# DISCUSSIONS RELATING TO AWARDING GRANT GR/S70623/01

# (i) Service Response dated 10th November 2003

## Introduction

There are five general areas of concern raised in the Grant Review Panel Report, which we understand require a formal response from the Service. They are *Overall & Longer Term Strategy, Service Staffing, Database Portfolio, Publicity*, and the *Management Advisory Panel.* We also include a revised *Funding Plan* and details of *New Databases*. Our responses are structured in line with the recommendations (bullet points in section 6) of the panel report.

## 1. Overall and Longer Term Strategy

1. *The panel recommended that the proposal be funded for the 3 years requested. Apart from the DETHERM element, the panel did not consider that the funds were sufficiently justified and brief focussed justification was required for all other costings.*
2. *The panel were concerned with the lack of awareness and potential low use in certain user areas, particularly in the synthetic organic, bioscience and chemical engineering communities. If after the enhanced programme of publicity, certain databases are still under used by the community then the issue of continuing to support them in the current manner must be addressed in any subsequent renewal.*
3. *The panel recommended that the service be reviewed one year into the grant. The review would assess the development of the service in light of recommendations and in particular how the publicity element of the service had developed. If the review panel do not consider the service to have satisfactorily addressed the issues raised in the renewal (particularly in terms of publicity) they can recommend that an open call be announced a year early, with the current service funding ending after two years.*

We will address questions of further justification for various elements of funding referred by *Bullet 1* within the later sections of this Response. In particular the *Publicity* Section is a key area, and we have outlined various spending revisions in line the Panel's concerns. *Bullet 2* relates to database use and support as well as publicity, and raises a whole range of issues which encompass most of the later Sections. In addition and because of the various revisions in proposed spending (particularly relating to publicity) we have produced a new *Funding Plan*.

We welcome the main recommendation in *Bullet 1* that the Chemical Database Service continues to be funded for 3 years, but would like to take the opportunity to address some of the wider issues implicit in *Bullet 2* and at other points in the full Report.

The grant we submitted was deliberated framed for a relatively short term. The reason for this was that we were given the reduced time frame of 3 years in which to operate, and there had been no specific commitment to the Service by the EPSRC beyond that point. Longer term developments within the chemical information world cannot be predicted in the medium to long term with any degree of certainty. But more importantly we had been lead to believe that, assuming a National Chemical Database Service was to continue beyond that point, it would be put out to competitive tender.

It is likely that Daresbury will wish to participate in the projected tender exercise. However, in an open forum we do not believe it appropriate to spell out details of any specific plans we are forming at this stage. We may wish to explore possible co-operative alliances and any tender submission is also likely to include elements of commercial confidentiality.

This said, the Service at DL is keen to interact with the EPSRC Chemistry Programme in initiatives to increase the scope and quality of chemistry related data available to the UK research community. The steer we have received has changed over time. We welcome the current emphasis on providing more bioscience related chemistry resources and those of value to the material sciences and engineering communities. This fits in well with what had been hoping to be able to achieve on previous occasions.

The question of CAS SciFinder has also been raised. DL has a proven track record in this type of venture. For instance, it played a key role in establishing the Beilstein/CrossFire service for the UK academic community. It is in many ways unfortunate that the Chemistry Programme was unable to find the resources to ensure the Beilstein/CrossFire could continue to operate by its own Service and thus ensure a more coherent and cost effective provision of chemical database resources for the UK community as a whole.

Don Parkin is the current chairman of the RSC Chemical Information Group. He will use this position and associated contacts to further details about current deals available, for instance to overseas academic communities, and make soundings about potential future co-operative deals for the UK. Here active involvement of the EPSRC will further strengthen his position.

## 2. Service Staffing

1. *The panel were concerned that all the staff employed on the grant were at senior salary grades and this may not be cost effective for the service; all but one of the staff members had worked on the database service for over 10 years. The panel noted that staff members had high technical expertise; however their skills and competencies may not be totally suited to current needs of the service. The issue of total numbers of staff and their competencies should be addressed in conjunction with EPSRC prior to the next renewal.*
2. *The panel considered the costs itemised under 'other staff' were not appropriate. The administrative support was not considered justified; with the management costs listed being ineligible. These costs could be covered through the overheads received through the grant.*

The presentation given by DL personnel on 4/8/03 was clearly not as informative as we might have hoped. We suspect that the Panel reached a view which does not reflect the range and depth of competencies which are required to operate the Service at its current state of development. For instance the statement "their roles included carrying out low level support work such as transferring data from one database to another" is not helpful.

It is true that certain tasks have become more straightforward with the advent of more modern, reliable hardware and various software innovations planned and implemented by the CDS team. This has been more than balanced by the need to embrace and implement new technologies. A few recent examples include the CrystalWeb initiative, implementation of LitLink, adopting relational database technology, reconfiguring computer hardware and software, and setting up trials for new database systems such as DETHERM. Computer system administration expertise is provided from within the CDS team, and a considerable degree of skill and innovation has been required to maintain and enhance what is widely recognised as a robust and reliable service.

