POSITION PAPER
Dataspace for Cultural and Creative Industries

v.2.0 – 10.10.2022

EBU
OPERATING EUROVISION AND EURORADIO

gaia-x

1 CCI DATA SPACE WORKING GROUP

On the request of the Technical Committee of the EBU, EBU T&I initiated a European working group on the creation of a Media/CCI data space. It is now chaired by Louis-Cyrille Trébuchet, CIO of France Télévisions and Véronique Demilly, project manager at France Télévisions and coordinated by Lucille Verbaere (EBU T&I), senior project manager at EBU T&I. Lucille also represents the Media/CCI vertical at Gaia-X Data Space Business Committee. Some of the group members that contributed to this document are listed below:

Company name	Description	Country
EBU	European Broadcast Union	International
France TV	Public broadcaster	France
VRT	Public broadcaster	Belgium
SSRSRG	Public broadcaster	Switzerland
Sitra	Innovation lab	Finland
DPA	Deutsch Press Agency	Deutschland
APA	Austria Press Agency	Austria
CBA	Cultural Broadcasting Archives	Austria
newsalliance.org	Alliance of news agencies	International
ATC	Technology Provider	Greece
Gaia-X French hub- CCI group members		France

2   DOCUMENT HISTORY

Version	date	Change Description	Change proposed by	Reviewed and approved by /When
V1.0	April 2022	listing use cases and projects in France cultural and Creative Sector. First version is available there: [0]	CCI WG French Gaia-X hub	CCI WG French Gaia-X hub
V1.2	June 2022	Additional use cases described	CCI WG French Gaia-X hub	CCI WG French Gaia-X hub
V2.0	Sept 2022	Defining a Media Dataspace and listing Use Cases and projects from Europe	EBU	CCI Dataspace working group And EBU Technical Committee

In May 2022, several workshops were organised with a broader community of EBU members, CCI working group of the French Gaia-X hub, and media companies from across Europe, to discuss Use Cases, data exchanges challenges, technical building blocks etc... Participants are listed in the minutes of the workshops [7]. The current version is based on inputs gathered during these workshops; use cases and projects from the CCI data space working group coordinated by EBU.

The first version of this positioning paper was published in April 2022 by the CCI working group of the French Gaia-X Hub, led by Louis-Cyrille Trébuchet, CIO of France Télévisions and Véronique Demilly, project manager at France Télévisions. It listed Use Cases and projects related to data sharing in the French Cultural and Creative Industries sector. As of 17 May 2022, the participants in the CCI working group of the French Gaia-X hub are listed in the document [0].

CULTURAL AND CREATIVE INDUSTRIES SECTOR

DEFINITION

The European Commission, in its proposal for a regulation establishing the Creative Europe programme for 2021 to 2027, considers that the Cultural and Creative Industry (CCI) sector includes all sectors whose activities are based on cultural values or on artistic and other creative expressions, whether individual or collective.

These activities may include the development, creation, production, dissemination and preservation of goods and services embodying cultural, artistic, or other creative expression, as well as related tasks, such as education or management. They will have the potential to create innovation and employment. Thus, these sectors include architecture, archives, libraries and museums, arts and crafts, audio-visual (including film, television, video games, multimedia) and advertising, tangible and intangible cultural heritage, design (including fashion), festivals, music, theatre, performing arts, literature, publishing, radio, and visual arts. **We consider that the whole journalistic sector is part of the CCIs, such as content creation, audio-visual, publishing, and radio.**

MARKET SIZE AND TRENDS

Pre-Covid, the European Cultural Sector was worth about 643 billion Euros[1]. **Figure 1** below provides the breakdown per subsector, and growth rate for the 2013-2019 period. The audio-visual subsector was weighting 119 billion Euros in 2020, an almost unchanged value compared to 2016: growth was
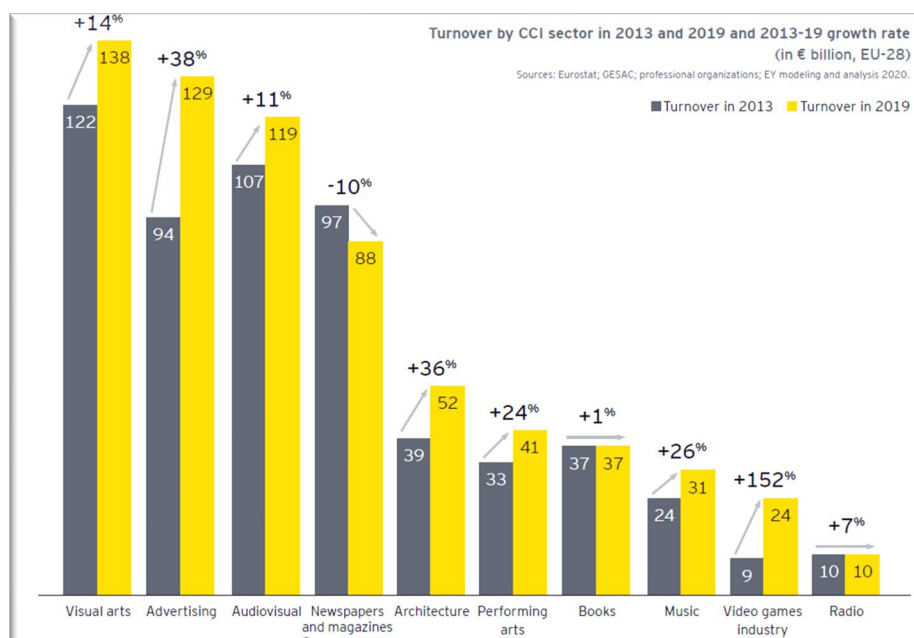
Figure 1:Turnover by CCI sector in 2013 and 2019 and 2013-2019 growth rate

steady but limited during 2016 and 2019 and Covid-19 resulted in a decrease of revenues of 5.6% from 2019 to 2020. Behind this market overall stagnation, a recomposition of the market has been observed

[1] Data covering EU27 and the UK. Detailed breakdown only available for 2019. Turnover is defined as the revenue at the end of the value chain (approach based on final consumer markets)
Source: EBU based on Ernst & Young

with a decrease of radio and TV advertising and cinema revenues. Subscription Video On-demand services (SVOD) were already quite important in Europe in 2016, especially in the Nordic countries. Netflix was available, as well as HBO and Amazon in some European countries. Today, on-demand ranks 4th in terms of revenues **(Figure 2).** [2]

Total: 120
EUR billion

Total: 119
EUR billion

PAY-TV

RADIO AND TV ADVERTISING

PUBLIC FUNDING

+235%
2016-2020

ON-DEMAND

CINEMA
PHYSICAL VIDEO

39   39
37   34

28   28

14

8   3
4   2
3

2016   2017   2018   2019   2020
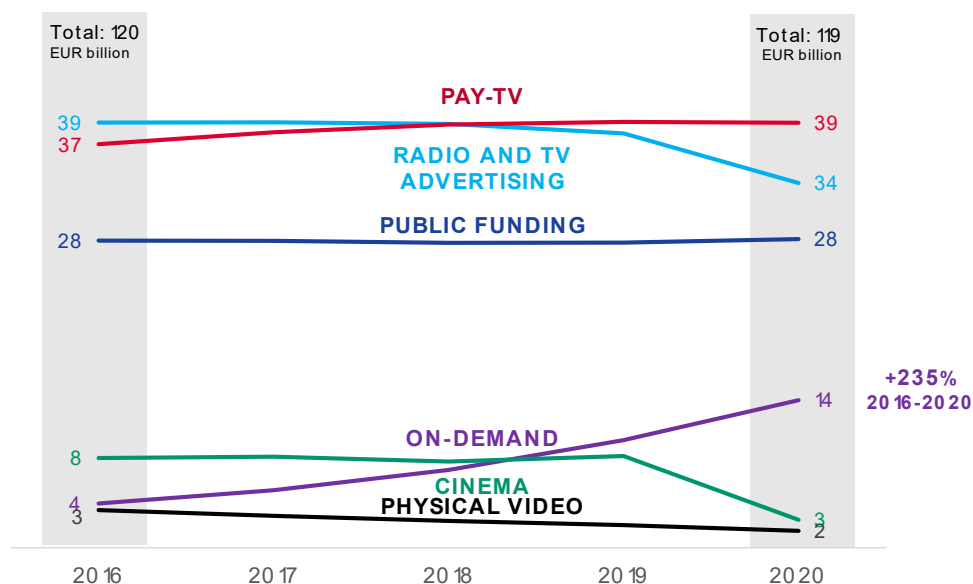
Figure 2: revenues of the European Audiovisual market by revenue streams (Eur Billion, 2016-2020)[2]

CCI Actors

CCI sector is made up of structures of very different sizes and operating methods. 1.2 million firms and over 8 million people contribute in European CCI sector. Alongside large private companies or large public institutions, there are many very small companies or structures based on associative models, with very few permanent staff, self-employed people, and casual entertainment workers. In some parts of the CCI sector, most of the structures are micro-enterprises.

European audio-visual sector faces strong competition from global and diversified actors, such as the FAANG (Facebook, Amazon, Apple, Netflix, Google), as shown in F**igure 3**. They invest in content creation massively, but public broadcasters still cover over 40% of all financing of Original European content **(Figure 4).**

---

[2] Based on 37 European market. Sources: EBU based on European Audio-visual Observatory
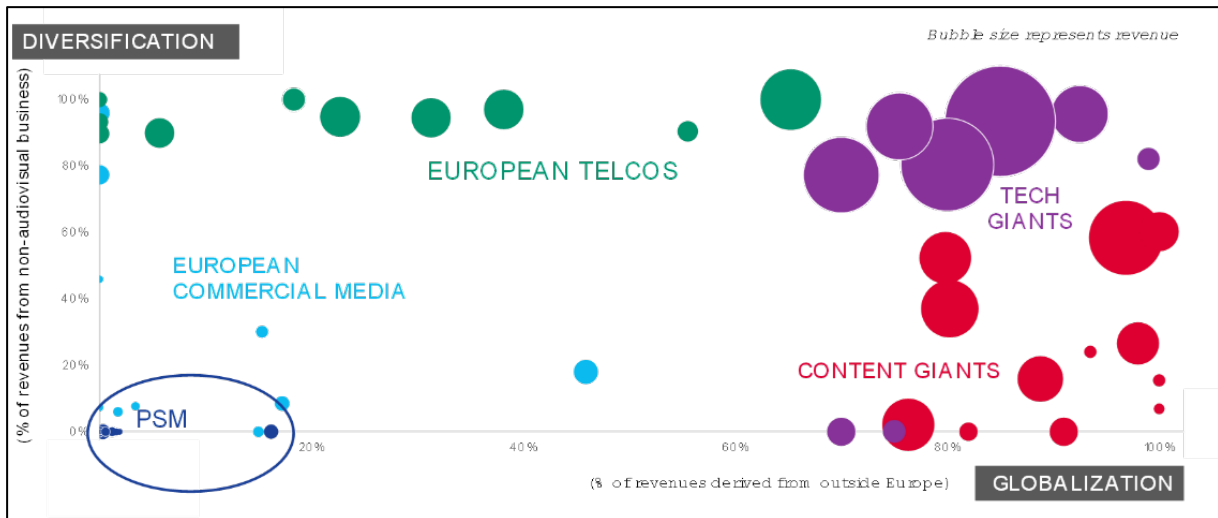
Figure 3: Revenues of selected audiovisual players (EUR billion, %, 2020)[3]
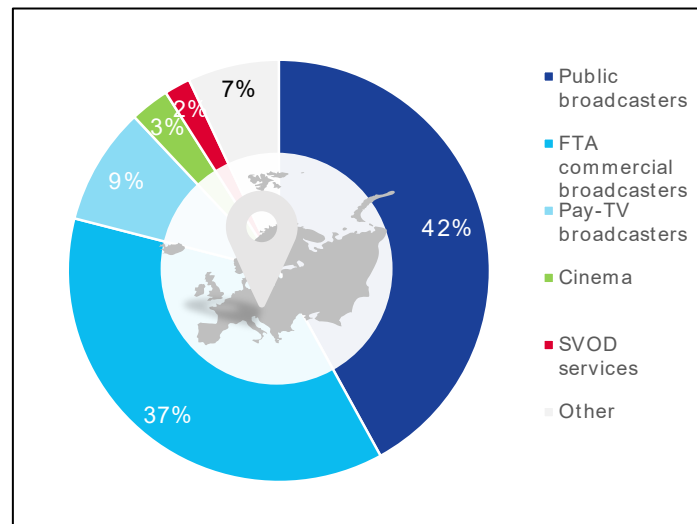
Figure 4: Share in total financing of original European content (excluding acquisitions, sports, and news) (EU27, 2019, %)[4]

OTHER CHARACTERISTICS

Revenues in the CCI sector come from a variety of sources. It can be income from commercial activity, as well as from public subsidies.

The CCI sector is subject to **regulations** specific to its different components, which respond to the objective of preserving cultural diversity, as well as other public interest objectives, such as the strengthening of social links, the development of educational activities, the leverage effects on tourism activity or territorial attractiveness.

