

Technical Developments in Support of Open Access: some recent UK initiatives

Moving Towards Open Access, 27 - 28 September 2006, Keble College, Oxford

Rachel Heery

Deputy Director R&D, UKOLN



UKOLN is supported by:



www.ukoln.ac.uk

A centre of expertise in digital information management



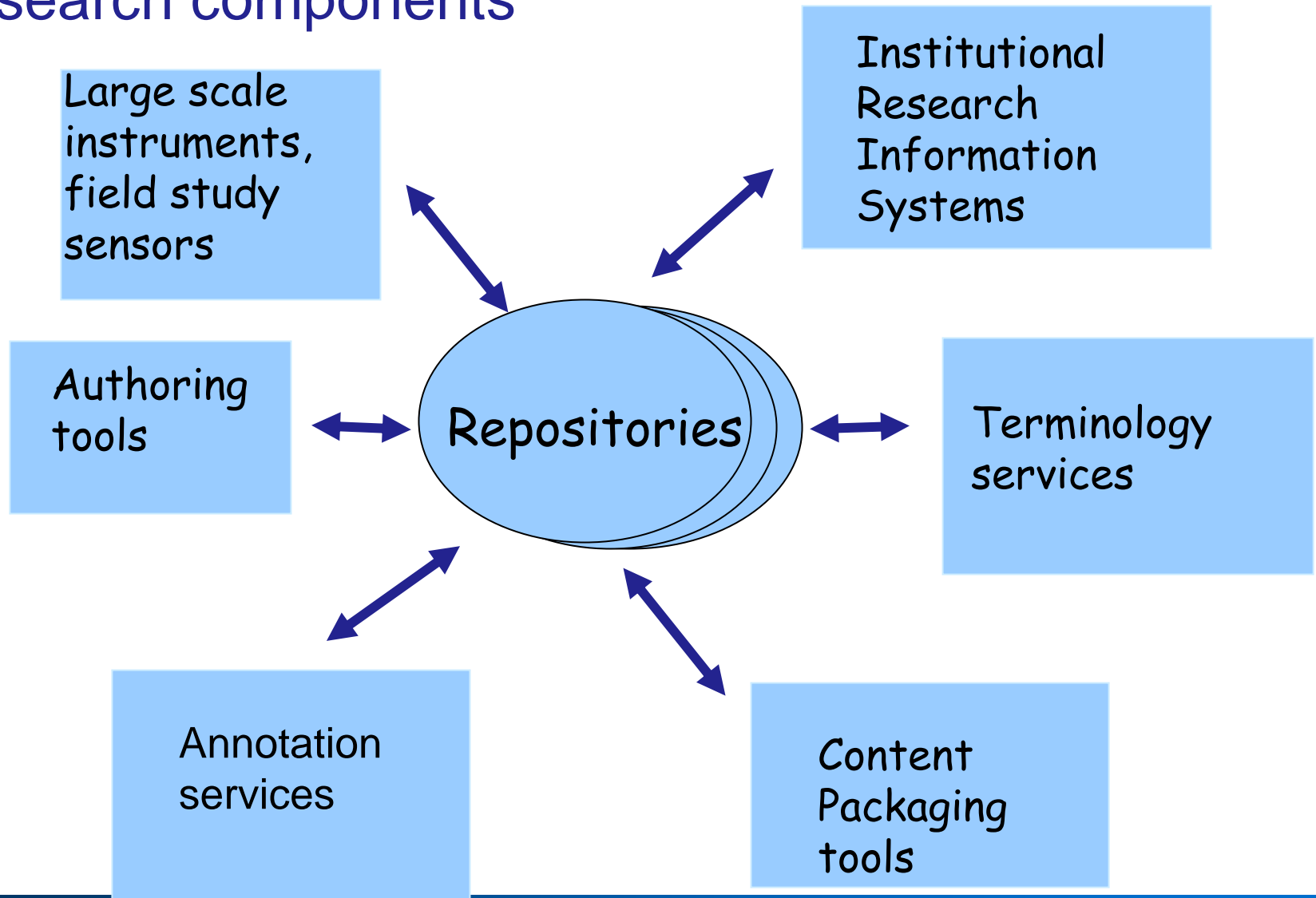
The OA technical challenge

- Managing diverse resource types
- Enabling rich data flow between
 - networks of repositories (federations or loosely distributed)
 - repositories and other eResearch components
- Enabling automated deposit
 - Of scientific data from observational, experimental, field sensor equipment
 - Of text from authoring tools
 - Of learning objects
- Incorporating knowledge management techniques
 - Subject search, re-use, linking, text mining

