Co-matching IVOA and DOI metadata: how do they complement each other?

Pierre Le Sidaner

IVOA (International Virtual Observatory Alliance) :

started in mid 2002 worldwide scientific organisation to facilitate sharing data in Astronomy :

- **Define standards to ensure interoperability.**
- **information system to distribute data and services in Astronomy and solar system data**

Datacite :

DataCite (end 2009) is an international non profit organization which aims to improve data citation in order to:

- **establish easier access to research data on the Internet**
- **increase acceptance of research data as legitimate, citable contributions to the scholarly record**
- **support data archiving that will permit results to be verified and re-purposed for future study**

IVOA is thematic (Astronomy), dedicated to share data and developed a set of Metadata and semantic focused on Interoperability
Funded by observatories and data center

Datacite : developed to handle digital objects

Funded by libraries to provide easier access to research data. now associated to Identifier DOI

IVOA eco-system

- Declare a Resource (VOResource) in yellow pages (Registry)

metadata inherited from dublin core

- Dublin Core was created in 1995 to describe networked resources (physical or electronic documents)
- web pages and library objects

First sets of 13 metadata in 1995, 15 in 1996 ...

TITLE: The name given to the resource by the CREATOR or PUBLISHER.
AUTHOR OR CREATOR: The person(s) or organization(s) primarily responsible for the intellectual content of the resource.
SUBJECT AND KEYWORDS: The topic of the resource, or keywords, phrases, or classification descriptors that describe the subject or content of the resource.
DESCRIPTION: A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources.
PUBLISHER: The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity.
OTHER CONTRIBUTORS: Person(s) or organization(s) in addition to those specified in the CREATOR element who have made significant intellectual contributions to the resource but whose contribution is secondary to the individuals or entities specified in the CREATOR element.
DATE: The date the resource was made available in its present form.
RESOURCE TYPE: The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. It is expected that RESOURCE TYPE will be chosen from an enumerated list of types.
FORMAT: The data representation of the resource, such as text/html, ASCII, Postscript file, executable application, or JPEG image. FORMAT will be assigned from enumerated lists such as registered Internet Media Types (MIME types).
RESOURCE IDENTIFIER: String or number used to uniquely identify the resource. Examples for networked resources include URLs and URNs (when implemented).
SOURCE: The work, either print or electronic, from which this resource is derived, if applicable.
LANGUAGE: Language(s) of the intellectual content of the resource.
RELATION: Relationship to other resources. Formal specification of RELATION is currently under development.
COVERAGE: The spatial locations and temporal durations characteristic of the resource. Formal specification of COVERAGE is currently under development.
RIGHTS MANAGEMENT: The content of this element is intended to be a link (a URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement, or perhaps a server that would provide such information in a dynamic way.

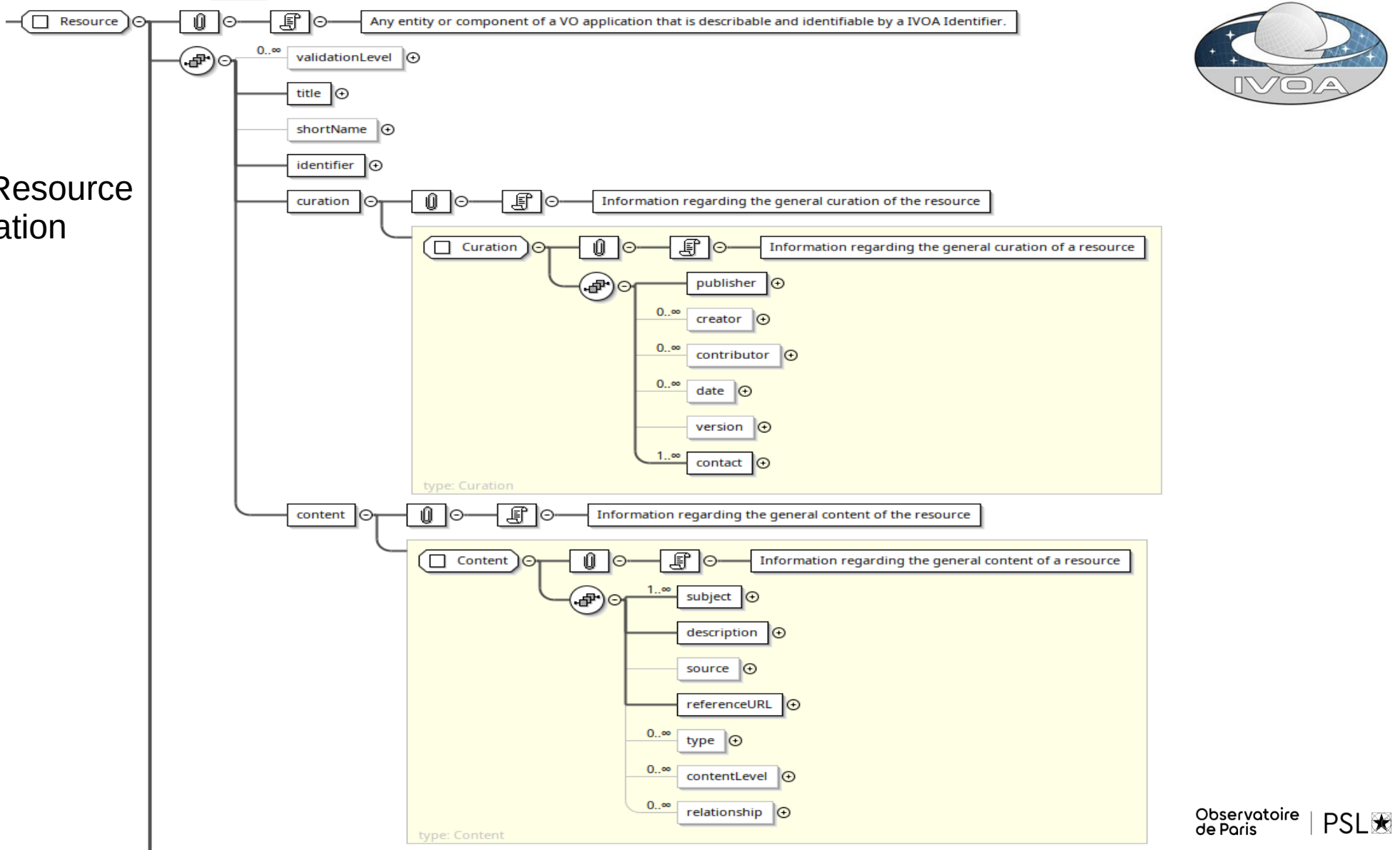
- VOResource schema split in different parts :

