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

Discussion of IUPAC Database Proposal
Teleconference
11:00am - 12:00pm EDST
October 10, 2014

Draft Minutes

****

IUPAC Attendees: Steven Bachrach, Jeremy Frey, Kazuhiro Hayashi, Colin Humphris, Mark Kinan, Bonnie Lawlor, James Liu, David Martinsen, Fabienne Meyers, Miloslav Nič, Lynn Soby

Discussion

Colin Humphris said that he would like guidance from the group on how to proceed. The existing *PAC* contract transferred the *PAC* back files to De Gruyter and they believed that they would control access to those back files as a revenue opportunity. However, the agreement with authors is that the articles will be available via open access after two years. At the Berlin meeting of the IUPAC Committee on Publications and Cheminformatics Data Standards (CPCDS), De Gruyter put forth the database concept as a *quid pro quo* for making the back files open. Both parties are operating outside of the current agreement and if we do move forward on this Colin will open discussion on a new *PAC* contract. Colin then asked a series of questions as follows:

1. Is there any reason why we would not want De Gruyter to create the basic database?

Jeremy Frey asked if allowing De Gruyter to create the database would exclude anyone else from following suit. For example, if De Gruyter creates a good search interface to the content, would that prevent someone (including IUPAC) from developing programmable access to IUPAC standards and recommendations? Others certainly can access the open access back files. Milo said that technically that would be possible. The terms of the database contract might make it a legal issue. James Liu noted that the back files are in PDF format, not XML, so that whoever accesses the back files would have to do a lot of work to create a useable database.
Colin noted that if IUPAC approves the database, the back file will be made open so that De Gruyter is taking a risk. He views that database as an opportunity to move forward because under the current contract De Gruyter owns the back files.

2. Is the group reasonably comfortable with the creation of the core database as long as it is contractually clear that the back file is open?

There was agreement. However, James Liu said that he would want to make sure that IUPAC controls the XML that De Gruyter creates. It was noted that since they are doing it at their sole cost this could be problematic and it is a point for the contract. (Note: It was also mentioned later in the discussion that perhaps IUPAC should consider making an investment in the core database, depending upon the business plan that De Gruyter develops – cost, revenues, return on investment).

3. If approval for the database is given, will this take us in a direction in which we might not want to go (e.g. the Color Books)? Can we isolate this project specifically to the *PAC* back files?

Jeremy said that IUPAC needs to have a choice about the future of the Color Books and we do not want to compromise our ability to take other IUPAC content forward. He questioned De Gruyter's

perception of the value of the *PAC* back file. He also said that if they provide a better system for access that would be great.

Bonnie Lawlor said that De Gruyter is very interested in the Color Books, but they backed off from them as the source of content for the initial database because there are too many issues to be resolved – use of the *PAC* back files is clearer cut. This initial database would be strictly for the *PAC* back files and would be operated under a new contract, not the existing *PAC* agreement. Any future database would also have its own agreement.

The consensus so far is that we are OK with the database in principal, but that the contractual issues will be key to moving forward.

4. What work will be required from IUPAC in the creation of the core database?

Bonnie said that the required effort is in the joint development of the tree structure that will serve as the foundation for the search and browse functions. Jeremy noted that this is something that we should do anyway. Bonnie also said that there is an option for the core database that De Gruyter put forward and this is to tag the full-text articles for which added information cannot be easily identified by the contractor who will convert the PDF into XML. These new tags can be created by IUPAC at its own expense or by a contractor for which IUPAC and De Grutyer will share the expense. The number of articles in question would have to be estimated along with the cost. Milo noted that De Gruyter files are not schema-based so that additional tags can be added after the fact.

Colin expressed concern about the commercial aspect of the arrangement. The original *PAC* agreement was a 50/50 cost sharing and the revenue flowed equally. He questioned how much IUPAC will have to invest to get to that point?

Bonnie said that if IUPAC does not make an upfront investment perhaps the agreement is that De Gruyter recoups its expenses first and then the revenue is split evenly.

Milo noted that we are offering something very important in that by allowing De Grutyer to create the database we are offering libraries and researchers an "official IUPAC-stamped database," unlike any database that someone else may choose to create from the open access files as noted earlier in the discussion. The IUPAC stamp has value.

5. Is this a good project from which to learn more about De Gruyter's abilities in database creation and marketing?

The group's consensus was "yes."

6. If having the Content Management System (CMS) will ultimately make the database better, why would we not want to give approval?

Milo said that he is all for the CMS, but noted that having it will make it easier for De Gruyter to create their own Color Books. It was also noted that the CMS is required if IUPAC wants to move onto the enriched database.

