



# Re-search Alps

Research laboratories in the Alpine Area

**INEA CEF-TELECOM Project**

AGREEMENT No INEA/CEF/ICT/A2016/1296967

## **Description of the metadata used in the Re-Search Alps Data Sources**

<b>Code</b>	v2.1
<b>Date</b>	30/06/2019
<b>Type</b>	Public
<b>Responsible</b>	Francesco Guerra
<b>Participants</b>	UNIMORE, SIDETRADE, MESRI, MIUR
<b>Authors</b>	Francesco Guerra (UNIMORE), Matteo Paganelli (UNIMORE), Paolo Sottovia, (UNIMORE), Laura Po (UNIMORE), Maurizio Vincini (UNIMORE), Francesco Pagliacci (UNIMORE), Margherita Russo (UNIMORE), François Bancilhon (SIDETRADE), Loic Petit (SIDETRADE), Emmanuel Weisenburger (MESRI), Frederic Olland (MESRI) Domenico De Martinis (MIUR)
<b>Corresponding Authors</b>	Francesco Guerra



<b>The integrated dataset</b>	<b>2</b>
Structures	4
Projects	11
Other outcomes	14



## The integrated dataset

The integrated dataset is released according to the best practices recommended by W3C for publishing data on the web<sup>1</sup>. In particular:

BEST PRACTICE	IMPLEMENTATION
<b>1. Provide metadata</b> <b>2. Provide descriptive metadata</b>	The dataset will be described according to the MQA standard, i.e., it will be DCAT-AP compliant. The idea is to deliver the final version via a CKAN system. For the moment, the page of the website describing the dataset will report the metadata of the dataset.
<b>3. Provide structural metadata</b>	The data are released in a JSON Format. This document provides a description of the internal structure of the distribution.
<b>4. Provide data license information</b>	Data are released according to a CC-BY 4.0 license.
<b>5. Provide data provenance information</b>	The data sources integrated are listed and described. For each data inserted in the dataset the provenance is described.
<b>6. Provide data quality information</b>	This version does not include any data quality information. This will be done in the final version.
<b>7. Provide a version indicator</b>	Versioning metadata are provided in the release.
<b>8. Provide version history</b>	This is the first official version of the dataset. This will be done for the next releases.
<b>9. Use persistent URIs as identifiers of datasets</b> <b>11. Assign URIs to dataset versions and series</b>	The <a href="http://researchalps.eu/index.php/re-search-alps-dataset/">http://researchalps.eu/index.php/re-search-alps-dataset/</a> is used for the versioned release of the dataset.
<b>10. Use persistent URIs as identifiers within datasets</b>	All the identifiers provided by the sources integrated in the release are reported.
<b>12. Use machine-readable standardized data formats</b> <b>13. Use locale-neutral data representations</b>	The data is released via a JSON file by using neutral data representations.

<sup>1</sup> <https://www.w3.org/TR/dwbp/>



<b>14. Provide data in multiple formats</b>	For the moment only a JSON release is provided.
<b>15. Reuse vocabularies, preferably standardized ones</b>	For the moment this is not adopted (no need) apart from the Country codes.
<b>16 Choose the right formalization level</b>	No need for this
<b>17. Provide bulk download 18 Provide Subsets for Large Datasets</b>	The release is composed of 3 JSON files describing structures, projects and other outcomes
<b>19. Use content negotiation for serving data available in multiple formats</b>	A CKAN server will be used for the final version of the dataset
<b>20. Provide real-time access 21. Provide data up to date</b>	The dataset includes data from other sources. When the sources are updated, the integrated source will be automatically updated. A crawl of the sites is periodically refreshed.
<b>22. Provide an explanation for data that is not available</b>	The data is self-contained. No need for this.
<b>23. Make data available through an API 24. Use Web Standards as the foundation of APIs 25. Provide complete documentation for your API 26. Avoid Breaking Changes to Your API</b>	The CKAN server we will deploy will have these features.
<b>27. Preserve identifiers</b>	Not relevant for the moment
<b>28. Assess dataset coverage</b>	Some statistics are reported in the release page.
<b>29. Gather feedback from data consumers 30. Make feedback available 33. Provide Feedback to the Original Publisher</b>	Feedbacks will be collected via email. A comment page will be developed in the future. When relevant, the comments will be sent to the original publisher.



