WorldFAIR

Project Title	Global cooperation on FAIR data policy and practice
Project Acronym	WorldFAIR
Grant Agreement No	101058393
Instrument	HORIZON-WIDERA-2021-ERA-01
Topic, type of action	HORIZON-WIDERA-2021-ERA-01-41 HORIZON Coordination and Support Actions
Start Date of Project	2022-06-01
Duration of Project	24 months
Project Website	http://worldfair-project.eu

MS9 WorldFAIR Chemistry Public Materials

Work Package	WP03
Lead Author (Org)	Leah McEwen (Cornell University)
Contributing Author(s) (Org)	Fatima Mustafa (IUPAC)
Due Date	31.08.2023
Date	29.08.2023
Version	1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

DOI https://doi.org/10.5281/zenodo.8296925

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Table of contents

'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.

1. WorldFAIR Chemistry outreach events

The WorldFAIR Chemistry case study has an objective of engaging stakeholders in articulating needs for chemical data sharing, identifying gaps that restrain chemistry related research and developing guidelines, tools and validation services that enable scientists to share and store data in a FAIR manner. The project has organised community discussions through a number of different venues and formats to engage stakeholders across roles, sectors and disciplines. This Milestone provides a listing of the workshops, a brief description of each and links to the publicly-available outputs.

1.1 What is a chemical? webinar series

A summary of the webinar series 'What is a Chemical?' is publicly available online[1]. This webinar series ran from September 2022 to February 2023 in virtual form. Organisers were Leah McEwen, Ian Bruno, Stuart Chalk and Fatima Mustafa.

1.1.1 Webinar series description

Chemical substances touch on all areas of laboratory science and chemistry underlies many critical worldwide issues, including climate, health, food availability and sustainable development. Increased reporting of machine-readable chemical data will support active research in chemistry and related sciences worldwide, and will be essential to the development of the interdisciplinary science critical to address the UN Sustainable Development Goals and UNESCO's priorities around Open Science. IUPAC is the world authority on chemical nomenclature, terminology, and standardised methods of measurement, and is engaging in a concerted effort through collaboration with the broader chemistry and data science communities to translate a range of assets and activities into the digital domain.

IUPAC is leading the chemistry work package WP03 of the EU-funded WorldFAIR project, aiming to align standards development and implementation with the FAIR data principles. This will facilitate development of guidelines, tools and validation services that support scientists to share and store chemical data in a FAIR manner and support the ability to compile and interpret data across scientific disciplines. In this webinar series, we aimed to:

1. Understand the chemical substance notations used within multiple disciplines *(geochemistry, nanochemistry, atmospheric chemistry, environmental chemistry, oceanography, crystallography, etc.)*.

2. Explore the data resources that are used in these applied areas *(life sciences and pharmaceuticals, agriculture and crop protection, dyes and pigments, and machine learning)*, and understand the current ways of communication and accessing data by other groups.

[1] https://doi.org/10.5281/zenodo.7903683 (accessed 20230505)

'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.

3. Investigate various digital and machine-readable depictions or notations of chemical substances, reactions and datasets (*InChI, HELM, SMILES, Graphical Representation, Systematic Representation, Media Types, ChEBI, Ontologies4Chem, and notations of Mixtures-Molecules and Complex Substance Schema)*.

Overall, the webinar series highlighted the current status of working with chemical notations, development of digital tools to transform chemical notations into digital entities and ways to implement the FAIR data principles across the chemical enterprise.

1.1.2 Overview of webinar sessions

1. "What is a chemical? Handling Chemical Data Across Disciplines" [2]

   In this webinar, seven experts from diverse backgrounds were invited: Emma Schymanski (Environmental Sciences), Iseult Lynch (Nanomaterials), Lesley Wyborn (Geochemistry), Kerstin Lehnert (Astromaterials), Marie-Lise Dubernet (Astrophysics), Ken Kroenlein (Material Science), and Dylan Walsh (Polymers). The panellists discussed the status of working with machine-readable formats of chemicals within their field and indicated challenges and future needs. The conversation revealed areas of emphasis that users and developers of machine-readable formats should work to advance and to which IUPAC can potentially contribute.

