

5 einfache Schritte, wie Kulturinstitutionen besser mit ihren Daten arbeiten können

Einleitung

Was ein Familienticket, der Kauf eines Kaffees im Museumsshop, die Digitalisierung von Exponaten und Social Media Likes gemeinsam haben? All diese Dinge produzieren Daten, die mithilfe richtiger Aufbereitung und Analyse wichtige Informationen über die Besucher*innen von Kulturinstitutionen geben. Entsprechend können etwa Besuchsangebote analysiert und zielgruppengerecht optimiert werden. Kulturdaten – etwa von eigenen digitalisierten Beständen – können aber auch auf vielfältige Art und Weise kreativ genutzt werden, wie bspw. die unzähligen Ideen im Rahmen des Kulturhackathon Coding da Vinci zeigen. Big Data stellt dabei schon längst nicht mehr ein zukünftiges Phänomen dar, das uns alle irgendwann betreffen wird. Die Sammlung, Nutzung und Auswertung großer und komplexer Daten ist im Zeitalter der Digitalisierung zur Normalität geworden, sodass ganze Geschäftsmodelle auf Daten als Schlüsselressource basieren. Mit dieser neuen Ära ergeben sich branchenübergreifend aber auch viele offene Fragen im richtigen und rechtssicheren Umgang mit Daten und dessen weiterer Nutzung, so auch für Kulturbetriebe. Welche Voraussetzung sind für die Arbeit mit Daten zu erfüllen? Welche Daten dürfen gesammelt werden? Wie sind diese aufzubereiten und zu analysieren?

Diese Broschüre soll Kulturinstitutionen einen kompakten Überblick über die Datenerhebung, -aufbereitung und -analyse geben und dabei rechtliche und technische Fragen klären.

1. Ziele definieren

Die wohl wichtigste Voraussetzung für den erfolgreichen Umgang mit Daten ist die Definition von Zielen. Kulturinstitutionen sollten eine klare Antwort auf die Frage, was sie mit den Daten erreichen wollen, geben können. Sollen die Daten bspw. für eine Zielgruppensegmentierung genutzt werden oder liegt der Fokus auf der Webseitenoptimierung? So vielfältig wie Daten sein können, so vielfältig ist dessen Einsatzgebiet. Umso wichtiger ist es, im Vorfeld die Ziele zu definieren, um eine strukturierte und umfassende Datengrundlage aufbauen zu können. Dabei gilt: Je detaillierter die Ziele formuliert werden, desto höher ist auch die Erfolgschance. Entsprechend ist es ratsam die SMART-Formel zu nutzen. Demnach sollten die Ziele spezifisch, messbar, attraktiv, realistisch und terminiert sein. In diesem Zusammenhang spielen sowohl die personellen als auch die technischen Ressourcen einer Kultureinrichtung eine wesentliche Rolle. Welche Bedeutung soll der Arbeit mit Daten zugeschrieben werden, welche Kapazitäten stehen zur Verfügung und wie viel ist man bereit zu investieren? All diese Fragen bilden die Grundlage des weiteren Vorhabens.

Die formulierten Ziele und deren Bedeutung sollten dabei offen und transparent innerhalb der Kultureinrichtung kommuniziert werden. Dies schafft ein besseres Verständnis für die etwaige Notwendigkeit der Datenerhebung und späteren Datenanalyse.

2. Vorbedingungen klären

Allgemein sind zwei Handlungsfelder in der Arbeit mit Daten zu unterscheiden: die Erhebung und Verwendung eigener Daten und/oder die Nutzung bestehender Datensets wie bspw. lokale Tourismuszahlen. Ersteres erfordert eine Bestandsaufnahme. Liegen bereits eigene Daten für die Erreichung der Ziele vor oder müssen diese noch erhoben werden? Dabei ist zwischen personenbezogenen und nicht personenbezogenen Daten zu unterscheiden. Bei der Nutzung fremder Datensets stellt sich die Frage nach Eigentums- und Nutzungsrechten. Diese gilt es zwingend im Vorfeld zu klären.

