	(CPCDS) Committee on Publications and Cheminformatics Data Standards
INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY	

IUPAC Database Tree Structure Discussion
Teleconference
10:40am - 11:30am EST
January 22, 2015

Draft Minutes

****

IUPAC Attendees: Bonnie Lawlor, Chair, Committee on Publications and Cheminformatics Data Standards (CPCDS), Hugh Burrows, Jeremy Frey, Kazuhiro Hayashi, Colin Humphries, Mark Kinnan, Robert Lancashire, James Liu, Bono Lučić, Dave Martinsen, Leah McEwen, Fabienne Meyers, Mila Nič, Lynn Soby

De Gruyter Attendees: Katharina Butsch

The meeting opened at 10:40am EST. The objective of today's meeting was to discuss possible changes to the tree structure drafted by Katharina Butsch for the IUPAC Database on Standards and Recommendations (see pages 6 - 9).

Database Objective

To put the discussion in context Bonnie Lawlor reiterated the purpose of this database as detailed in the minutes of the CPCDS meeting held on September 26, 2014 during which De Gruyter presented the database concept:

The added value of the database will be the ease of locating specific information of interest regarding IUPAC's standards and recommendations. Currently, the information seeker has three channels through which he/she can access this information: first, by going to IUPAC.org and searching the archives of standards and recommendations (to do this the searcher must know the year in which the material was published or the IUPAC Division that submitted the material); second, by searching *Pure and Applied Chemistry* (*PAC*) on IUPAC.org by refining the search to limit it solely to standards and recommendations; and third, through a Google search that may result in an unwieldy large number of results. Also, the Google search results may not be the most current documents, but rather those most frequently accessed and opened. It is to be noted that only about one-quarter of the documents containing standards and recommendations are short (1-2 pages). About half are full-article size (10-20 pages) and the remainder range from fifty-one pages to small booklets of more than one hundred pages. So searching for specific IUPAC Standards and Recommendations can be difficult and time-consuming. De Gruyter put forth the dedicated database concept as the most efficient and effective method for retrieving specific information on the IUPAC standards and recommendations.

Three database options were presented at the September meeting: 1) a basic core database created through the conversion to XML of the existing standards and recommendations PDF files as they were published in *PAC*; 2) a "living" database that would allow content to be entered, edited and updated via a content management system; and 3) an enriched database that would contain additional relevant information and more advanced search and retrieval features.

IUPAC has agreed in principle (subject to final contract and business plan) to initially go forth with the basic core database. IUPAC has also agreed to the inclusion of the content management system

option so that if the core database is successful a more advance version of the database could be more easily created.

Bonnie asked that everyone keep in mind that the tree structure under consideration is for the core basic database.

Tree Structure Discussion

The draft tree structure was then discussed. Leah McEwen had circulated her comments prior to today's meeting and Bonnie asked Leah to summarize her thoughts.

Leah said that she was trying to look at a variety of approaches - tree structure, flat structure, etc. – and that she is all about looking at a variety of ideas. If a tree structure approach is taken, she believes that the flexibility of assigning subcategories to more than one major heading must be permitted; e.g. "spectroscopy" certainly fits under "analytical methods," but is also applicable elsewhere. She also expressed concern that the classic disciplinary headings might be too broad and may not align with the users' needs if they are coming in from a different sub-discipline and/or with a specific measurement or process type question. Her final comment on the tree structure was regarding measurements and whether or not the term "Measurement" should be a main heading in and of itself.

Leah also put forth two ideas related to the categories listed in the draft tree structure, but that are slightly different approaches than a tree: 1) A two-phase faceting or pre-filtering where the user can select both a subject/process and a type of recommendation; and 2) A flat list of 20-30 topics not specifically listed by general chemistry areas since many topics and techniques are cross-disciplinary or not even associated with a classical research discipline (see Leah's examples on page 10).

