Enhancing and connecting performing arts information systems

Project Report

![CAPACOA logo]

CANADIAN ASSOCIATION FOR THE PERFORMING ARTS
ASSOCIATION CANADIENNE DES ORGANISMES ARTISTIQUES

Canada

Cover photo: The show *Glitch* is listed in Scène Pro's show directory, and its production company, Bouge de là, has been assigned Artsdata identifier K10-235. The show toured in several cities in Québec, as well as in British Columbia and Ontario. Photo credit: Suzane O'Neill.

Table of Contents

Definitions

Entity: Group of identifiable things that share the same attributes. These things can be tangible objects (e.g. books), people (e.g. artists), venues (e.g. concert halls) or intangible things (e.g. shows, musical genres).
Note: In RDF ontologies, the term "entity" refers to a class, an attribute or an instance.

Synonyms:

- class (term used in RDF ontologies)
- type (term used in Schema.org ontology)

Instance: Occurrence of an entity in a dataset. For example, the Salle Louis-Fréchette, at the Grand Théâtre de Québec, is an occurrence of the "salle de spectacle" entity in Scène Pro. Its Scène Pro ID is [272](#).

Synonyms:

- individual
- object
- named entity

Attribute: Characteristic or parameter of an entity and, by extension, of the individuals that are instances of the entity. For example, capacity is an attribute of the performance hall entity, and its expected value is a number. An attribute can also describe a relationship between two instances. For example, the instance Salle Louis-Fréchette is associated with the instance Grand Théâtre de Québec by the attribute containedInPlace in Artsdata. The association of an attribute with a particular value forms an "attribute-value pair."

Synonyms:

- property (term used in RDF ontologies)
- descriptive metadata

Cardinality: Number of occurrences of an entity that can be linked to another entity. For example, a performance can only be linked to a single venue. This cardinality is written as (1,1). Conversely, a venue can be linked to none or several performances, and this cardinality is written as (0,m).

The following is a possible illustration of this example:
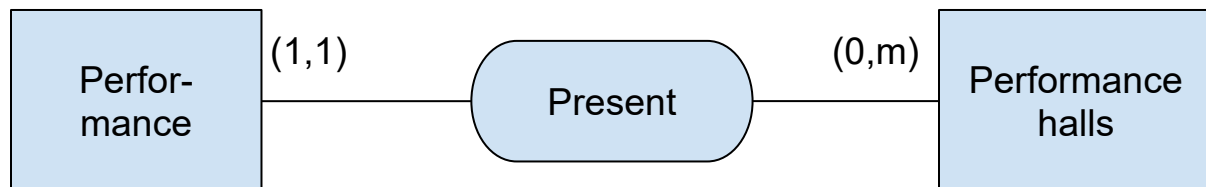
This representation constitutes a **conceptual data model**. The items in the rectangles are entities, the item in the rounded rectangle is the association linking the two entities, and the values in brackets are the cardinalities.

Ontology: Formal representation of a set of concepts and relationships within a specific domain. It is used to structure and organize knowledge to facilitate its interpretation and sharing.

Ontology alignment: Process of determining matches between concepts in different ontologies to enable interoperability and integration of data.

Synonyms:

- ontology matching
- mapping

Image 1: This is the artist profile image for Raphël Butler in Scènes francophones. This image was also reused for each of Raphaël's 14 performances presented by RADARTS members in February and March 2024. Photo credit: courtesy of the artist.

Context

CAPACOA

The Canadian Association for the Performing Arts/Association canadienne des organismes artistiques (CAPACOA) is a national arts service organization that supports performing arts touring and presenting. For over a decade, CAPACOA has been working to support the digital transformation of the performing arts sector, most notably through initiatives such as *A Linked Digital Future* and Artsdata.

Artsdata

Artsdata is a Canada-wide knowledge graph for the arts. Since 2018, Artsdata has been facilitating the sharing and reuse of open data between websites, databases and platforms in the performing arts sector.

The Project: Enhancing and Connecting Performing Arts Information Systems

Project Presentation

In connection with the Artsdata initiative, CAPACOA and its project partners analyzed the data structure of key information systems in the performing arts sector with a view to enable the efficient flow of information between these systems. This is a digital transformation project.

Project Objectives

The main objective is to map key performing arts information systems to document disparities in ontologies and recommend improvements leading to data openness, ontology alignment and data reuse. Performing arts data across Canada can be reused when open protocols are adopted and ontologies are interoperable (i.e. when information systems can understand each other).

Issues Related to Technological Projects

This section deals with the general issues surrounding technology projects and digital transformation initiatives that this type of project invariably faces. These challenges are social and technical in nature, and arise on management, organizational and technological levels. Although their intensity varies within and between organizations, these issues had a bearing on the operations of the information systems that the project analyzed.

Management Issues

Management issues involve planning, organizing, directing and controlling elements included in a project management process. When developing a technological project, we ask ourselves questions about the organization's context or external environment. Is the project's business model viable in the current environment? What needs to be done? What are the risks? Does the project fit in with the organization's current leadership and management style? Is the project part of a clear strategy? How will the project fit into IT governance? What organizational and/or technological need(s) does the project address? What objectives does it help achieve?

Within the context of the project, these challenges appeared in the form of the partners' management strategies, which both influenced the project and evolved with it.

Organizational Issues

Organizational issues are revealed when the internal environment of the organization and the current knowledge and skills of the staff are examined. Does the organization have the knowledge and skills required for the project? Is there an ongoing training program? The project's effects on staff organization and work methods, on business processes and on roles and responsibilities are also considered. How will staff react to these effects? How are information and knowledge managed within the organization? Is there an organizational change management plan?

