From Digitization to Digitalization: a Conceptual Model for Transforming the Bulgarian State Archives into a Socio-Technical Infrastructure

Yasena Chantova[0009-0003-8997-0838]

University of Library Studies and Information Technologies,
119 Tsarigradsko Shose Blvd., 1784 Sofia, Bulgaria
y.chantova@unibit.bg

Abstract. This research puts forward a conceptual model for transforming the Bulgarian State Archives from digitization to digitalization practices. Using conceptual analysis and archival standards it identifies systemic, metadata, and AI integration layers. It argues that sustainable digital archives demand an infrastructure based on interoperability, long-term preservation, and advanced access to archival knowledge.

Keywords: Digital Archives, Digitalization, Bulgarian State Archives, Digital Cultural Heritage.

1 Introduction

Analogue preservation practices have undergone considerable change at the start of the 21st Century. Now in the 2020s, these changes have been transforming memory institutions and their activities into hubs of digitally-mediated environments. As part of this transformation, it has become essential to make the difference between *digitization* and *digitalization*. (Brennen & Kreiss, 2016). What is understood by digitization is the conversion of analogue carriers into digital formats. This includes scanning a paper document into PDF, converting magnetic tapes into WAV files or capturing photographic images in TIFF format. A digital copy is the end product of this process. Digitalization in essence differs from digitization in that it comprises the integration of digital processes, infrastructures, metadata standards and automated workflows within the institutional practice. This means a broader organizational and technological transformation is enlisted to achieve a digital archival object embedded in the system managed by the archival institution.

What are the implications for the Bulgarian State Archives? The digitization-focused approach only represents the logic of analogue forms into digital formats. The broader and wider opportunities are offered by the digitalization-centered approach, where workflows, metadata models, access paradigms and preservation strategies are not left behind.

Digital Presentation and Preservation of Cultural and Scientific Heritage. Volume 16, pp. 423-432, 2026.
DOI: 10.55630/dipp.2026.16.35, ISSN 1314-4006, eISSN 2535-0366.

2 Exposition of the Investigation

2.1 Research Aim and Methodology

This research aims to put forward a conceptual model for transforming the Bulgarian State Archives from digitization to digitalization practices. To achieve this main aim, the methodology comprises conceptual analysis, archival theory and the technological infrastructure development, metadata management, preservation frameworks, institutional policies and user-oriented access mechanisms providing a unified digital ecosystem to ensure authenticity, contextualization, interoperability and sustainable long-term access.

2.2 The Emergence of the Concept of the Digital Archive

The modern vision of the digital archive came to existence (Petkova & Atanasov, 2023) during the 1990s and early 2000s, a period when deep technological and conceptual transformations occurred. During these years, archives gradually transitioned from traditional analogue models to digitally mediated working ambiences. It was not just the scanning technologies that sparked this transformation, rather, a redefinition of the archive as a system emerged.

The advent of the Internet changed everything (Borgman, 2015). Before the late 1980s, archives basically functioned as a service provider for information physically kept in a form of catalogues, indexes, and local databases. In the 1990s, the Internet allowed users to retrieve information remotely. As a result of this, simply creating archival descriptions was not enough anymore. There had to be an online publication of such descriptions. The digital archive became a public service and information structure.

However, a key problem persisted and that was the system changes that took place. The media carriers grew old with advanced technologies approaching and that posed problems to only scanning and publishing documents. Storing and maintaining digital information over time were recognized as important factors for long-term preservation and institutional framework strategies.

2.3 From Digital Collections to Online Archives

Three successive phases of digital archives evolution must be mentioned here (Conway, 2010). Digital collections began to be made; online archives permeated the professional area of archives (these were presented online as thematic exhibitions and visual resources raising the level of public visibility); and virtual reading rooms became popular. However, systematic archival contexts were missing: there were no links to archival fonds, provenance information or preservation policies whatsoever.

Online archival systems (Gilliland, 2014) marked the next stage in the development of archives. These highlighted the archival principals of provenance, original order, and hierarchical description. International archival standards were developed and approved,

probably the most popular of which is the General International Standard Archival Description (ISAD(G)). It was published in 1994 and then revised and amended in 2000. What made it special was the fact that it introduced the logic for archival description relying on hierarchical levels such as archival fonds, series, file, and item. Introducing this standard played an essential role because it helped what was lacking so far to become available like: consistency, comparability, interoperability and served as a conceptual foundation for digital archival systems.

