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

IUPAC Database Proposal
Teleconference
10:30am - 11:50am EST
November 26, 2014

Draft Minutes

****

IUPAC Attendees: Lynn Soby, Executive Director, Bonnie Lawlor, Chair, Committee on Publications and Cheminformatics Data Standards (CPCDS)

De Gruyter Attendees: Katharina Butsch, Konrad Kieling, Karin Sora

The meeting opened at 10:30am. The objective was to begin to address the questions provided by Katharina Butsch after IUPAC gave approval in principle to proceed with the database.

<u>Current Status</u>

Katharina reported that they have already been given approval to go ahead with the database and are in the process of initiating the activities that need to be done right away. This rapid movement ahead has resulted in a change in the milestone calendar that was circulated earlier; e.g. the section structure was not needed until June 2015, but is now needed asap so that the freelancer who will be doing the conversion from PDF to HTML can create the metadata that will be required to fill-in the structure. Katharina noted that the most important information that they need asap is as follows:

- Section structure (tree structure)
- Database functionalities
- Search criteria
- Database name

They also need to know who the IUPAC contacts will be for planning, content, etc. and whether or not IUPAC wants to name an Editor-in-Chief and Advisory Board for the database. Katharina also asked if IUPAC wants to create the metadata themselves or will give De Gruyter permission to hire a freelancer for this task.

<u>Discussion</u>

Bonnie Lawlor asked for clarification with regard to the metadata to which Katharina referred. Bonnie said that her understanding of the basic database (the one for which IUPAC gas given approval) is that metadata will be created by the freelancer who will be doing the conversion from PDF to HTML and that this will be done at no cost to IUPAC. The freelancer will only create metadata for a portion of the articles (those that are glossaries or encyclopedic in format). Bonnie noted that there is an option on the table to create additional metadata for the full-text articles that the freelancer will not be able to easily "cut," and asked if this is the metadata to which Katharina's question refers. The answer was "yes." Bonnie then mentioned that IUPAC has been discussing this option, but cannot make a decision unless the cost of creating the metadata is known. Katharina noted that they cannot quote a price until they know what the freelancer's fee will be.

Karin Sora said that she has an approved budget of 120,000 Euros for all processes related to the creation of the database (marketing is not included in this sum). She believes that this amount is more than sufficient as she was able to create a much larger database at Springer for one-third of that amount. She has a former colleague who worked on the database created from *Ullman's Encyclopedia of Industrial Chemistry* who is quite knowledgeable and experienced and who would be perfect for this project if she is available. She does have a science background (chemistry or chemical engineering) and it would be her responsibility to draft the instructions for De Gruyter's e-publishing group. This group is currently working to develop the "master table" hence the need for the tree structure and subject categories. This is especially important for any new subject(s) that will be added for searching (e.g. subject headings new to *Pure and Applied Chemistry* (*PAC*)) so that all content for each document can be added at the same time and multiple passes of the content are not required.

Katharina noted that there are keywords being added to PAC that were not used in the past and asked if De Gruyter has IUPAC's permission to add keywords to the older articles. Also, they would like to add new search categories such as "article type" that would enhance the database.

Lynn Soby noted that De Gruyter is much further ahead than we anticipated and that they have had the opportunity to look at the content and think about potential value-add enhancements. Therefore, with regard to this specific question she believes that it is in the interest of both parties that IUPAC give De Grutyer permission to add search terms while they are in the process of building the database (e.g. adding article/publication types, definitions, etc.). She also said that it would be of value if De Gruyter would provide a list of suggested additions which IUPAC can react, add, etc.

Katharina asked if IUPAC would want to have input to identifying the search criteria. Bonnie Lawlor asked exactly how De Gruyter was defining the term "search criteria." The response was that the search criteria are all of the categories required to identify an article and facilitate access and retrieval by the user: page numbers, authors, article descriptors such as "bibliography," IUPAC project numbers, funding information, etc. Bonnie said that she would like to get input from CPCDS. Katharina agreed to send a list of suggestions and asked that IUPAC respond if at all possible within a week. Bonnie said that she wants to have a teleconference with CPCDS and she has already sent out a survey to see when the maximum number of members can participate. She hopes to have a meeting no later than December 8th and will facilitate advanced discussion via email.

Katharina also asked if at the same time IUPAC could approve the section/tree structure that she presented in October or provide additional suggestions. Bonnie said that the tree structure concept was well-received, but that there had been no further discussion within IUPAC since the due date was originally so far out (June 2015). She will re-circulate what Katharina provided for input and CPCDS will certainly discuss it during the teleconference. Katharina said that she is willing to participate in the discussion as well.

Bonnie then asked how De Gruyter is defining the term "search functionalities." Katharina said that they mean any technical things that are required of the database from the user perspective; e.g. interlinkages, CrossRef, etc. (not technical things like user authentication, updating, etc.). Konrad Keiling said search functionalities should include anything related to the user search experience that needs to be identified before the database building actually begins. Bonnie requested that Katharina send the list of suggestions that De Gruyter presented and she will circulate to CPCDS for comment.

Katharina asked if De Gruyter has permission to add keywords to articles that do not have them. Lynn said that if the addition will improve search value, she has no problem and Bonnie concurred.

Lynn then noted that this project is moving forward much faster than anticipated and believes that there is a vacuum in the absence of an agreed-upon project plan, milestone calendar, and a mutually-acceptable business plan. Katharina said that she can send a copy of their standard contract that can serve as a foundation for discussion and the development of an agreement. The business plan will require ongoing discussion.

