



INTERNATIONAL UNION OF  
PURE AND APPLIED CHEMISTRY

## (SCDS) Subcommittee on Cheminformatics Data Standards

### IUPAC

Committee on Publications and Cheminformatics Data Standards (CPCDS)

SubCommittee on Cheminformatics Data Standards (SCDS)

February 9, 2017, 9am - 10am US-EST

(Web conference)

### Minutes

**SCDS Members Present:** Leah McEwen, Dave Martinsen, Ron Weir, Greg Banik, Ian Bruno, Andrey Yerin, Robert Lancashire, Stuart Chalk, Tony Davies

1. *Acceptance of Minutes of December 20, 2016*  
Approved
  
2. *Cheminformatics Color Book*
  - a. Thanks to all who participated in the informal SCDS brainstorm on 2017-01-27, raw notes are in the SCDS Google Drive, including links to other relevant activities
  - b. Many good questions and ideas were raised concerning:
    - i. Ideal handbook content for cheminformatics standards vis-a-vis IUPAC strengths
    - ii. Form of a cheminformatics handbook for machine accessibility
    - iii. Community engagement and dissemination methods
    - iv. Looking to exemplars in other fields
    - v. Propose preliminary scoping exercise to gauge interest of community and necessary expertise and commitment
  - c. This lead to a draft project proposal for developing content scope and work plan to address these considerations
    - i. Goal is to support a meeting of SCDS/the task group with key stakeholders within IUPAC at the General Assembly in Sao Paulo next July
    - ii. Engage broader broader chemistry community input and feedback
    - iii. Report back end of CY 2017

**ACTION:** SCDS members will consider if they or others in the chemical information community would be interested to join the task group

3. *JCAMP-DX status and future* (<http://old.iupac.org/jcamp/>)
  - a. There have been some inquiries on the status of JCAMP-DX availability of support mechanisms such as format validation software



INTERNATIONAL UNION OF  
PURE AND APPLIED CHEMISTRY

## (SCDS) Subcommittee on Cheminformatics Data Standards

- b. There is compelling reason to continue to support JCAMP-DX - there are millions of existing files available to the chemistry community in this format, and several spectra data publishers are using it.
- c. There is potential for JCAMP-DX to serve more broadly with updating and community engagement - it is available on most instruments and could potentially support an archival niche as a software-agnostic international publishing standard (for funder mandated research data deposit, for example).
- d. At minimum to start, the content of the former CPEP Subcommittee site for JCAMP-DX protocols needs to be reviewed and transferred onto the new IUPAC website. This will be an opportunity to re-familiarize us with the current status and information available regarding the specifications available for different types of spectra and where relevant software and data files might be.
- e. From an updated site we can start to get feedback on the standard in terms of things that it cannot accommodate as a way for us to get feedback on the needs of the community, relative to spectral data standards.
- f. The published specifications in PAC will remain openly available, for the present via the PAC archive: <https://www.iupac.org/publications/pac/index/>
- g. There is general consensus that a validator would greatly improve the utility of the existing standard as there are many challenges with syntax in writable files. The expertise required is likely present within SCDS and we can continue to discuss how to proceed with this, as well as strategies for garnering feedback on the needs of the community with respect to spectra data file formats.
- h. There is enough feedback among the SCDS members to indicate a potential project to formally update the specification and to connect it to other analytical data standards and initiatives (e.g., AnIML, Allotrope, etc.). The scope of this work could arise from further feedback and engagement from the community. A significant aspect of this work will need to involve business cases for 'packaging' and delivering the JCAMP-DX standard and other digital data standards. Buy-in from key stakeholders and influencers will be critical.
- i. There may be relevant lessons from crystallography and supporting the CIF standard format - journal guidelines as a significant driver, domain data repositories cleaning up files; checkCIF and others tools for supporting data generation, analysis, deposit and publication; requirements for processed and even raw data in addition to derived data to prevent fraud; chemical representation of the structure as a key metadata requirement.

**ACTION:** Robert and Leah will work through the material on the former SEDS site in the coming month and prepare a directory.



INTERNATIONAL UNION OF  
PURE AND APPLIED CHEMISTRY

## (SCDS) Subcommittee on Cheminformatics Data Standards

4. *Update on the status of the Gold Book*, reprinted from email update provided by Stuart Chalk on 2017-02-08:
- a. Conversion of existing static webpage site to HTML5 version as static pages (using XSLT and PHP scripts). This was implemented as an interim solution to stabilize the GoldBook site while the dynamic site is built (next two months)
  - b. HTML5 compliant pages that now work better for tablets and phones (though not as well as the dynamic site will)
  - c. Fixed a number of erroneous links that did not work on current site
  - d. Removed structure search and goldify functions (these will be added/improved on the dynamic website)
  - e. Integrated GoogleAnalytics into each page
  - f. Create new JavaScript search function (available on each page - can be reused on dynamic site)
  - g. Tested using Integrity v6 (webpage link checker) - 106553 links checked on 37955 pages, 0 bad
  - h. Files on site:
    - i. 43,930 HTML files
    - ii. 52,693 PNG files (mainly images for the maps)
    - iii. 11,862 SVG files (mainly images for the math displayed on the site)
    - iv. 7,070 PDF files (one for each of the terms + a few others)
  - i. Analysis of website files, including ancillary files stored on the site
  - j. Provided a clean picture of the structure and functionality to be built into the new dynamic website
  - k. Discovered a number of useful files for use in constructing the database
    - i. dois.xml (a full list of Gold Book DOI's in XML)
    - ii. fulltext\_source.xml (term titles and descriptions in XML - text only)
    - iii. query.log (log of searches done on the site since Sep 2011)
    - iv. data.js (terms and their descriptions in a JavaScript variable (as a text file))

**ACTION:** SCDS members will test out the new (interim site) and provide feedback:  
<http://dev.goldbook.iupac.org/gbook/>