

The eCrystals Model for Gathering and 'Publishing' Crystallographic Data

<http://wiki.ecrystals.chem.soton.ac.uk>

Simon Coles
EPSRC National Crystallography Service
School of Chemistry
University of Southampton

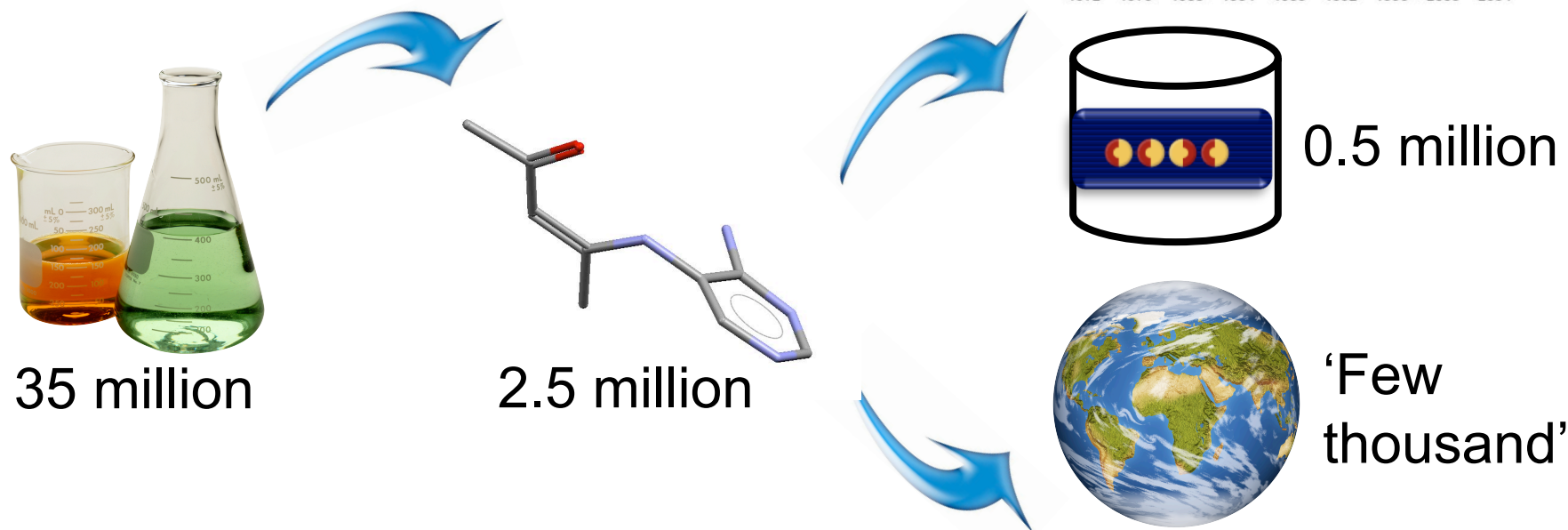
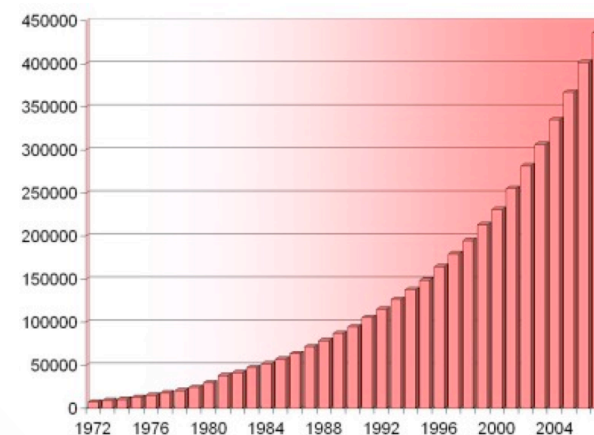


New Approaches to 'Publishing', 'Sharing' & 'Gathering' Data



Data Deluge

- 40 years ago a PhD student would determine about 3 crystal structures for their thesis – this can now be easily achieved in a day



The primary cause is the current data publication process, which is tied to journal articles and peer review

Current (Open?) Practice in Crystallography

To Share or not to Share: Publication and Quality Assurance of Research Data Outputs

Report commissioned by the Research
Information Network (RIN)

Annex: detailed findings for the
eight research areas

June 2008



www.rin.ac.uk

In association with:



CHEMICAL CRYSTALLOGRAPHY

Overview

Crystallography is the most-used unambiguous technique for identifying the structures of chemicals. This structural characterisation is done either for basic research purposes or as a service for other chemists who do not have their own crystallographic facility.

The chemical crystallography community has long had a relatively organised approach to data. Crystallographic data are highly-structured. Any heterogeneity resides in the instrumentation and the software this uses rather than in the data outputs. Outputs are in 3 or 4 possible formats but there is a *de facto* standard adopted by the bulk of the community, the CIF (Crystallographic Information File), and a standard protocol for making data available to others. Probably 95+% of crystallography data is in the form of CIF files.

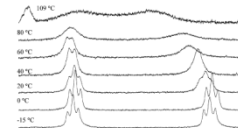
A CIF file represents derived data rather than raw data and chemists may derive further data from the CIF. Various players have made software available for generating and manipulating CIF files, including the Cambridge Crystallography Data Centre (CCDC) (see below for more on this organisation). Examples of such software are Mercury (from the CCDC)¹⁶, WinGX from Glasgow University¹⁷ and Olex, developed at Durham University¹⁸. All of them are free to the community. Additionally, researchers often write their own software to perform particular tasks. There have been people capable of writing software for crystallographic purposes in every major crystallography research group since the 1960s and frequently in smaller outfits everyone is a one-man-band, able to perform all roles in the data production process.

