

# Data dumps

---

The Wikimedia Foundation is requesting help to ensure that as many copies as possible are available of all Wikimedia database dumps. Please **volunteer to host a mirror** if you have access to sufficient storage and bandwidth.

## About Wikimedia Dumps

---

Wikimedia provides public dumps (<https://dumps.wikimedia.org>) of our wikis' content and of related data such as search indexes and short url mappings. The dumps are used by researchers and in offline reader projects, for archiving (<https://archive.org/details/wikimediadownloads?and%5B%5D=subject%3A%22dumps%22>), for bot editing of the wikis, and for provision of the data in an easily queryable format, among other things. The dumps are free to download and reuse (<https://dumps.wikimedia.org/legal.html>).

Note that the data dumps are not backups, not consistent, and not complete. They are still useful (<https://medium.com/@mjbaldwin/transforming-wikipedia-into-an-accurate-cultural-knowledge-quiz-b0a0f74877c>) even (<http://www.statmt.org/wmt19/translation-task.html>) so (<https://openreview.net/forum?id=r1173iRqKm>).

### What we dump and when

---

- Content and metadata of Wikimedia projects
- Cirrus search indexes of Wikimedia projects
- Wikidata entities
- Short url mappings
- Dump frequency
- More...

### Getting the dumps

---

- Warning about file sizes
- Mirrors for downloading, torrents
- download: XML/Sql dumps (<https://dumps.wikimedia.org/backup-index.html>) (wiki metadata and content)
- download: Wikidata entities (<https://dumps.wikimedia.org/other/wikibase/wikidata/wiki/>)
- download: other dumps and datasets (<https://dumps.wikimedia.org/other/>)
- Older dumps
- Tools for downloading
- Checking the status of a dump run

### Using and re-using the dumps

---

- XML/Sql dump format
- Wikidata entity dumps formats: JSON and RDF
- Other dump formats
- Importing the dumps

### Getting help

---

- Xmldatadumps-l mailing list (<https://lists.wikimedia.org/mailman/listinfo/xmldatadumps-l>) for general dumps questions
- wikitech mailing list (<https://lists.wikimedia.org/mailman/listinfo/wikitech-l>) for broader technical discussions

- [Tools](#) for working with the dumps
- [License](#) for text content
- [More \(https://dumps.wikimedia.org/legal.html\)](https://dumps.wikimedia.org/legal.html) license information

- [Phabricator project \(https://phabricator.wikimedia.org/maniphest/task/edit/form/1/?projects=dumps-generation\)](https://phabricator.wikimedia.org/maniphest/task/edit/form/1/?projects=dumps-generation) for bug reporting (requires [account](#))
- [wikimedia-tech irc channel \(https://webchat.freenode.net/?channels=#wikimedia-tech\)](https://webchat.freenode.net/?channels=#wikimedia-tech) for real-time chat, time zones permitting
- [Help with common import issues](#)
- [Help in other languages](#)

## Contributing

- [XML/sql dumps code \(https://gerrit.wikimedia.org/r/plugins/gitiles/operations/dumps+/master\)](https://gerrit.wikimedia.org/r/plugins/gitiles/operations/dumps+/master) besides [MediaWiki core \(https://gerrit.wikimedia.org/r/plugins/gitiles/mediawiki/core/\)](#)
- [Puppet code \(https://gerrit.wikimedia.org/r/plugins/gitiles/operations/puppet+/production/modules/snapshot/files/cron/\)](https://gerrit.wikimedia.org/r/plugins/gitiles/operations/puppet+/production/modules/snapshot/files/cron/) for other dumps
- [Technical docs](#) for the dumps
- [Developer docs](#) for the dumps
- [Contributing to Wikimedia repos](#)
- [Generating dumps yourself](#)

## FAQ, further reading

- [Dumps FAQ](#)
- [Wikipedia dumps help page](#)
- [Wikidata dumps information](#)
- [More...](#)

---

Retrieved from "[https://meta.wikimedia.org/w/index.php?title=Data\\_dumps&oldid=20357699](https://meta.wikimedia.org/w/index.php?title=Data_dumps&oldid=20357699)"

---

This page was last edited on 12 August 2020, at 08:12.

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. See Terms of Use for details.