



Deliverable 2.2
Overview of different
digital solutions to
generate and analyse
data



cci thrive



Project Information

“CCI Thrive – Bespoke Business Models and Innovative Practices of Cross-Sectoral Cultural and Creative Collaboration” (Project Nr. 101059476)

Project Website: <http://cci-thrive.eu/>

CCI Thrive will explore the exciting potential of cross-sectoral business co-operations and co-productions to shape their own digital transformation and future. The usage of data-driven technologies will attract new audiences and enliven existing ones.

Publisher & Project Coordinator

BGZ Berliner Gesellschaft
für internationale Zusammenarbeit mbH
www.bgz-berlin.de

Project Partners

SpielFabrique 360° UG (Haftungsbeschränkt)
Ars Electronica Linz GmbH & Co KG
IMZ Internationales Musik&Medienzentrum Wien
Stichting Nederlands Instituut Voor Beeld En Geluid
Centrica SRL
SAS BetaSeries

Authors

Rasa Bočytė (NISV)
Kelly Hazejager (NISV)
Johan Oomen (NISV)
Maarten Zeinstra (NISV)
Matthes Lindner (SPF)

Reviewer

Rémi Tereszkiewicz (BS)
Marco Cappellini (CEN)

Photo credits

@iStock-1336334833/carloscastilla

Berlin, 31.05.2023



Executive Summary

Data sharing between cultural and creative industries (CCIs) can stimulate collaboration and lead to new business opportunities, especially benefitting smaller actors. But it still comes with many reservations and barriers - fear of losing competitive advantage, lack of digital infrastructure to share data, lack of skills to draw insights from data, to name a few.

In order to overcome the barriers and risks of data-sharing, there is a need for compelling scenarios and proof of concepts that demonstrate how data-sharing can lead to new opportunities to generate and capture economic and societal value. This report presents cross-sectoral data-sharing scenarios and identifies digital infrastructure and tools needed to support them.

Using a design-based methodology, the CCI Thrive consortium partners identified three potential scenarios that introduce new opportunities for CCIs based around the aggregation of data:

1. Discovering Opportunities for Cross-Sectoral Collaborations
2. Matching Creative Products with Audiences
3. Identifying and Understanding Underserved Audiences

For each scenario, the report outlines its added value, data needed to realise it and an exemplary user journey illustrated through mock-ups. To support the implementation of these scenarios in real-world business context, a reference architecture for data aggregation, analysis and visualisation has been developed, including a description of digital tools that can be used for it.

Table of Contents

- 1. Introduction 4
- 2. Methodological Approach 6
- 3. User Stories..... 7
- 4. Scenarios for Cross-Sector Data-Sharing 11
 - 4.1 Scenario 1: Discovering Opportunities for Cross-Sectoral Collaborations 11
 - 4.1.1 Scenario Description 11
 - 4.1.2 Market Analysis & Current Challenges 12
 - 4.1.3 Value Proposition..... 13
 - 4.1.4 Cross-Sectoral Data..... 13
 - 4.1.5 User Journey 17
 - 4.2 Scenario 2: Matching Creative Products with Audiences 24
 - 4.2.1 Scenario Description 24
 - 4.2.2 Market Analysis & Current Challenges 24
 - 4.2.3 Value Proposition..... 25
 - 4.2.4 Cross-Sectoral Data..... 25
 - 4.3.5 User Journey 28
 - 4.3 Scenario 3: Identifying and Understanding Gaps in Audiences 32
 - 4.3.1 Scenario Description 32
 - 4.3.2 Market Analysis & Current Challenges 33
 - 4.3.3 Value Proposition..... 34
 - 4.3.4 Cross-Sectoral Data..... 34
 - 4.3.5 User Journey 35
- 5. Reference/Information Architecture 37
 - 5.1 Information Architecture Fundamentals 37
 - 5.1.1 Components in information architecture 38
 - 5.2 Requirements for Cross-Sectoral Collaboration Information Architecture 40
 - 5.3 Proposed Information Architectures for CCI..... 42
 - 5.3.1 Scenario – Discovering Opportunities for Cross-Sectoral Collaborations and Co-Productions 42
 - 5.3.2 Scenario – Matching Creative Products with Audiences & Identifying and Understanding Underserved Audiences..... 47
 - 5.4 Implementation and Integration 49
- 6. Conclusions and next steps..... 51
- Annex 1: User Stories..... 52

1. Introduction

In the last few years, the contribution of cultural and creative industries (CCIs) towards addressing complex societal, environmental and economic challenges has been gaining increasing recognition. The 'New European Bauhaus' movement¹ announced by the European Commission in 2020 puts cultural practices, aesthetics and creative processes at the centre of meeting the Green Deal goals at a local level. The launch of the EIT Culture & Creative programme² signals a strong belief from decision makers and investors that CCIs can significantly contribute to the 'triple transition' - digital, societal and environmental.

To fulfil this potential it is critical to implement strategies that help CCIs overcome challenges that are currently hampering their activities:

- **Fragmentation.** The diversity of European CCIs is the strength of the sector but it has also contributed to the fragmentation of policies, funding mechanisms, intellectual property regulations and data collection practices.³ This complexity has created barriers to collaboration among CCIs around the many shared values. Equally, it is hampering the formulation of evidence-based strategies that can effectively address existing gaps and challenges in the sector. These barriers are particularly detrimental to small companies and self-employed individuals who could benefit from the pooling of resources and increased visibility gained through collaboration and knowledge sharing.
- **Competition from non-European companies.** European CCIs face massive challenges when trying to gain a competitive advantage over non-European counterparts. The existing global imbalances have been further exacerbated by digital technologies and especially content platforms. While there is no shortage of European productions, they struggle to gather enough budget to gain visibility and reach their target market.
- **Lack of business intelligence.** Cultural and creative professionals still overwhelmingly lack business intelligence skills as well as tools and strategies to use data for business decisions. There are quite big differences in the capacity of CCIs to approach digital transformation and implement data-driven workflows that can bring competitive advantage and lead to new or improved products.
- **Lack of access to specific Venture Capital funds.** CCI companies lack dedicated funds to support the development phase. There are first signals of a more focused attention, for example, Vienna-based [CultTech](https://culttechaccelerator.org/) accelerator that is trying to create awareness around CCIs potential⁴, but there is a lot to do in this direction.

CCI Thrive believes that **cross-sectoral data sharing can lead to increased collaboration and knowledge transfer among CCIs which is crucial for a flourishing digital cultural and creative economy** in Europe. By sharing business intelligence data, CCIs can more easily enter into joint projects that benefit from cross-sectoral expertise to reach new markets or identify new business opportunities. This can especially benefit smaller actors who can gain exposure and ease their entry into the market by collaborating with others.

¹ https://new-european-bauhaus.europa.eu/index_en

² <https://eit-culture-creativity.eu/>

³ UNESCO, *Cutting Edge: The creative economy: moving in from the sidelines*, 2021. <https://www.unesco.org/en/articles/cutting-edge-creative-economy-moving-sidelines>

⁴ <https://culttechaccelerator.org/>

The goal of this report is to propose scenarios that demonstrate the benefits of such cross-sectoral data-sharing and illustrate how they could be implemented through digital technologies. They aim to address some of the current reservations and barriers that prevent CCIs from data-sharing - fear of losing competitive advantage, lack of digital infrastructure to share data, or lack of skills to draw insights from data, to name a few.

Using a design-based methodology, the CCI Thrive consortium partners identified three potential scenarios that introduce new opportunities for CCIs based on the aggregation and sharing of data: *Discovering Opportunities for Cross-Sectoral Collaborations*, *Matching Creative Products with Audiences*, and *Identifying and Understanding Gaps in Audiences*.

This report is aimed at CCI practitioners as well as CCI network organisations and decision-makers. CCI practitioners will get insight into how a cross-sectoral mindset could benefit their businesses and bring new opportunities through collaboration with other CCI actors. CCI network organisations and decision-makers will gain a better understanding of what support instruments and infrastructures are needed to propel cross-sectoral exchanges.

2. Methodological Approach

To arrive at the cross-sectoral data-sharing scenarios and design a reference architecture to support them, project partners used a design-driven methodology. Through interactive sessions, the aim was to tap into the expertise of consortium partners who represent five different CCI sectors - cultural heritage, gaming, film and television, (media) art, performing arts and music. These sectors are defined by different business models and varying degrees of maturity in relation to data-sharing. While acknowledging these differences, it was equally important to agree on shared values, priorities and business needs that could be translated into concrete scenarios for cross-sectoral synergies. The following key steps were taken in the process:

- By exchanging existing practices and barriers for data-sharing in their respective domains (this is presented in the accompanying deliverable [D2.1 Recommendations for shared cross-sectoral metadata](#)), partners identified a number of areas where data-sharing could benefit all participating CCI actors;
- Following this, partners produced personas and user stories describing specifically how cross-sectoral data-sharing would benefit CCI professionals - these are described in chapter 3;
- From the analysis of personas and users stories, three scenarios emerged that demonstrate how cross-sectoral insights can help CCI professionals benefit from different opportunities - (1) discovering opportunities for collaboration by finding complementary activities in other CCIs, (2) delivering products targeting better their audiences through a holistic understanding of cultural and creative consumption, and (3) identifying market opportunities and audience gaps in the offering of CCI products and services. Each scenario was illustrated with a user journey that practically demonstrates the use of cross-sectoral data. This is the focus of chapter 4.
- In parallel to the previous step, research was carried out on identifying an information architecture and digital tools that would support the implementation of these scenarios. This is presented in chapter 5.

This document also sets the stage for the next phase of the CCI Thrive project where project partners will build demonstrators that act as proof-of-concepts of the envisioned scenarios.

3. User Stories

This chapter presents the work carried out by the consortium to develop user stories exploring how a cultural and creative professional would make use of a cross-sectoral data-sharing platform. Nine user stories were developed based on the expertise of consortium partners and their professional knowledge of existing approaches to data in their sectors.

A design-driven methodology was applied to capture the demand side for cross-sectoral data-sharing. Earlier work by the consortium had explored what low-risk data each sector would be able and willing to provide.⁵ Low-risk data was defined as data that holds no critical competitive value, covers no sensitive issues, is publicly available, is reliable and remains valid.

The user stories were developed with the overview of low-risk data in mind, including creation and content data, consumption and engagement data, and data on external or high-level factors. Personas were created to facilitate imagining why a CCI professional would seek out cross-sectoral data and how the data should be presented to help them achieve their goals. The user stories describe what data insights are desired, therefore pointing to what databases and data processing would be needed to provide a relevant minimum viable product (MVP) for cross-sectoral data-sharing.

Each user story includes:

1. a profile of the imagined CCI professional,
2. their professional data needs,
3. their experience with data analysis,
4. and a narrative of what decision-making process(es) would be supported by data insights.

The full nine user stories can be referred to in Annex 1.

Professional Profiles

The nine user stories each described the professional profile of the imagined cultural or creative professional, listing names, job roles and the sector. The sectors represented matched the consortium partners and their expertise.

A wide range of job titles were described in the user stories, as illustrated in the word cloud (see Figure 1). The professionals imagined benefitting from a cross-sectoral data sharing initiative, were mainly producers and curators. These two job titles capture two main decision-making processes that could benefit from cross-sectoral data insights:

- what new content to create, produce, and/or invest in; and with whom;
- and what existing content to acquire and/or to promote from the catalogue.

⁵ See chapter 3 in [D2.1 Recommendations for shared cross-sectoral metadata](#)

Method of Interaction with Data

The nine user stories described what types of data, CCI professionals would expect to find on a cross-sectoral platform, as visualised in the word cloud in Figure 3. This takes into account the types of data that CCI professionals currently use, the comparisons they are looking for, and the gaps in data to address and support decision-making. Most of these data types correspond to the underlying goals for using cross-sectoral data outlined above.

The user stories include three main methods of interacting with data/data types:

- Data on past, current and future projects and/or content. These data are more valuable if enriched with topics, keywords and/or tags.
- Data on audiences and their market sizes in relation to specific sectors, genres and topics. These data would ideally include more granular information on characteristics such as age, preferences, price acceptance so that CCIs can segment for target audiences.
- Accessing analysis carried out on both these types of data to identify trends, potential projects, new target audiences, and to compare similar initiatives across CCIs.



Figure 3: word cloud capturing the types of data CCI professionals would expect to find on a cross-sectoral platform/database

Experience with Data

Previous experience with using data and data analysis to inform strategy and decision-making varies across CCI organisations and professionals. The user stories also imagined what the average CCI professional's capacity for using data and carrying out their own data analysis was. This was important to take further into the scenario development as the data have to be presented in a certain way to be accessible to a wide range of CCI professionals.

Most user stories described how CCI professionals currently collect information and insights by reading industry reports, attending conferences, taking advantage of peer and network knowledge, and reflecting on their personal professional experience.

As described in several user stories, some CCI professionals do have experience using databases such as internal databases that include sales and/or audience data. However, in most cases this experience will be limited, and a CCI-platform should not require any existing marketing or data skills.

Key Insights into CCI Professional Needs

The analysis of the four components in the user stories shows that there are two main types of data that CCI professionals need or want to access to inform their decision making.

The first main type of data is **industry data**, or data that describe activities within a CCI sector and is being collected and/or shared by CCI organisations themselves. This includes data on:

- Complementary activities across CCI sectors;
- Potential collaborators, possibilities for co-productions;
- Market comparables;
- What is being acquired and produced elsewhere;
- Other CCIs interested in your offering (content, skills).

The second main type of data of interest is **audience data**, or data that describe how many and what type of people engage with cultural content/events/service and trends in audience preferences. This includes data on:

- Target groups and their usages and motivations to help CCIs better understand them;
- What audiences are interested in and consuming as input for new productions;
- Topics, channels, type of interaction, technologies and platforms;
- Different types of content to demonstrate the potential of other media and methods of engagement via other assets;
- The performance of similar productions in the past to get more insight into successes and failures to learn from.

The next step within our design-driven methodology was to take the insights from the user stories to develop scenarios for cross-sectoral data-sharing that are feasible and relevant for the CCIs, as presented in the next chapter.

4. Scenarios for Cross-Sector Data-Sharing

The personas and user stories described above provided a broad overview of CCI professional needs that would be met with cross-sectoral data-sharing. From the discussions within the CCI Thrive consortium, it became evident that in many cases data collection and sharing practices in the industry are still relatively modest. While practices for data-sharing within a sector are more common (an example of are the European Data Spaces for Cultural Heritage and Media), there are still few examples of digital infrastructures that would support cross-sectoral data aggregation and analysis.

There are many barriers to data-sharing that require structural changes - for instance, large-scale industry investment in improving data collection and sharing infrastructures, revisiting data-sharing policies in CCI organisations, introduction of incentives to encourage data-sharing, etc. But given already existing infrastructures and databases, it is already possible to build proof-of-concepts that envision new business opportunities.

The goal of this chapter is to provide such illustrative examples that would demonstrate to the industry the benefits of cross-sectoral data sharing. It proposes three scenarios that respond to urgent needs in the sector: *Discovering Opportunities for Cross-Sectoral Collaborations*, *Matching Creative Products with Audiences*, and *Identifying and Understanding Gaps in Audiences*. The scenarios synthesise insights from the personas and user stories described in the previous chapter into concrete scenarios enabled by digital workflows. They have been designed to meet the needs of CCI professionals at different stages in the creative value chain - from the conceptualisation of the creative idea and finding funding opportunities, to production and internationalisation.

Each scenario describes what cultural and creative professionals can gain from using cross-sectoral data and what types of already existing data sources and databases can be used to support this. Scenarios are accompanied by illustrative user journeys that envision a digital platform providing CCI professionals with functionalities to gain meaningful insights into the aggregated cross-sectoral data. These speculative platform designs are meant to serve as inspiration for those who would like to start developing applications and data aggregation and analysis workflows with CCI data.