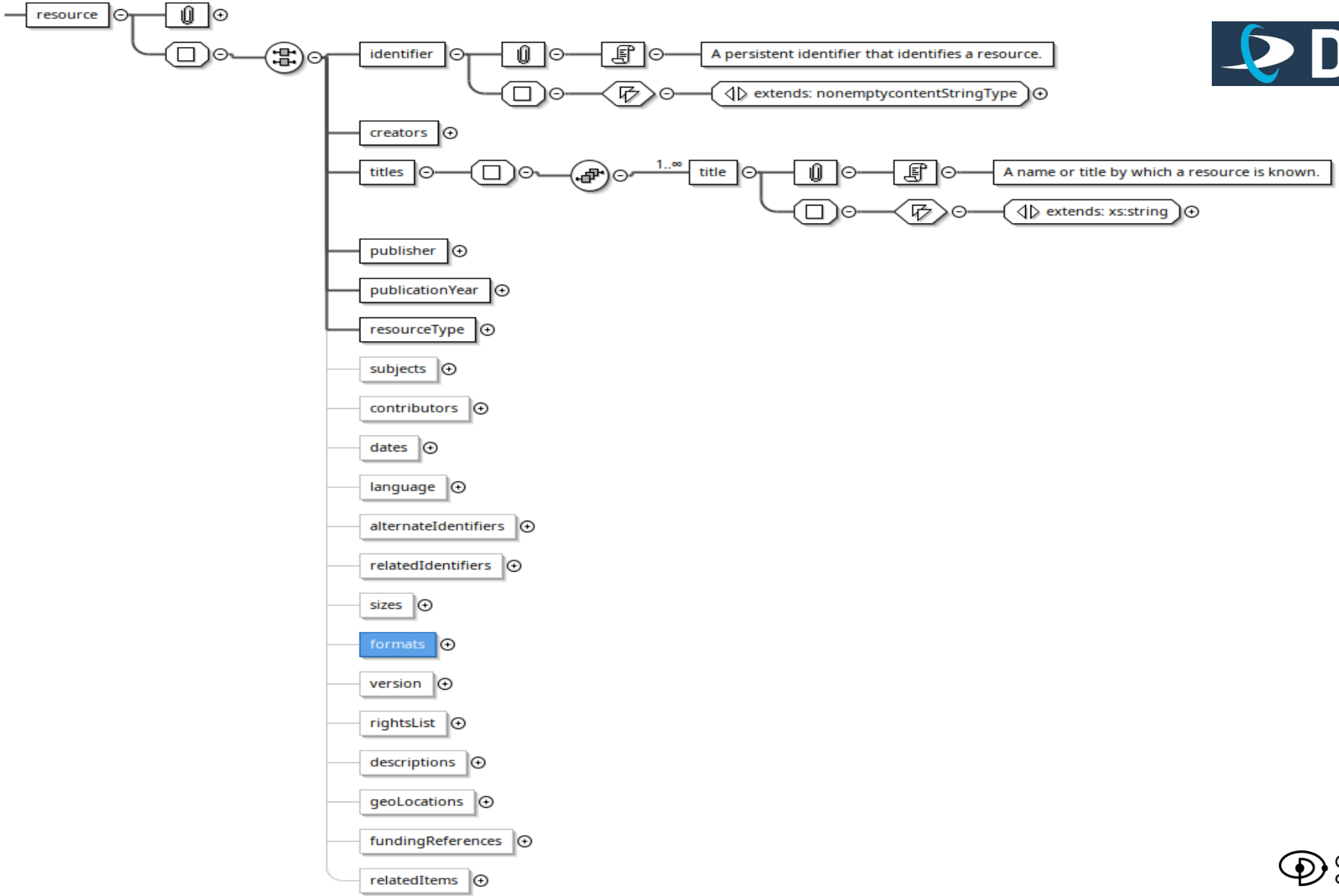
- Title + short name
- Identifier
- Curation : publisher, creator, contributor, date, version, contact
- Content : subject, description, source, referenceURL, type, content level, relationship
- Created, Updated, Status
- ValidationLevel, validation
- AuthorityID
- Right
- Capabilities, interface, accessURL

Stored in registry with standard query protocol
Harvested to avoid Single Point Of Failure
Findable-oriented for services or data collection then the next level is distributed.
User will query many data providers in an « **Interoperable** » way.

