The IUPAC FAIRSpec Project
FAIR Management of Spectroscopic Data in Chemistry

Bob Hanson

IUPAC Project 2019-031-1-024

ChemSpider 2025 Webinar 2: Data standards in chemistry
Oct. 21, 2025

Hey, AI!
What is *FAIR Data*?

Hey, AI!
What is *FAIR data*?

AI Overview

FAIR data are **digital assets that follow the Findable, Accessible, Interoperable, and Reusable principles, which act as guidelines for managing and sharing data to increase its utility and potential for reuse**. These principles aim to ensure data is not only available to humans but also easily processed by machines, facilitating scientific discovery, reproducibility, and transparency.

The four FAIR principles

Findable:

Data should be easy to locate through standardized metadata, a unique and persistent identifier (like a DOI), and registration in a searchable index.

Accessible:

Data access should be possible through clear protocols and standardized APIs, with open access where possible.

Interoperable:

Data and metadata should use common formats and ontologies to allow them to be used and combined with other data and systems.

Reusable:

Data should be well-described to ensure it can be reused in different contexts. This includes clear usage licenses.

Our approach:

Our approach:

Hey, AI!
What is FAIR data *management*?

Hey, AI!
What is FAIR data *management*?

AI Overview

FAIR data management is **a framework of principles—Findable, Accessible, Interoperable, and Reusable—for managing and sharing data to make it more effective for both humans and computers**. It ensures that data has unique identifiers, is described with sufficient metadata, is available in trusted repositories, uses common standards, and has clear usage licenses and provenance, maximizing its long-term value and impact.

Findable

- **Goal:** Make sure data can be discovered by others.
- **Methods:**
  - Assign a unique and persistent identifier (PID) to the data.
  - Describe the data with rich, searchable metadata.
  - Register the data in a searchable resource, like a data repository.

Hey, AI!
What is FAIR data *management*?

AI Overview

FAIR data management is **a framework of principles—Findable, Accessible, Interoperable, and Reusable—for managing and sharing data to make it more effective for both humans and computers**. It ensures that data has unique identifiers, is described with sufficient metadata, is available in trusted repositories, uses common standards, and has clear usage licenses and provenance, maximizing its long-term value and impact.

Findable

- **Goal:** Make sure data can be discovered by others.
- **Methods:**
  - Assign a unique and persistent identifier (PID) to the data.
  - Describe the data with rich, searchable metadata.
  - Register the data in a searchable resource, like a data repository.

Hey, AI!
What is FAIR data *management*?

Accessible

- **Goal:** Ensure data can be accessed by authorized users.
- **Methods:**
  - Make metadata and data retrievable via their PID using standard protocols, like HTTPS.
  - Deposit data in a trusted repository with long-term preservation goals.
  - Use clear usage licenses and access agreements.

Interoperable

- **Goal:** Enable data to be integrated with other data.
- **Methods:**
  - Use common data formats and standards.
  - Use recognized vocabularies, ontologies, or controlled terminologies.
  - Ensure metadata is in a common structure to allow for comparison.

Reusable

- **Goal:** Allow data to be reused by others in the future.
- **Methods:**
  - Provide clear usage licenses that state how the data can be reused.
  - Include rich documentation, like a README file, to help users understand and interpret the data.

Hey, AI!
What is FAIR data *management*?

Accessible

- **Goal:** Ensure data can be accessed by authorized users.
- **Methods:**
  - Make metadata and data retrievable via their PID using standard protocols, like HTTPS.
  - Deposit data in a trusted repository with long-term preservation goals.
  - Use clear usage licenses and access agreements.

Interoperable

- **Goal:** Enable data to be integrated with other data.
- **Methods:**
  - Use common data formats and standards.
  - Use recognized vocabularies, ontologies, or controlled terminologies.
  - Ensure metadata is in a common structure to allow for comparison.

Reusable

- **Goal:** Allow data to be reused by others in the future.
- **Methods:**
  - Provide clear usage licenses that state how the data can be reused.
  - Include rich documentation, like a README file, to help users understand and interpret the data.

Hey, AI!
What is *FAIRSpec*?

FAIRSpec is the name for two separate initiatives:

- An International Union of Pure and Applied Chemistry (IUPAC) project creating standards for managing chemical spectroscopic data.
- A Swiss-based performing arts group that promotes ethical work practices.

Hey, AI!
What is *IUPAC FAIRSpec*?

AI Overview

IUPAC FAIRSpec is **a set of guiding principles and standards from the International Union of Pure and Applied Chemistry (IUPAC) for the Findable, Accessible, Interoperable, and Reusable (FAIR) management of spectroscopic data in chemistry**. It aims to create standardized digital objects and a data model for describing the contents of spectroscopic data collections, which helps ensure that data can be easily found, shared, and reused by both humans and machines over time.

Key components and goals

- **Guiding principles:** Establishes rules for managing data throughout the research process, not just at publication.
- **Data model:** Defines the necessary digital objects and metadata needed to describe a spectroscopic dataset, including spectral data, chemical structures, samples, and analyses.
- **FAIR Data Collection:** A collection of digital objects that is organized according to the IUPAC FAIRSpec recommendations and is accompanied by a machine-actionable "FAIRData Finding Aid" that describes the collection.

Guiding Principles for the FAIR Management of Spectroscopic Data

IUPAC Specification for the FAIR Management of Spectroscopic Data in Chemistry (IUPAC FAIRSpec) - Guiding Principles

Robert M. Hanson, Damien Jeannerat, Mark Archibald, Ian Bruno, Stuart J. Chalk, Antony N. Davies, Robert J. Lancashire, Jeffrey Lang and Henry S. Rzepa

Pure and Applied Chemistry, 2022

https://doi.org/10.1515/pac-2021-2009

1. FAIR Management of data should be an ongoing concern.

A. FAIR management of data must be an explicit part of research culture.
B. FAIR management of data should be of intrinsic value.
C. Good data management requires distributed curation.
D. Experimental work is by nature iterative.

2. Context is important.

A. Digital objects are generally part of a collection.
B. Chemical properties are related to chemical structure.
C. Data relationships are diverse and develop over time.
D. FAIR management of data should allow for validation.

3. FAIR management of data requires curation

A. Data reuse relies upon practical findability.
B. Data has to be organized to be accessible.
C. Data interoperability requires well-designed metadata.
D. Value is in the eye of the reuser.

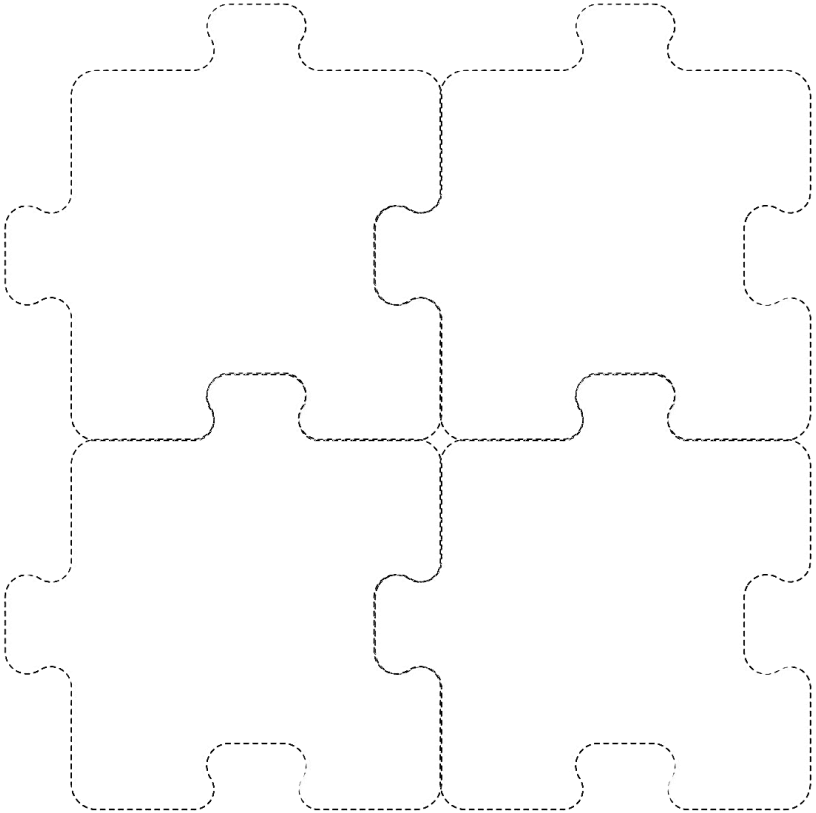
4. Metadata must be standardized and registered.