Technological Issues

Technological issues concern hardware and software infrastructure, maintenance and upgrading procedures, as well as administrative measures (including policies) governing information quality and data retention. What is the state of the organization's hardware and software infrastructure? How recent is it? How is it integrated? Is it compatible? Is it scalable? Is it interoperable? How often does it have to be maintained and upgraded? Finally, considering the nature of the data and personal information, is there any potential for an invasion of privacy?

Project Partners

Given the project's limited scope in terms of time and resources, CAPACOA focused on four digital platforms that centralize and facilitate business processes in the Canadian performing arts industry. The fact that CAPACOA was already associated with the organizations running these platforms factored into the project's success.

The project's selection is by no means an exhaustive inventory of platforms in Canada. There are many other platforms inside and outside Canada's live performance industry. In 2019, CAPACOA listed more than 40 different information systems related to 11 use cases.[1] Many of these are no longer in business: a fact which raises concerns about the long-term viability of the selected platforms.

Scène Pro

https://scenepro.ca/

Scène Pro is a platform designed to facilitate the coordination of shows for Québec's performing arts industry. Scène Pro generates a unified database using the data from the large number of shows producers have registered with the platform (over 2,000 registrations per year for RIDEAU, ROSEQ and Réseau Scènes). Scène Pro aims to be a source of structured data for the entire performing arts sector.[2] Scène Pro provides an API to enable third-party applications to access its data.

Référentiel des métadonnées descriptives du spectacle

The *Référentiel des métadonnées descriptives du spectacle* is a data model in Québec designed to be "a tool to facilitate the development and interoperability of information systems associated with live shows"[3] (this repository is referred to as the "Référentiel du Québec (RQ)" elsewhere in this report). The data structure of the *Référentiel* is aligned with that of Scène Pro. As the *Référentiel* model is well documented, it served as a benchmark reference for this project.

Scènes francophones

https://www.scenesfrancophones.ca/

Governed by *Alliance Scènes francophones*,[4] the Scènes francophones platform aims to be a "digital one-stop shop that facilitates encounters and exchanges to increase awareness of French language performing arts in Canada."[5] In addition to increasing the discoverability of the

---

[1] Estermann, B. and Julien, F. (2019). *A Linked Digital Future of the performing arts: Leveraging Synergies along the Value Chain* (pages 56, 59). CAPACOA. https://bac-lac.on.worldcat.org/oclc/1138886705

[2] Scène Pro. (2024). À propos. https://scenepro.ca/a-propos

[3] Référentiel des métadonnées descriptives du spectacle. (2024). Bienvenue. https://a10s-ca.github.io/referentiel/

[4] The Alliance Scènes francophones is made up of three networks supporting the presentation of performing arts in the Canadian *francophonie*: RADARTS, Réseau Ontario and Réseau des grands espaces.

[5] Scènes francophones (2024). À propos : scenesfrancophones.ca. https://www.scenesfrancophones.ca/%C3%A0-propos/scenesfrancophonesca

performing arts, this platform was set up to address various issues and needs, such as having to enter the same information in various systems.[6] Scènes francophones want to serve all performing arts industry stakeholders: networks, presenters, artists, artist agents/representatives and the general public.

I Want to Showcase

https://iwanttoshowcase.ca/

The I Want to Showcase platform provides artists with a simple way to apply to multiple performing arts showcases[7] across North America in one online location. This platform is used by showcase event organizers, and artists and their representatives. I Want to Showcase is a collaboration between 10 partner conferences in Canada and the United States.[8]

Arts Touring Connector

https://arts-tc.ca/

Arts Touring Connector (ATC) is a platform designed to facilitate the planning of show tours. This process, known as "block programming", is carried out by presenters and the networks that support them. It aims to make touring more efficient and increase its impact for artists, presenters and communities. Artists and shows are added to the platform when requested by a presenter. In addition, a tour proposal is posted when a minimum of three presenters have expressed interest in a show. At present, four presenter networks use ATC, and two more are in the process of being integrated into it. Each network approaches tour coordination differently, which means that the way they use the platform can vary, notably in the amount of information they disclose to agents, managers and artists.

---

[6] Scènes francophones. (2024). À propos. https://www.scenesfrancophones.ca/%C3%A0-propos/scenesfrancophones
[7] A showcase is an opportunity for an artist to present an excerpt (or a complete show) to an audience of presenters, who may then purchase the rights to present one or more performances of the show.
[8] I Want To Showcase. (2024). Our Partner Conferences. https://iwanttoshowcase.ca/partner-conferences.php

Overview of project flow

The following section describes the methodology of the project and concludes with a brief review of challenges encountered during the project.

Methodology

1. Conceptual Modelling of Information Systems

The first major step of a digital transformation analysis is creating a high-level model of the data structure of each information system included in the project. The team therefore undertook to interview the teams in charge of each system twice.

The initial interview sought information about the origin of the platform, the context that led to the original idea and the context in which the platform now finds itself. The analysis was looking for answers to the following questions:

- Where did the idea come from?
- What needs were behind the design of the platform?
- What were the initial intentions at inception?
- What is the current state of the platform vis-à-vis initial intentions?

The second interview was with the people in charge of the technical design of the platforms. Its purpose was to gain an understanding of the technological and technical assumptions that led to the development of the platforms.