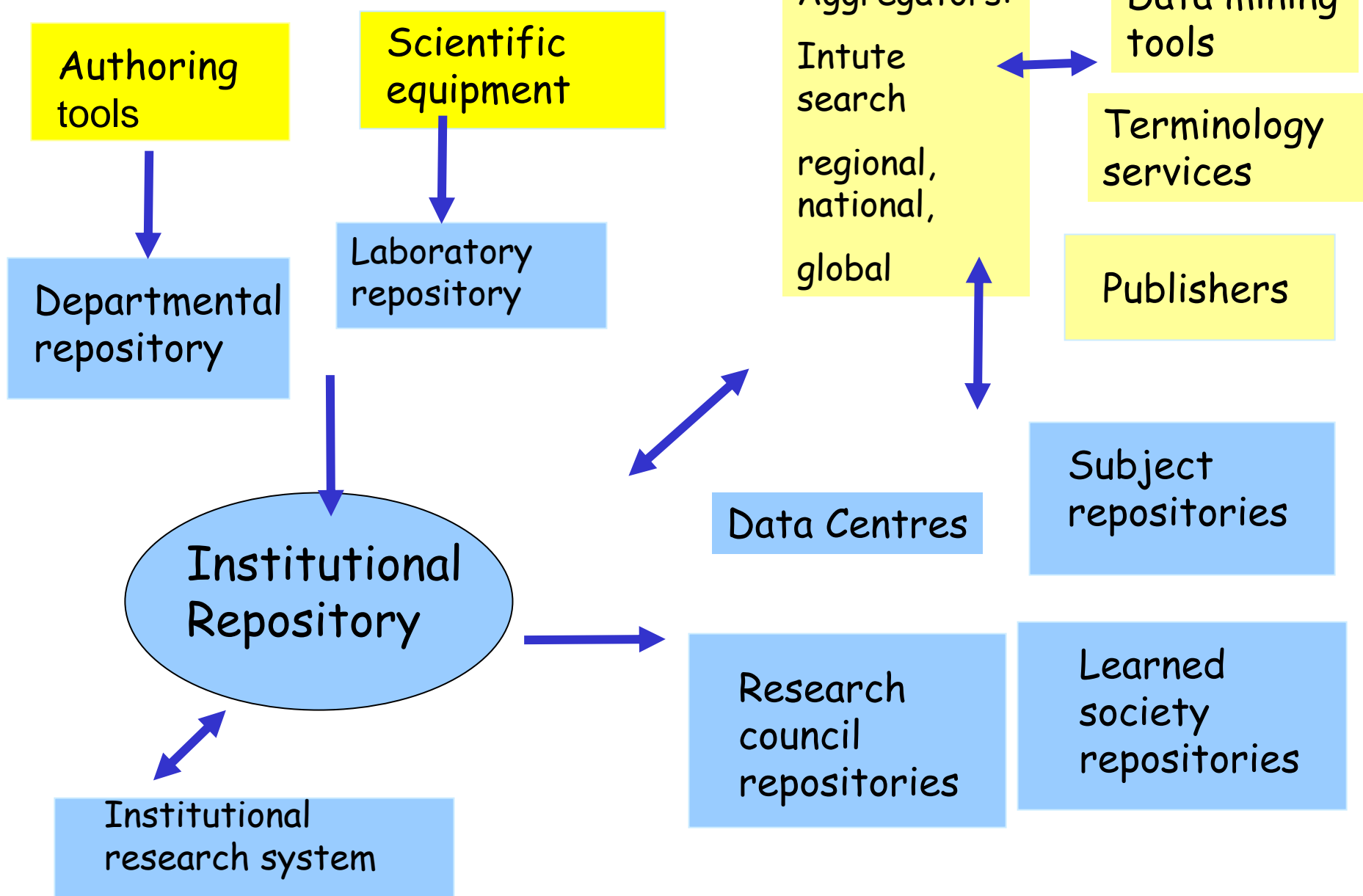
Increasingly complex environment



Repositories inter-working with other eResearch components



UK Repository network: an ecology



Defining workflows and dataflows

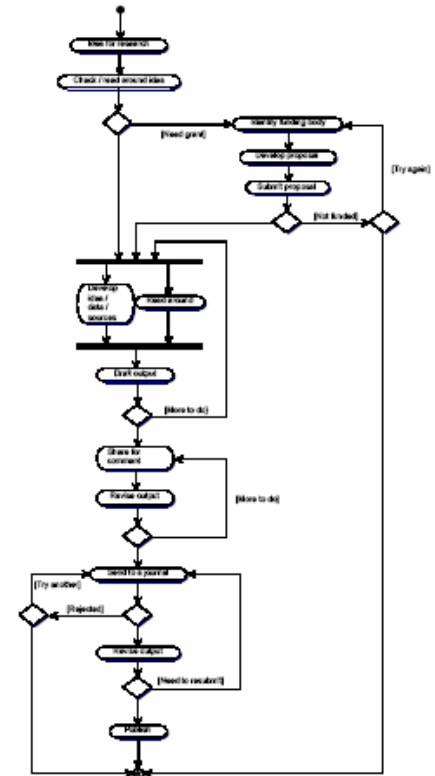
- Analysing roles and interactions within and between components of repository landscape
- Identifying and defining services
 - Potential for ‘shared services’, re-use of services
- Exploring potential dataflows
 - Aggregation, data exchange, metadata extraction and enhancement (lots of players!)
- Let’s not forget - what does the user want?

JISC projects: understanding the research process



Source-to-Output Repositories

- Project StORe: Source-to-Output Repositories (Edinburgh)
 - Relationship between primary data and research publications
- RepoMMan: Repository Metadata and Management (Hull)
 - Activity diagrams and workflows
- DCC SCARP
 - Curation staff working within research teams



Content creation

- Depositing content in repositories
- Creating metadata

..... automatically!

Deposit Service Description: aims

- Small scale collaboration between UK repository developers (ePrints, DSpace, Fedora, Intrallect)
- Enabling predictable interface between repositories and other applications
- Need a common Deposit interface
