

---

# **inspire-schemas Documentation**

*Release 61.4.8*

**CERN**

**Dec 13, 2021**



---

## Contents

---

<b>1</b>	<b>User's Guide</b>	<b>3</b>
1.1	Installation . . . . .	3
1.2	Usage . . . . .	3
<b>2</b>	<b>API Reference</b>	<b>5</b>
2.1	API Docs . . . . .	5
2.2	Utils Docs . . . . .	5
<b>3</b>	<b>Schemas Reference</b>	<b>11</b>
3.1	Schemas Documentation . . . . .	11
<b>4</b>	<b>Additional Notes</b>	<b>173</b>
4.1	Contributing . . . . .	173
4.2	License . . . . .	175
<b>5</b>	<b>Changelog</b>	<b>177</b>
	<b>Python Module Index</b>	<b>179</b>
	<b>Index</b>	<b>181</b>



Inspirehep schemas and related tools bundle.

- Free software: GPLv2 license
- Documentation: <https://inspire-schemas.readthedocs.io>



This part of the documentation will show you how to get started in using `inspire-schemas`.

**NOTE:** it is highly encouraged to pin the version of `inspire-schemas` that you use to the major number, for example using the `~=` version matcher in your `requirements.txt` or your `setup.py` like:

```
inspire-schemas~=1.0.0
```

That will prevent any major update (compatibility breaking) to be pulled automatically, and will only download minor and feature updates (1.X.Y).

## 1.1 Installation

```
pip install inspire-schemas
```

## 1.2 Usage

Schemas and utilities used in `inspirehep`.





If you are looking for information on a specific function, class or method, this part of the documentation is for you.

## 2.1 API Docs

Public api for methods and functions to handle/verify the jsonschemas.

## 2.2 Utils Docs

Public api for methods and functions to handle/verify the jsonschemas.

```
class inspire_schemas.utils.LocalRefResolver (base_uri, referrer, store=(),  
                                             cache_remote=True, handlers=(), urljoin_cache=None, remote_cache=None)
```

Bases: `jsonschema.validators.RefResolver`

Simple resolver to handle non-uri relative paths.

**resolve\_remote** (*uri*)

Resolve a uri or relative path to a schema.

`inspire_schemas.utils.author_id_normalize_and_schema` (*uid*, *schema=None*)

Detect and normalize an author UID schema.

### Parameters

- **uid** (*string*) – a UID string
- **schema** (*string*) – try to resolve to schema

**Returns** a tuple (uid, schema) where: - uid: the UID normalized to comply with the id.json schema  
- schema: a schema of the UID or *None* if not recognised

**Return type** Tuple[string, string]

**Raise:** UnknownUIDSchema: if UID is too little to definitively guess the schema SchemaUIDConflict: if specified schema is not matching the given UID

`inspire_schemas.utils.build_pubnote` (*title*, *volume*, *page\_start=None*, *page\_end=None*, *ar-tid=None*)  
 Build pubnote string from parts (reverse of `split_pubnote`).

`inspire_schemas.utils.classify_field` (*value*)  
 Normalize *value* to an Inspire category.

**Parameters** *value* (*str*) – an Inspire category to properly case, or an arXiv category to translate to the corresponding Inspire category.

**Returns**

**None** if *value* is not a non-empty string, otherwise the corresponding Inspire category.

**Return type** *str*

`inspire_schemas.utils.convert_new_publication_info_to_old` (*publication\_infos*)  
 Convert back a *publication\_info* value from the new format to the old.

Does the inverse transformation of `convert_old_publication_info_to_new()`, to be used whenever we are sending back records from Labs to Legacy.

**Parameters** *publication\_infos* – a *publication\_info* in the new format.

**Returns** a *publication\_info* in the old format.

**Return type** *list(dict)*

`inspire_schemas.utils.convert_old_publication_info_to_new` (*publication\_infos*)  
 Convert a *publication\_info* value from the old format to the new.

On Legacy different series of the same journal were modeled by adding the letter part of the name to the journal volume. For example, a paper published in Physical Review D contained:

```
{
  'publication_info': [
    {
      'journal_title': 'Phys.Rev.',
      'journal_volume': 'D43',
    },
  ],
}
```

On Labs we instead represent each series with a different journal record. As a consequence, the above example becomes:

```
{
  'publication_info': [
    {
      'journal_title': 'Phys.Rev.D',
      'journal_volume': '43',
    },
  ],
}
```

This function handles this translation from the old format to the new. Please also see the tests for various edge cases that this function also handles.

**Parameters** *publication\_infos* – a *publication\_info* in the old format.

**Returns** a `publication_info` in the new format.

**Return type** `list(dict)`

`inspire_schemas.utils.country_code_to_name` (*code*)

The country's name for the given code.

**Parameters** `code` – needs to be *alpha\_2* country code.

`inspire_schemas.utils.country_name_to_code` (*name*)

The country's code for the given name.

**Parameters** `name` – needs to be an *ISO 3166-1* or *ISO 3166-3* country name.

`inspire_schemas.utils.filter_empty_parameters` (*func*)

Decorator that is filtering empty parameters.

**Parameters** `func` (*function*) – function that you want wrapping

`inspire_schemas.utils.fix_reference_url` (*url*)

Used to parse an incorrect url to try to fix it with the most common occurrences for errors. If the fixed url is still incorrect, it returns `None`.

**Returns** String containing the fixed url or the original one if it could not be fixed.

`inspire_schemas.utils.fix_url_add_http_if_missing` (*string*)

Add the starting `http` to a url that is missing it

`inspire_schemas.utils.fix_url_bars_instead_of_slashes` (*string*)

A common error in urls is that all `/` have been changed for `|`, we fix that in this function

`inspire_schemas.utils.fix_url_replace_tilde` (*string*)

Replace unicode characters by their working equivalent

`inspire_schemas.utils.get_license_from_url` (*url*)

Get the license abbreviation from an URL.

**Parameters** `url` (*str*) – canonical url of the license.

**Returns** the corresponding license abbreviation.

**Return type** `str`

**Raises** `ValueError` – when the url is not recognized

`inspire_schemas.utils.get_paths` (*schema*, *previous\_node=None*)

`inspire_schemas.utils.get_refs_to_schemas` (*references={}*)

For every schema return path and index name for every referenced record :returns: index and path to the referenced record :rtype: dict(list(tuple))

`inspire_schemas.utils.get_schema_path` (*schema*, *resolved=False*)

Retrieve the installed path for the given schema.

**Parameters**

- **schema** (*str*) – relative or absolute url of the schema to validate, for example, 'records/authors.json' or 'jobs.json', or just the name of the schema, like 'jobs'.
- **resolved** (*bool*) – if True, the returned path points to a fully resolved schema, that is to the schema with all *\$ref* replaced by their targets.

**Returns** path to the given schema name.

**Return type** `str`

**Raises** `SchemaNotFound` – if no schema could be found.

`inspire_schemas.utils.get_validation_errors` (*data*, *schema=None*)

Validation errors for a given record.

**Parameters**

- **data** (*dict*) – record to validate.
- **schema** (*Union[dict, str]*) – schema to validate against. If it is a string, it is interpreted as the name of the schema to load (e.g. `authors` or `jobs`). If it is `None`, the schema is taken from `data['$schema']`. If it is a dictionary, it is used directly.

**Yields** *jsonschema.exceptions.ValidationError* – validation errors.

**Raises**

- `SchemaNotFound` – if the given schema was not found.
- `SchemaKeyNotFound` – if `schema` is `None` and no `$schema` key was found in `data`.
- `jsonschema.SchemaError` – if the schema is invalid.

`inspire_schemas.utils.is_arxiv` (*obj*)

Return `True` if `obj` contains an arXiv identifier.

The `idutils` library's `is_arxiv` function has been modified here to work with two regular expressions instead of three and adding a check for valid arxiv categories only

```
inspire_schemas.utils.load_schema (schema_name, resolved=False,
                                     _cache={'/home/docs/checkouts/readthedocs.org/user_builds/inspire-
                                     schemas/envs/latest/lib/python3.7/site-
                                     packages/inspire_schemas/records/elements/rank.json':
                                     {'enum': ['STAFF', 'SENIOR', 'JUNIOR', 'VISITOR',
                                     'POSTDOC', 'PHD', 'MASTER', 'UNDERGRADUATE',
                                     'OTHER', None], 'minLength': 1, 'title': 'Rank of aca-
                                     demic position', 'type': 'string'}, 'elements/rank': {'enum':
                                     ['STAFF', 'SENIOR', 'JUNIOR', 'VISITOR', 'POSTDOC',
                                     'PHD', 'MASTER', 'UNDERGRADUATE', 'OTHER',
                                     None], 'minLength': 1, 'title': 'Rank of academic position',
                                     'type': 'string'}}})
```

Load the given schema from wherever it's installed.

**Parameters**

- **schema\_name** (*str*) – Name of the schema to load, for example `authors`.
- **resolved** (*bool*) – If `True` will return the resolved schema, that is with all the `$refs` replaced by their targets.
- **\_cache** (*dict*) – Private argument used for memoization.

**Returns** the schema with the given name.

**Return type** `dict`

`inspire_schemas.utils.normalize_arxiv` (*obj*)

Return a normalized arXiv identifier from `obj`.

`inspire_schemas.utils.normalize_arxiv_category` (*category*)

Normalize arXiv category to be schema compliant.

This properly capitalizes the category and replaces the dash by a dot if needed. If the category is obsolete, it also gets converted it to its current equivalent.

## Example

```
>>> from inspire_schemas.utils import normalize_arxiv_category
>>> normalize_arxiv_category('funct-an') # doctest: +SKIP
u'math.FA'
```

`inspire_schemas.utils.normalize_collaboration` (*collaboration*)  
 Normalize collaboration string.

**Parameters** `collaboration` – a string containing collaboration(s) or None

**Returns** List of extracted and normalized collaborations

**Return type** list

## Examples

```
>>> from inspire_schemas.utils import normalize_collaboration
>>> normalize_collaboration('for the CMS and ATLAS Collaborations')
['CMS', 'ATLAS']
```

`inspire_schemas.utils.normalize_collaboration_name` (*full\_collaboration\_string*)

`inspire_schemas.utils.normalize_isbn` (*isbn*)  
 Normalize an ISBN in order to be schema-compliant.

