

Enabling Choice

## Decentralised Aggregation in the Data Space for Cultural Heritage: Opportunities, Challenges and Governance Implications

*Final report of the Europeana Aggregators' Forum's Task Force on Decentralised Approaches to Aggregation (version 1.0)*

Document history

Version	Author(s)	Comments
0.1	Kerstin Arnold Pavel Kats Marie-Véronique Leroi	Draft version incl. chapters 2 to 6 plus annexes ahead of final Task Force meeting
0.2	Kerstin Arnold Pavel Kats Marie-Véronique Leroi	Incorporation of comments and refinement of existing chapters Addition of chapters 7 and 8  Shared with Task Force members, ENA's Linked Data Task Force and DSSC for feedback
1.0	Kerstin Arnold Pavel Kats Marie-Véronique Leroi	Incorporation of comments Addition of chapters 1 and 9

## Table of contents

<b>1. Introduction</b>	<b>3</b>
<b>2. Premise of the Task Force</b>	<b>4</b>
<b>3. Preliminaries</b>	<b>7</b>
3.1. Definition of terms	7
3.2. Projects and organisations	12
<b>4. Approach</b>	<b>15</b>
4.1. Task Force meetings	15
4.2. Additional interactive sessions	16
4.3. Interviews	16
<b>5. Use cases and challenges from the current aggregation landscape</b>	<b>18</b>
5.1. Relevance of identified use cases across the Europeana Aggregators' Forum	18
5.2. Technical use cases from an aggregator's perspective	20
5.3. Organisational use cases from an aggregator's perspective	26
5.4. Use cases from the perspective of a cultural heritage institution (CHI)	28
<b>6. Decentralised approaches to aggregation</b>	<b>30</b>
6.1. Platform-centric approaches	31
6.2. Data-centric approaches	35
6.3. Other approaches	38
6.4. Summarising overview of technology offers	40
<b>7. Analysis</b>	<b>43</b>
7.1. General considerations	43
7.2. Blockchain: the appeal of technology-based trust	44
7.3. IDSA vs. Solid: platform-centric vs. data-centric views	44
7.4. To copy or not to copy - or: sharing data vs. visiting data	45
7.5. The applicability of it all: use cases revisited	46
<b>8. Recommendations</b>	<b>50</b>
8.1. Elevate continued exploration	50
8.2. Develop aggregation use cases further	50
8.3. Pilot a decentralised approach	51
8.4. Explore impact on aggregation landscape	52
8.5. Explore impact on data governance and governance of the data space	52
<b>9. Conclusions</b>	<b>54</b>
<b>Annex I - List of interviewees</b>	<b>55</b>
<b>Annex II - Interview questions</b>	<b>57</b>

## 1. Introduction

“Again and again we fall back on the folklore of the principles of good design. [...] Principles such as simplicity and modularity are the stuff of software engineering; decentralization [sic] and tolerance are the life and breath of [the] Internet.”  
Tim Berners-Lee, Principles of Design, 1998, last update on 4 March 2013<sup>1</sup>

This document is the final report of the Task Force on SOLID-based Decentralised Aggregation<sup>2</sup> initiated by the Europeana Aggregators' Forum (EAF) at its autumn meeting in 2022. The Task Force ran from November 2022 to July 2023 and included representatives from eleven accredited aggregators: Archives Portal Europe, Culturaltalia (Italy), Deutsche Digitale Bibliothek (Germany), Europeana Sounds, European Film Gateway, Federacja Bibliotek Cyfrowych (Poland), Institutul Național Al Patrimoniului (Romania), Jewish Heritage Network, Moteur Collections (France), MUSEU, and Photoconsortium. Furthermore, a prospective accredited aggregator (DigiPhil, Hungary), a strategic network on digital heritage (Netwerk Digitaal Erfgoed, Netherlands), Europeana Foundation, and external partners (Εθνικό Μετσόβιο Πολυτεχνείο / National Technical University of Athens, Pangeanic, and Post-Platforms Foundation) were represented.

While the current report will talk about technical and data-related aspects of aggregation as well as technological solutions promoting decentralisation, the authors would like to emphasise that they do not see a potential change towards a decentralised approach as the only or complete answer to challenges encountered by the aggregation landscape as it stands. Similarly, decentralisation should not be understood as an approach only to be applied to technology; there might be other aspects such as governance that would need to be considered as well. Last, it should be noted that the report does not aim at providing a detailed explanation of the technological solutions it discusses, i.e., not everything that might play a role in future will be mentioned. Especially more specifics around using Linked Data in general would be expected to rather be addressed by the Linked Data Task Force of the Europeana Network Association's (ENA) EuropeanaTech Community<sup>3</sup>.

The following chapters will first provide some of the [political and legal background](#) for the development of the common European data space for cultural heritage and, thereby, this task force, before defining some of the [central terminology](#) used in this report and presenting some of the [main actors](#) in the data spaces' realm. This will be followed by a brief introduction of the [approach](#) taken by the Task Force in their work. Afterwards a [set of use cases](#) identified as part of the Task Force activities will be presented, building on the status quo with a rather centralised aggregation workflow, as well as the [technological solutions](#) investigated by the Task Force. Both of these parts will then be brought together in a [summarising analysis](#) before closing with the [recommendations](#) of the Task Force and some [general conclusions](#) of this report.

---

<sup>1</sup> See online at <https://www.w3.org/DesignIssues/Principles.html> [last accessed: 11 July 2023]

<sup>2</sup> See online at <https://pro.europeana.eu/project/solid-based-decentralised-aggregation-task-force> [last accessed: 11 July 2023]

<sup>3</sup> See online at <https://pro.europeana.eu/project/linked-data-task-force> [last accessed: 11 July 2023]

## 2. Premise of the Task Force

The past five years have brought two major initiatives to the European stage in the context of digital and data-related work: in 2021, the digital decade<sup>4</sup> was launched with a strong intention of putting the citizen at the heart of the European Union's (EU) policies and actions for digital; one year earlier already, in 2020, the European Commission (EC) had published its data strategy<sup>5</sup> promoting EU values and rules.

This data strategy translates into the creation of common European data spaces for fourteen industry sectors. A horizontal infrastructure, governance rules and GDPR management will ensure the interoperability between those sectoral data spaces. The intention of creating a data space for cultural heritage alongside sectors such as health or mobility was welcomed with great enthusiasm.

The overall strategy and the building of the data spaces also come along with a harmonised legal framework thanks to several legislative acts which intend to unify the rules governing the data spaces and the data that are shared there. The Data Act<sup>6</sup>, the Interoperable Europe Act<sup>7</sup> and the Data Governance Act<sup>8</sup> are currently under negotiation within EU instances and with the Member States. In combination with each other, these acts will ensure trust and fairness in data sharing and access.

The data spaces are a means to support the data strategy from the EC and aim to be the European answer to the big US-American commercial platforms. The twelve data spaces officially recognised, launched, and deployed by the EC are mentioned in the EC staff working document<sup>9</sup> originally published in February 2022 and currently being updated.

This document also lists some of the key characteristics expected for a data space<sup>10</sup>:

- a secure and privacy-preserving infrastructure to pool, access, share, process and use data;
- a clear and practical structure for access to and use of data in a fair, transparent, proportionate, and non-discriminatory manner and clear and trustworthy data governance mechanisms;
- European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected;
- data holders will have the possibility, in the data space, to grant access to or to share certain personal or non-personal data under their control;

---

<sup>4</sup> See more information about Europe's Digital Decade online at [https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en) [last accessed: 11 July 2023]

<sup>5</sup> See the European Data Strategy online at [https://commission.europa.eu/document/download/cfd0e996-88e4-4525-8cb8-96c53a991270\\_en](https://commission.europa.eu/document/download/cfd0e996-88e4-4525-8cb8-96c53a991270_en) [last accessed: 11 July 2023]

<sup>6</sup> See the proposal from 23 February 2022 online at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0068> [last accessed: 11 July 2023]

<sup>7</sup> See the proposal from 18 November 2022 online at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0720> [last accessed: 11 July 2023]

<sup>8</sup> See the recommendation of 30 May 2022 online at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R0868> [last accessed: 11 July 2023]

<sup>9</sup> See the European Commission's staff working document on data spaces online at <https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces> [last accessed: 11 July 2023]

<sup>10</sup> Ibid, pages 2 and 3

- data that are made available can be reused against compensation, including remuneration, or for free;
- participation of an open number of organisations/individuals.

The EC is also deploying efforts to build a cloud-to-edge infrastructure that will provide technical support for the data spaces. The building of data spaces is funded under the Digital Europe Programme<sup>11</sup>, which also finances the Data Spaces Support Centre (DSSC).

The common European data space on cultural heritage (DS4CH) is one of the twelve data spaces officially recognised by the EC. As such the building of the DS4CH is in line with the overall digital strategy of the EC, but also framed by the recommendation 2021/1970 on a common European data space for cultural heritage published on 10 November 2021<sup>12</sup>. This recommendation addresses guidelines and targets for the Member States to ensure their contribution to the data space. And aggregators and CHIs are expected to share data in the data space based on this new political and strategic context.

While most of the data spaces start from scratch, the DS4CH is expected to build on and extend what the Europeana Initiative has developed and established throughout the past 15 years and was therefore considered to be one of the most advanced data spaces when it entered its deployment phase in September 2022 via a tender won by a consortium of 19 partners in total and led by the Europeana Foundation. In this context, the established infrastructure, a community of 40 accredited aggregators, a network of more than 4,000 professionals and 15 years of experience is reviewed accordingly and - where appropriate and useful - will be reused to put the Europeana Initiative at the core of the DS4CH.

The Europeana Foundation, the EAF and the ENA are the three pillars of the Europeana Initiative which work together with a community of cultural heritage professionals and (re)users to access and share data on the Europeana.eu website and via other aggregators and access points. Regarding the continuous development of the Europeana.eu website and the necessary steps to prepare and share data therein, the Europeana Initiative has - over the last 15 years - succeeded in overcoming the challenges of bringing digital cultural heritage data from all over Europe to one single access point. This has been achieved on the one hand thanks to a complex infrastructure that connects the central services provided by Europeana Foundation with the platforms and services developed and maintained by other aggregators on national and domain levels. More importantly, however, this was achieved thanks to a central data model, the Europeana Data Model (EDM), and common frameworks such as the Europeana Publishing Framework (EPF) and the Europeana Licensing Framework (ELF). The network of 40 accredited aggregators plays a key role as intermediary between cultural heritage institutions (CHIs) and Europeana by implementing EDM - or transforming data into EDM - and promoting those frameworks.

While the principle of not reinventing the wheel and building on the main achievements of the Europeana Initiative is a sustainable and rational approach, moving from a digital service infrastructure (DSI) to a data space brings up some new challenges and opportunities that not only urge the Europeana Initiative to review the current modus operandi, but also ask for

---

<sup>11</sup> See more about the Digital Europe Programme online at [https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/digital-europe-programme\\_en](https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/digital-europe-programme_en) [last accessed: 11 July 2023]

<sup>12</sup> See online at <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32021H1970> [last accessed: 11 July 2023]

some significant evolutions. Although technologies and processes currently in place have proven right for a website and a DSI, their adequacy for addressing the requirements of the field in the ever-evolving digital ecosystem has been questioned during recent years already. The deployment of the data space offers the opportunity to challenge, evaluate and review the current processes and models. This iterative review cycle is necessary to trigger changes in the short, medium, and longer term and in defining the most suitable way forward.

At the same time, information technology has come a long way since the launch of the very first databases. This is also very true in the cultural heritage sector where from some basic metadata on a cultural object one can reach the digital representation(s) of the actual object, whether that might be in 2D or 3D format(s). The continuous growth of data and the evolution of standards and technologies raise some new requirements for how data can be made available and accessible efficiently and effectively while considering needs and use cases from aggregators and CHIs but also global challenges such as digital sobriety.

Contributing to the data space: what does this exactly mean? Can the workflows, processes, and organisation of aggregation in place for the DSI still be considered as relevant in the context of the data space? Can the evolution from the DSI to the DS4CH foster further collaboration between aggregators and CHIs and further contribution of new and higher quality as well as other data? Can a more decentralised approach for aggregation help in achieving this? Would we still call the process aggregation? What would this mean for the role of the aggregators in a decentralised scenario? What would such an approach imply in terms of governance - of the data, but also of the data space? How far can a decentralised approach for data sharing foster interoperability with other sectoral data spaces and upcoming initiatives such as the European Collaborative Cloud for Cultural Heritage (ECCCH)?

The EAF's SOLID-based Decentralised Aggregation Task Force<sup>13</sup> was launched to tackle these as well as related questions in November 2022. It is thereby complementary to the Linked Data Task Force<sup>14</sup> organised within the ENA's EuropeanaTech Community. Both task forces aim to explore the opportunities and challenges of heading towards a decentralised approach for aggregation from different perspectives: the Linked Data Task Force explores possible models for aggregation in a more operational way using Linked Data standards and technologies, while this Task Force explores initiatives and technologies that might be relevant in the context of a data space in a more prospective way and investigates how these approaches could fit the requirements and constraints of the cultural heritage sector.

Both task forces will contribute to the work conducted in the context of the implementation plan of the data space deployment, especially to activities and tasks that are dedicated to investigating innovative scenarios for aggregation and their possible impact on data governance in the data space.

---

<sup>13</sup> See online at <https://pro.europeana.eu/project/solid-based-decentralised-aggregation-task-force> [last accessed: 11 July 2023]

<sup>14</sup> See online at <https://pro.europeana.eu/project/linked-data-task-force> [last accessed: 11 July 2023]

## 3. Preliminaries

### 3.1. Definition of terms

The terms defined here have proven central to the Task Force's investigation and activities. This does not mean that the following list is in any way conclusive and exhaustive in as far as terminology related to data spaces go. It should also be noted that the definitions below build on those used by some of the key players in the field with whom the Task Force interacted during their knowledge gathering. There might be other definitions used by other actors, or definitions might evolve as the deployment of the data spaces and their active usage move along. Definitions given in this chapter differing from the definitions of the same terms as used in the context of the Europeana Initiative for the past 15 years are highlighted accordingly.

#### 3.1.1. Aggregation

The dictionary definition of aggregation is “the act of putting together different items, amounts, etc. into a single group or total [as well as] the group that is formed”<sup>15</sup>. In the context of this report, data aggregation is the combination of various data from various sources and from different areas. The W3C glossary<sup>16</sup>, for example, defines aggregation simply as “the act of combining materials in various ways”.

In the Europeana ecosystem, aggregation is the process operated by an aggregator to gather data from various CHIs and publish them in their own context and/or on Europeana. The Europeana glossary<sup>17</sup> defines an aggregator as follows:

*“Aggregator: an organisation working with cultural institutions and collectors to gather authentic, trustworthy, and robust data. They make this content available to a broader audience via their own services, Europeana and other infrastructures, for example, for education and research.”*

This definition of aggregation as gathering data materialises practically - at the moment - in data being copied from an original source to the aggregator to then be made available in the context of data from other sources. It also includes the notion of data flowing in a specific direction (from the CHI to the aggregator). It should be noted that, in the Europeana ecosystem, currently only metadata is copied, simply including links to the digital objects or content, which remain within the system of the data provider (see the definition of these three terms in [chapter 3.1.4](#), below).

The Europeana website ([Europeana.eu](https://europeana.eu)) could hence be considered an aggregator of aggregators, like other examples especially of domain aggregators such as Archives Portal Europe ([archivesportaleurope.net](https://archivesportaleurope.net)), but also national aggregators in case they aggregate from regional levels or combine several domains.

In this context, aggregation is not strictly limited to the data provision process but also covers the collaboration between aggregators and CHIs to collect, process and publish data

<sup>15</sup> Oxford Advanced Learner's Dictionary, online at <https://www.oxfordlearnersdictionaries.com/definition/english/aggregation> [last accessed: 11 July 2023]

<sup>16</sup> Online at <https://www.w3.org/2003/glossary/keyword/All/aggregation.html?keywords=aggregation> [last accessed: 11 July 2023]

<sup>17</sup> Online at <https://pro.europeana.eu/page/glossary> [last accessed: 11 July 2023]

according to a shared and standardised framework. An aggregator therefore is not only a technical infrastructure but also an entity supporting the participation of CHIs in the Europeana Initiative. Aggregators may have their own publishing frameworks and comply with the EPF and the ELF mainly to publish data on Europeana.eu.

Aggregation can furthermore be understood more generally as data sharing as defined in the DSSC glossary<sup>18</sup>. Following on from this, an aggregator could possibly be considered a data space enabler - at least regarding parts of the definition - as the aggregator acts as a:

*“[...] data space member that provides a (technical or non-technical) service enabling data transactions for the transaction participants while not directly participating in that transaction itself. Examples of enabling services include identity provisioning, vocabulary provisioning, interconnecting, clearing, etc.”*

Considering further parallels between the roles defined by the DSSC glossary and the stakeholders in the Europeana Initiative context, a CHI and an aggregator can both be considered data space members:

*“Data Space Member: A data ecosystem party that has committed to the governance framework of data space and may have one or more roles in it.”*

Additionally, the CHI is a data space rights holder as it is the entity who will state the rights associated with the data that it will share to the data space. According to the DSSC glossary, the data space rights holder is defined as follow:

*“Data rights holder (role): A transaction participant that has the legal right to use, grant access to or share certain data.”*

### 3.1.2. (De)Centralisation

Decentralisation is a concept used in opposition to centralisation. These notions can cover technical aspects (infrastructure, data storage and flow, etc.) and organisational aspects related to governance and networks and any level in between. Centralisation implies a concentration of power, authority, and duties in one single entity, which can mean an agent (i.e., a person, organisation, initiative, etc.) or an asset (i.e., a database, an algorithm, another piece of software) whereas decentralisation implies the distribution of power and responsibilities among the members and stakeholders involved in a specific environment. Furthermore, centralised approaches tend to often be one-directional, while decentralised approaches allow for the possibility of multi-directional relationships.

