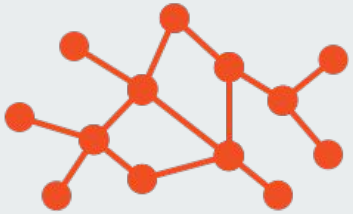
16/10/2020

Generic Dataset Metadata Template

N. Vasiljevic and J. Graybeal



FAIR
Data Collective

Problems to solve

1. Create generic metadata template that can be used across domains
2. Generic medatata template can be mapped to DataCite scheme to secure DOI generation (current currency for researchers)
3. Generic metadata template should be compatible with FDP and/or provide rationale for an FDP update
4. Generic metadata template that can be customized based on FIP inputs

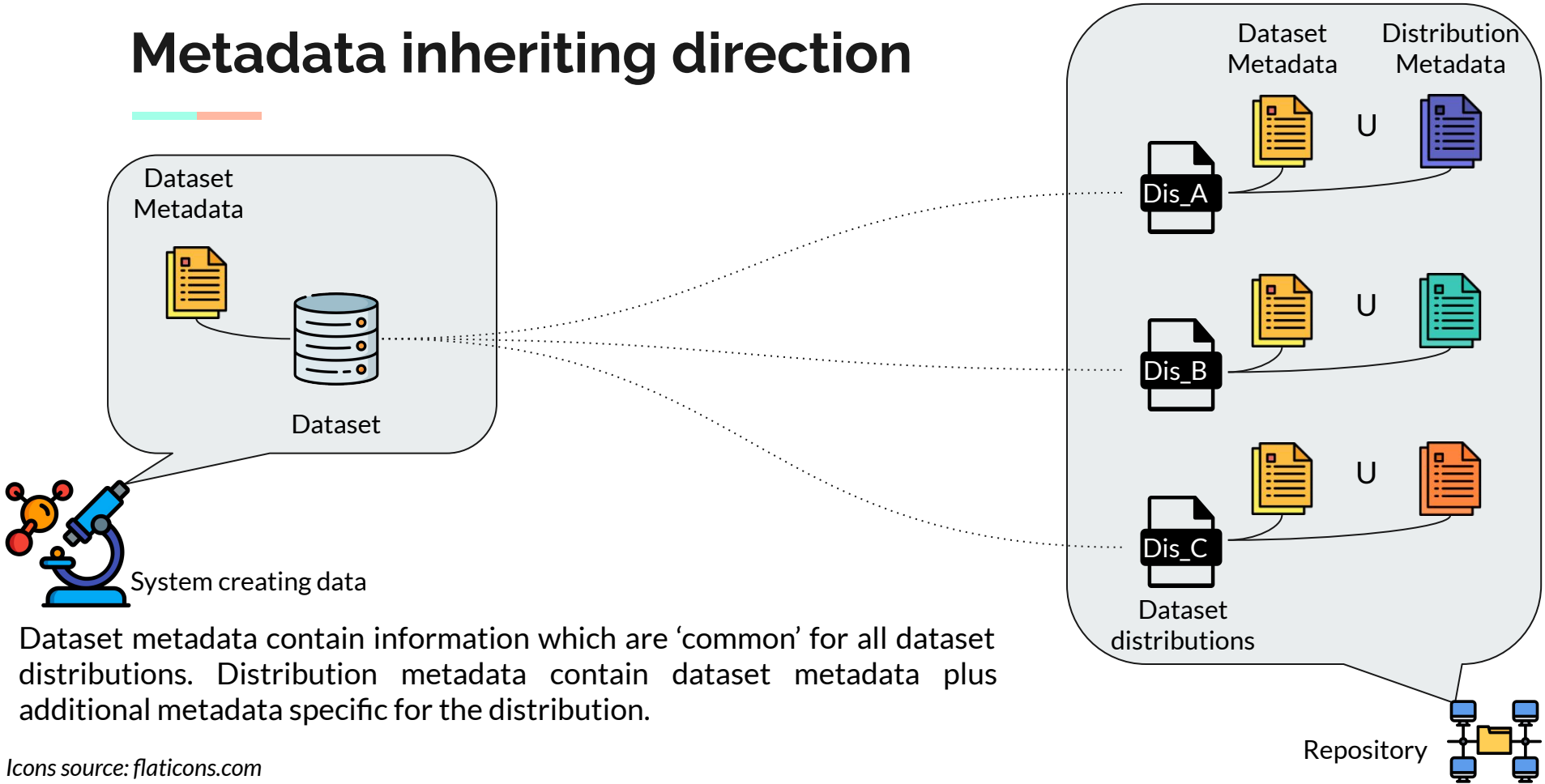
Definitions

- **Dataset** is a logical grouping of one or more **data streams**. It is can be provided in various data **distributions**. Dataset might be part of (parent) **data catalog**.

- **Data stream** represent a collection of one or more **variable records** originating from a single data source (e.g., sensor) which properties remain unchanged during the data stream existence. In some use-cases, the data source **location** is required to remain unchanged (e.g., sensors attached to a meteorological tower). Data stream can grow over time.

Icons source: flaticons.com

Metadata inheriting direction



Dataset metadata contain information which are ‘common’ for all dataset distributions. Distribution metadata contain dataset metadata plus additional metadata specific for the distribution.