Here are some other characteristics of the CCI sector:

---

[3] EBU based on Members' data and companies' financial statements. Bubble size represents revenue
[4] Source: EBU based on European Audio-visual Observatory

- The intersection between economy (generation of wealth and employment) and cultural (generation of values, meaning and identity)
- Creativity at the heart of the activity
- Artistic, cultural, or creative content, including content derived from past creation
- The production of goods and services frequently protected by intellectual property - copyright and related rights
- Innovation and creative renewal
- Public demand and behaviour that are difficult to anticipate
- A sector marked by the diversity of methods of remuneration of work, such as wage-earning and the payment of literary and artistic property rights, as well as by the predominance of micro-enterprises.

Given the different structures within the cultural and creative industries, many of them have few resources to develop their skills in the digital domain, particularly when this is not their core business. There are significant maturity gaps on data and IT issues within the sector.

3   CCI DIGITALISATION AND EUROPEAN CONTEXT

CCI DIGITALISATION

Digitalisation in all sectors is key to reach European Union's strategic autonomy. The Cultural and Creative Industries are no exception and all CCI sectors are now concerned by this transformation.

While for the recorded music, video game and audio-visual sectors, new creations have always been digital from their creation to their exploitation in all its forms, for some other industries, such as cinema or live music performances, still require the public to come to physical locations to access works, such as cinema screenings, concerts, or the sale of books in bookshops. However, digitalisation is also affecting these industries through the digitalization of theatrical projection tools, the development of live streaming or electronic book or the sale of physical books on online selling and/or referencing platforms.

Digitalisation does not necessarily mean total dematerialisation or the disappearance of physical modes of access to works. By digitalising, the industry takes advantage of data, networks, and computing power to transform industrial processes and improve the creation, exploitation, and discovery of works. It disrupts the relationship with the public.

BUSINESS MODELS

Today, the major business models adopted in the industry are all based on intellectual property rights and services offered to the public, such as:

- the sale of products, such as physical media (books, CDs or vinyl, DVDs, or Blu-Rays, etc.) or virtual media (NFT, etc.).
- the sale of access rights to a show, a work, a heritage object, etc., in a physical location or online, generally based on a « ticketing » system that is increasingly digitised and rich in usage data.
- revenues linked to subscriptions to access content aggregation services, linear or non-linear, centralised or not.
- revenues from public exposure to advertising, the targeting of which is a growing challenge, whether for linear services (radio, television, etc.) or not, sometimes combined with the subscription model (freemium).

The development of artificial intelligence, metaverse, NFTs, as well as immersive reality, offer new perspectives whose speed of realisation requires constant innovation, challenging the cultural industries to constantly adapt and adopt new business models.

PLATFORMS FOR ACCESSING WORKS

The « platformisation » of access to works consists of the aggregation, by a third party of creations in the form of digital files which are then made available to the public via indexing and recommendation tools. This trend is today driven to a large extent by global players with considerable resources and

representing real « gatekeepers » to works. Appropriate economic balances must therefore be constantly sought to reward creatives and distribute the value between intermediaries fairly.

An essential feature of this « platformisation » is the automation of part of the modalities of access to works and therefore of the public's capacity to discover them. Works must be accessible in the form of extracts or, in any case, referenced information, so that the public can become interested in them. Any bottleneck in access to works can create an imbalance. For ex., intermediaries may favour access to alternative works that bring more value to them. The purpose of legislation is to re-establish the balances necessary for a dynamic circulation of works compatible with the diversity of their creation and distribution.

One of the challenges is the availability of data on works. Data is needed by operators who want to reference, promote, and make works available to the public. Creative industries also need data on the uses of these works to guide their activity and develop their dynamism.

CURRENT CHALLENGES FOR DATA SHARING

Today, the large amount of data produced by CCIs is currently under-exploited and under-valued. Several financial, technical, legal, and commercial constraints hinder data exchanges within the CCI industries and prevent taking full advantage of data:

Market readiness
Companies and organisations are not always aware of the potential of data for the full development of their sector. They need to understand the impact of data sharing, for the development of their activity. Collective willingness to go forward is mandatory.

Competition
European competitors don't want to exchange their data with each other's. But coopetition is needed to join forces against their main competitors today, the major US- based media platforms (FAANGs).

Regulatory/ legal & compliance complexity
Many regulations are put in place at national and European levels to protect data: GDPR, Schrems2, DGA. They add complexity to data exchanges, while consent management is not currently straight forward. For content exchanges, digital rights need to be handled while there is no automated compensation solution.

Source Identification
Source identification and content provenance are not clear today. There is no simple way to exclude malicious sources or to assess the authenticity of any content.

Data structure and quality
To include new data feeds in existing work process, data suppliers expect that users know how data is structured and how to access it. But it is complex to handle the high number of existing content format multiplied by number of region-specific adaptations, languages etc. with many stakeholders in the supply chain. Moreover, it is difficult to maintain the quality of the data all along the supply chain, especially with 3rd party suppliers that do not benefit directly from it. For example, production companies providing comprehensive metadata, distribution platforms providing timely access to audience date.

Large files transfer and computing
The industry actors often need to handle large media files. It requires high-computation, large Bandwidth transfers and storage.

Interoperability
There is a lack of harmonised identifiers (cross-platforms) for media objects and persons, low adoption of open standards for data and metadata. For example, there are different standards for audience measurement, or definition of a user profile. Content versioning is not harmonised. There is no simple, established way to represent news items (that include formatted text, including audio, video etc), nor standard for exchanging securely fact checks.

Ethical considerations
Media companies have the mission to inform citizens and provide independent and plural views. There is an ethical risk to personalize content and provided biased views.

A EUROPEAN STRATEGY

The European data strategy should enable the EU to become a leading player in a data-driven society. The creation of a single market for data will make the flow of data within the EU and between sectors more fluid, to the benefit of businesses, creators, public administrations, and the public.

Data is at the heart of the digital transformation. It shapes the way we produce, consume and live. Access to and the ability to use an ever-increasing amount of data is essential for innovation and growth. Data-driven innovation can bring significant tangible benefits to citizens - for example in the form of improved mobility - and to the European economy, such as more efficient policy-making and better public services.

Several texts in preparation or recently adopted thus aim at establishing an open, secure, trusted, and transparent ecosystem for a better flow of data across Europe and the development of new business models. The ambition is to create a single market for data in line with the values of the European Union.

European regulation, through the Digital Services Act (DSA) and Digital Market Act (DMA) regulations, has a considerable role to play in inserting new players into the European ecosystem and ensuring fairer competition with traditional operators, notably the media. The Data Governance Act (DGA) will put in place robust mechanisms to facilitate the re-use of some categories of protected public sector data, build trust in data intermediation services and foster data altruism across the EU.

It creates a framework to promote a new business model - data intermediation services - that will provide a secure environment in which businesses or individuals can share data. For businesses, these services can take the form of digital platforms that will allow voluntary sharing of data between businesses or facilitate compliance with data sharing obligations set by law. These services will allow businesses to share their data without fear of misuse or loss of competitive advantage.

Gaia-X and Data Space organisations

Gaia-X is a European Association with 357 members from the industry and academyaimingto ensure European companies can access storage and computing capacities and services, data and data management services that are trustworthy, secure, open, and interoperable. A prime concern is also to guarantee Data privacy and European sovereignty. Gaia-X defines a reference architecture, trust model and federated services for Sovereign Data Exchange, Compliancy and sets up a catalogue of compliant services that make assets and resources easily searchable, discoverable, and consumable. Those services are provided by different entities but are centrally defined and certified by Gaia-X..

Gaia-X, IDSA (International Data Space Association), Fiware and BDVA (Big Data Value Association) have now created the Data Spaces Business Alliance (DSBA) to define a common reference model and join forces to accelerate the development of Data Spaces across Europe and beyond.

4   DEFINING A CCI DATA SPACE

By digitising and sharing data, CCI actors could better compete with global media platforms (FAANG) and foster innovation by stimulating the development of AI-based technologies. Depending on their size and cultural domains, companies and organisations do not all have the same means to invest in technological infrastructures or software required to digitalize or in the standardisation of their data. Nevertheless, there is an urgent need for interoperability, Trust and harmonized 'language', without which digitalisation cannot be achieved.

DATA SPACE DEFINITION

A data space is an ecosystem of exchange, processing, sharing and provision of data between trusted partners, for a fee or not. It is not about copying or repatriating data centrally, but about ensuring that each data holder has full control over the conditions (e.g., who, when, and under what condition) of access to their data. By giving visibility to existing data and facilitating access to different data sources, sectoral data spaces are intended to foster the creation of new services and generate value for Europeans, in Europe and beyond. To be successful and grow, data spaces should guarantee Interoperability, Portability, Security & Trust, and Data Sovereignty. Gaia-X supports the creation of sectoral data spaces and currently develops processes, architecture and/or software that could be used as core building blocks to guarantee these fundamental attributes of a data space.

THE DIFFERENT TYPES OF DATA

1. Digital Assets / Content

Digitised content, digital twins or natively digital content constitute a first type of data that should be easy to find and exchanged in a controlled and secured way. It could be any media files in the production chain, news stories, films, video files, pictures, paintings, photos, podcasts, audio tracks, texts, enriched text, composition playlists, but also subtitles, dubbings tracks, etc… Different versions of a work can exist, for ex. corresponding to each format published.

This data has an intrinsic value as it constitutes the content itself. It must be stored, uniquely identified, and protected. It should be possible to Find and Retrieve digitised content from decentralized storage using federated catalogues, under specific conditions (Rights) set by their owners, for free in case of open licensed content or under their copyright conditions.

For this type of data, key topics include digitalisation, identification and description of content, content metadata creation and enriching, reformatting to standard formats, reversibility, and maintenance of stored data.

2. Metadata

Metadata is data associated with another piece of data. It is essential to ensure the discoverability of the data to which it is associated and the remuneration of the rights holders. It can also provide key additional information such as fact checking info that ensure the provision of trusted content. Its quality is therefore essential. Metadata covers many fields [8]**:**

Category	Description	Examples
Unique Identifiers	Strings of alphanumeric characters that serve to stably identify a document, resource, or entity, regardless of its nature. In principle, an identifier should be unique for each resource.	ISBN (book) ISAN (film) ISRC (musical recording) ISWC (musical work) IPN (*International Performer Number*) ISNI (*International Standard Name Identifier* for authors, composers, performers, producers, etc.), Etc.
Descriptive metadata	Describe the content accurately and objectively.	Author's name Title of the work Materials Date of creation Author's date of birth, etc.
Administrative metadata	Provide information on how the files were created, saved, modified.	Date of creation of the file Name of the company creating the file Format used Identification of the source document that was scanned, etc.
Legal (Or intellectual property) metadata	Indicate whether the work is copyrighted, identify the rights holders and the collecting society, indicate whether the work is licensed for use, and determine the source statements that must appear with the work.	Name of the copyright owner or manager Owner's identifier Wording of the source statement to appear with the title of the work
Enrichment metadata	Provides additional and sometimes subjective information about content. The scope of enrichment metadata is vast. It can be used as an added value by digital cultural content providers to differentiate themselves from the competition. This enrichment data can be produced by artificial intelligence algorithms.	Biography of the artist Photographs Notes Song lyrics Artist's social network addresses Cover / poster image Text of a review, etc. Quality check info Fact checking info Data on investigation conditions

The use of common vocabularies and standards facilitate the development of applications and the exchange, sharing or enhancement of this data and increase efficiency. But they should be flexible enough to gradually cover any type of object, audio-visual, written, photographic or web, broadcast or not, and describe both the intellectual nature (content) and the materialisation, physical or digital (container).

Several standards exist today: the European Broadcasting Union (EBU) has developed and maintains EBUCore and EBU CCDM standards [15], which aim to facilitate the circulation of content between producers and broadcasters. INA (legal deposit of public broadcasting in France) has developed and deployed its own data model to harmonise its various content collection from public broadcasting. BBC also has developed its own models. At Data Space level, while a common ontology needs to be chosen, mappings can be used to harmonize and enable Find and Retrieve of content with heterogenous metadata models from different participants and locations.

3. Resonance data and audience data

Consumer data includes both usage data generated automatically when a given cultural content is consumed online and audience or sales data (or measurements) collected in a more traditional way, ex. panel data or content survey etc.

This data is essential for understanding uses and markets and for adapting offers to demand. They are also necessary for the valuation of advertising space, which is an essential source of revenue for certain CCI players.

There are different standards for audience measurement.

Today, it is difficult to get comprehensive resonance data on a given piece of content that is often published on several distribution platforms (including social medias, websites, events, apps etc).

The Digital Market Act (DMA) now guarantee access to content providers to certain consumption data collected by the major dominant platforms (« gatekeepers »). The Data Act (legislative proposal published in February 2022) aims to stimulate the sharing of consumer data generated using certain connected objects, as well as to encourage the development of standards for the establishment of « data spaces ».

4. Personal data

Personal data are data relating to an identifiable natural person. It is not forbidden to use this type of data, but special precautions must be taken, such as anonymisation. The processing of personal data is highly regulated. The European Union has developed a specific regulation dealing with the particularity of this data: the GDPR (General Data Protection Regulation) of 27 April 2016.