   - **Representation and finding chemical data** (If we can't represent materials, how are we going to be able to find them?) **:** Identifiers for small molecule chemicals such as InChI and InChIKey are widely used. Similiarialy for polymers, BigSmiles is in use; however stochasticity is a challenge to translate into machine-readable format. Polymer properties are based on assemblies, and assembling and processing needs to be known as they affect properties. Moreover, a notation for nanoparticles that builds on InChI is still in development. Linking as-synthesised material to various transformed forms (during storage, upon dispersion, in the environment, body etc.) to correlate actual form with effect is needed. This could help geochemists in identifying isotope ratios / crystal phases or domains etc.
   - **Real sample complexity** (Can we have chemical identifiers to reflect sample complexity?) **:** There are a variety of different types of chemical samples (cosmochemical, ceramic, alloys, nanomaterials, mixtures, etc.) with multiple factors (type, age, source, storage, uniformity, history, etc.) that play a role in sample representation.

[2] https://doi.org/10.5281/zenodo.7259101 (accessed 20230505)

'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.

2. "What is a chemical? Applying Chemical Data to Industrial Challenges" [3]

   In this webinar, six experts from multiple applied chemical areas were invited: Lutz Weber (AI & ML), Teodoro Laino (AI & ML), Yannick Djoumbou Feunang (Agir-Chem), Nelson Vinueza Benitez (Dyes and Pigments), Nick Lynch (Pharmaceuticals and life Science), Gunther Schadow (Pharmaceutical-Regulatory). The panellists discussed the challenges and forthcoming regulation requirements for using chemical data in machine-readable formats and provided many recommendations:

   - Need an easy to use more expressive molecular representation, across all nominal identifiers, extending InChI for bigger structures, and making molecules open for everyone!
   - Explore and adopt technologies to aid chemists as they generate data.
   - Define rules for structures, organising and sharing physical property data with structures.

3. "What is a chemical? User' Perspectives on Digital Machine Readable Depictions" [4]

   In this webinar, experts in five various digital machine-readable depictions were hosted: Greg Landrum (InChI), Dana Vanderwall (HELM), Jonathan Goodman (Graphical Representation), Michelle Rogers (Systematic Nomenclature), and Vincent Scalfani (SMILES). The panellists identified some gaps in chemical representation/identifiers such as:

   - Graphical representation is not standardised for small molecules including organometallics, and isomers.
   - The need to define/determine uniqueness, capturing process of macromolecules including glycans, and lipids.
   - Representation of created mixture of unique chemicals and differentiating these to a "natural product" such as canola oil.
   - Collecting critical information to make determinations on consistent representation is challenging.
   - Standardisation of description for non-standard structures.
   - Variability in interpretation of structures in different jurisdictions.

   The conversation invited users to be aware of common depictions, to use them in different applications and to understand what can and cannot pass from sketching tools into standard formats.

[3] https://doi.org/10.5281/zenodo.7259727 (accessed 20230505)
[4] https://doi.org/10.5281/zenodo.7435258 (accessed 20230505)

'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.

4. "What is a chemical? Innovations in Chemical Descriptions" [5]

   Experts of five creative descriptions were featured in this webinar: Henry Rzepa (Media Types), Adnan Malik (ChEBI), Oliver Koepler (Ontolgies4Chem), Alex Clark (Mixtures-Molecules), and Ken Kroenlein (Complex Substance Schemas). The panellists identified technical challenges that need to be focused on including:

   - Solubility: need for better solubility models;
   - Data capture: FAIR pushing forward, balance with informatics community building connections in parallel;
   - Data quality: particularly auto-validating before publication;
   - Community consensus: IUPAC chemical classification;
   - Many nuanced details arise in real world challenges:
     - What are critical variables in different use cases?
     - How many layers need to be captured in standard data models for interoperability?
     - How many layers arise empirically? How many can be computed?

