**CIEmetaDigitalProduct - Description**

Version 3

The CIEmetaDigitalProduct meta schema is based on the most recent version of the DataCite schema (<https://schema.datacite.org/meta/kernel-4.4/> ) . Additional, CIE relevant fields are defined as outlined below.

The JSON schema description can be found here: <https://doi.org/10.25039/CIE.SC.4taqevcd>

Following points have to be considered (the ID reference to the number in the Datacite 4.4 Schema), the obligation (mandatory, recommended, option may be different to Datacite 4.4):

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 1 | Identifier (with mandatory type sub-property) | M |

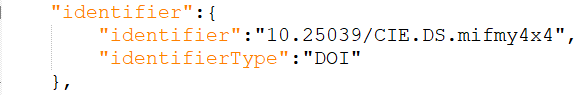
The standard identifier for CIE datasets is a DOI composed as following:

10.25039/CIE.DS.$$$$$$$$

$$$$$$$$ is a composition of random numbers and characterize excluding the following symbols (to avoid confusions): oOlL1I0

The DOI is typically attributed by CIE CB.

Example:



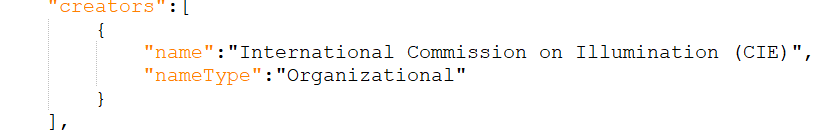
If a publication is translated into another language, usually same dataset is used however a "translated" metadata file may be generated, using the same number but the language code is added.

Example: "identifier":"10.25039/CIE.DS.mifmy4x4.ES"

Attention: not all terms shall be translated! In particular to JSON field names and datatableInfo (except the title and description) shall not be translated but the original english language shall be used.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 2 | Creator | M |

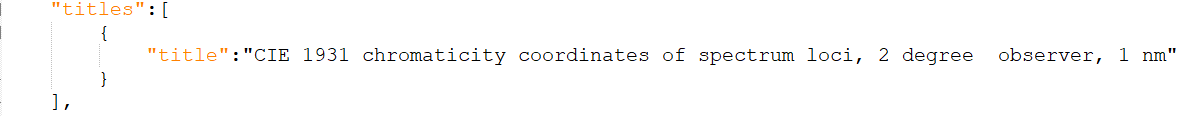
The default creator is:



|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 3 | Title (with optional type sub-properties) | M |

Usually only on title is given without sub-properties. For translated documents and additional title using the titleType "translatedType" can be used. However special characters (i.e. Chinese or russion letter) were not yet tested.

Example:



|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 4 | Publisher | M |

The default publisher is:

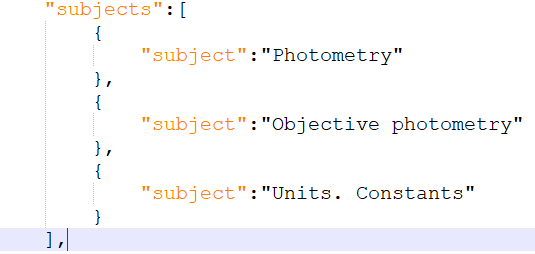


|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 5 | PublicationYear | M |

The year of publication of the original dataset shall be stated (in case that the related publication is translated into another language few years later than the original publication, to original publication shall be used).

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 6 | Subject (with scheme sub-property) | M |

Subject, keyword, classification code, or key phrase describing the resource.



Usually the keywords of the related publication are reused.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 7 | Contributor | O |

Usually not attributed in CIE datasets

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 8 | Date | O |

Usually not attributed in CIE datasets

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 9 | Language | R |

The primary language of the resource according ISO 639-1 language codes.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 10 | ResourceType | M |

For Datasets including "data tables" the ResourceTypeGeneral "Dataset" is used as the recourceType "dataTable". Other types could be "ComputationalNotebook" for excel calculation sheets, "Image", "Audiovisual ","Software",

Examples:

"types":{

"resourceType":"dataTable",

"resourceTypeGeneral":"Dataset"

}

or

"types":{

"resourceType":"ComputationalNotebook",

"resourceTypeGeneral":"Excel"

}

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 11 | AlternateIdentifier (with type sub-property) | R |

For datasets, computationalnotebooks typically the filename is given:

"alternateIdentifiers":[

{

"alternateIdentifier":"CIE\_cc\_1931\_2deg.csv",

"alternateIdentifierType":"fileName"