<b>31. Enrich data by generating new data</b>	This is done and described in the previous sections of this document
<b>32. Provide Complementary Presentations</b>	Some demos of the dataset are provided. A search engine for querying the data and visualizing the results is provided.
<b>34. Follow Licensing Terms</b>	We integrate only open dataset in this release.
<b>35. Cite the Original Publication</b>	For each item and for each field, the data provenance information is provided.

Table 15: Actual action done with respect to the W3C best practises.

The integrated dataset describing laboratories and innovation centres is composed of three parts describing the **organizations (structures.json)**, the **projects (project.json)** and the others research **outcomes (outcome.json)**.

## Structures

The file JSON describing the **structures**, i.e. the research centers, has the overall structure shown in Figure 8.

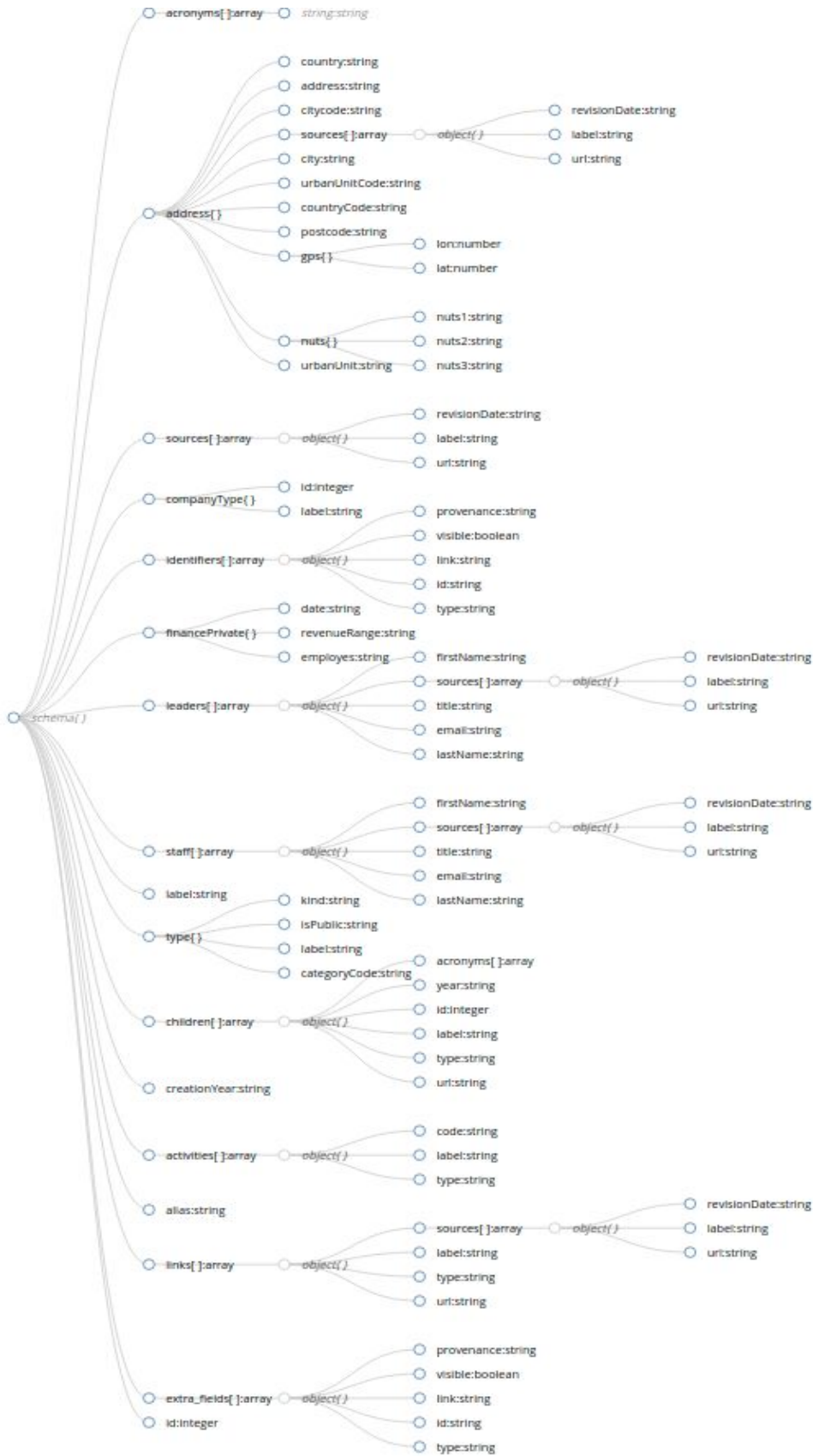


Fig. 8: Structure JSON schema



According to the schema, an enterprise is composed of the following fields:

<b>Organization</b>			
Name	Max Cardinality <sup>2</sup>	Type	Description (when needed)
acronyms	*	String	Short name representing the organization
address	1	Object	Address where the organization is located
sources	*	Object	Data sources describing the organization
companyType	1	Object	Organization legal form
identifiers	*	Object	Identifiers of the organization in the integrated data sources
financePrivate	1	Object	Some financial indicators for private organizations (e.g., Number of employees, revenue range)
leaders	*	Object	Details about the organization leaders
staff	*	Object	Details about the organization workers
label	1	String	Organization denomination
type	1	Object	Organization classification by dimension
children	*	Object	Reference to other " <b>Organization</b> " children of the one represented
creationYear	1	String	Year when the organization has been funded
