FAIR Metric FM-R1.2

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

November 21, 2017

FIELD	*DESCRIPTION*
Metric Identifier	FM-R1.2
Metric Name	Detailed Provenance
To which principle does it apply?	R1.2 - (meta)data are associated with detailed provenance
What is being measured?	That there is provenance information associated with the data, covering at least two primary types of provenance information: - Who/what/When produced the data (i.e. for citation) - Why/How was the data produced (i.e. to understand context and relevance of the data)
Why should we measure it?	Reusability is not only a technical issue; data can be discovered, retrieved, and even be machine-readable, but still not be reusable in any rational way. Reusability goes beyond "can I reuse this data?" to other important questions such as "may I reuse this data?", "should I reuse this data", and "who should I credit if I decide to use it?"
What must be provided?	Two URLs (GUIDs). One of these URLs points to one of the vocabularies used to describe citational provenance (e.g. dublin core). The second points to one of the vocabularies (likely domain-specific) that is used to describe contextual provenance (e.g. EDAM)
How do we measure it?	We resolve the URLs/GUIDs according to their associated protocols.
What is a valid result?	GUID 1 should resolve to a recognized citation provenance standard such as Dublin Core. GUID 2 should resolve to some vocabulary that itself passes basic tests of FAIRness
For which digital resource(s) is this relevant?	All
Examples of their application across types of digital resource	None

Comments	Many data formats have fields specifically for Provenance information. -> could fairsharing curate these 4 fields? for every format and vocabulary? Some formats do not have these fields. For example, although gff can have arbitrary headers, the standard itself does not provide specific fields to capture detailed provenance. It therefore would