



Committee on Publications and Cheminformatics Data Standards

Report to Council, August 2015

Bonnie Lawlor, Committee Chair

Executive Summary

The IUPAC Committee on Publications and Cheminformatics Data Standards (CPCDS) has spent the past eighteen months on the following initiatives:

- The finalization of the transition of IUPAC's publications, *Pure and Applied Chemistry (PAC)* and *Chemistry International (CI)*, to the Walter De Gruyter GmbH Publishing House. The transition for *Pure and Applied Chemistry (PAC)* is finished and production is now running smoothly (see page 2 for details). *CI* faces two challenges: 1) while production has improved, problems remain primarily because De Gruyter's production system was built to support the structure of traditional scholarly journals, not news journals (see page 3 for details); and 2) IUPAC needs to ensure that *CI* continues to meet the ever-changing needs and expectations of its readers, especially the younger generation of researchers who prefer digital information and mobile platforms. To that end CPCDS is currently analyzing the results of a *CI* user survey that they conducted earlier this year in order to develop recommendations regarding the future content, format, and mode of presentation of *CI* (see page 3 for details)
- The evaluation of a proposal from De Gruyter for the creation of a new database generated from IUPAC's Standards and Recommendations. CPCDS recommended that the proposal be accepted pending the completion of a satisfactory business plan. The latter was agreed upon earlier this year and CPCDS is actively involved in the development of the database. Product launch is expected in the first quarter of 2016 (see page 4 for details).
- The development of a vision for a new IUPAC website. A Task Force was established in early 2014 and over that summer the group conducted a survey of IUPAC members in order to learn why and how the current website is used and what features and functionalities are absolutely necessary for the provision of an efficient, smooth, and pleasant user experience. The Task Force recommendations were submitted to the IUPAC Executive Committee in November 2014 to help provide direction for the new site that is currently in development under the direction of the IUPAC Secretariat. A beta version will be available during the General Assembly in Busan, Korea in order to gather additional input (see page 5 for details).
- The drafting of new Terms of Reference for the committee. This was driven by the fact that both the committee's name and scope of activities were revised at the 2013 IUPAC meeting in Istanbul (the former name was the Committee on Print and Electronic

Publications (CPEP)). The new Terms of Reference will be presented for Council approval at the 2015 Council meeting in Busan, Korea (see page 9).

Additional details on the above highlights can be found on the following pages as noted. Committee and Task Force members are listed beginning on page 9 at the end of this report. They all deserve much thanks for the work that they have done over the past eighteen months.

Full Report

CPCDS supports all of the IUPAC Strategic Goals, either directly or tangentially. However, *all* of the activities discussed in this report are primarily related to Goals #2, #4 and #5 - dealing with issues of tools for communication (publications), standardization (IUPAC database on Standards), and the enhancement of chemical education (publications, databases), and the public appreciation of science (new web site). These strategic goals are as follows:

#2. IUPAC will facilitate the advancement of research in the chemical sciences through the tools it provides for international standardization and scientific discussion.

4. IUPAC will foster communication among individual chemists and scientific organizations, with special emphasis on the needs of chemists in developing nations.

#5. IUPAC will utilize its global perspective and network to contribute to the enhancement of chemistry education, the career development of young chemical scientists, and the public appreciation of chemistry

Transition of IUPAC Publications to De Gruyter

Pure and Applied Chemistry

The transition of *Pure and Applied Chemistry (PAC)* to the De Gruyter publication system began in 2013 in preparation for the first issue of 2014. In general terms the transition went smoothly and the first *PAC* issue was published on time. However, there were challenges, most of which have been resolved, including those related to conference issues and the workflow needed to ensure that the related papers appear in the same issue. The De Gruyter platform currently does not allow for a forward linking (“cited by”) feature. However, this feature will be added in August 2015 in the platform upgrade. The CrossMark functionality will not be added in 2015 and a date has not been provided for when this feature will be included on the platform. (Note: CrossMark is a service that gives scholars the information they need to verify that they are using the most recent and reliable versions of a document. Readers click on the CrossMark logos on PDF or HTML documents, and a status box tells them if the document is current or if updates are available).

