This purpose will be illustrated by two examples. The disorder previously evidenced either by the splitting of atoms over different sites in tunnels of $Sr_6(AlO_2)_{12}Bi_2O_3$ [1] or the large ADP of oxygen atoms of $CsMPO_4$ [2] (Fig. 1) can be elucidate just by observing satellite reflections (Fig. 2) and then solving the modulated structure [3] using superspace formalism.

[1] V. Bakakin, N. Podberezkaya, M. Mazus, L. Kozeeva, V. Shaburova, T. Malinovskii, *J. Struct. Chem.* **1994,** *35*, 229–235 [2] O.V. Yakubovich, M.A. Simonov, O.K. Mel'nikov, *Kritallographiya* **1990,** *35*, 42-43 [3] O. Pérez, S. Malo, M. Hervieu, *Acta Cryst* **2010,** *B66*, 585-593.

Keywords: aperiodicity, disorder

MS.88.5

Acta Cryst. (2011) A**67,** C193

Relationship of the aperiodic structure to the nanostructure

Nguyen Van Tri, *Institut of Engineering Physics, Hanoi University of Science and Technology, Hanoi, (Vietnam).* E-mail: nvtri@mail.hut.edu.vn

This paper aims to evidence that some typical aperiodic crystal materials, such as the composite and the modulated ones, show a *nanostructured material and just this nanostructure conditiones the unique properties of the aperiodic material.* The fragments, as the "basic units", that make up the material are nanoscaled and *aperiodically* arranged with respect to each other. To the preparation of such nanostructure, the consolidated and fashioned fragments (as the nanoparticles or nanolayers) can be composite or modulated in a bulk material. This process is widespread in biology with the variety of self-assembly [1].

The key problem is, the nanoworld is of *quantum nature* that creates the peculiar dynamics of the odd electrons (i.e. the active electrons, they show the unpaired spins) in the nanostructured material [2, 3]. And just this peculiar electron dynamics breeds the prominent features of the aperiodic material. For studying these effects, Electron Spin Resonance (ESR) can offer an especially efficacious help means.

From the experimental and theoretical results with ESR in combination with other methods, some peculiarities of dynamics of electrons in the real aperiodic materials can be summarized as follows:

1) The active electrons in the aperiodic structure can be considered as the Quasi-Free Electrons (QFEs) moving in a Nano Resonant Cavity (NRC) with nanoparticles or in a Nano Wave Guide (NWG) with nanowires and nanolayers of the nanoscaled Short-Range Order.

2) The QFEs must obey a "Compressed Distribution" as a normal Gauss form of the state density. This very outstanding peculiarity springs from the resonant character of the NRC and NWG and can strongly influence the ability of the aperiodic material to absorption and emission of radiation.

3) The NRC and NWG can essentially favour the strong super-exchange interactions of their QFEs. Thereby, two types of the spin coupled pairs (the Antiferromagnetic pair and the Ferromagnetic pair) can simultaneously occur and the [2A-2F] fivefold clusters can be spontaneously formed. This aperiodic structure can bring out an energy minimum, and therefore a stable existence of the fivefold clusters.

4) The spin coupling brings about different spin gaps of the QFEs in the NRC and NWG, thereby can cause the peculiar conductivity or the superconductivity of the aperiodic materials.

5) The interaction of the QFEs with the surroundings (Jahn-Teller effect) can generate a strong local crystal field of low symmetry in the aperiodic structure, escpecially in the aperiodic structure of the living body.

When the scale of the fragments (particles, wires or layers) extend out of the nanorange, the material becomes polycrystalline and its above mentioned peculiarities will disappear.

As illutrative examples, the results on *Composite Fivefold Aperiodic Structure of Dental Enamels* [4] and *Superconducting Nanomechanism in YBCO Compounds* [5] are briefly analysed.

[1] Ch.P. Poole Jr., J.F. Owens, *Introduction to Nanotechnology,* USA and CANADA, WILEY - INTERSCIENCE, **2003. [2]** E.L. Wolf, *Nanophysics and Nanotechnology,* WILEY – VCH **2005. [3]** N. Van Tri, S*pringer Proceedings in Physics, Springer – Verlag Berlin Heidelberg* **2009,** *127*, 51-60. [4] N. Van Tri, *Ferroelectrics, Gordon and Breach Science Publishers,* **2001,** *250,* 385-388. [5] N. Van Tri, *Ferroelectrics Inc., USA* **2004,** *305,* 141-145.

