



Julia Mayer, Martin Memmel

Mit Metadaten durch den urbanen Datendschungel

Verwaltung und Nutzung kommunaler Daten im Forschungsprojekt „Ageing Smart – Räume intelligent gestalten“

In öffentlichen Verwaltungen liegt eine Fülle von Daten vor, die zu zahlreichen Themengebieten wertvolle Informationen über die aktuelle Situation sowie Entwicklungen der Vergangenheit liefern. Wenn solche Daten mit adäquaten Methoden genutzt werden, können sie einen wichtigen Beitrag beim Treffen von raumplanerischen Entscheidungen leisten. Eine datengetriebene Entscheidungsunterstützung kann jedoch maximal so gut sein wie die Daten, die ihr zugrunde liegen. Daher ist es eine wichtige Aufgabe, Personen zu sensibilisieren und zu schulen, die Daten nutzen, verarbeiten, erzeugen oder in ihre Systeme integrieren. Dies betrifft sowohl Mitarbeitende in öffentlichen Verwaltungen als auch Forschende und andere Nutzergruppen. Sie benötigen leicht verständliche und umsetzbare Prozesse zur Dokumentation und Sicherstellung von ausreichender Datenqualität sowie intuitiv nutzbare Werkzeuge, die sie hierbei unterstützen. Eine Schlüsselrolle spielen in diesem Kontext Metadaten, die Informationen über die verwendeten Datensätze bieten. Sie sind bedeutend für die Auffindbarkeit, die Beurteilung der Relevanz, den Umgang und die Nutzung der Daten. Insbesondere der letztgenannte Punkt ist wichtig, wenn die Daten im Rahmen von algorithmischen Entscheidungsunterstützungssystemen verwendet werden. Anhand des Forschungsprojekts Ageing Smart wollen wir einen Ansatz vorstellen, der die Wichtigkeit einer guten Datenverwaltung im allgemeinen Bewusstsein aller Projektbeteiligten verankert und sowohl Bereitstellende als auch Nutzende von kommunalen Daten darin unterstützt.

Das Forschungsprojekt „Ageing Smart – Räume intelligent gestalten“

Das interdisziplinäre Forschungsprojekt „Ageing Smart – Räume intelligent gestalten“ widmet sich den Herausforderungen, die bei Entscheidungen auf kommunaler Ebene in Anbetracht des demografischen Wandels, der Digitalisierung und der räumlichen Polarisierung bewältigt werden müssen. Das Projektkonsortium besteht aus Mitarbeitenden der Fachbereiche Raum- und Umweltplanung, Mathematik und Informatik der RPTU Kaiserslautern-Landau, des Fraunhofer-Instituts für Experimentelles Software Engineering (IESE) und des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) und deckt verschiedene Expertisen und Herangehensweisen ab. Das gemeinsame Ziel ist zum einen, die Bedürfnisse und Wünsche der Kohorte der Baby-Boomer (Geburtsjahrgänge 1955–1969) hinsichtlich ihres bevorstehenden Ruhestands besser zu verstehen. Zum anderen sollen die damit verbundenen planerischen Herausforderungen für Kommunen – unter Berücksichtigung der oben genannten disruptiven Einflüsse – erforscht und mögliche Lösungsansätze aufgezeigt werden.

Die gewonnenen Erkenntnisse gehen in die Entwicklung eines datengestützten Systems ein, das primär kommunal agierenden Personen als Entscheidungshilfe in Planungsprozessen dient. Hierzu wurden sieben Modellkommunen

in unterschiedlichen Bundesländern und Räumen (urban, suburban und rural) ausgewählt, die ihre Daten zur Verfügung stellen. Dort werden zudem Befragungen unter den Ortsansässigen durchgeführt und mit Verwaltungsmitarbeitenden Workshops veranstaltet, bei denen Schwerpunktthemen ausgewählt und Anforderungen erhoben werden. Das Forschungsprojekt wird seit April 2021 über einen Zeitraum von fünf Jahren von der Carl-Zeiss-Stiftung mit rund 4,3 Mio. Euro sowie von der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau mit rund 0,9 Mio. Euro gefördert.

