International Union of Pure and Applied Chemistry
A member of the International Council of Scientific Unions

CPEP - Committee on Printed and Electronic Publications

IUPAC REPORT
to CODATA General Assembly Berlin 2004

Compiled by Tony Davies
Secretary,
IUPAC Standing Committee on Printed and Electronic Publications

CONTENTS

OPEN ACCESS POLICY TO REPORTS AND RECOMMENDATIONS

IUPAC has developed over the years a series of policies, not only with regard to "Data" initiatives but also with regard to overall policies of access to the activities of the Union, which are relevant to the ongoing discussions about the future direction, and activities of ICSU/CODATA.

Since opting for the self-publishing route and dropping the concept of the "official" publisher IUPAC has gained and exploited the freedom to make the official recommendations and technical reports as published in Pure and Applied Chemistry available for free over the IUPAC.ORG homepage. For an example of this policy in action see http://www.iupac.org/publications/pac/2004/7610/7610x1921.html .

We have followed and welcomed the ICSU initiatives around the World Summit on the Information Society (WSIS) but currently feel that the steps and positions that are currently "hot topics" have already been raised and resolved for IUPAC in the last 10 years. A good example of this has been the move to self-publishing. Interestingly enough although the "moral" aspects of such a move such as the freedom to provide far more open access to the results of Union were important, this has in fact also resulted in a financial model vastly more beneficial to the overall IUPAC budget than previously was the case.

This report deals mainly with "Data" aspects of IUPAC's activities, however a major part of the work carried out by IUPAC bodies is in the field of nomenclature, which has clear implications for computerised, automated search capabilities of information resources where a standardised use of language is an essential pre-requisite in any effective knowledge management or informatics system.

Furthermore, task groups working on the development of open, long-term stable data standards have always been required to work openly and with instrument vendors and user groups. We maintain an open source depository on SourceForge for computer code for handling these data formats. This tough thinking has continued into the new work in progress with ASTM International on the Analytical Mark-up Language AnIML where, to our knowledge for the first time in the history of ASTM International, the entire AnIML project documentation, discussions, presentations, test files and source code is available for everyone to review and challenge, also as a SourceForge project. (See http://animl.sourceforge.net/)

One further concrete step was contained in the joint IUPAC, IUBMB, IUCr, and ICSTI proposal on Networking Public Scientific Reference Databases which was rejected by ICSU for the 2005 funding round (see attachment). The project was rejected with 2A's and 2B's. The negative comments were *"This is a useful but routine project with little innovation"* and *"Although importance of ICSU support is underscored, it is not clear that it for more than a high level technical activity. The relationship to ICSU priorities is not clear."* I hope that the CODATA GA will present an opportunity to discuss this rejection alongside the other similar projects being proposed for CODATA support at this GA.

STANDARDIZATION OF SCIENTIFIC DATA EXCHANGE FORMATS

IUPAC has a long history in the development and support of scientific data standards. In 1995 we took over responsibility for the JCAMP-DX range of scientific standards from the Joint Committee on Atomic and Molecular Physical Data (JCAMP).

Since then a Working Party has had responsibility for the support and development of the JCAMP-DX family of scientific data standards. Key to the success of these projects has been the prerequisite involvement of the manufacturers in the respective fields ensuring relevance and implementation of the work. In 2003 due to the increasing interest in the use of XML for data exchange the Working Party evolved into the Subcommittee for Electronic Data Standards with oversight duties within IUPAC for all activities in either the JCAMP-DX sphere or the XML in Chemistry area. A number of projects around data standards are currently underway.

XML-based IUPAC Standard for Experimental and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML)

Create an XML-based dictionary for storage and exchange of thermophysical and thermochemical data based on fundamental principles of phenomenological thermodynamics covering a wide variety of systems such as pure chemical compounds, multi-component mixtures, and chemical reactions.
http://www.iupac.org/projects/2002/2002-055-3-024.html and for more up-to-date information http://www.iupac.org/namespaces/ThermoML/index.html

Data exchange standard for near infrared spectra and general spectroscopic calibration data types

To (a) develop, document and validate a new JCAMP-DX spectroscopic data exchange standard for Near Infrared Spectroscopic data sets, and (b) make recommendations for the storage and archival of spectroscopic calibration files. This task group will standardize the data dictionary for Near Infrared Spectroscopy as well as defining content for reporting chemometric models. This is a prerequisite for implementation of XML.
http://www.iupac.org/projects/2002/2002-020-2-024.html

Data exchange standard for electron paramagnetic resonance data types (incl. ESR EMR etc.)

