Artificial Intelligence as a Catalyzer for Open Government Data Ecosystems

Simonofski, Anthony; Nikiforova, Anastasija; Lnenicka, Martin; Bono Rossello, Nicolas

Published in:
Proceedings of the 58th Hawaii International Conference on System Sciences, HICSS 2025

DOI:
10.24251/hicss.2025.267

Publication date:
2025

Document Version
Publisher's PDF, also known as Version of record

Artificial Intelligence as a Catalyzer for Open Government Data Ecosystems: A Typological Theory Approach

Anthony Simonofski
University of Namur (Belgium)
anthony.simonofski@unamur.be

Anastasija Nikiforova
University of Tartu (Estonia)
anastasija.nikiforova@ut.ee

Martin Lnenicka
University of Hradec Kralove (Czech Republic)
martin.lnenicka@uhk.cz

Nicolas Bono Rossello
University of Namur (Belgium)
nicolas.bonorossello@unamur.be

Abstract

Artificial Intelligence (AI) within digital government has witnessed growing interest as it can improve governance processes and stimulate citizen engagement. Despite the rise of Generative AI, discussions on AI fusion with Open Government Data (OGD) remain limited to specific implementations and scattered across disciplines. Drawing from the synthesis of the literature through a systematic review, this study examines and structures how AI can enrich OGD initiatives. Employing a typological approach, ideal profiles of AI application within the OGD lifecycle are formalized, capturing varied roles across the portal and ecosystems perspectives. The resulting conceptual framework identifies eight ideal types of AI applications for OGD: AI as Portal Curator, Explorer, Linker, and Monitor, and AI as Ecosystem Data Retriever, Connecter, Value Developer and Engager. This theoretical foundation shows the under-investigation of some types and will inform policymakers, practitioners, and researchers in leveraging AI to cultivate OGD ecosystems.

Keywords: Open Government Data, Artificial Intelligence, Ecosystem, Portal, Typology

1. Introduction

In recent years, the utilization of Artificial Intelligence (AI) within the digital government context has garnered significant attention. Empirical evidence reports clear benefits of AI, including the development of AI-based technologies as a form of algorithmic bureaucracy. This interest has been further fueled by the emergence of Generative AI, which has introduced novel possibilities and challenges within the domain (Ugale, G. and Hall, 2024). Governments worldwide, at national and supra-national levels, with somehow less progress at the subnational and local levels, are striving to improve the findability, accessibility, interoperability and reusability (FAIR) of their data. In this context, AI technologies present unprecedented opportunities to unlock the full potential of Open Government Data (OGD). However, the existing body of research tends to focus on specific AI implementations (*problem 1*), in a specific setting of OGD - the portals, neglecting other critical aspects of the OGD ecosystem (*problem 2*).

Regarding *problem 1* and the scope of AI technologies examined, discussions surrounding the integration of AI and OGD have been limited with a fragmented literature consisting of studies around specific AI implementations in the OGD ecosystem. The use of AI in an OGD context is often assimilated studies exploring the use of chatbots for OGD exploration (Cortés-Cediel et al., 2023), browsing (Cantador et al., 2021) or interoperability (Barcellos et al., 2017)). Even though these studies have inherent merits, AI emerges as a broader promising avenue for transforming governance, enhancing transparency, and fostering citizen engagement in the OGD lifecycle and ecosystem (Li et al., 2024). Regarding *problem 2* - the narrow focus on OGD portals - this is even more so, considering the Open Data Charter (2024-2026) and the Open Data Policy lab activities, stressing that the ongoing Third Wave of Open Data[1] is expected to take a much more purpose-directed approach. In other words, the Third Wave is not about the open data as an artifact anymore but is about open data lifecycle and the entire open data ecosystem.

This study aims to delve into the multifaceted relationship between AI and OGD through the lens of a conceptual framework. By identifying and

[1] Third Wave of Open Data - Open Data Policy Lab

URI: https://hdl.handle.net/10125/109107
978-0-9981331-8-8

synthesizing the ways in which AI can be leveraged to enrich OGD ecosystems, this study seeks to provide clarity and insight into the potential synergies and implications inherent in this evolving landscape. Drawing upon existing literature, this research explores key conceptual lenses through which AI can enhance OGD initiatives. By systematically examining these conceptual lenses, we aim to formalize the diverse mechanisms by which AI can amplify the accessibility, value and relevance of OGD for various stakeholders. This research seeks to advance scholarly discourse and practical insights into the transformative potential of AI in shaping the future of OGD ecosystems. By synthesizing scattered literature and offering a coherent conceptual framework, this study aims to inform policymakers, practitioners, and researchers alike on how AI can be harnessed to foster transparency, accountability, and innovation in governance processes through the OGD lifecycle.