De Gruyter has hired staff member to work in their Boston, MA (USA) office, spending 85% of his time on IUPAC publications. He has years of publishing experience, including eight years with Wiley-Blackwell where he worked with Societies on their publications. He has learned the

PAC process and has established good working relationships with the *PAC* editors. In summary, the current status of the *PAC* transition is that it is complete and production is running smoothly.

Chemistry International

The transition of *Chemistry International (CI)* to the De Gruyter publication system also began in 2013. However, two factors contributed to making this transition less smooth than that of *PAC*. The first is that the De Gruyter production system was built to handle the traditional format of a scholarly journal not that of a news magazine with short articles, photos, etc. The second is that De Gruyter assumed that the *CI* contractor working for IUPAC served as a copy editor and when they took over the publication process they replaced the IUPAC contractor with one of their own. However, during the transition period it became quite apparent that the original *CI* contractor had offered a lot more added-value and actually served more as a co-editor than as a copy editor. These two factors resulted in the complete transformation of the *CI* production process. While it is certainly better now than it was a year ago, the process remains less than ideal.

In 2014 De Gruyter proposed a digital format for *CI* similar to one that they use for their digital publication *Public History Weekly* (<http://public-history-weekly.oldenbourg-verlag.de>). While CPCDS recommended against using the format, the proposal did raise a number of questions regarding *CI* format, content, features/functionality, etc. It was agreed that IUPAC had insufficient knowledge about how and why users read *CI* and that readers would be surveyed to learn more. The survey was done earlier this year (see below). The business plan for 2016 assumes no changes in format or distribution. Any recommended changes based upon the survey results will be considered for implementation in 2017.

Chemistry International Survey

As a result of the discussions with De Gruyter regarding the format and purpose of *Chemistry International (CI)*, it was decided that IUPAC needed to learn more about *CI* usage; e.g. how *CI* is accessed, read and used in its current print and digital versions; what content is most valuable; what, if anything, needs to be added or changed; and what features and functionalities are absolutely essential for the provision of an enjoyable and informative reader experience. A user survey was sent earlier this year to just under 2,000 IUPAC *CI* readers (141 or 7% responded) and it was found that:

- 70% of the respondents read some or all of each issue
- 53.57 % read *CI* in the office and 40.71% read it at home
- The most read sections are: Featured articles 77.14%; News 72.86%; IUPAC Projects 67.14%; Updates on units/symbols/terminology, etc. 50.00%; Conference reports 39.29%; and Forthcoming meetings 37.14%
- 70.71% do not believe that additional content was needed
- 80.71% read print only; 6.43% read digital only; and 12.86% read both formats
- The preferred format for reading is: Print 74.29%; Digital 13.57%; and Indifferent 12.14%

- If *CI* were only available in digital format 77.86% would read it; 22.14% said that they would not
- The preferred format for reading online articles is PDF 63.57% and HTML 22.14%
- The methodology used to access online content is: Computer browser (HTML-5) 77.14%; Computer (print out a PDF to read) 62.14%; Tablet 33.57%; Smartphone 30.00%; and other 4.29%
- 55.71% read other news magazines focused on the global chemistry community (e.g. *Chemical and Engineering News*, *Chemistry World*, etc.)
- 69.29% have no objection to having advertising in *CI*; 10.71% do object; and 20% had no opinion
- 79% of the respondents were age 46 or older; 10% were age 35 or younger

In addition to the above, survey respondents commented on possible content changes, new features, etc. The IUPAC Executive Committee has asked CPCDS to review the survey results with the objective of developing recommendations on the following issues: the need for any follow-up survey work; the future of *CI* (2017 and beyond) in terms of content, format, and mode of presentation; what *CI* content should be on IUPAC.org and what might available and sold in the public domain; how the IUPAC web site can be used to engage members and others with *CI*; and how *CI* can be used to engage members and others with IUPAC. The analysis is currently underway with the objective of having recommendations by the end of 2015.

