TOWARDS A UNIFIED CONCEPT FOR SHARING AND LINKING OF DATA RESOURCES IN COLLABORATIVE RESEARCH VIA WEB SERVICES

Plamen Milev

Associate Professor, PhD. Department of Information Technologies and Communications, University of National and World Economy, e-mail: pmilev@unwe.bg

Abstract

In an era where collaborative research is increasingly becoming the norm, the necessity for efficient and effective data sharing and linking mechanisms is paramount. This paper delves into the journey towards a unified concept for sharing and linking of data resources, emphasizing the pivotal role of web services in realizing this vision. From the challenges posed by diverse data formats and the need for interoperability to the potential for real-time collaboration, the research underscores the transformative potential of web services. Drawing on protocols such as OAI-PMH and the GO-FAIR initiative's principles, the study outlines the integral function of web services in adhering to standards of findability, accessibility, interoperability, and reusability (FAIR). Furthermore, platforms such as Zenodo, OpenAIRE, EOSC, and BPOS are discussed in light of their contributions to the open science movement. The research bridges the divide between the theoretical need for a unified data sharing concept and the practical technological solutions offered by web services, paving the way for a more integrated and collaborative future in scientific research.

Key words: Data sharing, data linking, web services, open science, collaborative research.

JEL: C88, L86.

Introduction

The modern scientific landscape thrives on collaboration. With researchers across the globe working together, sharing data efficiently and effectively has become crucial. This collaborative approach produces vast amounts of data from various sources, each with its unique format and structure. Navigating this complex environment without a consistent method for sharing and linking data can slow down progress, as researchers spend more time managing data than actually analyzing it. Web services offer a promising solution to this challenge. They are tools designed to operate over the internet, allowing different software applications to communicate seamlessly. For researchers, this means easier access to data, consistent communication between different research tools, and the potential for real-time collaboration. More than just a technical fix, web services could be the key to unlocking the full potential of collaborative research by streamlining data sharing and integration. It's in this context that the importance of a unified approach to data sharing emerges. Without standardized mechanisms to share and interlink datasets, researchers may find themselves mired in administrative tasks, struggling with data compatibility issues, or confronting barriers to accessing crucial datasets. These logistical hurdles can hinder the rapid exchange of insights, dampening the innovative spirit that collaborative research seeks to foster. Web services are designed to foster communication between different software applications, regardless of their underlying architecture or platform. In the research ecosystem, web services can act as vital bridges, connecting disparate data sources and ensuring that they speak a common language. As we delve deeper into this topic, we will explore the transformative potential of web services in the context of collaborative research. We will examine how they can address current challenges, highlight the platforms that exemplify their successful integration, and envision a future where collaborative research is truly synergized through the capabilities of web services.

This paper ventures into the realm of web services, aiming to highlight their instrumental role in forging a unified concept for sharing and linking data resources in collaborative research. We shall embark on a journey that traverses the challenges of the current research landscape, the transformative potential of web services, and the practical implications of integrating these services into the broader framework of scientific collaboration. Drawing from established protocols, open science platforms, and forward-thinking initiatives, this research seeks to pave the way for a future where collaborative research is not only facilitated but truly optimized through the power of web services.

Landscape of sharing and linking of data resources

In the contemporary digital era, the ability to share and link data resources effectively is pivotal in facilitating collaborative research efforts. This paper seeks to delineate the prevailing methodologies and frameworks that underscore data sharing and linking in collaborative research, exploring their evolution, advantages, and limitations. The importance of collaborative research cannot be understated in the modern academic and scientific landscape. Collaborative research, underpinned by the efficient sharing and linking of data resources, paves the way for multidisciplinary discoveries and innovations [1]. As research projects become increasingly complex and data-driven, the necessity for robust methodologies and frameworks for data sharing and linking becomes evident. This paper aims to scrutinize the current landscape of these methodologies and frameworks, providing a comprehensive overview of their characteristics and functionalities. In the nascent stages of collaborative research, data sharing was predominantly centralized, with data repositories being managed by a single entity. This approach, however, gradually gave way to decentralized and hybrid models as the volume of data increased and the necessity for multi-institutional collaboration became apparent [2]. Over the years, various frameworks have been developed to facilitate data sharing and linking, often characterized by evolving technological and policy considerations:

- Centralized approaches to data sharing and linking involve the consolidation of data resources into a single platform or repository. This approach ensures uniform data formats and standards, which can potentially simplify data management and analysis processes. However, it can also pose challenges in terms of data security and privacy [1].
- On the other hand, decentralized approaches allow for data resources to be stored and managed across various platforms and entities. While this approach fosters data ownership and autonomy, it can create hurdles in achieving data interoperability and integration [3].
- Hybrid approaches attempt to amalgamate the advantages of both centralized and decentralized approaches, fostering data sharing and linking while mitigating respective disadvantages. These approaches are often characterized by sophisticated technological infrastructures that facilitate data integration and interoperability [1].