7. Why would IUPAC not move forward with the database?

Mark Kinnan said that it is very important to keep in mind that a database is an asset. Will IUPAC own it? De Gruyter? Will there be joint ownership and if so how will this play out if for whatever

reason IUPAC does not want to renew the agreement? Ownership and re-use of content must be clearly spelled-out in the contract.

Milo noted that if IUPAC wants to create a database this proposal is a route to follow. In reality, IUPAC does not have the resources to create the necessary XML on its own.

Lynn noted that the value proposition may not solely reside with De Gruyter. There may be other entities that see value in IUPAC content and we need to keep that in mind.

Jeremy asked if IUPAC should reach out to other scientific unions such as ICSU to get their views and also questioned if IUPAC should try to get sponsorship so that the standards remain free. It was generally agreed that getting sponsorship can be difficult. Jeremy is concerned that De Gruyter will assume that they own any content that they create. IUPAC can take back the content that it originally gave to De Gruyter, but what De Gruyter adds is theirs.

Colin said that there is a difference between them owning the content and being given the opportunity to charge for it, so again the ownership issue will be a contractual one. He also said that IUPAC could decide to share in the costs from the beginning if it appears to be the sensible thing to do. Jeremy expressed the opinion that IUPAC should consider doing so in order to maintain a level of control over a key asset. Lynn said that we may have "defensive" reasons for such an investment, but questioned whether or not IUPAC will get a return on it.

In general, it was agreed that IUPAC has a business relationship with De Grutyer and is prepared to move forward on the core database depending upon the terms of the contract and the business plan that De Gruyter presents. Also, it is key to keep this agreement solely related to the *PAC* back file. It will be an opportunity to see if De Gruyter is a good partner databases.

8. Is there content and/or functionalities that IUPAC might want to add in the future for the creation of the enriched database?

Jeremy said that he believes that IUPAC should move towards the enriched database. The addition of standard identifies and links will be a major asset and he would like to see things made as automatic as possible.

Colin asked if CPCDS would take this on and Bonnie said "yes," but would want to throw a broader net across IUPAC for input.

Colin then asked meeting attendees for their final thoughts.

Dave Martinsen:	Overall it sounds good and he likes the direction that is being taken.
Mark Kinnan:	It will all come down to the contract (he also mention the possibility of licensing the data to De Gruyter).
Lynn Soby:	Agrees with Mark – the key is ownership.
James Liu:	Agrees to move ahead – still would like to get ownership of the XML.
Kazuhiro Hayashi:	Agrees, but the contract needs to be explicit about the re-use of content.
Jeremy Frey:	Agrees with the direction and that ownership is critical. He also expressed the fact that open access must be considered. Any work that he does under UK grants must be made open access. Also, many do free work for IUPAC and consider the results of their efforts to be open access for the scientific community. We need to make the status of future contributions clear. This is an opportunity to clarify all future open access issues.

Fabienne Meyers;   Nothing to add.

Bonnie Lawlor: This is a good opportunity for IUPAC and has no hesitation about moving on to the next step.

Summary of Approved Action:

IUPAC will agree "in principle" to the creation of the core database and the licensing of the CMS system. The latter will only be used when IUPAC chooses to write entries directly for the database or revise entries within the database or when IUPAC and De Gruyter decide to have a freelancer work directly within the database. The database will be limited solely to the PAC back files and will be governed by a totally new contract. Final agreement will be given only if the terms of the business plan and contract are acceptable to IUPAC.

Bonnie Lawlor will schedule a teleconference with De Grutyer, Colin Humphris, Lynn Soby and herself to begin discussions.

Actions for future consideration:

IUPAC will consider a possible investment in the core database once the business plan is presented. When/If a contract for the core database is signed, IUPAC will consider the possibility of moving towards the enriched version.

Colin Humphris thanked everyone for their time and input. There being no further discussion, the meeting was adjourned at 12:00pm.

Respectfully submitted,

Bonnie Lawlor
Chair CPCDS
October 14, 2014

****

Appendix 1

Points raised in discussions to date on which more information is needed from De Grutyer and/or issues for contractual consideration:

- The future re-use of any data that De Grutyer creates for the core product and any additional tags that IUPAC might choose to create or have created (Kazuhiro Hayashi)
- The internal process that De Grutyer will use to move forward (Lynn Soby)
- The workflow process that would be established if IUPAC chooses to move forward, not only the production process, but also the IUPAC-De Gruyter work relationship processes (Lynn Soby)
- Referrals – what databases have been created for others by De Grutyer and who can we learn from at those organizations (Lynn Soby)
- Database ownership (XML and additional tags)
- Open access of the back file and the open access of future contributions to the database and/or *PAC*
- Consider having a period of performance for the whole proposal and performance standards within the contract, (Lynn Soby)
- Renegotiate the current *PAC* agreement (Colin Humphris)