To achieve this goal, we employ a typological approach. This methodology allows for the formalization of "*ideal profiles*" of AI application within OGD lifecycle that emerge from two dimensions of the designed typology, capturing the varied and nuanced ways in which AI can support OGD through its lifecycle. As such, the first dimension evaluates the role of AI across tasks using the inward- (portal) and outward-looking (ecosystem) perspectives drawn from the Business Process Management (BPM) literature (Dumas et al., 2018). The second dimension of the OGD lifecycle adheres to fundamental stages proposed by Attard et al. (2015), including data pre-processing, exploitation, and maintenance, enriched by insights from Crusoe et al. (2019), thereby delineating between data exploration and transformation within the exploitation phase. By identifying distinct types of AI applications across OGD lifecycle, we can better understand the potential benefits for each profile. This, in turn, guides further actions to catalyze an efficient use of OGD through effective and efficient use of AI capabilities in scenarios that may remain unnoticed before. In this way we define 8 ideal profiles for AI uses in OGD: *AI as Portal Curator*, *Portal Explorer*, *Portal Linker*, *Portal Monitor*, *Ecosystem Data Retriever*, *Ecosystem Connecter*, *Ecosystem Value Developer* and *Ecosystem Engager*. This study thus contributes to a more nuanced understanding of the opportunities inherent in harnessing AI for OGD ecosystems.

The paper is organized as follows: Section 2 provides the methodology for the research, Section 3 presents the typology, whereas implications for research and practice, limitations and further research are brought forth in Sections 4 and 5, respectively, with Section 6 presenting conclusions.

2. Methodology

Typology is a top-down system of classification designed to categorize phenomena into distinct types based on a set of predefined criteria that emphasize notable differences and similarities (Bailey, 1994). Unlike taxonomies, which organize information based on bottom-up empirical observations, typologies are conceptual in nature and particularly useful in fields where phenomena do not fit neatly into predefined categories. Within the context of typologies, "*ideal profiles*" refer to conceptual models that represent the purest forms of identified categories. These profiles, though theoretical and as such may not be directly observable in their pure form, serve as benchmarks for comparing and understanding the empirical reality. "*First-order constructs*" are the foundational dimensions or criteria used to establish these profiles. They represent the essential attributes that define each type within the typology.

To ground our typology in the existing body of knowledge, we conducted a systematic literature review (SLR) examining the integration of AI within OGD lifecycle. Although typology development is primarily a conceptual approach, examining empirical examples reported in the literature can support the derivation of the ideal profiles by classifying selected studies using the first-order constructs (Stapley et al., 2022). Our search strategy combined a set of keywords related to OGD with terms associated with AI. AI entails several definitions, that keep evolving over time. Berente et al. (2020) suggest that AI is not a single technology but rather a frontier in computational advancements that references human intelligence in addressing ever more complex decision-making problems. This results in AI encompassing a wide spectrum of technologies from expert systems to Natural Language Processing (NLP), Large Language Models (LLM) and machine learning (ML). Thus, we position AI as computational elements automating or augmenting human tasks and an umbrella term for various techniques associated with the above. As such, we defined and used the following search string: *("open government* data" OR OGD OR "open data" OR "public sector information" OR "government transparency data" OR "data for* *democracy") AND (AI OR "artificial intelligence" OR "machine learning" OR "deep learning" OR "data mining" OR "text mining" OR NLP OR "natural language processing" OR reasoning OR "neural networks" OR LLM OR "large language model*" OR "knowledge representation" OR augmentation OR automation)).*

The defined search string was applied to the article title, keywords, and abstract to limit the number of papers to those where these objects were primary research objects rather than mentioned, e.g., as a future work. For the period covered by this search, we do not set a start or end date. We conducted this search on Scopus and Web of Science (Clarivate), ensuring a broad and peer-reviewed collection of academic literature. Additionally, we performed a coverage check on the first 20 pages of results from Google Scholar to verify the inclusivity and relevance of our corpus. To be included in our review, studies must satisfy the following inclusion criteria: (i) report on the use of AI as research object; (ii) explore an OGD lifecycle or its individual phase(s); (iii) focus on government applications; (iv) being empirical in nature; (v) be written in English; (vi) be journal articles, conference papers, or book chapters. This resulted in 209 articles in Scopus and 178 in Web of Science (as of April 2024). They were filtered into 33 papers after abstract analysis resulting from Scopus and 38 papers resulting from Web of Science. This was complemented by 18 papers resulting from a search performed in Google Scholar (out of 199 analyzed papers). In total, 71 papers were used to identify reported solutions in the literature and develop typology.

To design the typology, we followed the approach of Kulge (2000) consisting of four steps: (1) definition of analyzing dimensions, (2) grouping cases, (3) analyzing empirical regularities, and (4) constructing the "ideal profiles". The first analyzing dimension relates to the perspective taken that can be either inward-looking (portal) or outward-looking (ecosystem) as formalized in BPM by Dumas et al. (2018), whereas the second dimension relates to the phase of OGD lifecycle as formalized by Attard et al. (2015) and Crusoe et al. (2019). Following the constructs of these dimensions, we coded the papers based on the perspective reported and the impacted OGD lifecycle phase. This theoretical framing enabled us to analyze the regularities between the reported cases of the literature and to construct the ideal profiles of AI use in OGD lifecycle. Transversal to these four steps, our typology was iteratively evaluated using four criteria adapted from Scott & Davis (2015), ensuring its robustness and utility, namely:

– *intuitive sensibility*, indicating that the typology should be understandable and resonate with both academic and practical insights. This criterion informed the selection of analysis dimensions;
– *collective exhaustiveness and mutual exclusiveness*, ensuring that the typology covers all possible types without overlap. Our typology ensures this through a SLR, employing broad search terms to collect a wide range of reported AI uses in OGD;
– *construct validity*, confirming that the typology accurately reflects the theoretical concepts it is based on. As reported later, each dimension has defined spectrum values, well established in the literature;
– *conceptual elegance*, denoting the typology’s simplicity and clarity without sacrificing its depth and comprehensiveness. We keep this balance by providing an elegant summary matrix but delving into details when describing the identified types.

