# IUPAC International Chemical Identifier (InChI) – the compound identifier that makes molecule recognition FAIR

Gerd Blanke[1,2], Jonathan M. Goodman[3,4]

## 1. Introduction

The InChI (IUPAC International Chemical Identifier) is the unique identifier for chemical compounds that allows the precise identification of any molecule in each research step. The unique identification of compounds makes InChI the central pivot point for FAIR data handling in the pharmaceutical and chemical R&D. Recent developments of the InChI enhance the handling of inorganics and organometallics, add extended stereochemistry functionality, and will introduce additional normalization rules for tautomeric functional groups. InChI applications extend the usage of InChIs to reactions, mixtures, formulations, and nano materials.

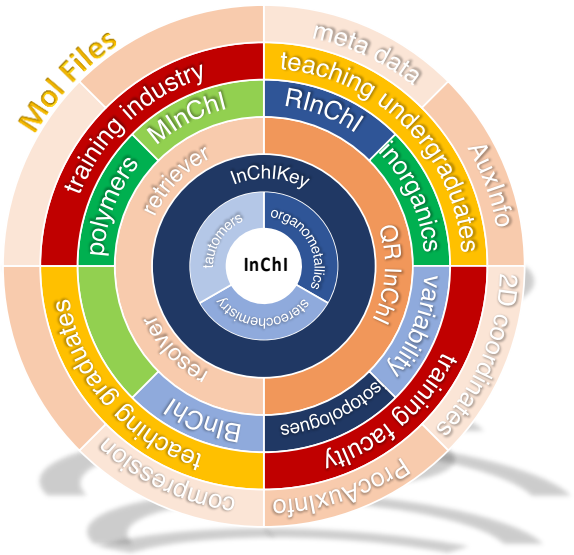


*Figure 1: The InChI universe*

[1] StructurePendium Technologies GmbH, Essen, Germany; Gerd.Blanke@structurependium.com
[2] InChI Trust, Cambridge, UK
[3] Yusuf Hamied Dept Chem, University of Cambridge, UK
[4] IUPAC

## 2. How do I find compound XYZ?

To search, compare, and analyse molecules over multiple sources from different resources you have to ensure the compound identifier follows the FAIR (Findable, Accessible, Interoperable and Re-Usable) principles. Independent of the actual compound depiction, the goal of an identifier for the chemical compound is to provide unique identification, making the compound findable and accessible within different research contexts. The accessibility and utility of public databases like PubChem relies on the unique identification of compounds. Research collaborations make it necessary to exchange data. Data interoperability is not possible without unambiguous compound IDs.

InChI fulfils all of these conditions. In the current version it handles more than 99.9% of all organic compounds correctly. (Goodman, Pletnev, Thiessen, Bolton, & Heller, 2021) The current development is focussed on the better representation of inorganics and organometallics and to fill open issues in the area of stereochemistry like MDL enhanced stereochemistry or atropisomers. In a second step, new transformation rules will be implemented to improve the unique representation of tautomeric groups. The support of polymers is already in the test phase. Last, but not least, InChI will provide the registration functionality supported by molfiles.

The InChI application framework addresses the needs of chemical reactions, mixtures and formulations, and nano materials. In all of these cases the identification is based on multiple compounds with each of them being represented by InChIs. In the case of reactions, InChIs represent each of the reactants, products and agents (catalysts, solvents). Additional string sorting rules make it possible that a unique identification called Reaction-InChI or RInChI is created. (IUPAC-InChI/RInChI) (Goodman, Blanke, & Kraut, Analysing a billion reactions with the RInChI, 2022) Mixtures and formulations seen as ordered mixtures consist of the InChIs of each component, the order of the component during the creation of the mixture and may be extended by other physical data and ensure that the unique identifier Mixture-InChI (MInChI) is built. (IUPAC/MInChI) With Nano-InChIs (NInChIs) the usage of InChI is extended to the world of materials where multiple material layers must be described together with physical data in a way that a unique identifier can be created out of it.