Icons source: flaticons.com

FAIR
Data Collective

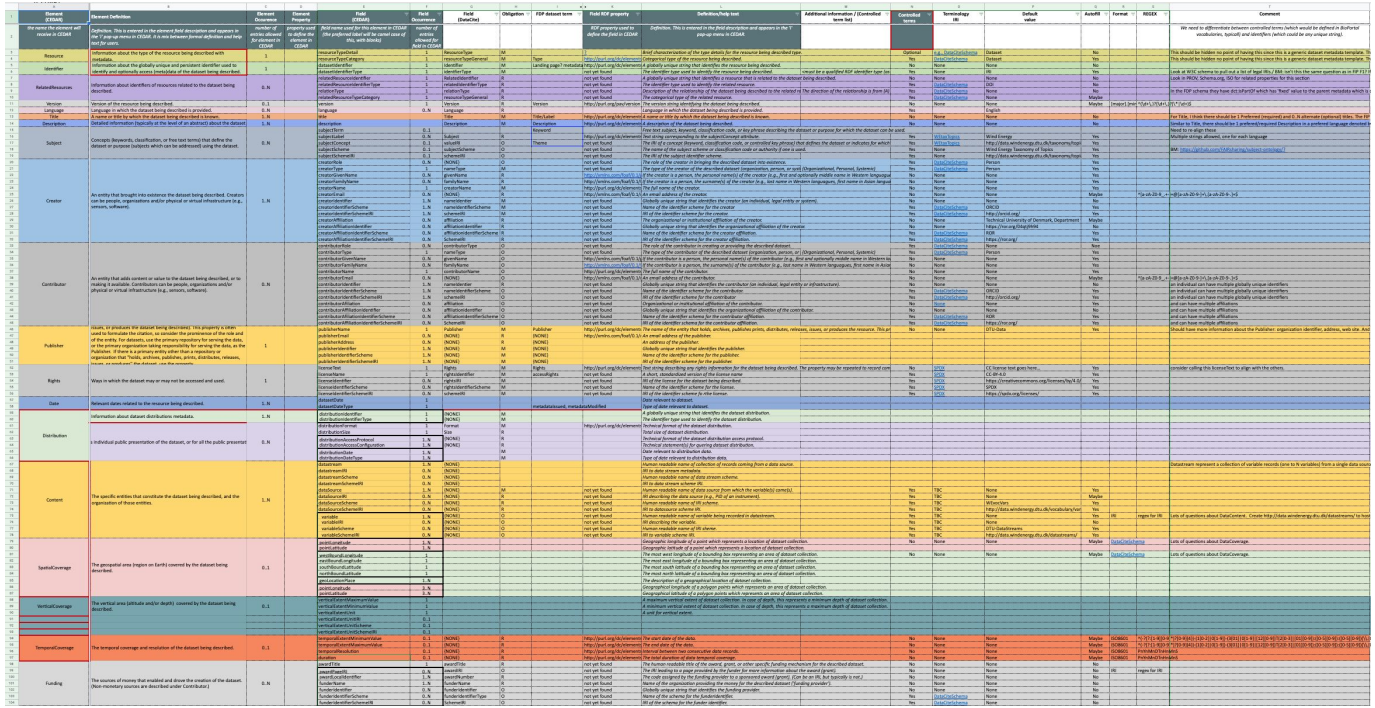
Relations

- DataCite
- FDP Specification
- FIPs
- nanoPubs
- ...