In terms of decentralised approaches, two main strands are often discussed: a federated systems design where components are developed in a way that independent but interconnected and interoperable instances can be used throughout the system while still providing a cohesive service to the users of the system as a whole; a distributed systems design, or distributed consensus technology (e.g., blockchain), where components and their functions are distributed to single members of the system, but are run based on consensus-agreed protocols and processes (e.g., cryptographic techniques).<sup>19</sup>

---

<sup>18</sup> Online at <https://dssc.eu/wp-content/uploads/2023/03/DSSC-Data-Spaces-Glossary-v1.0.pdf> [last accessed: 11 July 2023]

<sup>19</sup> For a more detailed overview see e.g., Mark Nottingham / Network Working Group: Centralization, Decentralization, and Internet Standards, online at



Tim Berners-Lee describes decentralisation as a “principle of the design of distributed systems, including societies. It points out that any single common point which is involved in any operation trends to limit the way the system scales and produces a single point of complete failure”<sup>20</sup>. This statement implies that centralisation and forcing a set of rules on participants may over time prevent an initiative to grow and scale up.

According to the International Data Space Association (IDSA), the notions of centralisation or decentralisation mainly apply to the authorities and governance aspects while the concepts of federated or distributed mainly apply to the technical organisation of data and data flow. Thereby, in an IDSA context, a data space with a distributed architecture might not necessarily go along with a decentralised organisation.

### 3.1.3. Interoperability

The Joinup<sup>21</sup> platform of the European Commission defines interoperability as follows:

*“Interoperability is the ability of organisations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organisations, through the business processes they support, by means of the exchange of data between their ICT systems.”<sup>22</sup>*

The Joinup initiative intends to set rules and frameworks for an interoperable Europe and has set up the Europe Interoperability Framework (EIF)<sup>23</sup> in this regard which defines basic interoperability guidelines in the form of common principles, models, and recommendations. This framework aims to support public administrations and institutions in improving interoperability and leading to an emerging network of sovereign and interconnected organisations that will bring more trust and increase possibilities for innovation.

The EIF defines four layers of interoperability: technical interoperability, legal interoperability, semantic interoperability, and organisational interoperability, which will all have an impact on the DS4CH and the common European data spaces in general, though the focus of this report - following its scope and mandate is likely going to be on the technical and semantic aspects, while touching on organisational and legal aspects along the way.

Additionally, ethical interoperability is to be considered in a world where data creation and flow has significantly increased and where Artificial Intelligence (AI) is more and more present. This layer relies on ethical principles that must be clearly stated and on shared codes of conducts.

### 3.1.4. Data Space

The Europeana glossary<sup>24</sup> defines a data space as:

---

<https://mnot.github.io/avoiding-internet-centralization/draft-nottingham-avoiding-internet-centralization.html#name-decentralization-strategies> [last accessed: 11 July 2023]

<sup>20</sup> See Tim Berners-Lee: Principles of Design, online at <https://www.w3.org/DesignIssues/Principles.html> [last accessed: 11 July 2023]

<sup>21</sup> See more online at <https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/1-introduction#1.1> [last accessed: 11 July 2023]

<sup>22</sup> It should be noted that, in this context, the notion of organisations refers to public administrations or EU institutions.

<sup>23</sup> [https://ec.europa.eu/isa2/eif\\_en/](https://ec.europa.eu/isa2/eif_en/) [last accessed: 11 July 2023]

<sup>24</sup> Online at <https://pro.europeana.eu/page/glossary> [last accessed: 11 July 2023]

*"[...] decentralised and standard-based infrastructure to enable trustworthy data sharing between the data space participants. Data spaces may be purpose- or sector-specific, or cross-sectoral. Common European data spaces are a subset of data spaces within the scope of EU policies."*

This definition is in line with the one provided by the DSSC Glossary<sup>25</sup> describing a data space as:

*"an infrastructure that enables data transactions between different data ecosystem parties based on the governance framework of that data space. Data space should be generic enough to support the implementation of multiple use cases."*

Professor Boris Otto, Director of the Institute for Software and Systems Engineering (ISST) and the Deputy Chair of IDSA's Executive Board, defines a data space as "a form of collaboration on data"<sup>26</sup> allowing practitioners of a domain, data space members and data space receivers, to collaborate on their data. He states that:

*"Data spaces and their underlying software infrastructures must support trust, interoperability, and portability of data and data sovereignty and must be nondiscriminatory [sic]. Thus, data spaces can be understood as intermediaries and data sharing service providers to which the EU Data Governance Act applies which is currently under review."<sup>27</sup>*

Otto follows up to say that a "data space is a distributed data integration concept. Thus, there is no central data store or data vault into which data providers deliver their data and from where it can be accessed and retrieved by data consumers."<sup>28</sup>

The key features that design such a data space are interoperability, data sovereignty and trust and security.<sup>29</sup>

In addition to data providers and data consumers, Otto also established the role of a federator. The Federator shown in figure 1 below might be the instance in charge of the governance of the data space initiative or, applying the concept to the Europeana Initiative context, might also be an aggregator. The federator only deals with metadata while the data provider and data consumer manage both data and metadata as part of their transactions.

---

<sup>25</sup> Online at <https://dssc.eu/wp-content/uploads/2023/03/DSSC-Data-Spaces-Glossary-v1.0.pdf> [last accessed: 11 July 2023]

<sup>26</sup> Boris Otto: The Evolution of Data Spaces. In: Boris Otto, Michael ten Hompel, Stefan Wrobel: Designing Data Spaces. The Ecosystem Approach to Competitive Advantage, 2022. See online at [https://link.springer.com/chapter/10.1007/978-3-030-93975-5\\_1](https://link.springer.com/chapter/10.1007/978-3-030-93975-5_1) [last accessed: 11 July 2023]

<sup>27</sup> Ibid.

<sup>28</sup> Ibid.

<sup>29</sup> For comparison see the EC Staff working document which also mentions some key characteristics of data spaces that are listed on [section 2](#) of this report.

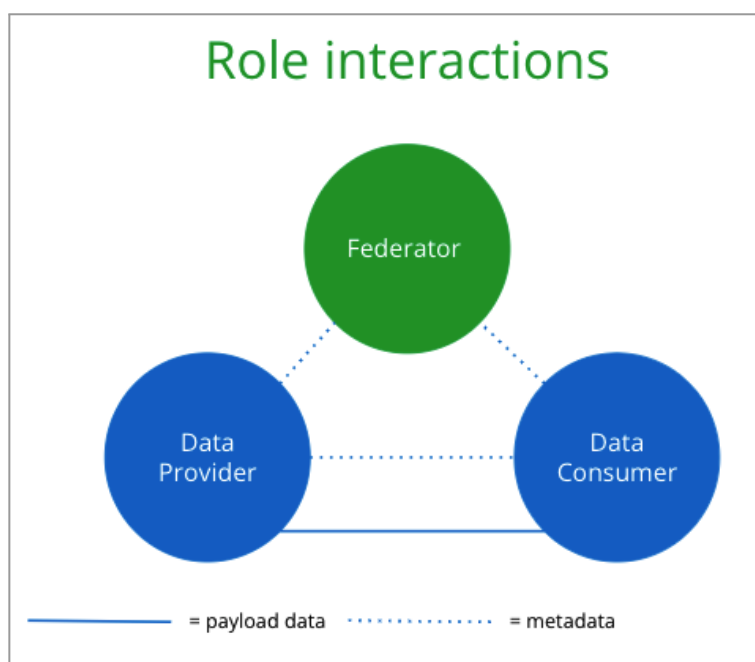


Figure 1: Role interactions in the data space; after Boris Otto: The Evolution of Data Spaces. In: Boris Otto, Michael ten Hompel, Stefan Wrobel: Designing Data Spaces. The Ecosystem Approach to Competitive Advantage, 2022. See online at [https://link.springer.com/chapter/10.1007/978-3-030-93975-5\\_1](https://link.springer.com/chapter/10.1007/978-3-030-93975-5_1) [last accessed: 11 July 2023]

Regarding what the concept of data covers in a data space, the Europeana glossary<sup>30</sup> defines this to cover the notions of “metadata”, “digital object” (incl. “object”) and “content”:

*“Metadata: the textual information and hyperlinks that serve to identify, discover, interpret and/or manage Content.*

*Digital Object: a digital representation of an object that is part of Europe's cultural and/or scientific heritage. The Digital Object can also be the original object when born digital.*

*Content: a physical or Digital Object that is part of Europe's cultural and/or scientific heritage, typically held by a Data Provider or by a data provider of the Data Partner.”*

While there is not a clear definition of what data is on the DSSC Glossary, the types of data that fall under the EU regulations are currently under negotiation with the Member States.

### 3.1.5. Platforms

The dictionary definition of a platform is “an application or website [or some type of computer hardware or operating system] that serves as a base from which a service is provided”<sup>31</sup>.

According to this general definition, an aggregating infrastructure or a data space can be considered a platform.

In everyday talk and in some of the interviews conducted by this Task Force, platforms are used in the narrower sense as referring to big commercial companies (usually from the USA)

<sup>30</sup> Online at <https://pro.europeana.eu/page/glossary> [last accessed: 11 July 2023]

<sup>31</sup> Merriam-Webster Dictionary, online at <https://www.merriam-webster.com/dictionary/platform>, and Collins Dictionary, online at <https://www.collinsdictionary.com/dictionary/english/platform> [both last accessed: 11 July 2023]

offering popular online business-to-business and business-to-consumer services. In this report, however, we are using the term in the more neutral and general sense cited at the beginning of this section.

### 3.2. Projects and organisations

Same as with the definitions in the previous chapter, the projects and organisations presented here only cover (1) those projects, initiatives, organisations, and networks that have had a role in the Task Force's knowledge gathering and (2) the bigger consortia and networks in which most other actors participate or to which these other actors are connected in some way. The Task Force acknowledges that there is more out there that would be useful to investigate further (see also the Task Force's recommendations in this regard in [section 8](#)).

#### 3.2.1. Data Spaces Support Centre<sup>32</sup>

The Data Spaces Support Centre (DSSC, <https://dssc.eu/>) was launched on 13 October 2022 as a first action in the project of the same name, which is currently funded by the European Commission under the Digital Europe Programme with a running time of three years. The project is coordinated by the Fraunhofer Gesellschaft zur Förderung der angewandten Forschung and its consortium includes, among others, the Big Data Value Association (BDVA), the IDSA, Gaia-x, and Capgemini Invent, with the latter also being a partner in the consortium for the common European data space for cultural heritage. The DSSC's associative and collaboration partners include the consultancy firm Innopay and the Big Data for Smart Society Institute (GATE), which is an autonomous structure of Sofia University "St. Kliment Ohridski", and others.

The DSSC "contributes to the creation of common data spaces, [...] that collectively create a data sovereign, interoperable and trustworthy data sharing environment, to enable data reuse within and across sectors, fully respecting EU values, and supporting the European economy and society." The Europeana Foundation as the coordinator of the project for deploying the common European data space for cultural heritage is one of 13 partners in the DSSC's Community of Practice.

#### 3.2.2. International Data Spaces Association<sup>33</sup>

The International Data Spaces Association (IDSA, <https://internationaldataspaces.org/>) is "an open, not-for-profit coalition" with its head office at the TechnologieZentrum Dortmund and more than 140 member companies representing a wide range of industries from over 25 countries (see an overview at <https://internationaldataspaces.org/we/members/>). IDSA was established on 26 January 2016 in Berlin. IDSA's goal is "a global standard for international data spaces (IDS) and interfaces, as well as fostering the related technologies and business models that will drive the data economy of the future across industries."

IDSA is a founding member of the Gaia-X AISBL, the not-for-profit association of the Gaia-X project (<https://gaia-x.eu/>). Gaia-X' mission is to "create an open, transparent, and secure

---

<sup>32</sup> Quotes are taken from DSSC's website at <https://dssc.eu/>. [last accessed: 11 July 2023]

<sup>33</sup> Quotes are taken from IDSA's website at <https://internationaldataspaces.org/> and the Gaia-X website at <https://gaia-x.eu/>. [both last accessed: 11 July 2023]

federated digital ecosystem, where data and services respond to common rules and can be and securely built, collated, and shared". For shared values and a comparison between the IDS Reference Architecture Model and the Gaia-X architecture see IDSA's position paper "Gaia-X and IDS", version 1.0, from January 2021<sup>34</sup>.

### 3.2.3. Big Data Value Association<sup>35</sup>

The Big Data Value Association (BDVA, <https://www.bdva.eu/>) is "an industry-driven organisation" with more than 230 members from all over Europe including companies from different industry sectors as well as research and user institutions. BDVA was established in 2014 and has its registered office in Brussels. In 2021, BDVA was given a new and extended mandate by its members, which also resulted in the change of its name to DAIRO Data, AI and Robotics). While BDVA's mission is to "develop an innovation ecosystem that enables the data-driven digital transformation of the economy and society in Europe", DAIRO has the more specific objective to "boost European Artificial Intelligence (AI), Data and Robotics research, development and innovation and to foster value creation for business, citizens and the environment".

Together with IDSA, Gaia-X, and the FIWARE Foundation (<https://www.fiware.org/>), BDVA has formed a new network in September 2021, the Data Spaces Business Alliance (DSBA, <https://data-spaces-business-alliance.eu/>). DSBA has been involved in the design of the Data Space Support Centre.

### 3.2.4. Solid Project<sup>36</sup>

The Solid (Social Linked Data) project (<https://solidproject.org/>) is led by Tim Berners-Lee and originated from a collaboration at the Massachusetts Institute of Technology (MIT) in 2015. Three years later, in 2018, Tim Berners-Lee co-founded the company Inrupt based on Solid and the W3C Solid Community Group (<https://www.w3.org/groups/cg/solid>) was established to support a process of open standardisation. Solid is "a specification that lets people store their data securely in decentralized [sic] data stores called Pods". To this aim, the Solid specification builds on existing web standards like HTTP, URI, or WebID.

The ideas represented by Solid have - among others - inspired the Post-Platforms Foundation (<https://postplatforms.org/>), a non-for-profit organisation under Dutch law that also is a member of IDSA. While the original idea of Solid does include organisational pods next to personal ones already, it is aimed at "restor[ing] the power and agency of individuals on the web [and removing such power from] digital giants"<sup>37</sup>, i.e., the big commercial platforms that dominate the internet. Post-Platforms' solution on the other hand keeps platforms as part of the picture but aims "to decouple data from platforms [and to have] a decentralized [sic] Web [3.0 Data Space] [...] where people, organizations [sic], and devices own their data and control access rights, whereas platforms and services access it on par

---

<sup>34</sup> Online at [https://internationaldataspaces.org/wp-content/uploads/dlm\\_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf](https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf) [last accessed: 11 July 2023]

<sup>35</sup> Quotes are taken from BDVA's/DAIRO's website at <https://www.bdva.eu/> [last accessed: 11 July 2023]

<sup>36</sup> Quotes are taken from the Solid project's website at <https://solidproject.org/> and Post-Platform's website at <https://postplatforms.org/> [both last accessed: 11 July 2023]

<sup>37</sup> <https://www.inrupt.com/blog/one-small-step-for-the-web> [last accessed: 11 July 2023]

with each other". Furthermore, Post-Platforms' approach includes increased security by providing all participants with keys and e-passports (verifiable credentials).

The Post-Platforms Foundation claims that it provides a bridge between the ideas of the Solid project and IDSA respectively by creating an ecosystem, where platforms indeed are interconnected, but via data remaining under the (personal) control of their owners. Furthermore, Post-Platforms was one of the initiators of the Personal Data Space task force at IDSA, thereby bringing this perspective from the Solid realm into the conversations at IDSA. More details on how Post-Platforms builds, but is rather different from the pure Solid approach, will be given in [chapter 6.2.2](#).

## 4. Approach

The Task Force followed a mainly qualitative approach in gathering insights and in identifying patterns relevant to its topic. This happened:

- in the form of
  - group discussions,
  - smaller working groups, and
  - individual interviews and
- in the context of
  - the Task Force's own meetings,
  - the Europeana Aggregators' Forum's meeting in spring 2023, and
  - scheduled interviews with representatives from technology projects, research institutions, strategic consultancies, aggregators, users of cultural heritage data, and opinion leaders of the cultural heritage field in general, the Europeana Initiative in particular, and the data space realm.

It should be noted that the Task Force's primary perspective has been the one of an aggregator, while the point of view of CHIs has been represented by aggregators based on their knowledge and experience in working closely with CHIs in their own countries or domains. The knowledge gathering done by the Task Force has, however, not yet included CHIs directly, nor users of cultural heritage data to a broader extent apart from the apps and platforms that have been interviewed. The Task Force's recommendations ([see section 8](#)) suggest further explorations in these directions to get a fuller picture. Similarly, not all the actors in the data space context could be interviewed due to time constraints and availability of potential interviewees, so that the interviews - while providing a stable and broad enough basis to draw conclusions from - are by no means exhaustive.

### 4.1. Task Force meetings

Between mid-November 2022 and mid-June 2023, the Task Force met six times. Next to general updates on the plan of actions, the progress of the interviews and the first draft of the report the Task Force conversations focused on identifying and developing the aggregation use cases. During the first Task Force meeting in November 2022, participating aggregators presented their current workflows and discussed existing challenges to their work in an open brainstorming session. While still at the beginning of the Task Force's exploration, this also included some initial thoughts on how a more decentralised approach to aggregation could potentially help in addressing these challenges.

Task Force members then had the opportunity between the first and second meeting to add to the initial collection of use cases by detailing their own contributions and confirming or supplementing use cases brought up by others. A first summary and different options to provide an overview of the information gathered were discussed back with the whole group during the second Task Force meeting in January 2023.

The third and fourth Task Force meetings in February 2023 were dedicated to preparing and reporting back on the interviews held with the group of participating aggregators themselves, which provided another opportunity for more of a deep dive into various aspects mentioned as part of the use cases during individual conversations.

Based on the insights from the group discussion and the respective interviews, a set of 17 summarising use cases was developed, mostly from an aggregator's perspective, but also including four from the perspective of a CHI ([see the following section 5](#)).

This was presented to the whole Europeana Aggregators' Forum (EAF) during their meeting at the beginning of April 2023 for a broader sense-check.