With regard to Leah's comments, the following was discussed. Bonnie noted that during the database discussion held by teleconference on December 8, 2014 most CPCDS members commented that subheadings might need to be assigned to more than one major heading and Bonnie asked Katharina if that is possible. Katharina said that it is technically feasible, but perhaps there should be a limit on the number of major headings to which a subcategory is assigned. Katharina also noted that "Measurement" is included as a subheading under "Analytical Chemistry," but it certainly could be pulled out as a major heading if CPCDS members believed that this would be important to those searching the database.

Bonnie also said that during the December discussion it was agreed that faceting would not be an option "out of the gate" with the core database, but both IUPAC and De Gruyter agreed that it is on the wish list for the future.

Katharina expressed concern that Leah's flat list of topics might be too general. She noted that those listed on the draft tree structure were chosen because they reflect categories used in major scientific databases and searchers are already familiar with them. However, she agreed that perhaps they may need to be revised to reflect the content of the IUPAC standards and recommendations. She did note that she is confident that there is IUPAC content for all of the categories listed on the draft tree as that was also a criteria for their selection. She did not want to set up categories for which searchers would not retrieve results. Katharina did say that any of Leah's topics can be added to the tree structure if CPCDS believes they are important for the user.

Jeremy Frey also submitted comments prior to the meeting (see page11). He said that a drawback of a tree structure is the difficulty of moving from the bottom of one tree across to another tree that might contain relevant content. Once a user drills down through the subheadings of one major topic, how

easy will it be for them to move from that tree to another? He also said that he might find it interesting to see what other documents searchers accessed/retrieved when performing a search on the same or similar topic. For example, if he was looking for information on how to calibrate a specific instrument it might be valuable to know what sources others used when seeking to answer the same question. A feature that allows the searcher to find relevant information in places that he/she might not have initially considered would certainly be of value.

Katharina said that she believes that the users of this database will be seeking very specific information. If after they find the answer to their question using the advanced search feature they were also interested in finding other relevant information, the browsing structure would offer that capability. She agreed that the ability to assign subheadings to more than one major topic would be of value in this situation. She also noted that in addition to these subheadings there is the possibility of linking between entries in the database and she noted that all of the existing metadata created for an article in *PAC* will also be retained in the database. Jeremy agreed with Katharina.

Bonnie asked that Katharina clarify what level of linking will be included in the core database because she thought that linking would not be possible until the enriched database was created. Katharina said that De Gruyter is considering linking terms to glossaries and/or encyclopedic entries. She said that such linking would be limited and very basic as they will not have the scientific editorship required to evaluate the scientific meaning of a term and identify other terms to which it could be linked. But basic links to provide some added-value to the user may be possible.

Mark Kinnan said that he understands the difficulty of building a tree structure. He is currently involved in building one for internal use in searching patent information in his company. He said that there are regular debates about headings, the location of subheadings, etc., and that they won't really know the ultimate effectiveness of the structure until it is implemented and used. He asked if De Gruyter will be tracking the use of the tree structure? If users have to drill down more than two clicks on average to find what they want, will we be able to revise it and move things closer to the top? Katharina commented that it might be good to have a certain amount of flexibility with the tree structure as the database is being built, at least with regards to subheadings, so that those processing the content can make any changes that they believe are necessary. She also said that in order to address Mark's question she needs to find out how flexible the tree structure will be once the database has been created.

Bonnie asked if the Committee believes that if the draft tree structure is revised 1) to reflect the recommended changes to the analytical section suggested by Brynn Hibbert (see page 12), 2) to incorporate some/all of the topics recommended by Leah McEwen, and 3) to allow the assignment of the same subheading to multiple topics, it will allow searchers to find the information that they need within the context of the core database as described above. Katharina said "yes" and Jeremey said that he believes that this would be the most practical solution.

