Deliverable 2.1

Recommendations for shared cross-sectoral metadata

Co-funded by
the European Union

Executive Summary

This document outlines the first step of a more elaborate investigation into the prerequisites and the benefits of sharing data across different sectors of the European Cultural and Creative Industries (CCI). The project *CCI Thrive - Bespoke Business Models and Innovative Practices of Cross-Sectoral Cultural and Creative Collaboration,* co-funded by the European Commission within the Creative Europe programme, undertakes an exploration of the potential of sharing data in a common data space (platform/infrastructure) across all CCI sectors for improving the competitiveness of the European CCI, thus empowering SMEs that characterise the fragmented and culturally rich landscape of the industry. Data sharing is the foundation for being able to exploit smart technologies and analytical tools to pave the way for increased success and new business opportunities. This document provides a first analysis of the current situation in the "data landscape" of the CCI and thus forms a building block in determining the best approach for a shared data space, underpinned with a demonstrator as proof of concept and leading to a set of recommendations on how to efficiently establish and cater for a shared CCI data space.

This report starts with a brief outline of most common data structures within each industry and the data-sharing conditions in each of the sectors to highlight the differences between them and to discuss the current constraints and obstacles or barriers to data-sharing in general. Commonalities have been researched across industries to find a common ground of metadata structure.

To overcome the most obvious obstacles of "proprietary or competitive" data we looked into presenting the findings of a joint consortium effort identifying so-called "low-risk data" in the members' respective creative industries represented in the project consortium.

Finally, the report suggests that for the identified low-risk data to be useful, it needs to be curated and embedded in interfaces that are based on its users' needs. These interfaces will in turn provide new pathways to yet unexploited business intelligence.

Table of Contents

1. Introduction

The CCI Thrive project is undertaking comprehensive evidence-based research on the question of status, requirements, and benefits of sharing data across sectors for SMEs in the European Cultural and Creative Industry (CCI). The consortium of seven organisations representing a large range of sectors (film, broadcasting, performance arts, music, games, art museums, media art) carry out their investigation on the firm conviction that for the CCI to remain competitive, they need to keep pace with the technological developments which are largely data driven. Thus, data availability and access to data will prove essential to be part of the next generation of CCI businesses, operators, and creators. In addition, the investigation also reflects the belief that cross-sectoral collaboration and co-production can open up new business opportunities and creative spaces.

A key objective of the CCI Thrive project is to provide a solid body of knowledge surrounding data and its usage, which can assist all sectors in improving their approach to data-driven business models. By leveraging the experiences and sector-specific knowledge of the consortium members, the project endeavours to find innovative ways of mutual benefit, fostering cross-sectoral collaboration and exploring the potential for creative co-productions within the CCI.

To harvest the full potential of these opportunities, the CCI Thrive project recognises that collaboration and the sharing of data, know-how, and technologies are paramount. By exploring new and state-of-the-art approaches to data sharing and to utilising cross-sectoral data, the project aims to develop innovative business models and enhance existing ones within the CCI (this part of the investigation is presented in the accompanying deliverable *D2.2 Overview of different digital solutions to generate and analyse data*). This entails understanding what type of metadata standards and specific ontologies are most commonly used in the different CCI sectors, in parallel to an ongoing assessment of databases for interoperability opportunities and improved workflows and models.

Complementary to the above research, this report reflects on the work undertaken by the CCI Thrive consortium in identifying current practices of data sharing across different CCI which underpinned the assumption that cross-sectoral data-sharing is not a common practice, as even intersectoral data-sharing seems not to be practised yet. The report provides a brief overview of the data structures and sharing conditions within each industry, highlighting the differences in managing data and obstacles to sharing data. The consortium also conducted research seeking to discover where similar or the same types of metadata are used across different sectors, in what way they provide important information that would be interesting for other sectors to harvest, and which might have potential benefit for effective data sharing practices (chapter 2).

The consortium considered the difficulties of matching different metadata standards or usages across sectors and the reluctance in sharing data (inter- or cross-sectoral) for fear of losing a competitive edge or because of ownership rights (this is in particular the issue with databases who have to permission to collect and work with data from third party sources, but not necessarily have the permission to share them with yet another third party). These considerations led to looking into what metadata would not be encumbered by neither of the two mentioned aspects, which we termed "low-risk data".

Hence, in chapter 3 of this report we present identified low-risk data within the creative industry sectors represented by the consortium members. This approach allows us to design a demonstrator as proof of concept for the benefit of sharing data, by using easily available and accessible data during the second year of the project. This will provide a compelling basis for understanding how much more benefit could be had if more data types which currently are perhaps considered "high-risk" in terms of competitive advantage or reliability or IP, could be shared. These considerations also entail proposing a business case for managing,

securing, and maintaining a reliable database, widely used sharing practice and an up-to-date platform. This part of the investigation will result in a white paper and a set of compelling recommendations informed by evidence-based business intelligence as the final output. Furthermore, the report emphasises the importance of curating the identified low-risk data and embedding it in user-centric interfaces that align with the specific needs of stakeholders (chapter 4).

By undertaking this research and implementing the recommended strategies, the CCI Thrive consortium aims to create a dynamic environment that fosters collaboration, innovation, and growth within the European Creative and Cultural Industries. Through the exploration of data-driven approaches and the effective utilisation of digital technologies, the project seeks to strengthen the competitiveness of CCI, enabling companies and service providers to make excellent use of digital tools and overcome the challenges posed by the international digital transition.

2. CCI & Metadata - Purpose and feasibility?

To begin with, there is a distinction between cultural heritage (CH) institutions, which serve as custodians of heritage assets and often benefit from public support, and organisations driven by new productions that must compete to secure capital for their projects. CH institutions, by nature of their mandate, work with archives and have been employing digital databases for at least the past decade, with a focus on the "care, protection, and preservation" of existing productions from recent to remote times. On the other hand, organisations primarily focused on new productions show less involvement in archiving past works (which they may not have the staff for), instead prioritising the analysis of current trends, audience preferences, and financial opportunities. Due to the distinct "business case" for each, their practices, and strategies in collecting and managing data are not easily comparable.

Most CCI sectors comprise both publicly funded or community-driven institutions responsible for archiving the sector's assets as cultural heritage and organisations focusing on new productions and capital raising. Additionally, each sector typically has industry associations that gather economic data, although these associations are often not publicly accessible, and it remains unclear whether they maintain databases employing metadata standards. Furthermore, there are market research businesses that operate with sectoral data, such as GEMA (DE) for music rights, Nielsen (global) for film, and Newzoo (NL) for the game industry.

To find out what data may be shared within this environment, the first approach taken by the consortium was to look at available metadata standards in each of their industries and seek for similarities. Before delving into the detailed analysis of data structures for potential sharing, it was essential to explore the project's context and summarise the research objectives and the necessary content stakeholders required to develop robust business cases.

Through extensive discussions among consortium members, the following content areas were unanimously agreed upon as pertinent to our respective stakeholders:

- Trend monitoring
- A safe trading/distribution environment
- Content monetization
- Decision making supported by sales data
- Audience/community understanding
- Franchising
- Discoverability
- Immersive content production
- Personalization
- Co-production/partnerships
- Rights management

With this consensus in mind, the next step was to formulate four distinct use cases that would serve as focal points for our project:

- Decision Making – trends, sales data
- Community – audience data, discoverability
- Content Visibility – distribution, monetization & rights management
- Co-production – partnerships, franchising

In order to prioritise our efforts for the initial phase of the project, the consortium members participated in a vote to determine the primary use case. The results of the vote are depicted in the figure below (see Figure 1), which clearly shows that the Decision-Making use case – focusing on trends and sales data – emerged as the top choice.

Figure 1: results of the vote to determine the primary use case

With the use case selected, our attention shifted towards understanding the key challenges faced by European CCI today. By identifying these challenges, we can effectively shape our objectives and strategies to address them. Some of the noteworthy challenges identified include:

- Attractiveness and discoverability of European content
- Diversity and inclusion (serving underrepresented groups and exposing audiences to more diverse content)
- Green transition and sustainability
- Boosting Small and Medium Enterprises (SMEs) within the CCI sector
- Transforming sector fragmentation into a competitive advantage

Building upon this framework established for our activities, this chapter offers a concise examination of the data architectures and sharing practices prevalent in each industry. It aims to illuminate the variations in data management approaches and identify the barriers that hinder data sharing. Moreover, this chapter presents research endeavours aimed at identifying commonalities and overlaps in the utilisation of metadata across diverse sectors. It explores the valuable insights that such metadata can offer, demonstrating their potential significance for other industries to harness and derive benefits for fostering efficient data sharing practices.

2.1 Overview of data management practices in different CCI sectors

2.1.1 Cultural Heritage Sector

Sharing data - more specifically providing access to digitised collections and metadata - is widespread in the cultural heritage sector. This is largely a result of large-scale investments by the European Commission and national governments in the digitisation of cultural heritage collections in galleries, libraries, archives and museums (so-called GLAMs).[1] In the context of CCI, GLAMs are rather advanced in their data sharing practices and especially in the development of shared digital infrastructures and standards.

The strong focus on shared infrastructures is partly due to the reliance on public funding in the sector, meaning that organisations are more likely to pool resources in order to benefit from economies of scale. Initiatives like Europeana[2] (which provides the basis for the new European Data Space for Cultural Heritage[3]) have invested significant resources in both (i) developing technological infrastructures to enable data aggregation, and (ii) support the upskilling of heritage professionals to ensure that they understand the benefits of providing online access to their collections and metadata. Both of these aspects are essential for creating an environment where data sharing is possible.

To support the data aggregation workflows, the sector has adopted Linked (Open) Data principles that enables the creation of links among GLAM data from different organisations as well as between GLAM data and external web data, thus providing much richer contextual information. Thesauri for different domains and topical areas[4] provide uniform vocabularies that are essential for ensuring the findability of relevant data among large (aggregated) datasets. Many online platforms for heritage collections provide APIs that offer access collections metadata in a live data format. This data can be easily used in other applications, for instance, for research purposes or educational apps.

Another important factor to consider here is the open access movement - a commitment to grant free and open access to collections and data in cases where there are no financial, legal, or technical barriers. In this context, the OPEN GLAM initiative[5] has gathered an overview of over 1.600 collections from around the world that are published under open licences (including public domain mark of Creative Commons).[6] Portals such as Trove, Japan Search and DPLA offer an interface to access aggregated heritage collections through advanced search functionalities, curatorials and APIs. While many GLAM organisations see it as their public role to democratise access to heritage collections, this is not always straightforward as many organisations have developed business models around content licensing. Because of this, some organisations only provide access to their data based on formal agreements, with APIs usually not being available (e.g Uffizi Galleries, Brera Museum).

As for metadata standards, many recommendations exist from international bodies like UNESCO or ICOM, but in reality, the work of cataloguing and creating metadata standards is managed at a national level and thus remains largely siloed. The Italian example for one such national organisation defining the local metadata standards is the ICCD (Institute of Cataloguing and Documentation).

