

**CCLRC DARESBUURY LABORATORY**  
**EPSRC NATIONAL CHEMICAL DATABASE SERVICE**  
**ANNUAL REPORT 2006/7**

**1. REPORT OF THE PAST YEAR**

1.1 The Cross National Services Panel which met in June 2006 decided that there was no longer a need for an EPSRC National Chemical Databases Service, funded by present arrangements. We accept that the Panel's decision reflects the perceived best interests of the community as seen by the EPSRC and that it was carried out with due process. However, we believe that certain aspects of the decision were in many ways flawed. Documents relating to the CDS submission are available at the URL <http://cds.dl.ac.uk/report/review2006.html>. In particular it includes the Response Documents prepared by the CDS and its Management Advisory Panel relating to the Panel Report.

1.2 The Panel's decision clearly had a negative impact on all involved in the Service, but the CDS group has continued to operate as a well integrated team. The previous complement of four full time staff was reduced to three after August 2006 (all have permanent contracts with the CCLRC – part of the Science and Technology Facilities Council since 1/4/07). They include the post doctoral members R.F. McMeeking (Group Leader), D. Parkin and D.A. Fletcher.

1.3 The Management Advisory Panel provides valuable input to the Service. Details of the membership of the current panel together with its Terms of Reference are given in Appendix 1. We make available online copies of the minutes of MAP meetings held since March 2004 at the URL [http://cds.dl.ac.uk/map/recent\\_mins.html](http://cds.dl.ac.uk/map/recent_mins.html). There were no full face to face meetings of the Panel during the Report Period. There were, however, a number of teleconference meetings throughout the Report Period with all panel members able to contribute.

1.4 University outreach through the programme of Roadshow visits continued, albeit at a lower level, during the Report Period. The total number of visits completed for the current Grant cycle has now reached 47. The quota stipulated by the Grant Appraisal Panel prior to awarding the current 3 years funding was 45-50 visits in all.

1.5 All major datasets collections have shown sustained growth during the Report period. Fuller information about components currently supported, including usage details, is provided in Section 2 and Appendix 4.

1.6 We have now removed all dependency on legacy codes. The old CSSR & ICSD still remain available to users but in a frozen state. The last update to CSSR was performed in August 2006, and the OpenVMS system on which the "inversion" process depended has been scrapped. However, in response to user requests the old QUEST retrieval system for the Cambridge Database System has been retained.

1.7 The CrystalWeb interface allows ready access to the full collection of crystallographic databases by a common web browser interface. It has been further refined and developed. In particular closer integration with the ICSD-WWW web interface is well established. As a result of this project a number of problems

associated with the ICSD-WWW system, as supplied by the producers, continue to be identified and corrected. A link from ICSD-WWW to the CrystalWeb structure display facilities has been added.

1.8 The Screening Compound database (compounds for activity screening and diversity synthesis) implemented by the CDS continues to grow. Catalogue information is provided by 24 individual suppliers and the number of available compounds has risen by 24% to 5.6M during the Report Period.

1.9 Access to the molecular databases ACD, SCD and NCI via a web interface has been improved by introducing a structure drawing java applet which replaces the use of the PC specific Chime Pro plugin. Mac users were specifically targeted in this development, but the system is now available to users of all platforms if they have access to a reasonable modern web browser.

1.10 As indicated in the 2005/6 Report, a number of trials for prospective new database systems have been in the pipeline. During the Report Period they reached their final stages of preparation. However, in the light of the decisions made by the Cross Services Panel it no longer seemed appropriate to continue with all of these.

1.11 It is a source of particular regret that we were not able to progress the trial of the SPRESI system for chemical structure and reaction information. This a large database system with over 5 million structures, 3.7 million reactions, and 28 million pieces of factual information. The software has been developed by InfoChem GmbH. It has a very impressive, easy to use, platform independent web browser interface and there is the potential for accessing 3<sup>rd</sup> party databases.

1.12 We were, however, able to continue with the trial of the ACD/Labs I-Lab spectral and properties predictor. This trial proved popular with a significant proportion of the user community, and there was an impressive amount of use made during the trial period. It showed that there was ample justification for the provision of a permanent national central service for this particular system at some point in the future.

1.13 The service computer systems all performed well with no major technical problems during the Report Period. We acquired a Microsoft Windows platform to run the software for the I-Lab trial. This was the first time that we have needed to include the Windows operating system within the Service environment, but all problems were successfully solved.