A. Register key metadata.
B. Assign a variety of persistent identifiers.
C. Enable metadata crosswalks.
D. Allow for value-added benefits.

5. FAIR data management standards should be *modular, extensible, and flexible*

A. Modularity allows specialization.
B. Allow for future needs.
C. Respect format and implementation diversity.
D. All data formats should be valued.

The data management puzzle

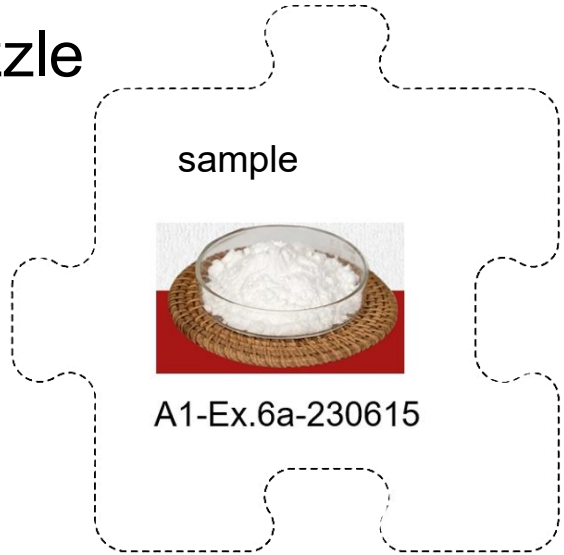


7	1 5 8 9	5 6 8	1 2 6 9	1 3 5 6 9	2 3 5 6 9	4	1 2 5	1 2 3
1 4 5 6 9	2	4 5 6	1 4 6 9	7	3 4 5 6 9	1 5 6	8	1 3
1 4 5 6	1 4 5	3	1 2 4 6	1 5 6	8	1 2 5 6 7	1 2 5 7	9
2 4 8 9	4 7 8 9	2 4 7 8	5	1 6 8 9	4 6 9	3	1 2 4 7	1 2 4 7 8
3 4 5 8	6	4 5 7 8	1 4 8	2	3 4	1 5 7 8	9	1 4 7 8
2 3 4 5 8 9	4 5 8 9	1	4 8 9	3 8 9	7	2 5 8	2 4 5	6
1 2 5 6 8	1 5 7 8	2 5 6 7 8	3	5 6 8	2 5 6	9	1 2 4 7	1 2 4 7 8
1 2 5 8	3	2 5 7 8	2 7 8 9	4	2 5 9	1 2 7 8	6	1 2 7 8
2 4 6 8	4 7 8	9	2 6 7 8	6 8	1	2 7 8	3	5

The data management puzzle

We start with a sample.

It has various characteristics, most importantly an ***identifier***

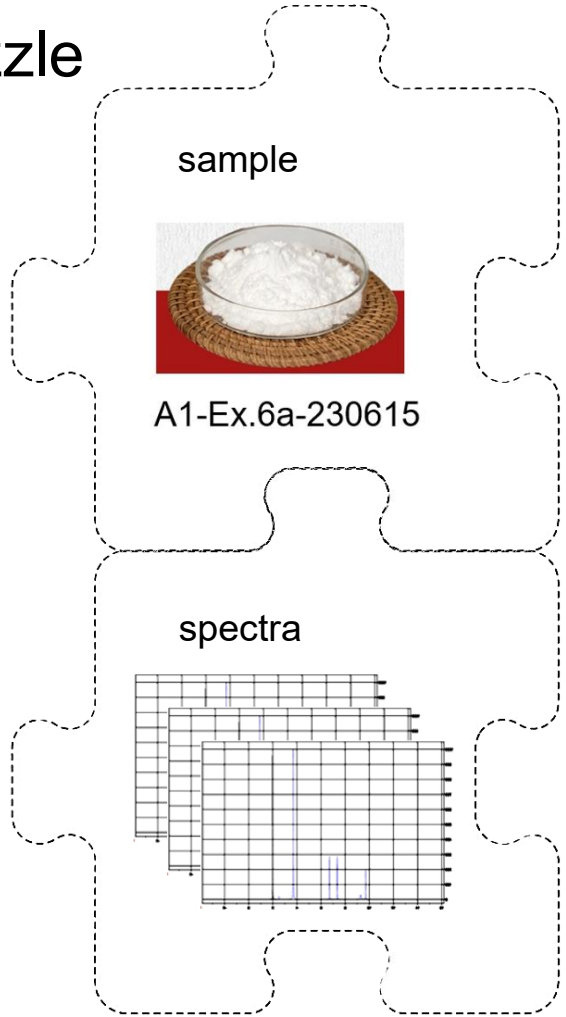


The data management puzzle

We run tests on it, creating instrument datasets and reports.

Now we have a ***data collection***.

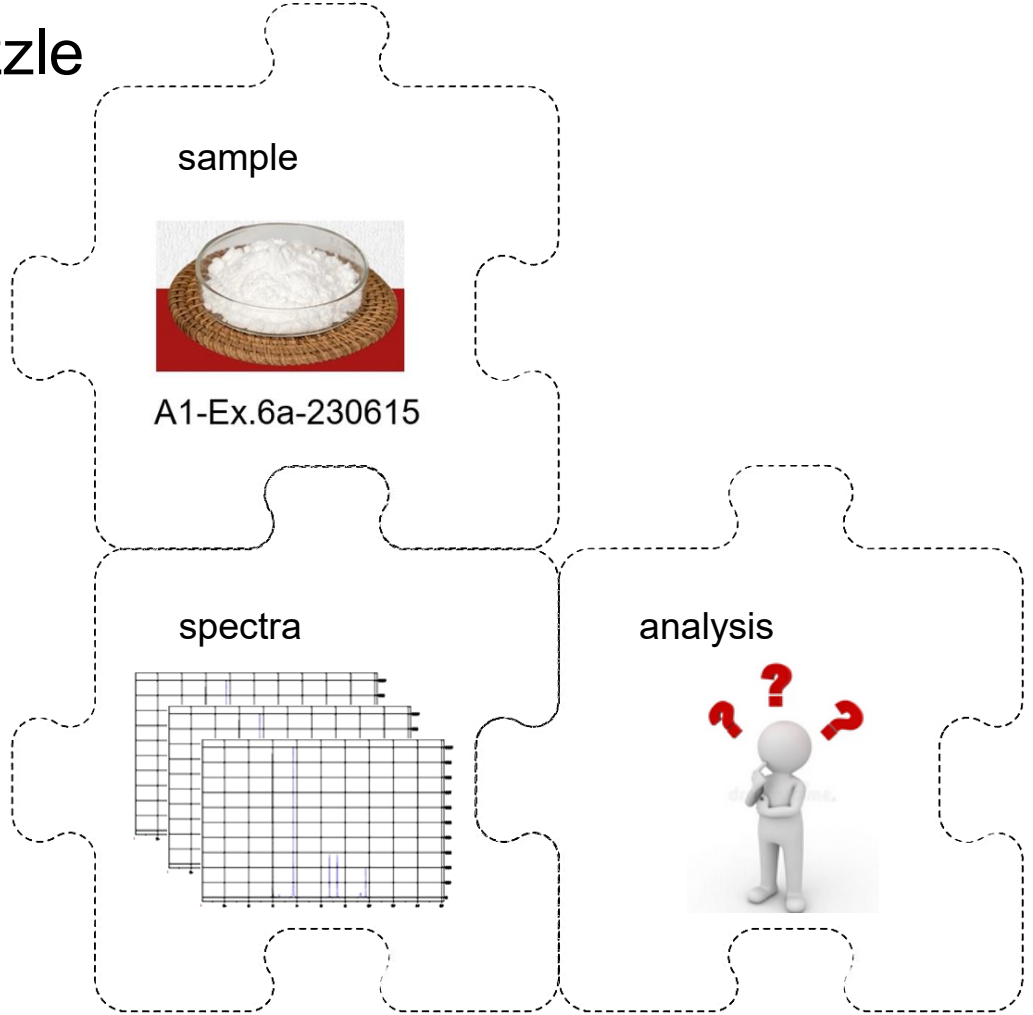
The sample identifier is the *metadata* that establishes the connection between the sample and its spectra.