---

[1] An example of this is the "Images of the Past" project that digitised over 90,000 hours of video, 20,000 hours of film, some 100,000 hours of audio and 2,500,000 photos held in four heritage organisations in the Netherlands. See https://publications.beeldengeluid.nl/pub/498

[2] https://pro.europeana.eu/

[3] https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage

[4] For instance, see Getty Art & Architecture Thesaurus https://www.getty.edu/research/tools/vocabularies/aat/

[5] https://medium.com/open-glam

[6] bit.ly/OpenGLAMsurvey

As far as business intelligence data is concerned (for instance, visitor numbers, data on engagement with content on online platforms), the situation is less advanced. GLAMs might have their own data collection workflows but the gathered data is mostly used for internal purposes (to improve organisational activities) and not shared with other organisations. Network organisations such as EGMUS and NEMO periodically conduct surveys that would gather such data but usually the results are presented in the shape of reports rather than raw data and it is difficult to get access to recent data as it is usually done on an annual or bi-annual basis.

2.1.2 Time-Based Arts

In the consortium, time-based art as a subsector of the fine arts sector, distinguished through its contemporaneity from heritage art, is represented by Ars Electronica who through their work and remit represent three perspectives:

- a festival - curation of artworks
- a collection - custody/care of artworks
- a community driver - advocacy for time-based art

Drawing on their experience provides us with insights from an organisation that has a long tradition of archiving the artworks and related documentation proposed and shown in the framework of the Ars Electronica projects. They thus represent the part of the sector that is familiar with databases and metadata. As such, they are distinct from e.g., the actual artist producing an artwork or organisations that exhibit or study the artworks (galleries, art institutes etc.).

One can assume that institutions like Ars Electronica have a better understanding of why using common metadata standards or developing a common ontology within a sector or subsector would also benefit the producers. The key driver is the support of the artists and the distribution of (in the sense of creating visibility for) their respective artworks. Though maintaining a database on the archived artworks is fairly common in this sector, the "sharing" remains highly complex: Basic information such as title, creator, year of creation (what we call "low-risk" data) are easily available, but problems arise because of the artists' need and expectation to earn money through the presentation of their art. Sharing, for instance, a video recording of a video artwork for free counteracts any exhibition value of the artwork. Additionally, in-house databases are usually not designed to distinguish between "low-risk" and "high- value" data or internal-external access levels and rights. Yet, artists wish for their work of art to be known which requires a certain visibility that in turn can be boosted through sharing data about the work.

There have been several endeavours to join forces for a joint operation on centralising information on time-based artworks such as the latest initiative by the organisers of the Summit for New Media Arts Archiving[7]. Standards like Dublin Core[8], the Europeana data scheme[9] and many others will fit the bill for the "low risk data" categories. However, establishing a common vocabulary to inform specific "time-based art/media art" metadata standards has been an ongoing discussion and ambition over the past decade. The challenge remains in particular with metadata describing content, genre, aesthetics of the actual works of arts, together with technical specifications on operations and technical requirements (e.g., which parts can be replaced by newer technology without compromising the integrity of the artworks). Instead, the archives harbour the video documentation with a restricted access level. The lack of these metadata does not only create gaps in knowledge of preservation and conservation but ignores the fact that today many artworks relying on and using technology is rarely a one-person-effort, but either a collective work or with commissioning parts to third parties.

---

[7] https://www.isea-archives.org/about-3rd_summit

[8] https://www.dublincore.org/

[9] https://pro.europeana.eu/page/edm-documentation

Ars Electronica developed a server-based digital archive[10] featuring a complex database structure with metadata and linked digitised materials (video, audio, images, texts). At present, the digital archive encompasses over 141,000 entries with a storage volume of approximately 61 terabytes (lossless compression) and over 4,5 million linked files. The database as well as all the files fed into and created by it are automatically transferred into and checked by a backup system. Parts of the database are already accessible for the general public via the online archive.

2.1.3 TV and Movie Industry

The TV and movie industry is highly dynamic, with several distribution channels (theatrical, television) now challenged with numerous streaming platforms (international and local new players), rich in online data potential. Additionally, there are strong historical silos of data sectors within the industry, with some publicly available databases (theatrical ecosystem, TV local audience measurement institutes). In the context of this environment, the following aspects of data are important to consider:

Descriptive

This type of data provides information about the content itself, such as the title, genre, synopsis, cast and crew details, production company, release date, and other relevant descriptive information. It helps users to quickly understand the nature of the content and make informed decisions about what to watch.

Technical

Technical data focuses on the technical aspects of TV series and films. It includes details like video format, audio format, resolution, aspect ratio, runtime, language, and other technical specifications. This information is important for ensuring compatibility across different platforms and devices.

Rights

Rights data encompasses information related to the rights and permissions associated with a TV series or film. It includes details about copyright, licensing agreements, distribution rights, and restrictions on the content. This metadata helps in managing and enforcing intellectual property rights and ensuring compliance with legal obligations.

Structural

Structural data defines the organisation and structure of a TV series or film. It includes information about the overall hierarchy, episode or scene breakdown, chapter markers, timecodes, and other structural elements. This metadata is used for navigation, indexing, and linking different parts of the content together.

User-generated

User-generated data refers to the annotations, tags, ratings, reviews, and other user-contributed information associated with TV series and films. Platforms often allow users to add their own metadata to enhance discoverability and provide personalised recommendations to other viewers.

Operational

Operational metadata is generated during the production and distribution processes of TV series and films. It includes information about editing decisions, post-production workflows, encoding settings, versioning, and more.

---

[10] Example of the Ars Electronic Archive: http://archive.aec.at

Between these data aspects, however, important weaknesses exist in terms of data usage and availability. Subscription Video on Demand (SVOD) platforms, for the most part, do not share their usage data due to confidentiality reasons (high risk data) and no standard exists so far to qualify an audience on platforms - contrary to linear TV that has been standardised for decades at this point. There is also a lack of standardised public identifiers for content, and de facto proprietary IDs like IMDB are used without a consensus on their accuracy and relevance. Moreover, there is a lack of international standardisation regarding B2B data within the industry, and data on future productions is limited. These factors contribute to a market with limited data-driven insights but with high potential needs.

2.1.4 Games Industry

In the games industry, the availability and sharing practices of metadata vary. The sector generally demonstrates a good level of metadata availability for the end user, as game developers and publishers often provide comprehensive information, feeded from their inhouse database, about their games, including titles, genres, descriptions, release dates, and platform compatibility. However, there is no common practice of data sharing with different platforms and storefronts using their own systems and prohibiting access to their users' or products' data.

We can distinguish following areas where data is used:

Game Development

A metadata structure is essential during the development process to organise and manage game assets. It helps developers categorise and tag various elements, such as characters, environments, items, animations, and sound effects. This enables efficient asset management, easy retrieval, and effective collaboration among the development team. In this field the only harmonisation offered comes from asset stores (e.g. Construct, Unity Asset Store, etc.) and technical engines (e.g. Unity, Unreal Engine, etc.)

Game Distribution

Metadata is used to provide vital information about games during distribution and marketing. This includes details such as game title, genre, developer, publisher, release date, supported platforms, system requirements, age rating, and languages available. Metadata ensures accurate and standardised information is displayed on digital storefronts, making it easier for players to discover and evaluate games.

User Experience

The provision of data enhances the user experience by improving game searchability and discoverability. For instance, game stores use metadata to enable filtering and sorting options based on various criteria like genre, rating, popularity, and release date. This helps players find games that match their preferences more effectively.

User-generated Content (UGC)

In games that support UGC, metadata is used to categorise and organise player-created content. This can include mods, custom levels, character skins, and other user-generated assets. Metadata helps players browse, rate, and find specific UGC that aligns with their interests, enriching the overall gaming experience.

Analytics and Personalisation

Game companies leverage metadata for analytics and personalisation purposes. By collecting and analysing metadata related to player behaviour, preferences, and engagement patterns, developers can gain valuable

insights into player demographics, game performance, popular features, and player retention. This data helps them make informed decisions about future updates, content creation, and personalised recommendations. Also data is used to track player progress, achievements, and milestones within games. This data is often associated with the player's profile and is used to provide a sense of accomplishment, offer rewards, and facilitate competition among players. Depending on the size of the company, these analyses vary in terms of depth.

Game Streaming and Broadcasting

Metadata is crucial for game streaming platforms and broadcasting services where a live audience watches a gamer playing live. It allows streamers to provide viewers with relevant information about the game they are playing, including the title, genre, developer, and any additional details. This helps viewers identify the game and engage with the content more effectively.

In all these domains, data is relevant for the game companies' business strategy, creating a reluctance towards sharing data and/or adopting metadata standards. Data analysis happens in closed ecosystems and this lack of accessibility renders the situation challenging for developers to efficiently use valuable metadata such as sales numbers or audience profiles from larger platforms. Additionally, the sharing of more sector-specific metadata, such as gameplay mechanics, detailed sales analysis, developer tools, or in-game performance data, does not happen because of their strategic importance, thus potentially hindering comprehensive analysis and insights within the industry.

2.1.5 Music Industry

In the ever-evolving landscape of the music industry, data has emerged as a powerful asset. However, data-sharing practices in this sector are often shrouded in secrecy, as private companies guard their valuable insights in the fiercely competitive environment. While there exists a willingness to share certain anonymized data for research purposes, protecting sensitive information remains a top priority for many players. Nonetheless, with advancements in technology, larger music sector players possess the capability to unlock the potential of data-sharing, while smaller independent entities face challenges due to limited resources and outdated data management systems.

With a long list of stakeholders, including labels, distributors, publishers, collecting societies, managements, live agencies, and sync agencies, data-sharing between these entities remains elusive, and much of the data remains hidden from the public eye. This lack of transparency and collaboration has been widely acknowledged, prompting calls for the establishment of a dedicated research department, such as a European Music Monitor[11].

Similarly, the music industry has already recognized the immense potential in reaching unconventional audiences beyond traditional distribution channels. However, accessing and understanding these audiences can be a complex endeavour and cross-sectoral data sharing may provide insights into user behaviour and trends from sectors such as social media, streaming platforms, and online marketplaces. This would enable music creators and rights holders to connect with untapped audiences, fostering a diverse and inclusive musical landscape.

The music sector is a very fast developing and (re)acting one, rather setting trends than following them. As such, data and insights within the music industry hold the potential to benefit other sectors.

---

[11] Example: https://culture.ec.europa.eu/news/the-european-commission-publishes-two-studies-of-music-moves-europe-feasibility-of-a-european-music-observatory-and-market-gaps-analysis

There already exists a history of fruitful collaborations with the realms of film, advertising, and dance. However, these collaborations often occur in a fragmented manner and fail to reach their full potential. By fostering greater data-sharing practices between these sectors, the music industry can contribute to the creation of captivating audio-visual experiences. Insights into trends and user behaviour can enable synchronised storytelling, where music serves as the thread that weaves together the emotions, visuals, and narratives. This synergy will not only enhance the impact of music in these sectors but also provide creators and artists with new avenues for expression and collaboration. There are also emerging signals for the potential of cross-sectoral collaboration between music and fashion or music and games.

2.2   How could data sharing work?

In the realm of data sharing, it is crucial to distinguish between metadata and the actual data itself. Metadata provides essential information about the data, such as its structure, format, source, and context. Sharing metadata is a foundational step in enabling efficient data interoperability and understanding the characteristics of the data. However, the true potential of data sharing lies in the exchange of the data itself. This encompasses the valuable insights, patterns, and trends embedded within the data. To fully harness the benefits of data sharing, it becomes imperative to establish standardised frameworks and protocols that facilitate the interoperability of content data. By defining common standards, organisations and sectors can overcome the challenges of data silos and unlock the power of data collaboration and analysis across diverse systems and platforms.

