The University of Pardubice
Faculty of Economics and Administration
Department of System Engineering and Informatics

Open and linked data management platforms

Bachelor's Thesis

2024 Md Deloar Hosen

University of Pardubice
Faculty of Economics and Administration
Academic year: 2023/2024

ASSIGNMENT OF BACHELOR THESIS

(project, art work, art performance)

Name and surname:	**Md Deloar Hosen**
Personal number:	**E21885**
Study programme:	**B0688A140005 Informatics and System Engineering**
Specialization:	**Informatics in Public Administration**
Work topic:	**Open and linked data management platforms**
Assigning department:	**Institute of System Engineering and Informatics**

Theses guidelines

The aim of the thesis is to identify and compare open and linked data management platforms and develop a model to select the most suitable one for a given case.
Outline:
– Reference to the theoretical background of the subject matter
– Identification and description of open and linked data management platforms
– Comparison of selected platforms
– Model development and evaluation
– Results and discussion

Extent of work report:	**approx. 35 pages**
Extent of graphics content:	
Form processing of bachelor thesis:	**printed/electronic**
Language of elaboration:	**English**

Recommended resources:

ALI, Mohsan; ALEXOPOULOS, Charalampos; CHARALABIDIS, Yannis. A comprehensive review of open data platforms, prevalent technologies, and functionalities. In: *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*. ACM, 2022. p. 203-214.
CHARALABIDIS, Yannis, et al. *The world of open data: Concepts, methods, tools and experiences*. Public Administration and Information Technology. Cham, Springer International Publishing, 2018.
HEATH, Tom; BIZER, Christian. *Linked data: Evolving the web into a global data space*. Cham, Springer, 2011.
MÁCHOVÁ, Renáta; LNĚNIČKA, Martin. A multi-criteria decision making model for the selection of open data management systems. *Electronic Government, an International Journal*, 2019, 15.4: 372-391.
WĘCEL, Krzysztof. *Big, open and linked data: Effects and value for the economy*. Business Information Systems. Cham, Springer, 2022.

Supervisors of bachelor thesis:	**Ing. et Ing. Martin Lněnička, PhD.** Institute of System Engineering and Informatics
Date of assignment of bachelor thesis:	**September 1, 2023**
Submission deadline of bachelor thesis:	**April 30, 2024**

L.S.

prof. Ing. Jan Stejskal, Ph.D. Dean	**Ing. et Ing. Martin Lněnička, PhD.** study programme guarantor

In Pardubice September 1, 2023

Declaration:

I declare:

The thesis entitled "*Open and linked data management platforms*" is my own work. All literary sources and information that I used in the thesis are referenced in the bibliography.

I have been acquainted with the fact that my work is subject to the rights and obligations arising from Act No. 121/2000 Sb., On Copyright, on Rights Related to Copyright and on Amendments to Certain Acts (Copyright Act), as amended, especially with the fact that the University of Pardubice has the right to conclude a license agreement for the use of this thesis as a school work under Section 60, Subsection 1 of the Copyright Act, and that if this thesis is used by me or a license to use it is granted to another entity, the University of Pardubice is entitled to request a reasonable fee from me to cover the costs incurred for the creation of the work, depending on the circumstances up to their actual amount.

I acknowledge that in accordance with Section 47b of Act No. 111/1998 Sb., On Higher Education Institutions and on Amendments to Other Acts (Higher Education Act), as amended, and the Directive of the University of Pardubice No. 7/2019 Rules for Submission, Publication and Layout of Theses, as amended, the thesis will be published through the Digital Library of the University of Pardubice.

In Pardubice on August 28, 2024 Md Deloar Hosen by own hand

ANNOTATION

The thesis paper summarizes the principles and practices of open and linked data management platforms. It explores the basic concepts of open data and linked data, explores their benefits and challenges, and offers practical use cases to illustrate the applications of these platforms. It provides a decision-making model consisting of criteria and alternative identified through a literature review. Using a public administration use case, the model is applied and the most suitable alternative recommended, including discussion of best practices.

KEYWORDS

Open data, linked data, data management platform, model, criteria, alternatives.

NÁZEV

Platformy pro správu otevřených a propojených dat

ANOTACE

Práce shrnuje principy a postupy otevřených a propojených platforem pro správu dat. Zkoumá základní koncepty otevřených dat a propojených dat, zkoumá jejich výhody a výzvy a nabízí praktické příklady použití pro ilustraci aplikací těchto platforem. Představuje rozhodovací model skládající se z kritérií a alternativ identifikovaných prostřednictvím přehledu literatury. Na případu užití z oblasti veřejné správy je pak model aplikován a doporučena nejvhodnější alternativa včetně diskuse osvědčených postupů.

KLÍČOVÁ SLOVA

Otevřená data, propojená data, systém pro správu dat, model, kritéria, alternativy.

Table of contents

List of figures

List of tables

List of abbreviations

AHP	Analytic Hierarchy Process
APIs	Application Programming Interfaces
AI	Artificial Intelligence
CKAN	Comprehensive Knowledge Archive Network
CR	Consistency Ratio
DKAN	Drupal Knowledge Archive Network
CSV	Comma-Separated Values
JSON	JavaScript Object Notation
ML	Machine Learning
NLP	Natural Language Processing
RDF	Resource Description Framework
XML	eXtensible Markup Language

INTRODUCTION

The increasing digitization of information and the rapid development of data generation have paved the way for innovative methods of data management, of which open and linked data management platforms are prominent. Open data refer to data that are freely available to anyone to use, edit, and share, while linked data is a way to structure and interconnect data on the web to improve its use and discovery. These concepts are very important in today's data-driven world, which offer potential for transparency, cooperation, and innovation in various fields, including government, healthcare, education, and business. The use of suitable open and linked data platforms enables the creation of a semantic web, where data are not only accessible on data portals but meaningfully interconnected through data ecosystems, allowing for more efficient retrieval and analysis of data.

Many open data portals and infrastructures have been launched as a part of the so-called open government movement, with the goal of offering a single point of access for government data and encourage greater public involvement, teamwork, and cooperation (Gray, 2023; Máchová and Lněnička, 2017; Máchová and Lněnička, 2019; Zuiderwijk et al., 2014). Data.gov first time was launched in USA government. Then around 150 governments and authorities have made mor then one million datasets available by open data portals (Gray, 2023). Open data generating a new business in the world to providing peoples with information that they can aggregate and combine in unique ways (Charalabidis et al., 2018; Lopez et al., 2012).

The aim of this thesis is to identify and compare open and linked data management platforms and develop a model to select the most suitable one for a given case. To address this aim, this thesis first presents a theoretical background on the topic. The next chapter is focused on description and comparison of open and linked data management platforms, including trends and innovations in these platforms. Model development and its evaluation are included in the next chapter. It also provides the research methodology and design. Results are discussed in the forthcoming chapter. Main contributions of this thesis lie in the development of a decision-making model consisting of criteria and its application for the evaluation of a use case using selected open and linked data management platforms (alternatives). This helps to identify the most suitable open and linked data management platforms.

1 THEORETICAL BACKGROUND

1.1 OPEN GOVERNMENT, OPENNESS AND TRANSPARENCY

Open government refers to the view that citizens have the right to access official documents and operations to allow for effective public surveillance. The principles of openness, accountability, and transparency are fundamental to open government, fostering trust and engagement while ensuring that government actions are visible and understandable to citizens (Wirtz and Birkmeyer, 2015).

The history of open government dates to the age of enlightenment, where transparency and public accountability began to take shape. Over the years, these ideas have evolved inspired by democratic movements and technological advances. Early milestones include Sweden's Freedom of the Press Act of 1766 and the United States Freedom of Information Act of 1966. The digital age has further accelerated the open government movement with initiatives such as different international and national open government strategies and polices as well as open data portals (Dawes and Helbig, 2010; Lněnička et al., 2021).

1.1.1 National strategies and policies

Many countries have adopted national strategies to promote open government and transparency. For example, the United States Open Government Directive from 2009 mandates federal agencies to publish official information online. The UK's Open Data White Paper from 2012 outlines promises to make public data openly available to promote innovation and economic growth. The Open Data Strategy of Germany from 2021 then provides a framework for action to improve the Federal Government's open data ecosystem. In the Czech Republic, open data are included in strategic framework Czech Republic 2030.

1.1.2 Open Government Partnership

The Open Government Partnership (OGP), launched in 2011, and in 2024 includes 75 countries, is a multilateral initiative promoting transparent, participatory and accountable governance. OGP provides a platform for governments to make concrete commitments to open up government reforms and be accountable for their implementation. Participating countries develop national action plans outlining specific commitments to improve transparency,

empower citizens, fight corruption, and harness new technologies to strengthen governance (Open Government Partnership, 2024).

1.2 OPEN AND LINKED DATA

Open data are data that anyone can access, use, and share. The concept has emerged from open access and open-source mobility, which advocates transparency, collaboration, and free flow of information. Linked data are a way of publishing structured data on the web that is interconnected and more useful, which forms the basic concept of the semantic web (Heath and Bizer, 2011; Wirtz and Birkmeyer, 2015). The open data movement gained momentum in the early 2000s with initiatives such as the Open Data Handbook and organizations such as the Open Knowledge Foundation. Important milestones include the launch of data.gov in the United States (2009) and similar initiatives in other countries (Zhu, 2017). The concept of linked data, introduced by Tim Berners-Lee, has evolved to facilitate the interconnection and use of data on the web (Heath and Bizer, 2011). Open data are guided by principles such as accessibility, machine readability, and reuse. These rules ensure that data are available and usable through a wide range of stakeholders. The Open Data Charter provides principles for making data by default, timely, comprehensive, accessible, usable, comparable, and actionable (Borgesius et al., 2015).

Connection between linked and open data can be created from open data to increase their value and utility. This makes it possible to connect and integrate various databases, producing new and richer knowledge. When released as open data, linked data optimize usability and accessibility while guaranteeing that the information is connected and compatible with various platforms and systems. (Fan and Zhao, 2017). Węcel (2022) emphasized that data themselves have no intrinsic value. Their worth is determined by their usage. The authors explored how big, open, and linked data can be combined to create new value and highlighted innovation as a primary driver for creating value (Janssen and Kuk, 2016).