4.1 Scenario 1: Discovering Opportunities for Cross-Sectoral Collaborations

4.1.1 Scenario Description

The fragmentation and diversity of European CCIs speaks to the existing cultural richness that should be encouraged. At the same time this poses a challenge for CCIs who struggle to establish themselves in a sustainable way and scale enough to find / reach their target audiences. Collective efforts can help to overcome this.

This scenario explores how cross-sectoral data can be used to identify and stimulate collaborations between CCIs. By looking at industry trends - past and future - joint interests and opportunities can be identified. This scenario particularly targets small and medium organisations in CCIs as well as self-employed professionals who can gain competitive advantage by forging connections with other actors in the sector.

This scenario proposes how CCIs can utilise cross-sectoral data to:

- Stimulate collaborations and knowledge transfer between CCIs with a focus on finding commonalities (based on topics, genres, etc.);
- Use experiences and knowledge from the past to improve your current “pitch”;

- Identify collaboration opportunities at different stages in the creative value chain, including co-production and marketing co-investment (from IP acquisition to media budget. This is particularly relevant for smaller actors who can gain competitive advantage by forging collaborations with others (economies of scale);

Identify trends across CCIs that can support different steps along the value chain of creating new projects and reveal new opportunities for collaboration.

4.1.2 Market Analysis & Current Challenges

The practice of collaborations and co-productions in European CCIs is not new. In fact, there are a number of well-established instruments and strategies - primarily in the form of industry events or digital matchmaking platforms - to facilitate networking and collaboration opportunities. However, in most cases they are limited in cross-sectoral exchanges.

For instance, SpielFabrique organises a [co-production market](#) that brings together indie game studios from different countries and excelling in different (complementary) foci of expertise. A market like [Avant Premiere](#), initiated by IMZ, is a good example of developing collaborations between two sectors - music and audiovisual industry. It creates a meeting place for players from both sectors to develop projects together and licence them to distributors. [Cartoon Next](#) event is unique in connecting designers and animators with the gaming sectors, offering different matching platforms to help producers to find co-producers as well as buyers and distributors.

Traditionally it has been difficult for CCIs to venture into other domains due to lack of commonalities in workflows, content formats and target audiences. But the dynamically changing markets, their internationalisation, consumption habits and possibilities introduced by the latest technological innovations invite cross-sectoral approaches. For instance, complex digital tools like game engines are no longer exclusively used in the gaming industry but are also commonly applied across other CCI domains to create various content formats (film, television, cultural heritage). Similarly, the traditional film distribution strategy that is based on territories is becoming obsolete with global platforms and the development of content franchises. This pushes the industry needs to find alternative collaborators and distribution channels in order to reach their audiences. The growing role of audiences and fan communities is also asking for more holistic strategies to deliver cultural and creative experiences that spill over sectoral borders and include user generated content in creative productions.

Based on the existing collaboration and co-production markets, the following challenges and gaps have been identified:

- Existing CCI collaboration strategies are limited in connecting two or three industries;
- Most co-production events focus on the distribution of already finished projects. There is no venue that would provide cross-sectoral data on ongoing or future projects that would benefit from finding collaborators from other sectors;
- Smaller CCIs encounter high entry barriers when seeking collaborators for co-productions as they lack industry insights and training;
- Some existing financing instruments are tied to a specific CCI sector and discourage co-productions;
- For an international collaboration project, clearing distribution rights in terms of territories, timing and sub-license agreements is often difficult at no or low costs and needs multi-national legal expertise. The clearance of rights in all CCIs is a big obstacle in the development of cooperation.

Accessing information on who owns what is time consuming and costly and can create barriers to collaboration.

As a conclusion, due to the lack of experience breaking sectoral silos, CCIs often lack a good understanding of what they can offer to other sectors. There is a likelihood of „blind spots“, i.e. not seeing opportunities because of the siloed perspective.

4.1.3 Value Proposition

The cross-sectoral collaboration scenario would allow to understand whether there is a comparable for a prospective or planned production by initiating a search by theme and moving forward from there to understand in which sector (e.g. book, game, movie, comic), what genre (e.g. historical, documentary, fiction, fantasy), and the release data of a comparable (theme) production (i.e. a recent release, recently finished, in the pipeline etc.). These first search results could assist in the decision-making process during the ideation phase of a project, whether continuing with the original idea, adapting it or giving up in case of competition warnings.

If the data supports the decision of moving on with the idea, then cross-referencing it with data from comparable productions would facilitate cross-sectoral collaboration and foster use of the same experts and supporting services during the planning and production.

As different CCI sectors excel in different business aspects (e.g. game sector in community engagement, film in digital storytelling), cross-sectoral projects can lead to learnings that help all involved parties acquire new competences. This can be especially beneficial in disseminating the latest technological advances or new content formats from CCI companies already experienced with them to those who are only starting.

4.1.4 Cross-Sectoral Data

The following is a minimum set of data about a given CCI production/service/event that would be required to produce a MVP in this scenario:

- Type of CCI production (e.g., game, video, exhibition) and channels of distribution;
- Companies involved (if applicable also the mother company) and contact information;
- Countries (production/release location);
- Dates (e.g. release date, running dates);
- Target audience.

Additionally, the following data would further improve decision making in the proposed scenario:

- Amount of audience / visitors;
- Received funding or investment (and pending);
- Revenue levels / sales etc.;
- Number of releases in cinemas (screens), performances or exhibitions;
- Means of reach-out / publicity amount of awards;
- Level of audience or expert acclaim;
- IP franchise extension (e.g. marketing products with games or movies);

- Any marketing opportunities to explore further.

The challenge again is where to get the data from as here the data is not with the production company as above, but needs to come from data collating organisations, public statistics bodies, broadcasting companies, streaming companies, game publishers and other exhibiting/presenting organisation such as museums, theatres, opera houses, ticket vendors, funding agencies etc. And the data need to be up-to-date and reliable.

An overview of data sources from several CCI sectors that can be used in the MVP is provided in Table 1 below. This overview was made in consultation with the project partners. They regularly use these data sources in their work. The overview does not attempt to be exhaustive, but is useful to gain insight in the type, format and access conditions of the datasets available to the CCIs:

- A significant amount of data sources (IMDB Pro, Rightscentre and Cinando and others) are only available after paying a fee. The fee varies considerably between sources and providers. Even after paying a fee it is still to be determined whether the user agreements allow for structured reuse of the data that is provided by the service.
- Some public organisations (e.g. EGMUS, European Media Observatory and Eurostat) offer unlimited, free access to data. The [CC-O Creative Commons Zero public domain dedication](#) is used by some.
- In many cases, these public organisations also offer easy downloads of data in structured formats (csv, xls).
- Only a few (Europeana, Wikidata, The Internet Archive) provide open API access to all data sources. That means in many cases available data are snapshots rather than machine-readable 'live' data.

These findings make clear that in designing services for cross-sectoral collaborations, it is important to take into account the varying degrees of openness and accessibility of the data, update cycles and consider the needs and requirements of different stakeholders involved. It is important in the development of a comprehensive and up-to-date register of data sources to ensure efficient and effective use of data in the development of the MVP.

Table 1: Overview of data sources from CCI

Sector	Name	Type of data available	Description / limitations
General	Eurostat	European statistics to policy makers, businesses, researchers and the public at large.	Link: Eurostat (europa.eu) and Database - Culture - Eurostat (europa.eu) Access: Free
	Wikidata	Collaborative, multilingual knowledge graph that serves as a central repository for structured data, enabling interconnected and accessible information across various Wikimedia projects and beyond.	Link: Wikidata Access: Free

Sector	Name	Type of data available	Description / limitations
	Content.Agent	The b2b trading platform from the IMZ: content / production / persons / versions / distribution	Link: Content.Agent - the global b2b film marketplace (contentagent.net) Access: 100€ per year
	European Audiovisual Observatory	Market and legal information on the audiovisual industry	European Audiovisual Observatory (coe.int) Access: Free
Design	Dribbble	A community-driven platform that allows designers, illustrators, and other creative professionals to share their work	Link: Dribbble Access: Free
Festivals	Ars Electronica Archive	Archive of media art works and festival content of Ars Electronica	https://ars.electronica.art/archive/en/
	Ars Electronica Festival catalogues	Documentation of Ars Electronica Festival and exhibited artworks. Festival catalogues (available in print and as online PDFs).	Link: https://archive.aec.at/print/ Access: Free
	Cinando	Festivals catalogues (created within the Festival de Cannes)	Link: Cinando Access: 129€ per year
Film	Lumiere (part of the OBS)	Available data on annual admissions to films released in European cinemas	Link: Lumiere (coe.int) Access: free
	The Swedish Film Database	Content / persons /	Link: The Swedish Film Database – the best source about Swedish film Access: free
	Europa Cinemas	Content	Link: https://www.europa-cinemas.org/en/ Access: free

Sector	Name	Type of data available	Description / limitations
Film and Television	IMDB Pro	content / production data incl. persons	Link: IMDbPro Official Site Start Your Free Trial Access: 20\$ per month subscription (?)
	The Movie Database (TMDB) and TVDB	Content / persons	Link: The Movie Database https://thetvdb.com/ Access: free
Photography	Behance	Creative professionals showcasing their portfolios, connect with other creatives across various disciplines such as graphic design, photography, illustration, etc.	Link: Behance Access: free
Games	Steams	Content	Link: Steam Access: fee
Heritage	Time Machine	Overview of projects of the Time Machine Partners	Link: www.timemachine.eu Access: free access
	Internet Archive	Large archive of webpages, photo's, books, audio recordings.	Link: Internet Archive: Search Engine Access: free
	The European Group on Museum Statistics EGMUS	Collection and publication of comparable statistical data on museums (visitors, exhibits etc.)	Link: EGMUS - Home Access: free access, data can be exported as in csv format.
	Europeana	Metadata and collections	Link: www.europeana.eu Access: free, API access
Music	DDEX	Information on standards	Link: DDEX

Sector	Name	Type of data available	Description / limitations
			Access: free access to overview of standards
	PPL	licensing / artists / content	Link: PPL Access: free access
	Digital Music Observatory	Links public datasets in order to provide a comprehensive view of the European music industry	Link: Digital Music Observatory (dataobservatory.eu) Access: free access
Publishing	Taleflick	Data on available books for film adaptation	Link: TaleFlick: Empower Your Story Access: 88\$ subscription / year
	Rightscentre	Film and television rights for over 125,000 books and other literary properties	https://www.rightscenter.com/ Access: various pricing levels

4.1.5 User Journey

To demonstrate how the above-described scenario could be practically implemented, a number of mock-ups have been co-created by the consortium to demonstrate a user journey engaging with cross-sectoral data in a web-based interface. They demonstrate how a CCI professional would be able to interact with data, what insights they could draw and how they could operationalise them to make business decisions.

The mock-ups described below try to answer three questions that a CCI professional would want to answer when looking for industry collaboration opportunities:

- *Who else is working / has worked on in the past / is planning to work on the topic I am interested in?* - this can be used to initiate co-productions between actors who are keen to work on one project or multiple complementary projects;
- *Who in other CCIs can offer knowledge on this topic based on past experiences?* - CCIs can use such knowledge to pool resources and built on expertise that other actors in the field already have;
- *What can I learn about the success of this topic in the past that I could use for my pitch?* - This question speaks to knowledge transfer between CCI activities, helping actors to identify how future projects in CCIs could benefit from lessons learned in the past.

The sections that follow describe four interfaces of an envisioned digital platform that aggregates data from existing databases or data sources and allows CCIs to interact with this data via various visualisations.

CCI Data Market

A prerequisite to facilitating future collaborations is not only an overview of past activities but also of projects currently under development as well as the possibility to work on projects on a given topic in the future. For past activities, we can primarily rely on existing databases that aggregate data about completed projects. Yet, there might be certain gaps to fill, especially about smaller-scale projects or specific types of activities about which there is little data. When it comes to projects that are currently being developed, such data is very rarely openly shared. An exception to this is the [IMDb](#) platform that aggregates film industry data on past and future productions, allowing subscribed companies to share and continuously update information about the status of their projects.⁶ Such present- and future-oriented information could be particularly useful in relation to CCI productions that take a long time to reach the market (e.g. games, exhibitions) as it would give other CCI actors a window of opportunity to identify potential synergies and initiate complementary projects.

An example of a more open alternative for gathering similar data would be matchmaking platforms, such as the one provided by the European Commission's funding portal where entities can express their desire to work on a given funding call or the [F6S](#) platform that helps start-ups find investors, collaborators and funders. Their limitation is that these matchmaking platforms are attached to a specific deadline of a funding scheme rather than being open-ended to opportunities that might emerge.

A "CCI Data Market" could fill this gap. Through a survey tool (see Figure 4), CCI professionals (individuals and organisations) would be invited to manually contribute data about one of the following categories:

- *Past activity* - a project that has not been documented in other databases;
- *Current activity* - an ongoing project or production under development;
- *Future activity* - a production planned in the future.

⁶ It should be noted that IMDb prohibits any kind of interaction with their data except by using their interface (see <https://pro.imdb.com/subagreement#copyright>) therefore it would not be possible to use its data for cross-sectoral data sharing scenarios.

Figure 4: Interface of a “CCI Collaboration Market” survey

The survey would require a set of minimum metadata and give an option to provide optional additional metadata (see table 2). The gathered data would be visualised alongside data gathered from other databases as described in sections below.


Table 2: Overview of data sources from CCI.

Required metadata:


- Title
- Type (selection from a list of categories - exhibition, video game, short film, etc.)
- Description
- Genre (list of options differs based on the Type selected)
- Creator (possible to add multiple companies/individuals and include URLs to each) and type of the creator (*Individual, Company, SME, Public Organisation, etc.*)
- Dates (start date (creation or opening) and end data (if application))
- Location
- URL
- Target Audience (possible to select multiple options from a list, including *local, international, youth, adults, elderly, LGBTQIA+, visually impaired, etc.*)
- Target Audience Size
- Image

<ul style="list-style-type: none"> - Keywords (free choice) - Tags (possible to select multiple options from a list, including <i>Female-Led</i>, <i>Bechdel Test</i>, <i>Carbon Neutral Production</i>, <i>Gender-Balance Production</i>, <i>Diversity Equity & Inclusion Label</i> (see Scenario 3), etc.) - Contact Details
<p>Optional Metadata:</p> <ul style="list-style-type: none"> - Funding source - Language - Identifier

Apart from more standard information about the project (title, creators, etc.), such a database provides an opportunity to gather societally relevant and urgently missing data related to diversity, sustainability and social inclusion. This could be represented in the choice of options when selected Target Audience and Tags (see suggested options in table 2). A database that could be followed as an example is [CherryPicks](#) - a film and TV recommendation platform where accredited female-identifying or non-binary critics provide reviews and assign criteria relevant to the assessment of gender representation during the production process and in the narrative features (see Figure 5).

 **THE CHERRY CHECK**

A film receives a Cherry Check if it passes at least one of these three tests.

 **Bechdel**

A film passes The Bechdel Test if it features two named women characters, who talk to each other about something besides a man.

 **ReFrame**

A film earns the ReFrame Stamp for demonstrating success in gender balance by hiring female-identifying people in key areas of production.

 **F-Rated**

A film is awarded an F-rating if it is written and/or directed by a woman.

(A film receives a CherryCheck according to the independent criteria of the external organizations listed. Note: Bechdel & ReFrame only apply to narrative features, not documentaries.)

Figure 5: *CherryPicks platform criteria for film and TV productions*

To motivate contributions and attract a critical mass of entries, such a CCI database should clearly articulate and demonstrate the benefits of data-sharing. A collection of success stories or case studies could be presented to demonstrate how such data-sharing can lead to collaborations that generate economic and societal impact.

