Enhancing Data Space Trust and Sustainability through Long-term Archiving Practices

White Paper by the eArchiving Initiative

July, 2025

Executive Summary

European data spaces are emerging as a foundational element of the EU's data strategy, designed to facilitate secure and trustworthy data exchange.

This white paper argues that integrating robust, long-term archiving practices into data spaces is not merely an optional addition, but a crucial, foundational service for enhancing trust and ensuring their sustainability.

By leveraging the established expertise, tools, and specifications of the EU eArchiving Initiative, data spaces can address the critical need for data preservation, authenticity, and long-term accessibility, thereby fostering a more resilient and trustworthy data ecosystem for all stakeholders. This paper outlines the challenges, identifies key synergies, and provides concrete recommendations for collaboration to achieve this vision.

1 Introduction

This white paper examines the relationships between data spaces and mature efforts in digital archiving. It makes concrete recommendations on how digital archiving methods and practices can benefit data spaces and provides the context for further discussion.

The European data strategy aims to make the EU the leading data-driven society. Creating a single market for data will allow it to flow freely within the EU and across sectors for the benefit of citizens, businesses, researchers, and public administrations.[1]

Within the overall strategy, the EU has fostered the development of digital archiving methods through the eArchiving Initiative. This work provides specifications, software, and training to ensure long-term preservation and reuse of digital information and data[2].

[1] A European Strategy for Data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data.
[2] The eArchiving Initiative. https://digital-strategy.ec.europa.eu/en/activities/earchiving.

More recently, the EU has established the data spaces programme. This substantial multi-sector activity aims to create a single market for data, fostering cross-sectoral data sharing within secure and trustworthy environments.[3]

Data spaces provide a next-generation approach to enabling services and businesses that use, provide, and rely on data. They go beyond prior approaches such as data warehouses or data lakes that consolidate data from multiple systems; and data federations that provide real-time interoperability between data-rich systems using high-performance APIs. Data spaces address the full range of requirements to implement complex use cases, including governance and legal frameworks in addition to more traditional technological components. This means that a data space can support cooperation, provide motivation and incentives, and enable trust among all the parties involved.

A data space is generally oriented towards a specific sector and set of foundational use cases. During their initial development, they typically on implementing these use cases integrating data and services from an initial set of participants. Participants in a data space may include data and service providers, intermediaries that deliver layered or bundled services, both public and private enterprises, and individuals. All of these may both provide and consume data and services.

Over time, however, a focus on immediate exchange can lead to data loss and undermine business continuity. For example, if a participant exits the data space, any services that depend on it may be degraded or compromised.

Data spaces are under active development across many sectors. We see them emerge in domains including skills, tourism, media, cultural heritage, and the European green deal. Taken together, this is a highly ambitious programme of work. Individual data spaces may involve hundreds or thousands of participating entities and face a wide range of challenges.

The eArchiving Initiative has emerged to address the needs of organisations that must provide access to data and digital information over the long term. Participants in the initiative and the associated communities have sometimes decades of experience and practice developed to address the challenges that emerge when ensuring trust in data over the long term.

For example, the Publications Office of the EU has used eArchiving standards, practices, and tools to realise cost savings, improve scalability, enhance flexibility and interoperability with other organisations and systems, and share knowledge with a community of experts.[4]

---

[3] The Data Spaces programme. https://digital-strategy.ec.europa.eu/en/policies/data-spaces. Second report on data spaces. https://digital-strategy.ec.europa.eu/en/library/second-staff-working-document-data-spaces.

[4] Nine years of eArchiving at the Publications Office of the EU: A reason to celebrate, Nov 2024.

To ground our discussion and illustrate some of the challenges, we consider two data spaces that that address problems that many of us have experience with.

The **Tourism Data Space** aims to accelerate the digital transformation of the European tourism sector.[5] It will help tourists plan trips and make the most of their experience; tour operators and related businesses to design more attractive products, identify emerging trends, improve their services, and make better strategic decisions. Regions and sites can use the data space to improve visitor management, routing, and sustainability.

