

Bensberg, Frank; Jede, Andreas

Article — Published Version

Open Data im Spiegel der wissenschaftlichen Literatur – Eine quantitative Literaturanalyse von 2005–2024

HMD Praxis der Wirtschaftsinformatik

Provided in Cooperation with:

Springer Nature

Suggested Citation: Bensberg, Frank; Jede, Andreas (2025) : Open Data im Spiegel der wissenschaftlichen Literatur – Eine quantitative Literaturanalyse von 2005–2024, HMD Praxis der Wirtschaftsinformatik, ISSN 2198-2775, Springer Fachmedien Wiesbaden GmbH, Wiesbaden, Vol. 62, Iss. 5, pp. 1051-1071,
<https://doi.org/10.1365/s40702-025-01195-3>

This Version is available at:

<https://hdl.handle.net/10419/330727>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/deed.de>



Open Data im Spiegel der wissenschaftlichen Literatur – Eine quantitative Literaturanalyse von 2005–2024

Frank Bensberg · Andreas Jede

Eingegangen: 13. Februar 2025 / Angenommen: 21. Mai 2025 / Online publiziert: 25. Juni 2025
© The Author(s) 2025

Zusammenfassung In den letzten zwanzig Jahren hat sich Open Data zu einem zentralen Forschungsbereich entwickelt, der eine Vielzahl von wissenschaftlichen Disziplinen berührt. Durch Analyse von 13.538 wissenschaftlichen Literaturquellen aus der EBSCOhost-Datenbank werden die Eigenschaften dieses Forschungsfelds transparent gemacht. Dabei werden zentrale Themenfelder, theoretische Ansätze, eingesetzte Forschungsmethoden und Entwicklungstrends identifiziert. Um dies zu erreichen, wurden unterschiedliche Verfahren und Anwendungssysteme zur Extraktion und Analyse bibliografischer Daten angewendet. Die Untersuchung zeigt, dass das Forschungsgebiet Open Data durch eine ausgeprägte thematische und methodologische Vielfalt gekennzeichnet ist. Thematisch erstreckt sich die Forschung von Fragen der staatlichen Transparenz über wirtschaftliche Anwendungen bis hin zu informationstechnischen Innovationen bei der Strukturierung und Integration offener Datenbestände. Dabei kommen Theorien aus unterschiedlichen Disziplinen zur Anwendung, wobei neben Ökonomie und Informatik auch psychologische und politikwissenschaftliche Theorieansätze zu verzeichnen sind. Methodologisch finden sowohl qualitative als auch quantitative Ansätze Anwendung, die von Fallstudien über Experimente bis hin zur Prototypenentwicklung reichen. Die Literaturanalyse deutet darauf hin, dass das Forschungsfeld Open Data künftig durch moderne Analysemethoden – insbesondere aus dem Umfeld der generativen KI – stimuliert wird.

Schlüsselwörter Open Data · Forschungsthemen · Forschungsmethoden · Literaturanalyse · Interdisziplinarität

✉ Frank Bensberg · Andreas Jede
Hochschule Osnabrück, Caprivistraße 30a, 49076 Osnabrück, Deutschland
E-Mail: f.bensberg@hs-osnabrueck.de

Andreas Jede
E-Mail: a.jede@hs-osnabrueck.de

Open Data in the Scientific Literature—A Quantitative Literature Analysis from 2005–2024

Abstract Over the last twenty years, open data has developed into a central field of research that touches on a variety of scientific disciplines. By analyzing 13,538 scientific literature sources from the EBSCOhost database, the characteristics of this field of research are made transparent. Central topics, theoretical approaches, research methods used and development trends are identified. To achieve this, various methods and application systems were used to extract and analyze bibliographic data. The study shows that the research field of open data is characterized by a pronounced thematic and methodological diversity. Thematically, the research ranges from government transparency and economic applications to information technology innovations in the structuring and integration of open data sets. Theories from various disciplines are applied, with psychological and political science theories being used alongside theories from economics and computer science. Methodologically, both qualitative and quantitative approaches are used, ranging from case studies and experiments to prototype development. The literature analysis indicates that the research field of open data will be stimulated in the future by modern analysis methods—particularly from the field of generative AI.

Keywords Open Data · Research Topics · Research Methods · Literature Review · Interdisciplinarity

1 Problemstellung

Open Data hat sich in den letzten zwanzig Jahren zu einem bedeutenden und dynamischen Forschungsfeld entwickelt, das zahlreiche Disziplinen betrifft und mittlerweile über eine reichhaltige Publikationslandschaft verfügt. Unbestritten sind die Vorteile von Open Data, die sich im Wesentlichen auf Faktoren wie effiziente Ressourcennutzung, erhöhte Transparenz, valide Reproduzierbarkeit aber auch Kollaboration und Innovationsförderung in Wissenschaft und Praxis beziehen.

Um Erkenntnisse über die Charakteristika dieses Forschungsfelds zu gewinnen und den Akteuren der wissenschaftlichen Community eine Orientierung in diesem Forschungsfeld zu bieten, sind mittlerweile zahlreiche Literaturanalysen durchgeführt worden. Allerdings weisen diese häufig die Merkmale auf, dass spezifische Themenfelder im Analysemittelpunkt stehen und eine kleinzahlige Quellenauswahl erfolgt (z. B. Hyseni und Bexheti 2024; Wirtz et al. 2022), um den Literaturbestand mit den traditionellen Techniken der Literaturanalyse handhabbar zu halten. Großzahlige, forschungsfeldübergreifende Studien zu Open Data, die den aktuellen Stand reflektieren, sind indes nur in geringem Ausmaß verfügbar. Exemplarisch sind hier zwei rezente Studien anzuführen. Die Studie von Catone (2023) untersucht 3110 Publikationen zu Open Data und Open Government Data im Zeitraum 2013–2022 aus dem Web of Science (WoS), während Shettar und Hadagali (2023) 3252 Literaturquellen zu Open Data aus der Scopus-Datenbank bis zum Jahr 2021 analysieren. Die Studie von Shettar und Hadagali (2023) verfolgt die Zielsetzung, etablierte Pu-

blikationsmuster, Forschungsnetzwerke und Institutionen zu identifizieren, die das Forschungsfeld prägen. Sie kommen zu dem Ergebnis, dass das Forschungsfeld überwiegend durch Konferenzbeiträge als Publikationstyp repräsentiert wird und die USA im Ländervergleich dominieren, während aus institutioneller Perspektive die Technische Universität in Delft (Niederlande) eine führende Rolle einnimmt. Dagegen untersucht Catone (2023) die Entwicklung von Forschungsthemen zu Open Data mit dem Fokus auf die Sozial- und Geisteswissenschaften und identifiziert auf Grundlage der in den Literaturdaten enthaltenen Schlagwörter drei unterschiedliche Gruppen von Publikationen sowie die hierfür charakteristischen Themenfelder. Eine erste Gruppe von Beiträgen fokussiert *grundlegende Themen* von Open Data, wie die gemeinsame Nutzung von Daten, Transparenz, Open Government und Zugänglichkeit (accessibility). Ein zweites Segment wird durch die Ausrichtung auf *methodische Anforderungen* im Kontext sozial- und geisteswissenschaftlicher Forschungsdesigns gebildet, etwa durch Fragestellungen der Replizierbarkeit, Reproduzierbarkeit und Evaluation. Schließlich konzentriert sich eine dritte Gruppe von Publikationen auf *Anwendungsbereiche* von Open Data im gesellschaftlichen Kontext (z. B. Klimawandel, Smart Cities, COVID-19).

Um einen Einblick in den aktuellen Stand und die Entwicklung des Forschungsfelds Open Data zu gewinnen, ist aus der interdisziplinären Perspektive der Wirtschaftsinformatik indes eine disziplinübergreifende Analyse als erforderlich zu erachten, mit der die Forschungsthemen und die theoretisch-methodische Verankerung des Forschungsfelds transparent gemacht werden kann. Dieser Forschungsbedarf motiviert die Durchführung einer aktuellen Literaturanalyse, die als Datengrundlage die Literaturdatenbank EBSCOhost verwendet. Diese verzeichnet zu Beginn des Jahres 2025 bereits mehr als 10.000 englischsprachige, wissenschaftliche Quellen mit dem Titelbegriff *Open Data*. Um die Charakteristika des Forschungsfelds transparent zu machen, sollen mit der Literaturanalyse die folgenden Forschungsfragen beantwortet werden:

1. Welche Themen prägen die Open Data-Forschungslandschaft aus inhaltlicher Perspektive?
2. Welche theoretischen und forschungsmethodischen Bezüge werden in der Open Data-Literatur expliziert?
3. Welche thematischen Entwicklungen und Trends können in dem Forschungsfeld Open Data identifiziert werden?