`inspire_schemas.utils.sanitize_html` (*text*)  
 Sanitize HTML for use inside records fields.

This strips most of the tags and attributes, only allowing a safe whitelisted subset.

`inspire_schemas.utils.split_page_artid` (*page\_artid*)  
 Split `page_artid` into `page_start/end` and `artid`.

`inspire_schemas.utils.split_pubnote` (*pubnote\_str*)  
 Split `pubnote` into journal information.

`inspire_schemas.utils.valid_arxiv_categories` ()  
 List of all arXiv categories that ever existed.

## Example

```
>>> from inspire_schemas.utils import valid_arxiv_categories
>>> 'funct-an' in valid_arxiv_categories()
True
```

`inspire_schemas.utils.validate` (*data*, *schema=None*)  
 Validate the given dictionary against the given schema.

### Parameters

- **data** (*dict*) – record to validate.
- **schema** (*Union[dict, str]*) – schema to validate against. If it is a string, it is interpreted as the name of the schema to load (e.g. `authors` or `jobs`). If it is None, the schema is taken from `data['$schema']`. If it is a dictionary, it is used directly.

### Raises

- `SchemaNotFound` – if the given schema was not found.
- `SchemaKeyNotFound` – if schema is `None` and no `$schema` key was found in data.
- `jsonschema.SchemaError` – if the schema is invalid.
- `jsonschema.ValidationError` – if the data is invalid.

Auto-generated documentation for Inspirehep schemas.

## 3.1 Schemas Documentation

### 3.1.1 authors

A record in the Authors collection

**type** object

**Required** *name, \_collections*

**Properties:** *\$schema, \_bucket, \_collections, \_private\_notes, acquisition\_source, advisors, arxiv\_categories, awards, birth\_date, control\_number, death\_date, deleted, deleted\_records, email\_addresses, ids, inspire\_categories, legacy\_creation\_date, legacy\_version, name, new\_record, positions, project\_membership, public\_notes, self, status, stub, urls*

#### **\$schema**

**type** string

**format** uri

**minLength** 1

#### **\_bucket**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string  
**minLength** 1

### **`_collections`**

Collections to which record belongs

**type** array  
**minItems** 1

Every element of **`_collections`** is:

**type** string  
**minLength** 1

**Allowed values:**

- Authors

### **`_private_notes`**

List of private notes

**MARC** 595

These notes are only visible to privileged users, not regular users.

**type** array  
**minItems** 1

Every element of **`_private_notes`** is:

**Reference** *sourced\_value*

### **`acquisition_source`**

**Reference** *acquisition\_source*

### **`advisors`**

List of advisors

**MARC** 701  
**type** array  
**minItems** 1

Every element of **`advisors`** is:

**type** object  
**Required** *name*

**Properties:** *curated\_relation, degree\_type, hidden, ids, name, record*



**curated\_relation**

**type** boolean

**degree\_type**

**MARC** 701\_\_g

**Reference** *degree\_type*

**hidden**

Whether this advisor should be hidden

**type** boolean

**ids**

List of advisor's IDs

**MARC** 701\_\_i

**type** array

**minItems** 1

Every element of **ids** is:

**Reference** *id*

**name**

**MARC** 701\_\_a

The full name of the advisor, written in the standard last names, first names form.

**example** Smith, John P.S.

**type** string

**pattern** ^[\^, ]+(, [\^, ]+)?(, ?[\^, ]+)?\$

**minLength** 1

**record**

**Reference** *author*

**arxiv\_categories**

List of research areas of author

**MARC** 65017a when it contains an arXiv category.

**type** array

**minItems** 1

Every element of **arxiv\_categories** is:

**Reference** *arxiv\_categories*

## awards

List of awards received by the author

**MARC** 678

**type** array

**minItems** 1

Every element of **awards** is:

An award

**type** object

**Properties:** *name, url, year*

## name

Name of the award

**MARC** 678\_\_a (also contains *year*).

**example** Nobel Prize in Physics

**type** string

**minLength** 1

## url

URL of award announcement

**MARC** 678\_\_u

**Reference** *url*

## year

Year of the award

**MARC** 678\_\_a (also contains *name*).

**type** integer

**minimum** 1000

**maximum** 2050

**birth\_date**

MARC 100\_\_d (first part)

**type** string

**format** date

**minLength** 1

**control\_number**

ID of current record

MARC 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

**death\_date**

MARC 100\_\_d (second part)

See also *status* deceased.

**type** string

**format** date

**minLength** 1

**deleted**

Whether this record has been deleted

MARC 980\_\_a/c:deleted

**type** boolean

**deleted\_records**

MARC 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *author*

## email\_addresses

List of email addresses

**MARC** 371\_\_m/o, 595\_\_m/o

**type** array

**minItems** 1

Every element of **email\_addresses** is:

**type** object

**Required** *value*

**Properties:** *current, hidden, value*

### current

Whether this address is still in use

**MARC** if *true*, the email comes from *m*, if *false*, from *o*.

**type** boolean

### hidden

Whether this address should be hidden

**MARC** if *true*, the email comes from 371, if *false*, from 595.

Email addresses might be considered sensitive information and hidden at the user's request.

**type** boolean

### value

An email address

**type** string

**format** email

**minLength** 1

### ids

**type** array

**minItems** 1

Every element of **ids** is:

**Reference** *id*

### inspire\_categories

**type** array

**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

### legacy\_creation\_date

**type** string

**format** date

**minLength** 1

### legacy\_version

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

### name

Name information of the author

**MARC** 100, 400, 880

**type** object

**Required** *value*

**Properties:** *name\_variants, native\_names, numeration, preferred\_name, previous\_names, title, value*

### name\_variants

List of name variants

Contains other variations (besides *preferred\_name*) of the author name in *value* that are in use.

These could be:

- other spellings;
- other transliterations from the *native\_names* to the Latin alphabet;
- other splittings among last names and first names;
- combinations with *previous\_names*.

**example** Smith-Taylor, Johnny

**MARC** 400\_\_a

```
type array
pattern ^[^\, ]+(, [^\, ]+)?(, ?[^\, ]+)?$
minItems 1
```

Every element of **name\_variants** is:

```
type string
minLength 1
```

### **native\_names**

Native names of the author, not using the Latin alphabet.

```
example
MARC 880__a
type array
minItems 1
```

Every element of **native\_names** is:

```
type string
minLength 1
```

### **numeration**

```
MARC 100__b
type string
minLength 1
```

**Allowed values:**

- Jr.
- Sr.
- I
- II
- III
- IV
- VI
- VII
- VIII

### **preferred\_name**

differs from the full name in *value*. This should use Latin alphabet.

```
example Smith, John
```

**MARC 100\_\_q**  
**type** string  
**minLength** 1

### previous\_names

List of previous names

Contains previous names of the author, e.g. the maiden name.

**example** Taylor, Jonathan Gerald C.T.  
**MARC** no field for this, was put into 667\_\_a as a note.  
**type** array  
**pattern** ^[\^, ]+(, [\^, ]+)?(, ?[\^, ]+)?\$  
**minItems** 1

Every element of **previous\_names** is:

**type** string  
**minLength** 1

### title

**MARC 100\_\_c**  
**type** string  
**minLength** 1

**Allowed values:**

- Sir

### value

Full name of the author

Author name in Latin alphabet (may contain diacritics). The format should be last names, first names. If not all names are known, initials should be followed by a . without space.

---

**Note:** Not all authors have two names, so only one name without a comma is perfectly valid.

**example** Jimmy

---

**example** Smith Davis, Jonathan Gerald C.T.  
**MARC 100\_\_a**  
**type** string  
**pattern** ^[\^, ]+(, [\^, ]+)?(, ?[\^, ]+)?\$  
**minLength** 1

### new\_record

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *author*

### positions

**MARC** 371

**type** array

**minItems** 1

Every element of **positions** is:

**type** object

**Required** *institution*

**Properties:** *curated\_relation, current, end\_date, hidden, institution, rank, record, start\_date*

### curated\_relation

**type** boolean

### current

**MARC** 371\_\_z:current corresponds to true

**type** boolean

### end\_date

**MARC** 371\_\_t

**type** string

**format** date

**minLength** 1

### hidden

Whether this position should be hidden

**type** boolean



**institution**

MARC 371\_\_a

**type** string

**minLength** 1

**rank**

MARC 371\_\_r

**Reference** *rank*

**record**

URI for the institution record

**Reference** *institution*

**start\_date**

MARC 371\_\_s

**type** string

**format** date

**minLength** 1

**project\_membership**

List of projects on which the author worked

MARC 693

**type** array

**minItems** 1

Every element of **project\_membership** is:

**type** object

**Required** *name*

**Properties:** *curated\_relation, current, end\_date, hidden, name, record, start\_date*

**curated\_relation**

**type** boolean

### current

Whether the author is still involved in it

**MARC** 693\_\_z corresponds to true

**type** boolean

### end\_date

**MARC** 693\_\_d

Date the author stopped working on this project.

**type** string

**format** date

**minLength** 1

### hidden

Whether this project should be hidden

**type** boolean

### name

**MARC** 693\_\_e

Name of the project.

**type** string

**minLength** 1

### record

URI for the experiment record

**Reference** *experiment*

### start\_date

**MARC** 693\_\_s

Date the author started working on this project.

**type** string

**format** date

**minLength** 1

**public\_notes**

**type** array

**minItems** 1

Every element of **public\_notes** is:

**Reference** *sourced\_value*

**self**

Url of the record

Url of the record itself

**Reference** *author*

**status**

The person's status

**type** string

**minLength** 1

**Allowed values:**

- active
- deceased
- departed
- retired

**stub**

**type** boolean

**urls**

**type** array

**minItems** 1

Every element of **urls** is:

**Reference** *url*

### 3.1.2 conferences

Conference

**type** object

**Required** *\_collections*

**Properties:** *\$schema, \_bucket, \_collections, \_private\_notes, acronyms, addresses, alternative\_titles, closing\_date, cnum, contact\_details, control\_number, core, deleted, deleted\_records, external\_system\_identifiers, inspire\_categories, keywords, legacy\_creation\_date, legacy\_version, new\_record, opening\_date, public\_notes, self, series, short\_description, titles, urls*

### **\$schema**

**type** string

**format** uri

**minLength** 1

### **\_bucket**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string

**minLength** 1

### **\_collections**

Collections to which this record belongs

**type** array

**minItems** 1

Every element of **\_collections** is:

**type** string

**minLength** 1

**Allowed values:**

- Conferences

### **\_private\_notes**

List of private notes

**MARC** 595

These notes are only visible to privileged users, not regular users.