}

],

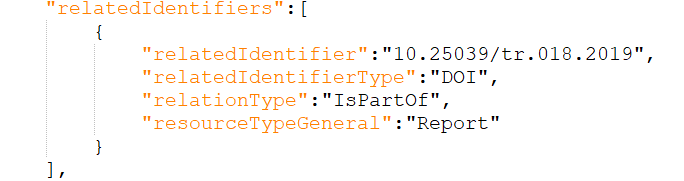
I short but meaningful file name shall be selected. White characters (space) shall be replaced by "\_" characters. English language shall be used.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 12 | RelatedIdentifier (with type and relation type sub-properties) | R |

The identifier (DOI) of the related publication(s) shall be specified. In case of a new publication relating to the existing dataproduct, the metadatafile needs to be updated. Also the relation to other datasets can be specified, typically:

IsSupplementTo, IsSupplementedBy, HasVersion, IsVersionOf, IsNewVersionOf, IsPreviousVersionOf IsPartOf, HasPart, IsPublishedIn, IsReferencedBy, References, IsVariantFormOf, IsOriginalFormOf, IsIdenticalTo, IsDerivedFrom, IsSourceOf, IsRequiredBy Requires.

As an example:



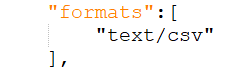
|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 13 | Size | O |

Optional, however care has to be taken that the actual size is used and not the size on the data storage (usually larger)

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 14 | Format | R |

Use MIME type where possible (see <https://www.iana.org/assignments/media-types/media-types.xhtml> for the official definition. Alternatively the file extension can be given.

Example:



Recommendation: The standard data format for CIE data table is CSV according RFC4180 (<https://datatracker.ietf.org/doc/html/rfc4180> ), without header line. The fields are separated by comma.

Other formats can be included however, a reference to the format description shall be provided.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 15 | Version | O |

Usually not used for dataset (in case of a revision, a new dataset should be provided), however for other dataProducts (i.e. file-format) this entry could be used.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 16 | Rights | R |

Needs to be define by CIE/CB

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 17 | Description (with type sub-property) | R |

Description of the dataproduct. Usually the sub-property "abstract" is used. Also the original source of the dataproduct shall be stated. Example:

"descriptions":[

{

"description": "locis in a chromaticity diagram of points that represent monochromatic stimuli, CIE 1931, 2degree observer, original source: CIE 018:2019, Table 6",

"descriptionType":"Abstract"

}

],

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 18 | GeoLocation | O |

Usually not used

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 19 | FundingReference | O |

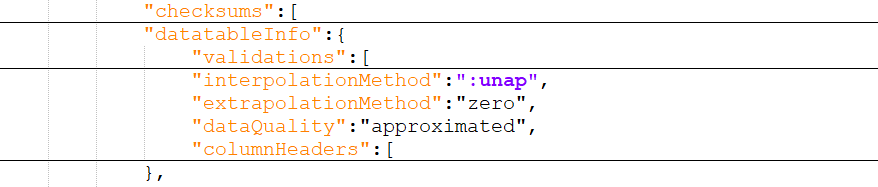
Usually not used

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| 20 | RelatedItem | O |

Usually not used

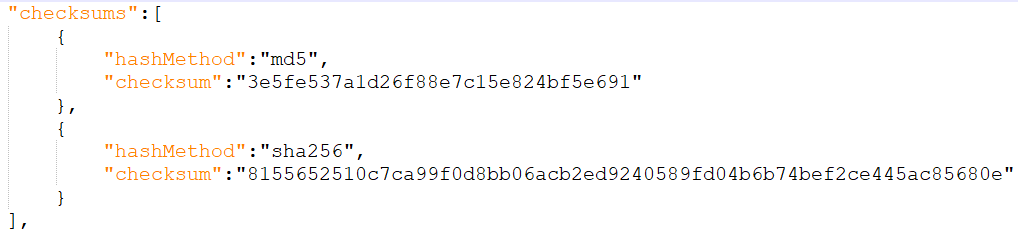
In addition to all fields defined above by the Datacite 4.4 schema, CIE dataProducts

The basic structure of a file that contains the metadata and is integrated into the document contains the following fields:



|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| CIE 1 | checksums | R |

An important point is to give the user the possibility to check the integrity of the data itself (as a human and as a machine). For this purpose, hash algorithms are usually used (a kind of checksum). There are different standards (e.g. md5 <https://en.wikipedia.org/wiki/MD5> or sha256 <https://en.wikipedia.org/wiki/SHA-2> ) If only a single letter or number in the document is changed, the checksum changes. These cryptographic algorithms are directly integrated in most operating systems and can be used from any software (Excel®, Python, Labview, Matlab). There is also freeware that can be used to generate the checksum. In the CIE metadata system this could be mapped as follows (example of 2 hash values):



The following CMD can be used on any WinOS to generate the hashes:

>certutil -hashfile "filename.exe" MD5

>certutil -hashfile "filename.exe" SHA256

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| CIE 2.1 | datatableInfo interpolationMethod | R |