activities	*	Object	Classification of organization economic activities (i.e., ATECO codes)
alias	1	String	Alternative organization denomination
extraFields	*	Object	Other details about of the organization where there is no specific field
id	1	Integer	Organization identifier

---

<sup>2</sup> All fields are optional



<b>Address</b>			
Name	Max Cardinality	Type	Description (when needed)
country	1	String	
address	1	String	
cityCode	1	String	City code (e.i., INSEE codes)
sources	*	Object	
city	1	String	
urbanUnitCode	1	String	Field used only for French Organization
countryCode	1	String	Country code (ISO 3166-1 alpha-2)
postCode	1	String	Zip code
urbanUnit	1	String	Field used only for French Organization
gps	1	Object	
nuts	1	Object	

<b>nuts</b>			
Name	Max Cardinality	Type	Description (when needed)
nuts1	1	String	High level geographical unit identifier
nuts2	1	String	Medium level geographical unit identifier
nuts3	1	String	Low level geographical unit identifier

<b>sources</b>			
Name	Max Cardinality	Type	Description (when needed)
revisionDate	1	String	
label	1	String	Internal Name of the data source





url	1	String	Link to the data source
-----	---	--------	-------------------------

<b>companyType</b>			
Name	Max Cardinality	Type	Description (when needed)
id	1	String	
type	1	String	Organization legal form

<b>identifiers</b>			
Name	Max Cardinality	Type	Description (when needed)
provenance	1	String	Data source where identifier comes from.
visible	1	String	Internal flag: deprecated
link	1	String	
id	1	String	Identifier value
type	1	String	Identifier name

<b>financePrivate</b>			
Name	Max Cardinality	Type	Description (when needed)
date	1	String	Temporal dimension to which the financial statistics refer
revenueRange	1	String	Revenue range
employes	1	String	Number of employees

<b>leaders</b>			
Name	Max Cardinality	Type	Description (when needed)



firstName	1	String	
lastName	1	String	
title	1	String	
email	1	String	
sources	*	Object	