Mit der Beantwortung dieser Forschungsfragen soll ein Beitrag zur Rationalitätssicherung von Forschungsprozessen (Auth et al. 2019) geleistet werden. So kann eine umfassende Einschätzung der thematischen und forschungsmethodischen Eigenschaften eines Forschungsfelds die informatorische Basis bereitstellen, um neue Forschungsideen zu generieren, bestehende Forschungsaktivitäten kritisch zu hinterfragen, und die Konfiguration von integrativen Forschungsdesigns in interdisziplinären Kontexten zu unterstützen. Aus wissenschaftsökonomischer Perspektive entsteht somit auch das Potenzial, den Einsatz knapper Forschungsressourcen effizienter zu lenken. Im Folgenden ist zunächst auf die Grundlagen und die Methodik zur Beantwortung der angeführten Forschungsfragen einzugehen.

2 Grundlagen und Methodik

Open Data bezeichnet Daten, die von jedermann frei genutzt, weiterverwendet und geteilt werden können, vorbehaltlich höchstens der Anforderung nach Quellenangabe und Weitergabe unter gleichen Bedingungen (Lucke und Geiger 2010). Die Idee hinter Open Data ist, Transparenz und Innovation zu fördern, indem man den Zugang zu wichtigen Informationen erleichtert und so die Zusammenarbeit und das Teilen von Wissen unterstützt (Catone 2023). Open Data findet Anwendung in zahlreichen Bereichen wie Wissenschaft, Verwaltung, Wirtschaft und im privaten Umfeld. Ein zentrales Ziel von Open Data ist es, Forschung und Entwicklung zu beschleunigen, indem redundante Datenerhebungen vermieden und vorhandene Datensätze für neue Analysen genutzt werden (Weerakkody et al. 2017). Zudem tragen offene Daten zur Demokratisierung von Wissen in Wissenschaft und Praxis bei, indem sie verhindern, dass Informationen in isolierten Silos verbleiben. Daraus ergeben sich neben den in Abschn. 1 genannten Vorteilen weitere indirekte Effekte, wie eine Beschleunigung des wissenschaftlichen Fortschritts, eine höhere Sichtbarkeit von Forschung oder auch optimierte Lern- und Lehrmöglichkeiten. Obgleich frühere Wachstumsprognosen von Open Data entlang des gesamten Spektrums vermutlich zu optimistisch waren, kristallisieren sich derzeit gewisse Schwerpunkte heraus, die mit unverändert hoher Dynamik zum Reifeprozess beitragen.

Zur quantitativen Literaturanalyse des Forschungsfelds Open Data wird als empirische Basis ein Datensatz zugrundegelegt, der mithilfe der Literaturdatenbank EBSCOhost generiert wurde. Dieser Datensatz enthält die bibliographischen Daten für 13.538 Literaturreferenzen, deren Titel den Begriff „Open Data“ aufweist. Die Suchabfrage wurde am 30.01.2025 ausgeführt, wobei nach englischsprachigen, wissenschaftlichen Literaturquellen gefiltert wurde. Die Begrenzung auf englischsprachige Literatur findet statt, um einen globalen Fokus sowie eine breite Datenbasis sicherzustellen. Da Englisch als internationale Wissenschaftssprache in der Informatik gebräuchlich ist, kann somit potenziell ein weitreichender Überblick über den Stand der Technik als auch über aktuelle Trends generiert werden. Zur Verarbeitung dieser Literaturdaten, die im RIS-Format vorliegen, wird auf unterschiedliche Me-

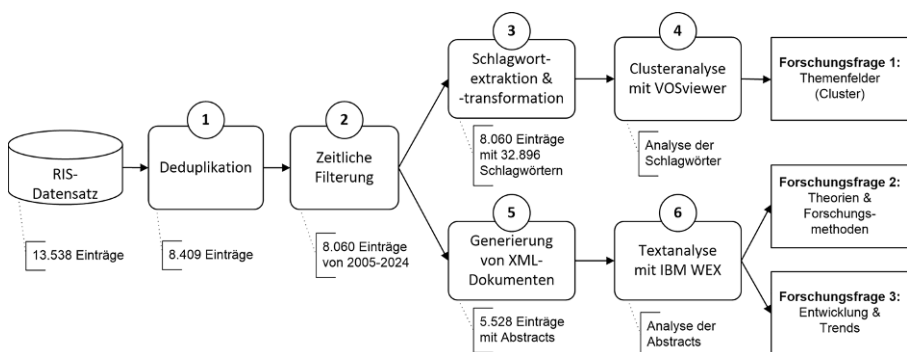


Abb. 1 Analyseprozess im Überblick

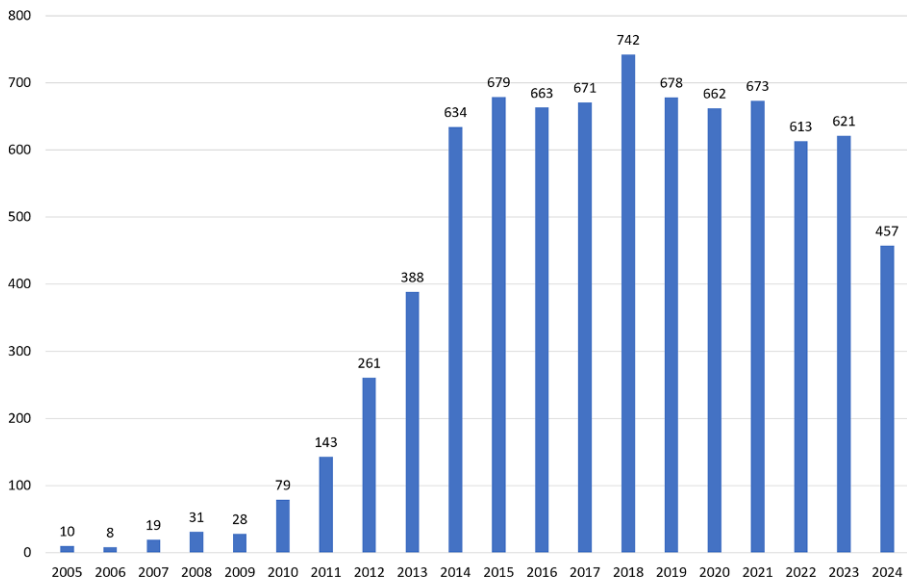


Abb. 2 Anzahl der wissenschaftlichen Publikationen im Forschungsfeld Open Data von 2005–2024 ($n = 8060$)

thoden und Softwaresysteme zurückgegriffen. Der Analyseprozess wird in Abb. 1 dargestellt und im Folgenden schrittweise erläutert.

Da RIS-Daten aus Literaturdatenbanken häufig Duplikate aufweisen, erfolgt in Schritt ① eine Deduplikation. Zu diesem Zweck wird der Online-Dienst *Deduplicator* eingesetzt, der speziell für die Erkennung von Duplikaten in Literaturbeständen entwickelt wurde (Forbes et al. 2024). Mithilfe dieses Dienstes sind 37,9 % der Einträge ($n = 5129$) als Duplikate erkannt und aus dem Datenbestand entfernt worden. Im Anschluss erfolgt in Schritt ② eine zeitliche Filterung der RIS-Daten anhand des Publikationsdatums für den Zeitraum 2005–2024, sodass 8060 Einträge zur weiteren Analyse zur Verfügung stehen. Die Anzahl der Publikationen nach Publikationsjahr zeigt Abb. 2. Die Festlegung des Zeithorizonts wird damit begründet, dass vor dem Jahr 2005 wissenschaftliche Publikationen mit dem inhaltlichen Fokus auf Open Data nur in einstelligem Umfang verfügbar sind (z. B. $n = 2$ Publikationen in 2004), sodass das Jahr 2005 mit $n = 10$ Publikationen als Diffusionsbeginn angesehen wird.

Zur Beantwortung der drei Forschungsfragen ist der Analyseprozess anschließend ausdifferenziert worden. Um die Themen des Forschungsfelds zu untersuchen (Forschungsfrage 1), wird auf die im RIS-Datensatz enthaltenen *Schlagwörter* zurückgegriffen. Dies erfolgt in Anlehnung an die Studie von Yoon et al. 2018, die die Themenstruktur des Forschungsfelds Internet of Things (IoT) durch Segmentierung der Schlagwörter von 12.600 Literaturreferenzen untersuchen. Indizierte Schlagwörter werden dabei als relevante Informationen angesehen, die die grundlegenden Konzepte eines wissenschaftlichen Beitrags reflektieren und von den Autoren meist sorgfältig kuratiert werden. Die Schlagwörter werden in Schritt ③ extrahiert und mithilfe regulärer Ausdrücke transformiert. Dabei werden z. B. die Schreib-

weisen frequenter Schlagwörter vereinheitlicht (z.B. Smart Cities – Smart City), überflüssige Leerzeichen entfernt, sowie Ausdrücke von Sektoren und Fachdisziplinen vereinheitlicht bzw. generalisiert (z.B. electrical & civil engineering – engineering). Anschließend wird in Schritt ④ das Werkzeug VOSviewer eingesetzt, das die Analyse von Kookkurrenzen von Schlagwörtern gestattet (Van Eck und Waltman 2010). Zu diesem Zweck wird das gemeinsame Auftreten von Schlüsselwörtern in den einzelnen Literaturdatensätzen ermittelt und mithilfe eines Ähnlichkeitsmaßes quantifiziert. Darauf aufbauend werden sämtliche Begriffe in einem zweidimensionalen Graphen angeordnet und disjunkten Clustern zugeordnet, sodass thematische Zusammenhänge zwischen den Konzepten erkennbar werden.

