Nuclear Instruments and Methods in Physics Research B 284 (2012) 10–14

Contents lists available at SciVerse ScienceDirect

Nuclear Instruments and Methods in Physics Research B

journal homepage: www.elsevier.com/locate/nimb

MPOD: A Material Property Open Database linked to structural information

Giancarlo Pepponi [a,*], Saulius Gražulis [b,c], Daniel Chateigner [d,e,f]

[a] MiNALab, CMM-irst, Fondazione Bruno Kessler, Via Sommarive 18, 38123 Povo, Trento, Italy
[b] Department of Protein – DNA Interactions, Vilnius University, Institute of Biotechnology, Graiciuno 8, LT-02241 Vilnius, Lithuania
[c] Department of Mathematical Computer Science, Vilnius University, Faculty of Mathematics and Informatics, Naugarduko 24, LT-03225 Vilnius, Lithuania
[d] Université de Caen-Basse Normandie, UMR 6508 CRISMAT, F-14032 Caen, France
[e] ENSICAEN, UMR 6508 CRISMAT, F-14050 Caen, France
[f] CNRS, UMR 6508 CRISMAT, F-14050 Caen, France

ARTICLE INFO

Article history:
Available online 16 September 2011

Keywords:
Material properties
Tensorial properties
Open database
CIF

ABSTRACT

Inspired by the Crystallography Open Database (COD), the Material Properties Open Database (MPOD) was given birth. MPOD aims at collecting and making publicly available at no charge tensorial properties (including scalar properties) of phases and linking such properties to structural information of the COD when available. MPOD files are written with the STAR file syntax, used and developed for the Crystallographic Information Files. A dictionary containing new definitions has been written according to the Dictionary Definition Language 1, although some tricks were adopted to allow for multiple entries still avoiding ambiguousness.

The initial set includes mechanical properties, elastic stiffness and compliance, internal friction; electrical properties, resistivity, dielectric permittivity and stiffness, thermodynamic properties, heat capacity, thermal conductivity, diffusivity and expansion; electromechanical properties, piezoelectricity, electrostriction, electromechanical coupling; optical properties; piezooptic and photoelastic properties; superconducting properties, critical fields, penetration and coherence lengths.

Properties are reported in MPOD files where the original published paper containing the data is cited and structural and experimental information is also given. One MPOD file contains information relative to only one publication and one phase. The files and the information contained therein can also be consulted on-line at http://www.materialproperties.org.

1. Introduction

The Crystallography Open Database, (COD) [1] collects and provides Crystallographic Information Files (CIF) [2] for phase structure definitions. Born in 2003, in 2008 it had already reached the milestone of 50,000 entries, and it currently hosts more than 140,000 files. The Predicted Crystallography Open Database (PCOD), an early fork of the COD providing predicted structures, can boast already over 1,000,000 files.

Crystallographic structures affect microscopic properties of the phases constituting a given material which has consequences at the macroscopic scale. Knowledge about the structure can also be used to make predictions about such properties, e.g. with a simulation approach like Density Functional Theory. Measured crystallite properties, however, are of key importance for counterchecking theoretical studies and also to better define fitting models for the structure determination itself.

Inspired by the success of the COD and PCOD and feeling the strong interconnection and complementarity between structural and functional information, the Material Properties Open Database was given birth in 2010. The link and motivation stands out in particular for tensorial properties that need structural information for their representation, in particular point group symmetry, otherwise at disposal in the famous Landold–Bornstein book series. Anisotropic materials properties are usually represented by tensors or matrices [3,4], and we used the property names as defined in [5].

CIF files are written according to the STAR file specifications [6,7]. Given the success and widespread use of CIF files and the connection of the material properties with the structural information it was decided to adopt the same format for MPOD files. CIF files in fact are recognised and used by most crystallographic software but also by other software where structural information is of key importance e.g. solid state simulation software that can predict material properties ab initio. The usage of the same syntax could then favour the implementation and usage of MPOD files in software taking into account not only structure but also other properties of the materials. Currently the database contains about 200 files with data related to over 40 different properties. In the MPOD files details about the experimental conditions at which the properties were measured are also given. The MPOD files are available

* Corresponding author.
E-mail address: pepponi@fbk.eu (G. Pepponi).

0168-583X/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.nimb.2011.08.070

for download but the information can also be viewed directly in a more reader-friendly style.

2. File structure and dictionary

The STAR file syntax and specifications were designed quite rigidly, specifically for the description of crystallographic properties and diffraction experiments. In the STAR/CIF design and concept a piece of information (data item) is given by the association of a tag (data name) and a data value. Data items are grouped in data blocks which are contained in a datafile, written in simple ASCII text to be human readable.

