InChI Trust Project Director's Report

July 2017

Summary:

Since the February 2017 report there continues to be good progress with InChI and the InChI Trust. The initial version 1.00 of RInChI was released in March 2017. The March EBI workshop on InChI was a great success and the upcoming August NIH InChI meeting is looking to be as good and useful. More organizations, databases, and publications continue to use the InChI algorithm, such as MDPI. Lastly, Guenter Grethe, who has led the RInChI project from the start has turned over the working group leadership to Gerd Blanke in March 2017. We are all very much indebted to Guenter for his considerable efforts over the years in getting RInChI working and proving a useful tool to chemists.

Items covered in this report:

Membership/Support
InChI RFP/Contracts
InChI development work
IUPAC InChI subcommittee and working groups
Meetings attended & Talks/ Posters given
Manuscripts
InChI Trust Web Site
InChI Usage
Technical Issues
Plans for 2017

Membership/Support:

There were no new memberships since the last report.

As of July 27, 2017

Existing Members and Associates: 16 (only 15 are listed on the web page)

Supporters: 47

InChI RFP/Contracts

As has been the case for a long time, the contract for Markush structures remains on hold waiting for funding. There was no particular interest in this or support for funding at the EBI meeting.

The contract with Dr. Gerd Blanke (Germany) for RInChI continues and has produced excellent results.

InChI Development Work

Igor Pletnev continues to do a superb and a very responsive job as the InChI programmer. With the release of version 1.05 there continues, as expected, to be useful feedback on minor issues and bugs.

IUPAC InChI subcommittee & working groups

InChI working Groups

Chemical mixture composition

Leah McEwen at Cornell University has initiated a working group for chemical mixture composition. As noted previously, recent highly damaging events in chemical laboratories and classrooms have led to increasing focus on chemical information management in laboratory organizations. The diverse teaching and research environment in the academic sector particularly is raising awareness of the complexity of chemical safety information resources and formats available. A key concern in this regard is that documentation of chemicals with current identifiers is a persistent challenge for tracking and managing chemicals across the chemical enterprise, from process planning to manufacture to waste disposal and emergency response.

The objective of this project is to establish requirements and guidelines for the generation of a unique identifier for all forms of a chemical (liquid, gas, solid, powder, etc.). Currently, many chemical identifiers exist, but very few reflect these bulk properties of substances, which may commonly exist in many forms and mixtures. Furthermore, most existing identifiers present cross-referencing challenges between systems designed around different initial applications and editorial principles.

The intended outcome of this project is global adoption of the InChI notation in chemical inventories and information systems across commercial, industrial, government, academic and educational sectors to facilitate accurate documentation, handling and exchange of chemical information in support of safer management and use of chemicals.

This project is complementary to another user-focused project that is developing a QR code version of the InChI to facilitate labeling and other communication of chemical safety information. That project will be consulting with global stakeholders to determine deployment and use approaches. This project will focus the specificity and usefulness of the information being encoded in the InChI.

This working group is probably unique for the InChI project in that it is of clear scientific value, but even of more importance and value to all the chemistry labs around the world. Safety is something that makes the front page of newspapers and TV news programs.

This project, entitled “InChI Extension for Mixture Composition” was funded by IUPAC in June 2016.

Positional Isomers

Considerable technical interest in positional isomers has developed in the past but no-one is willing to take the lead for this area. One person did offer to take the lead in June 2017, but was unable to get approval from their organization to spend the necessary time so the project remains leaderless.

The current members of this working group in limbo are:

Christoph Steinbeck
Egon Willighagen
John May
Steffen Neumann
Steve Stein
Roger Sayle
Evan Bolton
Oliver Fiehn

Resolver – No further progress report has been submitted since my last report. However I do expect an update prior to the August Trust Board meeting.

Polymers – With release of version 1.05 a limited area of polymer chemistry can now be handled by the InChI algorithm. A number of issues were found after release of 1.05 and Igor is working on these matters.

Reactions –Under the programming direction of Gerd Blanke this project has moved ahead very nicely. The RInChI 1.00 release was finalized in March 2017. This release has been tested against the US patent database 2008-2011 with 420.000 reactions (provided by NextMove and with thanks to Roger Sayle). The worst reaction they found had 95 educts and 22 products but the RInChI was calculated without any issues.

Guenter Grethe, who has chaired this working group from its inception, has had to turn over the lead to Gerd Blanke

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: Gerd Blanke

Members:
Gunther Grethe
Colin Batchelor
Jonathan Goodman
Hans Kraut
Martin Schmidt
Keith Taylor

Markush – With no interest from the US and other patent offices, this project remains on indefinite hold, but the possibility of starting work on it (mentioned above) could occur is there is sufficient interest and need and funding.