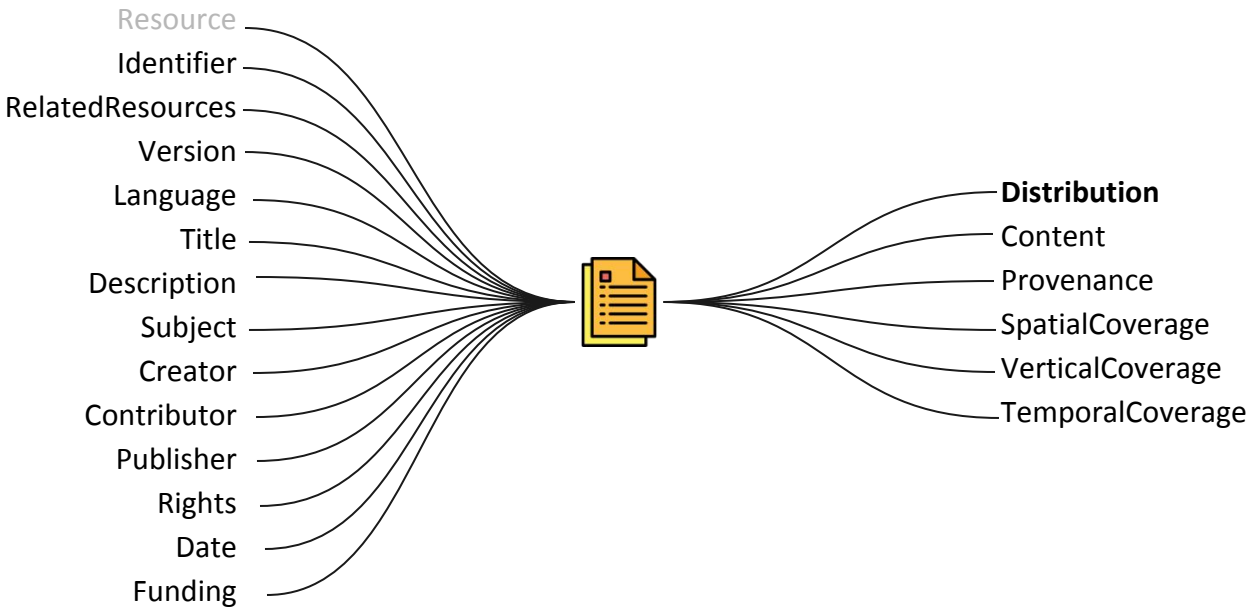
FAIR Data Collective

Source document

Template-description.xlsx



Dataset metadata elements



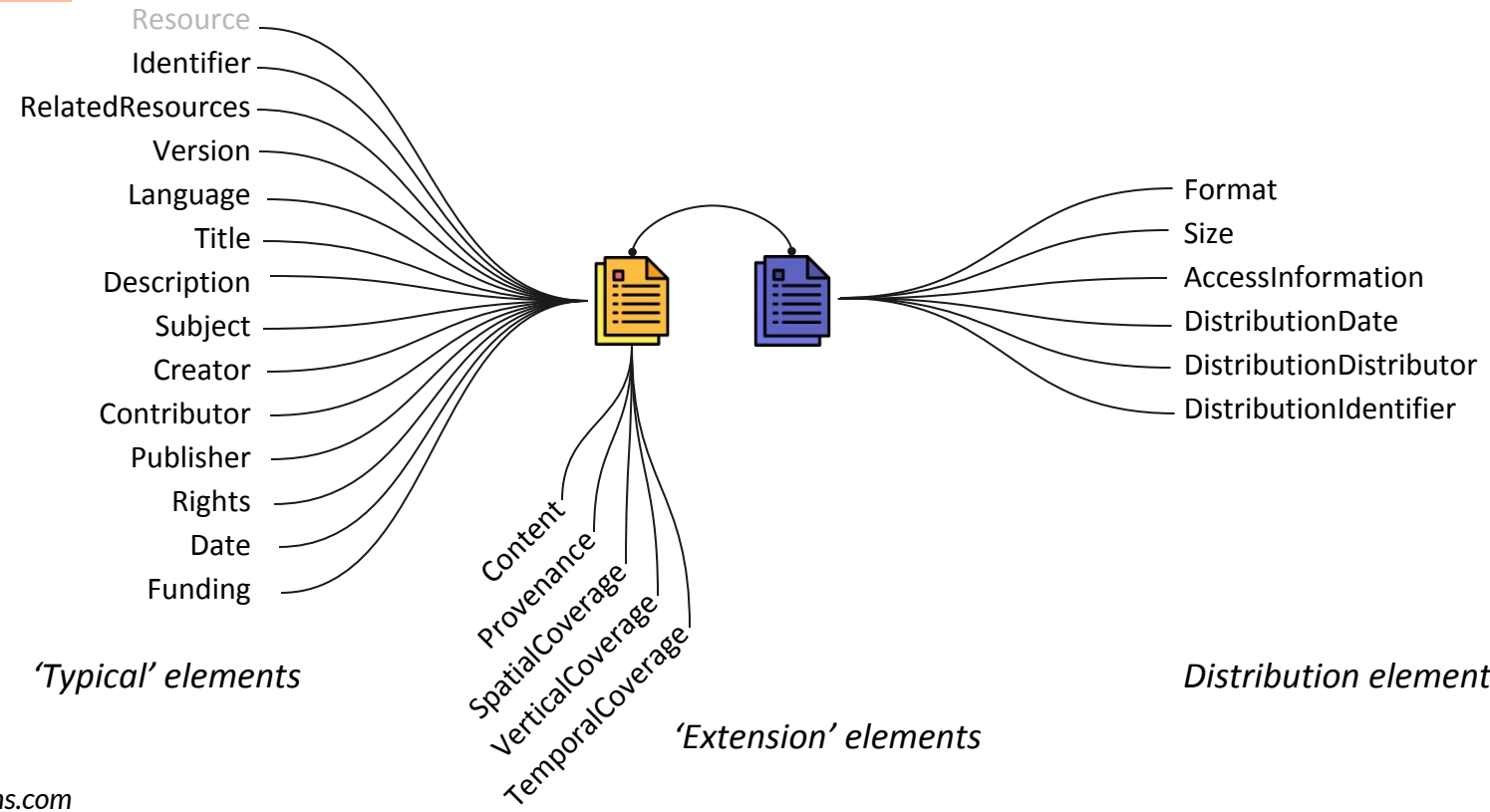
'Typical' elements

'Extension' elements

Icons source: flaticons.com

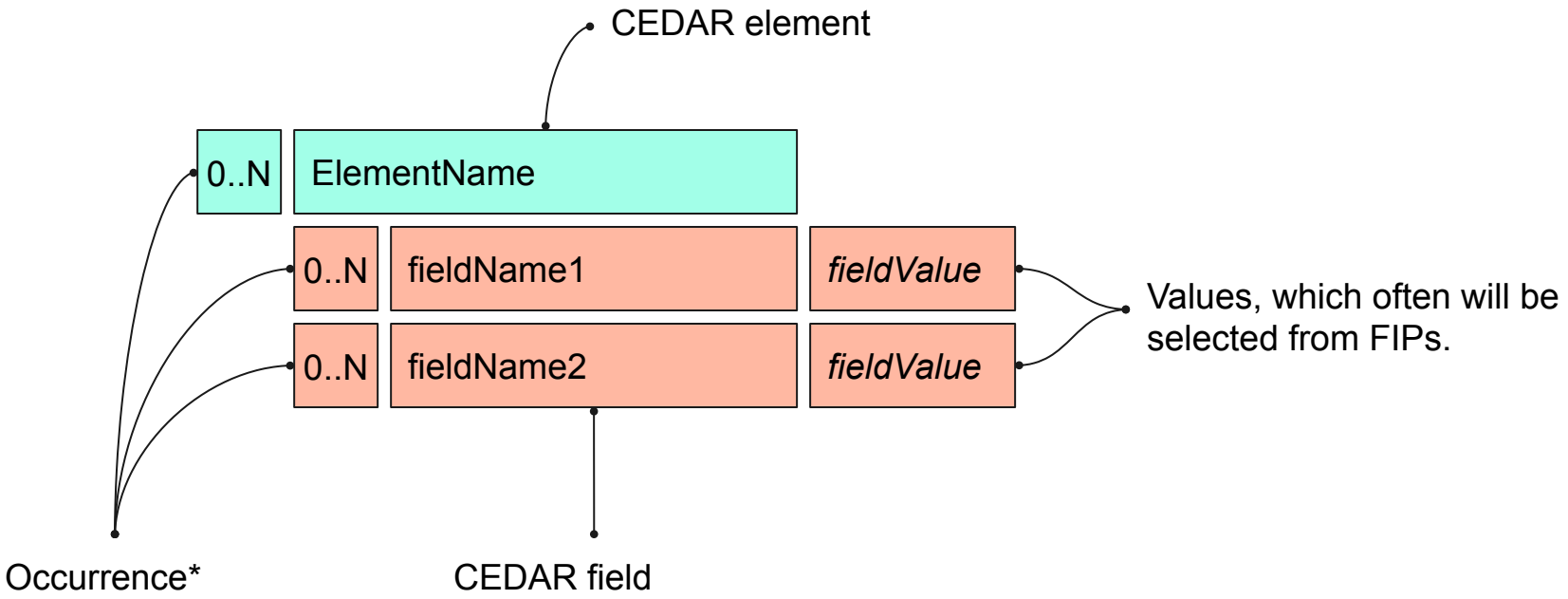
FAIR
Data Collective

Dataset metadata elements



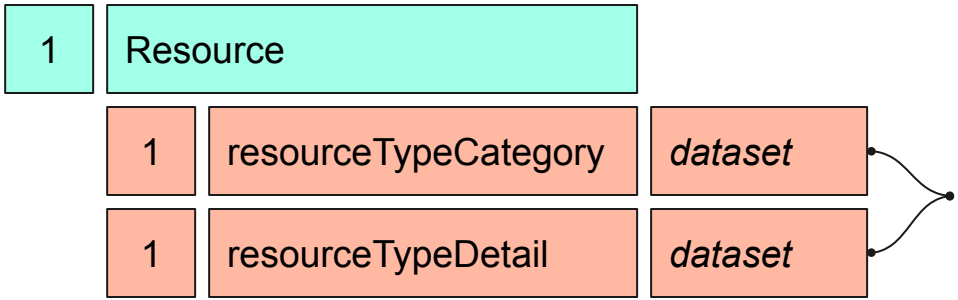
Icons source: flaticons.com