To develop, document and validate a new JCAMP-DX spectroscopic data exchange standard for Electron Paramagnetic Resonance (EPR), Electron Spin Resonance (ESR) and Electron Magnetic Resonance (EMR) data sets. This will standardize the data dictionary for this type of spectroscopy - a necessary prerequisite for any chemical mark-up language development work in this field.
http://www.iupac.org/projects/1999/1999-046-2-024.html

THE IUPAC XML IN CHEMISTRY INITIATIVE

Many scientific organisations and companies delivering scientific products have implemented or are looking at the use of the so-called eXtensible Mark-up Language or XML as a powerful alternative to conventional binary file storage and information exchange. XML can be regarded as an extension to the well-known HTML or Hyper Text Mark-up Language, which is the language, which is most frequently encountered when viewing web pages on the Internet.

As with a conventional web page it isn't the use of XML itself that is interesting or even particularly novel but the content stored within the XML files. In Chemistry and associated technical fields various groups, whether for commercial organisations, academic institutions or government bodies have been developing XML formats with similar content but differing data dictionaries and conventions so that they are not compatible with each other and, what is far worse, resources world-wide are being deployed to address problems already solved by other groups.

In order to support standardisation in the field of chemistry IUPAC decided to take an active role in helping to unify the various dictionaries and publicise their availability.

This started in the Brisbane GA in 2001 where the IUPAC XML in Chemistry - Do's and Don'ts were laid down as follows:

IUPAC should not:

- Commence activities better left to the computer scientists
- Re-invent the wheel - the current activities at various locations should be invited to contribute to a standardization process through IUPAC as long as their efforts remain in the public domain
- Become formal members of W3C, OMG or other similar organizations, however they should be informed of IUPAC activities in this area and we should continue to monitor their work.

IUPAC should:

- Establish 'ownership' of the definition of standard terms in chemistry to be used in digital communications through formal IUPAC recommendations.
- Generate a glossary of standard terms in chemistry for use in applications involved in digital communications such as scientific data exchange or electronic publishing.
- Locate potential interested parties within IUPAC who 'own' glossaries of terms or who are in the process of creating them (through IDCNS)
- Establish a method to identify and resolve problems in overlap of definitions (within IUPAC as well as with other scientific standards and other organizations)

This action was quickly followed up by the XML in Chemistry conference held in Cambridge UK – 24-25 January 2002. Delegates from outside IUPAC who are active in setting the direction in the handling of chemical objects within their organisations also attended from International Union of Crystallography, Chemical Abstracts

Service, US National Institute of Standards, European Bioinformatics Institute, and the International Union of Biochemistry and Molecular Biology.

Some of the first concrete actions arising from this initiative has been the establishment of the Namespace for Chemistry and the following projects:

Standard XML data dictionaries for chemistry

Translate existing IUPAC standard terminologies and related information to data dictionaries in eXtensible Markup Language (XML) format.
Establish a strategy for future IUPAC involvement in applications of XML for chemistry. Enable IUPAC to serve as the principal, authoritative source of basic chemical terminology for electronic communications.
http://www.iupac.org/projects/2002/2002-022-1-024.html

IUPAC-NIST Chemical Identifier

The objective of the IUPAC Chemical Identifier Project is to establish a unique text-based, machine-readable label, the IUPAC Chemical Identifier, which would be a non-proprietary identifier for chemical substances that could be used in printed and electronic data sources thus enabling easier linking of diverse data compilations.
http://www.iupac.org/projects/2000/2000-025-1-800.html

A new test version of the IUPAC-NIST Chemical Identifier (INChI) is now available. It replaces the previous test version issued last November. All features planned for inclusion in the final release have now been implemented and the final format for the Identifier has been proposed. The new name of the Identifier (formerly IUPAC Chemical Identifier, IChI) acknowledges the development work at NIST. The test program accepts input in the form of MOLFiles (or SDfiles) and CML files. An Application Program Interface (API) for communicating with external programs is under development.