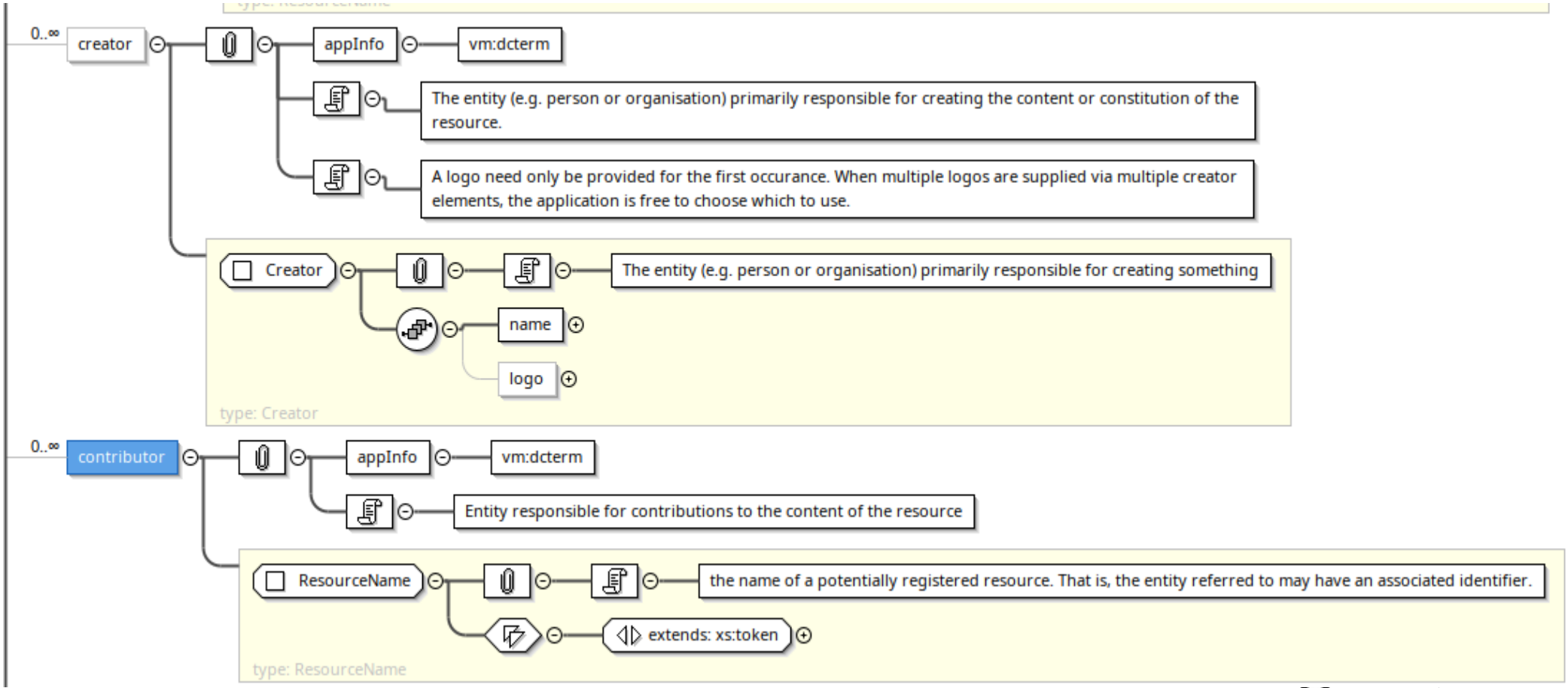
let's take a closer look at the terms

VOResource Curation

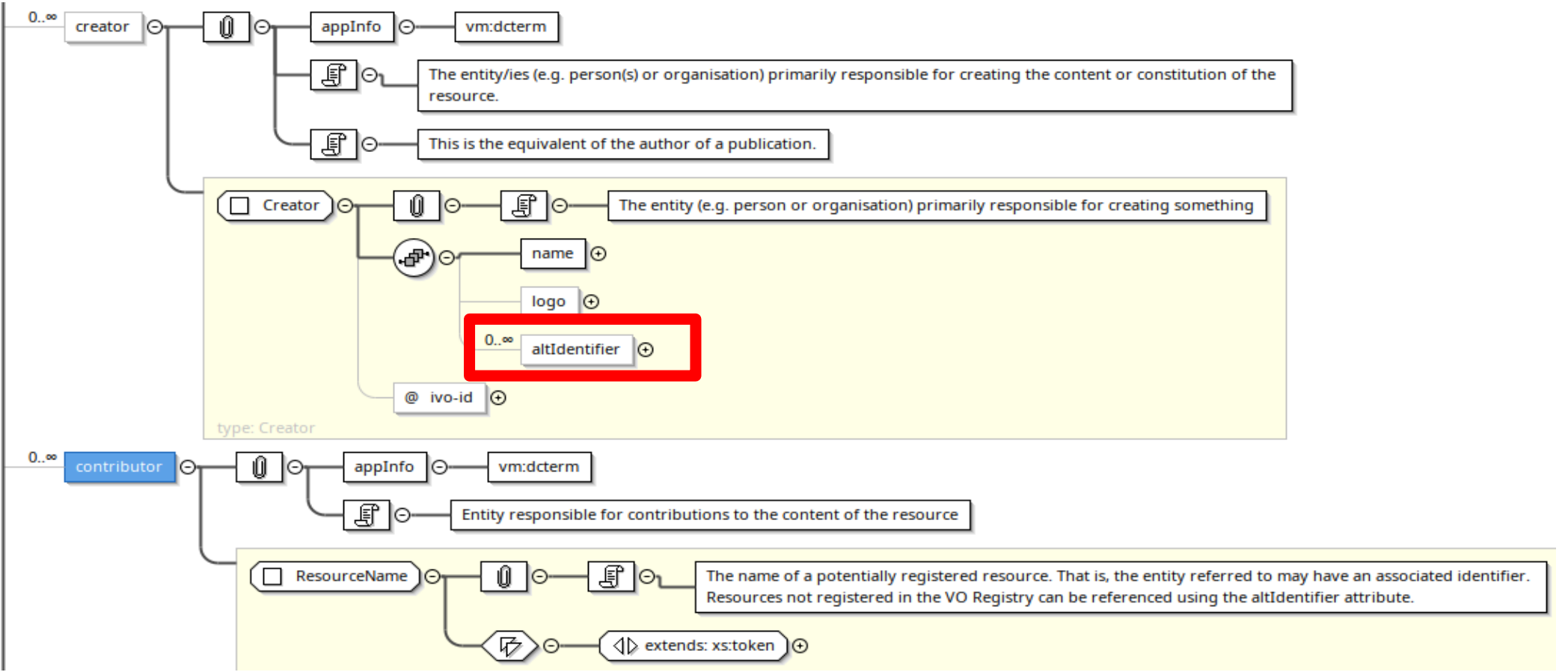



resource
identifier
A persistent identifier that identifies a resource.
extends: nonemptycontentStringType
creators
titles
1..∞
title
A name or title by which a resource is known.
extends: xs:string
publisher
publicationYear
resourceType
subjects
contributors
dates
language
alternateIdentifiers
relatedIdentifiers
sizes
formats
version
rightsList
descriptions
geoLocations
fundingReferences
relatedItems