Statements about the current management arrangements of the group are incorrect. The group does have a coherent management structure. McMeeking is the group leader with Parkin and Fletcher reporting to him, while Osa-Edoh reports to Parkin. This group structure does provide a framework within which high quality decisions are indeed made and future strategies developed and implemented.
There are also extensive and worthwhile interactions with the user community. For instance the development of the CrystalWeb project stemmed from solutions produced by the CDS team for a consistent and web-based interface. It arose as a response to explicit requests sent to us by members of the user community.
We recognise that inclusion of the cost item under the 'other staff' heading was inappropriate, and this request has been dropped. In addition the figures for salaries of named individuals were incorrect in Table M of Form Je-SRP1. The figures given under *the Basic Starting Salary* heading were, in fact, inflated to cover anticipated increments over the grant period. The figures given under the *Total on the Grant* heading presented in the form Je-SRP1 are correct. The amended starting salary figures based on basic starting salaries are: £104,572 (RFM), £93,963 (DAF), £89,106 (DP), and £75,549 (DOE). We have adjusted our costings spreadsheet to reflect these changes.
We were asked to comment on the increase in staff costs applied for in the proposal as compared to the one for 2001/5. Our amended figures make this discrepancy less marked, but there is still significant inflation of costs. The increase reflects standard salary increases at the CCLRC over the current grant period. In addition there have been increases related to the promotion of DP. The starting salary for DOE was greater than that we entered on the proposal application form for 2001/5 (here the figures followed the standard guidelines relating to unnamed staff). The salary offered follows standard Research Councils procedures and reflects the necessity to attract a suitable candidate to perform the duties required for the job.

## 3. Database Portfolio

1. *The panel considered that the service should be at the forefront of database provision and agreed that the addition of DETHERM to the portfolio of databases supported by the service was justified. The service should develop of strategy for negotiating a licence a national UK Sci Finder; the formal contact with the RSC information group could be a route for engaging with the key stakeholders.*
2. *The panel considered that the service should develop a formal strategy for the turnover of databases. The panel were concerned that there appeared to be no formal criteria for ceasing to support a database. In addition, criteria should also be drafted for acquisition of new databases, including smaller databases where no trials are to be carried out.*

Discussions by the Panel as reported in section 5.13 raise some interesting issues. We were already aware of the danger that a section of synthetic chemists might be unaware of or ignore the complementarity of the ISIS databases with SciFinder and Beilstein/CrossFire. This is clearly not the case for many highly regarded research groups as attested by the *Research Highlights* included in the *CDS Annual Reports* and available as supplementary material to the funding proposal. For specific examples of such work see the articles by Armstrong, Parkin, Davis, Brown & Moloney (we also make available the collection of highlights at http://cds.dl.ac.uk/report/res_high.html). The hope is that our enhanced publicity effort will re-enforce this message more effectively within the community.
We are also aware that sections of the physical chemistry community have access to much material electronically (e.g. some of the excellent and currently free material

provided by the NIST website). However, we do believe that the DETHERM database is a boon to many physical chemists and will provide data which cannot be obtained elsewhere. It will be a focus of attention and the CDS will become the logical place to house further physical chemistry and related datasets.

The *New Databases* costing item has been a source of confusion. Following a funding review held in 1994, the Panel instructed us to include extra in the funding line for further but as yet to be identified database systems. This amounted to £50K for the 2 years. Formal appraisal criteria outlined in Appendix L were framed to assess any relatively expensive items which might become available within this category. The *New Databases* spend was reviewed by the Panel again 2 years later and a similar budget confirmed We continue to develop a better feel for reasonable potential likely overall costings without knowing precise details of the specifics of such acquisitions.

DETHERM was discussed at the time of our 2001/5 proposal. In this case we had a rough idea of potential costings and would certainly have liked to have included this under the *New Databases* heading. This turned out to be unrealistic. It was considered not to be appropriate since the costs would put it in the *Large Capital* category. The steer was that funding for DETHERM should be by a specific Responsive Mode grant bid. This whole process increased the time and cost taken to establish DETHERM as part of the service, but allowed us to assemble a considerably more impressive case for this component.

Various items have been acquired using the *New Databases* allocation during the current funding cycle. They include LitLink, Spectroscopic Data, and DETHERM licensing to cover the extended trial. User evaluation trials have already been organised for a number of Accelrys systems (the Failed Reaction, Bioster and Metabolite databases). If these trials are successful and suitable terms are available the systems will be acquired using the *New Databases* allocation available during the next funding cycle. Various smaller items (e.g. spectroscopic datasets including $^{19}F$, $^{29}Si$, $^{31}P$, specialist $^{13}C$ collections, MS & IR) have also been identified which will probably not require user trials (in line with our current *Appraisal Protocol)*. For further details of past acquisitions in this category see the *New Databases* Table (page 8).

The issue of older and less intensively used database and related systems is indeed kept under review by the Service and has been a topic of discussion at Management Advisory Panel meetings. Some of the smaller and older systems are still of great value to a sector of the user community. Others will become totally redundant when we acquire new systems (we will remove ELYS when DETHERM comes on stream again). We will remove the FNMR database as part of the imminent transfer to the new CDS server. We will also get rid of a range of utility packages which have been superseded because of more recent improvements introduced by the Service.

We have been asked to comment on other database costing issues; in particular the rise in the cost of the MDL ISIS systems in the current proposal as compared to the one submitted for 2001/5. In the 2001/5 proposal we used the standard government inflator of ~2%. In fact MDL prices have been rising at ~4% (the level at which they are capped by our contracts). More importantly, CDS was previously billed by MDL from the USA and VAT was not included with the invoicing (but equivalent charges are recovered by HM Customs & Excise recovered from the CCLRC). In fact, we have modified these arrangements and are now invoiced from the UK. The new costings reflect the true VAT situation. In our current estimates we use 4% for MDL costs and the lower inflator for other systems.