A single INChI is generated for a single input structure, which can contain multiple components. Identifiers can be created for organic compounds with Z/E and sp3 stereochemistry, tautomers, and isotopes as well as salts, organometallic compounds and protonated forms of a compound.

Test programs (for Microsoft Windows), documentation and sample structure files are available upon request from Steve Stein <steve.stein@nist.gov>. The project team very much welcomes comments concerning the INChI and will be glad to assist in its testing or implementation. An implementation by a commercial company, Advanced Chemistry Development, Inc., is already available see:
http://www.acdlabs.com/download/technotes/80/draw_db/inchi.pdf

A SELECTION OF OTHER ONGOING IUPAC PROJECTS

This is a selection by the author of projects that may be of relevance to CODATA. The complete list of current and completed projects can be found at: http://www.iupac.org/projects/index.html

Physical and Biophysical Chemistry Division

Evaluated kinetic data for atmospheric chemistry
Enhance the accessibility and availability of the evaluated kinetic data base, which has been assembled by the IUPAC Subcommittee for Gas Kinetic Data evaluation for Atmospheric Chemistry,
http://www.iupac.org/projects/1999/1999-037-2-100.html and
http://www.iupac-kinetic.ch.cam.ac.uk/

Ionic liquids database
Create an open-access, free, on-line, comprehensive database for storage and retrieval of metadata and numerical data for ionic liquids, including their syntheses, structure, properties, and uses; lack of this information is impeding progress in a burgeoning field of significant current interest.
http://www.iupac.org/projects/2003/2003-020-2-100.html

Selected free radicals and critical intermediates: thermodynamic properties from theory and experiment
Compilation and critical evaluation of published thermodynamic properties, including the computation of accurate thermochemical data for selected free radicals, that are of importance in atmospheric and combustion chemistry.
http://www.iupac.org/projects/2003/2003-024-1-100.html

Heat capacity of liquids: critical review and recommended values for liquids with data published between 2000 and 2004
Update and extend two publications that contained recommended data on liquid heat capacities for almost 2000 mostly organic compounds, "Heat Capacity of Liquids: Critical Review and Recommended Values", and its "Supplement I" by M. Zábranský, V. Ruzicka, V. Majer (1st work only), and E.S. Domalski published in *Journal of Physical and Chemical Reference Data* in 1996 and 2001.
http://www.iupac.org/projects/2004/2004-010-3-100.html

Critical Compilation of Vapour Liquid Critical Properties
Values of vapour-liquid critical properties are essential in corresponding-states treatments of equilibrium thermodynamic and transport property data. The objective is to review all such measurements for pure organic compounds containing **nitrogen**, **halogen**(s), and **sulfur** and **silicon** and to recommend values for critical temperature, pressure and densities, with uncertainties.
http://www.iupac.org/projects/2000/2000-026-1-100.html

Inorganic Chemistry Division

Review of isotopic abundances in extraterrestrial materials
Extraterrestrial materials (meteorites, lunar and martian samples, cosmic dust) have significantly different isotopic abundances compared to terrestrial samples. The Subcommittee of Non-Terrestrial Isotopic Abundances has collected over 1,300 references on isotopic abundance variation in a variety of extraterrestrial materials. This project will focus on reviewing the differences for Rare Earth Elements and Noble Gases.
http://www.iupac.org/projects/2001/2001-042-1-200.html

A new comprehensive report on the isotopic compositions of the elements for global user communities
To design, compile and produce a new comprehensive Report on the Isotopic Compositions of the Elements, containing updated data and in a format suitable for the 21st Century.
http://www.iupac.org/projects/2002/2002-049-2-200.html

Isotopic compositions of selected elements
To design, compile and produce a new Report on the Isotopic Compositions of selected Elements, containing updated data with uncertainties that conform to ISO 9001 requirements.
http://www.iupac.org/projects/2003/2003-031-2-200.html

Determination of atomic weights using new analytical techniques
New analytical techniques, including inductively coupled plasma-mass spectrometry, have recently provided atomic weights with unparalleled precision. The purpose of this project is to assess the uncertainties in these new methods in atomic-weights work, evaluate published work, and, if possible, make recommendations to update IUPAC's Table of Standard Atomic Weights.
http://www.iupac.org/projects/2003/2003-033-1-200.html

