

FAIR Metric FM-F3

Mark D. Wilkinson, Susanna-Assunta Sansone,
Erik Schultes, Peter Doorn,
Luiz Olavo Bonino da Silva Santos, Michel Dumontier

November 21, 2017

<i>FIELD</i>	<i>DESCRIPTION</i>
Metric Identifier	FM-F4
Metric Name	Indexed in a searchable resource
To which principle does it apply?	F4 - (meta)data are registered or indexed in a searchable resource
What is being measured?	The degree to which the digital resource can be found using web-based search engines.
Why should we measure it?	Most people use a search engine to initiate a search for a particular digital resource of interest. If the resource or its metadata are not indexed by web search engines, then this would substantially diminish an individual's ability to find and reuse it. Thus, the ability to discover the resource should be tested using i) its identifier, ii) other text-based metadata.
What must be provided?	The persistent identifier of the resource and one or more URLs that give search results of different search engines.
How do we measure it?	We perform an HTTP GET on the URLs provided and attempt to find the persistent identifier in the page that is returned. A second step is to follow each of the top 10 hits and examine the resulting documents for presence of the identifier.
What is a valid result?	true -> the persistent identifier was found in the search results.
For which digital resource(s) is this relevant?	All

Examples of their application across types of digital resource	<p>- my Zenodo Deposit for polyA (http://doi.org/10.5281/zenodo.47641) Test Query: 10.5281/zenodo.47641 orthology GOOGLE: Pass (#1 hit); BING: Fail (no hits); Yahoo: Fail (no hits); Baidu: Pass (#1 hit) Test Query: “protein domain orthology RNA Processing” Google: Pass (Hit #13); BING: Fail (not in top 40); Yahoo: Fail: (Not in top 40); Baidu: Pass (#1 Hit)</p> <p>- myExperiment Workflow (http://www.myexperiment.org/workflows/2969.html) Test Query: “workflow common identifiers EMC ontology” GOOGLE: Pass (#2 and #5 hit); BING: Fail (not in top 40, though OTHER workflows were found in top 10!); Yahoo: Fail (not in top 40, though other workflows found in top 10); Baidu: Pass (5/10 pages contained a link to the workflow, but the workflow itself was not discovered)</p> <p>- Jupyter notebook on GitHub (https://github.com/VidhyasreeRamu/GlobalClimateChange/blob/master/GlobalWarmingAnalysis.ipynb) Test Query: “github python climate change earth surface temperature” Google: Fail (not in top 40; other similar Jupyter notebooks found in github); Bing: Fail (not in top 40... but MANY links to Microsoft Surface! LOL!); Yahoo: Fail (not in top 40); Baidu: Fail (not even a github hit in top 40!)</p>
Comments	None