Urbane Daten als Grundlage von Analysen und zur Entscheidungsunterstützung

Weil die Hintergründe, Perspektiven, Interessen und Fragestellungen der Forschenden im Projekt äußerst vielfältig sind, wird von den Kommunen auch eine sehr umfangreiche Auswahl an Daten benötigt. Beispielsweise werden für Erreichbarkeitsanalysen Informationen zu den Straßennetzen, dem ÖPNV, verschiedener Points of Interest, wie etwa Arztpraxen oder der Nahversorgung, benutzt. Für die stadtplanerischen Projektaspekte spielt die Art der Bebauung eine große Rolle. Die Lehrereinheit für Physische Geografie benötigt das Baumkataster und Sensordaten zur Luftqualität und die Mitarbeitenden der Raumökonomie analysieren verschiedenste Finanzdaten.



Um projektintern einen Überblick zu erhalten und die Modellkommunen nicht mit Anfragen zu überfrachten, sind generische Vorgehensweisen notwendig, die die Datenanfragen und -lieferungen, das Datenmanagement, den Zugang und die Verwendung regeln. Hierfür wurde eine Datenkommission aus Projektmitarbeitenden gegründet, die alle datenbezogenen Prozesse in enger Zusammenarbeit mit dem Projektkonsortium festlegt. Ihre Aufgabe ist es, transparente Ziele und Vorgehensweisen zu definieren, um so für Klarheit und Sicherheit beim Umgang mit Daten zu sorgen. Die Vielschichtigkeit der Anforderungen stellt eine große Herausforderung dar, die allerdings auch beispielhaft für die Komplexität bei der Bereitstellung solcher kommunalen Daten außerhalb eines eingeschränkten Projektkontextes ist. Des Weiteren sind adäquate Werkzeuge erforderlich, die diese Prozesse im Speziellen und das Datenmanagement im Allgemeinen unterstützen. Hier liegt der Fokus aktuell auf einem vom DFKI entwickelten Tool zur redaktionellen Bearbeitung von Metadaten.