As far as the CCI sector is concerned, this data comes into play in the management of subscriptions to a platform or in the targeting of advertising and the management of consents.

There are several standards, defining what is a user profile based on personal data.

5. Other data

Other type of data is also created, for ex by **each industry process**: logistic data, machine logs, energy consumption, execution reports, etc.

Administrative information is also used to describe entities (person, places, organisations incl suppliers, creatives) or news source (if they do not want to remain anonymous).

Lastly, Open data may also be needed, for ex. to train AI algorithms.

Data space Participants

Data space participants can generate, provide and/or consume data. They should have a verifiable identifier and self-description and agree to comply with Data Space governance rules to ensure trust within the data space.

1. Data holders

Data Holders have generated the data themselves or gained access to it through licensing or other contractual means. A Data Space guarantees data sovereignty for the data holder that defines the condition for sharing their data/content to Data Recipients through licensing transactions. While Data cannot be owned, Data Holders and Data Subjects have legal rights that are being progressively defined in the EU by instruments such as the GDPR, Data Governance Act and Data Act.

Data holders could be independent persons (ex: Freelance journalists, independent authors, fact checkers, end-user /citizens etc), private or public companies (ex: Producers, Publishers, Broadcasters, newsrooms, News agencies, social medias) or joint industry committees (ex. for audience calculation, protection of right holders etc).

2. Data providers

Data providers are collecting and distributing the data, under the specific conditions set by the data holders. They could be archiving companies, producers, editors, publishers/broadcasters, content delivery platforms, public medias, news agencies, Fact checking platforms, Data advertising platforms, global players including social medias, Civil society platforms and Repositories which offer content with open licenses (no copyright violation), NGOs, end-user devices that automatically collect data.

3. Data consumer

Many actors consume data all along the value chain: content creators, TV production, post-production-publishers, broadcasters, distribution managers, linear programming etc

For data on audience and end-users, and resonance data, data consumers could also be Collective Management organizations (CMO), Audience analysts, marketing experts, statistics providers, but also recommendation or personalisation service providers.

Data Consumers in a CCI data space could also include social medias, fact checkers, State agencies, NGOs and journalist protectors, Advertising platforms/AdServers. And of course, public/Citizens/audience in general.

Services, TOOLS, and functions

Many service providers can propose services and tools that provide, alter, or consume CCI data in the frame of the dataspace, for the creation, production, manufacturing, publishing, distribution and selling of content or other types of data.

1. Production and post-production

Many (AI–based) tools are used in production, post-production, publishing and could be data consumer and provider in a media data space:

Complex ingest (all formats and processing), AR/VR production, Translation / Transcription tools (incl. Real Time), Automatic dubbing, subtitling, OCR (Optical Character Recognition), Mining speech to text, Summarization tools, Recognition tools (face, objects, named entity etc).

2. Publication

Tools to improve efficiencies, help automatize publishing processes and compliancy checks could also interconnect with a media Data space to access and process data:

content curation / storytelling, content moderation, automatic detection of controversial information, Fact-checking tools, Federated content quality and filtering out offensive/violence, Automatic Right tracking, payment and clearance tools, AI-based self-regulation (for advertising content) etc.

Recommendation and personalisation tools can also leverage the access to usage/user data and metadata and content. For ex:

Recommendation tools and Automatic suggestions, identification of similar contents / Same speaker recognition, Accessibility services etc.

Lastly, any **content packaging, reformatting** could be optimized using data space effective interconnected (multi-cloud) solutions:

transcoding and packaging, reshaping for social net publication or live streaming or CDN, Formatting to all format (TV, mobile devices, etc) etc.

3. Distribution

Distribution functions can also be integrated with Data Space and benefit from the access to all data:

Search and Discovery / browsing, Automatic targeting of users (ads), Monetisation / Payments, Automatic plug-ins and platform that could be integrated in media website. More generally BI tools, dashboards, google analytics, user statistics (GDPR compliant) etc

4. Service Providers

Service providers could be:

- Cloud providers and local service providers
- Data collaboration platform
- DSP/SSP (Ads marketplace)
- Distribution platforms and CDNs
- Identity management organizations
- Fact-checking organizations
- DRM companies
- SaaS providers

Challenges in creating CCI Data Space

Technical challenges

Standardisation, data harmonization and common definitions are needed to make data meaningful, understandable, and to be able to aggregate data from different providers. It is a challenged today as many different standards and proprietary solutions exist that do not interoperate easily. Consensus to choose one standard is difficult to reach. Significant developments are needed for exhaustive mapping.

Key building blocks for Security, Id management, decentralized findability and traceability, data classification and distribution should be federated across different (existing) platforms.

AI-based algorithms: bias need to be monitored and corrected. Ex: recommendation tools, fact-checking etc. Large representative set of quality data are needed to train these algorithms.

Human resource with technical skills in data space implementation are scarce.

Fact checking relies on scientific method and critical thinking, but it is fundamentally complex to differentiate facts from opinions from conflicting world views.

Business challenges

There should also be a shift from the existing business models used by social media platforms. It is difficult as the current model with major actors (FAANGS) are already anchored in people habits.

Data providers will want to get as much as they give and be sure that their sensitive data is commercially protected.

It is difficult to find the right time to Market and ROI schema to make the business case work.

Governance challenges

Governance body should be a Trust anchor. It should be an independent, neutral organization, chosen and endorsed by all industry players.

Regulatory challenges

National and European regulations are complex and evolving. There are copyright restrictions for European use on content.

Environmental challenges

High computing and storage resources have significant carbon footprint. Blockchains technologies consume a large amount of energy today.

5 USE CASES

Newsroom Use Cases

1. Aggregating news: plurality of news

Goals: Creating a pan-European newsroom that provide trustworthy news from different European countries delivered in real-time to the public in the local language

Data: Trusted News from trusted organisations

Data Holders: News agencies, investigative journalists

Data Consumers: journalists, citizens

Tools and Functions: Transcription/Translation, Recommendations

Difficulties/barriers identified: Need to be independent, public, and free

Interested parties and current Projects: EBU: The European Perspective.

2. Fighting fake news

Goals: For the public, but also for the companies that produce information, the challenge is to distinguish real content from false or out-of-context content. Citizens and/or journalists could query the Data Space to find if a given content has been falsified, or used out of context, and access corresponding fact-checking observation. Goals are to:

- Share fact-checked data between media and fact-checking organisations
- Share content between investigating journalists and fact-checking organisations
- Share fact-checks/observations/knowledge coming from citizens/crowdsourcing/ volunteer factcheckers
- Verify content and tag it with its fingerprint or fact-checking info
- Avoid the spreading of false information

Data: Any stories incl. social media stories, unprofessional images, and video, etc…; fact-checking observations

Data Holders: Social medias, citizens, journalists, fact checking organisation

Data Consumers: journalists, Medias, end-consumers

Tools and functions: Translation/Transcription, fact-checking services and tools, fake news detection tools etc.

Difficulties/barriers identified:

- Need to put in place a federated identification of content and fact-checkers
- Need to engage as many fact-checkers as possible to make the database efficient enough.
- Need to be independent, public, and free.

Interested parties and current related projects: France Televisions, Bayerischer Rundfunk, Athens Technology Center (ATC)

FAKY: a search engine dedicated to fake news detection. Created by RTBF-English version is on its way [18]

WeVerify (Horizon 2020): provide various tools including a detector for Deep fakes, which are developed further in vera.ai project to start September 2022. DBKF is a platform where articles of debunking images are gathered

Bellingcat: independent international collective of researchers, investigators, and citizen journalists – developing tools for open source and social media investigation [19].

PRODUCTION AND PUBLISHING USE CASES

3.   Marketplace for digital works

Context: More and more productions today use digital works (for ex. digital sets such as monuments, landscapes, buildings, etc.) These assets are often created on demand for a given production and are archived at the end of the production.

Goals: Create a marketplace allowing the identification, description and discoverability of all works produced and distributed in Europe, to facilitate their re-use.

Data: Description of any digitized work – virtual sets, pictures, images, music, recordings, masters, PADs, subtitles, dubbing etc that could be used for creating new films, music, podcasts, books etc.

Data Holders: Creatives, Production companies, post-producers, etc.

Data Consumers: Creatives, Production companies, medias

Tools and Functions: DRMs, 3D/Virtual Reality /Augmented reality

Interested parties, current projects: France Televisions, TAMIS, IWA, CREA (for writings), MILC (for videos).

4. Co-Creation

Context: More and more productions are done by teams of creatives from different companies and locations.

Goals: Have creatives in different location, collaborate, and develop digital works to co-create content using a set of production tools, repository of Digital works, available from the Cloud.

Data: any digitised work – virtual sets, pictures, images, music, recordings, masters, PADs, subtitles, dubbing etc that could be used for creating new films, music, podcasts, books etc

Data Holders: Creatives, Production companies, post-producers, etc.

Data Consumers: Creatives, Production companies, medias

Tools and Functions: DRMs, 3D/Virtual Reality /Augmented reality

Interested parties, current projects: SSRSRG.

5. Catalogue of physical sets

Context: Audio-visual productions for television or cinema use many set elements, often manufactured on demand. However, reuse of these sets remains low because no database, to date, lists all the sets manufactured and available with the descriptive elements needed to make an informed choice (material, colour, storage location, condition, availability, etc.).

Goal: By referencing these sets with standardised metadata, Production companies could save raw materials and make the construction of the sets profitable thanks to a reinforced valorisation. A trans-European database would allow local know-how to be enhanced and would increase the base of available seats. The history of each set and its usage could be known and linked to all works that used it.

Data: description of physical sets

Data Holders: Production companies, set makers

Data Consumers: Production companies

Tools and Functions: Metadata enrichment

Interested parties, current projects: France Television, encouraged within the framework of the eco-responsible Ecoprod approach.

6. Producer Portal for Metadata

Context: Today, metadata is complex for broadcasters to collect and obtain. Producers and rights holders do not have systematic processes to create, store and distribute metadata.

Goal: The ***producer portal*** is intended to be a place where this metadata can be delivered at once or incrementally to support the production process. It also allows the broadcaster's teams to assess the quality and relevance of the metadata provided to correct or complete what needs to be corrected. Once validated, this data is transferred to broadcasters' data product Programme for use on each of the broadcasting vectors.

Data: Metadata on production files

Data Holders: Creatives, Production companies

Data Consumers: Broadcasters

Tools and Functions: Metadata enrichment, face recognition, tag generation, etc.

Interested parties, current projects: France Televisions, CST: TAMIS project, Media-cloud.ai (for AI-based services such as Metadata enrichment).

7. Content Production Validation

Context: Editorial teams perform manual checks to ensure compliancy to regulations and ethics and alignment with editorial goals and audience needs. They often lack visibility on digital rights that may lead to hidden costs to produce.

Goals: Automate editorial checks to improve efficiency and mitigate risks (ex. non compliancy to regulation, on-budget production, reputation)

Data: any content, including ads; editorial rules, legislations, data on actors, sets, etc.

Data Holders: any holder of content and production assets

Data Consumers: medias, publishers, etc.

Tools and Functions: AI-based compliancy checking tools

Interested parties, current projects: ARPP (France): Invenio project (for Ads)

8. Digital Right Management

Context: To ensure the collection and distribution of royalties for public broadcasting in the audio-visual media, Collective Management Organisations (CMOs) currently uses statements provided by each media. Each media invests dedicated resources in monitoring its channels and compiling these statements. Despite this, key information about the programmes and associated musical works is sometimes missing to allow for the fair distribution of the rights associated with the authors-composers and publishers of these works. This in turn creates a significant workload for the collecting societies to analyse the statements provided by broadcasters and to achieve the best possible distribution.

All stakeholders in the audio-visual creation ecosystem are impacted by this ineffective process:
- **Programme producers** have original data on the broadcast works but communicate them little or poorly to the other players.
- **Media** fulfil their legal obligation individually, potentially calling on service providers to produce and enrich the broadcast data, which results in significant human, technical and financial costs for them.
- **Collective Management Organisations (CMOs)** receive data that is often incomplete or inconsistent, which requires a huge amount of manual verification and enrichment work and results in a long processing time, whereas the accuracy and speed of payment of rights is a key issue for the creators' activity.
- **Rightsholders.** at the end of the chain, depend for their remuneration, both in terms of accuracy and timeliness, on the quality of the data and the implementation of efficient workflows between the stakeholders mentioned above.

While this is not a new issue, it is becoming more and more pressing for traditional players as the audio-visual sector is digitalising. Indeed, digital technology has generated a massive increase in the consumption of audio-visual content online, and therefore a colossal increase in the amount of data to be processed by CMOs. At the same time, digital technology provides a tremendous opportunity to facilitate the sharing and analysis of data.

Goals:

- **Technical:** speed-up data processing times, make them more reliable and simplify their reconciliation and enrichment; limit the production and manual processing of data.
- **Economical:** reduce the costs of these activities by pooling some of the processing and investment required to build the IT tools. Possibly, in a second phase, create new revenues from the new data thus generated.
- **Political:** develop a partnership strategy between the players in the sector to defend a common vision of cultural creation and technological sovereignty.