The tourism industry is highly fragmented with over 95% of businesses classified as SMEs. Participants rely on local and regional data, which must be integrated for effective planning and other services. The sector experiences both short-term and long-term shifts, requiring alignment with public sector data like weather, traffic, and transport. It must adapt to slow changes (e.g., climate), planned events (e.g., a football championship), and unplanned events (e.g., severe weather or viral video drawing visitors to a trail-side location).

From this white paper's perspective, key features include:

- Participant churn: The high number of small providers leads to considerable churn. Without a fundamental approach to ensuring long-term data access, services within the data space risk becoming brittle and unreliable as new participants join and others exit.
- Data heterogeneity: The wide range of information types demands a robust approach to resolving differences in data meaning. Without this, services may face complex data mapping efforts or become difficult to interpret.
- Public sector sources: Tourism benefits from diverse public sector and cultural data. The organisations managing this information, however, often lack the capacity to provide scalable, high-availability data services and may deliver data in formats that do not align with common commercial usage.

These features highlight the inherent risk of data loss and the need for robust solutions to prevent cascading failures that may compromise service availability and trust.

The **Skills Data Space** aims to empower individuals by allowing them to access and demonstrate their skills and qualifications throughout their lives, while

---

https://digital-strategy.ec.europa.eu/en/library/nine-years-earchiving-publications-office-eu-reason-celebrate

[5] For example, see the DeployTour Tourism Data Space. https://deploytour.eu/

ensuring these credentials are understood and recognised across Europe's diverse regions and industries.[6]

From this white paper's perspective, the key feature of this data space is that skills data maintains a lifelong relevance. It reflects an individual's educational and professional journey. A data space must address the differences in credentials across regions and borders, across time as standards evolve, and throughout the lifespan of the organisations that provide and manage them.

Addressing these challenges through strategic data preservation is vital for the Skills Data Space to fulfil its promise of lifelong accessibility and trust.

As these examples highlight, for data spaces to truly thrive and deliver on their potential, they must support the entire data lifecycle, including long-term preservation and accessibility.

Building resilient data spaces requires a holistic approach. Long-term preservation and access are not optional; they are critical to maintaining trust, ensuring sustainability, and enabling future innovation. Without mechanisms to guarantee the enduring availability and authenticity of data, the integrity of a data space is at risk. The eArchiving Initiative provides critical components and expertise to meet these needs. Long-term archiving capabilities are a strategic enabler for robust and future-proof data space ecosystems.

2 Enhancing trust in data spaces

Long-lived and sustainable data ecosystems fundamentally depend on the ability to preserve and access data over extended periods. The eArchiving Initiative delivers significant value by providing a suite of specifications, tools, and best practices that can be integrated into data spaces to build inherent trustworthiness. Emphasising the full data lifecycle, particularly long-term data availability, is crucial establishing and maintaining trust in these ecosystems.

Archiving by Design, also known as Sustainable Access, is a forward-looking approach that embeds long-term accessibility and preservation into the core architecture of digital information systems.[7] By aligning archiving objectives with system design, it ensures that valuable data remains usable, trustworthy, and resilient throughout its lifecycle. This approach is especially well suited to data spaces that are keen on increasing trust in their long-term viability.

The synergies between European Data Spaces and eArchiving are substantial and directly contribute to building trust and sustainability.

---

[6] Skills Data Space. https://skillsdataspace-blueprint.eu/

[7] Archiving by Design White Paper. https://commission.europa.eu/document/download/e3cf4d38-ee41-42f9-8994-f6589ffad458_en?filename=Whitepaper%20AbD_en.pdf

