

МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2024  
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2024  
*Proceedings of the Fifty-Third Spring Conference  
of the Union of Bulgarian Mathematicians  
Borovets, April 1–5, 2024*

**A PILOT CONCEPTUALISATION OF THE DATA SPACE  
ECOSYSTEMS FOR CULTURAL HERITAGE\***

**Kalina Sotirova-Valkova**

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences,  
Sofia, Bulgaria

e-mail: [kalina@math.bas.bg](mailto:kalina@math.bas.bg)

The digital heritage sector is among the nine common European data spaces outlined in *the EU Data Act*: Health, Industrial, Agriculture, Finance, Mobility, Green Deal, Energy, Public Administration, and Skills. *The data space* concept is complex, and being linked with the *data ecosystem* brings together methodologies and techniques from numerous domains: artificial intelligence (AI), text/data mining, data visualisation, mapping, image analysis, audio analysis, network analysis, and rights management. There is a need for clarifying data terminology in use for the GLAM field and transferable methodology to strengthen its' ongoing datafication and digital skill set. The paper offers a comprehensive literature review (incl. EU Data (space) policy documents) on *data spaces* and *data ecosystems* aimed at (1) formulating the main challenges in data collection, data processing, and data maturity for heritage-related information systems and (2) justification of the necessary mindset and digital skill set change for a mature museum. General conclusions are made as well as recommendations for the Bulgarian GLAM sector.

**Keywords:** data space, digital cultural heritage, data ecosystem, big data value.

**ЕКОСИСТЕМИ С ПРОСТРАНСТВА ОТ ДАННИ ЗА  
КУЛТУРНО НАСЛЕДСТВО – ОПИТ ЗА  
КОНЦЕПТУАЛИЗАЦИЯ**

**Калина Сотирова-Вълкова**

Институт по математика и информатика, Българска академия на науките,  
София, България

e-mail: [kalina@math.bas.bg](mailto:kalina@math.bas.bg)

---

\* <https://doi.org/10.55630/mem.2024.53.092-098>

2020 Mathematics Subject Classification: 94-XX.

Сектор дигитално наследство е едно от деветте общи европейски пространства за данни, посочени в *EU Data Act*: здравеопазване, промишленост, селско стопанство, финанси, мобилност, зелена сделка, енергетика, публична администрация и умения. Терминът *пространство на данни* е комплексен и свързан с *екосистема от данни*, като обединява методологии и техники от: изкуствен интелект (AI), извличане на текст/данни, визуализация на данни, *tapping*, анализ на изображения, аудио анализ, мрежов анализ, управление на права. GLAM екипите не само у нас, но и в Европа се нуждаят от дефиниране на ползваната терминология за данни, както и от методология, която да подобри както обработката на данни, така и набора от техните дигитални компетентности. Статията предлага изчерпателен преглед на литературата (вкл. документи на ЕС) относно пространствата от данни и екосистемите от данни, насочени към (1) формулиране на основните предизвикателства при събирането, обработката, и *зрелостта* на данните (*data maturity*) в информационните системи, свързани с наследство, както и (2) обосновка на нужната промяна в нагласите и дигиталните умения на *зрелия музей*. Направени са общи изводи и препоръки за българския GLAM сектор. **Ключови думи:** пространства от данни, дигитално културно наследство, екосистема от данни.

**Introduction.** GLAM sectors (galleries, libraries, archives, museums) are currently working on developing innovative ecosystems which would transform the use of digital collections. These developments should support researchers and educators, two important GLAM user communities interested in the application of available computational tools to ask questions on large and distributed collections and look for patterns. At the same time, the development of the data space is following the major trend set up by the EU Data Act [10]. The *Big Data Value Association* (BDVA), a counterpart of the European Commission in the Big Data Value Public-Private-Partnership, explains the new buzzword *datafication like this*: “rather than referring to data as the new oil, we think of *datafication* as a new currency underpinning a new data economy . . . Opportunities presented by Artificial Intelligence (AI) . . . cannot be fully exploited without wider access to large volumes of high-quality data; which remains severely limited. We consider widespread, secure and effective data sharing as the only solution” [1]. Limited high-quality interoperable data is the main obstacle for the heritage domain as well. Being a Digital Humanities researcher, ICOM and ENA member, the author’s goal in this paper is to summarise the emerging data space challenges for the heritage sector, showing what knowledge, expertise, and skills are necessary to make a successful GLAM digital transformation through data.