Data: usage data, broadcasting reports, rights associated to any content

Data Holders: Program producers, Media, CMOs

Data Consumers: CMOs, Right holders

Tools and Functions: Digital Right Management

Interested parties, current projects: SACEM: Media Hub project, MediaVerse project

9. Optimize carbon footprint

Context: The accurate measurement of the carbon footprint of the publishing, production and distribution workflows will progressively become mandatory. There is a need to bring together the actors in the value chain who have already calculated the carbon footprint of their activities.

Goals: Create a repository of measurements by type of activity, standardise measurement methodologies and implement a European service to help in computing these footprints, in the form of a software platform.

Data: Carbon footprint data

Data Holders: all actors in the supply chain

Data Consumers: all actors in the supply chain

Tools and Functions: Carbon measurement tools

Interested parties, current projects: Cap Digital (France).

10. Innovate

Context: AI based innovation require access to large pools of data to train and improve algorithms

Goals: Aggregate and give access to large sets of quality data to innovation labs and start-ups so that they can train and improve their AI-based solutions. Ex : translation / transcription tools- dubbing- face recognition etc.… for ex. starts.eu [20]

Data: any

Data Holders: all actors in the industry

Data Consumers: innovation labs and start-ups

Tools and Functions: Any AI-based SW

Interested parties, current projects: Media-cloud.ai

11. Metaverse

Context: Metaverses are under construction. It will be virtual places with often user generated contents. These virtual places will require also safe places to access trustworthy content. Traditional Media actors shall then be able to publish content in metaverses.

Goals: Share tools and digital assets to create and publish content

Data: digital assets, AI based tools

Data Holders: content producers, start-ups, etc.

Data Consumers: content producers

Tools and Functions: mapping of avatars, livestreaming solutions for metaverse, virtual production, immersive audio creation tools, etc...

Interested parties, current projects: France Télévisions, ….

DISTRIBUTION USE CASES

1. Pan European B2C platform for content on demand and live streaming

Context: Large global platforms such as Netflix or now dominating the market for on-demand content. European media of all kinds need to join forces to compete with such major actors.

Goals: Distribute content from all Europe in a B2C platform that manage digital rights and provide accessibility services.

Data: state of the art media content

Data Holders: Medias

Data Consumers: end-consumers

Tools and Functions: DRM, Recommendation tools, Translation/transcription, and any accessibility services

Interested parties, current projects: EBU, Swisstxt, SACEM: Media hub project, Mediaverse project, European Backbone.

2. Sustainability and promotion of (heritage) content/archives

Context: The audio-visual heritage (mainly cinema) remains largely stored on physical media. Access to these works remains difficult.

Goal: Digitize, promote, and distribute digitized (heritage) content to a wide audience. Digitalization obviously goes hand in hand with the creation of all the metadata necessary for referencing and discoverability of the works.

Data: audio-visual heritage content (mainly cinema?)

Data Holders: Archives, Production companies

Data Consumers: medias, end-consumers

Tools and Functions: Digitalization of archives, Metadata enrichment, formatting tools, colouring/video quality enhancement tools etc, federated catalogue

Interested parties, current projects: France Televisions, EBU: Balkans Archives project

3. Audience Analysis

Context: For any business model today, access to consumer usage data, data related to their Internet browsing habits or their tastes, and audience data, is essential. This data makes it possible to adapt content to consumer tastes and needs, improve discoverability by targeted audience, scale promotion, plan publication, or develop new offers. It is also required to optimise advertising and marketing campaigns and measure their effects.

When cultural players obtain access to their content usage data (either from platforms and service distributors or from their own distribution channels such as their website), they do not always have the means (technical, human, or financial) to harmonise, analyse and extract all the information they could benefit from.

Moreover, each player in the value chain currently collects and processes this data on their own, which only gives it a very fragmented view of users' tastes and needs.

Goals: Aggregate usage data and perform refined analysis on large pool of data to better target cultural offers. The public will get a better match between the proposals and their tastes, interests, and prospects for discovery.

Data: Usage data from different distribution channels on content and ads

Data Holders: distribution platforms, website, apps, social medias

Data Consumers: media companies, publishers, advertising agencies

Tools and Functions: BI analytics

Interested parties, current projects: DPA: DRIVE project, VRT: Cross-platform audience measurement (Flanders- starting 2023) [21]

4. Consolidated user profile and Consent management

Context: The obtention of advertising consent is currently given by the user for each data controller (often an advertising agency or a publisher). It is therefore complex and annoying for the consumer to trace/port his consent as the GDPR regulation allows.

Goal: By collecting all usage and user data, a data space can provide consolidated consent management in compliancy with GDPR. This C2B ecosystem would allow the user to manage his consent, profile, and interests for several publishers/ advertising agencies within the same interface.

Publishers and agencies can minimise legal risks linked to non-compliance with GDPR texts and be updated on possible changes in user consent over time.

Users have better visibility of the status of their consents and a better experience in updating their consents using a single interface. They could also have visibility of their predictive interests and ability to update them.

Data: User and usage Data

Data Holders: end-consumer

Data Consumers: Advertising agencies, publishers

Tools and Functions: consent management- GDPR compliant. Recommendation, personalization tools, BI analytics etc.

Interested parties, current projects: FTV, VRT

5. Personalisation

Context: As users have access to a huge amount of information, there is a need to provide trusted filters or personal guide to content. They become upset with traditional generic advertising and turn to ad-blocking Software.

Goals: Create recommendation engines that enrich cultural life of consumers, by encouraging discovery of new content. Provide users with useful personalized info or advice while promoting a brand or a product.

Data: Usage data, user data, context data; content and its metadata

Data Holders: any content owner, end-consumer,

Data Consumers: content distribution platforms, medias, end-consumers

Tools and Functions: recommendation tools, analytics, personalisation tools

Interested parties, current projects:

6. Targeted Advertising

Context: Advertising revenues are one of the three sources of revenues for CCI industries. They depend on the Audience size, the qualification of audiences and the ability for a content publisher to address an audience and expose them to an ad. Each publisher (or each advertising agency associated with a pool of publishers) builds and grows its own audience to monetise it. However, as advertisers want to reach targeted audience, publishers are now requested to segment their audiences. But each of their individual audience pools is often too small to accurately represent each segment.

Goals: Collect all (anonymized) usage and user data across publishers and agencies and develop new service offers to advertisers such as Cross-agency statistics, Cross-agency segmented audiences and Anonymised or pseudo-anonymised information on the entire sector.

Publishers and agencies have better knowledge of user behaviour and interests; They can also monetise the data and/or knowledge "as a service" through the shared ecosystem.

Advertisers could have visibility of audience volumes across agencies and better knowledge of user affinities.

Data: Usage, user data

Data Holders: End consumers, social media platforms, any distribution platform

Data Consumers: Advertisers, Advertising agencies, publishers

Tools and Functions: BI analytics, Recommendation, Personalization, etc

Interested parties, current projects:

7. Antipiracy

Context: Almost every creative work, including novels, songs, movies, software, and apps, may be pirated online. The global movie industry's revenue losses from digital piracy are between $40 and $97.1 billion per year. Uploads and Downloads of stolen content represent about ¼ of all worldwide internet traffic [11]. In Europe, a user accesses to illegal content in average 5.3 time per month [12]. It varies greatly depending on the country and the type of content (20% film, 10% music and 70% TV).

It impacts creative professionals, rights-holders, and their distribution partners. Anti-piracy measures require alignment and cooperation between all actors involved in the content lifecycle, including competing distributors, that do not always have the same incentives to fight against piracy.

Goals: A dataspace can enable trust and cooperation among all actors (including competitors) to fight against piracy. Thanks to consolidated and secured access to content metadata including identifiers, watermarking and fingerprints, Distributors, Social networks, and Search engines, can monitor any content published and verify real time its authenticity. AI-monitoring of video streams, watermarking or fingerprinting detection could be provided as a service within the Data Space. Once illegal content is found, and its source identified across distribution channels, real-time action can be taken to quickly stop the distribution.

Data: content metadata, fingerprints- content streams on distribution channels

Data Holders: creative professionals, rights-holders, and their distribution partners

Data Consumers: creative professionals, rights-holders, and their distribution partners

Tools and Functions: Watermarking, fingerprinting, AI-based monitoring tools or manual monitoring services.

Interested parties, current projects:

BUILDING BLOCKS FOR A CCI DATASPACE

This part has been developed based on initial inputs from some representatives from the CCI sector, gathered in workshops organized by EBU in May 2022. It is a starting point to future extensive analysis and work for the definition and development of the building blocks needed to support the Use cases listed previously.

GOVERNANCE

1. Trust Anchors

Trust is the foundation of a dataspace. To bring Trust, Trust anchors need to be engaged at different levels in the Data Space. Existing regulation define roles for Trust in data exchanges such as Data "intermediaries", ex. ledgers services, as defined in Data Governance Act (DGA) [9] or Trusted flaggers as defined in Digital Service Act (DSA). In the CCI industry, such Trust anchors could be:

- Traditional/ historical alliances, such as Identifiers registration authorities
- Existing Trusted associations /networks such as EBU and public Media companies
- Standards and certification companies
- Standards such as defined by Journalist Trust Initiative (JTI), C2PA, or the European Fact Checking Standards Network (EFCSN)
- Well known individuals (bloggers, influencers …), brands
- Journalists' reputation: News agencies are privately owned entities, and trust is given on an individual basis, but not always at organisation level
- Fact checking / Debunking organizations
- Moderation communities (Wikipedia)
- Open-source associations
- Data trusties that are obliged to the users to handle user data
- Consent management by data owners

2. Governance rules

General acceptance of clear common governance rules, together with measures for their monitoring by neutral parties, is the foundation for ensuring trust within the data space. These rules define functional, technical, operational, and legal aspects of a data space. They are set by an initial consortium of members, implemented, and monitored by a data space operator and maintained by a data space governance body [7]

Participants management:
- Each participant must agree to a Data Space charter at registration. Registration process should be controlled by a trusted authority (for audiovisual: CNC, ARCOM, ALPA etc).
- Close monitoring and analysis of participants' activity should be done. There could be reputation & evaluation of participants through user scoring.
- There should be possibility to exclude participants when they do not respect the rules. ex. sharing data to external organisation without authorization etc, spoiling analytics with wrong data sets etc.

- Participants should be able to know each other, thanks to available info in Self Descriptions and during industry events.
- Clear rules should be set to manage relationship between competitors within the data space.
  Ex: in DRIVE project, there is a common clear goal agreed by for all participants to combine data analytics with joint learning and practical application. But a participant can block competitors from accessing its data.

Safeguarding Trust
As explained in the next part, Technical Building blocks and procedures should be chosen and implemented to safeguard Trust& Security, Interoperability (data standards, Connectors with participants and external stakeholders e.g., other data spaces etc), traceability of transactions, Data sovereignty (consent management).

Technical rules
There should be a clear choice of Identifiers and data/metadata standards used for Search&Discovery. The quality level of metadata should be defined and monitored.
Anonymization/pseudonymization tools and procedures should be decided and monitored to mitigate risk and guarantee Security and Trust.

Continuity model
Continuity model also need to be defined for the management of changes, adoption of new standards and agreements, adoption of new technical building blocks or SW releases and procedures, risk mitigation of new threats etc. Exceptions should also be handled; disputes should be arbitrated.

Communication and Transparency
Clear communications, bi-annual reviews and setting clear goals and regular audits should be scheduled and communicated to participants in a transparent way.

3. Governance body

A Governance body should be appointed for decision making, steering and arbitration and conflict resolution. [5] In some initial discussions, some CCI actors mentioned the creation of a neutral consortium of members, with a non-for-profit status, with representatives from across the industry, including private and public stakeholders, data holders and data consumers[6]. The choice of the Governance body is a key topic that requires extensive study and industry buy-in without which a media dataspace will not be a success.

4. Data space operator

A Data Space Operator should ensure that the governance rules are implemented, and that data is easily discoverable (based on a set of standardized data models). It should be able to provide information of all data-related transactions within the data space (who used what data, when and under which conditions). Data space Operators also need to put in place mechanisms ensuring compliance with data usage policies. It should help in case of hostile activities.

5. Auditing

Auditing process and committee should be in place to guarantee Trust and Security in the data space. Auditing could be done by an 3rd party expert, external to the data space – ideal for transparency but

costly- or by experts from the auditing committee if it is neutral and trusted by participants because appointed by the governance body and a representative committee of participants.

