

Digital Standards: A Path to Sustainable and Interoperable Chemical Data Exchange

by Fatima Mustafa, Dana Vanderwall, Leah McEwen, Ray Boucher, Richard Hartshorn, Ian Bruno, Greta Heydenrych, and Gerd Blanke

In today's rapidly evolving digital landscape, the importance of standardized data exchange cannot be overstated, especially in fields as critical as chemistry that contribute hugely to many related disciplines. Recognizing this urgency and having the drive of applying Sustainable Development Goals (SDGs) especially strengthen global partnerships, IUPAC has led a workshop in collaboration with the InChI Trust, and the Cambridge Crystallographic Data Center (CCDC). This effort was in the form of a focused invitation-only workshop [1] that was held at the CCDC, Cambridge, UK during 25-26 March 2024 with the aim of charting a course for the development, adoption, and maintenance of digital standards within the Chemistry community. The workshop was held under the umbrella of the WorldFAIR Chemistry project coordinated by CODATA and RDA [2]*. It gathered participants from diverse backgrounds and areas of

expertise—academia, industry, standards organizations, publishing, and cheminformatics—who rarely have the opportunity to collaborate and be in one room. Using the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [3], the workshop sought to address fundamental questions surrounding sustainability, stakeholder and user engagement and the future of chemical data standards in an increasingly complex digital world.

The Importance for Digital Standards

For over a century, the Chemistry community has relied on established standards to communicate complex chemical information accurately and efficiently. IUPAC has played a key role in setting and disseminating these standards. Yet, with the advent of digital technologies, the landscape is shifting dramatically. Digital standards play a crucial role in enabling interoperability and data exchange across diverse scientific domains [4], and reinforce the digital economy as it is significantly enhanced by digital data [5]. While these technologies offer unprecedented opportunities for data exchange and collaboration, they also present new challenges, particularly in terms of implementation and scale. The need for consistent, machine-readable data exchange is more pressing than ever, necessitating a

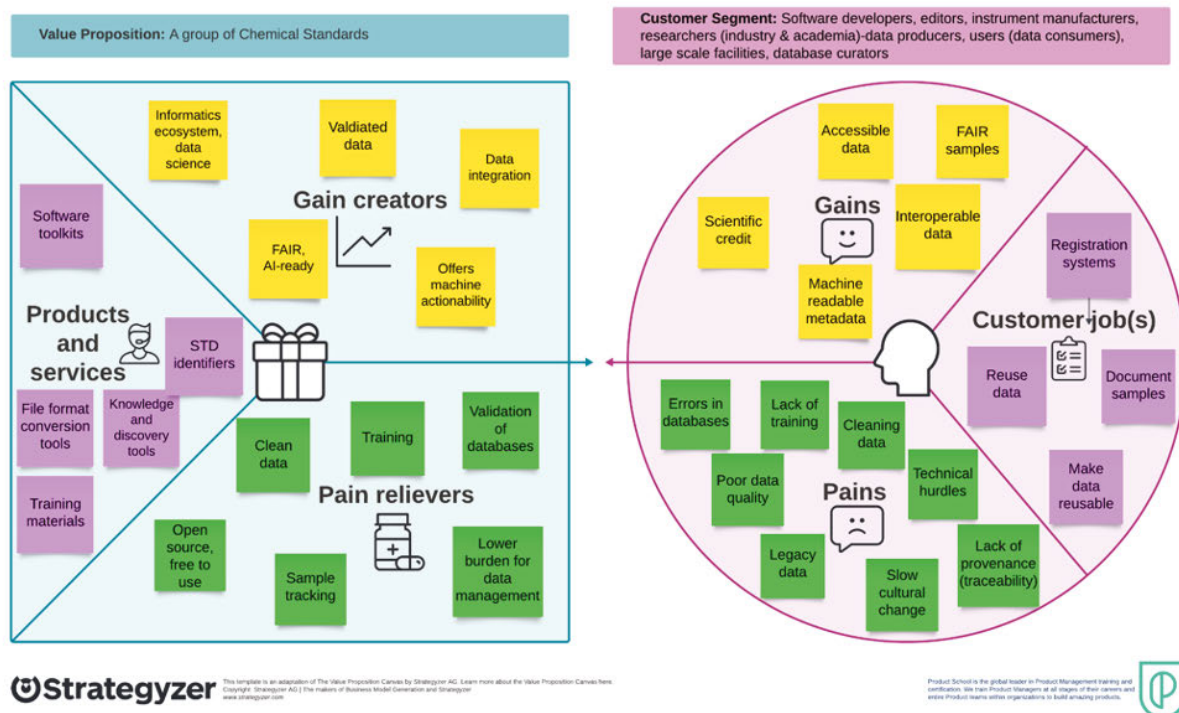


Figure 1. Business Model Canvas based on hands-on exercises in which participants examined multiple existing Chemical standards.