#### 4.2. Additional interactive sessions

Initially planned as part of the EAF's spring meeting but moved to an alternative date due to time constraints, two additional interactive sessions were held with aggregators, both Task Force members and from the wider EAF. These sessions looked at selected use cases, chosen by the participants of each session based on what they considered as the one or two use cases most important, most pressing, or simply most relatable in their specific context. In a guided discussion, attendees first mapped the use case with their experiences, then tried to anticipate how the workflows and processes involved in the use case would possibly change if they happened in a decentralised context.

#### 4.3. Interviews

Between December 2022 and June 2023, the Task Force conducted 45 interviews with 61 interviewees (see the complete list in [Annex I](#)). Next to the 13 participating aggregators and three non-European international aggregating services, this included:

- ten interviews with representatives of relevant technology projects and initiatives;
- three interviews each with representatives of research institutes, strategic consultancies, and apps and platforms using cultural heritage data;
- five interviews each with representatives of collection management system providers and opinion leaders.

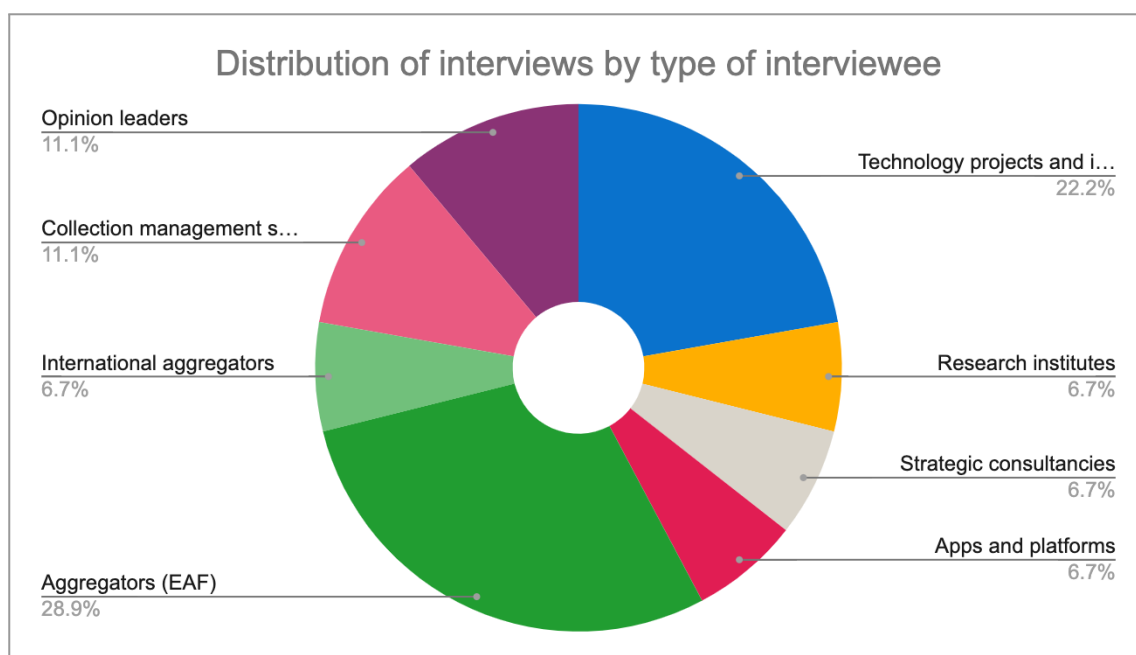


Figure 2: Distribution of interviews by type of interviewee with percentages



All interviews were held in an online or hybrid format and were recorded for the sole purpose of note keeping and with the agreement and approval of the interviewees.

For reasons of comparison, each interview was conducted with the same set of questions (see [Annex II](#) for details), though personalised to the interviewee and possibly adapted in sequence according to the natural flow of the interview. Not all questions applied to the same extent to each interviewee or type of interviewee, but a general baseline was kept across all interviews.

Each interview took between 45 and 90 minutes in total, i.e., including administrative aspects and introductions. Especially in interviews with persons from outside the more immediate Europeana context, the first part of the interview also included an introduction into the Europeana Initiative and its transformation into the common European data space for cultural heritage as the backdrop against which the Task Force was operating.

While all interviews with the same type of interviewee were initially planned to be held within the same timeframe of a few weeks per group of interviews, timetables and availability required a more mixed approach eventually. Similarly, not all intended interviews could be held - to some extent also because the intended interviewees did not feel they were at that stage in their own exploration journey yet when they would be able to provide valuable input - and not all additional suggestions for interviews could be incorporated into the overall schedule.

However, the mixed approach allowed bringing in thoughts from previous interviews, independent of the type of interviewee, into the ongoing interview process, so that later interviews benefitted from a kind of “cross-pollination” of ideas from various backgrounds. Furthermore, the group of interviewees proved substantial enough for the current remit of this Task Force and the anticipated balance between the different types of interviewees was kept in general.

Next to the group discussions and interactive sessions on the aggregation use cases, the interviews form the main basis for the insights gathered in this report and the conclusions drawn from them.

## 5. Use cases and challenges from the current aggregation landscape

As mentioned above, the use cases presented in this section stem from the conversations had within the Task Force in the form of brainstorming sessions and group discussions as well as the individual interviews with the participating aggregators. While not all details from these conversations will have found their way into the summarised version of the use cases as they are detailed below, these use cases still capture the main essence of what was said by various stakeholders in slightly different ways.

The 17 use cases covered here can broadly be categorised in two ways: whether they are presented from the perspective of an aggregator or the perspective of a CHI and whether they refer to a technical aspect of aggregation or an organisational one.

It should be noted that, while this set of use cases does include four which are described from the perspective of a CHI, this is based on aggregators' comments rooted in their knowledge and experience of working with CHIs in their own national or domain-specific contexts. An extension of the use cases presented here to also include other perspectives (CHIs and end-users in particular) is part of the Task Force's recommendations for next steps (see [section 8](#)).

Furthermore, it should be emphasised that the use cases detailed in this section are based on the status quo, while the prospective evolution of two selected use cases in a more decentralised approach will be described in [section 7](#). This also means that the terms "aggregator" and "aggregation" are mainly used in the way they are currently understood in the context of the Europeana Initiative; i.e., an aggregator is defined as "an organisation working with [CHIs] and collectors to gather authentic, trustworthy and robust data [...and] mak[ing] this content available to a broader audience via their own services, Europeana and other infrastructures, for example, for education and research"<sup>38</sup>, while the Task Force generally acknowledges that both these terms will have to be revisited as part of a potential move towards a decentralised approach.

Similarly, it is worth noting that while some of these use cases might possibly be addressed better in a decentralised approach, decentralisation will not be the (only) solution to all of them in the same way.

### 5.1. Relevance of identified use cases across the Europeana Aggregators' Forum

During the EAF meeting in April 2023, the Task Force presented its findings and conclusions so far, which included the summarised 17 use cases that had been identified and consolidated by the Task Force. Attendees of the meeting were asked to vote for those use cases that resonated most with them and their own experiences. The aim of this exercise was to confirm that the use cases were indeed representative and to get an understanding of a possible ranking of the use cases in anticipation of their prioritisation when addressing them and considering an alternative approach for them in a decentralised set-up.

Keeping use cases from an aggregator's point of view and those from the point of view of a CHI separate, three priority groups were identified for each of these two perspectives.

---

<sup>38</sup> See the definition of an aggregator as part of the Europeana Glossary of Terms, online at <https://pro.europeana.eu/page/glossary> [last accessed: 11 July 2023]

Use case number	Use case description
<b>Aggregators' use cases</b>	
<i>Considered relevant in more than 60% of votes</i>	
UC-1	As an aggregator, I want to make sure that sources I have aggregated stay available and accessible so that my users don't end up with broken links when using my services.
UC-2	As an aggregator, I want to always provide access to the latest version of the data available at the source.
UC-3	As an aggregator, I need to have a clear way to present the benefits of an aggregator so that I can successfully engage content partners.
UC-4	As an aggregator, I'd like all the enhancements (e.g., semantic enrichments) done by me to flow back to the source so that the process is not re-done repeatedly.
<i>Considered relevant in more than 25% but less than 50% of votes</i>	
UC-5	As an aggregator, I want to be able to easily publish subsets of my collection to Europeana and track which records are published.
UC-6	As an aggregator I want to have the necessary resources to provide additional services (e.g., storage), if needed, so that I can help in case CHIs do not have that capacity or projects end and their data are at risk of becoming unavailable.
UC-7	As an aggregator, I want to add URI-s from Linked (Open) Data vocabularies in case these are not available at source so that interoperability can be ensured more easily.
UC-8	As an aggregator, I want to use my own application profile in the data space as complementary to others so that I can better support my network's use case.
UC-9	As an aggregator, I'd like to set certain quality criteria for metadata and pick up (or reject) metadata according to these criteria so that I can provide my users with the best possible experience based on their needs.
UC-10	As an aggregator, I want to share data of a CHI according to the conditions the CHI has defined in a clear and easy way which doesn't impact the use of the data. This would ideally be done in an automated way rather than requiring an administrative step of signing an agreement.
UC-11	As an aggregator, I want to be able to incorporate "additional" data which is produced by projects on various platforms including my own and those of CHIs so that I can provide my users the broadest experience possible.
<i>Considered relevant in less than 25% of votes</i>	
UC-12	As an aggregator working in an ecosystem of aggregation, I want to complement the work of my peers in terms of data processing and

Use case number	Use case description
	publication. This is to avoid duplicated efforts at all stages of the process, from CHIs to aggregating platforms.
UC-13	As an aggregator, I want to have the possibility to receive data from CHIs' preservation systems if useful and required (e.g., for providing access to higher resolution digital representations) so that I can offer the most suitable services to my users.
<b>CHIs' use cases</b>	
<i>Considered relevant in more than 80% of votes</i>	
UC-14	As a CHI and a user of the data space, I want to make sure that the information on provenance is kept so that it is clear where original data and updates come from.
<i>Considered relevant in more than 50% but less than 80% of votes</i>	
UC-15	As a CHI, I want my changes to the data at source be reflected (automatically) on all aggregating platforms so that shared data do not get outdated.
UC-16	As a CHI, I want to be able to take down data from all the aggregating platforms within a certain amount of time.
<i>Considered relevant in less than 25% of votes</i>	
UC-17	As a CHI, I want to be able to modify data on the level of an aggregating platform and have the change(s) propagated to the source.

## 5.2. Technical use cases from an aggregator's perspective

Given the technical focus of the Task Force as well as the foundation of the aggregators' work in data processing and publication, about half of the use cases identified fall in the category of technical use cases from an aggregator's perspective.

The following sub-chapters present these use cases with their use case number and description as already introduced with the table above. Furthermore, an explainer has been added for each use case based on the knowledge gathered from the various interactions with aggregators as part of this Task Force's activities. The same applies for the sub-chapters of chapters 5.3 and 5.4. below.

The use cases detailed in 5.2.2. (combined with the one in 5.2.1.) and in 5.2.3. are the ones selected by the participants of the interactive sessions and therefore also include a visualisation of the current workflow along with an initial ideation of how this could be adapted to a decentralised approach.

### 5.2.1. UC-1 Avoid broken links

**Description:** As an aggregator, I want to make sure that sources I have aggregated stay available and accessible so that my users don't end up with broken links when using my services.

**Explainer:** CHIs as well as aggregating services might originally have shared data created, updated and/or made available in the context of funded projects. Once these projects end, the data might either not be updated anymore or the website where the data are held might be discontinued, so that data cannot be accessed anymore at all.

Another example are links to digitised or digital representations of the cultural heritage objects. Such representations, i.e., media files in a variety of formats, are often managed in a system different from the one used for managing the collections and their descriptions and might be shared online with a different base URL. When sharing their data, CHIs will employ workflows to create links to their media database or image viewer to be included in the metadata exported from their collection management system. However, when the media database or image viewer is moved to a different domain, data shared beforehand might not always be updated immediately, so that they still include the outdated resource links. If no redirection service exists, these links will appear as broken to the user. The same will happen if identifiers are missing completely or are changed or reused to describe another resource.

### 5.2.2. UC-2 Always provide access to latest version

**Description:** As an aggregator, I want to always provide access to the latest version of the data available at the source.

**Explainer:** The current aggregation landscape works with copies of data, from the CHI e.g., via a regional aggregator and a national aggregator to the Europeana.eu website. If something changes at the source, each copy needs to be updated in sequence, which takes time and might lead to data being out of date - or at least out of sync - on an aggregator level, even if only for some time.

Some aggregators might work with automated upload and update processes, but update intervals might not be daily, but only weekly, biweekly, or with longer periods of time or information related to an update (e.g., timestamps) is not always captured in a way that is easily reusable by humans or machines. Other aggregators might allow CHIs to run more immediate updates, e.g., via an HTTP upload option, but the person responsible for the updates towards the aggregator might not always be the same person as the one working on the collections and having done the update to the data and so the necessity for an update towards the aggregator might not always be known. And aggregators such as Europeana, where the actual ingest and data processing is done by a central team, might have other datasets up for processing already and with a higher priority, so that minor updates might not always be processed directly when they occur.

#### *Visualisation*

Figure 3 below shows a case where a CHI shares part of its metadata, with links to digitised or digital representations and managed with a focus on the object- or item-level, with a national aggregator (action 1a), while another part of its metadata, managed with a focus on

the collection-level, is shared with a domain aggregator (action 1b). The latter might also encompass data that does not include links to digitised or digital representations yet (action 1c). With this, the CHI might have three different packages of data to submit to two different aggregators. Each aggregator will most likely run some data processing of their own (actions 2a and 3a respectively), might it be for standardisation and transformation into a common data model or format applied by the aggregator in its own context, might it be for data enrichment, might it be for data indexing to enable discovery via a search engine as part of the aggregator's own services, etc.

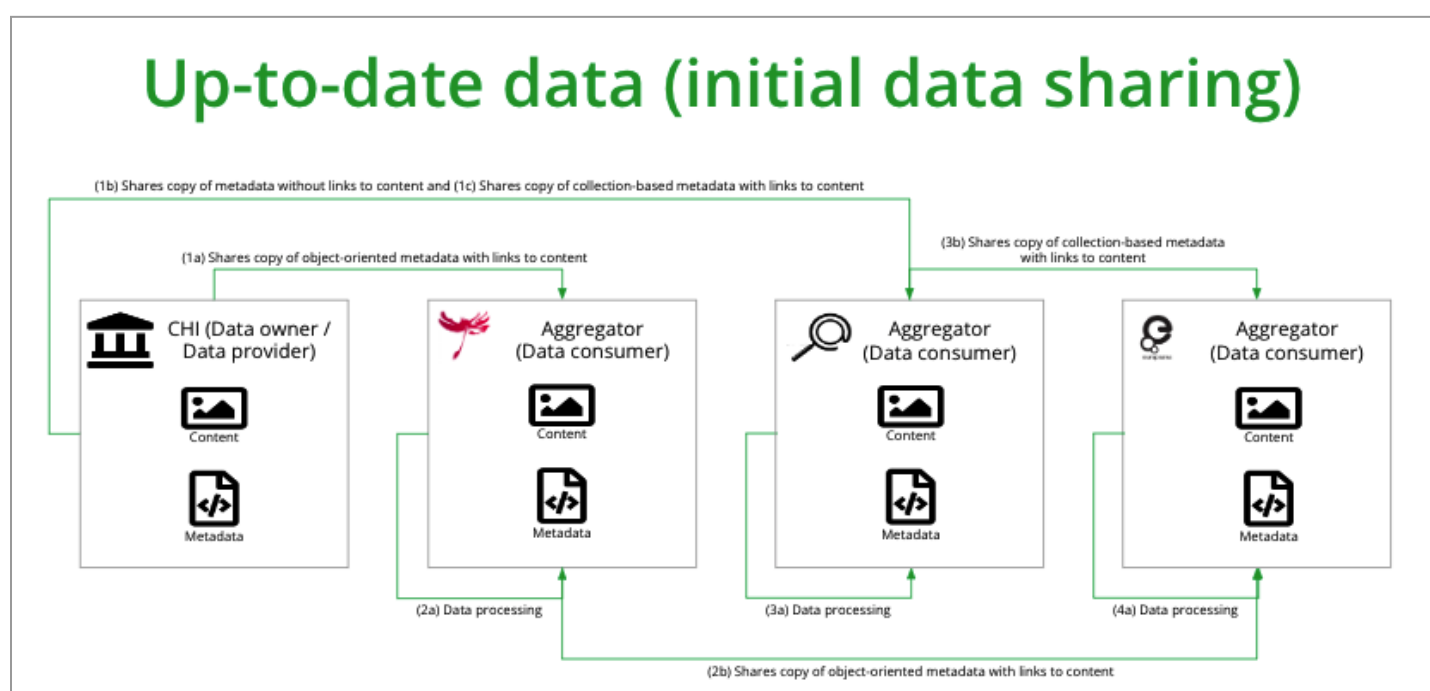


Figure 3: Visualisation for UC-2 Always provide access to latest version; here: current workflow for an initial data submission by a CHI

Both aggregators will furthermore run relevant processes in preparation for sharing the data via the Europeana.eu website (actions 2b and 3b respectively). Finally, Europeana Foundation as the operator of the Europeana.eu website might again do some data processing of their own (action 4a).

Figure 4 below then illustrates the process that needs to happen, when the CHI updates the data locally and wants these changes to be reflected in all aggregating services as well. With the current process of copying metadata from the source to every aggregator that a CHI wants to connect to, an update (action 5a) essentially requires the repetition of all the actions mentioned above.