**type** array

**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

## acronyms

List of acronyms of the conference

**type** array

**minItems** 1

Every element of **acronyms** is:

Conference acronym

**MARC** 111\_\_e

**example** SUSY 2018

**type** string

**minLength** 1

## addresses

List of locations of the conference

**MARC** 111\_\_c (in the form *city, country*), 270\_\_b

This field is a list because a conference can sometimes be split among several cities.

---

**Note:** The *postal\_address* should not be provided for a conference.

---

**type** array

**minItems** 1

Every element of **addresses** is:

**Reference** *address*

## alternative\_titles

**MARC** 711

Sometimes conferences are known by several titles/acronyms. These name variants can be provided here.

---

**Note:** If a conference has several titles or acronyms, it is more appropriate to use *titles* or *:acronyms*.

This field is meant mainly to port information from legacy, that was not categorized well into title/acronym variants and was used to boost search results.

---

**type** array

**minItems** 1

List of alternative titles

Every element of **alternative\_titles** is:

**Reference** *title*

### closing\_date

Conference closing date

111\_\_y

**type** string

**format** date

**minLength** 1

### cnum

CNUM of the conference

**Reference** *cnum*

### contact\_details

Contact information

**MARC** 270

**type** array

**minItems** 1

Every element of **contact\_details** is:

**Reference** *contact*

### control\_number

ID of current record

**MARC** 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

### core

Whether this record is CORE

**MARC** 980: CORE corresponds to `true`

**type** boolean

### deleted

Whether this record has been deleted

**MARC** 980\_\_a/c: deleted

**type** boolean

### deleted\_records

MARC 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *conference*

### external\_system\_identifiers

List of identifiers of this document on external systems.

**type** array

**minItems** 1

Every element of **external\_system\_identifiers** is:

May satisfy *any* of the following definitions:

**0**

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string

**minLength** 1

**Allowed values:**

- SPIRES

### value

MARC 970\_\_a

**type** string

**pattern** ^CONF-\d+\$

**minLength** 1

## inspire\_categories

**type** array

**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

## keywords

List of keywords

**MARC** 6531 and 695 (depending on presence of *schema*)

Keywords give information about the specific contents of the record, much more precisely than *inspire\_categories*.

**type** array

**minItems** 1

Every element of **keywords** is:

**type** object

**Required** *value*

**Properties:** *schema, source, value*

## schema

Keyword vocabulary

Describes to which vocabulary the keyword in *value* belongs.

### INSPIRE

**MARC** 695\_\_2:INSPIRE

The keyword has been assigned by Inspire, and belongs to its vocabulary.

---

**Note:** If not present, the keyword is a free-form keyword, not necessarily part of any vocabulary.

---

**type** string

**minLength** 1

**Allowed values:**

- INSPIRE

## source

**Reference** *source*



**value**

A keyword

**MARC** 6531\_\_a (depending on *schema*)

It belongs to the vocabulary specified by *schema*.

**example** black hole: mass (for *schema* = INSPIRE)

**type** string

**minLength** 1

**legacy\_creation\_date**

**type** string

**format** date

**minLength** 1

**legacy\_version**

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

**new\_record**

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *conference*

**opening\_date**

Conference opening date

**MARC** 111\_\_x

**type** string

**format** date

**minLength** 1

## public\_notes

List of public notes

**MARC** 500

Any notes about the document that do not fit into another field.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see *\_private\_notes*.

---

**type** array

**minItems** 1

Every element of **public\_notes** is:

**Reference** *sourced\_value*

## self

**Reference** *conference*

## series

Conference series that this is part of

**MARC** 411

**type** array

**minItems** 1

Every element of **series** is:

**type** object

**Required** *name*

**Properties:** *name, number*

## name

Name of the conference series

411\_\_a

**type** string

**minLength** 1

## number

Number of the conference in the series

411\_\_n

**type** integer

**short\_description****Reference** *sourced\_value***titles**

List of titles of the conference

**MARC** 111\_\_a/b**type** array**minItems** 1Every element of **titles** is:**Reference** *title***urls****type** array**minItems** 1Every element of **urls** is:**Reference** *url***3.1.3 data**

A data record

Schema is not yet specified

**type** object**Properties:** *\$schema, \_bucket, \_collections, control\_number, deleted, deleted\_records, dois, legacy\_version, new\_record, self***\$schema****type** string**format** uri**minLength** 1**\_bucket**Used by *invenio-records-files* to store information about the bucket attached to this record.**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.**type** string**minLength** 1

## **\_collections**

Collections to which record belongs

**MARC** 980\_\_a

**type** array

**minItems** 1

Every element of **\_collections** is:

**type** string

**minLength** 1

**Allowed values:**

- Data

## **control\_number**

ID of current record

**MARC** 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

## **deleted**

Whether this record has been deleted

**MARC** 980\_\_a/c:deleted

**type** boolean

## **deleted\_records**

**MARC** 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *data*

**dois**

List of DOIs

**type** array

**minItems** 1

Every element of **dois** is:

**type** object

**Required** *value*

**Properties:** *source*, *value*

**source**

**MARC** 0247\_9

**Reference** *source*

**value**

DOI

**MARC** 0247\_a

**example** 10.1023/A:1026654312961

**type** string

**pattern** ^10\.\d+(\.\d+)?/\S+\$

**minLength** 1

**legacy\_version**

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

**new\_record**

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *data*

**self**

**Reference** *data*

### 3.1.4 experiments

A collaboration/experiment/accelerator record

This record can describe different types of projects, depending on the value of *project\_type*.

**type** object

**Required** *project\_type, \_collections*

**Properties:** *\$schema, \_bucket, \_collections, \_full\_ingestion, \_private\_notes, accelerator, collaboration, control\_number, core, date\_approved, date\_cancelled, date\_completed, date\_proposed, date\_started, deleted, deleted\_records, description, experiment, external\_system\_identifiers, inspire\_categories, inspire\_classification, institutions, legacy\_creation\_date, legacy\_name, legacy\_version, long\_name, name\_variants, new\_record, project\_type, public\_notes, related\_records, self, urls*

#### **\$schema**

**type** string

**format** uri

**minLength** 1

#### **\_bucket**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string

**minLength** 1

#### **\_collections**

Collections to which this record belongs

**type** array

**minItems** 1

Every element of **\_collections** is:

**type** string

**minLength** 1

**Allowed values:**

- Experiments

### **\_full\_ingestion**

Whether to ingest all articles written by this project

**type** boolean

### **\_private\_notes**

List of private notes

**MARC** 595\_\_a, 667\_\_a

These notes are only visible to privileged users, not regular users.

**type** array

**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

### **accelerator**

- If `accelerator` is present in *project\_type*, this field contains information about this accelerator.
- If `experiment` is present in *project\_type*, this field refers to the accelerator that provides a beam to this experiment.

**type** object

**Required** *value*

**Properties:** *curated\_relation, record, value*

### **curated\_relation**

**type** boolean

### **record**

If `accelerator` is not present in *project\_type*, this links to the related accelerator record.

**Reference** *experiment*

### **value**

Accelerator name

**MARC** 119\_\_b

**example** LHC

**type** string

**minLength** 1

## collaboration

**MARC** 710

- If `collaboration` is present in *project\_type*, this field contains information about this collaboration.
- Otherwise, this field refers to the collaboration that runs this accelerator/experiment.

**type** object

**Required** *value*

**Properties:** *curated\_relation, record, subgroup\_names, value*

## curated\_relation

**type** boolean

## record

If `collaboration` is not present in *project\_type*, this links to the related collaboration record.

**Reference** *experiment*

## subgroup\_names

List of names of subgroups

**MARC** 710\_\_q

Sometimes papers are signed not by the full collaboration, but only by a more focused subgroup. This records the names of those groups.

**example** ATLAS TDAQ

**type** array

**minItems** 1

Every element of `subgroup_names` is:

**type** string

**minLength** 1

## value

Collaboration name

**MARC** 710\_\_g

**example** D0

**type** string

**minLength** 1



**control\_number**

ID of current record

**MARC** 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

**core**

**MARC** 980: CORE corresponds to `true`.

Whether the papers written by this project are generally *core*.

**type** boolean

**date\_approved**

Approval date of experiment/accelerator

**MARC** 046\_\_r

**type** string

**format** date

**minLength** 1

**date\_cancelled**

Cancellation date of experiment/accelerator

**MARC** 046\_\_c

**type** string

**format** date

**minLength** 1

**date\_completed**

Completion date of experiment/accelerator

**MARC** 046\_\_t

**type** string

**format** date

**minLength** 1

### date\_proposed

Proposal date of experiment/accelerator

**MARC** 046\_\_q

**type** string

**format** date

**minLength** 1

### date\_started

Starting date of experiment/accelerator

**MARC** 046\_\_s

**type** string

**format** date

**minLength** 1

### deleted

Whether this record has been deleted

**MARC** 980\_\_a/c:deleted

**type** boolean

### deleted\_records

**MARC** 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *experiment*

### description

Description of the project

**MARC** 520\_\_a

This should take the form of a short text.

**type** string

**minLength** 1

## experiment

If `experiment` is present in *project\_type*, this field contains information about this experiment.

**type** object

**Properties:** *short\_name, value*

### short\_name

**MARC** 119\_\_d

**example** SHINE

**type** string

**minLength** 1

### value

Experiment identifier

**MARC** 119\_\_c

**example** NA61

**type** string

**minLength** 1

## external\_system\_identifiers

List of identifiers of this document on external systems.

**type** array

**minItems** 1

Every element of `external_system_identifiers` is:

May satisfy *any* of the following definitions:

### SPIRES identifier

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string

**minLength** 1

**Allowed values:**

- SPIRES

#### value

MARC 970\_\_a

**type** string

**pattern** ^EXPERIMENT-\d+\$

**minLength** 1

#### inspire\_categories

**type** array

**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

#### inspire\_classification

**type** array

**minItems** 1

Every element of **inspire\_classification** is:

MARC 372\_\_a

Category to which this experiment belongs.

