



# DATA COLLECTION

Nguyen Khanh Trang Dang, Karine Léonard Brouillet and Nicholas Lobraico

## Table of contents

- Introduction
- Origins and list of institutions
- Data collection
- Additional data
- Reliability of the dataset
- The dynamic nature of websites combined with an extended data collection period
- The subjective nature of analysis
- Contact us
- Endnote

## Introduction

In this inaugural edition of the *LMCO Bulletin*, we will provide a brief introduction to the data collection process. More specifically, we will answer the following questions:

- How did we compile a list of institutions to analyze?
- What did we collect data on?
- How did we collect data?
- What other data did we use?

To obtain a PDF version, please consult [Government of Canada Publications](#).

To consult the data and its associated documentation, please visit the page [Landscape of museum collections online](#).



## Origins and list of institutions

In 2022, the Canadian Heritage Information Network (CHIN) commissioned a project to collect data on the online presence of heritage institutions in Canada. The project involved analyzing nearly 2000 institutions across the country. The list of institutions to be analyzed was initially derived from the [Open Database of Cultural and Art Facilities \(ODCAF\)](#), maintained by Statistics Canada, and internal lists used at the Department of Canadian Heritage, including the list of contributors to Artefacts Canada and the list of institutions associated with the [Government of Canada's Survey of Heritage Institutions](#).

The list was subsequently reviewed to focus on heritage institutions actively managing collections containing primarily cultural objects and belongings.<sup>1</sup> Exhibition centres, historic sites, and libraries and archives were thus not included, except in cases involving special collections or the merger of institutions (for example, a historic site in charge of a collection or a museum with a genealogical archive). Botanical gardens and zoos were also excluded because they manage living collections.

In addition, several institutions were removed, either because they were deemed unsuitable (for example, the institution was a library or another type of institution beyond the defined scope of interest), identified as a duplicate (in which case, the duplicate records were consolidated into one), or because the institution appeared to have closed. This initially resulted in a dataset of 1915 institutions (or records). However, further data cleaning to remove less reliable data, done in preparation for publication, resulted in a final dataset consisting of 1870 institutions (or records). Although the resulting list is by no means exhaustive, it is a large enough sample to be considered representative of heritage institutions managing collections in Canada.

## Data collection

Data collection was completed between July 2022 and September 2023. CHIN analyzed the online presence of listed institutions by answering the following questions:

- Does the institution operate seasonally?
- Does the institution report having information technology (IT) staff?
- Does the institution have a dedicated website?
- Does the institution's website have some form of online collection? If so, is it searchable?
- How many objects are online?
- What types of objects are in the collection?



- What disciplines does the institution align with (for example, natural history or fine art)?
- Does the institution have other online offers (for example, educational programs or virtual exhibitions)?
- Does the institution contribute to online collection portals (Artefacts Canada, Nova Muse, Données Québec, Sask Collections, OurOntario, Musetoba, etc.)
- Does the institution have a presence on social media (Facebook, X [Twitter], Instagram, YouTube)?
- When available, what is the institution's operational budget?

Data on these points was collected primarily via online searches and the review of institutional websites and annual reports. The website of the institution was the designated primary source of information, but other additional information sources such as directory listings from provincial and territorial museum associations, Canada Revenue Agency listings, select social media accounts or news articles were also used to complete each record.

## Additional data

In addition to online data collected, the project team has been able to source a handful of external datasets that further enhance our ability to use this data to analyze the digital capacity of heritage institutions in Canada. This includes data from the following sources:

- Statistics Canada
- [Open Database of Cultural and Art Facilities \(ODCAF\)](#) list of institutions (2020)
- [Census data](#) on population, language and census subdivisions (2021)
- [Innovation Science and Economic Development \(ISED\) Canada's National Broadband Map](#) (2024)
- [Canada Revenue Agency's list of registered charities](#) (2023)
- [Wikipedia](#)

Data from Statistics Canada's census program or ISED's national broadband map required our dataset to be mapped onto the same geographic reference model. This was done using the open-source software QGIS, which also allowed us to perform spatial analysis resulting in new data, such as the distance between institution and the density of institution in a given area.