The data manually submitted via the interface described above could be presented together with the data aggregated from various databases via a number of visualisations that allow CCIs to draw insights about collaboration opportunities. This is described in the sections that follow.

Visualisation of Related Activities and Collaboration Opportunities.

As users enter the platform, their starting point would be a search bar. For a specific query, users could enter a search keyword they are interested in (for instance, they might be interested to see all CCI productions related to “climate change”) or provide an advanced query (specify that they are interested in the topic “climate change” in location “Netherlands”). For a more exploratory search, they could choose from one of the provided categories that cross across CCIs (for instance, genre, target audience, type of funding). A number of filters (see left side of Figure 6) could help users to navigate the data and narrow down the search results:

- CCI sector - this would allow to compare productions in different sectors or zoom in to opportunities for collaboration with a specific sector;
- Audience size - this filter would help CCI find collaborators who aim for similar target audiences and reach, using a benchmark tool as developed further in scenario 2 (see section 4.2);
- Dates - this would allow to narrow down the search results to a specific timeframe, both in the past and near future (for productions currently under development).

The central element of an interface would be a visualisation presenting the filtered data. For this, we proposed three possible views that could support the CCI collaboration scenario: geographic mapping, timeline view and list view.

Geographic mapping

Geographical information is of particular interest to CCIs who are eager to adapt their productions to international markets across Europe. However, the specialisation and capacities of CCIs as well as the preferences of their audiences vary per country. Therefore, a cross-sectoral mapping would provide insights into trends around the continent, allowing CCIs to find the right collaborators for their productions. This could particularly bring visibility to regional and local productions that are not known outside of their immediate area of coverage.

A visual map (Figure 6) lends itself for a quick overview for a range of pertinent data. It could indicate whether specific regions have a higher potential for a specific topic or genre, audience interest in certain types of cultural or creative products, language presentation (e.g. are subtitles common or do you need to synchronise), cultural affinity (e.g. trend in Nordic or French thrillers or crime novels, Italian comedies etc.), public funding opportunities (e.g. regional priorities or focus such as indigenous movies, or gender-balance), reach-out of the distribution platform (existing number of cinemas, museums, theatres, concert halls etc.), internet connectivity, and a lot more.

Next to the visualisation, a user would benefit from key insights into data (see right side of figure 6). This would provide a high-level overview of activities related to the search query.

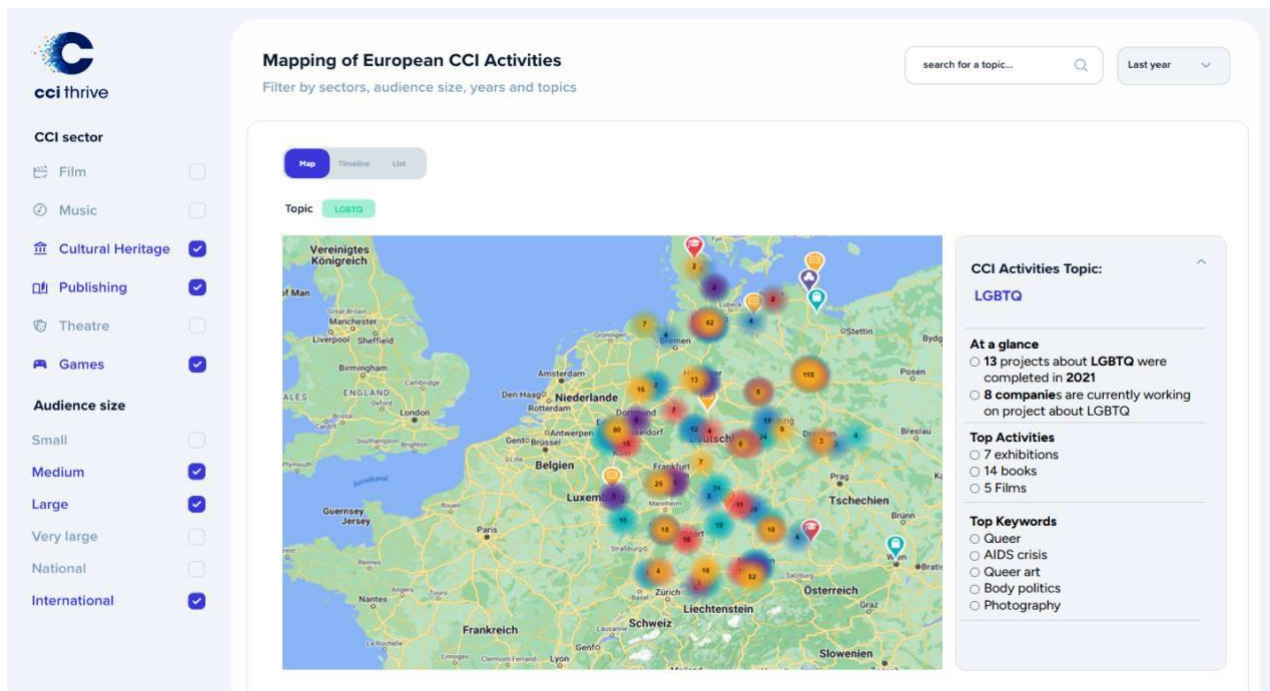


Figure 6: Visualisation of geographic distribution of CCI activities

Timeline view

Temporal information is equally pertinent when planning future activities. If a particular topic has been particularly successful in the recent past, a CCI actor might seek collaborators in other sectors who already have experience with it to work on a complementary production. Likewise, a timeline view could help to see if the market is already oversaturated with productions on a given topic.

The mock-up showcases a temporal view where CCI activities are mapped horizontally against a timeline (Figure 7). All activities could be colour coded to represent different selected filters, allowing users to visually assess the scale of existing offerings across different genres, CCI sectors, etc.



Figure 7: Temporal visualisation of CCI activities

List view

A list view provides yet another alternative way to query the data (see Figure 8). While the geographic and timeline views provide a great way to explore large amounts of data, once the user zooms in on a more specific data subset, a list view would serve as a way to learn more about individual productions.

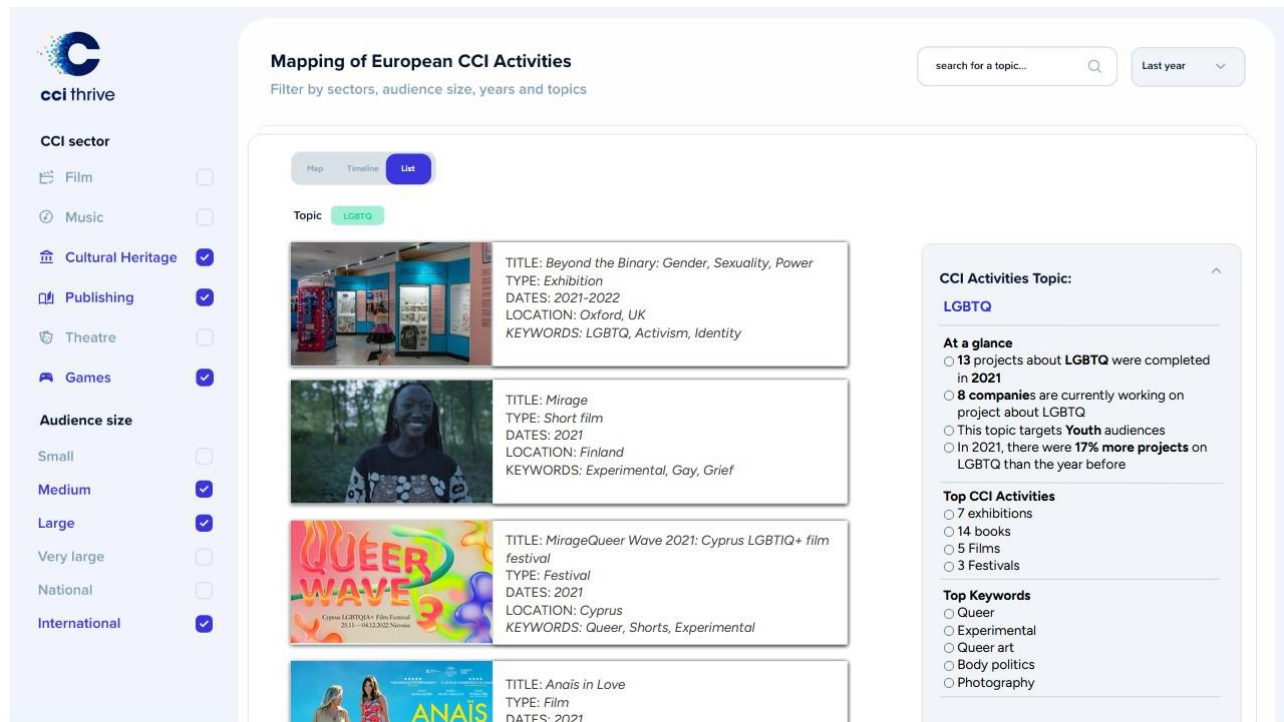


Figure 8: List view visualisation of CCI activities

The three visualisations should not be seen as separate but rather as complementary views on data between which users can seamlessly switch. For instance, using a timeline view, a CCI could identify an opportunity to work on a specific topic but the geographic view might reveal that the topic is not relevant in the geographic region where the CCI is active.

The challenge with the three visualisations presented is that if they reflect the market size, they need to rely on a benchmarking approach which in turn requires a quantifiable input. The more occurrences you have per region for a filtered search, the larger the circles shown on the map. However, those benchmarks are sector-specific. While a full house in a German cinema is approximately 200 seats, a full house in a concert house is easily 1.500 seats. A successful movie will, however, by far outreach the number of concert goers for several reasons which has nothing to do with the question whether the film was more successful than the concert. Thus, while 10,000 concert goers for a specific production might be a high success indicator, the same amount of movie goers would be a failure indicator.

Hence, each sector would need to define their success indicators and class them in categories (for instance, from 1 to 10 or A to F) to be able to compare them between CCIs. As not every sector is using standardised and comparable success indicators or categories, the way how the success of a creative or cultural production is measured is indicative of the business orientation.

When considering success metrics, another important aspect to consider are the differences between public and private CCIs. Organisations that act as curators and custodians of cultural heritage usually rely on public funding mechanisms to ensure that public values are prioritised over profit orientation. The metrics they deploy (e.g. educational goals, outreach to specific audiences groups such as refugee communities) might differ from a commercial understanding of success. Looking out of the box, the sectoral

differences in benchmarking strategies and assessing success could provide a valuable input for both the business oriented and public sector attitudes and operation strategies - for both are interested in a success based on the quality of their production and at the same time in their capacity to secure funds for the next high-quality production.

4.2 Scenario 2: Matching Creative Products with Audiences

4.2.1 Scenario Description

Content consumption in CCI sectors like film, television, gaming and music has become increasingly linked to a personalised profile. In this context, European CCIs are struggling to compete with CCI players that collect data on individual user consumption habits and use it to leverage insights for future investments and marketing. Cross-sectoral data sharing can provide CCIs with access to aggregated and enriched audience data.

This scenario explores how existing data on audiences can be connected across CCIs to form a more holistic (cross-sectoral) image of cultural consumption. As well as supporting CCIs in the identification of target audiences and the investment in future productions, data collection practices on audiences can be shared and standardised to improve insights as this is currently fragmented and unstandardised. This scenario also introduces how a cross-sectoral model of user-generated audience data could be developed to benefit both individual users and CCIs.

While this scenario argues that all CCIs can benefit from such cross-sectoral audience data-sharing, it particularly targets SMEs that do not work with/take advantage of specific user data, have a small catalogue on which to collect data and/or have a small target audience group.

This scenario proposes how CCIs can use cross-sectoral data to:

- Identify target audiences for content by understanding trends in cultural consumption across CCIs;
- Vice versa - for a specific audience, identify what content will be of interest and in what ways/formats these audiences can be most effectively reached;
- Create cross-sectoral personas on cultural consumption to provide a more holistic personas than currently used by the specific sectors to inform communication strategies and collaboration decisions.

4.2.2 Market Analysis & Current Challenges

Audience consumption data is a core component of how CCIs measure and communicate their success, using numbers such as (ticket) sales, physical visitors, viewers and/or social media followers. Many CCIs want to collect more in-depth knowledge on their audiences to support decision-making for future productions, marketing campaigns and strategic positioning. Some sectors currently use methods that involve carrying out user research, conducting surveys and developing personas to collect qualitative information on audiences. However, many CCIs are still starting or undergoing the transition to becoming more data-driven in understanding their audiences.

Some CCI players have platforms where a catalogue is connected to a specific user and their consumption.⁷ This data is leveraged to first offer attractive products to their audiences, such as tracking tools, recommendations, and crowdsourced reviews, incentivising audiences to keep contributing their

⁷ This includes consortium partner BetaSeries, and platforms such as Storygraph and Steam

consumption data. Secondly, the aggregation of these audience data and the ability to filter for combinations of parameters (such as age, gender, genre, average review, etc.) provides insights that can help predict future trends in consumption, identify opportunities for investment, and decide on directions for marketing. However, this potential for predicting trends, identifying new markets and targeting communication strategies is not accessible to all CCIs.

Collecting, connecting and analysing audience data is more challenging than working with the industry data explored in scenario 1 due to its fragmented and non-standardised state. CCIs have many different ways of measuring their audiences, with varying capacities to capture specific audience characteristics. When audience characteristics are recorded, different characteristics are prioritised and considered more useful, making it more difficult to compare data. CCI sectors and organisations also differ in their goals for audience engagement and their potential to reach audiences, presenting another challenge in directly comparing audience data.

The following gaps and challenges have been identified for how CCIs use data to understand audiences:

- Detailed trends, when available in a sector, are not accessible to all CCIs due to cost barriers on available studies or for some CCIs unavailability of this data for sharing (high risk data);
- CCIs have different ways of measuring audiences, prioritising different characteristics;
- Audience sizes vary per sector and within a sector. Different CCIs have different goals and potential to reach audiences, so data needs to reflect these differences;
- When collecting data on people and their consumption habits, GDPR and other ethical considerations must be complied with;
- Lack of skills and training in CCIs, especially small entities, to analyse and use these types of data.

4.2.3 Value Proposition

The matching audiences with creative products scenario will help small and medium-sized CCIs to better understand their audiences, their interests and the ways in which they engage with cultural and creative content. By sharing data on how many people are engaging with or consuming a specific instance of cultural content, trends in the popularity of topics and genres can be identified, contributing to investment and promotion decisions.

CCIs looking to identify a target audience for an existing or future production will be able to access data on the characteristics of the audiences interested in the relevant topic, such as age groups and gender. For CCIs who wish to target a specific audience, they can identify potential topics, platforms and other characteristics in order to develop a more relevant production/event/service.

4.2.4 Cross-Sectoral Data

This scenario makes use of four types of data to produce a MVP:

- Catalogue data of (recent) content per sector, with topics and genres extracted;
- Numbers of audience consumption of a given CCI production/service/event, organised in benchmarked categories to be comparable with other CCI's data and also to be available as low-risk data;
- Audience characteristics: age, gender, country, manner of consumption;

- Demographic data for audience comparable.

Additionally, the following data would further improve decision making in the proposed scenario:

- Language availability;
- Further audience characteristics: educational level, income, price acceptance, time budget, etc.;
- Cross-sectoral user-generated data linked to a profile: tracking, wish lists, reviews.

In addition to the data described in Scenario 1, the table below lists existing databases that can contribute to this scenario, the sectors they include, the types of data they offer, and their limitations. It is worth noting that while data required to implement Scenario 1 is already widely collected and made available across many CCIs, this is not the case with Scenario 2 as the data needed to implement an MVP are much more sparse and fragmented.

Table 3: Overview of data sources from CCI.