2.4 Standards that Stabilized the Digital Archive

Other international standards were developed in the 1990s and the early 2000s, whose role was to consolidate the digital archive as an institutional form (Baca, 2016). Open Archival Information System a.k.a. OAIS had a great impact for it was adopted as ISO 14721 in 2002. Among the merits of this standard should be mentioned its specific architecture, which comprises six key components: ingest, archival storage, data management, preservation planning, access and administration.

Encoded Archival Description (EAD) is an XML-based standard for encoding archival search. Its contribution to the development of digital archives is that now archival descriptions were machine-readable, which meant that their online publication and exchange between institutions were now made possible. The archival catalogues could now function in their new role as interfaces among archival institutions and the public.

At the same time, the emergence of Dublin Core and the OAI-PMH protocol provided mechanisms for metadata interoperability. Dublin Core gave birth to a minimal set of standardized elements for describing digital resources. OAI-PMH on the other hand, made possible the automated harvesting and aggregation of metadata from distributed repositories. These technologies working in unison created the possibility to build digital heritage unified portals integrating resources from different institutions.

Last but not least, standards such as METS (Metadata Encoding and Transmission Standard) allowed digital objects to be structured and packaged within digital repositories. METS's function is to organize files, structural relationships, and administrative metadata, which means that it assists the sustainability in managing and transferring digital archival resources.

2.5 Digital Archives as Information Systems

The 2000s and 2010s marked another stage in the development of digital archives. They came out as stable and intricate informing systems and not just isolated digital collections. It seems they reached a status where they were able to manage digital objects, metadata and workflows (Palmer et al., 2008). Archival resources were now managed by centralized infrastructures. Repository platforms and digital library systems began to support structured workflows for metadata management, indexing and online publication. What followed as a result of these developments was that digital archives were now functioning as infrastructures in aid of scientific research, cultural heritage platforms providing access to historically vital information.

2.6 Toward Intelligent and Open Digital Archives

Advanced computational technologies have been integrated into digital archives since the 2010s, which meant that large-scale digitization came about together with automation (Jaillant & Caputo, 2022). Digital materials suddenly grew in number due to the expansion of mass digitization, including born-digital records. Manual processing simply became difficult if not impossible to handle, which led to the embracement of automated pipelines and standardized workflows to be ingested, processed and accessed. The workload was assisted by introducing artificial intelligence technologies like OCR, handwriting recognition, named entity recognition and semantic search.

New forms of knowledge discovery (Smiraglia, 2014) were triggered by the next important development – semantic networks and knowledge graphs – which played the role of mediators between archival documents and people, institutions, places and events. Owing to these connections, we can now witness the transformation of the traditional hierarchical archival structure into a networked knowledge environment. In practical terms, this means users being able to reconstruct historical contexts and explore relationships among records. This has led to the reengineering and redefining the role of archives and downloadable datasets in society opening the door to digital humanities search, new educational initiatives and cultural innovation. Collective knowledge production is enhanced and archives have made yet another step toward becoming mediators between institutional memory and society.

3 Results

The history of the digital archives and the line of their technological development have demonstrated the prerequisites for their modern-day condition. In theory, however, their transformation into intelligent and networked infrastructures, has placed a large gap as the real-life maturity and practical realization of these systems lag behind significantly in many national archival systems, including Bulgaria. Despite the fact that digital platforms have been created, which counts as an enormous step forward, the transformation of the Bulgarian Archival State Agency (Darzhavna agentsia "Arhivi", n.d.)) into a fully integrated digital ecosystem remains partial. A conceptual framework is above all needed to align technological infrastructures, archival standards, institutional processes and user-centered services as internationally outlined (International Council on Archives & Australasian Digital Records Initiative, 2008). The next section of the article will therefore suggest a conceptual model for the transformation of the Bulgarian Archival State Agency pointing out the structural and functional layers needed to support sustainable digitalization to enable advanced access of archival knowledge.

3.1 Digitalization as Systemic Transformation in the Archival Environment

Speaking of digital archives, digitalization brings about a structural transformation in the way archival institutions organize, describe, manage, and provide access to documents. Digitalization stands for a shift from a material model of preservation and conservation to an informational model of management. The institutional architecture and the processes that sustain archival activity undergo a complete change. As a result, the archive begins to operate as a structured information system, and the digital archive becomes an integrated socio-technical system.

Implementation of Electronic Archival Management Systems. Electronic archival management systems are the first step toward the systemic change. The end result is the creation of an integrated relational or XML-based databases to change the dynamics of managing the archival descriptions. Compatibility with international standards such as ISAD(G) and Dublin Core ensures interoperability and alignment with global archival practices. Meanwhile, mechanisms are introduced by digital systems to access control traceability, and accountability aiming to strengthen the institutional responsibility and evidential reliability. Then we have an archive that is not just a repository of files, but a structural informing environment.

