Are Your NMR Data FAIRSpec-Ready?

IUPAC Project 2019-031-1-024, Development of a Standard for FAIR Data Management for Spectroscopic Data

https://github.com/IUPAC/IUPAC-FAIRSpec

Stefan Kuhn*, Mark Archibald, Ian Bruno, Stuart Chalk, John Cort, Tony Davies, Robert Hanson, Robert Lancashire, Henry Rzepa

"FAIRSpec" is an IUPAC standard in late-stage development for creating digital spectroscopic data packages for chemistry that are maximally *findable, accessible, interoperable,* and *reusable.* **Following the simple guidelines presented here**, you can ensure that your data are "FAIRSpec-Ready" – in a format that can be readily packaged into *IUPAC FAIRSpec Collections* and referenced by *IUPAC FAIRSpec Finding Aids,* as described in our recent prepublication:

https://chemrxiv.org/engage/chemrxiv/article-details/67725ce881d2151a025a3fc6

FAIRSpec Principles

The five principles that underlie development of the IUPAC FAIRSpec standard are given below.

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 registered and standardized.
- 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. Design to adapt to future needs.
- C. Respect digital diversity.
- D. All data formats should be valued.

FAIRSpec-Ready Guidelines

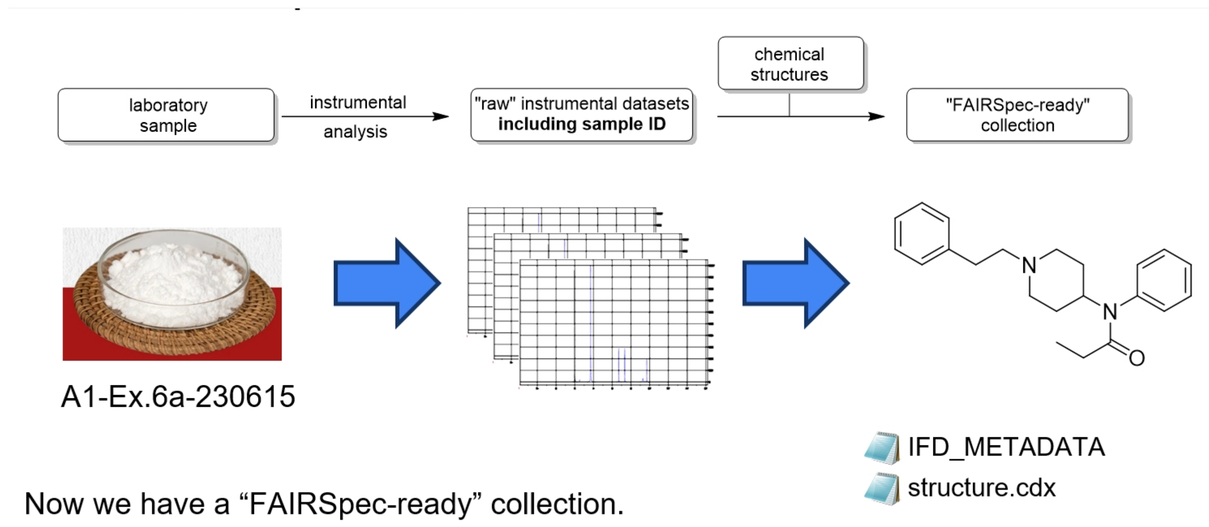
1. Systematically connect spectra to sample IDs. Make sure that sample IDs are placed in the dataset in a systematic manner (in a file, not just part of the title, for example.) See our publication for examples.

2. Encourage users to add CDX or CDXML structures. The key is to make a clear connection between structure and spectrum (if and when that is possible). ChemDraw format is ideal, in that it is easily created by users and provides all the information necessary to generate additional structure representations, such as SMILES, InChI, and MOL files. This significantly increases findability and interoperability.

3. Retain original data; don't rely only on post-processing software. For maximum reusability, provide all files (raw directory or optionally ZIP or TAR compressed) created by the instrument. Third-party analysis collections such as Mnova files can be provided *in addition* to instrument files, providing additional value, but should not constitute the *sole* form in which data are provided.

4. Provide IUPAC FAIRSpec Finding Aids and web portals for users. Use newly available tools to automatically create IUPAC FAIRSpec Finding Aids in JSON format from FAIRSpec-ready collections and associated web-based landing pages for use locally in the lab, added to publication SI packages, or available along with repository-based datasets.

5. Be proactive. Talk with us! Join or contact our IUPAC Project (hansonr@stolaf.edu) to get involved. Work with us to develop the standard and the Java tools we are developing to add value to your data packages. Tell us what you need; help us get it done.



Now we have a "FAIRSpec-ready" collection.

The IUPAC FAIRSpec Finding Aid

- Schema-based JSON document displayable in HTML
- Sample-based or compound-based
- Can be used at any stage in the process, privately or publicly
- Created by automated extraction of metadata from an IUPAC FAIRSpec-ready collection
- Collects and summarizes key metadata
- Creates additional value-added metadata (for example, cdx → MOL, SMILES, InChI)
- Can incorporate images and PDF files for quick review
- Searchable by substructure, full text, and property metadata
- Potential for assisting in the validation of spectroscopic data and structure-spectral analysis