InChI Trust Newsletter

March 2015

Summary:

Since the fall 2014 report there continues to be good progress with InChI and the InChI Trust in a number of areas. The latest InChI project, for large/biological molecules, had a very successful two day meeting at NIH in October 2014. In addition another new working group for positional isomers has been formed and an initial project proposal is being prepared to submit to IUPAC for approval and funding. 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 San Francisco ACS meeting was well attended and a report on the symposium was published in the CINF newsletter. A second US Government agency, NIH, has chosen to join the Trust. Lastly, the NIST members of the InChI team were awarded the Department of Commerce Gold medal as well as an internal NIST award.

Items covered in this report:

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. Bio-Rad has indicated they plan to join in 2015 and the necessary paperwork for them has been sent.

As of March 2015

Existing Members and Associates: 15 - NIH has joined.

Supporters: 48

InChI RFP/Contracts

The contract for Markush structures with Digital Chemistry remains on hold waiting potential funding. Related to this there are plans for a meeting with EBI/John Overington to look at possible EU support for this.

A contract for taking the RInChI work that Jonathan Goodman and Chad Allen did at Cambridge University was signed in August 2014 with Dr. Gerd Blanke (Germany).

A contract with Jeremy Frey to develop a proof of concept for an app for InChI QR codes is underway.

The InChI Trust web site was moved from the IUPAC web site and is now in the cloud.

InChI development work

Igor Pletnev continues to do a superb job as the InChI programmer. What follows is a summary of his recent work:

The New API procedure was added and extensively tested, MakeINCHIFromMolfileText (...)

It generates InChI directly from MOLfile or SDFile record (represented as a string).

It fully mirrors the parser of inchi-1 executable and so ensures that any correct caller of InChI library will behave exactly as inchi-1 executable (which was not true previously). This, for the first time, allows an end-user to avoid writing his/her own implementation of Molfile reader -- and to avoid possible incompatibilities with inchi-1 computed InChI strings.

Actually, there is one more new piece of software - demonstration program, mol2inchi, which uses inchi dll (libinchi.dll) and this new API call.

MOLfile V3000 support is added -

in the sense that V3000 molfiles (a) can be read and non-V3000-specific info (same as in V2000 but in changed format!) is treated correctly, and

(b) additional V3000-specific features like stereo collections and haptic bonds are understood and stored.

However, note that
(a) the testing is still far from extensive... I simply could not find enough V3000 molfiles;
(b) the new V3000 features are understood but not used -- as there are no finalized decisions/ideas/directives from InChI Subcommittee on how to treat them.

Code refactoring -

that is, rewriting parts of code to make the more structured/readable/modifiable. This work was performed throughout all the year, "in background".

It is not visible for most end users, but I consider it important for the future maintenance and development.

Support of simple polymers.

This has been delayed and is now expected to be finished early in 2015

For 2015:

Integration of Digital Chemistry (John Barnard with co-workers) API into InChI Software.

Investigation continues

Preparing the new (1.05) release -- with all API updates, added V3000 and polymers support.

This work is expected to be finished by mid-2015.

IUPAC InChI subcommittee & working groups

IUPAC Committees

The two IUPAC committees that I met with at the IUPAC International Chemical Education Conference in Toronto in July 2014 - CCE and COCI (Bernard West) have expressed interest in InChI and working together. To date nothing has materialized. At the CCE meeting was the incoming ACS/CHED chairperson from Purdue (Marcy Towns) who expressed interest in pursuing possible joint efforts to add InChI to the teaching content of chemistry courses, but nothing has yet has materialized from this..

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. Chris is now preparing a proposal for this working group to be submitted to IUPAC in early 2015.

As noted in the last report, 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 . Markus continues to work on this.

Polymers – This work was finished by the working group under Andrey Yerin. Igor has started programming this standard. The work will be completed and tested in 2015.

Reactions –With the August 2014 signing of a contract with Gerd Blanke this project is again 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[project_nr]=2009-043-2-800

Chairman: Grethe, Gunther

Members:

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

Markush – No change

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 spoke on this at the InChI symposium in San Francisco this past August.

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

Organometallics- Colin Batchelor and his working group expect a final report in the near future.

