

2018 Workshop: Supporting FAIR Exchange of Chemical Data through Standards Development

July 16 - 17, 2018, The Netherlands
Science Park, University of Amsterdam

Cosponsored by:

- International Union of Pure and Applied Chemistry (IUPAC)
 - Committee on Publications and Cheminformatics Data Standards (CPCDS)
 - CPCDS Subcommittee on Cheminformatics Data Standards (SCDS)
- CODATA, the Committee on Data of the International Council for Science

Organizing Team:

- | | |
|--|--|
| • Leah McEwen, SCDS Co-Chair,
<i>Workshop Chair</i> | • Jeremy Frey, SCDS member, GO
FAIR ChIN member |
| • Ian Bruno, SCDS member | • Simon Hodson, CODATA Executive
Director |
| • Stuart Chalk, SCDS member, GO
FAIR ChIN member | • Robert Lancashire, SCDS member |
| • Simon Coles, IUCr CommDat
member, GO FAIR ChIN lead | • David Martinsen, SCDS Co-Chair |
| • Antony Davies, SCDS member | • Egon Willighagen, GO FAIR ChIN
member, ELIXR |

Workshop Agenda:

- Iupac Interoperability Criteria for Spectroscopic Data Exchange
- GO FAIR Implementation Network for Chemistry

Description:

The GO FAIR initiative is intended to encourage and enable accurate and reproducible data exchange among and across disciplines. FAIR encapsulates the concepts that data should be Findable, Accessible, Interoperable and Reusable. This workshop will explore how these goals can be accomplished in the context of chemical data, addressing some of the issues present within chemistry, but also in the context of broader interaction with other disciplines. The topics have been chosen to begin a process of standards development and maintenance, tool development and maintenance, and addressing use cases from other domains. Key opportunities include interoperability criteria for spectra using IUPAC standards, prioritizing activities within the GO FAIR ecosystem, creating draft roadmaps with key milestones, and recruiting key leaders for ongoing activity.

Begins: 10:00 AM, Monday 16 July

Closes: 15:00 PM, Tuesday 17 July

Registration: [closed, contact Workshop Chair, lrn1@cornell.edu]

Hotel list: <https://drive.google.com/open?id=1GjbB4sU8PJkV5qt4-enzOg97ISY1leEzeV8RM2fEb90>

Directions: https://drive.google.com/open?id=1_ju1KSaLfP3W-4u_9WPfOpmkLqwNYHrbVcKnZLvCqwc

Schedule at-a-glance

Monday 16th AM	Background and context • IUPAC 100 and Beyond. Richard Hartshorn (IUPAC Secretary General) • CODATA Data Interoperability. Simon Hodson (CODATA Executive Director) FAIR chemical data • GO FAIR. Erik Schultes or Barend Mons (GO FAIR) • FAIRsharing: describing and linking standards, databases and policies across the natural sciences. Peter McQuilton (OeRC) • The WDS/RDA Assessment of Data Fitness for Use Working Group. Marina Soares E Silva (Elsevier) • That's not FAIR – why chemists don't share more. Martin Hicks (Beilstein Institut) • Community-enabled FAIR Crystallographic Data. Ian Bruno (CCDC) Spectroscopic data exchange • JCAMP-DX. Robert Lancashire (University of the West Indies), Antony Davies (AstraZeneca) • JMol. Robert Hanson (St. Olaf College) • Using JCAMP-DX. Greg Banik (BioRad) • NMReDATA. Damien Jeannerat (University of Geneva) • Allotrope Data Format, Henning Kayser (Bayer) • MPublish open spectra exchange efforts. Vadim Zorin (MestraLab)	
Monday 16th PM	FAIR Data strand**	Spectra Standards strand
	Developing a FAIR data guidance tool. - What Persona's do we need? - What should a Persona comprise? - Develop Persona's and link to a FAIRsharing collection of resources	JCAMP-DX review & future requirements - Benefits of JCAMP/agnostic data exchange - Deficiencies of JCAMP format - Requirements for evolving JCAMP (extensions, XML, etc.) - Community engagement - Validation requirements
Tuesday 17th AM	- FAIR data gap analysis of the chemical data landscape - Synergies between IUPAC, RDA & GO FAIR - Analysis of needs for chemical data in other disciplines - Scope a White Paper	Interoperability criteria for reporting and publishing spectroscopic data - Use cases: archive and audit raw data; communicate characterization processing - Description for raw data and subsequent processing - Workflow frameworks originating from instruments

Tuesday 17th PM	- Validating the FAIR landscape using the Spectra example - Prioritizing current efforts in IUPAC that support FAIR - Enabling and encouraging the wider community to engage
-----------------------	--

**** The discussions from this workshop will be recorded in a shared GoogleDrive - <https://drive.google.com/open?id=1jMMS4NwdQgYRPCxPpt3jOAHzrE8by2q> or <http://bit.ly/IUPACFAIR2018> for short!**

The slides are available here too. All participants are encouraged to co-author documents.