Um die Forschungsfragen 2 und 3 zu beantworten, werden indes die Abstracts der Literaturquellen herangezogen, die mithilfe des Text Mining-Systems IBM Watson Explorer (IBM WEX) analysiert werden. Zu diesem Zweck sind die RIS-Daten in XML-Dokumente zu transformieren. Da knapp ein Drittel der Literaturreferenzen über keinerlei Abstract verfügen, entsteht durch Schritt ⑤ ein Korpus mit 5528 XML-Dokumenten. Zur Untersuchung werden in Schritt ⑥ folgende Funktionalitäten von IBM WEX eingesetzt (Bensberg et al. 2019):

- Der Literaturkorpus wird durch IBM WEX linguistisch vorverarbeitet, indem Textelemente in ihre Grundform überführt werden (Stemming), die Wortart einzelner Wörter bestimmt wird (Part of Speech-Tagging, POS-Tagging) und eine Kompositazerlegung erfolgt.
- Dokumente des Literaturkorpus können mithilfe von *Abfragen* selektiert werden, wobei gezielt nach unterschiedliche Wortarten (z.B. Substantive, Verben) und Satzteilen (z.B. Substantivfolgen, Adjektiv-Substantiv-Kombinationen) gesucht werden kann. Diese Funktionalität ist relevant, da wissenschaftliche Fachterminologien in der Regel durch Substantive bzw. komplexere Substantivkonstruktionen gebildet werden.
- Mithilfe einer *Zeitreihenanalyse* kann die Fragestellung beantwortet werden, wie sich themenspezifische Publikationsaufkommen im Zeitablauf entwickelt haben. Dabei kann nach beliebigen Schlagwörtern bzw. Wortarten gefiltert werden, sodass die Diffusion einzelner Konzepte im Zeitablauf transparent wird.

Die Ergebnisse aus der Umsetzung des skizzierten Analyseprozesses werden im Folgenden dargestellt.

3 Ergebnisse der Literaturanalyse

3.1 Thematische Struktur der Open Data-Forschungslandschaft

Zur Beantwortung der ersten Forschungsfrage ist die thematische Struktur der Open Data-Forschungslandschaft transparent zu machen. Zu diesem Zweck werden die Ergebnisse der Clusteranalyse erörtert, die mit dem Werkzeug VOSviewer durchgeführt wurde, welches speziell für die Analyse und Visualisierung von bibliometrischen Netzwerken entwickelt wurde. Nach dem Datenimport werden in einem ersten Schritt die Schlagwörter aus den Literaturdaten durch Vektoren repräsen-

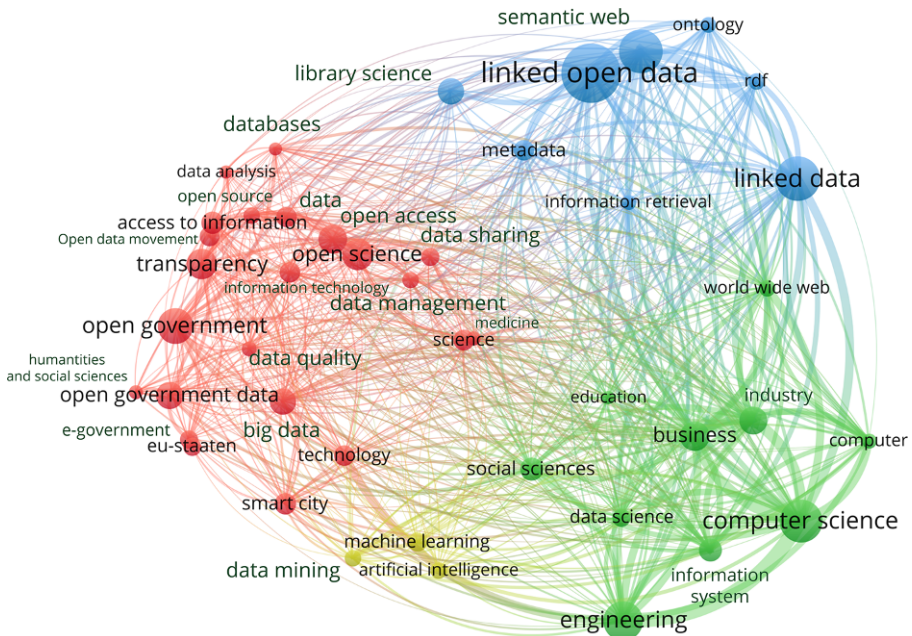


Abb. 3 Themenfelder der Open Data-Forschung. (Clusteranalyse modifiziert übernommen aus VOS-viewer)

tiert. Danach werden Ähnlichkeitsmessungen basierend auf der Vektorrepräsentation durchgeführt, wobei die Stärke der Kookkurrenz, also dem gemeinsamen Auftreten der Begriffe, entscheidend ist. Im dritten Schritt wird die Clusterbildung mit dem VOS-Algorithmus (Van Eck und Waltman 2007) vorgenommen, welcher einzelne Schlagwörter in einem multidimensionalen Raum so anordnet, dass oft zusammenhängende Schlagwörter innerhalb eines Netzwerks enger beieinander liegen. Daraus resultieren Schlagwortgruppen, die auch als thematische Cluster bezeichnet werden.

Abb. 3 zeigt die Ergebnisse der Clusteranalyse als Netzgraphen, dessen Knoten durch frequente Schlagwörter gebildet werden. Dabei sind sämtliche Schlagwörter selektiert worden, die mindestens 50 mal in der Datenbasis auftreten. Das Schlagwort *Open Data* ist vor der Clusteranalyse aus dem Datensatz entfernt worden, da dieses in nahezu sämtlichen Datensätzen auftritt und somit keinen Informationsbeitrag für die Segmentierung liefert. Die ungerichteten Kanten des Graphen zeigen die Verbindungen zwischen diesen Schlagwörtern, d. h., wie häufig sie gemeinsam in den betrachteten Datensätzen vorkommen. Diese Darstellung wird durch Tab. 1 ergänzt, welche die vier Cluster mit deren Schlagwörtern darstellt. Dabei steht jedes Cluster für einen thematischen Bereich innerhalb des Forschungsfelds, der durch eine Clusterbezeichnung prägnant benannt wird.

Der erste Cluster *Access & Transparency* umfasst Fachbegriffe, die sich generell an den Themen Daten und Offenheit orientieren, wobei aus technischer Perspektive insbesondere die Handhabung, Qualitätssicherung und Analyse von Daten im Mittelpunkt stehen. Dies wird durch die Themenfelder Big Data, Datenban-

Tab. 1 Ergebnisse der Clusteranalyse

Nr. u. Bezeichnung des Clusters	Anzahl Schlagwörter	Schlagwörter
1 Access & Transparency	23	Access to information, big data, data, data analysis, data management, data quality, data sharing, databases, e-government, eu-staaten, human and social science, information technology, medicine, open access, open data movement, open government, open government data, open science, open source, science, smart city, technology, transparency
2 Socioeconomic Practice	10	Business, computer, computer science, data science, education, engineering, industry, information system, social sciences, world wide web
3 Structure & Retrieval	8	Information retrieval, library science, linked data, linked open data, metadata, ontology, rdf, semantic web
4 Artificial Intelligence	3	Artificial intelligence, data mining, machine learning

ken, Datenanalyse, Datenmanagement und Datenaustausch (data sharing) nahegelegt. Die Offenheit und Transparenz wird durch die Ansammlung von Kernbegriffen der Open-X-Bewegung (Open Access, Open Data Movement, Open Government (Data), Open Science und Open Source) substantiiert. Außerdem weist der Begriff *EU-Staaten* darauf hin, dass eine geografische oder politische Kategorie in diesem Cluster verankert ist. Neben e-Government besitzt dieser Cluster weitere interdisziplinäre Praxisbezüge, indem auf Smart Cities und Medizin verwiesen wird.