Sector	Name	Type of data available	Description/Limitations
Publishing	Goodreads; StoryGraph, Bookly; etc.	User-generated tracking of consumption	Would have to get such a company onboard with sharing their data. Alternatively target users for cross-sectoral tracking.
Publishing	WordsRated https://wordsrated.com/books-stats/	Anonymised statistics of audience consumption	
Cultural heritage	https://www.art-rabbit.com/ ; TripAdvisor & Museums ; getyourguide (Culture experiences); civitatis (same as the previous more focus in the spanish market)	User-generated tracking of consumption	Would have to get such a company onboard with sharing their data. Alternatively target users for cross-sectoral tracking.
Cultural Heritage	Enumerate Europeana	Visitors flow (exhibition/concert/theatre)	For some heritage institutions this data is in annual reports

Sector	Name	Type of data available	Description/Limitations
	(other surveys: Art Newspaper)		
Cultural Heritage	https://www.theaudienceagency.org/resources/feature-insight-into-museum-visitor	visitor profile	
Film & TV	Betaserries.com and other TVdestination sites with tracked profiles	Audience preference profiles (gender, age-bracket, location,...)	Generic, not personalised
Film		Box office/tickets sold	
Film & TV	Rotten Tomatoes	User comments and ratings	
Games	Steam Review Explorer	User comments and ratings	Anonymised, not too much filtering based on audience.
Games	Newzoo	Audience profiles of all kinds	Pay-walled
Games	Sullygnome for Twitch stats	Statistics for Twitch	
Games	https://quanticfoundry.com/reports/	model of gamer motivations based on psychometric methods and data from over 1.25 million gamers.	Pay wall.
Games	GameStats	Sales Data from Steam	Based on publicly available data. Only estimations and very general, limited to one platform

Sector	Name	Type of data available	Description/Limitations
Games	HowLongToBeat	Sales Data and statistics across platforms fed by gamers.	Limited numbers of games, not representative of the market.
VOD/streaming		profile/account data	High risk data not shared by platforms
VOD/streaming		viewing habits/content preferences	High risk data not shared by platforms
VOD/streaming		ratings	subjective, reliability debatable
Music	www.last.fm	User tracking and ratings of music. Shows real time number of listeners for top music	Would have to get such a company onboard with sharing their data. OR target users for cross-sectoral tracking.

4.3.5 User Journey

Complementing the user journey in Scenario 1, we envision that audience data could be made available via the same digital platform, allowing users to switch between different data views. Given the above-mentioned sparsity of data to support Scenario 2, the proposed user journey below describes desirable functionalities that at the time of writing could be implemented only on a smaller scale - for instance, comparing data between two or three CCI domains that have mature audience data gathering practices.

Mapping Audience Sizes

Audience size data can provide insights into the popularity of topics, genres and platforms. While many sectors wish to understand their audiences at a more detailed level and gain insight into their specific characteristics, the fragmentation of this data across sectors means that a first step is to compare audience sizes cross-sectorally. To encourage CCIs to share data on their audience sizes and to support comparison, audience sizes can be segmented per CCI sector (categories 1-10 or A-F, for instance). This can be achieved by collecting data on audience sizes for specific instances of CCI productions. Comparing the popularity of topics cross-sectorally across time highlights topics that may already have found an audience in another sector. Using audience size data and presenting it cross-sectorally is a feasible first step to developing a better understanding of CCI audiences.

This interface (see Figure 9) answers: *How many people are engaging with or consuming a specific instance of cultural content, or category of content (genre, topic, country of production)?* The user would be able to adjust a number of filters (sector, country, language, date range) to visualise the popularity of a topic across time, broken down and benchmarked per sector. An explanation of the benchmarking would be necessary to enable users to draw meaningful insights from data. The interface below plots audience sizes from niche to mainstream as an example of how to communicate the scale, but other methods can be used to explain the benchmarking.



Figure 9: Visualisation of audience sizes per topic

Cross-Sectoral Audience Database

Capacities and skills to work with data and databases vary across CCIs. Each of the scenarios described takes this into account, and many of the interfaces attempt to be as accessible as possible to a wide range of users. The interface depicted in Figure 10 proposes how users could be given more control and freedom over the data, by enabling them to make a selection of the audience data available, view a data visualisation, and if they wish to, download the datasets themselves. This could support, for example, the development of cross-sectoral audience personas. The types of data that they would be able to access include:

- Sector;
- Popularity (aggregated per place);
- Audience characteristics connected to consumption numbers (age, gender);
- Audience retention.

As already outlined above, this scenario would heavily depend on data being made available for such uses, being standardised across sectors, and being valid and frequently collected/updated to be of use.

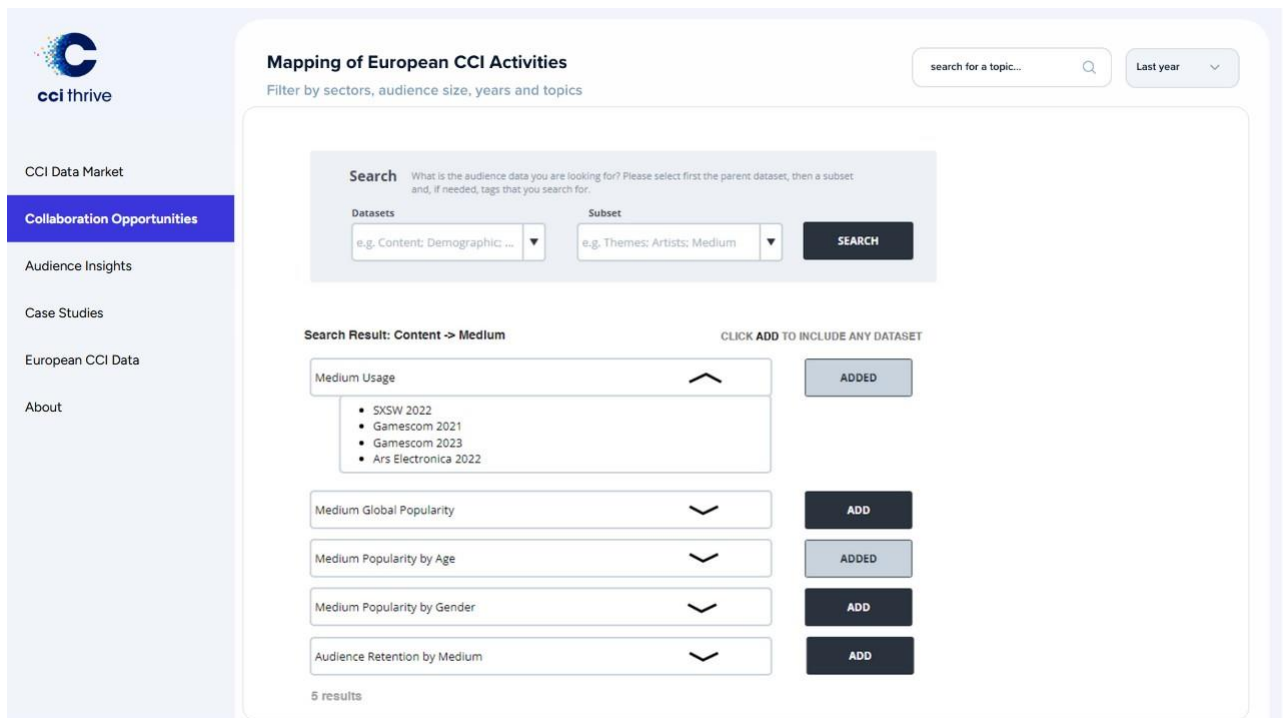


Figure 10: Interface showing how users could access available data on audiences

Cross-Sectoral Audience Characteristics

As mentioned above, there is the identified challenge of the monopoly on audience insights of CCI actors with large catalogues collecting data connected to a user profile. CCI Thrive proposes how existing data can be shared, connected and analysed to gather insights that are relevant cross-sectorally. Actors in sectors such as gaming and music deliver their content through B2C platforms where users must create a profile, enabling the tracking of consumption. This is also true for organisations that have shifted to a subscription business model tied to a specific user profile.

However, as not all CCIs collect detailed audience data, let alone tied to a specific person, the interface depicted in Figure 11 demonstrates how a cross-sectoral platform could present the insights gathered within each sector on audience characteristics. For each sector, available audience insights would be shared, identifying whether they are based on quantitative or qualitative analysis. For the sectors where it is available, insights gathered from user-generated data would be integrated, and would also be used to demonstrate the potential insights that other sectors can aspire to.

The interface allows the user to filter for:

- Audience characteristics such as gender, age, language; when available;
- A selection of one or more sectors;
- Dates to allow the user to choose a range of time.



Figure 11: Collection of Cross-sectoral Audience Insights

A cross-sectoral data-sharing platform could further enrich its data collection and analysis by targeting both B2B and B2C users in order to fill some of the gaps in audience data. One of CCI Thrive consortium members, BetaSeries, does this by enabling B2C users to track and rate their TV series consumption (mainly on SVOD & AVOD platforms), and then providing to B2B media players a paid access to the platforms audience insights that are today not shared directly by these players.

The consortium agreed that extending this model to other CCIs and building a B2C platform and targeting the general public to use it at a large enough scale to provide useful insights is a desirable future feature. A possible interface (see Figure 12) is still shared here to demonstrate what cross-sectoral data would be needed.

This interface makes use of three elements: (1) the cross-sectoral catalogue (data about CCI productions described in Scenario 1), (2) user-generated data (e.g. films and TV series views, festival or exhibitions attended, etc.), and (3) a recommendation technology that would match a person's profile with productions from across CCIs. Such an interface would be dependent on attracting enough users to provide their data in order to produce both relevant recommendations for the individual user and data insights for the business user.

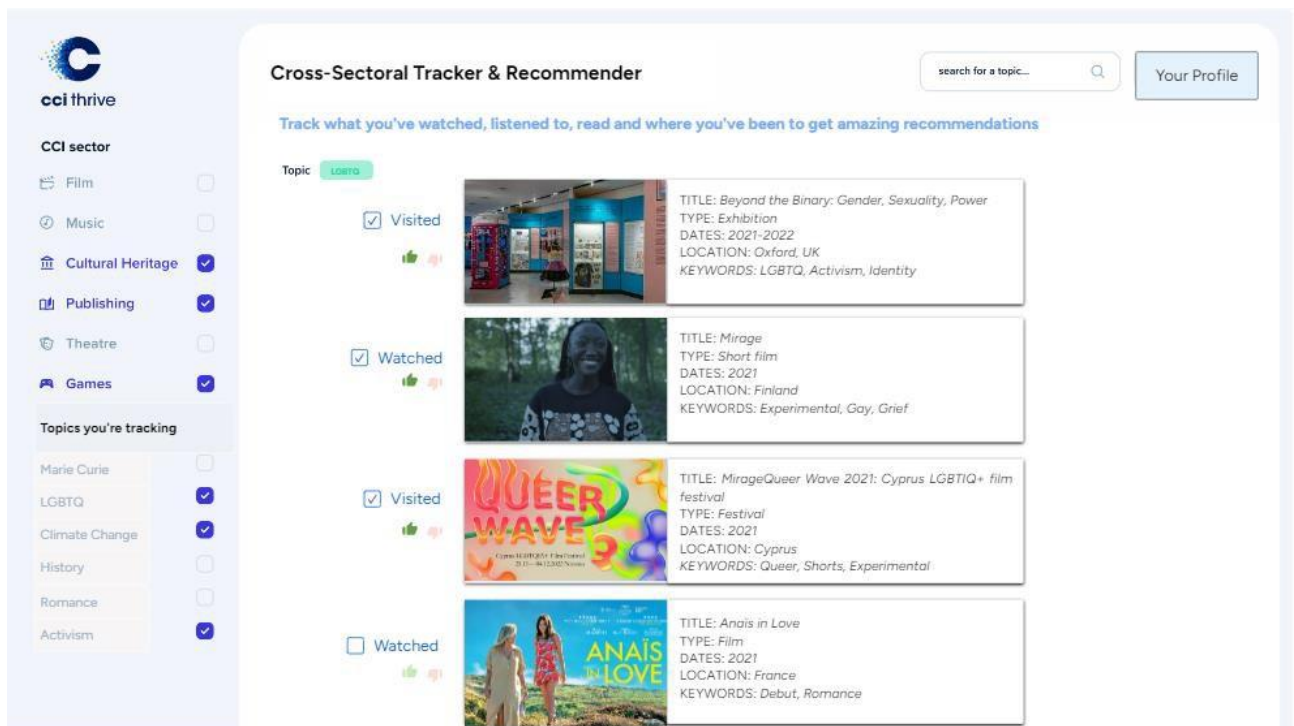


Figure 12: Interface of how a user would track their cultural consumption and get recommendations

4.3 Scenario 3: Identifying and Understanding Gaps in Audiences

4.3.1 Scenario Description

Many European CCIs recognise the importance of and are working to improve diversity, equity and inclusion (DEI) in their organisations and the content they produce. This is supported by the ratification of the idea that participation in the cultural life of the community and the arts is a universal human right in European policies.⁸ An important step to be able to address this is to identify across CCIs gaps in audiences who are not being reached and/or engaged with and understand if CCIs could bridge those gaps by tailoring their offerings to audiences who are currently underserved.

This scenario builds on Scenario 1 and 2 and explores how data on upcoming projects and cross-sectoral data on audiences and their engagement with a range of cultural content can be used to form a better understanding of who is underserved, underrepresented or missing. Underserved audiences in the context of CCIs refers to people that are not being reached by creative and cultural productions, whether this is due to lack of targeting, lack of offering that is relevant to certain communities or lack of access. CCIs can then respond by adapting their portfolios to these audiences, removing barriers to access or by initiating new productions that appeal to specific underserved groups. This use of cross-sectoral data can help CCIs fulfil their institutional social missions, identify important market opportunities, and respond to policies on DEI.

It should be noted that niche audiences are different from underserved audiences. A niche audience could be a small albeit a very active community whose niche interests are already being met by small-scale CCI productions. Underserved audiences, on the other hand, constitute groups of people whose cultural needs

⁸ For instance, see the FARO convention: <https://en.unesco.org/human-rights/cultural-life>

or interests are overlooked. An example of this could be refugee or expat communities who might lack access to cultural programming available in their language.

This scenario proposes how CCI can utilise cross-sectoral data to:

- Identify and remove barriers to access that audiences are experiencing in a specific CCI and/or across CCIs (i.e. language, subtitling, cost, technology);
- Respond to quotas or policies on DEI in both their production change and content production with relevant audience information;
- Incentivise investment in content that targets an underserved or underrepresented audience.

4.3.2 Market Analysis & Current Challenges

Awareness of the importance of identifying and catering for underserved and underrepresented audiences is growing across European CCIs, but when strategies to tackle DEI do exist within an organisation, they are rarely data-led. Direction may come from European or national policy, specific funding calls or an organisation's social mission. For example, TV channels and VOD platforms in Europe must fulfil the quota of having 30% of their content be European.⁹ Updates on policies and quotas are normally distributed via newsletters and articles.

Most CCIs do try to understand who they are not reaching and assess if they should fill this gap. The audiences who are the answer to this question could, however, be highly engaged within a different CCI, such as classical music. So to actually identify an underserved audience, a useful market opportunity and/or fulfil the social missions of institutions, it is important to know across CCIs who is not being reached, requiring access to cross-sectoral data and information.

Identifying and understanding gaps in audiences is an additional layer of analysis to the insights into audiences described in Scenario 2. This scenario therefore also faces the same challenges of different methods for measuring audiences, the relevance of certain audience characteristics, and the ethical questions of collecting data on personal consumption.