 - light-weight solution to assist populating repositories within timescales of JISC programme
- An 'Eazi-Deposit' service, 'Smart' deposit
- To capture content from desktop applications, experimental equipment (smart labs), learning content development tools etc
- To hide complexity from end-user

Deposit Service: issues

- Is there requirement for packaging service associated with deposit?
- Boundaries between deposit and ingest
- Data integrity (identifiers, provenance etc)
- Potential for storing repository policy rules within (repository) service registry?

Deposit Service Description: progress

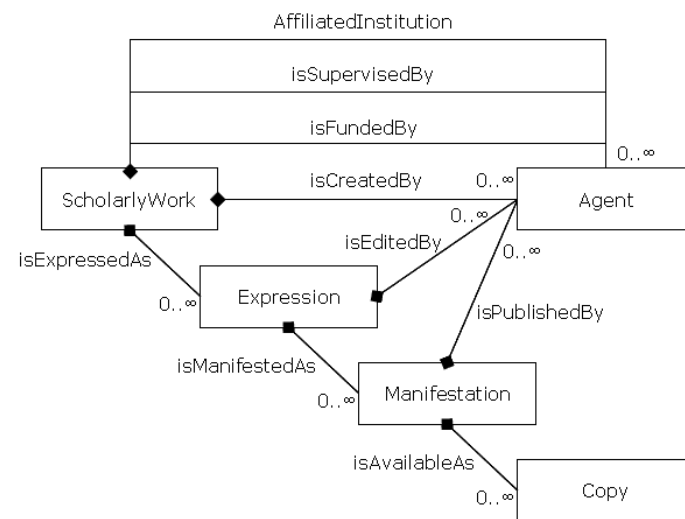
- Deposit explain, deposit, receipt interactions and parameters defined (tentatively)
- XML serialisations proposed
- .. Very draft, very tentative
- More work needed!
- <http://www.ukoln.ac.uk/repositories/digirep/index/Dep>
- <http://msc.mellon.org/Meetings/Interop>

ePrints DC Application Profile: aims

- Small scale six month JISC funded project
- Defining a Dublin Core application profile for ePrint repositories
- To expose richer metadata for harvesting
- Preliminary solution to version identification
- Guidance on differentiating jump-off pages and actual resources
- Recommendations for including bibliographic citations
- Distinguishing full text Open Access materials within repository

