

## **Europe PMC overview**

Europe PMC is an open repository of research publications, supporting the Open Access policies of 29 international funders of life sciences research, including several that have joined cOAlition S. The repository is built in collaboration with the PMC archive in the USA, and contains over 5 million full text articles and 35 million abstracts. Incoming full text articles are shared between the two sites daily. Europe PMC also recently started indexing preprint abstracts.

## **Europe PMC services**

The content in Europe PMC is made available via the website (<http://europepmc.org>). In addition to providing access to the core publications, Europe PMC adds value in a number of ways. For example, Europe PMC is a major integrator of ORCIDs, data, open citations, text-mined concepts and data citations, and grant information. All these integrations are applied to incoming content automatically and all are available via the website and APIs (see below).

Through indexing rich metadata and full text XML, it is possible to search transparently for full text articles in Europe PMC with a CC-BY license, and filter these by funder, publication date, presence of a data availability statement, and so on. Remixing open data allows Europe PMC to, for example, show open access publishing records for ORCIDs (see <http://europepmc.org/authors/0000-0002-3897-7955> for example).

A key part of Europe PMC's mission is to encourage innovation based on open access content. We therefore provide programmatic access to abstracts and full text via APIs, including RESTful (JSON, XML and Dublin Core) and OAI. Bulk download by FTP is also provided (<http://europepmc.org/developers>). Sharing content as widely as possible is our top priority.

Europe PMC also runs a grants metadata database for the funders (<http://europepmc.org/grantfinder>), so that incoming publications can be matched to grants.

Europe PMC is part of the global life science data infrastructure. We collaborate and share content with the USA National Library of Medicine, which runs PubMed and PMC. Europe PMC is an ELIXIR Core Data Resource and provides integration with over 40 critical data resources, such as the Protein Data Bank, the European Nucleotide Archive, UniProt and OMIM.

## **Technical considerations**

Europe PMC is built on the widely used data standard for research publications, JATS (<https://jats.nlm.nih.gov/index.html>). While Europe PMC has a life science focus, JATS is not discipline specific and therefore could be used for any research article.

Use of JATS (1) future-proofs the archive for long-term access (proven adherence to standards via validation against the data model and no reliance on proprietary formats) and (2) provides the ability to query across all content in a consistent manner. This is very important for deep queries, third party software development, and text and data mining. JATS, being an XML format, allows specific and important elements of an article (e.g.

ORCIDs, licences or article sections such as Data Availability Statements) to be identified. This kind of accurate deep indexing would be impossible for articles archived in an unstructured mixture of formats (PDF, Word, HTML).

### **Uploading publications**

Content is ingested into Europe PMC via two routes (1) journals that archive content in PMC (either in full or in part); and (2) via the Europe PMC or PMC manuscript submission systems. The manuscript submissions are overseen by Helpdesk staff who are trained in how to use the submission system and support authors in the submission process. Simply put, authors upload files (final accepted manuscripts, typically in Word or PDF), which, after various checks for integrity are converted to XML via a contract. The returned XML-formatted files are rendered to HTML for QA and sign-off by the contact author. It is possible to hold the article securely until such a time it can be made available, for example in the case that the article is submitted prior to an embargo date. The Helpdesk staff also handle incoming grant data from funders, in order to match incoming submissions with specific grants. These grant data are made public via the Grant Finder tool on the Europe PMC website, and via public APIs.

Europe PMC has recently been collaborating with the CoKo Foundation, Hindawi Publishing and eLife to develop an open source submission system for manuscripts. This will be released as a beta version very soon, and in full production shortly thereafter.