## **2. USAGE**

2.1 Use of the Service is closely monitored, as is its efficiency in terms of a number of defined performance indicators (see *Appendix 5*). A monthly digest is sent to our Management Advisory Panel (MAP) and the EPSRC. Performance targets are kept under review and tightened up where appropriate. An archive of the digests since January 2004 is available at <http://cds.dl.ac.uk/map/performance.html>.

2.2. 487 Additional users were registered during the period 01/04/06 to 31/03/07. This represents a 33% decrease in new registrations compared to the previous year but only a 12% decrease compared to 3 years ago (pre *Roadshow*). Two Roadshows were held prior to the announcement to cut the organic and spectroscopy databases and none were held after this announcement. After removing 18 “dead” usernames the total number of registered users at the end of March 2007 was 4,679 – an overall increase of 11%. For additional details see *Appendix 2*.

2.3 There has been sustained growth in the size of datasets. Numbers of entries continue to grow at the rate of around 10% per annum. For instance there were over 67K new crystal structures (12% increase) and over 106K new reactions (7% increase) for the year period 01/04/06 to 31/3/07. The Service now provides access to nearly 1.6M searchable reactions, and around 5.6M compounds. For a summary of currently available components see *Appendix 4*.

2.4 Use of all the databases has fallen over the year (see *Appendix 2*) - which may relate to the fewer Roadshows. Clearly another factor has been the uncertainty of future provision. It is also clear that, given the circumstances, fewer in house training sessions relating to components of the CDS have been given at universities as part of new graduate teaching programmes.

2.5 Although the total number of accesses to the crystallography databases has fallen (by 5%), the number of unique users of the main crystallography packages has gone up slightly (35 over the year). The ISIS databases saw a 12% overall reduction in users per month and 17% in accesses. Ironically the trial of the ACD Labs spectroscopy package probably accounted for 804 lower accesses (42%) to SpecInfo over the accounting year.

2.6 There has been a change in the pattern of usage with a greater uptake of the improved web browser access facilities which we have made available. This typically results in more intensive single sessions involving multiple searches. For example, CrystalWeb users and accesses have increased by around 15%. Following on the introduction of a structure drawing Java applet for use with the ISIS molecule databases, the ratio of web based usage as compared to that via ISIS/Base has changed from around 25:75 to 40:60.

2.7 All uptime on the Service computers was available for CDS use. In general, the servers run 24 hours per day and 7 days per week. During the full Report Period there was effectively 100% service availability with no unscheduled downtime. There were no major instances of note except for instances of external network (JANET) problems, which may have affected certain users.

2.8 Accesses per month to the core database systems oscillated between 4,089 and 7,193 with an average of 6,074, representing a decrease of 13% over the previous Report Period. The number of active users within any given month varied between 610 and 788 with an average of 713, representing a decrease of 8%. Figures for users do not include Detherm users, since the introduction of version 2.1 in June 2006, as these are not recorded as CDS ID's, but accesses to Detherm have risen by around 8%. Again fuller details are given in *Appendix 2*.

2.9 In this Report we do not attempt to match source of funding for individual users with their CDS access, but note that most use comes from departments with EPSRC Chemistry grants. The registration form now requests “main research funding source (or graduate funding body)”. Currently around 50% of new users provide their funding details on registration. Such details can now also be added or updated via a simple web interface, and we will issue a general call for all users to do so as part a general exercise to update all registration details.

2.10 The use of the CDS amongst all **chemistry** departments continues to correlate well with the grade allotted by the Research Assessment Exercise (see *Appendix 3*). Users from chemistry departments with a 2001 RAE rating of 4 or higher accounted for 87% of all registered and all active users and 91.1% of all accesses from chemistry departments.

2.11 Our policy continues to be to grant access for eligible use to all UK academics who apply. User registration is now exclusively via an online form on the CDS web site. This is efficient in term of operational procedure and there are no discernible bottlenecks relating to resource availability. Registrations are completed and returned on the same day as receipt of the associated authorisation slip. A fax option as an alternative to receiving registrations by posting was introduced in 2002 and now accounts for ~70% of all application received. This excludes registrations at *Roadshows* - where all applications are fully completed there and then.

2.12 A system whereby users who have a valid "ac.uk" address can choose to be sent their username and password by email rather than post was introduced in 2004. This further reduces the turn round time and is also proving very popular, with over 92% of users being dealt with this way. Around 40% of registrations during the Report Period were from staff or research fellows and just under half (45%) were from graduate students.