General-purpose vocabularies play a crucial role in structured data markup within CCI, enabling the effective organisation and representation of information. One such widely recognized and utilised vocabulary is schema.org. Schema.org provides a comprehensive set of schemas and terms that facilitate the description of various types of entities, such as artworks, events, organisations, and more, within the CCI domain. By employing schema.org vocabulary, content creators and organisations can ensure consistent and standardised representation of their data, making it more discoverable and interpretable by search engines   and   other   data   consumers.
As the previous chapter shows, however, there seem to be both no common standards and varying practices for data tracking and sharing across different cultural and creative industries. Notably, however, there also seems to be no particular sector that seems to put more emphasis than the others in terms of shared metadata standards.

In order to identify possible sharing opportunities, we first need to evaluate the current view on data usage in their sector by each consortium member. Looking at the various industries we detected a clear lack of uniform way of using existing metadata (see Figure 2).

Do you find metadata usage in you sector using/being close to one existing standard?
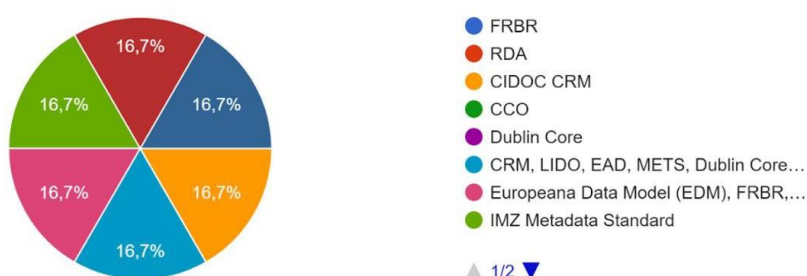6 réponses

- FRBR
- RDA
- CIDOC CRM
- CCO
- Dublin Core
- CRM, LIDO, EAD, METS, Dublin Core…
- Europeana Data Model (EDM), FRBR,…
- IMZ Metadata Standard

▲ 1/2 ▼

Figure 2: survey on metadata usage in different sectors

As the key to sharing data lies with interoperability which requires a) using metadata standards and b) providing access to the data (e.g., through an API or a downloadable CSV or other machine-readable file), we can assert that data sharing within CCI is far from being a common practice - both within the sectors and thus even less so across sectors. Other industries, but increasingly the film industry and the streaming services, are showcasing how much competitive advantage data-driven technologies could offer. Data is key, but the amount of data is crucial. Big data is needed and a joint effort across sectors would provide valuable insight that would help bridge the fragmentation of the European CCI landscape both in terms of national culture and of small companies.

The advocacy for sharing data remains the same whether for one sector or more: it mainly focuses on making a case for the application (or development) of metadata standards, widely employed in all sectors and along the value chain and domains (producers/curators/ custodians/communities etc.), and the understanding that providing access to a certain set of data is a long-term benefit not a severe drawback. The added challenge of the cross-sectoral approach would be to understand which data from other sectors might actually prove valuable and useful, and in what way.

For this purpose, CCI Thrive explores the status of commonly used metadata standards, the existence of accessible databases and the type of metadata that is most commonly published. The next step in moving forward towards this end was to set up a survey that would analyse the relevance of different data types per sector. Seeing that the issue is so complex, we limited the survey - intended to supply mere first insights - to the sectors known to the consortium members.

The consortium represents those actors in the sectors that either deal with existing productions (heritage or recently released/launched/published) or consult in business decisions based on data analytics of recent productions (e.g., comparables in the film business, sales figures in games, audience attendance in performance art, visitors in exhibitions etc.). This explains the positive fact that we actually received answers to questions regarding metadata, and yet it reflects the dismal situation in the lack of congruence between standards or the lack of applying standards across sectors.

As an intriguing exercise, the survey delved into the matter of metadata relevance by utilising a particular sector's metadata scheme and a core set of metadata derived from the game industry. The game industry's metadata was selected due to its lesser-known nature but potential alignment with audio-visual productions. The objective was to ascertain whether there exists cross-sector congruence regarding the significance of metadata. To facilitate this exercise, SpielFabrique provided concise descriptions of the metadata used (refer to Annex 2). The metadata schema employed was the Video Games Metadata Schema (VGMS) developed by the GAME research group at the University of Washington in collaboration with the Seattle Interactive Media Museum (see Figure 3).

The obtained results (additional results provided in Annex 1) proved to be surprising and simultaneously reinforced the advocacy for metadata standards (see Figure 4). Despite the glossary's elucidation of the metadata's meaning within this scheme, the fact that some metadata were deemed "not relevant" (although they are undoubtedly relevant in most, if not all, sectors) underscored the challenges associated with data transfer. For instance, the term "agent," which in the context of book publication would refer to an "author" or in music to a "composer," holds significant importance across all sectors. However, four out of six respondents marked it as "irrelevant," likely due to a lack of recognition regarding the term's significance within the game sector.

Conversely, the "agent" category appears to encompass a broader range of roles, including individuals, organisations, or groups responsible for the creation, realisation, manufacture, marketing, and/or distribution of a video game. Such a categorization may not be applicable to the same extent in other sectors utilising different metadata terms (see Figure 5). Only two metadata elements, namely "title" and

"language," garnered unanimous agreement as being relevant, most likely owing to their consistent application across all sectors. In contrast, the "genre" metadata received only two out of four votes for relevance, possibly because the game scheme referred to it as "narrative genre."

What in this scheme is relevant or not in your sector - Entity
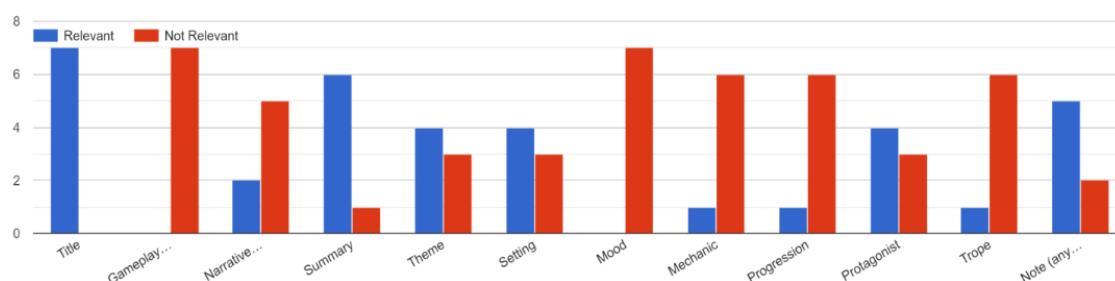
Figure 3: *Survey on the relevance of metadata using the Video Games Metadata Schema*

Additional content entity
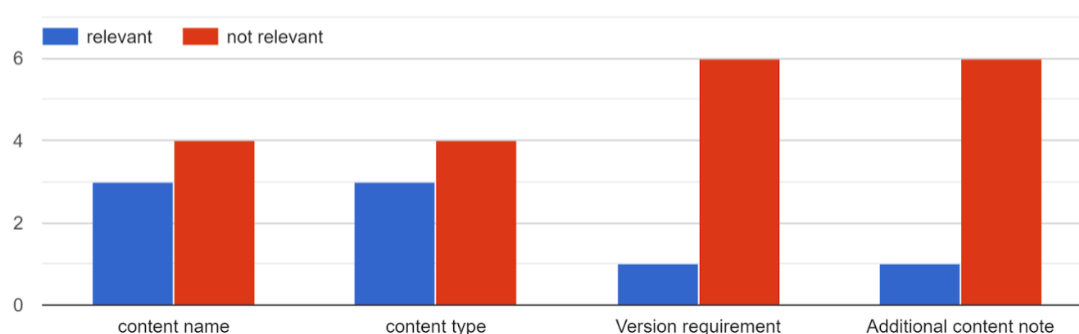
Figure 4: *Survey on the relevance of an additional content entity*

Agent entity
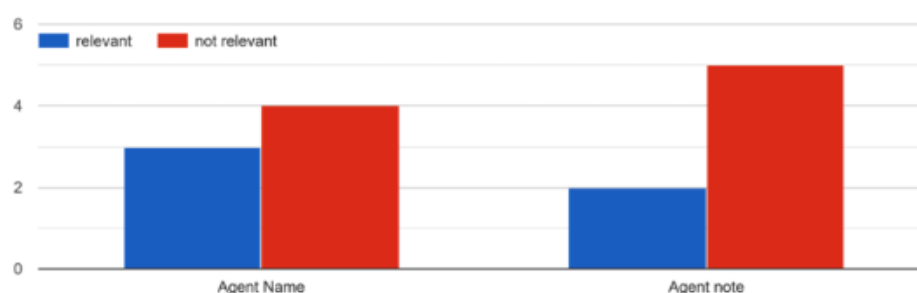
Figure 5: *Survey on the relevance of an agent entity*

Between 54 different data types, 17 were evaluated as predominantly relevant, 37 as predominantly irrelevant (see Figure 6). Of all of these results 16 data types were highly contested - in that the majority opinion on their relevance was contested by at least 3 other consortium members. 33 data types were relatively uncontested - with 1-2 consortium members voting against the majority opinion. And 5 data types were deemed relevant or irrelevant with a unanimous vote.
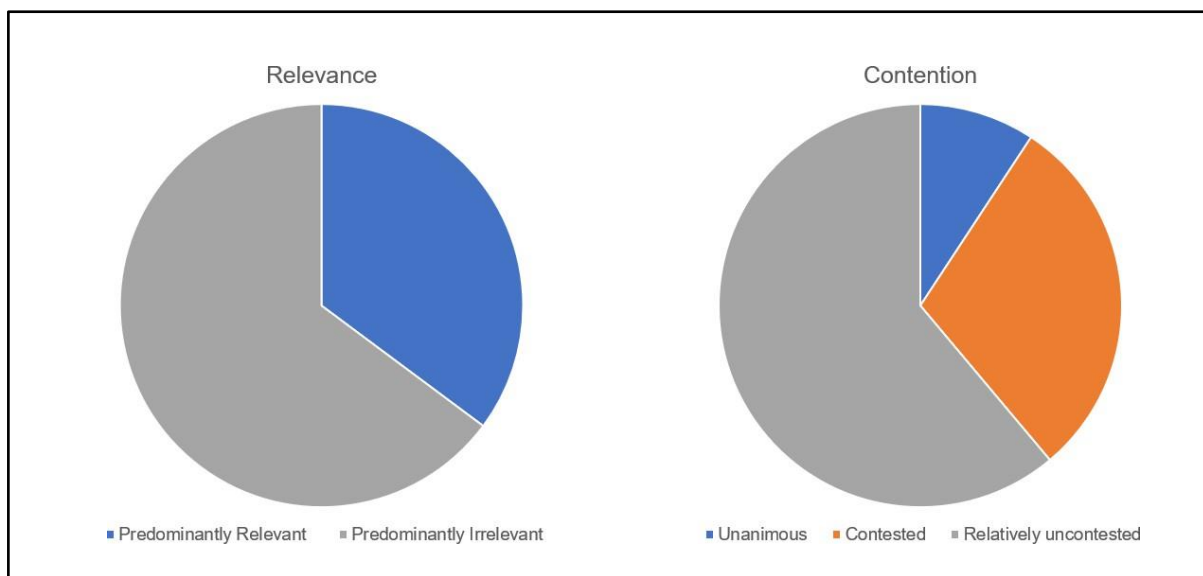
Figure 6: *Relevance of data profiles across CCI*