Open and linked data can be published in various formats to improve accessibility and usability. Common formats include (Charalabidis et al., 2018; Hert et al., 2011; Pezoa et al., 2016):

- CSV (Comma-Separated Values): is a popular and easy-to-use file format for tabular data storage. The format is particularly well-liked for data interchange across various applications or systems since it arranges data in a way that is simple to read and modify.

- JSON (JavaScript Object Notation): is a simple data-interchange format that is simple for computers to generate and parse as well as for humans to read and write. JSON is language-independent; thus, it works with more programming languages than simply JavaScript, despite its name.
- RDF (Resource Description Framework): is a standard model for describing and exchanging data on the web that was created by the World Wide Web Consortium. The Semantic Web, which allows data to be shared and reused across various applications, is built on this structured, machine-readable representation of resource information.
- XML (eXtensible Markup Language): is a versatile, organized text format for data storage and transfer. It is a commonly used format for data interchange between various systems, particularly in online applications, because it is both machine- and human-readable.

1.2.1 Benefits of open and linked data

Open data guarantees that information is available to the public, formerly inaccessible data can now be accessed and analyzed by citizens, journalists, researchers, and other stakeholders. As a result of people's ability to examine and assess the actions, regulations, and choices made by organizations promotes a culture of transparency (Dawes and Helbig, 2010; Lněnička and Nikiforova, 2021). Organizations can be held responsible for their activities by releasing publicly available data. Publicly available government expenditure data, for instance, enables citizens to monitor the use of tax dollars, which can lessen corruption and enhance governance (Lourenço et al., 2017). By allowing similar data from several sources to be connected, linked data increases transparency. Because of their interconnection, complicated situations can be understood more thoroughly (Janssen et al. 2012; Matheus and Janssen 2015).

Automated analysis and real-time transparency are made possible by linked data's easy-to-read format for machines. While linked data improve the ability to connect, interpret, and analyze this information, offering deeper insights and a more complete view of transparency, open data makes information freely available, creating trust and responsibility. When paired, they offer a potent toolkit for fostering transparency and accountability in numerous sectors (Janssen and Kuk, 2016; Lourenço et al., 2017; Matheus and Janssen 2015). Better decision-making is facilitated by open and linked data in several domains, including industry, government, academia, and non-profits. These are the ways that each lead to better decision-making (Węcel,

2022; Zuiderwijk and Janssen, 2013). Decision-making can be substantially improved when open data are published and linked using linked data approaches. A greater range of linked data is available to decision-makers, resulting in more thorough insights and efficient actions (Charalabidis et al., 2018; Heath and Bizer, 2011).

Open data give companies useful information about consumer behavior, market trends, and economic indicators. This data lowers uncertainty and promotes a more effective market by assisting businesses in making well-informed decisions (Jaakkola et al., 2014). Open data serve as a catalyst for economic growth by fostering new business opportunities and innovation. Businesses can leverage open data to develop new products and services, build applications, and improve existing solutions (Cruz and Lee 2015). Open and linked data are effective instruments for innovation and economic expansion. They facilitate more informed decision-making, encourage the development of new markets and services, and create an atmosphere that is conducive to innovation by making data openly accessible and interconnected (Janssen et al. 2012; Jetzek et al., 2014).

Connecting socioeconomic, health, and education data, for instance, can reveal insights into the variables influencing public health outcomes and allow for more comprehensive approaches to service delivery (Toots et al., 2017). Both open data and linked data have the potential to revolutionize public services by making data more accessible, interoperable, and actionable. They empower citizens and service providers alike to collaborate on creating more transparent, efficient, and innovative public services that better meet the needs of society. Additionally, open data initiatives can increase public participation by involving citizens in the data collection, analysis, and decision-making process (Lněnička and Nikiforova, 2021; Lourenço et al., 2017).

Open data initiatives can promote social equality by providing equal access to information for all citizens, regardless of their socioeconomic status (Cinnamon, 2020). Data that are open and linked can be very effective in promoting inclusion and social justice. These data techniques can assist in addressing systemic inequalities and advancing a more equitable and inclusive society by enhancing information access, strengthening advocacy, aiding targeted service provision, and encouraging inclusive decision-making. But achieving this potential will need paying close attention to moral issues, including the community, and creating instruments and regulations that guarantee data is used for everyone's advantage (Jaakkola et al., 2014; Toots et al., 2017).

1.2.2 Challenges of open and linked data

Open data quality and accuracy describe the dependability, accuracy, and usability of publicly accessible data. Governments, corporations, and other institutions may give open data, which is frequently utilized for public accountability, policy-making, and research. Because the judgments and conclusions made using the data are impacted, it is imperative to ensure the quality and veracity of open data. The quality and accuracy of linked data as well as open data are critical for facilitating successful data-driven research, innovation, and decision-making. However, because handling large-scale, interconnected information involves a variety of sources, formats, and complications, maintaining good standards in these areas can be difficult (Fan and Zhao, 2017; Jaakkola et al., 2014; Vetrò et al., 2016; Węcel, 2022).

Security and privacy of open and linked data raise a lot of concerns, particularly considering the possibility of abuse, data breaches, and unexpected outcomes. These data are intended to be shared and freely available, which promotes innovation and transparency but also increases privacy and security issues (Janssen and van den Hoven, 2015). It is possible for someone to be re-identified even after their personal information has been anonymized if datasets are combined or if advanced data analysis methods are used. For instance, to identify specific people, open datasets with location, health, or shopping patterns may be cross-referenced with other public data. Because linked data is designed to allow datasets to be joined and interlinked, correlated datasets may unintentionally reveal private information (Charalabidis et al., 2018; Heath and Bizer, 2011).

Accessing, integrating, and using open and connected data can be technically difficult. Users often need specialized knowledge and skills to work with different data formats, standards, and tools. Additionally, implementing the infrastructure needed to store, manage, and process large datasets can be expensive and complex (Janssen et al. 2012). The absence of standardization in data formats, metadata, and licensing can hinder the effective use and sharing of open data. Without common standards, it can be difficult to combine and compare datasets from different sources, limiting the potential benefits of open data (Jaakkola et al., 2014).

The release and use of open data involves navigating complex legal and ethical issues. Questions about data ownership, intellectual property rights, and the ethical use of data can complicate the sharing and use of open data. Organizations must carefully consider these factors to avoid legal disputes and ensure ethical practices (Charalabidis et al., 2018; Janssen et al. 2012).

1.3 DATA MANAGEMENT PLATFORMS

Open data management platforms are systems designed to publish, manage, and share open data. They provide tools for storing, retrieving, and visualization of data, increasing data access and usage (Máchová and Lněnička, 2019). These are made expressly to manage publicly accessible data, which is frequently made available to the public for usage by institutions, governments, and other organizations. These systems concentrate on making sure that a broad spectrum of users, including developers, academics, and the public, can readily access, understand, and utilize the data. The goal of open data publishing is to make data easily discoverable, useable, and comprehensible for the general audience. Promoting accountability, openness, and public involvement is the aim. Data must be structured for linked data publishing so that it may be related to other datasets via established web protocols. This method makes it possible to create a "web of data" where data from several sources may be combined and analysed simultaneously (Ali et al., 2022; Alexopoulos et al., 2014; Boch et al., 2022; Máchová and Lněnička, 2019).

Making publicly available data discoverable and accessible is the main goal of open data cataloging. It entails compiling datasets into an organized inventory that the public and organizations can browse and use with ease. Organizing and documenting data in a way that highlights the linkages between various datasets is known as linked data cataloging. It focuses on enabling querying and accessibility of datasets within a semantic web environment. Access to publicly available data in a structured, machine-readable manner is made possible by open data Application Programming Interfaces (APIs). Developers, scholars, and other users can access, modify, and examine publicly available datasets from different sources with the use of these APIs (Ali et al., 2022; Boch et al., 2022; Correa et al., 2018).

Alexopoulos et al. (2014) proposed a new generation of open data platforms that combine traditional data access features with social media elements such as data processing, feedback and collaboration, data quality ratings, and grouping and interaction capabilities. The study by Ali et al. (2022) identified a wide range of open data platforms, each with unique features and capabilities. They also highlighted emerging trends in open data platforms, such as increased focus on data quality and governance, integration with Artificial Intelligence (AI) and Machine Learning (ML), or development of more user-friendly interfaces. A systematic review by Boch et al. (2022) provides a comprehensive overview of existing data management platforms. The authors analysed various platforms, focusing on their functionalities, features, and potential use

cases. Most platforms share common features, such as data integration, metadata management, and security measures.

Correa et al. (2018) then proposed a methodology for automatically investigating open data portals to understand the characteristics, quality, and usability of these portals. They provided several recommendations for improving their design, functionality, and user experience. Ghahremanlou et al. (2019) provided an overview of open data platforms utilized in six United Kingdom smart city initiatives and analysed the platforms' functionalities, data types, and governance models. Máchová and Lněnička (2019) proposed a multi-criteria decision-making model to assist organizations in selecting appropriate open data management systems. The model aims to provide a systematic and objective approach to evaluating various options based on multiple criteria. Scanlon (2021) outlined a comprehensive framework for managing open data portals, designed to ensure sustainability and effectiveness.

Examples of open data management platforms include Comprehensive Knowledge Archive Network (CKAN), Drupal Knowledge Archive Network (DKAN), Socrata, OpenDataSoft, and ArcGIS Hub. Table 1 provides an overview of most common criteria used to evaluate and compare open data management platforms, including references.

Table 1. An overview of criteria used to compare open data management platforms. Source: own processing.