In addition to these, the following challenges have been identified for identifying and understanding audience gaps:

- Gaps in audiences will often be represented by gaps in data. However, just because there is no data on a specific audience group, it does not mean they are not engaging within that sector;
- Underserved audiences are often people that have been marginalised in society and might self-identify or be identified as minorities. Many data categorisation practices of audiences perpetuate harmful stereotypes or exclude these groups from being visible in the data. At the same time, certain groups might not want to be represented in data due to previous harms caused by unethical data extraction practices.¹⁰ Ways forward for collecting and sharing data must be both ethically sound and feasible, and designed only in consultation with the communities concerned;

⁹ <https://www.europarl.europa.eu/news/en/headlines/society/20180920STO14026/audiovisual-media-meps-approve-new-rules-fit-for-a-digital-age>

¹⁰ For instance, *Indigenous Artists and Wikidata: Explorations and Consultations Report* highlights that some indigenous artists might not want to be visible on Wikipedia and how the standard ways of describing Wikidata entries can impose colonial structures on indigenous communities. <https://linkdigitalfuture.ca/research/indigenous-artists-and-wikidata/>

- The ability of a business to serve (many) underserved target groups varies, and can depend on how varied their catalogue is, their relative resources, and whether many people are already familiar with them and consuming their content/using their services;
- Currently employed demographic data collection and analysis practices tend to be very binary, prescribing people into a single predefined categories, not allowing for more complex persona profiles to be encoded.

4.3.3 Value Proposition

The identifying and understanding audience gaps scenario will help CCI professionals to understand who is currently not or hardly engaging with cultural and creative productions in any or most CCIs, and what measures can be taken to engage them. This can include what platforms and mediums they are engaging with, the topics and genres that attract this audience group, and if barriers to engagement, such as cost or language, exist.

This scenario further highlights what data needs to be collected by the various CCIs in order to better identify and cater to audience gaps and underserved audiences. This will also enable CCIs to better respond to audiences currently highlighted in DEI policies and funding opportunities.

4.3.4 Cross-Sectoral Data

The following is a minimum set of data that would be required to produce an MVP in this scenario:

- Data on a specific CCI production/event/service and target audience from Scenario 1;
- Data on audience consumption (both anonymised and user-generated) from Scenario 2;
- Demographic data for audience comparables.

Additionally, the following data would further improve decision making in the proposed scenario:

- DEI policies and quotas per sector and country;
- Funding calls identifying specific audience groups, i.e. refugees, youth, migrants, LGBTQ+ (also per sector and country);
- User-generated data tracking interest in topics of social change.

In addition to the data mentioned in Scenarios 1 and 2, the table below lists existing databases that can further contribute to this scenario, the sectors they include, the types of data they offer, and their limitations. The brevity of this list points to how data on underrepresented audiences and DEI policies, quotas and funding is generally not aggregated, with CCIs often being informed through newsletters and articles.

Table 4: Overview of data sources from CCI

Sector	Name	Type of data available	Description/Limitations
	All databases mentioned in scenario 1 & 2		Would need to extract topics related to underserved audiences

Sector	Name	Type of data available	Description/Limitations
General	European funding portal https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/home	Search calls for proposal and tenders by keywords	Would need to use several related keywords to 'underserved'
General	Compendium of Cultural Policies and Trends (https://www.culturalpolicies.net/database/)	43 European cultural policy profiles; including statistics, knowledge in text-form, quotes.	Updated in the long-term. Mentions funding, but not specific to DEI.
Film	https://www.filmdaily.tv/grants/film-grants-worldwide		
Gaming	IndiePlaza (Release planed in Aug 2023)	Database of all european public funding available and most private funding.	Search engine for studio according to their profile and their financing goals. The updating of the data is very demanding and the structuration of the data demands deep knowledge of the subject. Can be only partially automated.

4.3.5 User Journey

Identifying and understanding audience gaps and underserved audiences is in many cases an additional layer to the data described in scenarios 1 and 2. The mock-ups described there have integrated features or dedicated filters that contribute to this scenario. The user journey for this scenario tries to answer four questions that a CCI professional would want to answer when trying to identify and understand underserved audiences:

- *Who is currently not/hardly engaging with cultural content in any/most CCIs?* This would highlight gaps in data which may be connected to an underserved audience
- *What audiences are currently highlighted in DEI policies and funding opportunities?* This would inform CCIs in their development of target audience (profiles). It could also contribute to investment strategies and co-production plans.
- *What data needs to be collected in order to identify and cater to underserved and underrepresented audiences?* Data-led decision making for DEI is still in early stages. By highlighting what is possible with existing data in a limited amount of sectors, CCI professionals or organisations might be influenced to collect similar data.

- *What topics/content could be invested in to target audience gaps?* Potentially untapped topics, when seen cross-sectorally, may overlap with audience gaps.

Identifying Niche Audiences

Taking into consideration the challenges described above for identifying audience gaps, one way to identify a potentially underserved audience is to understand what topics have a niche versus mainstream audience, as done in scenario 2 (figure 9). This can be achieved by collecting data on audience sizes for specific instances of cultural content. Comparing the popularity of topics cross-sectorally across time highlights topics that may seem unsuccessful or unimportant within a specific sector; but can capture a cross-sectoral niche audience for content/productions/events. This identified audience can then be the focus of further research to understand whether it is an underserved audience, and if there are barriers to access and mismatches in interest.

Cross-Sectoral Audience Characteristics Comparables

The interface described in scenario 2 (figure 11) where general insights into audience characteristics is combined with existing user-generated data is also of use in this scenario. The additional layer for understanding audience gaps is to compare it alongside demographic data to highlight where certain audience characteristics are underrepresented, underserved or missing. This information can be used to identify market opportunities and/or fulfil CCI's social missions.

As a user-generated tracking platform for all CCIs is not very feasible at the moment, this interface makes use of existing data and the varying forms it comes in. Some sectors such as film and television, gaming and music collect this type of user-generated granular audience information, though access to it is often behind a paywall. The interface described in scenario 2 would present the insights from each of the sectors across a selection of relevant audience characteristics, such as gender and age.

This would help answer the following two questions from the value proposition.

- *Who is currently not/hardly engaging with cultural content in any/most CCIs?* By comparing trends in audiences with existing demographic data; it can help CCIs understand what types of audiences are rarely if not engaging at all with the cultural content on offer;
- *What data needs to be collected in order to identify and cater to underserved audiences?* Not all sectors collect this type of data at regular intervals. This interface would demonstrate the benefits of being able to access these data, therefore encouraging other CCIs to start collecting this information.

Collaboration and Funding Opportunities for Underserved Audiences

Integrating DEI into the strategies, policies and productions of CCIs is a long-term endeavour. For organisations wanting to increase their impact in this area, it can be helpful to have a cross-sectoral overview of policies and funding that specify underserved audiences. In addition, the CCI Data Market described in Scenario 1 (see Figure 4) could also point to collaboration opportunities that target underserved audiences, by also allowing CCIs to specify if they have expertise in DEI. Past, ongoing and future productions could also be labelled with a DEI tag, similar to the Green Film certification. The methodology for creating such a label is, however, outside the scope of this project. An overview combining these elements would highlight high-level trends in CCIs in order to make CCIs more aware of what is happening with DEI, prepare them to include DEI in their strategies, and to find new production and co-production opportunities connected to underserved audiences. In addition to this, when describing their contribution to different productions, CCI companies could add tags to describe the expertise that they brought, including DEI.

5. Reference/Information Architecture

There is a lot of interest in exploring these opportunities, not only from the perspective of individual businesses and single economic sectors, but also to support the Digital Single Market, aimed at creating a unified online marketplace across all member states and across all economic sectors. Thus, CCI Thrive's work is closely aligned with the European Commission's European Strategy for data, which supports Europe's global competitiveness and data sovereignty. Specifically, our efforts are relevant to the establishment of Common European data spaces, a crucial component of the Strategy for data. These data spaces play a central role in achieving goals outlined in the Strategy for data.

The concept of Common European data spaces is integral to the Digital Single Market, which aims to create a unified online marketplace across all member states and economic sectors. By fostering an environment that encourages the sharing and reuse of data, these data spaces ensure that more data are made available for use in the economy and society. Importantly, the approach adopted by Common European data spaces prioritises maintaining control for companies and individuals who generate the data. It promotes data protection and privacy, allowing data owners to retain authority over their information while enabling its utilisation in a trusted and interoperable manner.

CCI Thrive recognises the significance of these opportunities and is committed to exploring them. Our work is driven by the aim of supporting the Digital Single Market and contributing to the creation of a thriving and interconnected European data ecosystem. Data spaces of particular relevance to the work of CCI Thrive are the European Data Space for Cultural Heritage Data Space for Media, the Data Space for Tourism and The European Open Science Cloud. By facilitating meaningful connections across and between data spaces, through the lens of the reference architecture outlined below, we strive to enhance innovation, competitiveness, and contribute to the overall growth of the European digital economy, especially the Cultural and Creative Industries.

5.1 Information Architecture Fundamentals

In order for the CCIs to collaborate with each other to create business cooperatives they need to have a common understanding of the principles of information architecture. This section will provide a short and global overview of those fundamentals.

Information architecture is the practice of organising, structuring, and presenting information in a way that facilitates the effective communication and retrieval of information, knowledge or culture. A good architecture helps the developers of services to communicate how the information can be accessed and understood by its users.

Designing an information architecture involves developing taxonomies, metadata schemas, navigation structures, and content organisation systems. These types of structures enable users to quickly find and understand the information they need. In cross-sectoral and inter-organisational collaborations this is of particular importance. A shared understanding of a common information architecture and collaborative systems enables organisations to structure and organise their information in a way that can generate cumulative value.

A well-designed information architecture helps to ensure that data is shared and used effectively across different sectors, enabling cross-sectoral collaborations and co-productions that can help to shape the digital transformation of the cultural and creative sector.

5.1.1 Components in information architecture

CCI Thrive has conducted an internal analysis and identified that CCIs encompass various types of information structures/systems like:

- Websites
- Databases
- Collection management systems
- Agendas
- Etc.

Each of these contain organised information. Not all of them are machine-readable. Some need processing before they can generate value when used in conjunction with other data, whereas others can easily be mixed with information from other types of information systems.

A lot of these sources of information that were discussed are not owned or controlled by the CCI consortium, they are third party information systems. Which has its legal and technical limitations in terms of access, knowledge about their organisation, and expectancy of long-term availability of the information.

A conceptual approach to understand these systems is to break them up into components, utilising terms like:

- source data,
- registers,
- linked data,
- Extract, Transform, Load/Enrich (ETL),
- indexes, and
- endpoints.

This framework serves as an ontology that describes the flow of information in these systems and ecosystems. The diagram below gives an abstract depiction on how these components are usually linked. Subsequently, there is a description of the diagram and a list containing definitions for each of the used components.

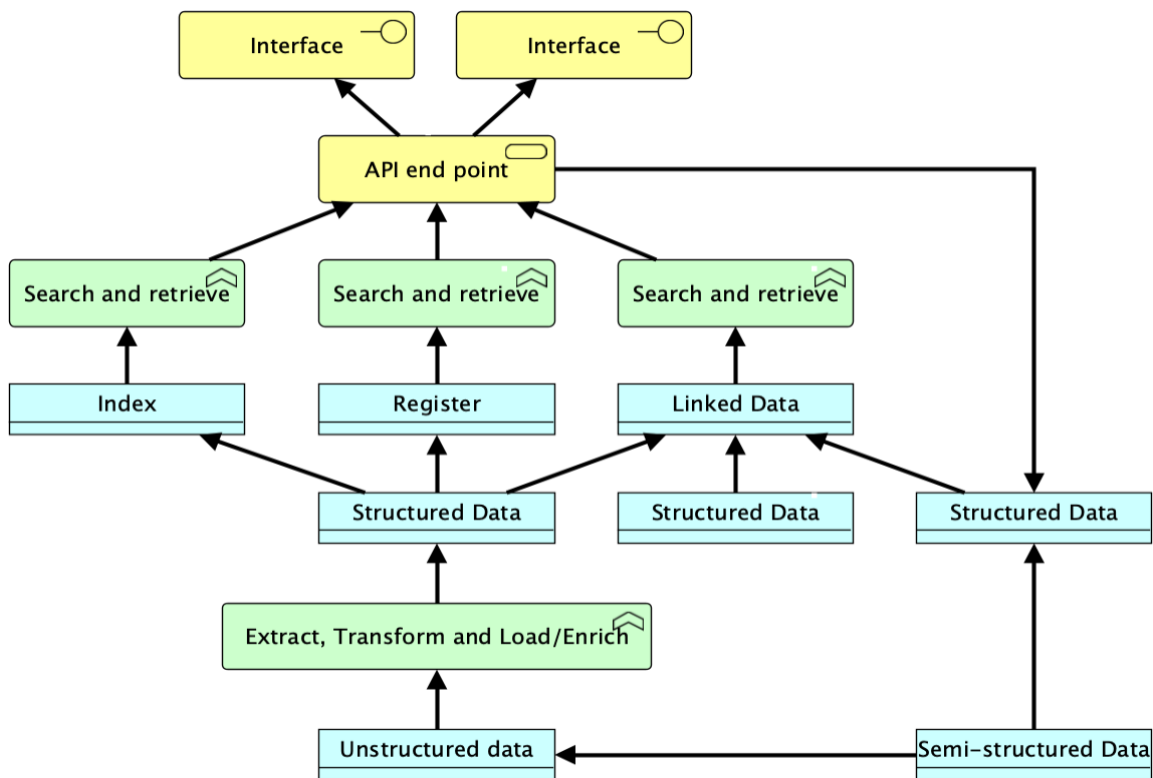


Figure 13: Ontology of different components within the above template information architecture

The various components of information architecture interact with each other to create an information ecosystem. Unstructured data can be linked and integrated with structured data using tools such as ETLs to extract and transform. Information from various sources, including structured data sources, indexes, and registers to create Linked data. User interfaces, including APIs and web services, enables users to interact with the information in the system and make use of the information. This then forms the abstract template for information architecture when collaborating with multiple parties that each contribute information to a centralised party.

These components that are used above follow the definition below:

Source data can be divided into three broad categories:

- structured,
- unstructured, and
- semi-structured data.

Structured data is organised according to a predefined format or schema, making it easier to search, analyse, and use. Examples of structured data include relational databases, spreadsheets, and XML documents.

Unstructured data, on the other hand, is not organised according to a predefined format and can be more difficult to navigate and use. Examples of unstructured data include text files, images, videos, social media feeds, and other types of media.

Semi structured data needs partial parsing using ETLs. For example if the source data contains a generic field call description or contains an image or video that can be mined for additional structured information.

Extract, Transform, Load/Enrich (ETL) – refer to software tools and applications that can be used to process and enhance data. These modules can be used to extract useful information from unstructured data,

transform data into a structured format, and enrich data with additional metadata. Information architecture can help to integrate these modules into a larger system and ensure that they are easily accessible for users.

Linked data – refers to a method of publishing and sharing data on the web. It involves creating relationships between pieces of data, making it easier to navigate and find related information.¹¹

Indexes – are points of access for users to find and interact with information. Indexes can include search engines, databases, and other types of information repositories.

Registers – do not describe the format of the data, rather its owner or maintainer. It are usually authoritative lists of linked data that are maintained and updated by an organisation or government agency. That also means that they are likely external sources and less mutable than other data sources. Registers can be used to ensure that data is accurate, consistent, and up-to-date. Registers can include information such as addresses, codes, classifications, and standards. Information architecture can help to organise registers and ensure that they are easily accessible and usable for users.

Examples of registers are the Netherlands '[termennetwerk](#)', which is both a search engine for terms and a register for other registers. It allows the user to search through different registers like city streets in Amsterdam, the Cultural Historic Thesaurus, etc.

API end-points – are points of access where users/machines can retrieve information, such as APIs, web services, and other types of interfaces. Information architecture can help to ensure that indexes and endpoints are properly connected to user interfaces, enabling users to easily search, retrieve, and interact with information.

