CORE Dataset

Download millions of research outputs for text and data analysis and process it directly in your own infrastructure.

REGISTER NOW

All the data in one single place
Download all the CORE data in a single package.

Matching your needs
Prototype, analyse and process your data directly on your infrastructure.

Largest full text collection
World's largest full text collection of scientific papers for machine processing
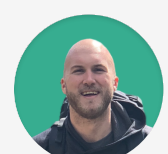
Simple to use
Accessible and easy to understand documentation and processes.

How it works?

CORE data can be downloaded as a bulk dataset, allowing you to process it on your own computer or within your infrastructure. The dataset provides a harmonised and enriched data format for access content from across our data providers. This is perfect for prototyping new methods, especially when intensive data processes need to be run. It is also a good choice for data analysis and text mining.

ACCESS DOCUMENTATION

Eric Olson
Consensus, co-founder and CEO

"To build the product we have always envisioned, having a robust and comprehensive dataset of machine-readable, peer-reviewed papers is absolutely essential. We are incredibly grateful to be able to partner with an organization like CORE that not only can meet our data needs, but also shares our vision of making science more accessible and consumable. This unique combination of best-in-class data-offering and mission-alignment makes CORE an ideal partner for Consensus."
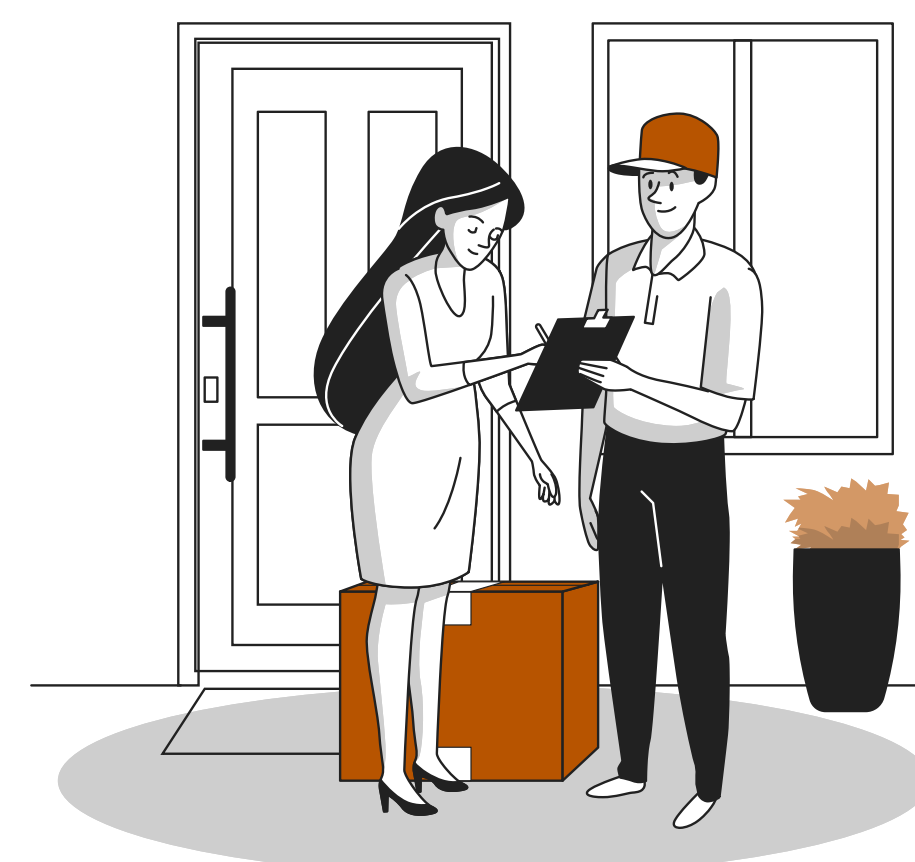
How often is the CORE dataset updated?

We aim to generate a new public dataset at least once a year. If you need a more recent dataset, please get in touch with us as we might be able to arrange it.

What's included?

The dataset provides you with:

- The entire CORE's corpus of both metadata and full texts in a machine processable format.
- Mappings of the CORE articles to entities in the Microsoft Academic Graph (MAG), enabling to access CORE fulltexts and use additional entities from MAG where available.
- Detailed documentation on how to download the CORE dataset and how data is organised.

Latest			
dataset	datasets	datasets	datasets
2022	**2021**	**2020**	**2019**
year	year	year	year

https://core.ac.uk/

theteam@core.ac.uk