

Approach to Open Data



Except for images or where otherwise stated this document is © National Library of Scotland and is licensed under the Creative Commons Attribution 4.0 International Licence. To view a copy of this licence, visit: <http://creativecommons.org/licenses/by/4.0/>

Cover image: [‘Superstructure, Fife, 15 Sept 1886’](#) in [Forth Bridge Illustrations 1886-1887 collection](#) by [National Library of Scotland](#). Licensed under Creative Commons Attribution Share-Alike 4.0 Generic Licence. To view a copy of this licence, visit: <https://creativecommons.org/licenses/by/4.0/>

1. Overview

This document sets out the Library's approach to open data, detailing the value of open data to the Library's core mission and strategic goals and the ways in which we make our data openly available for reuse.

In summary, we will:

- publish our non-personal and non-commercially sensitive data as [3-star open data](#). This includes out of copyright collections as data.
- use non-proprietary formats and licences that are conformant to the Open Definition.
- apply [FAIR](#) and [CARE](#) principles to our release of open data.
- continue to publish datasets and add these to our [Open Data Register](#) as they are identified.
- make our data available, in the first instance, through our open data website, [Data Foundry](#).

2. Context

In 2015 the Scottish Government released its [Open Data Strategy](#), asking public organisations in Scotland to prepare plans for publishing their open data. An independent report, '[Open Data in Scotland](#)', commissioned by the Scottish Government, published in 2024, provides recommendations for next steps around open data publication, with a view to Scotland positioning itself as a leader in open data. Additionally, the Scottish Government's [Artificial Intelligence Strategy \(2026-31\)](#) acknowledges the foundational need for responsibly shared, easy-to-access data, for the application of AI tools, and for these to generate value.

3. Open data vision

The National Library of Scotland recognises and welcomes the societal, economic, governance and institutional benefits of open data. The opening of public data aligns with our mission, detailed in the Library's 2025-2030 strategy, '[The Next Chapter](#)', to develop, protect and share the national collections widely and to a diverse range of users, for education, research and enjoyment. It also corresponds with our strategic themes of Responsible Stewardship, Connection and Sustainability.

4. Goals

The Library has published data openly for many years and, over the past five years, has dedicated resource to publishing and encouraging reuse of our open data collections.

By opening our data we aim to generate:

1. **Better data:** There will always be ways to improve our data and, in turn, our business and our public service. We aim to find improvements in our own creation, management, maintenance, release and use of our data by implementing this plan.
2. **Wider re-use and impact of our data:** We recognise that our data is valuable in a range of contexts and can be used in many ways, and to wide benefit. We aim to reduce barriers to access our open data collections and to encourage re-use of the Library's open data through implementation of this plan. This enables activities such as large-scale exploration of historical text for research; easy access to and use of cultural heritage data for teaching; creative use of historical data to make library data more accessible to wide audiences by digital artists; or use of spatial data for commercial purposes.

5. Approach

5.1. Data

We will continue to publish our non-personal and non-commercially sensitive data as open data in accordance with the Open Definition. We already make data available, for example on our websites, in our catalogues, and through third parties such as [The European Library](#). However, we have not published all of this data as open data.

We therefore have two priorities under this plan:

1. To publish as 3-star open data the data that we already make available;
2. To identify our non-personal and non-sensitive data that we do not currently make available, and to publish this data as 3-star open data.

Our work to release open data involves identifying datasets and data sources that can be published under this plan. We will add these datasets to the Open Data Register as they are published.

Certain datasets can be published more quickly than others. Some datasets only need to be licensed and encoded in an open format, while other datasets may need to be 'cleaned' to remove personal data or data contributed from other sources that we don't have permission to openly license.

In general, the data that we create and publish falls into two categories:

1. Collections as data
2. Organisational data

We will provide links to the data from relevant locations on our websites (e.g. we will link to open data related to our spending from our Freedom of Information (Scotland) Act publication scheme).

5.2. Licences and rights statements

We publish our data using rights statements or open licences that are open or conform to the principles of the [Open Definition](#). We will primarily publish our data using the following statements or licences:

Licence	Description	Link
Public Domain Mark	This statement means that the associated data is free of known copyright restrictions and is in the public domain.	https://creativecommons.org/publicdomain/mark/1.0/
No Known Copyright	This statement means that the associated data is likely to be free of known copyright restrictions, but that we have insufficient information to confirm that the material is in the public domain.	http://rightsstatements.org/vocab/NKC/1.0/
Copyright Undetermined	This statement means that the copyright status of the associated data has been evaluated, but this process was inconclusive. The associated data may be protected by copyright.	http://rightsstatements.org/vocab/UND/1.0/
CC0	The Creative Commons CC0 'No Rights Reserved' statement waives all rights in content, meaning that no legal restrictions are placed on the use and re-use of the content.	https://creativecommons.org/publicdomain/zero/1.0/

CC-BY	The Creative Commons Attribution (CC-BY) licence allows the free use and re-use of content for commercial or non-commercial purposes, subject only to the requirement to credit the source of the content.	https://creativecommons.org/licenses/by/4.0/
--------------	--	---

The Scottish Government's [Open Data Strategy](#) calls on organisations to publish data under the [Open Government Licence \(OGL\)](#), which is maintained by The National Archives.

However, we license our data with Creative Commons licences instead of the OGL for the following reasons:

- The OGL was designed for use with Crown copyright material, and the Library is not a Crown copyright body.
- The OGL is specific to the United Kingdom, whereas Creative Commons licences are used globally, reflecting the global audience using our data.
- The use of Creative Commons licences supports discoverability as they are used globally. This reflects our global audience who use our data.