## Up-to-date data (updating data)

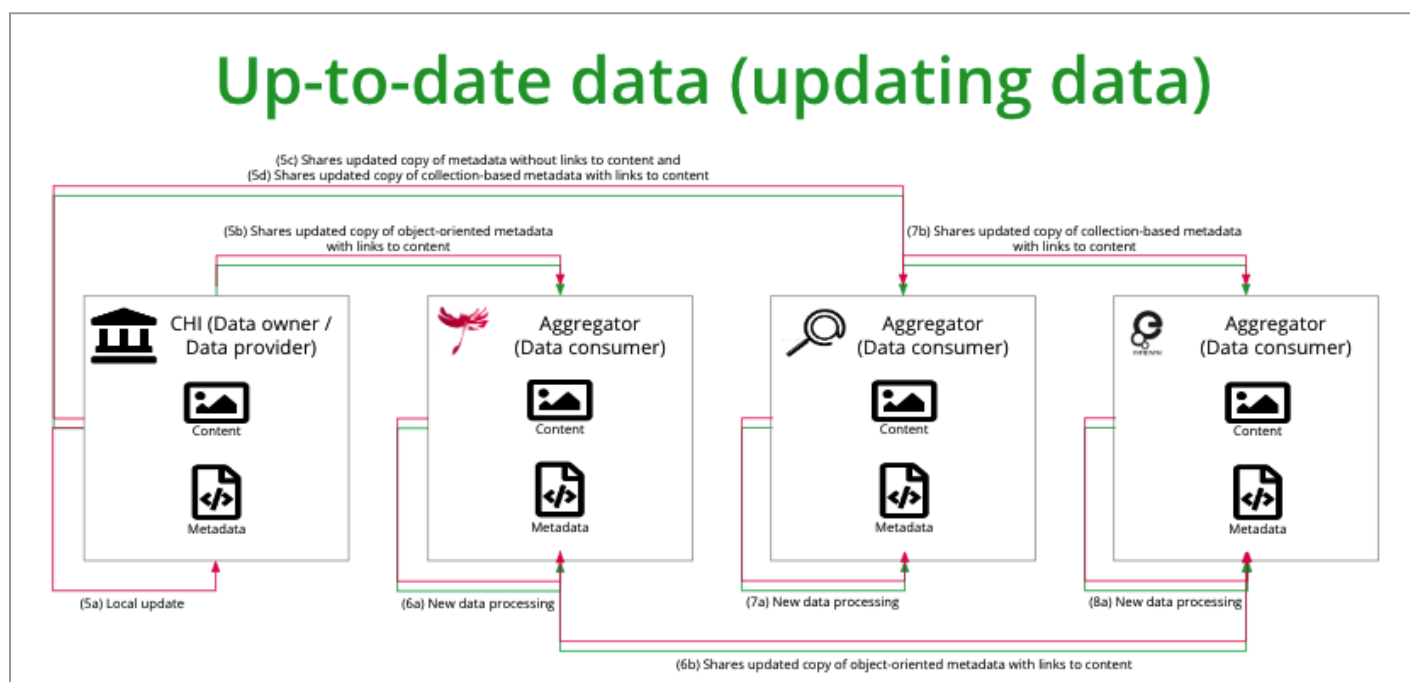


Figure 4: Visualisation for UC-2 Always provide access to latest version; here: current workflow for data being updated by the CHI

How this scenario might change in a decentralised approach, will be discussed in more detail in [section 7](#) below.

### 5.2.3. UC-4 Give back enrichments

**Description:** As an aggregator, I'd like all the enhancements (e.g., semantic enrichments) done by me to flow back to the source so that the process is not re-done repeatedly.

**Explainer:** As part of its data processing for publication on its own website and/or for onwards sharing with other services, an aggregator might enhance the data that are provided by the CHIs. This might include aspects of data standardisation (e.g., transforming a proprietary format into an international standard XML or RDF format), data normalisation (e.g., adding normalised dates to support discovery) or data enrichment, (e.g., adding Linked Open Data URI-s for the identification of entities, when these are not present in the source data). These enhancements currently stay on the aggregator level, without the option to feed them back to the source data and the data owner.

Visualisation

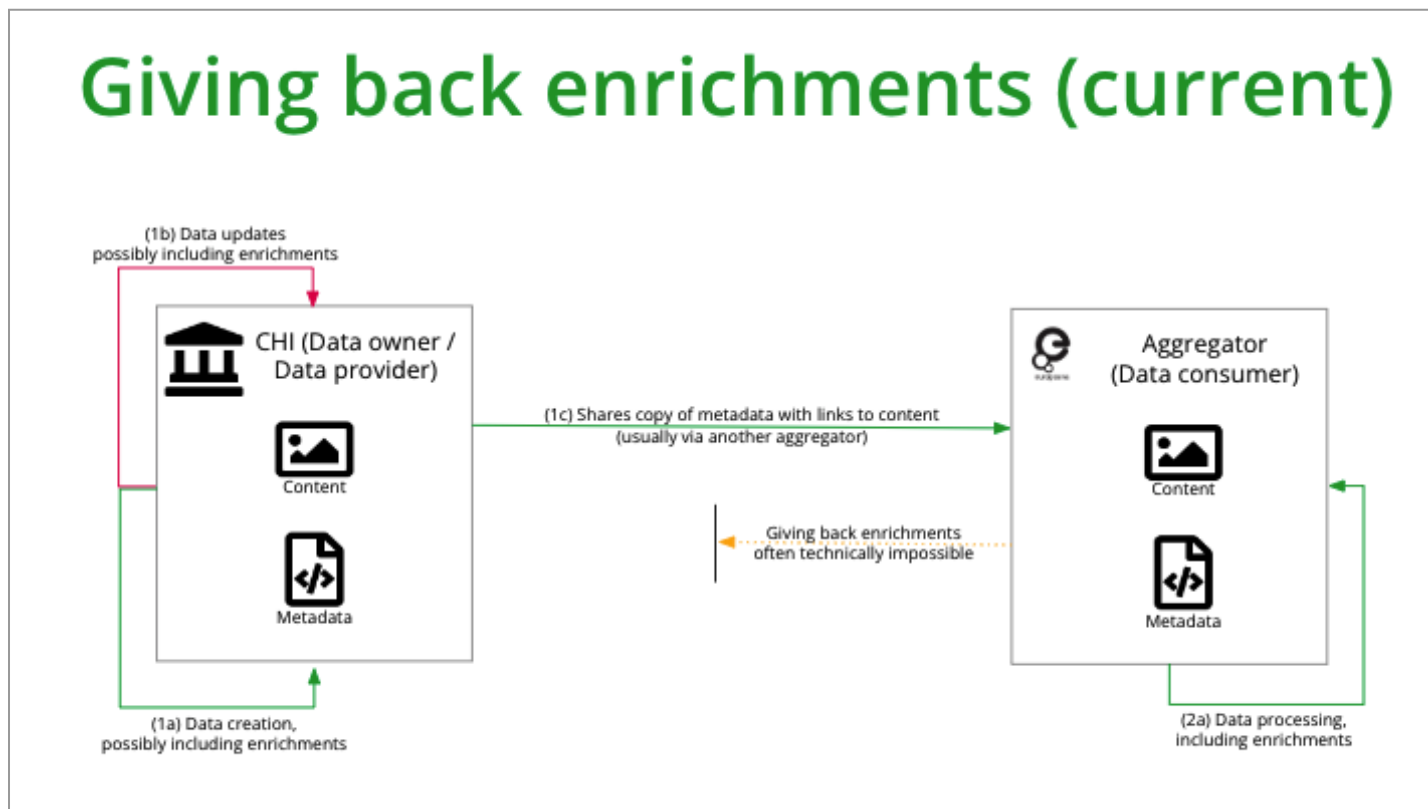


Figure 5: Visualisation for UC-4 Giving back enrichments; here: current workflow

Figure 5 above shows a case where a CHI creates and updates data in its own context (actions 1a and 1b), which might or might not include steps of data enrichment such as using OpenRefine to identify named entities and assign Linked (Open) Data (LOD) URI-s to them. The CHI then shares its metadata on Europeana.eu (action 1c), which usually happens via an aggregator. This intermediate step has been left out for simplicity of the visualisation and because it is not primarily relevant in the context of this use case. Instead of the Europeana.eu website any other national, domain or thematic aggregator could be represented on the right-hand side. The aggregator then runs some data processing of its own (action 2a), including data enrichment. As the aggregator might have transformed the submitted data into either a different standard (from a proprietary XML to LIDO<sup>39</sup> or EAD<sup>40</sup> or EDM) or a different data format (from CSV to XML or from XML to RDF) or might have combined pieces of data or separated pieces of data out, there usually is no technical direct way of integrating the enriched data back into a CHI's own system(s).

How this scenario might change in a decentralised approach, will be discussed in more detail in [section 7](#) below.

<sup>39</sup> Lightweight Information Describing Objects (LIDO), see more information online at <https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/> [last accessed: 11 July 2023]

<sup>40</sup> Encoded Archival Description (EAD), see more information online at <https://www.loc.gov/ead/> [last accessed: 11 July 2023]



#### 5.2.4. UC-5 Publish subsets and keep track

**Description:** As an aggregator, I want to be able to easily publish subsets of my collection to Europeana and track which records are published.

**Explainer:** Not all data held within a specific aggregator's context might be suitable for publication on Europeana.eu for a variety of reasons, e.g., when data does not (yet) include links to a digitised or digital representation of the cultural heritage objects or when digitised representations do not meet the quality criteria of the EPF. One example for the latter could be that digitised or digital representations are behind a login, which hinders reading out the necessary technical details as part of Europeana's own aggregation workflow. because they might be behind a login, which hinders the necessary technical details being read during the data processing by Europeana's own aggregation workflow. In such cases, there might be a need to identify and select the data according to their publication route.

#### 5.2.5. UC-7 Ensure interoperability via Linked Open Data URI-s

**Description:** As an aggregator, I want to add URI-s from LOD vocabularies in case these are not available at source so that interoperability can be ensured more easily.

**Explainer:** Source data might not always include LOD URI-s or controlled vocabulary terms in general, which play an essential part in connecting data from a variety of sources in the context of an aggregating service. Therefore, aggregators might want to further process the data submitted to them e.g., with Natural Language Processing (NLP) methods and algorithms to identify relevant entities and enrich the data by adding LOD URI-s.

When sharing data further with the Europeana.eu website, aggregators might want to ensure that the vocabularies they use are supported by Europeana<sup>41</sup> and that the URI-s follow the format of the example URI-s listed.

#### 5.2.6. UC-8 Use own application profile

**Description:** As an aggregator, I want to use my own application profile in the Data Space as complementary to others so that I can better support my network's use case.

**Explainer:** Aggregators might use domain metadata standards or national variations of a standard or they might be using their own implementation profile of the EDM to allow for more richness in the information they provide to their users. When sharing data on Europeana.eu, aggregators will need to then transform the data to EDM, which might go along with ignoring parts of more complex data models used by the aggregators in their own contexts or even tweaking some data to fit with the way the Europeana.eu website displays them, while ensuring that the most detailed information can be shared for the benefit of the user.

#### 5.2.7. UC-9 Set quality criteria for metadata

**Description:** As an aggregator, I'd like to set certain quality criteria for metadata and pick up (or reject) metadata according to these criteria so that I can provide my users with the best possible experience based on their needs.

---

<sup>41</sup> See details online at [https://docs.google.com/spreadsheets/d/1BoDNolkcp\\_qfvVShdOZyGcf61XslcwKF2MdGcigYs20/edit#gid=0](https://docs.google.com/spreadsheets/d/1BoDNolkcp_qfvVShdOZyGcf61XslcwKF2MdGcigYs20/edit#gid=0) [last accessed: 11 July 2023]

**Explainer:** While it could represent a threshold for some CHIs, an aggregator might want to require certain pieces of (meta)data to ensure the quality of the service they provide (like the EPF). Setting quality criteria for metadata is probably about striking the balance between making it easy for CHIs to participate and offering quality services to users. Aggregators might, in this context or more generally, have developed their own set of aggregation tools to support data processing in the form of extracting and moving metadata, adding metadata, or combining metadata with the aim of improving data quality.

#### 5.2.8. UC-11 Connect to additional data sources

**Description:** As an aggregator, I want to be able to incorporate "additional" data which is produced by projects on various platforms including my own and those of CHIs so that I can provide my users the broadest experience possible.

**Explainer:** There might be other data than LOD URI-s, which could be helpful to users of an aggregation service by contextualising the source data provided by CHIs. This could include crowdsourced data next to data that has been generated automatically via NLP and related methods, but also transcriptions, translations, annotations created and added by a variety of possible actors. An aggregator might want to present such data in combination with the source data for an enhanced user experience.

#### 5.2.9. UC-13 Include other types of data

**Description:** As an aggregator, I want to have the possibility to receive data from CHIs' preservation systems if useful and required (e.g., for providing access to higher resolution digital representations) so that I can offer the most suitable services to my users.

**Explainer:** Currently, most aggregators might receive data that is meant for publication, which usually sits in the collection management system of the CHI, while there might be parallel data specifically for preservation. This could be additional metadata, but also - potentially - higher resolution versions of digital objects. Depending on the services offered by an aggregator, they might be interested in presenting such other data either in addition to the data that is already made available or instead of them.

### 5.3. Organisational use cases from an aggregator's perspective

While data processing and publication is part of the aggregators' daily routine, there is a whole range of other activities on their task list as well: from capacity building via developing tools to support "their" CHIs to contributing to framework and policy decisions. Some of these might relate to, build on or influence the more technical activities, but others bring in a more administrative and organisational, a governing perspective.

#### 5.3.1. UC-3 Present benefits of an aggregator

**Description:** As an aggregator, I need to have a clear way to present the benefits of an aggregator so that I can successfully engage content partners.

**Explainer:** Aggregating services might have started in the context of funded projects and will now have to ensure continuous growth to stay relevant or they might currently receive funding bound to reaching certain key performance indicators regarding the number of (new) content partners they aggregate data from. Independent of whether the aggregator functions

on a regional or national level or builds on a thematic or domain approach, CHIs - often already having their own means of publishing and sharing their data - will ask for clear benefits of contributing and for ways to do so that least interfere with their workflows or add further workload to their staff.

#### 5.3.2. UC-6 Acquire necessary resources

**Description:** As an aggregator I want to have the necessary resources to provide additional services (e.g., storage), if needed, so that I can help in case CHIs do not have that capacity or projects end and their data is at risk of becoming unavailable.

**Explainer:** Aggregators are often run with a minimal number of staff. Not all might have full-time employees but might work with team members who are seconded by supporting institutions for a certain percentage of their time. Nevertheless, the ask towards aggregators in terms of the services they provide increases continuously. Especially when it comes to being as inclusive and diverse as possible, this also includes services aiming specifically at smaller CHIs, e.g., with just a single staff member or being run completely by volunteers that need support also in other areas than data processing per se. With aggregators being constrained themselves regarding resources, such requests often present a challenge.

#### 5.3.3. UC-10 Reduce administrative steps for the reuse of data

**Description:** As an aggregator, I want to share data of a CHI according to the conditions the CHI has defined in a clear and easy way which doesn't impact the use of the data. This would ideally be done in an automated way rather than requiring an administrative step of signing an agreement.

**Explainer:** Conditions to share data via the Europeana.eu website are currently defined as part of the Europeana Data Exchange Agreement<sup>42</sup>, which every data provider (CHI) or provider (aggregator) must sign. Among other aspects, this confirms the Creative Commons Universal Public Domain Dedication (CC0) licence<sup>43</sup> being applied to the metadata. In the data space for cultural heritage, however, data owners might want to make metadata and content available under open licensing models that are not covered by the DEA. Conditions for the reuse of data should be based on standardised licensing statements, so that they can be interpreted as part of data sharing mechanisms in a more automated way. Such a process would limit the necessity of more traditional agreements (e.g., requiring signatures of the parties involved).

#### 5.3.4. UC-12 Collaborate with other aggregators

**Description:** As an aggregator working in an ecosystem of aggregation, I want to complement the work of my peers in terms of data processing and publication. This is to avoid duplicated efforts at all stages of the process, from CHIs to aggregating platforms.

**Explainer:** Some aggregators are connected to each other in terms of data flow between their services, but potentially also organisationally. E.g., a regional aggregator might provide data to a national aggregator, which then shares the data with Europeana, but might also

---

<sup>42</sup> See more information online at <https://pro.europeana.eu/page/the-data-exchange-agreement> [last accessed: 11 July 2023]

<sup>43</sup> See online at <https://creativecommons.org/publicdomain/zero/1.0/> [last accessed: 11 July 2023]

share the data with a domain aggregator, which also shares data with Europeana. On the one hand, this allows CHIs to reduce their efforts by making use of existing products, services, and workflows between aggregators. It also allows aggregators to benefit from the work of other aggregating services, e.g., in terms of standardisation or enrichment of data. On the other hand, such an aggregation route might increase the risk of data being shared in duplicate if communication between aggregators fails or if the aggregators' technical workflows and processes do not include necessary selection of those data that would be shared via the one or the other channel.

#### 5.4. Use cases from the perspective of a cultural heritage institution (CHI)

Not only do aggregators interact directly with CHIs in the context of their own workflows and processes, but they also represent "their" CHIs in the communication and data publication process via the current Europeana.eu website. As such, the following use cases from the perspective of a CHI complement - to some extent - the use cases presented above from an aggregator's point of view. They thereby enhance the picture painted for a specific use case. However, there also are use cases presented here that speak directly to the needs of CHIs as recounted by the aggregators from their experience in working with CHIs.

##### 5.4.1. UC-14 Provide clear provenance information

**Description:** As a CHI and a user of the Data Space, I want to make sure that the information on provenance is kept so that it is clear where original data and updates come from.

**Explainer:** While the data provider and the provider are clearly mentioned in the display of the Europeana.eu website and can also be used as filters of search results, this attribution applies to the whole item description. This means that no distinction can be made on a more granular level to e.g., identify when an LOD URI for a person, a place, a subject, or some other entity has been added by an aggregator rather than having been included in the source data already.

In general, provenance information is an essential conveyor of transparency and trust and should be available as an attribute of the data at all levels, but also with statements on the data such as whether they were created manually by the data owner, crowd-sourced, or automatically generated.

##### 5.4.2. UC-15 Have local changes reflected on aggregator level automatically

**Description:** As a CHI, I want my changes to the data at source be reflected (automatically) on all aggregating platforms so that shared data do not get outdated.

