UNIVERSIDAD DE MÁLAGA

Powder diffraction data sharing and reuse: advantages and possible practical obstacles

Miguel A. G. Aranda

---

g_aranda@uma.es

Inorganic Chemistry Department, University of Malaga, Spain

UNIVERSIDAD DE MÁLAGA

Outline

1. Motivation
2. Raw powder diffraction data
3. FAIR & FACT
4. Sharing / reviewing process
5. Conclusions

UNIVERSIDAD DE MÁLAGA

1. Motivation

- We are in this WS, so little motivational discussion is needed for *worldwide* **Raw Diffraction Data Sharing and Reuse**

1. Helping experiment replication.

2. Likely better data analysis in the future (improvements)

3. New findings (and science) using ML and AI over many scientific datasets – our Big Data

UNIVERSIDAD DE MÁLAGA

2. Raw data - *Type of scientific raw data in PD*
Type of information to be extracted

Primary raw data → *(Ready-to-be-analyzed)* **Processed raw data** → **(Reduced data)** → **Derived data**

To be archived by the Facilities

To be shared by authors, along paper submission!

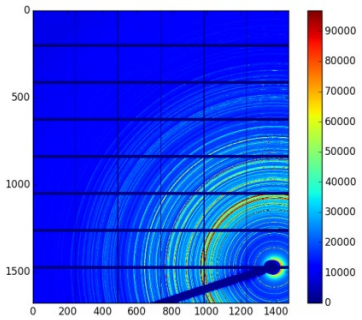
n point detectors

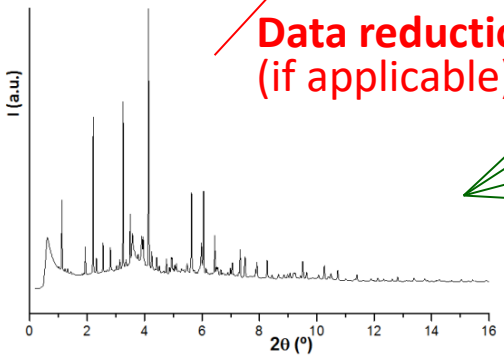
1D detector

2D detector

- Applying detector calibration
- Masking pixels (defective, etc.)
- Corrections (geometry, etc.)
- Radial integration
- Merging of data sets

hkl	I/I_0	d_{obs} (Å)	$2\theta_{obs}$ (°)
012	26	6.282	17.480
104	18	4.4681	24.670
110	5	4.4152	24.970
113	100	3.7841	29.220
006	10	3.6712	30.140
202	1	3.6127	30.640
024	24	3.1412	35.380





Data reduction
(if applicable)

- ♦ Atomic parameters
- ♦ Microstrain values
- ♦ Coherent diff. domain size
- ♦ QPA
- ♦ Total amorphous content
- ♦ Many other data (bulk modulus for high pressure, etc.)

Data processing
(detector & geometry dependent)

Data analysis
(several types of software & programs)

Sharing powder diffraction raw data: challenges and benefits *J. Appl. Cryst.* (2018). **51**, 1739–1744

UNIVERSIDAD DE MÁLAGA

Type of scientific raw data in PD

Primary raw data → **Processed raw data**

Data ready to be analyzed by common software(s) in the field (structure determination, RQPA, PDF

To be archived by the Facilities ?

To be shared by users, along with paper submission! (or as raw data in specific journals and/or platforms)

point detector

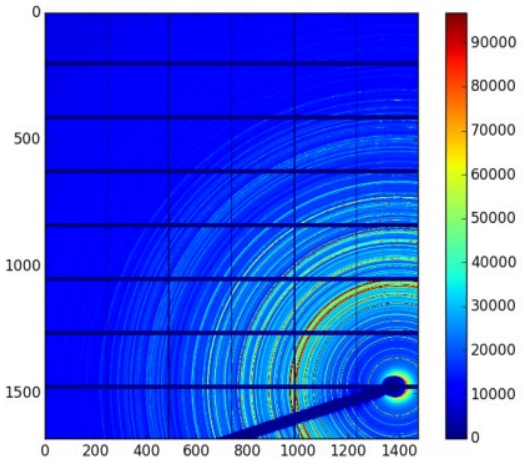
Mainly at laboratories. This is the easiest case as there is no processing of the data.
Just all metadata (instrument-dependent and sample-dependent) must be provided

1D detector

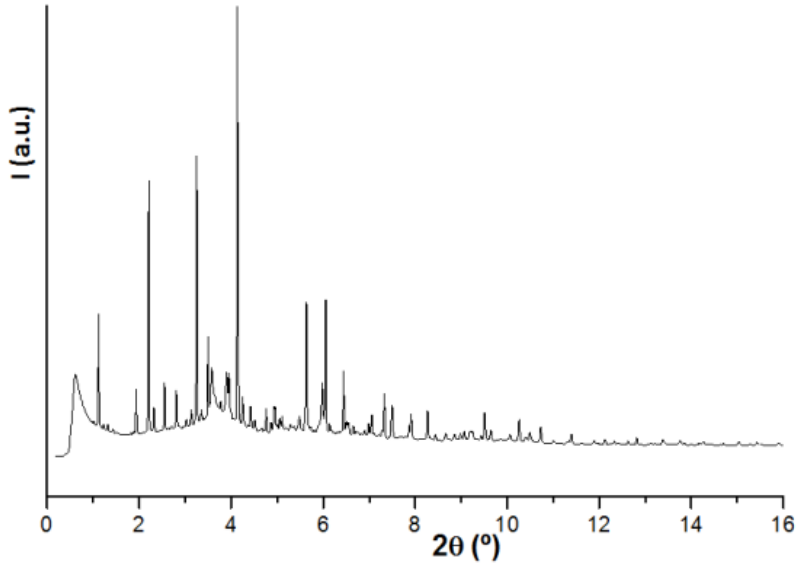
Mainly large Facilities (& some laboratories).
This is an intermediate case as the data processing is usually **'merging of data sets'**
Metadata (instrument-dependent and sample-dependent) must be provided

2D detector

- **Applying detector calibration**
- **Masking pixels (defective, etc.)**
- **Corrections (geometry, etc.)**
- **Radial integration**



Data processing
(detector & geometry dependent)



UNIVERSIDAD DE MÁLAGA

3. FAIR & FACT

FAIR : research data being findable, accessible, interoperable and reusable

Repositories and Large Facilities are addressing this.
Computer engineers are taking care.
To my understanding, not big challenges, just enough funding to accomplish the objectives.

BUT

Flooding the repositories with poor (raw) data could harm/delay research advancement

FACT : the shared data must have sufficient quality. They must be true facts.

HOWEVER

How to address/ensure this, in the publishing step(s) ?

4. Reviewing process – ideas

3. Pilot plan for Powder Diffraction?

3.1. To choose one subfield as standard as possible to test this strategy. In a first thought, to be more elaborated,

(i) PD, single phase for structure solution when unsuccessful

(ii) Quantitative Phase Analysis (cements or similar)

(iii) Pair Distribution Function – total scattering results

Will be used/reused deposited PD raw data?

UNIVERSIDAD DE MÁLAGA

5. Conclusions

Today, I advocate sharing processed raw 1D data in the powder diffraction field

BUT
(1) the 'acceptable' processing should be analysed-discussed-agreed
(2) clear (well-elaborated) guildeline(s) for metadata
sample_dependent
detector-geometry_dependent
processing_dependent

UNIVERSIDAD DE MÁLAGA

Acknowledgements

Thank you very much for your attention!

Our research work is supported by Spanish MINECO grants which are co-funded by FEDER