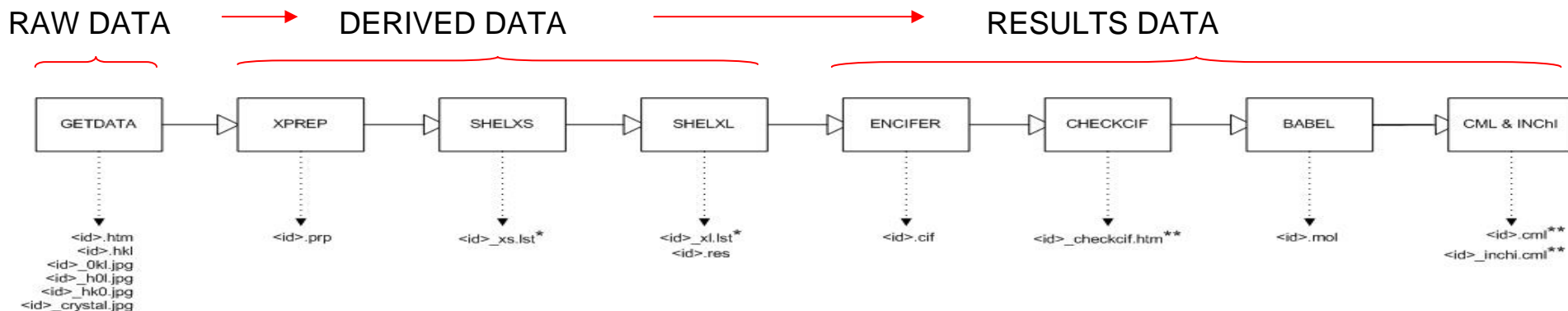
ePrints DC Application Profile: progress

- DC Application profile documented
- Contains recommendations, cataloguing and usage guidelines and examples
- Community acceptance plan drawn up
- Deployment planned by developers, repositories, services
 - Statements from Eprints, DSpace, Fedora, Intute, EDINA
- Workshops
 - DC 2006
 - Open Scholarship 2006
- Further work to agree XML schema



R4L Deposit scenario (...part of....)

1. Produce strategy for synthesis (=idea)
2. Submit plan to SmartTea system (incl. identifiers)
3. Retrieve and follow instructions (sub-workflow?)
4. Experimental synthesis metadata automatically recorded on instruments (Smart Lab)
5. Create record for synthesised sample (+ proposed chemical identifier) in R4L laboratory data management system
6. Run spectral analyses on sample capturing further analysis metadata (incl. time-stamp, analysis software version, researcher details etc.)
7. Save spectrum in native and common formats
8. Invoke R4L data capture service and deposit files + metadata in laboratory repository...



Repository software

ePrints repository software new release

- Improved author interface reducing the time and effort to ingest
- Tools and new XML format for improved import and export of data e.g. RSS feeds, export search results, interfaces to bibliographic management software
- Plugin architecture to make it easier for third parties to create and apply tools
- Request eprint and embargo features

Discovery and re-use

Intute search: aims

- JISC funded for 3 years
 - UK Repository Search Project
 - Facilitate discovery, access and retrieval of repository content
 - Raise visibility of repository content
 - Encourage deposit of more content
 - Adding value for UK HE stakeholders
 - Building on ePrints UK
- ... project start up imminent

Adding value: linking data to publications: eBank

eBank UK Demo

Crystal Structure Data Reports

[Crystal Structure Report of 2-\(N-Ferrocenyl\(methylcarbamoyl\)-5-\(N-phenylcarbamoyl\)-3,4-diphenyl pyrrole](#)

Creator(s): Hursthouse, Michael B., Light, Mark E., Coles, Simon J., Horton, Peter N., Gale, Phil A., Denuault, G., Wannier, C. N.