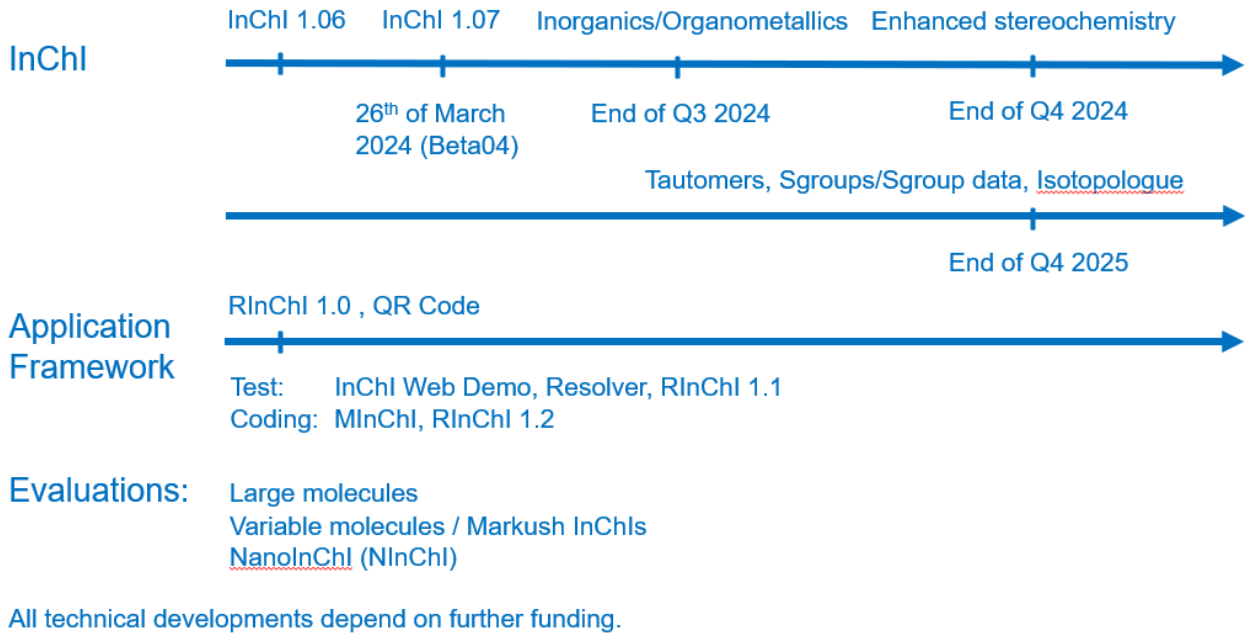


*Figure 2: The InChI roadmap, status beginning of April 2024*

InChI is an open-source application available on GitHub (/IUPAC-InChI/InChI) under the MIT license.

InChI is used in industry to identify compounds during the R&D processes,

InChI and its InChIKey are found in most of the publicly available databases like PubChem, CAS and ZINC. All together you find far more than one billion compounds that are publicly available via InChIs.

## 3. Conclusions

InChI has been developed as a unique universal identifier for molecules. The recent developments improve the treatment of inorganics, organometallics, complex stereochemistry and structure normalization. The InChI application framework extends the usage into reactions, mixtures, formulations, and nano materials.

That makes InChI to an integral part of FAIR data handling in R&D of chemical compounds.

## 4. References

*/IUPAC-InChI/InChI*. (n.d.). Retrieved from https://github.com/IUPAC-InChI/InChI

Goodman, J. M., Blanke, G., & Kraut, H. (2022). Analysing a billion reactions with the RInChI. *Pure Appl. Chem*. Retrieved from https://doi.org/10.1515/pac-2021-2008

Goodman, J. M., Pletnev, I., Thiessen, P., Bolton, E., & Heller, S. R. (2021). InChI version 1.06: now more than 99.99% reliable. *Journal of Cheminformatics*, 1340. doi:10.1186/s13321-021-00517-z

*IUPAC/MInChI*. (n.d.). Retrieved from https://github.com/IUPAC/MInChI

*IUPAC-InChI/RInChI*. (n.d.). Retrieved from https://github.com/IUPAC-InChI/RInChI