Das zweite Segment *Socioeconomic Practice* artikuliert Begriffe, die auf einen starken Bezug zur praktischen Anwendung von Informationssystemen im wirtschaftlichen und industriellen Bereich deuten: die Begriffe Computer, Informatik (Computer Science), Datenwissenschaft (Data Science) und Informationssysteme (Information System) konzentrieren sich auf technische und wissenschaftliche Themenfelder der Datenverarbeitung, wobei Wirtschaft (Business) und Industrie (Industry) wirtschaftliche Open Data-Anwendungen und die Relevanz dieser Technologie in industriellen Kontexten umfasst. Ferner zeigen die Begriffe Bildung (Education) und Sozialwissenschaften (Social Sciences) an, dass dieser Cluster auch die Rolle der Technologie im Bildungsbereich sowie in sozialwissenschaftlichen Disziplinen behandelt, sodass eine thematische Verbindung zur eingangs skizzierten Studie von Catone (2023) hergestellt wird. Durch den Begriff World Wide Web (WWW) wird der Bezug zur globalen digitalen Vernetzung aufgezeigt, die eine wesentliche Rolle bei der Verbreitung und Anwendung von Open Data-Technologien spielt.

Das dritte Cluster *Structure & Retrieval* verfügt über einen klaren Fokus auf die Organisation, Strukturierung und Nutzung von Open Data. So identifizieren die Begriffe Linked Data, Linked Open Data, Metadata und Resource Description Framework (RDF) solche Konzepte und Technologien, die zur Strukturierung und Verknüpfung von Daten im WWW verwendet werden. Dieses Instrumentarium ist für die Gestaltung eines semantisch angereicherten und interoperablen Datennetzes notwendig. Dabei wird insbesondere auch der Aspekt der Wissensrepräsentation berücksichtigt, die mithilfe von Ontologien und Techniken des Semantic Web realisiert werden kann. Aus fachdisziplinärer Perspektive weist dieses Segment einen deutlichen Bezug zu den Informations- und Bibliothekswissenschaften auf (Information Retrieval, Library Sciences), welche sich traditionell mit der Sammlung, Organisation und Zugänglichkeit von Informationen befassen. Cluster 3 konzentriert sich somit primär auf Technologien und Methoden zur besseren Organisation und Verknüpfung von Informationen im digitalen Raum, um den Zugang und die Nutzung zu verbessern.

Das letzte Segment *Artificial Intelligence* thematisiert fortschrittliche Techniken zur Datenanalyse: so umfasst das Instrumentarium der Künstlichen Intelligenz (Artificial Intelligence) solche Verfahren, die es Computern ermöglichen, Aufgaben zu erfüllen, die normalerweise menschliche Intelligenz erfordern. Diese Verfahren werden aus methodischer Perspektive häufig als maschinelles Lernen (Machine Learning) bezeichnet und können mithilfe von Data Mining-Systemen implementiert werden, die üblicherweise auch die notwendigen Werkzeuge zur Datenvorbereitung und -nachbereitung umfassen sowie die Möglichkeit zur Automatisierung von Analyseprozessen liefern.

Insgesamt verdeutlichen die vier Cluster, dass die englischsprachige Forschungslandschaft zu Open Data durch ein hohes Maß an fachlicher Interdisziplinarität und thematischer Heterogenität gekennzeichnet ist. Die Interdisziplinarität wird insbesondere in den Clustern 1–2 deutlich, die Bezüge zu unterschiedlichen Wissenschaftsdisziplinen herstellen. Die thematische Heterogenität wird hingegen durch die unterschiedlichen Handlungsfelder hervorgehoben, die durch die Schlagwörter der Cluster adressiert werden. Während Cluster 1 (*Access & Transparency*) den Schwerpunkt auf Datenzugang und Transparenz im primär gesellschaftlichen bzw. staatlichen Anwendungskontext legt, umfasst Cluster 2 (*Socioeconomic Practice*) praktische Anwendungen in Wirtschaft und Industrie. Demgegenüber zeigen Cluster 3 und 4 ein klares technologisches Profil. Während Cluster 3 (*Structure & Retrieval*) die technische Organisation und Strukturierung von Open Data fokussiert, liegt der Schwerpunkt bei Cluster 4 (*Artificial Intelligence*) auf innovativen Datenanalysetechnologien.

3.2 Theoretische und forschungsmethodische Bezüge

Um die theoretischen Bezüge der Forschungslandschaft zu explizieren, wurden die Titel und Abstracts der Datengrundlage in Bezug auf die ausdrückliche Nennung von Theorien untersucht (s. Abb. 1, Schritt ⑥). Dabei ist hervorzuheben, dass nur 68,6 % der gefilterten Datensätze über Abstracts verfügen, und somit eine Grundlage für diese Analyse zur Verfügung stellen. Zur Theorieanalyse sind in einem ersten Schritt diejenigen Literaturquellen identifiziert worden, in denen der Theoriebegriff (*theory*) als Substantiv mindestens einmalig auftritt. Die Abstracts und Titel der resultierenden 146 Quellen sind anschließend manuell auf die Nennung entsprechender Theorien überprüft worden, wodurch insgesamt 37 unterschiedliche Theorien identifiziert werden konnten. Die Ergebnisse dieser Analyse zeigt Tab. 2, wobei eine fachdisziplinäre Gruppierung der Theorien vorgenommen wird und jede Theoriereferenz anhand einer exemplarischen Literaturquelle belegt wird.

Insgesamt signalisiert das resultierende Theoriearsenal, dass Open Data ein interdisziplinäres und breit gefächertes Forschungsfeld bildet, das eine Analyse auf individueller, organisatorischer und gesellschaftlicher Ebene erfordert. Die identifizierten Theorien stammen aus verschiedenen wissenschaftlichen Disziplinen und betreffen eine Vielzahl von Themenfeldern. Dabei wird auch deutlich, dass wirtschaftswissenschaftliche und informationstechnische Theorieperspektiven das Forschungsfeld Open Data dominieren. Ein Schwerpunkt der informationstechnischen Perspektive scheint auf der Erforschung der Akzeptanz und Nutzung von Open Data-basierten Informationssystemen zu liegen (z.B. TAM, UTAUT, ECT), während aus wirtschaftswissenschaftlicher Sicht z.B. die organisatorischen Strukturen, das Ressourcen- und Change Management (z.B. RBT, RCT), sowie die Marktausbreitung (DOI) bedeutsam sind.

Anerkannte Theorien bieten die Möglichkeit, neues mit bestehendem Wissen zu verbinden. Offensichtlich bieten informatikorientierte und wirtschaftswissenschaftliche Theorien eine tragfähige Basis, um Forschungsergebnisse der einzelnen Studien in einem kohärenten Rahmen zu integrieren und Verständnis aufzubauen. Folglich ist Open Data ein Forschungsfeld, welches gerade in der Wirtschaftsinformatik

Tab. 2 Theorien des Forschungsfelds Open Data nach Fachdisziplin

Theorie	Exemplarische Open Data-Literaturquelle mit Theorieferenz
<i>1. Informatik/Informationstechnologie</i>	
3Q Model Theory (IS Success Model)	Muaafi und Priyanto (2024)
Coordination Theory	Zuiderwijk und Janssen (2013)
Dialogue Game Theory	Baker et al. (2017)
Expectation Confirmation Theory of Information Systems Continuance (ECT)	Zuiderwijk und Cligge (2016)
Information Theory	Meymandpour und Davis (2013)
Technology Acceptance Model (TAM)	Muaafi und Priyanto (2024)
Unified Theory of Acceptance and Use of Technology (UTAUT)	Zuiderwijk und Cligge (2016)
Value-Based Theory (Wertbasierte Theorie)	Wilms et al. (2018)
<i>2. Psychologie</i>	
Activity Theory (Tätigkeitstheorie)	Begany und Martin (2017)
Affordance Theory	Mohamad (2024)
Self-Determination Theory (SDT)	Zhang et al. (2023)
Flow Theory	Wang et al. (2024)
Protection Motivation Theory	Yang et al. (2020)
Theory of Planned Behavior	Yang und Wu (2022)
<i>3. Wirtschaftswissenschaften Ökonomie</i>	
Business Ecosystem Theory	Heimstädt et al. (2014)
Capability-Based Theory	Ahmadi Zeleti und Ojo (2019)
Dynamic Capability Theory	Ahmadi Zeleti und Ojo (2019)
Diffusion of Innovations Theory (DOI)	Weerakkody et al. (2017)
Information Asymmetry Theory	Washington (2020)
Innovation Resistance Theory (IRT)	Nikiforova und Zuiderwijk (2022)
Information Signaling Theory	Washington (2020)

Tab. 2 (Fortsetzung)

