

Data-Oriented EPrint Records

Dataset

- A one-off, independent collection of data.
- Contains one (or possibly more) data documents and probably README information to explain how to interpret the data document.
- Metadata reflects the collection and processing of the data (e.g. a range of contributors instead of a creator)

Experiment

- A structured process that may result in many datasets, intermediate analyses and final results.
- Contains various data documents and README information. The experimental methodology/trial protocol/observational process can be described in a normal descriptive text.
- Metadata reflects the experimental activities and context.

Data-Carrying EPrint Records

Publication

- Any publication-style eprint journal article,
 conference paper, report etc
- Publications may contain an supplemental data document with associated README information.
- Metadata reflects the publication, not the data gathering (e.g. list of authors rather than research technicians)

How A Data EPrint Appears

Citation

Carr, L and Takeda, K (2011) Weather Station Data: Roof of Southampton University. [Dataset] (Unpublished)



Generic Spreadsheet (Comma Separated Values) - Data Available under License <u>Creative Commons Attribution No Derivatives</u>. <u>Download (7Mb)</u>

List of data documents



Plain Text - README file
Available under License Creative Commons Attribution No Derivatives.

Download (8Kb)

Projects

[19] Intelligent Agents for Home Energy ManagementOfficial URL

Funders

[1] Engineering and Physical Sciences Research Council

Attributed projects & sponsoring funders

Item Type: Dataset

Subjects: G Geography. Anthropology. Recreation > GB Physical geography

Divisions: Faculty of Engineering, Science and Mathematics > School of Civil Engineering and the Environment

Depositing User: Unnamed user with email tdb2@ecs.soton.ac.uk

Date Deposited: 08 Apr 2011 15:44 Last Modified: 08 Apr 2011 15:44



Data Document Entry

Data documents can be attached to any any sort of eprint record (e.g. a journal article).

Format data explains what application is needed to render or edit the file contents.

Either:

README data is used to explain the layout of the information in the data document (*e.g.* how to interpret the row/columns in a spreadsheet).



Data Document Entry

Data documents can be attached to any any sort of eprint record (e.g. a journal article).

Format explains what application is needed to render or edit the file contents.

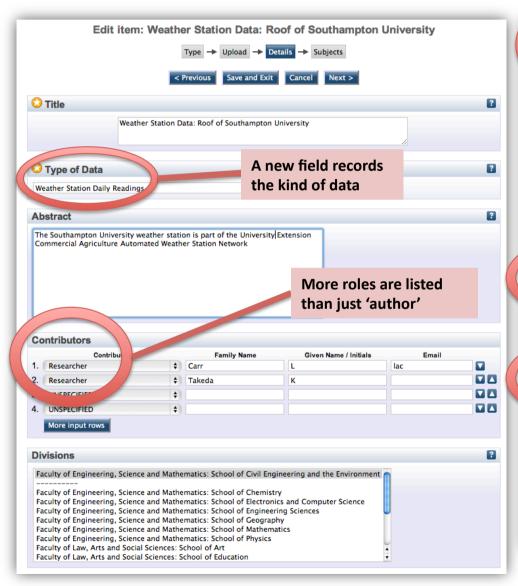
Or:

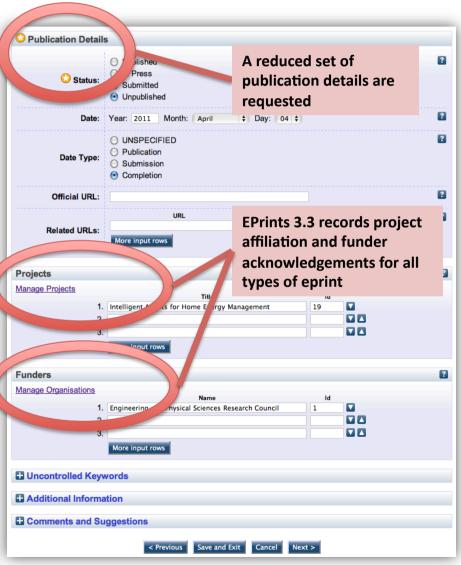
A separate **README file** is used to explain the layout of the information in the data files (*e.g.* how to interpret the row/columns in a spreadsheet).

Extended Range of Document Content

| draft | an early, unfinished or working version of a document |
|--------------|---|
| submitted | the version that was provided to an external agency for publication or evaluation |
| accepted | the version (with corrections and updates) that was accepted by the external agency |
| published | the "version of record", possibly with external agency updates and visual branding |
| updated | a version that has been updated post publication (ie post "version of record") |
| redacted | a version that has had sensitive material removed to allow dissemination |
| supplemental | material that supplements the main document |
| presentation | a specific kind of document used to give a presentation |
| coverimage | a visual representation of the material described in the EPrints record |
| metadata | additional metadata not covered by the EPrints schema |
| data | information provided for subsequent processing; e.g. a spreadsheet, database table |
| readme | an explanation of the contents or the interpretation of other documents |
| discussion | documents that contain discussions, commentary, review or response material |
| other | none of the above |
| other | none of the above |

Depositing a "Data" EPrint





Range of Data Contributors

Data Analyst

External Consultant

Data Contributor

Data Manager

Field Director

Geographic Information Specialist

Laboratory

√ Laboratory Director

Principal Investigator/Originator

Project Director

Photographer (Image Technician)

Experimental Process Contact

Programmer

Researcher

Research Team Head

Research Team Member

Data eprints don't have creators or authors, they have contributors.

The range of supported contributions is taken from the **MARC metadata** standard.

There are actually over 250 allowed contributor roles, but only a relevant subset are shown in the data depositing form.

Policies about Pragmatic Simplicity

- Tempting to get very explicit about subject data classification and control
 - this is a responsibility of data gathering/analysis, not archiving
 - An IR subject taxonomy cannot be augmented for data taxa
 - This is impractical there are hundreds just for archaeology!
- Keywords attribution should mainly be done by data capturing tools
 - it's too late by the time the repository is involved.
 - may be extracted from files (see the DepositMO project)
- Structured metadata can be provided
 - As an explicit document whose content is of type metadata
 - As a set of linked data relations (URIs)

Policies about Pragmatic Simplicity (2)

- Policy is that the metadata fields are HUMAN oriented and advisory.
- Explicit explanations can be given (should be!)
 provided in the README file(s) both of
 equipment used and processes undertaken.
 - Although they could be represented as complex metadata, they are unlikely to be successful in attracting any popular use
 - − Perhaps in 2 years ©
- EPrints will support the Open Provenance Model.