

C

A

P

U

I



WorldFAIR

Guidance on FAIR Chemical Data Reporting

Leah McEwen, Cornell University

Ian Bruno, Cambridge Crystallographic Data Centre

WorldFAIR Chemistry

Committee on Publications & Cheminformatics Data Standards

ACS National Meeting, 27 March 2023

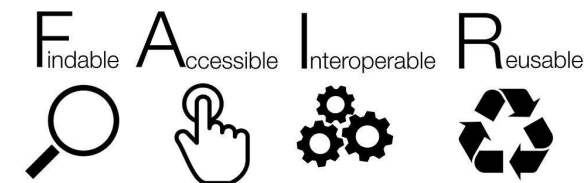
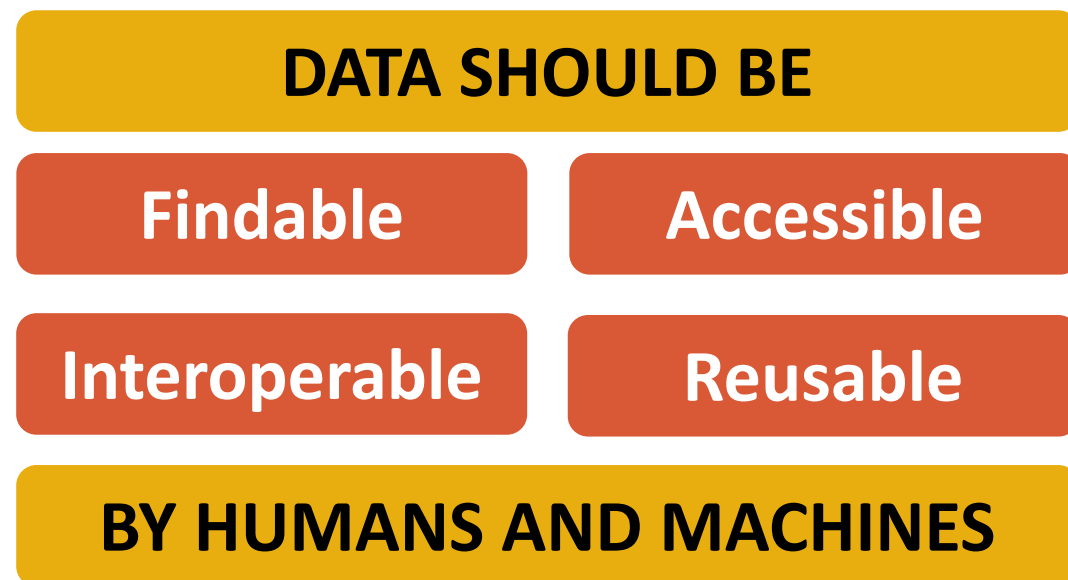
The FAIR Data Principles

Comment | [OPEN](#)

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier [...] Barend Mons 

Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).



“FULLY AI READY”

The machine knows what I mean



Image credit: Australian National Data Service, CC-BY 4.0
<https://www.ands.org.au/working-with-data/fairdata/training>

IUPAC World FAIR Chemistry

Making IUPAC Assets FAIR



Reporting Guidance

Develop Guidance Protocols
for Handling Chemical Data

FAIRly



Training Cookbook

Develop online materials on
how to manage digital data
files and content



Protocol Services

Develop web-based services to
confirm chemical identity &
machine-readability of
chemical data

Types of Guidance

Framing: Over-arching goals

- Addressing global sustainability challenges
- Providing global economic benefits
- Increased transparency and reproducibility
- Enabling inter-disciplinary collaboration
- Universal equitable access to knowledge



Types of Guidance

Perspectives: Desired end state

- Community standards for data and information
- Education in best practices
- Tools and infrastructure
- Mandatory data deposition
- Publication of failed experiments
- ELNs as enablers
- High-throughput computational platforms
- Consistency in policies and practice
- Liberate data – make it FAIR

Issue 3, 2022 Previous Article Next Article

From the journal:
Digital Discovery

Data management matters Check for updates

Cerys Willoughby *^a and Jeremy Graham Frey ^a

Israel Journal of Chemistry

Official Journal of the Israel Chemical Society

Essay | Open Access

The Long and Winding Road towards FAIR Data as an Integral Component of the Computational Modelling and Dissemination of Chemistry

Henry S. Rzepa

First published: 07 June 2021 | <https://doi.org/10.1002/ijch.202100034>

Perspective | Published: 04 April 2022

Making the collective knowledge of chemistry open and machine actionable

[Kevin Maik Jablonka](#), [Luc Patiny](#) & [Berend S.](#)