This evaluation process aimed to strengthen the validity and applicability of our typological framework.

3. AI x OGD Typology

Our typology introduces a categorization system anchored in two primary dimensions: the *perspective* and the *OGD lifecycle phases*. Regarding the **perspective**, we draw inspiration from the BPM field (Dumas et al., 2018) and distinguish between the "inward looking" and "outward looking" perspectives:

- **inward looking perspective**, or the portal perspective, focuses on the functionalities and operations/activities within OGD portals;
- **outward looking perspective** corresponds to the broader OGD ecosystem. This perspective extends beyond the boundaries of the portals themselves, encompassing the interaction between the OGD and its external stakeholders - businesses, governments, and the public.

Regarding the steps of the **OGD lifecycle**, we adhere to the fundamental lifecycle stages proposed by Attard et al. (2015), encompassing three stages, namely, data pre-processing, data exploitation and data maintenance. To facilitate a more fine-grained and detailed analysis of the literature and further classification, aligning with the user framework of (Crusoe et al., 2019), we further delineate between data exploration and data transformation within the exploitation phase. As such, OGD lifecycle encompasses:

- **data pre-processing**, consisting of data creation, data selection, data harmonization, data publishing;
- **data exploration,** consisting of data interlinking, data discovery, data exploration;
- **data transformation** consisting of datasets combination, and information/product/service development based on the aggregated data;
- **data maintenance** that involves data curation to ensure sustainability of the data by raising awareness, cleansing, enriching and updating.

Table 1 summarizes our typology, and the resulting 8 ideal types of AI uses in OGD, whereas we present every ideal type in the following sub-sections.

3.1. AI as Portal Curator

The first type deals with the **pre-processing stage** and is **inward-looking** as it deals with the portal perspective. The role of AI as a Portal Curator encapsulates a range of activities that facilitates the publication of data on OGD portals (e.g., tagging, labelling, categorization, quality check, modelling). In essence, this type automates or augments the role of data publishers and portal managers. Several examples can be found in the literature. AI can be used to anonymize data before publishing to protect individual privacy. This involves the development of a privacy risk model that detects personal and sensitive information to comply with legal and ethical standards, as seen in studies such as (Ali-Eldin et al., 2017).

Table 1. AI x OGD Typology

Perspective/OGD lifecycle	Pre-Processing	Exploration	Transformation	Maintenance
Inward Looking – Portal Perspective	**AI as Portal Curator** – AI facilitates the publication of OGD on the portal (e.g., automatic tagging or categorization (Yamada, 2023))	**AI as Portal Explorer** - AI helps users' exploration and interaction with the portal (e.g., natural language interfaces to access data (Albinali et al., 2023))	**AI as Portal Linker** – AI transforms and aggregates OGD to present it optimally to developers (e.g. transforming data into Linked OGD (Alsukhayri et al., 2020))	**AI as Portal Monitor** - AI monitors and ensures the compliance of the portal to standards (e.g., AI use to foster compliance and metadata quality (Šlibar & Mu, 2022))
Outward Looking - Ecosystem perspective	**AI as Ecosystem Data Retriever** – AI identifies and retrieves data from external sources (e.g., retrieving data from legal texts (Birghan et al., 2019))	**AI as Ecosystem Connector** – AI connects actors of the ecosystems based on their interests (e.g., recommending datasets of potential partners (Barcellos et al., 2017))	**AI as Ecosystem Value Developer** - AI supports the development of value based on OGD (e.g., suggesting appropriate analytics technique (Albinali et al., 2023))	**AI as Ecosystem Engager** - AI engages stakeholders in continued participation in the OGD ecosystem (e.g. through predicting the use of the portal (Wang et al., 2020))

Another example is AI-driven autocompletion, which assists less experienced users in data modeling by suggesting data structures and relationships (Pahl et al., 2021). AI technologies can also automate data cleaning and filtering (Afnan M. Alsukhayri et al., 2019). In feature extraction, AI technologies can help in handling issues related to missing data and the abundance of irrelevant features (Gazzaz & Rao, 2019). NLP techniques can also be applied to automatically link and index datasets based on descriptions provided by data publishers (Jiang et al., 2019). A last example resides in the automatic tagging and categorization of datasets, e.g., through ML based classification (Yamada, 2023). It must be noted that these examples sometimes automate the tasks of publishers (Sornkongdang et al., 2021), whereas sometimes augment the way they work by keeping them as key actors in the loop (Gazzaz & Rao, 2019). In our review, this type is generally well populated within the retrieved corpus.