Macromolecular Division

Recommendations for data presentation and storage applicable to mechanical and rheological measurements of polymers
Simplify the exchange and presentation of data from mechanical and rheological measurements. A document with broadly accepted recommendation based on ISO 67211 will be developed and published in cooperation with instrument manufacturers. The recommendations will be implemented into the software of the participating manufacturers.
http://www.iupac.org/projects/2003/2003-009-1-400.html

Critically evaluated termination rate coefficients for free-radical polymerisation. 1. Current status, evaluation of experimental methods, data for styrene and methyl methacrylate
To begin the task of providing critically evaluated ("benchmark") values of termination rate coefficients for free-radical polymerisation.
http://www.iupac.org/projects/2000/2000-028-1-400.html

Critically evaluated propagation rate coefficients for free-radical polymerisations: acrylic acid alkyl esters
Critical evaluation of propagation rate coefficients for free-radical polymerisation of alkyl acrylic acid esters with this k_p data being derived from pulsed-laser initiated polymerisations in conjunction with polymer molecular weight analysis by size-exclusion chromatography.
http://www.iupac.org/projects/2002/2002-023-1-400.html

Critically evaluated propagation rate coefficients for free-radical polymerisation of water-soluble monomers polymerised in the aqueous phase
To initiate critical evaluation of propagation rate coefficients for water-soluble monomers polymerised in the aqueous phase.
http://www.iupac.org/projects/2004/2004-034-1-400.html

Analytical Chemistry Division

Ionic Strength Corrections for Stability Constants
To facilitate the calculation of reliable stability constants for applications in an assortment of natural media (e.g. seawater, biological fluids).
http://www.iupac.org/projects/2000/2000-003-1-500.html

IUPAC Stability Constants Database - completion of data collection up to 2000+
To complete, as far as is practical, the collection of literature values for stability constants up to the year 2000.
http://www.iupac.org/projects/2000/2000-004-2-500.html

Critical evaluation of stability constants of metal complexes of complexones for biomedical and environmental applications
Critical review of stability constants concerning the complexation of all cations by complexones in aqueous solutions with an emphasis on physiological conditions (35-36 C, 0.15 M NaCl)
http://www.iupac.org/projects/2001/2001-055-1-500.html

Solubility Data Projects
A large number of projects on critical evaluated collections of solubility data

Solubility of volatile and gaseous fluorides in all solvents
http://www.iupac.org/projects/2001/2001-052-1-500.html

Solubility data of compounds relevant to mobility of metals in the environment. Inorganic actinide compounds
http://www.iupac.org/projects/2002/2002-025-1-500.html

Solubility data of compounds relevant to mobility of metals in the environment. Alkaline earth metal carbonates
http://www.iupac.org/projects/2002/2002-031-1-500.html

Solubility data of compounds relevant to mobility of metals in the environment.
Metal carbonates
http://www.iupac.org/projects/2002/2002-032-1-500.html

Solubility data related to oceanic salt systems.
Part I - Binary systems containing sodium, potassium, and ammonium sulfate
http://www.iupac.org/projects/2002/2002-033-1-500.html

Solubility data related to oceanic salt systems.
Part II - magnesium chloride-water and calcium chloride-water and their mixtures
http://www.iupac.org/projects/2002/2002-034-1-500.html

Solubility data of compounds relevant to human health.
Solubility of substances related to urolithiasis
http://www.iupac.org/projects/2002/2002-035-1-500.html

Solubility data of compounds relevant to human health.
Solubility of hydroxybenzoic acids and hydroxybenzoates
http://www.iupac.org/projects/2002/2002-036-1-500.html

Solubility data of compounds relevant to human health.
Solubility of haloganated aromatic hydrocarbons
http://www.iupac.org/projects/2002/2002-037-1-500.html

Solubility data of compounds relevant to human health.
Antibiotics: peptide antibiotics and macrocyclic lactone antibiotics
http://www.iupac.org/projects/2002/2002-038-1-500.html

Solubility data related to industrial processes.
Lead sulfate
http://www.iupac.org/projects/2002/2002-042-1-500.html

Solubility data related to industrial processes.
Carbon dioxide and the lower alkanes at pressures above 2 bar: methane to butane
http://www.iupac.org/projects/2002/2002-043-1-500.html

Solubility data related to industrial processes.
Carbon dioxide in aqueous non-electrolyte solutions
http://www.iupac.org/projects/2002/2002-044-1-500.html

