Bringing data together

The EPSRC National Chemical Database Service...
…and other eScience initiatives at the RSC

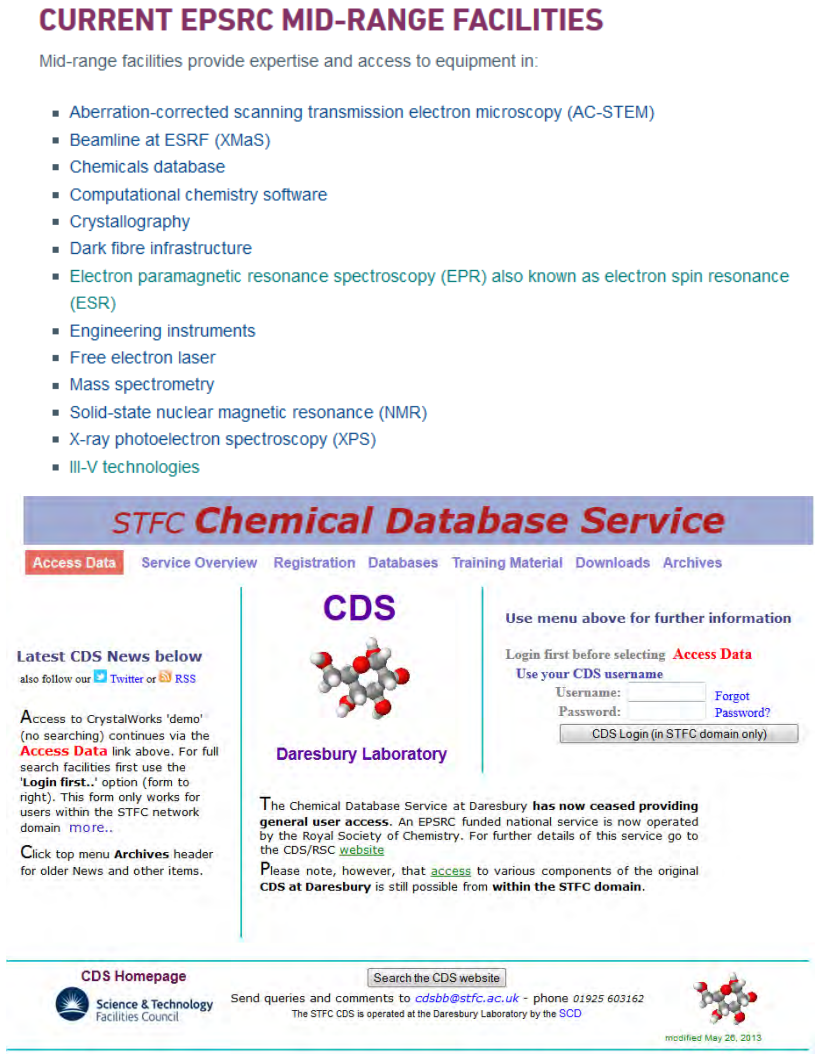
Richard Kidd
@rkiddr

The EPSRC National Chemical Database Service…

cds.rsc.org

History

- EPSRC mid-range facility
- CDS run from STFC Daresbury
- Tender in 2012 for 2013-2017



CURRENT EPSRC MID-RANGE FACILITIES

Mid-range facilities provide expertise and access to equipment in:

- Aberration-corrected scanning transmission electron microscopy (AC-STEM)
- Beamline at ESRF (XMaS)
- Chemicals database
- Computational chemistry software
- Crystallography
- Dark fibre infrastructure
- Electron paramagnetic resonance spectroscopy (EPR) also known as electron spin resonance (ESR)
- Engineering instruments
- Free electron laser
- Mass spectrometry
- Solid-state nuclear magnetic resonance (NMR)
- X-ray photoelectron spectroscopy (XPS)
- III-V technologies

STFC **Chemical Database Service**

Access Data | Service Overview | Registration | Databases | Training Material | Downloads | Archives

Latest CDS News below

also follow our Twitter or RSS

Access to CrystalWorks 'demo' (no searching) continues via the **Access Data** link above. For full search facilities first use the '**Login first..**' option (form to right). This form only works for users within the STFC network domain more..

Click top menu **Archives** header for older News and other items.

CDS

Daresbury Laboratory

Use menu above for further information

Login first before selecting Access Data

Use your CDS username

Username: Forgot

Password: Password?