3.2. AI as Ecosystem Data Retriever

This second type deals with the **pre-processing stage** and is **outward-looking** as it deals with the ecosystem perspective. The role of AI as Ecosystem Data Retriever enriches the data ecosystem by identifying, retrieving and integrating data from external sources in the larger OGD ecosystem, beyond the actual portal. In essence, this type augments or automates several phases often attributed to data publishers. Several examples can be found in the literature. Text mining can automatically analyze documents in ecosystems to extract data from these. Literature shows promising examples to simplify legal documents in structured data to facilitate public consumption (Birghan et al., 2019). Another type of document is online news where NLP techniques can help enrich data with this external source to provide a more comprehensive view to users (Sarmiento Suárez & Jiménez-Guarín, 2014). In urban environments, AI systems can support the gathering of essential data, such as pedestrian network data (Bolten & Caspi, 2022). A last example resides in the discovery and linking of heterogeneous datasets to ensure data integration within the actors of the OGD ecosystems (Sinif & Bounabat, 2018). It must be noted that these AI techniques mostly automate tasks as presented in the selected papers. This type is also well populated within our corpus. The readiness of the suggested technologies, however, remains at an earlier stage of maturity.

3.3. AI as Portal Explorer

This third type deals with the **exploration stage** and is **inward-looking** as it deals with the portal perspective. The role of AI as Portal Explorer includes opportunities to explore portal and its assets, including data available on it, along with facilitation of interaction with these artifacts by discovering and interlinking them. This is ensured predominantly using natural language interfaces to improve search capabilities, facilitate natural language interaction, and optimize data retrieval processes. In essence, this type augments the role of users (portal or data users depending on the artifact under exploration) and data publishers. Several examples can be found in the literature. One set of examples refers to enhancement of data accessibility and search. (Degbelo & Sherpa, 2020) proposes an approach for querying APIs through natural language terms to enhance API learnability and data retrieval. (Franciscatto et al., 2022), in turn, develops data search engine optimized for data retrieval and dataset integration based on metadata, utilizing word embeddings for semantic similarity. Alternatively, (Farazi et al., 2013) proposes semantic extension to a geo-catalogue that codifies knowledge about the geography of area created by reorganizing data extracted from the local geographical dataset with a semantic matching tool and a faceted ontology. Finally, Cantador et al. (2021) presented yet another example, which is the use of chatbots for searching and exploring OGD. Another set of examples stresses the role of NLP facilitating or enhancing interactions, especially when combined with advanced data analytics methods, including but not limited to ML (Albinali et al., 2023). Finally, in the light of current advances in the area and an increase in popularity of LLMs, first studies in this area emerged with (Mamalis et al., 2023), in which LLM are used to improve the access and retrieval of information from OGD portals. This type is very well populated within our corpus.

3.4. AI as Ecosystem Connector

This fourth type deals with the **exploration** stage of data lifecycle and is **outward-looking** as it deals with the ecosystem perspective. The role of AI as Ecosystem Connector is to explore the ecosystem to connect its actors by identifying and providing to the user recommendations of assets this actor can be interested in, such as datasets of potential partners (Barcellos et al., 2017; Gao & Janssen, 2020). Several examples can be found in the literature. The first set of examples delves into data exploration and analysis that goes beyond the OGD portal. Topic modeling and data visualization can be used to extract emerging topics from OGD for strategy formulation and identifying market opportunities, as exemplified by (Gottfried et al., 2021). Text mining proved to be instrumental in making complex open databases more effective for engineering design processes by automating data search and optimizing data transformation through optimal presentation (Giordano et al., 2022). Similar to (Gottfried et al., 2021), (Burkhardt et al., 2014) stress the importance and the role of visualization in data exploration, proposing an approach to handle multi-data through multi-visualization for a personalized information overview. It enables experts to investigate more information sources and generate a dashboard that allows more appropriate decision making. The exploration of external artifacts such as industrial innovation environments can also be facilitated by map-based dashboards designed with OGD, as showcased by Zuo et al. (2023). Finally, on a similar note, (Hsu & Lin, 2020) demonstrated how a linked data query platform can be used to personalize OGD suggestions by recommending them to Facebook users based on their social activity and personal preferences. This type is moderately populated within our corpus.

3.5 AI as Portal Linker

The fifth type is focused on the **data transformation stage** and is represented by the **inward-looking** perspective that encompasses features provided by OGD portals through which actors in different roles transform OGD. The role of AI as Portal Linker lies in providing, enhancing existing, or suggesting features and transformation tools and techniques, that transform, combine, and aggregate OGD to optimize their presentation for various user types such as developers. This type can be described as facilitator of the transformation of data into information, which can increase interest in datasets by automating and augmenting respective activities. This type primarily contributes to tasks performed by portal managers. Examples found in the literature show that AI can be used as machine reasoning technique to transform OGD into LOGD (to produce better knowledge graphs for customers) (Afnan Muhammadsalh Alsukhayri et al., 2020). Searching in datasets using semantics search is emphasized in (Birghan et al., 2019). AI (ML) can enhance the use of multivariate visualization of urban data, especially in tasks involving clustering. AI techniques can facilitate a deeper understanding of OGD by enabling more effective interpretation of complex information (Barcellos et al., 2017). The NLP techniques can be used to simplify and summarize the data available to make it more accessible to citizens (Lima et al., 2022). This type is not well populated within the retrieved corpus.

