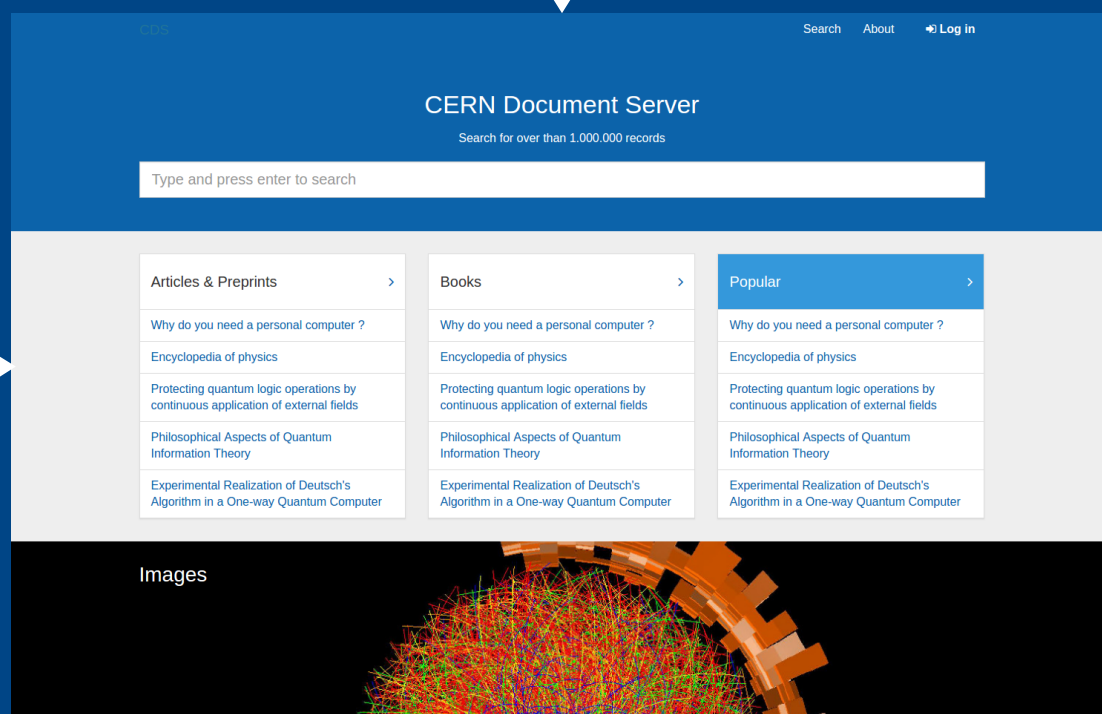# Publication Life Cycle at CERN Document Server

Ingestion

Data is ingested from multiple sources through various ways:

- 100+ submission workflows created during 15 years of CDS running – each tailored to meet the specific needs of a given user group.
- Single- and multi-records upload – "power tools" for catalogers to submit the desired metadata and files directly.
- Automatic import from other systems: via OAI-PMH protocol (e.g. arXiv.org) or import APIs (e.g. the CERN Experiments, digitisation projects, video lecture recordings).

Enrichment

- Automatic extraction of plots and captions, cross-reference with information from other collections or other systems (Inspire, arXiv), automatic DOI minting.
- Review process, in which the internal documents, after commenting rounds and revisions, are published.
- Authors disambiguation project using machine learning algorithms that enhanced the metadata of 67,000 records with more than 2,700,000 author signatures.
- Recommendation system suggesting relevant or similar documents based on previous users' interactions with CDS.

Dissemination

- Records can be exported in one of 7 formats: BibTeX, MARC, MARCXML, DC, EndNote, NLM and RefWorks.
- Support for open standards: oEmbed, OpenGraph, OpenSearch, OpenURL.
- Each collection of records defines it's own RSS channel.
- Users can set up notifications based on previous search queries.
- Videos can be exported to YouTube or embedded on any other website.
- Public CDS records can be harvested through OAI-PMH protocol.