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
1. Motivation
2. Raw powder diffraction data
3. FAIR & FACT
4. Sharing / reviewing process
5. Conclusions
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
2. Raw data - Type of scientific raw data in PD
Type of information to be extracted

Primary raw data
To be archived by the Facilities

Processed raw data
To be shared by authors, along paper submission!

Derived data

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

Data processing
(detector & geometry dependent)

Data reduction (if applicable)

Data analysis
(several types of software & programs)

n point detectors
1D detector
2D detector

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

<table>
<thead>
<tr>
<th>hkl</th>
<th>J</th>
<th>d_{ab} (Å)</th>
<th>2θ_{ab} (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>012</td>
<td>26</td>
<td>6.282</td>
<td>17.480</td>
</tr>
<tr>
<td>104</td>
<td>18</td>
<td>4.4681</td>
<td>24.670</td>
</tr>
<tr>
<td>110</td>
<td>5</td>
<td>4.4152</td>
<td>24.970</td>
</tr>
<tr>
<td>113</td>
<td>100</td>
<td>3.7841</td>
<td>29.220</td>
</tr>
<tr>
<td>006</td>
<td>10</td>
<td>3.6712</td>
<td>30.140</td>
</tr>
<tr>
<td>202</td>
<td>1</td>
<td>3.6127</td>
<td>30.640</td>
</tr>
<tr>
<td>024</td>
<td>24</td>
<td>3.1412</td>
<td>35.380</td>
</tr>
</tbody>
</table>

Data processing raw data: challenges and benefits  

IUCr-mtg “Raw diffraction data reuse: the good, the bad and the challenging” 22nd August, 2023
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)

IUCr-mtg “Raw diffraction data reuse: the good, the bad and the challenging” 22nd August, 2023
**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?
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) guideline(s) for metadata
   - sample-dependent
   - detector-geometry-dependent
   - processing-dependent
Thank you very much for your attention!

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