Relevance across different categories, called entities in this metadata standard, varied somewhat (see Figure 7). Data for some entities, like the Franchise Entity (e.g., franchise names) or the Platform Edition Entity (e.g., dimensions, system requirements) seemed to hold little value for many of the consortium's CCI. Conversely, the Series Entity (e.g., series titles), followed closely by the Collection Entity (e.g., collection title) was recognised to be quite a bit more relevant. It is worth noting, however, that no one entity showed a unanimous relevance and none a unanimous irrelevance.
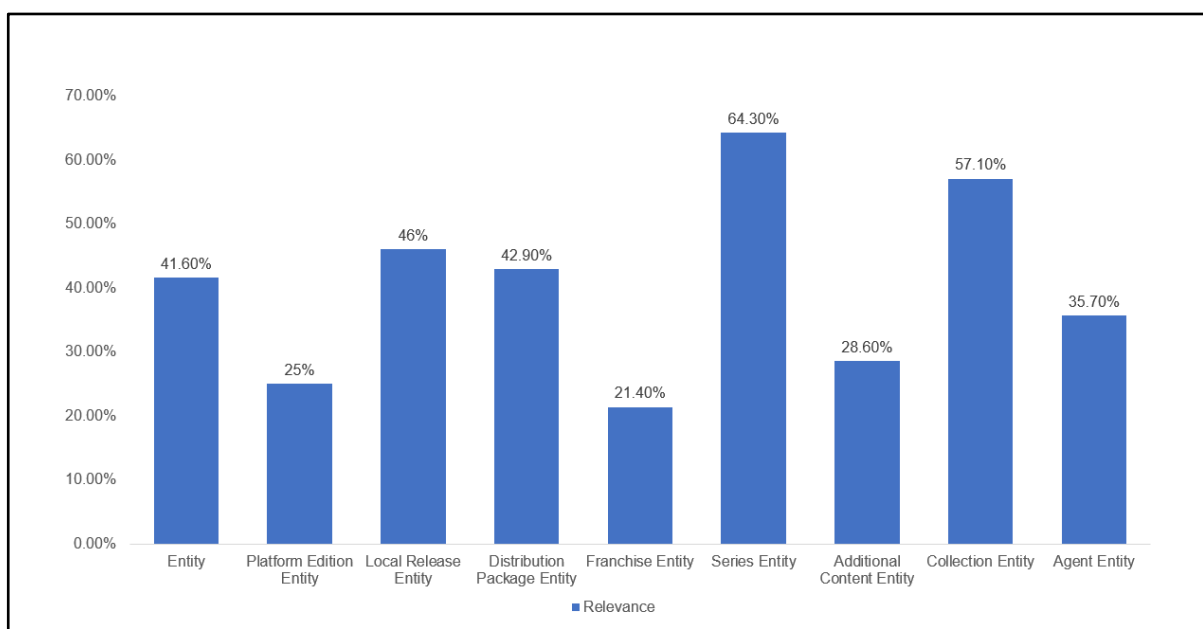
Figure 7: *Relevance of data profiles based on entities*

These findings largely corroborate the hypothesis that the varying standards in different industries may, at least in some part, stem from a differing understanding of what metadata holds as useful for the respective industries. It seems that in order to create shared metadata standards, some work would need to be put into negotiating the relevance of specific data types between industries. Alternatively, shared metadata

schemes may need to take account of the fact that some industries may lack and will - due to their inherent irrelevance - continue to disregard some metadata types.

It would be amiss, however, to ignore that the survey also laid open that a range of metadata types hold, indeed, a shared relevance for more than one CCI. Examples include data on series and collection entities. Leveraging these shared interests may already bear value for these CCI and may convince other sectors of the benefits of tracking and sharing this data themselves. Ultimately, even if the majority of data types hold value for merely a fraction of the consortium, the composition of this minority shifts from one data type to the other and the global interest in sharing *specific* data types remains.

So far, it has been understood that standards are important for data exchange, but what remains obscure is an understanding of how standards come about, e.g., which actors with which interests and with which market power enforce standards, and how these standards behave compatible with already existing quasi- or legally recognized standards. This issue is multiplied by a lack of vision about the economic appeal of data sharing.

The consortium took note of these insights and consequently conducted research on low-risk data profiles across CCI that will be presented in chapter three of this report. Evidently, some data types, however basic, remain uncontested in their relevance across otherwise very differently set up sectors.

2.3 Barriers to data sharing

Before investigating how to share low-risk data, it makes sense to analyse the barriers CCI face, across sectors and in their own ecosystems, when it comes to sharing metadata.

One of the major barriers is the lack of standardisation in metadata across different sectors. While standards exist, they are often not widely adopted, leading to incongruences and challenges in data sharing. Another barrier is the absence of licences that facilitate the legal sharing of metadata. This hinders the ability of CCI to freely exchange data and collaborate effectively.

Moreover, there is a lack of interoperability, making it difficult to reuse metadata across different platforms and industries. The absence of stable identifiers further complicates the situation, as there is no universally recognised system for identifying and referencing content. This lack of standardisation and stable identifiers hampers efficient data sharing and integration.

On a more fundamental level, there is a need for better understanding and education within the global creative market regarding data management and usage, including trend analysis and demand forecasting. Many stakeholders, especially smaller players, are unfamiliar with utilising data for strategic decision-making. Increasing familiarity and knowledge about data analytics would enable CCI to harness the full potential of data-driven approaches.

Across various domains, gatekeepers such as platforms and distributors control access to data. VOD platforms have user registration data, but it is mostly restricted by the gatekeepers. Audience research in, for example, the TV industry is generic and lacks personalisation. Physical products and pay-per-view models also suffer from limited data access controlled by gatekeepers. In the music industry, platforms, concerts, and radio/TV broadcasts possess data on consumption, but again, it is predominantly held by gatekeepers. Games on the other hand primarily rely on platform and console data, which varies in depth and availability based on platform requirements and subscriber information. The digital nature of content

consumption allows for some interactions to be monitored by platforms, but this data is not publicly accessible or freely available.

Simply put, the complexity and cost associated with accessing data pose significant barriers, particularly for small and independent players in the industry. The tools available for data analysis and utilisation are often complex and require substantial resources, preventing many stakeholders from benefiting from data-driven insights. Simplified and affordable tools tailored to the specific needs of the industry are necessary to democratise access to data.

In a nutshell, while data exists within CCI, its availability and accessibility are hindered by various barriers. Lack of standardisation, licensing constraints, limited interoperability, absence of stable identifiers, and proprietary control of data by gatekeepers pose significant challenges. Additionally, the complexity and cost associated with data access, along with a need for better education and understanding, further impede the effective sharing and utilisation of metadata within the industry. Overcoming these barriers is crucial for CCI to unlock the potential of data-driven insights and innovation.

2.4   Benefits of data sharing for CCI - What advantages can each sector see in data sharing?

Removing the aforementioned obstacles will require a significant amount of time. The initial step involves presenting a compelling argument in favour of data sharing. Subsequently, it is crucial not only to engage with the relevant stakeholders who can effectively act upon our proposals but also to support the argument with evidence-based proof of concept and a clear strategy for implementing feasible instruments that facilitate data sharing.

To embark on this process, a fundamental understanding of the prerequisites for data sharing is necessary. However, since the initial data would be provided by organisations that already collect and aggregate data, and therefore possess knowledge in data management, the focus shifts towards comprehending the potential and embracing the responsibility as data owners, custodians, or brokers to share data for the benefit of the entire industry. Consequently, the persuasive argument stems from the combination of data management expertise, infrastructure knowledge, and a business intelligence rationale that highlights the advantages of cross-sectoral data sharing. In pursuit of this, the consortium has developed realistic scenarios to illustrate how a shared platform could furnish the requisite knowledge to seize the opportunities presented through data sharing. The comprehensive report detailing this aspect of the project investigation constitutes a separate document[12]. Hence, in this section, we provide only a brief overview of the benefits we envision for the CCI sectors.

Based on group discussions by the consortium's members which contribute their sector-specific experience, the key benefits that data sharing across CCI would bring are the industry's enhanced overall efficiency, increased profits and optimised decision-making processes.

By leveraging shared data, CCI can gain valuable insights into global trends, enabling them to identify popular themes and preferences across different media. This knowledge empowers creators to produce content that aligns with viewers' interests, leading to more engaging and relevant offerings. Additionally, understanding franchise potential helps build sustainable business models and capitalise on opportunities

---

[12] https://cci-thrive.eu/media/d2-
2_overview_of_different_digital_solutions_to_generate_and_analyse_data.pdf

in over-the-top (OTT) and international markets. Data sharing also facilitates the identification of potential partners across different media, fostering collaborations and expanding creative possibilities.

A comprehensive understanding of distribution platforms is another advantage of data sharing. By analysing platform gaps and trends, CCI can strategically position their content to fill these gaps and maximise international financing opportunities. This knowledge supports effective engineering of distribution strategies and enables content creators to reach wider audiences. Moreover, data sharing facilitates trend monitoring, allowing CCI to stay ahead of the curve and adapt to changing consumer preferences, ensuring continued success in a rapidly evolving media landscape.

When it comes to fostering a more predictably successful economic environment, data sharing can also play a big role. By exchanging information on content performance, sales, and audience engagement, CCI can make informed decisions about content monetisation and distribution. This data-driven approach mitigates risks and increases the chances of success by providing insights into market demand, optimising content offerings, and enhancing revenue streams. Sales data in particular plays a pivotal role in decision making within CCI. By accessing and analysing sales data, creators can gain valuable insights into the performance of their content, enabling them to make informed decisions about future projects, marketing strategies, and audience targeting.

Understanding the audience and community, in turn, will be one of the largest benefits of sharing data cross-sectorally. By collecting and analysing data on audience preferences, behaviours, and engagement, CCI can gain a deeper understanding of their target demographic and create informed personas of their users and audience. This understanding enables more effective audience engagement strategies, personalised content recommendations, and the development of stronger relationships with viewers. Ultimately, it enhances the overall experience users and audiences of CCI products will have and will foster more cost-effective usages of both private capital and public funds.

On a similar note, discoverability is significantly improved through data sharing. By exchanging data on content metadata, tagging, and user behaviour, CCI can enhance content discoverability and recommend relevant content to users. This personalised approach increases engagement and will, in particular, help both creators and users to navigate through vast libraries of content and offers more efficiently, improving innovation and engagement levels on both sides.

Data sharing also facilitates franchising opportunities within CCI, i.e., working on similar content or the same IP across different sectors. Book adaptation by film or games are the most commonly practised "franchising" models. Sharing data on recent or planned production would unlock the potential for franchising across sectors beyond these more common examples. By identifying successful franchises and analysing their performance, CCI can replicate successful formulas and develop new content offerings with built-in brand recognition. This not only increases the chances of success but also supports the expansion of franchises across different media, fostering growth and diversification.