Criteria	**References**
Language versions	Ghahremanlou et al. (2019); Máchová and Lněnička (2019)
Search engine (filter)	Ali et al. (2022); Ghahremanlou et al. (2019); Máchová and Lněnička (2019)
User account	Máchová and Lněnička (2019)
Forum (feedback)	Alexopoulos et al. (2014); Máchová and Lněnička (2019); Scanlon (2021)
Help (documentation)	Ghahremanlou et al. (2019); Máchová and Lněnička (2019)
Social media (share and request)	Alexopoulos et al. (2014); Máchová and Lněnička (2019); Scanlon (2021)
Full data API	Ali et al. (2022); Boch et al. (2022); Ghahremanlou et al. (2019); Máchová and Lněnička (2019); Scanlon (2021)

Criteria	References
Linked data access	Ali et al. (2022); Ghahremanlou et al. (2019); Máchová and Lněnička (2019)
Data visualisation and analytics tools	Alexopoulos et al. (2014); Ali et al. (2022); Boch et al. (2022); Correa et al. (2018); Ghahremanlou et al. (2019); Máchová and Lněnička (2019); Scanlon (2021)
Thematic categories	Correa et al. (2018); Máchová and Lněnička (2019)
Tags (keywords)	Máchová and Lněnička (2019)
Dataset metadata (description)	Ali et al. (2022); Correa et al. (2018); Máchová and Lněnička (2019); Scanlon (2021)
Open data license	Ghahremanlou et al. (2019); Máchová and Lněnička (2019); Scanlon (2021)
User rating and comments	Máchová and Lněnička (2019); Scanlon (2021)
Data integration capabilities	Ali et al. (2022); Boch et al. (2022); Correa et al. (2018)
Data governance	Ali et al. (2022); Boch et al. (2022); Correa et al. (2018); Ghahremanlou et al. (2019); Scanlon (2021)
Security and compliance	Boch et al. (2022); Ghahremanlou et al. (2019)
Interoperability and flexibility	Correa et al. (2018); Ghahremanlou et al. (2019)
Data analytics and reporting	Ali et al. (2022); Boch et al. (2022); Ghahremanlou et al. (2019)
Cost and ROI	Scanlon (2021)
Reputation and track record	Boch et al. (2022)
Deployment options	Correa et al. (2018); Ghahremanlou et al. (2019)
Data backup and recovery	Ali et al. (2022); Boch et al. (2022)

2 DESCRIPTION AND COMPARISON OF OPEN AND LINKED DATA MANAGEMENT PLATFORMS

This section provides an overview of existing open and linked data management platforms.

2.1 IDENTIFICATION AND DESCRIPTION OF PLATFORMS

These platforms facilitate transparency, interoperability, and efficient data reuse across multiple domains by offering the fundamental infrastructure for the administration and distribution of open and linked data.

ArcGIS Hub

ArcGIS Hub is an open data platform developed by Esri that allows organizations to share data and engage with the community. It provides tools for data visualization, analysis, and collaboration, enabling users to take actions, track progress, and communicate outcomes. Organizations can promote openness and public participation by making their datasets publicly accessible through ArcGIS Hub. Dashboards, infographics, and interactive maps can be used to share these datasets. It provides an easily searchable, filterable, and downloadable dataset central data catalogue for users. Policymakers, researchers, and people can work together on data-driven projects as stakeholders. ArcGIS Hub facilitates the establishment of initiatives that allow these groups to collaborate on shared objectives, exchange insights, and offer feedback. Websites (Hub Sites) customized for particular campaigns or target audiences can be made by users. These websites can house a variety of tools, data sources, and interactive elements including feedback forms and surveys. Through integration with the larger ArcGIS platform, ArcGIS Hub users can take advantage of tools for mapping, spatial analysis, and visualization. Through APIs, it also facilitates integration with various systems and data sources. With ArcGIS tools, users may create custom mobile and web applications, opening new possibilities for data interaction (Butorac, 2021; Esri, 2024).

CKAN

CKAN is an open-source data management system that facilitates the publication, sharing, discovering, and use of data by offering tools. It offers a centralized platform for data sharing and is extensively utilized by governments, companies, and communities to manage and disseminate open data. Datasets with metadata can be readily published by users, improving

accessibility and usability. It supports a variety of data types, such as JSON and CSV, and offers an API for data management and automatic publishing. Users can locate pertinent datasets with the aid of sophisticated search and filtering features. Full-text search and faceted navigation are supported. With the data visualization capabilities included in CKAN, users may make charts and graphs right on the site (Kim et al., 2021).

DataPress

A program called DataPress was created to make data administration easier, especially for projects and businesses that rely heavily on data. Publishing, sharing, and organizing data are common uses for it. An overview of DataPress's approach to data management is provided by Halabi (2018). DataPress allows organizations to catalogue their datasets, making it easier to organize, search, and manage data. Datasets are stored with metadata, making them more discoverable and understandable for users. The platform enables controlled access to data, allowing users to share datasets with specific groups or keep them public. APIs are often provided for programmatic access to data, enhancing interoperability with other tools and systems. DataPress is a data management platform that offers tools for data publishing, visualization, and sharing. It supports various data formats and provides features for data discovery and exploration, making it suitable for open data initiatives (DataPress, 2024; Halabi, 2018).

DKAN

An open-source data management platform called DKAN was created to make it simple for businesses to store, publish, and exchange their data. Governments, non-profit organizations, and other organizations frequently utilize it to build open data portals. This is DKAN's approach to data management. Because it is open-source, developers can add to its features and combine it with other systems with greater freedom. Being open source, it provides flexibility for developers to extend its functionality and integrate it with other systems. DKAN provides a stable and adaptable solution for data management. Its open-source design allows it to be customized to meet different requirements, and its integrated features for data sharing, categorization, and visualization make it an all-inclusive platform for data management (CivicActions, 2023; Seto and Sekimoto, 2016).

OpenDataSoft

Being open source, it provides flexibility for developers to extend its functionality and integrate it with other systems. It was created to facilitate the effective administration, exchange, and display of data, especially when it comes to open data efforts. Governments, organizations, and companies use it extensively to publish and manage datasets so that the public or particular user groups can access them. Data can be imported into OpenDataSoft from a number of sources, such as databases, files (CSV, Excel, JSON, etc.), web scraping, and APIs. Because of the platform's support for real-time data integration, datasets can be updated continuously. Within the platform, users can clean, transform, and enrich datasets. This covers metadata addition, geocoding, and data purification. Data can be connected to other datasets to produce more useful and extensive resources. Because of its features for data intake, processing, visualization, and secure sharing, it is a well-liked option for businesses looking to leverage data for internal data management, smart city projects, or public transparency (Correa et al., 2018; Opendatasoft, 2024).

Socrata

A cloud-based data management tool called Socrata was created expressly to assist public sector and governmental entities in sharing, managing, and analyzing their data. It focuses on improving data's openness, usability, and accessibility for the public and government organizations. Governments can publish datasets for public access on open data portals created with Socrata, which is extensively used for this purpose. It makes it possible to publish a variety of data kinds, such as financial, transportation, medical, and public safety data. Additionally, it facilitates API connectivity, which makes it simpler for developers to create applications that use government data and access data programmatically. With a focus on open data, transparency, and public involvement, Socrata offers a potent platform for managing data in the public sector. It gives governments the instruments they need to properly manage their data while simultaneously opening it out to the public and making it valuable. With capabilities for data categorization, visualization, analytics, and governance, Socrata helps a variety of public sector data-driven projects (Ham and Crockett, 2021; Tyler Technologies, 2024).

2.2 COMPARISON OF SELECTED PLATFORMS

Table 1 in chapter 1.3 provided the overview of criteria used to compare open data management platforms. Based on these findings, the following descriptions were used to compare selected open and linked data management platforms. The results are in Table 2.

- Name of the platform.
- Is the data management system free or paid?
- Who is the developer?
- What features does the platform offer?
- What types of data formats are supported?
- Is the platform open-source or proprietary?
- Is there support and documentation available?

Table 2. Overview of open and linked data platforms. Source: own processing.

Platform	**Free / paid**	**Developer**	**Features**	**Data formats supported**	**Open source / proprietary**	**Support and documentation**
ArcGIS Hub	Paid	Esri	Data storage, management, sharing, visualization, APIs	CSV, JSON, Shapefiles, GeoJSON	Proprietary	Extensive documentation, community support, and forums
CKAN	Free	Open Knowledge Foundation	Data storage / management, customizable metadata, full-text search, data visualization, plugins – API, access for data integration	CSV, XLS, JSON, XML, RDF, TSV, SHP	Open-Source	Extensive documentation, community support, and forums
Datapress	Free/ Paid	Datapress	Data storage and sharing, version control, data validation, API access,	CSV, JSON, XLSX, TSV, XML	Open-Source	Community-driven support, detailed documentation, and tutorials

Platform	Free / paid	Developer	Features	Data formats supported	Open source / proprietary	Support and documentation
			integrations with various tools			
DKAN	Free	Community	Dataset management, APIs, visualization, customizability and flexibility	CSV, TSV, XLS/XLSX, shapefile, GeoJSON	Open-Source	Official documentation, community support
OpenDataSoft	Paid	Opendatasoft	Organizations to share, visualize, and analyse data	CSV, Excel (XLS, XLSX), TSV, SHP, SHX, DBF	Proprietary	Comprehensive documentation, community forum, training and webinars
Socrata	Paid	Tyler Technologies	Data hosting and visualization, built-in analytics, APIs, integration with other tools	CSV, JSON, XLS, XML, RDF	Proprietary	Official support, comprehensive documentation, and customer service

2.3 TRENDS AND INNOVATIONS IN OPEN DATA PLATFORMS

2.3.1 Emerging technologies in open data management

Emerging technologies are revolutionizing open data management, making data more efficient, secure, and intelligently handled. The integration of these technologies with open and linked data platforms can significantly improve their capabilities and broaden their applications.

First, data management is undergoing a transformation because to AI and ML, which improve analytics, automate procedures, and facilitate more thoughtful decision-making. Algorithms based on AI and ML automate the process of locating and fixing problems in datasets, including duplicate entries, missing values, and inconsistencies. As a result, there is less manual involvement and better data quality (Gupta and Kumar, 2023). Data can be tagged and classified by ML algorithms using patterns, context, and content. Open data can be categorized using Natural Language Processing (NLP) techniques into predetermined categories, which facilitates more effective information organization and retrieval (Aladakatti and Senthil Kumar, 2023).