Nature Chemistry **14**, 365–376 (2022) | [Cite](#)

14k Accesses | 8 Citations | 119 Altmetric

Abstract

Large amounts of data are generated in yet a considerable proportion is also ca

Angewandte Chemie International Edition / Volume 61, Issue 51 / e202203038
Viewpoint Article

Minimum Information Standards in Chemistry: A Call for Better Research Data Management Practices

Prof. Dr. Sonja Herres-Pawlis , Dr. Felix Bach, Dr. Ian J. Bruno, Dr. Stuart J. Chalk, Dr. Nicole Jung, Dr. Johannes C. Liermann, Leah R. McEwen, Dr. Steffen Neumann ... [See all authors](#)

First published: 08 November 2022
<https://doi.org/10.1002/anie.202203038>

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY Get e-Alerts

Into the Unknown: How Computation Can Help Explore Uncharted Material Space

Austin M. Mroz, Victor Posligua, Andrew Tarzia, Emma H. Wolpert, and Kim E. Jelfs*

Cite this: *J. Am. Chem. Soc.* 2022, 144, 41, 18730–18743
Publication Date: October 7, 2022
<https://doi.org/10.1021/jacs.2c06833>

Copyright © 2022 The Authors. Published by American Chemical Society
[RIGHTS & PERMISSIONS](#)

View PDF

Types of Guidance

Practical: Specific and targeted

- Professional Societies
- Scientific Unions
- Specific Communities
- Individual Researchers
- Institutional Policies

Research Data Management @Harvard

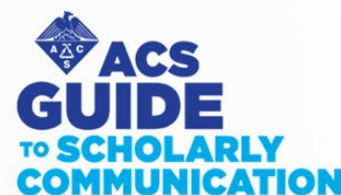
HOME /

Data Policies

There are a number of policies and regulations that apply to Harvard researchers working with data. Below are some commonly applicable internal and external

Harvard Policies

- [Data Ownership](#): Applies to research data resulting from the auspices of the University, or with University resources
- [Data Use Agreements](#): Policy and Guidance documents



3.1 Data Sharing ^

- 3.1.1 Why Share Data?
- 3.1.2 Reporting Data in the Digital Environment
- 3.1.3 FAIR Data Principles
- 3.1.4 Research Data Reporting in Chemistry
- 3.1.5 Preparing Your Data for Publication
- 3.1.6 Data Management
- 3.1.7 The Future of Data Sharing

3.2 Chemical Structures in the Google

3.3 Digital Chemical Data v

Open Data FAQs for chemists



The following FAQs have been asked by members of the Department of Chemistry and answered by members of the Open Data team at the University.

If you have any amendments or further questions you would like to ask please contact the Librarian at the Department of Chemistry, Clair Castle, at library@ch.cam.ac.uk, in the first instance.

The Open Data team can also be contacted at info@data.cam.ac.uk, <http://www.data.cam.ac.uk/>.

FAQs

What would open data for a typical synthetic organic chemistry paper look like?

For a synthetic paper you might include the output files from NMR, UV/Vis, and IR measurements (for example). These should be in a format that others can use, so the data should be in a form which can be manipulated (images of graphs, especially of NMR experiments, wouldn't meet this criteria). So for an NMR measurement you should include the processed data as .csv file (for example) so that future users can replot the data for themselves.

Lab books form an important record of the experiments, so where possible they should also be included in the data record (or at least the detailed methodology for the experiment, so that it can be reproduced).

However, if it would be too time consuming and costly to digitise the lab books then you can simply create a meta-data record on the repository so that future users can contact you to physically access your lab books.

What would open data for a typical molecular dynamics based paper look like?

Pure Appl. Chem., Vol. 80, No. 2, pp. 277–410, 2008.
doi:10.1351/pac200880020277
© 2008 IUPAC

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY
CHEMICAL NOMENCLATURE AND STRUCTURE REPRESENTATION DIVISION*

GRAPHICAL REPRESENTATION STANDARDS FOR CHEMICAL STRUCTURE DIAGRAMS**

(IUPAC Recommendations 2008)

Prepared for publication by
JONATHAN BRECHER

CambridgeSoft Corporation, 100 CambridgePark Drive, Cambridge, MA 02140, USA

Types of Guidance

Policies: Enforcing community norms

- Regulatory Guidance
- Government Policies
- Funder Policies
- Journal Policies

Wiley's Data Sharing Policies

	Data availability statement is published ¹	Data has been shared ²	Data has been peer reviewed ³	Example Wiley journals
Encourages Data Sharing	Optional	Optional	Optional	
Expects Data Sharing	Required	Optional	Optional	British Journal of Social Psychology
Mandates Data Sharing	Required	Required	Optional	Ecology and Evolution
Mandates Data Sharing and Peer Reviews Data	Required	Required	Required	Geoscience Data Journal American Journal of



Data sharing



Why is data sharing important?

