



---

# Chemical Database Service

## NEWSLETTER

Issue 12

May 2000

---

The **BioCatalysis** database, which contains organic reactions catalysed by biomolecules, is now available as a permanent part of the service. It is described in more detail on page 2. It is one of the growing number of **specialist databases** which are accessible using ISIS. The full collection is reviewed in the article on page 6.

Thanks to everyone who responded to the **survey** we ran last month. The results have given us information which will be extremely useful in planning the future direction of the service. A summary of the responses is given in the article on page 4.

**ISIS/Client** version 2.3 is now available for downloading from the CDS website. This is the last version of ISIS to be supported on the **Macintosh**. See page 9 for more details.

This year's **user meeting** will be held on Thursday 13th July at Daresbury. All users are invited to attend. More details on page 10.

Recent changes to the service include the introduction of a new crystallographic file format converter, **bedlam** (see page 8) and the new interface for the Cambridge Structural Database, **ConQuest** (see page 10). In addition, the CDS website now has a full text search ability, described on page 7.

The main CDS server will be upgraded next month. See page 7 for details.

### Contents

BioCatalysis database.....	2	ISIS/Client supported	
Survey results.....	4	platforms .....	9
Specialist ISIS databases.	6	User meeting.....	10
CDS server upgrade.....	7	ConQuest.....	10
Website search engine.....	7	Databases available.....	11
Bedlam.....	8	Useful information.....	12

The Chemical Database Service is funded by the Engineering and Physical Sciences Research Council



## BioCatalysis database

Following from the user trial of a number of life science related ISIS databases, we selected the BioCatalysis database from Synopsis to be a permanent component of the Service. The decision was reached on the basis of responses from the user questionnaire, expected usage, and cost.

Synopsis' BioCatalysis database has been developed to recognise the increasing importance of biomolecules as catalysts in organic synthesis. Users of the database will discover for themselves the wide range of possibilities offered by biocatalysts, used to effect novel transformations and as versatile replacements for traditional catalysts. The database focuses on the synthetic use of enzymes and micro-organisms, which can offer advantages of excellent chemo-, regio- and entantio-selectivity, coupled with important environmental benefits. It currently contains 26,479 reactions.

Selection of material which is included in the final databases is the joint responsibility of two internationally recognised experts, Professors Bryan Jones (who is based in the University of Toronto) and Herbert Holland (Brock University). The data is selected to illustrate the types of reactions catalysed, with an emphasis on preparatively useful transformations. In addition to enzymes and whole cells, the database also includes information on enzyme mimics, enzyme analogues and catalytic antibodies.

The database's selectivity ensures users are presented with relevant, up-to-date information that can be searched electronically. Specific enzymes or micro-organisms may be retrieved using accepted names or synonyms, CAS number and EC numbers. Further data includes the citation, title, biocatalyst supplier, yield, enantiomer excess, diastereoisomer excess, and an informative abstract, intended to highlight the scope and limitations of the reaction.

### Key features of the database

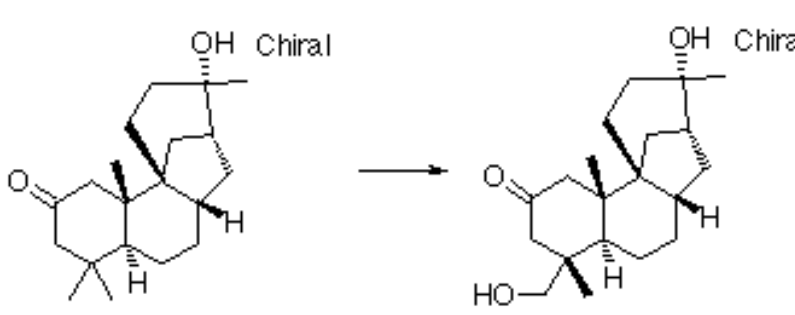
- Thorough coverage of the scientific literature, including patents, from 1903 to the present.
- Keyword and text options offered alongside user-friendly structure searching facilities.
- Regular updates, totaling approximately 2,000 reactions per year, keeping the database up to date.
- Developed in collaboration with academic experts in the field, Professors Bryan Jones (Toronto) and Herbert Holland (Brock).
- Examples include virtually every type of chemical transformation.