<b>staff</b>			
Name	Max Cardinality	Type	Description (when needed)
firstName	1	String	
lastName	1	String	
title	1	String	
email	1	String	
sources	*	Object	

<b>type</b>			
Name	Max Cardinality	Type	Description (when needed)
kind	1	String	Company type (i.e., "public" or "private")
isPublic	1	String	Boolean indicator (public or private)
label	1	String	Organization classification by dimension, extended name (e.i., small and medium enterprises)
categoryCode	1	String	Organization classification by dimension, code (e.i., SME)

<b>activities</b>			
Name	Max	Type	Description (when needed)



	Cardinality		
code	1	String	Activity classification, code
label	1	String	Activity classification, extended name
type	1	String	Activity classification system (i.e., ATECO system)

<b>links</b>			
Name	Max Cardinality	Type	Description (when needed)
sources	1	Object	
label	1	String	Explanation of what the website refers to
type	1	String	
url	1	String	

<b>extra_fields</b>			
Name	Max Cardinality	Type	Description (when needed)
provenance	1	String	Data source from which information derives
visible	1	boolean	
id	1	String	Field value
type	1	String	Field name
link	1	String	



## Projects

The file JSON describing the **projects** has the structure shown in Figure 9.

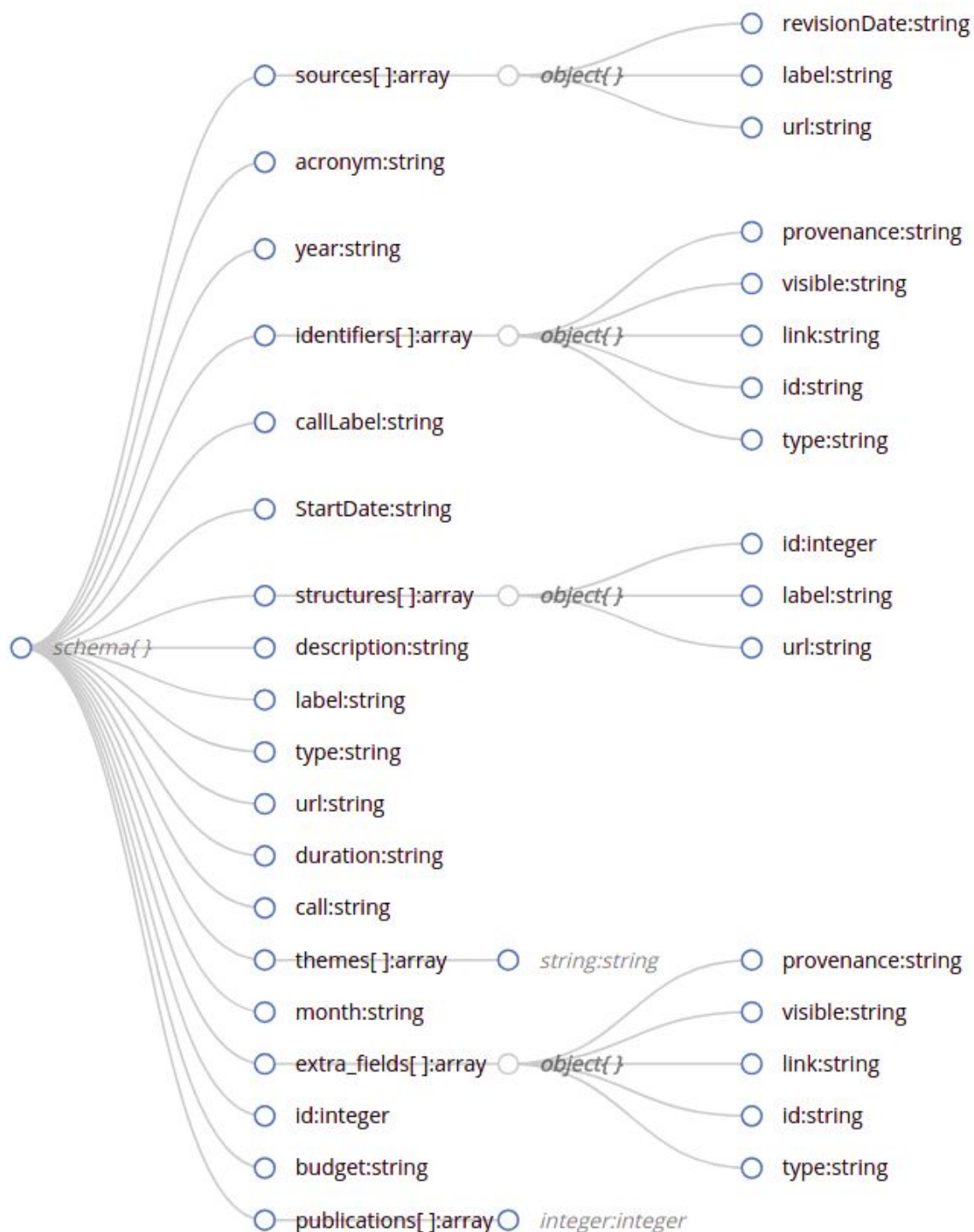


Fig. 9: Project JSON schema



<b>Project</b>			
Name	Max Cardinality <sup>3</sup>	Type	Description (when needed)
acronym	1	String	Short name representing the project