The data management puzzle

We think on this.

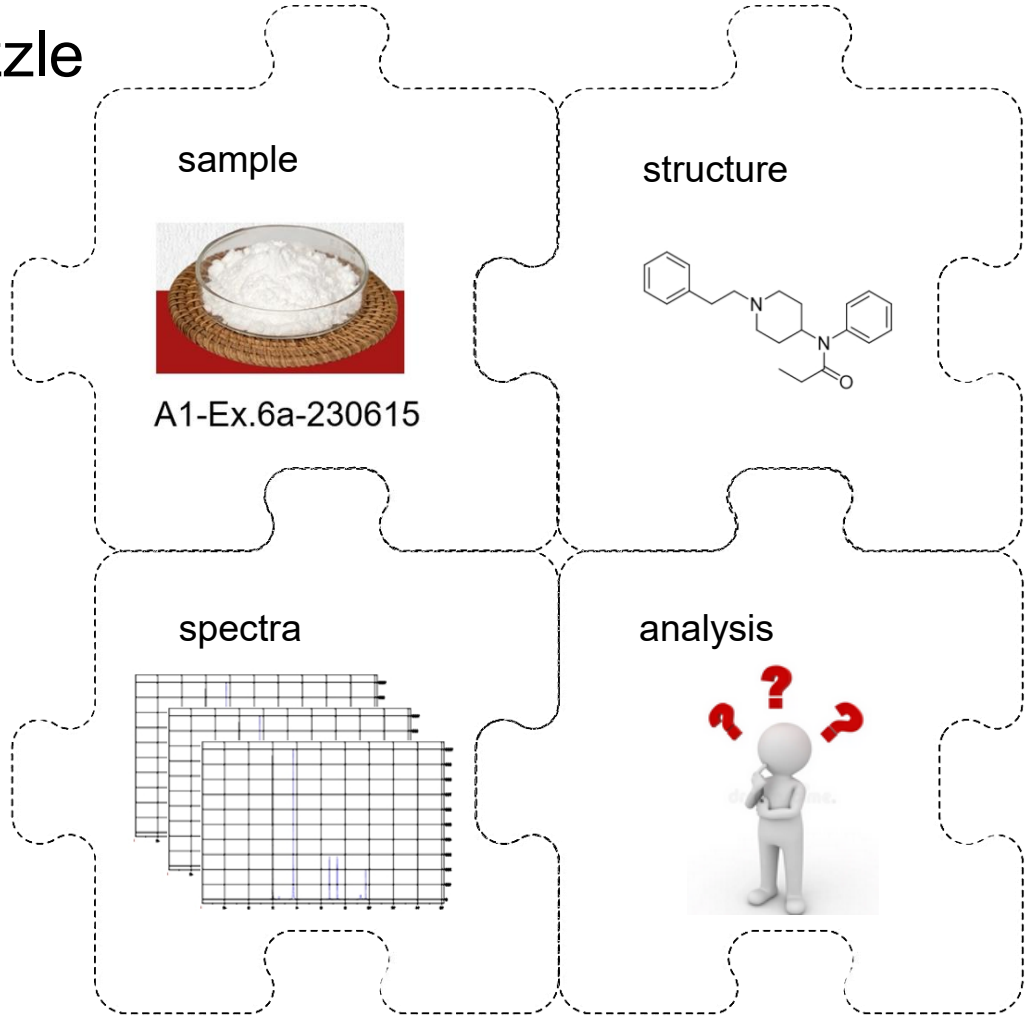
(Or pass it to AI??)



The data management puzzle

Our goal is to make a connection between sample and structure.

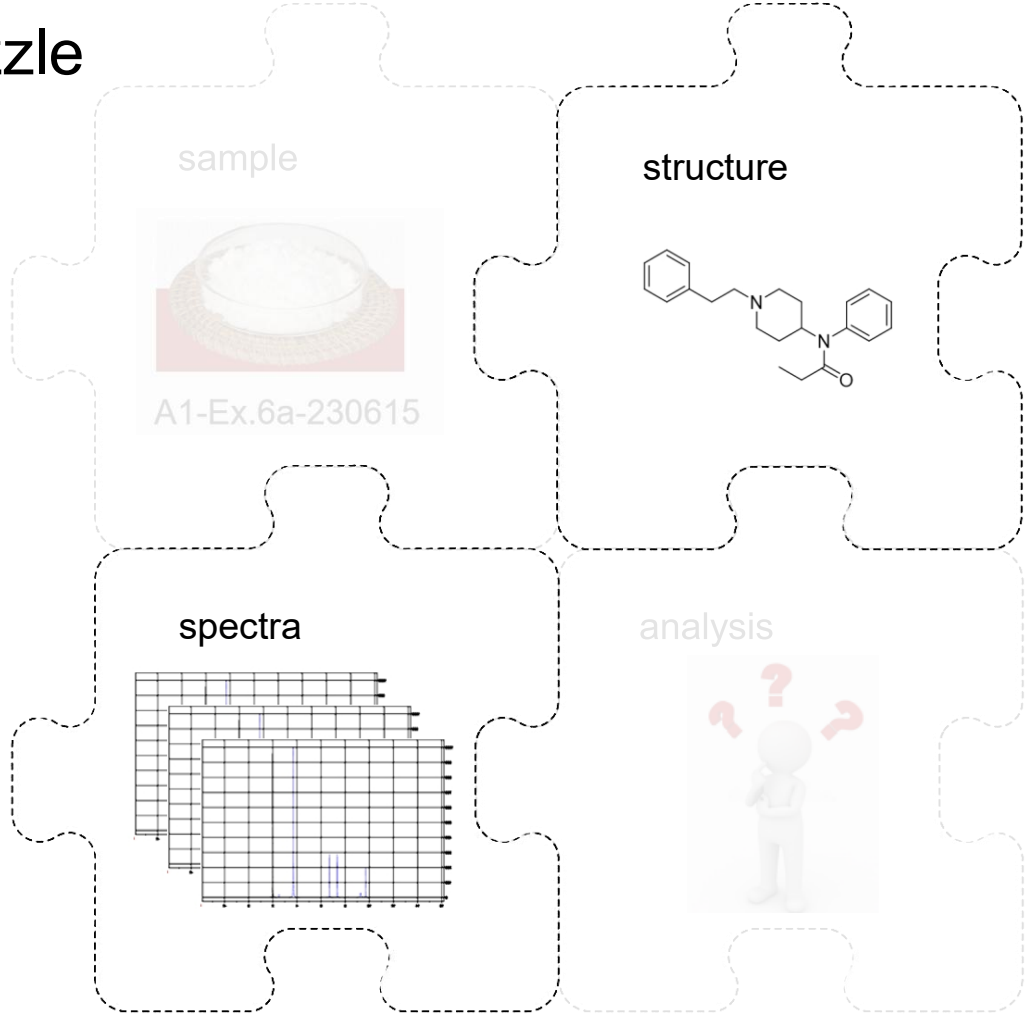
Note how indirect this is!