Keywords: aperiodic, nanostructured aperiodic, aperiodic structure and nanostructure

MS

MS.89.1

Acta Cryst. (2011) A**67,** C193-C194

Twenty years of the crystallographic information framework

Brian McMahon, *International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU (UK).* E-mail: bm@iucr.org

The Crystallographic Information File (CIF) format was adopted by the International Union of Crystallography (IUCr) as the standard for exchanging and archiving crystallographic information in 1991 [1] and has continued to be developed actively during the past two decades.

The initial specification used the Self-Defining Text Archive and Retrieval (STAR) format developed as a universal exchange mechanism [2] and subsequently used in a number of other scientific fields including botany, chemistry and NMR structure determination. Later developments including machine-readable dictionaries of data identifiers with associated attributes helped to separate the semantic content of CIF from its syntactic structure, allowing easy interchange with XML or other widespread formats. Work has continued to extend the dictionary definition language to include information about the permitted data content in a manner that computers can process directly. Latest work focuses on specifying algorithmic relationships between individual data items.

The result has been an information system that includes its own recipes for computational validation of related data items. This has allowed CIF to be used in electronic publishing and in database applications, where validation, integrity checking and format transformations can be performed automatically and with no or little information loss. Consequently, CIF underpins the entire publishing workflow of IUCr journal articles reporting crystal structures [3], and is an important intermediary in the management of the curated structural databases such as Protein Data Bank [4] and Cambridge Structural Database.

CIF separates semantics from syntax to a large degree, and has a rich granularity in its item descriptions. It adopts a uniform approach that treats experimental data and associated descriptive information (often called 'metadata') in the same way. It has come to be accepted as a standard across the entire field of crystallography that can handle raw image data, information about structure solution and refinement, the positional and displacement parameters of a derived structural model, and all aspects of an associated publication. In all these ways, it serves as a model for effective information flow – a model that is increasingly being used for case studies by other disciplines seeking to improve their data management strategies [5]. Predating the Web and XML, CIF remains at the leading edge of scientific information management systems.

[1] S.R Hall, F.H. Allen, I.D. Brown, *Acta Cryst.* **1991,** *A47*, 655-685. [2] S.R. Hall, *J. Chem. Inf. Comput. Sci.* **1991,** *31*, 326-333. [3] P.R. Strickland, B.

McMahon, *Acta Cryst.* **2008**, A*64*, 38-51. [4] J.D. Westbrook, H. Yang, Z. Feng, H.M. Berman, *International Tables for Crystallography* **2005**, *G*5.5, 539-543. [5] L. Lyon, *Consultancy Report* **200,** Bath, UK: UKOLN. http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

Keywords: CIF, electronic publishing, information management

MS.89.2

Acta Cryst. (2011) A**67,** C194

Data management for photon and neutron sources

Brian Matthews, *e-Science Centre, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot, OX11 0QX, (U.K.).* E-mail: brian.matthews@stfc.ac.uk

Photon and Neutron sources, such as the UK's Diamond Light Source and ISIS Spallation Neutron Source are large-scale facilities providing high resolution data for crystallography and other materials analysis techniques. Traditionally, the raw data generated from such facilities has been managed by the instrument and user scientists themselves. However, the current generations of such facilities can undertake a large number of experiments, and generate hugely increased volumes of data. As a consequence, the traditional approach has become unsustainable and a more automated approach to data management has had to be developed.

In this talk, I shall outline the data management infrastructure developed within STFC to manage raw data. This infrastructure takes an integrated approach to aggregate, store and catalogue data generated at ISIS and Diamond. In particular, I shall describe ICAT, a suite of tools which catalogues data as it is generated by beam lines, and provides access to that raw data to its user community, allowing them to search and retrieve their data, within the facilities themselves or within their home institution. This is provided using a service application programming interface so that a variety of different search and analysis tools can be interfaced to search and access the data, and also register and catalogue derived data.

