

InChI Trust Project Director's Report

August 2016

Summary:

Since the January 2016 report there continues to be good progress with InChI and the InChI Trust in a number of areas. Version 1.05 of the InChI algorithm was released for review in July/August 2016. Work on RInChI continues to move ahead. Related to the matter of the working groups, the main issue that, as usual, needs to be improved on is having the working groups be more active in moving towards their goals and getting more organizations, databases, and publications to use the InChI algorithm

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:

Two new organizations (University of California and Bio-Rad) have joined the Trust as Associate in 2016. Bio-Rad has been admitted in return for developing multithreading capability for the InChI software. IBM and PE have not paid their dues and have been removed. Discussions with EPA to join are proceeding.

As of August 1, 2016

Existing Members and Associates: 16

Supporters: 47

InChI RFP/Contracts

As has been the case for a long time, the contract for Markush structures remains on hold awaiting potential funding.



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.

InChI development work

Igor Pletnev continues to do a superb and a very responsive job as the InChI programmer.

IUPAC InChI subcommittee & working groups

IUPAC Committees

Chemical mixture composition

Leah McEwen at Cornell University has initiated a working group for chemical mixture composition. Recent highly damaging events in chemical laboratories and classrooms [Sheharbano (Sheri) Sangji, a 23-year-old chemistry research assistant, died from injuries sustained in a chemical fire on December 29, 2008, in a laboratory at UCLA] 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 few months but at the same time Chris Steinbeck is still looking for a new person to lead this working group.

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 – 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. He has put together a beta test version with some infrastructure and some test content .

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 is moving ahead nicely. There is an issue with how Google and other search engines index RInChIs.

Working group members are being asked to test the program with reactions from in-house databases or from RD files supplied on the web test site. After the successful conclusion of these tests and the incorporation of modifications, a beta test will be carried out by members of the SourceForge group. Thereafter, the first release of RInChI is projected to be in 2016.

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.

Electronic States – Don Burgess at NIST has developed plans for using InChI for Representations of Species at the Molecular Level. In 2014/2015 he published the 3 papers on this subject about InChI-ER (Elementary Reactions). The last two came out in the June 2015 issue of IJCK. Being manuscripts from a US Government employee PDF copies are freely available from Don. There still are no further developments here.

InChI for Materials – There is still no news from the NIST staff about this.

Organometallics- Colin Batchelor and his working group expect a final report in 2016. 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. –

Nothing has happened since the October 2014 working group meeting at NIH as Keith Taylor was waiting for the extensions of InChI past 1024 atoms. With this now accomplished it is hoped progress will follow.

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 Philadelphia ACS meeting in August 2016.

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) - 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

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.”

January 2016 – June 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.

While in London for the Trust Board meeting I met with the staff of J. Cheminformatics regarding InChI publications. The InChI publications have very much helped their impact factor. I also met with Henry



Rzepa. My EBI meeting was canceled at the last minute, but I continue to have productive interactions with EBI staff.

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

In April I attended the BioIT meeting in Boston and presented a talk on InChI as part of a session on InChI for large molecules. Others speaking there were Keith Taylor, Evan Bolton, Larry Callahan (FDA), and Tyler Peryea (NIH).

In May I visited the offices of John Wiley in Hoboken, New Jersey and gave a lecture on InChI.

We hope to have another similar session at the 2017 BioIT meeting.

Manuscripts

There was an announcement about the InChI project in the June 2016 issue of the CDISC newsletter.

<http://content.yudu.com/web/2htg1/0A2hthm/2016Q2/flash/resources/index.htm>, Pages 4-5.

I would like to express my thanks to Dominic Clark at EBI for helping to arrange this note in the CDSIC newsletter.

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

The Googable InChIKey – 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 - 1,233 (7/16); 1,084 (1/16); 984 views (7/15); 835 views (12/14); 687 views (7/14)

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



IUPAC InChI – 946 (7/16); 931 (1/16); 922 views (7/15)
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 – 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. 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 the second half of 2016 and 2017

For the second half of 2016 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 Philadelphia. Give an invited talk at the Skolnik Symposium, and meet with groups to discuss adoption and usage of InChI.
4. Attend the November 6-8, 2016 Fulda meeting and give either a poster talk or a lecture.
5. Attend the September 28-29, 2016 CDISC International Exchange meeting in Rockville Maryland (<https://www.cdisc.org/interchange>) compliments of CDISC which will cover the \$990 registration fee.

2017

The InChI Standard has celebrated its tenth anniversary in 2015, and, building on the past and ongoing work by its working groups, a three-day meeting will be held next year on 16-18 August 2017 (Wednesday-Friday) at the main National Institutes of Health (NIH) campus in Bethesda, MD (short Metro/Subway ride from downtown Washington DC) The meeting will bring together the current InChI community and working groups that define the current state of the InChI project, together with other interested stakeholders. The aim is to discuss what is needed for the chemical, biomedical, materials, and related academic and industry communities for proper and useful structure standard representation of both small and large molecules, and the future direction and activities of InChI development will be a major goal of the meeting.

Steve Heller