The project's analysts then set about modelling each platform, creating for each one an entity-attribute dictionary, and where possible, a dependency graph and a conceptual data model. This was done to highlight the types of entities (also called "classes") used by the information system, their attributes and possible values, as well as the relationships (including cardinalities) between these entities. The documents generated by the modelling were then used to analyze the ontological alignment of the different platforms.

2. Analysis of Ontological Alignments

Ontologies are aligned to minimize heterogeneity issues, which arise when data is exchanged between data sources. The project examined the syntactic, terminological, conceptual and semiotic aspects of data heterogeneity.

On a syntactic and terminological level, heterogeneity appears in the assignment of names to the entities (classes, properties, relations, etc.) that make up an ontology. On this level, the analysis consists of the following activities:

- Managing synonyms (when different words are used to refer to the same entity).

- Managing polysemy (when a name can refer to several different entities).

- Managing concepts originating in different languages (French, English) that refer to the same entity.

On a conceptual level, discrepancies observed in data sources are generally based on the following:

- Granularity: Some data sources describe the same concepts at different levels of detail.

- Coverage: Differences between ontologies can arise when they describe different or overlapping parts of the domain.

- Perspective: Even if two ontologies describe the exact same domain, they may have different points of views.

On a semiotic level, heterogeneity arises when different people or communities interpret the same ontology differently.

To resolve disparities between data sources, the project analyzed the existing sources before establishing matches (based on a relation of equivalence, of subsumption, etc.) between entities such as classes and properties, in order to reconcile the semantic viewpoints and opinions of the various experts interviewed.

After analyzing the structure, nomenclature and granularity of the data, two types of matches can be identified. One the one hand, simple matches of the (1,1) type, where each entity in one ontology is associated with a single entity in the other ontology. On the other hand, complex matches, which can be either of the (1,m) type, where an entity from one ontology is associated with several entities from the other ontology, or of the (n,m) type, where multiple entities from one ontology are associated with multiple entities from the other ontology.

The project relied on terminological (semantic and syntactic) alignment methods and on an analysis of the internal and external structure of ontologies to determine the entities, their relationships and the shared properties. Independent ontologies stemming from these different data sources were set up. This was followed by a merging of these sources, with each retaining its own context (IRI). Finally, we aligned the existing entities.

3. Implementation

After the project modelled and analyzed the alignment of each platform's ontology, a private report with observations and recommendations was submitted to the managing organization. This was followed by an offer to support platform managers who had undertaken the task of opening their data.

In the case of Scènes francophones, the project helped the managers of the platform implement most of the recommendations that appear at the end of this report. This was followed by the extraction of all of the open Scènes francophones data, which was then loaded into Artsdata. This data was subsequently reused by the New Brunswick government in its new Arts Culture NB platform (under development).

In the case of Scène Pro, the project extracted and loaded some Scène Pro data – on presenters and performance halls – into Artsdata.

In the case of I Want To Showcase, the platform's partners accepted the recommendations and initiated a process to open up their data. Their process was just getting underway when this project ended.

Project Challenges

Project Scope

The challenge of analyzing four platforms in one year, one after the other, was an ambitious one from the outset. Analyzing the first platform took longer than expected, but enabled us to identify methods that could be repeated afterwards. It also served as a kind of pilot project. Part of the analysis had to be carried out simultaneously on several platforms as dictated by the schedules of the partners and it wasn't always easy to find common timelines.

Partner Engagement

Some unplanned steps were added for communication with partners and their subcontractors, resulting in delays in obtaining data. The different governance structures of the various platforms and the outsourcing of the technical management of the platform to third parties were sometimes an obstacle to data collection. Except for governance elements required to open data, data governance (and platform governance) were not included in the scope of the project at the outset.

Technical Difficulties in Obtaining Data

In addition to the challenges of finding interested partners, technical difficulties were encountered in obtaining data. Some platforms did not have a relational database management system (RDBMS): software which greatly facilitates data structure modelling. Nevertheless, platforms based on a content management system (CMS) without a relational database were modelled with available information. Finally, intellectual property issues prevented the analysis of one of the platforms.

Highlights

Publishing Open Data is a Multi-Step Process

The value of open data for the performing arts sector lies in the many individual and collective benefits that can be derived from the reuse of this data,[9] notably on the efficiency front. For this project, opening data contained in the platforms will benefit the discoverability of artists and their shows, which is one of the objectives of the Scènes francophones platform.

For data to be "open" and give rise to these benefits, data must meet certain conditions.

The data must be made available under a license that allows anyone to copy, use and republish it, including the possibility of combining it with other open datasets, for any purpose, including commercial purposes.[10] This is what the Creative Commons CC0 license implies.

Data must also be easy to find. It must be freely accessible as a whole in a useful format (machine-readable and machine-processable). It must also be consistent with interoperability standards, such as those advocated by FAIR principles.[11]

Before they can open their data, organizations need to prepare the data to ensure its quality. Data quality can be assessed according to several criteria: accuracy/reliability, completeness, timeliness, consistency/comparability and comprehensibility.[12]

The quality of the data is improved through a management process that includes cleaning (identifying and removing duplicates, removing inaccurate or incomplete fields, etc.), standardization and, sometimes, anonymization (removing any personal information). This can be achieved by assigning responsibility for dataset quality to one or more employees. Scène Pro has taken this approach and has hired a data quality technician.

The process of opening data does not end with its publication. After having published open data, the organization must ensure that the data remains accessible over time. Practically, this means that the organization needs to secure a stable budget providing for sufficient resources to constantly update the platform and the data itself. In the case of I Want To Showcase, each partner makes an annual contribution to maintain the platform.

