Pure Appl. Chem., Vol. 71, No. 8, pp. 1587–1591, 1999.
Printed in Great Britain.

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

COMMITTEE ON PRINTED AND ELECTRONIC PUBLICATIONS
WORKING PARTY ON THE INTERNET AND IUPAC PUBLICATIONS*

GUIDELINES FOR THE USE OF THE INTERNET BY IUPAC BODIES

(Technical Report)

Prepared for publication by:
ANTONY N. DAVIES[1], STEPHEN R. HELLER[2] and JOHN W. JOST[3]
[1]ISAS, Institut für Spektrochemie, Postfach 10 13 52, 44013 Dortmund, Germany <davies@isas-dortmund.de>
[2]Guest Researcher NIST/SRD, Mail Stop: 820/113, 100 Bureau Drive, 820 Diamond Avenue, Room 101, Gaithersburg, MD 20899-2310 USA <steve@hellers.com>
[3]IUPAC Secretariat, PO Box 13757, Research Triangle Park, NC 27709-3757, USA <secretariat@iupac.org>

*Membership of the Working Party during the preparation of these guidelines (1997–99) was as follows:

J.W. Jost (USA), A.N. Davies (Germany), S.R. Heller (USA), H.V. Kehiaian (France), A.D. McNaught (UK).

Guidelines for the use of the Internet by IUPAC bodies (Technical Report)

Abstract: The rapid development of the Internet as a major communication tool between scientists has led to the need for a co-ordinated IUPAC presence. Many diverse groups have already initiated distribution of IUPAC related material via their Websites. These guidelines will provide the structure on which the official IUPAC Internet site maintained through the Secretariat will be based. Rules governing the interaction between this central site and various sites operated by other IUPAC bodies are published here as well as guidelines for the operation of sites maintained by other bodies which contain IUPAC related information. The need for special care when making provisional recommendations widely available on the Internet will be emphasised.

INTRODUCTION

The meeting of the Committee on Printed and Electronic Publications, 24–25 August 1997 in Geneva recognized the need for a set of guidelines for IUPAC bodies wishing to extend their activities onto the Internet. It was also clear that certain strategic decisions needed to be taken regarding the use and maintenance of the main IUPAC website. The Internet provides an excellent tool for increasing the profile of the Union.

The main product of the Union is the publications and recommendations of the different commissions. It was felt that in order to maintain the high quality of information emanating from the Union urgent steps needed to be taken to ensure that the rapid introduction of the Internet into scientific life did not result in a lowering of the quality of information linked to the IUPAC name. By their nature the membership of Commissions changes regularly and a mechanism needed to be found to ensure all work invested into developing electronic information had a lifetime well beyond the limited tenure of particular individuals in the commissions.

A further concern was the lack of guidance for commission members on how to implement use of the world wide web to improve efficiency within the commissions.

WWW.IUPAC.ORG

Location

The IUPAC Secretariat at Research Triangle Park, North Carolina, USA, maintains the 'official' IUPAC site on the World Wide Web (WWW). This is located on a very high speed server system linked to the Internet by a high speed high bandwidth professional network with 24 h-a-day availability.

Content

- All 'public' information made available by IUPAC bodies should be mounted on this website.
- Individual groups may wish to mirror their contribution on their group's own websites.
- Recommendations, technical reports and journal articles from IUPAC bodies will be available through the official website. These documents will initially be available in Portable Document Format (PDF) [1] to ensure no loss of format on distribution. Compatibility with the printed version of a document cannot currently be ensured by using the simple Hypertext Markup Language (HTML) encoding implemented in the current generation of Internet browsers. It is also not possible to maintain formatting if the original word processing files are distributed as these will reformat a document depending on the local computer hardware (including printers) and software (operating system, word processor version, installed fonts, etc.);

- Other information of a more transient nature such as commission or working party drafts will be made available in the most appropriate format depending on content. It is important to build in a formal hurdle to receiving preliminary or provisional information. Currently a request for the information needs to be made and distribution is controlled. This practice should continue in electronic form.
- Numerical data, which is produced by some commissions in large quantities, will also be made available through the website but the most appropriate method will be decided on in consultation with the various IUPAC bodies and may well vary depending on data type.
- An agreement with the IUPAC publishers will allow archiving of all IUPAC publications in SGML format [2] at the Secretariat to enable the Union to react to future developments in Internet browser technology. SGML is an ISO standard for content markup in electronic documentation and HTML is a very limited subset of this standard.

Electronic publication

Pure and Applied Chemistry

In the immediate future *Pure and Applied Chemistry* will be available in electronic form via the IUPAC Home Page. Technical details of the mode of access remain to be ironed out.

Chemistry International

The content of *Chemistry International* is usually of a more administrative or informative nature for people working within the Union as well as the world chemical community. As such it lends itself more to presentation on the Internet and is now available online at:

http://www.iupac.org/publications/ci/

Other publications

Other IUPAC publications such as the various color books will be made available in electronic form via the IUPAC website at various times following the release of the printed form. The exact timing of the release of the electronic version and the method of distribution will vary depending on the source of the publication, the content, and individual publishing agreements.

INTERNET HOMEPAGE DESIGN—SOME DO'S AND DON'TS

Basic principles

It is important not to lose sight of the reason for the presence of a particular IUPAC body on the Internet. The delivery of information is either to a restricted group of people or to a wider audience than would usually be the case. Whichever is the case, the presence of the required information should be clear and retrieval easy.