## 4. Publicity

1. *The publicity strategy proposed by the service was not considered sufficient. The service should develop a strategy involving visiting approximately 40-50 institutions during the lifetime of the grant; all the relevant departments within an institution should be included. The visits should involve both publicity through presentations of what the service is and what it can provide. The visits should also include training elements where new users are led through signing up to the service and introduced to the relevant databases.*
2. *The service should develop three publicity strategies; all three were considered realistic with the staffing levels of the service - General outreach to universities, Specific outreach to synthetic, bioscience and chemical engineering communities, Specific outreach to publicise DETHERM.*
3. *The panel did not consider that the CCPs or departmental representatives should be the primary routes for publicity. The road show idea developed by the service was also not considered appropriate to meet the publicity needs.*

We see the Panel's comments on the level, quality and effectiveness of the publicity and outreach achieved by the CDS as representing the most telling criticism of the Service at its current stage of development. We also acknowledge that there were similar criticisms made during the previous Panel visit in 2000. It is a little unfortunate that we were only presented with a copy of a report on that visit (in much abbreviated and highly edited form) just before we submitted the current proposal. We do, however, very much welcome the opportunity to study the report on the recent Panel visit.

The Panel expressed dissatisfaction with both the level of publicity achieved by the Service during the current grant cycle and our plans for future publicity included in the renewal bid. This is fair comment, and we are currently implementing plans to attain the suggested target of visits to 40-50 institutions during the lifetime of the grant. However, we do say in mitigation that it has made sense to allocate resources to developing high quality, enhanced extra web-based facilities and training/publicity material during the current grant cycle. Online material is, we believe, the most cost effective way of disseminating information and training material in many areas related to the Service. We are conscious that this has drawn much effort away from promoting site visits. However, our current portfolio of online material will considerably enhance the effectiveness of such visits in the future. We note the comments in *Bullet 3* and have been rethinking our strategy accordingly.

Feedback received during the recent CDS Users Forum held at DL has also brought to our attention the important position librarians hold in defining the scope of available chemical database resources seen by new graduate students at various university sites. Only a small number of science librarians have a specific chemistry background and many may not appreciate the full potential of the CDS in chemistry research, especially where SciFinder and Beilstein/CrossFire are also available. Don Parkin recently presented our introductory material to the graduate chemistry intake as part of the induction day given by the Library at Nottingham University. In addition new user were taken through the CDS registration process and over 40 new users signed up online. We see enhanced contacts with science librarians as an important part of our developing publicity strategy.

Librarian contacts potentially give access to new graduate students at the beginning of course, but clearly this may not always be practicable given manpower constraints. However, we will build on such contact. Contacts with librarians will also aid our planned penetration into the bioscience and engineering communities.

Another recent visit has been to UMIST (the Chemical Engineering and Chemistry Departments). Here Dave Fletcher gave a CDS presentation as part of the

departmental seminar series. It also provided an opportunity for fruitful discussions with staff members and graduate students. We have recently visited Liverpool (Chemistry and Biosciences) and Manchester (Chemistry), and will visit Loughborough (Chemistry and Library) and Dundee (Life Sciences) shortly. We are analysing feedback from these events, which is further helping us to refine our site visits strategy.

The conclusion that we have reached so far is that an effective publicity and education strategy involves two interacting strands. The first retains some aspects of our original roadshow ideas. Here typically a member of the CDS team presents poster and related publicity material at one or more departments on a given university site. He interacts with existing and prospective users and guides new ones through the registration process (they are where possible registered there and then).

The complementary strand consists of more formal presentations and training (in some cases for specific database components). This will happen either in parallel or subsequently. It will in many cases follow from an initial contact of type one. The recent visits to Liverpool and Manchester have been of type one, but the visit to Loughborough will involve both strands of the strategy. We have been invited back to give a further presentation at Liverpool Biosciences and are currently having follow up interactions with Manchester.

Our longer term strategy includes a programme of visits to the full range of university sites on a regional basis. We are currently working on the best way to timetable this, and will inform EPSRC as this timetable hardens up. The aim is to maximise effective coverage but minimise the overall travelling involved. Where possible a given outing will include several geographically related sites.

Publicity about the DETHERM and other major databases will be an integral part of presentations and other publicity material we produce. There will, of course, be a greater emphasis on DETHERM when we visit Chemical Engineering and other such departments. We will also send out a repeat mailshot to Department Heads prior to the formal relaunch of the DETHERM service.

We will arrange for new CDS summary material to be made available to various information portals such as the PSIgate and EEVL on the Resource Discovery Network and other JISC related facilities such their *Resource Guide for Physical Sciences*. In line with the recommendation in the report we expect the EPSRC will generate additional publicity in publications such as *Connect*. We are currently implementing our plans for transfer to a new server (10/11/03). As part of this process we are reviewing our website material which is being updated as appropriate.

The extra university visits which will take place during the lifetime of the new grant will demand further resources, but the precise amount is difficult to judge. We make an initial estimate of an extra 50 man visits over that included in the bid as initially presented (in some cases it will be appropriate to have more than one person involved for an individual site visit). Travel costs will vary and in some cases overnight accommodation will be required. We make an estimate of approximately £140 per man visit. We have, therefore, added £7000 to the travel budget, which is split equally over the three years.

We have also added and an extra £1500 under the *Exceptional Items* heading to cover increased expenditure on publicity related resources (e.g. maintaining updated poster and handout material) occasioned by the increased activity in this area. Funding to cover the cost of an extra laptop computer, which will be used during off-site visits, has also been added. Since the Service is to be reviewed after one year, it may be appropriate that the travel and related allocations also be reviewed at that stage.

## 5. Management Advisory Panel

1. *The panel considered that the MAP should provide more of an input to the service; this is particularly relevant in view of the appointment of a new service director. The MAP should formally meet twice a year to assist this process. There should also be a formal 4 year length of service on the MAP.*

We believe the MAP provides valuable input to the Service with much useful advice. It maintains scrutiny of the Service database and software portfolios and performance level measures (see Appendix G in the grant proposal). In fact they and the EPSRC are sent performance reports monthly and other material on an *ad hoc* basis.

The MAP has *Terms of Reference*, which have been agreed with the EPSRC (Appendix E). It has also developed additional working protocols. For instance there is an agreed convention on turn over of MAP membership; with a *de facto* period of tenure of 4-5 years. There are also agreed formal protocols to acquire new database systems (Appendix L). The issue of removing older systems is discussed at MAP meetings.