**Explainer:** This use case can be considered the extension to [UC-2 described above](#) from the CHI's perspective. CHIs might continuously work on improving existing descriptions, might add descriptions for existing cultural heritage objects, and might create descriptions for newly acquired cultural heritage objects and/or collections. Such changes can range from correcting a typo to adding a substantial amount of new information or including links to newly digitised representations or LOD vocabularies.

#### 5.4.3. UC-16 Unpublish or delete data from aggregator level easily

**Description:** As a CHI, I want to be able to take down data from all the aggregating platforms within a certain amount of time.

**Explainer:** While aggregated data usually are free from potential restrictions for publication, there still might be cases, when data need to be taken down due to including personal information or due to including information that might impact (inter)national security, e.g., in the case of current events changing the data's initial appraisal in this regard. Depending on the situation at hand, unpublishing or deleting data from all online services might need to happen as directly as possible.

#### 5.4.4. UC-17 Modify data on aggregator level

**Description:** As a CHI, I want to be able to modify data on the level of an aggregating platform and have the change(s) propagated to the source.

**Explainer:** This use case can be considered the inverse of [UC-15 described above](#) and a variation of [UC-4 described from the aggregator's perspective](#). In this case here, it is the CHI itself that - triggered by the presentation of its own collections in connection with the collections of other CHIs in the context of an aggregating service - might want to add new information to the description of its items, e.g., referencing related items held by other CHIs. The CHI might want to do this on the aggregator level at first, where the relationship between both items is more imminent, but it might also want to share this information back to its own systems for local use.

## 6. Decentralised approaches to aggregation

This report frames the topic of data aggregation within the broader concept of data sharing. Looking back on the past 15 years, the cultural heritage sector, particularly within the context of the Europeana Initiative, has traditionally used the term “aggregation” to talk about:

- giving access to data from one data owner or holder in combination and in connection with the data of any number of data owners or holders from the same geographical area or with the same thematic or domain interest;
- giving access to data via parties other than the primary data owner or holder; and
- giving access to data that is a copy of the source data, which might have seen various stages of transformation, standardisation, and enrichment.

For the Europeana Initiative, this went along with the development of common frameworks such as the ELF for standardised information about the licensing and rights status of specifically the digitised and digital representations of cultural heritage objects, and the EPF and EDM for metadata and content quality and for a common and interoperable data model for ingestion and central publication.

However, as new methods of giving access to data for usage have emerged and technology - as always - keeps evolving, the need to innovate how data is made available is also reflected in an evolving terminology and in the traditional understanding of the term “aggregation” and the role of the “aggregator” being challenged.

The following sub-chapters describe several decentralised approaches to securely and trustworthily making data available for usage by parties outside the data’s original jurisdiction. These approaches represent the initiatives that are mainly being discussed in the context of data spaces in Europe and have broadly been categorised in platform-centric approaches (see [chapter 6.1.](#)), i.e., initiatives that start from the concepts of data providers and data consumers first and concentrate on the network between organisations and platforms that make up a data space, and data-centric approaches (see [chapter 6.2.](#)), i.e., initiatives that start from the data, from where they are originally held, and from the processes and workflows in place to make data available. A third chapter (see [chapter 6.3.](#)) then gathers related approaches that might connect more easily to the one or the other of these categories, but do not necessarily fit in either of them themselves.

The exploration of these approaches below will lead to the introduction of suggestions for a new and extended terminology when it comes to making data available, which - among other aspects - will be detailed in [section 7](#) afterwards.

It should be noted that this report mainly looks at those approaches for general understanding of their direction and of their potential differences but will not go into further detail about their active implementation at this stage. Especially regarding the implementation of Linked Data based approaches, reference is to be made to the parallel Linked Data Task Force run by ENA’s Tech Community. The outcomes of that Task Force will become available in autumn 2023<sup>44</sup>.

---

<sup>44</sup> See online at <https://pro.europeana.eu/project/linked-data-task-force> [last accessed: 11 July 2023]

## 6.1. Platform-centric approaches

### 6.1.1. International Data Space Association (IDSA)

The primary goal of IDSA is to optimise the data flow from data providers (or data owners) to data consumers (or data users) by “establishing a standard for data sovereignty - for the trustworthy, self-determined exchange of data”<sup>45</sup>. The use cases provided on their website and in their documentation furthermore suggest that IDSA focuses especially on large organisational entities with extensive datasets, including historical data from various censuses, demographic or medical records, and transaction logs. This aligns with the background of the majority of IDSA members in business-to-business and business-to-government interactions. Such emphasis also comes along with an implicit assumption that data are present in their original form mainly in organisational enterprise resource planning systems which act as a proxy for the data owners.

In the IDS Reference Architecture Model (IDS-RAM)<sup>46</sup> data are transferred between data providers, who might or might not be the data owner, and data consumers, who request and use the data provided, through IDS Connectors. An IDS Connector aims to provide “standardized [sic] connectivity in the IDS Ecosystem, [is] responsible for [...] Usage Control [and] allow[s] the execution of trusted apps in an isolated Identity Provider environment.”<sup>47</sup>

With their IDS Connector, data providers define who can access which parts of their data and how often and set out the rules about what (specific) data consumers can or cannot do with the data once accessed. Being an essential part of the IDS-RAM, each IDS Connector “is certified against specific security standards and so is any participant in data sharing [creating a] basis of trust [for the] actual exchange of data”<sup>48</sup>.

With data providers (respectively data owners) and data consumers (respectively data users) being the only participants defined in the IDS-RAM, an adaptation to the data space for cultural heritage could see CHIs as data providers and aggregators as data consumers as indicated in figure 6 below. In parallel, aggregators who might create and manage their own data as well (from enrichments to editorials) could take on the role of a data provider as well.

---

<sup>45</sup> IDSA: International Data Spaces - Enabling Data Economy. IDSA brochure, page 9. Online at [https://internationaldataspaces.org/wp-content/uploads/dlm\\_uploads/IDSA-brochure-International-Data-Spaces-Enabling-Data-Economy.pdf](https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-brochure-International-Data-Spaces-Enabling-Data-Economy.pdf) [last accessed: 11 July 2023]

<sup>46</sup> See the latest working version of the IDS-RAM on GitHub at [https://github.com/International-Data-Spaces-Association/IDS-RAM\\_4\\_0](https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0) [last accessed: 11 July 2023]

<sup>47</sup> IDS infographic, online at [https://internationaldataspaces.org/wp-content/uploads/dlm\\_uploads/IDSA-Infografik-English.pdf](https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Infografik-English.pdf) [last accessed: 11 July 2023]

<sup>48</sup> See more about the certification process online at <https://internationaldataspaces.org/offers/certification/> [last accessed: 11 July 2023]

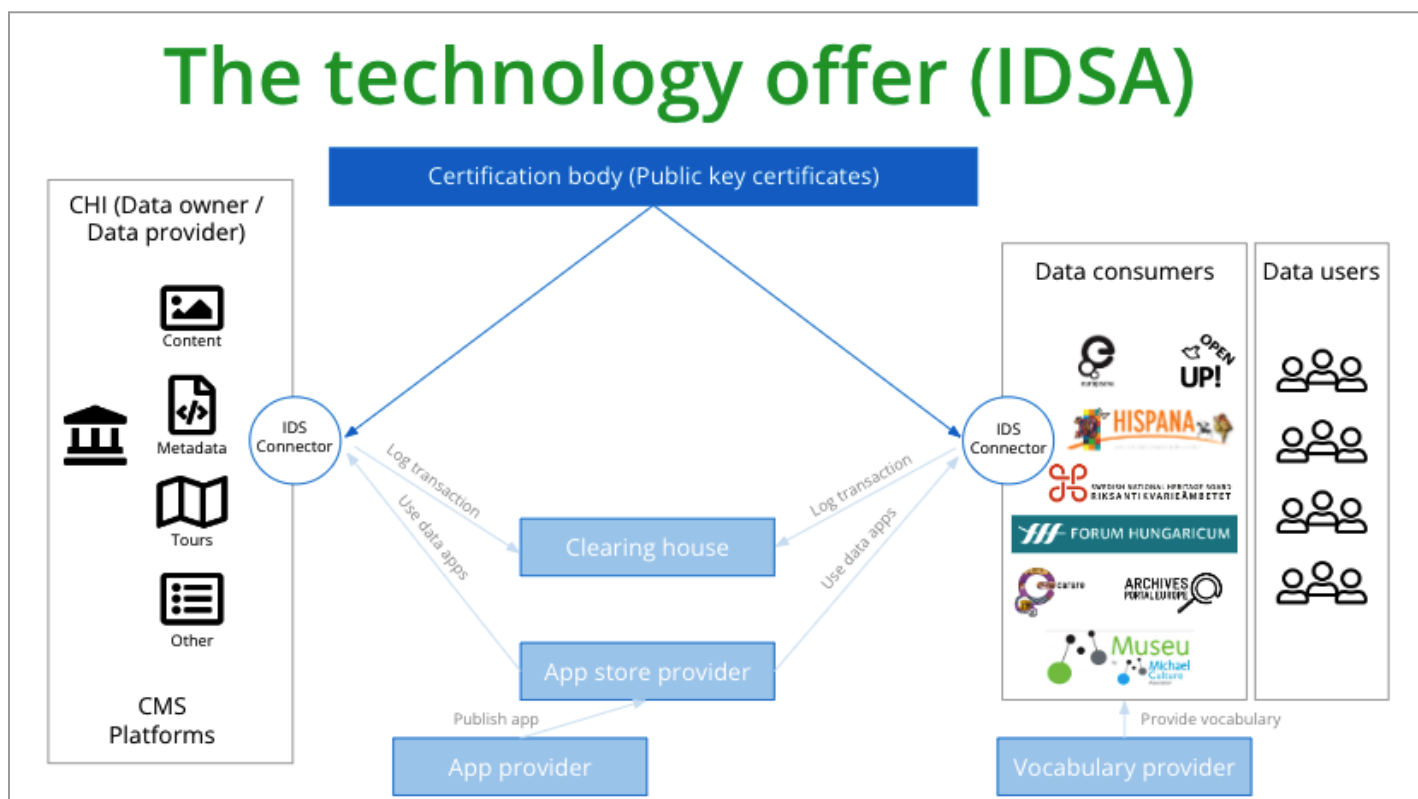


Figure 6: Visualisation of IDSA's technology offer and its potential application to the use case of the DS4CH; here: focus on the data provider on the one hand and the data consumer and the data users on the other hand, connected via their IDS Connectors

Next to the IDS connectors, the certification authority, ensuring that organisations are complying with the IDS standards before they are granted access to the data space ecosystem, and the certification body (see figure 6 above), granting IDSA certification to components and participants within the data space ecosystem, are essential in the IDS-RAM.

Within a limited group of entities, it is feasible to enforce comprehensive certification for every agent within the IDSA ecosystem. As mentioned, the origins of IDSA are in the corporate world where predetermined and contractually agreed-upon data transfers between a limited group of participants are common. Updating authorisation records by data providers becomes a manageable task.



In addition to the essential services<sup>49</sup>, there is a series of supporting services such as:

- Service providers, who are also called intermediaries, operators, or aggregators (see figure 7 below);
  - An aggregator (in the IDS-RAM sense) is meant to provide a service “combining data from multiple sources for computation at one partner (Specialization [sic]: data trustee)”<sup>50</sup>;
  - Note that aggregators as currently established in the Europeana Initiative are not intermediaries in the IDSA context, since intermediaries aim “to establish commercial relationships for data sharing”<sup>51</sup>.
- Clearing House (see figure 7 below), providing clearing and settlement services for financial and data exchange transactions;
- Broker (see figure 7 below), where IDS Connectors register the description of their data endpoints, so that potential data consumers can look up available data sources and data in terms of their content, structure, quality, actuality and other attributes;
- Identity providers, offering services to create, maintain, manage, and validate identity information of and for IDS participants and components;
- Vocabulary providers (see figure 7 below), managing and offering vocabularies which can be used to annotate and describe datasets based on agreed-upon, possibly domain-specific ontologies, reference data models, metadata elements, etc.;
- App stores (see figure 7 below), providing apps to be used in the IDS Connectors, e.g., executing tasks like transformation, aggregation, or analytics on the data.

---

<sup>49</sup> See complete list online at <https://internationaldataspaces.org/adopt/essential-services/> [last accessed: 11 July 2023]

<sup>50</sup> IDSA Rulebook, chapter 2 “Guiding principles”, section “Role models”, see online at [https://docs.internationaldataspaces.org/idsa-rulebook-v2/idsa-rulebook/2\\_guiding\\_principles](https://docs.internationaldataspaces.org/idsa-rulebook-v2/idsa-rulebook/2_guiding_principles) [last accessed: 11 July 2023]

<sup>51</sup> Ibid.

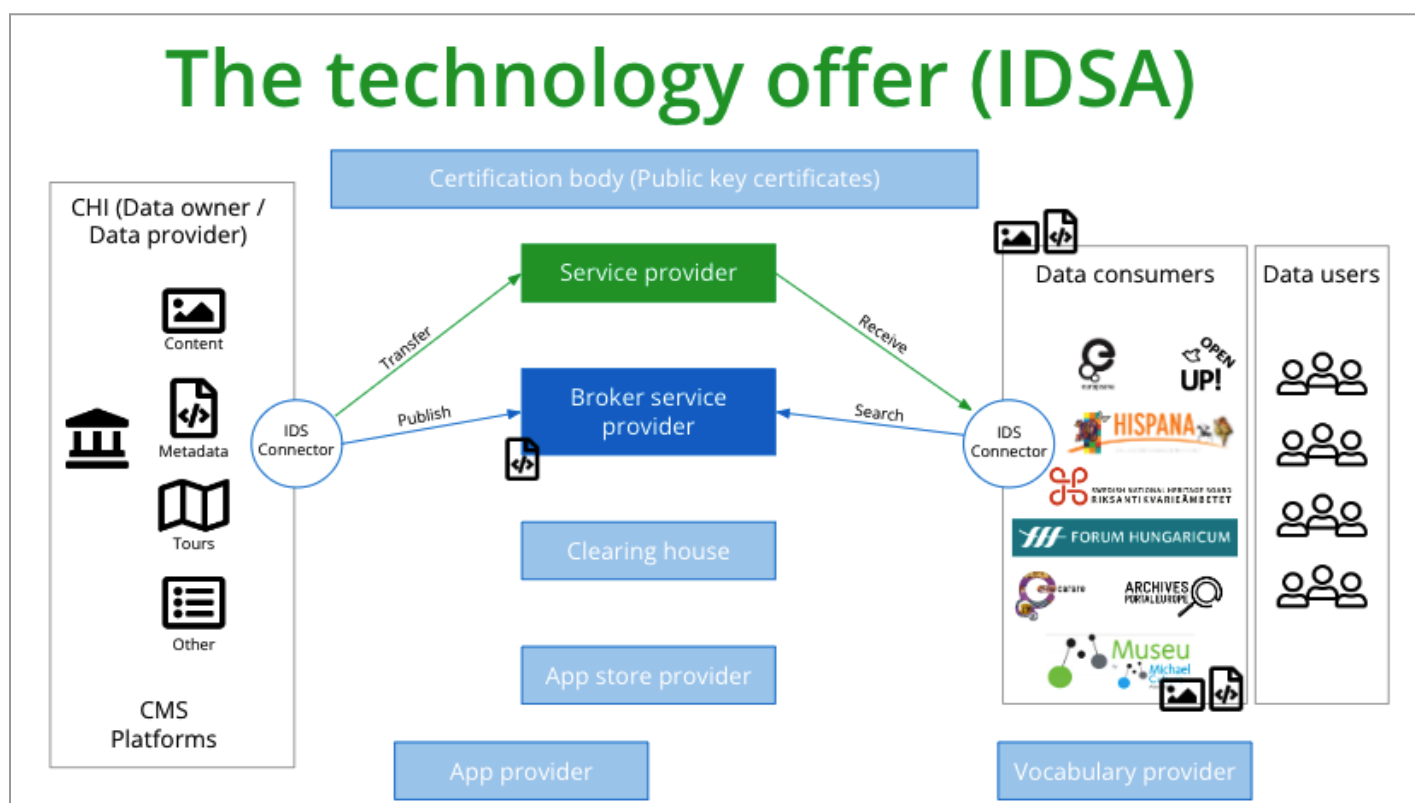


Figure 7: Visualisation of IDSA's technology offer and its potential application to the use case of the DS4CH; here: focus on the some of the additional services which might be of interest to the DS4CH

It remains somewhat unclear how contributions (of various kinds) and scaling are managed (especially, combined) within an IDSA ecosystem. If a data consumer or an intermediary service intends to add new data or enrich the available data, they would need to assume the role of data owner within the ecosystem. Nevertheless, technical protocols for non-owning producers are not provided, although the concept of shared ownership does exist. For example, consider an individual who wants to enhance their medical record with data from their wearable health tracker or privately obtained scan results. It is unclear how IDSA accommodates such scenarios.

The interoperability of different agents and roles within the IDSA ecosystem is both guaranteed and limited by the defined protocols. Data compatibility is derived from the use of ontologies and metadata accessible to all agents.

In summary, IDSA is designed to effectively address data flow considerations within limited-size networks of highly regulated agents. However, the scalability potential of this approach remains an open question.

### 6.1.2. Gaia-X

Gaia-X and IDSA share the common goals of data sovereignty, interoperability, and data sharing built on mutual trust between all participants of data exchange. Gaia-X' approach is aligned with the overall concepts of the IDS-RAM, but it adds further concepts regarding

“data storage and cloud-related elements, which can complement the IDS architecture”<sup>52</sup>. Gaia-X follows the principles of principles of:

- security-by-design;
- privacy-by-design;
- enabling federation, distribution and decentralisation;
- user-friendliness and simplicity;
- machine-processability;
- semantic representation.

Its data space ecosystem consists of an infrastructure ecosystem and a data ecosystem, which work independently from each other but cannot be viewed separately. Both ecosystems are connected via federation services. These can be infrastructure services or operational support functions, like Identity and Trust Services, Compliance Services, Federated Catalogues and Sovereign Data Exchange Services.