Legend



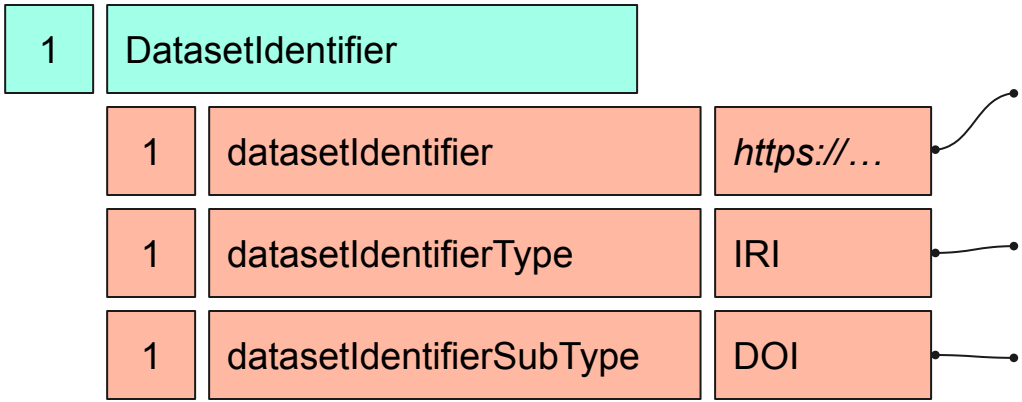
** if 1 or 1..N, then mandatory*

Resource



Fixed value since we know that the resource being described is dataset.

Dataset identifier



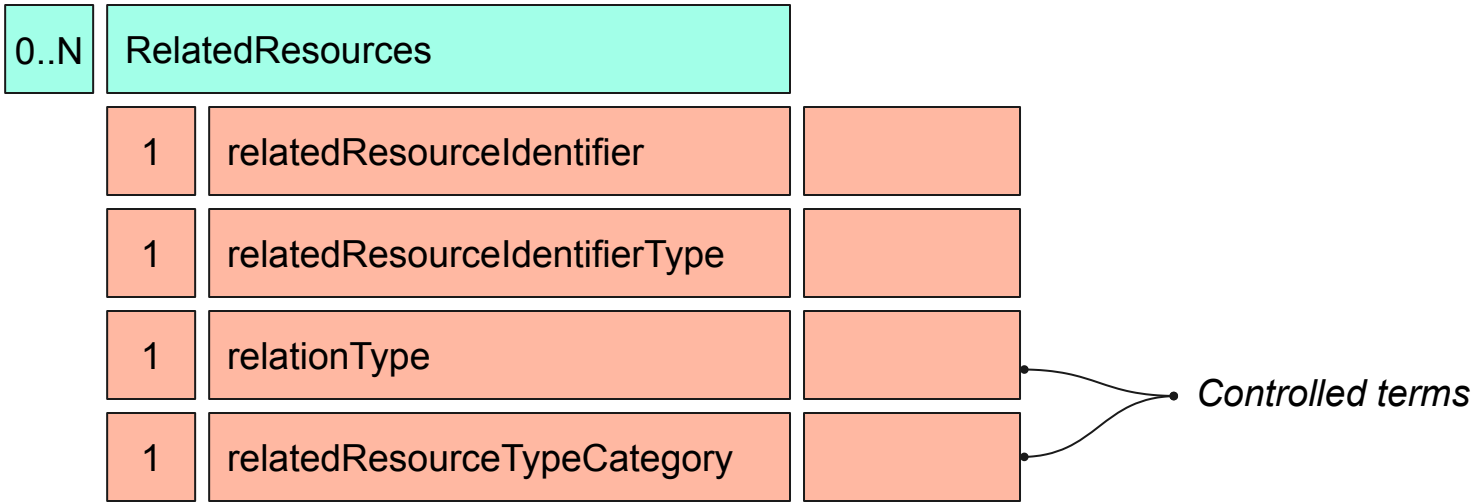
Value provided by identifier registration mechanism. This must be an IRI because FAIR principles require it to be resolvable.

This must always be an IRI.