**type** string

**Reference** *inspire\_experiment\_classification*

**minLength** 1

#### institutions

**type** array

**minItems** 1

Every element of **institutions** is:

Host institution

**type** object

**Properties:** *curated\_relation, record, value*

#### curated\_relation

**type** boolean

**record**

Record URI of the institution.

**Reference** *institution*

**value**

ICN of institution hosting the project

**MARC** 119\_\_u

Currently, the old *legacy\_ICN* is used here.

**type** string

**minLength** 1

**legacy\_creation\_date**

**type** string

**format** date

**minLength** 1

**legacy\_name**

project identifier on legacy INSPIRE

**MARC** 119\_\_a

**example** CERN-LHC-CMS

**type** string

**minLength** 1

**legacy\_version**

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

**long\_name**

long name of the project

**MARC** 245\_\_a

**example** Compact Muon Solenoid

**type** string

**minLength** 1

### name\_variants

Name variants of the project

**MARC** 419\_\_a

**type** array

**minItems** 1

Every element of **name\_variants** is:

**type** string

**minLength** 1

### new\_record

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *experiment*

### project\_type

Type of project that this record represents

**MARC** 980 :ACCELERATOR corresponds to an accelerator, otherwise it is an experiment.

A project can represent a combination of collaboration, accelerator and experiment, depending on the contents of this field. Typical cases are:

collaboration	experiment	accelerator	explanation
present	present	absent	typical case where a collaboration runs a single experiment (e.g. ALICE).
present	absent	absent	non-experimental collaboration (e.g. lattice) or experimental collaboration corresponding to several experiments (e.g. DO).
absent	present	absent	non-unique experiment among the experiments run by this collaboration (e.g. DO Run I).
absent	absent	present	accelerator that is not a collaboration (e.g. LHC).
present	absent	present	accelerator that is also a collaboration (e.g. CLIC).

- When `collaboration` is present, the *collaboration* field contains information about it. Otherwise, it might refer to the collaboration record corresponding to the current accelerator/experiment.
- When `experiment` is present, the *experiment* field contains information about it.
- When `accelerator` is present, the *accelerator* field contains information about it. If not, but `experiment` is present, it might refer to the accelerator record providing a beam to this experiment.

**type** array

**minItems** 1

**minLength** 1

Every element of **project\_type** is:

**type** string  
**minLength** 1

**Allowed values:**

- collaboration
- experiment
- accelerator

### **public\_notes**

List of public notes

**MARC** 500

Any notes about the document that do not fit into another field.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see *\_private\_notes*.

---

**type** array  
**minItems** 1

Every element of **public\_notes** is:

**Reference** *sourced\_value*

### **related\_records**

List of related records

**MARC** 510

**type** array  
**minItems** 1

Every element of **related\_records** is:

**Reference** *related\_record*

### **self**

**Reference** *experiment*

### **urls**

**type** array  
**minItems** 1

Every element of **urls** is:

**Reference** *url*

### 3.1.5 hep

A record in the Literature collection

**type** object

**Required** *document\_type, titles, \_collections*

**Properties:** *\$schema, \_bucket, \_collections, \_desy\_bookkeeping, \_export\_to, \_files, \_private\_notes, abstracts, accelerator\_experiments, acquisition\_source, arxiv\_eprints, authors, book\_series, citeable, collaborations, control\_number, copyright, core, corporate\_author, curated, deleted, deleted\_records, document\_type, documents, dois, editions, energy\_ranges, external\_system\_identifiers, figures, funding\_info, imprints, inspire\_categories, isbn, keywords, languages, legacy\_creation\_date, legacy\_version, license, new\_record, number\_of\_pages, persistent\_identifiers, preprint\_date, public\_notes, publication\_info, publication\_type, record\_affiliations, refereed, references, related\_records, report\_numbers, rpp, self, texkeys, thesis\_info, title\_translations, titles, urls, withdrawn*

#### **\$schema**

**type** string

**format** uri

**minLength** 1

#### **\_bucket**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string

**minLength** 1

#### **\_collections**

Collections to which this record belongs

**MARC** 980\_\_a (HEP maps to Literature)

Regular Literature records have collection Literature only.

The other collections are used by different groups to manage records using the Inspire infrastructure, but which do not belong to the Inspire literature collection. If a record belongs to one of these collections but not Literature, it is not shown in the standard search results.

**type** array

**minItems** 1

Every element of **\_collections** is:

**type** string

**minLength** 1

**Allowed values:**

- BABAR Analysis Documents



- BABAR Internal BAIS
- BABAR Internal Notes
- CDF Internal Notes
- CDF Notes
- CDS Hidden
- D0 Internal Notes
- D0 Preliminary Notes
- Fermilab
- H1 Internal Notes
- H1 Preliminary Notes
- HAL Hidden
- HEP Hidden
- HERMES Internal Notes
- LArSoft Internal Notes
- LArSoft Notes
- Literature
- SLAC
- ZEUS Internal Notes
- ZEUS Preliminary Notes

### **`_desy_bookkeeping`**

DESY keywording information

**MARC** 595\_D, 035\_\_9:DESY

Used by DESY to store information about the keyword-assignment process.

**type** array

**minItems** 1

Every element of **`_desy_bookkeeping`** is:

**type** object

**Properties:** *date, expert, identifier, status*

#### **date**

**MARC** 595\_Dd

**type** string

**minLength** 1

### expert

**MARC 595\_Da**  
**type** string  
**minLength** 1

### identifier

**MARC 035\_\_z** when 035\_\_9 is DESY  
**type** string  
**minLength** 1

### status

**MARC 595\_Ds**  
**type** string  
**minLength** 1

### \_export\_to

**MARC 595\_\_c**

Flags determining to which foreign databases this record should be automatically exported. Setting one of the keys to `true` means that the record must be exported to the corresponding system, setting it to `false` means that it must not be exported.

**type** object

**Properties:** *CDS*, *HAL*

### CDS

Controls automatic export to the [CERN Document Server](#).

**type** boolean

### HAL

Controls automatic export to [HAL](#), the French national repository.

**type** boolean

## **\_files**

*invenio-records-files* metadata

Used by *invenio-records-files* to store information about the files attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** array

**minItems** 0

Every element of **\_files** is:

**Reference** *records-files*

## **\_private\_notes**

List of private notes

**MARC** 595, 595\_H (when *source* is HAL)

These notes are only visible to privileged users, not regular users.

**type** array

**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

## **abstracts**

List of abstracts

**MARC** 520

The first abstract is the preferred one, to be displayed in priority.

**type** array

**minItems** 1

Every element of **abstracts** is:

**Reference** *sourced\_value*

## **accelerator\_experiments**

List of related accelerators/experiments

**type** array

**minItems** 1

Every element of **accelerator\_experiments** is:

**type** object

**Properties:** *accelerator, curated\_relation, experiment, institution, legacy\_name, record*

### accelerator

MARC 693\_\_a

If present, *institution* should contain the institution where this accelerator is located.

---

**Note:** Currently not used, see *legacy\_name*.

---

**type** string

**minLength** 1

### curated\_relation

**type** boolean

**default** False

### experiment

MARC not present.

If present, *institution* should contain the institution where this experiment is located and *accelerator* may contain the accelerator that this experiment is using (if appropriate).

---

**Note:** Currently not used, see *legacy\_name*.

---

**type** string

**minLength** 1

### institution

Institution hosting the experiment

MARC not present.

---

**Note:** Currently not used, see *legacy\_name*.

---

**type** string

**minLength** 1

### legacy\_name

Identifier of the experiment on legacy

MARC 693\_\_e

This field is used when migrating from legacy instead of separate *institution*, *accelerator* and *experiment*. In the future, it will be deprecated and the other fields will be used instead.

**example** CERN-LHC-CMS  
**type** string  
**minLength** 1

## record

**Reference** *experiment*

## acquisition\_source

**Reference** *acquisition\_source*

## arxiv\_eprints

**type** array  
**minItems** 1

Every element of **arxiv\_eprints** is:

arXiv metadata

Contains metadata of an arXiv eprint contained in this record.

---

**Note:** This metadata is copied over from arXiv and should not be edited manually.

---

**type** object  
**Required** *value*

**Properties:** *categories, value*

## categories

arXiv categories of the eprint

**MARC** first category in 037\_\_c with 9:arXiv, all categories in 65017a with 2:arXiv

The first category in the list has a special meaning: it is the primary category of the eprint on arXiv.

**type** array  
**minItems** 1

Every element of **categories** is:

**Reference** *arxiv\_categories*

## value

arXiv eprint identifier

**MARC** 037\_\_a with 9:arXiv

**example** math/0307245

**example** 1701.01431  
**type** string  
**pattern** ^\d{4}.\d{4,5}|[\w.]+(-[\w.]+)?/\d+\$  
**minLength** 1

## authors

Authors

**MARC** 100, 700 and 701

Besides authors, also contains editors and supervisors (see *inspire\_roles*).

**type** array  
**minItems** 1

Every element of **authors** is:

**type** object  
**Required** *full\_name*

**Properties:** *affiliations, affiliations\_identifiers, alternative\_names, credit\_roles, curated\_relation, emails, full\_name, ids, inspire\_roles, raw\_affiliations, record, signature\_block, uuid*

## affiliations

**type** array  
**minItems** 1

Every element of **affiliations** is:

**type** object  
**Required** *value*

**Properties:** *curated\_relation, record, value*

## curated\_relation

**type** boolean

## record

**Reference** *institution*

## value

ICN of affiliation

**MARC** 100/700/701\_\_u

Currently, the old *legacy\_ICN* is used here. In the future, this will change and become the new *ICN*.

**type** string

**minLength** 1

### **affiliations\_identifiers**

**MARC** 100/700/701\_\_t

List of identifiers of this institution on external systems.

