



**IUPAC Division (VIII) of Chemical Nomenclature and Structure Representation
International Chemical Identifier (InChI) Subcommittee**

Minutes of the meeting on 30th July 2009 at the Crowne Plaza Hotel, Glasgow, UK

Present: *Subcommittee members:*
Steve Heller (Chairman)
Evan Bolton (US National Center for Biotechnology Information)
Sandy Lawson (Elsevier, Frankfurt)
Alan McNaught (Secretary; InChI project coordinator, Cambridge, UK)
Igor Pletnev (ex-officio developer) (Moscow State University)
Tony Williams (ChemSpider)
Andrey Yerin (Advanced Chemistry Development, Moscow)

Observers:
John Barnard (Digital Chemistry Ltd)
David Black (IUPAC Secretary General)
Colin Bulpitt (Taylor & Francis)
Kirill Degtyarenko (IUPAC Division VIII)
René Deplanque (FIZ-Chemie, Berlin)
Richard Kidd (Royal Society of Chemistry)
Rachel Kirton (Taylor & Francis)
Dave Martinsen (American Chemical Society)
Tim Miller (Thomson Reuters)
Hinnerk Rey (Elsevier, Frankfurt)
Ulrich Roessler (FIZ-Chemie, Berlin)
Dick Wife (SORD B.V)

Apologies: *Subcommittee members:*
Colin Batchelor (Royal Society of Chemistry)
Marc Nicklaus (US National Cancer Institute)
Steve Stein (NIST)
Chris Steinbeck (European Bioinformatics Institute)
Keith Taylor (Symyx Technologies, CA)
Dmitrii Tchekhovskoi (ex-officio developer) (NIST)
Graeme Whitley (Wiley, New York)
Jason Wilde (Nature, London)

1.0 Minutes of the previous meeting

The minutes of the meeting in Salt Lake City on March 23rd 2008 were approved without comment.

2.0 InChI Trust status

Copies of the current Business Plan and the InChI Trust logo were circulated.



2.1 Background and current position

Steve Heller gave a brief overview of the status of the InChI project. The initial collaboration between IUPAC and NIST had had been remarkably successful; however NIST had taken the development of the Identifier to the point at which its needs were largely satisfied, and did not wish to provide resources for further development. There had been a need to find other sources of funding, for maintenance and for continuation of the project to deal with areas of chemistry not currently covered. Also it was apparent that the community of InChI users and potential users wished to see InChI maintenance and development placed on a firm business basis as a stable operation; it had become clear that this would be difficult to achieve under the auspices of IUPAC alone. The InChI Trust had been set up to address these concerns. Authority for the InChI standard would continue to be provided by IUPAC through the Division VIII InChI Subcommittee, and the Trust would be responsible for implementation of community requirements as approved by the Subcommittee.

Chemistry publishers in the UK (Nature and the Royal Society of Chemistry) had been the first to offer funding support for the Trust. RSC had provided administrative facilities to enable the Trust to be incorporated in the UK, and an application had been made for charitable status. Offices and computer facilities for the Trust had been provided free of charge by FIZ-Chemie Berlin Other organisations (Taylor & Francis, Symyx Technologies, OpenEye) had subsequently joined the Trust, and very recently final agreement to participate had been received from Thomson Reuters and Elsevier. Steve Heller would continue his publicity and awareness-raising activities in order to attract additional membership. It was intended to continue the development of InChI under the auspices of the Trust as a IUPAC-approved open-source algorithm, to be made freely available to the community for use as they see fit.

The Business Plan contained budget estimates for a five-year period; it was expected that the need for continuing development would decline after a few years and that funding requirements would eventually reduce to maintenance level. It was intended that all work would be carried out under contract: there would be no employees. Steve Heller would act as part-time Director, Igor Pletnev would continue as developer, and other developers would be added as required. It would be particularly important to spread knowledge of the InChI algorithm amongst several people, to ensure stability and continuity. It was noted that the logo currently displayed in the Business Plan was not that finally approved by the Trust Board.

