Research Article

On the Archiving and Visualisation of Scientific Data

J. W. Nolan*[,1], D. A. Gkika[1], N. Vordos[1], I. K. Kazanidis[1], A. Ch. Mitropoulos[1,2]

[1]Hephaestus Laboratory, Eastern Macedonia & Thrace Institute of Technology, Ag. Loukas, 65404, Kavala, Greece.
[2]Department of Petroleum & Mechanical Engineering, St. Lucas, 654 04 Kavala, Greece

___

Abstract

Porous materials are important in gas separations and catalysis. Processes involving the use of porous materials have significant potential for replacement of traditional and expensive industrial processes. This publication details the development of a comprehensive database for organising information related to porous materials. The data includes structural information on the material, its properties and the experimental data obtained for the material in the laboratory, using a wide variety of characterisation techniques. The rational behind the development, the database structure, and a technique for visualisation of the data are presented, as are examples of using the database.

Keywords: database, porous material, data visualisation

___

1. Introduction

The advent of the internet over two decades ago provided the facility to move Materials Databases from the then traditional handbook or excel file, to the online database where disseminating, searching and updating is quick and easy. An internet search on the term 'materials database' returns a significant number of links to databases that can provide mechanical engineers or designers with specifications on modern materials for designing and building structures. For example, *MatWeb* [1] is a free-to-access properties database on over 100,000 materials; this covers the details of mechanical, physical, chemical properties, crystal structure and phase transitions and a host of other properties associated with the materials. The *MatDat* [2] database contains over 1,500 mechanical properties datasets on materials, including alloys. The *AZoM* [3] database covers advanced materials for the worldwide engineering and design communities. Other databases include the *Engineering Materials Database,* which contains data on engineered alloys at different temperatures and radiation levels (for application in radioactive environments) according to international standards. The National Institute for Materials Science hosts the *NIMS MatiNavi Database* [4] which contains links to other collections of information on materials. More specific databases include the *Nesshy-DB* which stores hydrogen sorption and related laboratory data on specific materials, the *HTR-Fuel-DB* with data on materials for use as high temperature reactor fuel and similarly, the *HTR-Graphite-DB* for graphite. These databases are hosted by the *Joint Research Centre Institute for Energy and Transport* [5].

In the area of nanotechnology, the *Nanowerk Nanomaterials Database*, an online resource for researchers in the community, contains information on over 2,500 nanomaterials and lists hundreds of suppliers of nanomaterials by material type; i.e. nanoparticles, nanofibres, nanotubes, etc. [6]. Other online resources in the

area of nanomaterials include the Nanomaterial Registry [7], the *NanoPartikel Data and Knowledge on Nanomaterials* portal [8], and the Project of Emerging Nanotechnologies Nanotechnology Consumer Products Inventory [9].

In the area of porous materials, the *Database of Zeolite Structures* was developed on behalf of the *International Zeolite Association*'s *Structure Commission (IZA-SC)* [10] and was originally based on data from the *Atlas of Zeolite Structure Types* [11]. This includes information, such as space group, cell parameters, volume, framework density, topological density and ring sizes, on 218 different framework types. The database is generally limited to structural information on the materials.

The *Hephaestus Laboratory* at the *Eastern Macedonia & Thrace Institute of Technology* has a history of research on porous materials [12-15], particularly the characterisation of such materials. The Nanocapillary Database started life as a deliverable in a Greek/EU project (begun in 2012) by the same name [16]; initially the database was to be developed for collecting and cataloging Small Angle X-ray Scattering (SAXS) data generated in our laboratory for porous materials, but has since evolved into a more comprehensive database on the properties of micro- and mesoporous materials; it incorporates data collected on the materials from various laboratory instruments, which importantly for the laboratory, will provide open access to the data, and sets a solid foundation for introduction to the area of Big Data analysis and correlation.

Porous materials are classified according to IUPAC notation [17]; microporous materials are those with pore diameters less than 2nm. Mesoporous materials on the other hand have larger pores, with pore diameters between 2nm and 50nm. Zeolite, an aluminosilicate mineral, commonly used as an adsorbant or catalyst, is a well-known example of a microporous material. Porous glass and gels are examples of disordered mesoporous materials but synthetic mesoporous materials, such as MCM-41 [18] and SBA-15 [19], have been developed with a narrow pore distribution

and ordered pore arrangement. One of the most important properties of porous materials is their high surface area; a challenge is to increase surface area, because, for example, in the area of catalysis, an increase in surface area leads to an increased number of reactive sites. This in turns leads to lower reaction temperatures and the need for smaller quantity of catalyst.