**Research question and methodology.** Exploring the development of new infrastructures is a complex task with so many developments happening in parallel and in an area where a clear concept is still being coined. This paper is an initial attempt to collect and analyse different contributing perspectives and developments. The methodology used was a desk study. We did a comprehensive search for academic papers and policy documents, as well as examples of the most recent infrastructural efforts in this domain.

**Defining data spaces and data ecosystems for cultural heritage.** Conceptualising a data space is a cross-domain task, which includes a wide variety of computer science topics such as artificial intelligence, text/data mining, data visualization, mapping, image analysis, audio analysis, and network analysis. GLAM sector is one relatively

small sub-area among the nine common European Data Spaces outlined in the *European Strategy for Data* [8], *EU Data Act* [10], *European Data Governance Act* [9]: Health, Industrial, Agriculture, Finance, Mobility, Green Deal, Energy, Public Administration, and Skills. At the core of the heritage data space concept is improving *metadata quality in a way that adds value and increases inclusiveness, visibility and reusability*. This should include metadata and content improvement using new technologies like machine-learning algorithms to enrich metadata records. According to the *International Data Spaces Association* [13], the European data spaces ecosystems basically will bring “new services for users, based on enhanced transparency and data sovereignty; data sharing and exchange without dominance of, and dependency on, large, quasi-monopolistic players”. According to the IDSA, there are **four design principles for data spaces**: data sovereignty, data level playing field, decentralized soft infrastructure, and public-private governance (opendei.eu). These principles are based on a coordinated approach to establishing data space ecosystems with public-private governance and data space interoperability.

The conceptualisation of the new infrastructures is driven by two major initiatives: (1) *European Collaborative Cloud for Cultural Heritage* (ECCCH), aimed at the inclusive design of collections as data ecosystems equipped with AI potential for knowledge extraction, and (2) Europeana as the leader of the consortium implementing the European data space for cultural heritage. A major shift which needs to take place is to re-configure the data input in Europeana which should enhance aggregation with intelligent data analysis tools (AI, data mining, natural language processing). These tools “understand” collections of data owners/providers in a way that addresses specific user requests, which are not always using keyword searches. Europeana Foundation R&D team has created a methodology for evaluating and validating the results of (automatic) data enrichment efforts<sup>1</sup>.

A data space and a data ecosystem are both terms used to describe a collection of data-related resources and entities. While data space typically refers to a more bounded and controlled environment through which data is collected, stored, accessed, processed, managed and shared by a specific group of stakeholders and sharing data happens within agreed rules and policies, data ecosystem, on the other hand, refers to a larger, more open and interconnected environment where data is created, shared and used by multiple stakeholders across a variety of domains. Data ecosystem can include different types of data sources, such as open data, protected data, and user-generated data and is typically characterized by its diversity, complexity, and dynamism. According to BitCOM, research on a data ecosystem should include services as a preliminary step (e.g. generation and preparation) for sharing data in a data space, and services as a processing step after receiving data from a data space to generate concrete added value from it<sup>2</sup>. Among Data Ecosystem services should be mentioned: Analytics, Data Semantics and Validation, Anonymisation, and Compliance, while among Data Space services are authentication, data transfer, data catalogue, and application catalogue.

In summary, both data spaces and data ecosystems process data-related resources and

---

<sup>1</sup>Europeana R&D team, (2023) Methodology for Validating Enrichments [<https://pro.europeana.eu/post/methodology-for-validating-enrichments>] Last visited: 31-01-2024.

<sup>2</sup>Data Spaces and Data Ecosystems: First Explainer and Current Status, BitCOM Inc. 2022. [[https://www.bitkom.org/sites/main/files/2022-09/Bitkom\\_Data\\_Spaces\\_Data\\_Ecosystems](https://www.bitkom.org/sites/main/files/2022-09/Bitkom_Data_Spaces_Data_Ecosystems)] Last visited: 9-02-2024.

entities, but data space could be seen as the “inner core” of a data ecosystem. Edward Curry frames the theoretical foundations and principles of real-time linked data spaces within real-world smart ecosystems [4, 5] like this: “in its purest form, a data space is a set of participants and their inter-relations, where the participant is known as each data source (e.g., database, CSV, web service)”. Scerry [15] adds detail to this definition arguing that data space is: “any ecosystem of data models, datasets, ontologies, data sharing contracts and specialised management services (i.e., as often provided by data centres, stores, repositories, individually or within ‘data lakes’), together with soft competencies around it (i.e., governance, social interactions, business processes)”. This means that as for heritage domain soft competencies should be well combined initially with data infrastructure and management services, clear intellectual property rights (IPR) and copyright rules.