TECHNICAL BUILDING BLOCKS

1. Common building blocks and Gaia-X Federated services

As already mentioned, a data Space guarantee some core values, in particular interoperability, portability, trust, security, traceability, transparency, openness, data sovereignty. A data space organisation can already ensure generic compliancy to these values by adopting standard frameworks and certified building blocks such as Gaia-X federated services, similarly to any data space in other verticals. These (preferably open Source) building blocks will allow to:

- **Authenticate and authorise**, users and systems in decentralised and self-sovereign manner. ex. with digital signature, Verifiable Credentials, WebAccesscontrol [23], FIDO free and open standards for authentication [24]

- **Securely store and manage credentials** for authentication and authorisation

- **Establish trust** in participants and systems in decentralised ecosystems, by managing self-descriptions. Self-descriptions are machine readable descriptions of participants, resources, or service offerings. They can be digitally verified using Gaia-X Notarisation API.
- **Provide Trust Service APIs** that are implementation of **cryptographic functions** to ensure consistent level of trust between all components and participants, through Encryption and Hashing. Homomorphic encryption allows sharing and analysing data without revealing it.
- **Find resources, participants'** self-descriptions (machine readable) in Federated catalogue.
- **Negotiate and Define Data Asset Usage Policy** in Data Contract services for data exchanges in secure, trustworthy, and auditable way.
- **Log exchanges,** to check if data has been transmitted, received and if its data usage policies have been respected
- **Monitor continuously** and automatically if a service is compliant with data space rules and its self-descriptions.
- **Validate participants and offerings** before on-boarding them.
- **Interact via API calls with core service functions** via a Portal Service, with intuitive user interface and corresponding back-end functions.

In the next parts, we will focus on some specificities of a CCI dataspace that relate to the creation, handling, and discovery of digitised works.

2. Harmonising data

Whether a cultural piece of content is a show, a film, a television programme, a book, a museum object, a musical recording or a heritage item, a single naming and data description system would make it possible to identify the entities relating to this content in a unique way (name of the author, the place, the production company, etc.). It will allow discoverability across distributed systems and platforms through manual exploration of the data space using federated catalogues or through direct

or indirect recommendation and personalisation services. It could also help rationalising data flows as many players often collect and process the same data with slightly different objectives.

In media industry, many different standards or format are used for **identifiers, tags, taxonomies/ontologies, and metadata**, but also content consumption data, audience measurement metrics etc ... For ex. different initiatives to adopt common identifiers across the industry have not been a success so far: ISAN registration has been requested by the CNC in France since 2017, but its actual use remains low in audio-visual deliverables. Same with EIDR, that is requested by OTT platforms.

Each of these standards are used only by a limited number of compatible platforms and are sometimes incompatible with other sectors. Mappings for moving from one system to another are often developed for interoperability between systems.

The first challenge is therefore to choose among existing standards, which one(s) will be used as a basis within the dataspace, for federated Search and Discovery. Mapping solutions could then be put in place to ease the transitions from/to other proprietary or standard solutions. One or several **Open standards** should be used, such as:

- Ontology and linked data for content description: RDF [13] including FOAF describing persons [14]**,** EBU CCDM /EBU Core: [15]**,** schema.org**,** BBC work on ontology

- Standards for Ads businesses: World Federation of Advertisers (WFA)

- Specific models used in existing content sharing platforms. Ex: PEACH schema for stories, used by PEACH participants [16].

A Data Space should also ensure **quality data** is available by establishing requirements and good practice and enforce them within the dataspace with appropriate governance. Metadata associated with works should be complete and enriched, as it is the cornerstone for the discovery and sharing of data between the different actors as well as their governance. It also serves as a basis for the implementation of smart contracts between the various actors in the sector.

Following services could be used as mandatory key building block of a data space, to upgrade and align the quality level of data. Or they can be provided as a service that ease data discovery and sharing:
> *Automatic generation of metadata, pooling metadata, Automatic keyword creation, categorization, tagging, Automatic content recognition, detecting duplicates / copy of images, automatic creation of unique identifiers etc.*

3.   Data and Content Discoverability

As the amount of cultural content available is exploding, the ability to provide visibility and access to a piece of content to an ad-hoc audience are key. The World Wide Web is an almost infinite space of exploration in which the public evolves thanks to tools such as search engines that use algorithms leading to natural referencing (SEO) or paid referencing (SEM) of content.

In a Data Space, a Federated catalogue make sure participants can find best-matching data with up-to-date complete descriptions (Ex. Google data sets search [17]).

For content data, all digitized work should be uniquely identified (for ex. using fingerprinting) and associated with metadata, labels and indexing that will enable algorithms to detect its compatibility with users' tastes or to propose it as a discovery.

Ontology/linked data and Knowledge graph, Sparql, GraphQL, Linked Data fragments etc as well as Semantic AI, NLP vector embedding together with semantic data are key technologies for Federated catalogues.

4. Content Provenance, and Traceability in Data Space

As digitized work can be easily reproducible and modified, it is difficult to distinguish true information and misinformation, genuine works, and copies.

Traceability is a key function within a data space, that can be implemented using technical solutions such as immutable Databases or Distributed Ledger Technologies (DLTs), can be leveraged to provide trustworthy provenance information.

Content creators and editors, regardless of their geographic location or degree of access to technology, can disclose information about who created or changed an asset, what was changed and how it was changed. Content with provenance provides indicators of authenticity so that consumers can have awareness of who has altered content and what exactly has been changed.

This ability to provide provenance for creators, publishers and consumers is essential to facilitating trust online, fight against piracy and misinformation [10].

5. End to End Digital Right management

There are diverse rights, regulations, and economic models for exploiting a cultural content. For a global exploitation of a given content, one needs to comply with different legislation, at national and European levels, and keep up with new technical developments and use cases.

Consequently, the capacity to access and process data associated with works, associated rights, and their usage across distribution platforms in various countries, is crucial to ensure fair remuneration for right holders and their economic partners.

Data space can provide a protected environment where any type of digitized work can be licensed to its owner/creator, with automatic execution of contracts, for ex. through Smart Legal Contracts (SLCs), that rely on fresh and accurate usage data. A CCI data Space could also leveraging Distributed Ledger Technologies to notarise and execute transactions such acquisition or transfer of rights and provide auditability and revenue payment splitting. [3]

6. Preserving privacy

Media and advertising companies can no longer use third-party cookies to understand their customers' online habits. By collaborating and sharing their user and audience data, they could all better understand if they are reaching the right audience with the right content (ex. Drive project).

They cannot just share their user data as it is protected by GDPR though. There should be ways to "share without knowing" [22], in secure neutral environment such as a Data space. Data privacy tools and procedures can be performed either centrally as a building block of the data space or upstream by the data holder before transferring its data.

7. Data sovereignty: consolidated consent management for CCI use cases

Consent management is also a key building block for many data spaces. Its implementation can a priori be re-used by different verticals such as healthcare, banking etc. Several use cases in a CCI data space could benefit from it, such as advertising, recommendation/personalisation tools, audience analysis etc... (See Use Case part)

BUSINESS MODEL

.

1. Data space operating costs

Public sponsorship is needed to launch the initiative, design, and implement the initial soft shell, support operational costs while some lighthouse projects are executed and attract first data holders, data consumers, and develop additional use cases.

Data Space operating costs should be paid to Data Space Operator: it could be a fixed fee (for ex. paid by stakeholders that are pooling metadata) or (small) % of dataspace usage (content) costs. It could depend on size of participants (ex. audience size, financial resources of the CCI company) and be more expensive with SLAs for larger participants. It could also be financed by a share of Internet Access fee or by public European and national funds.

2. Commercial models for participants

It should be possible to have **different commercial models** in the data space depending on participants and use cases: share data for free or fair cost-sharing for non-commercials (ex. public sector or a defined group of members) and monetised data sharing (ex. Private sector).

Other commercial models may be used for some services, ex. crowdfunding tools for content creation, fact-checking as a Service etc. In Data Governance Act (DGA), data Altruism refers to people voluntarily donating their data for the public good.

Different monetisation models can be adopted, with a preference to direct monetization: Egress costs (per Mb, per month, per connection) or subscription based (with volumes included). Prices can also consider complexity of content and cost to produce.

Ex: Participants in DRIVE project pay a monthly service fee to access full services, incl. benchmark data from all participants and research on BI analytics.

3. Fact-checked news

Citizens should get free access to fact-checked news that should be omnipresent and easier to find than fake news. As it is in the interest of Democracy, it should be (at least partially) state funded.

Professionals would pay per use for access to data or services. Ex. fact checking for new website or online magazines that have an "elsewhere" section on their websites, and that is often full of misinformation today.

Media independence should be guaranteed in the data space.

6   EXISTING PROJECTS

The use cases listed in the previous section gave a large overview of how data space can benefit to the CCI industries. In this part, we list projects that are already on-track across Europe to develop platforms that will implement some of these use cases. Developing interoperability and a federating upper layer to these projects could constitute the basis for the creation of a sectoral data space, if they are ready to participate in the creation of an open, non-proprietary ecosystem without a monopolistic approach and agree on common rules and governance.

Projects	Data type	B2C /B2B	creation	Checks	Search& Discovery	Recom- mender	Consent mgt	DRM	Audience analysis
The European Perspective	Trusted News	both			x	X (PEACH)			
DRIVE	Usage data	B2B				x			x
MediaHub	Music	B2B						X	
CREA	writings	B2B	X					?	
TAMIS	Metadata	B2B	X						
MediaVerse	images, videos, 3D/VR	Both			X	x	x		
MILC	video	Both			X			X	Audience Feedback
C2PA/Origin Project	A/V/text	B2C and B2B		Content origin					
Balkan Archives	Archives Metadata	B2B			X				
ECB		Non-profit A/V content			X	X			
EDMO	Fact checking	B2B		Fact checking					
Alliance Culture Data	Any	B2B/B2C			x			x	
IWA	Digital twins	B2B							
Mediacloud.ai	AI-based services	B2B							

Invenio	Ads	B2B		Compliancy checks					

1. ## The European Perspective: a pan-european newsroom

Scope: B2B and B2C platform for trusted news exchanges

Link: https://www.ebu.ch/news/2022/05/a-european-perspective-aims-to-forge-stronger-bonds-between-europeans

Initiated by: EBU

Participants:

EBU Members from Belgium (RTBF), Finland (Yle), France (France Télévisions), Germany (BR-ARD), Ireland (RTÉ), Italy (RAI), Portugal (RTP), Spain (RTVE), Swedish Radio (SR)Switzerland (SWI swissinfo.ch) as well as ARTE, the Franco-German broadcaster.

Goal:

"A European Perspective" brings together public service media who share online news content that offers their audiences fresh insight into pan-European issues.

Participating public service media contribute to a news recommendation service that helps European citizens contextualise trending stories and better understand how neighbouring countries are coping with these challenges.

This initiative's backbone is supported by EBU technology. Each participating organisation shares its stories through a Digital News Hub, where they are translated into multiple languages using the EBU's EuroVOX tool and robust editorial workflows.

Then the Eurovision News Monitoring Tool allows participating broadcasters to monitor online content produced by their counterparts in real-time, using AI-assisted translation, automatic content recommendation tools (PEACH) and filters to identify items of interest that will resonate with national audiences.

2. ## Drive project: Exchanging publishers' BI analytics

Scope: BI analytics, Classification of content and automatically distribute content to target groups based on content models and user behaviour.

Link: https://innovation.dpa.com/2021/05/27/drive-daten-erfolgreich-teilen/

On the initiative of:

DPA and Schickler

Project participants:

The Digital Revenue Initiative (DRIVE), established by the German news agency DPA and Schickler consultancy in 2020, is built upon a shared data space of currently more than 20 newspaper publishers from Germany and Austria. In 2021 DRIVE received the Global Media Award by the world publishing organization INMA as the best industry innovation.

Solution provided:

The central project team of DPA and Schickler is operating an infrastructure which combines:

- Tracking of news usage on all publishers' websites
- Import of the collected usage data into a central data lake
- Application of AI analytics
- Visualisation in dashboards
- Resting of hypothesis in A/B experiments
- Benchmarking and collective learning in virtual teams to increase usage, user loyalty and the conversion into digital subscriptions.
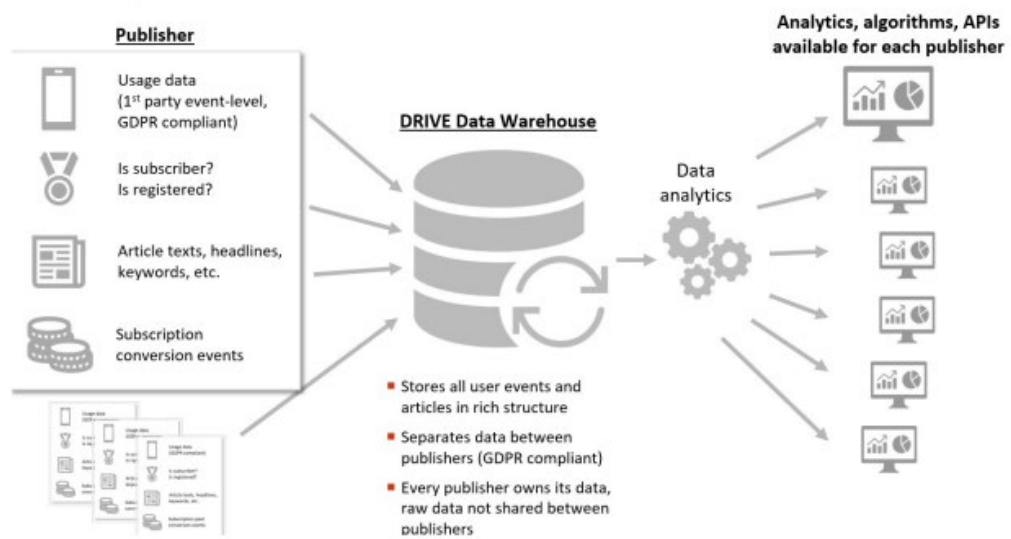

Figure 5: DRIVE Central Data/AI infrastructure

-

DRIVE is a subscription service. Publishers are charged with a monthly fee. The revenue is fully invested into technology, infrastructure and the recruiting of data analysts and digital publishing experts. Every

publisher is represented in a steering committee which decides not only on the general project strategy but also on budget issues and the monthly subscription fee.