Similarly, co-production and partnerships are facilitated through data sharing. By exchanging information on content production, performance, and audience engagement, CCI can identify potential collaboration opportunities and form strategic partnerships. These collaborations can leverage complementary strengths, shared resources, and access to new markets, resulting in innovative and mutually beneficial content offerings.

Finally, effective rights management is another benefit of data sharing. By maintaining a centralised and accessible database of rights information, CCI can streamline rights negotiation, track ownership, and manage licensing agreements more efficiently. This ensures that content rights are protected and appropriately monetised, reducing the risk of legal disputes and improving revenue streams.

To summarise this and the previous sections, we can conclude that data sharing across CCI, while facing a multitude of barriers both externally and within the CCI, holds a considerable potential. The consortium's CCI may hold different views on which particular type of data holds value for them specifically or their sector generally, but the inter-sectorial potential for cross-sectoral data sharing remains strong, nonetheless. It enables creators to identify global trends, produce relevant and engaging content, strategically distribute their offerings, and make informed decisions supported by sales data. It enhances audience understanding, content monetisation, franchising opportunities, discoverability and many more aspects that would deliver mutual win-wins across many CCI. Knowing this, the question of *how* to start building shared data platforms to unlock the potential of these use cases becomes even more pertinent.

3. Low-risk data profiles

Naturally, the real benefits of data sharing lie with the data that is currently either inaccessible or uncollected, such as audience data. However, to make a compelling case for data sharing, we require a persuasive scenario and a convincing demonstration of those benefits on a smaller scale. This demonstration should visualise and empower a scaled-up vision.

To achieve this, the consortium will develop a realistic demonstrator utilising available data and, in some cases, data that is typically not publicly accessible from our partners. The purpose is to showcase the value of the approach and to alleviate concerns about data sharing by starting with data that does not provide a competitive advantage or disclose intellectual property or proprietary information. For the sake of convenience, we refer to this data as "low-risk data." We adopt this approach to encourage partners and their sectors to reflect on which data can be shared without causing harm and to foster an understanding of any undue overprotectiveness of data.

3.1   Rationale behind investigating low-risk data profiles

As a result of the research into the value and relevance of metadata types, the CCI Thrive consortium ultimately concluded that an exploration of low-risk data profiles across CCI would be the most useful next steps. The previous sections highlighting the barriers to data sharing, the benefits of data sharing, and the challenges faced by CCI in accessing comprehensive and reliable data already set the stage for understanding the importance of exploring low-risk data. Between the many differences in practices and preferences between the consortium's CCI, the question remained how to create proof-of-concepts that may shed light on how data sharing on a cross-sectoral dimension could eventually work. As a consequence of the consortium's analysis and discussions, the decision was made to conduct an analysis into low-risk data profiles across sectors, so as to build a collection of data that could be relied on for the road ahead.

The consortium defined low-risk data profiles as data that is considered to
- hold no critical competitive value
- cover no sensitive issues

and
- is publicly available
- is reliable
- remains valid

One of the primary obstacles to data sharing is the lack of data availability and sharing practices, particularly in sensitive and competitive areas. However, low-risk data, as defined, circumvent these barriers by focusing on data that holds no or minimal competitive value, cover no sensitive issues, and is publicly available. By targeting such data, CCI can overcome the hurdles associated with confidentiality and proprietary restrictions, thereby fostering a more open and collaborative environment.

By investigating and utilising low-risk data, CCI can tap into its potential for various purposes. Being essentially a light-weight testbed for cross-sectoral data collaboration, low-risk data provide a valuable resource to understand potentials and barriers to data sharing when actually practised. Low risk data, in the consortium's rationale, may serve as a stepping stone for CCI to develop a culture of data usage and analysis. It can be seen as an entry point for the industries to become familiar with cross-sectoral data-driven practices and methodologies. By starting with low-risk data, CCI may gain confidence in utilising data

and gradually expand their data-driven capabilities. It is to emphasise that this is a project specific definition based on low-risk through greater availability.

Beyond that, curating a collection of low-risk data profiles across different CCI may lay the foundation for discovering the scope and utility of a potential cross-sectoral data sharing platform - informing the further direction of both the project on a micro level, and practical hurdles to overcome for any such endeavour on a macro level. Given this potential of low-risk data to lead the way towards further research and discussion, the consortium launched an investigation engaging all members in the crowdsourcing of available low-risk data.

3.2   Low risk data analysis results

Before moving into a detailed collection of low-risk data profiles, the consortium attempted to identify larger categories of data profiles where one would expect low-risk data to be present. In a joint effort, the members agreed on three larger data classification categories with several sub-categories:

Creation & Content Data:
- Geography (e.g., production country, language, subtitles)
- Topic (e.g., titles, synopsis, preview image)
- Genre/ Format (e.g. genre, format, IP)
- People (e.g., content creators, producers)
- Technology (e.g., medium, aspect ratio)
- Temporal (e.g., creation date, production status)
- Rights (e.g., rights holder, copyright)

Consumption & Engagement Data
- Audience (e.g., number of visitors/ users, social media followers)
- Financial (e.g., price, licence fee, number of downloads)
- Geography[13] (e.g., number of viewers per country, distribution type)
- Technology[14] (e.g., file format, file size)

External & High Level Factors Data
- Policies, guidelines & recommendations (e.g., streaming platforms, quotas, CO_2 emissions)
- Funding (e.g., public funding percentage, co-productions, production budget)

After defining these areas of investigations, the consortium members were tasked to research available low-risk data in each of them and document the following details:

● **Type of Data**: specific datasets that respond to one of the low-risk data categories (e.g., Genre, Temporal, People, etc.)

● **Relevant Sectors:** a CCI sector that considers this data type low risk.

● **Existing Data Source:** where does the data come from (e.g., online platforms, Content Management Systems, etc.).

---

[13] later discarded due to overlaps with Creation & Content Data's Geography section.

[14] later discarded due to overlaps with Creation & Content Data's Technology section.

- **Access Level:** is the data open or proprietary? Licence and conditions for sharing the data.
- **Formats / Standards[15]:** format in which the data is available (metadata standard).
- **Frequency:** how often can this data be updated (e.g., is it live data or are the numbers being reported on a weekly/monthly/quarterly/yearly basis).
- **Importance - 1 = lowest, 5 = highest:** the importance that the sector providing this data assigns to it (in terms of how much effort and resources they dedicate to maintain/improve this data).
- **Notes:** any additional information that might be relevant.

The resulting table of low-risk data profiles (see Annex 3) holds 102 entries from 8 different sectors. Between the aforementioned data categories, "Creation & Content Data" holds the most entries for low-risk data profiles, with "Topic" being the most fruitful sub-category in this larger section, holding 19 entries for low-risk data profiles. The category "External & High Level Factors Data", conversely, holds the least amount of entries, with only 9 entries in total and "Policies" only accounting for two of those entries, making it the sub-category with the least amount of entries.

Similarly, there is a discrepancy between the sectors on how many different low-risk data profiles were identified: Music and Film showed the highest diversity of low-risk data (32 different low-risk data profiles), followed by Games (23 different low-risk data profiles) and Fiction (16 different low-risk data profiles). On the low end of the spectrum were the Cultural Heritage and Arts sector.

However, tendencies as to which sub-categories had prominent representation in any given sector, differed somewhat, too: Most low-risk data for music and film, for instance were registered in the "Genre" sub-category, while for the runner-up, Games, low-risk data profiles in the sub-category "Topic" stood out.

As for which low-risk data were shared across the most CCI, sub-categories "Topic", "Genre", "People", and "Geography" performed best. The leading low-risk data profiles, "Work title", "Content creator name", "Genres", "Synopsis", and "Place of creation", shared five or six different CCI who identified low-risk data profiles for each of them. Notably, all these examples fall under the larger category "Creation & Content Data".

For the rest, 8 data profiles were shared across 3-4 sectors, 11 data profiles were shared across at least two sectors, and the 29 remaining data profiles had been identified in only one of the different sectors.

The consortium members were asked to rate the importance of the identified low-risk data, in terms of how much effort and time is dedicated to sourcing and maintaining this data. While not synonymous, this importance may provide clues for the value of the data from another sector's perspective.

Almost half of all identified data profile entries (48) were evaluated as being of highest importance (5). In total, 84% of data profile entries received a rating of 3-5 in terms of their importance. Of the remaining 16 entries, 7 were evaluated as too uncertain to receive a concrete score. The average importance rating was 4.1.

---

[15] later discarded due to a lack of data.

Category	Average importance
Creation & Content Data	**4.23**
– Geography	3.50
– Topic	4.26
– Genre/Format	4.63
– People	4.55
– Technology	3.71
– Temporal	4.28
– Rights Data	4.67
Consumption & Engagement Data	**3.94**
– Audience	4.40
– Financial	3.38
External & High Level Factors Data	**3.13**
– Policies, guidelines & recommendations	3.00
– Funding	3.14

Table 1: Survey on the average importance of the three categories of data profiles

While with a standard deviation of 1.14 there remains only little discrepancy across the different sub-categories (see Table 1), the importance of the identified low-risk data profiles seems to be at least somewhat lower in the larger category "External & High Level Factors Data", as compared to the other two.

When it comes to the accessibility of each of the data profiles, the consortium members rated them from "Publicly available", on the one end of the spectrum, to "Confidential", on the other end. Some data in between would be free, require a subscription model, or might be too dispersed to make one single claim and access levels would differ from one provider to the next.

Between all the profiles, a clear majority of 82 entries were evaluated to be publicly available or at least in part freely accessible. 10 entries were rated as confidential or pay-walled, and another 9 to be too dispersed to give a clear indication. The larger category "Creation & Content Data" ended up showing the largest

percentage of partially or fully freely available data with 90% of all entries being rated "Free", "Free/ subscription" or "Publicly available". Low-risk data on consumption and engagement, on the other hand, remained more contested, with 43% of entries being rated "Confidential" or "Dispersed".

Data availability	Average importance
Publicly available	3.79
Free/ subscription	4.40
Pay-walled	5.00
Confidential	3.67
Dispersed	4.13

Table 2: Average importance of data availability

When cross-referencing the importance of data profiles and their access level, publicly available and confidential data profiles both performed worst (see Table 2), albeit only with a difference of <0.5 to the total average importance. Pay-walled data performed best, with the caveat that this data set consists only of 4 entries.

Finally, the members rated the frequency of updates on low-risk data profiles. This frequency rating indicates how often low-risk data can or will be updated, ranging from "Live" for any real-time updated data to "Once" for data that will not receive any updates after being entered or submitted.

Here, the majority of data profiles were evaluated to receive no or only few updates. 58 data profiles were marked with "Once" or "Once - can be updated", whereas 21 profiles were updated on a regular basis (daily, monthly, yearly, or after certain recurring events). Five data profiles are registered to be updated in real time. The remaining 18 data profiles did not have any reliable update pattern, or their update schedules are unknown.

The discussion in the notes of the table used for collecting low-risk data also revealed some limitations and caveats:

One significant limitation is the dispersed nature of low-risk data, which often exists on local or national levels rather than being consolidated in transnational or even global bodies. This fragmentation poses challenges in accessing and aggregating data across different regions or countries and prevents data from being uniform and comparable.