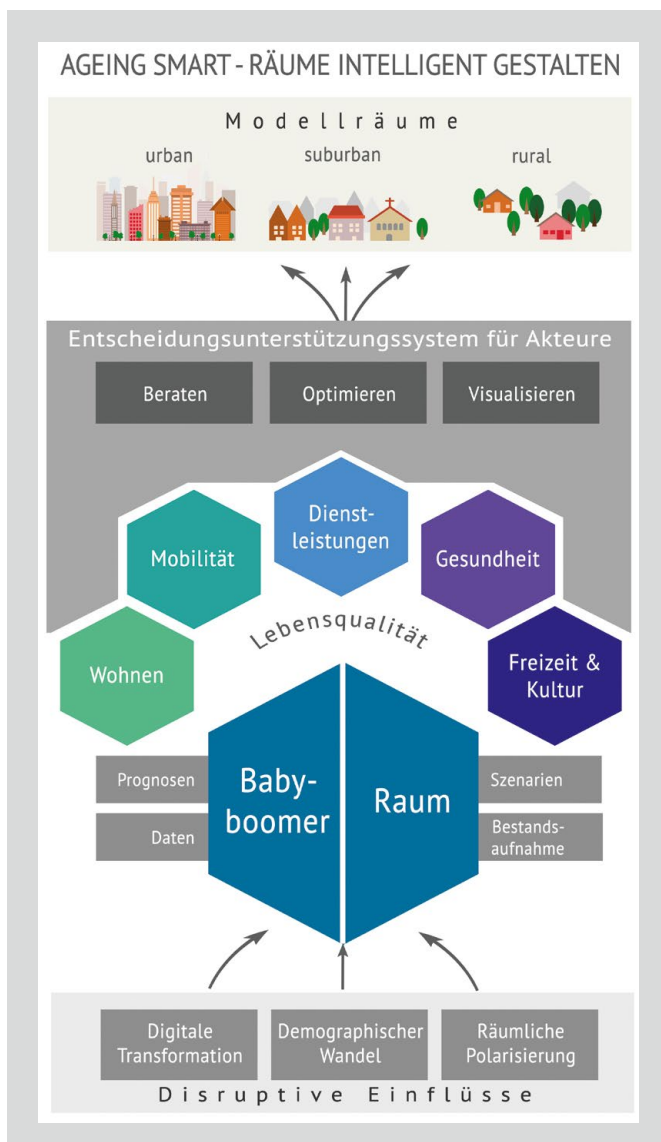


Abb. 1: Projektübersicht (Quelle: www.ageing-smart.de)

Datenqualität und Metadaten

Es gibt zahlreiche Dimensionen, die für die Qualität von Daten von Bedeutung sind. Dazu gehören Vollständigkeit, Fehlerfreiheit, Aktualität, Konsistenz, Transparenz, Vertrauenswürdigkeit, Verlässlichkeit, Genauigkeit, Konformität, Zugänglichkeit, Verfügbarkeit und Verständlichkeit. In der Praxis ist es nicht realistisch, dass seitens der Kommunen bereitgestellte Daten immer alle diese Qualitätsmerkmale erfüllen. Vielmehr ist es gelegentlich unvermeidbar, mit beispielsweise unvollständigen Datensätzen umzugehen oder mit Daten, in denen eine offensichtliche Verzerrung vorliegt.

Umso wichtiger sehen wir an dieser Stelle den Aspekt der Transparenz: Auch, wenn es nicht möglich ist, jederzeit in allen Merkmalen einwandfreie Daten zu nutzen, muss klar und offen kommuniziert werden, wo Probleme vorliegen und wo keine Aussage über die Qualität getroffen werden kann. Des Weiteren braucht es Verantwortliche und einen Prozess zur Freigabe derjenigen Datensätze, die zur Nutzung in einem Entscheidungsunterstützungssystem infrage kommen. Die Verwendung von Metadaten kann an dieser Stelle einen signifikanten Beitrag leisten.

Metadaten sind Informationen über Daten, die einige ihrer Eigenschaften beschreiben und dabei helfen, sie eindeutig zu identifizieren und ihre Integrität und Qualität sicherzustellen. Das geläufigste Beispiel aus der Praxis stellen Bibliothekskataloge dar, mit deren Hilfe man Bücher in einer Sammlung finden kann und die einige Informationen bereitstellen, ohne dass man ein Buch selbst in die Hand nehmen muss. Im Fall von Ageing Smart enthalten die Metadaten zusätzlich zu grundlegenden Informationen, wie Titel, einer Kurzbeschreibung und einer Kontaktperson, beispielsweise auch Hinweise zum Datenschutz, der räumlichen Verortung, zeitliche Aspekte und technische Angaben. Der konkrete Mehrwert der Nutzung von Metadaten im Forschungsprojekt entsteht also auf verschiedene Arten, die durch die Bereitstellung eines redaktionellen Bearbeitungstools noch unterstützt werden:

Kenntnis über die Existenz der Daten

Weil die Metadaten aller im Projekt verwendeten Datensätze in einem zentralen Metadatentool angelegt werden, erhält man einen Überblick, welche Daten im Projekt vorhanden sind. Auch wenn einige Datensätze (etwa aus datenschutzrechtlichen Gründen) nicht auf dem Projektserver abgelegt werden, informiert das Vorhandensein ihrer Metadaten im Metadatenkatalog die Forschenden über ihre Existenz.

Verständnis über die Daten

Schon ein aussagekräftiger Titel und eine gute Kurzbeschreibung können unterstützen, einen Datensatz besser zu verstehen. Alle weiteren Merkmale erlauben ein tieferes Verständnis darüber, was ein Datensatz enthält, ohne dass die Forschenden die Daten selbst einsehen müssten, und helfen somit, die Relevanz eines Datensatzes zu beurteilen.