1.1.3 IUPAC potential contributions

Overall, the four conversations reflected what the IUPAC can potentially contribute to. Below are some of the suggestions:

- Representations of challenging mixtures;
- How to capture distributions around samples e.g polymer length, phases,etc?
- Education of what machine-readable is and what FAIR is?
- Modularity of things: the sample space is incredibly complex. IUPAC can identify what the core commonalities are in sample space.
- Having a high level controlled vocabulary/machine actionable for the chemical space
- Unify data models that can be broadly used;
- There is a need for coordination of activities in enabling the community to not reinvent the wheel e.g vocabulary;
- IUPAC FAIR Cookbook! reach the base communities at the bench level;
- Facilitate Interoperability based on sharing across the communities and representations;
- Supporting interoperability of FAIR, similar to approach with SMILES+;
- Expanding InChI to other areas (e.g., organometallics);
- Education: creating tables and guides of what can be done with different representations, and documenting common use cases;
- Determining what is critical about different classes of entities (small well characterised, what do we need to know about proteins, other macromolecules, nanomaterials);

[5] https://doi.org/10.5281/zenodo.7683138 (accessed 20230505)

'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.

- What are the key parameters to capture in representations? May need different approaches for different applications;
- Classification is another area that intersects with describing chemical moieties;
- Coordinate common classification schemas for molecular classes and build active chemical and cross-disciplinary ontology community exchange and practice.

1.2 Research Data Alliance (RDA) P20 Plenary Symposium

A summary of the Symposium 'Describing diverse chemistry datasets across distributed data resources' is available online[6]. The Symposium took place on 23 March 2023 at the RDA 20th Plenary Meeting in Gothenburg, Sweden, in hybrid format. The session organisers were Ian Bruno, Stuart Chalk, and Leah McEwen. The group organising the session was the Chemistry Research Data Interest Group (CRDIG).

1.2.1 Speakers

- National Research Data Infrastructure for Chemistry: Johannes Hunold
- NFDI4Chem-Repositories: Felix Bach
- Implementing standards for sharing FAIR chemical data
  - Steffen Neumann, NFDI4Chem, and
  - Leah McEwen, IUPAC
- Physical Sciences Data Infrastructure (PSDI): Brian Mathews
- Catalysis Hub PathFinder: Abraham Nieva de la Hidalga
- Geochemistry Data Interoperability: Kerstin Lehnert

1.2.2 Session summary

This session provided updates and perspectives from regional and disciplinary initiatives relevant to chemistry, focussing on the challenge of describing chemistry data sets to enable interoperability and reuse across resources and domains. This was followed by a discussion that aimed to identify cross-community challenges that might be addressed through activities within the RDA. The discussion identified areas of focus that the group will aim to take forward in collaboration with other RDA groups and community initiatives.

1.2.3 Key outcomes/actions/takeaways

1. Opportunity for cross-community collaboration on agreeing standard approaches for describing samples in chemistry that align with wider initiatives.
2. Desire for a common approach to describing analytical techniques and a catalogue of available standards for representing analytical results in chemistry.

[6] https://www.rd-alliance.org/plenaries/rda-20th-plenary-meeting-gothenburg-hybrid/describing-diverse-chemistry-datasets-across (accessed 20230505)

'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.

3. Value in Identifying use cases that establish what is needed for discovery and interoperability across chemistry resources.

1.3 American Chemical Society (ACS) 2023 Spring Meeting Workshop

A summary of our workshop 'Advancing FAIR Chemistry: Developing New Services for Sharing Chemical Data' is available online [7]. The workshop occurred on 27 March 2023 at the ACS Spring Meeting in Indianapolis, USA. The organisers were Leah McEwen, Ian Bruno, Stuart Chalk, Evan Bolton and Fatima Mustafa.