Creator and contributor in VOResource 1.0



Creator and contributor in VOResource 1.2



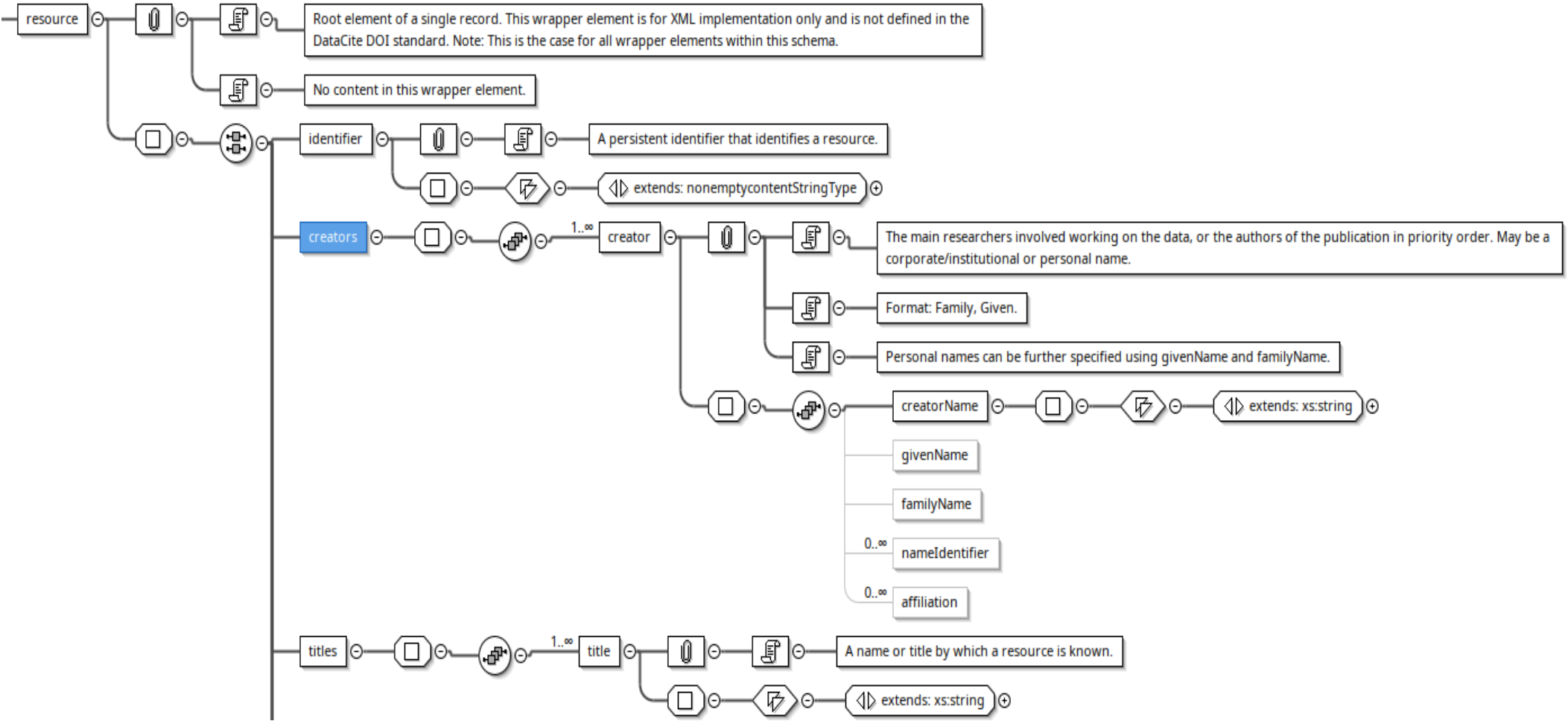
What alt identifier could be :

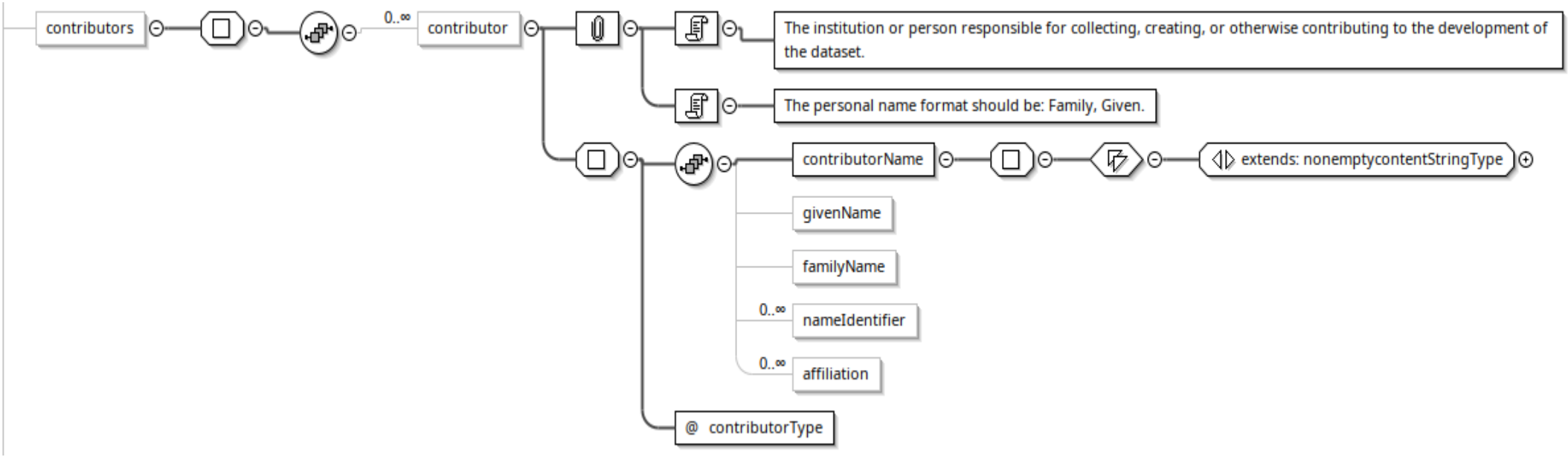
OrcID : https://orcid.org/
- remove homonym problems
- no need to care about institute affiliation …

ISNI (International Standard Name Identifier) ...