Data names are defined in a dictionary. A data file can then be checked for consistency against one or more dictionaries. The core CIF dictionary [8] is written specifically for single-crystal small-molecule and inorganic crystal studies; it contains however also general data names as for examples related to the article where the structure was published.

A new dictionary has been written to define the properties, experimental conditions and parameters not present in the CIF core dictionary and needed for the definition of the material properties and their measurements. MPOD datafiles are then checked against both the CIF core dictionary and the MPOD dictionary. Below an example of property definition taken from the MPOD dictionary is shown.

```
data_prop_piezoelectric_gij
_name                    '_prop_piezoelectric_gij'
_category                prop
_type                    char
_units                   C.m^-2
_units_detail            'coulomb per square metre'
_definition
;     piezoelectric gij
ij = 11-36
;
```

The STAR File grammar does not allow the repetition of a data name within one data block. For lists of data items the loop_ structure is foreseen. Nesting of loop_ structures is not allowed. Only data names that belong to the same category can be looped together. For this reason all properties and also experimental conditions and methods belong to the same category, ''prop'', in the material properties dictionary, even if initially they were divided into different categories for clarity.

A typical example is shown below for elastic stiffness values defined in the single crystal frame (1,2,3) and represented as a matrix. The values can be inserted in a loop_ structure where a data value takes into account the change of index:

```
loop_
_prop_data_tensorial_index
_prop_elastic_stiffness_cij
11 106.75
12 60.41
13 60.41
22 106.75
23 60.41
33 106.75
44 28.34
55 28.34
66 28.34
```

In this example one can see that for diagonally symmetric properties (tensors), as are elastic constants, we choose not to repeat the tensor components below the diagonal. Otherwise all the constants are explicitly written, so that software reading the files do not need to care about tensor symmetry among the components. A ''?'' is given for any missing value not reported in the source scientific paper.

Let us now consider the case where many properties have been determined in the same experiment and published in the same paper. In order to comply with the STAR syntax without repetition of any tag (_prop_data_tensorial_index for instance), and simultaneously to allow the insertion of all properties (with different index ranges in the general case) in the same loop, we chose to assign a label to a property and repeat the label in the loop_ structure. An example of simultaneous report of elastic stiffness and internal friction values is given below (dots stand to avoid a long list of constants):

```
_prop_measurement_method 'EMAR'
_prop_elastic_stiffness_cij 'cij'
_prop_internal_friction_Qij-1 'Qij-1'
loop_
_prop_data_label
_prop_data_tensorial_index
_prop_data_value
_prop_conditions_temperature
cij 11 229.0 290
...
cij 66 85.8 290
Qij-1 11 2.7 290
...
Qij-1 66 5.4 290
cij 11 235.9 5
...
cij 66 89.6 5
Qij-1 11 0.4 5
...
Qij-1 66 2.3 5
```

This choice gives rise to a reader-friendly list that can be extended also for changing experimental parameters (e.g. in this case the temperature), whereas in this example the method is indicated outside the loop structure since in this case it does not change for any of the property values reported.

3. Web interface

A web interface was created for searching and displaying the information contained in the MPOD files and it is available online at http://www.materialproperties.org. In the future mirrors are foreseen to warranty data availability with time, together within the COD actual sites. There is no restriction to the access to data, which can be reused in any other software and database for free.

The site provides general information about the MPOD, the dictionary and references to the STAR file and CIF file structure, syntax and core dictionary. Moreover it offers the possibility of searching files according to the name of the phase, the elements it contains, the author of the article where the data was originally published, or the code of the phase in the Crystallography Open Database. Fig. 1 shows a snapshot of the search web page after it was interrogated with the phase name aluminum: the MPOD code, filename, COD code, phase generic name, phase name, chemical formula and publication are indicated for each entry selected by the query. The MPOD code is linked to a page showing the properties and their values for the file. The filename is linked to the actual MPOD file, the COD code to the CIF structure file of the COD database, and finally there is a link to the details about the article where the data was originally published. A snapshot of the page with the article information is published in Fig. 2.

