

# InChI Trust Project Director's Report

February 19, 2017

## Summary:

Since the August 2016 report there continues to be good progress with InChI and the InChI Trust in a number of areas. The final version 1.05 of the InChI algorithm was released in January. Work on RInChI continues to move ahead smoothly. With the upcoming March EBI workshop on InChI and the August NIH InChI meeting movement with getting numerous working groups to be more active seems likely. More organizations, databases, and publications continue to use the InChI algorithm. Finally, ACS/Chemical Abstracts has joined the Trust.

## Items covered in this report:

- Membership/Support
- InChI RFP/Contracts
- InChI development work
- IUPAC InChI subcommittee and working parties/groups
- Meetings attended & Talks/ Posters given
- Manuscripts
- InChI Trust Web Site
- InChI Usage
- Technical Issues
- Plans for 2016/2017

## Membership/Support:

### Summary

ACS/Chemical Abstracts Service has joined the Trust for 2017, and Springer Nature have renewed their membership following the merger. Discussions with EPA to join are proceeding. PLOS has expressed interest in joining but feels their amount of chemistry publications does not yet justify membership. CDISC and GBSI have not yet responded to the follow up on their joining. The same is true for Prous Research, and others.

As of February 1, 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. I plan to have a short meeting on the priority of Markush at the March InChI workshop at EBI.

The contract for taking forward the RInChI work that Jonathan Goodman and Chad Allen did at Cambridge University with Dr. Gerd Blanke (Germany) is progressing well.

Testing of the code is now being undertaken by the RInChI working group and release should occur soon.

## **InChI development work**

Igor Pletnev continues to do a superb and a very responsive job as the InChI programmer. The work to program and respond to the many odds & ends bugs for version 1.05 took much more time than initially expected, but clearly the resulting product, while not perfect, is much better having had all the feedback and help from the community. The good news with this extensive and time consuming effort to get v 1.05 as correct as possible is that there was considerable feedback from the community in it being tested. And the bad news is that all this testing resulted in many things that needed to be fixed and hence the delay in the release.

## **IUPAC InChI subcommittee & working groups**

### **InChI work Groups**

#### **Chemical mixture composition**

Leah McEwen at Cornell University has initiated a working group for chemical mixture composition.

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 was willing to take the lead for this area. However Chris Steinbeck has just joined the University of Jena and has now agreed to lead the effort.

The current members of this working group 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 March Trust Board meeting.

**Polymers** – With release of version 1.05 a limited area of polymer chemistry can now be handled by the InChI algorithm.

**Reactions** – Under the programming direction of Gerd Blanke this project has moved ahead very nicely. The RInChI 1.0 release has been tested against the US patent database 2008-2011 with 420.000 reactions (provided by Nextmove) and they have not found anything that should stop the release.. The worst reaction they found had 95 educts and 22 products but the RInChI was calculated without any issues. They expect to be able to send the final release in February 2017.

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](http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx_wfqbe_pi1[project_nr]=2009-043-2-800)

Chairman: Gunther Grethe

Members:  
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 if there is sufficient interest and need and funding.

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

**InChI for Materials** – There is still no news from the NIST staff about this. This topic will be dropped in future reports.

**Organometallics**- Colin Batchelor and his working group expect a final report in 2017. They are having discussions with the Inorganic working group as there is considerable overlap.

**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 has 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 and Keith has tested it to his satisfaction. With this now accomplished progress will follow starting with the EBI InChI workshop in March 2017.

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](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" is 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](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 be 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 prioritise 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.”

## July 2016 – December 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.

I attended the fall ACS meeting in Philadelphia and had a number of productive conversations and meetings.

I was able to arrange for another InChI session at the May 2017 BioIT meeting.

## Manuscripts

No new manuscripts were published in the second half in 2016.

## 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 – 960 (1/17); 883 (7/16); 804 (1/16); 728 (7/15); 629 views (12/14); 526 views (7/14)

The Googable InChIKey – 1,379 (1/17); 1,203 (7/16); 1,037 (1/16) ; 915 views (7/15); 751 views (12/14); 597 views (7/14)

The Birth of the InChI - 1,365 (1/17); 1,233 (7/16); 1,084 (1/16); 984 views (7/15); 835 views (12/14); 687 views (7/14)

What on earth is InChI? - 4,188 (1/17); 3,762 (7/16) ; 3,331 (1/16); 2,956 (7/15); 2486 views (12/14); 1977 views (7/14)

IUPAC InChI (Google lecture)– 950 (1/17); 946 (7/16); 931 (1/16); 922 views (7/15)  
[https://www.youtube.com/watch?v=mH9fuspg\\_h0](https://www.youtube.com/watch?v=mH9fuspg_h0)

Representing Chemical Structures on computer – 675 (7/16); 546 (1/16); 390 views (7/15)  
<https://www.youtube.com/watch?v=uzXkJ9BsyHQ>  
(InChI part starts at about 14 ½ minutes into the video)

Scott Wiedemann

Cheminformatics, Encodings SMILES & InChI – 879 (1/17); 647 (7/16); 468 (1/16); 354 views (7/15)  
<https://www.youtube.com/watch?v=V9HHnRAS5BA>



## 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 of version 1.05 prior to the 1.05 release.

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 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 San Francisco. Meet with groups to discuss adoption and usage of InChI.
4. Attend the May 22-24, 2017 BioIT meeting and chair the InChI session at BioIT.
5. Attend the March 20-21 2017 EBI InChI industry workshop and give a lecture.
6. Attend the March 23, 2017 InChI Trust Board meeting.
7. Deliver an InChI manuscript for an upcoming issue of CI.
8. Prepare an invited talk for the ACS Washington DC meeting in August
9. Prepare a talk for Ray Boucher to deliver at the IUPAC General assembly in Sao Paulo.

Steve Heller