year	1	String	Year when the project has been funded
sources	*	Object	Data sources describing the organization
identifiers	*	Object	Identifiers of the project in the integrated sources
callLabel	1	String	
StartDate	1	String	Year when the project started
structures	*	Object	Organizations involved (reference to organization)
description	1	String	
label	1	String	Title of the project
type	*	String	
url	1	String	Project url
duration	1	String	
call	1	String	
themes	*	String	
month	1	String	Month extracted from project date
extra_fields	*	Object	
id	1	integer	Internal identifier of the project
budget	1	String	Budget of the project
publications	*	Integer	List of publication identifiers derived from the project

<sup>3</sup> All fields are optional



<b>sources</b>			
Name	Max Cardinality	Type	Description (when needed)
revisionDate	1	String	
label	1	String	Internal Name of the data source
url	1	String	Data source URL

<b>identifiers</b>			
Name	Max Cardinality	Type	Description (when needed)
provenance	1	String	Data source from which identifier comes from
visible	1	String	Internal flag: deprecated
link	1	String	
id	1	String	Identifier value
type	1	String	Identifier name

<b>extra_fields</b>			
Name	Max Cardinality	Type	Description (when needed)
provenance	1	String	Data source from which information derives
visible	1	boolean	
id	1	String	Field value
type	1	String	Field name
link	1	String	



## Other outcomes

The file JSON describing the other research **outcomes** has the structure shown in Figure 10.