## Reliability of the dataset

There are two important considerations that influence the reliability of this dataset as a whole:

- The dynamic nature of websites combined with an extended data collection period
- The subjective nature of the analysis

### The dynamic nature of websites combined with an extended data collection period

An unavoidable limitation of this data is that it is based on a survey of the Internet, which is inherently dynamic. An institution's website can be updated or changed at any time, and the same is true for creating or closing social media accounts. Because each institution was only surveyed at a given moment in time, it is likely that many of the websites analyzed have since been enhanced in some manner.

The dynamic nature of websites combined with the extended data collection period is cause for concern because establishing an online presence and offering access to collections online are increasingly common objectives for heritage institutions.

Therefore, institutions that were analyzed earlier on in our data collection process might be disproportionately labelled as having a less developed online presence and/or online collection than those analyzed several months later. These concerns can be alleviated, to some extent, by comparing records completed at different points of the reference period. The table below provides such an analysis and suggests that the extended data collection period does not, in fact, jeopardize the integrity of the dataset as a whole.

**Table 1: statistics on the institutions and their online presence by quarter (2022–2023)**

<b>Metrics</b>	<b>2022 Quarter 2</b>	<b>2022 Quarter 3</b>	<b>2022 Quarter 4</b>	<b>2023 Quarter 1</b>	<b>2023 Quarter 2</b>	<b>2023 Quarter 3</b>	<b>Averages</b>
<b>Count of institutions analyzed</b>	143	462	290	135	489	351	No data
<b>Average number of online records</b>	12,490	19,910	25,906	6,626	5,624	3,567	13,769
<b>Percentage of institutions with a website</b>	80%	75%	70%	70%	75%	75%	74%



<b>Percentage of institutions with no online collection</b>	55%	68%	83%	87%	83%	82%	77%
<b>Percentage of institutions with a searchable sample</b>	3%	6%	4%	4%	4%	5%	5%
<b>Percentage of institutions with a non-searchable sample</b>	26%	13%	7%	6%	8%	6%	10%
<b>Percentage of institutions with a searchable collection</b>	13%	11%	5%	3%	4%	4%	7%
<b>Percentage of institutions with a non-searchable collection</b>	4%	2%	1%	0%	1%	2%	1%
<b>Percentage of institutions offering exhibitions online</b>	9%	12%	13%	14%	15%	13%	13%
<b>Percentage of institutions offering educational programs</b>	14%	16%	9%	4%	6%	6%	9%
<b>Percentage of institutions offering specialized programs</b>	1%	3%	2%	1%	2%	1%	2%

### The subjective nature of the analysis

The subjective nature of the analysis primarily concerns the information recorded about the types of objects found in collections displayed online and the disciplines associated with each heritage institution. The “types of objects” and “disciplines” fields are populated using terms from controlled vocabularies. In some cases, an institution will describe itself as being within a certain discipline or as having a certain type of object in its collection; however, this is rare, and in most cases these labels are applied according to the surveyor’s own judgment when looking at the available data. This also means that in cases where only a small sample of the collection is online, our record might address only a partial view of the real holdings of the institution. This speaks to an inherent limitation of our data, in that it only concerns the online presence of institutions. An institution may have a much larger collection, diversity of disciplines or programming offers than what is mentioned online.



## Contact us

Despite the limitations outlined above, we hope this dataset will be useful to those interested in evaluating the digital capacity of heritage institutions in Canada, as well as enabling the survey of that capacity over time. We are committed to improving our methods to better support the information needs of the museum community in Canada.

With this in mind, do not hesitate to contact us with any comments, feedback, concerns or questions at [panoramamuseescip-chinmuseumslandscape@pch.gc.ca](mailto:panoramamuseescip-chinmuseumslandscape@pch.gc.ca).

© Government of Canada, Canadian Heritage Information Network, 2025

Published by:

Canadian Heritage Information Network  
Department of Canadian Heritage  
1030 Innes Road  
Ottawa ON K1B 4S7  
Canada

Cat. No.: CH12-3E-PDF

ISSN 3110-9381

### Endnote

1. Since this derives from an internal project, and CHIN's mission is directly tied to supporting collections management, managing a collection was a principal criterion for inclusion in this analysis. Therefore, any institution included on the list of institutions to be analyzed but not considered to be managing a collection was deemed out of scope and removed from the list.