Date released: 23/05/2004

Empirical Formula: C39H29F4N3O2

IUPAC name: 2-(N-Ferrocenyl(methylcarbamoyl)-5-(N-phenylcarbamoyl)-3,4-diphenyl pyrrole

CCDC code: XU25U

Compound Class: Organic

General keywords: Supramolecular Chemistry

Related article: [3A URL citation?](#)

Available Datafiles

CIF file

processing Dataset

refinement Dataset

solution Dataset

[Crystal Structure Report of 2-\(N-Ferrocenyl\(methylcarbamoyl\)-5-\(methoxycarbonyl\)-3,4-diphenylpyrrole](#)

Creator(s): Hursthouse, Michael B., Coles, Simon J., Light, Mark E., Horton, Peter N., Gale, Phil A., Denuault, G., Wannier, C. N.

Date released: 23/05/2004

Empirical Formula: C29H24F4N3O3

IUPAC name: 2-(N-Ferrocenyl(methylcarbamoyl)-5-(methoxycarbonyl)-3,4-diphenylpyrrole

CCDC code: XU25A

Compound Class: Organometallic

General keywords: Supramolecular Chemistry

Related article: [3A URL citation?](#)

Available Datafiles

CIF file

processing Dataset

refinement Dataset

solution Dataset

Publications

A supramolecular assembly: aquatris(pentafluorophenyl)borane as its mixed dimethyl sulfone and water solvate, (H2O)3(B(C6F5)3Me2SO)2H2O

The title compound, C18H20F15O2S2, obtained by crystallization of a product formed from a reaction mixture containing B(C6F5)3 and Me2SO2 (and H2O) in hexane, was characterized in the solid state as a supramolecular assembly containing water adducts of tris(pentafluorophenyl)borane, (H2O)3(B(C6F5)3), linked together by a network of hydrogen bonds involving one additional H2O and one additional Me2SO2 molecule per adduct molecule.

Creator(s): Coles, Simon J., Hursthouse, Michael B., Deckett, Michael A., Dutton, Michael Acta Crystallogr E Struct Rep Online Vol 59 Issue Pt 9 pp. o1354 - o1356

DOI:

Download from: <http://scripts.iucr.org/cgi-bin/getarticleid?issn=1502-3508&volume=59&page=1354&detail=yes>

Structural investigations of phosphorus-nitrogen compounds. 5. Relationships between molecular parameters of 2,2-diphenyl-4,6-cis-oxyltetra(ethyleneoxy)-4,6-R2-cyclotriphosphazinanes (R = Cl, OCH2CF3, OPh, OMe, NHPh, NHtBu) and substituent basicity constants

The syntheses and crystal structures of six new cis-ansa derivatives N3P3R2[O(CH2CH2O)4]R2 (R = Cl, OCH2CF3, OPh, OMe, NHPh, NHtBu) are reported and the observed relationship between molecular parameters of the N3P3 ring and substituent basicity constants is discussed.

Creator(s): Beek, S., Coles, S. J., Hursthouse, M. B., Kirc, A., Mayer, T. A., Strui, R. A. Acta Crystallogr B Vol 58 Pt 6 pp. 1067 - 1073

DOI: 10.1107/S0108768102018608

Download from: <http://scripts.iucr.org/cgi-bin/getarticleid?issn=0108-7681&volume=58&page=1067&detail=yes>

Related dataset: <http://crystals.chem.soton.ac.uk/archive/00000062/>

5 α -Cholestane

The title compound, C27H48, is a steroid derivative composed of a saturated carbon fused ring framework with two methyl substituents and an alkyl side chain

Creator(s): Coles, S. J., Hursthouse, M. B., Frampton, C. S. Acta Crystallogr E Struct Rep Online Vol 58 Issue Pt 4 pp. o445 - o446

DOI: 10.1107/S1502053602004786

Download from: <http://scripts.iucr.org/cgi-bin/getarticleid?issn=1502-3508&volume=58&page=445&detail=yes>

Related dataset: <http://crystals.chem.soton.ac.uk/archive/00000051/>

Ethyl (2S*)-2-[(2R*)-2R*)-5S*)-2'-5-dimethyl-5-oxopiperidino-[2,2']bifuranyl-5-yl]-2-hydroxyethanone

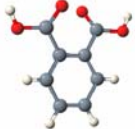
The framework of K2Zn(H2P2O7)2·2H2O contains acid diphosphate-metalate layers linked by KO interactions and weak hydrogen bonds. Zn²⁺ cations are coordinated octahedrally by O atoms from two bidentate H2P2O7²⁻ anions and two water molecules.