The data management puzzle

But, in the end, we often just deliver two of the four pieces – the structure and the spectra.

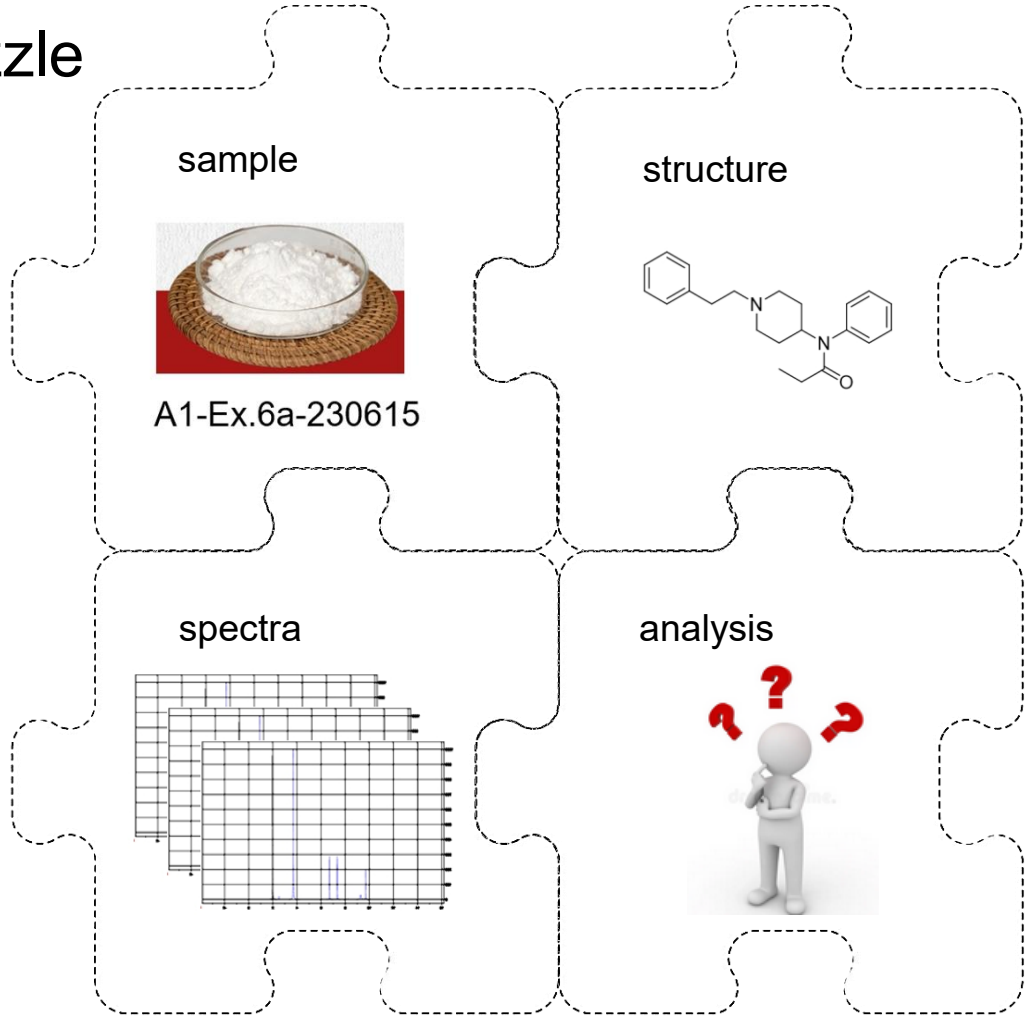
We call this a ***compound association***



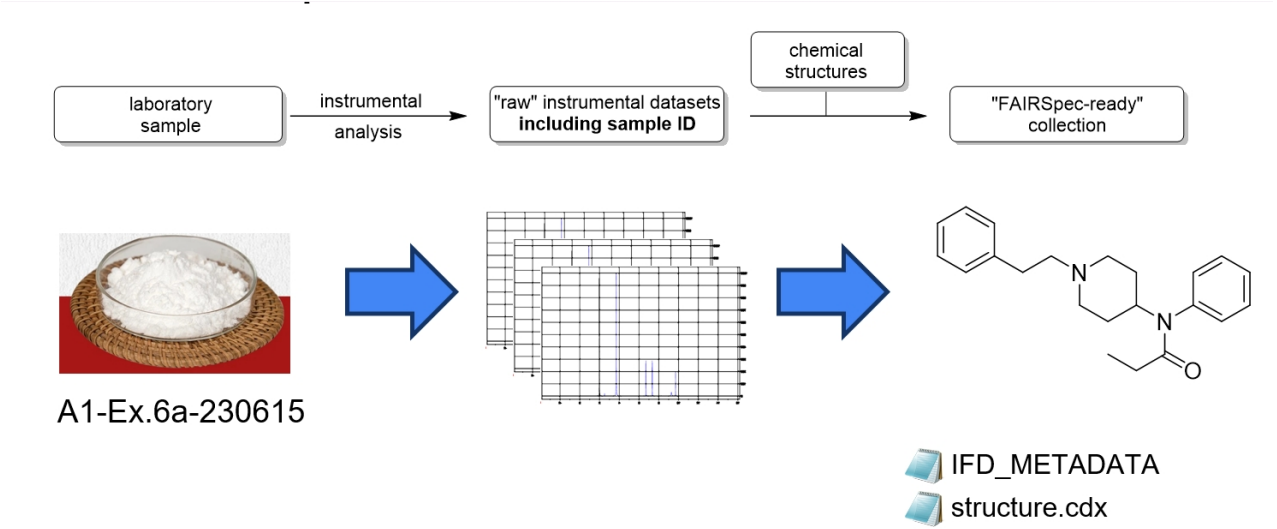
The data management puzzle

The goal of IUPAC Project 2019-031-1-024 is to develop standards for the maintenance and presentation of these relationships.

The overall context is much larger than just “SI for a paper”.



The data management puzzle



The goal of IUPAC Project 2019-031-1-024 is to develop standards for the maintenance and presentation of these relationships *throughout the entire research workflow*.

Guiding Principles for the FAIR Management of Spectroscopic Data

FAIRSpec-Ready Spectroscopic Data Collections – Advice for Researchers, Authors, and Data Managers (IUPAC Technical Report)

Mark Archibald, Ian Bruno, Stuart Chalk, Antony N. Davies, Robert M. Hanson, Stefan Kuhn, Robert J. Lancashire, and Henry S. Rzepa.

Pure and Applied Chemistry, 2025

https://doi.org/10.1515/pac-2025-0409

1. FAIR Management of data should be an ongoing concern.

- Don't wait until publication time to organize your data.
- Recognize the ongoing value of well-organized data.
- Allow for corrections and addition of new information.

2. Context is important.

- Associate spectra with chemical structure as much as possible.
- Allow for ambiguity and the reconsideration of these associations.
- Find ways to validate your structural and spectral analysis.

3. FAIR management of data requires curation.

- Accept that you are going to have to do part of the work.
- Optimize opportunities for data citation.
- Do not presume to know how people will utilize your data.

4. Metadata must be registered and standardized.

- Findability relies upon proper registration.
- Work with data management professionals in your organization.
- Include discipline-specific metadata.