Solubility data related to industrial processes.
Solids and liquids in supercritical carbon dioxide
http://www.iupac.org/projects/2002/2002-045-1-500.html

Solubility data related to industrial processes.
Acetonitrile: ternary and other multicomponent systems
http://www.iupac.org/projects/2002/2002-050-1-500.html

Mutual solubility of hydrocarbons and water (update of SDS Vol 37 & 38)
http://www.iupac.org/projects/2003/2003-018-1-500.html

Chemistry and the Environment Division

A critical compendium of pesticide physical chemistry data
Establish an IUPAC compendium of critically evaluated values for aqueous solubility, vapor pressure, octanol/water partition coefficient, and acid/base dissociation constant (pKa) for active ingredients of pesticides.
http://www.iupac.org/projects/2003/2003-011-3-600.html

Chemistry and the Environment Division

Global availability of information on agrochemicals
Increase the global availability of information on the chemistry of agrochemicals, including methods for testing and evaluation, summaries of properties for individual pesticides, and regulatory standards for pesticides. Information will be communicated via the world wide web and CD-ROM.
http://www.iupac.org/projects/2001/2001-022-1-600.html

Chemistry and the Human Health

Human drug metabolism database
Establish a model database for drug metabolism information that can be mounted on the WEB.
http://www.iupac.org/projects/2000/2000-010-1-700.html

Compendium of targets of the top 100 commercially important drugs
The intention is to provide a resource containing the following information on commercially important drug targets: type of target (enzyme, receptor etc); brief description of relevance to disease treated; reference to knockout information, if available; reference to structural information, if available; listing of the marketed drugs that interact with the target along with potency and relevant selectivity information.
http://www.iupac.org/projects/2004/2004-025-1-700.html

GRANTS PROGRAMME 2005

APPLICATION FORM

Applications must be submitted electronically to rohini@icsu.org
with the following subject line: *GRANTS PROGRAMME 2005*

Deadline for submission is 1 March 2004

Lead applicants' may submit up to two applications under Category 1 Grants (€ 50,000 – € 100,000) and two under Category 2 Grants (less than €50,000)

Project title: Networking Public Scientific Reference Databases

Requested amount (€): Category 1 100,000
Category 2

Applicants:

Lead Applicant (Organization): IUPAC

Contact name & Designation: Antony Neil Davies, Secretary Standing Committee on Printed and Electronic Publications

Email address: tony.davies@creonlabcontrol.com

Supporting Applicant(s) (Organization(s): IUCr, ICSTI, IUBMB

Contact name(s) & Designation: IUCr - Michael H. Dacombe, Executive Secretary
ICSTI - Wendy Warr - IUPAC Rep.
IUBMB - Richard Cammack - Nomenclature Chairman

Email address: tony.davies@creonlabcontrol.com

Priority area:

Please identify the category in which the proposal best fits

a. Science and technology for sustainable development	☐
b. Capacity building in science education	☐
c. Science / Policy Interface	☐
d. Dissemination of scientific information	☒
e. Emerging science and technology – creation of new knowledge	☐

How will ICSU support benefit the development of the selected priority area (max. 10 lines):
The proposed project has at its core the Principles of Science in the Information Society and specifically incorporates points 2, 4, 5 and 8 of the Agenda for Action. By working together a central system linking the publically available chemical, biological, crystallographic and physical property scientific databases will be created. ICSU is the ideal organisation to lead sponsor of this work as the umbrella organisation for the various associated scientific Unions. Not only will the impact of the project benefit science beyond the boundaries of the supporting Unions but will through its mere existance help to more widely publicise ICSU and the various activities of the collaborating Unions amongst each others members. It is the networking of the different disciplines that brings the true added value to the project.

Project plan (max 3 pages)
State clearly the objectives of the project and the beneficiaries. Elaborate on its relevance to the review criteria – innovative nature, interdisciplinary and international nature, visible and measurable outputs, etc. If the activity targets young scientists, women scientists and scientists from developing countries' – please refer to it here.

- Objectives (1/3 page)

A central public scientific reference access portal will be developed as a hub linking reference data collections backed by the participating ICSU ´scientific unions.
The project will document the different reference data collections of the participating unions and linking mechanisms established to generate a cross-referencing between the various reference collections either online (preferable) or cited as being available through the respective organisations.
This unified public scientific reference data portal will initially be based on the modern SQL multi-tier infrastructure used in the International Spectroscopic Data Bank.