A vast amount of research is being performed daily on porous materials, and a significant amount of that work is most likely generating data that already exists. This constant stream of data could be used as in Big Data [20], to be monitored, analysed and utilised by the scientific community, for the benefit of the scientific community, and society in general.

2. Requirements

The Nanocapillary Database is open access; users can register to use the database, and are then able to access its contents free of charge. The database users are mainly expected to be scientists interested in accessing properties or results on porous material. Users should be able to upload new data and modifying existing data. Therefore, interaction with the database through the frontend needs to be well structured and user friendly; during interaction with the database, the simplest form of interaction sees the user selecting a material from a list of materials stored in the database. The data available for that selected material is then displayed in an easily understood and user-friendly manner.

The fundamental requirement for the database is the creation, retrieval, update and deletion (CRUD) of information on materials. The user also needs to be able to upload and modify data pertaining to specific scientific characterisation instruments in the Hephaestus Laboratory. These are typically in the form of *ascii* data files, *jpg* and.*png* images. Where appropriate, experimental setup, and reference data is also considered for inclusion in the database. Due to the potentially large file sizes produced in Scanning Electron Microscopy and Transmission Electron Microscopy, it is not necessary to store all data in the database; in this case, links to files on the server are stored in the corresponding table's fields. The database and associated software for CRUD and data visualisation should be open source.

3. Implementation

The open source MySQL database was chosen for database development, with PHP selected as the preferred scripting language for performing the CRUD and visualisation operations. In MySQL, as all other types of database, the data is organised as ordered lists of elements grouped, in this case according to their relationship, using the relational database model [21]. The database and frontend are currently undergoing internal testing in the institute before public release in the next quarter.

3.1 Database Schema
The database, for which a simple representation of the schema is shown in figure 1, has been developed to store information on the basic material properties, synthesis recipe, relevant publications on the material, the costs involved in producing the material, and the experimental data collected from a number of instruments in the

laboratory, such as Scanning Electron Microscope, Transmission Electron Microscope, and Small Angle X-ray Scattering.

The main table at the top of the table hierarchy is *material*. The *recipe* table contains fields for description of the synthesis process and the *raw materials* used. The *publications* table is linked to tables for *publication type* and *people;* the authors. The *manufacturer's* table contains fields for standard information, such as *name*, *phone*, *address*, etc. The *people* table, which is used multiple times, contains fields for *name*, *email*, and *phone*. The *costs* table contains fields for the cost of *raw materials*, *services*, *person hours,* and *equipment costs.* The most complicated of the tables is for storing the *data_files,* which depending on the *data_type* links to fields for linking to either images or *ascii files.*

All tables consist of a primary id key used to reference the data in the table. An example of one of the research data schemas, for archiving SAXS data, is shown in figure 2. This schema includes tables for the *data* itself, the *data type* (for example, this can be the characterisation data itself, data for a standard samples used for calibration, mask data, etc.) and tables for logging the experimental setup, such as *saxs_experiment* which maintains data on the wavelength of the radiation used, the distance between the SAXS detector and the sample, the pinholes used for collimation of the X-ray beam, etc. This is standard data normally recorded in a log book, but is now part of an electronic log; a *Data Capture System* has been developed to automatically capture and archive laboratory data, and log the experimental setup. This system is described in a separate publication currently being prepared.

In addition, the schema includes tables for handling database users. There are currently two *usertypes: administrator* and *standard user.* New users can register in the database through the main site. These are automatically emailed a confirmation code and are then capable of accessing existing data and submitting a request to upload new data or modify existing data.
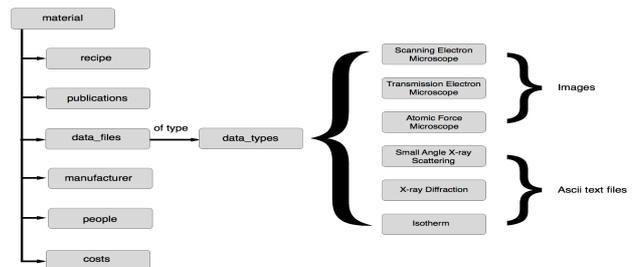

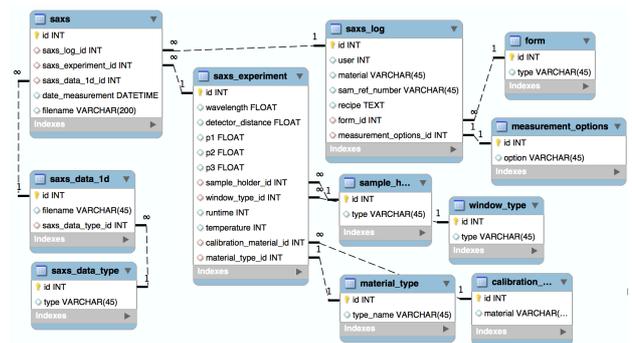**Fig. 1**. Simplified form of Nanocapillary dB schema