One result of this and the widespread lack of automatisation of data collection is that the availability and functionality of data and weblinks can become a limiting factor. Low-risk data sources may suffer from outdated information or broken links, rendering the data unreliable or inaccessible. This poses challenges in maintaining the accuracy and currency of the collected data.

Additionally, confirming this accuracy and transparency of low-risk data are being lamented as challenging by the contributors to the table. There may be limited transparency in terms of the data's sources, methodologies, or quality assurance processes. Ensuring the reliability and trustworthiness of low-risk data become challenging without proper validation mechanisms and transparency. Finally, to gain access to data, aggregators often have to rely on specific and granular terms in their sharing agreements, leading to further barriers to wide-spread data access and consolidation.

There are also limitations that rather concern the nature of data itself. CCI, for example, encompass a wide range of genres and formats, and the lack of standardised taxonomies makes it difficult to compare and integrate data from various sources. Similarly, identifying the roles of producers, co-producers, and contributors can be unclear, place names may change, and there may be either duplicate or vastly differing titles for the same content in different countries.

Lastly, the availability of data for analogue cultural goods heavily relies on their digital presence. If these goods lack a digital footprint, data availability becomes limited, hindering comprehensive analysis and insights.

3.3   Discussion of findings

The consortium's investigation into low-risk data profiles across CCI has provided valuable insights into the potential and challenges associated with utilising such data. The rationale behind exploring low-risk data was driven by the need to overcome barriers to data sharing, leverage reliable and publicly available information, and foster a collaborative environment among CCI. The findings highlight the potential of low-risk data for already providing a strong basis on which to build more cross-sectoral data collection efforts.

The analysis of low-risk data profiles revealed three major categories: Creation & Content Data, Consumption & Engagement Data, and External & High-Level Factors Data. These categories encompassed various sub-categories that provided a comprehensive framework for investigating and documenting low-risk data. Notably, Creation & Content Data emerged as the most extensive category - potentially reflecting an easier access to its data.

Within the Creation & Content Data category, sub-categories such as Topic, Genre/Format, People, and Geography stood out as the most prominent areas where low-risk data profiles were identified. These profiles, such as work titles, content creator names, genres, and synopses, were frequently shared across multiple CCI, indicating their importance and widespread applicability.

However, it is worth noting that the amount of low-risk data profiles varied across different sectors. Music and Film demonstrated the highest range, followed by Games and Fiction. On the other hand, Cultural Heritage and Arts sectors exhibited a lower range of low-risk data profiles. These variations pointedly reflect how differently the various sectors are positioned in terms of easily accessible low-risk data. Building cross-sectoral data collaborations, even with low-risk data, requires significantly more work in some sectors than in others.

The importance assigned to low-risk data profiles by the contributing members of the consortium provided insights into how the majority of identified data profiles were considered of high importance. This insight is promising as it shows the potential for easy access to valuable data. Paired with the insight that much of the data that was deemed important also remains publicly - or at the very least easily - accessible, this promises to remove barriers for using data cross-sectorally. Only few profiles were marked as confidential

or dispersed, indicating limitations in accessing certain types of low-risk data. Nonetheless, we need to consider that this research effort likely yielded low-hanging fruits in terms of valuable and accessible data - and that this may become more difficult if intensifying such research. We are also recognising that access to valuable low-risk data is not equally easy across all CCI and will likely result in an imbalanced sharing of data in the beginning of any cross-sectoral efforts.

Frequency of updates for low-risk data profiles varied widely, with a significant number of profiles receiving no or infrequent updates. This suggests that many low-risk data profiles are static or updated only periodically. While a subset of profiles was identified as updated regularly or in real time, for the majority of data this lack of up-to-date, and often even functioning, data presents a real problem for cross-sectoral aggregators or even users of this data.

The discussion surrounding the collected low-risk data profiles highlighted several other limitations and caveats that also need to be considered. The dispersed nature of low-risk data, primarily existing at local or national levels, poses challenges in aggregating and standardising data across regions. This fragmentation hinders data uniformity and comparability, limiting the insights gained from cross-industry analysis.

Data complexities and nuances within CCI, such as variations in producer roles, changing place names, and divergent content titles across countries, further complicate the collection and integration of low-risk data. Overcoming these challenges requires meticulous data management, harmonisation efforts, and standardised practices within the industry.

In conclusion, the investigation into low-risk data profiles across CCI has allowed a glimpse into their potential as an easily accessible and valuable tool for cross-sectoral sharing. What remains are concerns regarding an uneven distribution of such low-risk data across sectors, as well as concerns about the usability of existing data due to data-inherent or external barriers. Addressing these limitations will contribute to the broader utilisation of low-risk data. In the meantime, low-risk data still poses a good starting point to begin sharing data across CCI, develop proof-of-concepts, and gain insights on its transformative potential for the industry. The consortium members also see the potential for new business models in the aggregating of data and related services using specific data spaces (See digital marketplace). Also, it would be interesting to further discuss the self-motivated data sharing models in terms of business sustainability.

4. How to utilise cross-sectoral data in digital solutions to generate and analyse data

The research documented in the preceding chapters have provided valuable insights into the availability of low-risk data in Creative and Cultural Industries, highlighting the mixed state of data accessibility and the need for improved interoperability, access, and data sharing practices. While it is evident that most CCI do not currently share data across sectors, the research conducted by the consortium has revealed the existence of low-risk data, which makes data sharing between different CCI a realistic, albeit still challenging, objective. Moving forward, the focus should be on how to aggregate and curate this data to maximise its value for users and facilitate cross-sectoral collaboration.

The research has found that data availability in CCI is highly contextual, varying strongly across different sectors and even within them. Additionally, metadata standards are not uniform, further complicating data sharing efforts. Particularly when it comes to commercially used data, accessing, and sharing such data becomes increasingly difficult due to proprietary restrictions and confidentiality concerns. However, despite these challenges, the identification of low-risk data profiles through the consortium's research has opened new possibilities for data sharing between sectors. This provides a promising foundation upon which to build further exploration and collaboration.

To effectively leverage this low-risk data, the consortium recommends focusing efforts on the examining of database architecture, and specific use cases as described in more detail in the report "Overview of different digital solutions to generate and analyse data".[16]

While the benefits of data sharing are widely acknowledged in general terms, the specific use cases of data sharing in and across sectors still require a deeper understanding. A set of hypothetical use cases that demonstrate the utility of data sharing across industries could help bridge this gap. These use cases should showcase hands-on the practical applications and benefits of cross-sectoral data sharing, helping stakeholders understand the value it can bring to their respective fields.

Providing use cases will not only enhance our understanding of the utility of data sharing but also inform the development of interfaces and demonstrators which is the next step taken in the project. Creating such interfaces and demonstrators will serve as tangible examples and sources of inspiration for the design and implementation of data sharing platforms. By visualising and showcasing the benefits of data sharing, these tools can drive adoption and active participation from stakeholders across different CCI. For example, the BetaSeries platform opened new business opportunities on one side to creators by providing them insights of current watching trends, and on the other side to the platform holders by retaining the users and being able to offer to the community paying services.

Additionally, developing demonstrators will highlight the technical and legal challenges that are associated with data-sharing - such as establishing data-sharing protocols, ensuring data privacy, and developing mechanisms for data governance - and how to overcome them.

By taking these steps, CCI can unlock the full potential of cross-sectoral data and foster a data-driven ecosystem that drives innovation, collaboration, and growth across the industry. This research has shown that, while challenging in parts, the utilisation of shared data, even if low-risk, is both realistic and promising in terms of informing decision-making, enabling the identification of emerging trends and opportunities,

---

[16] https://cci-thrive.eu/media/d2-
2_overview_of_different_digital_solutions_to_generate_and_analyse_data.pdf

and supporting evidence-based policymaking. Ultimately, a robust data-sharing framework will pave the way for the development of data-driven strategies, products, and services that cater to evolving audience needs and contribute to the sustainable growth of CCI. By unlocking the full potential of cross-sectoral data, CCI can position themselves at the forefront of the digital era and drive the future of the creative and cultural landscape.