## Key advantages of using bio-molecules as catalysts

- High chemo-, regio- and enantioselectivity.
- Novel transformations become possible.
- Ecologically sound - non-toxic biomass, energy efficient
- Effective under mild conditions; ambient temperature & pressure, neutral pH, aqueous and inert solvents.

## Accessing the database

Database access files for PC, Mac and SGI are available to download from the CDS website. In addition, BioCatalysis is available in the RXN Browser, and is searched by default together with the main reaction databases.

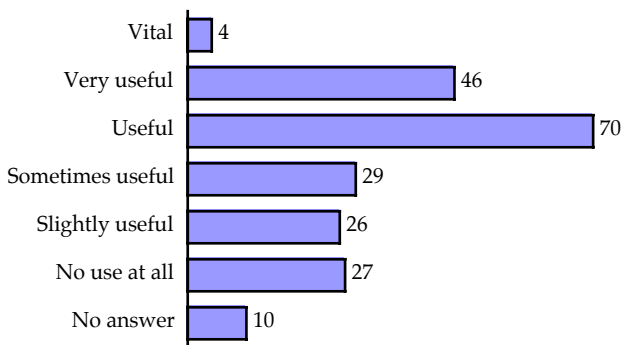
Main Form	Catsol	Biocat Data	Molecule Data	Other Data	Table
<b>Biocatalysis</b>		RXN 2 of 270	Variation 1 of 1	Path	Step
Cluster #	Item #	Value:			
					
Title      Biotransformation of some stemodane diterpenoids by <i>Cephalosporium aphidicola</i> .					
Conditions	Ref.      Hanson J R, Reese P B, Takahashi J A, Wilson M R, <i>Phytochemistry</i> , 36() p. 1391, 1994				
<i>Cephalosporium aphidicola</i>	Comments      A trace amount of a dihydroxylated metabolite of tentatively assigned structure was also isolated. In the absence of the keto-group a complex mixture of hydroxylated metabolites was obtained, each in a very low chemical yield. For experimental details and preparation of the culture medium see: <i>Phytochem</i> , 1992, 31, 3851).				
Chlorocholine chloride					

Retrieved reaction from the BioCatalysis database. Six different forms, showing different retrieved data, can be selected using the row of tabs.

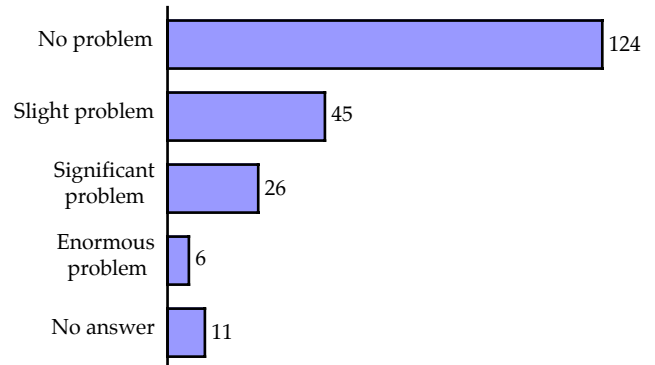
# Survey results

CDS ran a web based user survey during April of this year in order get feedback from the community on a number of issues relating to the future development of the service. In all there were 212 responses and the breakdown for each question is shown below. We would like to take this opportunity to thank everyone who took the time to fill in the survey.

