

## Differences between OpenAire export and new DataCite XML submitted to DataCite

| Change                                                                                                                                                                                                                                                                                                                                                                                      | Comment | Resolution | Fixed |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|------------|-------|
| New code uses DataCite 4.5 vs. OpenAire at 4.1                                                                                                                                                                                                                                                                                                                                              |         |            |       |
| OpenAire picks the max of distributionDate, publicationYear, and depositDate as the pub date and doesn't add the element if it is not set. DataCite(N (new), C (current)) picks the pub date or create date if not published.                                                                                                                                                               |         |            |       |
| OpenAire doesn't support PermaLinks and assumes identifier URLs can be parsed like DOIs and then submits the numeric part w/o the required identifierType value.                                                                                                                                                                                                                            |         |            |       |
| The new DataCite code assumes an identifier exists and parses it as a DOI, Handle, or URL, the last case handling PermaLinks and including the full URL as the identifier.<br>( <a href="https://guidelines.readthedocs.io/en/latest/data/field_identifier.html">https://guidelines.readthedocs.io/en/latest/data/field_identifier.html</a> does not specify a format for URL identifiers). |         |            |       |
| OpenAire did not check that a creator identifier URI matched the provided scheme. Now identifier will not be sent if it isn't valid (according to rules in edu.harvard.iq.dataverse.ExternalIdentifier)                                                                                                                                                                                     |         |            |       |
| OpenAire would send a creator identifier w/o a required nameIdentifierScheme if that was not provided. Now nameIdentifier won't be sent if the scheme is not specified (and the identifier is valid as above.)                                                                                                                                                                              |         |            |       |
| Affiliation is checked and if it is a ROR, identified by                                                                                                                                                                                                                                                                                                                                    |         |            |       |

|                                                                                                                                                                                                                                                                                                                                                                                                      |  |  |  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
| starting with <a href="https://ror.org/">https://ror.org/</a> , the affiliationIdentifierScheme and schemeURI for ROR are now sent.                                                                                                                                                                                                                                                                  |  |  |  |
| QDR has found names/affiliations that are not XML safe (e.g. with &) and has added escaping to the name/affiliation of creators. With OpenAire adding given/family names, these are also escaped.                                                                                                                                                                                                    |  |  |  |
| Currently the DataCite exporter reports the schemeUri for names as, e.g. <a href="http://isni.org/isni/">http://isni.org/isni/</a> whereas the isni example in the OpenAire guide uses <a href="http://isni.org/">http://isni.org/</a> . New code also uses the shorter form. (With ORCID and ROR, it doesn't matter.)                                                                               |  |  |  |
| OpenAire uses SchemeUri whereas the new code uses schemeUri as in the DataCite 4.5 schema                                                                                                                                                                                                                                                                                                            |  |  |  |
| OpenAire doesn't send the required creator element if no authors are listed, The original and new DataCite send a<br><creator><creatorName>:unav</creatorName></creator> element.                                                                                                                                                                                                                    |  |  |  |
| OpenAire sends the Installation Name rather than root collection name as the publisher - new DataCite does as well                                                                                                                                                                                                                                                                                   |  |  |  |
| OpenAire doesn't send publisher if it's empty but it looks like it is required. old/new DataCite send :unav in that case.                                                                                                                                                                                                                                                                            |  |  |  |
| OpenAire uses the latest of publication, distribution, or deposit year as the publicationYear. DataCite uses the production year if set or the distribution year for Harvested datasets, and for internal datasets picks the config defined citation date first, the later of the publication or embargo citation date (the latest date of embargoed files on the first version), or the last update |  |  |  |

|                                                                                                                                                                                                                                                                                                                                                                                           |  |                                                                                                                                      |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--------------------------------------------------------------------------------------------------------------------------------------|--|
| date if nothing else exists (will also pick 'now' if nothing is set but a last update date is now required). Keeping the DataCite logic for now.                                                                                                                                                                                                                                          |  |                                                                                                                                      |  |
| OpenAire doesn't handle files, the new/old DataCite do, new adds file tags as subjects for files                                                                                                                                                                                                                                                                                          |  |                                                                                                                                      |  |
| New DataCite xml escapes subjects and contributors                                                                                                                                                                                                                                                                                                                                        |  |                                                                                                                                      |  |
| OpenAire excludes "Funder" from the contributors list. QDR doesn't have Funder as an option, but the check has been added to the new DataCite code. FWIW: There's a note elsewhere that Funder is not i18n friendly - potentially should check for any variation of the Funder CVV/alternates available?                                                                                  |  |                                                                                                                                      |  |
| OpenAire maps release Time to type "Updated", new DataCite uses type "Available"                                                                                                                                                                                                                                                                                                          |  | New DataCite sends original (v1.0) publication date as Available and the version releaseDate as an 'Updated' date if this is > v1.0. |  |
| OpenAire doesn't write dateOfCollection if start or end is missing, new DataCite writes an open ended range per the RKMS-ISO8601 standard                                                                                                                                                                                                                                                 |  |                                                                                                                                      |  |
| OpenAire doesn't write a language element for the primary language of the resource, new DataCite doesn't either, but could pick the first entry in the languages field/add the element if there is only one value there? Or? FWIW: DataCite schema says 0 or more, OpenAire says 0 or 1, Allowed values are taken from IETF BCP 47, ISO 639-1 language codes (not 639-2 or 3 as we have). |  |                                                                                                                                      |  |
| OpenAire sent otherIds as alternatives w/o the mandatory alternativeIdentifierType attribute if the otherIdAgency wasn't set, new DataCite adds ":unav" in that case                                                                                                                                                                                                                      |  |                                                                                                                                      |  |