3.6 AI as Ecosystem Value Developer

The sixth type deals with **the data transformation stage** from the **outward-looking** perspective that provides a more complex view beyond the OGD portal and involves other elements, relationships between them, and dynamics of the ecosystem. The role of AI as Ecosystem Value Developer contributes most to value creation based on OGD through synergies between AI and other elements. This type augments or automates steps that should result in value creation through information/product/service development based on the aggregated data. It supports tasks such as assistance in coding, computing, ensuring correctness of transformed data, or distribution and presentation of outputs to target users. Both data publishers and data users are involved in and benefit from this type. Examples found in the literature show that AI can help by employing ML as a component of Query Editor Services that responsible for selecting the most suitable ML algorithm to perform on the data retrieved by the query (Albinali et al., 2023). Considering the outputs and data visualizations, creating dashboards that will be based on AI techniques (visual analytical methods) and tools can be a way (Zuo et al., 2023). This type is not well populated within the retrieved corpus.

3.7 AI as Portal Monitor

This seventh type deals with **the maintenance stage** and is **inward-looking** as it deals with the portal perspective. The role of AI as Portal Monitor deals with the compliance of the portal to standards (e.g., identifying anomalies, assessing data compliance, mitigating maintenance costs). In essence, this type automates or augments monitoring and maintenance activities. Several examples can be found in the literature. ML techniques can be used to identify data anomalies and missing values in OGD during the maintenance phase. Based on that, the portal manager can track these incidents and correct or delete the data (Karamanou et al., 2022). Similarly, metadata compliance can also be assessed by AI-driven computing techniques (Šlibar & Mu, 2022). Specifically, AI and cloud computing approaches can lower maintenance costs for OGD due to their automation and augmentation capabilities (Tan, 2023). This type is largely populated.

3.8 AI as Ecosystem Engager

This eighth type deals with **the maintenance stage** and is **outward-looking** as it deals with the ecosystem perspective. The role of AI as Ecosystem Engager deals with the engagement and participation in the OGD ecosystem. Some examples can be found in the literature. AI can be used to provide feedback and improve the usability of the portal, by providing methods and tools for engagement of users to audit data accuracy specifically centering the use of the data based on lived experience and information needs (Bolten & Caspi, 2022). In a different way, AI can be also used in a more proactive manner by predicting (based on population characteristics and other constructs) the usage of the portal (Wang et al., 2020). This type was the least populated within the retrieved corpus.

4. Theoretical and Practical Implications

The study's conceptual framework contributes to **theoretical advancements** of open government literature by identifying eight ideal types of AI applications. These applications can enhance the transparency, openness, accountability, and innovation of the OGD portals and the entire OGD ecosystem as one of open government pillars. It can be discussed how closely these concepts are interlinked and to what extent they require refinement of existing literature and open government initiatives. Our findings show that the key elements of the OGD ecosystem which benefit most from types of AI applications are *processes* and *stakeholders*.

The typological approach employed in the study offers a structured framework for categorizing and analyzing different roles of AI within OGD ecosystem, enhancing theoretical discourse on the integration of AI in governance processes and actor, including citizen engagement. Even though previous AI-focused typologies have been suggested in the literature, our study provides the first typology of AI in an OGD context. This framework also enables to identify the maturity of the literature in each type, showing that more work is needed to reach maturity in underexplored types such as AI as Ecosystem Value Developer or Ecosystem Engager. By formalizing ideal profiles of AI application within the OGD lifecycle, the study provides a theoretical foundation that can guide future research on leveraging AI to cultivate sustainable open data ecosystems and broader, namely, public data ecosystems (Lnenicka et al., 2024). This is particularly relevant given the rise of Generative AI, which has the potential to play a transformative role in the OGD ecosystem. We encourage future researchers to investigate AI in OGD beyond specific implements in a type such as in (Cortés-Cediel et al., 2023), but rather investigate the global architecture. We argue that an architectural view can combine several AI techniques to holistically support either a perspective (e.g., how to implement AI in a coherent manner on an OGD portal supporting all user phases), or a specific phase (e.g., how can AI support data transformation in OGD).

Our typology also contributes to research on skills and competencies associated with the work with OGD. Considering that AI-powered recommendation systems and virtual assistants permeate through all identified ideal profiles, they can be seen capable of bringing the OGD lifecycle phases closer to diverse groups of stakeholders. These tools can help overcome challenges that previously hindered certain stakeholders due to complexity, process clarity, or gaps in knowledge, skills, and competencies, effectively taking an open data intermediary role (Gonzalez-Zapata & Heeks, 2015). Through this lens, a specific AI technology can be investigated and developed to navigate several phases and perspectives in OGD (e.g., by designing an AI system enabling to balance the inward and outward perspectives when using an OGD portal). More generally, the proposed typology can be used in the Rigor cycle of Design Science Research (Schelhorn et al., 2024) to map the perspective and phases where the artefact can contribute.

Practitioners involved in OGD ecosystems can use the typology developed in the study to define their OGD ecosystem improvement agenda. By understanding phases of the OGD lifecycle with potentialities of AI within each phase, they can enhance OGD initiatives. This includes defining implementation scenarios and determining the appropriate level of autonomy—whether through augmentation or automation—at each step of the process. In addition, the proposed typology should help to guide the design and implementation of AI-driven solutions. E.g., OGD portal owners can consider implementation of AI as Portal Curator, Explorer, Linker, and Monitor, to optimize data retrieval, connectivity, and value creation. Policymakers, in turn, can leverage the identified ideal types of AI applications to inform strategic decisions on implementing AI in OGD initiatives, potentially enhancing the efficiency and effectiveness of governance processes. In addition, the proposed typology can help practitioners to better navigate and select among the myriads of AI technologies available, thanks to the formalization and classification of types of AI use in OGD.