Data stewards select values based on their FIPs. Typical values could include DOI, TrustyURL, or Handle; a regular IRI would be IRI, URI, URL, or blank.

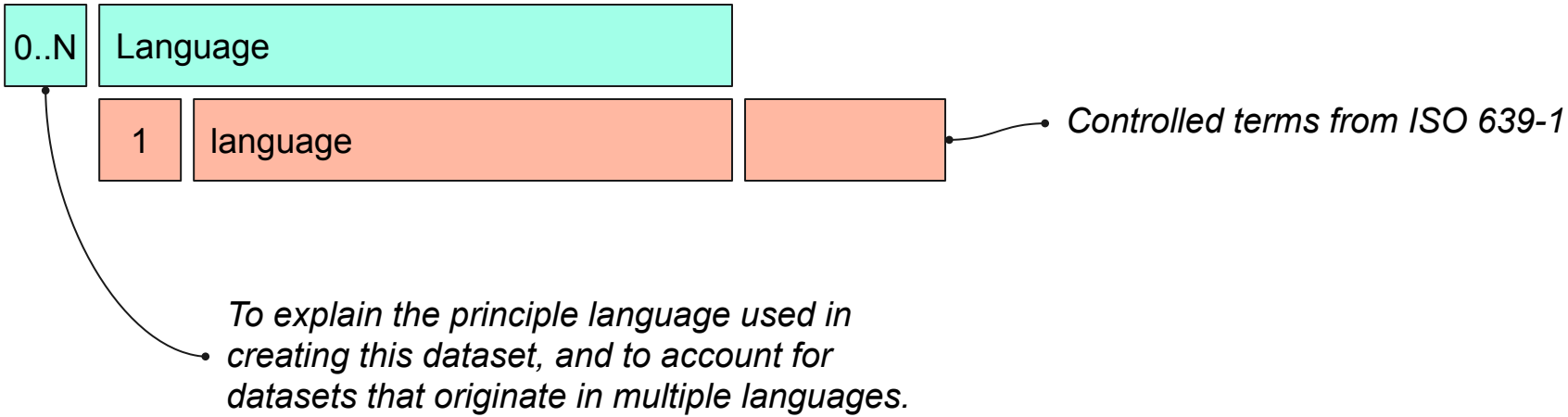
FAIR
Data Collective

Related resources



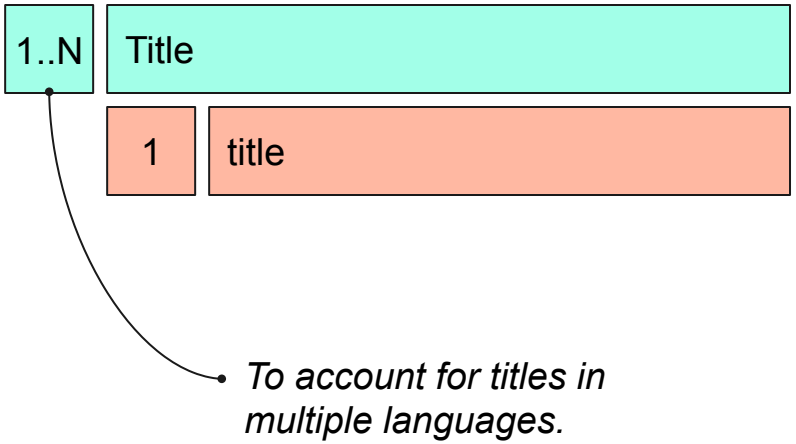
Version

Language

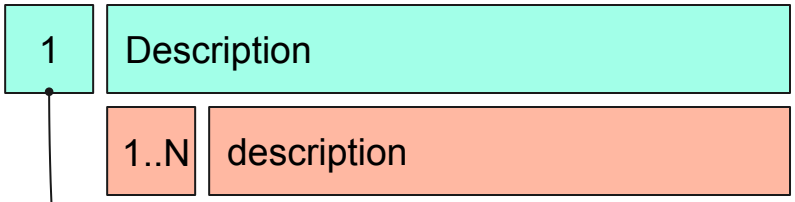

0..N
Language
1
language
Controlled terms from ISO 639-1
To explain the principle language used in creating this dataset, and to account for datasets that originate in multiple languages.

FAIR
Data Collective

Title



Description



To account for dataset descriptions in multiple languages.

FAIR
Data Collective

Subject

0..N	Subject

1..N	subjectTerm	*When subjects are free text.*

1..N	subjectLabel	*When subjects are sourced from controlled terminologies or external authorities.*
0..1	subjectConcept	
0..1	subjectScheme	
0..1	subjectSchemeIRI	

To account for labels in multiple languages.

Contributor fields/elements = Creator fields/elements

Creator

1..N	Creator	
1	creatorType	Creator can be a person, organization or system.
0..N	creatorRole	
	CreatorName	
	CreatorIdentifier	
	CreatorAffiliation	
0..N	creatorEmail	