What is research data?

Our data sharing policy

Recommended repositories

Choosing a repository

...t specific repositories

al repositories

ility Statements

ftware citation

data

software and code

NIH SCIENTIFIC DATA SHARING

DATA MANAGEMENT AND SHARING POLICY GENOMIC DATA SHARING POLICY OTHER

Home > Data Management and Sharing Policy > About Data Management & Sharing Policies > Data Management & Sharing Policy Overview

Data Management & Sharing Policy Overview

Learn what is expected of investigators and institutions under the 2003 NIH Data Sharing Policy and the 2023 NIH Data Management and Sharing (DMS) Policy.

Applications for Receipt Dates BEFORE Jan 25 2023 Applications for Receipt Dates ON/AFTER Jan 25 2023

NIH has issued the Data Management and Sharing (DMS) policy¹ effective January 25, 2023 to promote data accelerates biomedical research discovery, in part, by enabling validation of research results, promoting data reuse for future research studies.

Under the DMS policy, NIH expects that investigators and institutions:

- Plan and budget for the managing and sharing of data
- Submit a DMS plan for review when applying for funding
- Comply with the approved DMS plan

Individual NIH Institutes, Centers, or Offices may have additional policies and expectations (see NIH Data Management and Sharing Policy Overview Page).

Download a simplified version of the Data Management and Sharing Policy Overview Page.

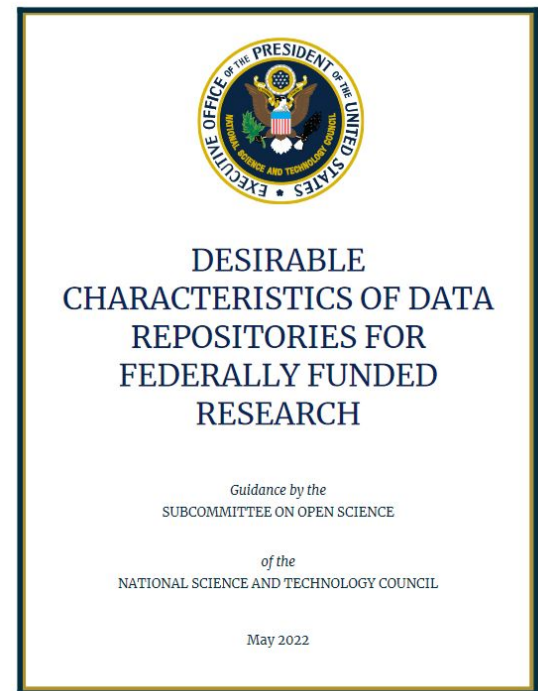
Select each step below to learn more.

PLANNING & BUDGETING SUBMISSION & REVIEW

Planning & Budgeting for Data Management & Sharing

Prospectively planning for how scientific data will be managed and shared is the first step in optimizing the reach of data generated from NIH-funded research. Institutions are encouraged to consider these crucial elements early in research planning.

Tip: Consider consulting institutional resources such as librarians and data managers to help plan effectively!



In Horizon Europe, *beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the **FAIR principles**, and should at least do the following:*

- Prepare a Data Management Plan (DMP) and keep it updated throughout the course of the project
- Deposit data in a trusted repository and provide open access to it ('as open as possible, as closed as necessary')
- Provide information (via the same repository) about any research output or any other tools and instruments needed to re-use or validate the data



Types of Guidance

Framing: Over-arching goals

Perspectives: Desired end state

Practical: Specific and targeted

Policies: Enforcing community norms

IG

Data policy standardisation and implementation IG

Taxonomy:



Developing a Research Data Policy Framework for All Journals and Publishers

Authors: Iain Hrynaszkiewicz [✉](#), Natasha Simons, Azhar Hussain, Rebecca Grant, Simon Goudie

Framework recognises that different journals and different domains may be at different levels of maturity

Different requirements (or features), varying different degrees of mandating ()

Progression from requiring information to be provided to checking and enforcing

14 journal research data policy features arranged as 6 policy types (tiers)

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exceptions to policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Embargoes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Supplementary materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data repositories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data licensing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Researcher/ author support	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data availability statements		<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data formats and standards				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (specific data types)				<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (all papers)				<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Peer review of data				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Data Management Plans (DMPs)				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Crystallography Journey