The main goal and purpose of data ecosystems in general is knowledge extraction and controlled data sharing administered in a way that allows constant and dynamic various data input. Wang, Y. [17] goes deeper into data integration, focusing on essential details: “A data space system processes data with various formats, accessible through many systems with different interfaces, such as relational, sequential, XML, RDF, etc. Unlike data integration over DBMS, a data space system does not have full control on its data, and gradually integrates data as necessary.” This means that analysis and knowledge extraction of various heritage data is at the core of the data spaces ecosystem. Such background can be very useful for the heritage domain, but conceptualization and contextual modelling need further research.

**European data (space) policy.** Successful digital transformation of the heritage sector, according to Europeana Strategy 2020-2025, is based on: (1) capacity building (Europeana Education Community), (2) strengthening the infrastructure, and (3) improving data quality. In this respect, data spaces are seen as the cornerstone of the EU data strategy, participating in one ongoing journey for excellence in smart ecosystems, altogether with the DEAI movement (Diversity, Equity, Accessibility, Inclusion), FAIR and CARE principles for open data and open science.

As for the policy, the European data space for cultural heritage is being developed with the support of the EC *Digital Europe Programme*. It affects the common digital market (tourism, education, creative industries) through modelling, experimentation, standardisation, deployment, and awareness, each targeted toward specific stakeholders in the data-sharing ecosystem of Europeana. The goal is to provide high-quality cultural content, particularly 3D (*Metaverse museums*), secure reuse of cultural resources, and enriched services. It is currently the main task of the Europeana Initiative, which, since September 2022, has been at the heart of the common European data space for cultural heritage, a flagship action of the EU, supporting GLAM sectors’ digital transformation. The expected outcomes include: new automated AI-based services for developing knowledge extraction over open linked data; more inclusive digital heritage environment, unified and jointly owned; new research opportunities delivered through new ways for analysing information about collections.

**Defining data maturity towards data mature GLAMs.** Significant challenges for a fully functional GLAM data spaces ecosystem, except for capacity upgrade, are technical: interoperability, quality and accuracy, decentralized data sharing and processing

architectures, *data maturity (DM)*, data verification and provenance support, strict rights management and IPR policies. If we move DM models from business analytics to heritage analytics their main goal should be to inform stakeholders whether they're taking the right approach to their intangible assets; and how reliable, efficient, and effective your organization is at managing data. According to business intelligence research, the four stages of data maturity are awareness, proficiency, savviness, and uncompromising data-driven decision-making<sup>3</sup>. The scale on which IDSA<sup>4</sup> measures the *maturity of data spaces* is a bit different: (1) lead-in, (2) committed case, (3) pilot, and (4) live [14]. Secure and sovereign data sharing is at the fourth live stage. Currently, data spaces in the domains mobility, manufacturing, energy, automotive, smart cities, and supply chain reached the pilot or live stages, but heritage is still in the initial conceptualizing phase. In the DM museum domain we have to mention the "Museum maturity framework" [12] which presents *five development stages of a mature museum*, which is a good tool to diagnose museum management requiring changes in the leadership, cultural and structural changes. Lauren Vargas proposes defining museum digital maturity through its data policy and practice [16]. She explores applying the maturity model for business processes to cultural institutions, showing that data collection and use requires a *well-nourished ecosystem of interconnected people, processes and technologies*; she claims that as soon as the GLAM sector understands the meaning and potential of digital skills as quickly will go digital transformation. At the same time, it is clear that there is no single or uniform path to digital maturity. Digital maturity is achieved when all aspects of the digital ecosystem operate uniformly as a system towards a future where cultural institutions may manage their growth through introspective study and predictive data modelling [16]. Innovation labs also contribute to more mature data services in GLAMs [7].