Interface – are ways for humans to interact with the available API endpoints. They can be tailored for specific use cases and specific branding to engage the user.

5.2 Requirements for Cross-Sectoral Collaboration Information Architecture

Collaborating parties that share information, knowledge and culture need to address the following basic requirements with either tools like agreements and standards, or physical tools like linked data stores. These basic requirements ensure that the information in a constellation of components described above can properly function and create value for its end-users. Any system that is built for CCIs will need to pay attention to these basic requirements.

Standardisation – Standardisation of information architecture elements, such as metadata and classification systems, helps ensure that information is consistently organised, reducing confusion and improving searchability. Information architecture should be compatible and interoperable across different organisations, allowing for seamless communication and exchange of information between organisations.

For example, use third party data models or ontologies to describe information. Schemas such as [Dublin Core](#) or [Schema.org](#) ensure consistent organisation and description of information across different aspects of an ecosystem. They also make it easier to search and acquire information and to build ETLs that extract and transform from a one-of-standard to the adopted standard.

¹¹ Linked data is based on the Resource Description Framework (RDF), which uses a subject-predicate-object structure to represent data. Linked data can help to improve the discoverability and accessibility of information by linking related data from multiple sources.

Internationalisation – Related to standardisation, the information architecture should be mindful of the language that the metadata is presented in. Even if source data can adhere to the same data standards, when they use different languages or describing rules than the data cannot easily be made interoperable.

For example, multiple data sources can refer to the same country, but adopt the name of that country in their own language. Data from the Netherlands, France and Germany describing activities in the Netherlands can then not be combined if the system tries to match the name of the country (i.e. Nederland, les Pays-Bas, and Die Niederlande). The same requirement is also applicable for any human readable user interface.

Linked data – The information architecture should enable linking and integration of data from various sources, including structured and unstructured data, to create linked data that can provide new insights and opportunities for innovation.

Most known Linked Data framework is the Resource Description Framework (RDF) and its query language SPARQL. Adopting these allows you to link and integrate data from various sources, creating a web of interconnected information. [Wikibase](#), [Triply](#) are examples of systems that can store RDF and use SPARQL to query them. There are plenty of additional database systems available that offer similar functionality.

Scalability and flexibility – The information architecture should be designed to scale and accommodate new information and data sources, and allow for flexibility in adapting to changing needs and technologies.

This is for example done by splitting the information system into its specific functions: data storage, logic for search, logic for access, and (human) interface to ensure that the underlying technology can be separately scaled depending on the need. Additionally, employing APIs allow you to expose the logic of the information system to multiple interfaces.

Usability and accessibility – The information architecture should be user-friendly and accessible to all stakeholders, regardless of their technical expertise or background. Information architecture should facilitate collaboration and sharing of information between different sectors, allowing for co-production and generation of value from collective information.

Legal and Governance – Proper security and privacy measures should be implemented to protect sensitive data and ensure compliance with relevant regulations and laws. Additionally, usage of CCI-data should have proper agreements on and reuse. This should be supported by a clear governance structure and processes for decision-making, maintenance, and updates to the information architecture should be established to ensure effective management of the system.

Specific focus should be placed on the intellectual property rights of the source data. It should be clear which data is used, which licences apply and how data is stored / used on the platforms in the scenarios. Ideally all data used is licensed under a Creative Commons Zero that allows for optimal reuse and remixing.

Store media – In order to create a comprehensive and enriched information architecture, it is necessary to incorporate media storage capabilities. This requirement entails the ability to store, manage, and retrieve various types of multimedia assets, including images, videos, audio files, and documents. The media storage component should provide seamless integration with the structured data within the information architecture, enabling users to access and utilise media resources alongside other relevant information.

To enable media storage, next to the storage of information the CCI can adopt a Digital Asset Management (DAM) system. These systems enable efficient organisation, categorisation, and storage of media. They should provide features such as tagging, metadata management, and search functionality to facilitate easy retrieval and utilisation of media resources. The specifics of these DAMs are out of scope for the remainder of this document. Mostly they can be considered as structured information in itself.

User-generated data – Systems that are developed need to take into account that users need to be able to add data in a set format or suggest a dataset to be adopted by the proposed platforms. This requirement entails providing mechanisms for users to contribute, submit, and incorporate their own data into the system. The information architecture should facilitate the inclusion of user-generated data in a standardised format, ensuring compatibility and consistency across diverse datasets contributed by different users.

Given the specific needs of an CCI system these user-generated systems can differ. Examples are given below for each of the scenarios.

The requirements outlined for cross-sectoral collaboration information architecture serve as the foundation for creating a system that enables sharing, collaboration, and value creation across diverse sectors. Careful implementation and adherence to these basic requirements are needed to fully realise the potential of cross-sectoral collaborations.

5.3 Proposed Information Architectures for CCI

The above-described template for a collaborative multi-input information architecture needs to be additionally configured and described for a proposed architecture. Here questions like who runs which parts of the architecture and what data standardisation is applied in the information architecture need to be answered.

5.3.1 Scenario – Discovering Opportunities for Cross-Sectoral Collaborations and Co-Productions

This scenario, described in section 4.1, targets small and medium organisations by offering a landscape of CCI companies and activities by collecting data from multiple third-party sources and making these searchable and explorable. The service also allows the user to add their own CCI Activities. The service is to be used to explore the market, find other people active in a similar field and explore new opportunities.

A given CCI production/service/event requires at least information on the:

- Type of CCI production
- Companies involved
- Applicable countries
- Relevant dates

Additionally, the following data would further improve decision making in the proposed scenario:

- Audience size category
- Revenue / funding
- funding information

This scenario calls for a centralised index that duplicates information from various sources and presents them in a uniform way. A centralised information hub of information from the identified example sources needs to be periodically harvested, and put in an index.

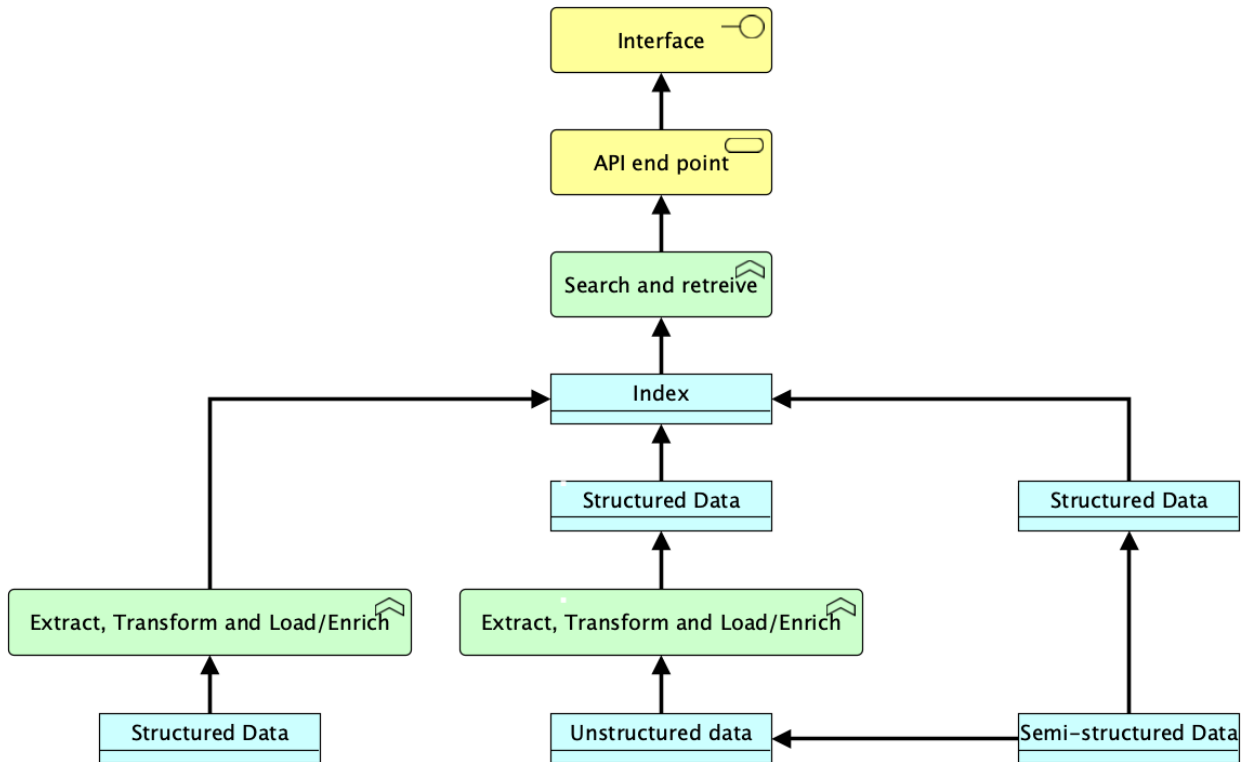


Figure 14: Information architecture for Scenario 1: Discovering Opportunities for Cross-Sectoral Collaborations and Co-Productions

Example platform

Consider [Snowman](#), for quick prototyping based on an API for linked data (SPARQL). This static site generator only requires access to a SPARQL-endpoint with the necessary data and allows the developer to quickly create extensive pages on that data model. When a search index is required for rapid prototyping based on linked data a library like [LunrJS](#) could be considered.

Snowman – is an innovative platform designed for quick prototyping and development based on linked data. It is built with the goal of enabling efficient creation of extensive repetitive web pages based on Linked Data database. With Snowman, developers can leverage Linked Data to generate dynamic and interconnected pages. The platform's static site generator simplifies the process of integrating data from a SPARQL-endpoint, allowing developers to focus on presenting comprehensive information.

LunrJS is a lightweight search library for web applications. With LunrJS, developers can easily implement search capabilities within their websites, it allows developers to quickly build a search interface that can support full-text search, enabling searches across multiple fields and providing accurate and speedy results. As it requires no server-side setup this library can be used for rapid prototyping of a concept.

Application profile for the Index

It is advised to use an existing external data model with a limited number of selected fields in order to make ingestion from third-party sources consistent and shareable. It also allows the index to be turned into Linked Data if needed.

A minimal data model based on schema.org, is chosen as an example here. The example fields below are based on the required fields from section 4.1. Each field also describes their cardinality in, whether they

should be required (1), can occur zero or more, or 1 or more times. Other data models for this scenario could also be developed.

Each [Event](#) should have the following information:

- [Identifier](#) – internal identifier of the event (1)
- [Name](#) – name of the event (1)
- [Audience](#) – type of audience (1 or more)
- [AdditionalType](#) – Type of event, e.g. project, performance, etc. (1 or more)
- [Funder](#) – name of the funder (0 or more)
- [Funding](#) – revenue or applicable funding (0 or more)
- [Location](#) – Location of the event (1 or more), with the following fields
 - [Name](#) – name of the location (1)
 - [Address](#) – The address of the location (0 or 1)
 - [Latitude](#) – The latitude of the location (1)
 - [Longitude](#) – The longitude of the location (1)
- [startDate](#) - [endDate](#) – the period of the event (0 or 1)
- [Organiser](#) of type [Organisation](#) (1 or more), with the following fields
 - [Name](#) – name of the organisation (1)
 - [Description](#) – A description of the organisation (1)
 - [URL](#) – A URL for the organisation's website (0 or 1)
 - [Logo](#) – An image of the organisation's logo (0 or 1)
- [sameAs](#) – Link to the source information if referenceable, e.g. IMDB link (0 or 1)
- [Image](#) – image that represents the event (0 or 1)

Technology – Technology plays a crucial role in designing and implementing the index for the information architecture. There are various options available, ranging from traditional SQL databases to triple or document stores. It is recommended that developers have the flexibility to choose the technology that best suits the specific requirements rather than imposing a predefined technology stack. By allowing developers to select the appropriate technology, they can leverage the strengths and features of different storage systems to optimise the performance and scalability of the index.

Regardless of the chosen technology, the index should include essential functionalities such as search and retrieval capabilities. These functionalities are vital for enabling users to efficiently explore and access the information stored within the index.

It's worth noting that if the index generated by this service is intended to be used in the subsequent scenario, it is advisable to select a compatible datastore that aligns with the technology choice in that scenario. This ensures seamless integration and interoperability between different components of the information architecture, facilitating data sharing and enhancing overall system efficiency and effectiveness.

Data entry guidelines – The above application profile should be expanded with additional data entry guidelines in order to ensure that data from multiple sources are mapped to this model in a uniform way. E.g. the maximum size of an image, that the URL of the organisation should be to their most relevant webpage, which fields are required, etc.

If required, an ETL needs to be set up to extract information from external sources in order to put them into a structured format compatible with the scenario.

Automation – The above information architecture does not take automation ingestion processes into account. External data sources change over time and the platform needs to be kept up-to-date. This will require additional technology considerations when developing the platform.

Application profile for each ETL

It is unlikely that the identified sources can be perfectly mapped on the data model of the index. This means that each of the data sources requires at least some ETL. It is likely that for each source a separate ETL needs to be developed.

Each ETL needs to be able to:

- Extract data from (external) sources (extract);
- Convert data delivered in external data model into an internal model (Transform);
- Act upon policies for Create, Update and Delete (Load).

Examples of platform that fulfil most of these main functionalities are:

[Apache Nifi](#) – Apache Nifi is an open-source data integration platform that allows you to automate the flow of data between systems. It provides a visual interface for designing ETL workflows, allowing you to extract data from various sources, transform it using a wide range of processors, and load it into the desired target system.

[Apache Kafka](#) – Apache Kafka is a distributed streaming platform that can also be used for ETL purposes. It allows you to stream and process large volumes of data in real-time. Kafka Connect, a component of Apache Kafka, provides connectors for extracting data from various sources and loading it into different sinks or target systems.

[Talend](#) – Talend is an open-source ETL tool that offers a set of data integration and transformation features. It provides a graphical interface for designing ETL processes and supports a wide range of data connectors and transformations. Talend allows you to extract data from multiple sources, apply transformations, and load it into the target system.

However it is important to note that for starting to work on these ETLs, it could also be beneficial to develop your own infrastructure if only limited ETL is needed. E.g. because the data only needs to be loaded once.

Application profile for the API endpoint and interface

The API endpoint and interface can – for a technology demonstrator – be seen as one technological component in the architecture.

The API endpoint needs to:

- Encode predefined search requests like
 - “Give me all events with a certain category X”
 - “Give me all events with company Y”
 - “Give me all events given a certain geographical box LAT and LON”

- Etc.
- Consider authorisation

The interface needs to:

- Provide an interface for the user to navigate the information;
- Take information from the API and place it within the structures of the visual design of the website.

Form based input – Users can also add information to the web platform. Adding user accounts can increase complexity of the platform, for a proof of concept it might be good to consider this an ingestion with human intervention. E.g. a form that sends an email with a request to add information to the index. Form-based input should be considered as Structured Data.

Ownership and organisation

As visualised in Figure 15, the components of this information architecture do not necessarily need to be solely owned by the maintainer of the platform. The data sources will most likely come from third party sources and thus fall outside of the ownership and organisation of the platform maintainers.

These maintainers need to be responsible for the main index, search and retrieve, API endpoint, and Interface. As well as any ETLs that will be needed to transform external structured and unstructured data into the format of its index.