Requirements for development would be established by the IUPAC InChI Subcommittee, working through appropriate subgroups; at present there were groups on requirements for organometallic structures, InChI/InChIKey resolver protocol, business rules for structure input control, and the use of InChI in description of chemical reactions.

The first meeting of the InChI Trust Board would take place on September 11th in Cambridge UK, and invitations to attend would be issued soon.

Steve Heller tabled his InChI report to IUPAC Division VIII for inspection.

2.2 Relationship with the American Chemical Society

Steve Heller had made contact with ACS on several occasions in connection with InChI, most



recently in March of this year, and had recently received a message from Publications Division in response to the press release announcing launch of the InChI Trust. Dave Martinsen outlined some of the concerns expressed:

2.2.1 *Continuing role of IUPAC as standards body*

There was some doubt as to whether the continuing role of IUPAC as standards-setting body for InChI was assured. For example, diagrammatic representations of relationships between various bodies could be taken to imply that IUPAC was being side-lined. Steve Heller emphasised the crucial role of the Division VIII InChI Subcommittee in this connection. All developments would need the approval of this body for incorporation in the InChI standard. It was therefore of prime importance for IUPAC to continue funding the Subcommittee at an appropriate level. Alan McNaught pointed out that the standard would lose all credibility if not authorised by IUPAC. It was agreed that the Business Plan and future documents on InChI development should give more emphasis to IUPAC's role.

2.2.2 *IUPAC participation in the InChI Trust*

It appeared that IUPAC was not at present participating directly in the Trust. Dr McNaught thought that an invitation for IUPAC to join as funding participant had been sent to IUPAC, but it appeared that this was not so; an invitation would be issued for consideration by the IUPAC Executive in October. Furthermore ACS considered that IUPAC should have an official role on the InChI Trust Board regardless of any paid-up Trust membership; this idea was supported and would be recommended to the Board meeting on September 11th. It was noted that the Trust's Articles of Incorporation were drawn quite widely, as is normal for such documents, but the primary objective was clearly specified as development and maintenance of the InChI standard.

David Black emphasised that IUPAC was delighted that the Trust had been established, though it would have been preferable for the set-up to have been carried out in a more independent way, through the IUPAC office. He noted that support would still be available for InChI-related projects through the IUPAC Project System.

2.2.3 *Future ACS involvement*

Steve Heller had written back to ACS suggesting a meeting in Washington. It was hoped that this correspondence and the present discussion would result in a fruitful continuing dialogue regarding possible participation in InChI developments.

2.3 **InChI Trust office facilities**

Ulrich Roessler and René Deplanque reported that the InChI Trust office had been set up at FIZ-Chemie at no cost to the Trust and the website inchi-trust.org had been registered. Ulrich Roessler would be managing the Trust facilities. e-Mail addresses (xxxx@inchi-trust.org) had been established for Board members and it was intended to set up wiki systems for ongoing discussions of working groups. Initial work on the website would concentrate on collecting available InChI information.

Igor Pletnev noted that a previous committee meeting had requested a development server,



including facilities for testing InChI implementations; he would discuss precise requirements with Ulrich Roessler and René Deplanque.

3.0 Developer's report

Igor Pletnev reported that, subject to agreement over various questions (specified in his circulated report), the next version of the InChI software would probably be available by mid-October. It would include one bug fix implemented since the release of version 1.02. It was agreed that the new version, which would include both standard and non-standard InChI generation in a single library, would be designated version 1.03. Preparation of a full technical paper on InChI for publication was in progress, and it was hoped to have a draft available by September/October. It was proving difficult to describe the algorithm completely, but it was hoped that a medium-high-level description would be sufficient, in conjunction with the source code. Some reservations were expressed by Kirill Degtyarenko and Dave Martinsen about the lack of a full description, in particular with regard to future portability, and it was agreed that this issue might benefit from further exploration by a contractor. However, John Barnard pointed out the likelihood that even a high-level description would probably not cover all aspects of detail.

3.1 Designation of non-standard InChI

It was agreed that a new (memorising) layer of several characters should be created within the non-standard InChI string to designate the options used in its creation, and that a single letter should be used to designate a non-standard InChIKey. Igor Pletnev would prepare a brief specification for circulation to the subcommittee for comments before implementation.*

3.2 Standard/Non-standard InChI compatibility

It was agreed not to worry about the possibility that a non-standard InChI string may appear identical with a standard string.