---

[9] Wylie, M., Sculthorp, M., Gagnon-Turcotte, S. and Chatwin, M. (2021). *A promising path to developing data governance in the performing arts sector: exploring charters and principles for data governance*. Open North and Canadian Association for the Performing Arts (CAPACOA), 24-26. https://bac-lac.on.worldcat.org/oclc/1264796109

[10] Open Knowledge Foundation. (n. d.). *What is Open Data? Open Data Handbook*. https://opendatahandbook.org/guide/en/what-is-open-data/

[11] FAIR principles summarize in four universal points the technical and commercial properties to be considered when generating and/or making data Findable, Accessible, Interoperable and Reusable, both for humans and for machines and algorithms.
Go FAIR. (n. d.). Fair Principles. https://www.go-fair.org/fair-principles/

[12] dia-log project. (2023). Cahier des recommandations. https://dia-log.ca/r%C3%A9sultats-et-bilan

Open Data Consent Considerations

Obviously, before opening any data, it is important to ensure that no personal information or sensitive data will be disclosed without the informed consent of the individuals involved.

Some data may be considered more or less sensitive, depending on the context. In the platforms analyzed, some data that might otherwise be considered personal information takes on a public dimension insofar as it describes artists in the context of their professional practice. For example, when an artist uses their first and last name as a stage name, this becomes public information. In the same way, an artist's region of origin can also be opened up in order to assert a sense of belonging to a territory. Such data is considered less sensitive than, for example, a self-declaration of belonging to an equity-seeking group. It's important to remember that data sensitivity is subjective and can vary from one person or context to the next.

A clear privacy policy is useful to clarify which data can be opened and which will remain private. An open data policy constitutes the second of two framework documents. This is the recommended approach which was implemented by Scènes francophones during the project. They supplemented their existing data management policy spelling out which data will be used and how it will be used with an open data policy.[13]

Also as part of the project, CAPACOA improved and translated the publication *Linking Your Data - Outreach Toolkit and Consent Guideline for Membership Based Arts Organizations.*[14] This resource details different strategies for obtaining informed consent and provides examples for each.

Linked Open Data and the Artsdata Knowledge Graph

Artsdata is a distributed database, i.e. a knowledge base linking together local databases in the arts sector, as well as other knowledge bases in the web of linked open data. Each data provider has its own graph within Artsdata. For example, Scène Pro data is loaded in this graph: http://kg.artsdata.ca/culture-creates/artsdata-planet-scenepro/scenepro-orgs.

When the Artsdata team undertakes the task of assisting a potential contributor in a data opening process, they begin by doing a quick analysis of the contributor's data, as previously described. If necessary, the team can also assist in improving data quality and implementing FAIR principles, by making data available using a machine-readable open protocol.

This is followed by an "extraction, transformation and loading" process:

- Data is extracted using the most appropriate technological option, depending on the state and structure of the website or the availability of an API.

---

[13] https://scenesfrancophones.ca/legal/politique-de-donn%C3%A9es-ouvertes

[14] MacIntosh, B. and Julien, F. (2024). Edition 1.1. https://linkeddigitalfuture.ca/2022/10/25/connecting-with-arts-service-organizations-a-key-to-opening-data/

- Data is transformed into RDF format, which involves mapping the contributor's data model to the Artsdata data model (based on the Schema.org ontology) and to other models loaded into Artsdata.
- The extracted and transformed data is loaded into a graph database instance dedicated to the contributor's data.
- Administrative traceability metadata indicating the contributor as the source of data is generated and published.
- The contributor's data is made available to data consumers via the SPARQL terminal and Artsdata APIs under the CC0 1.0 Universal license and/or other copyleft licenses employed by the contributor, in accordance with FAIR principles.

The Artsdata team then enriches the data in the following manner:

- The team reconciles instances of the contributor's dataset with the library of entities already present in Artsdata (where data quality allows).
- It retrieves persistent identifiers associated with these entities from Wikidata, Musicbrainz and other open knowledge bases (when data quality permits).
- It assigns persistent Artsdata identifiers to these entities (when the data meets the minimum requirements for identifier assignment).
- It infers new data from existing data, including from reconciled entities in other knowledge bases (when data quality allows).

As part of the project, the extraction, transformation and loading of Scènes francophones data was greatly facilitated by the fact that the platform already encoded its data in JSON-LD format, a common encoding format for linked open data.

Persistent Identifiers: The Key to Linking Data

In order to share clearly disambiguated data, Artsdata's data reconciliation process relies on persistent identifiers. This is not a requirement specific to Artsdata, but a key element of any data sharing process:

> It is absolutely necessary to assign unique and persistent identifiers to the entities being described, whatever their nature, in order to ensure the proper management, accessibility and reusability of the data and metadata.[15]

A persistent identifier is an alphanumeric string that designates an entity or instance (for example, an artist) in a permanent and global way, *independently of the database in which it is referenced*. When the same persistent identifier is referenced in several databases, it is sometimes called a "bridge identifier", as it facilitates the reconciliation and circulation of data between databases.[16]

---

[15] Ministère de la Culture et de la Communication de la France (2015). *Identifiants pérennes pour les ressources culturelles: Vade-mecum pour les producteurs de données.* https://www.culture.gouv.fr/Espace-documentation/Publications-revues/Identifiants-perennes-pour-les-ressources-numeriques

[16] The Artsdata data model provides recommendations for bridge identifiers, which can be accessed at https://culturecreates.github.io/artsdata-data-model/index.html#bridge-identifiers

Persistent identifiers are frequently expressed as **Uniform Resource Identifiers** (URI). A URI identifies a resource and provides access to its metadata. A URI is therefore actionable: it resolves to a location on the Web (just like a URL) where metadata is made available in a human-readable format (e.g., an HTML document) and/or in machine-readable format (e.g., a JSON-LD application or RDF document).

