

Kulturdaten im Open Data Portal Schleswig-Holstein

1 Auftrag

Im Rahmen eines auf sieben Monate angelegten Projektes der Schleswig-Holsteinischen Landesbibliothek (Kulturhackathon Coding da Vinci) in Kooperation mit dem ZIT war zu untersuchen, inwieweit Datenbestände bei Kulturinstitutionen des Landes Schleswig-Holstein zur Veröffentlichung im Open-Data-Portal des Landes geeignet sind und wie Kulturinstitutionen zur Veröffentlichung ihrer Datenbestände unter freien Lizenzen zu motivieren sind. Das Projekt lief vom 1. Dezember 2020 bis 30. Juni 2021.

Methode

Die Untersuchung wurde im Rahmen der von der Kulturstiftung des Bundes geförderten Veranstaltungsreihe „Coding da Vinci – der Hackathon für offen Kulturdaten“ durchgeführt, der von der Schleswig-Holsteinischen Landesbibliothek (federführend), mit Unterstützung der Christian-Albrechts-Universität zu Kiel, der Fachhochschule Kiel, dem Ministerium für Energiewende, Landwirtschaft, Umwelt, Natur und Digitalisierung Schleswig-Holstein, der Stiftung Schleswig-Holsteinische Landesmuseen, Toppoint e.V., IF(game)SH, Kultursphäre sh und der digiCULT Verbund eG ausgerichtet wurde.

Über dreißig Kulturinstitutionen wurden in Einzel-Videokonferenzen und telefonisch bezüglich ihrer Daten beraten. Dabei wurden die vorhandenen digitalen Datenbestände durchgesehen, ihre Eignung zur Veröffentlichung sowie, Hinderungsgründe und Verbesserungsvorschläge besprochen und Fragen der Kulturinstitutionen beantwortet. Außerdem wurden im Nachgang der Datenveröffentlichung zehn Interviews mit Kulturinstitutionen geführt.

Bei dieser Vorgehensweise wurden die Kulturinstitutionen nur wenig berücksichtigt, die sich aus dem Projekt im Laufe der Vorbereitung zurückgezogen haben.

Im Bericht wird das Open Data Portal Schleswig-Holstein mit ODPSH abgekürzt.

2 Bericht

2.1 Hohe Motivation bei Kulturinstitutionen

Bei den Kulturinstitutionen ließ sich insgesamt eine hohe Motivation zur Auseinandersetzung mit dem Thema „Digitalisierung“ feststellen. Es wird bei den Institutionen überwiegend als wichtig eingestuft. Damit ist die Hoffnung verbunden auf besseres Marketing im digitalen Raum, auf die Gewinnung neuer Kultur-Interessenten, auf mehr Kommunikation mit den Nutzer:innen, auf neue Kontakte und auf Anregungen für die eigene Datennutzung. Die Beratung über die mit offenen Daten verbundene Fernziele, etwa Ideen wie Linked Open Data oder Semantic Web, halfen bei der Einordnung.

Das persönliche Engagement der Mitarbeiter:innen, die aktiv an der Datenbereitstellung beteiligt waren, war insgesamt sehr hoch. Dabei spielte die Tatsache eine Rolle, dass die Datenbereitstellung im Rahmen der Veranstaltungsreihe erfolgte und deshalb unmittelbar Ergebnisse erhofft wurden. Dementsprechend haben Kulturinstitutionen, die im Rahmen der Veranstaltung einen Datennutzer fanden, die Datenbereitstellung insgesamt als besonders positiv beurteilt.

2.2 Arbeitsbelastung bei Mitarbeiter:innen

Probleme bereitet allerdings die überwiegend als unzureichend beurteilte Personalausstattung. Deshalb konnten die Mitarbeiter:innen nicht so viel Zeit investieren, wie er oder sie es vermutlich gerne getan hätten. Arbeitsüberlastung und Zeitdruck wirkten sich gelegentlich negativ auf die Motivation aus.

Die Teilnahme an „Coding da Vinci“ erforderte außerdem ein Mindestmaß an Datenqualität, Metadaten, Datensetbeschreibungen und Erläuterungen, was die Daten insgesamt qualitativ hochwertiger und attraktiver macht, aber auch mit zusätzlicher Belastung und Zeitdruck verbunden war. Andererseits erzwangen die Fristen der Veranstaltung aber auch eine zügige Datenbereitstellung.

2.3 Steile Lernkurve

Probleme bereiteten häufig auch mangelnde Kenntnisse im Umgang mit der Digitalisierung. Die Beratungsunterstützung wurde deshalb sehr positiv aufgenommen.

