Digital IUPAC Ten Years On

by Jeremy G. Frey

It has been 10 years since my article on Digital IUPAC was published in *CI* [1]. On the same pages in a text box the new CPCSD—Committee on Publications and Chemical Data Standards was introduced as IUPAC's initial formal recognition of, and foray into, the digital world.

A lot has happened in the world since then! Continuing fall out from financial crashes, globalisation (both continuing and unwinding), COVID-19 and lockdowns, remote teaching, and the rise in online digitally facilitated meetings (even as far as virtual assistants summarising meetings [2]). The UN Sustainable Development Goals arrived in 2015 and have underpinned many of the science objectives over the last decade. Digital, Computational, and AI areas have figured quite prominently in the IUPAC Top Ten Emerging Technologies [3].

While I could, and did, foresee the need to ensure that chemical data should be fully available in digital forms to meet the increasing digitalisation of chemistry, I certainly did not foresee that the huge rise in the importance of Machine Learning and in particular Deep Learning, as developments in AI, and the consequential demands on the need for large amounts of high-quality chemical data.

Chemists have made use of data driven science right from the foundations of chemistry, and the use of statistical modelling techniques, for example in drug discovery, using QSAR approaches and really examples of machine learning (and many of the lessons and principles from QSAR [4] apply to ML). What has changed is the exploitation of very large collections of data for deep learning—for example AlphaFold [5] making use of the pdb [6] resulting in the protein structure library held at the EBI [7], and the amazing abilities of the Large Language Models (LLMs) trained on the corpus of data available via the World Wide Web, which even if not specifically trained on chemistry are proving interesting and useful way to support chemistry research and (or subvert) teaching.

Some digital standards have been set (InChI, JCAMP-DX, Digital SI, CODATA's work). The InChI has been very successful in providing molecular identifiers and work is going ahead to extend the range of applicable chemistry. IUPAC has given significant support and recognition across a broad range of research to support the FAIR objectives as published in 2016 [8]. Interestingly FAIR is much better understood as a necessity by industry than in academia—poor exchange of data within a company comes with a major potential financial penalty! IUPAC's role in FAIR includes the major WorldFAIR project [9] which is developing examples and processes to ensure that Chemistry to move towards greater FAIRness.

In a major move towards a Digital IUPAC is the major project on an updated and fully digital version of the Gold Book so that in due course all major chemical terms will be available in a digital, human, and machine readable forms, and with unique IUPAC approved digital identifiers (several converging projects are managed by a dedicated subcommittee [10]). Work is also in progress to provide a digital version of future editions of the Green Book [11] and link these entries to those in the Gold Book.

So much has been achieved but having moved a little distance along the path to digitalisation, it is clear just how much still needs to be done and how complicated the work is. AI has moved up the pace given but also holds out the hope that it might facilitate some of the heavy lifting needed to digitalise processes and data. IUPAC's international standards setting agenda is as much in need as ever—*plus ça change, plus c'est la même chose*.

References

1. Frey, Jeremy G. Digital IUPAC: A Vision and a Necessity for the 21st Century. *Chem. Int.* 36(1), 14-16 (2014); https://doi.org/10.1515/ci.2014.36.1.14
2. https://zapier.com/blog/best-ai-meeting-assistant/, https://www.linkedin.com/pulse/leveraging-ai-virtual-assistants-enhance-business-ivo-ten-voorde-868qe/?trk=article-ssr-frontend-pulse_more-articles_related-content-card (accessed 8/04/2024)
3. https://iupac.org/iupac-2023-top-ten/
4. https://www.oecd.org/chemicalsafety/risk-assessment/qsar-assessment-framework.pdf (accessed 08/04/2024)
5. https://www.nature.com/articles/s41586-021-03819-2
6. https://www.rcsb.org
7. https://alphafold.ebi.ac.uk
8. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016); https://doi.org/10.1038/sdata.2016.18
9. WorldFAIR Chemistry: making IUPAC assets FAIR, Leah McEwen, IUPAC project 2022-012-1-024; https://iupac.org/project/2022-012-1-024/
10. IUPAC Joint Subcommittee on the IUPAC Gold Book, Stuart Chalk and Jan Kaiser (chairs); https://iupac.org/body/039/
11. Preparation of the 5th Edition of the IUPAC Green Book, Jeremy Frey, IUPAC project 2019-001-2-100/; https://iupac.org/project/2019-001-2-100/

Jeremy G. Frey, j.g.frey@soton.ac.uk, is professor at the School of Chemistry, University of Southampton, UK; ORCID 0000-0003-0842-4302