In addition to identifying and locating an entity or instance of an entity, a URI also *links* these resources together, the same way a hyperlink links web pages.

For all these reasons, FAIR Principles state the following:

> Principle F1 [(meta)data are assigned globally unique and persistent identifiers] is arguably the most important because it will be hard to achieve other aspects of FAIR without globally unique and persistent identifiers.[17]

An identifier in the form of a URI is to the web of linked open data what a primary key is to a relational database.

To facilitate the flow of data between a relational database and the web of linked open data, it is possible to define a naming convention that converts the primary key of an instance described in the database (for example, a presenter or an artist) into a URI designating the same entity across the web of data.

During the course of the project, the Scènes francophones platform adopted the following naming scheme recommended by the project team:

```
https://scenesfrancophones.ca/id/<instance primary key>
```

For example, the URI https://scenesfrancophones.ca/id/39 refers to the Centre culturel de Caraquet, whose primary key is 39 in the Scènes francophones database. When this URI is called in a browser, it resolves to the URL https://scenesfrancophones.ca/diffuseurs/centre-culturel-de-caraquet, which corresponds to the Centre culturel web page in the Scènes francophones directory of presenters.

Finally, in other circumstances, it is possible to name a URI by adding the # symbol to the URL of an existing web page, followed by a character string acting as a unique identifier. This type of character string is called a "URI fragment." This method is explained in the Artsdata documentation.[18]

Don't confuse the URI of something with the URL of the web page where that thing is described. A URL is a *digital medium* that contains content (textual and audiovisual) about things in the real world (people, organizations, events, etc.). A URI refers to *something* described in the content of the web page.

---

[17] Go FAIR (n.d.). *FAIR Principles*. https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/

[18] https://culturecreates.github.io/artsdata-data-model/gabarits-jsonld/README.html#fn:2

Methods for Publishing and Extracting Open Data

To encourage data reuse, data must be made available using machine-readable open protocols. As part of the project, four methods of publishing and extracting data were considered to enable Artsdata (or other parties) to access platform data.

Page-By-Page Publishing and Front-End Extraction

This method involves using a crawling "robot" to harvest structured data, mainly data that conforms to the Schema ontology,[19] and unstructured data in the body of the web page. The advantages of this method are:

- Ease of implementation, both for the data provider and for Artsdata
- Data security (because data is read-only)
- Process automation
- Positive impact of Schema structured data on search engine web crawlers

The disadvantage of this method is that few platforms are able to extract data in this way. What's more, as the configuration of a harvesting bot is specific to the site's structure, any change to the structure can disrupt the process.

This approach was initially considered for the extraction of Scènes francophones data. However, as no section of the public site had data on venues, the data was extracted using the following method.

Front-End Publishing and Extraction by MIME Media Type

With this extraction method, a dataset is sequentially exposed on a single URL in an appropriate MIME media type (for example, `application/json`, `application/ld+json`, `application/rdf+xml`, `text/xml`, `text/turtle` etc.). Sometimes, content negotiation[20] can be used to present the dataset according to different MIME types (with or without a human-readable HTML page). The advantages of this method are:

- Relative ease of implementation (compared to an API)
- Security (because data is read-only)
- Process automation
- Potential contribution to the discoverability of entities exposed using this method
- Content negotiation

The disadvantages of this method are:

- Potential for data updating latency, depending on the cache refresh rate
- Need for special technical skills to set up and transform data into the desired media format
- Need to update extraction and transformation scripts when changes are made to the database

---

[19] https://schema.org/docs/schemas.html
[20] https://en.wikipedia.org/wiki/Content_negotiation

This strategy was proposed and adopted by Scènes francophones as part of the project. The implementation by the Scènes francophones developers and data extraction by Artsdata were both straightforward and efficient.

▷ [Scènes francophones JSON-LD endpoints](#)

API Extraction

An API (Application Programming Interface) enables two or more software programs to communicate. Just as a user interface bridges the gap between a human and a software program, an API bridges the gap between different programs. In the context of this project, an API would be an interface for directly querying a database to extract data of interest. The advantages of this method are:

- Ability to be called by a large number of applications to perform a wide range of operations, in addition to reading data
- Automation

The disadvantages are:

- Difficult implementation (compared to other methods proposed in this document)
- Need to maintain the API

This strategy was put in place by Scène Pro during a previous project carried out with the support of CAPACOA's *Linked Digital Future* initiative.

▷ [Scène Pro API documentation](#)

Data Exporting in CSV Format

The easiest way to extract data is to export it as a CSV file. This type of file can easily be read by software like Excel to facilitate the transformation stage. The disadvantage of this method is that it requires manual intervention (unless you set up a script that generates the file itself at regular intervals). This extraction mode is thus well suited for one-off extractions, but less so when data extraction is a recurring process.

Observations on Information Systems Analyzed by the Project

The primary and common goal of the platforms that were analyzed is to centralize and facilitate business processes for the performing arts industry, for example applications for showcases at performing arts networking events (and the selection of these showcases for by the event organizers). Typically, these networking events allow artists to showcase their show in the hope that it will be purchased by one or more presenters. The application process allows artists to apply for a showcase at these networking events, i.e. a moment on stage during which they can present an excerpt or the entire show they wish to sell to presenters.