1.3.1 Speakers

- Cornell University: Leah McEwen
- Cambridge Crystallographic Data Center: Ian Bruno
- University of North Florida: Stuart Chalk
- U.S. National Center for BioInformatics: Evan Bolton
- St. Olaf College: Robert Hanson
- University of Alabama: Vincent Scalfani
- IUPAC: Fatima Mustafa

1.3.2 Workshop summary

The goal of the WorldFAIR Chemistry case study is to support the use of chemical data standards in research workflows to enable downstream data reuse through practical direction and resources. The aim of this workshop was to engage the input and expertise of stakeholders across the chemistry community on prototype services and other IUPAC standards activities in progress. We presented early work on each of these resources and invited the community to share feedback on what will help them to implement these in your workflows - What works well? What needs further refinement? What is missing?

WorldFAIR Chemistry prototype deliverables:

- **Guidance** [8]: recommendations for managing and sharing FAIR chemical data for various stakeholders;
- **Cookbook** [9]: recipes (Cookbook) for preparing and depositing FAIR machine-enabled chemical data;
- **Protocols** [10]: universal protocol for browser-based validation and lookup services.

[7] https://doi.org/10.5281/zenodo.7903727 (accessed 20230505)
[8] https://iupac.org/project/2022-027-1-024 (accessed 20230505)
[9] https://iupac.org/project/2022-028-1-024 (accessed 20230505)
[10] https://iupac.org/project/2022-029-1-024 (accessed 20230505)

'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.

IUPAC standards projects:

- **Doc-a-thon: Chemical representation best practices for humans and machines**. This session focused on reviewing the existing IUPAC graphical representation standards for chemical structure diagrams and stereochemical representation with considerations for machine-readability.
- **IUPAC FAIRSpec [11]-ready aggregations: Recommendations for researchers, authors, and publishers.** IUPAC is developing specifications for aggregating spectroscopy data and chemical structures for reporting spectroscopy measurements. The resulting data collections can be created either manually by a researcher or automatically by an electronic laboratory notebook (ELN) or laboratory instrument management system (LIMS) and the accompanying metadata can also be used as a finding aid.

1.3.3 Key discussion points

1. Ideally, data management resources should be capturing metadata from the point of sample identification, through experimental techniques, data collection, measurement parameters (including units), instruments, analysis, etc. through publication and re-use.
2. IUPAC standards need models for digital representation, including chemical systems, quantities, measurements and criteria for interoperability.
3. Ambiguity may not always be apparent until formats are moved from system to system, round-trip testing can help with review of specifications.
4. Consider data curators and others managing data who don't necessarily have extensive domain knowledge.

1.4 American Chemical Society (ACS) 2023 Fall Meeting Symposium

Our session, 'Helping Chemists Manage Their Data' took place on 13 - 14 August 2023 at the ACS Fall Meeting in San Francisco, USA. The organisers were Sonja Herres-Pawlis, John Jolliffe, Leah McEwen, Fatima Mustafa, Jochen Ortmeyer, and Nancy Washton.

More experiments than ever before are being conducted by scientists and researchers worldwide. On the other hand, producing a huge amount of data does not always result in its effective capture, dissemination, and reuse. Most labs still collect data manually as they are not digitised. The margin of error in data collection and its impact on the calibre of research are additional factors. The ability to instantly access matched data sets is another difficulty. More work is required to convert data types into common machine-readable formats rather than storing data in PDF files. These formats make it possible for users to find, access, interoperate, and reuse (FAIR) data and research findings.

Many scientists are reluctant to publish data - not for technical reasons but for psychosocial reasons that have various root causes: e.g. a lack of error cultures or pressures due to the publish or perish paradigm. Not only are technological solutions required for this paradigm shift but accompanying

[11] https://iupac.org/project/2019-031-1-024 (accessed 20230505)

'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.

the necessary cultural change and addressing the psychosocial root causes are at least equally as important. In this session, we are keen to discuss not only these issues but also other factors that may be holding researchers back from using new technologies for data management and publications: Lack of support by their organisation or local infrastructure or who pays for the implementation (including staff-hours)? Who pays for the training?

This session focused on using the FAIR data principles to advantage without placing an additional burden on existing workload. Implementation is anticipated to begin with the generation of experimental data, preparation of the results for publication, and consideration of the needs of funders and publishers in this area. Approaches, tools and technologies already exist and are being developed worldwide.