Some do's and don'ts

- Make sure you know exactly what information you wish to deliver, and how it will be delivered. If in doubt—ask for advice at the secretariat.
- Keep the information content up-to-date. There is nothing worse than a website full of links leading nowhere.
- Don't put information onto the website before it is complete. The use of such signs and phrases as 'Website under construction' should be avoided. If the page isn't ready. . .don't link to it!;
- If you have to alter the structure of your Website, ensure that all previously accessible http: addresses remain accessible. . .even if this requires extra messages such as 'this document has been superseded please now refer to http://ournewdomain.country/new_file_structure.html.';
- Although many commercial websites have a substantial amount of graphical gimmickry as standard,

this often distracts from the main aim of the website and is often used to hide a lack of significant content. Ensure that the primary aim of the site—information delivery—is achieved before starting to 'pretty up' the presentation.

- When you are preparing complex file structures keep the simplest naming rules and use only lower case letters to avoid problems which occur when moving between operating systems. A typical example of the problems that can occur is when web pages designed on a PC to be mounted later on a UNIX based server.
- When updating a web page include the date of the last revision.

RECOMMENDATION TO USE CHEMICAL MULTIPURPOSE INTERNET MAIL EXTENSIONS ON IUPAC INTERNET WEBSITES

A CPEP working party is finalizing a publication defining Multipurpose Internet Mail Extensions (MIME) types for chemistry and these should be implemented wherever possible on IUPAC Internet Websites.

A brief overview of the current status and a table of the data file extensions covered is given below but please refer to the website at the following URL: http://www.ch.ic.ac.uk/chemime/ for the latest information.

Multipurpose Internet mail extensions

In order to identify different types of data the Internet Engineering Task Force (IETF) has approved MIME protocols. MIME types are used to identify content of files or parts of files on the Internet. MIME types consist of a primary and subtype. The primary type defines the general type of data in the file, while the subtype defines the exact file format.

Examples of primary file types include text, image, audio and video. A couple of examples of MIME types are video/quicktime for QuickTime movies and audio/x-wav for wave audio files.

Table 1 Suggested Internet chemical mime types

Primary/subtype	Suggested qualifier(s)
chemical/x-cxf	cxf
chemical/x-mif	mif
chemical/x-pdb	pdb
chemical/x-cif	cif
chemical/x-mdl-molfile	mol
chemical/x-mdl-sdfile	sdf
chemical/x-mdl-rdfile	rdf
chemical/x-mdl-rxnfile	rxn
chemical/x-embl-dl-nucleotide	emb, embl
chemical/x-genbank	gen
chemical/x-ncbi-asn1	asn
chemical/x-gcg8-sequence	gcg
chemical/x-daylight-smiles	smi
chemical/x-rosdal	ros
chemical/x-macromodel-input	mmd, mmod
chemical/x-mopac-input	mop
chemical/x-gaussian-input	gau
chemical/x-jcamp-dx	dx, jdx
chemical/x-kinemage	kin

Chemical MIME

A proposal to include common scientific file formats in chemistry as a primary MIME type 'chemical' has been pioneered by Henry Rzepa of Imperial College, London, UK (rzepa@ic.ac.uk) and his colleagues and adopted as a topic for a IUPAC working party. A draft table of some of the chemical MIME types is given in Table 1 below.

What this basically means is that an Internet aware software should recognize a file called *mydata.gau* as a Gaussian input file. And a file called *myspectrum.dx* should be recognized as a JCAMP-DX spectrum and treated as such. The 'x-' refers to the MIME type as being experimental. This will be dropped when adopted by the IETF.

CHEMICAL MARKUP LANGUAGE

It has been suggested that a short description of the Chemical Markup Language (CML) would also be appropriate here although its implementation on IUPAC Internet sites cannot currently be recommended due to a lack of available software support. This is a developing field which may have relevance in the future for reporting chemistry on the Internet and IUPAC bodies should be aware of its existence, especially in case of any overlap with current or future projects.

CML is an application of the extensible markup language (XML) and is being developed to contain chemical information within electronic documents. Currently work is concentrated on chemical structure information but IUPAC bodies should watch the website at: http://www.xml-cml.org/ for developments. There is also a large amount of background information to this project available here.

Many large mainstream development projects on the Internet involve the use of XML owing to the need for and value in electronic commerce. It is therefore only a matter of time before XML, and its chemical child, CML, will find widespread use see: http://www.xml.org/

GLOSSARY

CML	Chemical Markup Language
CPEP	Committee on Printed and Electronic Publications
DTD	Document Type Definition is the formal definition of the elements, structures, and rules for marking up a given type of SGML document. You can store a DTD at the beginning of the document or externally in a separate file.
HTML	Hypertext Markup Language
IETF	Internet Engineering Task Force
MIME	Multipurpose Internet Mail Extensions
PDF	Adobe Portable Document Format
SGML	Standard Generalized Markup Language is an international standard (ISO 8879) published in 1986.
WWW	World Wide Web
XML	Extensible Markup Language

REFERENCES

1 For more information on the Adobe Portable Document Format see http://www.adobe.com/prodindex/acrobat/adobepdf.html

2 On publishing in Standard Generalized Markup Language 'SGML' see http://www.oasis-open.org/cover/sgml-xml.html