Some platforms also manage the registration process for networking events, including accreditation of participants. Finally, not all platforms manage the payment process.

Some platforms use the information they collect about presenters' scheduling to build a calendar of cultural events for public consumption, which helps increase the discoverability of artists and shows.

Table 1: Overview of platforms and their business processes

Business process	Scène Pro	Scènes francophones	I Want to Showcase	Arts Touring Connector
Showcase application	✔	✔	✔	
Directory of show offers	✔	✔		
Directories of presenters and venues	✔	✔		✔
Block booking				✔
Events calendar		✔		

Finally, some platforms cover specific regions or linguistic communities of Canada.

Given Artsdata's Canada-wide focus, the presence of these platforms is an opportunity to access a wide range of data on the country's performing arts ecosystem.

However, the analysis has revealed a high level of heterogeneity in the nature and structure of the data about artists, shows, presenters and venues collected by the platforms. Both the main entities and their attributes differ considerably between platforms. This is partly due to the variation and scope of the business processes integrated into the platforms (application, registration, payment, calendar), as well as the regional or linguistic contexts of the platform managers and users.

This heterogeneity hinders the implementation of a uniform, standard process to open data. Artsdata cannot apply the same formula to every platform, a fact which limits what gains can be made in terms of effort and time spent.

Observations on the Platforms' Ontologies

Our analysis identified matches based on both entities (or classes) and their attributes (or properties).

The project identified four main groups of entities present in the data sources of most of the platforms:

- Entities designating actors, whether individuals, groups or organizations.
- Entities designating shows, in the sense of works intended to be performed on stage.
- Entities designating events and, more specifically, show performances.

- Entities designating venues, such as performance halls.

Observations:

- Although the ontologies have many equivalent classes, these classes have few common properties. When equivalent classes share few properties, aligned ontologies may be less useful in different settings. Interoperability between systems using these ontologies will be limited because of the lack of common properties. The exchange of data between systems could be incomplete. One way to remedy this would be to enrich the ontologies by including missing properties or by making them more compatible. This usually involves adding new properties or adapting existing definitions.

- Data disparities occur because of synonymy and of differences on a conceptual level between languages. For example, the platform I Want To Showcase uses the term "Performer" to describe a performing arts actor who presents an excerpt of a show in a showcase event. Arts Touring Connector and Scènes francophones use the term "Artist" / "Artiste" (respectively) to express a similar concept regarding stage performances. Each platform's term can refer to an individual artist, a group, a collective or a company.

- There are cardinality disparities. One the one hand, ATC established a many-to-many cardinality (0,m:1,m) for the relationship between the "Artist" class and the "Show" class: an artist can be linked to zero, one, or many shows, while a show must be linked to at least one or more artists. With this particular cardinality, one can represent two theatre companies joining forces to co-produce a show or else two musicians touring as a duo. On the other hand, Scènes francophones had built its user interface in such a way that users were forced into a one-to-one cardinality: each artist could only be linked to a single show, and each show could only be linked to a single artist. This led users to create either duplicate artist instances (two entries about the same artist) or else a single artist instance describing two distinct theatre companies. When these data quality issues came to light during the project, Scènes francophones implemented fixes: they implemented a many-to-many cardinality, like the ATC, and they modified their user interface accordingly.

- There are also conceptual differences in granularity. For example, in the Référentiel du Québec ontology, the term "Contributeur" is much more generic and has a variety of meanings. In contrast, the Scènes francophones ontology offers more precise details on this generic concept, using specific terms such as "Artiste" (artist), "Diffuseur" (presenter) and "Réseau" (network). Precise and meaningful matches are difficult to achieve when generic concepts in one ontology need to be aligned with more specific concepts in another ontology. More complex and precise mapping rules are required to manage differences in granularity. In the case of the "Contributeur" concept, the matches will be complex and of a (1, m) nature, where **1** represents a single entity of a data source (IWTS, ATC, Scènes francophones) and **m** its equivalent as expressed by a logic formula with the meaning of the term "Contributeur" in the Référentiel du Québec. The following table

illustrates how matches can be established between the generic concept "Contributeur" in the Référentiel du Québec and the three concepts "Artiste," "Diffuseur" and "Réseau" in Scènes francophones.

Table 2: Examples of complex matches

Generic concept (Référentiel du Québec)	Specific concept (Scènes francophones)	Equivalence in propositional logic	Description of equivalence
Contributeur	Artiste	$\forall x(Artiste(x) \iff (Contributeur(x) \land typeContributeur(x, "Artiste")))$	"Artiste" is equivalent to Contributeur where typeContributeur is set to "Artiste".
Contributeur	Diffuseur	$\forall x(Diffuseur(x) \iff (Contributeur(x) \land typeContributeur(x, "Diffuseur")))$	"Diffuseur" is equivalent to Contributeur where typeContributeur Is set to "Diffuseur".
Contributeur	Réseau	$\forall x(Réseau(x) \iff (Contributeur(x) \land typeContributeur(x, "Réseau")))$	"Réseau" is equivalent to Contributeur where typeContributeur is set to "Réseau".to

- As was seen in the observations on the information systems analyzed, there are differences in the coverage of the ontologies in addition to polysemy of their terms. Automating the alignment process becomes more complex when there are homonyms and differences in definitions. Attempts to merge or interoperate data from such platforms can lead to incoherent results. Human intervention will be needed to interpret these discrepancies and, above all, domain experts will be needed to determine the matches between entities.