**type** array

**minItems** 1

Every element of **affiliations\_identifiers** is:

May satisfy *any* of the following definitions:

#### **GRID identifier**

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

#### **schema**

**MARC** first part of 100/700/701\_\_t before the :

**type** string

**minLength** 1

**Allowed values:**

- GRID

#### **value**

**MARC** last part of 100/700/701\_\_t after the :

**type** string

**pattern** ^grid\.\d+\.\w+\$

**minLength** 1

#### **ROR identifier**

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**MARC** first part of 100/700/701\_\_t before the :

**type** string

**minLength** 1

#### Allowed values:

- ROR

### value

**MARC** last part of 100/700/701\_\_t after the :

**type** string

**pattern** ^https://ror\.org/0\w{6}\d{2}\$

**minLength** 1

### alternative\_names

**type** array

**minItems** 1

Every element of **alternative\_names** is:

**MARC** 100/700/701\_\_q

#### example

**example** Smith, Joe John

**type** string

**minLength** 1

### credit\_roles

**type** array

**minItems** 1

Every element of **credit\_roles** is:

CRedit role of author

**MARC** not present.

Role of the author according to the [Contributor Roles Taxonomy \(CRediT\)](#)

**type** string

**minLength** 1

#### Allowed values:

- Conceptualization
- Data curation



- Formal analysis
- Funding acquisition
- Investigation
- Methodology
- Project administration
- Resources
- Software
- Supervision
- Validation
- Visualization
- Writing - original draft
- Writing - review & editing

### curated\_relation

**type** boolean  
**default** False

### emails

**MARC** 100/700/701\_\_m  
**type** array  
**minItems** 1

Every element of **emails** is:

**type** string  
**format** email  
**minLength** 1

### full\_name

Author name

**MARC** 100/700/701\_\_a

Author name in Latin alphabet (may contain diacritics). Should be of the form ‘family names, first names’, and, except for a possible reordering, be exactly as on the paper.

**example** Smith Davis, John F.K.  
**type** string  
**pattern** ^[\^, ]+(, [\^, ]+)?(, ?[\^, ]+)?  
**minLength** 1

## ids

Identifiers of the author

**MARC** 100/700/701\_\_i/j

**type** array

**minItems** 1

Every element of **ids** is:

**Reference** *id*

## inspire\_roles

**type** array

**minItems** 1

Every element of **inspire\_roles** is:

### supervisor

**MARC** 701 is used for supervisor metadata.

This person is a thesis supervisor. Should be used together with the thesis *document\_type*.

### editor

**MARC** 100/700\_\_e:ed.

This person is an editor of the conference proceedings. Should be used together with the proceedings *document\_type*.

**type** string

**default** author

**minLength** 1

### Allowed values:

- author
- supervisor
- editor

## raw\_affiliations

**MARC** 100/700/701\_\_v

List of full affiliations, as written on the paper.

**type** array

**minItems** 1

Every element of **raw\_affiliations** is:

**Reference** *sourced\_value*

**record****Reference** *author***signature\_block**

Phonetic representation of the name, needed for author disambiguation. This field should not be edited manually.

**type** string**minLength** 1**uuid**

UUID

Used by the BEARD author disambiguation tool. This field should not be edited manually.

**type** string**pattern** `^[0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{12}$`**minLength** 1**book\_series****MARC** 490

List of book series in which this record has been published.

---

**Note:** This field should only be present if *book*, *thesis* or *proceedings* are among *document\_type*.

---

**type** array**minItems** 1Every element of **book\_series** is:

---

**Note:** If the book series has a Journal record, *hep.json#/properties/publication\_info* should be used instead.

---

**type** object**Required** *title***Properties:** *title*, *volume***title**

Title of the book series

**MARC** 490\_\_a**type** string**minLength** 1

## volume

Volume of the book in the series

**MARC** 490\_\_v

**type** string

**minLength** 1

## citeable

**MARC** true corresponds to 980\_\_a:Citeable.

The main effect of setting this field to `true` is that the record is taken into account in the citation summary. The need for this field arises because of limitations in the automatic recognition of cited papers in references: a record is flagged as citeable only if the common way to cite this record can be reliably recognized. Currently, a record is automatically flagged as citeable if it has

- **sufficient information in *publication\_info***: *journal\_title*, *journal\_volume* and (*artid* or *page\_start*), or
- an *arxiv\_eprints* field.

Besides, a record may be manually flagged as citeable by a curator if a user tracks and reports citations to that record.

**type** boolean

Whether this record should be considered as citeable

## collaborations

List of collaborations

**type** array

**minItems** 1

**Required** *value*

Every element of **collaborations** is:

**type** object

**Required** *value*

**Properties:** *record*, *value*

## record

**Reference** *experiment*

## value

Collaboration name

**MARC** 710\_\_g

Collaboration name as it appears on the paper (with the `Collaboration` suffix dropped).

**example** ATLAS instead of ATLAS Collaboration

**example** Particle Data Group

**type** string

**minLength** 1

### control\_number

ID of current record

**MARC** 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

### copyright

List of copyrights to documents in this record

**MARC** 542

**type** array

**minItems** 1

Every element of **copyright** is:

**type** object

**Properties:** *holder, material, statement, url, year*

### holder

Copyright holder

**MARC** 542\_\_d

**example** American Physical Society

**type** string

**minLength** 1

### material

**MARC** 542\_\_e

**Reference** *material*

### statement

Copyright notice

**MARC** 542\_\_f

See [Wikipedia](#) for explanations.

Alternatively to this string, a URL to the copyright notice may be provided in *url*.

**example** © The Author(s) 2015. Published by Oxford University Press on behalf of the Physical Society of Japan.

**type** string

**minLength** 1

## url

Copyright notice URL

**type** string

**format** uri

**minLength** 1

## year

**MARC** 542\_\_g

**type** integer

**minimum** 1000

**maximum** 2050

## core

Whether this record is CORE

Inspire ingests metadata and creates records for many documents. Some of them are relevant for High-Energy Physics, whereas others might be only tangentially related. In the first case, the record is said to be CORE, whereas in the second case it is non-core. CORE records are manually curated to ensure their metadata are up to Inspire's quality standards, but non-core records only receive automatic treatment.

---

**Note:** If this field is not present or set to `false`, the record is considered as non-core. The latter is useful to override automatic core assignments based on other properties of the record.

---

**type** boolean

## corporate\_author

**type** array

**minItems** 1

Every element of **corporate\_author** is:

List of corporate authors

**MARC** 110\_\_a

In case the document has not been signed by a real author, but only mentions the name of an organization.

**example** CERN

**type** string

**minLength** 1

### curated

**MARC 500\_\_a** containing *Temporary entry*, *Temporary record* or *Brief entry* correspond to *false*, otherwise it is *true*.

Whether this record has been curated by a human, to ensure the quality standards of Inspire. Records having the *core* flag are all curated eventually, whereas non-core records are not systematically curated.

**type** boolean

### deleted

Whether this record has been deleted

**MARC 980\_\_a/c:deleted**

**type** boolean

### deleted\_records

**MARC 981\_\_a**

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *literature*

### document\_type

**MARC 980\_\_a**, with corresponding values.

Types of document this record represents.

---

**Note:** As a record aggregates information from multiple sources, it can have multiple document types simultaneously.

---

**article** An article, whether it is published or only a preprint. In the first case, *publication\_info* contains information about the journal in which it was published.

**book** A monograph that does not fit into any other *document\_type*. The *editions*, *imprints*, *isbns* and *book\_series* fields may contain specific metadata about this book.

---

**Note:** A proceeding or a thesis is not considered as a book.

---

**book chapter** A chapter in a book. The field *publication\_info* contains information about the book of which this record is a chapter.

**conference paper** A paper that part of a conference proceedings. The field *publication\_info* contain information about the conference proceedings. In particular, *cnum* contains the identifier of the conference.

**note** A note, not meant for publication.

**proceedings** Proceedings of a conference. The *cnum* field contains the identifier of the conference. If published in a journal, *publication\_info* also contains the journal information. If published as a book, the *editions*, *imprints*, *isbns* and *book\_series* fields may contain specific metadata about this book.

---

**Note:** This is not considered as a book.

---

**report** A report.

**thesis** A thesis. The *thesis\_info* field contains additional information about the thesis. If published as a book, the *editions*, *imprints*, *isbns* and *book\_series* fields may contain specific metadata about this book.

---

**Note:** This is not considered as a book.

---

**type** array

**minItems** 1

Every element of **document\_type** is:

**Reference** *document\_type*

## documents

List of documents attached to the record

**MARC** 8564 (if the URL points to the inspire server)

**type** array

**minItems** 1

Every element of **documents** is:

**type** object

**Required** *key, url*

**Properties:** *description, filename, fulltext, hidden, key, material, original\_url, source, url*

## description

**MARC** 8564\_y

Describes how this document relates to the document that this record represents. :example: table of contents

**type** string

**minLength** 1



**filename**

Filename

Proper name of the file

**type** string

**minLength** 1

**fulltext**

Whether this document contains the fulltext

**type** boolean

**hidden**

Whether this document is hidden

This is `true` if a document is attached to the record for automatic extraction of metadata, but Inspire is not allowed to make it publicly available.

**type** boolean

**key**

Key to the corresponding file, which must match the key in *\_files*.

**type** string

**minLength** 1

**material**

**Reference** *material*

**original\_url**

Original URL of the document

**type** string

**format** uri

**minLength** 1

**source**

**Reference** *source*

## url

MARC 8564\_u

Relative URL to the file containing the document. Its form is `/files/bucket/key`. It can temporarily be the url to download the document from, until actually downloaded.

**type** string

**format** uri-reference

**pattern** `^(https?://|/api/files/).*`

**minLength** 1

## dois

List of DOIs

**type** array

**minItems** 1

Every element of **dois** is:

**type** object

**Required** *value*

**Properties:** *material*, *source*, *value*

## material

MARC 0247\_q

**Reference** *material*

## source

MARC 0247\_9

**Reference** *source*

## value

DOI

MARC 0247\_a

**example** 10.1023/A:1026654312961

**type** string

**pattern** `^10\.\d+(\.\d+)?/\S+$`

**minLength** 1

## editions

List of book editions

**MARC** 250\_\_a

**type** array

**minItems** 1

Every element of **editions** is:

**type** string

**minLength** 1

## energy\_ranges

**MARC** 695\_\_e:n corresponds to the n-th range in the enum, e.g. 695\_\_e:2 is *3-10 GeV*.

Ranges of energies the record refers to. The ranges are inclusive on the right.