Many dataTables represent data as a function of wavelength. If the wavelength spacing is different to the values given in the dataset, interpolation algorithm is typically used. The following methods can be used:

|  |  |
| --- | --- |
| Value | Description |
| "nearest" | Chooses the Y value corresponding to the X value that is nearest to the current xi value |
| "linear" | Sets the interpolated values to points along the line segments connecting the X and Y data points. |
| "cubic-spline" | Guarantees that the first and second derivatives of the cubic interpolating polynomials are continuous, even at the data points. |
| "cubic-Hermite" | Guarantees that the first derivative of the cubic interpolating polynomials is continuous and sets the derivative at the endpoints to certain values in order to preserve the original shape and monotonicity of the **Y** data. |
| "Sprague" | Sprague 5 point interpolation as outlined in CIE 167:2005 |
| " Lagrange" | Lagrange Interpolation |
| "useRelatedDataset" | In some cases interpolation is not recommended but a dataset with different wavelength range is recommended. Example is the 5 nm spectral data given in CIE 015. For those the 1 nm data should be used as published with CIE 018:2019. The related dataset should be stated in the corresponding relatedIdentifier field. |
| "useRelatedFormula" | In some cases interpolation is not recommended but an explicitely formula shall be used. Example is the Standard illuminant A. The reference to the formula shall be described in the description of the dataset |
| ":unal" | unallowed, suppressed intentionally |
| ":unap" | not applicable, makes no sense |
| ":unas" | value unassigned (e.g., Untitled) |
| "other" | The method of interpolation shall be stated in the description |

The same interpolationMethod parameter applies to all data provided by the given dataset. If different interpolation methods have to be used, the dataset shall be splited.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| CIE 2.2 | datatableInfo extrapolationMethod | R |

The property can have the following values:

|  |  |
| --- | --- |
| Value | Description |
| "nearest" | Chooses the Y value corresponding to the X value that is nearest to the current xi value (i.e. Y stays constant to the first/last value |
| "zero" | Values are set to zero for X values outside the xi range |
| "useRelatedDataset" | In some cases interpolation is not recommended but a dataset with different wavelength range is recommended. |
| "useRelatedFormula" | In some cases interpolation is not recommended but an explicitely formula shall be used. |
| ":unal" | unallowed, suppressed intentionally |
| ":unap" | not applicable, makes no sense |
| ":unas" | value unassigned (e.g., Untitled) |
| "other" | The method of interpolation shall be stated in the description |

The same extrapolationMethod parameter applies to all data provided by the given dataset. If different extrapolation methods have to be used, the dataset shall be splited.

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| CIE 2.3 | datatableInfo dataQuality | R |

The property can have the following values:

|  |  |
| --- | --- |
| Value | Description |
| "nominal" | The stated values are nominal values (i.e. without uncertainty, etc) |
| "approximated" | The stated values are approximated values (typically the result of a computation) |
| ":unap" | not applicable, makes no sense |
| ":unas" | value unassigned (e.g., Untitled) |
| "other" | The method of interpolation shall be stated in the description |

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| CIE 2.4 | datatableInfo columnHeaders | R |
| CIE 2.4.1 | title | R |
| CIE 2.4.2 | Quantity | R |
| CIE 2.4.3 | Unit | R |
| CIE 2.4.4 | Description | O |

For data tables that are typically distributed CIE publication, additional information shall be provided to increase the machine interpretability of the datasets. In particular, for each column:

Title: The title includes typically the quantity and additional relevant information

Quantity: The generic quantity

Unit: the unit associated with the quantity ("dimensionless" for relative quantity or those that are based on counting).

Description: Additional information about the column can be stated. This information can also be given in a structure JSON string. Example

Example for CIE 241:2020 (data sets for spectral distributions):

"description":"{"airmass":1.0,"water vapour": 1.42,"O3": 0.43,"AOD":0.1,"Albedo":0.2}"

Example:

"columnHeaders":[

{

"title":"wavelength",

"unit":"nm",

"quantity":"wavelength"

},

{

"title":"x(lambda)",

"unit":"dimensionless",

"quantity":"chromaticity coordinates"

},

{

"title":"y(lambda)",

"unit":"dimensionless",

"quantity":"chromaticity coordinates"

}

|  |  |  |
| --- | --- | --- |
| ID | Property | Obligation |
| CIE 2.5 | datatableInfo validation | R |

Additional validation information can be provided (the field can be repeated):

|  |  |
| --- | --- |
| Value | Description |
| "sumOfColumns" | States the sum of each column, in JSON string notation (see example below) |
| "sampleRow" | States a sample row in JSON notation, the row number has to be specified |
| "numberOfRows" |  |
| "numberOfColumns" |  |
| ":unap" | not applicable, makes no sense |
| "other" | The method of interpolation shall be stated in the description |