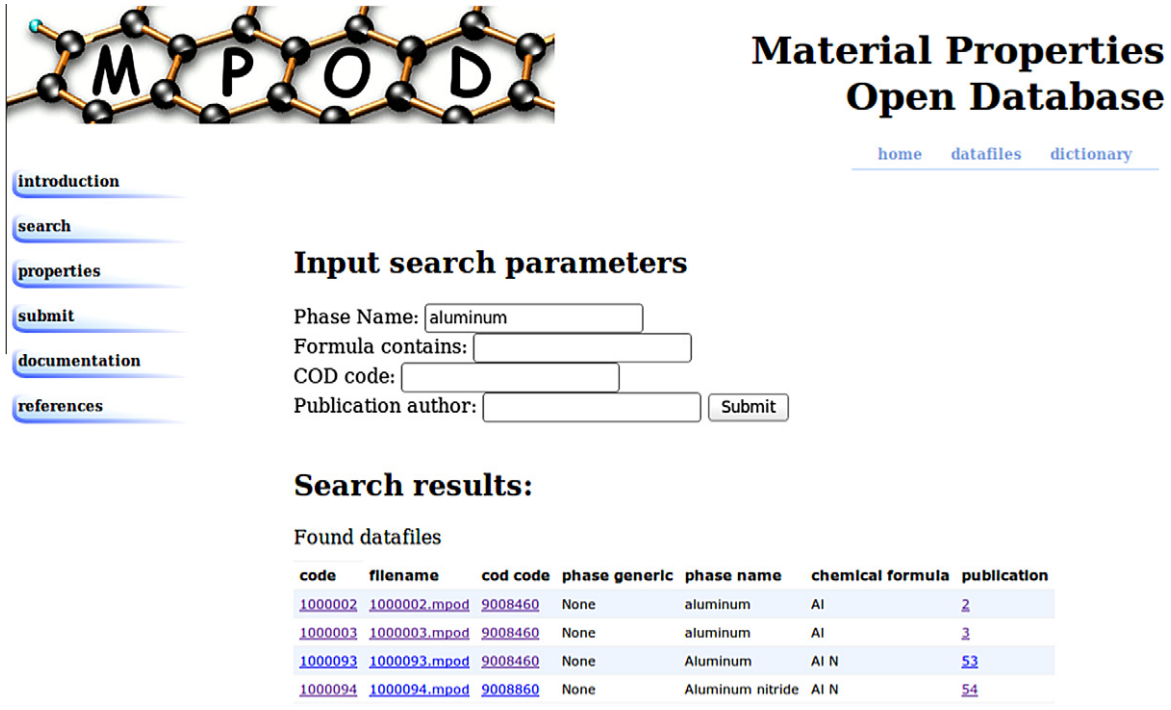

Fig. 1. Snapshot of the search page of the MPOD website after a search with the keyword aluminum.