We are, of course, happy to discuss with the MAP & EPSRC any proposed modifications in *Terms of Reference* and work practices, which are deemed appropriate. These will include a formal protocol for removing redundant and little used systems.

We will organise a MAP panel meeting with representation from the EPSRC at a time shortly after details of our new grant are announced. We will present further information on our recent publicity visits and have the opportunity of discussing our developing strategies with the MAP. The CAS SciFinder issue is an area of great importance, and hopefully this meeting will also provide a forum for discussion with clarification of any direct involvement and commitment the EPSRC is likely to be able to make.

## Table: New Databases/Software Acquired by CDS

| | | £ k |
|---|---|---|
| **1994** | Protecting Groups | 35.3 |
| **1995** | ORAC Core | 11.8 |
| | ACD (2nd copy) | 3.7 |
| | Beilstein disks | 5.5 |
| **1996** | Chirbase | 27.9 |
| **1998** | Solid Phase Synthesis | 22.1 |
| **1999** | NIST 98 mass spectra | 2.0 |
| | Proton NMR spectra | 3.8 |
| | Netscape Enterprise Server | 1.2 |
| **2000** | Biocatalysis | 35.3 |
| | Oracle RDBMS | 8.8 |
| **2001** | Natural Products CNMR spectra | 1.4 |
| | Fluka HNMR spectra | 1.2 |
| | Detherm trial | 4.0 |
| | Chemscape | 12.4 |
| | LitLink | 18.6 |
| **2002** | Aldrich HNMR spectra | 3.7 |
| | SpecSurf | 9.3 |
| | IPlanet webserver | 2.5 |
| | Detherm extension | 15.1 |
| | National Cancer Institute database | 0.0 |
| **2003** | Screening compounds | 0.0 |
| | ACD 3D structure upgrade | 0.0 |
| | Organic Compounds HNMR Vol 2 and 3 | 6.2 |
| **Total** | | **231.8** |

| **Money for new data:** | |
|---|---|
| 1995 | 50.0 |
| 1997 | 40.4 |
| 2001 | 100.0 |
| **Total** | **190.4** |

# DISCUSSIONS RELATING TO AWARDING GRANT GR/S70623/01

## (ii) Outstanding Issues following Video Meeting held with the EPSRC on 16th January 2004

### 1. Introduction

The following notes address the seven outstanding issues raised in the letter of 5 December 2003, and are intended to act as a focus for discussion by video conference on the 16th January 2004 involving the chemistry programme manager. These notes are intended to help progress towards funding of the current service proposal. The seven issues raised are summarised in turn, followed by our current thoughts on how these might be best addressed.

### The Issues

**Service Awareness**

*A process to be put in place to ensure the service is aware of current developments within the chemical community and to actively find out how databases are used in modern chemical research both in universities and in industry.*

This process comprises a number of components that are designed to establish both a greater interaction with key figures in the field, and to ensure a greater degree of involvement with, and feedback from, the community at large. The various anticipated components are itemised below:

1. Review membership of CDS MAP, preferably in conjunction with EPSRC

2. Receive views and feedback from individuals encountered during our programme of university publicity/training visits.

3. Establish and/or maintain various experts contact e.g., Peter Willett (Sheffield, Information Studies), Robert Glen, Peter Murray-Rust (Cambridge), Henry Rzepa (Imperial), Mike Hursthouse (Southampton), Figures with industry connections...., Wendy Warr (Chemical Information Consultant), Steve Davies (Oxford), Mark Bradley (Southampton), Gordon Tiddy (UMIST, Chemical Engineering), Frank Allen (CCDC), Andy Vinter (Cresset Biomolecular Discovery), Dominic Tilsdley (Unilever).

We aim to make personal visits to these figures, with the director of the service attending whenever possible.

4. Conduct a major survey of all CDS users. This will be an online survey. It is proposed that this takes place around May 2004.

5. Review product line and where appropriate arrange meetings with major chemical database supplier, e.g.: MDL, Accelry, CAS, Tripos, CCDC, DECHEMA, FIZ, NIST, ICDD Daylight, ISI, etc

6. Keep watching brief on Cheminfomatics and Library discussion lists (cheminf, lis-scitech, chemweb) and various chemistry sites (e.g. orgnet, inorganic-chemistry, molecular-dynamics).

7. Establish a CDS User Group. This may be partly modelled on the CCP concept if appropriate. A major aim is to boost the attendance at the CDS User Forum. Ideally the User Group would organise forum events and/or have more input into determining format and agenda of meetings.

8. Make a concerted effort to revive the CDS-UK Jiscmail discussion list.

9. Have group members attend appropriate conferences and workshops e.g. BCA, RSC, etc. events such as the recent BBSRC/RSC Chemical Genetics Workshop at Hinxton and the forthcoming Joint Sheffield Conference on Chemoinformatics.

**Implementation of Publicity Elements**

*The service to provide increased detail on the implementation of the publicity elements.The additional details should include the format of the visits, how the visits will be divided up amongst service staff and how institutions will be prioritised. A draft timetable for the first year of institution visits should also be included.*

We propose organising visits of two main types, broadly classified as *Publicity and Awareness* (P&A) and *Specific Training* (ST) visits:

**1. Publicity and Awareness (P&A) Visits**

These will consist of two parts:

(a) Poster Demonstration in foyer of Chemistry Department and/or Library - typically during the morning

- The Service has purchased a portable poster display stand and has generated appropriate publicity/information material. This material will be updated and modified to take account of Service developments and in the light of feedback as the visits programme proceeds.
- We will ensure that tables are available near the poster display. This will be manned by two CDS members (DOE + a longer established member). We will ensure that network connectivity is available for (preferably) two laptops computers. This will allow CDS personnel to demonstrate aspects of the Service arising from discussions with current and potential users who visit the stand. In addition we will endeavour to allow all users have an interest to register via the online mechanism there and then.