IUPAC Database on Standards and Recommendations

When De Gruyter assumed the publishing role for *Pure and Applied Chemistry (PAC)* they were given contractual rights to sell the *PAC* back files as an avenue for generating additional revenue. However, IUPAC allows that material to be accessed via open access. After much discussion on this issue, De Gruyter put forth a proposal in September 2014 to create a database derived from the IUPAC Standards and Recommendations that are published in *PAC*. This would be an alternative a new revenue-generating opportunity for both IUPAC and De Gruyter even though the Standard and Recommendations are freely-available through the back file. De Gruyter believes that the database will make it much easier and quicker to locate and use the content, and that users will be willing to pay for this added-value.

Three database options were presented: 1) the first was a core database that would be created by extracting the Standards and Recommendations from the *PAC* content that De Gruyter already has in house. Required actions are the conversion of the PDF files to XML, the “cutting” of some of the articles into appropriate database entries, and the uploading and hosting of the database. By “cutting” the articles it is meant that there is content (e.g. definitions) that can very easily be identified during the PDF conversion to XML for documents that have formats similar to glossaries or encyclopedias. This content would provide added-value to the database. However, this identification and cutting could not be done easily for full-text articles due to the time and subject expertise required. De Gruyter would cover the cost of creating the core database. However, an option for this core database was to create metadata for the full-text articles that could not be cut. IUPAC could do this on its own or share the cost with De Gruyter; 2) the second database option was a more advanced version of the core database that would involve the use of a Content Management System (CMS) so that database entries could be worked on within

the database itself. The CMS and a technical help desk would be supplied and paid for by De Gruyter; and 3) the final database option was an enriched database that would contain additional information that complements the Standards and Recommendations; e.g. InChI codes, links to related databases, etc. This would require the use of a freelancer and the costs would be shared by IUPAC and De Gruyter.

After much discussion, CPCDS recommended that IUPAC start with the basic, core database (not including the enhancement of full-text articles) and retain the right to upgrade the product if warranted by market acceptance. An agreement in principal pending the development of a mutually-satisfactory business plan was reached in December 2014. While discussions on the business plan took place early in 2015, CPCDS worked in parallel with De Gruyter to develop search criteria, system features/functionalities, and the basic database tree structure (the framework for searching). The database production agreement was signed at the end of April 2015 and CPCDS continues to work with De Gruyter as the development work progresses.

It should be noted that IUPAC has given approval “in principle” for the CMS system (second database option) so that De Gruyter can obtain the necessary license. The CMS system will not be used until IUPAC chooses to initiate the next step in which database entries will be directly written or revised by IUPAC or until IUPAC and De Gruyter decide to have a freelancer work directly within the database for writing and revision of entries. IUPAC will not be charged for the CMS. It was approved to facilitate a seamless move to upgrade the database should IUPAC desire to do so.

It is important to note that the database will not include all *PAC* content, only the Standards and Recommendations. Also, the content will be as of the immediate prior year (e.g. the most recent content in the 2016 database will be from 2015). The database will be updated annually with more frequent updates if the scientific content requires it. The launch will be in the first quarter of 2016 and IUPAC will receive a royalty on sales with payments made annually. The Secretariat will have free access to use the database and to monitor its quality.

CPCDS will monitor the success of the database and consider how the database could evolve into the enriched version if deemed appropriate. The vision for this database needs to be IUPAC-driven as only IUPAC knows what content will best complement the existing standards and recommendations. No commitment has been made to move to this level, but if the initiative goes well, CPCDS members did concur that this enriched database should be pursued and with the CMS already approved, IUPAC and De Gruyter can move more easily in that direction.