The major public funder of chemical crystallographic research in the UK is the Engineering and Physical Sciences Research Council (EPSRC), but a considerable amount of other crystallography work carried out in universities is paid for by pharmaceutical or chemical companies. This of course has some repercussions on what can be made publicly-available for sharing, but in general if a project is a purely academic one, even if funded by industry, then the funder places no barriers to public revelation of the data outputs. At the other end of the scale, some industry funded research is completely client-confidential, with the university crystallographic centres providing a bespoke service accordingly.

<http://www.rin.ac.uk/data-publication>

Intellect & Interpretation
(Journal article, report, etc)

Fig. 5 Examples of compounds featuring C–F...H–N interactions, giving scalar H...F coupling constants



$[\text{Na}(\text{EtO})_2\text{N}(\text{CN})\text{BAr}_2']][\text{Ar}^+ = \text{C}_6\text{F}_5$ and $\text{C}_6\text{F}_4\text{C}_6\text{F}_5]$ [5a]. Although dicyanide has two lone pairs and is therefore theoretically capable of binding to three or possibly four borane molecules, only the diborates were obtained, even in the presence of excess BAr_2' (Scheme 5). Colourless crystals of the sodium salts **4a** and **5a** were grown from diethyl ether but proved unsuitable for X-ray diffraction studies. Further reaction of the sodium salts with HgCl_2 in dichloromethane afforded the triyl derivatives, $[\text{Ph}_3\text{C}(\text{Hg})\text{N}(\text{CN})\text{BAr}_2']][\text{Ar}^+ = \text{C}_6\text{F}_5$ and $\text{C}_6\text{F}_4\text{C}_6\text{F}_5]$ [5b]. The structure of **5b** was confirmed by X-ray diffraction analysis. The triyls **5** eventually yielded orange crystals suitable for X-ray analysis which, however, once again turned out to be $[\text{C}_6\text{H}_5\text{IP-PBB}]$.

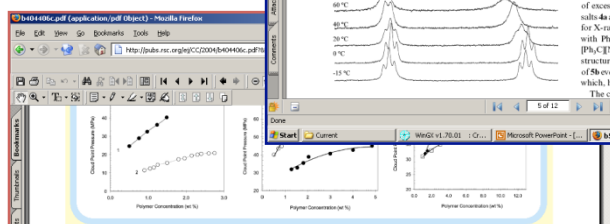
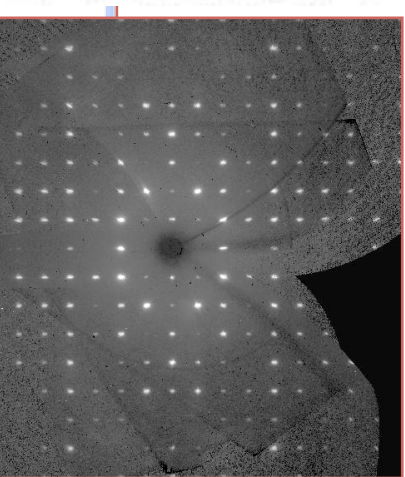
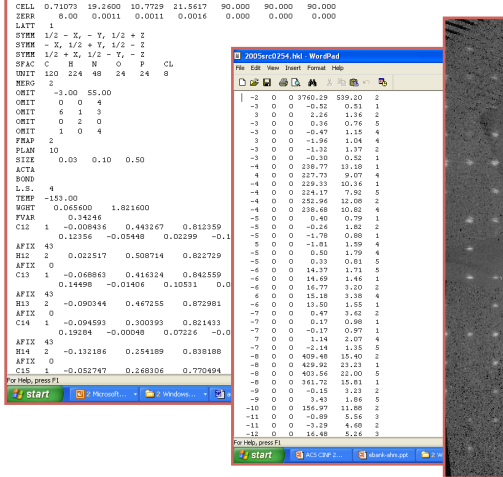
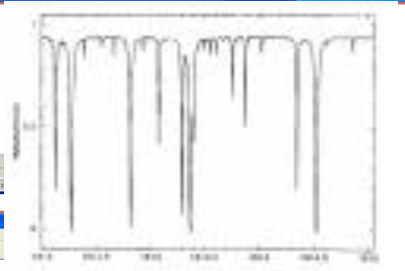
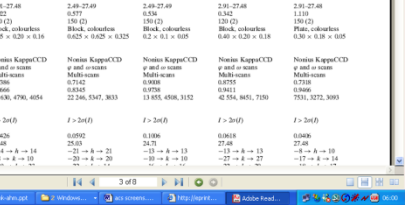
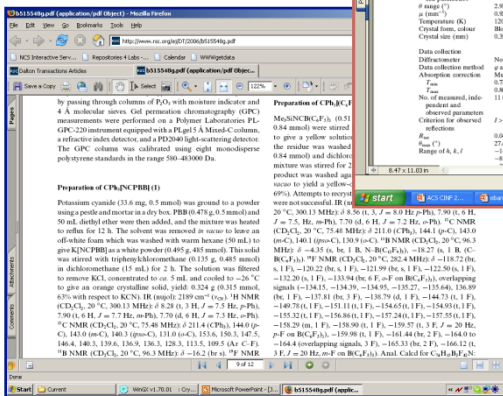
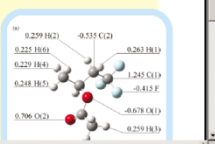


Fig. 2. (a) Phase behavior of (1) atactic polypropylene ($M_w = 425$), (2) polypropylene glycol monomethylether ($M_w = 1000$), in carbon dioxide at 295 K. (b) Phase behavior of oxygen-containing polymers in carbon dioxide at 295 K. (1) polypropylene glycol dimethyl ether, $M_w = 3500$, (2) polyvinyl methyl ether, $M_w = 3850$, (3) polyvinyl ethyl ether, $M_w = 3800$. (c) Phase behavior of oxygen-containing polymers in carbon dioxide at 295 K. (1) polyvinyl acetate, $M_w = 7700$, (2) polyvinyl ethyl ether, $M_w = 3800$, (3) polyvinyl acetate, $M_w = 3090$.