CDS Login (in STFC domain only)

The Chemical Database Service at Daresbury **has now ceased providing general user access**. An EPSRC funded national service is now operated by the Royal Society of Chemistry. For further details of this service go to the CDS/RSC website

Please note, however, that access to various components of the original **CDS at Daresbury** is still possible from **within the STFC domain**.

CDS Homepage

Science & Technology Facilities Council

Search the CDS website

Send queries and comments to cdsbb@stfc.ac.uk - phone 01925 603162

The STFC CDS is operated at the Daresbury Laboratory by the SCD

modified May 26, 2013

Vision for the Service

- An aggregate collection of best-in-class databases
- An environment for data validation, curation and model building
- A platform for open collaboration
- An environment for micropublishing
- A unique opportunity

Progress in 2013

- Set up and launch for 2 January
- Addition of databases and services
  - WebCSD, DETHERM, ArChem, Chemicalize
- Matching and exceeding remote access
- Access: Shibboleth (coming very soon)
- Advisory Board in place
- 900+ registered CSD users

Disambiguation…

- CDS (and NCDS, and EPSRC CDS/NCDS)
- CSD from CCDC
- ICSD
- I-Lab from ACD/Labs
- ACD from Accelrys

(thanks to DETHERM, SPRESIweb)

Building a repository

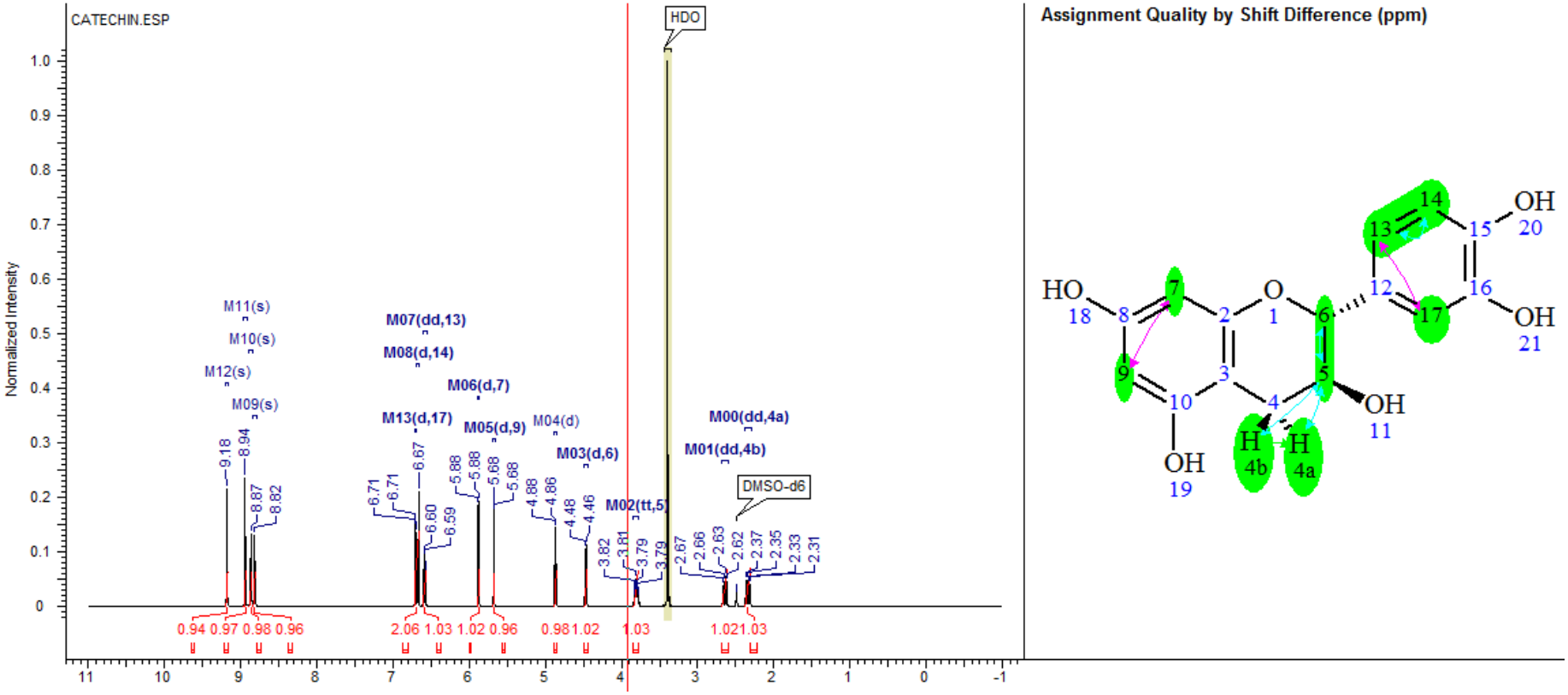
- Depositions of data
- DOI'ed data objects for citation purposes
- Other national services to feed the repository
- Integration APIs to blogging tools for chemistry and Electronic Lab Notebooks
- Linking mandates to value and credit