- Ensuring long-term trustworthiness and authenticity: eArchiving's expertise in digital preservation ensures that data remains reliable and verifiable over time - an essential requirement, for any data space where data may be used for decades. This involves applying eArchiving specifications like Submission Information Packages (SIPs) and Archival Information Packages (AIPs) to guarantee data integrity and authenticity even if the original data provider ceases operations.
- Providing trusted archival services: Integrating certified eArchiving service providers into data spaces offers the necessary infrastructure and expertise for long-term preservation, ensuring data is managed professionally beyond the operational lifespan of initial systems.
- Establishing trust in historical and orphaned data: eArchiving delivers solutions for ingesting and providing access to data from defunct organizations or legacy systems, thereby preserving valuable historical context and minimising data loss within data spaces.
- Promoting interoperability: eArchiving specifications and especially the Common Specifications for Information Packages (CSIP) provide a framework for packaging and exchanging digital information. They serve to facilitate interoperability between data spaces and archival systems. They have been designed to be sector independent.
- Developing Content Information Type Specifications (CITS) for data spaces: By extending existing eArchiving models and metadata, CITS can provide high-quality descriptions of data that support both immediate reuse and long-term preservation, enhancing trust and simplifying compliance for data space participants. Designed around a common core, CITS incorporate detailed analysis specific to sector-relevant data. CITS also support semantic interoperability and domain specific metadata.
- Leveraging existing standards and best practices: Data spaces can adopt eArchiving's mature standards (e.g., METS, PREMIS) and best practices for digital preservation, ensuring that data is managed in a way that is compliant with international norms. This can reduce the risk of data loss or corruption.
- Providing proven tools and software: The eArchiving Initiative offers a suite of open-source tools (e.g., for validation, packaging) that data spaces can directly leverage to streamline data management and prepare data for long-term archiving, reducing development costs and ensuring reliability.
- Providing both modelling expertise and ready-made architecture models through the eArchiving Reference Architecture. The Reference Architecture is mature tool that supports European interoperability. It is based upon leading-edge standards, principles, and approaches (e.g., TOGAF®, ArchiMate®, the FAIR principles, archiving by design, and

capability maturity models). The reference architecture also includes business process scenarios and real-world application examples.

- Offering training and expertise: the eArchiving community provides extensive training materials and expert knowledge in digital preservation, which are invaluable for data space participants in understanding and implementing long-term data management strategies.
- Supporting regulatory compliance and legal certainty: By enhancing data preservation and authenticity, eArchiving helps data spaces meet regulatory requirements for data retention, auditability, and legal certainty. These are crucial factors in finance, healthcare, public administration, and other sectors.
- Establishing best practices for data governance: The mature governance frameworks and best practices developed by eArchiving for managing digital assets over time can inform and strengthen the governance models within data spaces, ensuring responsible and sustainable data stewardship.
- Facilitating long-term accessibility and reuse: Beyond preservation, eArchiving ensures archived data remains accessible and usable over time, enabling future innovation, research, and public services that rely on historical data.

Data spaces can move beyond transient data exchange to establish truly long-lived, sustainable ecosystems where data is a durable asset, fostering continuous innovation and building enduring trust among all participants.

3 Opportunities for collaboration

The synergy between European Data Spaces and the eArchiving Initiative presents a unique opportunity to build more robust, trustworthy, and sustainable data ecosystems. This collaboration can manifest across several key areas, leveraging eArchiving's established frameworks and tools to enhance the foundational services of data spaces.

The Data Space Support Centre (DSSC) **Blueprint** provides a foundational framework for designing and implementing data spaces, covering everything from governance to technical interoperability. Integrating eArchiving considerations into this blueprint is essential to ensure that data lifecycle management, particularly long-term preservation, is embedded as a core feature rather than treated as an afterthought.[8]

Key engagement points for eArchiving within the DSSC Blueprint include:

[8] Data Space Support Centre. https://dssc.eu/blueprint/

- Governance model: eArchiving's expertise in digital preservation governance can support the establishment of clear rules, roles, and processes for data retention, archiving, and access within a data space. This ensures legal compliance and addresses issues such as data ownership and accountability and stewardship.
- Contractual frameworks: Incorporating provisions related to long-term data preservation, archival service provision, and data exit strategies into contractual agreements between data space participants will provide legal certainty and protects critical data assets over time.
- Service model: Recognising archiving as a core service on par with identity management or secure data exchange, ensures that mechanisms for long-term data preservation are readily available and integrated into the operational fabric of the data space. This includes specifying the roles of certified archival service providers.
- Metadata specifications: eArchiving's rich metadata standards (e.g., PREMIS for preservation metadata; CITS for sector specific data) can be adopted and adapted within data spaces to provide comprehensive descriptions of data objects. These include provenance, authenticity, and preservation history. The result is to enhance data discoverability, usability, and long-term trustworthiness.