Creator contd. /1

1	CreatorName
0..1	creatorGivenName
0..1	creatorFamilyName
1	creatorName

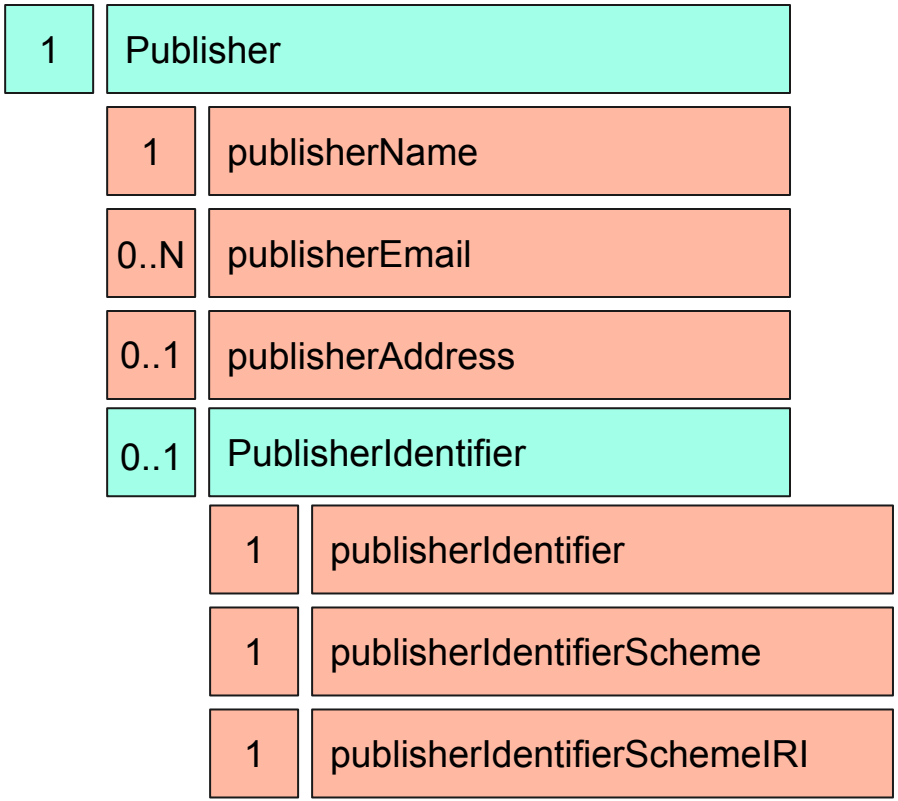
Creator contd. /2

1..N	CreatorIdentifier
1	creatorIdentifier
0..1	creatorIdentifierScheme
0..1	creatorIdentifierSchemeIRI

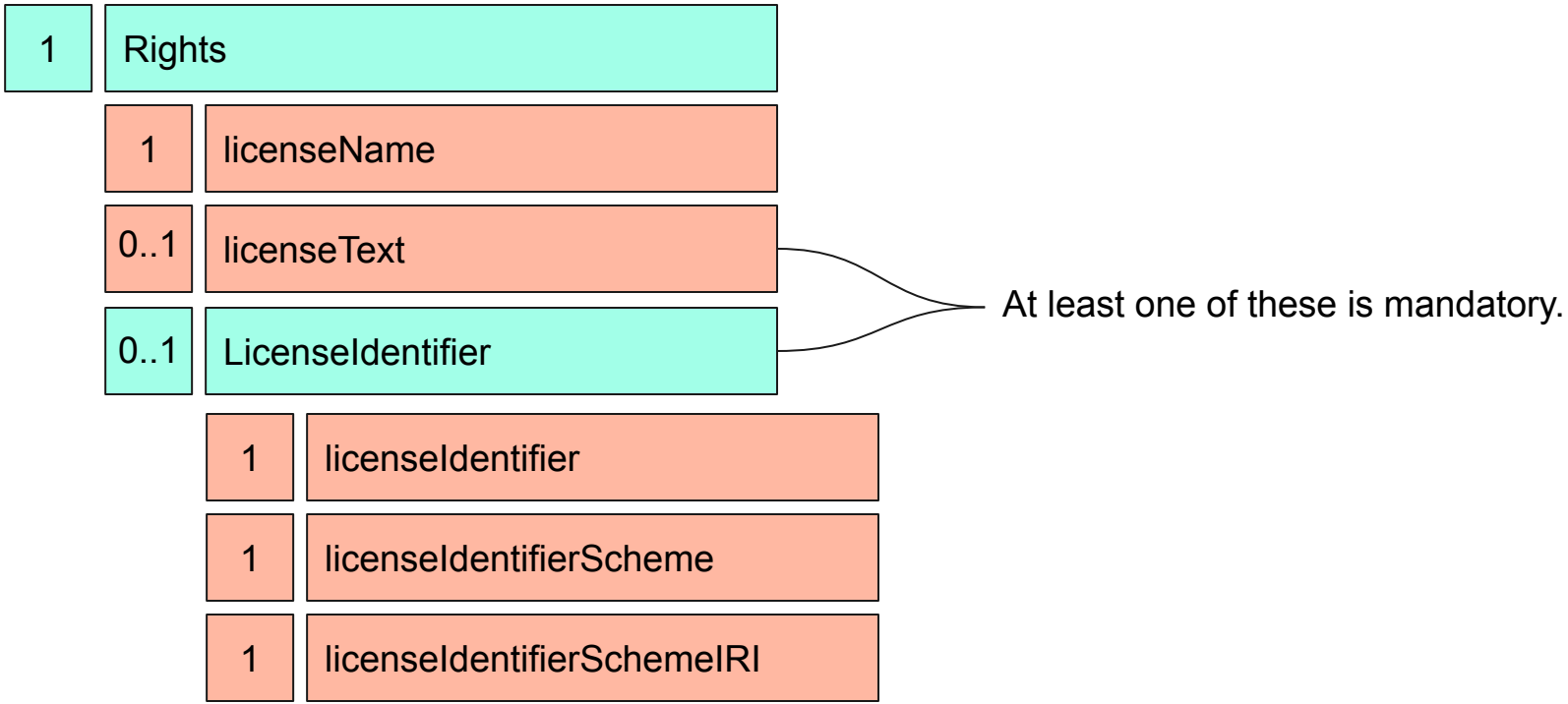
Creator contd. /3

0..N	CreatorAffiliation
1	creatorAffiliationIdentifier
0..1	creatorAffiliationIdentifierScheme
0..1	creatorAffiliationIdentifierSchemeIRI

Publisher


1
Publisher
1
publisherName
0..N
publisherEmail
0..1
publisherAddress
0..1
PublisherIdentifier
1
publisherIdentifier
1
publisherIdentifierScheme
1
publisherIdentifierSchemeIRI