Table 3 below contains the alignment of the different ontology classes. For the sake of brevity, properties have been excluded from the table. For more details, please refer to the following spreadsheet:

▷Mapping of performing arts ontologies

Interpreting the data in the table: Perfect matches represent classes with identical or very similar meanings (in this case, equivalent matches). When one class is more generic than another, the relationship between these two classes should be that of superclass and subclass. We would therefore say that this superclass subsumes the other, and this relationship becomes one of subsumption.

Prefixes of ontologies referenced in the table:

- *adr:*   stands for   `PREFIX adr: <http://kg.artsdata.ca/resource/>`
- *schema:*   stands for   `PREFIX schema: <http://schema.org/>`

Table 3: Matches identified during ontology analysis

Référentiel du Québec (RQ)		Scènes francophones (SF)		Artsdata (AD)		IWTS		ATC	
Entities	Type of match	Entities	Type of match	Entities	Type of match	Entities	Type of match	Entities	Type of match
Contributeur and typeContributeur value Artiste or groupe	Contributeur subsumes entities from (SF, ATC,IWTS)	Artiste	Perfect match of entities- (ATC,IWTS,SF) -Subclass of Contributeur (RQ)	PerformingArtGroup or schema:Person (and profession value performer).	*PerformingArtGroup is a subclass of Artiste (SF)* PERFORMER (IWTS) and Artist (ATC)	PERFORMER	Perfect match of entities- (ATC,IWTS,SF) - Subclass of Contributeur (RQ)	ARTIST	Perfect match of entities- (ATC,IWTS,SF) -Subclass of Contributeur (RQ)
Location	Subclass of 'Place' in AD	/	/	schema: Place	*Subsumes RQ entity "lieu"*		/	/	/
Performance hall	"Perfect Match (RQ,ATC,SF)" subclass of Lieu(RQ)	SallesDeSpectacles	"Perfect Match (RQ,ATC,SF)" subclass of Lieu(RQ)	/	/	/	/	PERFORMANCE_SPACE	"Perfect Match (RQ,ATC,SF)" subclass of Lieu(RQ)
Spectacle	"Perfect Match (RQ,ATC,SF)" and Subclass of Oeuvre (RQ)	OffreSpectacle (Spectacle)	"Perfect Match (RQ,ATC,SF)" and Subclass of Oeuvre (RQ)	schema:CreativeWork	For CreativeWork, this entity is not clearly specified, which complicates the matching process.	/	/	SHOW	"Perfect Match (RQ,ATC,SF)" and Subclass of Oeuvre (RQ)
Offre				schema:Offer	*Subsumes RQ entity "Offre"*				
Contributeur and typeContributeur value Diffuseur	Contributeur Subsumes entities from (SF,AD,ATC)	Diffuseur	"Perfect Match entities from (ATC,SF,AD)" Subclass of Contributeur (RQ) and adr:ArtsOrganization(AD)	adr:PresentingOrganization	"Perfect Match entities from (ATC,SF,AD)" Subclass of Contributeur (RQ) and adr:ArtsOrganization(AD)	/	/	MEMBER	"Perfect Match entities from (ATC,SF,AD)" Subclass of Contributeur (RQ) and adr:ArtsOrganization(AD)
/	/	EvenementContact	"Perfect Match(SF,AD)" Subclass of schema:event(AD)	adr:Conference	"Perfect Match(SF,AD)" Subclass of schema:event(AD)	/	/	/	/
/	/	RepresentantArtiste	"Perfect Match (ATC,SF,AD)" Subclass of adr:ArtsOrganization (AD)	adr:ArtistRepresentative	"Perfect Match (ATC,SF,AD)" Subclass of adr:ArtsOrganization (AD)	/	/	REPRESENTATIVE	"Perfect Match (ATC,SF,AD)"Subclass of adr: ArtsOrganisation (AD)
Représentation	"Perfect Match(RQ,SF,AD)"Subclass of schema:event(AD)	Évènement	"Perfect Match(RQ,SF,AD)" Subclass of schema:event(AD)	adr:PerformingArtsEvent	"Perfect Match(RQ,SF,AD)" Subclass of schema:event(AD)	/	/	/	/
SÉRIE	"Superclass of adr:EventSerie(AD)			adr: EventSeries	"Perfect Match(AD,ATC)" Subclass of SÉRIE(RQ) and schema:event(AD)	/	/	TOUR	"Perfect Match(AD,ATC)"Subclass of SÉRIE(RQ) and schema:event(AD)
Contributeur and typeContributeur value Distribution network	Contributeur Subsumes entities from (SF,AD,ATC)	Réseau	"Perfect Match entities from (ATC,SF,AD) Subclass of Contributeur (RQ) and ArtsOrganization(AD)	adr:ArtsServiceOrganization	"Perfect Match entities from (ATC,SF,AD)" Subclass of Contributeur (RQ) and adr:ArtsOrganization(AD)	/	/	ORGANIZATION	"Perfect Match entities from (ATC,SF,AD)" Subclass of Contributeur (RQ) and adr:ArtsOrganization(AD)

Observations on Equity, Diversity and Inclusion

Existing data on target audiences does not often indicate whether or not the shows being showcased can be enjoyed by deaf, hard-of-hearing or blind audiences.