particular, we noted that very small changes to structure led to dramatic and unpredictable changes in phase behavior, a frustrating situation. For example, it has been known for almost a decade²¹ that polymethyl acrylate (PMA) and polyvinyl acetate (PVAc) exhibit miscibility pressures in CO₂ that differ by hundreds of bar—this result would not be predicted by any group contribution thermodynamic model currently in use without purely empirical adjustments. Indeed, the PMA/PVAc effect is preserved even when

We believe that at least part of the answer to these puzzles lies in CO₂'s ability to act on both Lewis acid and Lewis base, coupled with subtle effects of neighboring substituents on the acidity of certain protons. For example, Wallen and colleagues²⁷ in an analysis of interactions between acetate groups and CO₂ found that the acidity of the methyl acetate



The eCrystals Data Repository

- Quick & simple to deposit
- Software tools
- Laboratory archive
- Community involvement
- 'Embargo' facility
- Structured foundations
- Discoverable & harvestable

eCrystals
 Home | About | Browse by Year | Browse by People
 Login | Create Account

6,7,9,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin

Sample Originator: Esther Rousay* and Jeremy G. Frey*
 Data Collection: Simon J. Coles*
 Structure Determination: Simon J. Coles* and Michael B. Hursthouse*
 University of Southampton*

C14H20O5
 InChI=1/C14H20O5/c1-2-4-14-13(3-1)18-11-9-16-7-5-15-6-8-17-10-12-19-14/h1-4H,5-12H2

Identification Number: 10.3737/ecrystals.chem.soton.ac.uk/145
 Controlled Keywords: crown ethers, crown
 Date Created: 07 October 2004
 Deposited On: 21 Jan 2008 15:29
 Deposited By: Dr Simon J Coles

Available Files

Final Result
[04sjc0831.cif](#)
[04sjc0831.cml](#)
[04sjc0831.fcf.txt](#)

Collection parameters	
Chemical formula	C14 H20 O5
Crystallisation Solvent	
Crystal morphology	Plate
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	16.4963(18)
Cell length b	8.325(3)
Cell length c	20.061(6)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00
Data collection temperature	120(2)
Refinement results	
Solution figure of merit	0.0409
R Factor (Obs)	0.0487
R Factor (All)	0.0977
Weighted R Factor (Obs)	0.1008
Weighted R Factor (All)	0.1192
Solution	
04sjc0831.res	6k
04sjc0831.x1.lst	34k
Processing	
04sjc0831.psp	6k
04sjc0831.xls.lst	39k
04sjc0831.hkl	702k
04sjc0831.htm	10k
04sjc0831.0kl.jpg	57k
04sjc0831.h0l.jpg	85k
04sjc0831.hk0.jpg	88k
Data Collection	
04sjc0831.crystal.jpg	17k
Other Files	
04sjc0831.doc	
04sjc0831.ins	78k
04sjc0831.mol	5k
04sjc0831.p4p	3k
04sjc0831.pcf.txt	1k
04sjc0831.ellipsoid.gif	2k
	19k



<http://ecrystals.chem.soton.ac.uk>

A Thorough Approach to Dissemination

- Using simple Dublin Core protocol (OAI-PMH)
 - Crystal structure
 - Title (Systematic IUPAC Name)
 - Authors
 - Affiliation
 - Creation Date
- Additional **chemical** information through Qualified Dublin Core
 - Empirical formula
 - International Chemical Identifier (InChI)
 - Compound Class & Keywords
- Specifies which 'datasets' are present in an entry
- Application Profile <http://www.ukoln.ac.uk/projects/ebank-uk/schemas/>
- DOI links <http://dx.doi.org/10.1594/ecrystals.chem.soton.ac.uk/145>
- Rights & Citation <http://ecrystals.chem.soton.ac.uk/rights.html>

Data Harvesting

The image displays five screenshots from the CrystalWeb database interface, illustrating the search and data retrieval process for the compound 3-furoic acid.

Top Left Screenshot: Search results for "3-furoic acid". It shows sample origin (Donated Sample), data collection details (Susanne L. Huth and Michael B. Hursthouse), structure determination (Susanne L. Huth), and a list of available files (2007may0020.cif, 2007may0020.cml, 2007may0020.fcf, 2007may0020_checks.cif, 2007may0020.res).

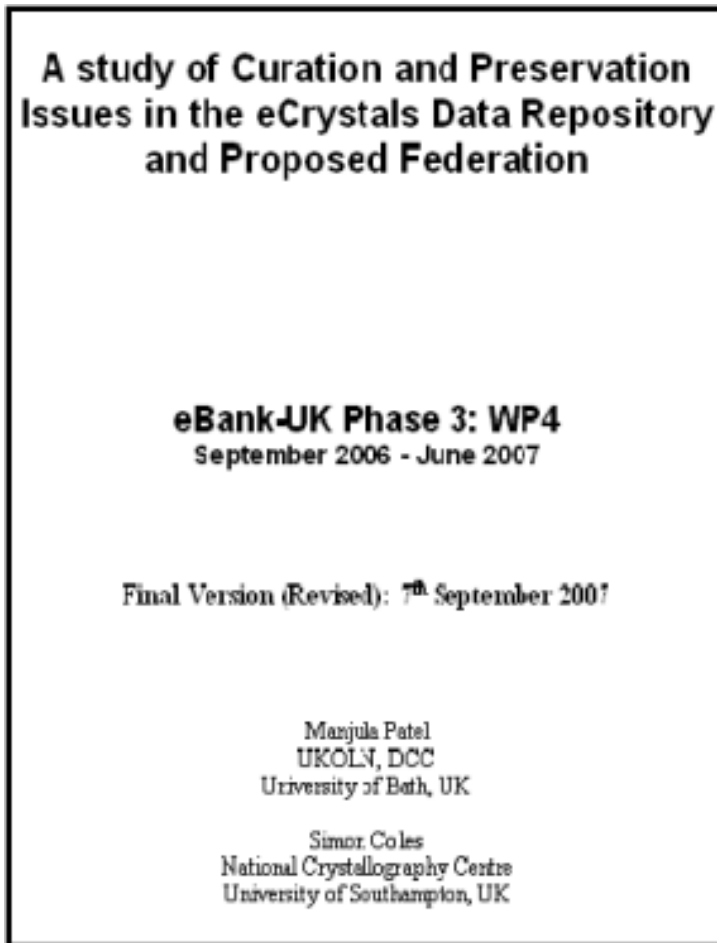
Top Right Screenshot: A detailed view of the "3-furoic acid" entry, showing the chemical structure, unit cell parameters (a = 10.1107 Å, b = 5.8094 Å, c = 16.223 Å, α = 90.0°, β = 91.1°, γ = 90.0°), and a 3D molecular model.