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[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 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. Major changes will be done by sending out a RFP. 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[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 was held at NIH on October 27 & 28, 2014. The meeting was sponsored and hosted by NLM/NCBI, but they provided no financial support – only the meeting room space. I should point out that Phi Ngo at NLM/NCBI was particularly helpful in organizing the logistics for the meeting. There was no registration fee for anyone who attended. This was 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, Rutgers
Keith Taylor, Working Group Chair
Evan Bolton, PubChem
Yulia Borodina, FDA
Peryea Tyler, NIH/NCATS
Valery Tkachenko, RSC
Sameer Verlanker, EBI
Roger Sayle, Nextmovesoftware
Sergio Rotstein, HELM
Matt Sage, BIOVIA

35 people registered for the meeting and most all attended, except f2-3 who had travel or related issues.

After the meeting Keith Taylor provided access to the collection of presentations from the InChI for Large Molecules Symposium. They will be uploaded shortly.

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[project_nr]=2012-023-2-800

Chairman: Marc Nicklaus

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.

September 2014 – December 2014 activities

Meetings Attended; Talks/Posters Presented

One person (Ghislain Deslongchamps, Chairman, Chemistry Department, University of New Brunswick) came up to me after this talk to discuss the possibility of helping to create InChI teaching materials. I am hoping we can follow up on this and get something done here.
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) were given the Mike Lynch award for the work on developing InChI.

Steve Heller
Alan McNaught
Igor Pletnev
Steve Stein
Dmitrii Tchekhovskoi

Awards

The InChI team received two awards from NIST. One was a Gold Medal from the Department of Commerce. Here is the announcement of the award sent to Steve & Dmitrii:

===
Dmitri, Steve:

I'd like to congratulate you on winning the Department of Commerce Gold Medal this year for your development of InChI. As you know, InChI has had a world-wide impact on the dissemination of chemical information and I am gratified to see that the DoC is recognizing your achievement by giving

you its highest honor. You will be notified by MML shortly with more details and I hope that you will be able to accept your award at the DoC ceremony in Washington, DC.

Best regards
Mike

Michael J. Tarlov, Ph.D.
Chief, Biomolecular Measurement Division

====

The second award was an internal NIST laboratory award:

Dear Stephen and Igor,

Congratulations on being awarded the MML Distinguished Associates Award! This award recognizes associates or previous staff members who have made outstanding contributions to MML goals. This is a prestigious award with only six award winners recognized yearly.
You were nominated for developing InChI, an algorithm that turns 3D chemical structure into a unique string of characters, enabling chemical data search and exchange.

Again, congratulations on this award, and a sincere thank you for your work for MML.

Laurie

Laurie E. Locascio, Ph.D.
Director, Material Measurement Laboratory

Manuscripts

In late 2014 Igor delivered the long awaited InChI manuscript and has been sent to the J. Cheminformatics for publication.

InChI Trust web site

The Trust web site has left the IUPAC server is now up on the InChI Trust cloud server. Aletia Ray who was hired to maintain and add content to the web site is doing a nice job.

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.

InChI Trust Videos - Access numbers:

InChI & the Islands – 629 views (12/14); 526 views (7/14)

The Googlable InChIKey – 751 views (12/14); 597 views (7/14)

The Birth of the InChI - 835 views (12/14); 687 views (7/14)

What on earth is InChI? - 2486 views (12/14); 1977 views (7/14)

Two other YouTube InChI videos are available:

Mcule video:

2012 San Diego ACS presentation: Registration system of mcule: InChI is the key
122 Views (12/14 – uploaded 2012)

Audiopedia video
International Chemical Identifier
5 views (12/14 – uploaded 11/14)
(1 minute video):
http://youtu.be/MG49gn_CdUE

Published on/Uploaded to YouTube Nov 30, 2014

The IUPAC International Chemical Identifier (InChI /ˈɪntʃiː/ IN-chee or /ˈɪŋkiː/ ING-kee) is a textual identifier for chemical substances, designed to provide a standard and human-readable way to encode molecular information and to facilitate the search for such information in databases and on the web. Initially developed by IUPAC and NIST during 2000–2005, the format and algorithms are non-proprietary. The continuing development of the standard has been supported since 2010 by the not-for-profit InChI Trust, of which IUPAC is a member. The current version is 1.04 and was released in September 2011.
Prior to 1.04, the software was freely available under the open source LGPL license, but it now uses a custom license called IUPAC-InChI Trust License.

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.

Plans for 2015

For 2015 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.

.

Steve Heller