In addition, the data does not often indicate whether or not the venues can accommodate disabled persons without placing them in a disabling situation.[21]

The platforms analyzed include one or more fields that allow individuals to self-declare their cultural affiliation or membership in the LGBTQIA2+ community. The options available for self-declaring one's cultural affiliation are not standard or uniform on the different platforms.

Not all platforms include a field for self-declaration of gender identity.

Not all platforms make it easy to distinguish solo artists from groups or collectives, making it more difficult for individuals to self-declare their cultural affiliation and gender identity.

More EDI resources are available on the Mass Culture website here.[22]

---

[21] "A person with [a disability] is referred to as a disabled person. This person is said to be in a disabling situation when their autonomy is limited because they are unable to fully carry out an everyday activity or their social role due to the absence of facilitators designed to remove obstacles in their environment. For example, if a person with a mobility-related disability needs to go to a higher floor, they are not in a disabling situation if they use a wheelchair and have access to an elevator. However, they will be in a disabling situation if they can't take an elevator". https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8385452/handicap

[22] Bernicky, S. (2024) Spiralling Outwardly for Equity in Public Arts. https://massculture.ca/research-in-residence/spiralling-outwardly-for-equity-in-public-arts/

Recommendations

General Recommendation Regarding Information System Design and Ontological Alignment

Conceptual modelling should receive special attention when designing and maintaining an information system. No matter what technological solution is chosen, conceptual data modelling makes it possible to identify, define and visualize all the information elements and relationships that need to be represented in the information system. This modelling should be carried out early in the design process of an information system such as a web platform. It must be based on an in-depth knowledge of the domain rather than on technical expertise. Finally, it needs to be documented – notably in an entity-attribute dictionary and in schemas – and updated as understanding of the environment evolves.

New conceptual models should build off of existing ontologies, wherever possible, in order to achieve harmonization. Ontologically aligning conceptual modelling with existing ontologies will help future-proof the digital platforms that rely on them. Ontological alignment involves revising the way classes are defined, so that they better match the definitions of existing ontologies. Ontological alignment can also involve integrating additional attributes to describe certain classes. Finally, it may be necessary to revise the names given to classes and attributes to avoid problems of synonymy, homonymy and polysemy.

Conceptual modelling is not just good design practice. In the event of a security or confidentiality incident, it can also provide various documents that may be requested by authorities in charge of protecting privacy and personal information in Canada.

It is important to note that changes to an information system can trigger the need to revise the business processes expressed by the information system. In this case, the scope of the initiative will be quite different. It's advisable to review the definition of the project in which the changes to the information system are to be made to validate timelines, budgets, constraints, expectations and the availability of appropriate resources.

Recommendation Regarding the Granularity of the Artist Type

Data collection carried out by platforms would benefit from having finer entity types to handle the following instances of the artist type:

- Solo artist (i.e. an individual)
- Group or collective (for example, a duo or a music group)
- Organization (for example, a theatre or dance company)

Increasing the granularity, i.e. the degree of precision, of the artist entity has several advantages:

- Disambiguating featured artists from accompanying performers
- Enabling the proper use of persistent identifiers
- Allowing groups and collectives to self-declare their cultural affiliation independently of individuals

Ultimately, this kind of granularity will have a positive impact on the level of interoperability of different datasets, encouraging them to be opened and shared via a third-party system such as Artsdata.

Recommendation Regarding EDI

Data concerning gender identity and cultural affiliation should only be made open if, and only if, platforms are able to guarantee anonymization of the data. As the content of this data may be harmful to the people concerned, anonymization is essential.

In the long term, platforms might benefit from having accessibility modules included in their interface.

Image 2: The show Êtres de bois *by the dance company Fleuve Espace Danse is one of the shows that can be booked on Scène Pro. Performer: Victoria Coté. Photo credit: Emie-Liza Caron St-Pierre.*

Conclusion

In the aftermath of the pandemic, it has become clear that improving and consolidating Canada's performing arts information systems would help the sector become more resilient. To this end, and given its strategic positioning within the performing arts sector in Canada, CAPACOA submitted a project proposal to the Community Services Recovery Fund for the purpose of mapping key performing arts information systems, documenting disparities in ontologies and recommending improvements geared towards data openness and reuse.

Four of CAPACOA's partners, each the owner of a digital platform that collects, stores and uses data on the performing arts in various regions of Canada, participated in the project. The information system underpinning each of these platforms was modelled and analyzed, which revealed the heterogeneity of the ontological models used by these platforms. This is one of the main takeaways of the project. The project also examined what needs to be done and tried out ways to reduce ontological disparities between the platforms in order to initiate open data processes.

The other main takeaway is the impact of a platform's governance on its operations and, within the context of the project, on processes to open and share data. The aim here is not to criticize governance choices made by organizations that own the platforms. It is rather to acknowledge the importance and validity of the various modes of governance and to ensure that they are considered during any subsequent activities following this project or in any other similar project.

Finally, it's important to categorize the initiative as a digital transformation project, since in essence, the project aims to open performing arts data and encourage its reuse. As a digital transformation project, it must be part of a strategic planning process that goes beyond the decision whether or not to acquire the "latest trending application"[23], a process that must always consider the various social and technical challenges of this type of project.

---

[23] Pelletier, C. and Cloutier, L. M. (2019). *Challenges of digital transformation in SMEs: Exploration of IT-related perceptions in a service ecosystem*. In Proceedings of 52nd Hawaii International Conference on System Sciences (HICSS), Grand Wailea, HI.
https://scholarspace.manoa.hawaii.edu/bitstreaam/10125/59934/0494.pdf