In terms of frameworks, a considerable focus has been placed on both technical and policy aspects:

- From a technical standpoint, considerations such as data formats, integration, and interoperability are pivotal. Emerging frameworks are increasingly leveraging technologies such as blockchain to facilitate secure and efficient data sharing and linking [4].
- In parallel, policy and governance aspects have garnered attention, with emphasis on data privacy, security, and ownership. Various frameworks have emerged to navigate the complex landscape of data governance, focusing on fostering collaboration while safeguarding data rights and privacy [5].

The landscape of data sharing and linking in collaborative research is characterized by a rich tapestry of methodologies and frameworks. As research endeavors continue to evolve, the development of robust, secure, and efficient frameworks for data sharing and linking will remain a focal point. This analysis underscores the importance of ongoing efforts to refine and optimize these frameworks, fostering a collaborative research environment that is both innovative and secure. Data sharing and linking are central processes in collaborative research, aiding in the development of a cohesive and comprehensive research landscape. Despite the advancements, several challenges persist, hindering the

full realization of a collaborative research ecosystem. Concurrently, the evolution of a unified concept for data sharing and linking presents promising opportunities for overcoming these challenges and fostering enhanced collaboration in research. We delineate the prominent challenges and explore the potential opportunities that a unified concept could usher. Data sharing and linking in collaborative research face a myriad of challenges that span across technical, policy, and ethical dimensions. These challenges are elaborated as follows:

- Technical challenges comprise issues related to data interoperability, integration, and standardization. Disparate data formats and inconsistencies in data structures often impede seamless data sharing and linking [3].
- Policy challenges encompass complexities in data governance, including issues surrounding data ownership, privacy, and security. The diverse regulatory landscapes across different regions further complicate policy adherence [5].
- Ethical challenges involve safeguarding the privacy of individuals and ensuring the responsible use of data. Ethical considerations necessitate frameworks that promote transparency and consent in data usage [1].

The conceptualization and implementation of a unified concept for data sharing and linking portend significant opportunities in fostering collaborative research:

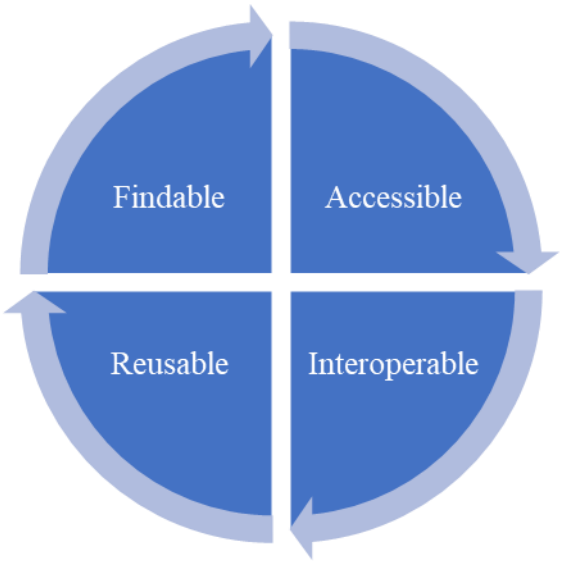
- A unified concept can facilitate greater collaboration by streamlining data sharing processes and fostering inter-disciplinary research endeavors [2].
- Technological advancements such as the integration of blockchain technology can offer secure and efficient mechanisms for data sharing and linking, paving the way for innovative research approaches [4].
- A unified concept can aid in the harmonization of policy frameworks, fostering a cohesive approach to data governance and facilitating compliance with diverse regulatory mandates [5].