Tree Structure Terms

In addition to the terms discussed above ("Measurements," and Leah's suggested topics), there were other terms/heading that had come under scrutiny during the December discussion for which Katharina requested clarification. For example, in a discussion of the section on "Industrial Products and Materials," Colin Humphries had suggested that the heading changed to "Materials," that pesticides be moved under "Environmental Chemistry," and that the entry "chemical weapons" be deleted. He had also suggested that "food chemistry" be moved, but it was unclear where Colin wanted it placed. James Liu asked it should be a main heading or if it should be placed under a new heading, "Agriculture." Colin said that he thinks we are struggling with concepts related to pure

chemistry versus applied chemistry and Katharina asked if there should then be a separate heading entitled "Applied Chemistry." After some discussion, the following was agreed upon:

- Applied chemistry topics would be placed for the most part under pure chemistry headings
- Agriculture would be a subheading under "Environmental Chemistry"
- Food chemistry, fertilizers, pesticides, etc. would be under "Agriculture"

James Liu asked two questions: 1) Will we be able to revise the tree structure once the database is implemented and 2) can an article be assigned to more than one major heading? Katharina said that if IUPAC decides to move forward beyond the core database and the content management system is implemented we can look at modifying the tree structure. This would have to be done for the enriched database especially if new content is added. And she said that technically an article can be placed under more than one major heading, but that we might initially want to limit the number of placements to two or three. Bonnie responded by saying that we can place an arbitrary limit initially, but might have to adjust it if we find that the number is too limiting.

Katharina said that De Gruyter has identified a freelancer who will be doing the article tagging and said that if we have any general rules that we would like to have that person follow (e.g. the number of headings to which they can assign an article) we should let Katharina know.

Bonnie asked if IUPAC will have an opportunity to review some of the freelancer's work to see if the tree structure is working, since the freelancer will be assigning terms based upon that structure. Katharina said that she did not believe that this was in the schedule, but that she will see what is possible.

Bonnie said that she will circulate a revised tree structure based upon today's discussion and asked that Committee members review the document and submit their final set of revisions by Friday, January 30th (now revised to Sunday, February 1st) regarding:

- Major headings (e.g. should "Measurement" be a main heading?)
- Subheadings (a subheading can appear under more than one major heading and this will facilitate cross-tree searching so add as you see fit):
- Inclusion and placement of Leah's suggested topics

Also, she asked that Committee members recommend any "general rules" that they believe the freelancer should follow.

There being no further discussion the meeting was adjourned at 11:30am EST.

Respectfully submitted,

Bonnie Lawlor, Chair
IUPAC Committee on Publications and Cheminformatics Data Standards (CPCDS)
January 24, 2015

The following action items resulted from today's meeting:

ACTION ITEMS

1. **Bonnie Lawlor** will schedule the next IUPAC-De Gruyter teleconference for the week of February 1, 2015. It will be cancelled if not needed.

2. **Bonnie Lawlor** will circulate a revised tree structure for review and comment by the Committee.

3**. Bonnie Lawlor** will draft minutes of today's meeting and circulate to the Committee.

4. **Katharina Butsch** will find out if De Gruyter will review user activity on the tree structure to determine if it needs to be adjusted (e.g. how many clicks is it taking for the user to get the information that they need?). She will also see if such adjustment is possible after the database is implemented.

Answer: Yes, it is possible to shift sections within the database to another position after publication.

5. **Katharina Butsch** will find out if IUPAC will have an opportunity to review the freelancer's work to see if the proposed tree structure is working.

Answer: Yes, but we need to discuss the workflow implications.

6. **All Committee members** will review the revised tree structure and submit their final set of revisions to Bonnie by Sunday, February 1st) regarding:

- Major headings (e.g. should "measurement" be a main heading?)
- Subheadings (a subheading can appear under more than one major heading and this will facilitate cross-tree searching so add as you see fit):
- Inclusion and placement of Leah's suggested topics

7. **All Committee members** will submit their suggestions regarding any "general rule" that they believe the freelancer should follow when reviewing articles and assigning terms to the tree structure.