Bottom Left Screenshot: A "View Results" page showing a list of search hits (33 hits) and a 3D molecular model. The list includes entries like ADEP002, RESU001, CERWA001, FEYPAB01, SULAMD07, SULAMD08, SULAMD09, SULDAZ04, XIFJUS, XIFKAZ, XIFKED, XIFKH, XIFKOH, XIFKUT, XIFLAA, XIFLEE, XIFLII, XIFLI01, XIFLI02, XIFLOO, and XIFLUU.

Bottom Center Screenshot: A table showing the "Reference" and "Cell data" for the compound. The reference is "Sample, Donated, Huth, Susanne L., Hursthouse, Michael B., University of Southampton, Crystal Structure Report Archive, (2007)". The cell data includes lengths (a = 10.1078 Å, b = 5.8535 Å, c = 16.0142 Å) and angles (α = 90°, β = 90.624°, γ = 90°).

Bottom Right Screenshot: A "ChemSpider" entry page for "3-furoic acid". It shows the chemical structure, molecular weight (112.0835 Da), and a list of references (e.g., CHEBI:30846, CHEBI:30847, CHEBI:30848, CHEBI:30849, CHEBI:30850, CHEBI:30851, CHEBI:30852, CHEBI:30853, CHEBI:30854, CHEBI:30855, CHEBI:30856, CHEBI:30857, CHEBI:30858, CHEBI:30859, CHEBI:30860, CHEBI:30861, CHEBI:30862, CHEBI:30863, CHEBI:30864, CHEBI:30865, CHEBI:30866, CHEBI:30867, CHEBI:30868, CHEBI:30869, CHEBI:30870, CHEBI:30871, CHEBI:30872, CHEBI:30873, CHEBI:30874, CHEBI:30875, CHEBI:30876, CHEBI:30877, CHEBI:30878, CHEBI:30879, CHEBI:30880, CHEBI:30881, CHEBI:30882, CHEBI:30883, CHEBI:30884, CHEBI:30885, CHEBI:30886, CHEBI:30887, CHEBI:30888, CHEBI:30889, CHEBI:30890, CHEBI:30891, CHEBI:30892, CHEBI:30893, CHEBI:30894, CHEBI:30895, CHEBI:30896, CHEBI:30897, CHEBI:30898, CHEBI:30899, CHEBI:30900, CHEBI:30901, CHEBI:30902, CHEBI:30903, CHEBI:30904, CHEBI:30905, CHEBI:30906, CHEBI:30907, CHEBI:30908, CHEBI:30909, CHEBI:30910, CHEBI:30911, CHEBI:30912, CHEBI:30913, CHEBI:30914, CHEBI:30915, CHEBI:30916, CHEBI:30917, CHEBI:30918, CHEBI:30919, CHEBI:30920, CHEBI:30921, CHEBI:30922, CHEBI:30923, CHEBI:30924, CHEBI:30925, CHEBI:30926, CHEBI:30927, CHEBI:30928, CHEBI:30929, CHEBI:30930, CHEBI:30931, CHEBI:30932, CHEBI:30933, CHEBI:30934, CHEBI:30935, CHEBI:30936, CHEBI:30937, CHEBI:30938, CHEBI:30939, CHEBI:30940, CHEBI:30941, CHEBI:30942, CHEBI:30943, CHEBI:30944, CHEBI:30945, CHEBI:30946, CHEBI:30947, CHEBI:30948, CHEBI:30949, CHEBI:30950, CHEBI:30951, CHEBI:30952, CHEBI:30953, CHEBI:30954, CHEBI:30955, CHEBI:30956, CHEBI:30957, CHEBI:30958, CHEBI:30959, CHEBI:30960, CHEBI:30961, CHEBI:30962, CHEBI:30963, CHEBI:30964, CHEBI:30965, CHEBI:30966, CHEBI:30967, CHEBI:30968, CHEBI:30969, CHEBI:30970, CHEBI:30971, CHEBI:30972, CHEBI:30973, CHEBI:30974, CHEBI:30975, CHEBI:30976, CHEBI:30977, CHEBI:30978, CHEBI:30979, CHEBI:30980, CHEBI:30981, CHEBI:30982, CHEBI:30983, CHEBI:30984, CHEBI:30985, CHEBI:30986, CHEBI:30987, CHEBI:30988, CHEBI:30989, CHEBI:30990, CHEBI:30991, CHEBI:30992, CHEBI:30993, CHEBI:30994, CHEBI:30995, CHEBI:30996, CHEBI:30997, CHEBI:30998, CHEBI:30999, CHEBI:31000, CHEBI:31001, CHEBI:31002, CHEBI:31003, CHEBI:31004, CHEBI:31005, CHEBI:31006, CHEBI:31007, CHEBI:31008, CHEBI:31009, CHEBI:31010, CHEBI:31011, CHEBI:31012, CHEBI:31013, CHEBI:31014, CHEBI:31015, CHEBI:31016, CHEBI:31017, CHEBI:31018, CHEBI:31019, CHEBI:31020, CHEBI:31021, CHEBI:31022, CHEBI:31023, CHEBI:31024, CHEBI:31025, CHEBI:31026, CHEBI:31027, CHEBI:31028, CHEBI:31029, CHEBI:31030, CHEBI:31031, CHEBI:31032, CHEBI:31033, CHEBI:31034, CHEBI:31035, CHEBI:31036, CHEBI:31037, CHEBI:31038, CHEBI:31039, CHEBI:31040, CHEBI:31041, CHEBI:31042, CHEBI:31043, CHEBI:31044, CHEBI:31045, CHEBI:31046, CHEBI:31047, CHEBI:31048, CHEBI:31049, CHEBI:31050, CHEBI:31051, CHEBI:31052, CHEBI:31053, CHEBI:31054, CHEBI:31