A Solid Research Infrastructure Leveraging DMA Data Portability

LK Seiling[1,†], Jake Stein[2,†], Simon Mayer[3] and Luka Bekavac[3,*]

[1] *Weizenbaum Institute, Berlin, Germany*
[2] *Department of Computer Science, University of Oxford, Oxford, United Kingdom*
[3] *University of St. Gallen, St. Gallen, Switzerland*

Abstract

Article 6(9) of the EU's Digital Markets Act (DMA) grants users the right to port their data from gatekeeper platforms to authorized third parties. We propose a multi-institution Solid-based research infrastructure in which universities act as such authorized third parties, storing platform data in user-owned Solid pods. Rather than aggregating data into centralized repositories, this model preserves user-level ownership and fine-grained access control, enabling consent-based , cross-institutional research within a decentralized architecture. The proposed infrastructure addresses a key implementation gap in DMA portability: the absence of standardized receiving environments and interoperable schemas. The standardisation of incoming platform data and the associated schema extraction could additionally facilitate broader usage beyond academic fields of application. While gatekeeper vetting procedures, export format heterogeneity and enforcement uncertainties pose practical challenges, a university consortium offers institutional credibility and governance safeguards. Beyond academic research, such an infrastructure may contribute to broader interoperability and portability ecosystems under EU data regulation, creating a convergence point for DMA, DSA, and GDPR-based access mechanisms.

Keywords

Solid Pods, Digital Markets Act, data portability, personal data spaces

1. Vision

The Digital Markets Act (DMA)[1] sets out to ensure the proper functioning, contestability, and fairness of the digital sector internal market in the European Union. To do so, it introduces interoperability and data portability requirements. Specifically, Article 6(9) DMA obliges designated gatekeepers (currently: Alphabet, Amazon, Apple, ByteDance, Booking, Meta, and Microsoft) to provide end users and their authorized third parties with effective, free, continuous, real-time data portability and the tools to effectively exercise it. This sort of direct, consent-driven access would ameliorate many of the existing issues in current data donation research [1] and provide an effective means to promote agency over personal data. The scope of accessible data would additionally go beyond the data portability provisions of the General Data Protection Regulation (GDPR)[2], extending portability beyond user-provided to *user-generated* data (e.g., data on their activities on a platform). Yet, the right remains largely unexercised: users lack awareness and, critically, infrastructure to receive and govern their data as services attempting to make use of that obligation face a fragmented landscape of platform-specific export tools that deposit data in heterogeneous formats [2].

We propose a Solid-based [3] consortial infrastructure that makes use of Article 6(9). Consortium members (we primarily expect universities and research institutions), would then set up projects that involve the transfer of gatekeeper data into individual Solid pods, where users retain ownership and fine-grained access control over their pod [4]. Researchers could then request project-specific access to

Solid Symposium'26: THE 4TH SOLID SYMPOSIUM, April 30 - May 1, 2026, London, UK

[*] Corresponding author.

[†] These authors share first authorship.

lukas.seiling@weizenbaum-institut.de (L. Seiling); jake.stein@cs.ox.ac.uk (J. Stein); simon.mayer@unisg.ch (S. Mayer); luka.bekavac@unisg.ch (L. Bekavac)

[1] See https://eur-lex.europa.eu/eli/reg/2022/1925/oj

[2] See https://eur-lex.europa.eu/eli/reg/2016/679/oj

query the pods in order to answer their research questions.

This initiative pursues five goals:

- **Research Data Sharing**: With user consent and ethics approval, researchers across institutions can access pod data through Solid's native access control mechanisms [5], profiting from a distributed alternative to centralized institutional research data repositories. With user consent, data might even be shared between studies, fostering cross-institutional collaboration.
- **Schema Extraction**: By systematically ingesting gatekeeper exports, the consortium can extract, publish, and version the de-facto data schemas. Today, this is largely undocumented; schema extraction would simplify access, permit interoperability across institutions, and perspectively even allow integration across different data access mechanisms (e.g., from the EU Digital Services Act or from the EU General Data Protection Regulation).
- **No Single Point of Failure**: A distributed data repository logic avoids single points of failure, thereby better serving users' privacy protection requirements.
- **Broadening Solid Use**: Universities actively providing users with pods that are pre-populated with the user's own platform data would circumvent the cold-start problem that currently limits Solid adoption.
- **Promote Individual Data Sovereignty**: Initiating Pods and populating them with data improves users' ability to steward their own data as a result of research interests, enabling them to access, manage, and port their own data across services, thereby reducing vendor lock-in, lowering switching costs, and facilitating the exercise of their personal data rights.