Using statistical methods and ML algorithms, predictive analytics makes predictions about future patterns based on past data. Predictive analytics may improve the value and usefulness of data in open data management platforms by offering forecasts and actionable insights (Balbin et al., 2020). Using open data, predictive analytics may find and examine trends to help users make sense of patterns and future developments. For example, examining past traffic data can assist in predicting the patterns of congestion in the future (Derguech et al., 2014).

Large volumes of unstructured data can be made accessible, searchable, and comprehensible with the help of NLP, which can be quite helpful in managing open data. Text data can be normalized by NLP approaches by eliminating extraneous punctuation, reducing text to lowercase, and fixing spelling mistakes (Li and Rafiei, 2017). Put in place NLP-based tools to provide a more interactive and user-friendly experience by directly responding to user questions from the data. NLP has the power to completely change the management of open data, improving its usability, accessibility, and actionability for a broad spectrum of users (Witte et al., 2011).

A novel method to data management is provided by blockchain technology, which enables transparent, safe, and decentralized means of storing, sharing, and validating data. Blockchain relies on a distributed ledger that is kept up to date by several network nodes, or computers. Because there is no longer a requirement for a central authority, there is less chance of data manipulation or single points of failure thanks to this decentralization. Blockchain technology creates a verifiable audit trail by transparently recording all transactions and data changes (Fujihara, 2020). Blockchain technology has applications in many different industries and provides a transparent, safe, and decentralized method of managing data. Although scalability, regulatory compliance, and complexity are some of its obstacles, its potential to improve data

integrity, security, and collaboration makes it a useful tool for modern data management (Zhang et al., 2019).

Data integrity and security are critical aspects of managing open and linked data, particularly because such data is often widely accessible and used across various platforms. Ensuring that data remains accurate, consistent, and secure while being openly available requires a careful balance between openness and protection. refers to the reliability, consistency, and correctness of data during its entire existence. To guarantee that data is reliable and useful in the context of open and linked data, integrity must be upheld (Truong et al., 2019).

To manage and extract value from open and linked data, big data analytics is essential (Lněnička and Komárková, 2019). Open data pertains to datasets that are publicly accessible, useable, and shareable, whereas linked data is structured and connected data that is made possible by standards such as RDF. These datasets can be used in conjunction with big data analytics to produce insights, facilitate decision-making, and enable creative applications. Open data can be gathered and combined using big data analytics technologies from a variety of sources, including social media, government databases, and research archives, onto a single platform. However, working with a variety of unstructured open data sources can be very difficult (Matheus and Janssen, 2015). To comprehend the context, origin, and structure of open and linked data, big data systems frequently come with tools for managing metadata. Analysis of streaming data as it is generated is made possible by tools like Apache Spark and Apache Flink, which are used to process massive datasets in batch mode or in real-time. Through data portals and APIs, open data can be made available to the public or to certain users, enabling others to query and study the data (Lněnička and Komárková, 2019; Węcel, 2022).

2.3.2 Future directions for open and linked data

The future of open and connected data is poised for significant progress driven by technological innovations, the development of policy frameworks, and increased awareness of the benefits of data openness (Fayyaz et al., 2018). Better standards compliance and interoperability are essential for using and managing open and linked data efficiently. They guarantee the smooth integration, comprehension, and repurposing of data from many sources across various platforms, applications, and domains. This is a more thorough look at how standards and interoperability can be improved for open and linked data (Binding and Tudhope, 2016). This covers the application of standard interfaces, protocols, and formats. Guaranteeing that data

meaning is maintained and comprehensible across various platforms, data models, ontologies, and standardized vocabularies must be used for this. Coordinating organizational procedures, guidelines, and contracts to enable smooth data consumption and sharing (Guyo et al., 2021).

Efficiency, accuracy, and scalability in open and linked data processing can be greatly improved by increasing automation. Data handling processes are made more efficient by automation at every stage, from collection and integration to analysis and visualization. Utilize API calls and automatic web scraping programs (like Scrapy and Beautiful Soup) to get data from public sources regularly. Automation minimizes manual labor and guarantees that data are always updated (Quinn et al., 2020). Utilize ML techniques and tools (like OpenRefine and Dedupe) to automate the process of finding and eliminating duplicate items in different datasets. This is essential for connecting data from many sources. Use tools like RDF to automate data linkage based on relationships and shared concepts. The construction and querying of linked data graphs can be automated using programs like RDFLib and Apache Jena (Mohamed et al., 2022).

Utilize AI to anticipate and manage bottlenecks or breakdowns in the data pipeline, ensuring efficient data processing without the need for human involvement. Automated processes require regular monitoring and maintenance to ensure they function correctly over time. Organizations may more effectively utilize massive datasets by improving efficiency, consistency, and scalability through the automation of open and linked data processing. Data processing can be made more efficient by using the appropriate tools and techniques, which can lead to quicker insights, better judgment, and more creative applications (Obitko et al., 2013).

Better data discovery tools let users locate, access, and use pertinent information more quickly, which is crucial for managing open and linked data. These technologies apply AI, NLP, and semantic web standards to improve data search, integration, and analysis. Natural language queries can be used to search for datasets with NLP-enabled tools, improving the usability and accessibility of data discovery for non-technical users. Using controlled vocabularies and ontologies aids in comprehending the context and meaning of search terms, producing more precise search results (Li and Rafiei, 2017). Users can visually explore datasets with the aid of interactive tools, which facilitates the understanding of intricate data relationships and trends. Users are interacting with and using data in new ways thanks to enhanced data discovery tools for open and linked data management. These tools facilitate the discovery, integration, and analysis of pertinent data by utilizing AI, ML, semantic web technologies, and user-friendly interfaces (Berkley et al., 2009).

To maximize the value and reliability of data, open and linked data management must ensure data quality and usability. For precise analysis, well-informed decision-making, and efficient integration across several datasets, high-quality data is necessary (Gupta and Kumar, 2023). Data uniformity and consistency both inside and between datasets as well as throughout time are guaranteed by consistency (Zuiderwijk and Janssen, 2013). Create user-friendly data portals with download, filtering, and search functions that satisfy the needs of both experienced and inexperienced users (Lněnička and Nikiforova, 2021).

Provide consumers with lessons, examples, and explicit directions so they may utilize the data in an efficient manner (Kontokostas et al., 2014). Give consumers access to interactive visualization tools so they can examine and understand data without requiring complex technical knowledge. Provide dashboards that are customisable so users can make their own visualizations that are tailored to their own requirements (Lněnička and Nikiforova, 2021). For data profiling, cleaning, and enrichment, use Informatica Data Quality, Apache Griffin, and Talend Data Quality. Protégé for developing ontologies, OpenRefine for transforming and cleaning data, and Apache Jena for creating applications for the semantic web and linked data. DataHub and CKAN are tools for organizing and disseminating open data with thorough metadata. Making data relevant, reliable, and accessible requires improving data quality and usability in open and connected data management. Organizations may guarantee that their data effectively fulfills its intended purpose and helps wider data-driven initiatives by emphasizing correctness, consistency, completeness, and user-oriented characteristics (Martin et al., 2017).

3 MODEL DEVELOPMENT AND EVALUATION

3.1 RESEARCH METHODOLOGY AND DESIGN

To guarantee a thorough evaluation of open and linked data management platforms, this thesis employs a structured research methodology. The process of selecting the most suitable criteria was the first step, and it was guided by a literature review. Considering aspects like support, performance, or data formats supported, criteria for evaluating the suitability of open data management platforms for different use cases were selected by examining existing studies and industry best practices. These results are described in previous chapters, i.e., the list of criteria that are most commonly used to compare open data management platforms can be found in chapter 1.3 (Table 1) and the list of alternatives representing the open data management platforms that are used to power open data portals can be found in chapters 2.1 and 2.2 (Table 2).

This choice was not made haphazardly; rather, it was influenced by the need to satisfy the specific requirements of specified use cases, thereby guaranteeing that the chosen criteria would be reliable and pertinent. The next step was to identify and describe the alternative platforms that were going to be evaluated after the criteria were established. The available platforms were thoroughly examined during this procedure, with a focus on those that have gained widespread industry recognition for their efficiency and dependability. The alternatives were chosen based on how well they met the criteria.

This step was very important in getting ready for a good comparison and analysis. Primary data was gathered to compare each platform to the established standards after the alternatives were identified. Utilizing a variety of resources and tools, this data collection process gathered data on the platforms' performance, user experiences, and technical specifications. First, a scoring method (on a scale of 1 to 5, where 1 denoted the worst performance and 5 the best) was used to show how different requirements put one three use cases result in selection of different open data management platform. Then, one use case was selected, and an Analytic Hierarchy Process (AHP) approach was used to find the most suitable open data management platform for this use case.

3.2 CRITERIA AND THEIR DESCRIPTION

A careful process that combined theoretical research with relevant practical considerations for the use case was used to identify the evaluation criteria for the platforms. The selection of these criteria was based on existing literature and industry standards because they are crucial in determining whether or not a platform is suitable for managing open and linked data (Ali et al., 2022; Boch et al., 2022; Máchová and Lněnička, 2019).

Performance and scalability, for instance, were included as important criteria due to their significance in efficiently handling large datasets. Research demonstrates that scalability has a significant impact on the overall effectiveness of data management platforms making it an essential consideration in the evaluation process, supporting this decision. To ensure that the evaluation was both objective and comprehensive, each criterion was carefully defined. For instance, *Cost* was treated as a maximization criterion, with higher performance and better scalability being more desirable, while *Performance and Scalability* was treated as a minimization criterion. On a scale of 1 to 5, where 1 denoted the worst performance and 5 the best, the criteria were evaluated. This scale ensured that the evaluation was sensitive to platform-specific variations in performance by allowing for a nuanced assessment of each platform's capabilities. The criteria consider not only technical aspects like performance and scalability, but also key features, supported data formats, cost, developer reputation, and availability of support and documentation. These factors were chosen to give a complete picture of each platform's capabilities and to make sure the evaluation considered all relevant aspects.

These criteria were used to score each platform in the evaluation table, revealing the platforms' strengths and weaknesses in relation to the specified use case. The recommendations in this thesis are solidly supported by this comprehensive approach, which ensured that the final platform selection was based on a balanced consideration of all important criteria.

