern weil er einen seltenen Übergang schafft: von strukturiertem Corpus Engineering zu einer epistemisch kontrollierten, statisch publizierten Wissensoberfläche.

7. Grenzen des aktuellen Projektstands

Gerade weil der Demonstrator bereits funktioniert, lassen sich seine Grenzen nun präzise benennen.

Die erste Grenze liegt in der **Kurationsdichte**. Ein großer Teil der zentralen Knoten ist noch mit Platzhaltern oder vorläufigen Kurzdefinitionen versehen. Das Produkt kann also bereits Struktur erzeugen, aber noch nicht überall hinreichend starke editorische Verdichtung.

Die zweite Grenze liegt in der **Typ- und Relationstiefe**. Nach dem dokumentierten Entwicklungsstand sind 106 validierte Knoten vorhanden, aber nur vier vorsichtige Relationsvorschläge. Das ist für einen ersten konservativen Lauf methodisch legitim, reicht jedoch noch nicht aus, um eine wirklich starke begriffliche Navigation, Pfadbildung oder dichtes semantisches Querlesen zu ermöglichen.

Die dritte Grenze liegt in der **Startseitendidaktik**. Die Demo eröffnet den Raum derzeit noch stark von der Architektur und Taxonomie her. Für Entwicklungszwecke ist das sinnvoll; für spätere Leserführung braucht es stärker kuratierte Einstiege, in denen Methode, Fallraum und Werkraum nicht nur gelistet, sondern didaktisch geschichtet werden.

Die vierte Grenze liegt in der **Werkstattdimension**. Die Ausgabe ist publizierbar, aber die eigentliche lokale Review- und Bearbeitungsoberfläche steht noch aus. Der Prototyp kann also bereits zeigen, was das Produkt ist; er kann aber die editorische Weiterarbeit noch nicht vollständig in einer integrierten Oberfläche tragen.

8. Ausblick

Der nächste Entwicklungsschritt sollte nicht in eine vorschnelle Plattformisierung führen, sondern in eine Phase kontrollierter Qualitätsverdichtung. Zunächst bedarf es einer kuratierten Konsolidierung der Zentralknoten, einer Schärfung der Taxonomie, einer Verbesserung der Relationslogik und einer didaktisch stärkeren Startseitenarchitektur. Erst darauf aufbauend wird eine eigentliche Werkstattoberfläche sinnvoll, in der Knoten, Slots und Relationenvorschläge lokal redigiert und anschließend neu publiziert werden können.

Mittelfristig liegt die eigentliche Tragweite des Produkts in seiner **Verallgemeinerbarkeit**. Das AS-Institut ist hierfür ein anspruchsvoller Erstkorpus, aber nicht der begriffliche Endpunkt. Der Demonstrator zeigt bereits jetzt, dass die Grundidee universalisierbar ist: Wer über heterogene Wissensbestände verfügt, kann daraus mit einer geeigneten Pipeline einen begehbaren Lernraum erzeugen. Langfristig wäre daraus eine neue Klasse von Werkzeugen denkbar, die zwischen Archivwissenschaft, Knowledge Organization, strukturierter KI-Assistenz und didaktischer Publikationsarchitektur vermittelt. Im Licht der aktuellen Entwicklung von RAG und GraphRAG wäre dies keine bloße Variation, sondern eine Verschiebung des Fokus: weg vom reinen Antwortsystem, hin zu einer publikationsfähigen, überprüfbaren und editorisch verantwortbaren Wissensinfrastruktur.

9. Schluss

Der bisherige Projektstand belegt, dass ein Archive-to-Learning-Compiler nicht bloß als Konzeptpapier tragfähig ist, sondern bereits in einer ersten realen, online einsehbaren Form existiert. Die Demo des AS-Instituts zeigt eine funktionierende Transformation vom Dokumentenkorpus zum statischen Lernraum, einschließlich Knotenstruktur, Slotlogik, Search Index, Sitemap und typisierter Navigation. Sie ist damit weder bloß ein Dateiverzeichnis noch schon die Endform einer digitalen Schule. Sie ist etwas methodisch Präziseres: **ein belastbarer Prototyp einer quellengebundenen, statischen und interaktiven Wissensarchitektur**.