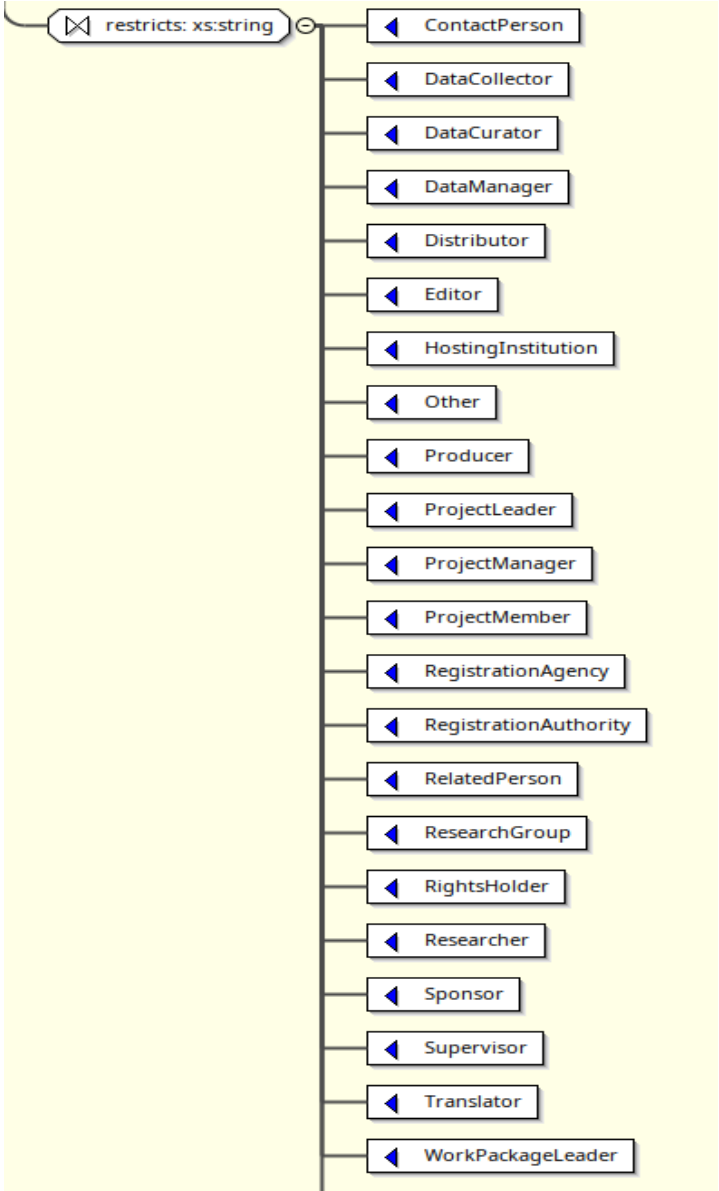
For example association on one person member of one lab at Observatoire de Paris

LIRA, Observatoire de Paris, Université PSL, Sorbonne Université, Université Paris Cité, CY Cergy Paris Université, CNRS, 92190 Meudon, France

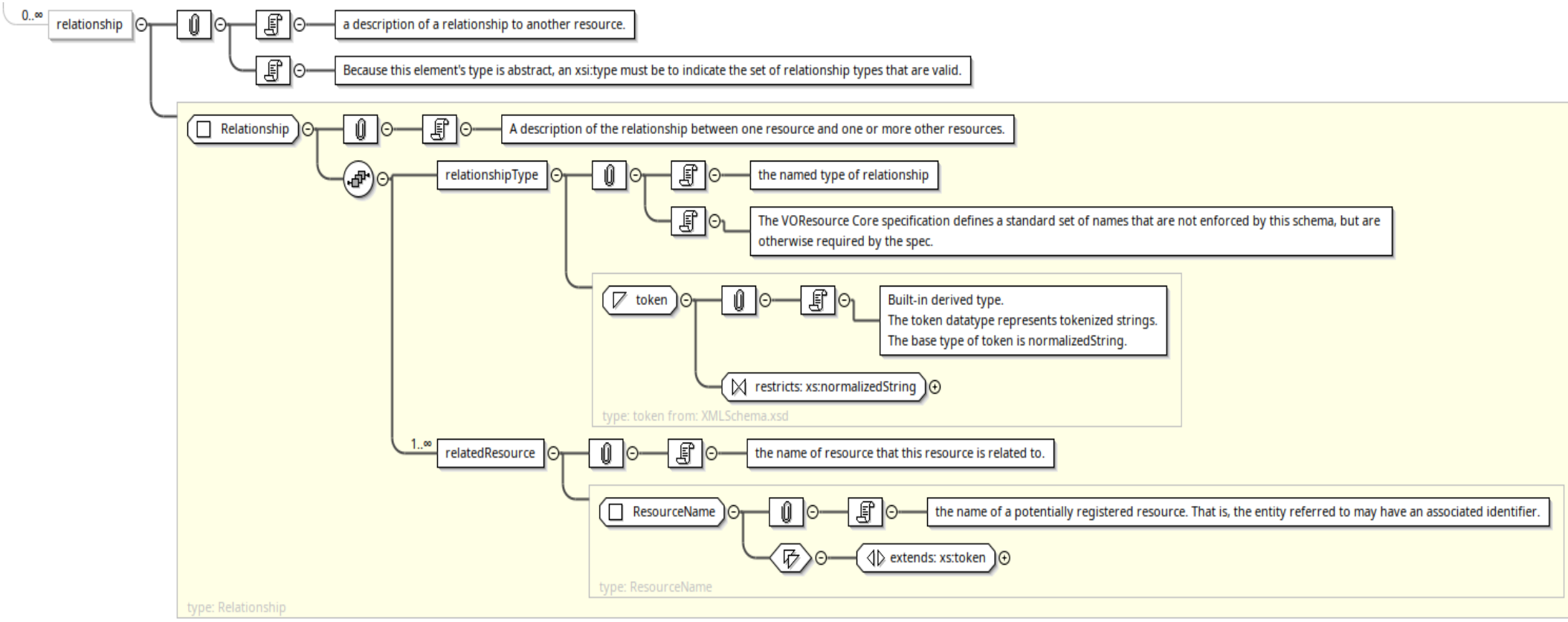

resource
Root element of a single record. This wrapper element is for XML implementation only and is not defined in the DataCite DOI standard. Note: This is the case for all wrapper elements within this schema.
No content in this wrapper element.
identifier
A persistent identifier that identifies a resource.
extends: nonemptycontentStringType
creators
1..∞
creator
The main researchers involved working on the data, or the authors of the publication in priority order. May be a corporate/institutional or personal name.
Format: Family, Given.
Personal names can be further specified using givenName and familyName.
creatorName
extends: xs:string
givenName
familyName
0..∞
nameIdentifier
0..∞
affiliation
titles
1..∞
title
A name or title by which a resource is known.
extends: xs:string


contributors
0..∞
contributor
The institution or person responsible for collecting, creating, or otherwise contributing to the development of the dataset.
The personal name format should be: Family, Given.
contributorName
extends: nonemptycontentStringType
givenName
familyName
0..∞
nameIdentifier
0..∞
affiliation
@ contributorType

