

BioStudies database: aggregating all outputs of a life sciences study

BioStudies is a database at EMBL-EBI that facilitates transparency and reproducibility of research, by aggregating all the outputs of a study (a 'data package') in a single place.

Time

Data generation

- Linking data in community archives
- Support for data without a community archive
- Support for project data sharing and transition to publication

Article publication

- Direct deposition
- Integration with journals (aspiration)

Post-publication

- Import from Europe PMC
- Makes data citable and findable independent of the article

Data rescue

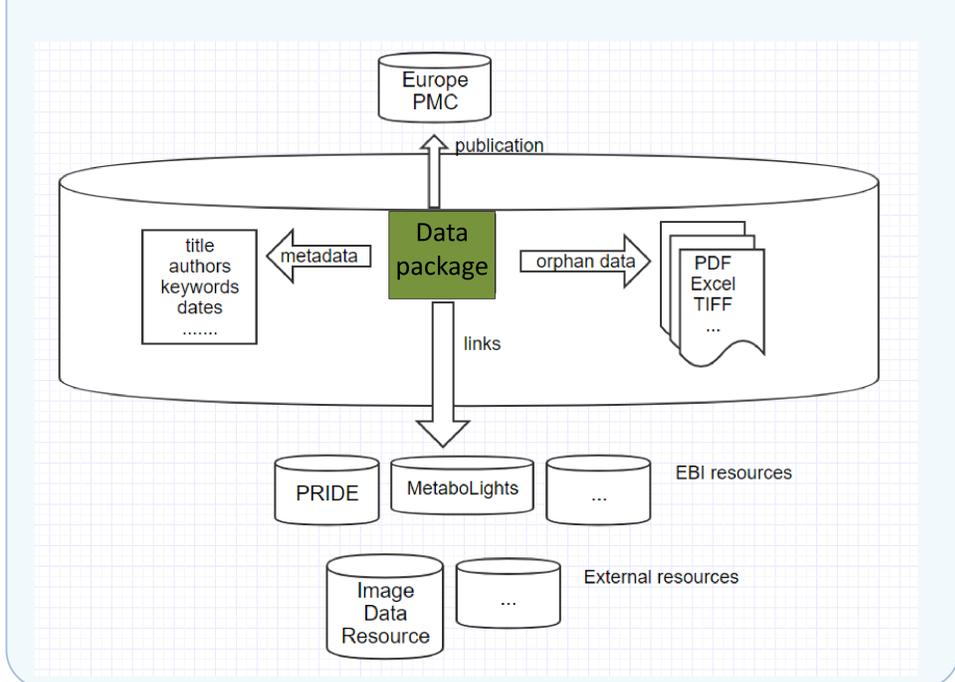
- Home for large datasets from retired projects
- Cost-efficient data sustainability

Better for: reproducibility, open science, FAIR data

Data submissions

- Automatic ingest pipelines
 - Europe PMC
 - SOURCE DATA
- Programmatic submission of JSON, XML and TSV
 - A simple data format for flexibility and fast prototyping
- Online web submission tool

Generic data model



Data access

- Text search
- Unique accessions
- Ability to share studies privately
- Downloads available in a number of formats
- Data files and links
- Metadata
- Data searchable via facets

Some future developments

- More data packages created pre-publication:
 - Links from paper to data can be created
 - Mechanisms to ensure better annotation for supplementary data, specific for journal/community/project

BioStudies as a pilot infrastructure for light microscopy data management, providing capabilities of visual representations for images and easy means to explore and download large datasets