Benzene 1,2dicarboxylic acid

Simon J Coles, Michael B Hursthouse, Claire L Taylor and Peter N Horton

University of Southampton

CaH4O4



ICM Code: NCH=1.12Beta(C8H6O4)cd.7(10)5-3.1-2-4.6(5)(11)2h1-4h(19,10)(11,12) (google for icm)

Compound Class: Organic

Keywords: Phthalic acid

Creation Date: 15 February 2005

Deposited By: Dr Simon J Coles

Deposited On: 21 February 2005

Data collection parameters

Chemical formula	C8H6O4
Crystallisation Solvent	
Crystal morphology	Prism
Crystal system	monoclinic
Space group symbol	C2/c
Cell length a	5.0016(10)
Cell length b	14.214(3)
Cell length c	9.5196(19)
Cell angle alpha	90.00
Cell angle beta	94.33(3)
Cell angle gamma	90.00
Data collection temperature	120(2)

Available Files

File Name	Size
Final Result	
05neh1006.cml	3k
05neh100605neh1006.cif	9k
05neh100605neh1006_checked.htm	7k
05neh1006_nctv.cml	1k
Refinement	
05neh100605neh1006.res	3k
05neh100605neh1006.in	21k

research paper

Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Structural investigations of phosphorus-nitrogen compounds. 6. Relationships between molecular parameters in per-X-substituted bridged spermine derivatives and basicity constants ΣaR of substituents

Received 8 July 2004
Accepted 13 October 2004

Simon J Coles,* David B. Davies,^a Michael B. Hursthouse,^a Adem Kirc,^b Thomas A. Mayer,^c Robert A. Shaw^b and Gonul Yenilmez Ciftci^d

^aSchool of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, England, ^bSchool of Biological and Chemical Sciences, Birkbeck College, University of London, Gordon House, 29 Gordon Square, London WC1H 0PP, England, and ^cDepartment of Chemistry, Gaziye Institute of Technology, Gebze, Turkey

*Correspondence e-mail: s.j.coles@soton.ac.uk

A systematic study is reported of the products of the nucleophilic substitution reactions of the spermine-bridged cyclotriphosphazene, [N₃P₃X₆(NHCH₂CH₂CH₂N)CH₂CH₂]₃ [where X = Cl (2a)], to give a number of new structures [(2a)-(2g)] in which X = OPh, (spiro-O(CH₂)₃O)₂, Ph, NHPh, NC₂H₅ and NHBA^f, respectively. A comparison has been made between the sum of the substituent basicity constants, ΣaR , obtained in nitrobenzene solution, and ten molecular parameters of the N₃P₃ ring (the internal bond angles α , β , γ , δ and θ , and the P-N bond lengths α , β , γ , δ and ϵ) as well as the difference between the bond lengths a and b , $\Delta(P-N)$. It is found that the systematic change in molecular parameters of compounds (2a)-(2g) is in line with changes in ΣaR values, indicating the similarity in relative electron-releasing capacity of substituents X in the solid state and in solution. It is also found that the effect on molecular parameters of (2a)-(2g) with two X' substituents in P(O,R)₂ groups is greater than that for one X' substituent in P(O,R)₂ groups in an analogous series of compounds observed previously [Bejli et al. (2002), Acta Cryst. B58, 1067-1073].

Linking research to learning e.g. embedding eBank aggregator service in a science portal for student learners



PSigate
Physical Sciences Information Gateway

[PSigate Home](#) > [eBank](#)

This is a prototype test interface to the eBank UK service providing access to data in the University of Southampton eCrystal data repository and elsewhere. eBank UK is a JISC-funded project which is a part of the Semantic Grid Programme. The project is being led by UKOLN in partnership with the Combechem project at the University of Southampton and PSigate.