5. FAIR data management standards should be modular, extensible, and flexible.

- FAIR data management should be as simple as possible.
- Find (or create!) the right tools for the job.
- Find ways to make data management useful to you and your project *now.*

IUPAC FAIRSpec Principles

1. FAIR Management of data should be an ongoing concern.

A. FAIR management of data must be an explicit part of research culture.
B. FAIR management of data should be of intrinsic value.
C. Good data management requires distributed curation.
D. Experimental work is by nature iterative.

What it means to be *FAIRSpec Ready:*

- Don’t wait until publication time to organize your data!
- Recognize the ongoing value of well-organized data.
- Find (or create!) the right tools for the job.
- Allow for corrections and addition of new information.

IUPAC FAIRSpec Principles

2. Context is important.

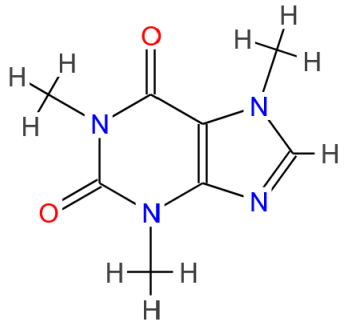
A. Digital objects are generally part of a collection.
B. Chemical properties are related to chemical structure.
C. Data relationships are diverse and develop over time.
D. FAIR management of data should allow for validation.

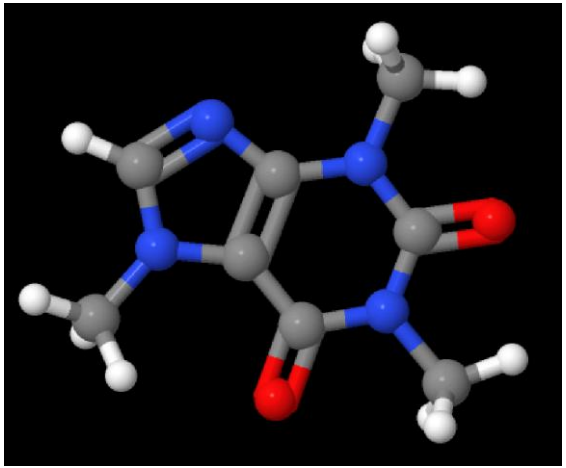
What it means to be *FAIRSpec Ready:*

- Recognize context – a day's work, a project, a team effort.
- Associate spectra with chemical structure, if you can.
- Allow for ambiguity and reconsideration of these associations.
- Find ways to validate your structural and spectral analysis.

Examples of Stucture Representations

… a 2D or 3D image





Examples of Stucture Representations

… a MOL file

```
C8H10N4O2
APtclcactv03202207183D 0   0.00000     0.00000

 24 25  0  0  0  0  0  0  0  0999 V2000
    1.3120   -1.0479    0.0025 N   0  0  0  0  0  0  0  0  0  0  0  0
    2.2465   -2.1762    0.0031 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.7906    0.2081    0.0010 C   0  0  0  0  0  0  0  0  0  0  0  0
    2.9938    0.3838    0.0002 O   0  0  0  0  0  0  0  0  0  0  0  0
    0.9714    1.2767   -0.0001 N   0  0  0  0  0  0  0  0  0  0  0  0
    1.5339    2.6294   -0.0017 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.4026    1.0989   -0.0001 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.4446    1.9342   -0.0010 N   0  0  0  0  0  0  0  0  0  0  0  0
   -2.5608    1.2510   -0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -2.2862   -0.0680    0.0015 N   0  0  0  0  0  0  0  0  0  0  0  0
   -3.2614   -1.1612    0.0029 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.9114   -0.1939    0.0014 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.0163   -1.2853   -0.0022 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.4380   -2.4279   -0.0068 O   0  0  0  0  0  0  0  0  0  0  0  0
    3.2697   -1.8004    0.0022 H   0  0  0  0  0  0  0  0  0  0  0  0
```

Examples of Stucture Representations

… a CDXML (ChemDraw) file

```
<?xml version="1.0"?>
<!DOCTYPE CDXML SYSTEM "http://www.cambridgesoft.co
<CDXML HashSpacing="2.50" MarginWidth="1.60" LineWi
<page>
<fragment id="1">
<n id="2" p="256.73 207.15"/>
<n id="3" p="256.73 223.22"/>
<n id="4" p="242.82 231.25"/>
<n id="5" p="242.82 247.32"/>
<n id="6" p="256.73 255.36"/>
<n id="7" p="256.73 271.43" Element="16"/>
<n id="8" p="272.8 271.43" Element="8"/>
<n id="9" p="240.66 271.43" Element="8"/>
<n id="10" p="256.73 287.5" Element="7"/>
<n id="11" p="242.82 295.54"/>
<n id="12" p="242.82 311.61"/>
<n id="13" p="256.73 319.64" Element="7"/>
<n id="14" p="256.73 335.72" Element="16"/>
<n id="15" p="240.66 335.72" Element="8"/>
<n id="16" p="272.8 335.72" Element="8"/>
<n id="17" p="256.73 351.79"/>
<n id="18" p="242.82 359.82"/>
<n id="19" p="242.82 375.89"/>
<n id="20" p="256.73 383.93"/>
<n id="21" p="256.73 400.0"/>
<n id="22" p="270.65 375.89"/>
<n id="23" p="270.65 359.82"/>
<n id="24" p="228.45 320.2"/>
```