Nonetheless, the terms of the OGL and the CC-BY licence are compatible. Because we will primarily use the CC-BY licence or the more open CC0 mark wherever possible, we are confident that our open data licences conform to the Scottish Government's strategic objectives for open data.

5.3. Formats

We will publish our data as machine-readable structured data using non-proprietary formats. Although we may use other open formats as required in respect of any particular set of data, we will primarily publish our data using the following formats:

Format	Description	Link
Plain text	Plain text files (.txt) contain only text, without formatting.	
CSV	Comma Separated Values (CSV) provides a file format for storing tabular data as plain text, using commas to separate fields. CSV is an open alternative to proprietary file formats (e.g. Excel).	https://tools.ietf.org/html/rfc4180
MARCXML	<p>Machine-Readable Categories (MARC) is a set of open standards for describing library items. The MARC standards are maintained by the Library of Congress.</p> <p>Extensible Markup Language (XML) is a markup language that allows data to be encoded in a format that is both human-readable and machine-readable. XML is defined by the W3C.</p> <p>MARCXML is a framework developed by the Library of Congress for using MARC data in the XML language.</p>	http://www.loc.gov/marc/ http://www.w3.org/TR/2008/REC-xml-20081126/ https://www.loc.gov/standards/marcxml/
ALTO XML	Analysed Layout and text Object (ALTO) is an open XML schema, maintained by Library of Congress.	https://www.loc.gov/standards/alto/
Dublin Core	Dublin Core is a metadata schema used to describe library collections, as well as	https://dublincore.org/

	other forms of digital and physical resources.	
--	--	--

5.4. Level of openness

We have adopted [Sir Tim Berners-Lee's 5-star deployment scheme](#) for linked open data in assessing our openness. We will publish our data as 3-star open data or above, in accordance with the Scottish Government's [Open Data Strategy](#). We will publish our data as 3-star open data by:

1. Making data available on the web with open licences (see 3.1 and 3.2).
2. Publishing machine-readable structured data (see 3.3).
3. Using non-proprietary formats (see 3.3).

5.5. FAIR and CARE

We align with the [FAIR](#) principles by making our data:

- Findable: our open data is published, in the first instance, on our [Data Foundry](#) website.
- Accessible: to encourage wide reuse, we have aimed to reduce barriers to access by making data available via bulk download, with tiered download options including small, 'trial' datasets; plain text only (where relevant); and full, structured datasets.
- Interoperable: our data formats comply with international standards and are available in non-proprietary formats.
- Reusable: rights or licensing information is clearly applied to our open data and our datasets are released with relevant contextual information and technical metadata.

Where relevant, we will align with the [CARE](#) principles for indigenous data governance, recognising the need for collective governance, authority to control, responsibility and ethics.

5.6. Publication

We publish our open data for download on our [Data Foundry](#) website. We also publish some data specifically aimed at AI-related use on the external platform, Hugging Face.

We will continue to assess the necessity and benefit of making our data available on further discovery sites or portals, for example at the UK or EU levels.

5.7. Resourcing

Preparing and publishing open data is resource intensive, requiring action, intervention and transformation from areas across the Library, including roles such as Curators, Rights, Metadata, Digitisation, Digital Asset Management and Digital Development, as well as choices of appropriate software, file formats and storage/publication platforms. Many of these are existing activities, which need some modification or additional effort to enable the publication of data openly. We will continue to work towards ways to improve these cross-team workflows and, in the longer-term, aim to automate as much of this activity as possible, to enable the publication of open data at scale.

With this in mind, we currently publish and update our data on an ad-hoc basis. We normally indicate planned future releases of data and regularity of updates on our [Data Foundry](#) website.

5.8 Artificial Intelligence and open data

Open data is increasingly valued for training AI models, and we recognise the resource, infrastructure and ethical implications of extractivist data mining practices. However, the longer-term benefits of releasing data as openly as possible to enable broad reuse and analysis remain, and align with the Library's mission to 'share the national collections widely' and our strategic goal to remain a 'responsible steward', by enabling wide access to the collections and serving the needs of current and future audiences, as outlined in '[The Next Chapter](#)'.

Our [AI Statement](#) outlines the Library's principles and commitments relating to the responsible adoption of AI, including that we will 'open up access to collections to enable public use and research alongside safeguarding sensitive data from automated access, upholding the rights of authors and protecting copyright'.

6. Regulatory environment

The Scottish Government's Open Data Strategy (February 2015) called on public organisations in Scotland to have Open Data publication plans in place and published online by December 2015.

The National Library of Scotland is subject to the Freedom of Information (Scotland) Act 2002 and the Environmental Information (Scotland) Regulations 2004, which require the making available of certain information. The Library is also subject to the Re-use of Public Sector Information Regulations 2015, which requires the making of public information useable and re-useable.

The European Commission's [2011 report on open data](#) identified the opening of public information as an 'engine for innovation, growth and transparent governance'. In 2013, the members of the G8, including the United Kingdom, signed an [Open Data Charter](#), which calls for 'Open Data by Default'.

7. Roles and responsibilities

7.1. Overall responsibility

The Director of Collections, Access and Research and the Director of Digital and Service Transformation have overall responsibility for the Approach to Open Data.

7.2. Operational responsibility

The Rights and Provenance Officer and the Digital Scholarship Librarian have operational responsibility for the Approach to Open Data.

8. Related documents

- **Open Data Register:** The [Open Data Register](#), available to download on [Data Foundry](#), is a tabular list of the datasets we have published openly. The Register is a live document, made available on the [Data Foundry](#) website, and datasets will be added to it as they are published.

9. Review

This plan will be reviewed every two years.