By embedding these considerations within the DSSC Blueprint, data spaces can be proactively designed for longevity and trust from their inception, rather than retrofitted archival solutions later.

The DSSC **Toolbox** provides practical tools and software components to support the implementation of data spaces. The eArchiving Initiative offers a mature suite of tools and capabilities that can directly augment this toolbox, providing essential functionalities for data preservation and management.

Key eArchiving capabilities and tools that can be integrated into the DSSC Toolbox include:

- Packaging tools: Software for creating standardised information packages (SIPs), Archival Information Packages (AIPs), and Dissemination Information Packages (DIPs) streamlines the process of preparing data for transfer to archival systems, ensuring consistency and interoperability.
- Package validators: Tools for validating the structure and content of SIPs and AIPs help ensure that data conforms to specified standards. This guarantees integrity and facilitates long-term accessibility and reuse.
- Archiving services: The DSSC Toolbox can point to or facilitate access to certified eArchiving service providers, offering data spaces a reliable and secure means to offload their long-term preservation needs to expert organisations.

- Metadata specifications and Tools: In addition to standardised formats practical tools for generating, managing, and exchanging preservation metadata based on eArchiving specifications such as E-ARK Content Information Type Specifications (CITS) can be integrated. This will ensure that all necessary information for future reuse and understanding of the data is captured and maintained.

Integrating these tools and capabilities directly into the DSSC Toolbox reduces technical barriers for data space participants, making it easier to adopt robust long-term archiving practices and enhance the overall trustworthiness of the data ecosystem.

Beyond integrating existing blueprints and tools, both European Data Spaces and the eArchiving Initiative can make **complementary investments** to further strengthen the overall data ecosystem.

Key areas for complementary investment include:

- Developing content information type specifications (CITS) for Data Spaces: While eArchiving already provides valuable CITS, there is a growing need to develop more domain-specific CITS tailored to the unique data types and requirements of various data spaces. This collaboration would involve experts from both initiatives working together to define high-quality descriptions of data that support both immediate reuse within a data space and long-term preservation in archives. This effort will enhance trust and simplify compliance for data space participants, especially in regulated sectors.
- Integrating archiving into data space use cases: Demonstrator projects and pilot programs that explicitly incorporate long-term archiving into their design can showcase the value and feasibility of this integration. These investments can illustrate the benefits of persistent data for analytics, AI training, regulatory compliance, and historical research within specific data space contexts (e.g., health, mobility, energy).
- Research into automated preservation workflows: Continued research and development are needed to automate aspects of data preservation within dynamic data space environments. This includes exploring AI-driven metadata extraction, automated data curation, and robust provenance management to support scalable, sustainable preservation practices.
- Establishing cross-sectoral archiving hubs: Investing in the creation of cross-sectoral archiving hubs that serve multiple data spaces can create efficiencies and ensure the consistent application of preservation standards. These hubs could act as trusted intermediaries, providing specialised long-term storage and access services, and digital preservation expertise across domains.

- Building a community of practice: Establishing a joint community of practice between data space developers, data providers, and archiving professionals will facilitate knowledge exchange, disseminate best practice, and support collaborative problem-solving. This includes joint workshops, training programs, and shared platforms for innovation.

By making these complementary investments, European Data Spaces and the eArchiving Initiative can accelerate the evolution of data ecosystems, ensuring they are not only functional for immediate needs but also resilient, trustworthy, and valuable for future generations.

4 Recommendations

To effectively integrate eArchiving into European Data Spaces and realise the benefits of enhanced trust and sustainability, a series of strategic recommendations can be implemented. These recommendations span policy, technical integration, and community building, ensuring a holistic approach to long-term data lifecycle management.

In Section 2, we discussed how eArchiving capabilities can enhance trust in data spaces. In Section 3, we outlined how and where these capabilities fit with the DSSC Blueprint, which provides a framework for thinking about data spaces. This provided a broad range of possible activities and complementary investments. Now we turn to specific recommendations which could be carried forward.