055, CHEBI:31056, CHEBI:31057, CHEBI:31058, CHEBI:31059, CHEBI:31060, CHEBI:31061, CHEBI:31062, CHEBI:31063, CHEBI:31064, CHEBI:31065, CHEBI:31066, CHEBI:31067, CHEBI:31068, CHEBI:31069, CHEBI:31070, CHEBI:31071, CHEBI:31072, CHEBI:31073, CHEBI:31074, CHEBI:31075, CHEBI:31076, CHEBI:31077, CHEBI:31078, CHEBI:31079, CHEBI:31080, CHEBI:31081, CHEBI:31082, CHEBI:31083, CHEBI:31084, CHEBI:31085, CHEBI:31086, CHEBI:31087, CHEBI:31088, CHEBI:31089, CHEBI:31090, CHEBI:31091, CHEBI:31092, CHEBI:31093, CHEBI:31094, CHEBI:31095, CHEBI:31096, CHEBI:31097, CHEBI:31098, CHEBI:31099, CHEBI:31100, CHEBI:31101, CHEBI:31102, CHEBI:31103, CHEBI:31104, CHEBI:31105, CHEBI:31106, CHEBI:31107, CHEBI:31108, CHEBI:31109, CHEBI:31110, CHEBI:31111, CHEBI:31112, CHEBI:31113, CHEBI:31114, CHEBI:31115, CHEBI:31116, CHEBI:31117, CHEBI:31118, CHEBI:31119, CHEBI:31120, CHEBI:31121, CHEBI:31122, CHEBI:31123, CHEBI:31124, CHEBI:31125, CHEBI:31126, CHEBI:31127, CHEBI:31128, CHEBI:31129, CHEBI:31130, CHEBI:31131, CHEBI:31132, CHEBI:31133, CHEBI:31134, CHEBI:31135, CHEBI:31136, CHEBI:31137, CHEBI:31138, CHEBI:31139, CHEBI:31140, CHEBI:31141, CHEBI:31142, CHEBI:31143, CHEBI:31144, CHEBI:31145, CHEBI:31146, CHEBI:31147, CHEBI:31148, CHEBI:31149, CHEBI:31150, CHEBI:31151, CHEBI:31152, CHEBI:31153, CHEBI:31154, CHEBI:31155, CHEBI:31156, CHEBI:31157, CHEBI:31158, CHEBI:31159, CHEBI:31160, CHEBI:31161, CHEBI:31162, CHEBI:31163, CHEBI:31164, CHEBI:31165, CHEBI:31166, CHEBI:31167, CHEBI:31168, CHEBI:31169, CHEBI:31170, CHEBI:31171, CHEBI:31172, CHEBI:31173, CHEBI:31174, CHEBI:31175, CHEBI:31176, CHEBI:31177, CHEBI:31178, CHEBI:31179, CHEBI:31180, CHEBI:31181, CHEBI:31182, CHEBI:31183, CHEBI:31184, CHEBI:31185, CHEBI:31186, CHEBI:31187, CHEBI:31188, CHEBI:31189, CHEBI:31190, CHEBI:31191, CHEBI:31192, CHEBI:31193, CHEBI:31194, CHEBI:31195, CHEBI:31196, CHEBI:31197, CHEBI:31198, CHEBI:31199, CHEBI:31200, CHEBI:31201, CHEBI:31202, CHEBI:31203, CHEBI:31204, CHEBI:31205, CHEBI:31206, CHEBI:31207, CHEBI:31208, CHEBI:31209, CHEBI:31210, CHEBI:31211, CHEBI:31212, CHEBI:31213, CHEBI:31214, CHEBI:31215, CHEBI:31216, CHEBI

A Thorough Approach to Preservation



Project Honeycomb: Intelligent Storage for Massive Data Volumes

Project Honeycomb is an innovative storage system that introduces a whole new level of intelligence and programmability to the category. To learn more about it we talked to Mike Davis, senior product manager for Honeycomb in the Storage Group.

The second level of programmability is what I'm most excited about — the realization of an intelligent storage system. Honeycomb can manipulate data on the fly as it comes in and out of the system, and that's where it really becomes strategic for a lot of customers and application partners.

There used to be a very clear distinction between what was above the wire and what was below the wire or what was executed in the application server and what was executed in the storage system. Honeycomb blurs that distinction.

Q: How does Honeycomb work?

MIKE: There are two elements to Honeycomb: a highly reliable Serial ATA-based storage system based on a clustered architecture and a rich extensibility framework for access and management of data. Embedded in the cluster is fully distributed high performance database technology that is aligned very well with the storage system.

Honeycomb's architecture is designed to reduce or eliminate bottlenecks and single points of failure that exist in legacy architectures. Built-in parallelism improves performance and provides outstanding performance for query as well as data I/O.