Interoperability Criteria for Spectroscopic Data Exchange

FAIR data principles require that data be Findable, Accessible, Interoperable, and Reusable. The IUPAC JCAMP-DX data standard has become a critical piece of this FAIR data exchange for spectroscopic data. It satisfies a number of critical criteria in that JCAMP-DX file export is available in nearly all software packages for spectroscopic instruments, it is ASCII not Binary, it is non-proprietary and there has been a large amount of data already generated. IUPAC is reviewing the current status of the JCAMP-DX format, including the extent to which it is being used, what enhancements users would like to see, and the extent to which the files that are generated in "JCAMP-DX format" adhere to the JCAMP-DX standards. Another key step towards FAIR spectroscopic data is the development of standard criteria for publishing spectroscopic data that will optimize data use, reuse, and interoperability across domain repositories.

GO FAIR Implementation Network for Chemistry

The GO FAIR Chemistry Implementation Network (<https://www.go-fair.org/implementation-networks/>) is in the initial phase of being established. This involves a working group and champions promoting the initiative, as well as beginning to build a repository of FAIR resources that are of use to chemists and those who wish to use or manage chemistry data. The next phase of work will involve understanding the landscape of FAIR resources that already exist for this group and performing a GAP analysis to highlight where short, or long, term work is required. The purpose of the GO FAIR aspect of this workshop is to establish a mechanism to achieve this and begin the work.

Key Themes to Address in Supporting FAIR Data:

- I. Use cases and interoperability needs for chemical data and information across the enterprise and related disciplines
- II. Development of tools for researchers and other expert users to support application and use of standards for chemical data
- III. Mechanisms for validation and curation of standard representation of chemical data

I. Use Cases for Interoperability of Chemical Data

As a central science, chemistry underpins many other disciplines and chemical data are used broadly across the pure and applied sciences. Implementation of standard formats and representation of chemical data could greatly facilitate data management for professionals across and beyond chemistry and expand connections to data in other disciplines. Understanding the contexts in which chemical data are referenced in other disciplines and what chemical aspects these users are interested in can inform how chemical data standards are developed, promoted and supported; for example, guides on using and interpreting chemical data formats based on examples from related disciplines, or bridging across types of data packages by including chemical notation as metadata. Developing and understanding use cases required by other disciplines cannot be done from within the chemistry domain. Discussions with other researchers outside the domain, enabled through contact with CODATA and the other Unions, is required.

II. Development of Tools to Support Use and Reuse of Chemical Data

Specification for standard chemical data reporting is not enough to enable wide adoption in the digital environment. Tools are necessary for development, implementation and re-use of digital formats for seamless and accurate data exchange between systems. Syntax checkers and standard test suites are important for validating machine readable data formats. Prototype implementations and input formatters can facilitate uptake and incorporation into workflows for databases, publishers, instrument laboratories and other large data facilities. Decoders and resolvers can support review and use of digital data by chemists and other scientists. Surveying the landscape of existing toolkits and identifying gaps for different stakeholders can inform next steps for enabling FAIR chemical data. There has been much focus on digital chemical representation, but other areas, such as spectral data package specifications and chemical metadata description, also are in need of definition and support.

III. Mechanisms for Validation and Curation of Chemical Data

Critical evaluation is an important function for re-use of chemical data. Consistent description of chemical data, including units, uncertainties and other measurement conditions, are crucial for the evaluation process. Additionally, in digital environments, checking file syntax and algorithm efficacy are needed to ensure accurate data representation in automated workflows. Identifying validation requirements for digital data files and formulating these processes through IUPAC is a key step for the utility of existing and emerging chemical data formats.

Communicating use of official and validated IUPAC standard formats will be a significant value-add for sharing FAIR chemical data.