Annex 1: Metadata types analysis

What in this scheme is relevant or not in your sector - Entity
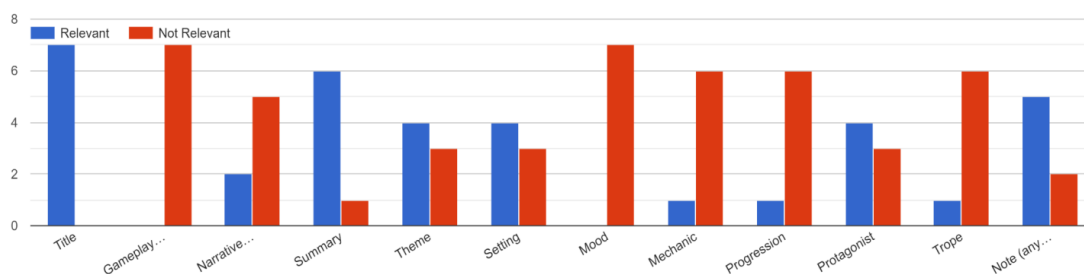
What in this scheme is relevant or not in your sector - Platform Edition Entity
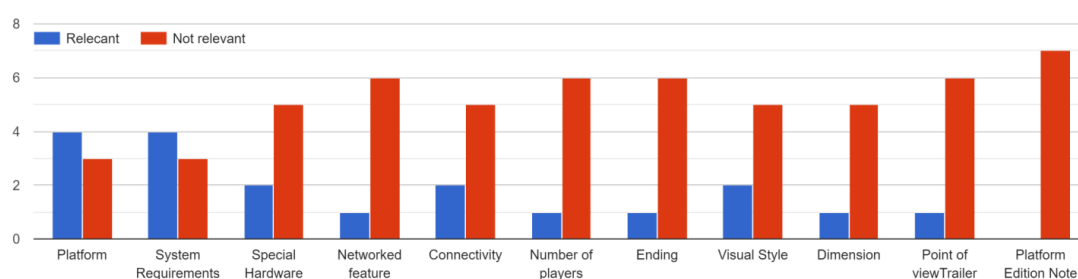
Local release Entinty
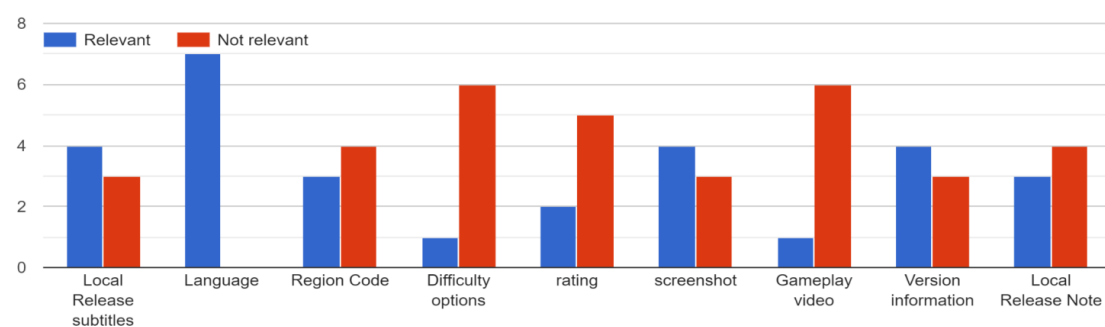
Distribution package entity
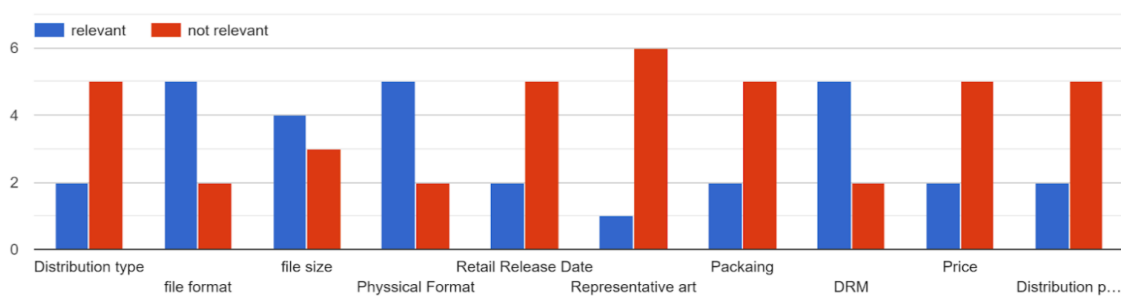
Franchise entity
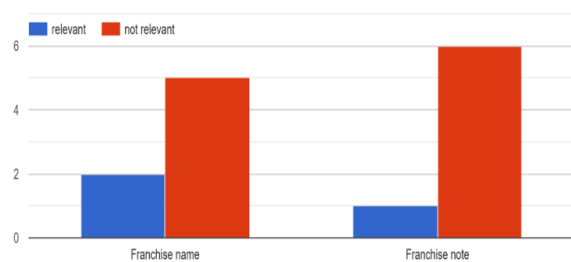
Series entity
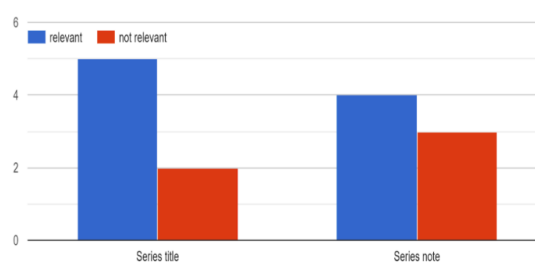
Additional content entity
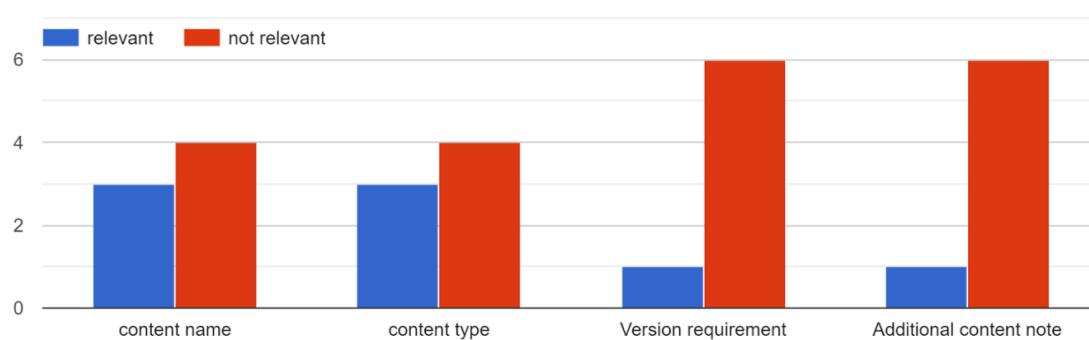
Collection entity
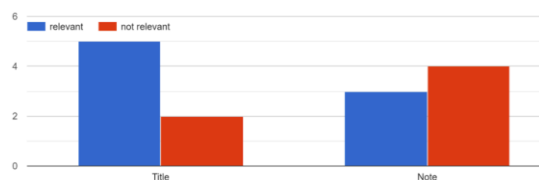
Agent entity
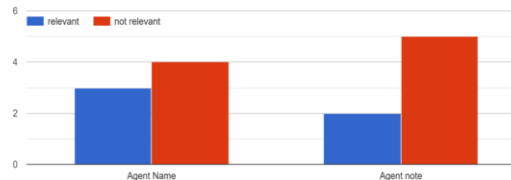
Annex 2: VGMS metadata glossary

Here are the definitions of the different element of the schema, as described in the 4.1 version of the VGMS from the Game Research (GAMER) Group.

Game Entity

Title: Proper names used to refer to a video game, assigned by the creator

Gameplay Genre: The overall nature of a video game's interactivity based on its objectives, types of rules, distinctive characteristics, modes of action, and manners of gameplay.

Narrative Genre: The narrative genre describes the type of game world and plot, similar to literary genres such as science fiction and fantasy

Summary: A brief statement or account of events in the video game.

Theme: A common thread, motif, subject, or idea that recurs in the video Game

Setting: The type of world, location, and time period in which the video game takes place.

Mood: The pervading atmosphere or tone of the video game which evokes or recalls a certain emotion or state of mind.

Mechanic: A primary method or rule used to interact with the game state and progress through the game content

Progression: A description of how the player progresses or advances through the video game.

Protagonist: The nature of the main character(s) in the game, from the viewpoint of the characters(s) the player assumes

*Trope*: A generally recognisable narrative device or convention. Tropes rely on culturally mediated expectations to expedite the development of events, characterisation, or narrative. When overused or oversimplified, tropes often come at the expense of fair and inclusive representations of vulnerable identities.

Note: Any other notable characteristics of the video game.

Platform Edition Entity

Platform: The operating system or device on which the video game was designed to be played.

System Requirements: Hardware, firmware, and/or software components that are prerequisites for running the video game on a particular platform.

Special Hardware: The additional hardware devices that are recommended or required for playing the video game (e.g., motion controller; gaming headset)

Networked Feature: The ways in which the video game can be experienced through connection(s) to other entities (e.g., companies, third-party organisations, other players).

Connectivity: The technology through which the online features of the video game are realised.

Number of players: The number or range of the number of players the video game

can accommodate either separately or concurrently.

Ending: How the video game ends and/or the inclusion of post-game content.

Visual Style: The predominant and recognizable visual appearance of a video game as originally intended by its creator, and/or determined in the context of creation.

Dimension: The intended perception of the depth of the represented entities inside the video game.

Point of view: The perspective from which the player experiences the gameplay.

Trailer: The video footage released and/or endorsed by the developer/publisher of the video game for promotional purposes.

Platform Edition Note: Any other notable characteristics of the platform edition.

Local release Entity

Local Release subtitles: A word or phrase appearing in the video game denoting the uniqueness of the local release (e.g., second edition, greatest hits, collector edition limited edition)

Language: The classification code for the language(s) in which the video game conveys information.

Region Code: The classification code that indicates the video encoding and regional hardware necessary to realise the video game.

Difficulty options: The in-game options for difficulty of gameplay.

Rating: The classification of the content in the video game used to inform decision making about the video game, provided by organisations such as professional associations, video game distributors, or creators.

Screenshot: Still images taken during the gameplay.

Gameplay video: Video footage of the gameplay excluding such things as introductions, cutscenes, or trailers.

Version information: Information related to the particular version of the video game being catalogued.

Local Release Note: Any other notable characteristics of the local release of the video game edition.

Distribution package entity

Distribution type: The manner in which the video game is distributed.

File format: The file format for digitally distributed video games

File size: The size of the file for digitally distributed video games

Physical Format: The distribution medium that provides the executable code of a video game with physical media.

Retail Release Date: The date of the public/commercial release of the video game

Representative art: The officially released image that is representative of the video game, prominently featured in a physical or digital distribution package.

Packaging: All items included in the original packaging of the video game

DRM: Digital rights management technologies intended to control the use of the video game.

Price/MSRP: The manufacturer's suggested retail price (MSRP) at time of initial release in the region where the video game was released.

Distribution package note: Any other notable characteristics of the distribution entity of the video game local release.

Franchise entity

Franchise name: A commonly used name referring to the intellectual property, related data, and content shared among a group of cultural objects to which the video game being described belongs.

Franchise note: Any other notable characteristics of the franchise of the video game

Series entity

Series title: Proper name of a set of related video games, often indicated by consecutive numbering, continuing narrative, or similarities in gameplay and themes, to which multiple video games belong.

Series note: Any other notable characteristics of the series of the video game.

Additional content entity

Content name: Proper name used to refer to additional content for video games, assigned by the creator.

Content type: The type of additional content.

Version requirement: Any version requirement to play the video game with the additional content.

Additional content note: Any other notable characteristics of the additional content.

Collection entity

Title: Proper name used to refer to the collection of video games, assigned by the creator.

Note: Any other notable characteristics of the collection the video game is part of, that are not captured in other fields

Agent entity

Agent Name: An individual, organisation, or group of individuals or organisations responsible for creation, realisation, manufacture, marketing, and/or distribution of a video game

Agent note: Any other notable characteristics of the agent.

Annex 3: Low-risk data analysis

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Geography						
Place of creation/ production	Cultural Heritage (NISV)	In-house collection catalogue metadata (as part of provenance)	Publicly available	Once - can be updated	5	Sometimes unknown; complicated by differences in granularity of place names, and historical changes
	Fiction (Betaseries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Complicated by nuances like who is the lead, places of shooting or place of production lead/ several leads and place of creation (co-producers)
	MIC (Ministry of Culture)	B2B	Free	Continuous	4	very complicated, some links do not work
	Game (SPF)	Online media & owner website	Publicly available	Once	1	Places of creation, realisation, manufacture, marketing, and/or distribution
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising	Publicly available	Continuous	3	Artists' country - prix map section inside the Ars Electronica archive

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Regional rights availability	Music and Film (IMZ)	Producer/ distributor	Free/ subscription	Once	4	Where the asset is available as a general product (basic rights)
Language	Music and Film (IMZ)	Producer/ distributor	Free/ subscription	Once - can be updated	4	Dubbing/subtitles
	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	4	
Rights owner country	Music and Film (IMZ)	Owners legal domicile	Free/ subscription	Once	4	Main producer who owns the rights
Subtitles	Game (SPF)	Online media & owner website	Publicly available	Once	1	

Note: The top portion of the first table continues from the previous page, with the partial cell:
"contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Topic						
Work Title	Cultural Heritage (NISV)	In-house collection catalogue	Publicly available	Once	5	
	Fiction (BetaSeries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Complicated by same titles for different contents and very different title in other countries for same content

	Music and Film (IMZ)	Producer/ distributor	Free/ subscription	Once	5	
	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	
	Artwork (painting, sculpture) (Centrica)	In-house collection catalogue (ArtCentrica), B2B	Pay-walled	Once	5	Complicated by different sources of data (Open Access, based on specific agreements)
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca	Publicly available	Continuous	5	In the different sections of the archive Prix , Pic archive, Prix Map, talks and lectures, S+T+ARTS and web specific projects art& science and AI LAB
Synopsis	Fiction (BetaSeries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Complicated by several length formats, official/used versions
	Music and Film (IMZ)	Producer/ distributor	Free/ subscription	Once	5	Short description or summary of the asset's content