Same as IDSA, Gaia-X works with providers and consumers as the main roles for participants in the data space, though Gaia-X explicitly foresees a participant taking on both roles. Participants in the Gaia-X context are mainly understood as business organisations, another similarity to the IDSA approach. Where IDSA uses certification, Gaia-X uses a mandatory self-description for each participant and asset within the data space. This self-description includes some elements which are required depending on the type of participant or asset being described and which will need to be confirmed by signature by trusted parties, before a participant or an asset is granted access to the data space.

Gaia-X' main types of assets are:

- nodes, a general term for any computational resource or infrastructure element placed in the infrastructure ecosystem;
- services, to be understood as any cloud service offered by a service provider within the data space; services can be combined with each other, and they are - in general - portable from one node to another;
- service instances, meaning a service executed on one or more specific nodes;
- data assets, or simply datasets, which are made available or hosted on nodes and provided to consumers via specific services; in theory, a data asset could also be hosted privately; same as services, data assets can be combined with each other.

With the organisational connections between IDSA and Gaia-X (see [chapter 3.2.2.](#)) and the already mentioned commonalities of both initiatives' approaches to data spaces, further alignment between both is likely to be expected in the future.

## 6.2. Data-centric approaches

### 6.2.1. Solid

Solid stands for Social Linked Data and is a specification that allows individuals to securely store their data in decentralised data stores called pods. Pods function as “personal online data stores”. Users have complete control over their pods, including deciding what data to

---

<sup>52</sup> IDSA: Position paper. Gaia-X and IDS. Version 1.0 from January 2021, page 4, online at [https://internationaldataspaces.org/wp-content/uploads/dlm\\_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf](https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf) [last accessed: 11 July 2023]

share, with whom, and for which purposes (including potential read-and-write access). Access to data can be granted or revoked at any time. Applications interact with pods using standard, open, and interoperable data formats and the Solid protocol<sup>53</sup>. Solid builds on existing standards rather than inventing anything new. Contrary to IDSA, however, Solid starts from a personal - and only as a secondary step an organisational - perspective on data sharing.

Solid enables storing various types of data, with a focus on supporting interoperability through Linked Data, thereby allowing different applications to work with the same data. With its decentralised approach, Solid allows for the possibility of anyone being a data provider and a data consumer, to stick with the terminology introduced above. It does not see these categories as separate from each other as IDSA does. Thereby, Solid - in its conceptual state - seems to enable more of a web of relationships and interactions between the various players in a data space.

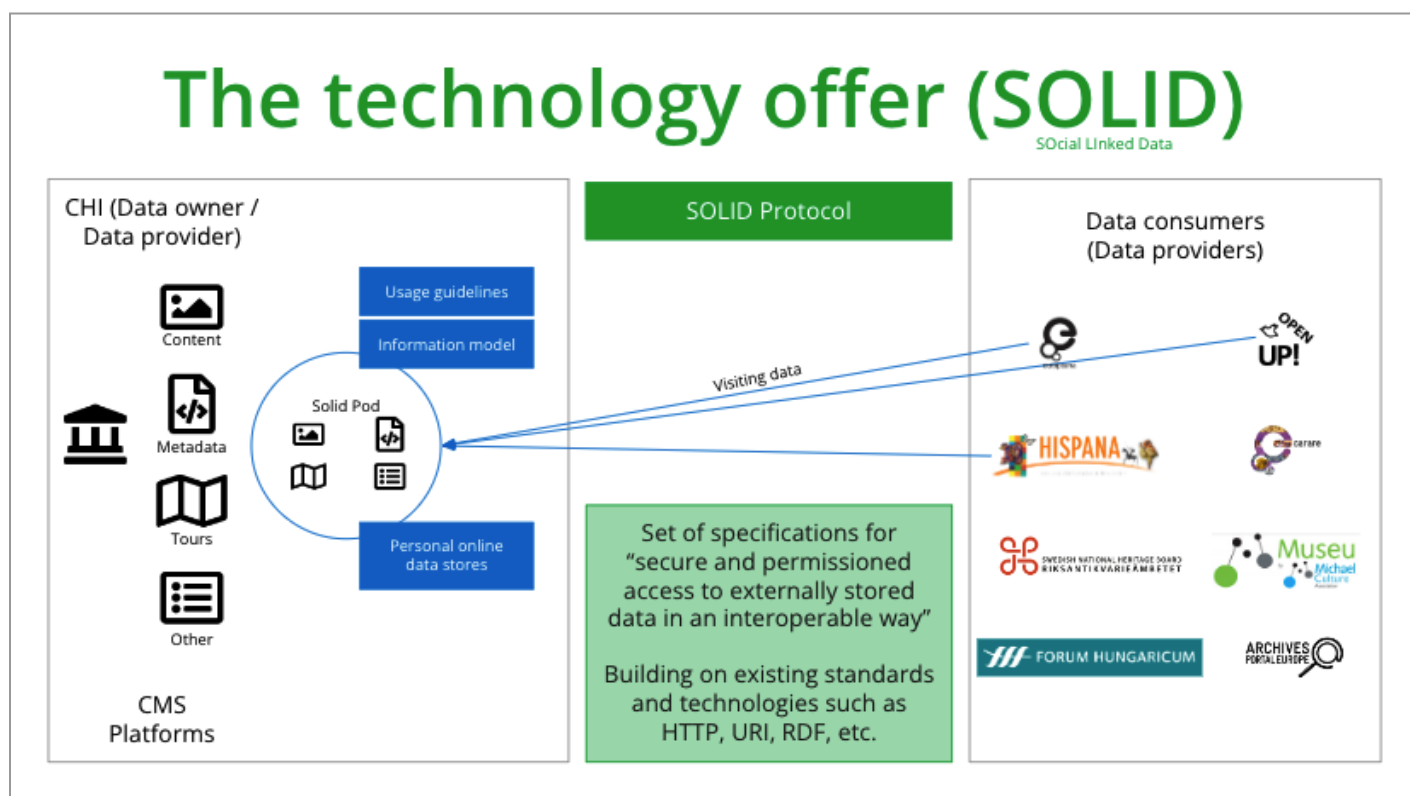


Figure 8: Visualisation of Solid's technology offer and its potential application to the use case of the DS4CH; here: focus on a data provider's pod being visited by several data consumers

Users (i.e., data providers and data consumers) can obtain pods from pod providers or choose to self-host them, which requires running a pod server. Currently, pod providers mainly come from the technology sector, but one could imagine that in the future providers of e.g., collection management systems might become pod providers, too, offering pods as an additional or an alternative service to their customers.

<sup>53</sup> See more online at <https://solidproject.org/TR/protocol> [last accessed: 11 July 2023]

A user can have multiple pods in place, hosted by the same or by different providers, and/or by themselves. Similarly, a pod server hosts one or more pods, with each pod being controlled by its owner and the data and access rules for each pod being separate from those of other pods.

Solid applications utilise the Solid protocol to store and retrieve data from pods. Thereby the data remain in the pods, i.e., at source, while they are visited by whoever has been granted the relevant access rights. This promotes efficiency and eliminates the need to provide redundant information to different services.

Each user and each application in the Solid ecosystem are identified by a WebID, which references an RDF document describing the agent that is being identified. Furthermore, Solid applies the Solid OpenID Connect (Solid OIDC) specification to authenticate the identity of each user and each application involved.

While Solid has existed for several years, its usage during this time has been primarily limited to research and early-stage innovation cases. The scalability of the Solid approach has been traditionally questioned due to its reduced ability to suggest a practical alternative to existing data ecosystems, which could represent data silos in the worst case, or multiple limited internally interoperable segments in other cases.

However, the project and its approach have gained increasing popularity again during the past two years. This popularity can be seen in both the business environment, with the funding of several start-ups and commercial projects based on Solid, and in the broader European conversation surrounding personal data. Solid is often informally associated with the concept of personal data spaces, which draws inspiration from the common European data spaces, although it should be noted that there is no universally agreed-upon taxonomy for defining common data spaces and personal data spaces. This lack of consensus has caused confusion for many.

#### *Solid use case: Solid-CRS (NDE)*

The project Solid-CRS<sup>54</sup>, conducted by the Dutch Digital Heritage Network (Netwerk Digitaal Erfgoed, NDE), showcases one of the most extensive applications of Solid in the cultural heritage field so far. The NDE is a collaboration between cultural heritage institutions in the Netherlands, and one of their core strategies is the implementation of Linked Data and semantic technologies to enhance standardisation, data integration, and reuse. While larger institutions and their IT vendors have been working with NDE to enable data sharing and publication within the network, there are also numerous small organisations and individuals contributing digital information. To facilitate the interoperable sharing of data from these smaller entities, NDE initiated a project to explore the feasibility of a Solid-based collection registration system (Solid-CRS) with basic functionality<sup>55</sup>. The Solid-CRS application was developed by the NDE team in partnership with Digita<sup>56</sup>.

---

<sup>54</sup> See more on GitHub online at <https://github.com/netwerk-digitaal-erfgoed/solid-crs> [last accessed: 11 July 2023]

<sup>55</sup> Test environment available online at <https://solid-crs.netwerkdigitaal.erfgoed.nl/login>, login required [last accessed: 11 July 2023]

<sup>56</sup> See online at <https://www.digita.ai/> [last accessed: 11 July 2023]

### 6.2.2. Post-Platforms

Post-Platforms is an architectural approach adding to the concept of Solid pods the idea that platforms - in this case understood broadly as apps, services, and essentially any application using data - should be redesigned to be working with data stored on a variety of Solid pods instead of these platforms storing (copies of) data internally. Implementing such changes transforms platforms into “post-platforms”, or “platforms without data”.

The architecture of Post-Platforms comprises several components. Pods and the Solid protocol keeping data about people, organisations, and objects in Linked Data format and accessible through the Solid protocol, are inherited from Solid. Platforms evolve into post-platforms accessing data from a distributed ecosystem of pods storing data. These platforms continue to act as service providers, like traditional platforms, but they read and write their data directly on users' pods, provided they have the relevant access permissions.

Furthermore, Post-Platforms operates on a security-by-design basis, same as Gaia-X, but utilising a global Public Key Infrastructure (PKI), which provides users with keys, certificates, cryptography, and secure data transmission protocols, and a register similar to the domain name system (DNS) establishing a list of all available pods within the data space environment, ensuring their discoverability and indexing by semantic search engines. The register thereby acts as a directory for efficient location of pods.

The parallel to the DNS also affects the maintenance of this Post-Platforms register, which could potentially be run by a public non-profit as an independent body within the ecosystem.

Incorporating Solid, Linked Data plays a critical role also in the architecture of Post-Platforms and in its vision of enabling scalability, machine-readability, and enhanced interoperability of platforms. Linked Data is seen as an enabler that allows platforms to interact effectively with the pods through semantic reasoning<sup>57</sup>.

## 6.3. Other approaches

### 6.3.1. Blockchain

Blockchain-based projects utilise blockchain, or distributed ledger, technology to safeguard data by protecting them from tampering and by establishing a decentralised way of organising trust. Blockchain serves as a database maintained by multiple parties, recording transactional data, though not aiming at storing larger amounts of data. It allows anyone to verify and add to the database but prevents unauthorised changes. Unique fingerprints for documents are created upon their arrival in the database or digital archive and are stored immutably in the blockchain, and even as the file format might change over time, the fingerprint remains consistent.

Blockchain-based projects posit that blockchain is preferred over a centralised database managed by a single actor because this choice aligns with addressing the shifting basis of public trust. No actor within a blockchain-based context is considered more trustworthy than another and all are identified and authenticated via private and public keys. However, blockchain does not work with a traditional PKI, thereby creating opportunity for security

---

<sup>57</sup> Cf. section “Linked Data” on Post-Platforms website under Technology, online at <https://postplatforms.org/technology> [last accessed: 11 July 2023]

breaches such as manipulator-in-the-middle attacks<sup>58</sup> on public keys. Similarly, actors are presented with the need to keep their one-and-only private key throughout their life, as the loss of the private key automatically would ruin all their work.

Adapted to the cultural heritage sector, one could say that the authoritative power and institutional legitimacy of cultural heritage institutions, which was seen as a given in the past, but is increasingly questioned nowadays, is legitimised by establishing and confirming it via a network of various actors.

According to this vision, cultural heritage institutions might generate and register digital fingerprints of documents within, for example, a permissioned blockchain. Such a blockchain would be collaboratively maintained by its participants, even across international borders, and access would be managed via an extra access control layer. At any given time, documents could be verified against their original signatures, including during their release, guaranteeing the integrity of the records.

The global network of actors helps protect against accidental or intentional data corruption, allowing all participants worldwide to verify and safeguard each other's data. With this, participating cultural heritage institutions could (re)establish indefinite trust, not only for research and reference but also for critical matters such as legal cases, official investigations, and holding organisations and governments accountable.

### 6.3.2. International Image Interoperability Framework (IIIF)

The International Image Interoperability Framework (IIIF)<sup>59</sup> is a set of open standards for delivering high-quality, attributed digitised and digital representations online at scale. It is backed by a consortium of leading cultural institutions. First launched by the Stanford University, IIIF has gathered a strong international community of cultural heritage institutions, developers and vendors developing and implementing the six IIIF APIs.

With its approach to align with general web standards, IIIF has established a consistent way of sharing media files once and making them available in various contexts, across different types of viewers, and across institutional boundaries.

Everything needed to display a digitised or digital representation of an object is in its IIIF Image API resource URL, which can be addressed from anywhere and in any context, with a dedicated IIIF Viewer or without. On the other hand, while the IIIF Presentation API allows for the provision of some basic metadata and structure to digitised or digital representations and while it is possible to reference structured metadata from the IIIF manifest, metadata in itself is not necessarily in IIIF's focus.

With this, IIIF is not a direct fit with this Task Force's investigations, which concentrate on the metadata first with the links to digitised and digital representations being part of those. However, specifications such as the IIIF Authentication API to manage access and the IIIF Change Discovery API to manage new publications and updates, together with having structured metadata referenced in the IIIF manifest, might make the IIIF environment interesting to consider in future elaborations, especially once the IIIF Change Discovery API is adopted more widely. For IIIF data sources that do not yet implement the IIIF Change

<sup>58</sup> See explanation online at [https://owasp.org/www-community/attacks/Manipulator-in-the-middle\\_attack#](https://owasp.org/www-community/attacks/Manipulator-in-the-middle_attack#) [last accessed: 11 July 2023]

<sup>59</sup> See more online at <https://iiif.io/> [last accessed: 11 July 2023]

Discovery API, it is still possible to combine IIIF standards and protocols with other ones such as Linked Data standards or the OAI-PMH protocol making it a relevant initiative to follow.

### 6.3.3. InterPlanetary File System (IPFS)

IPFS, short for InterPlanetary File System, is an open-source protocol and network that introduces decentralisation into file storage and retrieval. IPFS is the underlying technology of the activities of the Filecoin Foundation<sup>60</sup> and its not-for-profit branch, the Filecoin Foundation for the Decentralized Web (FFDW)<sup>61</sup>. Its approach utilises unique content-based hashes to address and retrieve files, eliminating the reliance on centralised servers.

Files are divided into smaller chunks, distributed across the network, and accessed based on their cryptographic hashes. The latter remain findable in a given environment, whether that is a concrete storage environment or the internet, even if the address or URL of a resource has changed. This decentralised nature avoids redundancy and ensures availability and resistance to censorship. IPFS supports features like content addressing, peer-to-peer communication, versioning, and integrity checks.

Filecoin Foundation creates the incentive for everyone involved in their ecosystem with the currency of filecoins. Storage providers get them from making deals to store the data; developers building tools on top of the storage providers get a cut when they make storing data easier; data providers and consumers use filecoin as an investment.

As such, IPFS is close to the concept of bitcoin and the implementation of blockchain in the context of cryptocurrencies. On the other hand, there also are relations to members of the Solid community. However, with IPFS focusing on decentralised data storage, it might rather be understood as a basis to build on and to, for example, be extended by the application of Solid on top of it - provided a transformation between IPFS' Content Identifiers<sup>62</sup> and Solid's Linked Data URI-s can be established as a first step.

## 6.4. Summarising overview of technology offers

While platform-centric and data-centric approaches aim to transform data sharing paradigms within existing power structures, blockchain-based approaches seek to fundamentally redesign relationships and trust in the digital and in the real world. They propose replacing trust based on traditional authorities with trust based on technology and automation. However, it is now widely recognised that blockchain applications have limitations and may not have an impact as broad as initially expected. They are most suitable for specific contexts where technology can mediate between parties, but they face challenges in terms of performance, scalability, and privacy.

Platform-centric approaches offer an alternative for data sharing in controlled environments with limited participants. They follow a proven reference architecture that has been implemented in production settings. Data-centric approaches on the other hand require a more fundamental redesign of data architecture, treating data as a primary entity on the web and developing new systems for its storage and management. While both approaches

---

<sup>60</sup> See online at <https://fil.org/> [last accessed: 11 July 2023]

<sup>61</sup> See online at <https://ffdweb.org/> [last accessed: 11 July 2023]

<sup>62</sup> See more online at <https://docs.ipfs.tech/concepts/content-addressing/> [last accessed: 11 July 2023]



include the concept of data sovereignty, data-centric approaches prioritise the autonomy and control of data also independent of a data owner or provider, emphasising the data's significance in the digital landscape.