Spätestens mit Inkrafttreten der Datenschutz-Grundverordnung der Europäischen Union (DSGVO) im Mai 2018 ist die Erhebung und Verarbeitung personenbezogener Daten mit einigen Auflagen verbunden. Demnach dürfen Informationen, die sich direkt oder indirekt auf eine identifizierbare natürliche Person beziehen, nur dann verwendet werden, wenn dies ausdrücklich von der betreffenden Person erlaubt wurde. Bei Personen, die das 16. Lebensjahr noch nicht vollendet haben, bedarf es der Einwilligung der Erziehungsberechtigten bzw. muss die Einwilligung mit Zu-

stimmung der Erziehungsberechtigten erfolgen. Diese muss noch vor der Datenerhebung erfolgen, eine nachträgliche Einwilligung ist nicht ausreichend.

Da fast jede Kulturinstitution über eine eigene Webseite verfügt und im Zuge dessen auch personenbezogene Daten erhebt, übermittelt, verarbeitet oder nutzt, schreibt die DSGVO eine Datenschutzerklärung vor. Diese soll die Nutzer*innen umfassend über die Verarbeitung von Informationen und Daten durch die Institution informieren. Zu den wichtigsten Inhalten zählen:

1. Allgemeine Datenschutzhinweise
2. Verantwortliche Stelle
3. Datenschutzbeauftragte*r
4. Art der Datenerhebung
5. Zweck der Datenerhebung
6. Rechte der Nutzer*innen
 - Auskunft
 - Berichtigung
 - Löschung/Einschränkung der Verarbeitung
 - Widerspruch
 - Recht auf Datenherausgabe und Übertragbarkeit

Sofern Cookies, Tracking- und Analyse Tools, Social Media Plugins und/oder Newsletter verwendet werden, sind diese separat mit einer umfassenden Erläuterung in die Datenschutzerklärung aufzunehmen, pauschale Hinweise sind nicht ausreichend.

Allgemein gilt es Folgendes zu beachten:

- Die Datenschutzerklärung muss von jeder Unterseite zu erreichen sein (z.B. Verlinkung im Footer).
- Die Datenschutzerklärung muss in der gleichen Schriftgröße und -art gestaltet werden wie die gesamte Webseite.
- Die Datenschutzerklärung ist von dem Impressum und den Allgemeinen Geschäftsbedingungen (AGB) abzugrenzen.
- Die Datenschutzerklärung ist modular aufgebaut und variiert entsprechend der Webseiten-Funktionen.

Neben einer Datenschutzerklärung müssen die personenbezogenen Daten anonymisiert, pseudonymisiert oder verschlüsselt werden, sodass keine Rückschlüsse auf die natürliche Person erfolgen kann. Die drei Verfahren unterscheiden sich dabei in der Art der Unkenntlichmachung. An dieser Stelle ist allerdings darauf hinzuweisen,

dass bei weitem nicht alle Daten personenbezogen sind. So unterliegen Bilder, Töne und Texte nicht der DSGVO. Werden bspw. Daten in Form von Fotos von Gebäuden aus dem öffentlichen Raum gesammelt, ist dies nicht in die Datenschutzerklärung aufzunehmen. Lediglich die Rechte an den Bildern sind zu klären.

Darüber hinaus ist es wichtig, das Eigentum an den erhobenen Daten sicherzustellen. Es ist zu klären, wo die Daten gespeichert werden, wer Zugriff darauf hat und ob dieser Zugriff unbeschränkt ist. Dies ist notwendig, um jederzeit Änderungen und Anpassungen vornehmen zu können und im späteren Verlauf Komplikationen mit der Datenanalyse zu vermeiden.

Weiter ist festzulegen, ob die Daten nur für interne Zwecke genutzt werden sollen oder ob sie Dritten bspw. über das Online-Portal „Open Data Schleswig-Holstein“ für weitere Zwecke zugänglich gemacht werden. Auch dies entscheidet über die Art und Weise der Datenerhebung.