1940s

Data published in tables in journal articles

1960s

Data abstracted into electronic databases

1990s

Data standards (CIF) established

2000s

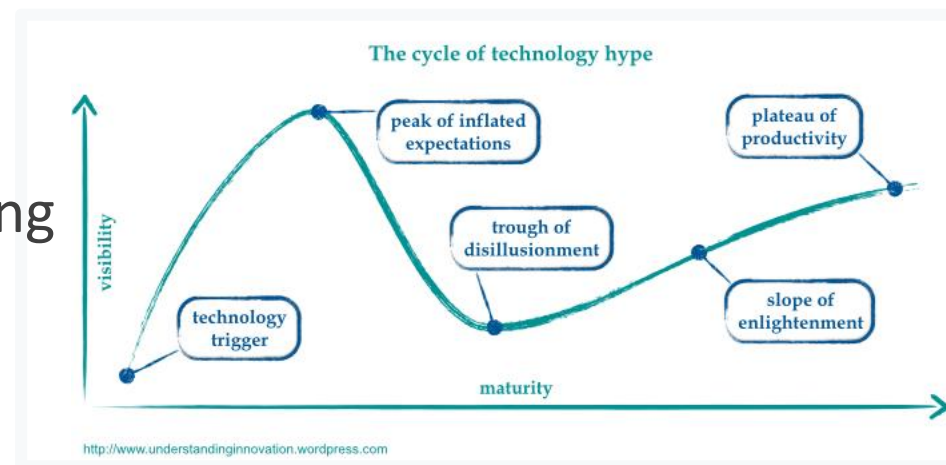
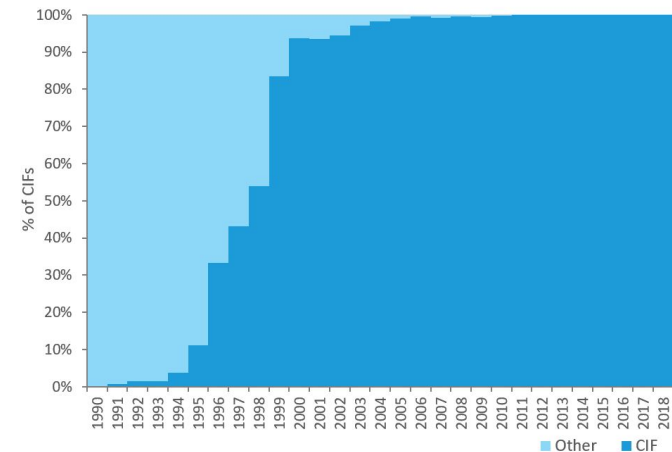
Cultivation of data validation and publication workflows