3.3 Non-standard InChIKey format

It was agreed that the non-standard InChIKey string should adopt the same format as the standard InChIKey in not containing a check character and using the same proton counting scheme.

3.4 Software release

As noted above, the next software release would be version 1.03 and provide a single package to generate both standard and non-standard forms, specified as 'universal' in the developer's report. However, the already introduced std-InChI API calls should remain, with the code collapsed where possible; Igor Pletnev should feel free to deal with this in whatever way he considered most appropriate.

3.5 Bug fixes and software options

* *Secretary's note:* Following subsequent e-mail discussions amongst members of the subcommittee, it has been decided that options memorisation recorded in non-standard InChI will not be retained in the corresponding non-standard InChIKey.



It was agreed that bug fixes and the drawing inconsistency/new normalisation features should be turned on unconditionally. The software options specified in the developer's report were also agreed (subject to the decisions noted in item 3.1). The possibility of introducing annual software releases should be considered seriously.

3.6 *Validation test-bed.*

It was noted that an InChI/InChIKey validation test bed was required; this would be developed as soon as resource for providing it becomes available.

4.0 **Reports from Working Groups**

4.1 *Resolver*

Tony Williams reported that a Google group had been set up to deal with the Resolver discussions. Discussions with National Cancer Institute, RSC and Nature Publishing Group (Tony Hammond) to define a protocol were ongoing. The need for multiple resolvers according to the needs of individual databases was recognised. A draft architecture was currently under discussion, and it was hoped that an example of a functioning resolver protocol would be available by the end of 2009. Sandy Lawson would provide an Elsevier contact for this work, and Chris Steinbeck would be asked whether EBI wished to participate.

Steve Heller noted that not all organisations with databases will want to set up their own resolvers; we should therefore provide information to users about organisations with resolver facilities that will accept database deposition. The need for a resolver on the Trust website should be considered.

4.2 *InChI for Organometallics*

The working group proposed that (1) the possibility of carrying out normalisation before metal disconnection be considered; (2) the molfile format be extended to encode more coordination and organometallic structures losslessly. Sandy Lawson agreed to provide details of decisions made for Reaxys.

Igor Pletnev pointed out that implementation of normalisation before disconnection would be a huge job, and it was agreed that it would be preferable to attack the problem initially by developing business rules for structure representation. Further consideration would be needed to (a) define practical options, (b) discuss encoding protocols and (c) look for any convenient ways of modifying disconnection procedures. In Colin Batchelor's absence, Andrey Yerin thought that a further report could be produced by October 1st.

4.3 *Business rules for structure input*

Andrey Yerin reported that the working group had paid specific attention to stereochemistry, since this area was the most problematic. The need to enforce correct input, with existing IUPAC recommendations as the basis, was clear. It would be desirable to introduce functionality into the InChI algorithm that would warn users of incorrect input, and it was agreed that rules specific to InChI should be extracted from the IUPAC recommendations with this in mind. A status report



would be provided by October 1st. Consideration would be given to the possibility of a face-to-face meeting, perhaps alongside the San Francisco ACS meeting.

4.4 *InChI for reactions*

Preliminary work was proceeding at Cambridge University under the direction of Jonathan Goodman, with funding from RSC. The work was being carried out by summer students, and would continue up to the end of August. A test website allowing conversion from .rxn to RInChI had been established at <http://www-rinchi.ch.cam.ac.uk/>. A project report would be prepared in September for distribution to those Trust members with reaction databases, for comment.

5.0 **Future needs**

It was envisaged that most of the work of the Subcommittee would be carried out by working groups, and that eventually the Subcommittee would not need to meet more than once a year. The Trust Board would be asked to consider what additional working groups should be established on what timescale. It was noted that stereo-encoding should be added to the list of topics in the Business Plan.

6.0 **Next meeting**

It was suggested that this take place adjacent to one of the 2010 ACS meetings (March, San Francisco or September, Boston), probably the former.

Alan McNaught

13 August 2009