Topics included:

- Developer/researcher practices in data acquisition and storage (e.g. file naming and folder hierarchy), Electronic Lab Notebook (ELN) and integration with instruments, data processing softwares, and transferring data and metadata to repositories (Smart Lab).
- Error in data generation: Trust and error cultures in working groups and its effect on error in data collection, and processing.
- National and international initiatives in science data management.

1.5 In progress and future outreach events

1.5.1 SciDataCon Symposium

Beyond FAIR: Reusing Chemical Data Across-disciplines with CARE, TRUST, and Openness

25 October 2023. International Data Week. Salzburg, Austria. Organisers: Anjana Elapavalore, Hiba Mohammed, Iseult Lynch, Thomas Exner, Lesley Wyborn, Alexander Prent, Ian Bruno, Leah McEwen, Fatima Mustafa.

The UN Sustainable Development Goals (SDGs) are intended to promote awareness and adoption of sustainable practices around the world. There is a huge momentum of data around topics such as deleterious elements in the environment (plastics, water contamination, fossil fuels by-products, electronic waste, climate change), health care, food, biodiversity, oceanography, etc. Assuming these data are made FAIR, how can we re-use multidisciplinary scientific data in addressing these grand challenges of our global society?

Scientific data management has recently received more attention as a result of conversations around Open Science. The adoption of data management practices is still in its infancy. The FAIR Data Principles[12] emphasise the need of adopting best practices by outlining crucial attributes of data objects to guarantee that data be reused by both humans and machines. They ought to be

[12] https://www.nature.com/articles/sdata201618

'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.

Findable, Interoperable, Accessible and Reusable. The FAIR Principles and the CARE Principles[13] (Collective benefit, Authority to control, Responsibility, Ethics) are meant to work in tandem. On the other hand, FAIR data preservation requires trustworthy digital repositories function with the TRUST principles[14] (Transparency, Responsibility, User Focus, Sustainability, Technology).

In this session, we aim to look beyond FAIR data from reusable to reused. Through real-world research projects we will hear how FAIR data can be integrated across disciplines to enable innovative analysis - what opportunities were realised by having access to FAIR data? What challenges continue to arise in working with heterogeneous data? How usable are chemistry data in different disciplinary contexts? We will explore use cases of interoperability around chemistry data in the broader environment and the range of issues that may arise in applying chemical notation to nanomaterials, therapeutics, geochemistry, atmospheric research and beyond.

We will concentrate on how data concepts such as FAIR, CARE, TRUST, openness and synergies among them can maximise data applicability to timely real world problems. Putting these principles into practice in everyday research is critical for increasing data sharing and availability. We will also explore different approaches to support research data workflows through several large data initiatives.

Invited talks will focus along these axes:

- Data concepts beyond FAIR: CARE, TRUST, and openness.
- Cross-disciplinary: Environment, Health Care, Food, Energy.
- Used cases for re-use: demonstrating how FAIR chemical data have been applied. Not only reusable but reused.

1.5.2 Research Data Alliance (RDA) P21 Plenary 'Birds of a Feather' session

Describing Chemical, Physical and Biological samples digitally: Seeking harmonization

26 October 2023. International Data Week. Salzburg, Austria. Organisers: Kerstin Lehnert, Alexander Prent, Leah McEwen, Fatima Mustafa, Debora Pignatari Drucker, Rolf Krahl, Lesley Wyborn, Ian Bruno, Iseult Lynch, Stuart Chalk.

Samples are taken in every field of study, but they vary widely in terms of type, e.g., single crystals, powder, complex structures, proteins and other biological (macro)molecules, cells, tissues, organisms, archeological artefacts, fossils, artwork, etc. Different fields may categorise samples from multiple perspectives simultaneously (e.g., nanomaterials are considered both physical particles and molecular entities, proteins are molecular entities of biological origin). Samples may consist of multiple components, in multiple phases; samples may represent collections of multiple entities, or single entities.

[13] https://www.gida-global.org/care
[14] https://www.nature.com/articles/s41597-020-0486-7

'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.

