# ISTC Data Dictionary

Table of Contents

[1. Full ISTC download (istc\_clean.yaml)](#_Toc187744018)

[**2.** Core data (csv)](#_Toc187744019)

[**3.** Imprints data (csv)](#_Toc187744020)

[**4.** Holdings data (csv)](#_Toc187744021)

[**5.** References](#_Toc187744022)

[Document versions and credits](#_Toc187744023)

The Incunabula Short Title Catalogue (ISTC) is the international database of 15th-century European printing created by the British Library with contributions from institutions worldwide. In 2024, British Library staff began preparing a version of the ISTC data for public release in January 2025. This document was last updated in January 2025.

This document contains an overview of all the available data: the full ISTC download in a YAML file and versions of the same data in two CSV files, which combine the core bibliographical fields with the imprint and holdings data respectively.

The README sections for the csv files contain explainers for all of the fields as well as snapshots of how the information is displayed in the public interface in the [Incunabula Short Title Catalogue](https://data.cerl.org/istc/_search) (all examples are taken from [ISTC ic00432000](https://data.cerl.org/istc/ic00432000)).

Background information on ISTC is available in the separate ISTC background document. Contact [incunabula@bl.uk](mailto:incunabula@bl.uk) if you have further questions.

### Full ISTC download (istc\_clean.yaml)

This is a full unedited download of the ISTC data in human-readable markup language format. The download was done in late 2024. The YAML file contains non-repeatable core bibliographic data and repeatable fields: imprints, holdings, references and related resources. The csv files were made from this file by joining the repeatable imprints and holdings data onto the core bibliographic data.

### Core data (csv)

Core data is included in both datasets (istc\_core\_imprints.csv and istc\_core\_holdings.csv). This is the core bibliographical data, which only appears once for every ISTC record.

|  |  |
| --- | --- |
| ISTC nr. | Alphanumeric ID with 2 letters and 8 numbers. Based on Goff numbers (e.g. Goff C432 = ISTC ic00432000) |
| Author | Author name (some variations might be present i.e. name of one author given in two different forms) |
| cataloguing\_level | Minimal or full |
| date\_of\_item\_from | If we have a range this is the earlier date |
| date\_of\_item\_to | If we have a range this is the later date |
| date\_of\_item\_single\_date | Single date |
| Dimensions | Open field for format information. Frequent formats include Bdsde (broadside), 2°/f° (folio), 4° (quarto), 8° (octavo). Sometimes the type of paper is indicated as well e.g. Royal 4°. |
| id\_assigned\_by | UK country code |
| language\_of\_item | Languages appear in the MARC language codes e.g. lat=Latin, frm=Medieval French. |
| material\_type | Monograph or blank |
| title | Open field containing title of the work. The same work should normally have the same title but there are some inconsistencies. |
| date\_of\_cataloguing | Usually given in 6-digit numbers e.g. 800501 (1 May 1980). Sometimes a 0 might be missing and only 5 digits display in the dataset. |
| notes | Open field for notes referring to the edition as a whole. Often used for references to research publications or notes on dating based on circumstantial evidence like water marks or type analysis. |
| woodcut | This field indicates the presence of one or more woodcut illustrations (TRUE or FALSE). This field was created based on the notes field, from which all notes indicating the presence of woodcuts were extracted. |

**User view**



### Imprints data (csv)

The imprints data alongside the core are included in istc.core.imprints.csv.

Imprint refers to publication data like place and date as well as name of printer/publisher. This information is usually contained in the colophon, a short statement of responsibility at the end of the book. If no publication information is given in the colophon, ISTC may still give imprint information based on circumstantial evidence e.g. water marks or type analysis. In that case, the publication information is given in square brackets. Decision usually justified in a note in core data.

Most records only have one set of imprint information but some editions might have multiple. This can occur when no clear attribution is possible.

For most 15th-century books, there is no distinction between the printer and the publisher. In some cases, a printer prints something on behalf of another printer.

All imprints contain a country code and a geonames ID, which allows for mapping of the place of publication.

|  |  |
| --- | --- |
| imprint\_country\_code | Alpha-2 country code e.g. GB |
| imprint\_date | Open field containing date of printing or date range. Square brackets added if needed. This corresponds with the date given in the date\_of\_item fields in the core bibliographical data |
| imprint\_name | Open field. Name of printer/publisher |
| imprint\_place | Open field. Place of publication |
| geonames\_id | GeoNames ID corresponds with place of publication |
| lat | Latitude (derived from GeoNames) |
| lon | Longitude (derived from GeoNames) |

**User view in ISTC:**



### Holdings data (csv)

The holdings data alongside the core are included in istc\_core\_holdings.csv.

Most ISTC records have more than one holding and it’s the part of the ISTC data that changes most frequently as we continuously add new holdings to existing records. We do not usually include editions for which there is no surviving copy unless there is clear evidence for a copy e.g. reproduction of a now destroyed or missing copy.

Each holding consists of five fields, of which country\_code and holding\_insitution\_name are essential. However, if the holding also includes a holding\_instiution\_id, it then links to a holding institution record in the [Holding Institution Database](https://data.cerl.org/holdinst/). Many institution records contain further information like Geoname IDs, collection notes, or ISIL references. Throughout 2024 a significant amount of work has been put into linking more than 90,0000 holdings in ISTC with the corresponding Holding Institution Database access points.

Both the shelfmark and the copy\_note field are open and optional fields.

Currently, if an institution holds more than one copy of an edition, the copies are usually aggregated into one holding with the number of copies indicated in the copy\_note field. Recently added collections like the Solothurn, Zentralbibliothek collection have been disaggregated with each copy displaying as a separate holding.

Sometimes shelf marks are in the copy\_note field. We are currently planning a project which will allow us to identify shelf marks and put them in the correct field. In addition, we are also planning to disaggregate multiple copies at one institution into multiple holdings.

|  |  |
| --- | --- |
| country\_code | Alpha-2 country code e.g. GB |
| holding\_institution\_id | This links an ISTC holding to a record in the [Holding Institution Database](https://data.cerl.org/holdinst/). If linked the full Institution name displays in ISTC. If empty the holding\_institution\_name will display instead |
| holding\_institution\_name | Open field. Only displays if holding\_institution\_id field is empty. Many entries in holding\_institution\_name contain abbreviated or outdated forms of institutional names. |
| shelfmark | Open field. Can be empty, however for some institutions like the British Library we have a full range of shelf marks. In ISTC it displays in first brackets. |
| copy\_note | Open field, can be empty. Contains information about condition, sometimes provenance information. Quite a lot of variation with regard to the level of detail as well as formatting (e.g. imperfect is given as ‘(-)’, ‘imp.’, ‘imperfect’, and ‘Imperfect’). In ISTC it displays in second brackets in italics. |

**User view**