**1. How useful would it be to you if the CDS resources were available using Athens authentication**

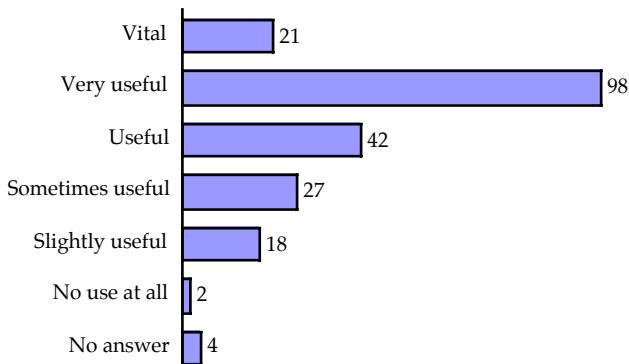


**2 Would it be a problem to you if the CDS was only accessible using Athens authentication**

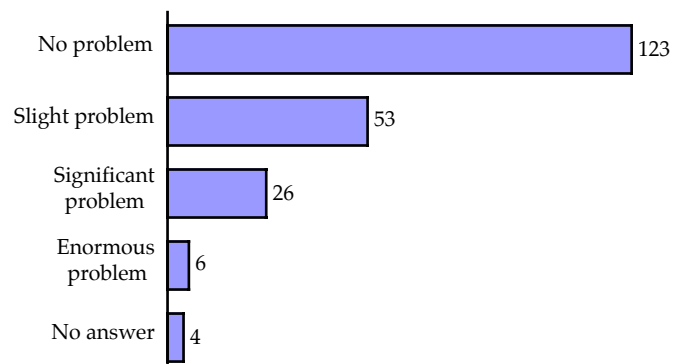


The first two questions were included since we have had a number of requests to use Athens authentication (the Athens system is a single username/password used to access a variety of different information sources, mostly JISC funded services - see <http://www.athens.ac.uk> for more details). We are currently investigating the technical issues with using Athens. It is possible that some, or maybe all, of the CDS services will become accessible using Athens. We do not intend that Athens authentication will become the sole means of access any of the CDS resources.

**3 How useful is it to you to have the CDS databases available via the web**

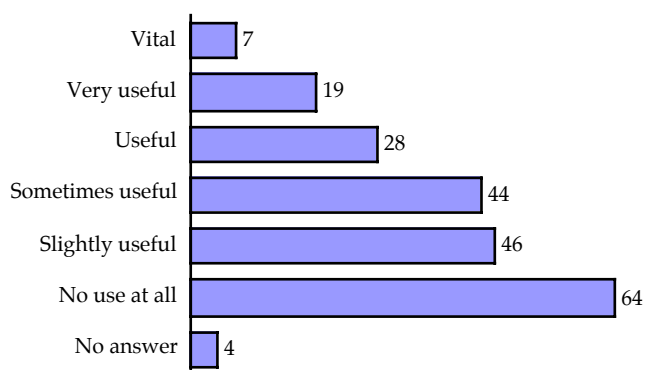


**4 Would it be a problem to you if the CDS databases were only available via the web**

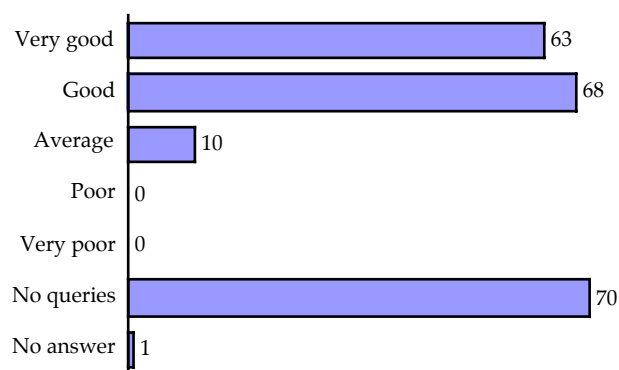


The second two questions were to check out user response to the current trend towards web interfaces. Whilst there was a very favourable view of web access, there was also a significant minority with reservations. From the comments received, these related mostly to the speed of web based interfaces. Introduction of web interfaces to the CDS databases will continue, but we see them as complementary to, rather than replacing, the current interfaces.

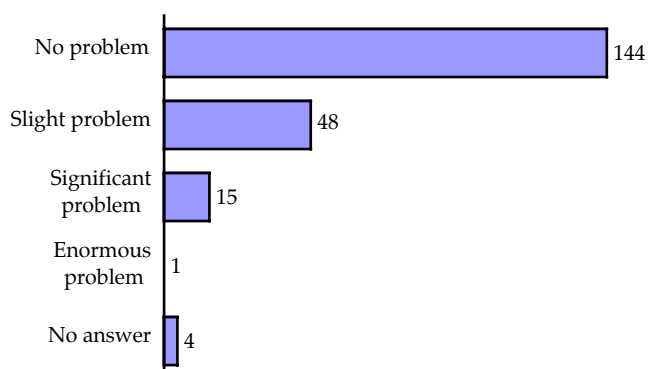
**5 How useful would a database of physical and thermodynamic properties of mixtures be to you**