Für konkrete datenschutzrechtliche Fragen sollten sich Kulturinstitutionen an eine Rechtsberatung wenden, sofern kein*e Datenschutzbeauftragte*r vorhanden ist. Dies ist unabdingbar, da fehlende oder falsche Datenschutzerklärungen mit Bußgeldern geahndet werden können. Darüber hinaus stellt das Unabhängige Landeszentrum für Datenschutz Schleswig-Holstein (ULD) auf der eigenen Webseite Informationsmaterialien zum Umgang mit Daten zur Verfügung. Allgemeine Fragen können auch per E-Mail geklärt werden, wobei eine umfangliche Beratung aus Kapazitätsgründen nicht realisierbar ist.

3. Datenqualität und Datenverarbeitung prüfen

Die Datenqualität ist das A und O einer jeden Analyse und steht damit über der Quantität. Je besser die Qualität, desto aussagekräftiger sind die Ergebnisse und die Handlungsempfehlungen, die abgeleitet werden können. Um die Datenanalyse individuell und zielgerichtet zu gestalten, ist der Zugriff auf die Rohdaten unerlässlich. Rohdaten sind dabei diejenigen Daten, die nach einer Erhebung in unverarbeiteter Form vorliegen. Zur besseren Einordnung und zum besseren Verständnis dieser Daten können Metadaten herangezogen werden. Diese beinhalten übergreifende Informationen zu der Datenquelle und können Antworten auf die sechs W-Fragen (Wer, Was, Wann, Wo, Wie und Warum) geben. Aggregierte Daten dagegen eignen sich in den seltensten Fällen für eine umfassende Datenanalyse, da Informationen durch das Zusammenfassen einer Vielzahl von Beobachtungen verloren gehen, die nicht wiederherzustellen sind. Darüber hinaus sollten die Daten in digitaler, maschinenlesbarer Form vorliegen. Die gängigsten Formate sind csv,

xml und.xlsx. Nur so ist es möglich, die Daten in verschiedensten Programmen wie Excel, SPSS oder R-Studio je nach Bedarf und Zielsetzung zu importieren, wobei R-Studio auf der gängigen Programmiersprache R basiert. Auch Python gilt als verbreitete Sprache. Neben diesen zwei genannten Voraussetzungen bildet das sogenannte Codebuch die dritte Grundlagenkomponente der Datenanalyse. Ein Codebuch ist eine Art Verzeichnis, die den numerischen Variablen und Ausprägungen im Datensatz eine Bedeutung zuschreibt. Somit ist der Datensatz für jeden, der darauf Zugriff hat, verständlich. Andernfalls könnten die Daten nicht interpretiert werden.

Beispiel:

Eine Kulturinstitution führt eine Befragung zur Zielgruppensegmentierung durch, wobei das Codebuch wie folgt aussieht:

Fragestellung	Variable	Merkmalsausprägung	Codierung
Was ist der Grund Ihres Aufenthalts?	v_21	Besuch von Freunden und Familie	1
		Tagesausflug	2
		Mehrtägiger Ausflug	3
		Urlaub	4
		Sonstiges	5

Variable:	Beobachtbare oder messbare Merkmale, die verschiedene Werte annehmen können
Merkmalsausprägung:	Werte, die eine Variable annehmen kann
Codierung:	Den Merkmalsausprägungen zugeordnete Ziffer

Pro-Tipp: Daten sollten in einer einheitlichen Struktur abgespeichert werden, sodass bereits am Dateinamen das Datum sowie der Inhalt zu erkennen ist. Die einmal festgelegte Struktur ist schließlich von allen anzuwenden, um versehentliche Dopplungen, Überschreibungen oder Löschungen zu vermeiden.

4. Daten analysieren

Steht nach der Datenvorbereitung die Analyse an, stellt sich die Frage, in wessen Aufgabenbereich dies fällt. Ist die Aufgabe mit interner Expertise zu bewältigen oder bedarf es für zufriedenstellende Ergebnisse externes Know-how? Dies sollte dringend vor Beginn der Analyse geklärt werden, um ressourceneffizient zu arbeiten. Auch die Auswertungsmethode ist im Vorfeld festzulegen. Sind einfache statistische Verfahren wie die deskriptive Statistik, Boxplots oder Korrelationsanalysen ausreichend oder müssen umfassendere Verfahren oder sogar KI-gestützte Methoden wie self-organizing maps (SOM) ausgewählt werden? Das Festlegen der Methoden bildet die Grundlage für die Auswahl des geeigneten Tools. Während Excel und SPSS für umfangreiche statistische Verfahren geeignet sind, können in Python und R auch KI-Analysen programmiert werden. In diesem Zusammenhang sollte auch geklärt werden, wer in welcher Form Zugriff auf die Analysen hat. Ist es sinnvoll die Analyse lokal über einen Rechner laufen zu lassen oder ist eine Cloud-Lösung für kollaboratives Arbeiten vorteilhafter?