Examples of Stucture Representations

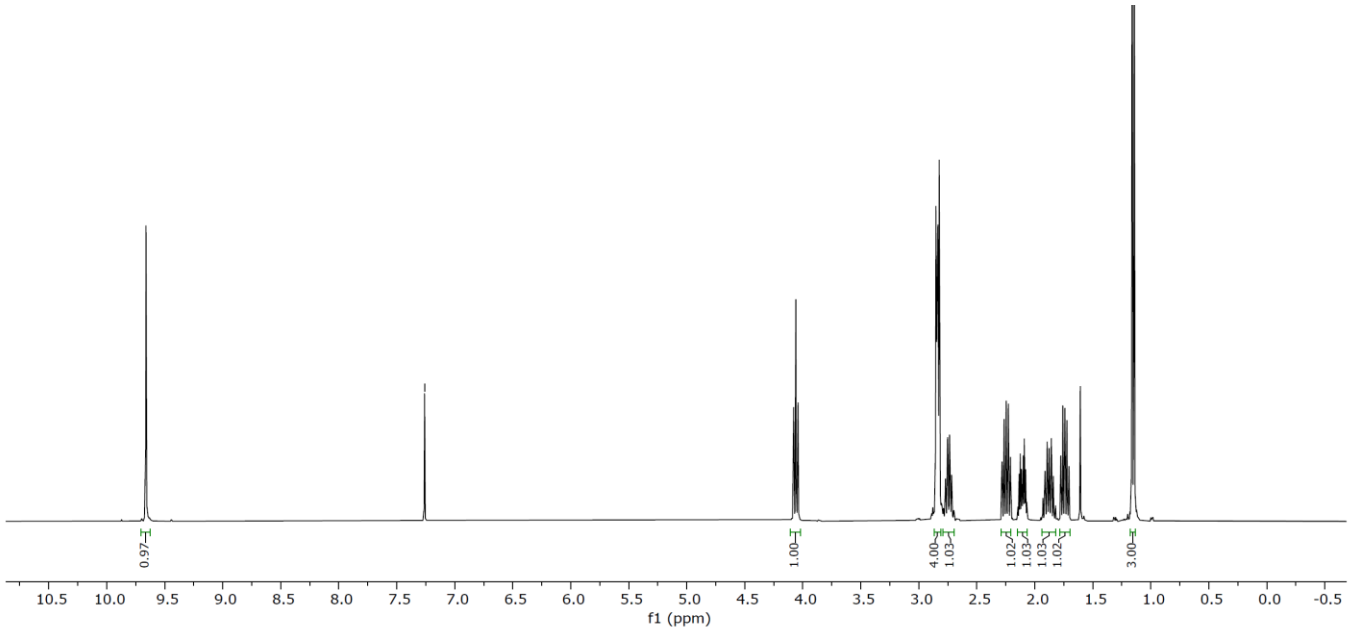
… a *SMILES* or InChI string

C[N]1C=NC2=C1C(=O)N(C)C(=O)N2C

InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3

Examples of Spectroscopic Representations

… a spectrum image



Examples of Spectroscopic Representations

… a linear description

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.45 (ddq, *J* = 4.3, 2.9, 1.4 Hz, 1H), 4.09 (t, *J* = 5.94 Hz, 1H), 3.13 – 3.02 (m, 1H), 2.98 (s, 1H), 2.59 (ddtd, *J* = 16.1, 5.2, 2.4, 1.3 Hz, 1H), 2.34 (ddd, *J* = 11.5, 5.4, 1.9 Hz, 1H), 1.87 (tq, *J* = 6.1, 4.0 Hz, 1H), 1.79 (ddd, *J* = 14.4, 8.5, 4.9 Hz, 1H), 1.72 – 1.64 (m, 4H), 1.63 – 1.58 (m, 1H), 1.57 – 1.49 (m, 1H), 1.37 (dtd, *J* = 12.0, 5.6, 0.6 Hz, 1H), 1.05 (d, *J* = 6.5 Hz, 3H), 1.02 (s, 3H), 0.99 – 0.94 (m, 12H), 0.94 (s, 3H), 0.65 – 0.56 (m, 7H), 0.52 (td, *J* = 9.3, 5.0 Hz, 1H) ppm;

Examples of Spectroscopic Representations

… a JCAMP-DX file

```
##TITLE= Beta_Pinene
##JCAMP-DX= 6.0 $$ MestReNova 14.0.1-23559
##DATA TYPE= NMR SPECTRUM
##DATA CLASS= XYDATA
##ORIGIN= Mestrelab Research S.L.
##OWNER= skim592
…
```

Examples of Spectroscopic Representations

… an instrument dataset

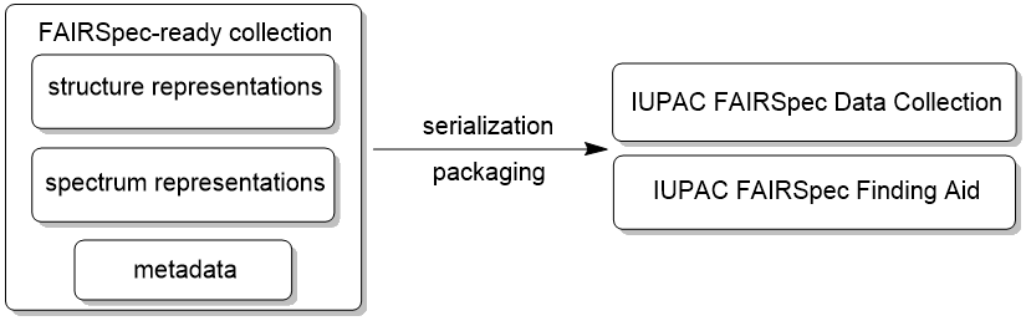
This PC › Windows (C:) › temp › bruker › 10

sonal

Name

- pdata
- acqu
- acqus
- audita.txt
- fid
- orig
- prosol_History
- pulseprogram
- scon
- uxnmr.par

The IUPAC FAIRSpec Solution



Our solution is to enable authors and data managers to create what we are calling a **“FAIRSpec-ready” collection.** Then let automation take that the rest of the way to an **IUPAC FAIRSpec Collection** with its associated **IUPAC FAIRSpec Finding Aid.**

W**e have tested this with over a dozen datasets**, *and it works*.

The IUPAC FAIRSpec Finding Aid

JSON serialization

The IUPAC FAIRSpec Finding Aid

Web page interpretation

fully customizable display

This page and its associated IUPAC FAIRSpec Finding Aid were automatically generated by IFDExtractor.java (GitHub site). It is a demonstration page for IUPAC Project 2019-031-1-024, *Development of a Standard for FAIR Data Management of Spectroscopic Data.*

Show Finding Aid

IFD.collection.zip
(193.2 MB)

summary search

Collections: Compounds(44) Structures(90) Spectra(374)

Title	*Syntheses and Characterization of Main Group, Transition Metal, Lanthanide, and Actinide Complexes of Bidentate Acylpyrazolone Ligands*
Authors	**Thomas Mies, Andrew J. P. White, Henry S. Rzepa, Luciano Barluzzi, Mohit Devgan, Richard A. Layfield, Anthony G. M. Barrett**
DOI	https://doi.org/10.1021/acs.inorgchem.3c01506 (metadata)
Data Title	*Syntheses and Characterization of Main Group, Transition Metal, Lanthanide and Actinide Complexes of Bidentate Acylpyrazolone Ligands*
Data DOI	https://doi.org/10.14469/HPC/10386 (metadata)

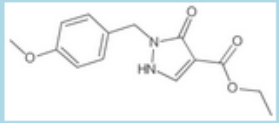
IFD 0.1.0-beta+2025.06.25;FAIRSpec 0.1.0-beta+2025.06.25

The IUPAC FAIRSpec Finding Aid

Web page interpretation

value-added content

Structure 8



Predicted Spectra
1H 13C
COSY HMBC

cdxml 8_8.cdxml (7.2 KB)
mol_2d 8_8.cdxml.mol (1.5 KB)
standard_inchi
fixedh_inchi
smiles

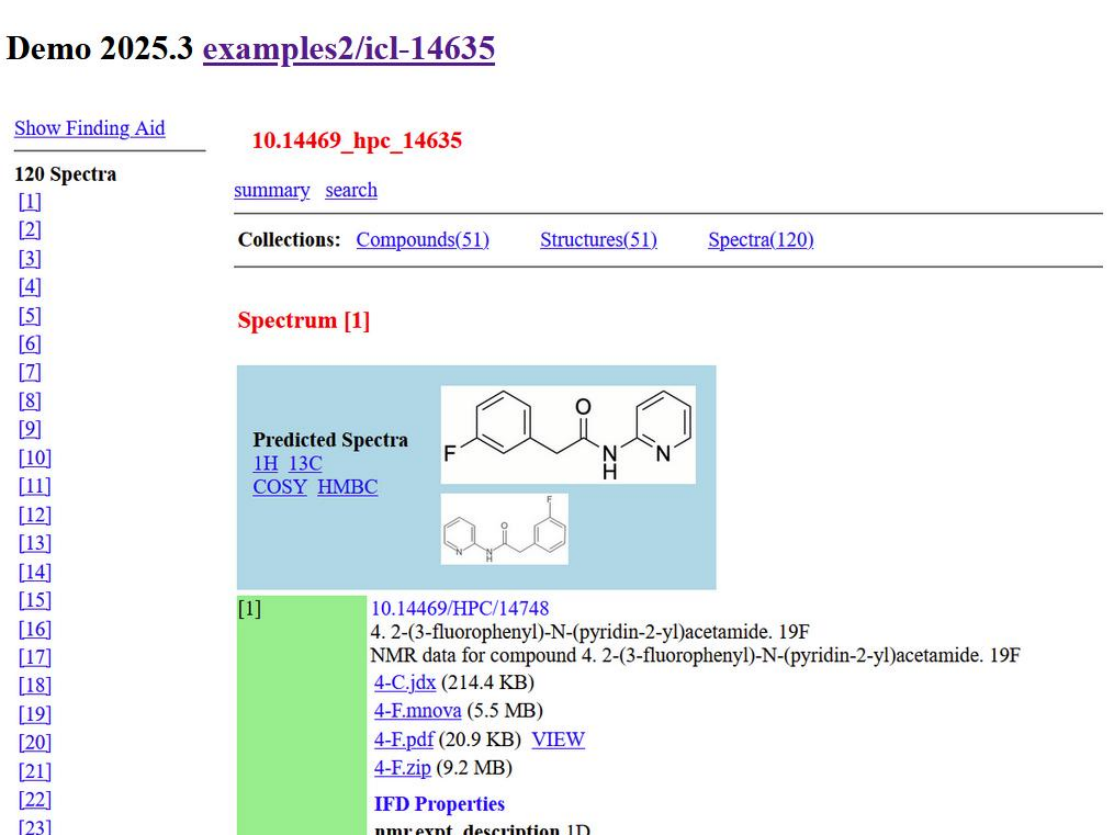
inchikey NQMLPZGANVLZAJ-UHFFFAOYSA-N
molecular_formula H 16 C 14 N 2 O 4

