

InChI Trust Newsletter

August 2013

Summary:

There continues to be good progress with InChI and the InChI Trust in a number of areas. In the non-technical area of support, new memberships continue to be very slow. The various IUPAC InChI working groups are continuing to move ahead at a good pace. A new project proposal for large/biological molecules has been submitted to IUPAC. Publicity for the project remains very good and has resulted in considerable increase in the uptake and usage of the InChI algorithm. The 4 InChI videos were finished in March 2013 and had a public screening at a number of exhibitor booths at the spring ACS meeting. The manuscripts from the J. Cheminformatics InChI thematic issue are now online. Work on transferring the current InChI Trust web site to a new server has been started.

Items covered in this newsletter:

Membership/Support
IUPAC InChI subcommittee and working parties/groups
Meetings attended & Talks/ Posters given
Manuscripts & Other Publications
InChI Usage
Technical Issues
Plans for 2013

Membership/Support:

.As of August 1, 2013

Existing Members and Associates: 15 (a loss of 2 members – ZINC and FIZ Chemie; a gain of 1 – Perkin-Elmer)

Supporters: 41 (ZINC & Scientific Thinking LLC)

Certification Suite purchases: 3 (Novartis, NCI, & FDA)

IUPAC InChI subcommittee working parties



Resolver - work is now being done at NCI under Markus Sitzmann, with assistance from Evan Bolton at NIH/NLM/NCBI/PubChem. They have released a draft of the resolver protocol.

Polymers – This work was finished by the working group under Andrey Yerin. Igor Pletnev has started programming this standard. The work should be completed and tested in 2013.

Reactions - ChemInform declined to provide resources to take on the responsibility of making the code and web site developed by Jonathan Goodman and Chad Allen at Cambridge University operational and maintaining the system. Hence as noted above an RFP has gone out to undertake this effort.

Electronic States - Don Burgess at NIST has developed a draft plan, as there is now interest at NIST. NIST has recently established an Office of Data and Informatics. With this draft plan Don should be able to gather together the right people to start a working group for extending InChI for molecular entities in different electronic states

InChI for Materials – The original initial interest at NIST in undertaking this extension has vanished with a recent reorganization. I recently came across a publication by a different group at NIST interested in developing and defining a unique material identifier:

[Materials Genome III: Building CALPHAD File and Data Repositories](#)

Scripta Materialia, Available online 18 June 2013, Pages

C.E. Campbell, U.R. Kattner, Z.-K. Liu

Their initial thoughts, as noted (below) in their paper indicate it would be worthwhile to try to work with them:

“Within a given materials community materials may have their own identifiers; however, in a broad common data space a single unique identifier should be used. One method to uniquely identify a material is to follow a format similar to the International Chemical Identifier (InChI) used to identify chemicals [16], which is a set of predicable series of ASCII characters. This type of format uses a layer structure to define specific levels of details. A unique materials identifier should include layers for alloy name, composition, processing/form, and the heat-treatment. If the information for any given layer is not available, that layer would be left blank.”

Organometallics- A very difficult area owing to, amongst other things, variable attachments, but it is expected to have something by the fall of 2013 from Colin Batchelor and his working group.

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 early 2013 and is IUPAC Project 2012-0462-800.

Biopolymers/Proteins/biological polymers/macromolecules/biomolecules etc. - Keith Taylor has submitted a project proposal to IUPAC for travel support for the group he has put together to undertake this work. The project has not yet been approved, but Keith and his working group members have started



to do what they can while they wait for funding. 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, with their latest project activity is designing a logo for their project. 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.

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.

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.

QR Codes – Jeremy Frey is proceeding with a proof-of-concept plan for developing InChI QR codes.

January 2013 – June 2013 activities

Meetings Attended; Talks/Posters Presented

I attended the week long NIH/FDA/USP chemical structure standards meeting, Rockville, MD in February 2013. I presented a talk on InChI which received a warm welcome.