Enter your search term(s)

Author:

CCDC Code:

IUPAC name:

Empirical Formula:

Compound Class:

General keywords:

Date released:

OR published in the last:

Search within: Data Reports Publications e.g. journal articles

Copyright © PSigate. All rights reserved. PSigate is a service of the [Resource Discovery Network \(RDN\)](#). feedback@psigate.ac.uk

PSigate
Physical Sciences Information Gateway

[About Us](#) | [Contacts](#) | [Site Map](#) | [Help](#)

SEARCH SUBJECTS

- [Astronomy](#)
- [Chemistry](#)
- [Earth sciences](#)
- [Materials science](#)
- [Physics](#)
- [History & policy](#)

FEATURES

- [PSigate home](#)
- [Search PSigate](#)
- [Subject Headings A-Z](#)
- [About PSigate](#)
- [Site map](#)
- [New additions](#)
- [Feedback form](#)
- [Suggest site](#)
- [News services](#)
- [Reference](#)
- [Spotlight](#)
- [Science Timelines](#)
- [Hot Topics](#)
- [Science Data](#)
- [Courses](#)
- [Jobs](#)
- [Learning & Teaching](#)

[PSigate Home](#) > [eBank](#) > **Search Results**

Your search returned 28 data reports and 4 publications. Viewing 1 to 10

[Next](#)
[New search](#)

Crystal Structure Data Reports

[Crystal Structure Report of 2-\(N-Ferrocenylmethylcarbamoyl\)-5-\(N-phenylcarbamoyl\)-3,4-diphenyl pyrrole](#)

Creator(s): Hursthouse, Michael B., Light, Mark E., Coles, Simon J., Horton, Peter N., Gale, Phil A., Denuault, G., Warriner, C. N.

Date released: 23/05/2004

Empirical Formula: C35H29FeN3O2

IUPAC name: 2-(N-Ferrocenylmethylcarbamoyl)-5-(N-phenylcarbamoyl)-3,4-diphenyl pyrrole

Compound Class: Organic

General keywords: Supramolecular Chemistry


Related article: [2A URL citation?](#)

Adding value:

- Annotation, visualisation

Knowledge extraction

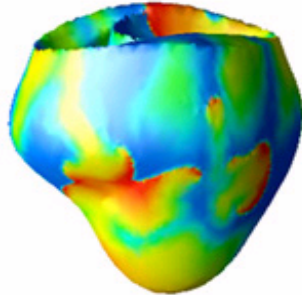
- Mining (data, text, structures)



Integrative Biology
Exploiting e-Science to combat fatal diseases


[\[Content\]](#) [About the Project](#) [About the Heart](#) [About](#)

Modelling the heart



Computer simulation of beating heart.
© Alan Garfinkel - UCLA

the national centre for text mining



You are in: [Home](#) | [Welcome to NaCTeM](#)

Home
News
Events
People
Vacancies
Seminars
Research
FAQ
Software
Resources
Location



Welcome to NaCTeM

The National Centre for Text Mining is the first publicly-funded text mining centre in the world. We provide text mining services in response to the requirements of the UK academic community. We also make significant contributions to the text mining research community, both nationally and internationally. You can find a fuller description of the aims and objectives of the Centre in our [article](#) in Ariadne Magazine.

On our website, you can find pointers to more sources of information about text mining, including links to tutorials and publications (see [Resources](#)). There is more information on available software tools, both those provided by the Centre and by others, under [Software](#). Our calendar contains information on text mining conferences and workshops (please let us know if you would like an event you are organising included). Check back regularly as we are continually adding to this information.

Consortium Partners

The Centre is operated by two Universities:

THE UNIVERSITY of LIVERPOOL

News & Events

- [IBM UIMA Innovation Award](#)
Top award for Dr Ananiadou
- [Vacancy at NaCTeM](#)
Research Associate
- [Workshop on Text Mining in Social Sciences](#)

NaCTeM

<http://www.nactem.ac.uk/>

Emerging tools: TerMine, GENIA, Cafetiere

Nature 23 March 2006

OTMI: Open Text Mining Interface

nature.com

Nascent

Nature's blog on web technology and science

[» P2P in science](#) | [Main](#) | [Linda Stone visits Nature](#) »

Open Text Mining Interface

Every now and then a scientist contacts Nature asking for a machine-readable copy of our content (i.e., the XML) to use in text-mining research. We're usually happy to oblige, but there has to be a better way for everyone concerned, not least the poor researcher, who might have to contact any number of publishers and deal with many different content formats to conduct their work. Much better, surely, to have a common format in which all publishers can issue their content for text-mining and indexing purposes.

Gaps... (some of them)

- Automated metadata creation (for text documents)
- Integrating access to OA material and traditionally published material
 - Using OA material as resource for OpenURL resolvers
- Identifiers, provenance

Thank you!

<http://www.ukoln.ac.uk/repositories/digirep/>