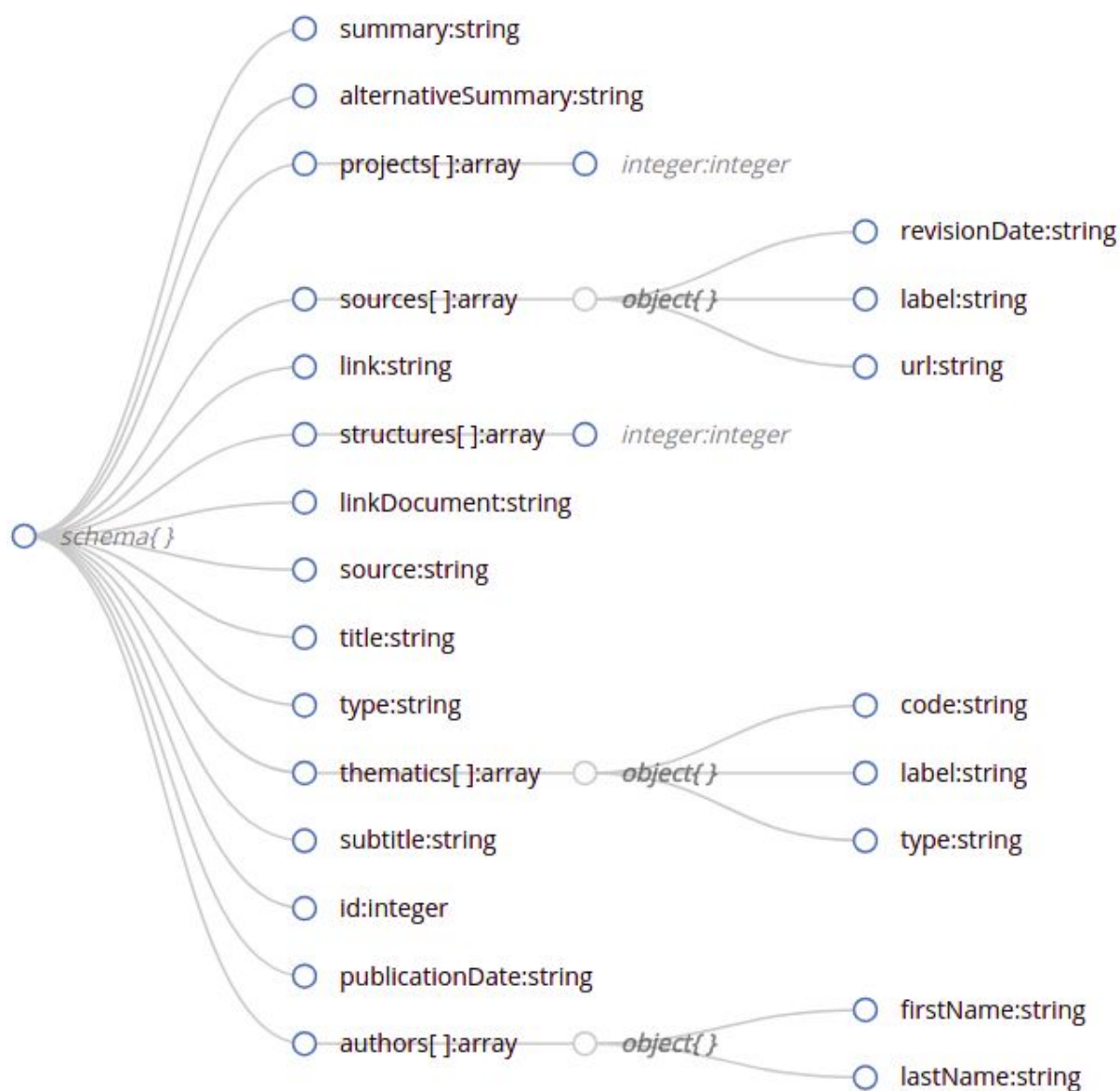


Fig. 10: Outcome JSON schema



<b>Outcome</b>			
Name	Max Cardinality <sup>4</sup>	Type	Description (when needed)
summary	1	String	Short name representing the project
alternativeSummary	1	String	
projects	*	Object	Reference to <b>projects</b> connected to the outcome
sources	*	Object	
link	1	String	
structures	*	Object	Organizations involved (reference to <b>organization</b> )
linkDocument	1	String	Link to outcome PDF
source	1	String	
title	1	String	
type	1	String	Outcome type (i.e., publication, patent)
thematics	*	Object	Outcome thematic keywords
subtitle	1	String	
id	1	integer	Internal identifier of the outcome
publicationDate	1	String	Budget of the project
authors	*	Object	

<b>sources</b>			
Name	Max Cardinality	Type	Description (when needed)
revisionDate	1	String	
label	1	String	Internal Name of the data source
url	1	String	Link to the data source

<sup>4</sup> All fields are optional





<b>thematics</b>			
Name	Max Cardinality	Type	Description (when needed)
code	1	String	Thematic classification code
label	1	String	Thematic classification extended name
type	1	String	Thematic classification system

<b>authors</b>			
Name	Max Cardinality	Type	Description (when needed)
firstName	1	String	
lastName	1	String	