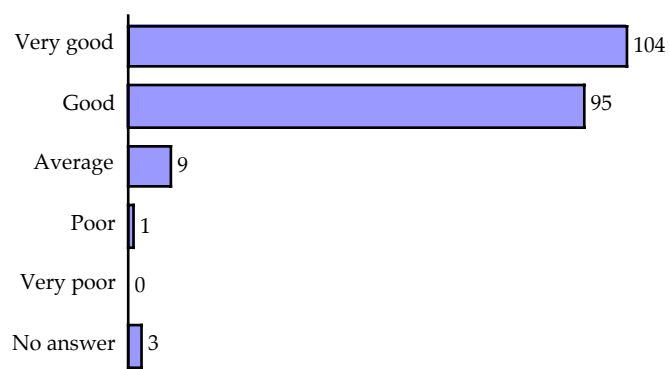
**6 How would you rate the support (response to queries) provided by the service**



**7 Would it be a problem to you if you could only submit queries by email and not by phone**



**8 How would you rate the Chemical Database Service overall**



Question 5 was to get some idea of the potential takeup from the current user community for the DETHERM database, a trial of which we are hoping to run, or other thermophysical data.

The other questions were to get some benchmark numbers. Note that we have no plans to stop phone support.

Most of the comments received were either enthusiastic about web interfaces or expressing reservations about them, particularly with respect to speed of use.

## Specialist ISIS databases

There are five specialist reaction databases currently available under ISIS. These are smaller datasets dealing with a particular area or technique. A brief description of each is given below.

There is a separate database access file available for each of the databases from the download section of the CDS websites since in many cases it makes sense to restrict a search to a specific database. In addition, SPG, SPS and BioCatalysis are included in the Reaction Browser, and are therefore searched by default along with the main reaction databases.

### **Protecting Groups (SPG)**

The Synopsys' Protecting Groups database provides access to selected information on the full range of protecting group chemistry. The focus is on selected protection and deprotection reactions and the conditions under which the protecting groups are stable or labile.

The database currently contains 31,080 reactions, covering the literature to January 2000. Stability data is now present for over 63% of the entries, while lability data is present for over 85% of the entries.

### **Solid Phase Synthesis (SPS)**

The Solid-Phase Synthesis database from Synopsys contains a selection of relevant material from primary literature, reviews and patents (1963 - 2000) from over 2,000 publications. The Focus is on Chemical transformations, selected to illustrate the variety of reaction types which have been demonstrated on the solid phase. Emphasis is placed on 'small-molecule' transformations but 'state of the art' methods and new developments in the areas of peptide, oligosaccharide and oligonucleotide synthesis are also covered. The latest release contains 12,639 reactions, abstracted from 2,268 citations, covering the literature to 2000.

### **Chirbase**

The Chirbase database contains data (chiral stationary phase and operating conditions) for some 60,000 chiral chromatographic separations of enantiomers (almost 40,000 entries in the database) dealing with more than 18,000 independent structures.

## BioCatalysis

The BioCatalysis database contains some 26,500 organic reactions which use biomolecules as catalysts. It is described more fully in the article earlier in this newsletter.

## Molecular Diversity Preservation International (MDPI)

We now make the database of over 8,000 compounds compiled by Molecular Diversity Preservation International (MDPI) available. Samples of these rare compounds, which are not available through normal commercial channels can be exchanged, deposited or sold through MDPI.

## CDS server upgrade

The main CDS server, cds3, is going to be upgraded at the start of next month. In effect it will be replaced by a more powerful machine, however the name and IP address will not be changed so the upgrade will appear transparent to users. We anticipate that there will be less than one hour of downtime required for the upgrade and this will be advertised in advance.

The upgraded cds3 will have two R12k 270 MHz processors (compared with the current R10k 180 MHz processors) and memory is being doubled to 512 MB. This should produce a 70% increase in cpu power. The 'old cds3' will be used to provide an additional 'back end' server which will, initially, handle some of the ISIS load. Again, this process will be transparent to users. The combination should significantly improve the throughput of searches and response time during peak periods.

## Website search engine

The CDS website has been enhanced with the addition of a search facility. It uses the ht://Dig search engine and provides a full text search service for the whole of the publicly viewable site. Password protected sections, such as the ISIS download area and ICSD web interface, are not currently searched.

Once an initial search has been made, the hit list returned allows for refinement of the search using all or any of the search words specified or more complex Boolean expressions.

# Bedlam

Bedlam is a generalised crystallographic file format converter which in many ways supersedes existing CDS utilities (e.g. `xr2cam`, `cam2xr`, `xr2pdb`, `xr2shx`). The syntax for using bedlam has been designed to be very similar to the widely used converter program, `babel`. The range of file types currently supported by bedlam is much less than that supported by `babel` (and is likely to remain so). To handle formats for a whole range of molecular modeling packages `babel` remains the package of choice.