Kompatibilität verschiedener Ressourcen

Unter anderem kann es bei der räumlichen Abdeckung zu Inkompatibilität von Datensätzen kommen. Wenn man beispielsweise Bevölkerungsdaten auf Stadtteilebene, aber Daten über Schulen auf Schulbezirksebene hat, kann man diese nicht ohne Weiteres kombinieren. Metadaten liefern also Informationen zum Austausch und der (Wieder-)Verwendung von Datensätzen aus unterschiedlichen Kontexten und Anwendungen und können diese somit erleichtern.

Auffinden und Sortieren der Daten

Es kann sehr erkenntnisreich sein, in den vorhandenen Metadaten zu stöbern. Man kann sich beispielsweise alle vorhandenen Datensätze aus einer bestimmten Modellkommune, zu einem oder mehreren Stichworten, Kategorien oder Autoren ansehen. Hierdurch können bisher unbekannte Datensätze entdeckt, ungeeignete Datensätze aussortiert oder Prioritätenlisten erstellt werden. Weiterhin können Metadaten helfen, die am besten geeigneten Versionen eines Datensatzes bereitzustellen. Darüber hinaus bieten die gespeicherten Metadaten beispielsweise Informationen darüber, wer als Ansprechperson fungiert, um etwa Auswertungen durchzuführen oder den Zugriff auf die Daten zu klären.

Ort der Ablage

Die Metadaten enthalten auch Informationen zum Ort der Ablage der Daten. Die Forschenden wissen also unmittelbar, ob sie auf dem Projektserver liegen, auf einem Open-Data-Portal einer der Modellkommunen oder gar im Fall von sehr sensiblen Daten auf einem Rechner ohne Internetzugang. Hinterlegte URLs ermöglichen gegebenenfalls einen direkten Zugriff auf die Datensätze für (autorisierte) Personen.

Glaubwürdigkeit der Quelle

Datensätze, die im Entscheidungsunterstützungssystem genutzt werden, müssen in einem wohldefinierten Prozess dafür freigegeben werden. Der Ursprung und somit ihre Vertrauenswürdigkeit ist hier zusammen mit einer nachvollziehbaren Dokumentation von etwaigen Änderungen ein erheblicher Faktor.

Organisation

Je mehr Datensätze vorhanden sind, desto hilfreicher kann die Möglichkeit einer Sortierung oder einer Filterfunktion nach unterschiedlichen Metadatenmerkmalen sein.

Archivierung, Pflege und Erhalt

Zum Ende des Forschungsprojekts wird der Projektserver, auf dem fast alle Datensätze gespeichert sind, nicht weiter zur Verfügung stehen. Mithilfe der Metadaten kann dann überblickt werden, welche Daten endgültig gelöscht oder archiviert werden müssen.

Um von den möglichen Mehrwerten von Metadaten profitieren zu können, ist ein fachkundiger Umgang notwendig. Dieser umfasst die Verwendung eines entsprechenden Vokabulars mit ausreichender Ausdrucksmöglichkeit. In Ageing Smart wird hierfür der Metadatenstandard DCAT-AP.de genutzt.

Der Metadaten-Standard DCAT-AP.de

DCAT-AP (Data Catalogue Vocabulary – Application Profile) wurde auf Initiative der Europäischen Kommission entwickelt. Es ist ein Standardvokabular zur Beschreibung von Datenkatalogen und soll einen Austausch von offenen allgemeinen Verwaltungsdaten vereinfachen, indem es Bezeichnungen vereinheitlicht. Ein Datensatztitel ist hier also immer ein „title“ und niemals eine „headline“ oder gar „überschr“. DCAT-AP.de ist eine deutsche Erweiterung von DCAT-AP und wurde vom IT-Planungsrat als formaler Austauschstandard für Verwaltungsdaten festgelegt.