As the landscape of collaborative research evolves, addressing existing challenges and leveraging emergent opportunities becomes critical. A unified concept for data sharing and linking stands as an important stage, promising to revolutionize the collaborative research arena by fostering enhanced collaboration, leveraging technological advancements, and harmonizing policy frameworks. The journey towards realizing this unified concept necessitates concerted efforts from all stakeholders, paving the path for a collaborative and innovative research ecosystem.

Sharing data can lead to new insights and discoveries, often beyond the initial scope of the original research. Open datasets can be used by multiple researchers to validate findings, explore new hypotheses, or combine with other data sources for richer analyses. Reducing duplication of efforts by making use of already collected data can save time and resources. This is particularly important in expensive domains, like large-scale experiments or medical research, where reproducing data collection can be cost-prohibitive. Open sharing of data can increase trust in scientific results, as independent parties can validate and reproduce findings. Shared data facilitates collaborations across disciplines, institutions, and even countries. These collaborations can lead to interdisciplinary projects and breakthroughs.

Technological aspects of sharing and linking of data resources

The advent of the digital era has brought forth an unprecedented surge in data generation and consumption. Ensuring the effective management, sharing, and utilization of this vast ocean of data, particularly in research, is paramount. Central to this pursuit are the GO-FAIR principles, which lay out a comprehensive roadmap for outlining fundamental principles that aim to bolster the discoverability and utility of data [3]. Specifically, the principles highlight the importance of making data findable, accessible, interoperable, and reusable – often abbreviated as FAIR (Fig. 1) [6].



Source: GO-FAIR Initiative (2023)

Figure 1: FAIR data principles

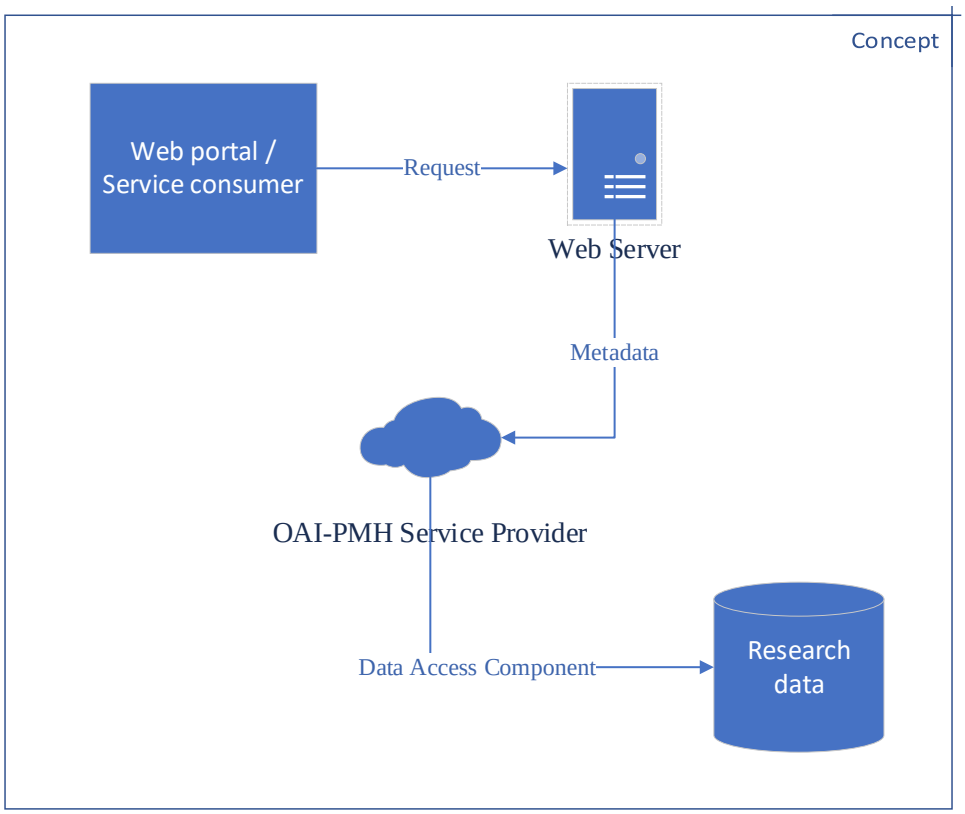
Platforms such as Zenodo, OpenAIRE, EOSC, and BPOS have been at the forefront in actualizing these principles, albeit with tailored approaches suited to their specific user demographics and regional mandates. Positioned as a universal repository, Zenodo is more than just a storage platform [7]. Its significance is accentuated when viewed through the prism of the FAIR principles. Each piece of data or research output uploaded to Zenodo is endowed with a unique Digital Object Identifier (DOI). This not only heightens the findability of the data but also assures its longevity and traceability. Open access, a defining feature of Zenodo, ensures that data remains accessible to researchers globally. Furthermore, its capability to support an array of data formats ensures that data from disparate sources can be integrated, promoting interoperability. By clearly delineating licensing terms, Zenodo also champions the reusability of data, ensuring researchers can build upon previous works with clarity on rights and attributions. Serving as a beacon in the open science movement in Europe, OpenAIRE is a testament to the power of collaboration [8]. Its vast repository, teeming with publications, datasets, and other research products, is a treasure trove for scholars. Advanced search algorithms ensure data findability, while its commitment to open access aligns with accessibility mandates. What sets OpenAIRE apart is its proactive approach to promoting interoperability. By providing tools and guidelines for data harmonization, it ensures that datasets from varied domains can be seamlessly merged. Moreover, by setting standards and offering data management recommendations, OpenAIRE fosters data reusability, ensuring the longevity of research outputs. The European Open Science Cloud (EOSC), an ambitious venture by the European Commission, is a paradigm shift in how research data is viewed and managed [9]. Its mission transcends mere storage. It aims to provide a cohesive environment where data from various research infrastructures across Europe can be harmonized and utilized. By creating a network of interconnected data sources, EOSC amplifies data findability. A robust governance structure, complemented by clear policies, ensures data accessibility while minimizing misuse. Efforts to integrate resources from diverse research domains underscore its commitment to interoperability. Finally, by championing open science and emphasizing proper data attribution, EOSC ensures that data remains reusable. As Bulgaria's foray into the open science arena, BPOS is a reflection of the country's commitment to democratizing research [10]. Its centralized structure enhances data findability, acting as a one-stop portal for Bulgarian research outputs. The commitment to standardizing data formats and aligning with broader European directives ensures interoperability. By providing clear guidelines on data usage and rights, BPOS ensures that research outputs are not only accessible but also primed for reuse. The essence of the GO-FAIR principles is to facilitate a seamless research ecosystem where data is not just a byproduct but a valuable asset. Platforms like Zenodo, OpenAIRE, EOSC, and BPOS, each in its unique way, are emblematic of these principles in action. As research continues to break