(b) Introductory Lecture - typically during the afternoon

- This will take place in a lecture theatre/seminar room. It will make use of PC-based presentation material and will require the availability of a PC projector and network connectivity.
- The lecture starts with a run through all of the main CDS facilities. This will typically be given by DOE. A standard presentation has been prepared but may be updated and modified to take account of Service developments and in the light of feedback as the visits programme proceeds. This presentation can if needs be used in standalone mode, but includes interactive access to the CDS server. We can provide copies on CD or other media, which will be left at the site. It is also downloadable from the CDS website.
- This demonstration makes a point of demonstrating aspects of various "flash" training movies which been produced. The aim is not to go through these movies in their entirety, but rather to point out their availability and encourage users to explore them later.
- There will then be a questions session. This is likely to require the greater involvement of the other CDS member. Depending on available time there will then be further more free ranging demonstrations of aspects of the system. The focus of these will depend on (a) topics brought up during previous discussion (b) known topics of interest highlighted when planning the visit.

**2. Specific Training (ST) Courses**

These can be tailored to specific departmental requirements, but are likely to cover the main subject areas of the Service. They will be presented a local computer training facility, and will involve a large component of hands on use. It will also be important to ensure a PC projector is available of effective demonstration of material to the entire class

(a) **Structural Chemistry**: This is likely to cover the main crystallography databases. In the past request have been mainly for coverage of the Cambridge system, but we aim to increase awareness of ICSD-WWW and CrystalWeb. The lead trainer is likely to be RFM.
(b) **Organic Reactions and ACD**: This is likely to cover the ISIS system with increasing emphasis on the web-based interfaces. Emphasis will be placed on specific advantages of the ISIS system. In particular SPS, SPG, Biocatalysis, Chirbase. It will point out the wide applicability of ACD (beyond the use by the synthetic chemists). The availability of the Screening Compounds database will be emphasised. We see particular value of this in the field of chemical biology and diversity-oriented synthesis. The lead trainer is likely to be DP.
(c) **Spectroscopy**: This will cover the functionality available via the SpecInfo system. It will emphasise the multi-technique coverage available (nmr, IR & MS), and demonstrate how to use the system for both peak searching and spectral prediction. We expect to restrict training to use of the newer web interface, which has effectively superseded the older X-Windows system. The lead trainer is likely to be DAF.
(d) **Thermophysical data**: This will cover the scope and use of the DETHERM database system. It will survey the full scope of the database components and give training in effective use of the Client software. It may also include a review of other free access web-based systems such as the various NIST web components. Lead trainers are likely to be RFM, DAF & DP.

We expect to achieve 15 P&A visits during the first year 2004/5. We aim to cover all 5*, 5 & 4 grade chemistry departments during the 3 years period of the grant. We will concentrate effort on the 5/5* for the Chemical Engineering and Biosciences departments.

We do not feel able to commit to specific site visit itineraries at this stage (see below) but, as already indicated, hope to be able make as many as possible regionally based to minimise travelling involved.

Our experience to date indicates that it is most important to establish good contacts before to ensure (a) room/space availability (b) appropriate network connects and local technical support and notification to allow access by our travelling laptop systems (c) locally publicity including tailored flyers produced at DL.

We do not think it appropriate to attempt to mix P&A and ST events. The latter will require a training suite. This will not be able to comfortably accommodate as many attendees as a lecture/seminar room. Places will generally be booked in advance

for an ST event. It is expected that interest ST events will be triggered by the programme of P&A events.

The launching of these events will proceed along the following lines:

- We will generate a draft timetable of institution visits for the 1st year after the discussions on the 16th. A specific issue for which we appreciate a steer is whether to concentrate effort on (a) 5/5* departments, (b) biosciences and chemical engineering departments.

- We will write to all Chemical Engineering department heads (as well as department representatives where available) when we are in a position to formally announce the DETHERM service. This will include material for display.

- Similarly we will write to Bioscience department heads in advance of our planned trial for several new Accelrys database. This will place emphasis of the Bioster, Metabolism database and relevant established components such as Biocatalysis, Chirbase, Protecting Groups and Solid Phase Synthesis.

**Service Development Strategy**

*The service to outline its development strategy for the short to medium term. At present the service strategy appears to be solely based on the implementation of DETHERM. Additional details are specifically required on how the service will ascertain the needs of the bioscience and engineering communities. The strategy should go beyond the scope of the publicity elements already incorporated.*
The Service's stated development aims include the implementation and promotion of the DETHERM system as well other aspects which may not have been quite so clearly stated. These include aspects which are highly important but are easy to take for granted such as maintaining and enhancing the most viable aspects of the current system. Specifically these include transfer to more modern, efficient and cost effective hardware. We also propose creating more seamless links for downloading database information to (in particular) quantum chemistry and other codes using modern data structures with a view to taking full advantage of emerging semantic grid technologies.

The current database portfolio represents a secondary information system. Recent technology developments have included links to primary information sources. Within CDS we have implemented the LitLink system which provides links to the primary electronic journal and patent literature. We have used this technology to extend these linkages specifically to our crystallographic

databases. We propose to extend these ideas in the light of available technologies. Tertiary information sources are increasingly becoming available in electronic format. Cross-linking to major tertiary source such as Heuben Weyl and its planned successors will become increasingly important.