- Project description (2 pages)

The idea for this project was conceived during formal discussion sessions and informal meetings with representatives of the Scientific Unions at the ICSTI/CODATA/ICSU "Seminar on Preserving the Record of Science", hosted at UNESCO, in Paris, Feb 14/15 2002. The design and planning of the International Spectroscopic Data Bank was discussed along with the ability to use the available technology as a route through to other public scientific reference databases thereby increasing their availability and profiles.

In the last two years we have made a specific point of bringing the interested parties from the other Unions together at a series of meetings on how to handle reference data starting with the IUPAC XML in Chemistry meetings in Cambridge UK in Jan 2002, CAS/IUPAC meeting on Chemical Identifiers and XML for Chemistry in July 2002 and the two-day XML in Chemistry project meeting in November 2003.

The Science in the Information Society - Agenda for Action has a number points that are extremely relevant to this discussion.
2, Promote sustainable capacity building and education initiatives to ensure that all countries can benefit from the new opportunities offered by information and communication technologies for the production and sharing of scientific information and data.

4, Promote interoperability principles and metadata standards to facilitate cooperation and effective use of collected information and data.

5, Provide long-term support for the systematic collection, preservation and provision of essential digital data in all countries.
7, Encourage initiatives to increase scientific literacy and awareness of how to interpret web-based scientific information.
8, Support urgently needed research on the use of information technologies in key areas, such as geographical information systems and telemedicine, and on the socio-economic value of public domain information and open access systems.

The following project description is the fulfilment of the first steps in generating such a public reference scientific data portal. The project planned in 4 stages. In Stage One the available reference data collections from the participating unions will be reviewed and documented. A technical design will be made for one of the selected databases. During Stage Two this selected database will be linked to the IS-DB system as a prototype for other implementations. Stage Three will see all the reference databases selected in Stage One implemented and Stage Four will review the project and look to the future.

STAGE 1.
At the start of the project the partner Unions and organisations in this project will be required to nominate technical representatives to form the Technical Management Team (TMT).
This team will identify key public reference database resources from the area of responsibility of their Unions. Dr Steve Stein of NIST has been actively involved at all stages of the work cited above and was present at all the meetings and organised the last two. He will bring the knowhow to the project gained through work on the NIST Chemistry Webbook.
The criteria for identifying these resources, which will be named as the Phase One databases, should include public accessibility as well as impact in the worldwide scientific community and an assessment of potential value of synergistic linking to the other resources.
These discussions will be widely publicised and made available for rapid scrutiny within the partner Unions and must include a brief outline of how the various systems will be integrated.
Particular weight will be given to the use of international data standards which are described in various recommendations from IUPAC and this approach will be maintained throughout the project. The use of the IUPAC Chemical Identifier standard as a possible central key will also be tested in Stage 1.
Programing and database enhancement activities will be carried out by the team under Dr. Peter Lampen at the Institute of Spectrochemistry and Applied Spectroscopy - one of the leading scientists worldwide on the standardisation and transport of scientific information, Secretary of the IUPAC CPEP subcommittee on Electronic Data Standards as well as the ASTM international subcommittee E13.15 on Analytical Information Exchange.

STAGE 2
With the action plan agreed for the Phase One databases, a prototype implementation will be carried out for a database selected as being most typical of those in the Phase One list. This selection will be based on issues such as accessibility, infrastructure compatibility, use of international standards within the application layer etc.
This initial work will lay the foundation for the implementation of the cross-linking to the other resources later in the project and technology issues will be clarified on this prototype rather than worrying about all the Phase One databases at the same time.
This demonstration implementation will be shown to all appropriate bodies within the partnering Unions for review and comment on issues which might arise with their own

special needs. This review will be published.

STAGE 3
Following the review the other Phase One databases will be implemented with special interest being paid on identifying the best synergistic features generated by the new ability to provide reference data from these different sources. It might be the case, for example, that a slight change in presentation of the data from a particular source will make the information relevant to a much broader field of use than foreseen by the original stand-alone system.
It is intended during the 3rd quarter of 2005 to present these systems to a wider public within the participating Unions to gain input not only from the respective data specialists but also, and probably more importantly, from the real users.
By the end of stage 3 we will have all the Phase One databases intelligently linked and the various end user reviews which we will publish.