When comparing the three main approaches detailed in the previous chapters - IDSA, Solid, and Post-Platforms - based on six essential aspects, the following can be stated:

	IDSA	Solid	Post-Platforms
Actors / nodes	Mostly platforms	Persons and organisations	Persons, organisations, devices
Data Location	Different copies in various locations	Original on platform / CMS and pods	Single original on pods
Connection	Platform-2-Platform	App-2-pod Platform-2-pod	Platform-2-Pod
Major Architecture Elements	IDS Connectors, Clearing House, Broker, etc.	Pods Apps for Browsers	Pods Registry, Ontology Provider
Security	Certification body, Certification authority	WebID Single-sign-on	Cryptographic keys PKI hierarchy
Technology Readiness Level Strategy	? Focus on network	Assume TRL4 Focus on R&D	? First implementation use cases

Figure 9: Comparison between the approaches of IDSA, Solid and Post-Platforms

- **Actors and/or nodes:** while IDSA mainly looks at platforms or systems in proxy for a data provider or data consumer, Solid and Post-Platforms aim more specifically at persons and organisations, with Post-Platforms adding platforms in a redefined way;
- **Connection:** subsequently, IDSA establishes connections between platforms, while Solid connects apps and platforms to personal online data stores; Post-Platforms subsumes apps under the broader concept of (post-)platforms in its context;
- **Data location:** this relates to the question of where the original data are held; while IDSA's architecture allows for different copies of the data being held in various locations, Solid and Post-Platforms see the only (and original) version of the data in the Pod, where the data are accessed directly by others in the ecosystem who have the respective access rights;
- **Major architecture elements:** both, IDSA and Post-Platforms, have a more architectural approach and include not only the IDS Connectors respectively pods, but also e.g., the Broker respectively the Registry and a Vocabulary Provider respectively an Ontology Provider; Solid in its initial form on the other hand is mainly represented by a set of specifications and protocols;

- **Security:** while all three approaches address security in terms of identification and authentication of all actors and assets, there are varying degrees of enforcement from the more organisational approach in IDSA via the use of WebID and single-sign-on in Solid to the use of cryptographic keys and a PKI in Post-Platforms;
- **Technology Readiness Level (TRL) and Focus:** in terms of readiness, Solid is probably the most advanced, even though all three approaches present several use cases; however, none of them have fully proven their implementability yet; each initiative's focus reflects its main aims and background.

## 7. Analysis

By bringing the use cases of the current aggregation workflow together with the possibilities offered by decentralised approaches as described in the two previous sections, this current chapter considers some of the general themes identified in the discussion around decentralisation and how a more decentralised approach might help solving some of the current challenges.

### 7.1. General considerations

Technologies and organisational models of decentralised networks have been the subject of research and practise for many years. This research has been motivated by various use cases where a centralised resource within an ecosystem became a hindrance to its growth or democratic development. These situations do not always involve traditional market monopolies. Sometimes even non-profit entities, serving the greater good but in control of a centralised resource, have acted to maintain their monopolistic position in the ecosystem against the direction of progress. It seems to be inherent in networks where members with a strong position control access to a crucial resource and tend to reinforce this position at the expense of the common interest.

Another recurring factor is the impossibility of achieving a completely decentralised solution. As a member of the decentralised research community eloquently put it: "If you push centralisation through the door, it gets back through the window." This implies that some level of centralisation is inevitably required to decentralise other components of the environment in which we operate. It is worth noting that even the internet, often considered a successful example of a decentralised network - and despite its creator's contrary position on the matter - has a centralised component: the DNS, or more specifically the Internet Assigned Numbers Authority (IANA), which is maintained and administered centrally by the Internet Corporation for Assigned Names and Numbers (ICANN). This interplay between centralisation and decentralisation means that any solution will (have to) involve a combination of both elements, and decisions should be made based on a practical evaluation of the pros and cons of each approach, rather than out of unrealistic belief in a purely decentralised architecture.

Furthermore, various researchers and practitioners of decentralisation emphasise that technology is just one aspect of decentralisation. Organisational and governance aspects are equally important. Perceived centralisation in an ecosystem is not solely a result of lacking decentralised technologies but also a symptom of the participating parties lacking the necessary motivation and power to create and sustain the organisational changes needed for a more decentralised design. Technologists alone do not hold the key to decentralisation; software design must be complemented by ecosystem design.

In this regard, it is useful to view centralisation and decentralisation as dynamic characteristics of an ecosystem that change over time. They reflect the balance of power among participating entities and the availability and maturity of technological solutions capable of supporting various architectural designs of the network. Additionally, it is essential to consider the various alternatives available in the market, as mentioned in this report, in the context of historical changes as reactions to a status-quo and specific conditions imposed by different players, such as business platforms or central regulators or funders.

Last, the balance between decentralised and centralised aspects of an ecosystem will also have to be looked at from a cost-saving perspective and from the perspective of digital sobriety: where can decentralisation result in reducing our carbon footprint in the digital environment we operate in and under which conditions might a central service still provide the ecologically more suitable solution? How does, for example, the expected positive impact on the digital carbon footprint of cutting back on - unnecessarily - copying data from one place to another and yet another in a decentralised approach to aggregation compare to the possibly negative impact on the digital carbon footprint that comes with the need of deploying and maintaining a central service such as a Broker or a registry to identify the immense number of decentralised data sources?

## 7.2. Blockchain: the appeal of technology-based trust

In this context, it is natural to recognise the appeal of blockchain-based solutions that aim to replace trust relationships based on existing societal institutions and conventions with trust based on technology. Blockchain enthusiasts paint a bright picture of a future where expensive and sometimes ineffective facilitation of trust-based economies by institutions such as banks or courts, but also cultural heritage institutions is replaced by a decentralised ledger distributed across multiple nodes.

While a detailed analysis of blockchain technology is beyond the scope of this report, interested readers are encouraged to explore the abundant online resources available on the topic. However, it is important to stress that contrary to the desired perception blockchain-based solutions are not completely devoid of centralised elements either. For instance, one commonly raised critique of blockchain is that, while it is distributed across many nodes, the ledger itself is centralised, leading to infamous performance and scalability issues.

Another example pertains to the identification of parties within the blockchain system. Since trust is not based on real-world identities, parties are identified through a cryptographic key pair. However, the loss of a private key can result in the party's inability to regain ownership of their assets, which seems to be an unrealistic design for real-world scenarios. Finding a solution to this challenge would inevitably involve a traditional, centralised authentication party.

## 7.3. IDSA vs. Solid: platform-centric vs. data-centric views

While blockchain-based solutions can be broadly categorised as attempts to address the challenge of - in its perception - oftentimes financial interactions requiring the parties involved to trust each other even if they have never interconnected with each other before, other approaches outlined in this report stem from various research and related communities and from business contexts, resulting in diverse perspectives on the problem at hand.

IDSA, which emerged nearly a decade ago, originated from the requirements of large European manufacturers seeking to manage and exchange data stored in various repositories in a trusted manner. The necessity to repeatedly scale ad-hoc projects for data exchange led to the development of a reference architecture that outlines the key elements to be implemented in any project. The IDS-based or IDS-related approaches adopt a platform-centric view, focusing on the need to exchange data between big(ger)

organisations. The strength of such approaches lies in their practical roots and evolution through implemented projects. However, the scalability of these approaches to larger ecosystems, including a diverse group of small(er) players and individuals, and the realistic performance and cost considerations remain open questions.

On the other hand, Solid, which began around the same time, started as a research project with a distinct data-centric perspective, free from the need to implement a specific business environment or incorporate real-world platforms or applications. Solid has been developed over several years in a more flexible and theoretical manner, offering an alternative vision to today's data architecture. Its proponents have mainly experimented with various aspects of artificially constructed ecosystems. The attractiveness of Solid lies in its long-term data-centric vision, but its weakness lies in the circumstance that the major players in the current digital economy, such as platforms and apps, are not part of the Solid ecosystem at all - at least not in its original conceptualisation.

IDSA and Solid thereby seem to form a complementary duo, with the former viewing the world through the lens of platforms and applications ("algorithms") that require data connectivity but do not prioritise data as a first-class citizen, while the latter views the world through the lens of data but does not encompass platforms within its framework.

Post-Platforms, as a slightly more recent initiative, incorporates the concepts of Solid regarding its data-centric perspective, but also takes platforms into account again, though decoupling them from the data, which currently establish their monopoly. This approach goes along with the application of a likewise distributed public key infrastructure, aiming at a scalable security solution. A member of IDSA, the Post-Platforms foundation actively seeks to consolidate and collaborate with the other two approaches, most prominently resulting in the establishment of a personal data task force within IDS. However, the challenge remains that Post-Platforms' approach currently represents the most comprehensive change requiring a complete transformation of existing workflows and processes, which will only be possible with the availability of significant resources for the transition and advocacy process.

#### 7.4. To copy or not to copy - or: sharing data vs. visiting data

The accounts provided highlight a crucial distinction between the different approaches, focusing on the location of data and the mechanisms employed to ensure necessary workflows. The IDS-family approaches establish infrastructures to maintain data sovereignty and control through data replication across multiple repositories via the IDS Connectors. On the other hand, the Solid-family approaches utilise Solid pods as the primary source of data, facilitating diverse workflows and access models without data replication. In these Linked Data scenarios, the data owners retain full control over who can access their (personal) data and with which rights, while sharing them in a fully interoperable way. Post-Platforms explicitly redesign modern platforms to integrate with pods, while pure Solid approaches offer developers of Solid apps more flexibility in making that decision. Blockchain-based approaches fundamentally re-evaluate the concepts of location and jurisdiction by utilising distributed ledgers to store records or, at the very least, their fingerprints. It is important for communities of practice to recognise these fundamental differences between the approaches, as many communication materials tend to mix specific data architecture principles with general ideas about data ecosystems. A helpful distinction can be made when thinking of it as the difference between "data sharing" and "data visiting," where the first

follows the notion of giving (parts - or copies - of) one's data to others, while the latter assumes the idea of going to the location where the data are to view them.

### 7.5. The applicability of it all: use cases revisited

With all of the above to consider, there are still a range of aspects to explore in more detail to be in a position for a more encompassing analysis of how well the different technology offers might serve the use cases currently identified (see also the recommendation to explore further in [section 8](#)).

Apart from deciding for or against a specific approach this will also have to include a general review of the services we might want to offer to participants of the future DS4CH as a first step, so that considerations of the best data (work)flows to support these services can follow in the next stages. Similarly, such an evaluation will have to look at options pertaining to metadata on the one hand and to digitised and digital representations on the other hand as well as potentially considering types of data, which might currently not be present in the ecosystem of the Europeana Initiative but could be in the DS4CH.

This being said, this last chapter of the current section will look again at the two use cases illustrated previously (see chapters [5.2.2](#) and [5.2.3](#)). It should be noted that the term "aggregator" is still used more in its current sense, even though the authors of this report acknowledge that this is likely to shift in a decentralised environment.

Furthermore, the below visualisations only explore one possible route of addressing these use cases, while others could provide solutions in a similar way, but would require further explorations. This could include more detailed investigations in the current work on Linked Data Event Streams<sup>63</sup>, Event Notifications<sup>64</sup>, or Activity Streams<sup>65</sup> among others.

---

<sup>63</sup> See more online at <https://semiceu.github.io/LinkedDataEventStreams/> [last accessed: 11 July 2023]

<sup>64</sup> See more online at <https://www.eventnotifications.net/> [last accessed: 11 July 2023]

<sup>65</sup> See more online at <https://www.w3.org/TR/activitystreams-core/> [last accessed: 11 July 2023]

### 7.5.1. How a decentralised approach could help in UC-2

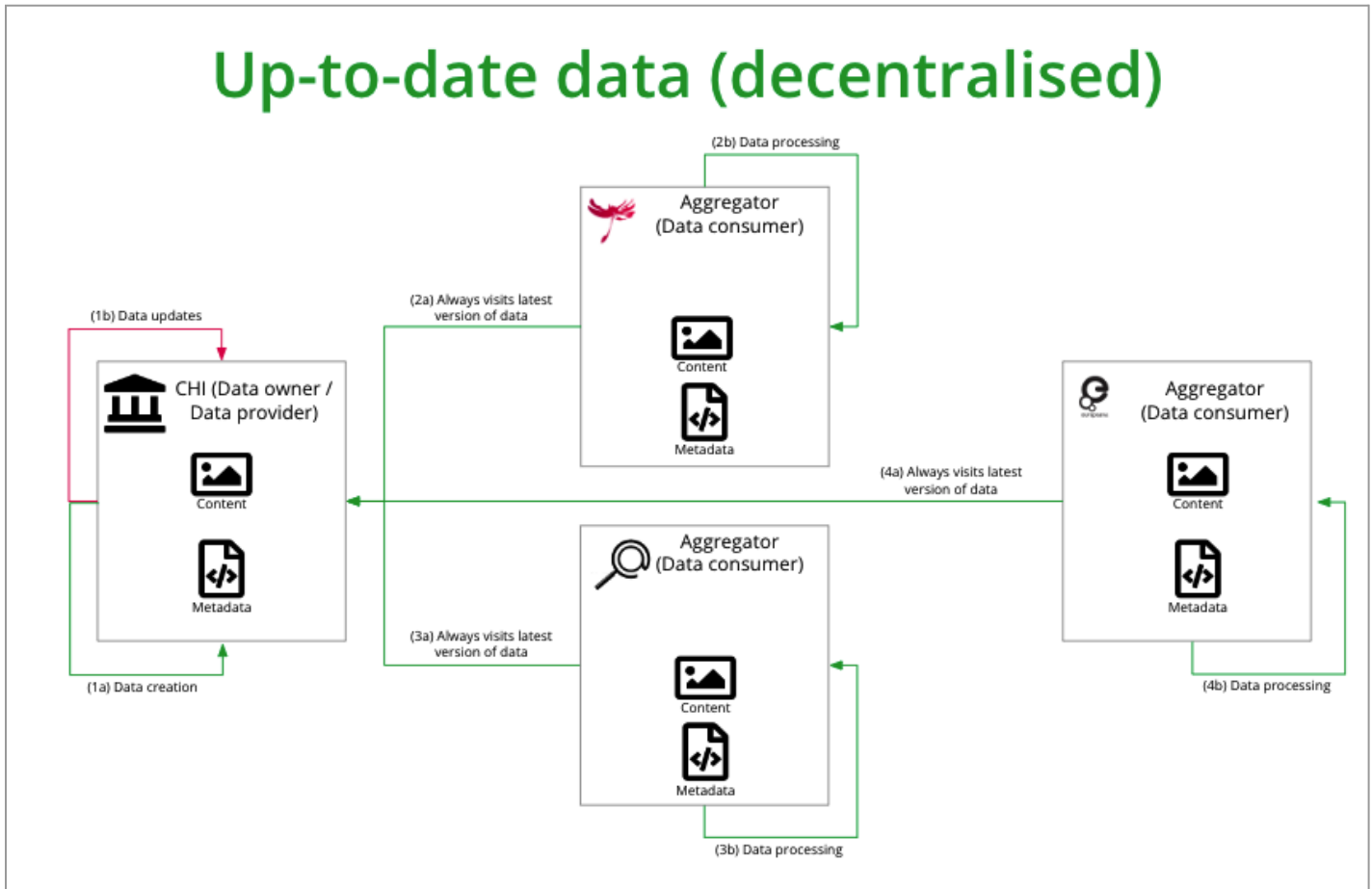


Figure 10: Visualisation for UC-2 Always provide access to latest version; here: prospective workflow in a decentralised approach with the focus on how data from a CHI would be made available in several aggregators' contexts

Figure 10 above shows a case where a CHI works in its own environment in terms of creating (action 1a) and updating (action 1b) data and then makes all or parts of these data available via several aggregators. In a decentralised approach, it would be the aggregators that visit the CHI's data regularly and thereby always get the latest version (actions 2a, 3a, and 4a). Each aggregator might then do some data processing of their own (actions 2b, 3b, and 4b) to present the data in their own contexts.

Figure 11 below builds on this and adds further interactions between all actors that materialise more easily in a decentralised approach. It might, e.g., be the case that the CHI regularly visits the data as processed by one of the aggregators (action 1c) to benefit from any data enrichment that the aggregator might have applied. This could cover only the CHI's own data in an enriched version, but could also include data from other CHIs, provided the licensing and sharing options allow for such transactions. Similarly, the three aggregators in this example might visit each other's data pool (actions 2c and 2d, 3c and 3d, and 4c and 4d) in addition to getting the source data from the CHI. With this, the aggregators also become data owners or providers and are not only data consumers.

## Up-to-date data (decentralised, extended)

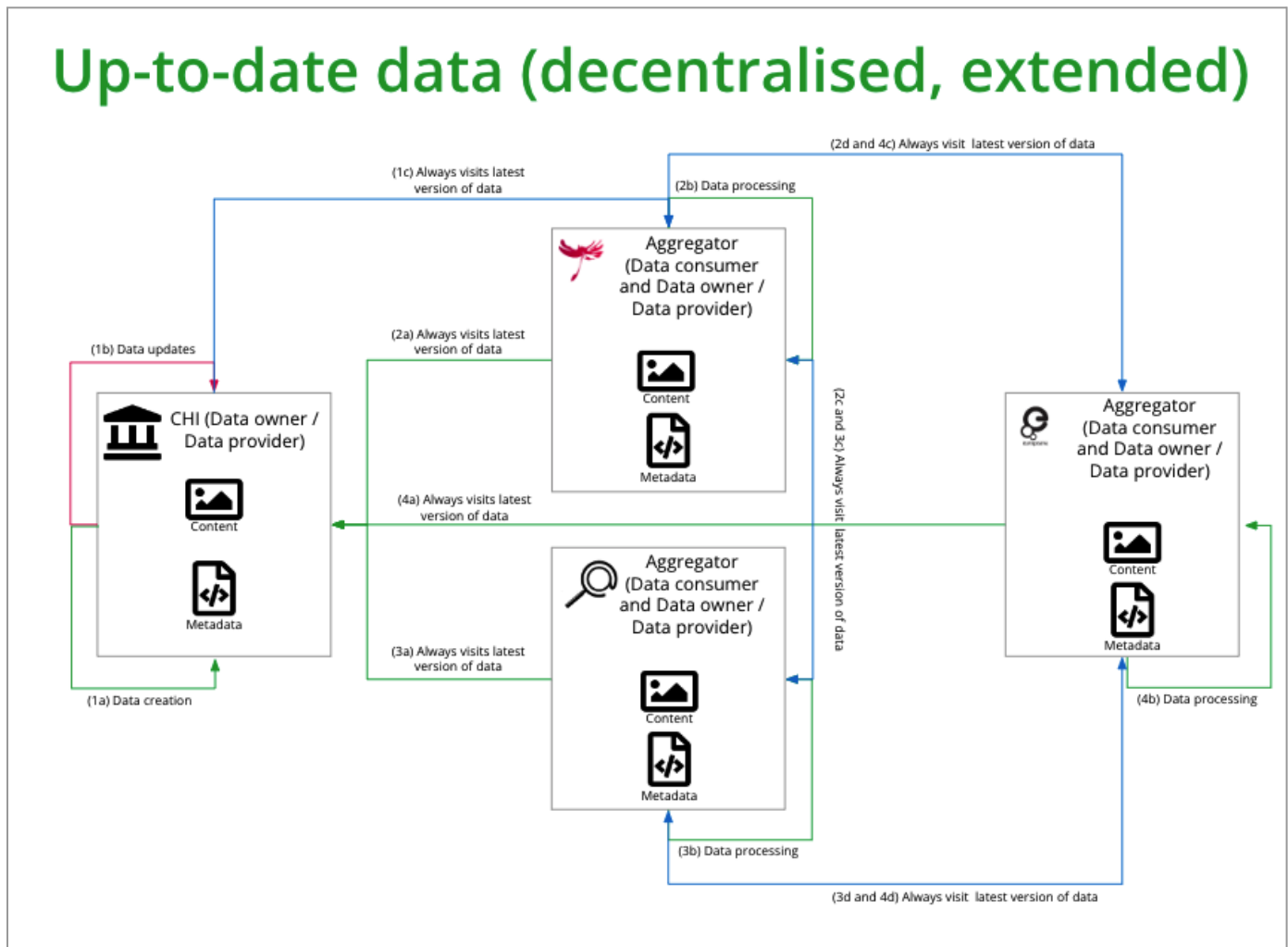


Figure 11: Visualisation for UC-2 Always provide access to latest version; here: prospective workflow in a decentralised approach including interactions between all actors

However: for relations to be kept, all entities in the data will need to have persistent identifiers (PIDs) available at CHI- and at aggregator-level. With this comes not only the question of whether such PIDs would be created and assigned centrally or decentrally, but also the question of which entities - on the data side - will (need to) be identified, e.g., whether we would mainly be looking at identifying the end points, the individual objects and their digitised or digital representations, or the collections and datasets as a whole as well.