Fig. 2. Database schema for cataloguing SAXS data

3.2 Data Input

A web interface frontend has been developed for inputting and retrieving data from the database. This has been written using a combination of HTML, PHP and JavaScript. An example of one of the data input forms is shown in figure 3, for inserting basic information on the porous material into the database. To meet the open access requirement, all users can input new data and modify existing data. However, to ensure the quality of the data, this is done through a Quality Layer, in which the quality of the data is examined according to predefined standards by a dedicated committee. Final upload and modification of existing data is done by System Administrators.

Fig. 3. Form for inputting data on materials

3.3 Data Retrieval and Visualisation

The interface has been developed using the HTML5 canvas feature for displaying data, in combination with PHP and MySQL for accessing the database. The jQuery JavaScript (JS) Library is used to make Asynchronous JavaScript and XML [22, 23] (Ajax) calls to PHP files which interface with the database through SQL queries [24], as illustrated in figure 4. The returned data is displayed on a canvas using the Fabric JS canvas library [25].
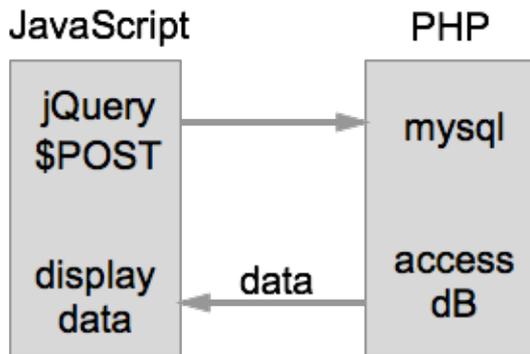

Fig. 4. Illustration of JavaScript and PHP interaction through Ajax

Screenshots of the web interface for retrieval and visualisation of the data for the SBA-15 porous material stored in the database is presented in figure 5. Search methods have been developed in PHP to dynamically iterate through the data available on a specific material, and graphically display the data for the user. Unavailable data is omitted from the displayed content. As illustrated in the figure 5, when a user clicks on an icon representing a material on the *materials.php*, the selected material is passed to *material.php* in which a joined MySQL search query is used to determine the availability of data under different categories, such as recipe, properties, related publications, research data, etc. The data display is setup and displayed through a JavaScript file using Fabric JS.
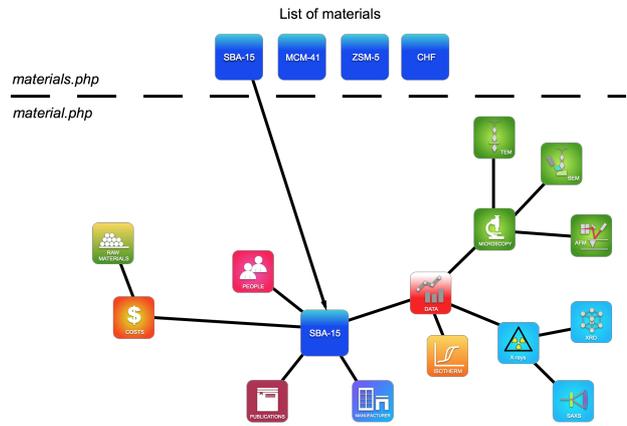

Fig. 5. Web interface for viewing the specific information on a selected material from the dB

The presented data is user-interactive; the user can click on any of the icons representing specific data, and obtain a detailed overview of the items available at that level on the data hierarchy. In addition to the visual representation of data in the database, the user has the option to search the database by inputting search terms into a form, with filters for the types of data available in the database.

4. Using Nanocapillary Database

An example of the typical procedures followed when using the database is illustrated in figure 6. On the home page, the user selects a material by clicking on an icon representing the material (click A). On the next page, the user then has a visual overview of the data available for that material. The user can then obtain more information on for example, at click B, the recipe for synthesising the material, at click C, Transmission Electron Microscope Images, or click D, publications relevant to the material.
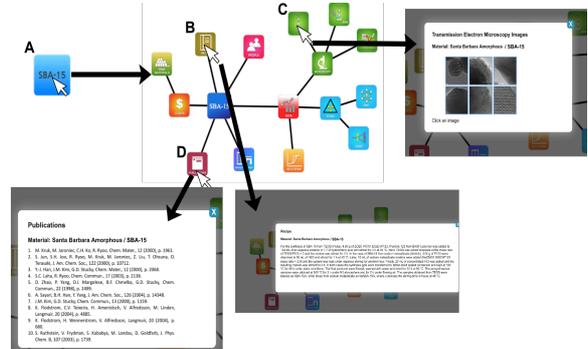