**example** 100-300 GeV means that  $100 \text{ GeV} < \text{Energy} \leq 300 \text{ GeV}$ .

**type** array

**minItems** 1

Every element of **energy\_ranges** is:

**type** string

**minLength** 1

### Allowed values:

- 0-3 GeV
- 3-10 GeV
- 10-30 GeV
- 30-100 GeV
- 100-300 GeV
- 300-1000 GeV
- 1-10 TeV
- > 10 TeV

## external\_system\_identifiers

**MARC** 035

List of identifiers of this document on external systems.

**type** array

**minItems** 1

Every element of **external\_system\_identifiers** is:

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

MARC 035\_\_9

Identifies the external system, and allows to interpret unambiguously the *value*.

**example** ADS

**type** string

**pattern** `^\w+$`

**minLength** 1

### value

External identifier

MARC 035\_\_a

Identifies the record in the external system specified by *schema*.

**example** 1999IJTP...38.1113M

**type** string

**pattern** `^\S+$`

**minLength** 1

### figures

List of figures attached to the record

---

**Note:** The order in which the figures appear in this list is the same as in the fulltext.

---

**type** array

**minItems** 1

Every element of **figures** is:

**type** object

**Required** *key, url*

**Properties:** *caption, filename, key, label, material, original\_url, source, url*

### caption

MARC 8564\_y (without the five-digit counter)

**type** string

**minLength** 1

**filename**

Filename

Proper name of the file

**type** string

**minLength** 1

**key**

Key to the corresponding file, which matches a *key* in *\_files*.

**type** string

**minLength** 1

**label**

Label used to refer to this figure in the document.

**type** string

**minLength** 1

**material**

**MARC** not present (figures were extracted only from arXiv preprints)

**Reference** *material*

**original\_url**

Original URL of the document

**type** string

**format** uri

**minLength** 1

**source**

**Reference** *source*

**url**

**MARC** 8564\_u

Relative URL to the file containing the figure. Its form is */files/bucket/key*.

**type** string

**format** uri-reference

**pattern** ^(https?://|/api/files/).\*

**minLength** 1

## funding\_info

Funding information

**MARC** 536

Information about the sources of funding for the research performed in this record.

**type** array

**minItems** 1

Every element of **funding\_info** is:

**type** object

**Properties:** *agency, grant\_number, project\_number*

## agency

Funding agency

**MARC** 536\_\_a

**type** string

**minLength** 1

## grant\_number

Grant number

**MARC** 536\_\_c

**type** string

**minLength** 1

## project\_number

Project number

**MARC** 536\_\_f

**type** string

**minLength** 1

## imprints

List of imprints

**type** array

**minItems** 1

Every element of **imprints** is:

**MARC** 260

When/where/by whom this record was published.

**type** object

**Properties:** *date, place, publisher*

### date

Date of publication

**MARC** 260\_\_c

**type** string

**format** date

**minLength** 1

### place

Place of publication

**MARC** 260\_\_a

---

**Note:** This field is populated if this record has been published in a book, i.e. it makes sense if book, thesis or proceedings are among *document\_type*.

---

**example** Paris

**type** string

**minLength** 1

### publisher

**MARC** 260\_\_b

**example** Springer

**type** string

**minLength** 1

### inspire\_categories

List of Inspire categories

**type** array

**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

## isbns

List of ISBNs

**MARC** 020

**type** array

**minItems** 1

Every element of **isbns** is:

**type** object

**Required** *value*

**Properties:** *medium, value*

## medium

Physical medium to which this ISBN refers

**MARC** 020\_\_b

**type** string

**minLength** 1

**Allowed values:**

- hardcover
- online
- print
- softcover

## value

**MARC** 020\_\_a

**example** 0201021153

**type** string

**pattern** `^\d*[0-9X]$`

**minLength** 1

## keywords

List of keywords

**MARC** 084, 6531 and 695 (corresponding to different *schema*)

Keywords give information about the specific contents of the record, much more precisely than *inspire\_categories*.

**type** array

**minItems** 1

Every element of **keywords** is:



**type** object

**Required** *value*

**Properties:** *schema, source, value*

### schema

Keyword vocabulary

Describes to which vocabulary the keyword in *value* belongs.

#### INIS

**MARC** 695\_\_2:INIS

The keyword is part of the [International Nuclear Information System \(INIS\)](#) vocabulary.

#### INSPIRE

**MARC** 695\_\_2:INSPIRE

The keyword has been assigned by Inspire, and belongs to its vocabulary.

#### JACOW

**MARC** 6531\_\_2:JACOW

The keyword is part of the [Joint Accelerator Conference Website \(JACoW\)](#) vocabulary.

#### PACS

**MARC** 084\_\_2:PACS

The keyword is a number from the [Physics and Astronomy Classification Scheme \(PACS\)](#).

#### PDG

**MARC** 084\_\_2:PDG

The keyword is a [PDG Identifier](#).

---

**Note:** If not present, the keyword is a free-form keyword, not necessarily part of any vocabulary.

---

**type** string

**minLength** 1

#### Allowed values:

- INIS
- INSPIRE
- JACOW
- PACS
- PDG

### source

**Reference** *source*

## value

A keyword

**MARC** 084/6531/695\_\_a (depending on *schema*)

It belongs to the vocabulary specified by *schema*.

**example** black hole: mass (for *schema* = INSPIRE)

**example** 29.27.Fh (for *schema* = PACS)

**example** G033M (for *schema* = PDG)

**type** string

**minLength** 1

## languages

List of languages of the document

**MARC** 041\_\_a

---

**Note:** If this field is not present, the language of the document is assumed to be English.

---

**type** array

**minItems** 1

Every element of **languages** is:

**Reference** *language\_code*

## legacy\_creation\_date

Date of record creation on legacy

**MARC** 961\_\_x

Only present if the record already existed on legacy Inspire.

**type** string

**format** date

**minLength** 1

## legacy\_version

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

## license

List of licenses

**MARC** 540

**type** array

**minItems** 1

Every element of **license** is:

**type** object

**Properties:** *imposing, license, material, url*

## imposing

Organization/person imposing the license

**MARC** 540\_\_b

**example** arXiv

**type** string

**minLength** 1

## license

License statement

**MARC** 540\_\_a

Either the short name of the license or the full license statement.

**example** CC-BY-4.0

**type** string

**minLength** 1

## material

**MARC** 540\_\_3

**Reference** *material*

## url

URL of the license

**MARC** 540\_\_u

URL where the full license statement may be found, if only a short name is provided in `license`.

**type** string

**format** uri

**minLength** 1

### **new\_record**

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *literature*

### **number\_of\_pages**

Number of pages of the document

**MARC** 300

**type** integer

**minimum** 1

### **persistent\_identifiers**

List of persistent identifiers

**MARC** 0247

**type** array

**minItems** 1

Every element of **persistent\_identifiers** is:

**type** object

**Required** *schema, value*

**Properties:** *material, schema, source, value*

### **material**

**Reference** *material*

### **schema**

Type of identifier

**MARC** 0247\_2

Currently, the following identifiers are recognized:

- HDL
- URN

**Note:** DOIs should be put into *dois*, not here.

Although ISBNs can also be mapped to a URN (by prepending `urn:isbn:` to the ISBN), they should be put in the *isbns* field instead.

---

**type** string

**minLength** 1

**Allowed values:**

- HDL
- URN

### source

**Reference** *source*

### value

Persistent identifier

**MARC** 0247\_a

This value gets interpreted as an identifier of the type specified by *schema*.

**example** 10138/168995 (when *schema* = HDL)

**example** urn:nbn:de:hebis:77-diss-1000009520 (when *schema* = URN)

**type** string

**pattern** ^\S+\$

**minLength** 1

### preprint\_date

Preprint release date

**MARC** 269\_\_c

---

**Note:** This is only for preprints. For the publication date of published documents, see *imprints*.

---

**type** string

**format** date

**minLength** 1

## public\_notes

List of public notes

**MARC** 500

Any notes about the document that do not fit into another field. For arXiv eprints, also contains the contents of the comment field on arXiv.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see *\_private\_notes*.

---

**type** array

**minItems** 1

Every element of **public\_notes** is:

**Reference** *sourced\_value*

## publication\_info

List of information on publication

**MARC** 773

**type** array

**minItems** 1

Every element of **publication\_info** is:

**type** object

**Properties:** *artid, cnum, conf\_acronym, conference\_record, curated\_relation, hidden, journal\_issue, journal\_record, journal\_title, journal\_volume, material, page\_end, page\_start, parent\_isbn, parent\_record, parent\_report\_number, pubinfo\_freetext, year*

## artid

Article ID

**MARC** 773\_\_c (when it contains an ID instead of starting page or page range).

For journals which issue IDs for articles instead of (or in complement to) continuous numbers within a volume.

---

**Note:** On legacy, this was treated as a starting page.

---

**type** string

**minLength** 1

## cnum

**MARC** 773\_\_w

When *conference` paper` or ``proceedings* is among *document\_type*, contains the CNUM of the corresponding conference.

**Reference** *cnum*

### conf\_acronym

Conference acronym

**MARC** 773\_\_q

**type** string

**minLength** 1

### conference\_record

record of the conference

**Reference** *conference*

### curated\_relation

**type** boolean

### hidden

Whether this *publication\_info* should be hidden

**MARC** If `true`, this comes from 7731 instead of 773.

This is useful when:

- there are variants in the way this publication is cited,
- **an error was present, has been fixed, but should be preserved** in order to find the record.

**type** boolean

### journal\_issue

**MARC** 773\_\_n

**type** string

**minLength** 1

### journal\_record

Record of the corresponding Journal

**Reference** *journal*

### journal\_title

**MARC** 773\_\_p

Journal title abbreviated as in the corresponding entry in the Journals collection.

**example** Phys.Rev.

**type** string

**minLength** 1

### journal\_volume

**MARC** 773\_\_v

**type** string

**minLength** 1

### material

**MARC** 773\_\_m

**Reference** *material*

### page\_end

Last page of document

**MARC** last page in 773\_\_c (if present)

**type** string

**minLength** 1

### page\_start

First page of document

**MARC** first page in 773\_\_c (if not an *artid*)

**type** string

**minLength** 1

### parent\_isbn

**MARC** 773\_\_z

ISBN of the book of which this document is a part.

**type** string

**pattern** ^\d\*[0-9X]\$

**minLength** 1



## parent\_record

Reference *literature*

## parent\_report\_number

MARC 773\_\_r

Report number of the document of which this record is a part. This applies e.g. for large reports containing several chapters that have been split into individual records.