Theorie	Exemplarische Open Data-Literaturquelle mit Theorieferenz
Naive Quantity Theory of Money	McGinn et al. (2018)
Organizational Change Theory	Enders et al. (2020)
Platform Theory	Lnenicka et al. (2024)
Performance Feedback Theory	Meng und Fan (2024)
Random Utility Maximization Theory (RUM)	Ciang (2018)
Resource-Based Theory (RBT)	Zuiderwijk et al. (2015)
Resource Dependency Theory (RDT)	Zuiderwijk et al. (2015)
4. <i>Soziologie/Politikwissenschaft:</i>	
Actor Network Theory (ANT)	McLeod und McNaughton (2016)
Social Capital Theory	O'Hara (2012)
Rational Choice Theory (RCT)	O'Hara (2012)
Deliberative Democracy Theory	O'Hara (2012)
5. <i>Rechtswissenschaft:</i>	
Binding Precedent Theory	Capasso (2024)
Sources of Law Theory	Capasso (2024)
6. <i>Physik/Mathematik:</i>	
Evolutionary Game Theory	Li et al. (2024)
Systems Theory (Systemtheorie)	McBride et al. (2020)
7. <i>Kommunikationswissenschaft:</i>	
Media Theory	Yoon et al. (2019)

auf besonders fruchtbaren Boden stößt, und aufgrund der Anwendungsorientierung insbesondere durch das Cluster 2 (*Socioeconomic Practice*, s. Tab. 1) thematisch repräsentiert wird.

Dagegen ist die Referenz psychologischer Theoriebeiträge deutlich geringer ausgeprägt. Während im psychologischen Forschungskontext menschliches Verhalten, Motivation und Handlungsregulation betrachtet wird (z. B. SDT, Tätigkeitstheorie), stehen aus rein soziologischer Perspektive soziale Netzwerke, Machtstrukturen und soziale Dynamiken im Fokus (z. B. ANT, FT). Schließlich sind auch interdisziplinäre und systemübergreifende Theoriebeiträge von Bedeutung, wie etwa die Systemtheorie, die in verschiedenen Forschungsfeldern zur Modellierung und Analyse komplexer Systeme eingesetzt wird.

In Summe überwiegen bei den untersuchten Veröffentlichungen deutlich die Erklärungstheorien und die konstruktiven Theorien. Während Erklärungstheorien Forscher und Praktiker dahingehend unterstützen, Ursachen und Prozesse für bestimmte Realphänomene besser zu verstehen, liefern konstruktive Theorien designbasierte Anleitungen, wie bestimmte Forschungsfragen und Ziele besser erreicht werden können. Weitere Theoriegruppen wie normative Theorien, deskriptive Theorien, Makrotheorien oder auch Vorhersagetheorien sind im Forschungsfeld eher unterrepräsentiert. Daher lässt sich hier schlussfolgern, dass das Forschungsfeld Open Data ein breites Spektrum von Ansätzen zur Erklärung seiner Kernphänomene entwickelt hat, um unterschiedliche Mechanismen und Perspektiven zu beleuchten. Zugleich zeigt die Existenz der konstruktiven Theorien in der Breite, dass das Forschungsfeld einen starken praktischen Fokus hat, aktiv nach Lösungen sucht und versucht, reale Herausforderungen professionell zu bewältigen. In dieser Kombination der Theorien

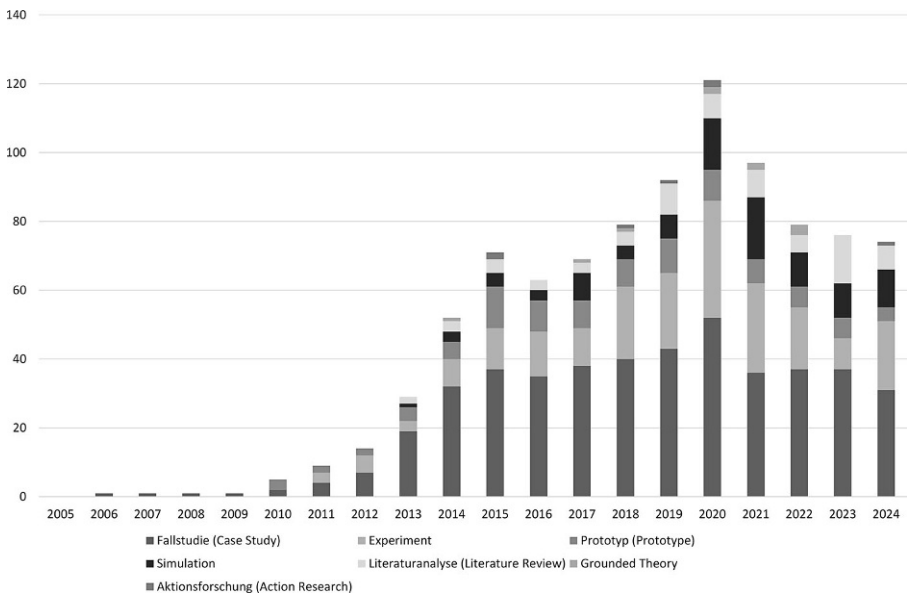
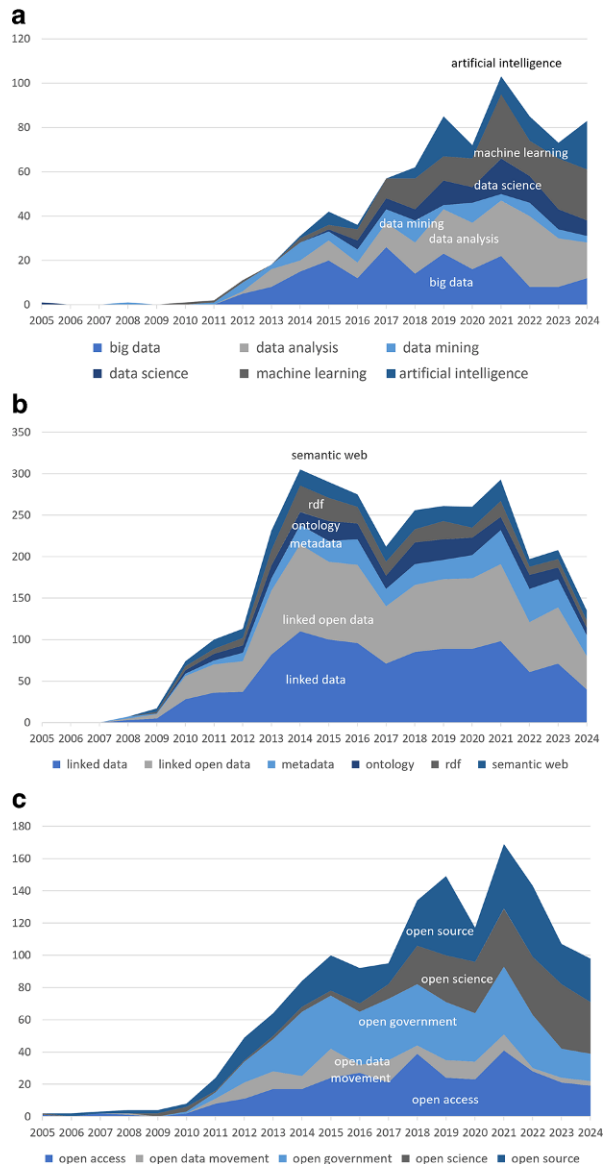


Abb. 4 Forschungsmethoden im Forschungsfeld Open Data (Anzahl Publikationen mit benannter Forschungsmethode)

Abb. 5 Entwicklung ausgewählter Schlagwortgruppen (Anzahl Publikationen mit benanntem Schlagwort). **a** Datenanalyse, **b** Semantische Technologien und Standards, **c** Open X-Konzepte



kann davon ausgegangen werden, dass das Forschungsfeld einen methodologischen Reifegrad erreicht hat, bei dem sowohl Grundlagenforschung als auch robuste angewandte Forschung vorhanden sind.

