

You are a materials researcher, and you openly publish your datasets.

But publishing data seems like throwing it over a fence. Is your data being found and used? What's the point of publishing it if no one finds and uses it? You want your research to have a positive impact, to be part of a growing web of knowledge that is actually used, but you don't have a sense of what you can do apart from tossing your data online in some repository and hoping people will stumble across it.

With this self-assessment tool, you will discover specific gaps you can close relating to the Findability of one of your datasets. You will also gain terminology for understanding this first element of the FAIR guiding principles (Findability, Accessibility, Interoperability, and Reusability) by applying specific questions to your dataset. *Load your dataset, and answer the following:*

<p>Does the dataset have any identifiers assigned?</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Globally Unique, citable and persistent (e.g. DOI, PURL, ARK or Handle)</li> <li><input type="checkbox"/> Web address (URL)</li> <li><input type="checkbox"/> Local identifier</li> <li><input type="checkbox"/> No identifier</li> </ul>	<p>Identifiers are essential for identifying, finding, retrieving, linking and citing datasets. A Web address (URL) can be used to specify the online location of a resource but over time URLs tend to change which leads to broken links to the data. To be useful, identifiers need to be persistent and unique.</p>
<p>Is the dataset identifier included in all metadata records/files describing the data?</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Yes</li> <li><input type="checkbox"/> No</li> </ul>	<p>The identifier (preferably a persistent identifier) needs to be clearly stated in the metadata record describing the data collection, and also in any associated data files or metadata.</p>
<p>How is the data described with metadata?</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Comprehensively, using a recognised formal machine-readable metadata schema.</li> <li><input type="checkbox"/> Comprehensively, but in a text-based, non-standard format.</li> <li><input type="checkbox"/> Brief title and description</li> <li><input type="checkbox"/> The data is not described</li> </ul>	<p>Comprehensive metadata records will include descriptive content that facilitates discovery, access and reuse of the data being described.</p> <p>Providing metadata in a standard schema allows it to be read and used by machines as well as humans.</p>
<p>What type of repository or registry is the metadata record in?</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Data is in one place but discoverable through several registries</li> <li><input type="checkbox"/> Generalist public repository</li> <li><input type="checkbox"/> Domain-specific repository</li> <li><input type="checkbox"/> Local institutional repository</li> <li><input type="checkbox"/> The data is not described in any repository</li> </ul>	<p>A rich metadata description alone does not ensure a dataset's 'findability' on the internet; the dataset needs to be registered or indexed in a searchable resource, such as a generalist (e.g. Figshare, Dryad, Zenodo), domain-specific (e.g. COD, Materials Data Facility, NOMAD), or institutional data repository or registry. Ideally, these repositories/registries are indexed by search engines such as Google, Google Dataset Search / Scholar, etc.</p>

If you have any questions about dataset findability, please don't hesitate to contact me at [donny@polyneme.xyz](mailto:donny@polyneme.xyz).

Donny Winston is a knowledge scientist who helps materials researchers build sustainable data infrastructure to accelerate discovery. You can reach him at [donny@polyneme.xyz](mailto:donny@polyneme.xyz).