**type** string

**minLength** 1

## pubinfo\_freetext

MARC 773\_\_x

Unstructured text describing the publication information.

---

**Note:** This field is used when provided with unstructured publication information, e.g. from arXiv. If known, the other fields should be used instead.

---

**type** string

**minLength** 1

## year

Publication year

MARC 773\_\_y

**type** integer

**minimum** 1000

**maximum** 2050

## publication\_type

**type** array

**minItems** 1

Every element of **publication\_type** is:

**type** string

**minLength** 1

**Allowed values:**

- introductory
- lectures

- manual
- review

## record\_affiliations

Affiliations not associated to authors

**MARC** 902

Because of technical limitations, for some older records the affiliations were associated globally to the record instead of its individual authors (using the *affiliations* field).

---

**Note:** This field is present for legacy records and should not be used for new records.

---

**type** array

**minItems** 1

Every element of **record\_affiliations** is:

**type** object

**Required** *value*

**Properties:** *curated\_relation, record, value*

## curated\_relation

**type** boolean

## record

**Reference** *institution*

## value

ICN of affiliation

**MARC** 902\_\_a

Currently, the old *legacy\_ICN* is used here. In the future, this will change and become the new *ICN*.

**type** string

**minLength** 1

## refereed

Whether the record has been peer reviewed

**MARC** true corresponds to 980\_\_a:published

This asserts that the document is considered as peer reviewed. This assessment might differ from the journal's claim to do peer review. As a consequence, most conference papers are not considered as refereed.

**type** boolean

### references

List of references

**MARC** 999C5

**type** array

**minItems** 1

Every element of **references** is:

**type** object

**Properties:** *curated\_relation, legacy\_curated, raw\_refs, record, reference*

### curated\_relation

**type** boolean

### legacy\_curated

**MARC** 999C59:CURATOR corresponds to True

Whether the reference has been modified by a curator to correct errors in reference extraction from the document.

**type** boolean

### raw\_refs

List of raw references

**MARC** 999C5x

These references are unparsed and as they appear in the document.

**type** array

**minItems** 1

Every element of **raw\_refs** is:

**type** object

**Required** *value, schema*

**Properties:** *schema, source, value*

### schema

Content type of the raw reference

**example** text

**example** Elsevier XML

**type** string

**minLength** 1

### source

**Reference** *source*

### value

Content of the raw reference

**type** string

**minLength** 1

### record

**MARC** 999C50

**Reference** *citeable*

### reference

**Reference** *reference*

### related\_records

List of related records

**MARC** 78002, 78502, 78708

**type** array

**minItems** 1

Every element of **related\_records** is:

**Reference** *related\_record*

### report\_numbers

List of report numbers assigned to the record

**MARC** 037

**example** DESY-17-036

**type** array

**minItems** 1

Every element of **report\_numbers** is:

**type** object

**Required** *value*

**Properties:** *hidden, source, value*

## hidden

**MARC** if `true`, the *value* is in `037__z` instead of `037__a`.

Whether this report number should be hidden from the display. This is appropriate for:

- a report number that appears on the document but is not the one officially assigned;
- a report number that has been assigned by one of the authors' institutions, but is not displayed on the document (typically for large collaborations).

**type** `boolean`

## source

**Reference** *source*

## value

**type** `string`

**minLength** `1`

## rpp

**MARC** `210__a:RPP` corresponds to `true`

Whether this record is (part of) the Review of Particle Physics

**type** `boolean`

## self

**Reference** *literature*

## texkeys

List of TeX keys

**MARC** `035 with 9:SPIRESTeX` or `9:INSPIRETeX`

These keys are used to cite this record using TeX/LaTeX/BibTeX. The first one is the valid one being shown in the TeX display formats, the others are the obsolete ones which are kept in order to identify this record by those texkeys.

**type** `array`

**minItems** `1`

Every element of **texkeys** is:

**type** `string`

**minLength** `1`

## thesis\_info

MARC 502

Information on a thesis (degree, date, university)

**type** object

**Properties:** *date, defense\_date, degree\_type, institutions*

### date

MARC 502\_\_d

Publication date of the thesis

**type** string

**format** date

**minLength** 1

### defense\_date

MARC 500\_\_a

Date of the thesis defense. On legacy, this was put in the note field as 'presented on ...'.

**type** string

**format** date

**minLength** 1

### degree\_type

MARC 502\_\_b

**Reference** *degree\_type*

### institutions

MARC 502\_\_c

List of institutions issuing the degree.

**type** array

**minItems** 1

Every element of **institutions** is:

**type** object

**Properties:** *curated\_relation, name, record*

**curated\_relation**

**type** boolean

**name**

**type** string

**minLength** 1

**record**

**Reference** *institution*

**title\_translations**

MARC 242

Translations of the titles in a language that is not the language of the document (see *languages*). Usually, this contains an English translation of the title of a non-English paper, but sometimes a native language if the paper is in English but the title has been translated.

**type** array

**minItems** 1

Every element of **title\_translations** is:

**type** object

**Required** *language, title*

**Properties:** *language, source, subtitle, title*

**language**

**Reference** *language\_code*

**source**

**Reference** *source*

**subtitle**

**type** string

**minLength** 1

## title

**type** string

**minLength** 1

## titles

**MARC** 245

There can be several because the preprint title often differs from the published version. The title should be kept as on the document and in the same language as the document (see *languages*).

Translations go into *title\_translations*.

**type** array

**minItems** 1

List of titles

Every element of **titles** is:

**Reference** *title*

## urls

**MARC** 8564

URLs to external resources related to this record.

---

**Note:** Links to internal resources are found in *figures* or *documents*.

---

**type** array

**minItems** 1

Every element of **urls** is:

**Reference** *url*

## withdrawn

**MARC** true corresponds to 980\_\_a:withdrawn

Whether the paper has been withdrawn (mainly occurs for arXiv preprints). It shouldn't appear in author profiles.

**type** boolean

## 3.1.6 institutions

A record representing an Institution

**type** object

**Required** *\_collections*



**Properties:** *\$schema, ICN, \_bucket, \_collections, \_private\_notes, addresses, control\_number, core, deleted, deleted\_records, external\_system\_identifiers, extra\_words, historical\_data, inactive, inspire\_categories, institution\_hierarchy, institution\_type, legacy\_ICN, legacy\_creation\_date, legacy\_version, name\_variants, new\_record, public\_notes, related\_records, self, urls*

### **\$schema**

**type** string

**format** uri

**minLength** 1

### **ICN**

List of affiliation identifiers

**MARC** 110\_\_t

HEP affiliation unique identifier following the general pattern institution name, city, department. This was called the “new ICN” on legacy.

---

**Note:** This is a list in case the ICN needs to be updated because of a mistake in its assignment. The current one is the first element of the list.

---

**type** array

**minItems** 1

Every element of **ICN** is:

**type** string

**minLength** 1

### **\_bucket**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string

**minLength** 1

### **\_collections**

Collections to which this record belongs

**type** array

**minItems** 1

Every element of **\_collections** is:

**type** string

**minLength** 1

**Allowed values:**

- Institutions

### **\_private\_notes**

List of private notes

**MARC** 595, 667

These notes are only visible to privileged users, not regular users.

**type** array

**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

### **addresses**

List of addresses

**MARC** 371 (secondary addresses contain an x subfield)

The primary address is the first in the list, the secondary addresses, if present, come next. This might happen for instance for institutions distributed on several campuses.

**type** array

**minItems** 1

Every element of **addresses** is:

**Reference** *address*

### **control\_number**

ID of current record

**MARC** 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

### **core**

Whether this institution is CORE

**MARC** 980: CORE corresponds to `true`.

**type** boolean

## deleted

Whether this record has been deleted

**MARC** 980\_\_a/c:deleted

**type** boolean

## deleted\_records

**MARC** 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *institution*

## external\_system\_identifiers

**MARC** 035, 970

List of identifiers of this institution on external systems.

**type** array

**minItems** 1

Every element of **external\_system\_identifiers** is:

May satisfy *any* of the following definitions:

## GRID identifier

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

## schema

**MARC** 035\_\_9:GRID

**type** string

**minLength** 1

**Allowed values:**

- GRID

## value

**MARC 035\_\_a**  
**type** string  
**pattern** ^grid\.\d+\.\w+\$  
**minLength** 1

## HAL identifier

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

## schema

**MARC 035\_\_9:HAL**  
**type** string  
**minLength** 1

**Allowed values:**

- HAL

## value

**MARC 035\_\_a**  
**type** string  
**pattern** ^\d+\$  
**minLength** 1

## ROR identifier

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

## schema

**MARC 035\_\_9:ROR**  
**type** string  
**minLength** 1

**Allowed values:**

- ROR

**value**

**MARC 035\_\_a**  
**type** string  
**pattern** ^https://ror\.org/0\w{6}\d{2}\$  
**minLength** 1

**SPIRES identifier**

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

**schema**

**type** string  
**minLength** 1

**Allowed values:**

- SPIRES

**value**

**MARC 970\_\_a**  
**type** string  
**pattern** ^INST-\d+\$  
**minLength** 1

**extra\_words**

Extra words

**MARC 410\_\_g**

Contains extra words that might be used when searching for this record, but that are not contained in the other visible metadata.

**type** array  
**minItems** 1

Every element of **extra\_words** is:

**type** string  
**minLength** 1

## historical\_data

Historical information about institution

**MARC** 6781\_a

**type** array

**minItems** 1

Every element of **historical\_data** is:

**type** string

**minLength** 1

## inactive

**MARC** 980:DEAD corresponds to `true`

Whether this institution is not active any more, and hence authors of recent articles cannot have it as an affiliation.

**type** boolean

## inspire\_categories

List of Inspire categories

**type** array

**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

## institution\_hierarchy

**MARC** 110\_\_a/b

This record can represent a department, or a group within a department, as opposed to a complete institution. To represent that fact, this field contains the names and acronyms of the specific institution that this record represents, followed by its parents, from most specific to least specific.