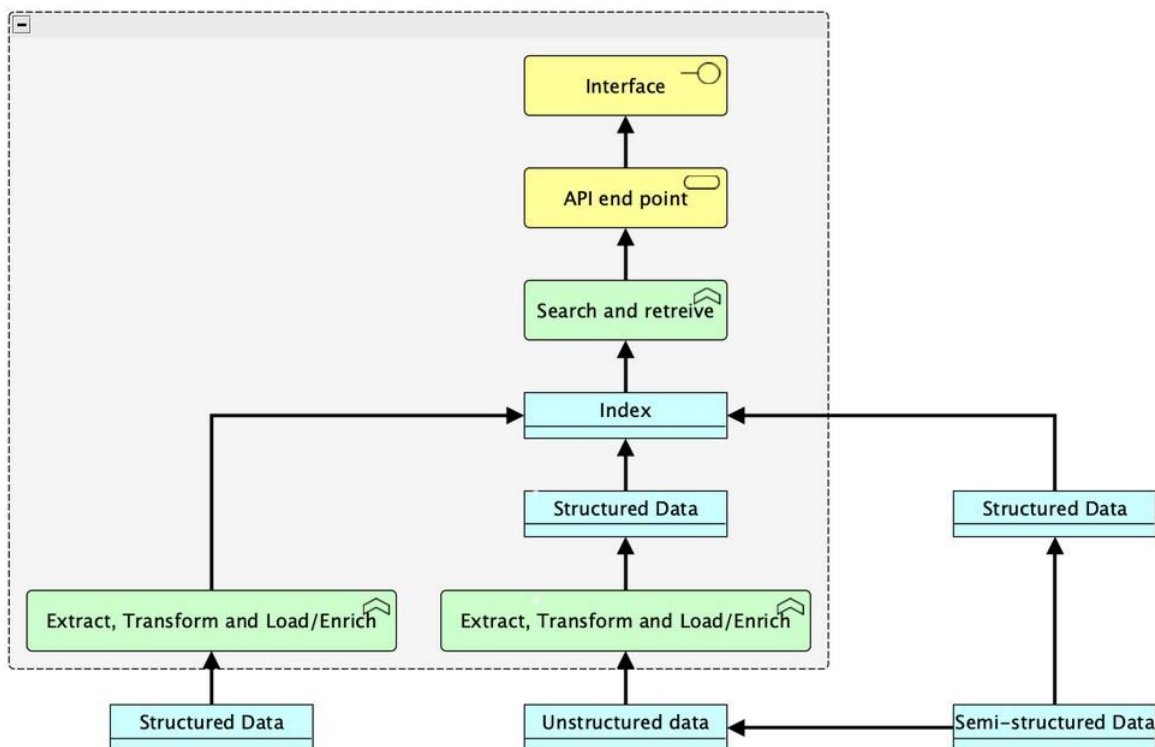


Figure 15: Ownership of data and components for Scenario 1: Discovering Opportunities for Cross-Sectoral Collaborations and Co-Productions

5.3.2 Scenario – Matching Creative Products with Audiences & Identifying and Understanding Underserved Audiences

Both the scenarios “Matching Creative Products with Audiences” & “Identifying and Understanding Gaps in Audiences” are exploration and visualisation services on multiple (curated) datasets. Where these datasets are either duplicated on the platform or linked via linked data. In these scenarios it is up to the user to explore and combine data together into interesting and new perspectives on the collected sources.

The service looks similar to other statistical services like Eurostat, destatis.de, and CBS.nl.

The CCI services for these scenarios combine (arbitrary) data from multiple sources, each using their own data models.

A MVP for these scenario need to:

- Allow for the storage or linking of multiple datasets;
- Allow for explorations and visualisations of data;
- Be agnostic to external data models;
- Be publicly available;
- Need to be able to do federated searches.

A baseline model for the information flow in this scenario would be.

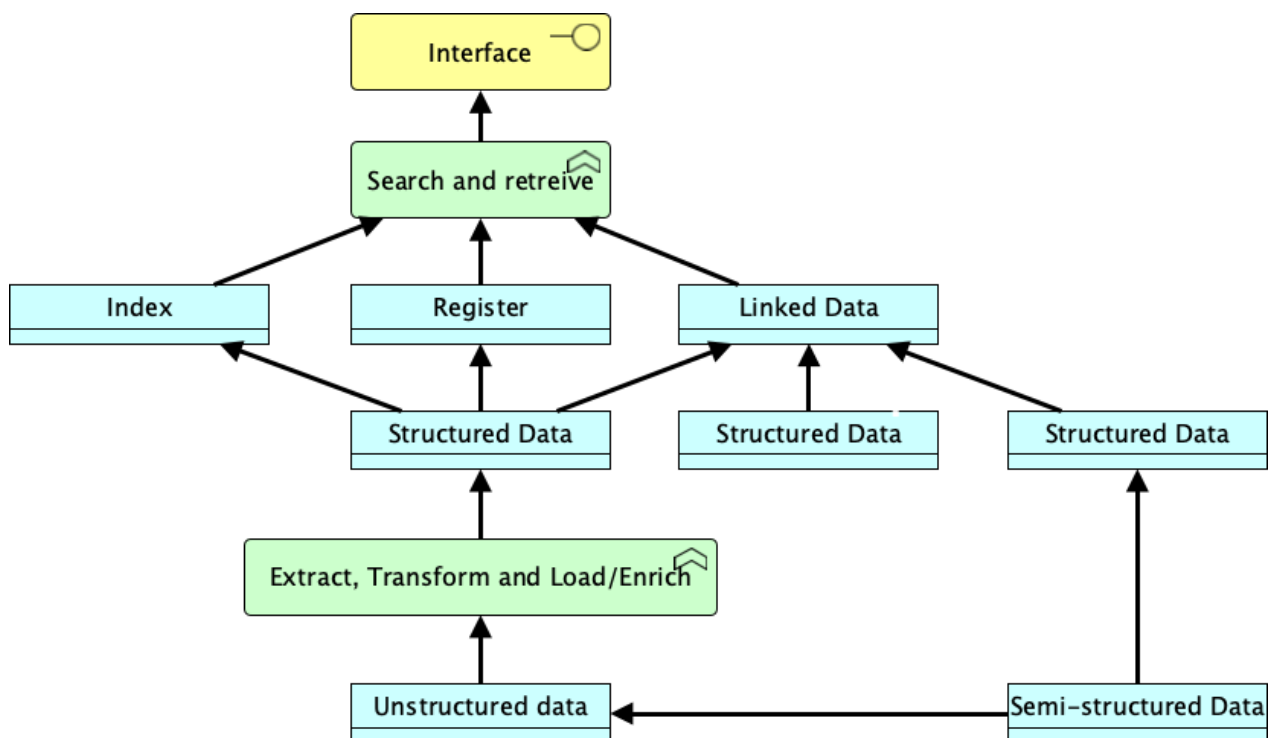


Figure 16: Information Architecture for Scenario 2: Matching Creative Products with Audiences & Scenario 3: Identifying and Understanding Gaps in Audiences

Example platforms

Two open-source platforms that fulfil the basic concepts behind these scenarios are [Apache Superset](#) and [CKAN](#).

- **Apache Superset** is an open-source data exploration and visualisation platform that allows users to create interactive dashboards, reports, and charts. It offers a flexible interface that allows users to connect to multiple data technologies, including popular databases. It has an extensive library of visualisations and powerful querying capabilities. It supports collaborative data exploration, sharing and embedding of visualisations.
- **CKAN** is an open-source data management platform that allows for the storage, sharing, and discovery of datasets. CKAN provides means to publish, manage, and access data resources. It offers features such as data organisation, metadata management, and version control, allowing users to maintain a comprehensive and up-to-date catalogue of datasets. CKAN also supports data visualisation through integration with various visualisation tools, making it easier for users to understand and analyse the data they work with. CKAN does require [additional extensions](#) to be able to work with various types of data. Depending on the particular use case or dataset these need to be added if CKAN is chosen as a main platform.

Application profile for each ETL

It is unlikely that the identified sources can be perfectly mapped on the data model of the index. In contrast to the previous scenario, this does not mean that each of the data sources needs to be extracted with ETL. However, they will likely need to be transformed or be placed as a separate interface (e.g. SPARQL) in order to be able to integrate with other data sources.

Application profile for Search and Retrieve

A comprehensive data management and publishing platform that enables organisations to:

- store,
- share,
- discover datasets effectively.

It should also support metadata management, including version control.

Application profile for interface

A user-friendly interface for cataloguing and organising data resources. The platform supports extensibility through plugins, allowing customisation to meet specific data management requirements. Including an interactive data exploration and visualisation platform that allows you to create dashboards, reports, and charts.

Ownership and organisation

In this scenario the maintainer of the platform is responsible for hosting, linking and providing an interface for exploring a multitude of datasets.

Note that the structured data of the previous scenario can be included into this scenario as well. This would make it possible to combine technologies and only build an additional interface for each of these two infrastructures.

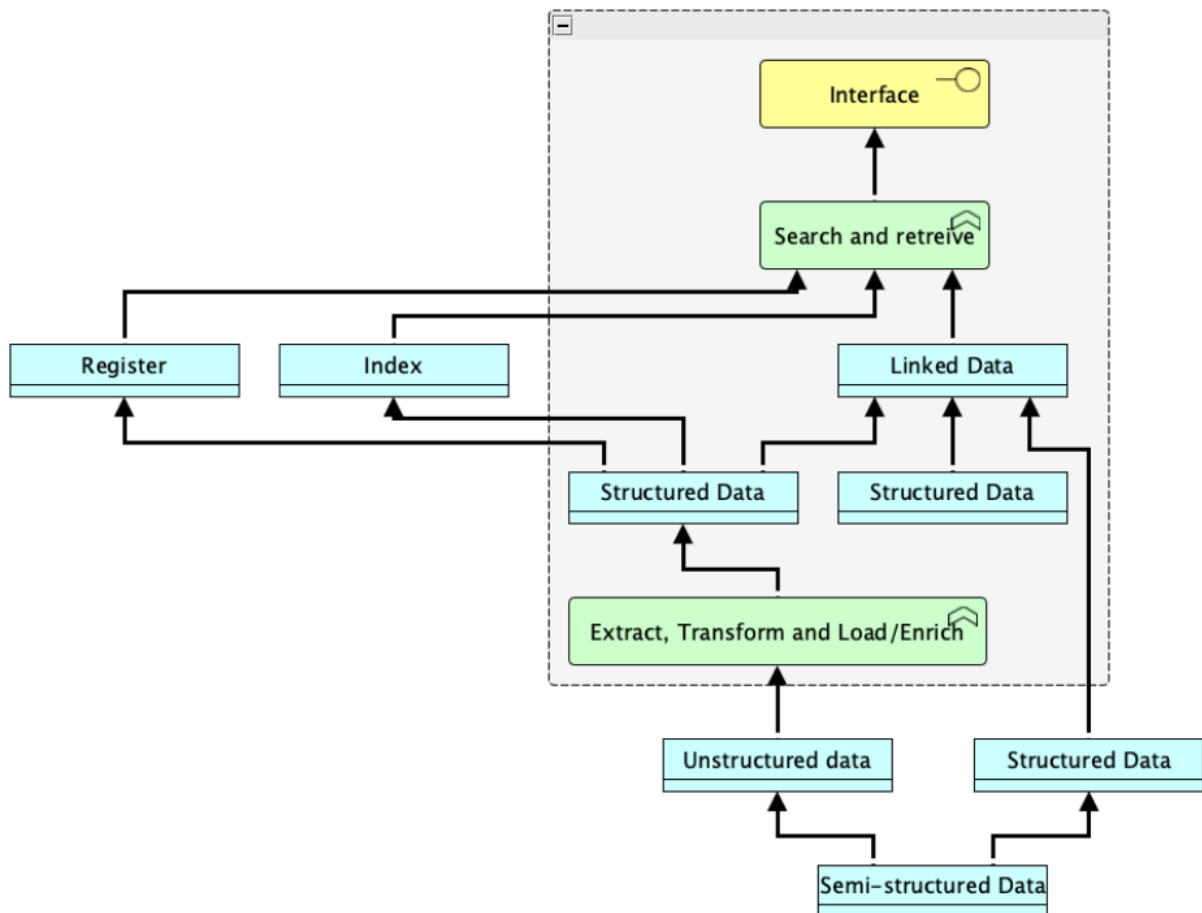


Figure 17: Integrated information architecture for all scenarios

5.4 Implementation and Integration

Careful attention needs to be given to the following steps, before starting to develop information services for the scenarios described above:

- **Identify Data Sources** – Determine the relevant definite third-party data sources and content that will be crucial for the services. It is important here to carefully document the licence, data standards, and interfaces to the data. E.g. whether the data can be copied, whether you do a federated search, and if these actions are allowed;
- **Data Acquisition and Integration** – Establish mechanisms to acquire and integrate the data from the identified sources into the information systems. This could involve developing data pipelines, setting up data connectors, or utilising ETL (Extract, Transform, Load) processes to gather and consolidate the required data. Take into account that you might want to update data from the same source multiple times over the lifetime of the project;
- **Data Modelling and Storage** – Formalise the proposed data model for discovering opportunities for cross-sectoral collaborations and co-productions that aligns with the specific requirements of that service;
- For Matching Creative Products with Audiences & Identifying and Understanding Gaps in Audiences take care of documenting any data model that you include in your service;

- **Data Processing and Enrichment** – For the first scenario, apply data processing techniques to clean, normalise, and enrich the acquired data. This may involve data cleansing, deduplication, standardisation, and enrichment with additional metadata or contextual information. Use data transformation methods to convert the data into a format suitable for analysis and exploration. Take into account any data that needs to be frequently updated;
- **Application Development** – Choose the primary platform for building both scenarios. For “*Matching Creative Products with Audiences & Identifying and Understanding Gaps in Audiences*” this could for example be CKAN or Apache Superset. For the *Discovering Opportunities for Cross-Sectoral Collaborations and Co-Productions* this could be the same platform, but used as a Linked Data interface itself.

To be able to realise both scenarios, separate interfaces can be built on the search and retrieve functionality that one of both of these platforms have to offer. This leads to an integrated ontological information architecture that is visualised in the diagram below (figure 18).

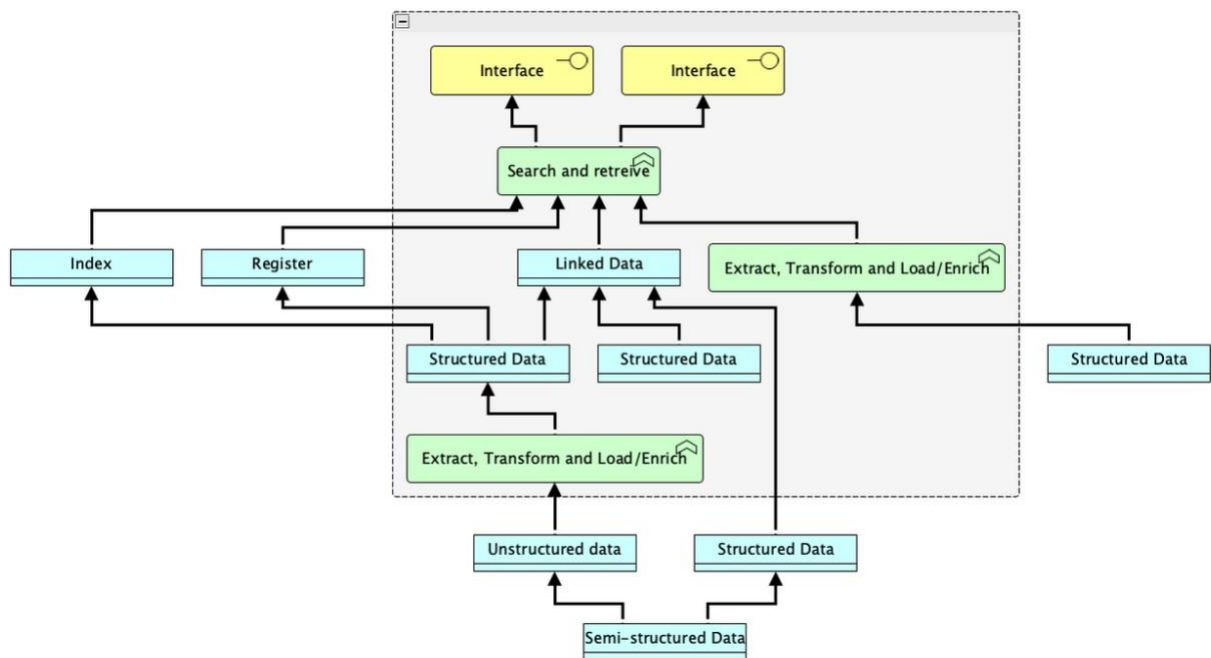


Figure 18: integrated ontological information architecture

6. Conclusions and next steps

The CCI Thrive project employed a design-driven methodology to establish cross-sectoral data-sharing scenarios and develop a reference architecture to support them. By leveraging the expertise of consortium partners this work identified shared values, priorities, and business needs that could be translated into tangible cross-sectoral synergies. Also, areas were identified where data sharing could benefit all actors in the CCI sector. Personas and user stories were created to illustrate the specific advantages of cross-sectoral data sharing for CCI professionals, leading to three scenarios that showcases how cross-sectoral insights can enable collaboration, deliver better products, and identify market opportunities: (1) Discovering opportunities for collaboration by finding complementary activities in adjacent CCIs (2) Delivering better products to target audiences through a holistic understanding of cultural and creative consumption (3) Identifying market opportunities and audience gaps in the offering of CCI products and services. These scenarios offer practical examples of how collaboration and data exchange can enhance the competitiveness of European CCIs. We illustrated them through bespoke user journeys, providing a tangible visualisation of their potential impact.