The IUPAC FAIRSpec Finding Aid

Web page interpretation

selective access to representations

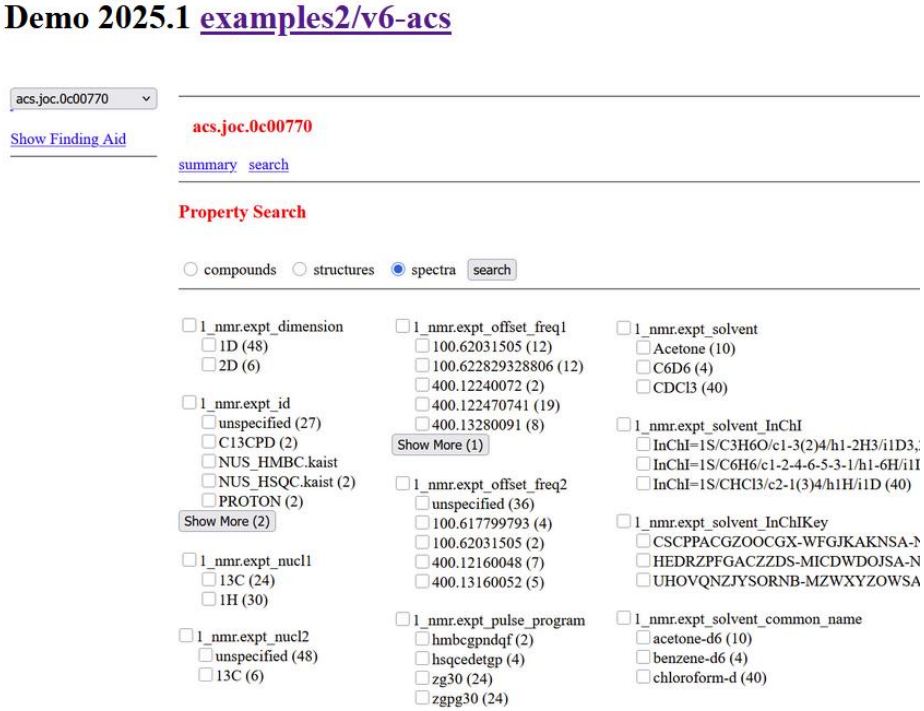
[examples2/icl-14635]