STAGE 4
Based on the various reviews a single strategy document will be generated by the Technical Management Team. In Stage 4 further databases will be selected for Phase Two. Enhancements and necessary revision of the service provided by the system will be made based on the End User reviews. Decisions will be made on the source and mode of further funding should it be required. Appropriate applications will be made at this time.

- Relevance to review criteria 1/3 page)

This project has arisen as a direct result of the actions and discussions around the World Summit on the Information Society. The Internation Spectroscopic Data Bank has at its core the ICSU principles on the Dissemination of Scientific Information with free and open access to all data deposited. The choice of Web technology ensures maximum availability and the infrastructure has been chosen specifically for the ease of deployment to users of differing technology levels and the simplicity of interfacing between different disciplines. By linking the resources available within various Unions the impact of this interdisciplinary project will be far greater than that of the various single programmes on their own.
The deliverables are clear and will be available from anywhere in the world with minimal access ability to the internet. This catalytic activity will encourage other groups to participate in what will hopefully become the central public reference data portal worldwide for freely available reference scientific data.
The Science in the Information Society - Agenda for Action has a number points that are extremely relevant. 2, Promote sustainable capacity building… for the production and sharing of scientific information and data. 4, Promote interoperability principles and metadata standards to facilitate cooperation and effective use of collected information and data. 5, Provide long-term support for the systematic collection, preservation and provision of essential digital data in all countries. As well as 7, Encourage initiatives to increase scientific literacy and awareness of how to interpret web-based scientific information and 8, Open Access content.

- Targeting of priority groups (1/3 page)

Locating and evaluating reference data available through the internet is becoming increasingly difficult as different organisations offer data of varying quality through differing interfaces. Many scientists come to rely on data portals provided by commercial organisations which suddenly become "pay-for-view" services or dissappear entirely due to changes in corporate policy.
For the normal user of public reference data a central portal supported by respected and trusted representative international scientific bodies would be of significant help in their daily work. The increased profile of the individual reference databases achieved through the

proposed networking will enhance their general availability and of the various public reference data activities of the Unions within ICSU

Work plan (max 1 page)
Specify time schedule, major events, methodologies to be used, leadership and management structure, and key milestones in the implementation process, etc.

1st Quarter 2005 - Stage 1.
Identify the central technical contact within the particpating unions to form the Technical Management Team. Team will generate a catalog of the key collections, method of storage and potential interfacing strategies in the reference databases selected by the collaborating scientific Unions as well as the NIST Chemistry WebBook for the first phase of the project. Applicability of the IUPAC Chemical Identifier as the central key to be assessed
MILESTONE ONE CATALOGUE PUBLISHED

2nd Quarter 2005 - Stage 2.
Implementation of the initial cross-referencing infratstructure to a public online reference database and the NIST Chemistry WebBook chosen from the list generated in Stage one.
End of 2nd Quarter - Review of prototype implementation by all collaborating Unions with special concentration on issues arising with their own reference collections and the parameters identified in the Catalogue.
MILESTONE TWO PROTOTYPE RUNNING FOR REVIEW

3rd Quarter 2005 - Stage 3.
Modify the prototype implementation based on the results of the review. Carry out similar prototyping for the remainder of the first Phase reference data collections. Present systems to the participating Unions main yearly meetings for wider comment.
End of 3rd Quarter. Review new prototypes with the various Unions. Review the end-user responses.
MILESTONE THREE PUBLIC PRESENTATION TO THE PARTCIPITATING UNIONS
MILESTONE FOUR ALL SELECTED PHASE ONE DATABASES NOW LINKED IN PROTOTYPE

4th Quarter 2005 - Stage 4.
Update prototypes following extensive review.
Technical Management Team to investigate extending the list of selected databases.
Following a review of the project further funding to be sought to enhance the systems in place and extend the support to the Phase Two public reference databases identified by the Technical Management Team. Return to ICSU and the other Unions not currently in the project for input as to other reference collections.
End of 4th Quarter. Review of project and finalise mid- and long-term strategies for the gradual enhancement and expansion of the systems prototyped during the project.
MILESTONE FIVE PUBLICATION OF PROJECT REPORT

Expected results (max 0.5 page)
What outcomes are expected from the project: publications (including audience and dissemination plan), new programme initiatives, etc., Explain how an ICSU grants can strengthen your own overall programme of work, e.g., leveraging funds from other sources, enhancing visibility, enhancing impact or role of your organisation. Assess potential follow-

on action which may result from the activity.