Darüber hinaus sind die Literaturquellen auch in Bezug auf die Nennung von *Forschungsmethoden* untersucht worden. Zu diesem Zweck wurden die aus den Texten per Part of Speech-Tagging generierten Nominalphrasen herangezogen bzw. eine Volltextsuche durchgeführt. Die quantitativen Ergebnisse der Analyse zeigt das Balkendiagramm in Abb. 4. Daraus wird deutlich, dass Open Data stark durch Fall-

studienforschung geprägt wird, die insbesondere im Kontext von Open Government Data (OGD)-Initiativen auftritt und aus zeitlicher Perspektive bereits zu Beginn des Untersuchungszeitraums Anwendung findet. Als weitere, frequente Forschungsmethode werden in den Literaturquellen ab 2011 auch *Experimente* angeführt, die aufgrund der disziplinären Vielfalt im Forschungsfeld allerdings sehr unterschiedlich ausgeprägt sein können. So finden sich in der Literaturbasis zum Beispiel naturwissenschaftliche Experimente, die mithilfe von Open Data dokumentiert werden (u. a. des CERN), aber auch softwaregestützte Experimente zur algorithmischen Evaluation von Open Data. An dritter Stelle der Forschungsmethoden reiht sich ab 2010 die *Prototypenentwicklung* ein, die häufig im Umfeld von OGD-Infrastrukturen stattfindet und an der Nachnutzung bereits existierender Datenbestände ansetzt. Auch *simulative Forschungsdesigns* profitieren von der Verfügbarkeit offener Datenbestände und kommen in der Literatur für die unterschiedlichsten Disziplinen zum Vorschein. Besonders häufig finden sich in dem Literaturkorpus Verweise auf Klima-, Wetter- und Verkehrsdaten, die zur Simulation in den unterschiedlichen Anwendungsdomänen eingesetzt werden. Die *Literaturanalyse* wird als Forschungsmethode im Open Data-Kontext insbesondere seit 2019 genutzt, was angesichts des vorangehenden Publikationswachstums (s. Abb. 2) plausibel erscheint. Sie wird häufig im thematischen Umfeld des Open Government bzw. des eGovernment verwendet. Die Forschungsmethoden der *Aktionsforschung* und der *Grounded Theory* thematisieren schwerpunktmäßig die Nutzeffekte und sozialen Implikationen von Open Data, besitzen in der Literaturbasis allerdings nur eine geringe Verbreitung.

3.3 Entwicklungstrends

Zur Analyse von thematischen Trends im Forschungsfeld wurde die zeitliche Entwicklung ausgewählter Schlagwörter (s. Tab. 1) in den Titeln und Abstracts der Datenbasis untersucht. Dabei sind solche Schlagwörter der vier Cluster selektiert worden, die sich auf die Bereitstellung und Nutzung von Open Data konzentrieren, während Begriffe für die disziplinäre Einordnung (z. B. Business, Medicine) oder generische Begriffe (z. B. Computer) ausgeschlossen wurden. Durch diese Analyse konnten interessante Entwicklungsverläufe identifiziert werden, die im Folgenden dargestellt werden.

Zunächst sind aus den Schlagwörtern diejenigen ausgewählt worden, die sich mit der Datenanalyse beschäftigen (s. Abb. 5a). Hierzu gehören die Begriffe *Artificial Intelligence*, *Big Data*, *Data Analysis*, *Data Mining*, *Data Science*, sowie *Machine Learning* (Cluster 1, 2 u. 4). Die Abbildung verdeutlicht, dass die Frequenz analytischer Themen im Open Data-Umfeld seit 2010 deutlich zugenommen hat. Während in jüngerer Zeit ein Bedeutungsrückgang von *Data Mining* und *Data Science* zu identifizieren ist, konnten indes die Themen *Artificial Intelligence* und *Machine Learning* deutlich zunehmen. Diese Entwicklung erscheint vor dem Hintergrund plausibel, dass Künstliche Intelligenz bzw. maschinelles Lernen als KI-Methodik in der wissenschaftlichen Literatur bereits seit mehreren Jahren einem positiven Trend folgen. In diesem Zusammenhang ist auch darauf hinzuweisen, dass Bezüge auf Large Language Models (LLM) in der Datenbasis bislang nur in geringem Umfang vorhanden sind. Es konnten zum Forschungsfeld LLM lediglich acht Quellen iden-

tifiziert werden, die im Zeitraum von 2022–2024 publiziert wurden. Aufgrund der intensiven Weiterentwicklung des Forschungsfelds der generativen KI ist künftig mit einem anhaltenden Bedeutungszuwachs zu rechnen.

Ein weiterer, interessanter Entwicklungsverlauf konnte für die Gruppe der *semantischen Technologien und Standards* identifiziert werden, die im Cluster 3 (*Structure & Retrieval*) verankert sind (Linked (Open) Data, Metadata, Ontology, RDF, Semantic Web). Die Themen dieser Gruppe (s. Abb. 5b) haben sich in der Literatur seit 2007 überaus rasant entwickelt, zeigen aber seit 2020 einen deutlichen Rückgang. Zwar ist die Häufigkeit dieser Themen vergleichsweise hoch ausgeprägt, indes scheint hier die Forschungstätigkeit zurückzugehen. Es ist daher zu vermuten, dass im Umfeld semantischer Technologien aufgrund des erzielten Reifegrads mittlerweile die laufenden Arbeiten zur Standardisierung, zur Verbesserung der Interoperabilität und zur Bewältigung von Datenschutz- und Sicherheitsanforderungen im Mittelpunkt stehen.

Schließlich ist auch von Interesse, wie sich die unterschiedlichen Open-X-Konzepte im Zeitablauf entwickelt haben, die im Cluster 1 (*Access & Transparency*) verankert sind. Wie aus Abb. 5c deutlich wird, ist auch hier nach einem anfänglich starken Wachstum ab 2021 ein Rückgang zu verzeichnen. Dieser entfällt insbesondere auf die Open Data-Bewegung (Open Data Movement) und Open Government, was auf mögliche Sättigung in der Forschung oder veränderte Prioritäten hinweisen könnte. Indes ist positiv zu verzeichnen, dass Open Science und – mit Abstrichen – Open Access im Open Data-Umfeld weiterhin eine deutliche Rolle spielen, was ihre anhaltende Relevanz in der wissenschaftlichen Community unterstreicht.

4 Fazit und Ausblick

Zusammenfassung: Die Untersuchung hat gezeigt, dass sich die Open Data-Forschung durch ein hohes Maß thematischer Heterogenität auszeichnet. Eine Clusteranalyse auf Grundlage der Schlagwörter von 8060 Literaturdatensätzen hat vier thematische Hauptsegmente identifiziert, die sich deutlich voneinander abgrenzen lassen (Forschungsfrage 1). Während im ersten Segment *Access & Transparency* die Bedeutung der Open-X-Bewegungen und der europäische Bezug in der Open Data-Nutzung betont wird, fokussiert das zweite Segment *Socioeconomic Practice* Anwendungen in Wirtschaft und Industrie, ergänzt durch Bezüge zur Bildung. Mit dem dritten Cluster *Structure & Retrieval* ist die Thematik der Organisation und Verknüpfung von Daten im Web – einschließlich Linked Data und semantischer Technologien – erfasst worden, sodass starke Verbindungen zu den Informations- und Bibliothekswissenschaften bestehen. Im letzten Cluster *Artificial Intelligence* werden moderne Analysetechniken subsumiert, die zur produktiven Nutzung offener Datenbestände notwendig sind. Aus methodologischer Perspektive konnte mit der Literaturanalyse deutlich gemacht werden, dass ein breites Spektrum von Theorien aus unterschiedlichen Disziplinen eingesetzt wird und die Erkenntnisgewinnung mithilfe unterschiedlicher Forschungsmethoden stattfindet (Forschungsfrage 2), wobei das Feld methodisch stark von Fallstudien geprägt wird. Die Analyse der Entwicklung von Schlagwörtern in Titeln und Abstracts von Forschungsarbeiten verdeutlicht

zudem, dass insbesondere Themen wie Artificial Intelligence und Machine Learning in jüngster Zeit an Bedeutung gewonnen haben, während semantische Technologien und Standards an Frequenz verloren haben, was auf einen Reifegrad des Teilgebiets schließen lässt (Forschungsfrage 3).

Insgesamt ist mit der vorgelegten Literaturanalyse ein großzahliger, englischsprachiger Datenbestand aus einer etablierten Literaturdatenbank verarbeitet worden, der eine Publikationsperiode von 20 Jahren abdeckt. Auf diese Weise wird nicht nur ein längerer Zeitraum berücksichtigt, als dies bei den eingangs skizzierten Studien der Fall ist, sondern auch eine aktuelle, disziplinübergreifende Betrachtung durchgeführt, die auf einem größeren Publikationsvolumen basiert und auch die in den Literaturdaten enthaltenen Abstracts zur Erkenntnisgewinnung erschließt.

Limitationen: Um das wissenschaftliche Verständnis von Open Data weiterhin zu verbessern, bietet sich die Optimierung der hier dargestellten Literaturanalyse an. Wie jede wissenschaftliche Arbeit, weist auch diese potenzielle Limitationen auf. So besteht die Möglichkeit, dass nicht sämtliche relevanten Literaturquellen durch den Analyseprozess (s. Abb. 1) identifiziert wurden. Die Ursachen hierfür können z. B. in der Unvollständigkeit der definierten Schlagwörter, in alternativen Begriffsbezeichnungen (Synonymen) in den Literaturquellen sowie in der Qualität der gewählten Literaturdatenbank bzw. des eingesetzten Deduplikationsdienstes liegen. In diesem Zusammenhang ist auch darauf hinzuweisen, dass die genutzten Literaturdaten in Bezug auf die Abstracts unvollständig sind – so verfügen nur 68,6 % der Datensätze über Abstracts, sodass die daraus abgeleiteten Erkenntnisse (Forschungsfrage 2 u. 3) nicht die gesamte empirische Basis repräsentieren.