At the time of framing our original bid the technology offered by the MDL DiscoveryGate system seemed to offer the most promising way forward. DiscoveryGate offers access to many layers of chemical information including the major electronic reference works, ISIS database, the Beilstein/Gmelin databases as well as (potentially) the CAS database system. We were advised that DiscoveryGate technology would not be available for "in-house" use during the lifetime of the proposed grant. Also it only make strategic and economic sense if a DiscoveryGate systems included Beilstein, etc. This would involve potential strategic alliances with MIMAS/JISC, etc. In our opinion, it made little sense to raise these issues at the time of the application.

Since that time the possibility has arisen that Elsevier will not release DiscoveryGate technology for in-house use. This may make necessary a radical rethink of aspects of our database provision. Moving away from MDL to another software provider (such as Accelrys) may then make sense.

We are currently arranging the implementation of a whole series of bioscience and related database systems as an integrated package from Accelrys. Our record indicates that we are likely to be able to acquire and trial other new databases if and when they become available during the lifetime of the forthcoming grant (c.f. Beilstein, Solid Phase Synthesis, Chirbase, etc.)

**SciFinder Developments**

*There is a requirement for a strategy for co-ordinating the development of a national agreement for SciFinder in the medium term. While additional stakeholder consultation would be required, it is considered that the service is in a unique position to drive this agenda forward.*
For discussion during video meeting

**Database Portfolio**

*A procedure and criteria to be detailed for the acquisition of new databases or the removal of low use ones which no longer justify their upkeep.*
A formal procedure for acquiring new database already exists. It was included as Appendix L in the original funding bid and is listed below.

We maintain usage logs of all databases and utility programs provided by the Service. We are thus able to identify those which appear to have been superseded or of little value. We do not belief that a simple metric based solely on usage should be used. The cost, support requirements and (in some cases) potential use should also be taken into account.

The Service will compile a list of candidates for removal from the Service. This will be put before the Management Advisory Panel and its next meeting and annually thereafter. Suggestions for revision of the appraisal criteria will also be elicited from the MAP.

*Appendix L: Protocol for Appraising and Acquiring New Databases and Systems*

1. Confirm that the prospective new database contains high quality data that would be of value to a reasonably large number of CDS users at reasonable cost.

2. Inform the CDS MAP of the potentially valuable database.

3. Assuming that the necessary systems are available, attempt to obtain a free trial of the database for a period of several months.

4. Inform the user community of the trial, inviting them to try out the database.

5. Record usage of the database during the trial (as a measure of the community's interest).

6. Towards the end of the trial, evaluate the user community's perception of the database via a survey. Encompass both the number of users who want the database to be available as well as its value to their research.

7. Submit a proposal to obtain the database to the CDS MAP, including details of funding.

8. If all the above conditions have been met, obtain the database and release it to the community.

**Staff Competencies**

*The competencies of the staff requirements of the national service are to be addressed during the lifetime of this award. The assessment of staff competencies required for this national service will be carried out in collaboration with the service. The service should commit to participate in such a review and to implement any recommendations*

*including redeployment, redevelopment and re-training as appropriate.*

The Service is keen to participate in any review procedure put in place by the EPSRC. The competencies and indeed level of staffing is likely to depend on explicit plans for Service development. What developments are possible or desirable are likely to depend on external circumstance such as technological innovations and policy decisions made by database and software suppliers. It is anticipated that the director of the service will further address this issue during the proposed visit to Polaris House.

**MAP TORs**

*The terms of reference of the Management Advisory Panel to be redrafted to include the amendments detailed in Annex 1*

We are happy to redraft the Management Advisory Panel terms of reference according to the amendments detailed in Annex 1. Should we seek agreement from the chair, whole panel or do this unilaterally?

# DISCUSSIONS RELATING TO AWARDING GRANT GR/S70623/01

## (iii) Items requested following 16/1/2004 meeting via video link

## Introduction

Following the meeting on 16 January, the Chemical Database Service agreed to provide additional material relating to ***1.*** *Termination of Database/Software Support;* ***2.*** *Amplification of selection Criteria for prospective new Systems;* ***3.*** *Modified Terms of Reference for Management Advisory Panel;* ***4.*** *Additional Material related to Staff Competencies, and* ***5.*** *A Timetable of Site Publicity & Awareness Visits during the first Year of the proposed Grant.*

After feedback from Drs. Carmine Ruggiero (January 29) and Alison Wall (February 13), we have provided additional clarification in a number of the sections below, specifically sections 2.1-2.3 and 2.5. We trust that this additional material will address the issues raised by EPSRC, hence enabling the announcement of the CDS grant in the very near future.

## REQUESTED MATERIAL

**Termination of Database/Software Support**

*The EPSRC has requested details of the formal procedure the CDS personnel will implement for deciding which current databases and software utilities should cease to be supported and removed from the Service.*

*Protocol for Termination of Database/Software Support*

*1.* The full database/utilities portfolio will be formally reviewed periodically by the Service:

*a)* Every 12 months

*b)* When there is a major system change (e.g. transfer to new server)

*c)* Individual items may be reviewed on an *ad hoc* basis where appropriate.

*2.* Candidates for removal will fall into the categories

*a)* Limited Usage

- Less than on average 5 accesses per month or low usage out of balance with the ongoing support costs

*b)* System Superseded

- Better system available on the CDS (e.g. various file conversion utilities now superseded by BEDLAM; FNMR data accessible via SpecInfo)

*c)* Readily available Elsewhere

- Other Services/Sites provide ready access (e.g. PDB protein datafiles on EBI, and other web sites)
- Software can be downloaded and run more effectively on users' local workstation (e.g. molecular display packages such as RasMol, the crystallographic package, PLATON)

*d)* Not Maintainable

- System requires hardware and/or software which is now redundant - would require excessive effort for continued support (e.g. in the future this will increasingly apply to the various legacy codes – indeed certain functionality already dropped for CSSR)

*3.* Where a specific item meets one or more of the removal criteria it will be examined in more detail. Any contentious cases will be sent to the MAP for discussion.