	Artwork (painting, sculpture) (Centrica)	In-house collection catalogue (ArtCentrica), B2B	Pay-walled	Once	5	Complicated by different sources of data (Open Access or lacking, based on specific agreements)
	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca	Publicly available	Continuous	5	Archive section Prix (awarded artists). Catalog description
Preview Image	Music and Film (IMZ)	Producer/ distributor	Free/ subscription	Once - can be updated	4	At least one image + credits
	Artwork (painting, sculpture) (Centrica)	In-house collection catalogue (ArtCentrica), B2B	Pay-walled	Once	5	Complicated by different sources of data (Open Access or lacking, based on specific agreements)

		In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca	Publicly available	Continuous	3	Archive section Prix (awarded artists).
Version Type/ title/ description	Music and Film (IMZ)	Producer/ distributor	Free/ subscription	Once - can be updated	3	One name + other info if necessary
Setting	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	2	The type of world, location and time period in which the video game takes place.
Franchise	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	
Series	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	2	
Age Rating	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	2	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Genre / format						
Genres	Fiction (BetaSeries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Complicated by several taxonomies, some errors (mix of genre and formats), need to be further detailed for most titles
	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Applies to concerts; clear genre to be communicated
	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Unknown	Uncertain	Applies to documentaries; clear genre to be communicated
	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	Type of game world and plot
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising contemporary Artworks project"	Publicly available	Continuous	3	It´s a video selection from the conference talks we have during the Ars Electronica Festival

		2011-2013: https://pro.europeana.eu/project/dca				
Format	Fiction (BetaSeries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Complicated by several format taxonomies, mix formats/genres
IP (on which content is based)	Fiction (BetaSeries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	
	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Based on which music piece
Resolution	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once - can be updated	4	Available Resolutions
	Media Art (AE)	The image resolution is describe in the file section example Prix award artists and Pic Archive	Publicly available	Unknown	Uncertain	Applies to images archive; file names contain the resolution and event the size Prix archive and Pic archive
Duration	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once	5	

	Media Art (AE)	The videos contain the information with the duration	Publicly available	Unknown	4	Applies to talks and lectures archive
Gameplay genre	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	Overall nature of a video game's interactivity based on its objectives, types of rules, distinctive characteristics, modes of action, and manners of gameplay

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: People (creators, producers, employees)						
Content creator name	Cultural Heritage (NISV)	In-house collection catalogue (i.e. as artist, author, director)	Publicly available	Once - can be updated	5	The lines can be blurred between creators, producers, co-creators, contributors
	Fiction (BetaSeries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	The lines can be blurred between creators, showrunners, directors, producers, co-creators
	Music and Film (IMZ)	Producer and distributor	Free/ subscription	Once	5	Production company in that case

	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	Agents of creation/ realisation
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising Contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca	Publicly available	Unknown	Uncertain	The artists names appears in all the archives with the work title, Prix , pic archive (photographer), prix map, talks and lectures (speaker/s), S+T+ARTS and web specific projects art & science and AI LAB There are in the Prix the awarded artists and the ones that apply to the Prix Ars Electronica
Producers	Fiction (Betaseries)	B2B, B2C online media & owner website	Free/ subscription	Once	4	Complicated by nuances like who is the lead, several leads co-producers
	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once	4	Several co-producers
	Game (SPF)	online media & owner website	Publicly available	Once - can be updated	3	Agents of marketing, publishing
Talents	Fiction (Betaseries)	B2B, B2C online media	Free/ subscription	Once	5	Complicated by criteria to

		& owner website				include or not in the list
	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Music artists featured, musicians, director, etc.

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Technology						
Medium	Cultural Heritage (NISV)	In-house collection catalogue metadata	Publicly available	Once	5	
	Fiction (Betaseries)	B2B, B2C online media & owner website	Free/ subscription	Variable	5	Complicated by dynamic change not tracked systematically and media description level of detail
Aspect Ratio	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once	4	
Audio tracks channels	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Once - can be updated	4	One entry per language
Identification code	Music and Film (IMZ)	Depends who generates the code	Free/ subscription	Once	4	
Technology used (games motors, VR…)	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	3	
Dimension (2D/3D…)	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	1	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Temporal (dates, production status)						
Creation date	Cultural Heritage (NISV)	In-house collection catalogue metadata	Free/ subscription	Once	4	
	Fiction (Betaseries)	B2B, B2C online media & owner website	Free/ subscription	Once	5	Complicated as not precisely communicated before first release
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitising contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca	Confidential	Unknown	3	Internal information we ask in the applications to the Prix award
Production status	Music and Film (IMZ)	B2B, B2C online media & owner website	Free/ subscription	Continuous	3	Different names: „ideation, in progress, in co-production, completed…"
Date of announcement	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
	Media Art (AE)	In-house PostgreSQL database solution based on django as application framework; metadata format based on Europeana's Digitizing contemporary Artworks project" 2011-2013: https://pro.europeana.eu/project/dca	Confidential	Continuous	5	
Retail release date	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Rights Data						
Rights holder	Cultural Heritage (NISV)	In-house collection catalogue metadata	Publicly available	Unknown	5	Dependent on the digital availability of the heritage
	Fiction (Betaseries)	B2B, B2C online media & owner website	Free/ subscription	Variable	4	Complicated by long windowing/ representations, groups communication, TV/producers confusions
	Music and Film (IMZ)	B2B, B2C online media	Free/ subscription	Variable	5	Key data to develop sales

		& owner website				
Type of copyright	Cultural Heritage (NISV)	In-house collection catalogue metadata	Publicly available	Unknown	5	Dependent on the digital availability of the heritage
	Music and Film (IMZ)	Producer/Distributor	Dispersed	Variable	4	Allow faster and easier sales
Image	Artwork (painting, sculpture) (Centrica)	In-house collection catalogue (ArtCentrica), B2B	Pay-walled	Once	5	Complicated by different sources of data (Open Access or lacking, based on specific agreements)

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Audience						
Number of visitors (physical)	Cultural Heritage (NISV)	Annual reports, surveys of visitor counts	Dispersed	Variable	Uncertain	
	Media Art (AE)	Blog post in our website	Publicly available	Unknown	Uncertain	Applies to visitors for Ars Electronica Festival
Number of visitors/ users (digital = online viewers)	Music and Film (IMZ)	Scarce B2B data availability & owner website	Dispersed	Continuous	5	Difficult because of large number of media and different ways to count

	Fiction (Betaseries)	Scarce B2B data availability & owner website	Dispersed	Continuous	5	Confidential but getting low-risk/ public with the development of AVOD & advertising need to share data
	Media Art (AE)	Matamo (statistic tool)	Confidential	Unknown	5	Applies to archive visitors; includes numbers, systems, which device and geographic location of the visitors; it is possible to find what artworks they visited
Number of visitors/ users (TV = viewers)	Media Art (AE)	Matamo (statistic tool)	Confidential	Unknown	5	Applies to talks and lecture section; includes numbers, systems, which device and geographic location of the visitors
	Fiction (Betaseries)	B2B for free TV, almost unavailable for PayTV	Dispersed	Continuous	5	National measurement systems, not all synched. No low-risk measure on PayTV

	Music and Film (IMZ)	B2B for free TV, almost unavailable for PayTV	Dispersed	Continuous	5	National measurement systems, not all synched. No low-risk measure on PayTV
Social Media followers & 'engagement'	Cultural Heritage (NISV)	Followers publicly available, engagement can be calculated	Publicly available	Live	Uncertain	
	Fiction (Betaseries)	Followers publicly available, engagement can be calculated	Publicly available	Live	4	Growing interest to capture and analyse impact and sequel potentials
	Music and Film (IMZ)	Mostly based on music artists follow ups	Publicly available	Live	4	Most interesting for contemporary artists, what about Public domain artists? Role of fans, interest groups
	Game (SPF)	Online media & owner website	Publicly available	Live	4	
Distribution platforms wishlists	Game (SPF)	Online media & owner website	Publicly available	Live	2	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Financial (sales, revenue, prices)						
Licence fee	Music and Film (IMZ)	Confidential	Confidential	Continuous	2	No transparency
Number of downloads	Music and Film (IMZ)	Confidential	Confidential	Continuous	2	No transparency; private platforms do not share numbers
Download price	Music and Film (IMZ)	Public on platform	Publicly available	Variable	4	Platform like JustWatch show price offers
	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	
Retail price	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	
Ticket price	Cultural Heritage (NISV)	Public on website	Publicly available	Continuous	2	Relatively similar to download and retail price
Business Model: FtP, subscription, single sale, episodic	Game (SPF)	Online media & owner website	Publicly available	Once - can be updated	5	
Festival ticket price	Media Art (AE)	Public website	Publicly available	Continuous	2	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Geography (consumption in different locations)						

	Archive (AE)	Matamo (statistic tool)	Confidential	Unknown	Uncertain	Geographic location
Number of viewers (online/ TV)	Music and Film (IMZ)	Producers and distributors	Dispersed	Continuous	3	Depends on available language, platforms for distribution, rights (audio and video)
Distribution type (Physical/ Digital)	Game (SPF)	Online media & owner website	Publicly available	Once	1	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Technology (medium, platform)						
File format	Game (SPF)	Platform data	Publicly available	One time, can be updated	1	
File size	Game (SPF)	Platform data	Publicly available	One time, can be updated	1	

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Policies, Guidelines & Recommendations (quotas, Green Deal)						
Streaming platform data	Music and Film (IMZ)	Not automatic connected, requires manual research	Publicly available	Continuous	3	Trend for interest in musicians
Carbon Footprint	Cultural Heritage (NISV)	Publications by some museums; some national aggregations	Publicly available	Variable	Uncertain	

		like Julie's Bicycle				

Type of Data	Relevant Sectors	Existing Data Source	Access Level	Frequency	Importance - 1 = lowest, 5 = highest	Notes
Data Category: Money (funding)						
Production budget & funding part	Fiction (BetaSeries)	Institutions' annual reports, B2B sites	Free/ subscriptions	Unknown	3	Partially available, manual work so far
	Music and Film (IMZ)	Institutions' annual reports, B2B sites	Free/ subscriptions	Once	3	Information difficult to check - European observatory
Co-production	Music and Film (IMZ)	B2B Markets	Dispersed	Continuous	3	Markets and festival follow ups
Pre-sales (MG Levels)	Music and Film (IMZ)	B2B Markets	Dispersed	Continuous	3	Markets and festival follow ups
Public funding %	Music and Film (IMZ)	AV observatory/ National funding	Dispersed	Continuous	3	Research, update after each committee
	Game (SPF)	Online media & institution website	Publicly available	One time, can be updated	4	
	Cultural Heritage (NISV)	Annual reports for institutions, might be aggregated nationally for a sector/subsector, i.e. Museumcijfers	Publicly available	Published once a year	3	Tracking crowdfunding is also interesting, see Erfgoedmonitor

Project Consortium:

Germany

Project lead:

BGZ Berliner Gesellschaft für internationale Zusammenarbeit mbH

Pohlstraße 67

DE – 10785 Berlin

+49 (30) 80 99 41 11

info@bgz-berlin.de

www.bgz-berlin.de

https://spielfabrique.eu/

Austria

http://www.imz.at/

http://ars.electronica.art/

France

Italy

www.betaseries.com/

www.centrica.it

The Netherlands

www.beeldengeluid.nl