A key element of DRIVE is the translation of data analytics into practical recommendations for editors following the slogan "Make the data work". In weekly video meetings of the "Newsroom Track" data analysts and journalists exchange the latest findings and match data theory with operative reality to jointly improve results.
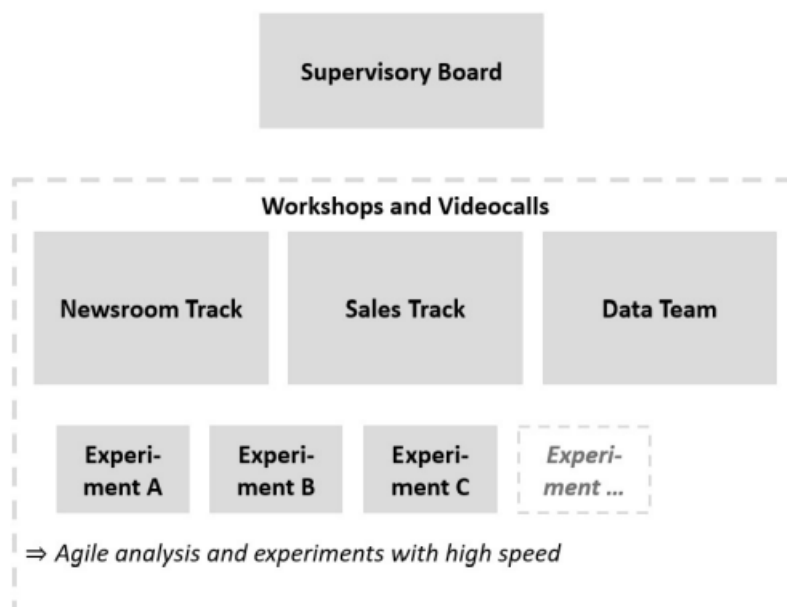

Figure 6: DRIVE translating data analytics into operations

The Basis for this exchange is the central DRIVE metric media time. On the hand DRIVE is measuring the performance of stories, topics or formats by the time users spend to consume them. On the other hand, media time is applied to compare the individual user engagement. The more time an individual spends on a media site the higher the tendency to convert him/her into a subscription and build a loyal customer relationship.

3.   Media Hub Project: Accelerate remuneration of rights holders

Scope: broadcasting report collection and automatic Digital Right Management

On the initiative of: SACEM

Potential participants: France Télévisions

Background: (see Use Case: Digital Right Management")

Sacem protects, represents, and defends the interests of over 182,520 members in France and abroad, authors, composers, and publishers. In the service of creation, it has the largest repertoire in the world and is committed to musical diversity by supporting numerous projects. Sacem is a private, not-for-profit company managed by its members, whom it supports through 3 essential missions:
- Collecting and distributing royalties
- To promote and support creators
- To defend and protect its members

Sacem collects and distributes royalties for public broadcasting (audio-visual media, concert halls, festivals, internet services, cinemas, shops, etc.) and for the reproduction of the works it represents on media (discs, videos, legal digital files, DVDs, CD-ROMs, video games, etc.). Copyright is distributed as closely as possible to the actual use of the works: 80% work by work, 9% by survey and 11% taking into account consumption habits.

Project goals:

A "Media Hub" is positioned transversally within the ecosystem. Its mission will be to provide producers, media and CMOs with a common technological platform, allowing for the acceleration and reliability of the process of declaring data concerning works and their broadcasting, by:

- Facilitating the production of broadcasting reports by the media
- Ensuring the quality and completeness of the data, in particular the data relating to the works, included in these statements
- Sharing the necessary data between the stakeholders in the rights distribution process
- Pooling and industrialising technical resources between the partners in the form of a platform
- Ensuring the security and sustainability of the platform by allowing partners to choose the level of confidentiality of the data shared within the Hub
- Enabling the reproduction of processes that can eventually be used to offer derived commercial services.

Expected benefits:

The deployment of the "media hub" will bring several benefits:

- Fairer and faster remuneration of rights holders and thus overall support for creation in a responsible and sustainable model
- Overall cost reduction and productivity gains for all stakeholders
- Simplification of the processes of collecting and processing the data necessary for the distribution of rights for the CMOs and broadcasters
- Taking into account new uses and innovation in the field of collective management, in order to perpetuate the French leadership in this sector.

Main Gaia-X technologies at stake:

Data platform: sharing, exchange and potentially intelligence of data between several players in the CCI sector, with the potential involvement of public authorities, to defend a French-style model for defending copyright and creation, while taking into account the evolution of uses in the face of the risk of disintermediation.

Security and confidentiality: authentication, management of identity federation, management, securing of data, management of consent and control of use.

Compliance: data portability, rights management, onboarding, and certification.

4. CREA project: Collaborative platform for the creation and distribution of editorial content

Scope: Creative Room European Alliance (CREA), a place to co-create, publish, and monetize quality content (writings)

Link: https://panodyssey.com/fr/article/tech/press-release-panodyssey-8djesa5hkadp

On the initiative of:

Agencia EFE, BSC SME, Cap Digital, Ceipes, Panodyssey, Pro Progressione, Venetian Cluster, Voxeurop, Worldcrunch.

Consortium members:

The consortium (CREA) gathers 9 international partners and 5 countries (Bulgaria, Spain, France, Hungary, Italy) in the media, culture, education, and technology sectors. It is created and supported by the European Commission as part of the Creative Europe Lab Innovation programme.

Members of the Creative Room European Alliance consortium:

- **AGENCIA EFE:** Spanish press agency, European leader in the audiovisual sector (1939)
- **BSC SME:** Bulgarian non-governmental organisation in the cultural sector (1996)
- **CAP DIGITAL:** European cluster for digital & ecological transformation (2006)
- **CEIPES:** Italian institution for education and development (2007)
- **PANODYSSEY:** French start-up in the digital content sector (2018)
- **PRO PROGRESSIONE:** Multidisciplinary Hungarian organization for the arts and culture (2008)
- **VENETIAN CLUSTER:** Italian cluster for culture and innovation, based in Venice (2014)
- **VOXEUROP:** Multilingual online media with a European reach (2014)
- **WORLDCRUNCH:** International English-language media (2011)

This consortium is open to new associated partners.

Goals:

Panodyssey is a creative and collaborative platform to meet professionals and non-professionals creators to share their fiction or nonfiction content. It gathers writers and readers around common values ethics and the digital world.

The platform is designed to let authors develop their creativity by exploring new narrative formats. Panodyssey also integrates native features that make it possible for authors to foster their community and manage their earnings based on their objectives - while keeping copyrights in mind and ensuring transactions are secured thanks to blockchain technology.

By 2030, Panodyssey aims to reach 10 million digital creators and 100 million online users and become a local alternative to the GAFAM.

Constantly evolving, the consortium builds new European economic partnerships to become a reference ecosystem in the audio-visual, cultural and creative sectors.

Project funding:

This project is fully funded for the duration of its activities for 2022 and 2023. The European Commission contributes 60% and the founding partners of CREA finance the remaining 40%.

Issue and ambition:

The objective of the CREA, backed by the entire sector, is to create a European alternative to American platforms. It is essential to bring out local operators to promote and economically enhance the content of creators in Europe.

New forms of hybrid creations and new creative formats are a major challenge for innovation in the cultural sectors.

Vision and values of the European collective is to offer professional solutions, in line with European standards, and facilitate cross-border cooperation in the sectors of cultural and creative industries.

To combat disinformation, Panodyssey has set up a secure system to guarantee the certification of the identity (ID Check) of content creators. Therefore, fake profiles, fake content and bots do not have access to the platform.

The lack of advertising is a complementary and effective way to prevent the economic mechanisms that finance disinformation. The action of CREA will continue this work of education, in particular with its activity "Media Education Laboratory", by addressing societal themes, such as disinformation, inclusion, and digital accessibility, through various initiatives, involving all cultural, audio-visual, and creative sectors.

These sectors are particularly sensitive to the protection and enhancement of intellectual property. To control the use of content, without the authorization of the rights holders, the solution is equipped with processes of awareness and management of intellectual property. Indeed, the certification of identity makes it possible to guarantee the authorship of the works and to fulfil the legal obligations in terms of financial transactions in complete transparency.

To industrialise the valuation and protection of intellectual property, the content platform includes blockchain technology, IP expertise and an IP Smart Contract.

Expected benefits:

- **Expected benefits for content creators**: Enhance the creation of European content through the protection of intellectual property and the enhancement of works of the mind. Foster new online experiences around creative and collaborative writing.
- **Benefits for users**: Allow citizens to have a tool that is easy to use and guarantees the information disseminated. Promote accessibility to quality content around literature and knowledge. Deploy a new vision around an algorithm chosen by the user.
- **Benefits on the regulatory side**: Anticipation of the implementation of new European standards such as the Digital Services Act.

Strengths:

- An ambitious project, bringing together quality European players and specialists in their sector.
- A team with technological and/or sectoral expertise
- A project supported and co-financed by the European Commission
- An ability to bring together stakeholders in the cultural and creative sectors across Europe.

5. TAMIS project: optimize data flows in Production chain

Scope: Optimised sharing of media files in production chain, using a decentralised architecture

On the initiative of: CST (France), which supports the project and its service providers

Participants: MediaADN, Lum::Invent (formerly Media-IO), Startin'blox with the support of ISAN-IA agency, The Procirep, Cap Digital (Paris Region competitiveness cluster), Titra

Context:

In a context of increasing dematerialised deliveries of audio-visual works, the number and variety of associated media files are multiplying, requiring increasing efforts in referencing and monitoring by the various players in the sector. There is no single standard for identifiers and metadata.

The validation of versions is done in an artisanal manner (email, SMS, etc.), generating misunderstandings and confusion in the ordering and delivery process. The registration of variants and

events is centralised in the hands of the right holder, making it more difficult for the other actors in the chain to declare new deliverables.

Goal:

The **TAMIS project** aims to promote the exchange of manufacturing metadata, beyond identifiers, in order to optimise exchanges between actors in the manufacturing chain.

The envisaged solution is based on the Solid architecture, which is based on a stack of existing Web technologies. This technology is promoted in particular by Tim Berners Lee to return to a truly decentralised web of data. The Solid technology stack is based on established web technologies, semantic modelling, such as RDF, the use of restful APIs, WebId authentication based on OpenID, to name a few.

For semantic modelling, the project envisages the use of EBU-CCDM, a generalisation of EBUCore proposed by the EBU [15].

The **TAMIS-1** project provides for an evaluation of the options considered and a fine-tuned modelling of the use cases.

The objective is to continue with a TAMIS-2 project, which would be a proof of concept (PoC) involving typical representatives of the sector's value chain, with a laboratory, a licensing entity, a distributor, and a rights holder.

The deliverables would be, in addition to detailed documentation, an SDK to facilitate the development of connectors, and an easy-to-implement "plug and play" brick with a graphical interface allowing low-tech actors to do manual input to enter the TAMIS loop.

Expected benefits:

- The effective circulation of identifiers is a systemic change. The massive adoption of this distributed solution by the audio-visual industry will optimise the exchange of these identifiers and the metadata that must be associated with them within the audio-visual production chain. Tamis will thus make it possible to:**Streamline the production process of a work:**
  – Facilitate collaboration between actors in the chain by reducing the effort required to adapt information systems.
  – Generate productivity gains through reliable automation of the various processes and by reducing input errors.
  – Optimise the organisation of the schedules of the various stages of production of a work
- **Reduce the "Time to Market"**
  - Contribute to the monitoring of the marketing and exploitation of audio-visual works produced.
  - Promote the circulation of information related to the remuneration of rights holders.
  - Highlight the skills of the market by improving the visibility of the players.
  - Contribute to digital sobriety

Difficulties/barriers identified:

- – Attract film producers to a technical project.
- – Mobilise investments from companies in the sector to finance a connector.
- – Aggregate the different controlled vocabularies used in the industry (EBUCore or EBU-CCDM [15], or the MDDF ontology proposed by MovieLabs, the Hollywood studios' laboratory).

Main Dataspace technologies at stake:

The architecture of the Gaia-X federation of services proposes identity and trust services, which are a priori resolved by the Solid architecture chosen, but a more in-depth study may reveal technological and architectural convergences.

On the other hand, the cataloguing service must be able to index and expose entry points to the service.

At this stage, this approach is rather envisaged technically as a Solid connector to the Gaia-X federated services.