Web Vision Task Force

In order to have a new or updated website in place for the IUPAC General Assembly in August of 2015, IUPAC President, Mark Cesa, established an *ad hoc* Task Force under CPCDS in early 2014. The charter of the Task Force was to assess the state of the current IUPAC Website and recommend improvements that would bring the site quickly and inexpensively to a state that is satisfactory for IUPAC National Adhering Organizations, volunteers, staff, and the chemical and general publics. Once the Task Force had completed the vision, a second Task Force would be formed with the goal of implementing the necessary changes, within IUPAC’s financial means, to bring the site to the recommended standard.

The Task Force first met in April 2014 during the IUPAC Bureau meeting in Coimbra, Portugal. At that time it was agreed that broader input was desirable and to that end an online survey was developed to collect information on why people use the web site, how they use it, with what frequency, what they like, don't like, etc. The ultimate goal is to create an attractive, intuitive "task-oriented" website, so that the user can get in, get what they need, and get-out with a minimum number of click-throughs and distractions.

The Web Vision Survey was distributed on June 30, 2014. It was sent to all Division Presidents for distribution, to a list of active committee and project members, and was posted on the IUPAC Discussion Board. Reminders were sent out in August. The number of responses was far less than desirable (45). While some interesting data was gained that re-enforced individual opinions on the website, it is hoped that additional input is gained as the actual project moves forward, especially during the General Assembly in Busan, Korea, where a beta web site will be used to gather feedback.

The following is a summary of the survey responses:

- 60% of the respondents use the web site often.
- 68% of the usage is with regard to governance activities
- 62.86% use the site to access *Chemistry International* (*important as the future format/distribution of CI is being considered*).
- 50% regard the speed of downloading as average (11% don't download at all)
- 38.64% want the site updated weekly
- 55% said that content curation needs to be done by both those who create the content and IUPAC staff
- 35% said that the home page is intuitive (29% believe just the opposite - the remainder have no opinion)
- 23.5% think that the web site is easy to navigate (28% believe just the opposite - the remainder have no opinion)
- 10.41% cannot find information easily (24% believe just the opposite - the remainder have no opinion)
- 11.44% do not believe that the web site has all of the IUPAC information that they need (32% believe just the opposite - the remainder have no opinion)
- 12.38% do not believe that the website promotes a positive image of IUPAC (29% believe just the opposite - the remainder have no opinion)
- 13.47% do not believe that the website is effective with the general public (23% believe just the opposite - the remainder have no opinion)

The features given the highest number of top 5 rankings for inclusion in a new website are (in order of priority):

- Ease of Navigation (clear winner)
- Content for the General Public to promote the image of chemistry
- Linking to related sites
- The ability to upload/update and edit
- More content/Group workspace for sharing content (equal rankings)

Other possible features, e.g. social media, mobile access, job postings and a “My IUPAC” personal section shared similar rankings, but all were well below the top five. In the total rankings they remained at the bottom of the list in the following order: mobile access, social media, job postings and My IUPAC.

Most of those who responded to the survey were in governance; 47% were between the ages of 25 and 55; and 18 countries were represented.

The Task Force submitted the following recommendations to the Secretariat and IUPAC Executive Committee in November 2014:

1. Platform Recommendation

That the new IUPAC website be based upon a platform other than the current TYPO3 as that system is neither easy-to-use nor readily supported within the United States. Recognizing its own lack of technical expertise the Task Force did not suggest a specific platform, but recommended that a new platform be selected after an external third-party technical assessment was conducted that looked at 1) the recommended website features and 2) how the new website will need to be integrated with the IUPAC in-house database. It was also recommended that the platform support user-authentication to facilitate member control over the input of their own personal information and that if IUPAC needs to implement the recommended new website features and functionalities in phases, the platform should provide the flexibility to add such features relatively easily and at a reasonable cost.