The sampling scheme is a critical aspect of designing any experiment to yield informative and reproducible results. A number of factors around sample collection, storage and processing are relevant for interpretation of measurement data derived from those samples. Different samples may be collected for different purposes: for example, biological specimens (or parts of specimens such as leaf for plants and tissue for animals), soil, and even air samples. Samples may be dependent on conditions of handling and storage (e.g., humidity, temperature), and may also be subject to further processing workflows (e.g., dispersion, mixing, plating, staining etc.). Samples may have spatial, temporal or other relationships that need to be articulated. A macro sample may be collected, with subsequent subsamples taken at increasing granularity down to the nanoscale, and multiple series of parent-child relationships need to be documented.

There are many well developed identifiers and other semantic descriptions used to describe different facets of sample provenance. A few cross-domain community endorsed examples include ISO 19156: 2013 Observations, Measurements and Samples and the W3C/OGC Semantic Sensor Network Ontology which includes the core SOSA (Sensor, Observation, Sample, and Actuator) Ontology for its elementary classes and properties. Other more domain specific approaches include iSamples and GBIF. How widely known and used are these existing cross domain ontologies and models? New international cross-domain ontologies are being published, as well as community driven ontologies, such as in earth sciences and biodiversity; how can we adapt these to be suitable for additional disciplines?

The ability to compile data from disparate disciplines will greatly facilitate the opportunity to answer broader, global challenges. Harmonisation of sample descriptions will also facilitate the workflow of instrument facilities that apply physical measurement techniques to extremely diverse sample types and need to meet a broad range of user needs for documentation. This session brings together a variety of disciplines including geochemistry, biodiversity, nanomaterials, analytical chemistry, and crystallography, among others, to explore approaches to harmonisation around sample description and provenance.

The expected outcomes of this discussion will be to:

- Compile a list of needs for describing sample types, origin, processing workflows and other requirements across disciplines
- Identify existing identifiers, classifications, ontologies and terminologies that support these descriptions
- Outline a proposal for an RDA Working Group project to develop best practices for sample data model specifications.

1.5.3 Pistoia Alliance - IUPAC joint workshop

Sustainable business models for open source standards development

'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.

14-15 November 2023. Pistoia User Group Meeting. Boston, USA. Organisers: Ian Bruno, Leah McEwen, Carmen Nitsche, Dana Vanderwall

In the broader world of open source software development, major tech firms both use AND contribute to projects, their revenue producing products or their own infrastructure rely on the software from many open source projects. New open source projects are commonly seeded with something developed at one company, put into the public domain to be enhanced . Their investment in these with both funding and commitment of people resources has a clear return on investment.

Many projects or collaborations in the chemistry, life sciences, or pharmaceutical domain have delivered useful output that solves or helps solve a common problem. However, few of the projects have been driven or even resourced by anyone for which the output is intrinsic to their "product", and the software development resources at many pharma companies are minimal, given the prevalence of commercial off the shelf software. Further, many of the projects solve very specific niche problems so are not often going to be regarded as business critical for a pharma company (for example).

Finally, many of these efforts are a project, with a distinct funding cycle and resource model. Consequently there is rarely enough critical mass of software developers to sustain an active open source development project or maintain the code produced anywhere near the level of something like a project at the Apache Foundation. These small efforts can subsequently struggle to maintain engagement, keep the resulting output up to date (technically or scientifically), and if they fall into obsolescence they no longer add value and slowly fall out of use.

The workshop will build on a business model canvas to foster interactive collaboration across multiple non-profit or volunteer projects/communities involved in developing software, standards or other content made available in the public domain. The objectives of the workshop are to work collaboratively to:

- identify opportunities for sharing resources or infrastructure;
- identify ideas for funding for shared projects.

2. WorldFAIR Chemistry public materials index

In this section we provide an index of the publicly available materials so far produced by the WorldFAIR Chemistry team.

2.1 WorldFAIR Deliverables

1. WorldFAIR Project (D3.1) Digital recommendations for Chemistry FAIR data policy and practice (https://doi.org/10.5281/zenodo.7887283)

'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.

