

CERN Dataset

The dataset currently consists of:

- 60306 Image Files
- 17424 Records
- 1125 Authors
- 13 Collections
- 25 Copyright-Holders
- 6018 Keywords
- 16 Subjects

Data types and their fields:

Image:

- id
- comment
- description
- eformat (type of file, most will be ".jpg;icon-1440")
- full_name (original file name)
- localFile (path to file in local filesystem hierarchy)
- record id

Record:

- id
- title
- authors (list of *Author* ids)
- collections (list of *Collection* ids)
- copyrightDate
- copyrightHolder (*CopyrightHolder* id)
- copyrightStatement
- files (list of *Image* ids)
- keywords (list of *Keyword* ids)
- subjects (list of *Subject* ids)

Author:

- id
- first name
- last name
- full name
- records (list of *Record* ids)

Keyword and Subject:

- id
- term
- records (list of *Record* ids)

CopyrightHolder:

- id
- name
- records (list of *Record* ids)

Collection:

- id
- primary
- secondary
- records (list of *Record* ids)