Wissen fehlte etwa in folgenden Bereichen:

- Grundsätzliche Unterscheidung der Daten in Metadaten / Mediendateien / sonstige Daten.
- Wie speichert man Daten, so dass sie gut nutzbar sind? Welche Datenformate sind möglichst offen / zugänglich? Wie kann man die Daten maschinenlesbar und -bearbeitbar machen? Welche Struktur (tabellarisch, hierarchisch) ist sinnvoll? Was ist bei nicht-typisierten Formaten (wie CSV-Tabellen) zu beachten?
- Welche Eigenschaften sind wichtig bei Mediendateien, welche Struktur und Benennung?

Viele Weichen wurden schon in der Vergangenheit ungünstig gestellt und lassen sich kurzfristig nur mit großem Aufwand korrigieren. Dazu gehören:

- Benutzung proprietärer Formate,
- uneinheitliche Struktur und Benennung von Mediendateien,
- völliges Fehlen von Metadaten.

Mit der individuellen Beratung war ein hoher Lerneffekt in Bezug auf Digitalisierung bei den Kulturinstitutionen verbunden. Dieser Lerneffekt war für die Kulturinstitutionen maßgeblich für eine positive Gesamtbeurteilung, auch dann, wenn sich direkt keine Datennutzer:innen im Rahmen von Coding da Vinci fanden.

2.4 Die Veröffentlichung hindernde Befürchtungen

Mit der Veröffentlichung von Kulturdaten – egal ob gemeinfrei oder unter freien Lizenzen – sind bei den Kulturinstitutionen häufig Befürchtungen verbunden:

- a) Einige Institutionen möchten wissen, in welchen Zusammenhängen ihre Daten verwendet werden, auch einfach deshalb, um mit den Nutzern in Kontakt treten zu können.
- b) Einige Institutionen befürchten Einnahmeverluste aus Gebühren oder Lizenzentgelten. Häufig wird deshalb eine NC-Lizenz gewollt.
- c) Einige Institutionen zögern mit der Veröffentlichung, weil dieser Schritt nicht rückgängig gemacht werden kann.
- d) Einige Institutionen haben eigene Vorstellungen darüber, welche Verwendung ihrer Daten angemessen ist und welche nicht und möchten eine Lizenz, die inhaltlich bestimmte Verwendungen ausschließt oder eingrenzt.
- e) Einige Institutionen befürchten, dass ihre Daten in verbotenen oder ethisch zweifelhaften Zusammenhängen verwendet werden und dies auf sie zurückfällt.
- f) Einige Institutionen sind unsicher, ob die Veröffentlichung ihrer Daten (z.B. historische Fotos mit heute verbotenen Symbolen) rechtlich zulässig ist.
- g) Mögliche negative Reaktionen in den „Sozialen Medien“ waren in den Vorbereitungen kein wesentliches Thema, können aber zum Problem werden. Im Rahmen des Hackathons wurde ein Datenset aus der Kolonialzeit nach einem Rassismuskvorwurf zurückgezogen.

Wie viele Institutionen sich letztlich deswegen zurückgezogen haben, ließ sich nicht klären. Bei den Informationsveranstaltungen stellten diese Fragen jedenfalls den überwiegenden Anteil.

2.5 Lizenzwahl

Ein Schwerpunkt der Beratung lag bei der Information über die verschiedenen freien Lizenzen, deren Wahl den Kulturinstitutionen aufgrund mangelnder Kenntnisse Schwierigkeiten bereitet. Im Nachhinein lässt sich sagen, dass ein großer Teil der für die Veröffentlichung vorgesehenen originären Inhalte ohnehin durch Zeitablauf gemeinfrei ist, so dass es für diese auf die Lizenzwahl rechtlich gesehen nicht ankommt.

Für Daten (meist Mediendateien), deren Rechte bei den Kulturinstitutionen selbst liegen, wurde – nach Ausräumung der unter 3.5 dargestellten Bedenken – meist eine Lizenzierung unter CC-BY(-SA) gewählt. Bei Daten anderer Rechteinhaber:innen, deren Schutzfrist nicht abgelaufen war, kam eine Veröffentlichung ganz überwiegend von vornherein nicht in Betracht. Für die von den Kulturinstitutionen erstellten Metadaten bereitete eine Lizenzierung unter CC – zum Teil auch unter CC-0 - regelmäßig keine Schwierigkeiten.