Contributor Type



Relation between resources



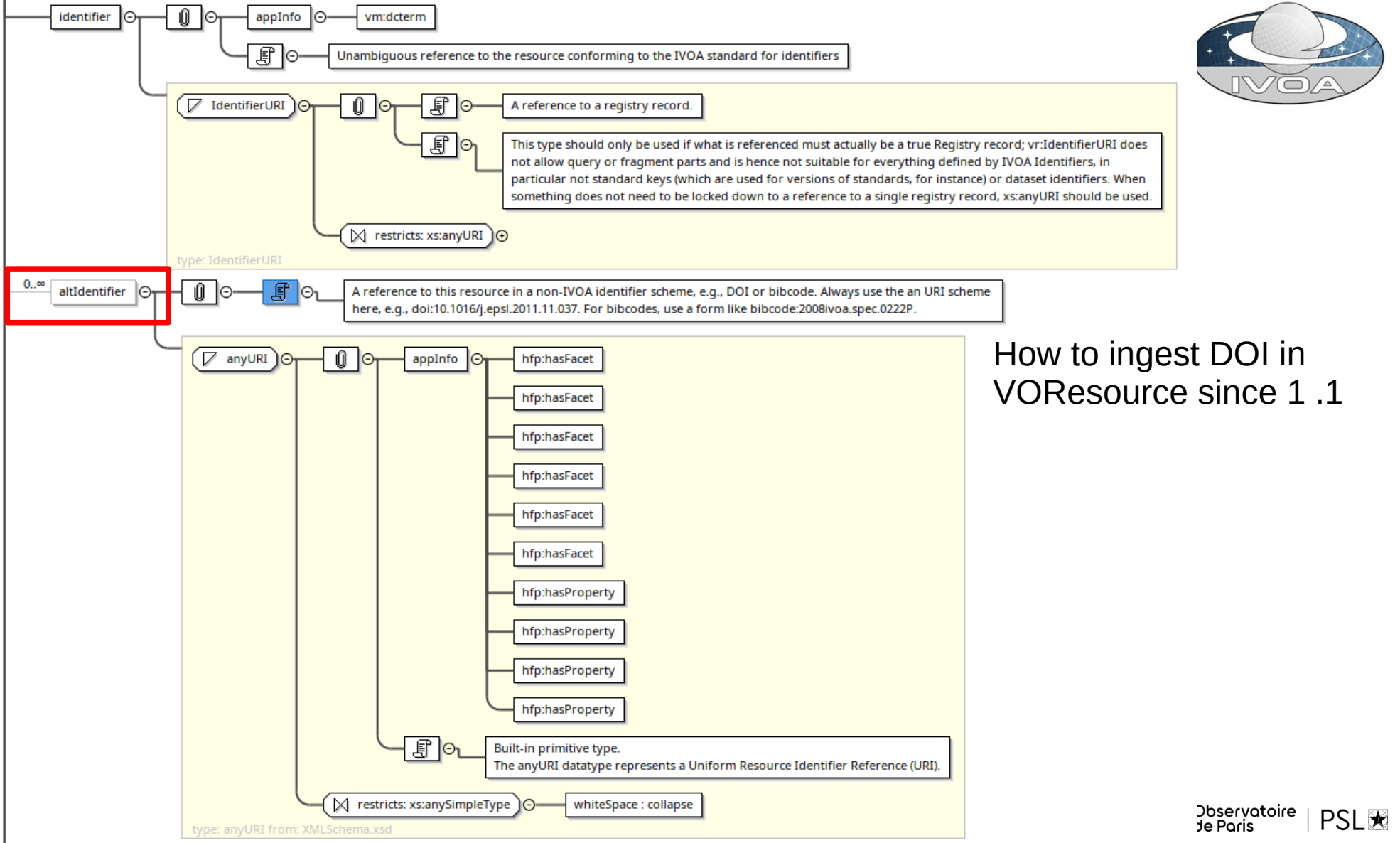
Cites	cites	This resource makes use of the referenced resource. This is in particular to reference tools like programs from educational material like tutorials. It is not intended to track citations in a bibliographic sense.
Continues	continues	This resource has taken up the service of the referenced resource (inverse of IsContinuedBy)
HasPart	has part	This is resource is a collection of other resources, and this relationship points to the individual parts (inverse of IsPartOf)
IsContinuedBy	is continued by	This resource has been or will be discontinued, and clients should turn to the referenced resource (inverse of Continues)
IsDerivedFrom	is derived from	This resource was produced using data from the referenced resource (inverse of IsSourceOf)
IsIdenticalTo	is a identical to	This resource is identical to the referenced resource, for use when there is a need to register two separate instances of the same resource. In a VO context, this will typcially indicate mirrors operated by different publishers (mirrors with identical publishers use interface/mirrorURL elements).
IsNewVersionOf	is new version of	The referenced resource has been developed on, and this resource should be used in preference (inverse of IsPreviousVersionOf)
IsPartOf	is part of	The referenced resource is a collection of some sort, perhaps a series, and this resource is a part of that collection (inverse of HasPart)
IsPreviousVersionOf	is previous version of	This resource has been developed on, and the referenced resource should be used for more up-to-date information (inverse of NewVersionOf)
IsServedBy	is served by	This resource can be accessed or otherwise used through the referenced service (inverse of IsServiceFor)
IsServiceFor	is a service for	This resource makes data from the referenced resource (typcally some sort for data collection) available (inverse of IsServedBy)
IsSourceOf	is source of	This is resource has been used to produce the referenced resource (inverse of derived-from)
IsSupplementTo	is supplement to	This resource augments the referenced resource (inverse of IsSupplementedBy)
IsSupplementedBy	is supplemented by	Additional material pertaining to this resource is available at the referenced resource (inverse of IsSupplementedTo)

Relation between resources

IsCitedBy
Cites
IsSupplementTo
IsSupplementedBy
IsContinuedBy
Continues
IsDescribedBy
Describes
HasMetadata
IsMetadataFor
HasVersion
IsVersionOf
IsNewVersionOf
IsPreviousVersionOf
IsPartOf
HasPart
IsPublishedIn
IsReferencedBy

References
IsDocumentedBy
Documents
IsCompiledBy
Compiles
IsVariantFormOf
IsOriginalFormOf
IsIdenticalTo
IsReviewedBy
Reviews
IsDerivedFrom
IsSourceOf
IsRequiredBy
Requires
IsObsoletedBy
Obsoletes
IsCollectedBy
Collects