6. MediaVerse project: decentralized network for distinct digital asset repositories

Field: Distributed content Search, Discovery & Access, and DRM

Link: https://mediaverse-project.eu/

Initiated by:

Research institutions and universities (CERTH, LINKS, UAB) join forces with ICT companies and SMEs (ATOS, ATC, FIN, VRAG) to provide the needed research innovations and technological bricks, while traditional industry stakeholders (STXT, DW) provide valuable input to shape new age industrial requirements.

Goal:

MediaVerse (MV) enables the deployment of a decentralised network for distinct digital asset repositories, where corresponding stakeholders and media owners can share and monetize multimedia content. With this, MV aspires to disrupt current practices of working in isolated silos, while enabling secure and traceable media content sharing. As soon as a digital asset repository is integrated in the MV network as a distinct MV node, the cross-network rights negotiation and content monetization will be enabled via the MV blockchain-enabled rights management solution.

Following the European data sharing vision, MV aspires to help content owners to share content within a network of other content owners, while keeping control over the content shared. For this, several AI-driven features are implemented towards making content accessible. Integrated content adaptation services ensure the distribution of the created media over several networks, platforms, and end-devices. The project is to be validated through large scale pilots in the field of citizen journalism and co-creation of immersive and inclusive media, also foreseeing a STARTS-inspired use case that involves artistic experiments and practices, pushing forward creativity for a critical approach to user-centric media [20].

Implementation/Architecture:

As part of the need for transparency, especially in rights and assets management, the aim of MediaVerse is to constitute a federated network of nodes - each node belonging to a different organisation. To connect to the network, an organisation can deploy their own node, or just join one as a member.

A deployed MV Node is the core element of the system. The process of installing a new MV node is easy, and the setup is agnostic of the underlying infrastructure. Beyond the general components of each node, a variety of connected services and tools make possible to connect to each node as needed to support individual requirements like automatic translation and subtitle generation, content moderation and verification and many others. Each MV Node is responsible for the following tasks:

- Digital Asset Management
- Media library: supports images, videos, 3D/VR models & scenes
- User management: each user manages his/her own media library
- Digital Rights Management: each user can apply licenses to his/her assets
- Group Management: user and access rights management in the context of larger organisations
- Distribution
- Adapt content and share it to other users
- Share to social media or use links to deliver the content
- Content Retrieval
- Search and filter content within your media library
- Discover content from other users across the MV networkAutomatic negotiations based on permissions

7. MILC: a marketplace for all Video Content

Scope: Global availability of all Video Content, Discovery process for Buyers, Complex licensing for the entire industry, and Early feedback from its audience

Link: https://project.milc.global/

Initiated by: Welt der Wunder TV, SRG

Goal:

The MILC Project aims at providing the industry with a new, open marketplace for every professional and non-professional content provider and buyer, through the Media Industry Licensing Content (MILC) platform and the use of the Media Licensing Token (MLT). MILC includes the offering, marketing, rights management, contracting and the final delivery of the content in any imaginable technical format. It will also build a bridge between the audience and the content creators, buyers, and distributors. This connection will allow the audience to directly impact and receive benefits from the creation & distribution of quality new content.

8. C2PA/Origin Project: authenticating media provenance

Scope: Fighting misinformation - certifying the source and history (or provenance) of media content

Link: https://c2pa.org/about/about/ and https://www.originproject.info/about

Initiated by: C2PA unifies the efforts of the Project Origin, initiated by **BBC**, NYT, **CBC/Radio Canada**, Microsoft, and that tackles disinformation in the digital news ecosystem, and Adobe-led Content Authenticity Initiative (CAI) which focuses on systems to provide context and history for digital media.

Participants: This specification has been, and continues to be, informed by scenarios, workflows and requirements gathered from industry experts and partner organizations, including the Project Origin Alliance and the Content Authenticity Initiative (CAI). It is also possible that regulatory bodies and governmental agencies could utilize this specification to establish standards for digital provenance.

Project: C2PA addresses the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content.

The goal is to establish an end-to-end process for the publishing, distribution, and presentation of provenance enhanced media. This tamper proof indication can be adapted for audio, video, images, and text-based digital media. C2PA has developed technical specifications, architecture, and a model for storing and accessing cryptographically verifiable information whose trustworthiness can be assessed based on a defined trust model. It also provides information about how to create and process a C2PA Manifest and its components, including the use of digital signature technology for enabling tamper-evidence as well as establishing trust. It is designed to enable global, opt-in, adoption of digital provenance techniques through the creation of a rich ecosystem of digital provenance enabled applications for a wide range of individuals and organizations while meeting appropriate security requirements. [10]

9. Western Balkan Archives project: platform for Archive description sharing

Scope: Search and Discovery of archives in the Balkans

Initiated by:

Service Contract of the European Commission that has been awarded to a Consortium led by the International Federation of Journalists (IFJ) together with the EBU, the European Federation of Journalists (EFJ), the Austrian Public Broadcaster (ORF), the Balkan Investigative Reporting Network (BIRN), and the Eurovision News Exchange for Southeast Europe (ERNO).

Participants:

The beneficiaries are the six public service media organizations in the Western Balkans: RTSH (Albania), BHRT (Bosnia-Herzegovina), RTCG (Montenegro), RTK (Kosovo), MKRTV (North Macedonia) and RTS (Serbia). Other stakeholders include the parliamentary committees responsible for media, the media regulatory bodies and the PSM governing bodies.

What:

The platform developed by EBU provides a view to Western Balkans PSMs archive material, giving access to information on available content, title, format, source, date, rights, and description, and should serve as a basis for the exchange of material on a bilateral or multilateral basis.

By connecting and sharing metadata and knowledge, this platform enables the development of joint actions and creations between PSMs. For its launch, the platform is initialized with the metadata of 884 hours of archive content, representing 1302 items. This metadata list was prepared by the Western Balkans PSMs in cooperation with ORF, the project's consortium partner. From a technological point of view the platform is a graph database exposed via a user interface that generates pre-coded queries. The platform provides an efficient way to share digitised content and explore the associated metadata.

In terms of hosting, the exchange platform is deployed on AWS but does not use the AWS database, user interface and user management services that form the core service of the platform. The metadata exchange platform can then be moved to other cloud providers or on-premises servers without development effort.

In terms of usability, the architecture has been designed to make interaction with the graph database easy for non-experts. For reading, the user interface generates pre-coded queries that are sent to the graph database to generate data. For writing, a formatted Excel file is used to generate queries that feed the graph database.

10. European Cultural Backbone: a federated platform for European civil society media and content producers

Scope: Federated platform for European Civil society media and content producers

Link: https://cba.fro.at/news/european-cultural-backbone-2-0

Initiated by: ECB is an initiative by Cultural Broadcasting Archive (A), freie-radios.net and the development collective arso.xyz (D).

Participants: Emerging from the non-profit media sector, the founding consortium is formed from a diverse set of technologists and content producers with a long history of being rooted in the civil societies of Europe. We're currently reaching out to partners in France, Spain, Ireland, Hungary, and Poland.

Goals:

We seek to provide a technological basis for the long-standing desire in using the Internet as an open space with a shared commitment to open and collaborative exchange of knowledge and art. Big Tech and commercial platforms cannot properly fulfil these needs, because often, their business model is at odds with the equally important values of data sovereignty, privacy, and self-determined governance. ECB will create self-governed public open sphere for the European civil society and thereby provide a much-needed alternative to existing platforms controlled by for-profit entities.

Whereas commercial providers invest billions into their digital strategies and technology, technologists in independent outlets have long struggled with keeping up to the pace of progress in media technologies. It is time for a resilient, self-governed infrastructure for the media-technological future of the European civil society.

ECB is an initiative to build a federated platform for European civil society media and content producers. Started by technologists, archivists, and community media outlets, we want to connect communities across Europe by developing a sustainable and self-governed alternative to centralized commercial services for publication and preservation of non-profit audio-visual content.

As a decentralised network of interconnected platforms, ECB will be resilient towards single points of failure.

ECB integrates novel approaches to recommendation, search, and moderation, and ongoing research into fair artificial intelligence algorithms.

Through intelligent, cross-network recommendation and public indexes, individual communities are empowered to expand their reach through shared nodes and databases. Through secure, decentralised software tools we want to allow the existing trust networks in the civil society to expand into the digital realm.

ECB will provide recommendation and moderation tools that shall be controlled by self-determined governance structures and not the maximisation of advertisement value of a platform or the sale value of user data. Moderation can be delegated and discussed in a bottom-up process, empowering communities, and networks to provide a self-governed space of publication.

By encoding existing networks into distributed platforms of shared nodes, we create an efficient and secure way to share content between collaborators and like-minded communities.

Project plan:

Development of an open-source toolset that creates an interconnected network of ECB nodes, which import and link existing European non-profit media platforms. Each node is a repository of metadata for public-reachable media files. Nodes are designed to exchange data with each other through a secure exchange network, building on trust networks between nodes. Each node exposes an API to integrate with existing infrastructure.

- Development of an intelligent recommendation and categorization framework to be employed by ECB nodes. Nodes provide a search index for any number of connected nodes. Nodes may use Open Audio Search to make audio-visual content available for full-text search and automatic categorisation.
- ECB plans to build up a trust-based moderation system that secures free speech while trying to prevent hate speech. We are thinking about using social trust networks to delegate moderation responsibility between groups and connected nodes.
- Integration of ECB to existing content platforms which are managed by the initial partner organisation in the consortium. Setup of a cross-outlet recommendation system. Support for adopting partners. Integration with in-house tools and websites.
- Collaboration on business models / monetisation for civil society media creators and technologists

11. EDMO: European platform against Disinformation

Scope: The European Digital Media Observatory (EDMO) is advancing the development of EU fact-checking services. It is a multi-disciplinary independent response to the phenomenon of disinformation in Europe.

Link: https://edmo.eu/fact-checking/

Initiated by: Under the leadership of the European University Institute in Florence (Italy), which relies on the expertise of its School of Translational Governance and Centre for Media Pluralism and Media Freedom, EDMO is a partnership that also includes Datalab at Aarhus University, Athens Technology Center, which provides the technological support and is also coordinating the Social Observatory for Disinformation and Social Media Analysis (SOMA); and Pagella Politica.

Participants: The European Digital Media Observatory (EDMO) brings together fact-checkers, media literacy experts, and academic researchers to understand and analyse disinformation, in collaboration with media organisations, online platforms and media literacy practitioners.

Goals:

The European Digital Media Observatory (EDMO) addresses the phenomenon of disinformation in Europe and contribute to a deeper understanding of disinformation relevant actors, vectors, tools, methods, dissemination dynamics, targets, and impact on society.

The EDMO secure online platform is built on the existing collaborative verification platform **Truly Media** (www.trulymedia.com) and Twitter analytics tool **TruthNest** (https://www.truthnest.com/) and facilitates the **verification of online multimedia content** (text, audio and visual material), including support for investigating video content (such as deepfakes).

The system includes **tools for the exchange of information**, and allows for **joint fact-checking activities**, peer-review of fact-checks, and real-time collaboration among team members and/or across different teams. Such activities are facilitated by the use of integrated real-time communication

capabilities (directory of contacts, individual and group chats, activity logs, etc.) and translation tools (Google Translate is currently used for this purpose).

The EDMO platform has already integrated **access to data repositories of the European Union and international institutions** (such as Eurostat, EU Open Data Portal, European Parliament repositories, OECD repositories). Access to Member States databases (e.g., national statistical offices) will be added. Facebook has already allowed Truly Media access the API of **CrowdTangle** (providing access to Facebook, Reddit, and Instagram public content) which has already been integrated.

State-of the art tools and services for fact-checking activities are integrated in the platform that are able to facilitate the identification of relevant and trending news content published and shared on the Internet and on major social media platforms. In particular, the **European Media Monitoring** (EMM) trends are integrated and easily shown to users, as well as access to **CrowdTangle**, Facebook's internal monitoring and analytics tools for Facebook, Instagram, and Reddit, through the CrowdTangle API.

A **repository of EU based fact-checking organisations** and their professional characteristics has been created and is regularly updated.

A **searchable directory of online fact-checks** is available in several EU languages through automated translation. Monthly fact-checking briefs picturing the main disinformation trends are also published.

The EDMO platform is supporting and facilitating the coordination of academic research activities on disinformation in Europe, creating a searchable repository with relevant peer-reviewed scientific literature, and publishing cooperative investigations focused on specific disinformation topics.

EDMO also provides support to public authorities in the **monitoring of the policies put in place by online platforms** to limit the spread and the impact of disinformation.

EDMO governance body ensures public trust regarding the work of the platform and establishing a framework to provide secure access to data of online platforms for research purposes. Under the guidance of an independent Advisory Board and the Executive Board, which includes the five heads responsible for each of the core tasks listed below, the consortium brings together stakeholders in the field of disinformation, while **maintaining the independence and impartiality of the partners** in the execution of the project.

All members of the consortium have a strong experience in analysing the online disinformation phenomenon and are used to working with policymakers, public and private actors (including media organisations, journalists, media literacy experts, and online platforms) with a multi-disciplinary applied approach.