****

Initial Draft Tree Structure

> **Analytical Chemistry**

>> Actinometry
>> Calibration
>> Isotopic and Nuclear Techniques
>> Microscopy
>> Purification
>> Quantitative Measurements
>>> Calorimetry
>>> Conductometry
>>> pH-Measurements
>> Reagents
>> Sensors
>> Separation
>>> Adsorption
>>> Chromatography
>> Spectroscopy
>>> Electron Spectroscopy
>>> EPR Spectroscopy
>>> Flame Spectroscopy
>>> Massspectroscopy
>>> Mössbauer Spectroscopy
>>> NMR Spectroscopy
>>> Optical Spectroscopy
>>>> Absorption Spectroscopy
>>>> Circular Dichroism Measurement
>>>> Fluorescence Spectroscopy
>>>> Interferrometry
>>>> Spectroelectrochemistry
>>> Rheology
>>> Vibrational Spectroscopy
>>>> IR-Spectroscopy
>>>> Raman-Spectroscopy
>>> X-Ray Spectroscopy
>> Stability Constant Measurements
>> Standards
>> Trace Analytics

> **Biochemistry**

> Bioengineering and Biotechnology
> Biotransformation
> Enzymatic Processes
> Microbial Processes
> Biomolecules

> **Environmental Chemistry**
> Atmospheric Chemistry
> Pollution
> Geo-Chemistry
> Soil
> Oils
> Water Chemistry
> Concentrations
> Ground Water treatment
> Pollution

> **Industrial Products and Materials**

- > Chemical Weapons
- > Food Chemistry
  - > Additives
  - > Contaminants
  - > Fats
- > Materials
  - > Ceramic Materials
  - > Coatings
  - > Colloids
  - > Composites
  - > Fullerenes
  - > MOFs and Coordination Polymers
  - > Porous Materials
    - > Macroporous Materials
    - > Microporous Materials
  - > Thin Films
- > Pesticides
- > Polymer Chemistry
  - > Nomenclature
  - > Structure
    - > Stereochemistry
  - > Polymerization
  - > Properties
  - > Substances
    - > Copolymers
    - > Crystalline Polymers
    - > Ionic Polymers
    - > Membranes
    - > Nanotubes, Fibers
    - > PET
    - > Polystyrene
    - > PVC
    - > Resines
    - > Rotaxanes
    - > Rubbers

	(CPCDS) Committee on Publications and Cheminformatics Data Standards
INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY	

> **Inorganic Chemistry**

- > Elements
  - > Atomic Weights
  - > Isotopes
  - > Nomenclature
- > Inorganic Substances
  - > Boron Substances
  - > Coordination Compounds
  - > Metals
- > Processes
  - > Solid State
  - > High Temperature

> **Medicinal Chemistry**

- > Clinical Chemistry
- > Diagnostics
- > Laboratory Technology
- > Properties and Units

> **Organic Chemistry**

- > Catalysis
  - > Biocatalysis
  - > Catalysts
- > Organic Molecules
  - > Carbenes
  - > Cyclic Compounds
    - > Heterocycles
  - > Protecting Groups
  - > Radicals
- > Organic Reactions
- > Stereochemistry

Continued

- **Physical Chemistry**
  - Colloids and Surfaces
  - Electrochemistry
    - Corrosion
    - Electrodes
    - Electrolytes
    - Electroanalysis
      - Potentiometric Measurements
      - Voltammetrie
      - Electrochemical Stripping Analysis
    - Piezoelectricity
  - Magnetochemistry
    - Processes
      - Non-Aqueous Reactions
      - Transport Phenomena
  - Photochemistry
    - Luminescence
      - Quantum Yields
  - Phase Equilibria
  - Thermodynamics
  - Kinetics

- **Theoretical Chemistry**
  - Force-Field Calculation
  - Molecular Design
  - Quantum Chemical Calculation
    - Ground-State Structures
    - Properties
  - Quantum Mechanics