Building the data repository

Chemistry data of interest

- Chemical compounds
- Chemical reactions
- Analytical data
- Alphanumeric text – logP, pKa, solubility
- Graphs and plots
- Figures

This is what we really want…



We have chosen for now….

Focus on supporting transformed and structured data rather than raw data or dumb files

- Chemicals
- Chemical reactions
- Measured properties
- Analytical data in useful form

What we will deliver…

- Simple interfaces for uploading of data
- Embeddable widgets and programming interfaces to utilise in in-house systems, ELNs
- Automated harvesting approaches – sweeping directories for data
- Data validation where possible
- Embargos

What we will deliver…

- Micropublishing platform for submission of
  - Protocols and procedures
  - Reactions
  - Safety and hazard data
- Template-based submissions of procedures

Work in Progress

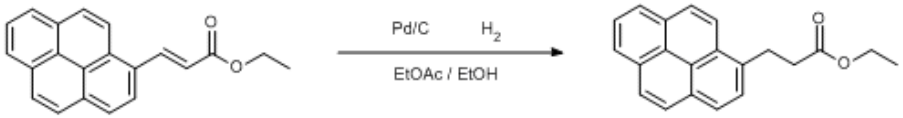
- Micropublishing platform for ChemSpider SyntheticPages (CSSP) already exists
- Work in progress
  - ChemSpider Reactions in development
  - Integration of ELN templates to CSSP – full protocol captured in ELN and procedure mapped to CSSP for deposition
  - Word-based templates to match CSSP depositions

Micropublishing Syntheses

ChemSpider SyntheticPages

Hydrogenation of Ethyl 3-(1-pyrenyl)acrylate; Ethyl 3-(1-pyrenyl)propanoate

SyntheticPage 510
DOI: 10.1039/SP510
Submitted Sep 29, 2011, published Oct 06, 2011
Anish Mistry (a.mistry@warwick.ac.uk)
A contribution from Fox Group, Warwick University



Chemicals Used

Ethyl 3-(1-pyrenyl)acrylate (1 equiv, prepared)
10% Pd/C (2 mol%, Alfa Aeser)
Ethanol
Ethyl Acetate
Hydrogen gas

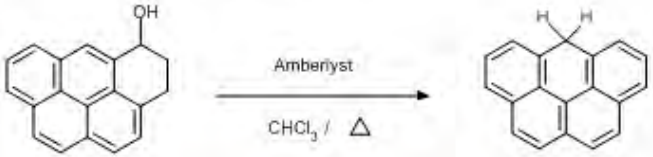
Procedure

Ethyl 3-(1-pyrenyl)acrylate (3.02 g, 10.1 mmol) was firstly dissolved in ethyl acetate (70 ml). Ethanol (70 ml) and the 10% Pd/C (2 mol%) were then added to the mixture and the air evacuated out the system and replaced with hydrogen. The reaction was left to stir under a hydrogen baloon at room temperature for 29 hours. The reaction mixture was then filtered through a pad of celite with ethyl acetate and the solvent removed under *vacuo* to yield a dark yellow solid (3.02 g, 99%).

Dehydration of 3,4-dihydro-5H-Benzo[cd]pyren-5-ol

Anish Mistry

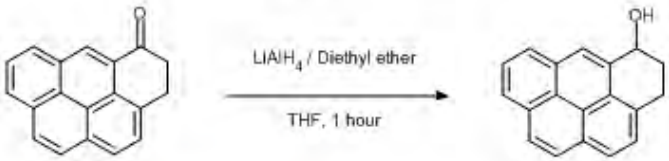
Published: May 31 2012



Reduction of 3,4-dihydro-5H-benzo[cd]pyren-5-one.