boundaries, both geographically and disciplinarily, such platforms will be instrumental in ensuring that data remains a unifying force, driving collaboration, innovation, and discovery.

The technological landscape has seen rapid developments that hold the potential to revolutionize the domain of data sharing and linking in collaborative research. The advent of sophisticated technologies could pave the way for a unified concept that fosters efficiency, security, and interoperability in data sharing and linking. The integration of Artificial Intelligence (AI) and Machine Learning (ML) in data sharing and linking has the potential to revolutionize data analysis and interpretation. AI and ML can facilitate the automation of data processing, helping in the identification of patterns and trends, which can significantly enhance collaborative research efforts [11]. Cloud computing offers a scalable and flexible platform for data storage and sharing. It enables real-time access to data from anywhere, fostering collaboration and facilitating seamless data integration across various research groups. Furthermore, cloud computing can aid in efficient data management through services like Data as a Service (DaaS), which can streamline data access and usage [12]. The Internet of Things (IoT) can be leveraged to enhance data collection and sharing in research. Big Data and IoT are one of the most promising technologies in our time [13]. IoT is a modern paradigm that describes the connection of almost everything to the Internet [14]. IoT devices can generate a vast amount of real-time data, which can be integrated into research databases for comprehensive analyses. This technology fosters connectivity and data exchange, opening new avenues for collaborative research [15]. The amalgamation of the IoT revolution with platforms upholding the GO-FAIR principles is a testament to the evolving nature of research.