Lynn said that there are certain things that are very important to IUPAC such as a very clear understanding by both parties with regard to any assumptions related to content ownership and copyright protection; what level of money and effort will be required from IUPAC as the database evolves; how IUPAC will interface with any freelancers that are hired to work on the project (the freelancers will be contracted by De Gruyter, not IUPAC), etc.

Karin said that they plan to release the database in either January or April of 2016 (the latter is most likely) and therefore the work is budgeted for 2015. They budgeted an estimated ball park figure of 120,000 Euros to create the database of which 30,000 Euros has been allocated for various efforts related to metadata creation. They have already begun looking for a freelancer and testing applicants due to time constraints as such freelancers are hard to find and are not always readily-available due to their prior commitments. Only people with STM experience are being considered. The cost of the Content Management System (CMS) is included in the 120,000 Euros for budget purposes, but it will not be used until needed. Once De Gruyter finds a freelancer and has a handle on that person's fee, they will be able to give IUPAC an estimate on the cost of creating metadata for the full-text articles.

Lynn asked if De Gruyter has made any pricing assumptions or revenue projections and the answer was "no." However, they will use two models: a sales model under which an organization can buy the content that is available at launch (with no requirement to take updates); and a subscription (renting) model that most likely will be based upon FTE's.

Bonnie asked if they will be providing usage statistics on the database. Karin said that they do provide them for journals, but she knows that the platforms for journals and databases differ and she does not know if usage statistics are provided for databases. Katharina said that there would be some level of usage statistics gathered when the CMS system is actually used. Konrad said that database hits versus usage at an article level differ as well. Karin said that she will try to get an answer to Bonnie's question.

Lynn asked if they submitted a 2015 project plan when they sought approval for the database. She would like to have a better understanding of the workflow and timeline. Karin said that the upfront editorial work needs to be completed by April 2015 and then the database will move into testing. Katharina will provide an updated workflow and timeline.

Lynn asked if there will be a beta version available in August and the answer was that this would be highly unlikely. However, De Grutyer will see if it is at all possible to have a marketing piece or database mock-up for distribution at the IUPAC meeting next August. Bonnie noted that we will probably need to have such a piece so that libraries can budget for the database in 2015.

Lynn asked if De Gruyter thinks that the database will cannibalize *PAC* sales and the answer was "no," the rationale being that the database is only a subset of *PAC* and it will not have current data.

Katharina asked about IUPAC's long-term vision for the database and Karin noted that she believes that it is important to add additional content such as InChI codes. Bonnie said that the first commitment is to the core database and if that proves to be successful, IUPAC will consider moving to the enriched version. This will be a more sophisticated version of the database that includes content

3

that is not currently in *PAC,* such as InChI codes, links to relevant internal and external content, etc. CPCDS will identify what additional complementary content should be included, but this action item does not have a near-term due date. Lynn noted that IUPAC does not hold the rights to InChI codes and that De Gruyter will need to talk to the InChI Trust and others who own the rights to the codes. IUPAC has a good relationship with the Trust and will help to facilitate those discussions when the time is right.

Bonnie said that while we are working to resolve the issues that are essential to actually creating the database, there are other issues that, as Lynn noted, need to be resolved in parallel. With De Gruyter adding additional content there is the natural question as to who will own that content and how revenue-sharing will be resolved. Karin said that her goal is to draft a clear, uncomplicated contract with the conditions spelled-out in a very straightforward manner. Their standard contract does not clearly spell out ownership rights and copyright and it will need to be adapted to the IUPAC conditions. She has not discussed revenue-sharing with her management so cannot be held to this, but their usual position is to provide a 10% royalty to their partners who choose not to invest in database development. If the partner does make an investment, the royalty increases. Lynn noted that she and Karin will meet next week and that they can begin discussions on the contract and business plan.

Action Items

Before closing, the following action items were noted:

1. IUPAC needs to provide De Gruyter with the following information asap and by the middle of December at the latest:

- The database title
- Search criteria
- Search functionalities
- Section structure (tree structure)
- Contacts for database planning, content, etc.
- Names for the database Editor-in-Chief and/or Advisory Board

2. IUPAC needs to provide De Gruyter with the database aims and scope by June 30, 2015. This is essentially a database description that will be used for promotion on the De Gruyter website and in related marketing collateral.

3. Lynn Soby and Karin Sora will meet in Boston next week to begin discussion of the contract and business plan.

4. De Grutyer needs to provide IUPAC with the following information within a week:

- An updated timeline for the creation of the database
- The standard De Grutyer contract noting the permissions given by IUPAC today as an appendix
- Suggestions for search criteria
- Suggestions for search functionalities

5. Katharina Butsch will provide Lynn Soby with the workflow/timeline that has been developed in cooperation with the De Grutyer e-publishing group.

6. Once known, De Grutyer will provide IUPAC with the cost of creating additional metadata for the full-text articles.

7. Once provided with the costs of creating additional metadata for the full-text articles, IUPAC will make a decision regarding potential investment in the core database.

8. Karin Sora will find out I De Gruyter provides usage statistics for databases and if so, at what level.

9. De Gruyter will consider creating a promotional piece on the database for distribution at the IUPAC conference in August 2015 and to libraries for budgeting purposes.

10. Bonnie Lawlor will schedule a teleconference with CPCDS for early December and will notify Katharina Butsch so that she can join in.

11. Bonnie Lawlor will schedule the next IUPAC-De Gruyter teleconference for the week of December 5th.

There being no further discussion, the meeting was adjourned at 11:50am.

Respectfully submitted,

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

************************************** n