Fig. 6. Illustration of typical dB usage

5. Conclusions

Materials' databases constitute an important resource for engineers during the design and construction of buildings and other structures. Despite the vast amount of information, including research data, that is generated and published yearly on porous materials, no equivalent open access database exists in the scientific community for porous materials. This is a missed opportunity for the scientific community; such data could be monitored and analysed using advanced algorithms. The present article provides details on a database for archiving such data and a frontend for visually presenting the data to the user in a clear and user-friendly manner.

_____

References

1. Matweb Material Property Data at http://www.matweb.com
2. Material Properties Database at http://www.matdat.com
3. The A to Z of Materials at http://www.azom.com
4. Materials Database Group of NIMS at http://mits.nims.go.jp/
5. Online Data & Information Network of the Institute for Energy and Transport at https://odin.jrc.ec.europa.eu
6. Nanowerk at http://www.nanowerk.com
7. The Nanomaterial Registry at www.nanomaterialregistry.org
8. The Data and knowledge on nanomaterials (DaNa2.0) website at http://www.nanopartikel.info
9. Nanotechnology in the real world: Redeveloping the nanomaterial consumer products inventory, M. E. Vance, T. Kuiken, E P. Vejerano, S P. McGinnis, M. F. Hochella Jr., D. Rejeski, M. S. Hull, Beilstein J. Nanotechnol. 6, 1769–1780 (2015).
10. Ch. Baerlocher and L.B. McCusker, Database of Zeolite Structures: http://www.iza-structure.org/databases/
11. W. M. Meier, D. H. Olson and Ch. Baerlocher, Atlas of Zeolite Structure Types, 4th edition (1996).
12. "Preparation, characterization and gas permeation properties of carbon hollow fiber membranes based on Matrimid (R) 5218 precursors", E.P. Favvas, G.C. Kapantaidakis, J.W. Nolan, A. Ch. Mitropoulos and N.K. Kanellopoulos, Journal of Materials Processing Technology, 186, 102-110 (2007).
13. What is spinodal decomposition? EP Favvas, AC Mitropoulos, Journal of Engineering Science and Technology Review 01 (2008).
14. Mixed matrix hollow fiber membranes with enhanced gas permeation properties, Evangelos P. Favvas, Konstantinos L. Stefanopoulos, John W. Nolan, Sergios K. Papageorgiou, Athanasios Ch. Mitropoulos, Didier Lairez, Separation and Purification Technology 08 (2014).
15. Effect of air gap on gas permeance/selectivity performance of BTDA-TDI/MDI copolyimide hollow fiber membranes, Evangelos P. Favvas, Sergios K. Papageorgiou, John W. Nolan, Konstantinos L. Stefanopoulos, Athanasios Ch. Mitropoulos, Journal of Applied Polymer Science, 130(6) (2013).
16. The project's website can be found at http://nanocapillary.teikav.edu.gr.
17. Recommendations for the characterization of porous solids (Technical Report), J. Rouquerol, D. Avnir, C. W. Fairbridge, D. H. Everett, J. M. Haynes, N. Pernicone, J. D. F. Ramsay, K. S. W. Sing, K. K. Unger, Pure and Applied Chemistry, Volume 66, Issue 8, Pages 1739–1758 (1994).
18. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism, C.T. Kresge, M.E. Leonowicz, W.J. Roth, J.C. Vartuli, J.S. Beck, Nature 359, 710 - 712 (1992).
19. Direct Syntheses of Ordered SBA-15 Mesoporous Silica Containing Sulfonic Acid Groups, D. Margolese, J. A. Melero, S. C. Christiansen, B. F. Chmelka, and G. D. Stucky, Chem. Mater., 12, 2448-2459 (2000).
20. http://ec.europa.eu/digital-agenda/en/news/worldwide-big-data-technology-and-services-2012-2015-forecast
21. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 1.0, 11 December 2013.
22. Ajax: A New Approach to Web Applications, Jesse James Garrett (2005). http://www.adaptivepath.com/ideas/ajax-new-approach-web-applications/
23. AJAX and PHP: Building Modern Web Applications 2nd Edition, A. Hendrix, B. Brinzarea, C. Darie, Packt Publishing, December 2009.
24. SQL Queries for Mere Mortals: A Hands-On Guide to Data Manipulation in SQL 2nd Edition, J. Viescas, M. J. Hernandez, Pearson Education, September 2014, US.
25. http://fabricjs.com