Neben der Unterstützung von spezialisierten Unternehmen bieten Kooperationen mit Hochschulen, Verbänden oder anderen Kultureinrichtungen im Bereich der Datenanalyse eine interessante Möglichkeit, geplante Vorhaben in gemeinsamen Projekten zu realisieren. Dabei können alle Beteiligten von der Zusammenarbeit und dem Austausch profitieren.

5. Handlungsempfehlungen ableiten

Nachdem die Daten rechtskonform und zielgerichtet erhoben und ausgewertet worden sind, folgt schließlich die Ableitung der Handlungsempfehlungen. Dazu werden die Ergebnisse aus der Analyse für das weitere Vorgehen interpretiert. Es leiten sich so strategische und operative Empfehlungen ab. Beispielsweise zeigt eine Zielgruppenanalyse mittels einer SOM, dass eine Kulturinstitution insgesamt vier Besuchergruppen aufweist. Während drei der Besuchergruppen bereits als Zielgruppen im Marketingkonzept identifiziert sind, blieb die vierte Gruppe bis dahin unentdeckt und kann nun mit den Interessen in das Marketingkonzept aufgenommen werden. So kann sich die Kultureinrichtung strategisch anpassen, indem in Zukunft ein spezifisches Angebot erstellt wird und entsprechende Werbemaßnahmen ergriffen werden.

Ein gutes Beispiel dafür, dass im Kulturbereich erfolgreich mit verschiedensten Daten gearbeitet werden kann, stellt der seit 2014 veranstaltete Kulturhackathon „Coding da Vinci“ dar. Das von der Kulturstiftung des Bundes geförderte Format bringt Kultur- und Technikwelt zusammen und gemeinsam werden auf Grundlage

offener Kulturdaten unter anderem Apps, Spiele und interaktive Installationen entwickelt. Die bisher hervorgebrachten digitalen Kultur Anwendungen zeigen dabei, dass verschiedenste Bereiche mit Daten erschlossen werden können.

Allgemein ist festzuhalten, dass im Umgang mit Daten der Fokus nicht nur auf dem Ergebnis liegen sollte. Viel bedeutender ist der Weg zum Ziel. Die aus den Datenanalysen gesammelten Erfahrungen sind wesentlich für die erfolgreiche Arbeit. Hinzu kommt der aktive Austausch mit anderen Kultureinrichtungen, um voneinander lernen zu können und Probleme gemeinsam anzugehen.

Kultureinrichtungen als gesellschaftlich anerkannte Einrichtungen sind im gegenwärtigen Diskurs zum Umgang mit Künstlicher Intelligenz prädestiniert dafür, öffentliches Vertrauen auf diesem Themengebiet zu schaffen. Es sind die idealen Orte, um mit dem verantwortungsvollen Datenumgang KI in der Gesellschaft zu etablieren und vermeintliche Vorurteile abzubauen.

Haftungsausschluss: Trotz aller Sorgfalt ändern sich technische Spezifikationen und rechtliche Rahmenbedingungen stetig, weshalb keine Gewähr für die Richtigkeit der Angaben in dieser Handreichung gegeben werden kann.

Diese Handreichung entstand im Rahmen des Projektes „KI in der kulturellen Infrastruktur“, gefördert vom Ministerium für Bildung, Wissenschaft und Kultur des Landes Schleswig-Holstein.

Text: Iwana Schmidt, Nina Thomsen



kultursphäre.sh

Prof. Dr. Tobias Hochscherf FH Kiel |

Fachbereich Medien

Grenzstr. 3 | 24149 Kiel

info@kultursphaere.sh