2. Architecture and Legal Basis

The system comprises per-gatekeeper adapters that retrieve raw exports (in a diversity of formats) on behalf of consenting users and insert the raw data into per-user Solid pods on Community Solid Server instances that are operated by the consortium. Schemas that are defined based on these exports are then used to transform these data (e.g., using RML mappings or through bridge ontologies) to produce RDF triples conforming to (to-be-published) shared vocabularies. SHACL shapes could be used to validate transformed data and serve as extracted schema artifacts.

The DMA does not explicitly address vetting procedures, meaning gatekeepers currently have discretion to implement their own proportionate checks on third parties seeking to use data portability tools[3]. Prior to this proposal being implemented, such vetting would hence need to take place, where we believe that a university consortium which puts forth this proposal possesses inherent credibility and would (possibly after a considerable time lag) be fit to pass this vetting. This is backed by findings of the Data Transfer Initiative that has identified the lack of standardized third-party trust frameworks as a key DMA implementation gap[4]. Receiving personal data triggers controller obligations under the GDPR, and we argue that Solid's architecture aligns well with these obligations: Per-resource access control supports data minimization, access grants tied to specific projects enforce purpose limitation, and pod or resource deletion implements the right to erasure. In our proposal, each institution operates under its own data protection and ethics framework. The consortium relies on shared technical standards and we foresee the proposed infrastructure to later integrate data obtained through other mechanisms, such as the EU Digital Services Act (DSA)'s provision 40(12) and the GDPR's data access (Article 15) and data portability (Article 20) provisions, making it a convergence point for multiple EU data access regimes. We invite the Solid community to discuss the feasibility of this proposal and explore pathways toward a pilot implementation.

[3] For example LinkedIn and Meta conduct extensive vetting and proving that third parties meet their (legal, technical, and organisational) requirements.

[4] See https://dtinit.org/blog/2024/04/29/supporting-effective-portability

Declaration on Generative AI. During the preparation of this work, the author(s) used Claude Opus 4.6 in order to draft and structure the manuscript text based on information about the given proposal. The authors made substantial changes to the resulting text and no sentence in this text is the direct output of Opus 4.6. They take full responsibility for the publication's content.

References

[1] V. Hase, J. Ausloos, L. Boeschoten, N. Pfiffner, H. Janssen, T. Araujo, T. Carrière, C. De Vreese, J. Haßler, F. Loecherbach, Z. Kmetty, J. Möller, J. Ohme, E. Schmidbauer, B. Struminskaya, D. Trilling, K. Welbers, M. Haim, Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies?, Internet Policy Review 13 (2024). doi:10.14763/2024.3.1793.

[2] M. Kirkwood, Interoperability and the DMA in action: Developers experiences of data portability API access, 2024. URL: https://mydata.org/wp-content/uploads/2024/11/Interoperability-and-the-DMA-in-Action-Developers-Experiences-of-Data-Portability-API-Access.pdf.

[3] S. Capadisli, T. Berners-Lee, K. Kjernsmo, et al., Solid Protocol, W3C Solid Community Group, 2024. URL: https://solidproject.org/TR/protocol, draft Community Group Report, Version 0.11.0.

[4] C. Esposito, R. Horne, L. Robaldo, B. Buelens, E. Goesaert, Assessing the solid protocol in relation to security and privacy obligations, Information 14 (2023). doi:10.3390/info14070411.

[5] J. Bingham, E. Prud'hommeaux, elf Pavlik, Solid Application Interoperability, Solid Community Group, 2025. URL: https://solid.github.io/data-interoperability-panel/specification/, draft Community Group Report, Version 0.1.