12. Alliance Culture Data: a data exchange platform for the Cultural and Creative Industries

Scope: a data exchange platform for the Cultural and Creative Industries

Contact: Veronique Demilly (chair GT CCI , French Gaia-X hub)

On the initiative of:
Initially initiated and formalised by BnF-Partenariats in Alliance Culture Data, this project now needs one or more players capable of ensuring its support and development.

Targeted project ecosystem:

The CCI sector.

Alliance Culture Data project:
ACD is defined as a platform allowing its actors to provide or acquire, free of charge or against payment, any type of data. The platform constitutes itself as a "data intermediary" within the meaning of the Data Governance Regulation.
The technical solution of the platform would be provided by the the French company Dawex, whose technology is already used by several industries.

Specific services will be provided:
• **Upstream** (services independent of data exchanges):
  – Structuring, quality control and cleaning of data
  – Training in data governance
  – Estimation of the value of the data
  – Packaging of the data to meet specific use cases.
• **Downstream** (services following the exchange of data with buyers and offered by specialised companies via the platform):
  – o Analysis of quantitative and qualitative data
  – o Provision of data skills, data marketing and data strategy services
  – o Packaging of data to meet specific needs.

Strong positive externalities for the sector:
The positive externalities linked to the exchange and sharing of data made possible by the platform are numerous, both for the sector and the players in it, but also for sectors outside the CCI, thus making it possible to multiply the links and exchanges between the CCIs and other industries. More than thirty possible use cases and applications facilitated by ACD were identified in interviews conducted by BnF-Partenariats with CCIs. ACD also appears to be a lobbying tool that can unite CCIs and influence strategy and regulation on data exchange with the web giants.

The economic model:
An economic model built to allow Alliance Culture Data to be self-financing beyond its initial Financing. It is based on two sources of revenue:
  • Sales of subscriptions to the platform plus onboarding services
  • Sales of ancillary services offered, upstream and downstream of data exchange on the Platform

Project status:
The project requires an estimated €6M in seed funding.
The project is intended to be financially autonomous from public support at the end of the start-up phase. The deployment trajectory planned to date allows ACD to be self-financing after 5 years following its launch and to generate a positive return on investment after 8 years

13. IWA project: a repository for digital twins

Scope: IWA= interoperability With the Ark. Repository of Digital twins

contact: Jérôme Battistelli from Digital twin agency Monark (Mnrk.io)

Initiated by: Monark, agency for the creation of digital twins of architectural heritage, sensitive natural sites, and remarkable objects.

Potential participants: Garou, multi-user XR platform for designers and companies, metaverse.

And Project partners:
- Software publisher, specialising in 3D audio capture and playback
- Provider of scalable, interoperable, sovereign, and integrated Blockchain solutions
- 3D animation production company
- Research institutes (possibly)

Context:

The digitalisation of real estate, furniture, and remarkable objects for the purposes of conservation, transmission, enhancement, and communication makes sense in our contemporary digital society. The creation of digital twins of our architectural or pictorial heritage, of certain typical places, held by the owners of the works, could constitute a new source of income for the owners of the goods and of inspiration for the Cultural and Creative Industries sector.

To date, there is no professional platform to enable the CCI sector to fully exploit the resources linked to digital twins. Legal uncertainties, lack of sharing, interoperability, dialogue, or lack of visibility often lead to the re-digitalization of works.

The major hyperscalers are investing massively in new distribution media such as virtual reality (e.g., Facebook with Meta), augmented reality or video mapping (architectural projection).

Goals:

To avoid the capture of value linked to digital twins by these actors, the IWA project aims at the emergence of a French or European actor who would organise the collection, storage, description, visibility, and marketing of digital twins' data according to precise and transparent specifications.

The project also aims to raise the awareness of CCI players on the importance of transposing their material assets into the virtual world and the benefits they could gain from it.

The IWA project proposes to:
- Set up a platform for remarkable digital twins, so that they can be used as virtual sets by film, gaming, and contemporary art professionals
- Create digital twins.
- Secure the use of these twins through fingerprinting and NFTs in a blockchain.
- Commercialise (license) the use of digital twins by paying the owner of the physical asset.

Through a platform using WebGL (GPU) technology, all actors in the cultural and creative industries will be able to consult the available digital twins.

Expected benefits:

Consolidate the intangible heritage of the cultural and creative industries and enhance it internationally.

Monark allows institutional and private investors to invest in a digital twin and to monetise it through scenarios open to numerous opportunities. IWA is an alternative to the GAFAs to keep control of the content and retain its value in Europe.

Difficulties/barriers identified:

There are several factors that make the project more complex to implement:
- The lack of standardisation of digital twin formats
- Owners and institutions are unfamiliar with this new market
- The absence of a common platform for the CCI (Cultural and Creative Industries) sector.
- The hegemony of the GAFA (Google, Apple, Facebook, Amazon...).

Main Data space technologies at stake:

- **Data platform:** sharing, exchange of data between several players in the ICC sector
- **Security and confidentiality:** authentication, identity federation management, data management, data security, consent management and usage control.
- **Compliance:** data portability, rights management, and certification.

14. Media-cloud.ai: A platform for Multimedia data processing services

Scope: Multimedia data processing services

Link: https://media-cloud.ai

Initiated by: France Télévisions, INA

Potential participants: lum::invent (ex-Media-IO), Audiovisual Public Services

Goals:

Based on an open-source project initiated in 2018, the objective is to offer multimedia data processing services (images, moving images, sound, subtitles, interactive data, etc.) that can be used on an existing open-source platform instance made available as a Cloud service, for users who do not have the software engineering skills required to implement it. Beyond the classical processing of audiovisual files and streams, this platform allows the orchestration of processing based on Artificial Intelligence from explicable models. This platform could be developed jointly by various players in the sector, particularly public players in France and Europe.

15. Invenio project: Artificial intelligence and self-regulation of advertising content

Initiated by:
ARPP (Autorité de Régulation Professionnelle de la Publicité) is the French advertising self- regulatory body. Founded in 1935, the organisation now has nearly 700 paying members, brands, agencies, media, platforms, advertising technology solution providers (AdTech), etc., who agree to adopt and comply with ethical rules, and entrust ARPP with the task of ensuring that these rules are properly applied.

Participants:
ARPP and its European counterparts federated by EASA (European Advertising Standards Alliance) exchange information within the framework of a "Data driven self-regulation Committee" working group. The tech partners of the European organisations are invited to attend and source codes are shared. The aim is to make these tools available on a European scale for common infringements.

Project ecosystem:

- Digital advertising agencies
- Website publishers
- Inventory monetisation platforms
- Platforms/Social Networks

Context:

One of the main challenges facing the ARPP today is the multiplication of advertising channels and creations. Moreover, programmatic advertising and the automation of campaigns, particularly in RTB (real-time bidding system) have made it complex to identify and manually filter advertising practices that are not compliant with legal or ethical rules.

Faced with the expectations of civil society and public authorities, professionals must demonstrate their ability to implement an efficient self-regulatory system based on modern technologies. It should be able to adapt to market changes, representative of all players in the advertising industry, able to control advertising all along the value chain. Goal is to maintain high ethical standards in a market that is complex and undergoing rapid digital change, and continue to work towards legal, ethical, fair, and respectful advertising for all audiences, in closer consultation with civil society.

Goals:

As advertising is a combination of images, text and/or sound, technologies based on machine learning such as automatic natural language processing, computer vision, convolutional neural networks, etc., can help detect basic elements in advertising content, such as alcohol (cf. the "Evin" law), size of characters (legibility of legal notices), men and women (contemporary representation of genders, ages, diversity in advertising), specific textual or audio references (health claims, environmental claims, etc.), etc.

The goal of such a **"Compliance as a Service"** solution is to provide digital advertising market players with a catalogue of detection services to help ensure the legal and ethical compliance of advertising content.

Expected benefits:

The provision of automatic detection models for suspected non-compliance will help:

- Secure the responsibility of the players in the value chain and enable them to demonstrate their ethical approach to all stakeholders (consumers, public authorities, brands, etc.).
- Implement a healthy and trusted digital environment, allowing the market to develop over the long term.

Difficulties/barriers identified:

The accuracy of the models depends on the availability of large-enough sets of correctly annotated data, corresponding to each "non-compliance" use cases (e.g., misleading environmental claims, vehicle on a natural area, excessive food consumption, etc.).

16. French ministry of culture project: Aggregation of ticketing data

Initiated by: Ministry of Culture. Arenametrix

Participants: Any Producers, Broadcasters and Ticketing intermediaries

Context:

Ticketing is a strategic element in the cultural field, essential for the business models of operators, but also for public knowledge, communication, and loyalty.

The dematerialisation of ticketing systems raises multiple issues and the sector is currently lacking a player capable of bringing together and making visible the offers made by cultural structures with

multiple statuses, operating in very diverse fields of activity and generating a very high number of transactions (more than 50 million transactions for the main French museums in 2019 alone, to which should be added theatres, monuments and other cultural venues with a dedicated ticketing system).

In addition, the emergence of new operators specialising in the electronic sale and distribution of tickets and the development of online purchasing and service platforms have significantly modified the competitive landscape (verticalization of certain groups, concentration and emergence of global players, new intermediaries from online communication or payment, etc.). There is a large range of maturity in terms of dematerialisation, the smallest players are currently the least mature.

The evolution of uses in favour of the large ticket distribution or accommodation and tourism platforms poses a significant risk to the visibility of the cultural diversity of the territories. For a large majority of producers (CCI, tourism, leisure, etc.), ticket distribution, accommodation and tourism platforms have become essential for customer acquisition and economic development. This model naturally leads to a fragmentation of the value of the offers and has harmful consequences for producers and users:

- The battle for direct access to consumers and their data
- Large mark-ups on ticket prices
- Lack of visibility for small players without solutions.

Goal:

The pooling and aggregation of data from the various ticketing systems would enable the entire sector to better understand the behaviour and aspirations of the public and thus be able to adapt the programming of cultural events and promote encounters between artists and the public. This data could also be shared with other sectors (tourism, transport, etc.) to adapt their own offers.

The data to be reconciled are:
- Sales data
- Sales locations
- Available customer information.

Expected benefits:

- Better knowledge of audiences
- Adaptation of cultural programming
- Visibility of the offers of small structure.
- Possibility to create combined offers with tourist offers
- Traceability of ticketing and fight against fraud and the black market.

FURTHER WORKS

This position paper is far from being an exhaustive description of all the use cases and projects relevant to CCI data space. Based on first feedbacks gathered from a limited number of CCI organisations, it is not a detailed and comprehensive study on what should be a European Media Dataspace. Many topics, such as governance, business models, require to be investigated in much more details. Members of the sector are invited to amend and complete it and to inform the working group's sponsors of any interest in any of the initiatives presented, either as a contributor or as a project leader.

Contact details:

Lucille Verbaere
Cloud and Cybersecurity Senior Project Manager
email: verbaere@ebu.ch
Mobile phone: +41 79 376 49 62
Technology & Innovation department
European Broadcast Union
L'Ancienne-Route 17A | 1218 Le Grand-Saconnex | Geneva, Switzerland
Mobile phone: +41 (0) 79 376 49 62

REFERENCES

[0]: Media Vertical Position paper version 1 (CCI group from French Gaia-X hub):

[1] Gaia-X Architecture documents: https://docs.gaia-x.eu/technical-committee/architecture-document/22.04/ecosystem/#goals-of-federation-services

[2] IDSA Reference Architecture- Avril 2019

[3] Mediaverse D4.1: Copyright and Procedure for IPR Definition

[4] Building a European Digital Public Space (Alexander Baratsits)

[5] Design Principles for Data Spaces (OpenDEI)

[6] Summary of results from workshops about the governance structure of a media data space (Carsa, July 2022)

[7] EBU Workshops (May 2022): Defining a Media Data Space

[8] *Etat des lieux sur les métadonnées relatives aux contenus culturels* » de l'Observatoire de la culture et des communications du Québec (2017).

[9] https://www.twobirds.com/en/insights/2022/global/the-eu-data-governance-act-what-privacy-professionals-need-to-know)

[10] https://c2pa.org/specifications/specifications/1.0/specs/C2PA_Specification.html#_overview

[11] Dataprot.net. "*Piracy is back: piracy statistics for 2022*" (May 2022)

[12] EU Intellectual Property Office, "Online Copyright infringement in the European Union, Music, Films and TV (2017-2020), trends and drivers" (Dec 2021)

[13] https://www.w3.org/RDF/

[14] http://xmlns.com/foaf/spec/

[15] https://www.ebu.ch/metadata/ontologies/ebucore

[16] https://peach.ebu.io/technical/collect_format/

[17] https://datasetsearch.research.google.com/

[18] https://faky.be/fr

[19] www.bellingcat.org

[20] https://starts.eu/about/

[21] https://www.vlaanderen.be/cjm/nl/nieuws/vlaamse-mediasector-bundelt-krachten-voor-nieuw-crossmediaal-meetsysteem

[22] https://www.snowflake.com/blog/world-data-collaboration-snowflake-summit/

[23] https://www.w3.org/wiki/WebAccessControl

[24] https://fidoalliance.org/