In common with the various other CDS utilities, bedlam handles crystallographic data better than `babel`. Thus fractional coordinates are retained when output is to a crystallographic format file (`babel` always converts to orthogonalised coordinates). Also, crystallographic symmetry operation information, when present, is transformed and carried over to the output format.

The basic syntax for using the program is

```
bedlam -i<input-type> [input-file] -o<output-type> [output-file] [options]
```

## Currently supported input types

<code>xr</code>	.xr file (generated by <code>cssr</code> , <code>icsd</code> , <code>mdf</code> , etc)
<code>cssr</code>	same as <code>xr</code> (included for consistency with <code>babel</code> )
<code>fdat</code>	CSD FDAT file
<code>shelx</code>	ShelX file
<code>free</code>	CDS free format file

## Currently supported output types

<code>xr</code>	.xr file (generated by <code>cssr</code> , <code>icsd</code> , <code>mdf</code> , etc)
<code>cssr</code>	same as <code>xr</code> (included for consistency with <code>babel</code> )
<code>fdat</code>	CSD FDAT file
<code>shelx</code>	ShelX file
<code>free</code>	CDS free format file
<code>bccab</code>	CSD BCCAB file
<code>pluto</code>	CSD PLUTO free format file
<code>xyz</code>	XYZ file
<code>pdb</code>	PDB file
<code>alc</code>	Alchemy file
<code>mdl</code>	MDL Molfile file

## Options

-h	Display the current Help File
-c [utility-program]	Apply a conversion program. If no qualifier is given the default is to use the umol package (calculates connectivity using the standard CSD distance criteria)

## Further details

If no input or output filenames are specified bedlam will use standard input or standard output. If no specific input or output format is given .xr format is taken as default. Bedlam supports multiple entry files for all formats. Thus if you wish to convert a series of .xr files to FDAT format, you should first concatenate them to a single file.

By default bedlam does not recalculate connectivity data, but retains any such data present in the input file. Use the "-c" option to recalculate connectivity.

Bedlam is still under development. If you encounter problems or require clarification about details of current features or those being developed contact [r.f.mcmeeking@dl.ac.uk](mailto:r.f.mcmeeking@dl.ac.uk).

## ISIS/Client supported platforms

ISIS/Client (includes ISIS/Base and ISIS/Draw) version 2.3 has just been released for PC and Macintosh and is available to download from the CDS website.

This will be the last version to be supported on the Macintosh. We believe that it will run under Mac OS9, but have not been able to test this. We expect that version 2.3 will continue to work fine with ISIS/Host on the CDS server for some time (probably several years). Support for the SGI client was dropped over 2 years ago, but the last version still works with the current version of ISIS/Host.

In the longer run, access to the ISIS data is likely to be via web-based clients. We hope to have a system running, using Chemscape and Chime Pro on the CDS server, later this year which will provide an alternative to ISIS/Base for searching the organic reaction data.

Although MDL have stopped the support of ISIS/Client for Macs, they remain committed to supporting web-based clients based on Chime Pro and Java on the Macintosh.

## User meeting

This year's user meeting will be held on Thursday 13th July at Daresbury. All users are invited to attend and Department reps are particularly encouraged to attend.

Current topics for discussion this year include:

- Web based interfaces
- Athens authentication for CDS
- Future funding of the service
- New systems and future developments

As in previous years we will be able to cover the travelling expenses of attendees. If you would like to attend, please contact us at [cdsbb@dl.ac.uk](mailto:cdsbb@dl.ac.uk)

## ConQuest

ConQuest is the new graphical user interface (GUI) for the Cambridge Structural Database (CSD). Like its predecessor, quest, ConQuest is an X-Windows application.

ConQuest is a much more intuitive and easier to use interface than quest. Features include:

- Intuitive sketcher for drawing chemical substructures and defining geometrical parameters.
- Straightforward mechanisms for combining queries to generate complex searches.
- Interactive browsing of search results while searches are running.
- Extensive range of output file formats (CIF, PDB, MOL2, ShelX, etc.).
- Context dependent help, full documentation and tutorials.

The ConQuest interface has just been installed with the latest CSD release. It can be started by logging on the cds3 and typing 'conquest'.