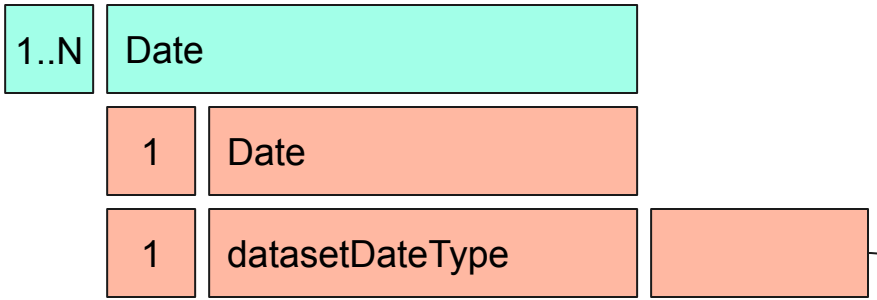
Rights


1
Rights
1
licenseName
0..1
licenseText
0..1
LicenseIdentifier
1
licenseIdentifier
1
licenseIdentifierScheme
1
licenseIdentifierSchemeIRI
At least one of these is mandatory.

Dataset Publication Date

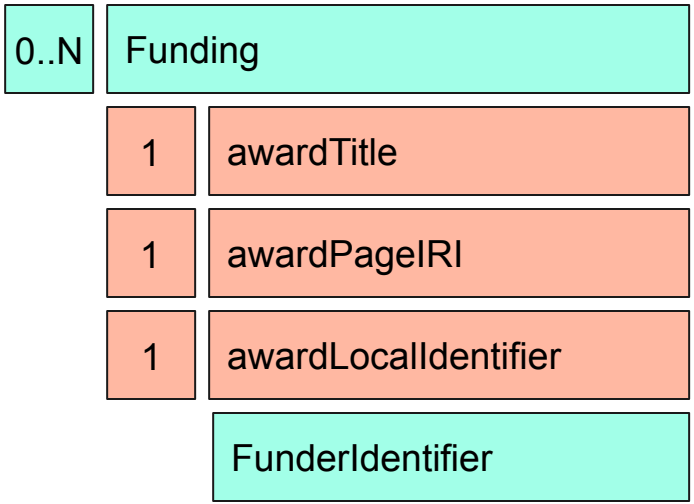
1	DatasetPublicationDate	
1	datasetDate	
1	datasetDateType	*Published*

Date Element (reusable)