**The Bulgarian case.** GLAM sector in Bulgaria lacks a focused national policy and roadmap exploring how data space concept and initiatives like ECCCH will affect the way we preserve, research and present cultural assets. The other important challenges can be systematized as follows: mindset change and clear understanding of what means in practice to present collections as data; low or average level of digital competencies; lack of funding for specialized training.

As for digital skills, the *DigiComp* EU framework offers many resources, which could be seen as a good base for further specialization for heritage specialists. There are good educational resources for citizens for main and *specialized digital skills*<sup>5</sup>, but not all are available in Bulgarian language (aged Bulgarian museum staff does not know foreign languages). In 2023 Europeana team created a framework<sup>6</sup>, which defines 9 types of digital skills needed for educators working in the cultural heritage sector and creative industries;

---

<sup>3</sup>Google and Deloitte *Data Maturity Benchmark* [<https://digitalmaturitybenchmark.withgoogle.com/retaildatatransformation>], [<https://www2.deloitte.com/us/en/pages/consulting/articles/digital-transformation-through-data-for-retail.html>] Last visited: 31-01-2024.

<sup>4</sup>Cultural heritage data spaces are still not mapped on the IDSA radar (they are supposed to be placed in the so-called "Cross-Domain" category).

<sup>5</sup>NEMO TOOLKIT [<https://www.ne-mo.org/news-events/article/toolkit-aims-to-empower-museum-professionals-with-digital-skills>] Last visited: 31-01-2024.

<sup>6</sup>Europeana digital skills framework for CH educators and museum teams [<https://pro.europeana.eu/post/what-digital-skills-are-needed-for-educators-working-in-cultural-heritage-institutions>] Last visited: 31-01-2024.

among them are inclusive learning, skills making easier communication, collaboration, and project management, digital interactive storytelling tools, virtual exhibitions, digital research skills, strong digital literacy skills to facilitate digital curation and present collections as data, metadata management of images, videos and documents; digital preservation techniques and digital storage systems, incl. creating back up strategies. In this respect, we have to mention NGOs like CHARTER (European Cultural Heritage Skills Alliance) [2], which work towards creating a comprehensive strategy and resources that will guarantee that cultural heritage professionals are keeping their knowledge up to date with their digital natives' user level. In the last year, there has been huge ongoing work on preparing qualitative and up-to-date<sup>7</sup> resources that are not yet available in Bulgarian. As for data spaces and collections as data realities, there is still no representative analysis for the Bulgarian GLAM sector to be mentioned here.

The analysis from 2022 of the current state of digitization in Bulgarian museums<sup>8</sup> among 109 museum teams shows that for 54% of them less than 15% of their collections have been digitized; at the same time for 13% of respondents more than 50% of their collections are digital already. Only 6% of the respondents, museum staff, claim that are using web platforms, metadata, and cloud technologies for digitization, which shows a sad overall picture. In addition, the answers to the question "Do the representatives of your team have the necessary digital competencies for cultural assets digitization" show that most of them can create and process photographic images, video and scanning software/hardware, but only 4% of the respondents are familiar with 3D scanning, less than 2% claim that their team does not have digital competences. The good news is that most Bulgarian museum teams are ready to achieve new practical knowledge on digital skills.

*Concluding*, methodology for generation and work with data spaces is a work in progress in Bulgaria. The status quo at the beginning of 2024 is based on the rich know-how gained from the national Europeana aggregator (Public Library Pencho Slaveykov, Varna), the National Archives, and research and digitization activities of national infrastructures CLADA-BG and HERITAGE BG. The expected digital transformation of the heritage sector is linked with Bulgaria's Recovery and Resilience Plan (Digital Transformation Branch), in which frame the Ministry of Culture is managing a mega-project for the creation of a national repository. All these initiatives increased common knowledge level in the sector which raised awareness for applying contemporary data policy trends.

**Conclusion.** This paper looked at a number of carefully curated sources in the emerging domain of data spaces for cultural heritage. In this domain, no single sector can resolve the complexity of building data spaces, and the overview of recent developments shows several developments with diverse stakeholders. The research question for Digital Humanists, GLAM and IT sectors concerning the data space concept is how all actors involved in the creation and management of the digital heritage ecosystem should work together on delivering a large-scale data-rich environment. The technology and ambition are here, but the GLAM sectors are not fully engaged in this development yet. To present, structure and analyse heritage collections within data spaces requires rethinking

---

<sup>7</sup> "Up-to-date" resource to be understood as resource using *data spaces* and *data collections* concepts.