- **Toxicology**
  - Ecotoxicology
  - Exposure to Chemicals
  - Immunotoxicology
  - Metals
  - Toxicokinetics

Leah's Topics

A flat list of 20-30 topics not specifically listed by general chemistry area since many topics and techniques are cross-disciplinary or not even associated with a classical research discipline. Essentially most of the topics at the second level of the tree, my rough take below. More specific terms could be applied as keyword tags (such as specific spectroscopic techniques) and appear as the user clicks down, similar to the Amazon categories (this shows up much more clearly on the UK view than the U.S. view I noticed, http://www.amazon.co.uk/b/ref=amb_link_21604265_150?ie=UTF8&node=564334&pf_rd_m=A3P5 ROKL5A1OLE&pf_rd_s=center-4&pf_rd_r=0PSA0F42563YHQ2JDZ4A&pf_rd_t=101&pf_rd_p=204406667&pf_rd_i=1025612)

Atmospheric Chemistry
Bioengineering and Biotechnology
Catalysis
Colloids and Surfaces
Electrochemistry and Magnetochemistry
Elements
Equilibrium
Food Chemistry
Geo-Chemistry
Iorganic Substances
Isotopic and Nuclear Techniques
Kinetics
Nomenclature
Materials
Measurement and Instrumentation
Medicinal Chemistry
Microscopy
Molecular Structure
Organic Molecules
Pesticides and Agrochemicals
Photochemistry
Polymer Chemistry
Processes
Purification and Separation
Quantum Chemistry and Molecular Mechanics
Reactions and Reagents
Sensors
Solubility
Spectroscopy
Stereochemistry
Thermodynamics
Toxicology
Trace Analytics
Water Chemistry

INTERNATIONAL UNION OF
PURE AND APPLIED CHEMISTRY

Initial Comments from Jeremy Frey

While I quite understand the wish to limit the number of categories per article to one or at least just a few I would question if this is really necessary. This is certainly the way I would see a conventional database working, but I would like to see a system much more like that we get when buying from say Amazon.

As I understand it, any item there has lots of metadata attached to it in a "flat" manner and the organisation that a viewer sees if driven as much by they way they select the data as by any preconceived view of the "metadata tree." This then allows the "others who viewed this document also looked at…" Perhaps this is not practical but is would be good. See for example to very interesting discussion in the book http://www.amazon.co.uk/Everything-Is-Miscellaneous-Digital-Disorder/dp/0805088113.

The suggested tree structure is all about the way IUPAC is structured rather than how chemistry is used in the modern day of interdisciplinary work.

For example suppose I wanted to look up recommendations on how to make a measurement using STM (something I happen to know is there) I would initially have no idea where that would fit into the suggested tree structure. After a while I would realise that these are classified under Analytical chemistry as that seems to be where all physical measurements have ended up. I guess I would have resorted to a general text search, but that being so, why would we even bother to set up these categories? Just make sure we can text search all of PAC?

I would argue that having an upper level category of Measurement would be a good idea – but this is on par with Analytical not necessarily containing analytical, which is why I suggest a much more flat structure

****

Sub disciplines of Analytical Chemistry – based on techniques (not analytes). One level of subheading is respected. Brynn Hibbert (President Division V)

- ➢ General methods of analysis
    - o Titrimetry
    - o Gravimetry
- ➢ Chemometrics
- ➢ Metrology
    - o Calibration
    - o Standards
    - o Quality Assurance/Quality Control
- ➢ Sampling
- ➢ Spectrometry/Spectroscopy
    - o Atomic spectrometry
    - o Molecular spectrometry
- ➢ Mass spectrometry
- ➢ Separation methods
    - o Gas chromatography
    - o Liquid chromatography
    - o Electrophoresis
    - o Other principles of separation
- ➢ Electrochemical methods of analysis
- ➢ Radiochemical method of analysis
- ➢ Surface analysis
- ➢ Thermal methods of analysis
- ➢ Analysis using biological systems

****