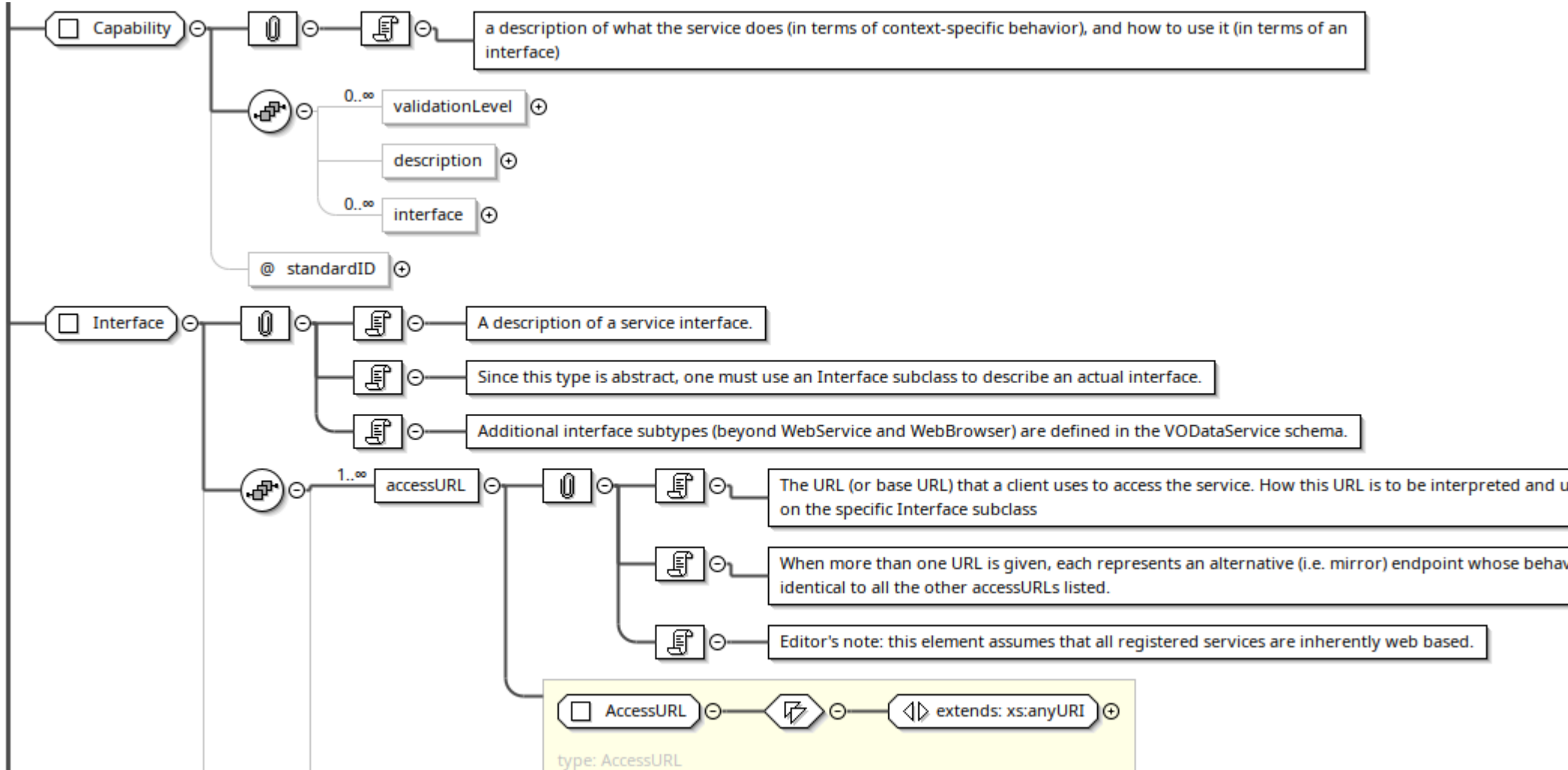
IsTranslationOf
HasTranslation

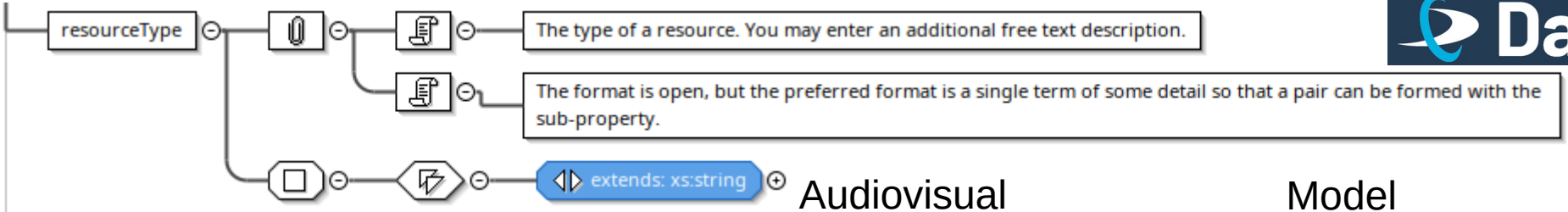


How to ingest DOI in VOResource since 1 .1

Observatoire de Paris | PSL

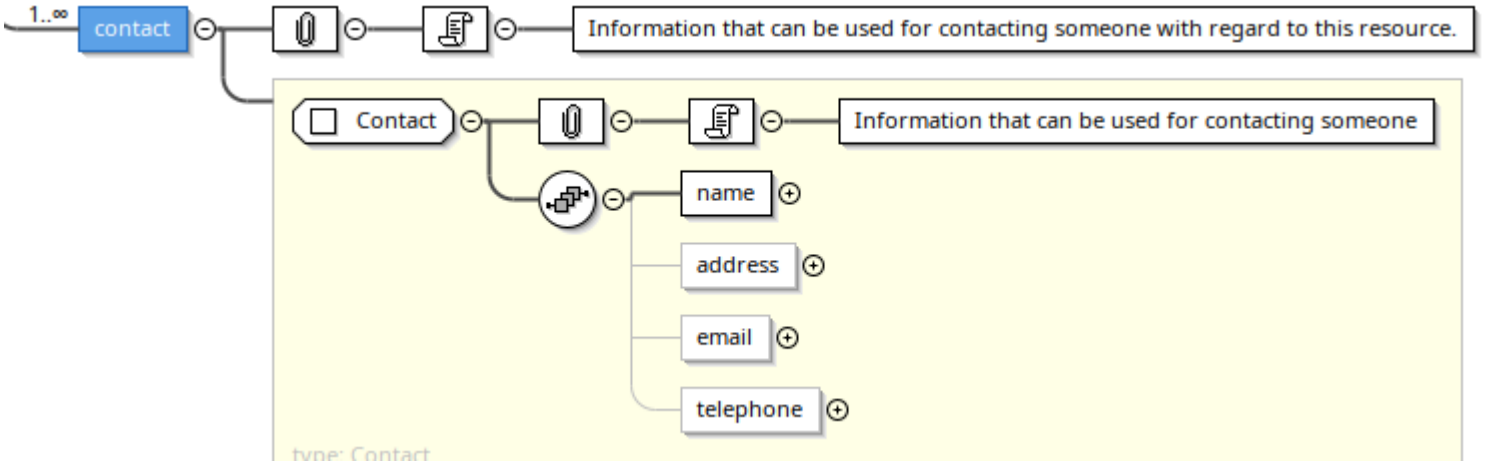
Capability / Type of resource



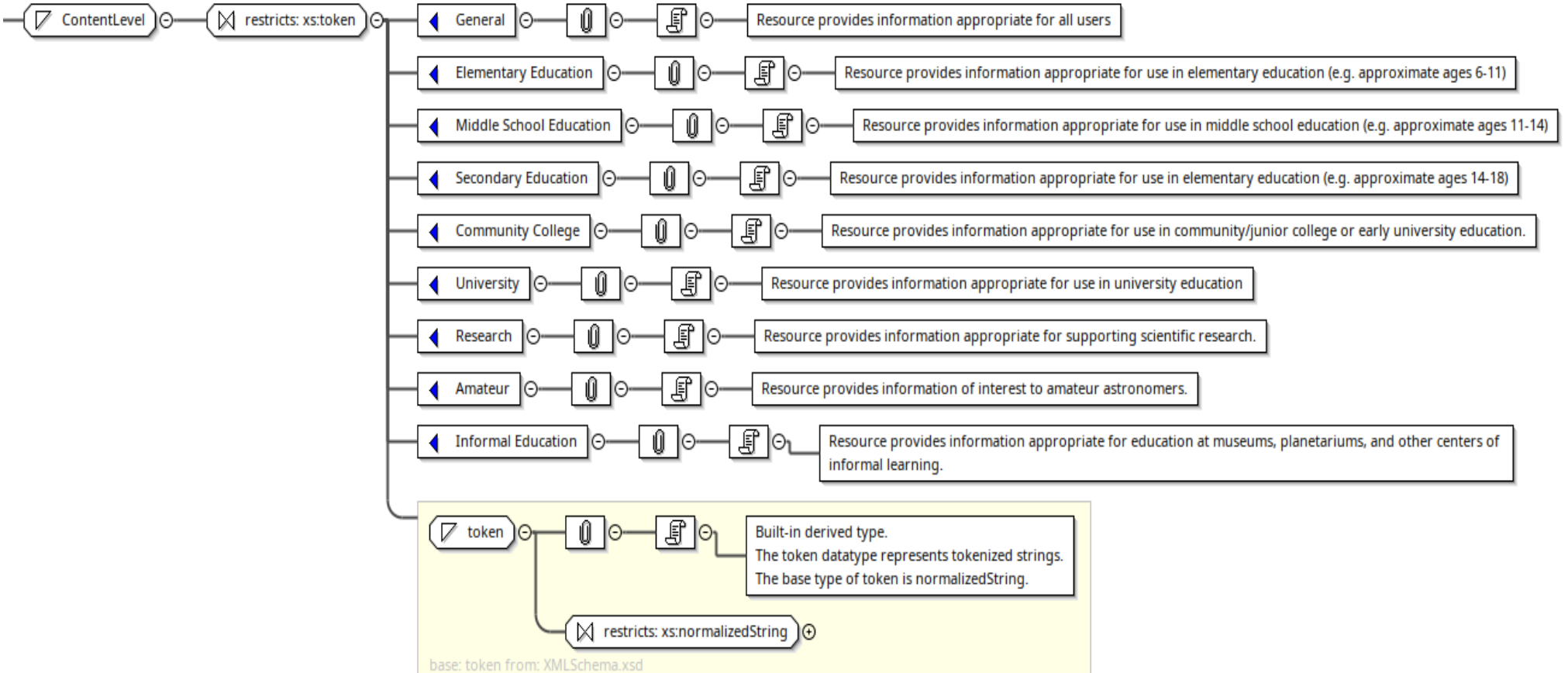


- Audiovisual
- Award
- Book
- BookChapter
- Collection
- ComputationalNotebook
- ConferencePaper
- ConferenceProceeding
- DataPaper
- Dataset
- Dissertation
- Event
- Image
- InteractiveResource
- Instrument
- Journal
- JournalArticle
- Model
- OutputManagementPlan
- PeerReview
- PhysicalObject
- Preprint
- Project
- Report
- Service
- Software
- Sound
- Standard
- StudyRegistration
- Text
- Workflow
- Other

Whom to ask if something get wrong ?



Target population



Conclusion

IVOA Voresource is dedicated to Interoperability and specially to services for accessing the data

Metadata is the key in IVOA to fulfill the FAIR principle using VO Standards for searching, retrieving, compare and dedicated tools to analyses data.

DataCite is dedicated to identify « permanently » the resource and associated the participant for proper citation

ID (DOI) is the key point to attribute correct identifier for citation and give access (not specifically in a standard way and format).

For citation / acknowledgments

IVOA → Creator definition : “Users of the creation should use this name in subsequent credits and acknowledgments This should be exactly one name”,
IVOA organize Metadata mainly for services even associated to data collection