The IUPAC FAIRSpec Finding Aid

Web page interpretation

metadata searchability

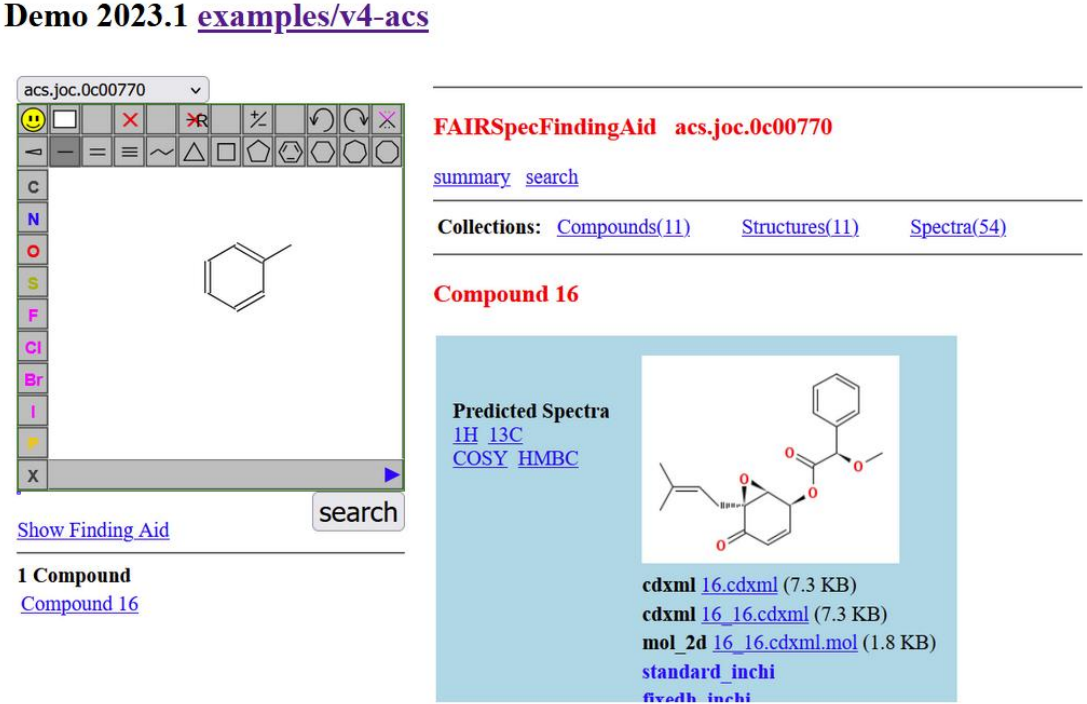


examples2/v6-acs

The IUPAC FAIRSpec Finding Aid

Web page interpretation

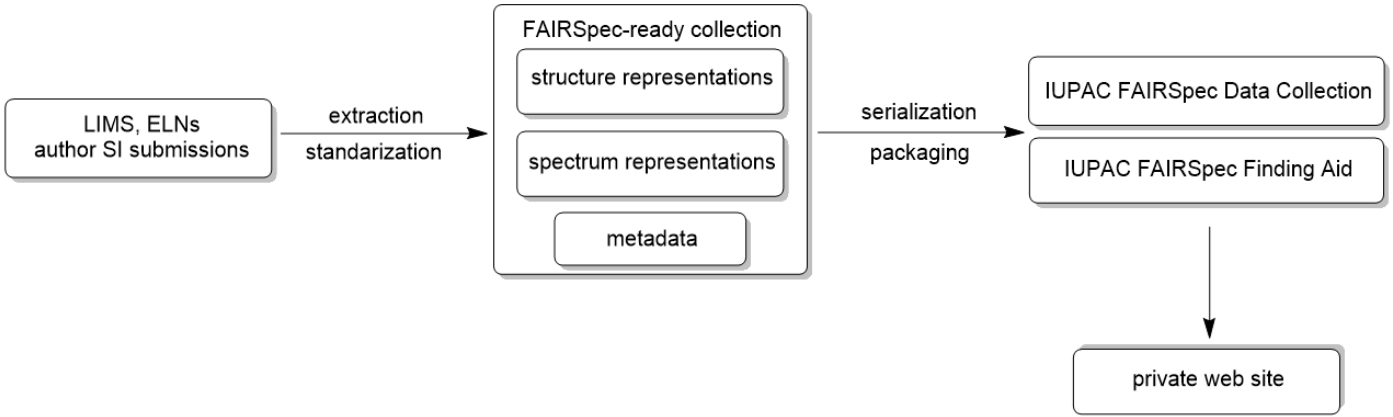
substructure searchability



examples/v4-acs

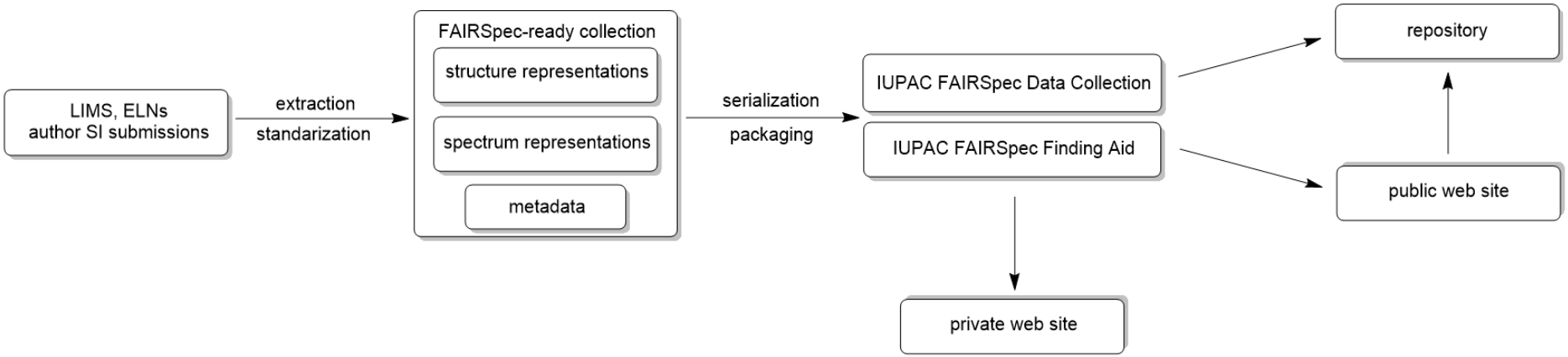
The IUPAC FAIRSpec Collection – Local Implementation

allows for immediate customized in-lab access to project data and analyses



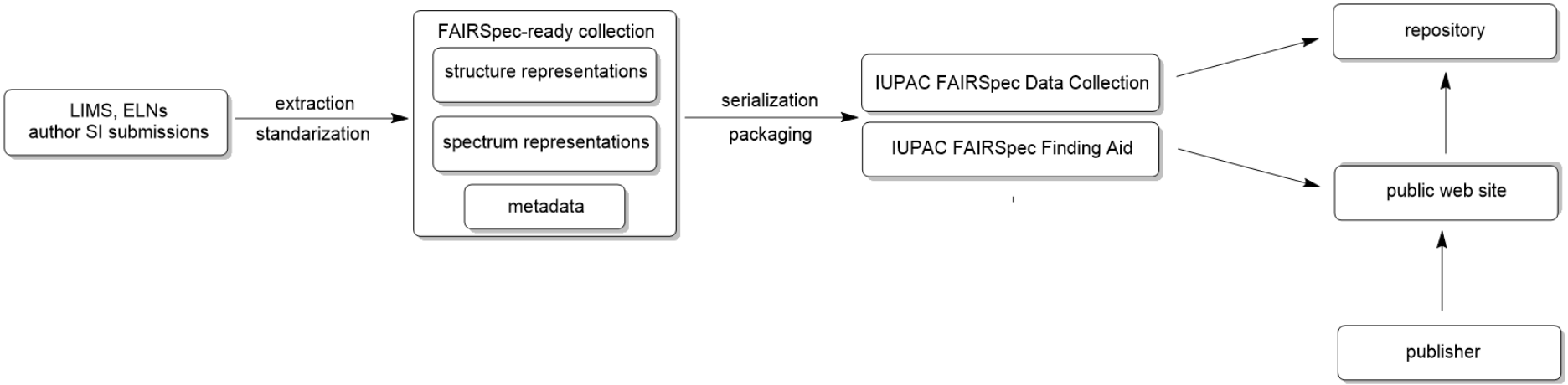
The IUPAC FAIRSpec Collection – Public Implementation

allows for controlled access
via a repository and web-based portal



The IUPAC FAIRSpec Collection – Publication Implementation

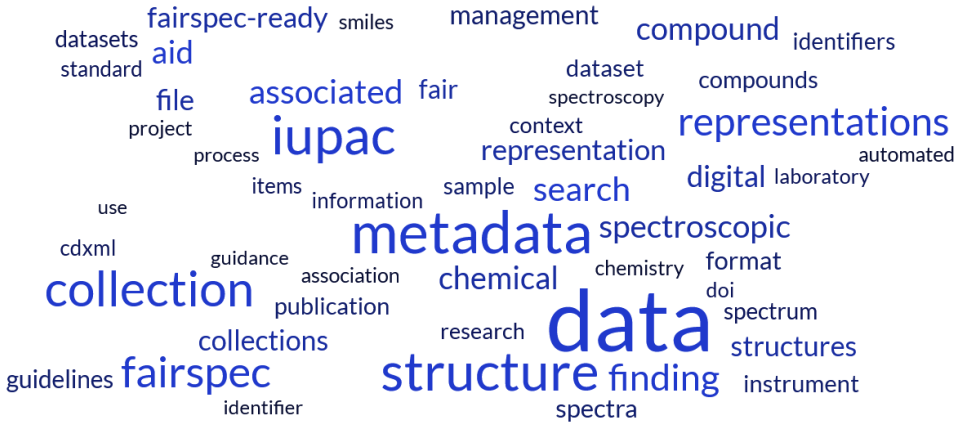
allows for customized post-publication
interactive access to supporting information

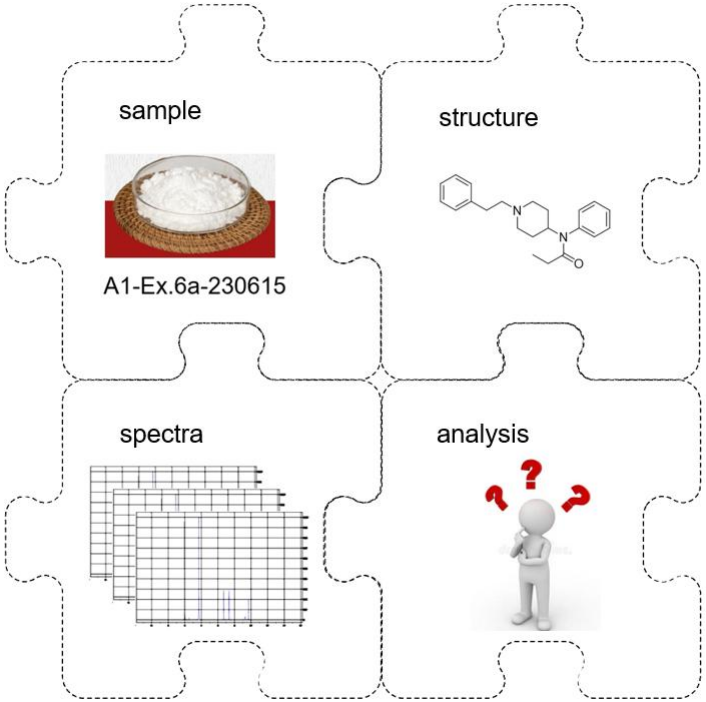


IUPAC FAIRSpec – six take-aways

- We have created a standard for spectroscopic data management.
- The standard focuses on *data collections* and their *finding aids.*
- “FAIRSpec-ready” collections are easy to create and manage.
- IUPAC FAIRSpec Finding Aids are created using automation.
- Collections can be private (sample-based) or public (compound-based).
- Standardization provides a basis for value-added opportunities that can be customized to suit the context.

Thank you for your attention





Bob Hanson hansonr@stolaf.edu
IUPAC Project https://iupac.org/project/2019-031-1-024
GitHub Project https://github.com/IUPAC/IUPAC-FAIRSpec