I attended the spring ACS meeting, April 2013. I met with LSU, InfoChem, IUPAC, Chemistry Central, and others to discuss various aspects of the project.

Attended the BioIT meeting in Boston, April 2013. I presented a poster on InChI which attracted some attention. Chris Southan and I met with Cindy Crowninshield, the Conference Director and proposed an InChI workshop for the 2014 BioIT meeting. I have prepared a draft for the workshop, including suggested speakers. A decision on this workshop is expected this summer.

I attended the SLA June 2013 meeting in San Diego and presented a poster on InChI training materials for researchers and students, in collaboration with the LSU staff (Bill Armstrong and Karen Salazar). I have provided Judith Currano (U of Penn) with InChI information for her "Chemistry for the Non-Chemist" workshop class at SLA. The poster was well attended with a paper likely to come from this (see Other Publications section of this report). I also met with Nicole Boris, the Managing Editor of Powder Diffraction from the International Centre for Diffraction Data (ICDD). Nicole has expressed an interest in adding InChIs to their database – the Powder Diffraction File (PDF) which contains 760,019 unique material data sets. (<http://www.icdd.com/>)



I gave a lecture on InChI at Syngenta in the UK.

Manuscripts

The Journal of Cheminformatics issue on InChI that Tony Williams has organized includes the following two manuscripts based on the ACS San Diego InChI symposium. A number of other manuscripts have also been published, with more to come. See:

(<http://www.jcheminf.com/series/InChI>)

- 1. InChIKey collision resistance: an experimental testing** - Igor Pletnev, Andrey Erin, Alan McNaught, Kirill Blinov, Dmitrii Tchekhovskoi, Steve Heller *Journal of Cheminformatics* 2012, **4**:39 (20 December 2012)

Abstract: <http://www.jcheminf.com/content/4/1/39/abstract>

PDF version - <http://www.jcheminf.com/content/pdf/1758-2946-4-39.pdf>

Manuscript ID: 2052653967827646

Submitted: 19 October 2012

Published: 20 December 2012

DOI: 10.1186/1758-2946-4-3923256896

As of 6/1/13 - 1750 accesses

- 2. InChI - The Worldwide Chemical Structure Identifier Standard** - Stephen Heller, Alan McNaught, Stephen Stein, Dmitrii Tchekhovskoi, and Igor Pletnev, *Journal of Cheminformatics* 2013, **5**:7 doi:10.1186/1758-2946-5-7

<http://www.jcheminf.com/content/5/1/7>

Manuscript ID: 1564938902827637

Submitted: 19 October 2012

Published: 24 January 2013

DOI: 10.1186/1758-2946-5-723343401

As of 6/1/13 - 2161 accesses



Other Publications

Andrey Yerin has prepared an InChI paper for Chemistry International. I expect this paper will be sufficient for the update needed to be presented to IUPAC Division VIII and CPEP at the General Assembly in Istanbul in August 2013.

A paper on RInChI has been prepared by Guenter Grethe and will shortly be submitted to J. Cheminformatics. I expect it to be published before the end of 2013.

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.

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 Ulrich and he 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.

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.

Work continues slowly by Igor on the full documentation paper on InChI. This should, finally, be completed by the end of 2013 and will be submitted to J. Cheminformatics.

Plans for 2013

For 2013 my overall plans and goals for the remainder of 2013 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. . I will give an invited paper at the ACS meeting in Indianapolis in September. I will give a poster at the November Goslar meeting (which is moving to Fulda).
3. FDA and NIH are starting to work on a joint project to create a chemical identifier for chemicals as part of the NIH NCATS program. I am working to be part of their planning activities to see how much InChI can solve their needs.
4. Work with Chris Southan to offer an InChI workshop at the April 2014 BioIT meeting. An initial proposal has been submitted to the conference chair – Cindy Crowninshield at CHI (Cambridge Heath Institute).
5. Chair, oversee, and coordinate the IUPAC InChI subcommittee activities and working groups. Some face-to-face meetings are currently being considered for 2013.

Steve Heller