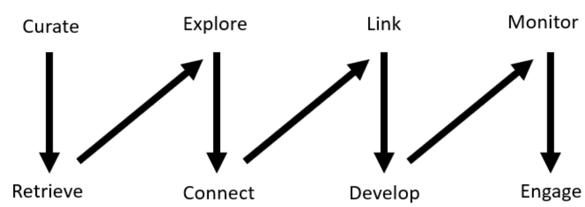

Figure 1. AI for OGD Process Model

Based on the typology developed, we suggest a process model (Figure 1) for OGD practitioners regarding the use of AI. The proposed sequence begins with *AI as Portal Curator*, which addresses the pre-processing stage by organizing and preparing data within the portal. Following this, *AI as Ecosystem Data Retriever* extends the scope outward by integrating external data sources into the ecosystem. The process then returns inward with *AI as Portal Explorer*, facilitating user interaction with the data. Next, *AI as Ecosystem Connector* bridges internal and external elements by linking actors within the ecosystem based on their interests. Subsequently, *AI as Portal Linker* transforms and optimizes data presentation within the portal, while *AI as Ecosystem Value Developer* focuses on creating value from these transformations through external partnerships and collaborations. The process continues with *AI as*

Portal Monitor, ensuring data quality and compliance within the portal. Finally, *AI as Ecosystem Engager* fosters ongoing engagement and participation across the ecosystem. This iterative approach, alternating between internal and external focus, ensures a comprehensive integration of AI across all phases of the OGD lifecycle. The proposed structured process should be empirically validated in future research to ensure that AI capabilities are developed in a balanced manner, enhancing both internal operations and the broader OGD ecosystem.

5. Limitations and Further Research

This study has several inherent limitations related to the scope of the work: the choice of the dimensions, the choice of the SLR methodology, and the choice of the application domain of OGD. First, the selected dimensions can be considered as limitations, where one of dimensions was tailored to inward and outward-looking perspectives. Future research could consider broadening this context to cover other dimensions (as additional or substitute for the selected), such as the level of AI implementation autonomy. The two archetype values of automation and augmentation surrounding OGD ecosystem (Rossello et al., 2024) can be used within this range (Johnson et al., 2022). Indeed, behind each ideal type, there are several tasks. These tasks can be either automated or augmented. However, the distinction is unclear in the literature, and thus our typology. In further research, we suggest expanding the typology by focusing on the role of AI across a continuum of tasks, ranging from entirely automated processes to those relying on humans.

Second, the study's reliance on a systematic review approach may introduce limitations related to the availability and quality of the included papers. Moreover, some studies are predominantly conceptual or at relatively early stage of development, whereas others, even if demonstrated in use, are not necessarily implemented. Instead, they primarily demonstrate the potential capabilities of AI within the OGD ecosystem. As such, this typology is rather instrumental to demonstrate not only actual but also most potential AI uses to enrich OGD initiative, thereby contributing to the resilience and sustainability of the OGD ecosystem. Furthermore, our systematic approach may still have omitted AI uses in the OGD lifecycle and OGD ecosystem due to (1) lack of research and articles on them, (2) fact that some implementations are being used internally in administrations. To address these limitations – at least in part – our future research will validate the inward-looking perspective of the designed typology by applying it to OGD portals to identify the presence and implementation of AI features corresponding to the identified "ideal profiles". Through this empirical examination, we aim to assess how the theoretical constructs manifest in real-world settings, providing insights into the current state of AI in OGD portals and identifying areas for future enhancement and research. The outward-looking perspective of typology will be validated through interviews with OGD ecosystem stakeholders. Since this perspective extends beyond the portals, encompassing the interaction between the OGD and its external stakeholders-such as businesses, governments, and the general population- a stakeholder analysis will be performed, with further interviews with respective stakeholders. As part of the typology validation, we will also identify scenarios where AI could be applied— even if not yet implemented— to enhance the typology and make our ideal profiles more forward-looking rather than reflecting current implementations. These steps aim to bridge the gap between theoretical typology and practical application through examining how AI is being or can be used in OGD.

Third, our focus has primarily been on the OGD ecosystem. However, the role of AI should also be examined within the broader context of the public data ecosystem, where the OGD ecosystem is just one component of a significantly larger network of public data sub-ecosystems (Lnenicka et al., 2024). This could provide more insights into the potentialities of AI. This broader examination is expected to be explored in future research.

Fourth, our paper does not explicitly discuss the challenges or limitations of AI within the OGD context. Instead, we focus on describing the capabilities of AI applications. However, AI risks should be also counted with and brought to the discussion within the community (Harrison & Luna-Reyes, 2022). To enrich our typology, we plan to incorporate key challenges, risks, and limitations associated with each ideal type. This will be part of our future work, which will proceed following the validation of the typology in line with the strategy outlined above.