The selected open data management platforms are evaluated based on the following criteria:

1. **Cost (free / paid)**: Whether the platform is free to use or requires a subscription or one-time payment.
2. **Key features**: The main functionalities and tools offered by the platform, such as data visualization, API access, and user interface capabilities.
3. **Data formats supported**: The types of data formats that the platform can handle, such as CSV, JSON, XML, etc.

4. **Open-source or proprietary**: Whether the platform's source code is openly available for modification and distribution, or if it is proprietary with restricted access.
5. **Support and documentation**: The availability and quality of support services and documentation, including user guides, tutorials, and customer service.
6. **Performance and scalability**: The platform's ability to handle large datasets efficiently and its performance under varying loads.

3.3 ALTERNATIVES AND THEIR DESCRIPTION

The following alternatives were chosen for assessment based on the aforementioned criteria:

1. **ArcGIS Hub**
    - **Cost**: Paid
    - **Developer**: Esri
    - **Key Features**: Comprehensive mapping and spatial analysis tools, community engagement tools, data visualization, and sharing capabilities.
    - **Data Formats Supported**: CSV, JSON, Shapefiles, GeoJSON
    - **Open-Source or Proprietary**: Proprietary
    - **Support and Documentation**: Extensive official documentation, tutorials, and customer support
    - **Performance and Scalability**: High performance and scalability for spatial data
2. **CKAN**
    - **Cost**: Free
    - **Developer**: Open Knowledge Foundation
    - **Key Features**: Data storage and management, customizable metadata, full-text search, data visualization plugins, API access
    - **Data Formats Supported**: CSV, XLS, JSON, XML, RDF, TSV, SHP
    - **Open-Source or Proprietary**: Open-Source
    - **Support and Documentation**: Extensive documentation, community support, forums

- **Performance and Scalability**: Scalable and robust performance

3. **DataPress**
    - **Cost**: Paid
    - **Developer**: DataPress Ltd.
    - **Key Features**: Data publishing, cataloguing, visualization, API access, integration with other tools
    - **Data Formats Supported**: CSV, JSON, XML
    - **Open-Source or Proprietary**: Proprietary
    - **Support and Documentation**: Comprehensive support and documentation, user guides, and tutorials
    - **Performance and Scalability**: Reliable performance with moderate scalability
4. **DKAN**
    - **Cost**: Free
    - **Developer**: CivicActions
    - **Key Features**: Data cataloging, visualization, API access, customizable, built on Drupal CMS
    - **Data Formats Supported**: CSV, JSON, XML, RDF
    - **Open-Source or Proprietary**: Open-Source
    - **Support and Documentation**: Community-driven support, detailed documentation
    - **Performance and Scalability**: Good performance with high scalability
5. **OpenDataSoft**
    - **Cost**: Paid
    - **Developer**: Opendatasoft
    - **Key Features**: Data visualization, API management, data sharing, integration with various data sources
    - **Data Formats Supported**: CSV, JSON, XML, GeoJSON

- **Open-Source or Proprietary**: Proprietary
- **Support and Documentation**: High-quality support and detailed documentation
- **Performance and Scalability**: Excellent performance with high scalability

6. **Socrata**
    - **Cost**: Paid
    - **Developer**: Tyler Technologies
    - **Key Features**: Data hosting and visualization, built-in analytics, API for data access, integration with BI tools, user-friendly interface
    - **Data Formats Supported**: CSV, JSON, XLS, XML, RDF
    - **Open-Source or Proprietary**: Proprietary
    - **Support and Documentation**: Official support, comprehensive documentation, customer service
    - **Performance and Scalability**: High performance and scalability

3.3.1 Rationale for choosing the alternatives

The rationale for choosing these specific alternatives is based on their importance, characteristics, and applicability to a wide range of data management needs. Each platform represents a unique approach to handling open and linked data, and together they provide a comprehensive overview of the current landscape.

ArcGIS Hub was chosen because of its strong mapping and spatial analysis capabilities, which makes it ideal for organizations that need to visualize and analyse geographic data. Its comprehensive collection of tools and strong community support make it a leading choice in the field of local data management.

CKAN is a widely used open-source data management system known for its flexibility and scalability. It is supported by a large community and offers a rich set of features, making it a suitable choice for a wide range of data management needs.

DataPress was chosen to focus on data publishing and visualization, offering a user-friendly interface and strong integration capabilities. This is especially useful for organizations that want to publish and share data with minimal setup.

DKAN provides a customizable and open-source solution built on Drupal CMS, making it a good option for those who need a flexible platform with extensive community support. Its features are designed to meet the needs of public sector organizations and data publishers.

Opendatasoft has been included because of its high-performance data visualization and API management capabilities, which are essential for organizations that need to effectively integrate and display data from multiple sources.

Socrata was selected for its user-friendly interface and powerful analytical tools. Its integration with business intelligence tools makes it a preferred choice for organizations that aim to gain actionable insights from their data.

3.3.2 Relating the alternatives to open and linked data

Each of the selected platforms plays an important role in the realm of open and linked data, which plays a unique role in the data ecosystem.

ArcGIS Hub expands the reach and use of local data, allowing geographic datasets to be combined with other types of data. This capability is essential for applications in urban planning, environmental monitoring, and disaster management where integrated datasets provide comprehensive insights.

CKAN is the cornerstone of the Open Data Movement, which enables governments, organizations, and communities to publish, share, and link datasets. Its open-source nature allows for widespread customization and integration with other linked data technologies promoting transparency and data reuse.

DataPress simplifies the process of publishing and visualizing data, making it easier for organizations to share relevant datasets with stakeholders. Its user-friendly interface and strong API support facilitate the creation of interactive data-driven applications.

DKAN, being open source and built on popular Drupal system, allows seamless integration of linked data. It supports different data formats and metadata standards, ensures interoperability and improves the discovery of datasets.

Opendatasoft specializes in connecting diverse data sources through its powerful API management and data visualization capabilities. This makes it a valuable tool for these organizations that aims to create linked data ecosystems and develop data-driven services.

Socrata focuses on making open data accessible and useful through its analytics and visualization tools. By integrating with business intelligence platforms, Socrata enables open data to be linked to organizational data, driving insights, and informed decision-making.

These platforms collectively contribute to the development of open and linked data by providing strong tools for data publication, management, and integration. They support the principles of open data by increasing accessibility, interoperability, and use, thus empowering organizations and individuals to leverage data for a variety of applications.

3.4 EVALUATION OF USE CASES USING A SCORING METHOD

To illustrate the various requirements that organizations have when it comes to managing open and linked data, three distinct use cases are discussed in this section. The selection of a suitable platform that is in line with the requirements of the organization is necessary because each use case highlights distinct difficulties and requirements.

3.4.1 Use case 1: Public administration transparency and engagement

The first use case involves a public administration organization that publishes a variety of datasets with the intention of increasing citizen engagement and transparency. Data on public transportation schedules, environmental monitoring, and the census are all included in these datasets. The main problem here is how to manage a lot of data at the same time as making sure the platform has good tools for data visualization and is easy for everyone to use. The organization needs a platform that can handle large datasets well, provides options for easy visualization, and ensures that the data can be used by a wide range of people.

Considering the evaluation performed in Table 3, **ArcGIS Hub** may be the best choice for this use case. This stage is known for its exhaustive set-up of apparatuses intended for information representation and public commitment. Its strong support services guarantee that the organization will be able to continue publishing data in an efficient and continuous manner, and its high-performance capabilities make it ideal for managing large datasets. ArcGIS Hub's suitability for this use case is further enhanced by its capacity to handle local data and integrate

a variety of data sources, ensuring that the platform satisfies the organization's objectives of transparency and citizen engagement.

Table 3. Comparison of alternatives based on criteria for use case 1. Source: own processing.

Criterion	**ArcGIS Hub**	**CKAN**	**Datapress**	**DKAN**	**Open DataSoft**	**Socrata**
Cost	3.0	4.0	2.5	4.5	3.5	2.5
Developer	4.0	3.5	3.5	4.0	3.0	4.5
Features	5.0	3.0	3.0	3.5	4.5	4.0
Data Formats	4.5	5.0	4.0	4.0	3.5	4.0
Open-Source	2.0	5.0	2.0	1.5	3.0	2.5
Support	4.5	3.0	4.0	4.5	3.5	4.0
Performance	4.0	3.5	3.5	3.5	3.0	4.5

3.4.2 Use case 2: Data analytics and integration for a startup

A startup that focuses on publishing data analytics and LinkedIn open data is the subject of the second use case. A platform that not only supports cutting-edge data visualization and analytics tools but also offers robust data integration capabilities is necessary for this startup. In order to keep their data services flexible and innovative, the startup is looking for a solution that can scale with its datasets and provide API access for integration with other applications.

Considering the evaluation performed in Table 4, a platform like **Socrata** might be a good choice for this use case. Scalability is a feature of this platform that enables startups to expand their data processing and storage capacities as needed without incurring significant costs. Additionally, strong API integration support on Socrata platforms makes it simpler for the startup to connect their analytics tools to other services and applications. Socrata is an appealing option for the startup's data analytics and integration requirements because of its adaptability and the platform's capacity to handle expanding datasets.

Table 4. Comparison of alternatives based on criteria for use case 2. Source: own processing.

Criterion	**ArcGIS Hub**	**CKAN**	**Datapress**	**DKAN**	**Open DataSoft**	**Socrata**
Cost	4.5	4.0	3.5	4.0	3.0	3.5
Developer	4.0	4.5	3.5	3.5	4.5	4.0
Features	4.0	5.0	3.0	4.5	4.0	4.5
Data Formats	4.0	4.5	4.0	4.5	3.5	4.0
Open-Source	2.0	1.5	3.0	2.0	2.5	3.0
Support	4.5	4.0	4.0	4.0	4.5	4.0
Performance	3.5	4.0	4.5	4.5	4.0	4.5

3.4.3 Use case 3: High-performance data management for a financial institution

A significant financial institution with a substantial budget is the subject of the third use case. Its objectives are to enhance data-driven decision-making, strengthen data management practices, and guarantee compliance with stringent regulatory requirements. A platform that can handle complicated data formats, offer robust security features, and support high performance is required by this organization. In addition, the platform needs to have advanced analytics capabilities to help the institution make decisions.