Creating a Metadata Architecture. Metadata architecture is the next area of importance. Descriptive, administrative and technical metadata in standardized formats (XML, TEI, METS) needs to be systematically implemented. Digitalization is impossible without metadata structures consisting of for instance named entities, ontologies, controlled vocabularies, and semantic linking mechanisms, including Linked Open Data approaches. These mechanisms are a guarantee that the archival document becomes an information object to be searched, linked, analyzed and contextualized as part of wider knowledge networks. The metadata serves as a mediator between archival theory and digital infrastructure, operationalizing in computational form the principles of provenance, authenticity, and context.

Integration of Artificial Intelligence Tools. The third inseparable and indispensable component is the integration of artificial intelligence tools. Such integrated tools intensify the archive's functionality beyond simple storage and retrieval. Artificial Intelligence Tools might include OCR and HTR for text recognition, named entity recognition (NER), thematic classification, even extraction, and multimodal visual-text models. The digital archive, having embraced the idea of integrating AI into its functional components, starts to act as an intelligent informing environment capable of analytical operations. It is important to point out that AI does not replace archival principles; it amplifies them when scalable interpretation is introduced, discovery is made, and semantic enrichment is achieved.

3.2 Conceptual Model of Transformation

In the suggested model, three levels of transformation (Table 1) can be mentioned within digital archives. The *technical level* (digitization) modifies the archival carrier and produces a digital file. The *systemic level* (digitalization) transforms the archival institution and its core processes, making a digital archive. The *Intelligent level* (AI integration) enhances information processing and helps to create an analytical environment. These levels happen in a sequence and yet they are interdependent as they form a developmental trajectory from technical conversion to socio-technical infrastructure and ultimately to an intelligent archival ecosystem.

Table 1. Conceptual Model of Transformation.

Level	Process	Object of Transformation	Result
Technical Level	Digitization	Archival carrier	Digital file
Systemic Level	Digitalization	Archival institution and its processes	Digital archive
Intelligent Level	AI Integration	Information processing	Analytical environment

What defines success is the establishment of a coherent digital system capable of ensuring long-term preservation, semantic interoperability, controlled access, and analytical potential. For the Bulgarian Archival State Agency, recognizing digitalization as systemic transformation lays the foundations toward becoming a fully articulated socio-technical infrastructure.

4 Discussion

4.1 Current State of the Digitalization of Archival Documents in Bulgaria

For the last 20 years, the Bulgarian Archival State Agency has developed its digital infrastructure step by step in line with wider European policies on digital cultural heritage (Tsvetkova & Petrova, 2015). A key component of this process is the e-Archiving System SEA, which features structured online access to the National Digital Archives. The platform also has digitized audiovisual materials – such as magnetophone and video recordings, negatives, and film reels – as well as electronic records coming from state and municipal institutions and from private collections of historical or/and cultural value.

An important addition to the platform above is the Information System of State Archives (ISDA) (Informatsionna sistema na darzhavnite arhivi, n.d.), which began in 2012 as the main digital infrastructure of the agency. ISDA offers researchers to consult archival entries from regional state archives through a unified online interface and in this way it facilitates nationwide access to archival information. Moreover, in 2008, a

consistent series of national and European initiatives laid the foundations for a systematic digitization of textual archival documents, an example of which is the Project 43 initiative "Digitization of Museums, Library and Archival Fonds" of the Ministry of Culture (Proekt 43, n.d.). Archival digitization was enhanced through technological infrastructure and methodological support and in the end, document retrieval was enabled and the visibility of archival fonds, archival units and digitized registers was organized and improved.

Despite these developments, the general level of digitization remains constrained. Only about 1% of the total archival holdings – approximately 200,000 documents out of nearly 94 linear kilometers of records – have been digitized. Similar omissions between digitization ambitions and results have been observed worldwide due to the high costs and complexity of large-scale projects (Perry, 2014).

4.2 Challenges in the Digitalization of Bulgarian Archival Documents

It is no surprise that the management of the Bulgarian Archival State Agency are obliged to emphasize that digitization should not be understood merely by scanning documents. Sustainable digitalization demands institutional adaptation, steady funding, and appropriate regulatory mechanisms.

A major obstructions occur in the sphere of financial and technical support, and human resources. Among the issues are the limited funding opportunities, the outdated equipment, and a shortage of well-trained experts, which hinders the process of digitalization.

Many archival repositories are also facing infrastructural problems, including inadequate storage conditions and insufficient climate control. This in its behalf leads to threats to the safeguarding of archival fonds and complications in the efforts for digitization. Comparable resource obstacles have been found in other national archival systems (Colavizza et al. 2022).