Concept of sharing and linking of data resources via web services

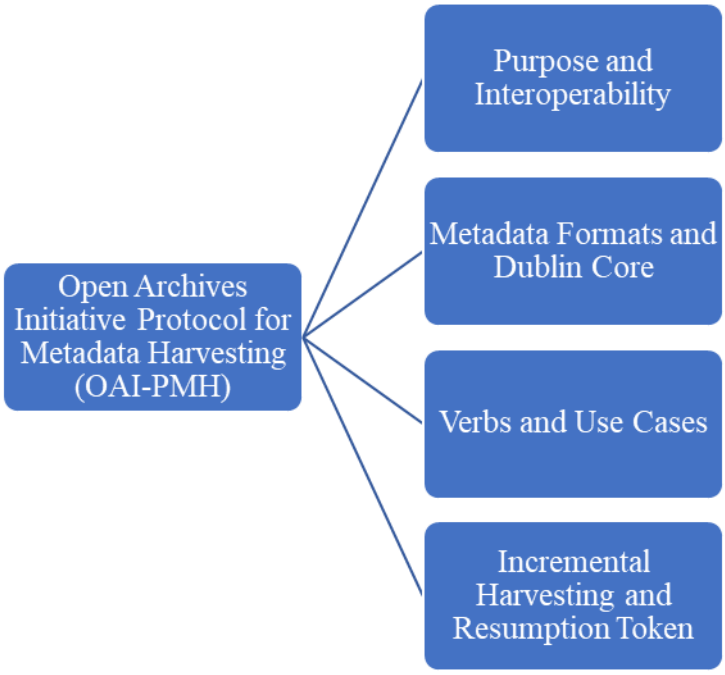
Technological advancements stand as a cornerstone in the evolution of a unified concept for data sharing and linking. Concept development is a prior process to the technological implementation of a corresponding prototype solution [16]. Sharing and linking of data resources in a unified way also requires the corresponding reorganization of business process management in organizations in the context of their digital transformation [17]. In the context of the present study, we will propose a concept for sharing and linking of data resources within the alliance of leading European universities in business, economics, and social sciences with rich experience in analyzing societal change, known as ENGAGE.EU (Fig. 2).



Source: Author

Figure 2: Conceptual approach for sharing and linking of data resources via web services

The objective of the concept is to show and share data resources on topics of societal change. We collect data of the repositories and existing data centers of different partner universities. Sharing and linking data resources will foster open science strategies in research [18]. For the purposes of sharing and linking of data resources, the OAI-PMH protocol is used. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a protocol developed for harvesting (or collecting) metadata descriptions of records in archives. It allows automated agents to access metadata from repositories that support the protocol (Fig. 3) [19].

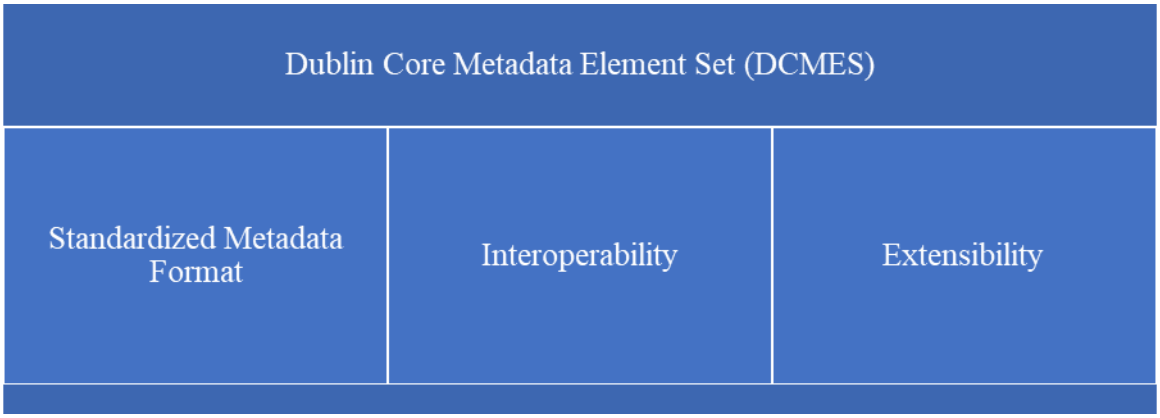


Source: Open Archives Initiative Protocol for Metadata Harvesting (2023)

