**InChI Trust Newsletter**

**August 2014**

**Summary:**

Since the February 2014 report there continues to be good progress with InChI and the InChI Trust in a number of areas. A new project proposal for large/biological molecules has been approved by IUPAC. In addition another new working group for positional isomers has been formed. Publicity for the project remains very good and has resulted in considerable increase in the uptake and usage of the InChI algorithm. The InChI Symposium at the 2014 ACS Fall meeting took place as two half-day sessions. The US Federal Drug Administration has chosen to join the Trust. Lastly, the InChI team was awarded the Chemical Structure Association (CSA) Mike Lynch Award.

**Items covered in this report:**

Membership/Support

InChI RFP/Contracts

IUPAC InChI subcommittee and working parties/groups

Meetings attended & Talks/ Posters given

Manuscripts & Other Publications

InChI Trust Web Site InChI Usage Technical Issues

Plans for 2014

**Membership/Support:**

Summary: A number of organizations are still in the process of joining or talking about joining, but there still is little progress. Like most organizations, since InChI works and it is not high on their immediate priority lists, actual real progress is slow.

As of July 18, 2014

Existing Members and Associates: 14 - FDA has joined. Supporters: 46

Certification Suite purchases: 2 - NCI and Novartis (in process)

**InChI RFP/Contracts**

The contract for Markush structures with Digital Chemistry remains on hold awaiting funding.

A contract for taking the RInChI work that Jonathan Goodman and Chad Allen did at Cambridge

University to public release is expected to be signed by the fall of 2014.

A contract with Jeremy Frey to develop a proof of concept for developing an app for InChI QR

codes is underway. Progress was reported at the ACS San Francisco InChI Symposium

A contract for an InChI Trust web site in the cloud is underway and expected by fall 2014.

**IUPAC InChI subcommittee working groups**

**Positional Isomers**

Considerable interest in positional isomers has developed in the past few months. Chris Steinbeck at EBI has agreed to chair the working group for this effort. This is background for this working group that Chris has drafted:

There is a need for a canonical representation to encode positional isomers of (bio-)organic molecules. Use cases include the concise representation of positional isomers that were either systematically synthesized or isolated from a biological system.

Depending on the amount of evidence available, the identification of (bio-)organic molecules can also be ambiguous, in which case one will want to represent a number of possibilities for true identity of the

partially identified compound.

This working group will produce a specification for the extension of the IUPAC InChI to allow for the canonical encoding of positional isomers. To this end, it will gather user requirements from a number of subspaces of the applicable chemistry community and ensure that the specifications are in line with the strategic value to IUPAC.

The white paper with the specifications will be discussed with experts from the InChI Trust and prototypes will be produced, tested and refined in an agile development process.

Chris has put together the following experts from 5 countries as members for his working group: Christoph Steinbeck, Chair

Egon Willighagen

John May

Steffen Neumann

Steve.Stein

Roger Sayle

Evan Bolton

Oliver Fiehn

**Resolver** – The work is now being done under Markus Sitzmann, with assistance from Evan Bolton at

NIH/NLM/NCBI/PubChem.

**Polymers** – This work was finished by the working group under Andrey Yerin. Igor Pletnev has started programming this standard. The work will be completed and tested in 2014.

**Reactions** – Guenter Grethe has been unable to move forward with his working group owing to the lack of a contract to clean up and do further testing of the current software. With the expectation that a contract will be signed by the fall of 2014 this project should again be moving forward.

009-043-2-800 Standard InChI-based Representation of Chemical Reactions [http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx\_wfqbe\_pi1[](http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx_wfqbe_pi1)project\_nr]=2009-

043-2-800

Chairman: Grethe, Gunther

Members: Batchelor, Colin Goodman, Jonathan Kraut, Hans Lawson, Alexander Schmidt, Martin

**Markush** - We are still waiting for the needed financial support before proceeding.

**Electronic States** - Don Burgess at NIST has developed two draft plans (available upon request) for InChI for Representations of Species at the Molecular Level. Don gave a talk on this at the InChI Symposium in San Francisco.