The project has also conducted research to identify an information architecture and digital tools that can be broadly applicable across the sector. This work, along with Deliverable [D2.1 Recommendations for shared cross-sectoral metadata](#), forms the foundation for the next phase of the project. In this phase, demonstrators will be built to validate the envisioned scenarios as proof-of-concepts. Project partners will develop fully functional demonstrators, while also inviting third parties to participate in hackathons and create simple demonstrators using provided back-end infrastructure and example data. In these activities, consortium partners will use their own databases and publicly available data as well as engage other CCI companies who might want to bring their data to see how it could be valorised to support the sector.

The scenarios presented in this report serve as a powerful demonstration of the transformative potential of the Digital Single Market. This is relevant for two reasons:

- First, they highlight how CCIs can leverage data and collaborative efforts to unlock new opportunities, drive innovation, and strengthen their global market position. This work serves as inspiration for ongoing efforts in the sector, such as the EIT Culture and Creativity and the data spaces being developed as part of the EU's Strategy for Data, with a specific focus on the Cultural Heritage, Media, and Tourism data spaces.
- Second, by showcasing practical examples, recommendations and demonstrator development, CCI Thrive demonstrates how its work can complement and support the implementation of EU regulations in data-sharing. In the context of the Data Governance Act, it demonstrates how collaboration and data exchange can be facilitated within the cultural sector while complying with data protection and governance principles. And aligned with the Digital Services Act, CCI Thrive will contribute to the development of guidelines and best practices for digital service providers operating within the CCI.

Through this work, CCI Thrive showcases how data exchange can enhance competitiveness, drive innovation and inspire initiatives in the field, while aligning with and informing relevant EU regulations.

Annex 1: User Stories

	<p>Name: Michelle Role & sector: museum curator Goals / needs / outcomes: attract more people to the museum’s platforms (physical or digital); understand audiences that their museum is underserving (gap analysis); diversify their current offering to respond to policies on increasing youth cultural engagement (what new activities they need to add to attract such audiences). Experience with data analysis: looks at annual reports from the sector, mainly looking at pre-analysed data and trend analysis, their museum outsources audience analysis to a consulting company.</p>
---	--

User story:

A report from the consulting company revealed that their museum is currently serving families and young children well but not reaching other age groups. Michelle wants to find out how to attract more people aged 18-30 (young adults) to the museum.

She visits the cross-CCI data online platform where she finds data on the engagement of 18-30 year-olds in cultural activities. She is specifically interested to find data on:

- Which sectors in CCIs are provide offering for (and are successful in attracting) this age group;
- Genres and topics that attract these audiences across CCIs and what genres/topics are failing to attract them;
- Digital platforms that they use (Tiktok, Twitch/streaming, forums);
- Census data on the size/diversity of this age group in each country.

When looking at data on the platform, Michelle can filter it by multiple filters on age, gender, country, CCI sector, time period. The platform offers them a number of visualisations that they can choose from (e.g. a geo map, a graph, list, etc.). Using the filters, they can select which types of data they would like to see. For instance, Michelle chooses to visualise and compare [1], [2] and [3]. Inspecting this data overlay, they can look for correlations. Michelle notices that in the last two years this age group in other sectors have been consuming a lot of interactive, choose-your-own-narrative type of content and that platforms like Discord are a popular way to communicate with these audiences. It is evident from the data that the gaming sector has been particularly good at this. There is an explanation on the interface about the filters, how the data was collected, what are potential gaps - which makes it easier to make informed decisions. Based on this, Michelle pitches an idea to the curatorial team about a potential collaboration with a game design company on a choose-your-own narrative activity based on the museum's collections. They also decide to do research on how museums can use communication channels such as Discord.

	<p>Name: Kirstin Role & sector: Coordinator Digital Projects / Opera House Goals / needs / outcomes: Create or assign digital projects to attract new audiences, retain existing audiences during non-play seasons, to incite new productions by fostering new creative expressions and experiments Experience with data analysis: a large database on visitors</p>
---	---

User story:

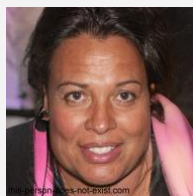
Kirstin read a report from a consulting company that revealed that during the Covid-19 pandemic the interest in classic music of the age group of XXX, which is underrepresented in the usual audience group, has significantly increased. Of course, this doesn’t automatically mean that they would come to a live concert or even to an opera or ballet.

Kirsten therefore wants to understand this age group XX better. She would use a cross-CCI database/platform to identify the channels to reach them through, the obstacles for coming to a live performance, and their general preferences and dislikes.

She would be looking to answer questions such as:

- Does age group X prefer interaction, immersion, group experiences, experimental or 'perfect'?
- What does age group X generally think about opera? Do they see it as outdated or elitist?
- What time of day does age group XXX prefer to go to cultural performances?
- For age group X, what is their price acceptance? Would they be able and willing to pay the cost of a ticket?

With this information, Kirstin and her Opera House colleagues could better decide which projects to invest in digital methods and means of engagement.



Name: Malou | **Role & sector:** Director of a theatre ensemble | **Goals / needs / outcomes:** Determining which topics are currently funded by which programme (national, regional and European-wide), which topics are successful, hence possibly favoured and popular with the programmes | **Experience with data analysis:** not much, own figures on amount of tickets sold per show.

User story:

Playwrights send their works to Malou and her colleagues. Malou and her team then brainstorm on existing scripts for their next theatre performance. The theatre needs to secure funding for the adaptation of plays, rehearsals, and touring them around. Therefore Malou wants to know what topics are of interest and highlighted in funding programmes, both internationally as well as regionally.

With this information Malou can further develop the theatre's programming. Once potential topics have been identified and narrowed down, opportunities for cross-sectoral collaboration, to access for example European funding or cooperate on shared marketing between a film and theatre play, will be of interest. Malou would also be interested in identifying regional funding to support touring of a play.

E.g. in the year of van Gogh's 65th death anniversary, Julian Schnabel made his movie on Gogh (Eternity's Gate), there were possibly some exhibitions in Provence and Amsterdam, so showing a drama play on Van Gogh might get some funding in those regions.



Name: Felix | **Role & sector:** Independent Video Game Developer, already released 1 game with a publisher. Is working on his second project | **Goals / needs / outcomes:** Has a new concept for a PC/Console Game. Wants to know how big the market for his concept is and adapt accordingly to the scope of the game. Also wants to find out if certain markets are more or less relevant for this kind of game. **Experience with data analysis:** Knows his ways around free tools for game developers like <https://vginsights.com/> or <https://indiebi.com/about-us>. He also listens to a lot of podcasts and recordings of gaming conferences.

User story:

Felix is developing a new game and intends to build on his current small community of gamers. To do that, he needs to find out what other media his target users consume, so he can connect with them early and make sure that they are following my developer blog/channels. What kind of music do they like? Are they going to festivals? Do they visit museums or literary events? Or what kind of video/film consumption do they have? What kind of societal issues are they engaged with? To know these things will help him find his future player base.

He will build a persona to go deep into the motivations of his target group, so he is interested in qualitative information, motivations and aspirations. He will also need to provide potential publishers with estimated sales targets, so it's important for him to have quantitative information on how many people my competitors reach.

Additionally, data like this will give him insights that feed backs into his game development: What kind of art style does his community prefer? How complex can his storytelling be? What societal trends, narratives, and issues does he need to be mindful of when developing his game? He wants his user base to identify with the game and for them to find real value for their lives in it. So he needs to have qualitative information on these issues.

Finally, gaining access to data from other creative industries will help him understand the resources that he has at his disposal for game development. Are there video/image/audio/animation assets that he could use or be inspired by from other industries? Are there local events that are not tailored to video games but where he could still meet interesting people who could help him write scripts, compose music, build assets, co-produce, market or publish the game? Or do his skills apply to other creative sectors, so he could contribute to their projects?



Name: Stéphane S. | **Role & sector:** producer of TV series | **Goals / needs / outcomes:** select within my author's pitches the projects I'll invest in more to sell them with minimum effort on the growing OTT market | **Experience with data analysis:** none, I read the trade press like anyone else but have no clear idea of the purpose and importance of data for my clients (the commissioners) as they almost don't share with me. And I have no insights on other media fields except sometimes books or video game adaptations.

User Story:

Stephane has heard that the 'platforms' have a huge need for local content, a peak time for producers. But most of his projects are rejected as they 'aren't developed enough', 'don't fit the need', have no 'potential', or a new season is not ordered because it did not perform well enough in season 1. He does not know what this 'need' or needed performance is. From 9 scouted projects by his team, he needs to make an informed decision in which project to bet and invest more in the development phase to deliver a high quality project matching the demand of his commissioners.

To do this he would need to access the following information :

- access to existing stats of TV series and platforms with their target and performances explained to teach him about his market and clients (sectoral data);
- find proper comparables on each of his projects to identify competition, his potential audience target and most suitable commissioners to pitch to (sectoral data);
- evaluate a user interest for his subject today and be able to project in 2 years from now to check if the interest is the same (cross-sectoral data);
- identify other similar projects on other media segments to build a more powerful franchise, maybe buy some

IPs together, share marketing tools and budgets in complementarity (cross-sectoral data).

He'd like very easy-to use SaaS tools requiring no marketing & data skills that he doesn't have himself or in his team. He can pay for it, but it should be cheap as he has today no budget on „strategic marketing“. He needs tutorials with simple explanations. With this data and data analysis he will select 3 projects on which I'll put three times more development money than usual (securing IPs, writing room, securing a talent).



Name: Katheryn | **Role & sector:** Film Distributor | **Goals :** promote her catalogue to relevant customers/ **needs:** find new buyers, extend catalogue, increase sales opportunities / **outcomes:** Increase turnover, profit margin and better use of existing catalogue | **Experience with data analysis:** Sales data analysis but little experience of market analysis

User story:

Katheryn's main goal is to find out how she can increase the value of her catalogue sales.

She would consult a cross-sectoral database to find data on:

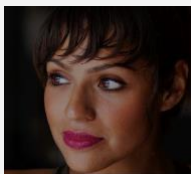
- Who are potential clients for my current catalogue?
- Where are they based and how can I track them?
- How can I find information about their acquisition needs?
- Where can I find other assets to add to my catalogue?
- Where can I measure the impact of my promotion /marketing?
- How do I know what is successful in other creative assets catalogues?

She would prefer to access these forms of data

- On a global platform with information on acquisition managers and their editorial line
- On a aggregated platform of social networks and online marketing based on tags/Key about existing content - ready to be sold
- Content Agent like Avant Premiere
- With the help of an AI like Chat GPT
- Database based on topics and not on media

With this data she could support her decision making in:

- Efficient identification of potential buyers
- Support to analyse the marketing mix expenses to promote catalogue
- Comparison with other content assets and competitive catalogues in terms of price



Name: Charlie | **Role & sector:** Buyer Audiovisual for channel | **Goals :** curate and acquire relevant content at right price **needs:** find easy to use access to secure content / **outcomes:** Develop high quality program at the right price | **Experience with data analysis:** Very little as business is based on networks

User story:

Charlie wants to know where she can find secured content that fits her editorial line. She needs to access a

database of secured catalogues of assets with editorial information. She would also like to access data that helps her answer the following questions:

- How can I see the content and have a business discussion with the right holder?
- How can I better understand trends in other sectors to move forward with my acquisitions?
- How can I measure interest from target groups?

She would expect to interact with this data:

- Online platform offering content of all kinds
- Access to cross recommendation tools between sectors
- Data on audience penetration and sales of comparable projects or topics
- Using a Content Agent like Avant Premiere
- With the help of an AI technology like Chat GPT

With this data Charlie would be able to make decisions about

- Acquisition strategy + channel positioning
- Identify general content topics to be researched
- Focus on relevant distributors with a more efficiency
- Identify relevant content before competition



Name: Anton | **Role & sector:** Movie Director | **Goals / needs / outcomes:** Making his ideas/stories heard and understood by producers and financiers- Need to create clear stories based on both artistic vision and market interest- Need market data | **Experience with data analysis:** very little experience with data , work mostly done based on life experience, peer experience, Film festival visit.

User story:

Anton needs to make sure the project he is pitching is unique and presented to the relevant market players.

Data needs:

- Market contacts with information on their editorial lines to identify possible producers and /or financiers;
- Market data about projects that are pitched at festivals and co-production markets . Access to data on cast, locations, target audience and script developers;
- Market data about budget of existing comparable projects;
- Data on general trends about his story-line - Data analysis of similar storyline based on different criteria - Spotify numbers for music for example- Books sales on similar topics;
- Data on script writers to look for partners to develop the story.

Access to data:

- First, analysis of his project through AI to understand the tags and keywords attached to his project;
- Then use of these tags and keywords in different databases to access inspiring feedback and comments;
- Once the project finds its space on the market - access to databases based on storyline to look for producers and other financial partners.

By having access to this cross-sectoral data, Anton will decide if a project is worth developing further based on its uniqueness, whether it's currently on trend, and whether he has identified possible partners.



Name: Eva | **Role & sector:** Curator & Producer Media art | **Goals / needs / outcomes:** platform that help to find artists / creators and content to create the next media art exhibition | **Experience with data analysis:** relies heavily on personal knowledge and networks, research through catalogues and online sources

User story:

Eva is a producer in the media art sector. In the early stages of the content research to produce the exhibition, she would like to have a place with data about current artists and other creators in the field of digital media art and related fields as well as their work. This can help her in the early phases of the project because it's really complicated to find resources with accurate information to start developing the exhibition project (normally she has her own database for this type of research).

To answer the question, 'what are the currently most interesting creators and works in the wider field of (digital) media art?'; Eva would like to access data:

- Data regarding the content that is important for the exhibition topic; data about the creators, data about the works (ideally including metadata, image material, documentation, previous showings, etc.)

She would want to access the data using:

- features that can help to filter the data with keywords;
- Artist agenda (exhibitions current or past);
- artwork year of production; artwork format (audio, video, print, photo, etc.);
- thematic and methodological clusters (art queer, AI, AR..)

With this data she can make a selection of artworks for a given exhibition. The exhibition can be much richer in term of narratives and artworks if the research phase is accurate in content and information to produce the exhibition, and negotiations with creators can be streamlined if the basic information is already available in a reliable form.

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



www.betaseries.com/

Italy



www.centrica.it

The Netherlands



www.beeldengeluid.nl

Disclaimer: Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EACEA. Neither the European Union nor the granting authority can be held responsible for them.