Considering the evaluation performed in Table 5, **ArcGIS Hub** can be considered once more for this use case due to its capacity to manage complex data requirements in a secure environment. However, due to the platform's advanced analytics and security features, the financial institution may also take into consideration Microsoft Azure or Amazon Web Services. These platforms are well-suited for large businesses with stringent security and compliance requirements because of their high performance and ability to manage complex data formats. The financial institution will be able to effectively manage its data while also complying with regulatory requirements because of these platforms' extensive features and robust support.

Table 5. Comparison of alternatives based on criteria for use case 3. Source: own processing.

Criterion	ArcGIS Hub	CKAN	Datapress	DKAN	Open DataSoft	Socrata
Cost	4.5	3.0	3.5	3.0	3.0	3.5
Developer	4.0	3.5	3.5	3.5	3.5	3.0
Features	4.0	4.0	3.0	4.5	4.0	4.5
Data Formats	4.0	3.5	4.0	3.5	3.5	3.0
Open-Source	3.0	1.5	3.0	2.0	2.5	3.0
Support	4.5	4.0	3.0	4.0	3.5	4.0
Performance	3.5	3.0	3.5	4.5	4.0	3.5

3.5 EVALUATION OF USE CASE 1 USING AHP

Thes use case 1 was selected for a more detailed analysis using the AHP approach, now using a scale from 1 to 9. Figure 1 shows the preference weights assigned to each criterion based on pairwise comparisons. Quality is classified in descending order of importance, with "cost" being most important at 32.9%, followed by "features" at 28.6%, "developer" at 16%, etc. The figure also includes a Consistency Ratio (CR), which is 6.5%, indicating that decisions are consistent.

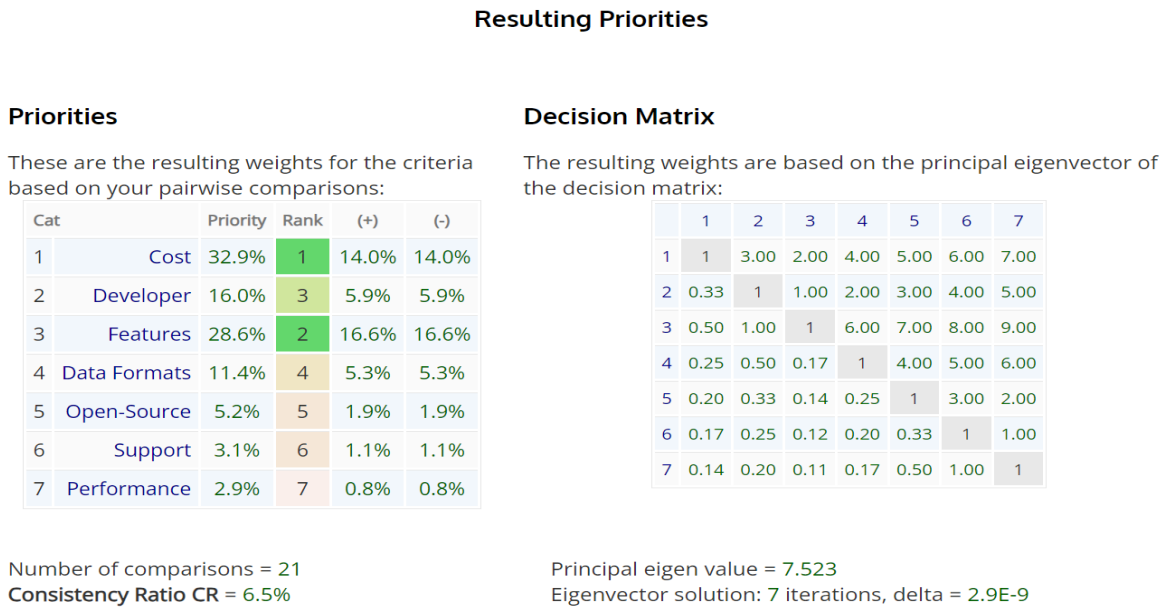

Figure 1. Comparison of criteria with each other for use case 1. Source: own processing.

The bar chart in Figure 2 visualizes the preferred weight for each criterion, showing the minimum, the result, and the maximum values. It provides a visual representation of how much weight each criterion has in the decision-making process. "Cost" and "Features" are the most important criteria, while "Support" and "Performance" are considered the least important.

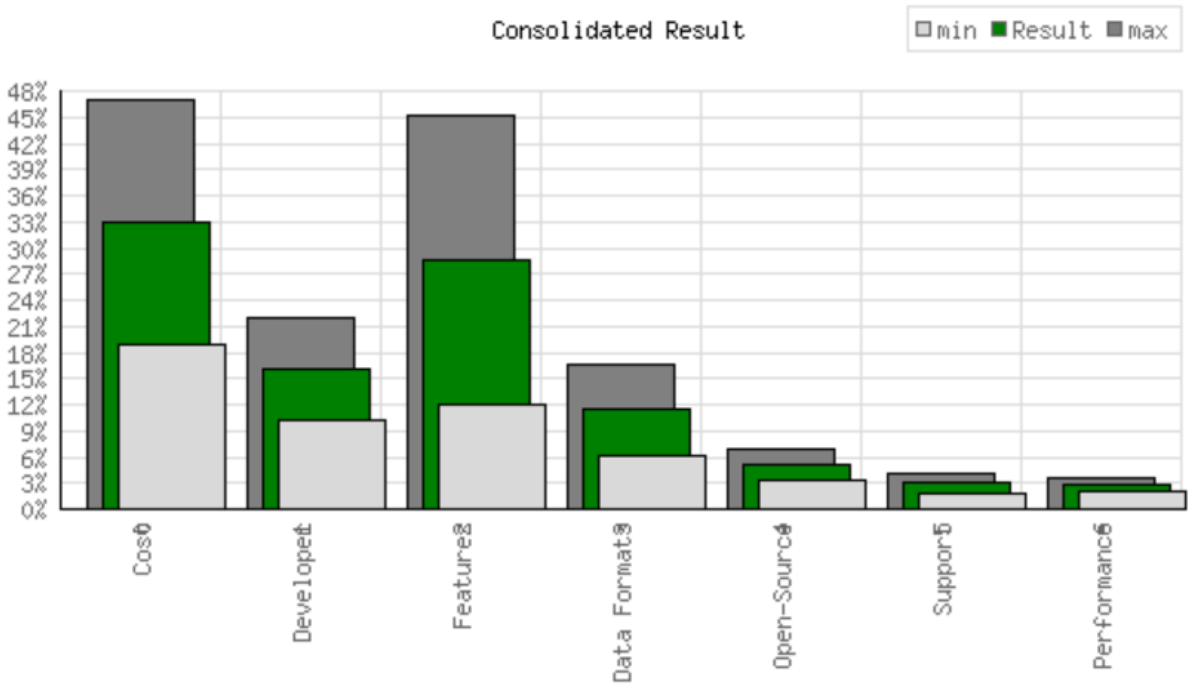

Figure 2. Consolidated result for use case 1. Source: own processing.

Based on predetermined criteria, the AHP analysis provides a methodical and quantitative evaluation of selected open and linked data platforms. The results are discussed, the advantages and disadvantages of each platform are emphasized, and a final ranking is provided in this section. The AHP tool was used to calculate the quality weights and overall scores for each platform. Cost, developer, key features, supported data formats, open source or proprietary, support and documentation, performance, and scalability are some of the criteria. The order of platforms based on their weights is displayed in Figure 3. ArcGIS Hub can be recommended as the most suitable open data management platform for the use case 1. It is followed by CKAN, which gained its weight mostly because it is free and open source.

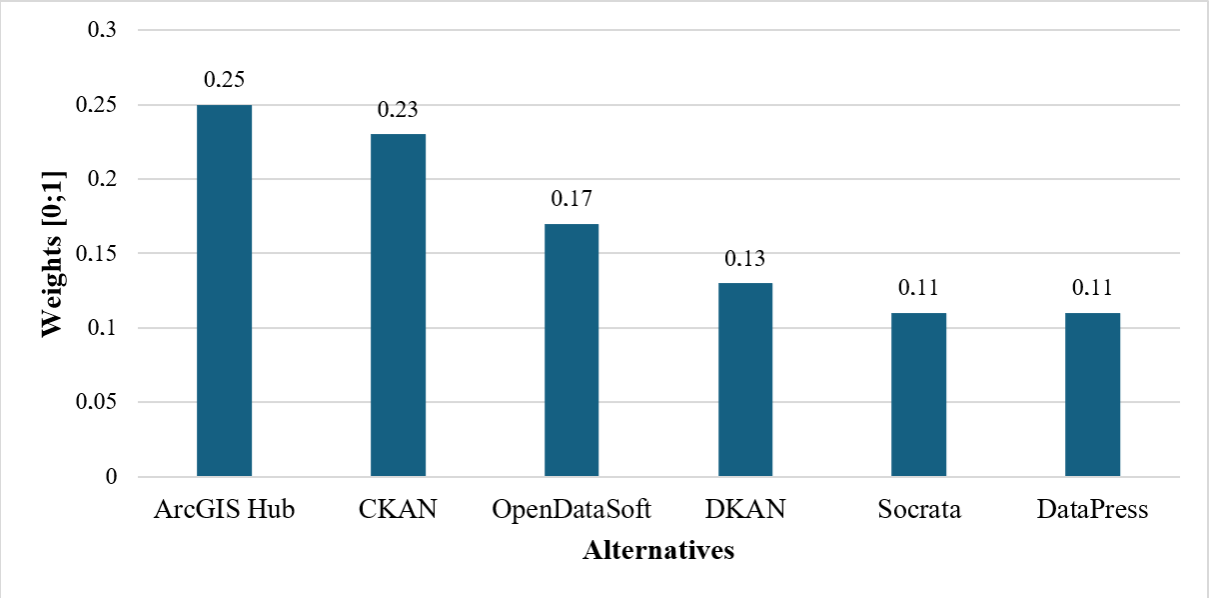

Figure 3. Final scores for each platform for the use case 1. Source: own processing.