Im Rahmen künftiger Forschungsarbeiten eröffnet sich die Möglichkeit, auch deutschsprachige Publikationen aufzunehmen und im Zuge komparativer Analysen methodologische bzw. thematische Unterschiede aufzudecken. Außerdem können die Daten aus weiteren Literaturdatenbanken erschlossen werden (z. B. Web of Science, Scopus), sodass durch eine breitere Literaturbasis potenziell vorhandene blinde Flecken abgebaut werden können. Eine Erweiterung der Literaturanalyse um eine institutionelle bzw. autorenbezogene Perspektive ist ebenfalls möglich. So unterstützt das eingesetzte Werkzeug VOSviewer auch die Analyse von Autorennetzen. In Bezug auf das methodische Arsenal zur Analyse von Literaturdaten bietet sich eine Ergänzung um natürliche Verfahren der Textdatenverarbeitung (Natural Language Processing, NLP) an, die in der Lage sind, komplexere sprachliche Muster zu erkennen, indem Wortbedeutungen und -beziehungen analysiert werden. Mit diesen neueren Methoden (z. B. Word2Vec, GloVe) geht das Potenzial einher, Synonyme und Kontexte in Literaturdaten besser zu erkennen, sodass die Validität der identifizierten Zusammenhänge gesteigert werden kann. Eine breitere Verfügbarkeit von Volltexten in der wissenschaftlichen Publikationslandschaft – etwa durch die stringente Anwendung von Open Access-Lizenzen – würde dabei eine detaillierte Analyse über den reinen Titel oder Abstract hinaus gestatten, sodass komplexere Forschungsfragen bearbeitet bzw. die Aussagekraft der Ergebnisse gesteigert werden können. Ferner ist bei der Interpretation der Ergebnisse insbesondere zu Abb. 3 zu berücksichtigen, dass keine Unterscheidung vorgenommen wurde zwischen Publikationen, die Open Data als eigenständigen Erkenntnis- oder Gestaltungsgegenstand

fixieren, und Publikationen, die Open Data lediglich als Quelle bzw. Instrument zur Beantwortung spezifischer Forschungsfragen nutzen, wodurch die Clusterbildung potenziell beeinflusst wird. Des Weiteren ist es denkbar, dass aufgrund der bewusst breit ausgewählten Datenbasis auch Literaturdaten in die Analyse eingeflossen sind, die als falsch-positiv zu werten sind, z.B. aufgrund kompositen Titelbegriffe wie *Open Data Center*.

Kritische Diskussion: Trotz dieser Limitationen zeigt die Literaturanalyse zum Forschungsfeld Open Data eindrücklich dessen interdisziplinäre Natur und thematische Vielfalt, was sowohl als Stärke als auch als Herausforderung gesehen werden kann. Einerseits ermöglicht die Integration verschiedener Disziplinen, wie der Informatik, der Ökonomie und der Psychologie, ein umfassendes Verständnis und die Entwicklung innovativer Anwendungen. Andererseits kann die hohe thematische Heterogenität zu einer Fragmentierung des Wissens führen, wobei Synergiepotenziale möglicherweise ungenutzt bleiben. Während viele Erklärungstheorien und konstruktive Theorien zur Verfügung stehen, sind andere theoretische Ansätze, wie normative oder Vorhersagetheorien, unterrepräsentiert, was ein Ungleichgewicht in der theoretischen Fundierung des Feldes suggerieren könnte. Der prognostizierte Einfluss moderner Analysemethoden, insbesondere generativer KI, deutet auf eine spannende, zukünftige Entwicklung hin, birgt jedoch auch potenzielle Risiken, die sich aus ethischen und datenschutzrechtlichen Herausforderungen ergeben können. Schließlich zeigt die Studie auf, dass gezielte Bemühungen erforderlich sind, um die Informationssilos zu überwinden und die vollständige Bandbreite der Potenziale von Open Data auszuschöpfen, was eine strategische Neuorientierung in Forschung und Praxis erfordern kann.

Implikationen für Theorie und Praxis: (1) Obwohl die Untersuchung gezeigt hat, dass viele Erklärungstheorien und konstruktive Theorien im Einsatz sind, gibt es eine Unterrepräsentation von theoretischen Ansätzen wie normativen oder Vorhersagetheorien. Zukünftige Forschung sollte daher verstärkt darauf abzielen, diese Lücken zu schließen, indem sie eine breitere Palette theoretischer Rahmenwerke in Betracht zieht. Dies kann insbesondere dazu beitragen, politisch-gesellschaftliche und ethische Dimensionen von Open Data umfassender zu beleuchten. Aus Sicht einer gestaltungsorientierten Wirtschaftsinformatik bietet sich somit noch Potenzial für die Entfaltung von Forschungsaktivitäten. So finden sich in der untersuchten Literaturbasis nur 15 Beiträge, die auf die Design Science Research (DSR) Bezug nehmen. Insgesamt zeigt die Untersuchung der theoretischen Landschaft, dass Open Data ein komplexes Forschungsfeld ist, das auf unterschiedlichen Ebenen und aus unterschiedlichen Perspektiven analysiert und weiterentwickelt werden kann. (2) Die interdisziplinäre Natur des Open Data-Feldes sollte weiter gestärkt werden, um Synergien zwischen den verschiedenen Disziplinen besser zu nutzen. Forschende sollten gezielt kollaborative Projekte initiieren, die Expertise aus Informatik, Wirtschaft, Sozialwissenschaften und weiteren Bereichen integrieren. Dies kann helfen, die Fragmentierung des Wissens zu überwinden und umfassendere Lösungen für die Herausforderungen im Bereich Open Data zu entwickeln. (3) Angesichts der prognostizierten Bedeutung generativer KI-Methoden sollten künftige Forschungspro-

jekte stärker auf die Integration moderner Analysemethoden setzen. Dies beinhaltet sowohl die technische Erforschung dieser Technologien als auch die Erforschung ihrer ethischen und regulatorischen Rahmenbedingungen. Forscher und Praktiker sollten die Entwicklungen in diesem Bereich aufmerksam verfolgen, um das Potenzial von Open Data zu erschließen. Die aus vielerlei Hinsicht komplementäre Beziehung zwischen Open Data und generativer KI betrifft unterschiedliche anwendungsorientierte Bereiche. Die Bereicherung kann z. B. die Menge der Trainingsdaten, die Qualität bzw. Vielfalt der Daten oder auch eine transparentere KI-Entwicklung betreffen. Solche Untersuchungen könnten die Entwicklung verantwortungsvoller und effektiver Anwendungen von Open Data fördern und sicherstellen, dass die hieraus hervorgehenden Innovationen sozialverträglich und nachhaltig sind.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Ahmadi Zeleti F, Ojo A (2019) Qualitative structural model for capabilities in open data organizations. In: Bui TX (Hrsg) Proceedings of the 52nd Hawaii International Conference on System Sciences 2019 Maui, January 8–11 2019, S 2902–2911 <https://doi.org/10.24251/hicss.2019.351>
- Auth G, Czarnecki C, Bensberg F, Thor A (2019) Digitalisierung des Forschungsprozesses aus Sicht von Forschenden – durch Serviceintegration zum persönlichen Forschungsinformationssystem. In: Barton T, Müller C, Seel C (Hrsg) Hochschulen in Zeiten der Digitalisierung. Springer, Wiesbaden, S 287–307 https://doi.org/10.1007/978-3-658-26618-9_17
- Baker M, Groff J, Détienne F et al (2017) Technology-supported effective transparency around open data: a dialogue game analysis. In: Proceedings of the European Conference on Cognitive Ergonomics (ECCE) 2017 Umeå, September 19–22 2017, S 39–42 <https://doi.org/10.1145/3121283.3121293>
- Begany GM, Martin EG (2017) An open health data engagement ecosystem model: Are facilitators the key to open data success? In: Erdelez S, Agarwal N (Hrsg) Proceedings of the Association for Information Science and Technology, Proceedings of the 80th ASIS&T Annual Meeting Washington DC, October 27–November 1 2017. Bd. 54(1), S 621–623 <https://doi.org/10.1002/pra2.2017.14505401091>
- Bensberg F, Auth G, Czarnecki C, Wördle C (2019) In: İter HK (Hrsg) IMISC, Bd. 2018. Ankara, S 24–26 <https://doi.org/10.6084/m9.figshare.7582073.v1>
- Capasso V (2024) Open data of judicial decision in comparative perspective: towards a reshaping of source theory? In: AIDC/DPCE Colloquium on Itineraries of Legal Comparison—New Issues and New Perspectives Rome, May 16–17 2024 (<https://www.iris.unina.it/handle/11588/960484>). Gesehen 11. Mai 2025)
- Catone MC (2023) The role of open data in digital society: the analysis of scientific trending topics through a bibliometric approach. Front Sociol 8:1134518. <https://doi.org/10.3389/fsoc.2023.1134518>
- Ciang D (2018) Measuring citizen preferences for the Dutch education open data policy: a path towards citizen-informed decision making. Delft university of technology. <https://resolver.tudelft.nl/uuid:f0f3d793-f34d-4fbb-98a4-f41d937c4f89>. Zugegriffen: 10. Mai 2025