1. Prioritise data longevity in data space design: From the outset, data spaces should embed long-term preservation as a core design principle, rather than an afterthought. This involves explicitly defining data retention policies, identifying critical data assets for long-term archiving, and allocating necessary resources.
2. Integrate eArchiving specifications into data space blueprints: The DSSC Blueprint and other relevant frameworks should formally reference and require the adoption of eArchiving specifications (e.g., CSIP, CITS, and PREMIS). This ensures a standardised approach to information packaging and metadata management as well as facilitating interoperability and long-term access.
3. Leverage eArchiving Tools and Services: Data space developers should be encouraged to use existing open source eArchiving tools for data packaging, validation, and metadata generation. Furthermore, certified eArchiving service providers should be recognised and integrated into data space service models, offering reliable and scalable options for long-term storage and access.
4. Develop sector-specific archiving guidelines and use cases: Collaboration between eArchiving and individual data spaces (e.g., health, mobility,

energy, skills) is crucial to develop tailored guidelines and demonstrator projects. These efforts can showcase the practical application of archiving practices within specific domains, highlighting benefits like regulatory compliance, historical research, and AI training with persistent data.

5. Promote policy and legal alignment: Advocate for policies that clarify responsibilities and incentives for long-term data preservation within data spaces. Ensure legal certainty around data authenticity and accessibility over time, by leveraging eArchiving's experience in this domain.
6. Foster a shared community of practice: Establish regular forums, workshops, and training sessions that bring together data space architects, data providers, legal experts, and archiving professionals. This will facilitate knowledge exchange, build expertise, and identify common challenges and solutions for long-term data stewardship.
7. Invest in research and innovation: Support research into advanced techniques for automated data preservation, including the use of AI for metadata extraction, content analysis, and the development of self-preserving data objects. Explore methods for enhanced provenance and integrity tracking in long-term archives.
8. Establish funding mechanisms for long-term data stewardship: acknowledge that long-term data archiving requires sustained funding. Explore funding models where data space participants contribute to shared archiving infrastructure, shared risk models, or where public funding supports essential preservation services for common good data.

5 Conclusion

The journey towards a robust and trustworthy European data economy hinges not only on facilitating immediate data exchange but also on ensuring the long-term availability and authenticity of digital assets. This white paper has argued that the integration of the EU eArchiving Initiative's expertise, specifications, and tools is a critical enabler for the sustainability and trustworthiness of European Data Spaces.

By embracing long-term archiving as a foundational service, data spaces can overcome the challenges of data loss, ensure business continuity, and build profound trust among end-users, service providers, and the organisations nurturing these ecosystems. The identified synergies - ranging from ensuring authenticity and interoperability to providing proven tools and supporting regulatory compliance - demonstrate the immense potential for a mutually beneficial collaboration.

The blueprint for this integration is clear, as outlined in the opportunities for collaboration around the DSSC Blueprint, Toolbox, and complementary investments. By proactively designing for longevity, leveraging existing best practices, and investing in continuous innovation, European Data Spaces can

become resilient, future-proof infrastructures capable of supporting generations of innovation, research, and public service.

Ultimately, the vision is of a European data ecosystem where data is not transient but a durable, trusted asset. This requires a commitment to the full data lifecycle, with long-term archiving serving as the bedrock upon which truly thriving and sustainable data spaces can be built.

6 Acknowledgements

Adam Farquhar was the primary author. Contributors to this white paper include members of the eArchiving Initiative: Jože Škofljanec, István Alföldi, David Anderson, Janet Anderson, Žiga Koncilja , Sven Schlarb, and Gregor Završnik.

In addition, it benefited from discussions with Kuldar Aas and Sylvia Petrovic-Majer.

Appendix - Concrete changes

Engaging with the eArchiving community and adopting relevant practices could result in the following concrete changes within data spaces:

- Continuity of data access reflected in legal framework.
- Snap-in adaptors for public sector and archival data.
- Service providers that enable access to public sector and archival data.
- Service providers that ensure access continuity.
- Shared metadata specifications (eg CITS) to simplify data access and use.
- Lightweight standards processes to establish and maintain shared metadata specifications.
- Business models, perhaps shared-risk/insurance based, for access and data continuity.
- Increased availability of data space services
- Increased use of archival data services within data spaces.
- Increased value from public sector and archival data.
- Increased trust in and reliability of layered data space services.
- Simplified pathways for data and service providers to enter data spaces.