3.6 SENSITIVITY ANALYSIS

Sensitivity analysis looks at how a model's output results are affected by changes in its input parameters. By looking at how changes in the criteria weights affect the final scores, sensitivity analysis aids in determining the robustness and reliability of the ranking of open and linked data platforms in the AHP analysis. We can examine how fluctuations in the importance of various criteria, such as cost, developer support, or performance, affect the platforms' overall ranking by conducting sensitivity analysis.

We can determine through this analysis whether or not the final rankings remain stable across a variety of scenarios and which criteria have the greatest impact on the ranking. For instance, we can observe how this adjustment affects the platforms' overall ranking if "performance and scalability" gets more weight. The results are affected by this criterion if the ranking changes significantly. On the other hand, rankings that maintain consistency despite changes in criteria weights indicate greater robustness.

3.6.1 Impact of criteria weights on rankings

Sensitivity analysis involves examining how changes in the weight of different standards affect the overall ranking of open and connected data management platforms. This process helps identify the robustness of the rating and helps to understand which criteria have the most impact on the decision-making process.

Methodology:

1. **Criteria Adjustment:** Incrementally adjust the weights of each criterion (e.g., Cost, Features, Developer, etc.) while keeping the total weight sum constant.
2. **Recalculate Rankings:** Recalculate the overall rankings of the platforms for each adjustment scenario.
3. **Analyse Results:** Identify how the rankings change in response to the weight adjustments.

Results: The sensitivity analysis results reveal the following key insights:

- **Cost Sensitivity:** Increasing the weight of the Cost criterion causes significant shifts in rankings, highlighting its critical impact on the overall decision. Platforms with lower costs rank higher when Cost weight is increased.
- **Features Sensitivity:** Similarly, increasing the weight of Features alters the rankings, with feature-rich platforms climbing higher. This suggests that Features is another pivotal criterion in the evaluation.
- **Developer and Support Sensitivity:** Changes in weights for Developer and Support criteria result in moderate ranking shifts, indicating these factors have a less dominant, but still notable, influence on the decision.
- **Minimal Impact Criteria:** Criteria such as Open-Source and Performance show minimal impact on rankings when their weights are adjusted, suggesting these are less critical in this analysis.

Implications:

- **Key Decision Drivers:** Cost and Features are the primary drivers of platform selection. Decision-makers should carefully consider these criteria when making final choices.
- **Robust Rankings:** The overall rankings are relatively robust to minor adjustments in less influential criteria, providing confidence in the initial evaluation's validity.

3.6.2 Scenario analysis

Scenario analysis explores how different platforms perform in different situations, such as budget constraints or the need for specific features. This analysis helps to develop recommendations tailored to specific organizational needs and constraints.

Scenario 1: Budget constraints

- **Description:** Analyse platform performance with a high emphasis on low cost due to budget limitations.
- **Findings:** Platforms with lower costs, such as CKAN and DataPress, rank higher under this scenario, making them suitable choices for budget-conscious organizations.
- **Recommendation:** Organizations with stringent budget constraints should prioritize platforms like CKAN and DataPress, which offer essential functionalities at a lower cost.

Scenario 2: Need for specific features

- **Description:** Evaluate platforms based on the need for advanced features and functionalities, such as robust data visualization or extensive API support.
- **Findings:** Platforms like Socrata and OpenDataSoft, which offer comprehensive feature sets, rank higher under this scenario.
- **Recommendation:** Organizations requiring advanced features should consider Socrata and OpenDataSoft to leverage their extensive capabilities.

Scenario 3: Focus on developer support

- **Description:** Examine platform performance with a high emphasis on developer support and ease of integration.
- **Findings:** Platforms with strong developer support, such as ArcGIS Hub and DKAN, perform well in this scenario.
- **Recommendation:** For organizations that prioritize developer support and seamless integration, ArcGIS Hub and DKAN are recommended.

Scenario 4: Open-source preference

- **Description:** Assess platform performance with a preference for open-source solutions due to customization needs and community support.
- **Findings:** Open-source platforms like CKAN and DKAN rank higher under this scenario.

- **Recommendation:** Organizations favouring open-source solutions should consider CKAN and DKAN for their flexibility and strong community support.

4 DISCUSSION AND POLICY RECOMMENDATIONS

Implementing open data platforms successfully necessitates adhering to best practices that guarantee their efficiency and longevity. Key guidelines and policy suggestions for making open data initiatives work best are listed below.

Organizations must clearly articulate the goals of the open data initiative, such as increasing transparency, enhancing public services, or stimulating innovation. Ensure that the platform's features align with these goals to maximize its relevance and impact. The need to involve a wide range of stakeholders, including government entities, private sector partners, and civil society organizations, from the outset. Gather and incorporate their feedback to meet diverse needs and encourage widespread acceptance. They must establish stringent data quality standards to maintain accuracy, completeness, and timeliness. Regularly update and validate datasets to uphold their reliability and trustworthiness. They need to use open standards and protocols to enable seamless data sharing and integration across different systems. Ensure the platform supports a variety of data formats and APIs to facilitate interoperability.

Organizations should also prioritize design the platform with transparency in mind (Lněnička and Nikiforova, 2021), ensuring it is intuitive, accessible, and easy to navigate. Organizations should offer training and resources, include e-learning courses, to improve users' ability to interpret and utilize data effectively. Organize workshops, hackathons, and collaborative projects to encourage the practical use of open data. In the whole data ecosystem, organizations must implement robust security measures to protect data from unauthorized access and breaches. Adhere to data privacy regulations and best practices to safeguard user information. Finally, organizations need to monitor and evaluate performance by developing metrics and key performance indicators to track the platform's performance and impact, and regularly assess and adjust the platform to optimize its effectiveness.

Based on the findings, in which ArcGIS Hub and CKAN appeared to be the most suitable open data management platforms for public administration, **we recommend** developing a national or regional policy that outlines clear goals, principles, and strategies for open data initiatives. The policy should be inclusive, addressing the needs of various stakeholders and ensuring equitable data access. There also must be legal and regulatory frameworks for releasing non-sensitive public data in open formats, and regulations that address data privacy and security concerns effectively. Also, public administration agencies and institutions must secure funding

for the development, maintenance, and expansion of open data platforms, and invest in the necessary infrastructure, technology, and human resources to support these initiatives sustainably. They should foster collaboration between government agencies, the private sector, academia, and civil society to leverage collective expertise and resources. Support public-private partnerships to drive innovation and enhance the value of open data. Develop and enforce standards for data collection, storage, and dissemination to ensure consistency and interoperability. Promote the adoption of open standards and protocols across all sectors. Promote a culture of openness and transparency within government agencies and among the public. Use awareness campaigns, success stories, and case studies to highlight the benefits of open data. Offer training and capacity-building programs for government officials, data managers, and end-users. Encourage ongoing professional development in data management and analytics. Implement mechanisms to regularly track and report the progress of open data initiatives. Utilize data-driven insights to inform policy decisions and enhance the effectiveness of open data programs.

Finally, although this thesis compared open data management platforms that can be used or buy by everyone, it must be also noted that a lot of national governments or cities worldwide are using their own data management systems. One of the reasons it the topic of data security and privacy, resulting from current cybersecurity threats faced by public administration agencies and institutions.

CONCLUSIONS

This thesis provides a comprehensive evaluation of open and linked data platforms through the AHP and sensitivity analysis. The AHP analysis revealed that CKAN and ArcGIS Hub are the top-performing platforms based on criteria such as cost, developer support, and performance, while platforms like DataPress lagged in several areas. Sensitivity analysis demonstrated the impact of varying criteria weights on platform rankings, indicating that some platforms are more sensitive to changes in criteria importance than others.

The discussion highlighted best practices for implementing open data platforms, including defining clear objectives, ensuring high data quality, fostering interoperability, and addressing security concerns. Policy recommendations emphasized the need for comprehensive policies, legal frameworks, and collaboration to support open data initiatives. Overall, this thesis underscores the importance of selecting and managing open data platforms strategically to enhance transparency, accessibility, and innovation.

REFERENCES

Aladakatti, S. S., & Senthil Kumar, S. (2023). Exploring natural language processing techniques to extract semantics from unstructured dataset which will aid in effective semantic interlinking. *International Journal of Modeling, Simulation, and Scientific Computing*, 14(1), 2243004.

Alexopoulos, C., Zuiderwijk, A., Charapabidis, Y., Loukis, E., & Janssen, M. (2014). Designing a second generation of open data platforms: Integrating open data and social media. In *Electronic Government: 13th IFIP WG 8.5 International Conference, EGOV 2014, Dublin, Ireland, September 1-3, 2014. Proceedings 13* (pp. 230-241). Springer, Berlin Heidelberg.

Ali, M., Alexopoulos, C., & Charalabidis, Y. (2022). A comprehensive review of open data platforms, prevalent technologies, and functionalities. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance* (pp. 203-214). ACM.

Balbin, P. P. F., Barker, J. C., Leung, C. K., Tran, M., Wall, R. P., & Cuzzocrea, A. (2020). Predictive analytics on open big data for supporting smart transportation services. *Procedia Computer Science*, *176*, 3009-3018.

Berkley, C., Bowers, S., Jones, M. B., Madin, J. S., & Schildhauer, M. (2009). Improving data discovery for metadata repositories through semantic search. In *2009 International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 1152-1159). IEEE.

Binding, C., & Tudhope, D. (2016). Improving interoperability using vocabulary linked data. *International Journal on Digital Libraries*, 17, 5-21.

Boch, M., Gindl, S., Barnett, A., Margetis, G., Mireles, V., Adamakis, E., & Knoth, P. (2022). A systematic review of data management platforms. In *World Conference on Information Systems and Technologies* (pp. 15-24). Cham: Springer International Publishing.

Borgesius, F. Z., Gray, J., & Van Eechoud, M. (2015). Open data, privacy, and fair information principles: Towards a balancing framework. *Berkeley Technology Law Journal*, 30(3), 2073-2131.

Butorac, V. (2021). ArcGIS Hub. *Geografski horizont*, 67(2), 60-62.