Q: What customer challenges does Honeycomb address?

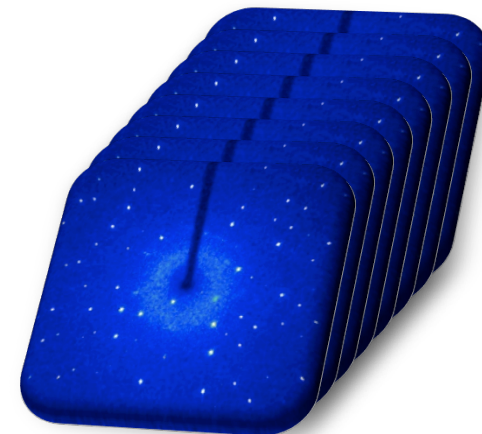
MIKE: Honeycomb was developed to solve continuing problems in the management of large-scale repositories, and by that I mean any large collection of unstructured data that tends to grow over time. This design concentrates on reducing administrative and service costs. We know that IT budgets tend to be static though data sets are growing exponentially, and our goal is to enable a single system administrator to manage a petabyte of storage.

We know that IT budgets tend to be static though data sets are growing exponentially, and our goal is to enable a single system administrator to manage a petabyte of storage.

Inexpensive storage hardware is easy to find, the trick is reducing the need to configure, repair, provision, service, migrate data, and other tasks that contribute to the true cost of the system.

Large repositories need failure resilience — the ability to suffer multiple types of failures without risk of losing data integrity. Through parallelism, Honeycomb can provide a level of reliability that is arguably better than what customers can get in any SAN environment.

The last major unsolved challenge for these customers is the management of metadata, the rich set of attributes that describe the data and allow it to be recalled instantly.



<http://www.ukoln.ac.uk/projects/ebank-uk/curation/>

So How Much Does it Cost?

APPENDIX 4- CASE STUDY : THE UNIVERSITY OF SOUTHAMPTON

BACKGROUND

Information to be considered as a case study for the cost model will be collected from the UK National Crystallography Service (NCS), The School of Chemistry and the National Oceanography Centre, Southampton (NOCS).

The National Crystallography Service

History

The National Crystallography Service has been operation since 1981, firstly at Queen Mary College London, then at the University of Wales College of Cardiff and since 1998 at the University of Southampton. For the whole period of its existence the NCS has always been housed within the respective Schools or Departments of Chemistry at these institutions and has been funded by a succession of research grants under both the rolling grant and responsive mode funding schemes of the EPSRC. The NCS provides an analytical service for UK chemists, based on state of the art experimental data collection facilities. This service includes the provision of raw data for those 'skilled in the art', who wish to work up a crystal structure themselves but don't have experimental facilities available to them or the provision of fully analysed crystal structures for chemists who do not have the necessary training or facilities to conduct these experiments.



KEEPING RESEARCH DATA

SAFE

A COST MODEL AND GUIDANCE FOR UK UNIVERSITIES

Neil Beagrie, Julia Chruszcz, and Brian Lavoie

with case studies contributed by the Universities of Cambridge, Southampton, King's College London, and the Archaeology Data Service University of York.

Final Report - April 2008

Prepared by:

Charles Beagrie Limited

www.beagrie.com

A study funded by

JISC

<http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx>

Scaling Up: A Community Solution

Interviews & analysis

Synthesis: IR Policy & Practice, Laboratory Practice & Workflows, Technical Interoperability & Standards, Metadata Schema & Application Profiles, Semantic Interoperability, Data Citation, Identifiers & Linking, Federation Architectures & Third Party Services, Rights & Licensing, Data Quality & Validation, Preservation, Curation & Sustainability

Recommendations, commentary

Matters Arising: Diverse lab practice, LIMS and proprietary formats, Data policy should reflect lab practice & institutional model, Data quality criteria/validation, “Prior publication” problem, We need scalable assignment of “terms” for data discovery, No discipline preservation model



Scaling Up: Towards a Federation of Crystallography Data Repositories

Document details

Author:	Liz Lyon, Simon Coles, Monica Duke, Traugott Koch
Date:	12th May 2008
Version:	1.0 Final
Document Name:	ebank-phase3-report-final.doc
Notes:	

<http://wiki.ecrystals.chem.soton.ac.uk>

Building a Federation of Repositories

General Chemistry | [HOME](#) | [ABOUT](#) | [CONTACT](#) | [MOLECULES](#) | [VIEWS](#) | [EXPERIMENTS](#)

Search

This is an open source science project in chemistry. Post specific problems in chemistry that need to be solved. Post specific partial solutions to these problems. Or execute a suggested step. NOTE: ANYTHING POSTED HERE IS SUBJECT TO A SHARE-A-LIKE WITH ATTRIBUTION CREATIVE COMMONS LICENSE (see bottom of page for link)

SATURDAY, JUNE 07, 2008

Drexel eCrystals Repository

Richard Stephenson from Southampton University has kindly added Drexel to their eCrystals X-Ray crystal structure repository, now located at: <http://ecrystals.chemistry.drexel.edu>



Contributors

- Met Todd
- Jean-Claude Bradley
- David Bradley
- James G
- Beth Ritter-Guth
- Khalid Mirza
- Trider Jimm
- Guido
- Michele Aquino
- David Berke-Schlesel
- Yana
- Elmo
- Sean Gardner
- Brett
- David J. Strumfels
- Jeremy Frey

Links

- Drexel CoAS E-Learning Blog
- Drexel CoAS E-Learning Podcast
- Drexel CoAS E4 Transcript
- Orgo I

Since we are operating under Open Notebook Science conditions, all of our structures will be made immediately available. However, the system is designed to allow contributors to select an embargo date if desired. Anyone from the Drexel community is welcome to participate.