2010s

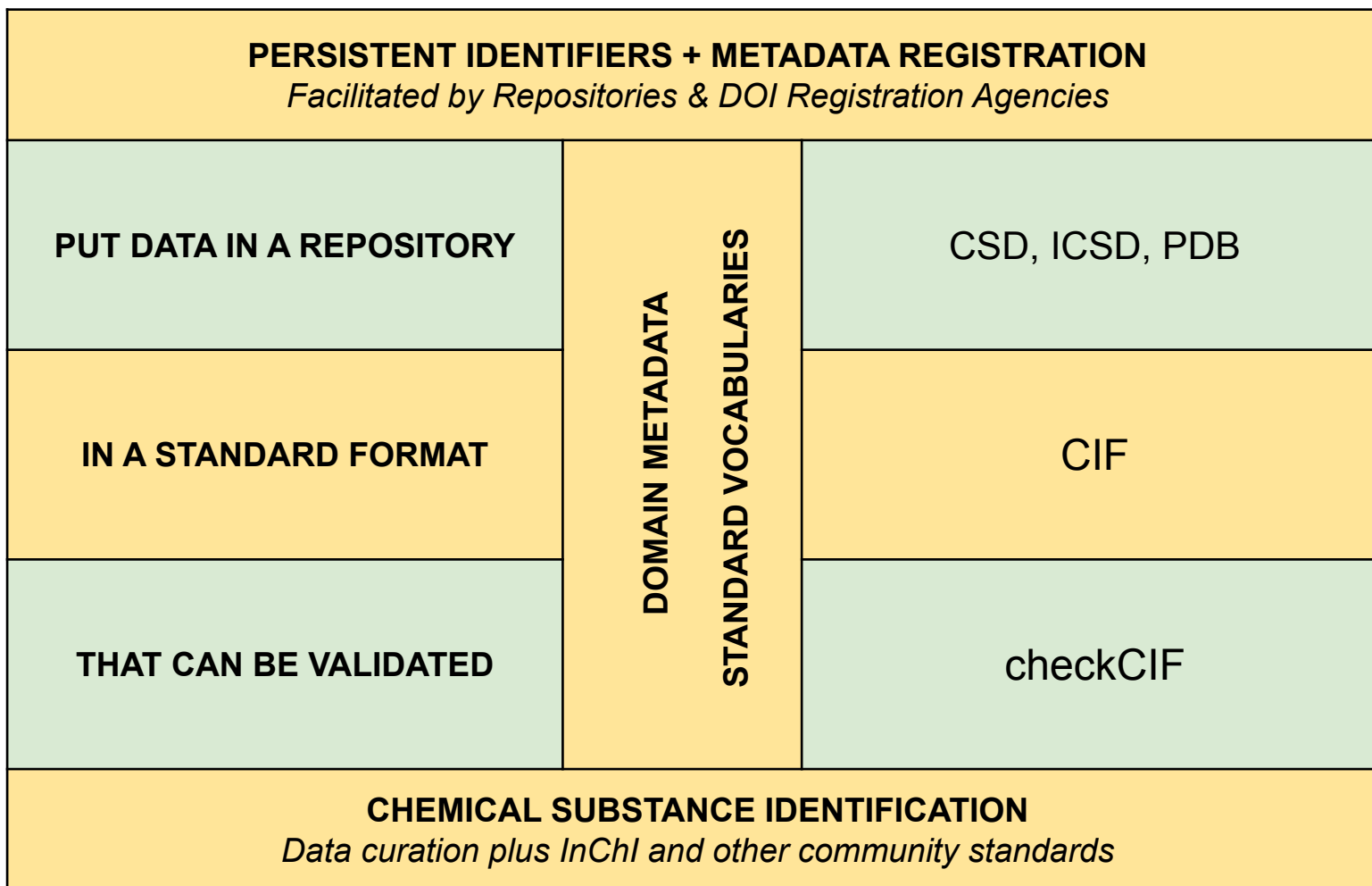
Deposition in data repository pre-publication the default

Crystallography Journey

- Initially CIFs were embedded in PDFs and Word documents
- Lasting uptake of CIF and digital deposition catalysed by:
 - Leadership by IUCr – principles and practice
 - Adoption of CIF by software tools
 - Validation services
 - Joined up repository/publisher workflows
 - Aggregation of data enabling reuse for generating new knowledge



FAIR Data Implementation in Crystallography



Stakeholders

Researchers
 Institutions
 Funders

Publishers
 Editors
 Reviewers

Repositories
 Tool Providers
 Instrument Manufacturers

Scientific Unions
 Professional Societies

What has adopting FAIR enabled for crystallography?

- Validation services checking for consistency and completeness
- Extension of deposition services from organic to inorganic small molecules
- Aggregation and linking of data across general information resources

Satisfied by established community practices and workflows

Enabled by adoption of established CIF, DOIs

- Chemistry-based discovery, reuse and linking

Enabled as a result of further enrichment by repository data curators

A **chemical crystallography** dataset has FAIR **crystallographic** attributes when deposited – it rarely has FAIR **chemical** attributes.

Describing Diverse Chemistry Datasets Across Distributed Data Resources

RDA Plenary 20, Gothenburg, Sweden, 23 March 2023

Key themes and ideas to emerge:

- Minimal information standards – what is enough for future reuse?
- Data born FAIR (future smart labs) vs made FAIR when published
- Importance of sample – common challenge across domains
- Capturing provenance and process – enabling reproducibility
- Characterising analytical techniques – cataloguing existing standards
- Use cases to drive the need for discovery and interoperability across resources



Points of integration

Chemical substance: integration by chemical identification

↳ *standard chemical identifier*

Chemical property: integration of property values ↳ *standard property terms*

Measurement: integration by technique, by conditions ↳ *standard definitions*

Units: integration of quantities ↳ *standard units of measure*

Material sample: integration by composition, state of matter, space group

↳ *standard classifications/descriptions*

Origin of sample: integration by location, source (e.g., species), named reactions

