

# IUPAC InChI - Recent Expansion in the Global Chemical Structure Identifier Standard



Stephen R Heller<sup>2,3,4</sup>, Gerd Blanke<sup>2,6</sup>, Ray Boucher<sup>2</sup>, Ian Bruno<sup>1,2,3</sup>, Jonathan M Goodman<sup>3,7</sup>, Richard Kidd<sup>2,5</sup>, and Rudy Potenzone<sup>2</sup>

<sup>1</sup>Cambridge Crystallographic Data Centre; <sup>2</sup>InChI Trust; <sup>3</sup>IUPAC; <sup>4</sup>National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Department of Health and Human Services; <sup>5</sup>Royal Society of Chem; <sup>6</sup>StructurePendium Technologies GmbH; <sup>7</sup>Yusuf Hamied Dept Chem, University of Cambridge

## InChI Project Goals

Create a **free, non-proprietary chemical substance identifier**:

Turn chemical structures into unique, canonical text strings

Enable easier finding and linking of chemical and biological data

- Stakeholder collaboration and partnership
- Educational resources



- New uses and extensions of the InChI standard
- Promote the use of InChI and raise awareness

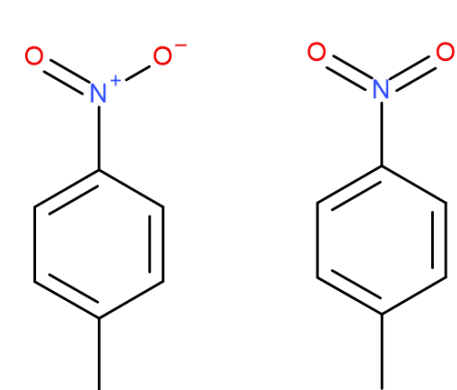
## InChI Supports FAIR Principles

- Findability, accessibility, interoperability, and reusability
- In particular– **findability** and **reusability**
- The first step in (re)using data is to find and link it
- InChIs allow **interoperability** between structures and related (meta)data including qualified references to other (meta)data

InChI is critical to enable references based on chemical identity

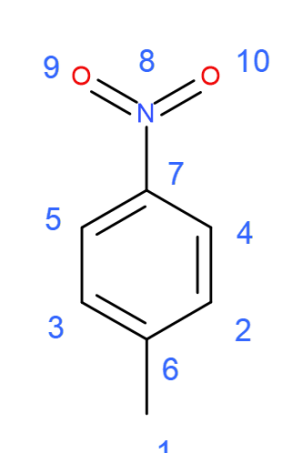
We address this global challenge with international cooperation and support the needs of industry, academia, governments and individuals

## InChI and InChIKey



**InChI:** Built to work as a chemical identifier

Different representations must lead to same InChI



/C7H7NO2 (Formula C: 1-7, N:8, O: 9-10)  
/c1-6-2-4-7(5-3-6)8(9)10 (Connectivity)  
/h2-5H,1H3 (Hydrogen layer)

Stereochemistry also supported

InChI = 1S/C7H7NO2/c1-6-2-4-7(5-3-6)8(9)10/h2-5H,1H3



**InChIKey:** Fixed-length hash of InChI: identity and web searches

InChIKEY= ZPTVNYMIQHSSEA (formula, connectivity, hydrogens (hashed))  
UHFFFAOY (remaining layers: stereochemistry, isotopes, etc. (hashed))  
SA (Standard InChI, version 1 (= A))  
N (Neutral / total charged atoms)

InChIKey = ZPTVNYMIQHSSEA-UHFFFAOYSA-N



## Examples of InChI Adoption

### Databases

- Elsevier/Reaxys –179 million
- EBI UniChem –177 million
- NIH/NCI/CADD/iRL – 140 million
- RSC/ChemSpider –114 million
- NIH/PubChem - 111 million
- Chemical Abstracts Service – 264 million

### Virtual repositories

- GDB17 (The Human Genome Database) – 166 Billion
- NIH/NCI/Argonne/SAVI – 1.75 Billion

## InChI Working Groups and IUPAC Projects

- Extended Stereochemistry
- Extended Tautomers
- Isotopologues
- Large Molecules
- Monomer Atoms
- Organometallics/Inorganics
- Variable Structure (Markush)
- GitHub code development
- Mixtures
- Nanomolecules
- QR Codes
- Open Education Resource (OER)
- Reactions
- Resolver

**InChI release status:** <https://github.com/IUPAC-InChI>

### InChI versions:

#### Released

- Current version 1.06  
[https://github.com/IUPAC-InChI/InChI\\_1\\_06](https://github.com/IUPAC-InChI/InChI_1_06)
- Latest implemented enhancements
- Polymers (in beta status)

#### Under development

- Organometallics
- Enhanced Stereochemistry
- Tautomers
- Isotopologues
- Full V2/V3 molfile support (for registration)
- Sgroups and Sgroup data

#### Longer evaluations

- Large molecules
- Variable molecules / Markush InChIs

### Applied InChI versions:

#### Released

- RInChI 1.0 for Reactions

#### Test Phase

- Resolver  
<https://github.com/inchiresolver/inchiresolver>

- QR code

- RInChI 1.1

Test phase

<https://github.com/IUPAC-InChI/RInChI>

#### Awaiting coding

- MInChI for Mixtures

Prototype released:

<https://github.com/cdd/mixtures>

- RInChI (1.2) for Reactions

Multithreading

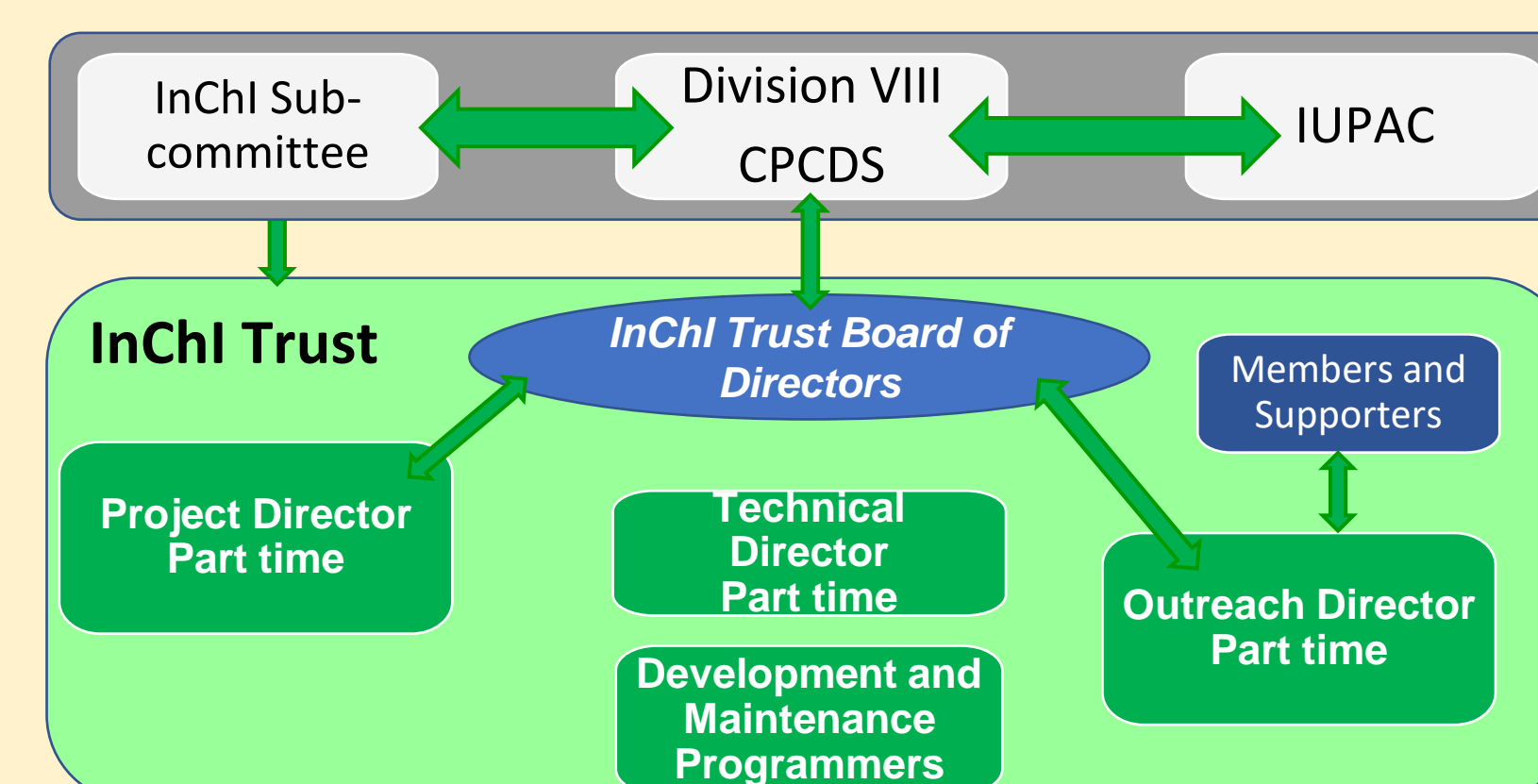
Atom mapping

#### Under development

- NanoInChI (NInChI) for Nanomaterials



## InChI Organization



## InChI Trust Was Created to Oversee InChI

- UK Public Benefit Charity
- Funded by Members who pay an annual fee
- Some development activities supported by funding from the VolkswagenStiftung secured by RWTH Aachen University



## InChI Trust Financials

- Core revenue \$160k from Member fees
- Revenue has been enough for core maintenance, incremental development and outreach
- The current budget is insufficient for all priority enhancements: stereochemistry extensions, organometallics, tautomers, isotopologues, mixtures etc.

## InChI Future

How to support the further development of InChI:

- Contribution in kind based on open-source development on GitHub
- Donations to the InChI Trust
- Partnership with the InChI Trust and/or IUPAC

If you are interested in helping to ensure the future sustainability of InChI then become a member of the InChI Trust

✉ [member-info@inchi-trust.org](mailto:member-info@inchi-trust.org)

🌐 <https://inchi-trust.org/>