There are access and quality issues as well. Some audiovisual resources are available only in partial or demonstrative versions, which diminishes their research value. The reliability of optical character recognition (OCR) for textual documents is often limited due to historical orthography, deteriorating paper, or irregular typography. A number of scientific studies indicate that OCR quality has its negative impact on the way search texts are used and the computational analysis is made. (Memon et al., 2020; AlKendi et al., 2024).

Inconsistent metadata practices are the main reason for further hindrances. Despite the fact that international standards are adopted on a regular basis, many older archival fonds are still poorly described, which is the reason for their reduced visibility and interoperability in digital systems (Ogilvie, 2016).

And finally, legal and institutional constraints in the form of copyright issues and unclear access policies have caused issues related to limitation in the availability of particular digitized collections. Without heavier investment and timely reforms in the systems, the pace of digitalization and the transformation of archives into a socio-technical infrastructure will continue to be barely enough to meet the long-term needs of preservation, accessibility, and digital research (Friedewald at al., 2024).

4.3 Possible Practical Implementations of the Three-level Model

The following pipeline is proposed to ensure the practical implementation of the conceptual three-level model of transformation (s. Table 1) into a workflow (Fig. 1). Each level has its own inputs, processes and outputs with their main tasks assigned to them. This turns the conceptual model into an operational sequence.

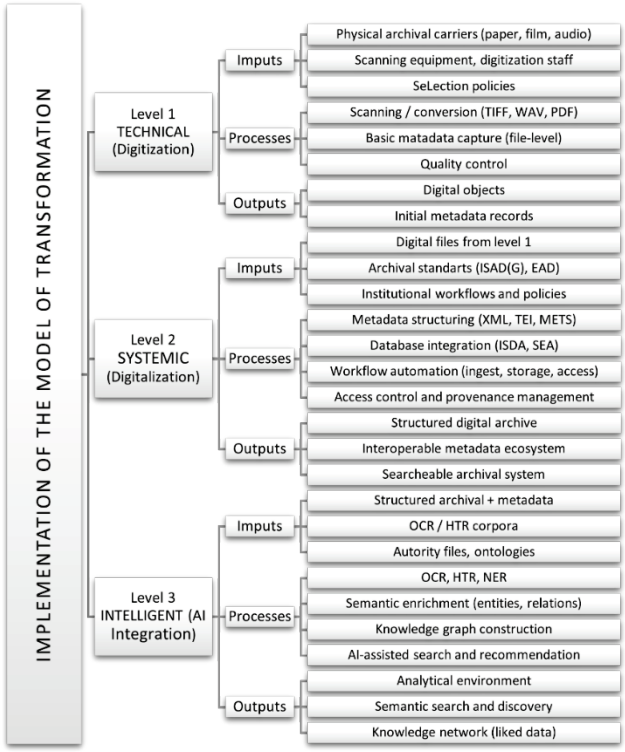

Fig. 1. Workflow of the implementation of the three-level model of transformation.

At least four milestones are possible to be pointed out. At the stage of foundation, the creation of the first collection of standardized digital archival objects is of high importance. This also involves an institutionally approved metadata schema based on the EAD standard. At the stage of system integration, the national archives must be transformed into a unified database. This will allow cross-archive search functionality. At the stage of AI implementation, a full-text search across collections will be possible due to the integration of NER for Bulgarian, of knowledge graphs (people, places, organizations, events) and of ontologies. Finally, an entire ecosystem will take place as a result of the creation of socio-technical archival infrastructure.

The major advantage of the stages of the proposed implementation is its practical aspect. This will allow a pilot realization to be started. However, the implementation of the model needs human resources, which are not limited to only archivists, but also involves other specialists such as IT ones, legal experts etc.

5 Conclusions

The outcome of this research demonstrates that the transformation of archival institutions cannot be limited to the technical process of digitization only. Digitalization on the other hand represents a change in the systems of memory institutions. That change manifests itself in the integration of technological infrastructures, metadata standards, preservation strategies, and user-oriented access. The Bulgarian experience in the field could boast platforms such as the e-Archiving System (SEA) and the Information System of State Archives (ISDA) but despite the improvements made, the digitization levels show limited progress and significant challenges, among which insufficient resources, infrastructure constraints, uneven metadata quality and legal barriers can be seen. A conceptual model to deal with these challenges has been proposed. Three interrelated levels of the model contribute. These are in particular: digitization, digitalization and AI integration. Ideally, together they contribute to a sustainable socio-technical infrastructure. It has been concluded that the future perspectives of the Bulgarian Archival State Agency are in the integrated digital systems which guarantee long-term preservation, interoperability, and meaningful access to archival knowledge.