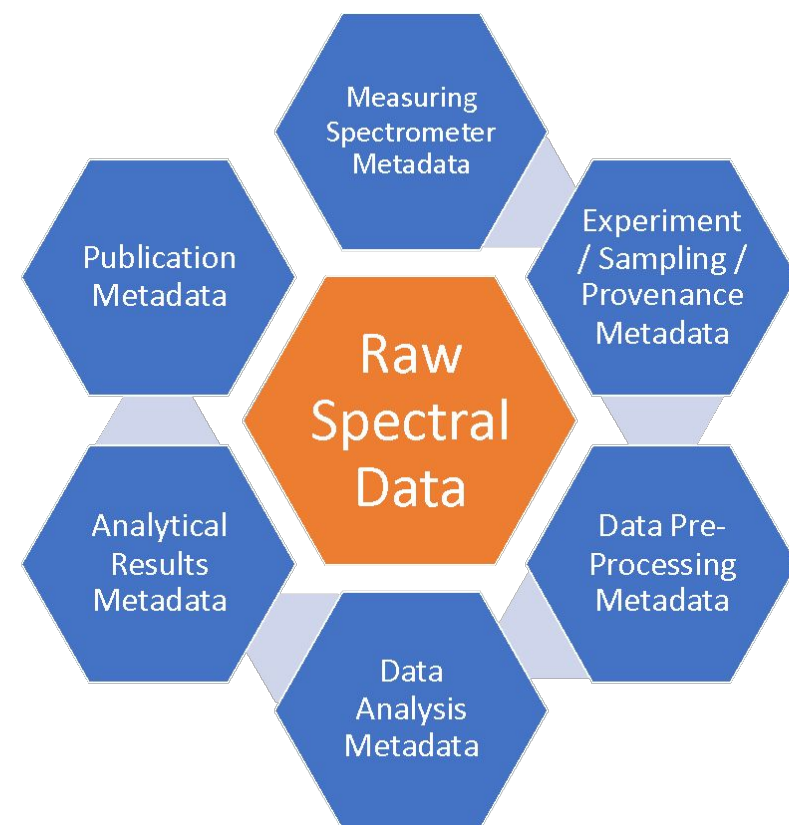
↳ *standard location metadata, species classification, reaction classification*