Anish Mistry

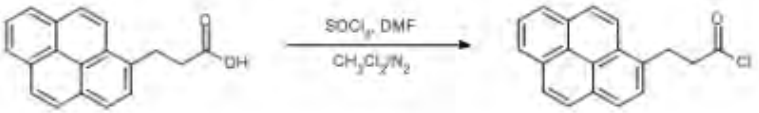
Published: Mar 12 2012



Chlorination of a carboxylic acid

Anish Mistry

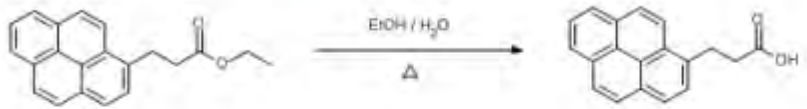
Published: Nov 27 2011



Hydrolysis of Ethyl 3-(1-pyrenyl)propanoate

Anish Mistry

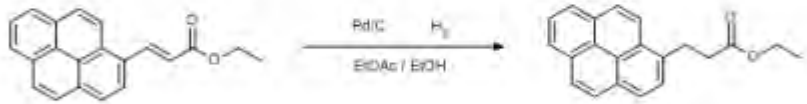
Published: Oct 11 2011



Hydrogenation of Ethyl 3-(1-pyrenyl)acrylate

Anish Mistry

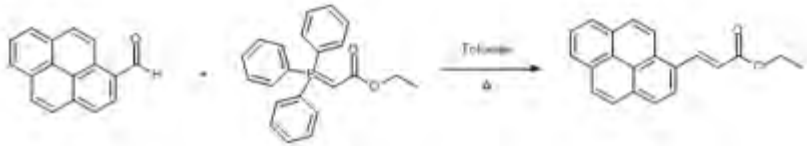
Published: Oct 06 2011



Wittig Reaction

Anish Mistry

Published: Sep 14 2011



Requirements

- Build templates for CSSP/Reactions deposition
- Supporting other EPSRC initiatives, storing successful and failed reactions

Dial-a-Molecule
An EPSRC Grand Challenge Network

Model building with community data

- Community data as a basis of model building
  - Consume newly deposited data
  - Will include us adding in data from publications
  - Build predictive algorithms for the community
  - Regular updating, version stamped prediction
  - Will be using Accelrys Pipeline Pilot

- Provides testing and validation for data

Data validation and curation required

Encouraging participation with

Rewards and **RECOGNITION**

Manual Curation

- Integrated commenting, curating and validation across all RSC platforms
- All integrated to a central RSC profile and feeding altmetrics

Future recognition in altmetrics?

Chem. Sci., 2010, **1**, 561-566 | DOI: 10.1039/c0sc00351d | Edge Article

Total synthesis of all (−)-agelastatin alkaloids†

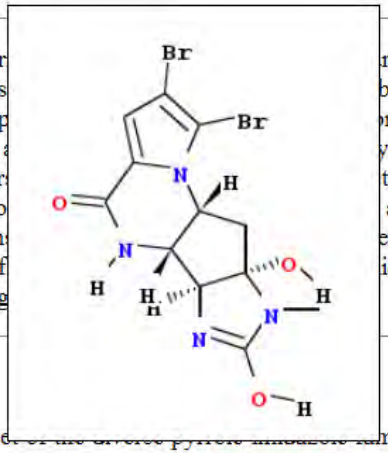
Mohammad Movassaghi *, Dustin S. Siegel and Sunkyu Han

Massachusetts Institute of Technology, Department of Chemistry, 77 Massachusetts Avenue 18-292, Cambridge, MA 02139-4307, USA. E-mail: movassag@mit.edu

Received 2nd July 2010 , Accepted 20th July 2010

First published on the web 16th August 2010

The pyrrole-imidazole family of marine alkaloids, der rsors, constitutes a diverse array of structurally complex natural products bers of this family that possess a tetracyclic molecular framework incorp nnectivities. We provide a hypothesis for the formation of the unique y exploits the intrinsic chemistry of plausible biosynthetic precursor tive total syntheses of all known agelastatin alkaloids including the first to and F. Our gram-scale chemical synthesis of agelastatin A was in enesis of the cyclopentane C-ring and required the development of idazolone-forming annulation reaction and a carbohydroxylative trapping