Für die redaktionelle Bearbeitung der Metadaten im Projekt Ageing Smart wird dieser Standard eingehalten. Zum einen ist er für alle Zwecke im Projekt ausdrucksstark genug, zum anderen ermöglicht er eine spätere Integration oder Übertragbarkeit einiger oder aller Metadatenätze oder -kataloge. Es wird jedoch nur eine Teilmenge aller Attribute aus DCAT-AP.de erfasst, weil nicht alle relevant oder verfügbar sind.

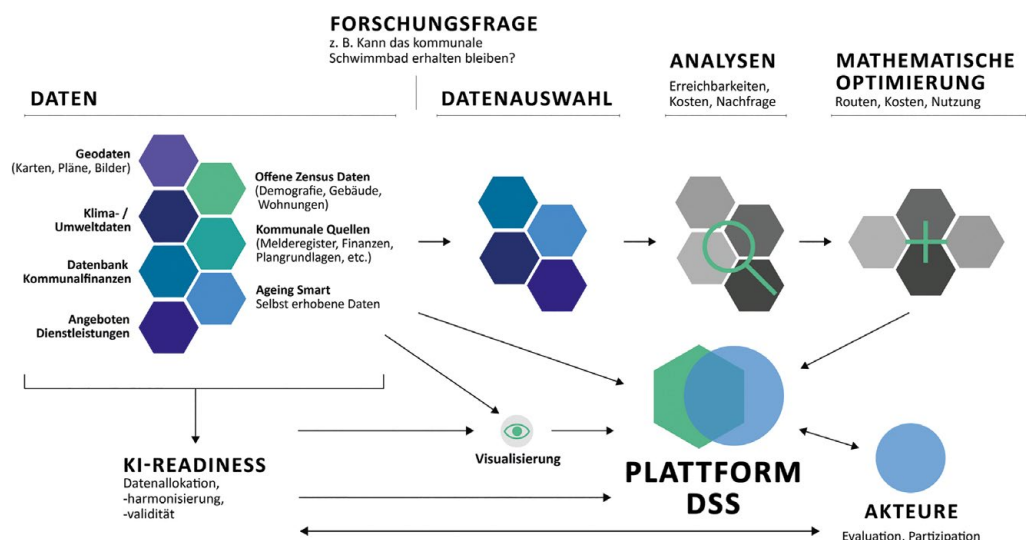


Abb. 2: Projektverlauf (Quelle: www.ageing-smart.de)



Redaktionelle Bearbeitung der Metadaten

Um die Metadaten im Projekt zu erfassen und zu verwalten, entwickelt das DFKI ein webbasiertes Tool, das zusammen mit den Forschenden sukzessive erprobt und an ihre Bedürfnisse angepasst wird. Eng begleitet wird diese Entwicklung von der projektinternen Datenkommission. Hierbei müssen die Entwicklerinnen und Entwickler berücksichtigen, dass das Bearbeiten der Metadaten möglichst niedrigschwellig erfolgen kann. Weil sehr viele Merkmale angegeben werden könnten, von denen manche nicht trivial sind, kann es leicht zu einer Überforderung der Eintragenden kommen. Hinzu kommt, dass es oft keinen unmittelbar erkennbaren Mehrwert durch das Anlegen und Editieren der Metadatenätze gibt. Dieser ergibt sich vielfach erst im weiteren Lauf des Projekts und vielleicht auch nicht für die Person selbst, sondern erst, wenn Transparenz, Erklärbarkeit und Nachvollziehbarkeit von Ergebnissen und Verfahren eine Rolle spielen. Daher ist es umso wichtiger, die Forschenden für die Relevanz von Metadaten zu sensibilisieren und ihre Kompetenzen im Bereich Datenmanagement zu stärken.