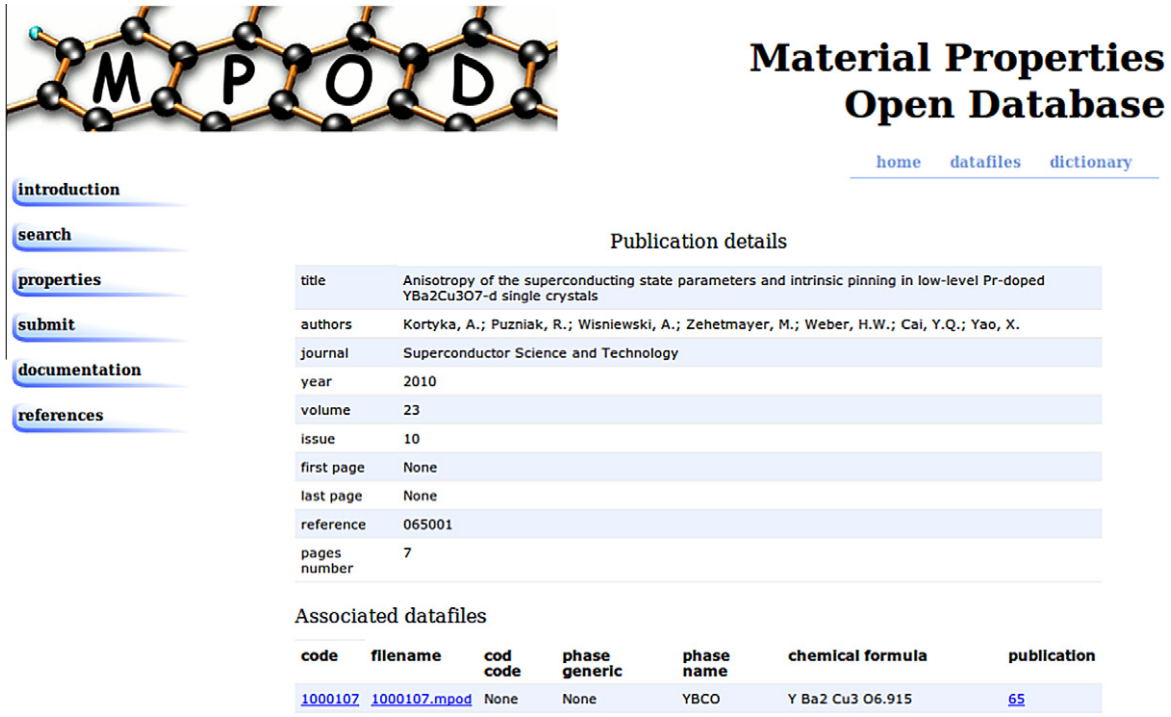

Fig. 2. Snapshot of the article details page with the list of datafiles containing values originally published in the article.

In Fig. 3 the visualisation of a MPOD datafile with the matrix representation of elastic stiffness values is shown; the property name is linked to a detailed view of the property and a list of the datafiles where values related to this property are reported as shown in Fig. 4. On the website, the matrix values are reported below the general datafile information and experimental conditions but they were placed on the right side of the picture for size reduction.

4. Summary and further implementations

Over 40 property data names have already been defined. An indicative list is given in the abstract, the full list being easily checkable on the MPOD site. A significant set of experimental parameters and measurement methods are included in the dictionary. The website provides not only the files but also the dictionary and information on the software used for the syntactic correctness

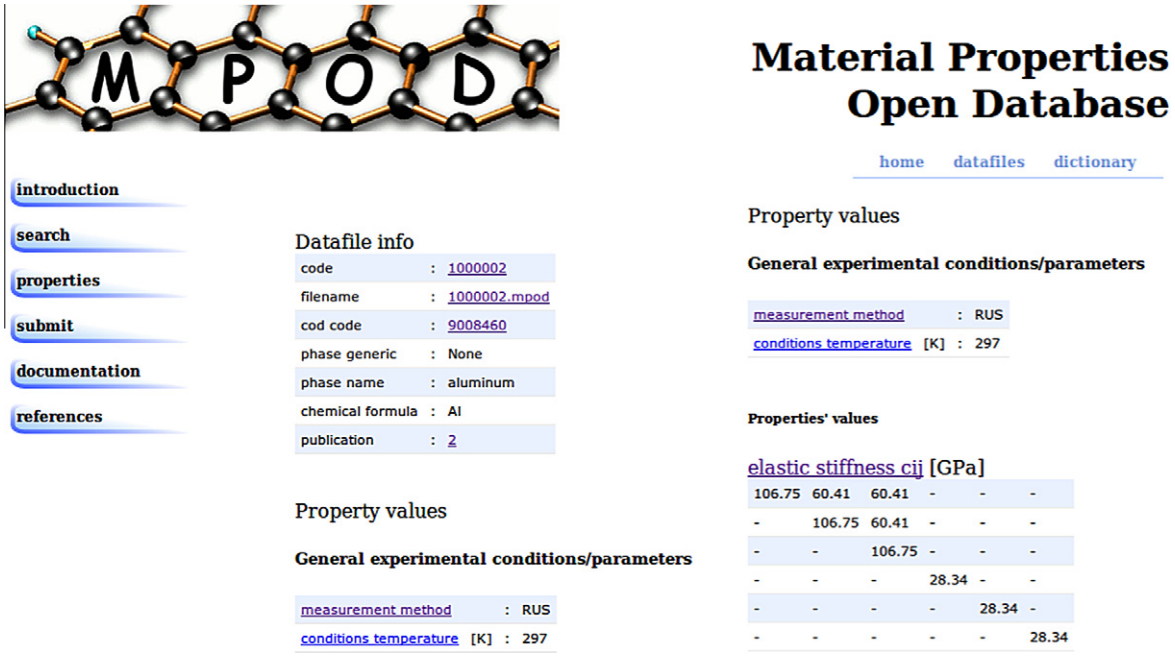

Fig. 3. Snapshot of the datafile view where property values for a given file are displayed in matrix form.