Introduction

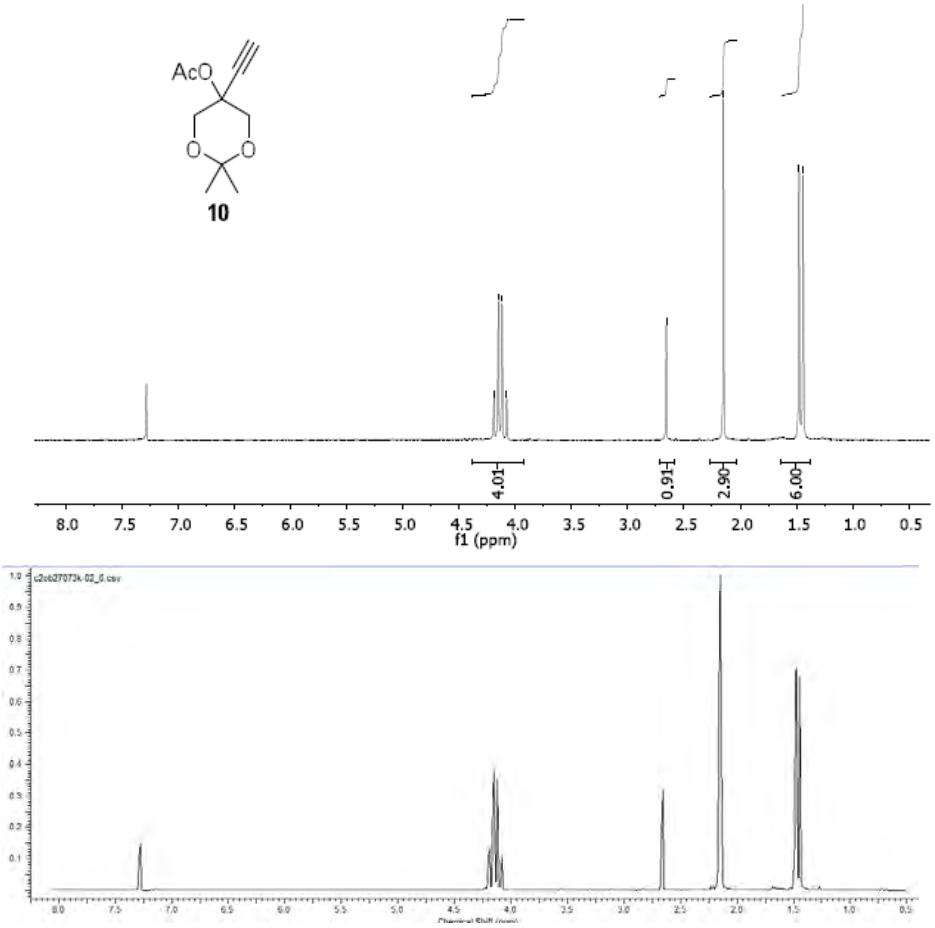
The agelastatin alkaloids constitute an intriguing subse nily of marine alkaloids that are likely derived from linear biogenetic precursors such as clathrodin (**7**),[1] hymenidin (**8**),[2] and oroidin (**9**, Fig. 1).[3,4] (−)-Agelastatins A (**1**) and B (**2**) were first isolated from the Coral Sea sponge *Agelas dendromorpha* by Pietra *et al*. in 1993 who successfully identified and chemically studied their unique

DERA : data enable the RSC Archive

How much **data** is in the archive, in the publications and in the supplementary info?

- How many compounds?
- How many syntheses?
- How many characterisation measurements?
  - Property data
  - Spectral data
  - Graphs and charts to be used for modelling