**InChI for Materials** – No news from the NIST staff about this.

**Organometallics**- Colin Batchelor and his working group expect a final report by the fall 0f 2014.

**Inorganics** - Hinnerk Rey from Elsevier/Frankfurt has replaced Nigel Wheatley to head up this working group. His working group proposal to IUPAC for funding was approved in 2013 but there has not been any progress report yet.

2012-046-2-800: Handling of Inorganic compounds for InChI V2 [http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx\_wfqbe\_pi1[](http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx_wfqbe_pi1)project\_nr]=2012-

046-2-800

Chairman: Rey, Hinnerk

Members: Damhus, Ture Druckenbrodt, Christian Hartshorn, Richard Schenk, Roger Sitzmann, Markus

**Large molecules, biopolymers/ Proteins/ biological polymers/ macromolecules/ biomolecules** etc. - Keith Taylor, who has just retired from Accelrys, had his proposal to IUPAC approved in October 2013. The work will be based on the work he has done at Accelrys and has made publicly available. The Pistoia HELM project, while lacking a number of technical capabilities (e.g., it does not support variably attached drugs or variable groups and it has limited canonicalization), is moving forward. There are many issues with the project, which seem to be ignored by those running the project. There is no one responsible for any bug fixes. Major changes will be done by sending out a RFP. Those

at Pfizer who developed HELM have been let go or reassigned, but seem to be involved without management knowledge. How this can function, let alone become a real standard is beyond me. They

have expressed no interest in working with IUPAC or the InChI Trust. Keith, Evan Bolton (NIH), and I

remain on the HELM mailing list to monitor what is going on.

2013-010-1-800: Implementation of InChI for chemically modified large biomolecules [http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx\_wfqbe\_pi1[](http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx_wfqbe_pi1)project\_nr]=2013-

010-1-800

Chairman: Taylor, Keith

Members: Blanke, Gerd Bolton, Evan Chalon, Didier Drijver, Alex Jensen, Jan Yerin, Andrey Helen Berman

The first meeting of this working group will be held at NIH on October 27 & 28, 2014. The meeting is be sponsored and hosted by NLM/NCBI, but they are providing no financial support – only the meeting room space. There will be no registration fee for anyone who wishes to attend. This will be a two day meeting with short presentations the first morning by a superb set of speakers with knowledge and experience in the area:

Philip Bourne, NIH Associate Director for Data Science

Helen Berman, IUBMB & PDB, Rutgers

Keith Taylor, Working Group Chair

Evan Bolton, PubChem

FDA (Larry Callahan or Yulia Borodina) NIH/NCATS (Tyler Peryea)

EBI (Sameer Verlanker)

Roger Sayle (nextmovesoftware) HELM (Sergio Rotstein) BIOVIA (Matt Sage)

**Tautomers**. Under the leadership of Marc Nicklaus, NIH/NCI, InChI project #2012-023-2-800, "Redesign of Handling of Tautomerism for InChI V2” was approved for funding by IUPAC. Marc held his first working group meeting at the ACS New Orleans meeting.

2012-023-2-800: Redesign of Handling of Tautomerism for InChI V2 [http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx\_wfqbe\_pi1[](http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx_wfqbe_pi1)project\_nr]=2012-

023-2-800

Chairman: Nicklaus, Marc

Members: Bolton, Evan

Ihlenfeldt, Wolf-Dietrich

Peryea, Tyler

Pletnev, Igor Rey, Hinnerk Sitzmann, Markus Tchekhovskoi, Dmitrii

**Interlocking structures** (rotaxanes) - Andrey Yerin will consider starting a project/working group

(soon).

**Extended Stereochemistry** - Evan Bolton still thinking about what to do in the area of stereogenic centers such as cumulenes.

**January 2014 – August 2014 activities**

**Meetings Attended; Talks/Posters Presented**

I gave an InChI talk at the BioIT InChI workshop in April. About 10 people paid to attend. Other speakers included Tony Williams, Steve Boyer, Chris Southan, and Evan Bolton

I gave an InChI presentation at the APHL *(*Association of Public Health Laboratories) in April. About 40 people attended.