Participants of the “Sustainable Business Modeling for Digital Standards Development” workshop, CCDC, Cambridge, UK, 25-26 March 2024. Image credited to Michael Webb ©University of Cambridge

robust and proactive approach to standards development and adoption.

Paving the Path for Sustainability

While numerous valuable standards have emerged from collaborative efforts in Chemistry and the Life Sciences, many of these projects have lacked a sustainable plan or resources for long-term viability. Establishing a sustainable framework for projects targeting specific demands around digital standards for chemistry, thorough comprehension of the economic landscape, operational dynamics, and resources essential for developing, sustaining, and maintaining a solution (digital standard and supporting tools). This area may frequently be unfamiliar to scientists or developers who initially conceive solutions for common problems. Based on Peter Drucker's definition, a good business model answers the questions: “Who is the customer? And what does the customer value? It also provides the fundamental questions every manager must ask: How do we make money in this business? What is the underlying economic logic that explains how we can deliver value to customers at an appropriate cost?” [6].

A critical focus of the workshop was the challenge

of sustainability. Developing and maintaining digital standards requires significant ongoing investment of resources and expertise, making long-term sustainable business models a pressing concern. Through hands-on exercises, participants examined multiple existing standards such as:

- InChI - Identifier for chemical structures
- IGSN - Persistent identifier for samples
- HELM - Notation for biological macromolecules
- SMILES - Notation for chemical structures
- MOL/SDfile - Representation format for chemical structure
- UDM - Unified data model for chemical reactions
- CIF - Exchange format for crystallographic data + checkCIF
- ChEBI - Ontology for chemical entities of biological interest
- IDMP-o - Ontology for identification of medical products
- ADF - Allotrope Data Framework for analytical data.

Throughout the activity, the business model (Figure 1) and value proposition canvas were used to define community, contributors and users, and

other stakeholders; define parameters around values, both open and revenue-generating; articulate target resources and services to serve the broader community in using standards; initiate market analysis and identify funding streams which will feed into the IUPAC roadmap for digital Chemistry standards.

The main themes derived from these discussions were:

- **Value Proposition:** at first standards might seem similar in benefits; however, it would be helpful to communicate with potential adopters to help differentiate the utility of each one in certain use cases.
- **Relationships with customers:** It was common to require extra activities and resources to build and keep relationships, bring more awareness and understanding which are a precondition to adoption.
- **Resources:** All the standards we studied commonly require technical resources, although to a lesser degree for operational support tasks.
- **Activities:** Various aspects such as payment innovation, operational support including licenses, travel, and software development alongside strategic planning, software development sustainability, and infrastructure maintenance and construction are crucial considerations. Additionally, efforts directed towards driving adoption through innovation workshops and bringing subject matter experts together. These activities collectively contribute to the comprehensive ecosystem supporting advancement and adoption.
- **Revenue Models:** There's a widespread interest in assessing license and revenue models that achieve a balance between accessibility and sustainability.

Stakeholder Engagement

Central to the workshop's agenda was the exploration of strategies to maximize the opportunities presented by digital technologies. Participants delved into questions of stakeholder engagement, identifying key players in the Chemistry data economy, and fostering collaborative approaches to standards development. Stakeholder engagement is essential for the effective development and implementation of digital standards, ensuring that they meet the diverse needs of the user community [7]. By aligning efforts across disciplines and sectors, the workshop aimed to ensure that digital standards remain relevant, accessible, and interoperable in an increasingly interconnected world.

Looking Ahead

The collective goals for Chemistry digital standards encompass building relationships, establishing new partnerships, enhancing outreach and engagement, managing technical and coordination resources, and understanding business models. The discussions reflected a shared interest in clarifying common objectives and limitations of collaborative efforts, identifying key stakeholders, and extending the reach of the initiative. Participants emphasized the importance of a well-articulated mission, potentially through an extended white paper [8] (which was initially drafted to outline the scope of the workshop), to guide future actions and recruit additional contributors. Additionally, there was a recognition of the need to address resource challenges, explore alternative models for sustainability, and connect with relevant international organizations such as CODATA and RDA, as well as national organizations, the private sector, and funding organizations. The dialogue also touched on the potential impact of digital standards on related disciplines and industries, particularly in the context of artificial intelligence and machine learning applications.

Following the workshop, there will be an 18-month plan, involving contributions from various participants to continue working on the white paper and to convene for another workshop. The plan includes the establishment of a coalition, association, or federation dedicated to sustaining digital standards in the chemical sciences and adjacent disciplines. Workstreams will be discussed in June 2024, and a workshop focusing on the sustainability of standards organizations themselves will be in near future.

