**Big Data Catalogue Data Schema**

version: 1.1.1

Last updated: Oct. 3rd, 2019

**Table I-1 Catalogue overview**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Name/Role name** | **Description** | **M/R/O** | **Cardinality** | **Domain** |
| 1 | BDC\_Catalogue | root entity which defines metadata about dataset resource | M | 1 | Line 2 to 12 |
| 2 | role: responsibleParty | responsible party information for the data catalogue | M | N | BDC\_ResponsibleParty |
| 3 | role: dataset | dataset information for the data catalogue | M | N | BDC\_Dataset |
| 4 | role: sourceDomain | source domain information for the data catalogue | M | 1 | BDC\_Domain |
| 5 | role: usageDomain | usage domain information for the data catalogue | R | N | BDC\_Domain |
| 6 | role: dataQuality | data quality information for the data catalogue | R | N | BDC\_DataQuality |
| 7 | role: record | record information for the data catalogue | R | N | BDC\_record |
| 8 | role: hasRelation | related data catalogue of the data catalogue | O | N | BDC\_Catalogue |
| 9 | title | title of data catalogue | M | 1 | free text |
| 10 | description | a short introduction to data catalogue | M | 1 | free text |
| 11 | language | language used within the data catalogue | R | 1 | RFC 5646 |
| 12 | versionInfo | version of data catalogue specification | R | 1 | free text  (major.minor.patch) |

**Table I-2 Responsible party information**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Name/Role name** | **Description** | **M/R/O** | **Cardinality** | **Domain** |
| 13 | BDC\_ResponsibleParty | name and position information for an individual or organization that is responsible for the data catalogue or dataset | M | N | Line 14 to 18 |
| 14 | indivisualName | name of the responsible person- surname, given name, title | M | 1 | free text |
| 15 | organisationName | name of the responsible organization | M | 1 | free text |
| 16 | positionName | role or position of the responsible person | M | 1 | free text |
| 17 | function | function performed by the responsible party | M | 1 | free text |
| 18 | contact | information required to enable contact with the responsible person and/or organization | R | N | BDC\_ContactType |
| 19 | BDC\_ContactType | information required to enable contact with the responsible person and/or organization | R | N | Line 20 to 23 |
| 20 | phoneNumber | telephone numbers at which the organization or individual may be contacted | R | N | free text |
| 21 | address | physical and email address at which the organization or individual may be contacted | R | 1 | free text |
| 22 | onlineResource | on-line information that can be used to contact the individual or organization | R | N | URI |
| 23 | hoursOfServices | time period (including time zone) when individuals can contact the organization or individual | R | 1 | free text |

**Table I-3 Dataset information**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Name/Role name** | **Description** | **M/R/O** | **Cardinality** | **Domain** |
| 24 | BDC\_Dataset | a collection of data which is available for access and/or download | M | N | Line 25 to 32 |
| 25 | role: contactPoint | contact point information for the dataset | R | N | BDC\_ResponsibleParty |
| 26 | role: distribution | information about the distributor of and options for obtaining the dataset | R | N | BDC\_Distribution |
| 27 | role: hasRelation | related dataset of the dataset | O | N | BDC\_Dataset |
| 28 | title | title of the dataset | M | 1 | free text |
| 29 | description | a short introduction to the dataset | M | 1 | free text |
| 30 | keyword | commonly used word(s) used to describe the dataset | R | N | free text |
| 31 | temporalCoverage | time period (including time zone) when the dataset was captured | O | 1 | free text |
| 32 | updateFrequency | rate of occurrence of dataset update | O | 1 | free text |
| 33 | BDC\_Distribution | information about the distributor of and options for obtaining the dataset | R | N | Line 34 to39 |
| 34 | role: dataRight | Information about limitation placed upon the use of the dataset | R | 1 | BDC\_DataRight |
| 35 | dataType | the type information of the dataset | M | 1 | BDC\_DataTypeCode |
| 36 | format | the format information of the dataset including a version | M | 1 | free text |
| 37 | byteSize | information about the size of dataset in bytes | M | 1 | int |
| 38 | accessInformation | access information to dataset | M | 1 | free text |
| 39 | sampleData | location information about a sample data including URL | R | N | URI |
| 40 | BDC\_DataRight | information about limitation placed upon the use of the dataset | R | 1 | Line 41 to 42 |
| 41 | rightStatement | information to the usage permissions being agreed to | R | 1 | free text |
| 42 | license | This links to the license document under which the catalog is made available and not the datasets. | R | N | URI |

**Table I-4 Domain information**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Name/Role name** | **Description** | **M/R/O** | **Cardinality** | **Domain** |
| 43 | BDC\_Domain | source and/ or usage information for data catalogue | M | 1 | Line 44 to 46 |
| 44 | title | a member item of taxonomy which used for identifying the domain of source and /or usage | M | 1 | free text |
| 45 | taxonomy | taxonomy information including classification system and authority | M | N | URI |
| 46 | comment | an additional information for domain | O | N | free text |

**Table I-5 Data quality information**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Name/Role name** | **Description** | **M/R/O** | **Cardinality** | **Domain** |
| 47 | BDC\_DataQuality | data quality information for the data catalogue | R | N | Line 48 to 50 |
| 48 | qualityRequirement | requirement statement for evaluating quality of dataset | M | 1 | free text |
| 49 | unitOfMeausre | standard unit or system of units by means of which a quantity is accounted for and expressed. | M | 1 | free text |
| 50 | qualityResult | the result of quality evaluation described by the unit of measure | M | 1 | free text |

**Table I-6 Catalogue record information**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Name/Role name** | **Description** | **M/R/O** | **Cardinality** | **Domain** |
| 51 | BDC\_Record | record information for the data catalogue | R | N | Line 52 to 53 |
| 52 | modifiedDate | the most resent modified date of data catalogue | M | N | ISO 8601 |
| 53 | issuedDate | the issued date of listing the dataset in the catalogue | R | 1 | ISO 8601 |

**Table I-7 <<Enumeration>> DataTypeCode**

|  |  |  |
| --- | --- | --- |
| **Concept name** | **Domain code** | **Definition** |
| STURECTURED | 001 | data that is organized in a format easily used by formatted repository |
| SEMI\_STRUCTURED | 002 | data that is neither raw data, nor typed data in a conventional database system (e.g. XML, JSON) |
| UNDTRUCTURED | 003 | data that either does not have a pre-defined data model or is not organized in a pre-defined manner |