Top: Figure; Bottom: Converted

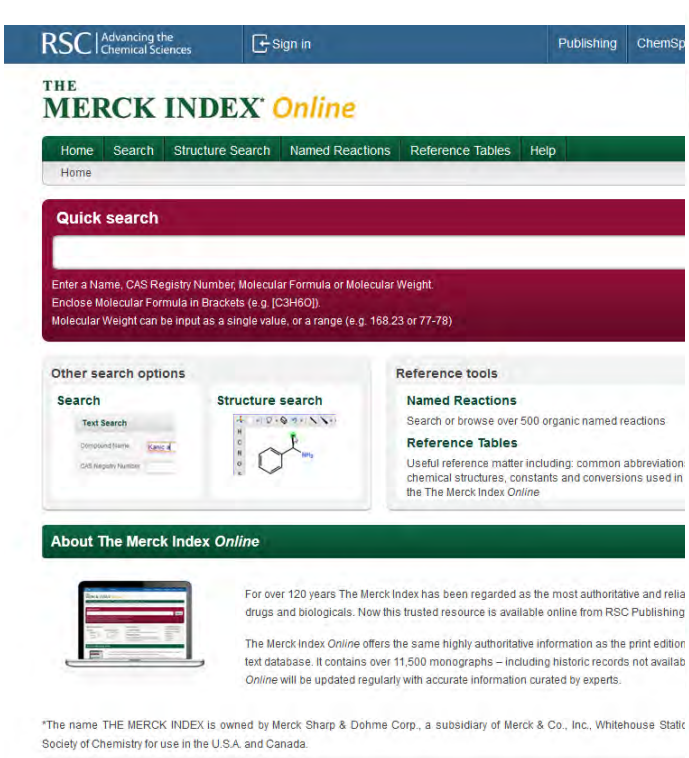


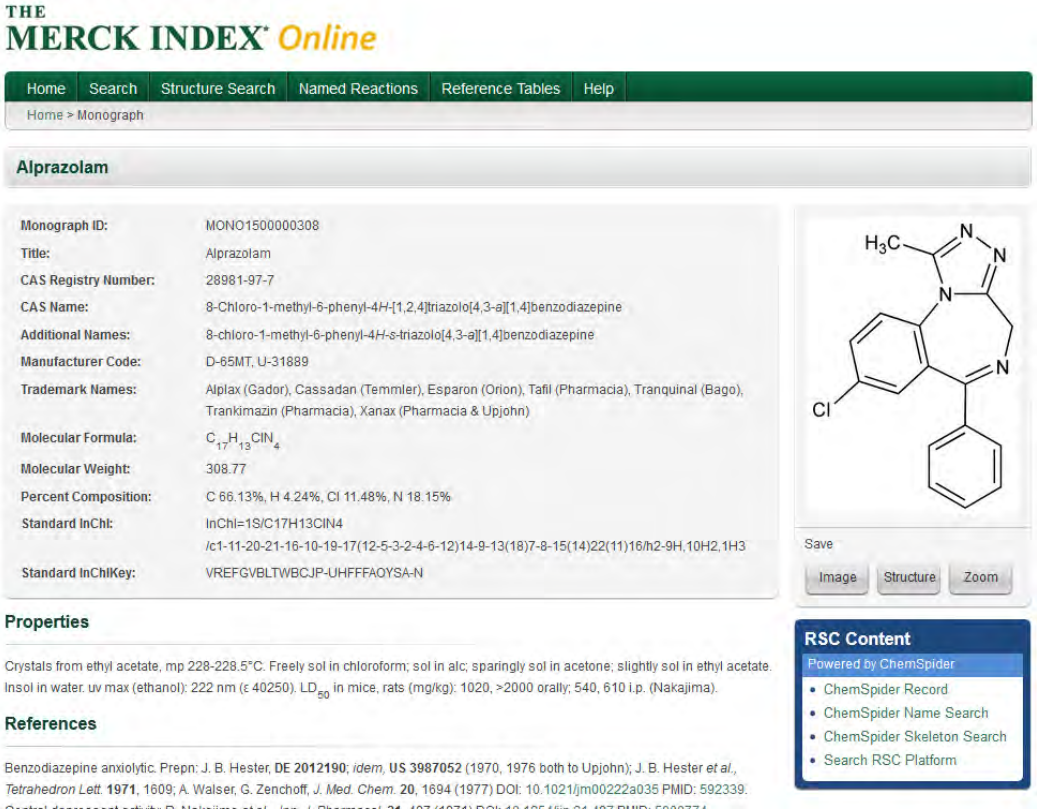
What if we could capture it all?

Aims in summary

- Provide best-in-class data resources to (initially) UK academia
- Provide domain data repository to support models, grand challenges, reuse and analysis
- Support micropublications
- Support curation
- Supporting chemical data interoperability

The Merck Index* Online





*The name THE MERCK INDEX is owned by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Whitehouse Station, N.J., U.S.A., and is licensed to The Royal Society of Chemistry for use in the U.S.A. and Canada.

See also…

- Open PHACTS data interoperability
  (see Lee's talk)
- PharmaSea (Natural Products to drug leads)
  (www.pharma-sea.eu)
- Topic modelling for chemistry
  (ask me)

Richard Kidd
@rkiddr
@cds_rsc