6. Conclusion

Discussions surrounding the integration of AI and OGD have been limited to only a few studies around specific AI implementations in the OGD ecosystem. However, it emerges as a promising avenue for transforming governance, enhancing transparency, and fostering citizen engagement. To this end, in this study we explored the relationship between AI and OGD, aiming to advance scholarly discourse and practical insights. Drawing from the synthesis of the scattered literature through a systematic review, we examined how AI can enrich OGD initiatives. Employing a typological approach, we formalized ideal profiles of AI application within the OGD lifecycle, capturing varied roles across the portal and ecosystems perspectives. The resulting conceptual framework identifies eight ideal types of AI applications. This study aspires to contribute to a more nuanced understanding of the opportunities and challenges inherent in harnessing AI for OGD to tackle those ethical dilemmas. This theoretical foundation is aimed to inform policymakers, practitioners, and researchers in leveraging AI to cultivate sustainable open data ecosystem.

References

Albinali, A. A., Lock, R., & Phillips, I. (2023). The next generation of open data platform (ODP+): use case of Qatar. *Transforming Government: People, Process and Policy*.

Ali-Eldin, A., Zuiderwijk, A., & Janssen, M. (2017). Opening more data: A new privacy risk scoring model for open data. *BMSD 2017 - Proceedings of the 7th International Symposium on Business Modeling and Software Design*, 146–154.

Alsukhayri, Afnan M., Aslam, M. A., Arafat, S., & Aljohani, N. R. (2019). Leveraging the Saudi Linked Open Government Data: A Framework and Potential Benefits. *International Journal of Modern Education and Computer Science*, *11*(7), 14–22. https://doi.org/10.5815/ijmecs.2019.07.02

Alsukhayri, Afnan Muhammadsalh, Aslam, M. A., Saeedi, K., & Malik, M. S. A. (2020). A linked open data-oriented sustainable system for transparency and open access to government data: A case study of the public's response to women's driving in Saudi Arabia. *Sustainability (Switzerland)*, *12*(20), 1–26.

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418.

Bailey, K. (1994). *Typologies and taxonomies: An introduction to classification techniques*. SAGE.

Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., & Trevisan, D. (2017). Transparency in practice: using visualization to enhance the interpretability of open data. *In Proceedings of the 18th Annual International Conference on Digital Government Research (Dg.o '17), New York, NY, USA*, 139–148.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2020). Managing Artifical Intelligence. *Management Information Systems Quarterly*, *45*(3), 109–115.

Birghan, F., Hettenhausen, R., Siebenlist, T., & Meschede, C. (2019). Informing citizens via council information systems. *ACM International Conference Proceeding Series*, 280–286.

Bolten, N. J., & Caspi, A. (2022). Towards operationalizing the communal production and management of public (open) data: a pedestrian network case study: A pedestrian network case study in operationalizing communal open data. *ACM International Conference Proceeding Series*, *Par F18047*, 232–247.

Burkhardt, D., Nazemi, K., Retz, W., & Kohlhammer, J. (2014). Visual explanation of government-data for policy making through open-data inclusion. *2014 9th International Conference for Internet Technology and Secured Transactions, ICITST 2014*, 83–89.

Cantador, I., Viejo-Tardío, J., Cortés-Cediel, M. E., & Rodríguez Bolívar, M. P. (2021). A Chatbot for Searching and Exploring Open Data: Implementation and Evaluation in E-Government. *ACM International Conference Proceeding Series*, 168–179.

Cortés-Cediel, M. E., Segura-Tinoco, A., Cantador, I., & Rodríguez Bolívar, M. P. (2023). Trends and challenges of e-government chatbots: Advances in exploring open government data and citizen participation content. *Government Information Quarterly*, *40*(4).

Crusoe, J., Simonofski, A., Clarinval, A., & Gebka, E. (2019). The Impact of Impediments on Open Government Data Use: Insights from Users. *Proceedings of the 13th International Conference on Research Challenges in Information Science (RCIS)*.

Degbelo, A., & Sherpa, A. (2020). Open Geodata Reuse: Towards Natural Language Interfaces to Web APIs. *The Web Conference 2020 , WWW 2020*, *2*, 703–710.

Dumas, M., Rosa, L. M., Mendling, J., & Reijers, A. (2018). *Fundamentals of business process management*. Springer.

Farazi, F., Maltese, V., Dutta, B., Ivanyukovich, A., & Rizzi, V. (2013). A semantic geo-catalogue for a local administration. *Artificial Intelligence Review*, *40*(2), 193–212.

Franciscatto, M. H., Del Fabro, M. D., de Bona, L. C. E., Trois, C., & Tissot, H. (2022). Blending Topic-based Embeddings and Cosine Similarity for Open Data Discovery. *International Conference on Enterprise Information Systems, ICEIS - Proceedings*, *1*(Iceis), 163–170.

Gao, Y., & Janssen, M. (2020). Generating Value from Government Data Using AI: An Exploratory Study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12219 LNCS*, 319–331.