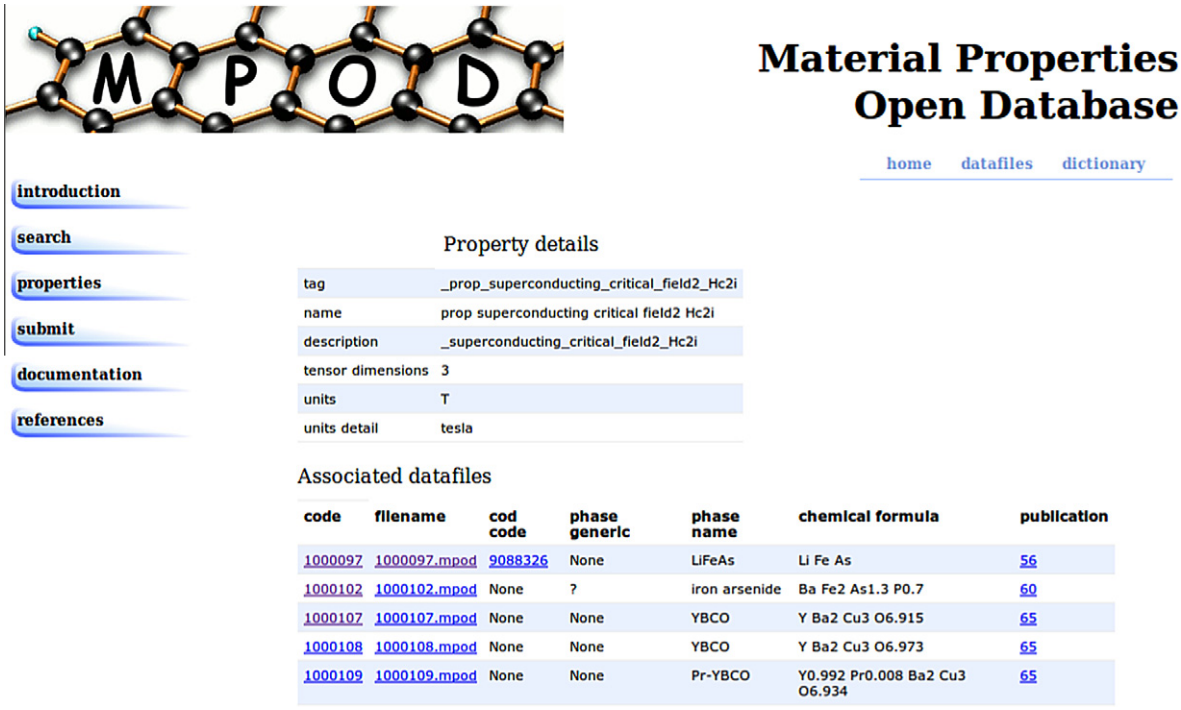

Fig. 4. Snapshot of the properties details page with the list of all files containing values related to the property (in this case superconducting critical field).

of the files. Further validation algorithms will be implemented in the near future taking into account the validity of the matrix indices indicated and the labels for the properties. A search tool helps consulting the properties.

A basic form for submitting MPOD files is available. Putting together MPOD files from publication is a tedious work if done repeatedly by few people but it is matter of a few minutes for scientists after they have accomplished publication. The advantage for the authors is visibility of their work and data that can be searched through the web interface of the MPOD. The website in fact provides reader-friendly views of the properties values, with links to the properties' definitions and the details of the publication where the data are released.

Acknowledgements

This work was partially carried out within the Xmat project ("Combination of X-Ray diffraction and X-Ray Fluorescence techniques in material science"), supported by the Provincia Autonoma di Trento and the European Union in the framework of the Marie

Curie COFUND program - Call for proposals 4 - researcher 2009 - Outgoing.

References

[1] S. Grazulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quiros, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, A. Le Bail. J. Appl. Cryst. Vol. 42 (2009), p. 726–729, 10.1107/S0021889809016690.
[2] S.R. Hall, F.H. Allen, I.D. Brown, Acta Cryst. A47 (1991) 655–685.
[3] J.-F. Nye, Physical Properties of Crystals: Their Representation by Tensors and Matrices, Lavoisier Publishers, 1985. p. 330.
[4] A. Sirotine, M. Chaskolskaia, Fondements de la Physique des Cristaux, MIR Publishers, Moscow, 1984. p. 695.
[5] D. Chateigner, Combined Analysis, Wiley-ISTE, 2010. p. 496.
[6] S.R. Hall, J. Chem. Inf. Comput. Sci. Vol 31 (1991) 326–333.
[7] S. R. Hall, A. P. F. Cook, in: S. R. Hall, B. McMahon (Eds.), International Tables for Crystallography Volume G: Definition and Exchange of Crystallographic Data, International Union of Crystallography, 2006 (Chapter 2.5).
[8] Information on http://www.iucr.org/resources/cif/dictionaries/cif_core.