2.6 Probleme im Open Data Portal Schleswig-Holstein

2.6.1 Technische Basis

Das ODPSH folgt der der genutzten Software (CKAN, Comprehensive Knowledge Archive Network) zugrundeliegenden Logik und dessen Aufbau, so dass die Kulturdaten, die dazu schlecht passen, Schwierigkeiten bereiten.

Kulturdaten bestehen häufig aus vielen Mediendateien und einer dazugehörigen Metadatendatei, was bei CKAN nicht vorgesehen ist. Die vielen Einzeldateien einzeln hochzuladen würde weder beim Hochladen noch beim Download viel Freude bereiten, so dass man sie in einer gepackten Datei zusammenfassen müsste, was aber ungünstig ist, wenn man als potentielle Datennutzer:in nur eine oder wenige Mediendateien sichten oder verlinken möchte. Außerdem erschwert es Aktualisierungen und Ergänzungen des Datenbestandes. Es ließ sich dankenswerterweise eine pragmatische Lösung improvisieren, notwendigerweise an CKAN vorbei, durch Anlage eines gesonderten Verzeichnissystems.

Außerdem sollen laut CKAN nur alternative Versionen derselben Daten in einem Datensatz bereitgestellt werden. Im Kulturbereich sind aber für das bessere Verständnis oft unterschiedliche Daten und Datenformate, die inhaltlich zusammengehören, in einem Datensatz zusammengefasst, was bei CKAN so nicht vorgesehen ist. Man kann dafür eine Verbindung mehrerer Datensätze über gemeinsame Attribute konstruieren, aber das ist eigentlich impraktikabel.

2.6.2 Mangelnde Benutzerführung

Aus Sicht der Kulturinstitutionen bietet das OPDSH keine inhaltliche Ansprache oder Führung durch die Datenlandschaft in Bezug auf das Thema „Kultur“, sondern nur eine technische Suchschnittstelle ohne Themenbezug. Aus diesem Grund erschien vielen Kulturinstitutionen das Portal als nicht attraktiv für ihre Daten, weil potentielle Interessenten sie dort aus ihrer Sicht gar nicht erst suchen würden.

3 Handlungsempfehlungen

3.1 Leitfaden zur Standardisierung

Es ist wichtig, dass bereits bei der Datenerfassung auf die Einhaltung von offenen Standards, die Nutzung offener Datenformate und auf eine geeignete Struktur und Systematik (etwa bei der Speicherung und Benennung von Mediendateien) geachtet wird. Bei diesen Entscheidungen sind die Kulturinstitutionen häufig überfordert.

Bei den Metadaten besteht die zusätzliche Schwierigkeit darin, dass es viele konkurrierende Standards gibt, in denen unterschiedliche Vokabulare genutzt werden können. In den Metadatenformaten können außerdem verschiedene Datenfelder gewählt werden sowie individuelle Erweiterungen je nach Standard genutzt werden. Quintessenz ist, dass für Metadaten von Kulturinstitutionen in der Regel nur in einfachem Tabellenformat nach Vernunft-bezogener Auswahl von geeigneten Spalten in Betracht kommen, sofern man nicht eine umfangreiche Komplett-Software hat oder sich in der Tiefe mit dem sehr komplexen Thema auseinandersetzen will, was kurzfristig kaum geht.

Hierzu wäre die Erstellung eines Leitfadens durch die Open Data-Leitstelle sinnvoll, der niedrigschwellig grundlegende Prinzipien bei der Arbeit mit Daten und Mediendateien darstellt und Empfehlungen zur Formatwahl je nach Einsatzgebiet enthält. Dabei sollte die Darstellung vor allem anhand praktischer Beispiele erfolgen.

3.2 Bildungs- und Beratungsangebote

Bildungsangebote zum Umgang mit Kulturdaten, die eine eingehendere, ausführliche Beschäftigung mit dem Thema der Veröffentlichung ermöglichen, wurden sehr häufig gewünscht. In Zeiten der Digitalisierung muss es sich dabei keineswegs um Präsenzveranstaltungen handeln, „Tutorials“ sind in anderen Lebensbereichen eine gängige Informationsquelle, die auch hier genutzt werden könnte.

Zur Lösung konkreter Problem und zur Klärung von Verständnisfragen, wäre eine Anlaufstelle zur persönlichen Beratung bei der konkreten Umsetzung der Leitfadens-Empfehlungen sinnvoll und notwendig. Ungeeignete Weichenstellungen schon im Vorfeld zu erkennen und zu vermeiden, ist langfristig ökonomischer als umfangreiche Digitalisierungsprojekte im Nachhinein.