- Enders T, Benz C, Schüritz R, Lujan P (2020) How to implement an open data strategy? Analyzing organizational change processes to enable value creation by revealing data. In: Proceedings of the 28th European Conference on Information Systems (ECIS) Marrakesch, June 15–17 2020 (https://aisel.aisnet.org/ecis2020_rp/204)
- Forbes C, Greenwood H, Carter M, Clark J (2024) Automation of duplicate record detection for systematic reviews: Deduplicator. *Syst Rev*. <https://doi.org/10.1186/s13643-024-02619-9>
- Heimstädt M, Saunderson F, Heath T (2014) From toddler to teen: growth of an open data ecosystem. *JeDEM Ejournal Edemocracy Open Gov* 6(2):123–135. <https://doi.org/10.29379/jedem.v6i2.330>
- Hyseni B, Bexheti LA (2024) Synchronization and standardization of open data platforms: a systematic literature review. *TEM J* 13(2):1265–1276. <https://doi.org/10.18421/TEM132-40>
- Li Q, Pi B, Feng M, Kurths J (2024) Open data in the digital economy: an evolutionary game theory perspective. *IEEE Trans Comput Soc Syst* 11(3):3780–3791. <https://doi.org/10.1109/TCSS.2023.3324087>
- Lnenicka M, Nikiforova A, Clarinval A et al (2024) Sustainable open data ecosystems in smart cities: a platform theory-based analysis of 19 European cities. *Cities* 148:104851. <https://doi.org/10.1016/j.cities.2024.104851>
- Lucke J, Geiger CP (2010) Open Government Data. Zeppelin University, Friedrichshafen. <https://www.zu.de/institute/togi/assets/pdf/TICC-101203-OpenGovernmentData-V1.pdf>. Zugegriffen: 10. Mai 2025
- McBride K, Olesk M, Kütt A, Shysh D (2020) Systemic change, open data ecosystem performance improvements, and empirical insights from Estonia: a country-level action research study. *Inf Polity* 25(3):377–402. <https://doi.org/10.3233/IP-190195>
- McGinn D, McIlwraith D, Guo Y (2018) Towards open data blockchain analytics: a Bitcoin perspective. *R Soc open sci* 5(8):180298. <https://doi.org/10.1098/rsos.180298>
- McLeod M, McNaughton M (2016) Mapping an emergent Open Data eco-system. *J Community Informatics* 12(2):26–46. <https://doi.org/10.15353/joci.v12i2.3220>
- Meng X, Fan B (2024) Historical performance gap, readiness of government, and open data quality: transnational empirical analysis based on performance feedback theory. *ACM Sigmis Database* 55(3):50–69. <https://doi.org/10.1145/3685235.3685239>
- Meymandpour R, Davis JG (2013) Ranking universities using linked open data. In: linked data on the web (LDOW 2013), May 14 2013, Rio de Janeiro, (Brazil). <https://ceur-ws.org/Vol-996/papers/ldow2013-paper-09.pdf>. Zugegriffen: 10. Mai 2025
- Mohamad AN (2024) Investigating the actualization of open data affordances for start-up entrepreneurs. *Inf Discov Deliv*. <https://doi.org/10.1108/IDD-03-2024-0050>
- Muaafii YH, Priyanto P (2024) Optimizing „open data Jawa Tengah“ through technology acceptance model (TAM). *ELINVO* 9(1):11–23. <https://doi.org/10.21831/elinvo.v9i1.67506>
- Nikiforova A, Zuiderwijk A (2022) Barriers to openly sharing government data: towards an open data-adapted innovation resistance theory. In: Amaral L et al (Hrsg) Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance Guimarães, October 4–7 2022, S 215–220 <https://doi.org/10.1145/3560107.3560143>
- O'Hara K (2012) Transparency, open data and trust in government: shaping the infosphere. In: WebSci '12: Proceedings of the 4th Annual ACM Web Science Conference Evanston, June 22–24 2012, S 223–232 <https://doi.org/10.1145/2380718.2380747>
- Shettar I, Hadagali GS (2023) Global research publications on open data: a scientometric analysis. *RBU J Libr Inf Sci* 24:112–119. <https://doi.org/10.2139/ssrn.4344330>
- Van Eck NJ, Waltman L (2007) VOS: a new method for visualizing similarities between objects. In: Decker R, Lenz H-J (Hrsg) Advances in data analysis. Springer, Berlin Heidelberg, S 299–306 https://doi.org/10.1007/978-3-540-70981-7_34
- Van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84:523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Wang P, Xie Y, Li X, Li Q (2024) Allocation of attention to metadata and retrieval functions: Implications for perceived value and open data discovery and reuse. *J Librariansh Inf Sci* 56(3):607–622. <https://doi.org/10.1177/09610006231154529>
- Washington AL (2020) Uncertain risk: assessing open data signals. *Transform Gov People Process Policy* 14(4):623–637. <https://doi.org/10.1108/TG-09-2019-0086>
- Weerakkody V, Irani Z, Kapoor K et al (2017) Open data and its usability: an empirical view from the Citizen's perspective. *Inf Syst Front* 19:285–300. <https://doi.org/10.1007/s10796-016-9679-1>
- Wilms K, Brenger B, López A, Rehwal S (2018) Open data in higher education—what prevents researchers from sharing research data? In: Proceedings of the International Conference on Information Systems (ICIS) 2018 San Francisco, December 13–16 2018 (<https://aisel.aisnet.org/icis2018/>)

- Wirtz BW, Weyerer JC, Becker M, Müller WM (2022) Open government data: a systematic literature review of empirical research. *Electron Markets* 32:2381–2404. <https://doi.org/10.1007/s12525-022-00582-8>
- Yang T-M, Wu Y-J (2022) To use or not to use? Exploring the factors influencing professional Reusers' intention to adopt and utilize governmental open data in Taiwan. *J Educ Media Libr Sci* 59(2):101–135
- Yang N, Singh T, Johnston AC (2020) Open data: a replication study of user motivation in protecting information security using protection motivation theory and self-determination theory. *AIS Trans Replication Res* 6(12):1–5 (<https://aisel.aisnet.org/trr/vol6/iss1/12>)
- Yoon S-P, Joo M-H, Kwon H-Y (2019) How to guarantee the right to use PSI in the age of open data: lessons from the data policy of South Korea. *Inf Polity* 24(2):131–146. <https://doi.org/10.3233/IP-180103>
- Yoon YS, Zo H, Choi M, Lee D, Lee H-W (2018) Exploring the dynamic knowledge structure of studies on the Internet of things: keyword analysis. *ETRI J* 40(6):745–758. <https://doi.org/10.4218/etrij.2018-0059>
- Zhang Y, Liu Z, Zhao YC, Wu D (2023) Understanding the motivation of participants in innovation open data contests: a task presentation affordance perspective. *Proc Assoc Inf Sci Technol* 60(1):1218–1220. <https://doi.org/10.1002/ptra2.997>
- Zuiderwijk A, Cligge M (2016) The acceptance and use of open data infrastructures—drawing upon UTAUT and ECT. In: *Innovation and the public sector: electronic government and electronic participation*, Bd. 23. IOS Press, S 91–98 <https://doi.org/10.3233/978-1-61499-670-5-91>
- Zuiderwijk A, Janssen M (2013) A coordination theory perspective to improve the use of open data in policy-making. In: Wimmer MA, Janssen M, Scholl HJ (Hrsg) *Electronic Government. EGOV 2013. Lecture Notes in Computer Science* (8074). Springer, Berlin, Heidelberg, S 38–49 https://doi.org/10.1007/978-3-642-40358-3_4
- Zuiderwijk A, Janssen M, Poulis K, Van De Kaa G (2015) Open data for competitive advantage: insights from open data use by companies. In: *Proceedings of the 16th Annual International Conference on Digital Government Research Phoenix*, May 27–30 2015, S 79–88 <https://doi.org/10.1145/2757401.2757411>

Hinweis des Verlags Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.