Origin of measurement: integration by analyst or lab, by instrument

↳ *PIDs: ORCID, ROR, etc.*

Temporal: integration by date of sample collection, date of measurement

↳ *standard date format*



Adapted from A. Davies, IUPAC FAIRSpec

Adapted from K. Lehnert, OneGeochemistry

FAIR Chemistry Data – General Concepts

FINDABILITY

Generic object identifiers

DOIs for example

Metadata Registries

With high-level chemistry metadata

E.g. InChIs in DataCite

Data Repositories

Ideally chemistry-aware

ACCESSIBILITY

Open Protocols

With chemistry-aware APIs

Authentication and Authorisation where needed

REUSABILITY

Community Licences

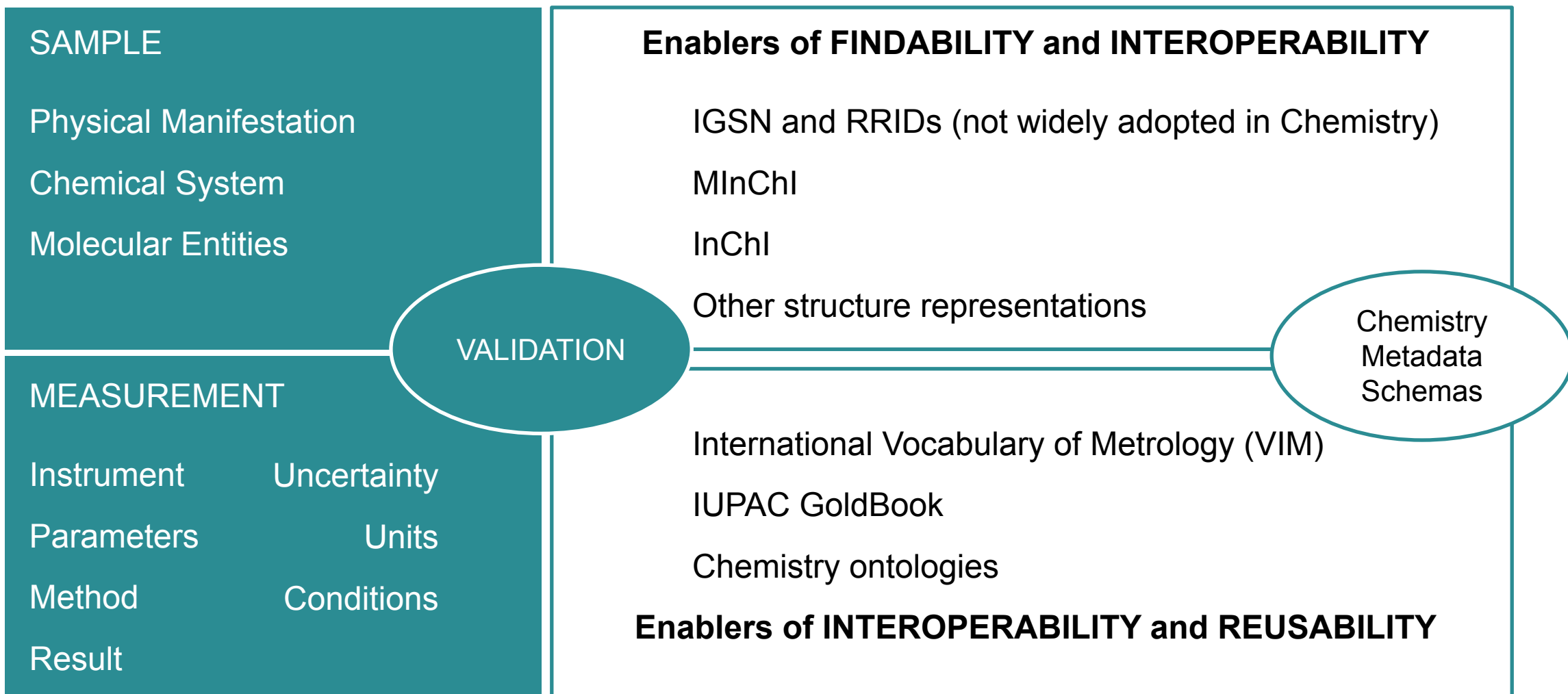
Ideally CC0 or CC-BY

Non-derivative (ND) for reference data

INTEROPERABILITY

Cross-domain Interoperability Framework (CDIF): A WorldFAIR Focus

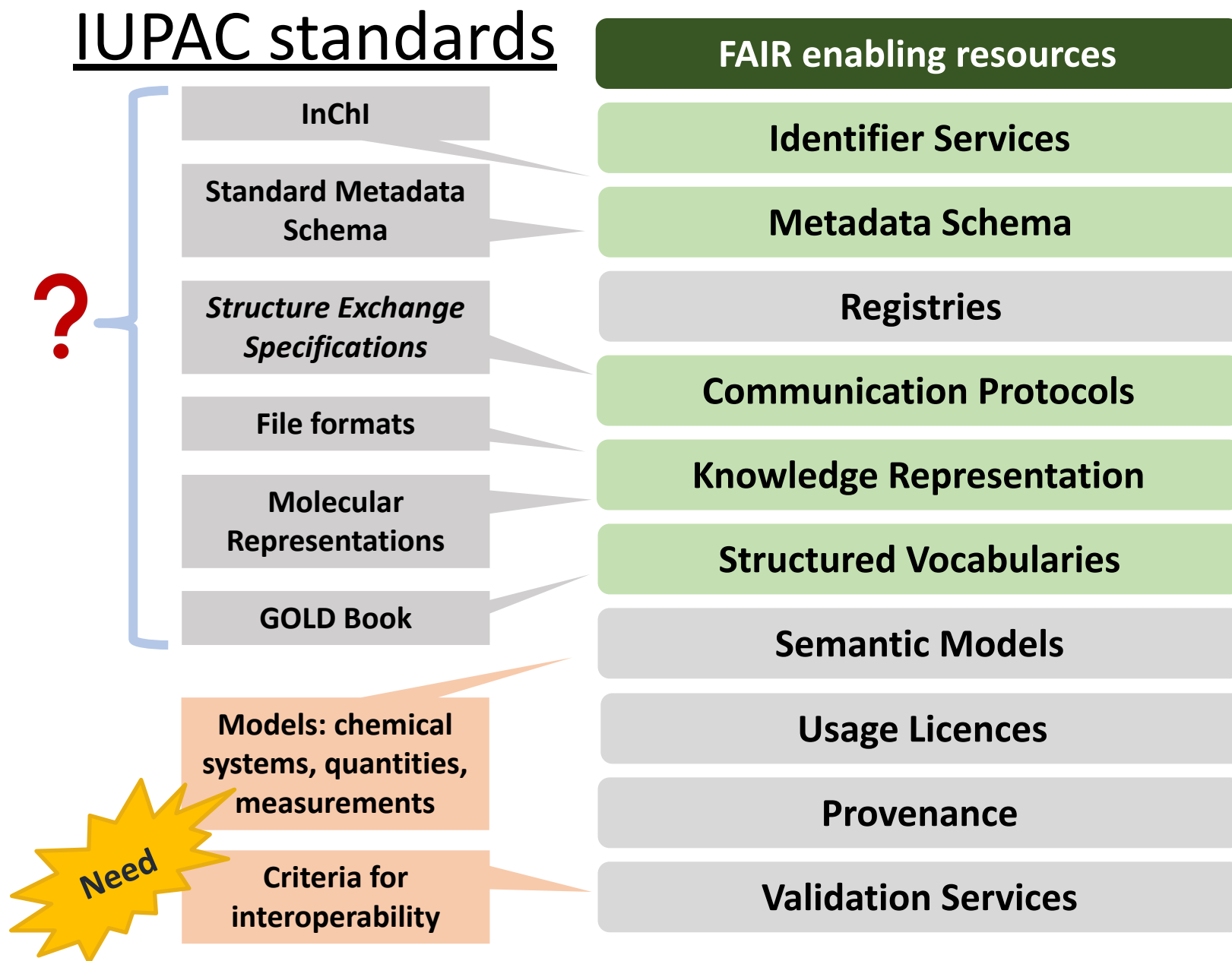
FAIR Chemistry Data – Key Concepts



Well defined chemical data are broadly reusable

RIPE for sharing	Chemical data	Standard definitions (examples)
Reliable information for samples & measurements	Samples: identity of substance(s), sample description (provenance, purity, state) Measurements: techniques, conditions, calibrations, uncertainties	nomenclature (Blue/Red/Purple books), graphical representation, InChI Terminology for analytical chemistry (Orange book), metrology (VIM)
Interpretable scientific expression	Results: quantities, units, calculations, dependencies, processing/derivation	Notations, symbols, terminology for physical chemistry (Green book)
Processable formatted for machines	File formats, validation Referrable terms, ontologies Data models, metadata schema	SDF, CIF, ThermoML, JCAMP-DX, mzML Gold Book, CHMO, RXNO, ChEBI FAIRSpec, <i>Solubility</i> , <i>Periodic Table</i>
Exchangeable metadata online	Registered metadata for indexing chemicals Standardized exchange APIs for chemicals	InChIs, standard terms/notations <i>Chemical structure API specification</i>

Are these digital standards FAIR for programmatic access and reuse?



FAIR for machines



Persistent Identifiers



Rich Metadata

Data Repositories



Standard Open Protocols

Knowledge Representation



FAIR Vocabularies