2.2 Webinars

2.2.1 Webinar series “What is a chemical?”

- Summary (https://doi.org/10.5281/zenodo.7903683)

- “What is a chemical?”- Webinar 01: Handling Chemical Data Across Disciplines" (https://doi.org/10.5281/zenodo.7259102). Poll results and attended notes are included.
    a. Flyer (https://doi.org/10.5281/zenodo.7114395)
    b. Recording (https://www.youtube.com/watch?v=I9m-xkrWNPE)
    c. Speakers and Presentations

1	Emma Schymanski	Environmental Sciences (https://doi.org/10.5281/zenodo.7096404)
2	Iseult Lynch	Nanomaterials: Chemicals and Particles
3	Lesley Wyborn	What is a (Geo)Chemical Analysis: A High Level View (https://doi.org/10.5281/zenodo.7255025)
4	Kerstin Lehnert	Chemicals in Astromaterials
5	Marie-Lise Dubernet	Chemistry in astrophysics : Atoms and Molecules (Slides and Video) https://doi.org/10.5281/zenodo.7101252 https://doi.org/10.5281/zenodo.7101347

‘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.

6	Ken Kroenlein	Chemical Identification in Materials Informatics (https://doi.org/10.5281/zenodo.7255015)
7	Dylan Walsh	Community Resource for Innovation in Polymer Technology - Mission: Open-Source Polymer Data Ecosystem (https://doi.org/10.5281/zenodo.7255001)

- "What is a chemical?"- Webinar 02: Applying Chemical Data to Industry Challenges" (https://doi.org/10.5281/zenodo.7259728). Poll results and attended notes are included.
    a. Flyer (https://doi.org/10.5281/zenodo.7259509)
    b. Recording (https://www.youtube.com/watch?v=I3_ZZy0mR0Q&t=1943s)
    c. Speakers and Presentations

1	Lutz Weber	Using Large Scale Data & Chemical Ontologies to Predict Any Molecular Property (https://doi.org/10.5281/zenodo.7259334)
2	Teodoro Laino	From a Combination of Chemical Synthesis and Automation to Enzymatic Design: the Many Opportunities of Language Models in Chemistry
3	Yannick Djoumbou Feunang	Applying Chemical Data to Crop Protection Discovery (https://doi.org/10.5281/zenodo.7259454)
4	Nelson Vinueza Benitez	Colorants: Dyes, Pigments and Databases (https://doi.org/10.5281/zenodo.7259980)

'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.

5	Nick Lynch	FAIR Data in Life Sciences – Chemistry Focus (https://doi.org/10.5281/zenodo.7259464)
6	Gunther Schadow	Pharmaceutical-Regulatory Perspective (https://doi.org/10.5281/zenodo.7259474)

- "What is a chemical?"- Webinar 03: User Perspectives on Digital Machine-Readable Depictions" (https://doi.org/10.5281/zenodo.7435259). Attended notes are included.
    a. Flyer (https://doi.org/10.5281/zenodo.7435185)
    b. Recording (https://www.youtube.com/watch?v=7i7WdBbaEEs&list=PLNX2RnWLBzAU-DWHX6PUnKLsmr29mn59Q&index=3)
    c. Poll results (https://doi.org/10.5281/zenodo.7689655)
    d. Speakers and presentations

1	Greg Landrum	What is InChI? A user's perspective on identifiers and representations (https://doi.org/10.5281/zenodo.7435207)
2	Dana Vanderwall	HELM: Hierarchical Editing Language for Macromolecules (https://doi.org/10.5281/zenodo.7435223)
3	Jonathan Goodman	What is a chemical? Graphical Representation (https://doi.org/10.5281/zenodo.7435249)

'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.