For example, the record for one of our anti-malarial compounds UC-1500 can be shared through a simple URL:

Reciprocal Net

a distributed crystallography network for researchers, students and the general public

- About Reciprocal Net
- Site Network
- Learn About
- Help

Highlight

Methanol (H4-phthalocyaninato)-magnesium (II)

Dr. Ilya Guzei
Molecular Structure Lab, Univ. of Wisconsin-Madison
Prof. Robert McGaff
Univ. of Wisconsin-La Crosse



As a class of compounds, the metallophthalocyanines exhibit a wide array of physical and chemical properties that can be profoundly affected by subtle changes in their structures. Shown here is methanol(H4-phthalocyaninato), a recent addition to the list of compounds that fit this classification. This compound adopts a crystal structure that results from two different types of interactions between molecules. These are known as p-p dimerization and hydrogen bonding. One interesting subtlety in the structure of the molecule shown can be described as a "hat visor" shape, a description put forth for the first time in a very recent paper describing this structure. This molecule is only one of several related compounds under investigation through a collaboration involving Dr. Ilya Guzei, Director of the Molecular Structure Laboratory at the University of Wisconsin-Madison and Robert McGaff, an Associate Professor of Chemistry at the University of Wisconsin-La Crosse.

Welcome to Reciprocal Net

The Reciprocal Net is a distributed database used by research crystallographers to store information about molecular structures; much of the data is available to the general public. The Reciprocal Net project is still under development. Currently, we have 18 participating crystallography laboratories online. The project is funded by the National Science Foundation (NSF) and part of the National Science Digital Library. [More about Reciprocal Net.](#)

Partners

- Indiana University
- Consortium for Advanced Radiation Sources
- Los Alamos National Lab
- Massachusetts Institute of Technology
- McMaster University
- Northwestern University
- Ohio State University
- Princeton University
- Purdue University
- University of California, San Diego
- University of Cincinnati
- University of Iowa
- University of Kansas

News

The Crystal-Grid Collaboratory
held its second workshop in Bloomington, Indiana, USA, April 26-28, 2007.

Events

The Nobel Prize in Chemistry 2007 was awarded to Gerhard Ertl for his studies of chemical processes on solid surfaces.

The Future of Energy Storage: Challenges and Opportunities are discussed by Nicola Armaroli and Vincenzo Balzani

[Home](#) | [About](#) | [Browse by Year](#) | [Browse by People](#)

[Login](#) | [Create Account](#)

Welcome to eCrystals - University of Sydney

eCrystals - Sydney is the archive for Crystal Structures generated by The Crystal Structure Analysis Facility of The University of Sydney.



The information contained within each entry of this archive is all the fundamental and derived data resulting from a single crystal X-ray structure determination, but excluding the raw images. The results have not been externally refereed, but the information supplied should enable any reader to check the reliability and validity directly, since all the files provided are freely available for download. Should any error be detected, we would appreciate receiving suitable comments, and we will make any necessary amendments, and include a note to that effect. Any reader wishing to have access to the raw images is welcome to contact us, and we will make arrangements for these to be made available.

Readers are welcome to use these results for further studies, but such re-use should be fully acknowledged by a suitable reference to the specific archive entry, as in normal scientific publication practice.

eCrystals - University of Sydney supports [DOI 2.0](#) with a base URL of <http://sydney.e-crystals.chem.usyd.edu.au/>

[eCrystals](#) is powered by [eprints](#), which has been customized by [bluewin.ch](#) in collaboration with the [University of Southampton](#).

[eprints](#)

Reciprocal Net

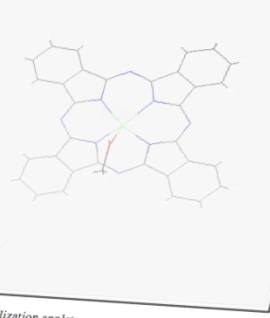
Part of the Reciprocal Net Site Network

Site Info | Search

Wisconsin #mcg015

sample # (local search) [Go](#)

[Log in](#)



Empirical formula (observed): $C_{13}H_{20}MgN_8O$

a: 13.1896 Å
b: 24.3104 Å
c: 8.0185 Å
alpha: 90.000 °
beta: 100.103 °
gamma: 90.000 °
space group: P2(1)/c
calculated density: 1.493 g/cm³
Z: 4
temperature: -173.0 °C
volume: 2531.20 Å³
R(F): 0.043
R(F²): 0.066
R_w(F²): 0.120
goodness of fit: 1.007
Crystallographer name: Ilya A. Guzei
Citation of a publication: I. A. Guzei, R. W. McGaff and H. M. Kieler. Acta Cryst. (2005). C61, m472-m475.
Layman's explanation: As a class of compounds, the metallophthalocyanines exhibit a wide array of physical and chemical properties that can be profoundly affected by subtle changes in their structures. Shown here is methanol(H4-phthalocyaninato)magnesium(II), a recent addition to the list of compounds that fit this classification. This compound adopts a crystal structure that results from two different types of interactions between molecules. These are known as p-p dimerization and hydrogen bonding. One interesting subtlety in the

switch to another visualization applet:

[miniJAMM](#) [open in new window...](#)

eCrystals



ChemSeer

ChemSpider Beta



CrystalEye (beta)

eBank UK



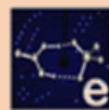
Chemical Database Service

RSC



Chemistry Central

Unilever
Cambridge
Centre For Molecular Science Informatics



eCrystals

Newcastle
University



The University of Sydney



UNIVERSITY OF
OXFORD



The University of
Nottingham



University
of Glasgow

Reciprocal Net



Science & Technology
Facilities Council

Useful Chemistry

JISC



EPSRC

Engineering and Physical Sciences
Research Council

$$n\lambda = 2d \sin\theta$$

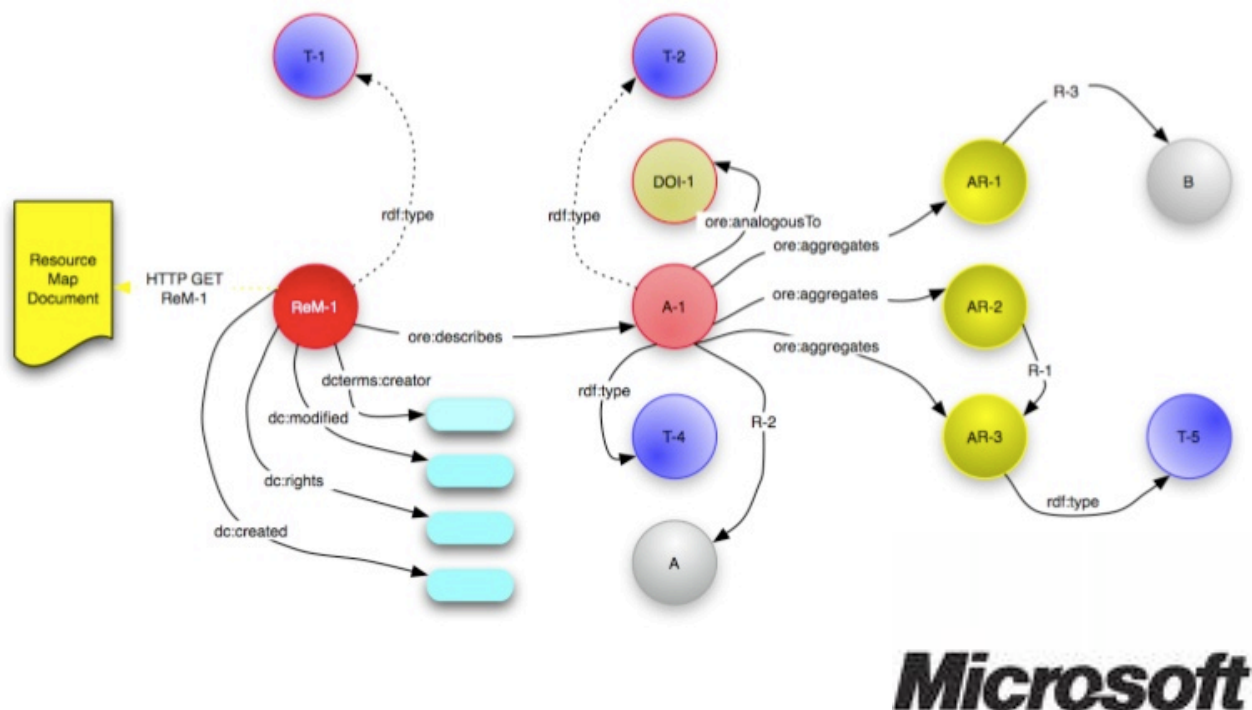
blogs@ChemTools



DCC

Packaging and Interoperability

- New moves in Digital Libraries community to enable distributed repositories to fully describe and exchange content
- OAI-ORE (Open Archives Initiative – Object Reuse and Exchange)
- <http://www.openarchives.org/ore/>
- Describes an aggregation of objects in an exchangeable format
- OREChem testbed project



Big Business

Google to Host Terabytes of Open-Source Science Data

By Alexis Madrigal  January 18, 2008 | 2:23:21 PM Categories: [Dataset](#), [Research](#)

Sources at Google have disclosed that the humble domain, <http://research.google.com>, will soon provide a home for terabytes of open-source scientific datasets. The storage will be free to scientists and access to the data will be free for all. The project, known as Palimpsest and [first previewed](#) to the scientific community at the [Science Foo camp](#) at the Googleplex last August, missed its original launch date this week, but will debut soon.



Building on the company's acquisition of the data visualization technology, [Trendalyzer](#), from the oft-lauded, [TED presenting Gapminder](#) team, Google will also be offering algorithms for the examination and probing of the information. The new site will have YouTube-style annotating and commenting features.

The storage would fill a major need for scientists who want to openly share their data, and would allow citizen scientists access to an unprecedented amount of data to explore. For example, two planned datasets are all 120 terabytes of Hubble Space Telescope data and the images from the [Archimedes Palimpsest](#), the 10th century manuscript that [inspired the Google dataset storage project](#).

UPDATE (12:01pm): Attila Csordas of [Pimm](#) has [a lot more details](#) on the project, including a set of slides that Jon Trowbridge of Google gave at a presentation in Paris last year. WIRED's own Thomas Goetz also mentioned the project in his fantastic piece of [freeing dark data](#).

One major issue with science's huge datasets is how to get them to Google. In this post by a [SciFoo attendee](#) over at [business|bytes|genes|molecules](#), the collection plan was described:

(Google people) are providing a 3TB drive array (Linux RAID5). The array is provided in "suitcase" and shipped to anyone who wants to send they data to Google. Anyone interested gives Google the file tree, and they SLURP the data off the drive. I believe they can extend this to a larger array (my memory says 20TB).

You can check out more details on why hard drives are the preferred distribution method at Pimm. And we hear that Google is hunting for cool datasets, so if you have one, it might pay to get in touch with them.

Image: flickr/[dannysullivan](#)

A solid foundation for Open/Self-Publishing of Chemistry Data???

Thanks to:

- **Mike Hursthouse**, **Jeremy Frey**, Andrew Milsted, Richard Stephenson, Mark Borkum



- **Dave DeRoure**, **Les Carr**, Chris Gutteridge, Tim Myles-Board, Arouna Woukei, Dave Tarrant



- **Liz Lyon**, Rachel Heery, Monica Duke, Michael Day, Traugott Koch, Manjula Patel, Pete Cliff



JISC

Microsoft