Although ConQuest will be the primary GUI for accessing the CSD, quest will remain available for some time, though it will not be developed further.

## Summary of databases available from the Chemical Database Service

A brief description of all the databases currently available from the CDS at Daresbury.

**ISIS** A chemical reaction information management system allowing search, retrieval and display of molecules, reactions and their associated data. Currently contains 960,000 searchable reactions. Specialist databases available include Protecting Groups, Solid Phase Synthesis, BioCatalysis and Chiral Separations.

In addition there is the Available Chemicals Directory (ACD) which contains supplier information on 270,000 different compounds from over 500 different suppliers.

**SPEC** SpecInfo is a multi-technique spectroscopic database package intended to cover NMR, IR and mass spectra. The package is designed to aid the chemist in spectral interpretation and structure elucidation problems. The database currently contains 99,059  $^{13}\text{C}$  NMR; 999  $^{15}\text{N}$  NMR; 856  $^{17}\text{O}$  NMR; 2,183  $^{31}\text{P}$  NMR; 1,825  $^{19}\text{F}$  NMR; 5,000  $^1\text{H}$  NMR; 20,898 infra-red spectra and 128,000 mass spectra.

**CSD** The Cambridge Structural Database. Crystal structure data for over 215,000 organic and organometallic compounds. Currently available under Cambridge's Quest and ConQuest retrieval software and our in-house software (CSSR).

**IsoStar** A database of non-bonded interactions taken from the CSD, PDB and theoretical calculations.

**ICSD** Inorganic Crystal Structure Data File. About 53,000 inorganic structures - the companion file to the Cambridge organic file.

**MDF** Metals Data File. Crystal structure data for nearly 56,000 metals, alloys and intermetallics.

**CDIF** Crystal Data Identification File. Crystal class and unit cell data for over 237,000 crystal structures.

**FNMR** A databank of 6,000  $^{19}\text{F}$  NMR spectra and coupling constants.

**ELYS** Electrolyte Solutions Database. Thermodynamic and transport property data such as density, viscosity and diffusion coefficients. Currently contains about 10,000 entries.

# Information

## General queries:

Computer Help Desk/User Interface  
Group (UIG)  
Daresbury Laboratory  
Daresbury, Warrington  
WA4 4AD  
(01925) 603351  
Email [uig@dl.ac.uk](mailto:uig@dl.ac.uk)

Hours are:  
Mon-Thur 9:00-12:00, 13:00-17:00  
Fri 9:00-12:00, 13:00-16:00

Otherwise 24 hour answer phone.

## General database queries:

Chemical Database Service (CDS)  
Room C18, Daresbury Laboratory  
Daresbury, Warrington  
WA4 4AD  
(01925) 603162  
Email [cdsbb@dl.ac.uk](mailto:cdsbb@dl.ac.uk)

## For ISIS specific problems:

Dr. D. Parkin  
Email [D.Parkin@dl.ac.uk](mailto:D.Parkin@dl.ac.uk)  
(01925) 603162

## For SpecInfo specific problems:

Dr. D.A. Fletcher  
Email [D.A.Fletcher@dl.ac.uk](mailto:D.A.Fletcher@dl.ac.uk)  
(01925) 603162

## For crystallography specific problems:

Dr. R. F. McMeeking  
Email [R.F.McMeeking@dl.ac.uk](mailto:R.F.McMeeking@dl.ac.uk)  
(01925) 603162

## World Wide Web Site

Information about CDS, including  
online help and documentation is  
available over the World Wide Web at  
the CDS website, URL:  
<http://cds.dl.ac.uk/cds>

## Registration of new accounts:

You can now register online for a CDS  
account from the CDS website.

## Documentation:

Most documentation is available online  
and some can be downloaded from the  
CDS web site. Printed manuals can still  
be ordered from UIG at the above  
address.

## Access:

The service is available free of charge to  
UK academic research groups for non-  
commercial work. Each individual user  
will be issued with a unique ID. It is not  
our policy to allow shared IDs.

## cds (main service machine):

Internet number: 193.62.124.35  
Internet name: [cds.dl.ac.uk](http://cds.dl.ac.uk)

## Comments:

All comments, questions and  
suggestions about this newsletter  
should be sent to:

Dr. D.A. Fletcher  
Email [D.A.Fletcher@dl.ac.uk](mailto:D.A.Fletcher@dl.ac.uk)