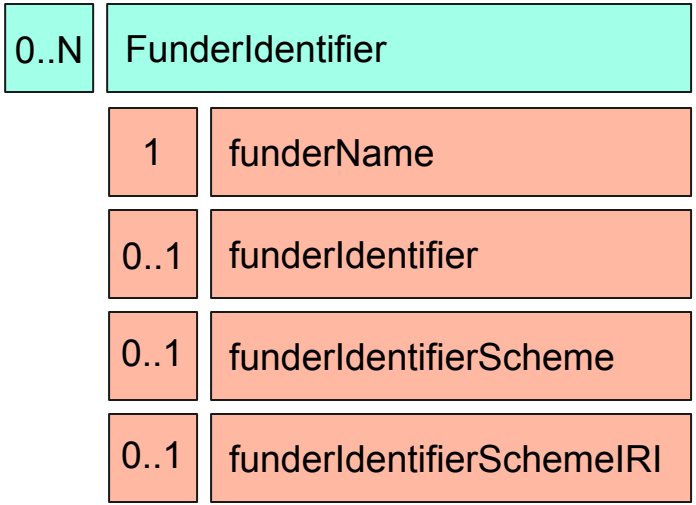


Describes the role this date represents.
One of Published, Updated, Created, …

Funding

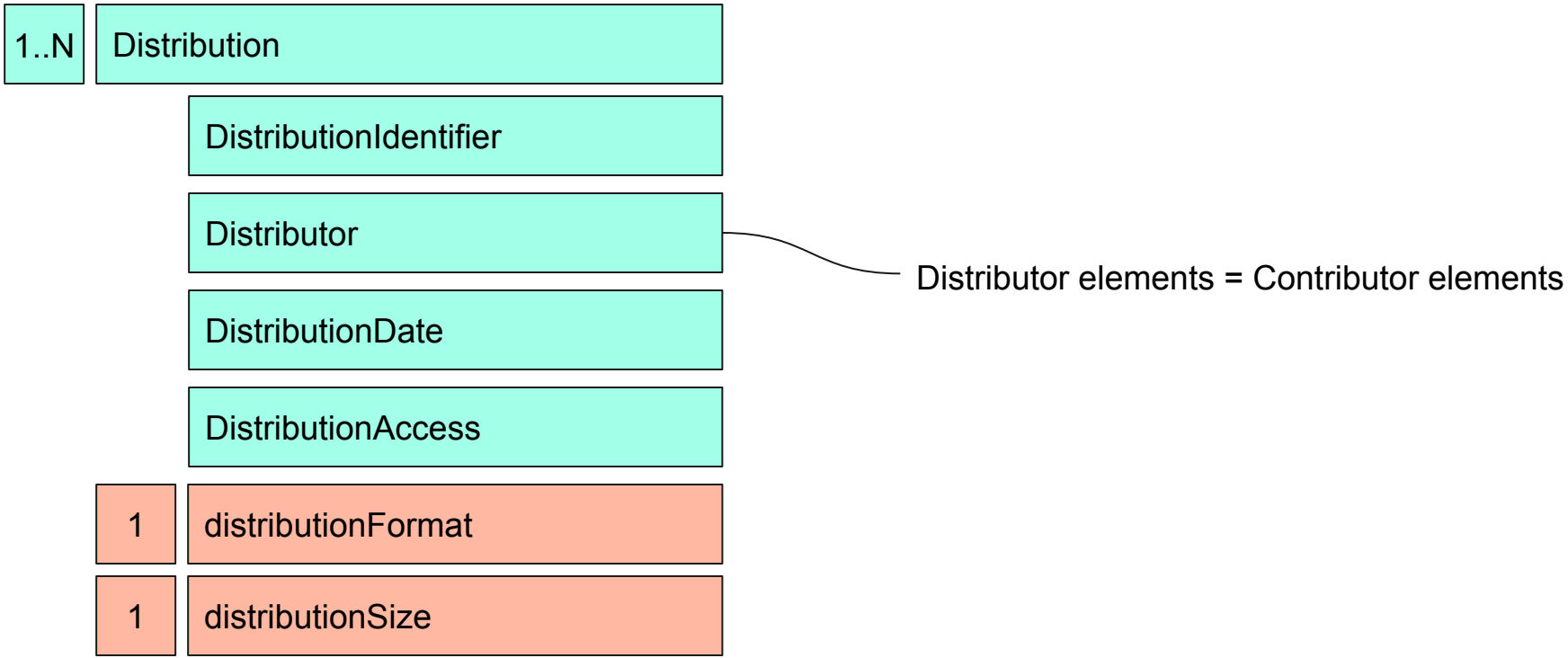

0..N
Funding
1
awardTitle
1
awardPageIRI
1
awardLocalIdentifier
FunderIdentifier

Funding contd.



FAIR
Data Collective

Distribution



FAIR
Data Collective

Distribution contd. /1

1	DistributionIdentifier
1	distributionIdentifier
1	distributionIdentifierType

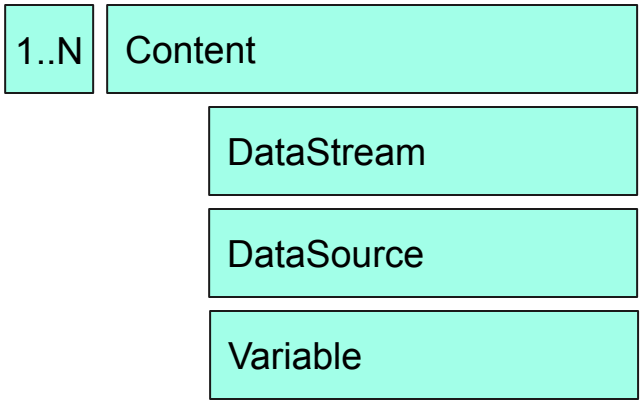
Distribution contd. /2

1..N	DistributionDate
1	distributionDate
1	distributionDateType

Distribution contd. /3

1..N	DistributionAccess
1	distributionAccessProtocol
1..N	distributionAccessConfiguration

Content



Content contd. /1

1	DataStream
1	dataStream
0..1	dataStreamIRI
0..1	dataStreamScheme
0..1	dataStreamSchemeIRI

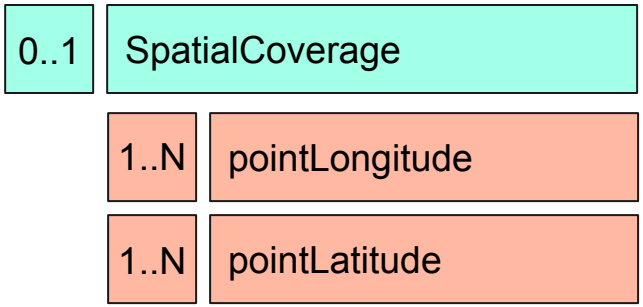
Content contd. /2

1	DataSource
1	dataSource
0..1	dataSourceIRI
0..1	dataSourceScheme
0..1	dataSourceShemeIRI

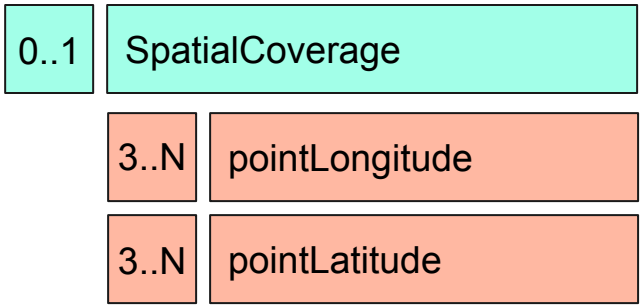
Content contd. /3

1..N	Variable
1	variable
0..1	variableIRI
0..1	variableScheme
0..1	variableSchemeIRI

SpatialCoverage via collection of points



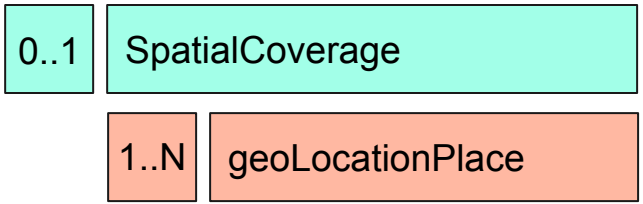
SpatialCoverage via polygon



SpatialCoverage via bounding box

0..1	SpatialCoverage
1	westBoundLongitude
1	eastBoundLongitude
1	southBoundLatitude
1	northBoundLatitude

SpatialCoverage via geo places



VerticalCoverage

0..1	VerticalCoverage
1	verticalExtentMaximumValue
1	verticalExtentMinimumValue
1	verticalExtentUnit
0..1	verticalExtentUnitIRI
0..1	verticalExtentUnitScheme
0..1	verticalExtentUnitSchemeIRI

TemporalCoverage

0..1	TemporalCoverage
0..1	temporalExtentMinimumValue
0..1	temporalExtentMaximumValue
0..1	duration
0..1	temporalResolution

Work to be done

- Assign RDF properties to elements and fields
- Verify mapping of the generic template to FDP / DataCite
- Create CEDAR elements, fields and templates
- Make Github repo public, potentially move it from user account to Github org
- Nanopublications
- Provide inputs for improvements of FDP
- Improve slide deck for M4M

Thank you.

FAIR
Data Collective