Linked Data



Usage Licences



Provenance

Community Standards

IUPAC standards

InChI

Standard Metadata
Schema*Structure Exchange
Specifications*

File formats

Molecular
Representations

GOLD Book

Models: chemical
systems, quantites,
measurementsCriteria for
interoperability

Need

FAIR enabling resources

Identifier Services

Metadata Schema

Registries

Communication Protocols

Knowledge Representation

Structured Vocabularies

Semantic Models

Usage Licences

Provenance

Validation Services

**WE ARE
FAIR
ENABLERS**



**PIDs &
registered
metadata**

- standard chemistry descriptors
- key metadata



**Domain
repositories**

- standard chemistry APIs
- authentication and authorization
- standard chemistry APIs (e.g., instrument to ELN)
- facilitate deposit
- data preparation checklist
- select repository & upload



**Open standard
formats**

- standard formats, terminology, ontologies
- metadata relationships
- standard descriptors in native formats
- link data/metadata
- how-to support for using file formats
- metadata templates
- assemble data files
- document which formats used



Verified, licensed

- standardized validation
- transparent licensing
- metadata extraction
- data review
- process guide
- prepare ReadMe

Repositories

**Software
(tools)**

**Support
services**

Researchers

What Next for Guidance

- **Where do you or your researchers feel all at sea?**
For managing? For sharing? For validation? For reuse?
- **What new chemistry guidance would help you in your role?**
- **What existing chemistry guidance are you aware of?**
- **Are there specific communities we should be engaging with?**
Engaged: IUPAC, RDA, NFDI4Chem, PSDI, Geochemists, Nano – who else?
E.g. Data Curation Network? Library communities? Industry groups?



WorldFAIR

Guidance on FAIR Chemical Data Reporting

WorldFAIR Chemistry

ACS National Meeting, 27 March 2023



Contact & Follow US

iupac.org@FAIRChemistry
@iupactinyurl.com/iupacchanFAIRChemistry@iupac.org

@iupac.org



FAIRChemistry Community



Guidance on FAIR Chemical Data Reporting

WorldFAIR Chemistry

ACS National Meeting, 27 March 2023



Discussion

Shared notes link:

<https://bit.ly/ACSChemNotes>