Gazzaz, S., & Rao, P. (2019). Extracting meaningful insights on city and zone levels utilizing US open government data. *The Web Conference 2019 - Companion of the World Wide Web Conference*,

WWW 2019, *2*, 1279–1285.
Giordano, V., Coli, E., & Martini, A. (2022). An Open Data Repository for Engineering Design: Using Text Mining with Open Government Data. *Computers in Industry*, *142*, 103738.
Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, *32*(4), 441–452.
Gottfried, A., Hartmann, C., & Yates, D. (2021). Mining open government data for business intelligence using data visualization: A two-industry case study. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(4), 1042–1065.
Harrison, T. M., & Luna-Reyes, L. F. (2022). Cultivating trustworthy artificial intelligence in digital government. *Social Science Computer Review*, *40*(2), 494–511.
Hsu, I. C., & Lin, Y. H. (2020). Integrated machine learning with semantic web for open government data recommendation based on cloud computing. *Software - Practice and Experience*, *50*(12), 2293–2312.
Jiang, S., Hagelien, T. F., Natvig, M., & Li, J. (2019). Ontology-Based Semantic Search for Open Government Data. *Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019*, 7–15.
Johnson, P. C., Laurell, C., Ots, M., & Sandström, C. (2022). Digital innovation and the effects of artificial intelligence on firms' research and development – Automation or augmentation, exploration or exploitation? *Technological Forecasting and Social Change*, *179*(April 2021).
Karamanou, A., Brimos, P., Kalampokis, E., & Tarabanis, K. (2022). Exploring the Quality of Dynamic Open Government Data Using Statistical and Machine Learning Methods. *Sensors*, *22*(24).
Li, Y., Yang, R., & Lu, Y. (2024). A privacy risk identification framework of open government data: A mixed-method study in China. *Government Information Quarterly*, *41*(1), 101916.
Lima, T., Brito, K., Nascimento, A., Valença, G., & Pedrosa, F. (2022). Using Natural Language Processing to Improve Transparency by Enhancing the Understanding of Legal Decisions. *CEUR Workshop Proceedings*, *3399*.
Lnenicka, M., Nikiforova, A., Luterek, M., Milic, P., Rudmark, D., Neumaier, S., ... & Rodríguez Bolívar, M. P. (2023). Understanding the development of public data ecosystems: from a conceptual model to a six-generation model of the evolution of public data ecosystems. *Available at SSRN 4831881*.
Mamalis, M. E., Kalampokis, E., Karamanou, A., Brimos, P., & Tarabanis, K. (2023). Can Large Language Models Revolutionalize Open Government Data Portals? A Case of Using ChatGPT in statistics.gov.scot. *ACM International Conference Proceeding Series*, 53–59.
Pahl, M. O., Bauer, F., & Luubben, C. (2021). Pipeline for Crowdsourced IoT Data-Modeling with AI-Supported Convergence. *Proceedings of the IM 2021 - 2021 IFIP/IEEE International Symposium on Integrated Network Management*, 702–706.
Rossello, N. B., Simonofski, A., Clarinval, A., & Castiaux, A. (2024). A Typology for AI-enhanced Online Ideation : Application to Digital Participation Platforms. *Proceedings of the 57th Annual Hawaii International Conference on System Sciences*.
Sarmiento Suárez, D., & Jiménez-Guarín, C. (2014). Natural language processing for linking online news and open government data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8823*, 243–252.
Schelhorn, T. C., Gnewuch, U., & Maedche, A. (2024). Designing a Large Language Model Based Open Data Assistant for Effective Use. *19th International Conference on Design Science Research in Information Systems and Technology*, *May*.
Scott, W. R., & Davis, G. F. (2015). *Organizations and organizing: Rational, natural and open systems perspectives*. Routledge.
Sinif, L., & Bounabat, B. (2018). Approaching an optimizing open linked government data portal. *ACM International Conference Proceeding Series*, 135–139.
Šlibar, B., & Mu, E. (2022). OGD metadata country portal publishing guidelines compliance: A multi-case study search for completeness and consistency. *Government Information Quarterly*, *39*(4).
Sornkongdang, N., Sanglerdsinlapachai, N., & Anutariya, C. (2021). DataCat: Attention-based Open Government Data (OGD) Category Recommendation Framework. *16th International Joint Symposium on Artificial Intelligence and Natural Language Processing, ISAI-NLP 2021*, 1–6.
Stapley, E., O'Keeffe, S., & Midgley, N. (2022). Developing Typologies in Qualitative Research: The Use of Ideal-type Analysis. *International Journal of Qualitative Methods*, *21*, 1–9.
Tan, E. (2023). Designing an AI compatible open government data ecosystem for public governance. *Information Polity*, *28*(4), 541–557.
Ugale, G. and Hall, C. (2024). *Generative AI for anti-corruption and integrity in government: Taking stock of promise, perils and practice*.
Wang, D., Richards, D., Bilgin, A. A., & Chen, C. (2020). A Predictive Model for Citizens' Utilization of Open Government Data Portals. *International Conference on Asian Digital Libraries*, *2*, 159–175.
Yamada, Y. (2023). Tag Recommendation System for Data Catalog Site of Japanese Government. *International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K - Proceedings*, *3*(Ic3k), 325–331.
Zuo, C., Ding, L., Liu, X., Zhang, H., & Meng, L. (2023). Map-based dashboard design with open government data for learning and analysis of industrial innovation environment. *International Journal of Cartography*, *9*(1), 97–113.