Leider fehlen gute kleine Open-Source-Tools, die die Heranführung an die Arbeit mit Metadaten erleichtern würden - die vorhandenen sind eher große Komplettlösungen, wo die Kulturinstitution weitgehend von den zugrundeliegenden Metadatenstandards abgeschirmt wird. Die resultierende Datei kann die Kulturinstitution dann selbst nicht genauer prüfen.

Eine große Motivation, sich trotz Komplexität mit dem Thema zu beschäftigen gibt es allemal, weil sich die weltweite, auf Metadatenstandards basierte Vernetzung der Daten (Semantic Web, Linked Open Data, 5 Sterne Modell) am Horizont abzeichnet und sich die Web- und Suchmaschinenlandschaft verändern wird.

Hinsichtlich der vielfach gewünschten Fortbildung sind die Dachverbände (etwa der Museumsverband Schleswig-Holstein und Hamburg e. V.), nachgeordnete Behörden (wie die Schleswig-Holsteinische Landesbibliothek) oder Unternehmen (wie die digiCULT Verbund eG) gefordert.

3.3 Unterstützung bei juristischen Fragen

3.3.1 Lizenzen

Der Leitfaden, der vom ODPSH derzeit entwickelt wird, ist ein guter Ansatz. Eine Weiterentwicklung mit mehr Details und Beispielen könnte schon eine gute Lösung sein. Viele praktische Beispiele erleichtern es dem Laien, sich das Thema zu erschließen.

Dabei ist auch die Urheberrechtsreform von 2021 berücksichtigen, nach der Vervielfältigungen eines gemeinfreien visuellen Werks keinen Leistungsschutz mehr genießen (§ 68 UrhG).

3.3.2 Andere juristische bzw. ethische Probleme

Grundlegende juristische Probleme bei offenen Kulturdaten (insbesondere historischen) können nicht von Kulturinstitutionen selbst geklärt werden. Es ist auch weder sinnvoll noch notwendig, dass jede Institution dies für den eigenen Datenbestand gesondert untersuchen lässt. Im Rahmen dieses Projektes tauchten drei Problembereiche auf:

- Veröffentlichung von Bildern oder anderen Dokumenten, die heute verbotene Symbole (vor allem aus der Zeit des Nationalsozialismus) enthalten.
- Veröffentlichung von Bildern oder anderen Dokumenten, die nach heutiger Einschätzung rassistisch sind oder sonst gegen die Menschenwürde verstoßen (vor allem aus der Zeit des Kolonialismus).
- Veröffentlichung von Bildern oder anderen Dokumenten, die personenbezogene Daten oder Bilder verstorbener Personen enthalten (hinsichtlich lebender Personen ist die Rechtslage ohnehin klar).

Hierzu sollte das ODPSH eine Stellungnahme der zuständigen Fachbehörden oder eines externen Gutachters einholen und zur Verfügung stellen, auf die sich Kulturinstitutionen berufen können.

3.4 Probleme im Open Data Portal Schleswig-Holstein

3.4.1 Technische Basis

CKAN lässt in der Flexibilität zu wünschen übrig. Es nützt wenig, wenn die eventuelle Möglichkeit besteht, dass Daten aufgrund einfacherer logischer Struktur besser automatisiert nutzbar sein könnten, wenn dann viele Daten wegen konzeptuellen Problemen gar nicht eingestellt werden können. Auf Dauer sollte der technische Aufbau des ODPSH unter diesem Gesichtspunkt überdacht werden.

Für das hilfweise angelegte Verzeichnissystem, in dem derzeit Mediendateien abgespeichert werden, muss eine geeignete Dauerlösung gefunden werden, etwa in Form eines Medienservers.

Darüber hinaus sollte die Möglichkeit bestehen, Datensätze automatisiert aus anderen Datenbanken (etwa DigiCULT) zu importieren, ohne dass unnötige Download- und Uploadschritte anfallen.

3.4.2 Mangelnde Benutzerführung

Es sollte ein Weg gefunden werden, Kulturdaten direkt auffindbar zu machen. Die Kategorie „Bildung, Kultur und Sport“ ist sachlich wenig hilfreich. Die vorhandenen Einträge im Kulturbereich sind kaum sichtbar.

Inhaltlich zusammenhängende Datensammlungen, die aufgrund der Struktur des CKAN in Einzelstücke aufgeteilt werden müssen, sollten auch als Ganzes beschrieben werden können.

Jede weitere Zugänglichkeit für Kulturinstitutionen wäre zu begrüßen, etwa ein Text zur direkten einladenden Ansprache als Kultur-Wegweiser durch das Portal mit einfachen praktischen Beispielen und Hinweisen, sowohl für datensuchende wie datengebende Kulturinstitutionen.