*4.* Where it has been determined that an item will be removed

*a)* Specific active users of the system will be contacted where relevant

*b)* References to item will be removed from relevant documentation Web site, Promotional Material, Reference Sheets, Online Help, etc.

*b)* Notice will be given to the general user community by E-mail and/or the CDS Newsletter where relevant

*5.* Access to database/utility by users will be barred after completion of the above procedures

*6.* Total Removal from the system will occur

*a)* After 6 months

*b)* If there would be a major support overhead in restoring system to usability

**Amplification of selection Criteria for prospective new Systems**

*The EPSRC requested that the Protocol for selecting prospective new systems be amplified.*

*Protocol for Appraising and Acquiring New Databases and Systems*

An extended Protocol is listed below. Additional clarification is given with the addition of sub-sections ***1a*** to ***1c.***

*1.* Confirm that the prospective new database contains high quality data that would be of value to a reasonably large number of CDS users at reasonable cost. Criteria for deciding this will include:

*a)* Prospective systems regarded as of high potential value if they occur frequently amongst suggestions for new data in responses to CDS surveys of user requirements.

*b)* The quality of new data is likely to be guaranteed if it is supplied by a known and reliable source. Such sources currently include: MDL, Accelrys, CCDC, NIST/FIZ, Tripos, Daylight, etc.

*c)* Where a prospective source is new or not well known the Service will solicit advice from the MAP, other experts, discussion lists, etc. In some cases a specific user survey will be conducted under Item *2*. Particular note will be made of the tests for quality and usefulness under Items *5* & *6*.

*2.* Inform the CDS MAP of the potentially valuable database. Add "new databases" as a standard item on agenda for MAP meeting.

*3.* Assuming that the necessary systems are available, attempt to obtain a free trial of the database for a period of several months.

*4.* Inform the user community of the trial, inviting them to try out the database.

*5*. Record usage of the database during the trial (as a measure of the community's interest).

*6.* Towards the end of the trial, evaluate the user community's perception of the database via a survey. Encompass both the number of users who want the database to be available as well as its value to their research.

*7.* Submit a proposal to obtain the database to the CDS MAP, including details of funding.

*8.* If all the above conditions have been met, obtain the database and release it to the community

**Modified Terms of Reference for Management Advisory Panel**

*The modified terms of reference of the Management Advisory Panel as included in the amendments detailed in Annex 1 of the letter from the EPSRC dated 6/12/03 are acceptable to the Service.*

*Management Advisory Panel Terms of Reference*

In line with the discussions on 16/1/04 an extra item relating to representation by the UK academic librarian community has been added. This is present as *Section 5d.*

1. The Management Advisory Panel (MAP) exists to assist the Service Director in the effective operation of the National Service by:

*a)* Ensuring Service is fully utilised in supporting the highest quality science.
*b)* Advising on the special and changing research needs of the communities using the Service and how the Service might be developed to meet these needs.
*c)* Advising on how the Service is perceived both scientifically and organisationally by its user communities so that timely action may be taken to build on strengths and address weaknesses.
*d)* Assisting in the promotion of the Service to ensure that as many as possible of the researchers who might benefit from it are aware of its existence and technical capabilities.

2. The full MAP will normally meet twice a year, and members may be asked to participate in additional meetings involving users or EPSRC as necessary.
3. Members of the MAP should declare any personal interests and not participate in discussions where there would be a conflict of interest.
4. Membership of the Management Advisory Panel should ensure that the MAP has representatives from each main user community and should be reviewed on an annual basis to reflect changes in the user base.
5. The composition of the MAP will be based on the following criteria:

*a)* MAP membership should be for a fixed three year term.
*b)* New MAP members to be chosen in consultation with EPSRC.
*c)* MAP to have at least one member from industry.
*d)* MAP to have a member from the UK academic library community
*e)* MAP to include no more than one member from any single institution.

**Additional Material related to Staff Competencies**

*The EPSRC asked for a summary of functions provided by the Service personnel and the associated skills involved:*

## *Competencies and Time required for Support of the Chemical Database Service*

A mapping of time involved for the individual staff members with respect to core CDS functions is given in *Appendix A*. Some clarification of the skills involved for these functions is given below.

1. **User support**: This includes answering queries, registering users and the preparation of relevant sections of the website. It requires knowledge of the various service components, relevant chemistry disciplines, client computer systems and web authoring skills.
2. **Training**: This includes delivering courses and preparing online material. It requires detailed knowledge of the appropriate service component and chemistry discipline and web authoring skills.
3. **Publicity**: This includes site visits, mailshots, newsletters, the user meeting and web-based material. It requires knowledge of the various service components and relevant chemistry disciplines as well as web authoring, communication and presentation skills.
4. **Database administration and maintenance**: This includes database updates, website maintenance, the installation of fixes and of new releases. It requires detailed knowledge of the database systems, scripting, programming and Oracle DBA skills.
5. **System administration and development**: This includes keeping the systems secure, running and healthy as well as improving the system to user interface. It requires system administration, scripting and programming skills as well as technical knowledge of the various CDS packages and utilities and the client systems which interact with them.
6. **Database development**: This includes the installation of new databases, the building and deployment of new interfaces, the installation and configuration of new hardware as well as the improvement of the CrystalWeb interface and the ongoing migration away from legacy code. It requires detailed technical knowledge of the relevant packages and legacy code, scripting, programming and system administration skills. It also requires knowledge of the current state of chemical information systems and user requirements.
7. **Strategic development**: This includes maintaining a watching brief on developments, attendance at conferences, collecting user requirements and planning. It requires knowledge of the current state and of user requirements, contacts with vendors and others in the chemical information field, communication and negotiation skills.