Charalabidis, Y., Alexopoulos, C., Ferro, E., Janssen, M., Lampoltshammer, T., & Zuiderwijk, A. (2018). *The World of Open Data: Concepts, Methods, Tools and Experiences*. Public

Administration and Information Technology. Springer International Publishing, Cham.

Cinnamon, J. (2020). Data inequalities and why they matter for development. *Information Technology for Development*, 26(2), 214-233.

CivicActions (2023). *DKAN Open Data Portal*. Available at: https://dkan.readthedocs.io/en/latest/

Correa, A. S., Zander, P. O., & Da Silva, F. S. C. (2018). Investigating open data portals automatically: a methodology and some illustrations. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age* (pp. 1-10). ACM.

Cruz, R. A. B., & Lee, H. J. (2015). Opening the Nation: Leveraging Open Data to Create New Business and Provide Services. *Knowledge Management Research*, 16(4), 157-168.

DataPress (2024). *DataPress*. Available at: https://datapress.com/

Dawes, S. S., & Helbig, N. (2010). Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In *Electronic Government: 9th IFIP WG 8.5 International Conference, EGOV 2010, Lausanne, Switzerland, August 29-September 2, 2010. Proceedings 9* (pp. 50-60). Springer, Berlin Heidelberg.

Derguech, W., Bruke, E., & Curry, E. (2014). An autonomic approach to real-time predictive analytics using open data and internet of things. In *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops* (pp. 204-211). IEEE.

Esri (2024). *ArcGIS Hub*. Available at: https://hub.arcgis.com/

Fan, B., & Zhao, Y. (2017). The moderating effect of external pressure on the relationship between internal organizational factors and the quality of open government data. *Government Information Quarterly*, 34(3), 396-405.

Fayyaz, N., Ullah, I., & Khusro, S. (2018). On the current state of linked open data: Issues, challenges, and future directions. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 14(4), 110-128.

Fujihara, A. (2020). Proposing a blockchain-based open data platform and its decentralized oracle. In *Advances in Intelligent Networking and Collaborative Systems: The 11th*

International Conference on Intelligent Networking and Collaborative Systems (INCoS-2019) (pp. 190-201). Springer International Publishing.

Ghahremanlou, L., H Tawil, A. R., Kearney, P., Nevisi, H., Zhao, X., & Abdallah, A. (2019). A survey of open data platforms in six UK smart city initiatives. *The Computer Journal*, 62(7), 961-976.

Guyo, E., Hartmann, T., & Ungureanu, L. (2021). Interoperability between BIM and GIS through open data standards: An overview of current literature. *sign*, 3, 5-9.

Gray, J. W. (2023). What do data portals do? Tracing the politics of online devices for making data public. *Data & Policy*, 5, e10.

Gupta, A., & Kumar, R. (2023). Modelling the Barriers of Artificial Intelligence Adoption in the Organisations: An Interpretive Structural Modelling and MICMAC Analysis. In *Smart Analytics, Artificial Intelligence and Sustainable Performance Management in a Global Digitalised Economy* (pp. 45-66). Emerald Publishing Limited.

Halabi, A. S. E. H. (2018). Saudi public uses and gratifications of the data press. *The Egyptian Journal of Media Research*, 2018(65), 319-381.

Ham, K. D., & Crockett, D. L. (2021). *Evaluation of Data Catalog Software for Hanford Site Environmental Datasets* (No. PNNL-31960; DVZ-RPT-066). Pacific Northwest National Lab. (PNNL), Richland, WA (United States).

Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a global data space*. Springer, Cham.

Hert, M., Reif, G., & Gall, H. C. (2011). A comparison of RDB-to-RDF mapping languages. In *Proceedings of the 7th International Conference on Semantic Systems* (pp. 25-32). ACM.

Jaakkola, H., Mäkinen, T., & Eteläaho, A. (2014). Open data: opportunities and challenges. In *Proceedings of the 15th International Conference on Computer Systems and Technologies* (pp. 25-39). ACM.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258-268.

Janssen, M., & Kuk, G. (2016). Big and open linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 3-13.

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?. *Government Information Quarterly*, 32(4), 363-368.

Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2), 100-120.

Kim, D., Gil, M. S., Nguyen, M. C., Won, H., & Moon, Y. S. (2021). Comprehensive Knowledge Archive Network harvester improvement for efficient open-data collection and management. *ETRI Journal*, 43(5), 835-855.

Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R., & Zaveri, A. (2014). Test-driven evaluation of linked data quality. In *Proceedings of the 23rd international conference on World Wide Web* (pp. 747-758). ACM.

Li, Y., & Rafiei, D. (2017). Natural language data management and interfaces: Recent development and open challenges. In *Proceedings of the 2017 ACM International Conference on Management of Data* (pp. 1765-1770). ACM.

Lněnička, M., & Komárková, J. (2019). Big and open linked data analytics ecosystem: Theoretical background and essential elements. *Government Information Quarterly*, 36(1), 129-144.

Lněnička, M., Machova, R., Volejníková, J., Linhartová, V., Knezackova, R., & Hub, M. (2021). Enhancing transparency through open government data: The case of data portals and their features and capabilities. *Online Information Review*, 45(6), 1021-1038.

Lněnička, M., & Nikiforova, A. (2021). Transparency-by-design: What is the role of open data portals?. *Telematics and Informatics*, 61, 101605.

Lopez, V., Kotoulas, S., Sbodio, M. L., Stephenson, M., Gkoulalas-Divanis, A., & Aonghusa, P. M. (2012). Queriocity: A linked data platform for urban information management. In *International Semantic Web Conference* (pp. 148-163). Springer, Berlin Heidelberg.

Lourenço, R. P., Piotrowski, S., & Ingrams, A. (2017). Open data driven public accountability. *Transforming Government: People, Process and Policy*, 11(1), 42-57.

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, 12(1), 21-41.

Máchová, R., & Lněnička, M. (2019). A multi-criteria decision making model for the selection of open data management systems. *Electronic Government, an International Journal*, 15(4), 372-391.

Martin, E. G., Law, J., Ran, W., Helbig, N., & Birkhead, G. S. (2017). Evaluating the quality and usability of open data for public health research: a systematic review of data offerings on 3 open data platforms. *Journal of Public Health Management and Practice*, 23(4), e5-e13.

Matheus, R., & Janssen, M. (2015). Transparency dimensions of big and open linked data: Transparency as being synonymous with accountability and openness. In *Open and Big Data Management and Innovation: 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015, Delft, The Netherlands, October 13-15, 2015, Proceedings 14* (pp. 236-246). Springer International Publishing.

Mohamed, A., Abuoda, G., Ghanem, A., Kaoudi, Z., & Aboulnaga, A. (2022). RDFFrames: knowledge graph access for machine learning tools. *The VLDB Journal*, 31(2), 321-346.

Obitko, M., Jirkovský, V., & Bezdíček, J. (2013). Big data challenges in industrial automation. In *Industrial Applications of Holonic and Multi-Agent Systems: 6th International Conference, HoloMAS 2013, Prague, Czech Republic, August 26-28, 2013. Proceedings* (pp. 305-316). Springer Berlin Heidelberg.

Opendatasoft (2024). *Opendatasoft*. Available at: https://www.opendatasoft.com/en/

Open Government Partnership (2024). *About Open Government Partnership* [online]. Available at: https://www.opengovpartnership.org/about/

Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., & Vrgoč, D. (2016). Foundations of JSON schema. In *Proceedings of the 25th international conference on World Wide Web* (pp. 263-273). ACM.

Quinn, C., Shabestari, A. Z., Misic, T., Gilani, S., Litoiu, M., & McArthur, J. J. (2020). Building automation system-BIM integration using a linked data structure. *Automation in Construction*, 118, 103257.

Scanlon, A. (2021). *Open Data Portal Management Framework*. MBA Thesis. [online]. Available at: https://penni.wu.ac.at/supervision/Anne%20Scanlon%20Thesis%202021.pdf

Seto, T., & Sekimoto, Y. (2016). The construction of open data portal using DKAN for integrate to multiple Japanese local government open data. In *Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings* (Vol. 16, No. 1, p. 17).

Toots, M., Mcbride, K., Kalvet, T., & Krimmer, R. (2017). Open data as enabler of public service co-creation: Exploring the drivers and barriers. In *2017 Conference for E-Democracy and Open Government (CeDEM)* (pp. 102-112). IEEE.

Truong, D. D., Nguyen-Van, T., Nguyen, Q. B., Huy, N. H., Tran, T. A., Le, N. Q., & Nguyen-An, K. (2019). Blockchain-based open data: An approach for resolving data integrity and transparency. In *International Conference on Future Data and Security Engineering* (pp. 526-541). Springer International Publishing, Cham.

Tyler Technologies (2024). *SODA Developers*. Available at: https://dev.socrata.com/

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337.

Węcel, K. (2022). *Big, Open and Linked Data: Effects and Value for the Economy*. Business Information Systems. Springer, Cham.

Wirtz, B. W., & Birkmeyer, S. (2015). Open government: Origin, development, and conceptual perspectives. *International Journal of Public Administration*, 38(5), 381-396.

Witte, R., Kappler, T., Krestel, R., & Lockemann, P. C. (2011). Integrating wiki systems, natural language processing, and semantic technologies for cultural heritage data management. In *Language Technology for Cultural Heritage: Selected Papers from the LaTeCH Workshop Series* (pp. 213-230). Springer, Berlin Heidelberg.

Zhang, L., Luo, M., Li, J., Au, M. H., Choo, K. K. R., Chen, T., & Tian, S. (2019). Blockchain based secure data sharing system for Internet of vehicles: A position paper. *Vehicular Communications*, 16, 85-93.

Zhu, X. (2017). The failure of an early episode in the open government data movement: A historical case study. *Government Information Quarterly*, 34(2), 256-269.

Zuiderwijk, A., & Janssen, M. (2013). A coordination theory perspective to improve the use of open data in policy-making. In *Electronic Government: 12th IFIP WG 8.5 International*

Conference, EGOV 2013, Koblenz, Germany, September 16-19, 2013. Proceedings 12 (pp. 38-49). Springer, Berlin Heidelberg.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 19(1-2), 17-33.