DataCite - DOI allow proper credit for the data for creator, contributor, associated institute and funder.

describes the data for access but not specifically the access protocols, IVOA call it capability

Resource in SPASE
- Resource Name, Alternate Name
- Release Date
- Description
- **Acknowledgement**
- Contact, Information URL, Association ID
- Created, Updated, Status ...

For service of data
- Repository ID spase://authority/path
- Availability, Access Rights, Access URL
- **Caveats :** info for avoidance of the misuse of the resource, for contamination ….
- Format
- Encoding
- AcknowledgementL

SPASE (Space Physics Archive Search and Extract)

For data

- Observatory ID (link to Resource description)
- Instrument ID (also a link)
- Measurement Type (the general category, such as Thermal Plasma)
- Temporal Description (time range available, resolution)
- Observed Region (the source of data)
- Physical Parameter (very useful)

note that the difference with IVOA is not the parameter, but the quantity or nature.

Categories are :

- Photons: which are electromagnetic fields
- Fields: distinguished from Photons by being measured as time series
- Particles: which are forms of matter; and
- Mixed: which are composite or derived quantities.

Comparison to ivoa and datacite, roles associated to person (contact)

The assigned or assumed function or position of an individual.

Allowed Values:

- Archive Specialist
- Co-Investigator
- Data Producer
- Deputy-PI
- General Contact
- Metadata Contact
- Principal Investigator
- Project Scientist
- Scientist
- Team Leader
- Team Member
- Technical Contact.