Figure 3: Key points about OAI-PMH

Its primary goal is to facilitate the efficient dissemination of content. OAI-PMH allows for the exposure and sharing of metadata, making it easier for search engines or other services to find and access content from various repositories. One of the significant advantages of OAI-PMH is that it promotes interoperability. Repositories can expose their metadata in a standard format, allowing service providers to aggregate this metadata from various sources easily. While OAI-PMH can support multiple metadata formats, the only format that repositories are required to support is Dublin Core [20], a simple and standardized metadata schema. However, repositories can also offer their metadata in other formats. OAI-PMH operations are based on the use of verbs. These verbs allow users to get information about the repository. The protocol is used in various applications, especially in the realm of digital libraries, institutional repositories, and scholarly communication. OAI-PMH supports incremental harvesting, allowing service providers to request only those records that have been added or changed since a previous harvest. For situations where the response to a request is too large to be returned in one go, OAI-PMH uses the concept of a resumption token. This token can be used to continue a request from where it left off. The connection between OAI-PMH and Dublin Core can be understood through their intertwined roles in enhancing the accessibility and interoperability of digital resources (Fig. 4).

For repositories just starting with metadata exposure via OAI-PMH, the Dublin Core's simplicity offers a straightforward path. They can begin with basic metadata descriptions and later opt for more detailed or specialized schemas as needed. Given the widespread adoption of both Dublin Core and OAI-PMH in the digital library and repository communities, their combined use has become something of a de facto standard. This widespread acceptance ensures that tools, best practices, and community support are readily available. When repositories expose their metadata via OAI-PMH using Dublin Core, they ensure that their digital resources can be more easily discovered and accessed by a global audience. Aggregators, search engines, and other digital services can harvest this metadata, making the content of these repositories more widely known and accessible.



Source: Dublin Core (2023)

Figure 4: Connection of DCMES to OAI-PMH

At its core, OAI-PMH is a protocol developed for collecting (or "harvesting") metadata descriptions from repositories. It is based on the widespread use of web standards, particularly HTTP and XML. OAI-PMH itself is a web service, allowing different software on different servers to interact with each other over the internet. It uses HTTP for the transport of requests and responses, making it accessible and straightforward to integrate with various platforms. One of the foundational principles of web services is the standardization of requests and responses to ensure smooth communication between different software applications. OAI-PMH adopts this principle by defining specific "verbs" (like GetRecord, ListIdentifiers, etc.) that dictate the type of requests and expected responses. This ensures that any OAI-PMH compliant harvester knows exactly how to request data and what kind of response to expect, regardless of the specific repository it's communicating with. Web services need a standardized way to represent data that both the requester and the responder can understand. OAI-PMH uses XML for this purpose, a common choice for web services. XML allows for structured representation of metadata, which can be easily parsed and processed by different software applications. One of the main reasons for using web services is to achieve interoperability between different software systems, often running on different platforms and written in different programming languages. OAI-PMH is designed with this in mind. By adhering to the OAI-PMH protocol, repositories ensure that their metadata can be harvested and utilized by a diverse range of systems. OAI-PMH is flexible regarding the metadata formats it supports. While it mandates the support for a simple Dublin Core format, repositories can offer their metadata in multiple formats. This extensibility is a key feature of many web services, ensuring they can adapt to evolving requirements without overhauling the entire system. Web services are often used to automate tasks that would be tedious or impractical for humans to perform manually. In the context of OAI-PMH, automation is evident in how harvesters can periodically query repositories for new or updated records, ensuring metadata collections are up-to-date. This automated process can scale to accommodate large repositories or frequent updates. The proposed concept assumes the presence of an intermediate layer between the web server and data repositories to implement the OAI-PMH protocol. The use of an additional layer is common for concepts that include a metadata model [21]. Within concepts that involve data, it is a common principle to treat data as a very valuable asset and to occupy a central place in the respective concept [22]. In the context of sharing and linking of data resources, data should be available in real time from any point in the world, which is also considered one of the main reasons for the emergence of new directions in databases [23].

The principles that underpin OAI-PMH – standardization, interoperability, and efficient data dissemination – align closely with the broader goals of creating a unified concept for data sharing and linking. By leveraging such protocols, the research community can move towards a more collaborative and integrated environment where data resources from various sources can be effortlessly shared, linked, and utilized. The connection between OAI-PMH and Dublin Core represents a synergy between a protocol designed for efficient metadata harvesting and a standardized, simple metadata schema. Their combined use has greatly facilitated the growth and success of digital repositories, ensuring that digital

resources are both accessible and discoverable on a global scale. OAI-PMH is a shining example of web services applied to the domain of metadata harvesting. Its design principles, rooted in web standards, facilitate easy and standardized access to metadata across diverse repositories. This ensures that metadata from different sources can be aggregated, searched, and utilized efficiently, promoting the broader goals of open access and interoperability in the digital world.