<sup>8</sup> InterDIGImuseum EU funded Project 2021-1-BG01-KA220-ADU-000035298 [<https://interdigimuseum.eu/>] Last visited: 31-01-2024

data input, AI-based tech infrastructure and consensus between GLAM specialists on their concrete role in the process. Large quantities of high-quality data are crucial for AI solutions and assume serious upfront costs associated with data collection, integration, and sharing. This limits large-scale data-driven projects to those with the necessary expertise. For the European GLAM sector achieving such expertise is becoming a must.

## REFERENCES

- [1] BigDataValue Association. Towards a European-Governed Data Sharing Space - Enabling data exchange and unlocking AI potential. Position Paper (2020). [https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2\\_2020\\_Final.pdf](https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf), Last visited: 12-12-2023.
- [2] CHARTER Alliance. <https://charter-alliance.eu/about-us>, Last visited: 12-12-2023.
- [3] Collections as Data: Part to Whole (Andrew Mellon Project, 2021-2023) Deliverables. <https://collectionsasdata.github.io/part2whole>, Last visited: 12-12-2023.
- [4] E. CURRY. Real Time Linked DataSpaces – Enabling Data Ecosystems for Intelligent Systems, Springer Open, 333 pp, 2020. <https://doi.org/10.1007/978-3-030-29665-0>
- [5] ED. CURRY et al. Common European Data Spaces: Challenges and Opportunities. Cham, Springer, 2022 [https://doi.org/10.1007/978-3-030-98636-0\\_16](https://doi.org/10.1007/978-3-030-98636-0_16).
- [6] Design Principles for Data Spaces. International Data Spaces Association, Dortmund, 111 pp, 2021, <https://doi.org/10.5281/zenodo.5244997>.
- [7] M. DOBREVA, K. STEFANOV, K. IVANOVA. Data Spaces for Cultural Heritage: Insights from GLAM Innovation Labs. In: ICADL 2022 Proceedings, 492–500. Springer-Verlag, [https://doi.org/10.1007/978-3-031-21756-2\\_41](https://doi.org/10.1007/978-3-031-21756-2_41).
- [8] EU Strategy for Data, February 2020. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>, Last visited: 12-12-2023.
- [9] European Data Governance Act, June 2022. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>, Last visited: 12-12-2023.
- [10] EU Data Act, February 2022. <https://digital-strategy.ec.europa.eu/en/policies/data-act>, Last visited: 12-12-2023.
- [11] Deployment of a common European data space for cultural heritage: PM. Annual report M12. Europeana Foundation, 2023, <https://pro.europeana.eu/page/data-space-deployment>, Last visited: 12-12-2023.
- [12] P. Humphrey, J. Yochim. The Museum Maturity Framework: A Path to Purpose, Meaning, and Values. *Curator – The Museum Journal*, **43**, no. 4 (2000), 313–323, <https://doi.org/10.1111/j.2151-6952.2000.tb00015.x>.
- [13] IDSA: Implementation of European Strategy on Data Role of the International Data Spaces (IDS). Position Paper of IDS Association, 2020. <https://internationaldataspaces.org/publications/most-important-documents/>, Last visited 12-12-2023.
- [14] IDSA Radar: Maturity of Data Assess. [https://internationaldataspaces.org/wp-content/uploads/dlm\\_uploads/Data\\_Spaces\\_Radar\\_Web\\_230307.pdf](https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/Data_Spaces_Radar_Web_230307.pdf), Last visited: 12-12-2023.
- [15] S. Scerri, T. Tuikka, I. Lopez de Vallejoan (Eds). Towards a European Governed data sharing space, BDVA Position paper, 2020, [https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2\\_2020\\_Final.pdf](https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf), Last visited 12-12-2023.
- [16] L. VARGAS. Understanding museum digital maturity through data collection and use. Doctoral dissertation, University of Leicester, 2019.
- [17] Y. WANG et al. A survey on accessing dataspace. *ACM SIGMOD Record*, **45**, no. 2 (2016), 33–44. <https://doi.org/10.1145/3003665.3003672>.