I gave two InChI seminars in Singapore at the NUS and NTU universities on May 15 and 16. About a dozen people attended each talk.

I gave three presentations at the July meeting in Toronto, which included a formal talk and a workshop presentation on how to create InChIs. I was invited to these sessions by Bob Belford, who is trying very hard to get more university and other teaching staff to incorporate InChI in their teachings.

On behalf of the entire InChI team I gave the CSA Mike Lynch Trust award talk at the 10th ICCS meeting in Noordwijkerhout in June. The five InChI team members (below in alphabetical order) was given the Mike Lynch award for the work on developing InChI.:

Steve Heller Alan McNaught Igor Pletnev Steve Stein

Dmitrii Tchekhovskoi

**Manuscripts**

No additional manuscripts were published by the InChI team. Igor Pletnev has produced a first draft of a full paper on InChI.

**Other Publications**

As a result of the InChI poster (co-authored by the folks at LSU) given at SLA 2013 in San Diego in

June, I was asked by the journal editor, Tony Stankus, to prepare a paper for publication in the “Science

& Technology Libraries” journal. To be sure this paper is sufficiently and appropriately different from those already written by myself, Alan McNaught and others, I have asked Bill Armstrong at LSU to take

the lead in writing this paper. So far Bill has still not gotten very far with preparing the manuscript in

spite of a number of reminders.

**InChI Trust web site**

The Trust web site is now up on the IUPAC server. There have been a number of problems with the web site and plans are underway to move to the cloud. Aletia Rey has been hired to maintain and add content to the web site.

**InChI Usage**

For lack of a better a better term, I use InChI Usage to refer to publications and blogs about InChI. Alan and I have been passing these on to Aletia and she has added these to the web site. There have been quite a number of publications using InChI. The numbers continue to grow. Searches on Google (and other search engines) continue to have more hits for InChI strings and InChIKey strings.

In February 2014 the ACS *J. Med. Chem*. came out with an editorial encouraging authors to submit SMILES as structure representation for molecules in their manuscripts.

<http://pubs.acs.org/doi/abs/10.1021/jm5002056>

In a series of conversations - public on the CINF list and in private, we were unable to get the

Journal to alter its current plan and add InChIs and InChIKeys. The main concern seems to be that converting an InChI string back to a structure does not give a pretty as picture and accurate picture as the

author might wish. Since the drawing is actually a function of the structure drawing programs, not

SMILES or InChI, this continues to puzzle me. Specifically they said (in an email): “Our major concern when adding a line notation representation of chemical structures to JMC articles was to faithfully reproduce the structures drawn in the data tables in the paper” The fact that SMILES are not unique (as pointed out in the case of Warfarin that Marc Nicklaus found on the web) does not seem to have bothered these people. I referred to this approach as “art” vs. “science”. Also they believe, against all evidence,

that authors will do what they wish and there is no need for the publisher to get involved. At the same time they admit the publisher has shown no interest in helping. Lastly, since I was told any SMILES entered in the manuscript will be supplemental information, I was told it will not be searchable. While we were not able to change anything at this time, virtually every response to the CINF list was in support of InChI and not SMILES. So, at least, we came out of this looking very good.

**Technical Issues**

The mechanism to discuss and resolve technical issues continues to work well. Most issues seem to be able to be resolved by email and phone calls, but face-to-face meetings are still very critical as there are some very strongly held opinions that do not get resolved by emails. My regular meetings with NIH (PubChem , NCI, and FDA) staff have been very useful.

**Plans for 2014**

For the remainder of 2014 my overall plans and goals are as follows:

1. Work to expand the current membership with two basic classes of members – Full and Associate, and add to the number of Supporters. Work to sign up more organizations for the Certification Suite.

2. Continue to attend meetings and give talks on InChI where useful and appropriate.

3. Help to organize the InChI Large Molecules meeting at NIH/NLM/NCBI in October 2014.

Steve Heller. Chair

Alan MeNaught Secretary

**lnChiTRUST**