4	Michelle Rogers	Systematic Nomenclature (https://doi.org/10.5281/zenodo.7435251)
5	Vincent Scalfani	What are SMILES? (https://doi.org/10.5281/zenodo.7435254)

- "What is a chemical?"- Webinar 04: Innovation in Chemical Descriptions" (https://doi.org/10.5281/zenodo.7683138)
    a. Flyer (https://doi.org/10.5281/zenodo.7683142)
    b. Recording (https://www.youtube.com/watch?v=wsURb99NV_0&list=PLNX2RnWLBzAU-DWHX6PUnKLsmr29mn59Q&index=4)
    c. Speakers and presentations

1	Henry Rzepa	Media types: Interesting Enhancers of the F. A. I. and R. of FAIR (https://doi.org/jvk9)
2	Adnan Malik	ChEBI: Chemical Entities of Biological Interest (https://doi.org/10.5281/zenodo.7683072)
3	Oliver Koepler	Ontologies4Chem (https://doi.org/10.5281/zenodo.7683092)
4	Alex Clark	Mixtures-Molecules (https://doi.org/10.5281/zenodo.7683101)
5	Ken Kroenlein	Complex Substance Schemas: A Case Study (https://doi.org/10.5281/zenodo.7683116)

'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.

2.2.2 Webinar titled "What is digital IUPAC?"

- Bruno, Ian, McEwen, Leah, & Mustafa, Fatima. (2023, May 31). Webinar: What is digital IUPAC? Tools for an increasingly digital research culture. Zenodo.
- https://doi.org/10.5281/zenodo.7992804
- Hosted by ChemVoices on May 31, 2023.

2.3 Workshops

1. Advancing FAIR Chemistry: Developing New Services for Sharing Chemical Data (https://doi.org/10.5281/zenodo.7903727)
    a. Flyer (https://doi.org/10.5281/zenodo.7803680)
    b. Agenda (https://doi.org/10.5281/zenodo.7803789)
    c. Speakers and Presentations

1	Leah McEwen	Guidance on FAIR Chemical Data Reporting (https://doi.org/10.5281/zenodo.7803831)
2	Ian Bruno	Guidance on FAIR Chemical Data Reporting (https://doi.org/10.5281/zenodo.7803831)
3	Stuart Chalk	Digital recipes for managing chemical data (Cookbook) (https://doi.org/10.5281/zenodo.7803867)
4	Evan Bolton	Standardized programmatic access to chemical information (Protocol Services) (https://doi.org/10.5281/zenodo.7803871)
5	Robert Hanson	FAIRSpec
6	Vincent Scalfani	Doc-a-thon: Chemical representation best practices for humans and machines (https://doi.org/10.5281/zenodo.7803914)

'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.

7	Fatima Mustafa	Advancing FAIR Chemistry: Developing New Services for Sharing Chemical Data: Program (https://doi.org/10.5281/zenodo.7803795)

2. RDA P20: CRDIG session on "Describing diverse chemistry datasets across distributed data resources".

1	Chemistry Research Data Interest Group (CRDIG)	Leah McEwen	Advancing FAIR with data standards for chemical description (https://doi.org/10.5281/zenodo.7789071)
2	Data representation in materials and chemicals based on harmonized domain ontologies	Stuart Chalk	The IUPAC Gold Book: A Compendium of Chemical Terminology (https://doi.org/10.5281/zenodo.7789101)
3	The Way to FAIR: from data collection to citation	Stuart Chalk	Chemical Analysis Use Case (https://doi.org/10.5281/zenodo.7789085)
4	Interoperable Descriptions of Observable Property Terminology WG (I-ADOPT WG)	Stuart Chalk	iAdopt in SciData (https://doi.org/10.5281/zenodo.7789111)

'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.

5	The Cross-Domain Interoperability Framework (CDIF) Workshop	NA	NA
6	Chemistry Research Data Interest Group (CRDIG) Session	Leah McEwen and Ian Bruno	Organised by Ian Bruno

2.4 Invited presentations

1. **Presentation** titled "WorldFAIR Chemistry: FAIR-enabling resources" (https://doi.org/10.5281/zenodo.7262368). Presented at the FAIR Convergence Symposium, 24-26 October, Leiden 2022.
2. **Presentation** titled " The IUPAC Gold Book: Compendium of Chemical Terminology" (https://doi.org/10.5281/zenodo.7331968). Presented at the Australian Vocabulary Symposium, 14-15 November 2022.

'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.