### 7.5.2. How a decentralised approach could help in UC-4

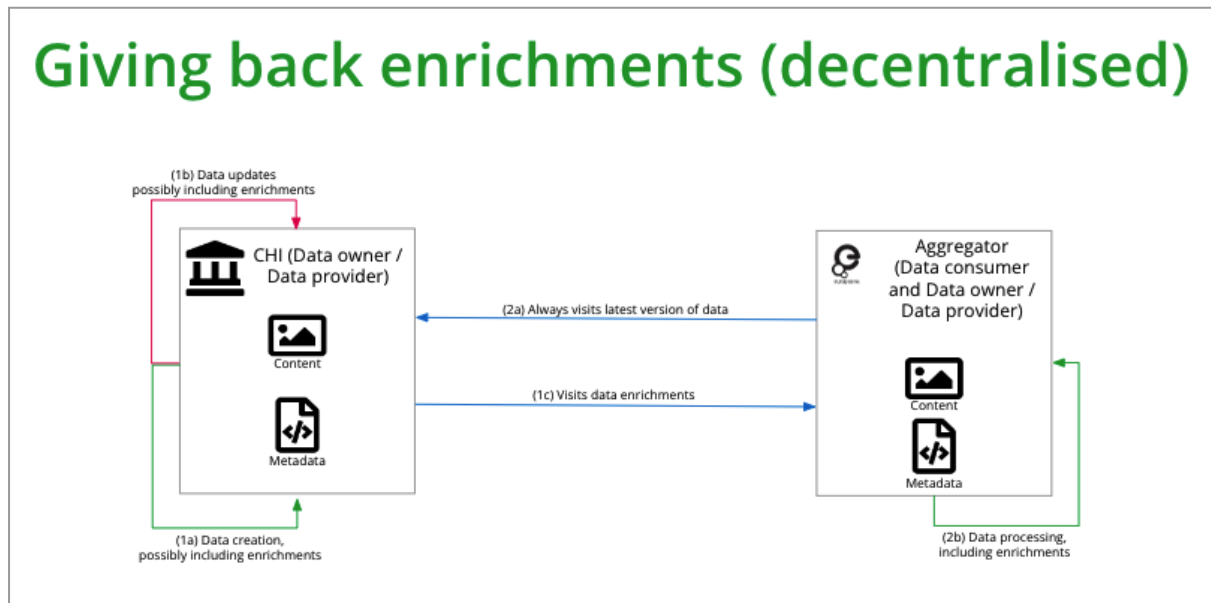


Figure 12: Visualisation for UC-4 Give back enrichments; here: prospective workflow in a decentralised approach

Figure 12 above shows a case where a CHI works in its own environment in terms of creating (action 1a) and updating (action 1b) data, which might or might not include data enrichments. The CHI's data is then visited by an aggregator (action 2a) to be surfaced - together with the data of others - in the aggregator's context. For this, the aggregator runs some data processing of its own (action 2b), including data enrichment. The CHI is then able to - in turn - visit the data enrichments done by the aggregator and connect these to the data held in its own context. Same as in the prospective workflow for [UC-2 described](#) above, making connections between the data owned and held by the CHI and the enrichments to that data owned and held by the aggregator, having persistent identifiers and clear provenance metadata available for all entities at CHI- and at aggregator-level will be a crucial precondition.

## 8. Recommendations

### 8.1. Elevate continued exploration

While the insights gained due to the knowledge gathering conducted by this Task Force are promising and enlightening in certain aspects, there are other areas which - at the current stage of development not only of the technologies investigated, but also of the DS4CH and the data spaces in general - cannot be evaluated to their full extent yet. We therefore recommend continuing the exploration work initiated by this Task Force at minimum in the context of the Implementation Plan for Year 2 of the DS4CH deployment contract, but with the potential of this being continued also during Year 3 and 4.

This exploration should be considered as an extension of the existing activities aimed at investigating innovative scenarios for aggregation models, with a link to the anticipated update of the Aggregation Strategy in 2024. Additionally, it should align with the ongoing efforts to develop and manage a data governance mechanism, which extends into the second year of the DS deployment contract as well.

To enhance the exploration process, we suggest conducting a more in-depth investigation of various technological approaches and fostering deeper interactions with the respective initiatives, also including initiatives which could - in the restricted time frame of the Task Force - not be investigated yet such as MyData Global (<https://www.mydata.org/>).

This involvement should go beyond interviews and involve more active engagement with relevant stakeholders. It is important to leverage existing organisational affiliations to facilitate this work. Specifically, IDSA's role in the DSSC, Post-Platforms' membership in IDSA, and the Post-Platforms-led Personal Data Space Task Force involving SolidLabs within IDSA should serve as the foundation for these collaborative efforts.

As an organisational format we suggest the establishment of a longer-term Working Group, ideally Initiative-wide to represent other perspectives and requirements from within the future DS4CH as well. In addition to the technological investigation, this Working Group should aim at proposing a conceptual framework for centralisation vs decentralisation in the cultural heritage domain and at providing decision-making support for stakeholders in this field.

To get to this, a more informal approach, e.g., with a conceptual blog series to elaborate on the more [general questions](#) mentioned above and maybe an informal events series to discuss them could be considered.

### 8.2. Develop aggregation use cases further

Use cases from the cultural heritage sector form a critical foundation for the design and construction of the DS4CH and have hence also been a central aspect of this Task Force's work. Drawing from the experience of other data spaces, it is evident that well-defined and effectively communicated use cases are the key component in driving the development of data spaces. Given this importance, Europeana Initiative and its stakeholders should prioritise this work.

We recommend that the start made with the [use cases described in this report](#) should be continued, deepened, and expanded in its efforts. On the one hand, it is essential to involve other stakeholders in addition to aggregators such as CHIs, users, authorities, creative

industries, and researchers; on the other hand, it will be important to evaluate all these potential use cases for their completeness and to make them fit for purpose for the roadmap towards the data space for cultural heritage. As such, the development of use cases should be viewed as an ongoing, collaborative process involving multiple stakeholders and should also consider existing use cases, e.g., the ones Europeana Foundation has for the users of its services.

Clear communication and reporting mechanisms should be established to facilitate this process and an open platform for doing so should be found; to effectively showcase the use cases and engage the wider community, we recommend creating a dedicated section on Europeana Pro.

This section should present the use cases in an engaging and easily understandable manner, encourage contributions from the community, and provide effective reporting. This webpage can serve as a democratic tool, involving a broad community of heritage professionals and users in the design of the data space.

Next to including more perspectives and more stakeholders, those new and extended use cases should - apart from their description and explainers - prospectively also include an evaluation of their cost-effectiveness and their ecological impact as well as a possible transition plan establishing a baseline of what would need to change in the current workflows and processes. Furthermore, new and other types of data, e.g., personal data, should be considered in such a broader approach.

### 8.3. Pilot a decentralised approach

Pilots will play a crucial role in complementing theoretical explorations by offering rapid prototypes of different approaches in environments that closely resemble production settings. Following up on the proof of concept developed by the ENA Linked Data Task Force and in relation to any potential next steps resulting from this, we recommend that the application of a decentralised approach to aggregation is piloted as a next stage of this work.

Considering the significant investment not only in the data spaces, but also in related initiatives such as the European Collaborative Cloud for Cultural Heritage (ECCCH), these pilots should not only provide solutions to the use cases within the DS4CH, but similarly prioritise aligning the DS4CH with other data spaces and the ECCCH on various levels, from data sharing - or data visiting, via mechanisms and frameworks, to semantic standards, Linked Data, and workflows.

Furthermore, pilots should aim to establish an understanding of how much of a change a decentralised approach would bring for the current stakeholders and how many resources this might require especially from aggregators and CHIs. Ways should be investigated that would allow to minimise interference as much as possible, while still enabling the intended positive impact overall.

Next to the general approaches to decentralised aggregation as presented in this report, a pilot should also include a more detailed evaluation - in practice - of different technologies that might play a supporting and/or extending role such as the work on notifications mentioned under [chapter 7.5](#) above.

#### 8.4. Explore impact on aggregation landscape

Technology, which has been the focus of the investigation in this report, cannot be analysed in isolation from its wider application contexts, including business, vision, and society. Technological advancements enable new forms of work, but they are also shaped and driven by changes in collaboration and communication models. Recognising the vast opportunities presented by data exchange, sharing, and (re)use, also thanks to significant public investment in European innovation, it is essential to consider how initiatives like Europeana can evolve in terms of their overall vision and operations. We recommend taking the opportunity of the work of this Task Force, alongside parallel efforts in the DS4CH deployment project, to challenge the existing operational models - and definitions - of aggregation and aggregators and envision them in the future DS4CH. The work on developing pilots as suggested in 8.3. above will support this exploration, providing insights into the applicability of different forms of design to the aggregation landscape.

More concretely, this work should entail reassessing assumptions that have been accepted by the community of the Europeana Initiative regarding the role of aggregators as physical intermediaries, data facilitators, and knowledge hubs. This reassessment should consider societal trends that lean towards decentralised modes of engagement, action, communication, and resource distribution. While decentralisation can increase autonomy, inclusivity, resource allocation efficiency, and governance effectiveness, it is not a one-size-fits-all solution. Decentralisation must be complemented by centralisation in certain areas, as we suggested above, and it will be a continued task to define the right balance between the two for the future DS4CH.

With the increasing affordability and commoditisation of software, computation, and AI capabilities, hierarchical models of content aggregation need critical review. It is crucial to determine what elements should be centralised in the new model, such as semantic standards, governance aspects, or membership structures. These questions should be explored to lay the foundation for the renewed Europeana as part of the DS4CH.

Clearly, this work resonates and aligns with the strategic planning that the Europeana Initiative is undertaking as part of its transformative process. Aggregators should play a vital role in this process, as they have throughout Europeana's history. However, a clear division of scope and alignment with other strands of work must be established. Researching and ideating new work models can serve as an example of co-creation between the Europeana Foundation as a central actor and the EAF and the ENA as representatives of the domain, with potential involvement of additional stakeholders from the cultural heritage community. Successful co-creation can provide inspiration and guidance for the forthcoming redesign in the years to come.

#### 8.5. Explore impact on data governance and governance of the data space

Bringing the development of technology pilots (8.3.) and ecosystem design (8.4.) to its culmination, it is crucial to focus on the design of a governance approach for the data space in combination with the more specific aspect of data governance. As above, we recommend that both centralised and decentralised elements of governance within the European cultural heritage ecosystem should be explored, and their applicability to the DS4CH should be discussed. Europeana Initiative's previous experiences with different governance phases and

structures can serve as a valuable reference for this process. Examining other ecosystem examples, such as Wikipedia and other decentralised autonomous organisations (DAO-s), can also provide insights and inspiration for the discussion. It is important to involve various stakeholders, including CHIs, national platforms, the European Commission, and the public, to ensure a comprehensive perspective on governance considerations.

## 9. Conclusions

This report, while already quite extensive, is - as commented in the recommendations above - not to be understood as complete in the evaluation of decentralised approaches to aggregation and their potential impact on existing infrastructures, frameworks, governance models, roles, and relationships between the various stakeholders involved. The Task Force's work has shown that there are a lot more aspects to consider, which require further exploration and practical trials that were not possible to cover in the limited period assigned to the Task Force.

The authors of this report and the Task Force members hence are looking forward to continuing and building on the work done so far as part of the future steps towards deploying the common European data space for cultural heritage.

## Annex I - List of interviewees

Organisation	Interviewee(s)
Centre for Vision Speech and Signal Processing, University of Surrey	John Collomosse
Digita	Tom Haegemans
Filecoin Foundation	Danny O'Brien
Gaia-X	Francesco Bonfiglio
International Data Spaces Association (IDSA)	Sebastian Steinbuss
International Image Interoperability Framework (IIIF)	Régis Robineau
Post-Platforms	Alex Tourski, Sergey Nepomnyashi
SolidLab Flanders	Esther de Loof
Solid Project	(see Ghent University below)
Startin'blox	Sylvain Le Bon
GATE Institute	Milena Dobрева, Seamus Ross
Ghent University	Brecht Van de Vyvere, Olivier Van D'Huynslager
Institut national de recherche en sciences et technologies du numérique (Inria)	Fabien Gandon
Artefactual	Jenn Roberts, Kelly Stewart
Axiell	Rob Jacobs, René van den Heuvel
Lyrasis	Christina di Bella
Preservica	Peter Anderton, David Clipsham
Zetcom	Norbert Kanter
ArtCentrica	Marco Cappellini, Paolo de Rocco, Paolo Romoli
AskMona	Marion Carré
Historiana	Alice Modena
Big Data Value Association (BDVA)	Ana Garcia Robles, Tuomo Tuikka
Innopay	Mariane ter Veen
Team Data Spaces	Emre Bayamlıoğlu
Archives Portal Europe	Kerstin Arnold
CulturalItalia	Antonio Davide Madonna

<b>Organisation</b>	<b>Interviewee(s)</b>
Deutsche Digitale Bibliothek	Cosmina Berta
DigiPhil	Gábor Palkó
Europeana Foundation	Valentine Charles
Europeana Sounds	Tom Miles
European Film Gateway	Kerstin Herlt, Kristina Rose
Federacja Bibliotek Cyfrowych	Aleksandra Nowak, Tomasz Parkola
Jewish Heritage Network	Pavel Kats
Moteur Collections	Katell Briatte, Thierry Bultingaire, Marie-Véronique Leroi
MUSEU	Maria Teresa Natale, Marco Scarbaci
Netwerk Digitaal Erfgoed	Enno Meijers
Photoconsortium	Valentina Bachi
Statens Museum for Kunst	Nikolaj Erichsen
Digital Public Library of America (DPLA)	Scott Williams
Trove	Pi Klinjun, Amy Mathers, Aileen Weir
United Nations Educational, Scientific and Cultural Organization (UNESCO)	Denis Pitzalis
Europeana Foundation Advisory Board (chair)	Michael Peter Edson
Europeana Foundation Advisory Board	Merete Sanderhoff
Europeana Network Association Management Board (chair)	Rob Davies
Europeana Network Association Management Board (vice-chair)	Sofie Taes
CapGemini / Data Spaces Support Centre	Esther Huyer



## Annex II - Interview questions

Each interview started with an introduction to the Europeana Initiative (if necessary) and to the Task Force's specific remit. This was followed by a set of questions, which - for reasons of comparison - was the same for most interviews held. For the interviews with the aggregators, this set was slightly adapted to put more emphasis on the role of an aggregator within the common European data space for cultural heritage. Similarly, the set of questions for consultancies and opinion leaders was adapted according to the background and strategic positions of the interviewees. Questions only used with certain interviewees have been marked accordingly below. While the questions are numbered here, the sequence of questions might have been changed on the spot to adapt to the natural flow of the interview:

1. Short overview of the aggregator/initiative/organisation/project
  - a. *Was not asked in interviews with opinion leaders (instead the question on the interviewee's role was adapted and extended)*
2. Role of the interviewee(s)
3. Background (founding of the aggregator/initiative/organisation/project)
  - a. *Was not asked in interviews with opinion leaders*
4. What would you consider the Europeana Initiative's biggest success and what is its biggest ongoing challenge?
  - a. *Was only asked in interviews with opinion leaders*
5. Organisation and governance
  - a. *Was not asked in interviews with opinion leaders*
6. More detailed picture of the aggregator/initiative/organisation/project's daily activities
  - a. *Was not asked in interviews with opinion leaders*
7. Traction and progress
  - a. *Was not asked in interviews with opinion leaders*
8. Use cases (with special focus on use cases from the cultural heritage sector)
  - a. *Mainly asked for technology projects and providers, research institutions, and consultancies*
9. Understanding/image of the data space for cultural heritage
  - a. *Was only asked in interviews with European aggregators, consultancies, and opinion leaders*
10. How to proceed in setting up the data space for cultural heritage
  - a. *In interviews with European aggregators this was combined with some visualisations of the technology approaches*
11. Communities and stakeholders of the aggregator/initiative/organisation/project
  - a. *Was not asked specifically when interviewing collection management and preservation system providers (only if covered by the question on organisation and governance)*
  - b. *Was not asked in interviews with opinion leaders*
12. Benefits of sharing data
  - a. *Aggregators were particularly asked about improvements to sharing data*
13. Vision for the future
14. Final statement