Electronic States –. There still are no further developments here.

Organometallics- Colin Batchelor and his working group will release a final report in 2017. They are having discussions with the Inorganic and RInChI working groups to be sure there are no problems in overlapping areas.

Inorganics - A decision on how to proceed with this awaits the outcome of the Organometallics work

Large molecules, biopolymers/Proteins/biological polymers/macromolecules/biomolecules etc. –

Little had happened since the October 2014 working group meeting at NIH as Keith Taylor was waiting for the extensions of InChI past 1024 atoms. Igor has now accomplished this in version 1.05 and Keith has tested it to his satisfaction. With this now accomplished progress will follow discussions at the EBI InChI workshop in March 2017 and the InChI workshop at NIH in August 2017. A representative from the HELM project may join the working group.

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
Berman, Helen

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 plans to hold a working group meeting on this at the August 2017 InChI meeting at NIH.

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) - There has yet to be any effort to look into how to handle these structures. This topic area will dropped in future reports until there is movement.

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

QR Codes

The InChI QR code consultation workshop IUPAC project was approved in June 2015. Richard Hartshorn is leading this project. This is the announcement for this project:

"The InChI Trust (http://www.inchi-trust.org/) is examining development of a QR code (2D bar code) version of the InChI. We wish to consult with industry/regulatory/academic sector users to identify and prioritize additional information that could/should be included in the QR code to enhance the value and commercial utility of the QR InChI. Possibilities to be evaluated and elaborated upon include: health/safety information (hazard code and/or safety data URL); catalog code; batch number; inventory information; sample composition/purity. This project is complementary to another user-focused project that is developing InChI for states and mixtures."

Education/Academic/Training

I have started working with Bob Belford (University of Arkansas at Little Rook) to lead an effort for educational and training materials, such as an InChI primer, to help educate and make easier use of InChI. The initial effort will begin at the August InChI meeting at NIH. Bob is looking into assembling a

working group of academics to go forward with this idea. They would include Vincent Scalfani, Henry Rzepa (UK), Jordi Cuadros (Spain), Bob Hanson, and Martin Walker.

January 2017 – July 2016 activities

Meetings Attended; Talks/Posters Presented

A number of conference call meetings with David Evans, Richard Kidd, and Alan McNaught were held over the past six months to deal with issues that needed to be addressed between Board meetings.

I met on a regular basis with members of NIH/NCBI, particularly Evan Bolton, to discuss InChI issues. In particular I worked to arrange for the facilities and support for the August 16-18, 2017 InChI meeting at NIH.

I attended the March 2017 InChI meeting at EBI in Hinxton UK.

I attended the March 2017 InChI Trust Board meeting in Cambridge, UK

I attended the Spring ACS meeting in San Francisco and had a number of productive conversations and meetings.

I chaired the InChI session at the May 2017 BioIT meeting.

I visited and gave a lecture at the New York offices of Schrodinger. I was told they have started to work on incorporating InChI into their Live Design product. (https://www.schrodinger.com/livedesign)

An invited InChI lecture I prepared was given by Ray Boucher at the Sao Paulo meeting in July 2017,

An InChI poster I prepared was presented at this IUPAC meeting in Sao Paulo by Ray Boucher.

Manuscripts

Ray Boucher, Stephen Heller, and Alan McNaught, The Status of the IUPAC InChI Chemical Structure Standard, *Chemistry International*, July - September 2017, page 48.

InChI Usage

I met with MDPI (Molecular Diversity Preservation Institute – www.mpdi.org) staff at the ACS meeting in San Francisco and suggested they look into using InChIs in their publications. Much to my pleasant surprise after talks in May 2017 with Martyn Rittman (Publishing Services Manager) they have added InChIs to their Molbank (http://www.mdpi.com/journal/molbank) publication.

Leah McEwan pointed out to me that WIPO is using InChIs in the PatentScope database and search system. (http://www.wipo.int/patentscope/en/)

There are no doubt more organizations, databases, and publications using InChIs but I have yet to discover an easy way find any of them

Technical Issues

The mechanism to discuss and resolve technical issues continues to work well, as evidenced by the activities from the community during the testing and release of version 1.05.

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 2017

For the remainder of 2017 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. Attend ACS meeting in Washington DC. Give an invited InChI lecture. Also I will meet with groups to discuss adoption and usage of InChI.
4. Attend the August 19, 2017 InChI Trust Board meeting.
5. Attend the annual GDCh meeting in Mainz, Germany and present either a poster or a short lecture.

Steve Heller