2. Feature/ Functionality Recommendation

That all of the features be included in the new website. However, if a phase-in is required, only “Job postings” and “My IUPAC” should be relegated to a later phase. The rationale for including social media and mobile access in the first phase is that these features are very closely-tied to promoting a positive, modern image of IUPAC to both the general public and others within the scientific community. In addition, they are essential communication tools for today’s young generation of researchers whom IUPAC wishes to attract and retain as members. The Task Force did not recommend a specific structure or site map for the new website, but said that a professional and experienced website designer undertake an unbiased assessment of the desired website features, its usage, and its management requirements with the goal of providing a proposed structure to which IUPAC leadership and this Task Force can react and ultimately revise.

3. General Guidelines

- That the Homepage should be visually attractive and provide prominent search starting points that will facilitate quick access to the most frequently-used IUPAC content with a minimum number of clicks. Site navigation must be quick, easy, and intuitive for both experienced and novice users. The IUPAC Home Page must also appeal to varied audiences. While the primary audience is IUPAC members and others within the scientific community, the secondary audience is the general public, including high school

and college students who may be seeking information on chemistry-related topics. Reaching out to this secondary audience has been identified as a key requirement for the IUPAC website of the future. IUPAC's stated Vision is as follows: "IUPAC advances the worldwide role of chemistry for the benefit of Mankind." The website must demonstrate those benefits and provide access to information that will be of interest to the secondary audience.

- That Project information, the second most sought-after information on the IUPAC website, be highly-visible and searchable on multiple data points (e.g. project number, Division, Committee, members, etc.).
- That the future of IUPAC publications and conferences be considered. These represent the third most sought-after information on the web site. The publications most used on the IUPAC site in the past have been *Pure and Applied Chemistry (PAC)* and *Chemistry International (CI)*. However, while these publications are now being published by De Gruyter and by contract must be accessible only their site, IUPAC should have the option to make them accessible on the IUPAC site in the future 1) if the De Gruyter relationship ends, 2) if IUPAC chooses to host the open access files in order to drive traffic to the site, and/or 3) if IUPAC chooses to again host *CI* as part of the public face of chemistry. Also, *CI*, through May 2013 is currently on the IUPAC site behind a fire wall and not accessible from the outside. This was done with De Gruyter's approval in order to preserve the links from *CI* to other parts of the site (e.g. projects, conferences, etc.) that had been built manually over a decade. This data and the links should be preserved in the new site.
- That Management develop content curation policies to ensure quality and consistency while meeting member needs and expectations. One of the top five requirements of the new website is the ease of uploading, updating, and editing content, but opinions differ on how these efforts should be accomplished; e.g., 23% of survey respondents believe that those who create the content (Divisions, Committees, etc.) should be responsible for data curation; 14% believe that the responsibility lies with IUPAC staff; and 55% believe that the responsibility is a combination of those two efforts in order to maintain uniformity.
- That Management must decide what content will be accessible on the IUPAC website moving forward. Many survey respondents do not believe that the IUPAC website has all of the information that they need. However, the information may be on the website, but buried too deeply to be found or the links are broken, etc.
- That Management decide what existing information can be used to inform non-scientists, including students (high school, undergraduate, other) who may be searching for information on a current issue or of a historical nature (e.g. periodic table), and what additional content, if any, needs to be added over time. Can information from *CI* be deployed here?

- That Management consider if it wants to use the website for either one- or two-way communication with members. The use of social media and the creation of group workspaces have been identified as desired features for the new website. However, there are other possible communication mechanisms such as group e-mails, listservs, etc. that can be used to disseminate surveys, important information, etc. If the website is to serve in a communication role it should be decided upfront. Note that survey respondents want a replacement for the current discussion boards and group workspaces may serve that role as noted earlier.
- That Management decide where and how the new website is hosted. However the Task Force did recommend that IUPAC consider keeping an arms-length relationship with the organization that is selected and not have a member organization host the site as in the past.
- That in seeking proposals for the development of a new web site, other issues to be considered are: search speed; web analytics to identify usage, trends, etc; search engine optimization (SEO) to ensure that the IUPAC site is easily found; and frequency of updating.