3.5 Zukünftige Anreize zur Datenveröffentlichung

Der Hackathon hat sich als starker Anreiz zur Datenveröffentlichung erwiesen. Zukünftig muss die Attraktivität des ODPSH durch andere Maßnahmen erhöht werden. Hierfür eignet sich ein inhaltsbezogenes Portal für Schleswig-Holstein zum Themenbereich Kulturdaten, das nicht selbst die eigentlichen Daten hält, sondern eine zentrale Anlaufstelle für Datensuchende darstellt mit Verweisen auf die eigentlichen Daten, die auf diversen Portalen liegen.

Das ODPSH ist ein guter und kostenlos nutzbarer Ort, wo Datengeber ihre Daten halten können. Es erscheint nur leider aufgrund der Ausrichtung unwahrscheinlich, dass es sich als zentrale Anlaufstelle für alle Datengeber und Datennutzer zum Thema Kulturdaten etablieren könnte, weil Interessierte aus dem Kulturbereich Wert auf inhaltliche Ansprache und Führung durch die komplexe Themenwelt der Kulturdaten legen. Datengeber wollen ihre Daten auch dort sichtbar machen, wo sie Datensuchende am ehesten vermuten.

Heutzutage ist die Suche nach Daten anstrengend, da diverse Datenportale, wie z.B. verschiedene Bibliotheken, versuchen, im Kulturbereich ein Renommee aufzubauen und man jedes Portal und eine Vielzahl an verstreuten Webseiten aufsuchen muss, um herauszufinden, welche Daten zu finden sind, aber auch wie zugänglich und verwendbar diese sind. Ein zentraler Hub böte hier eine Lösung, der als Wegweiser mit mittlerem Detailgrad dient und zusätzlich auch alle Portale auflisten könnte.

Das Portal auf Open Data einzuschränken, würde bewirken, dass die Fragmentierung des Zugangs zu Kulturdaten bestehen bliebe und dass das Portal nicht als zentral empfunden werden würde. Folglich sollte das Portal nicht eingeschränkt sein, damit eine zentrale

Vollständigkeit angestrebt werden kann, jedoch sollte es die (Freiheit der) Lizenzen prominent hervorheben und standardmäßig nach Freiheit der Lizenzen bevorzugend Listen sortieren und danach sortierte Suche anbieten, um Datengeber zu offenen Lizenzen zu ermuntern. Die Listen sollten aber auch nach Unterthemen aufgegliedert werden. Kulturdaten-Interessierte aus dem privaten bis unternehmerischen Bereich wie z.B. Vereine sollten ebenso wie die staatlichen Kulturinstitutionen bei den Themen Open Data und Vernetzung der Daten (u.a. mit Hilfe von Metadatenstandards) unterstützt werden können.

Einzeldatensätze auflisten zu können, ist weniger wichtig als die vergrößernde Klammer Datenset (Sammlung); dort wo nur Einzeldatensätze vorhanden sind, müssen Datenset-Klammern notfalls willkürlich gebildet werden (z.B. durch Datensätze von gleichem Datengeber und gleicher Struktur). Es ergeben sich zwei Hierarchieebenen: Datenset und Datensatz, letzteres ist jeweils nur optional verfügbar. Zu den Datensets muss die Zugänglichkeit (Downloadbedingungen, Datenformate, ...) der Daten bestmöglich erfasst werden.

Das Portal soll auch verschiedene Service-Angebote enthalten, insbesondere zur pragmatischen Wissensvermittlung. Informationen für Datengeber, wie Leitfäden zur Einführung in eine einfache, strukturierte Datenerfassung oder Übersichten zum komplexen Metadatenbereich, und welche Software es dazu gibt. Auch die Datennutzer:innen sollen in ähnlicher Weise Überblick und Leitfäden/Einführungen erhalten. Hier könnte das ODPSH z.B. mit Leitfäden unterstützen, das Portal kann das ODPSH und die Open Data Idee besonders befördern, Austausch in diversen Bereichen wäre möglich wie das Abstimmen der Strukturen (z.B. Granularität der Datensets) und bei den Abläufen.

Das Portal soll allgemein die digitale Transformation befördern, indem es diese aus Themenperspektive für alle verständlich und zugänglich macht: Wissen erwerben und gleich auch selbst ausprobieren mit Hilfe von Beispiel-orientierten Einführungen und Daten, die man sich auch gleich für den Zweck als Übungsmaterial und zur Nutzung leicht herausuchen kann und gleichzeitig ein für Datengeber attraktives Umfeld für ihre Kulturdaten.