### **3. PUBLICITY**

3.1 The CDS web site (<http://cds.dl.ac.uk>) continues to be a key component of the Service and a major vehicle for publicity. It served on average, 9,174 distinct requesting sites per month during the period 01/04/06 to 31/03/07, an increase of 6% over the previous year. There was, on average, 10,637 successful requests per day with an average of 2,710 successful requests for pages per day.

3.2 Revisions of the web site continue. In particular these have including upgrades to the homepage. As part of this we have improved our procedures to display current and relevant news items.

3.3 Another important way of distributing information about new Service features to registered users is via broadcast email messages. Broadcast messages usually include links to fuller details on our web site. We invest significant effort in keeping our email lists accurate and up-to-date (using "ac.uk" addresses where practicable). This helps us to identify those users who move site and/or leave the academic community.

3.4 The Service produces a biannual CDS Newsletters (latest editions dated April 2006 and Winter 2006). The current version and back issues from 1994 are also available on the CDS web site (<http://cds.dl.ac.uk/newsletters>). All users are sent email notification of the release of a new Newsletter which includes a link for electronic download. Less than 10% now elect to receive it in print format.

3.5 Publicity material continues to be sent to User Representatives for display within departments. There is a CDS Overview PowerPoint show (this was developed for use in conjunction with Roadshow visits), which is available for display and/or downloading from the CDS web site at <http://cds.dl.ac.uk/overview.ppt>. This is advertised as being made free for copying and distributing within departments.

#### **4. RESEARCH HIGHLIGHTS**

4.1 We include research highlights to provide a brief snapshot of *some* of the fields of study which have benefited from use of the Service. The articles themselves are included in the separate *Research Highlights* document with this Report. A complete archive of the research highlights from previous years is available at [http://cds.dl.ac.uk/report/res\\_high.html](http://cds.dl.ac.uk/report/res_high.html).

4.2 It is in the nature of things that Service personnel do not have direct contact with the majority of the work carried out. We ask users to acknowledge the Service in their published work, and we also have a 1996 review article which we encourage them to cite. This allows us to track published paper using the *ISI Web of Science*. There were 82 citations from journals published in April 2006 to March 2007. Further details are given in *Appendix 6*.

4.3 A User Survey conducted last year implied that we are considerable under represented by the actual citations made, at least twice as many articles are produced by users who use the Service but do not cite the Service.

#### **5. FUTURE PLANS**

5.1 We are looking at the possibility of making the web interfaces accessible via Shibboleth technology for members of the UK Access Management Federation for Education and Research.

5.2 LitLink will no longer be available for literature linking in CrystalWeb as it was part of the software licensed through MDL, which we no longer have funding for. However, this function will now be handled using the CrossRef system, or directly with Digital Object Identifiers (DOIs) where they are available.

5.3 CDS have had discussions with people from the Southampton University Crystal Structure Report Archive and it is proposed to include entries from this archive into CrystalWeb.

5.4 We plan to continue improving CrystalWeb. The procedure for calculating bond connectivity in the CrystalWeb structure display will be modified for ICSD and

CrystMet entries so that it gives a closer match to "chemical reality". We also plan to introduce features which allow the user to modify the connectivity criterion and also add or delete individual connections. Work on the ability to selection individual structural moieties for output in the full range of available molecular modeling formats is also well advances. On the fly selection of individual atoms for radial distribution calculation will also be added.

5.5 Work on setting up a site subscription service for the organic chemistry components, which have been dropped from the current CDS portfolio, will be progressed. This was discussed at a meeting in late march 2007. This involved the CDS, JISC, CHEST/EduServ, EPSRC and representation from the academic community.

5.6 Such an arrangement would be sponsored by JISC, who are keen to go ahead subject to sufficient formal expression of interest by the community. The canvassing process is planned to be completed by the end of June. However, we have encountered potential contractual problems with one of our current software suppliers, Elsevier MDL. They have the power to obstruct the deal, as currently envisaged, if they believe it to be against their commercial interests.

5.7 The EPSRC has committed to putting the Crystallographic and Thermophysical components of the full service out to open competitive tender. This process is scheduled to be completed and new arrangements in place on 1<sup>st</sup> April 2008. The contract will be for a 3 years period in the first instance. The CDS will continue to provide these components until that time with funding at the level of two posts The CDS will also tender to provide the new service.