It should be noted that the final Task Force report included an excellent and poetic history of the current IUPAC website written by Fabienne Meyers. It has provided significant input into how the new web site should (and should not) be built.

CPCDS Terms of Reference

It was decided in 2013 during the General Assembly held in Istanbul, Turkey that the scope of activities of the IUPAC Committee on Print and Electronic Publications (CPEP) be expanded to cover cheminformatics data standards. This change was approved by the IUPAC Bureau and the committee's name was changed to the IUPAC Committee on Publications and Cheminformatics Data Standards (CPCDS). The Committee drafted new Terms of Reference that were approved by the IUPAC Bureau in November 2014.

The current approved Terms of Reference that were originally developed for to CPEP are as follows:

(i) To advise the President, Executive Committee, other Standing Committees, Divisions, and Commissions on all aspects of the design and implementation of printed and electronic publications, including computerized databases of all sorts, and to promote the compatibility of electronic transmission and storage of information,

(ii) To make recommendations to the President and the Executive Committee on matters of policy and procedures related to the production and dissemination of printed and electronic publications,

(iii) To advise the Secretary General and the Executive Director on hardware and software requirements for the Secretariat and on the development and operation of its computer systems,

(iv) Subject to approval by the President and the Executive Committee, to establish Advisory Boards, Subcommittees, and Working Groups as needed to carry out specific functions of the Committee.

The proposed revision (to section (i) only, changes are bolded).

(i) To advise the President, Executive Committee, other Standing Committees, Divisions, and Commissions on all aspects of the design and implementation of **publications and data-sharing**, including computerized databases of all sorts, and to promote the compatibility of **the** electronic transmission, **storage, and management of digital content through the development of standards for the creation of a consistent and interoperable global framework for human and machine-readable chemical information.**

Changes

On line 2 of the current section (i) the differentiation between print and electronic publications has been removed since the committee believe that this distinction is no longer necessary in today's publishing environment. "Data- sharing" has been included to reflect the current need for this activity in e-science. The wording that has been added following the term "storage on line 3 reflects the approved breadth of the committee's activities as had been suggested at the Istanbul meeting; i.e. to support the development of standards that will support the efficient exchange/communication of content.

Members of the Committee on Publications and Cheminformatics Data Standards (CPCDS 2014-2015

Chair: Bonnie Lawlor (USA)
Secretary: James Liu (USA)

Titular Members:

Jeremy Frey (UK)
Kazuhiro Hayashi (Japan)
Lene Hviid (The Netherlands)
Wolfram Koch (Germany)
Robert Lancashire (Jamaica)
Bono Lučić (Croatia)
Miloslav Nic (Czech Republic)
Leah McEwen (USA)

Associate Members:

Mark Kinnan (USA)

Ex Officio Members:

Hugh Burrows (Portugal)

Web Vision Task Force Members

Christopher Brett (Portugal)
Mei-Hung Chiu (China)
Karl-Heinz Hellwich (Germany)
Fabienne Meyers (USA – IUPAC Associate Director)
Tom Perun (USA)

Ex Officio:

Brynn Hibbert (Australia)
Roberto Marquardt (France)

IUPAC Database Task Force

CPCDS Members (see above)
Steve Bachrach (USA)
Brynn Hibbert (Australia)
Colin Humphris, IUPAC Acting Secretary General
Dave Martinsen (USA)
Fabienne Meyers, IUPAC Associate Director
Lynn Soby, IUPAC Executive Director
Ron Weir (Canada)

Publications Transition Task Force

CPCDS Members (see above)
Steve Bachrach (USA)
Javier Garcia-Martinez (Spain)
Colin Humphris, (UK - IUPAC Acting Secretary General)
Dave Martinsen (USA)
Fabienne Meyers, (USA - IUPAC Associate Director)
Lynn Soby, (USA - IUPAC Executive Director)
Ron Weir (Canada)
