

Home

On this page

- [How to use this portal](#)
- [Definitions](#)
- [Licensing](#)

Welcome to the Internet Archive Developer Portal.

What is the Internet Archive?

Once upon a time, there was the Library of Alexandria, which housed thousands of papyrus scrolls and codices. It's resources were available to scholars and students.

Today, there's the Internet Archive, which contains thousands of books, videos, pictures, music, games, websites, cultural artifacts, and more. All of these are in digital form and freely available to anyone with an internet connection. [More about the Internet Archive...](#)

This portal contains information to help you access data from, integrate with, or contribute to the Internet Archive. Some of the things that you can do with the information on this Developer Portal are:

- Use the APIs and services to read data, for example, see what the first edition of Shakespeare's *Hamlet* looked like.
- Update or modify data, for example, change the metadata of an item.
- Upload items, for example, add a video file to the Internet Archive.

Tip

- A complete list of the Internet Archive APIs and services is at [Tools and APIs](#).
- Some step-by-step How-Tos are at [Tutorials](#).
- Quick-start guides are at [Quick start](#).

How to use this portal

The navigation pane at the left contains all the topics. Every page also has a navigation list near the top right.

To locate something on this portal, type a phrase in the Search bar near the top of the navigation pane at the left, and press Enter. You're shown a list of pages that contain the search term.

Definitions

Before you start using the Internet Archive APIs, you might want to become familiar with some terms that are used frequently on this portal.

Item

Things on the Internet Archive are called *items*. A song, a book, or a video is an item. Every item can have one or more files. For example, an item called *Euclid's Geometry* can have a *PDF* file, an *HTML* file, and a *TXT* file. See [Internet Archive Items](#).

Collection

Items can be placed in *collections*. For example, a collection called *European Libraries* can contain

several items, one of which can be *Euclid's Geometry*. An item can belong to more than one collection. See [Internet Archive Items](#).

Metadata

Information that describes an item is called *metadata*. For example, the name of the writer is metadata for a book. The Internet Archive has a schema for the metadata for files and also for items. Additionally, you can specify your own metadata when creating or updating an item. See [metadata-schema/index.rst](#).

WARC (Web ARChive)

Web pages crawled by the Internet Archive are stored as *WARC*. This is a file format for concatenating several resources, each consisting of a set of simple text headers and an arbitrary data block, into one long file. The WARC format is an extension of the ARC file format (ARC) that has traditionally been used to store web crawls as sequences of content blocks harvested from the World Wide Web. Each capture in an ARC file is preceded by a one-line header that briefly describes the harvested content and its length. This header is directly followed by the retrieval protocol response messages and content. See [The WARC Format 1.1](#).

Licensing

The Internet Archive does not assert any new copyright or other proprietary rights over any of the material in its database. The legal issues for community projects like the Internet Archive can be confusing, but the Internet Archive aims to make available database that can be freely used.