Overall, the focus was on fostering collaboration, defining strategic priorities, and leveraging collective expertise to advance the development and adoption of Chemistry digital standards. The sustainability of chemical informatics standards depends on active engagement, collaboration, and support from the broader community of stakeholders. By following best practices in standards development, collaboration, adoption, and long-term planning for maintenance and resourcing, we can ensure that these standards continue to support research, education, and innovation in Chemistry and related fields for years to come.

References

1. Vanderwall, D., & Leah, M. (2024). Workshop: Sustainable Business Modeling for Digital Standards Development. Zenodo. <https://doi.org/10.5281/zenodo.10966075>
2. WorldFAIR Chemistry: making IUPAC assets FAIR, <https://iupac.org/project/2022-012-1-024/>

3. 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>
4. NIST (2018), The Role of Standards in Digital Transformation. National Institute of Standards and Technology
5. OECD Roadmap Towards a Common Framework for Measuring the Digital Economy (2020)
6. <https://hbr.org/2002/05/why-business-models-matter>
7. OECD Roadmap Towards a Common Framework for Measuring the Digital Economy (2020)
8. Vanderwall, D., & Leah, M. (2024). Workshop: Sustainable Business Modeling for Digital Standards Development. Zenodo. <https://doi.org/10.5281/zenodo.10966075>

Acknowledgment

The organizing team extends heartfelt gratitude to all workshop participants, and panelists whose enthusiastic engagement and insightful contributions were crucial to the event's success. Many of them have expressed their willingness to further collaboration and commitment.

Workshop organizing team and authors

Dana Vanderwall <dvanderwall@dlc.io> is Senior Director of Digital Transformation at Digital Lab Consulting, a ProPharma Group Company. He is a member of the IUPAC Committee on Publications and Cheminformatics Data Standards, Chair of the IUPAC Subcommittee on HELM, a Fellow of the Society for Laboratory Automation and Screening and contributes to projects within the Pistoia Alliance. ORCID 0000-0003-1076-8788

Leah McEwen <lrm1@cornell.edu> is a Chemistry Librarian at Cornell University, in Ithaca, NY, USA. She is currently chair of the Committee on Publications and Cheminformatics Data Standards (CPCDS) of IUPAC, responsible for the design and implementation of digital standards and lead on the WorldFAIR Chemistry project led by IUPAC and to advance FAIR data practices in Chemistry. ORCID 0000-0003-2968-1674

Gerd Blanke <Gerd.Blanke@structurependium.com> is a principal Consultant and Managing Director, StructurePendium Technologies GmbH and a member of the IUPAC Subcommittee on InChI. ORCID 0000-0001-8147-7000

Ray Boucher <ray@inchi-trust.org> is a retired Publishing professional, and currently chairs the InChI Trust Board. ORCID 0000-0002-4786-4223

Ian Bruno <bruno@ccdc.cam.ac.uk> is Director of Data Initiatives at The Cambridge Crystallographic Data Centre, Chair of the WorldFAIR Chemistry subproject - Reporting Guidance: Recommendations for FAIR chemical data reporting, and Secretary of the InChI Trust. ORCID 0000-0003-4901-9936

Greta Heydenrych <gheydenrych@iupac.org> is the Executive Director of IUPAC and collaborates closely with CPCDS to define IUPAC's role in chemistry's digital future. ORCID 0000-0003-3329-5901

Fatima Mustafa <fatima.mustafa@utsa.edu> is a Chemistry lecturer at the University of Texas in San Antonio (UTSA), Texas, USA. She is the coordinator of the WorldFAIR Chemistry project. ORCID 0000-0001-6754-7375

Richard Hartshorn <richard.hartshorn@canterbury.ac.nz> is a professor of Chemistry at the University of Canterbury | Te Whare Wananga o Waitaha in Christchurch, New Zealand. He recently completed eight years as Secretary General of IUPAC, is a member of the International Chemical Identifier (InChI) Trust Board, and is Vice President of CODATA, the Committee on Data of the International Science Council. ORCID 0000-0002-6737-6200

* 'Global cooperation on FAIR data policy and practice' (WorldFAIR) has received funding from the European Union's Horizon Europe project call HORIZON-WIDERA-2021-ERA-01-01, grant agreement 101058393.



Feature Articles Wanted

Contact the editor for more information at <edit.ci@iupac.org>.

GREEN CHEMISTRY POSTGRADUATE SUMMER SCHOOL

In memory of Professor Pietro Tundo

30th of June - 5th of July 2024

Venice, Italy

Organizers:

Francesco Trotta Chairman

Fabio Aricò

Aurelia Visa

Mirabbos Hojamberdiev

Graziana Gigliuto Secretary

Topics:

Benign synthesis routes

Green catalysis

Alternative solvents

Renewable and green raw materials

Green chemistry for energy

Clean processes

Green Chemistry education

Info:

www.greenchemistry.school

www.gssd-foundation.org

Contacts:

postmaster@pec.gssd-foundation.org

secretariat@gssd-foundation.org