|                                                                                                                                                                                                                                                                                                                                                                                                |  |  |  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
| OpenAire did not consider AlternativePersistentIdentifiers as alternates, new DataCite adds them                                                                                                                                                                                                                                                                                               |  |  |  |
| OpenAire doesn't add IsPart or HasPart relationships between a dataset and datafile(s), new DataCite does                                                                                                                                                                                                                                                                                      |  |  |  |
| OpenAire used IsCitedBy for relatedidentifiers, new DataCite uses IsSupplementTo (until the relationship is added to the citation block)                                                                                                                                                                                                                                                       |  |  |  |
| OpenAire assumes the related IDs publicationURL field is only used for type url and all other types use the publicationIDNumber field, new DataCite checks both, preferring the URL if set, and then tries to canonicalize DOIs and Handles, and, like OpenAire does for the publicationUrl field only, tries to extract a schemeURI for any non DOI/Handle/Url types using an "http..." entry |  |  |  |
| OpenAire uses "SchemeURI" for relatedIdentifiers, new DataCite uses "schemeURI" per the schema                                                                                                                                                                                                                                                                                                 |  |  |  |
| new DataCite writes size, format for a file DOI                                                                                                                                                                                                                                                                                                                                                |  |  |  |
| Neither OpenAire or new DataCite currently report info about original files (format, size, ...)                                                                                                                                                                                                                                                                                                |  |  |  |
| New DataCite reports rights for a file, and provides a URL/text for custom terms. Both assume that as long as access requests are enabled, the license/terms allow some access (so the info:eu-repo/semantics/closedAccess doesn't apply, restrictedAccess does)                                                                                                                               |  |  |  |
| New DataCite follows mappings of software, etc. in descriptions - are these all acceptable? New DataCite will send software when version is not supplied                                                                                                                                                                                                                                       |  |  |  |

|                                                                                                                                                                                                                                                                      |  |  |  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
| OpenAire only sent productionPlace as a geoLocationPlace and assumed it was a single value, new DataCite sends all values, and any geographicCoverage entry (as a comma separate string of country, state, city, other)                                              |  |  |  |
| OpenAire assumes a geo box, new DataCite checks if n/s and e/w vals are equal, sends a geoLocationPoint instead                                                                                                                                                      |  |  |  |
| New DataCite includes funder info for files                                                                                                                                                                                                                          |  |  |  |
| New DataCite xml escapes description fields, funder fields                                                                                                                                                                                                           |  |  |  |
| New DataCite xml includes ROR when used in the grantAgency field (e.g. with the ROR external vocab script) - as with ORCID, the ROR, schemaUri are included as attributes and the funder name is sent.                                                               |  |  |  |
| OpenAire sent multiple resourceType entries if there were several kindOfData entries, new DataCite sends one element (which is what the DataCite and OpenAire schemas require) with the kindOfData entries written as the element value in a comma separated string. |  |  |  |
| New DataCite adds a date/range for the citation block "Time Period" as a "Date" of type "Other" with a dateInformation attribute of "Time period covered by the data"                                                                                                |  |  |  |
| OpenAire doesn't recognize ROR URLs as URIs in author affiliation, new DataCite does, and will leverage external cvv to get the institution name if configured                                                                                                       |  |  |  |
| Bigger:                                                                                                                                                                                                                                                              |  |  |  |

|                                                                                                                                                                                                                                                                                                                                                               |  |  |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
| OpenAire and the old DataCite ignore the dataset metadata language and the machine's default language. The new code could use it to add xml:lang tags where appropriate (nominally for user entered metadata fields, not on all xml elements, e.g. PIDs, using the machine's lang for global vals like the publisher (which is the installation brand name)). |  |  |  |
| OpenAire doesn't handle file PIDs or draft datasets (since you can't export metadata for those) - new DataCite does (and can't be a stand-alone exporter for this reason)                                                                                                                                                                                     |  |  |  |
| What should files inherit w.r.t. all new fields, e.g. contact, etc.(Cache dataset vals to avoid recreating for each file?)                                                                                                                                                                                                                                    |  |  |  |