Von technischer Seite wurden mehrere Ideen umgesetzt, die eine Nutzung des Tools vereinfachen sollen: Begriffe, die für die meisten Menschen vermutlich unverständlich sind, werden auf der Nutzeroberfläche durch leichter greifbare Begriffe ersetzt. So wurde beispielsweise aus „Distribution“ die gewöhnlichere und in DCAT anders verwendete Bezeichnung „Datensatz“. Die interne Logik und die Bezeichnungen der einzelnen Spezifikation von DCAT-AP.de bleibt im Hintergrund jedoch erhalten, sodass ein standardkonformer Datenaustausch weiterhin möglich bleibt. Darüber hinaus gibt es ausführliche Hilfetexte mit einfachen und sprechenden Beispielen. Durch eine möglichst übersichtliche Anordnung der Merkmale in zusammenhängende Themenblöcke mit grafischen Elementen, die diese gedanklichen Blöcke unterstreichen, soll eine einfachere Eingabe ermöglicht und Klarheit gewährleistet werden. Nicht zuletzt werden die in der Spezifikation als obligatorisch definierten Merkmale an den Anfang gestellt. Nutzende können also allein durch Eingabe von nur wenigen Merkmalen einen neuen Metadatenatz anlegen. Die empfohlenen und optionalen Felder können bei Bedarf auch später und von anderen Projektbeteiligten nachgetragen werden.

Der Ageing-Smart-spezifische Metadatenkatalog wird zunächst nur intern Verwendung finden. Durch die Nutzung von DCAT-AP.de ist allerdings eine Übertragung in kommunale Metadatenportale möglich. Das entwickelte Tool kann in Zukunft auch in anderen Forschungsprojekten oder in kommunalen Verwaltungen zum Einsatz kommen.

Fazit und Ausblick

Aktuell werden die Daten aus den Modellkommunen eingesammelt. Die Prozesse und Werkzeuge werden dabei erprobt und in enger Zusammenarbeit mit den Akteurinnen und Akteuren weiterentwickelt. Gleichzeitig finden durch Workshops eine Sensibilisierung und ein Kompetenzaufbau bei den Forschenden statt. Dabei muss man bemerken, dass die Entwickelnden und die Datenkommission eine Gratwanderung durchlaufen. Man muss Prozesse etablieren und Strukturen vorgeben, die es erlauben, einen Überblick über die Datensätze und ihre Metadatenätze zu behalten. Gleichzeitig darf man die Projektbeteiligten nicht durch organisatorische Aufgaben überlasten oder sogar überfordern, indem man zeitlich exzessive und inhaltlich zu komplexe Vorgehensweisen erwartet. Weitere notwendige Schritte zur Verbesserung des Metadaten-Tools liegen daher in einer weiteren Vereinfachung der Nutzung.

Bei bestimmten Merkmalen kann eine Autofill-Funktion durchaus hilfreich sein. Eine Plausibilitätsprüfung kann zukünftig an bestimmten Stellen verhindern, dass falsche Metadaten ihren Weg in die Metadatenbank finden. Darüber hinaus muss ein Prozess konzipiert und integriert werden, der die Freigabe von Datensätzen zur Verwendung im Entscheidungsunterstützungssystem regelt. Neben dem reinen Sammeln der Daten mit Fokus auf Übersicht und Auffindbarkeit liegt in der nächsten Phase der Schwerpunkt auf der Frage der Nutzung: Wie verweist man auf konkret verwendete Daten? Wie stellt man Eigenschaften dar, die für Transparenz und Erklärungsfähigkeit relevant sind?

Unser Ziel ist es, dass die im Projekt erarbeiteten Prozesse und Werkzeuge auch außerhalb von Ageing Smart einen Beitrag zum besseren Umgang mit Daten leisten – sowohl bei der Bereitstellung als auch bei der Nutzung der Daten.



Julia Mayer

wissenschaftliche Mitarbeiterin zum Smart-City Living Lab des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI), Kaiserslautern



Dr. Martin Memmel

Wissenschaftler und Berater, Leitung des Smart-City Living Lab im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI), Kaiserslautern

Quellen/Links:

<https://www.ageing-smart.de>

Normentwurf für qualitativ hochwertige Daten und Metadaten: <https://www.nqdm-projekt.de>

<https://www.dcat-ap.de>

<https://scll.dfki.de>