8. **Service management**: This includes regular reports, MAP meetings, financial management and other administrative tasks.

**A Timetable of Site Publicity & Awareness Visits during the first Year of the Grant.**

*The EPSRC asked the Service to produce a timetable for university site visits for the first year of the grant. A total of between 20 and 30 visits should be scheduled.*

## *Site Visits Strategy*

We have produced a draft timetable, which includes 26 Publicity and Awareness (P&A) visits during this period and is included as *Appendix B*. Note that discussions with local representatives indicate that arrangements for such visits are by and large only practicable within a window from October to March. There is a distinct preference for around October corresponding to the arrival of the new graduate student intake. The time after March presents difficulties because of Easter, examination commitments, conferences and the summer vacations.

In *Appendix B* we include a single named department for each visit. This is where we expect the P&A event to take place, but will make efforts for it to be also publicised within relevant nearby departments. We have not included any Specific Training (ST) Courses in *Appendix B*. Experience indicates that constraints for scheduling these will be less demanding.

Where possible we intent to further refine our timetable with a view to *a)* bringing forward visits to within the spring/summer "dead period", *b*) rescheduling the order of some visits so that regional visits are further clustered together, thus minimising the travel involved. We do not include any visits before 1 April 2004 but understand that the impact of these will be taken into consideration by the *2005 Review.* A high priority in trying to add extra visits during the "dead period" is to target Chemical Engineering departments. We will also try to get a news/publicity article about the DETHERM release into relevant **I**Chem**E** or other journals. We believe such prior publicity will help enhance the impact of the subsequent site visits.

Prior to visits the members of the Service will examine the major research interests of staff at the sites involved. The main source of information is likely to be departmental websites. We expect to change the emphasis of the material presented in the light of department requirements. These will also be discussed with the CDS departmental representatives and a *pro forma* will be devised to help capture this information. In addition we will identify some key individuals within the departments (e.g. chair of Organic Chemistry, librarians), and will write to these people (a couple of weeks beforehand) informing them of the site visits and asking them to encourage their students and colleagues to attend. This will be in addition to publicity material (flyers, etc.) distributed through the departmental representatives.

The Service will gauge the success of the site visits. This should be reflected in increased usage of key components. Immediate measures of success during the site visits themselves will be:

a) the number of people attending and having discussions at the poster based presentations;

b) the number of new users who registered online for the Service;
c) the number of people who attended the presentation/lecture segment; and
d) whether there were requests for follow up training courses.

In addition to recording the number of attendees at the events we will estimate the distribution across subject areas. In particular the online registration procedure now includes a section specifically recording this information. The nature of questions asked plus the degree of interest and level of requests for follow up courses will also give us a measure of impact by subject area. Particular emphasis will be placed on gauging demand for follow up courses, establishing local contacts to help us organise these, and recording this information to help us plan our follow up course programme.

## *APPENDIX A*

| Function | **"Ideal" case estimate (FTE)** | RFM | DP | DAF | DO-E | **Total (FTE)** |
|---|---|---|---|---|---|---|
| User support | **0.75** | 5% | 25% | 10% | 25% | **0.65** |
| Training | **0.5** | 5% | 20% | 5% | 10% | **0.4** |
| Publicity | **1** | 5% | 30% | 10% | 45% | **0.9** |
| Database administration and maintenance | **0.8** | 20% | 10% | 20% | 10% | **0.6** |
| System administration and development | **0.4** | 5% | - | 20% | 5% | **0.3** |
| Database development | **1** | 40% | 5% | 30% | 5% | **0.8** |
| Strategic development | **0.25** | 10% | 5% | 5% | - | **0.2** |
| Service management | **0.15** | 10% | 5% | - | - | **0.15** |

The "Ideal" case corresponds to the resources we would hope to be able commit to the various individual tasks in the absence of the inevitable constraints related to available manpower

## *APPENDIX B*

| University | Department | RAE Grade | Provisional Date |
|---|---|---|---|
| Imperial | Chemical Engineering | 5* | 21/04/2004 |
| Sussex | Chemistry | 5 | 20/04/2004 |
| Warwick | Chemistry | 5 | 05/10/2004 |
| UCL | Chemistry | 5* | 06/10/2004 |
| Swansea | Chemistry | 4 | 12/10/2004 |
| Cardiff | Chemistry | 4 | 13/10/2004 |
| Nottingham | Chemistry | 5 | 14/10/2004 |
| York | Biological Sciences | 5 | 19/10/2004 |
| Leeds | Chemistry | 5 | 21/10/2004 |
| Sheffield | Chemistry | 5 | 26/10/2004 |
| Newcastle | Chemical Engineering | 5 | 28/10/2004 |
| Bath | Chemical Engineering | 4 | 30/10/2004 |
| Durham | Chemistry | 5* | 02/11/2004 |
| Belfast Queen's | Chemical Engineering | 4 | 04/11/2004 |
| St. Andrews | Chemistry | 5 | 09/11/2004 |
| Glasgow | Biological Sciences | 5 | 11/112004 |
| Edinburgh | Chemistry | 5 | 16/11/2004 |
| Strathclyde | Chemistry | 4 | 17/11/2004 |
| Dundee | Biological Sciences | 5* | 18/11/2004 |
| Liverpool | Chemistry | 5 | 25/11/2004 |
| Manchester | Chemistry | 5 | 30/11/2004 |
| Southampton | Biological Sciences | 5 | 25/01/2005 |
| Bristol | Chemistry | 5 | 27/01/2005 |
| Cranfield | Biological Sciences | 4 | 14/02/2005 |
| Loughborough | Chemical Engineering | 4 | 17/02/2005 |
| Leicester | Chemistry | 4 | 24/02/2005 |
| Birmingham | Chemistry | 5 | 01/03/2005 |