The management of raw data is part of a wider scientific process, starting from proposals for research through to the publication of results. We shall further discuss how the ICAT and similar tools can be extended to support this wider process by allowing data to be federated across a number of different data sources and also linking the raw data to analysed and published data so that the provenance of data can be tracked; this is being considered in the project Integrated Infrastructure in Structural Sciences (I2S2). This allows data to be formally cited and reused, and results to be validated. We relate this work to the publication process being developed by the International Union of Crystallography, tracing the relationship between raw data generated from beam lines, and the CIF files lodged during the publication process.

This integrated data infrastructure is being taken forward by the European Photon and Neutron Data Infrastructure initiative (PaNData), a consortium of European photon and neutron sources serving an expanding user community of tens of thousands of scientists across Europe. The experiments in these facilities are of increasing complexity, they are increasingly done by international research groups and many of them will be done in more than one laboratory. The resulting data needs to be accessible over the Internet and remain on-line until the results are published and in many cases much longer to allow re-processing and to allow for the preservation of knowledge. PaNData is developing common data formats, data and software catalogues within the framework of a common data policy.

Keywords: data management, information management, large-scale facilities

MS.89.3

Acta Cryst. (2011) A**67,** C194

Crystaleye: Publication and re-use of open semantic crystallographic data

Peter Murray-Rust, *Unilever Centre for Molecular Sciences Informatics, Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, (UK).* E-mail: pm286@cam.ac.uk.

Berners-Lee's vision of the Semantic Web (SW) is now a reality in many scientific fields (including macromolecular structures and much bioscience). The SW is based on Linked Open Data (LOD) where each component of information is Openly available with a published unique identifier scheme. The LOD are linked together through RDF triples where the semantics are provided by published ontologies or dictionaries. This creates a graph (or "cloud") of data on web sites and in triple stores that can be explored by the current generation of SW tools.

In our Crystaleye system we have applied this approach to "small molecule" crystal structures (organic, inorganic and organometallic) by extracting Open Data from published CIFs, mainly on publishers' websites. The extraction is performed daily by our "Pubcrawler" system and any new CIFs are added and processed. Each CIF, split into individual data blocks if necessary, is automatically processed into semantic form (using Chemical Markup Language (CML) and RDF). During this process many validity checks are applied, in particular to extract and check the chemistry. After any reported disorder is processed the chemical connection table (CT) is created and checked against any reported formula and chemical names. The CT and compositional formula are then re-usable as primary indexes and search terms. All reported data in the CIF are translated to RDF, stored in our Chempound (chem#) repository where they can be searched through a SPARQL endpoint.

Where the full-text of the article is Open (as in Acta Cryst. E) we extract information from the text such as methods of preparation and crystallization as well as citations. This enhances the data in the CIF and creates a potentially valuable node in the LOD cloud. By comparing the deduced CT with the images and names in Acta E papers we show that the automatic generation of CTs has a precision/recall > 99%

Crystaleye (http://wwmm.ch.cam.ac.uk/crystaleye) provides a natural browsing interface to the crystal structure which includes interactive exploration and search. All bond lengths are indexed and can be searched by element types. Readers can link back to the original splah page and article if it is published on the web.

All data and software is fully Open (i.e re-usable for any purpose without further permission). Crystaleye, whose maintenance cost is near-zero, shows that it is possible to create a global knowledge base of crystallography simply by publishing CIFs to the open web and letting machines do the rest. The technology is also applicable to theses (which are currently under-used) and for departments to expose their unpublished data. Unfortunately restrictions imposed by some publishers and some data aggregators mean that the current coverage of Crystaleye is only partially complete. Besides the technology the presentation will address aspects of Openness in crystallography and low-cost approaches to sustainability.

Keywords: semantic, linkeddata, open

MS.89.4

Acta Cryst. (2011) A**67,** C194-C195

The wwPDB Working Format: A Simplified Application of CIF Technology

John Westbrook,[a] Helen M. Berman,[a] Jasmine Young,[a] Gerard J. Kleywegt,[b] [a]*RCSB PDB, Rutgers University, Piscataway, NJ (USA).*