A new central reference data portal will be established as an Open and Trustworthy resource for the international scientific community. This portal will operate as a central information node to link various public reference data collections supported by the participating ICSU Scientific Unions.

Through the use of unique keys and also full text indexing a simple search mechanism will identify public reference information available at the various locations and lead the user to the desired information whilst making them aware of other relevant data sources on the searched topic or substance.

The experience gained by this project will also have benefits within the partcipitating Unions for their own work as a number of new hurdles such as incompatible ontologies and methods of working between the different disciplines must be overcome.

The portal will build on the state-of-the-art International Spectroscopic Data Bank infrastructure which has been established through a 1.2 Mio € grant from the European Union.

ICSU is the ideal umbrella organisation to support this process. We hope to build a community spirit amongst the various scientific unions promoting the availability and dissemination of public open-access reference quality data. Once the Phase One interconnectivity has been demonstrated in 2005 we will encourage other groups and Unions to come forward with the necessary interest and essential funding to include their own reference data systems into this central ICSU portal.

Other collaborative partners (max 0.5 page per partner)
The role of each partner should be clearly described. New partnerships, between organisations that do not routinely collaborate, are particularly encouraged and should be clearly identified as such.

- Partner 1

International Union of Pure and Applied Chemistry, Committee on Printed and Electronic Publications, Subcommittee on Electronic Data Standards - Overall Project Oversight Body
Programing and database enhancement activities will be carried out by the team under Dr. Peter Lampen at the Institute of Spectrochemistry and Applied Spectroscopy - one of the leading scientists worldwide on the standardisation and transport of scientific information, Secretary of the IUPAC CPEP subcommittee on Electronic Data Standards as well as the ASTM international subcommittee E13.15 on Analytical Information Exchange. Dr Steve Stein of NIST and IUPAC, CPEP and the XML in Chemistry Project Chairman has been actively involved at all stages of the work cited above and was present at all the meetings and organised the mostrecent ones. He will help to integrate the work with the NIST Chemistry Webbook initiative.

- Partner 2

IUBMB - Richard Cammack - Chairman - Nomenclature Committee of IUBMB (NC-IUBMB) and Chairman IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN)
The scientific fields covered by the IUBMB are currently producing some of the larger (size of data sets), and more difficult questions surrounding long-term archiving and open availability of scientific data. The need for action in this area was highlighted for example in

the Standards and Ontologies for Functional Genomics November 2002 at the Wellcome Trust Genome Campus, Hinxton, UK, where the interoperability of systems and nomenclature problems dominated the meeting.

- Partner 3

IUCr - The availability of crystallographic reference information through the Protein Data Bank PDB and the small molecule crystal collection are excellent examples of the need for the scientific community to take responsibility for the collection and dissemination of reference quality data. The experiences gained in the crystallography field will be invaluable to the project technical team. Meetings have already taken place between the IS-DB developers and the small molecule Cambrideg Crystallographic Data Centre CCDC team under Frank Allen in Cambridge UK and the NIST team working on the PDB.

- Partner 4

ICSTI - As a leading international body and a scientific associate body of ICSU, is active in the area of open access, while not taking a firm position on the various options and models under way and the ideas under investigation. ICSTI will support the activities of the project through peer review of the work being carried out and can provide advise on overlap with other initiatives currently underway amongst its member organisations. They have experience outside the analytical science reference data field which could be of use broadening the applicability of the portal. ICSTI is therefore in a unique position to help in the networking of the data sources. They will also form the link to the CODATA organisation who have expressed interest in entering the project at a later phase (John Rumble)

Project budget

Amount requested from the ICSU Grants Programme:		€ 100000
Estimated breakdown of cost		
Research / Content	€ 90000	
Travel / Accommodation for Meetings	€ 10000	
Training / Teaching	€	
Planning / Coordination	€	
Other (specify):	€	
Amount provided by the applicants:		€20000
Amount provided from other sources (specify): ISAS Dortmund		€50000
Total cost of project:		€170000