Declaration on Generative AI

AI was only used for proofreading in order to avoid repeated phrases, grammar and spelling mistakes.

References

AlKendi, W., Gechter, F., Heyberger, L., & Guyeux, C. (2024). Advancements and Challenges in Handwritten Text Recognition: A Comprehensive Survey. *Journal of Imaging*, *10*(1), Article 18. https://doi.org/10.3390/jimaging10010018

Baca, M. (Ed.). (2016). *Introduction to metadata* (3rd ed.). Getty Publications. http://www.getty.edu/publications/intrometadata/

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.

Brennen, J. S., & Kreiss, D. (2016). Digitalization. In K. B. Jensen, E. W. Rothenbuhler, J. D. Pooley, & R. T. Craig (Eds.), *The international encyclopedia of communication theory and philosophy*. Wiley. https://doi.org/10.1002/9781118766804.wbiect111

Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2022). Archives and AI: An overview of current debates and future perspectives. *Journal on Computing and Cultural Heritage*, *15*(1), Article 4. https://doi.org/10.1145/3479010

Conway, P. (2010). Preservation in the age of Google: Digitization, digital preservation, and dilemmas. *The Library Quarterly: Information, Community, Policy*, *80*(1), 61–79. https://doi.org/10.1086/648463

Darzhavna agentsia "Arhivi". (n.d.). *Home page*. Retrieved February 28, 2026, from https://www.archives.government.bg

Friedewald, M., Székely, I., & Karaboga, M. (2024). Preserving the Past, Enabling the Future: Assessing the European Policy on Access to Archives in the Digital Age. *Preservation, Digital Technology & Culture*, *53*(2), 61–71. https://doi.org/10.1515/pdtc-2024-0003

Gilliland, A. J. (2014). *Conceptualizing 21st-century archives*. Society of American Archivists.

Informatsionna sistema na darzhavnite arhivi (ISDA). (n.d.). *Fund search portal*. Retrieved February 5, 2026, from https://isda.archives.government.bg:84/FundSearch.aspx

International Council on Archives & Australasian Digital Records Initiative. (2008). *Principles and functional requirements for records in electronic office environments – Module 1: Overview and statement of principles*. Retrieved May 14, 2025, from https://www.naa.gov.au/sites/default/files/2019-09/m1-ica-overview-principle-and-functional-requirements_tcm16-95418.pdf

Jaillant, L., & Caputo, A. (2022). Unlocking digital archives: Cross-disciplinary perspectives on AI and born-digital data. *AI & Society*, *37*, 823–835. https://doi.org/10.1007/s00146-021-01367-x

Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR). *IEEE Access*, *8*, 142642–142668. https://doi.org/10.1109/ACCESS.2020.3012542

Ogilvie, B. (2016). Scientific archives in the age of digitization. *Isis*, *107*(1), 64–72. https://doi.org/10.1086/686075

Palmer, C. L., Teffeau, L. C., & Newton, M. P. (2008). Strategies for institutional repository development. *Library Trends*, *57*(2), 142–167. https://doi.org/10.1353/lib.0.0033

Perry, S. R. (2014). Digitization and digital preservation: A review of the literature. *School of Information Student Research Journal*, *4*(1), Article 4. https://doi.org/10.31979/2575-2499.040104

Petkova, S., & Atanasov, H. (2023). *IKT, digitalizatsiyata i dostapat do arhivnite dokumenti v Bulgaria (misli, praktiki, nadezhdi)* [ICT, digitization and access to archival documents in Bulgaria (thoughts, practices, expectations)]. Akademichno izdatelstvo "Za bukvite - O pismeneh".

Proekt 43 "Digitalizatsia na kolektsii na muzei, biblioteki i arhivi" (investitsia S11.I7 ot Plana za vazstanovyavane i ustoychivost na Republika Bulgaria). (n.d.). *Project info portal*. Retrieved February 11, 2026, from https://mc.government.bg/page.php?p=46&s=48&sp=1027&t=0&z=0

Smiraglia, R. (2014). *The elements of knowledge organization*. Springer.

Tsvetkova, N., & Petrova, E. (2015). Izgrazhdane na digitalnia arhiv na Darzhavna agentsia "Arhivi" na Republika Bulgaria kato tsentralizpована baza ot digitalni obekti [Creating the digital archive of the State Archives Agency of the Republic of Bulgaria as a centralized repository of digital objects]. *Godishnik na Natsionalnia voenen universitet "Vasil Levski"*, 13–20.

Received: March 12, 2026 /
Reviewed: May 05, 2026 / Finally Accepted: May 20, 2026