Conclusion

In an increasingly digitalized research ecosystem, the imperative for effective data sharing and linking has never been more pronounced. The evolution of the research domain, with advancements like IoT, reinforces the necessity for protocols and platforms that facilitate seamless, efficient, and interoperable data management. Drawing from established concepts like OAI-PMH and the principles of web services, we glean insights into how such mechanisms can be developed, standardized, and universally adopted. Central platforms like Zenodo, OpenAIRE, EOSC, and BPOS, while varying in their scope and functionality, share a common goal – promoting the open sharing of data and research outputs. Their alignment with the GO-FAIR principles ensures not just the availability of data but its accessibility, interoperability, and reusability. The application of these platforms, bolstered by national initiatives like BPOS in Bulgaria, underscores the global push towards open science and research transparency. The integration of web services, especially in the context of OAI-PMH, exemplifies how technology can be harnessed to address the challenges of data sharing. The future of research collaboration will inevitably rest on the confluence of technology and policy, ensuring that data isn't just shared, but shared with purpose, clarity, and a commitment to advancing the collective knowledge of the global research community. As we look forward, it's evident that a unified concept for data sharing and linking, underpinned by robust technological infrastructures and harmonized policies, will be pivotal.

Acknowledgement

This work was supported by the Erasmus+ project "The European University engaged in societal change – ENGACE.EU" (Research Grant No. 101004071).

References

1. Borgman, C. L. (2017). Big data, little data, no data: Scholarship in the networked world. MIT Press.
2. Kahn, R., & Wilensky, R. (2006). A framework for distributed digital object services. *International Journal on Digital Libraries*, 6(2), 115-123.
3. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.
4. Zyskind, G., Nathan, O., & Pentland, A. (2015). Decentralizing privacy: Using blockchain to protect personal data. In *2015 IEEE Security and Privacy Workshops* (pp. 180-184). IEEE.
5. European Commission. (2018). Turning FAIR into reality: Final report and action plan from the European Commission expert group on FAIR data.
6. GO-FAIR Initiative. (2023). Retrieved from GO-FAIR Website: https://www.go-fair.org
7. Zenodo. (2023). Retrieved from Zenodo Website: https://zenodo.org/
8. OpenAIRE. (2023). Retrieved from OpenAIRE Website: https://www.openaire.eu/
9. European Open Science Cloud. (2023). Retrieved from EOSC Website: https://eosc-portal.eu/
10. Bulgarian Portal for Open Science. (2023). Retrieved from BPOS Website: https://bpos.bg/
11. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
12. Mell, P., & Grance, T. (2011). The NIST definition of cloud computing.

13. Stefanov, G. (2019). Analysis of Cloud based ETL in the Era of IoT and Big Data. In *Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education.* 198-202. ICAICTSEE.
14. Boyanov, L. (2021). A Conceptual Approach for Industrial Internet of Things Assessment. *Godishnik na UNSS*, (2), 97-107.
15. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660.
16. Tabov, Y. (2019). Application of Concept for Structuring of Web Data. In *Proceedings of International Conference on Application of Information and Communication Technology and Statistics in Economy and Education.* 266-270. ICAICTSEE.
17. Belev, I. (2018). Software Business Process Management Approaches for Digital Transformation. *Godishnik na UNSS*, (1), 109-119.
18. ENGAGE.EU. (2023). The European University Engaged in Societal Change. Retrieved from ENGAGE.EU Website: https://www.engageuniversity.eu/
19. Open Archives Initiative Protocol for Metadata Harvesting. (2023). Retrieved from OAI-PMH Website: https://www.openarchives.org/pmh/
20. Dublin Core. (2023). Retrieved from Dublin Core Metadata Element Set (DCMES) Website: https://www.dublincore.org/
21. Mihova, V. (2015). Methods of Using Business Intelligence Technologies for Dynamic Database Performance Administration. *Economic Alternatives*, (3), 105-116.
22. Delchev, D., & Lazarova, V. (2021). Big Data Analysis Architecture. *Economic Alternatives*, (2), 315-328.
23. Radoev, M. (2021). Contemporary Trends in the Development of Databases. *Ikonomiceski i Sotsialni Alternativi*, (1), 5-15.