### example

```
[
  {
    "name": "Stanford Synchrotron Radiation Lightsource",
    "acronym": "SSRL"
  },
  {
    "name": "SLAC National Accelerator Laboratory",
    "acronym": "SLAC"
  }
]
```

**type** array

**minItems** 1

Every element of **institution\_hierarchy** is:

**type** object

**Properties:** *acronym, name*

### **acronym**

MARC in parenthesis at the end of 110\_a/b

**type** string

**minLength** 1

### **name**

MARC value in 110\_a/b, excluding final parenthesis

**type** string

**minLength** 1

### **institution\_type**

MARC 372\_\_a

**type** array

**minItems** 1

Every element of **institution\_type** is:

**type** string

**minLength** 1

#### **Allowed values:**

- University
- Research Center
- Company
- Other

### **legacy\_ICN**

MARC 110\_\_u

This was called simply “ICN” on legacy.

**type** string

**minLength** 1

### legacy\_creation\_date

Date of record creation on legacy

**MARC** 961\_\_x

Only present if the record already existed on legacy Inspire.

**type** string

**format** date

**minLength** 1

### legacy\_version

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

### name\_variants

**MARC** 410\_\_a

Variants of the institution name. This is used when alternative names are in common use, possibly in English when the official name is not.

**type** array

**minItems** 1

Every element of **name\_variants** is:

**Reference** *sourced\_value*

### new\_record

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *institution*

### public\_notes

List of public notes

**MARC** 680



Any notes about the document that do not fit into another field. For arXiv eprints, also contains the contents of the comment field on arXiv.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see *\_private\_notes*.

---

**type** array

**minItems** 1

Every element of **public\_notes** is:

**Reference** *sourced\_value*

### related\_records

List of related records

**MARC** 510

**type** array

**minItems** 1

Every element of **related\_records** is:

**Reference** *related\_record*

### self

**Reference** *institution*

### urls

**type** array

**minItems** 1

Every element of **urls** is:

**Reference** *url*

## 3.1.7 jobs

A record in the Jobs collection

**type** object

**Required** *\_collections, deadline\_date, description, position, regions, status*

**Properties:** *\$schema, \_bucket, \_collections, \_private\_notes, accelerator\_experiments, acquisition\_source, arxiv\_categories, contact\_details, control\_number, deadline\_date, deleted, deleted\_records, description, external\_job\_identifer, external\_system\_identifiers, institutions, legacy\_creation\_date, legacy\_version, new\_record, position, public\_notes, ranks, reference\_letters, regions, self, status, urls*

## **\$schema**

**type** string  
**format** uri  
**minLength** 1

## **\_bucket**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string  
**minLength** 1

## **\_collections**

Collections to which this record belongs

**type** array  
**minItems** 1

Every element of **\_collections** is:

**type** string  
**minLength** 1

### **Allowed values:**

- Jobs

## **\_private\_notes**

List of private notes

**MARC** 595

These notes are only visible to privileged users, not regular users.

**type** array  
**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

## **accelerator\_experiments**

List of related accelerators/experiments

**type** array  
**minItems** 1

Every element of **accelerator\_experiments** is:

**type** object

**Properties:** *accelerator*, *curated\_relation*, *experiment*, *institution*, *legacy\_name*, *record*

### **accelerator**

**MARC** 693\_\_a

If present, *institution* should contain the institution where this accelerator is located.

---

**Note:** Currently not used, see *legacy\_name*.

---

**type** string

**minLength** 1

### **curated\_relation**

**type** boolean

**default** False

### **experiment**

**MARC** not present.

If present, *institution* should contain the institution where this experiment is located and *accelerator* may contain the accelerator that this experiment is using (if appropriate).

---

**Note:** Currently not used, see *legacy\_name*.

---

**type** string

**minLength** 1

### **institution**

Institution hosting the experiment

**MARC** not present.

---

**Note:** Currently not used, see *legacy\_name*.

---

**type** string

**minLength** 1

## legacy\_name

Identifier of the experiment on legacy

**MARC** 693\_\_e

This field is used when migrating from legacy instead of separate *institution*, *accelerator* and *experiment*. In the future, it will be deprecated and the other fields will be used instead.

**example** CERN-LHC-CMS

**type** string

**minLength** 1

## record

**Reference** *experiment*

## acquisition\_source

**Reference** *acquisition\_source*

## arxiv\_categories

List of research areas for the job

**MARC** 65017a when it contains an arXiv category.

**type** array

**minItems** 1

Every element of **arxiv\_categories** is:

**Reference** *arxiv\_categories*

## contact\_details

**MARC** 270\_\_m/p

Contact information for questions from applicants regarding the job opening.

**type** array

**minItems** 1

Every element of **contact\_details** is:

**Reference** *contact*

## control\_number

ID of current record

**MARC** 001

Read-only field. Also called `record ID` or `recid`.

**type** integer

### deadline\_date

Application deadline

**MARC** 046\_\_i

**type** string

**format** date

**minLength** 1

### deleted

Whether this record has been deleted

**MARC** 980\_\_a/c:deleted

**type** boolean

### deleted\_records

**MARC** 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *job*

### description

Job description

**MARC** 520\_\_a

This may contain rich text in a subset of HTML.

**type** string

**minLength** 1

### external\_job\_identifier

**MARC** 245\_\_a (in parentheses)

Identifier used for the job posting.

**example** TH-SP-2019-LD

**type** string

**minLength** 1

### **external\_system\_identifiers**

**type** array

**minItems** 1

Every element of **external\_system\_identifiers** is:

May satisfy *any* of the following definitions:

**0**

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### **schema**

**type** string

**minLength** 1

**Allowed values:**

- SPIRES

### **value**

**type** string

**pattern** ^JOBS-\d+\$

**minLength** 1

### **institutions**

Institutions opening a position

**type** array

**minItems** 1

Every element of **institutions** is:

**type** object

**Required** *value*

**Properties:** *curated\_relation, record, value*

### **curated\_relation**

**type** boolean

**record**

**Reference** *institution*

**value**

ICN of institution

**MARC** 110\_\_a

Currently, the old *legacy\_ICN* is used here. In the future, this will change and become the new *ICN*.

**type** string

**minLength** 1

**legacy\_creation\_date**

Date of record creation on legacy

**MARC** 961\_\_x

Only present if the record already existed on legacy Inspire.

**type** string

**format** date

**minLength** 1

**legacy\_version**

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

**new\_record**

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *job*

## position

Position title

**MARC** 245\_\_a

**type** string

**minLength** 1

## public\_notes

List of public notes

**MARC** 500

Any notes about the document that do not fit into another field.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see [\\_private\\_notes](#).

---

**type** array

**minItems** 1

Every element of **public\_notes** is:

**Reference** [sourced\\_value](#)

## ranks

**MARC** 656\_\_a

Ranks at which the successful applicant may be hired.

**type** array

**minItems** 1

Every element of **ranks** is:

**Reference** [rank](#)

## reference\_letters

Where to send the reference letters

**type** object

**Properties:** [emails](#), [urls](#)

## emails

**MARC** 270\_\_o (if email address)

Email addresses where reference letters should be sent.

**type** array

**minItems** 1



Every element of **emails** is:

**type** string  
**format** email  
**minLength** 1

## urls

**MARC 270\_\_o** (if URL)

URLs to websites where the reference letters should be uploaded.

**type** array  
**minItems** 1

Every element of **urls** is:

**Reference** *url*

## regions

**MARC 043\_\_a**

World regions where the position is located.

**type** array  
**minItems** 1

Every element of **regions** is:

**type** string  
**minLength** 1

### Allowed values:

- Africa
- Australasia
- Asia
- Europe
- Middle East
- North America
- South America

## self

Url of the record

Url of the record itself

**Reference** *job*

## status

**MARC** Not present

Status of the current job posting. This can be:

**open** The job posting is open and welcomes applications.

**closed** The job posting is closed and it is not possible to apply.

**pending** The job posting has been submitted to INSPIRE and needs to be approved for display.

**type** string

**minLength** 1

### Allowed values:

- closed
- open
- pending

## urls

**MARC** 8564

URLs of web pages related to this job posting.

**type** array

**minItems** 1

Every element of **urls** is:

**Reference** *url*

## 3.1.8 journals

A record representing a Journal

**type** object

**Required** *journal\_title, short\_title, \_collections*

**Properties:** *\$schema, \_bucket, \_collections, \_harvesting\_info, \_private\_notes, book\_series, control\_number, date\_ended, date\_started, deleted, deleted\_records, doi\_prefixes, inspire\_categories, issns, journal\_title, legacy\_creation\_date, legacy\_version, license, new\_record, proceedings, public\_notes, publisher, refereed, related\_records, self, short\_title, title\_variants, urls*

### \$schema

**type** string

**format** uri

**minLength** 1

### **`_bucket`**

Used by *invenio-records-files* to store information about the bucket attached to this record.

---

**Note:** This field is maintained by *invenio-records-files* and should not be edited manually.

---

**type** string

**minLength** 1

### **`_collections`**

Collections to which this record belongs

**type** array

**minItems** 1

Every element of `_collections` is:

**type** string

**minLength** 1

**Allowed values:**

- Journals

### **`_harvesting_info`**

**MARC** 583

Metadata about the harvesting process of this journal.

**type** object

**Properties:** *coverage, date\_last\_harvest, last\_seen\_item, method*

### **`coverage`**

**MARC** 583\_\_a

Whether all articles are automatically added to Inspire (*full*) or a selection is made (*partial*).

**type** string

**minLength** 1

**Allowed values:**

- full
- partial

### date\_last\_harvest

MARC 583\_\_c

Date on which the most recent harvest was performed.

---

**Note:** This date does not necessarily mean that any records were created on that day. This can happen if there was no journal update since the previous time, or that the update did not have any relevant effect for Inspire.

---

**type** string

**format** date

**minLength** 1

### last\_seen\_item

MARC 583\_\_3

Information about last processed item in the harvest. This item can be a volume, an issue or even a specific article.

**type** string

**minLength** 1

### method

MARC 583\_\_i

How the harvesting is performed. Possible values are:

**feed** Inspire receives a feed with publisher updates.

**harvest** harvesting is done through webscraping.

**print** articles are picked manually from the printed journal.

**hepcrawl** harvesting is done through a native `hepcrawl` spider.

**type** string

**minLength** 1

#### Allowed values:

- feed
- harvest
- print
- hepcrawl

### \_private\_notes

List of private notes

MARC 595\_\_a, 667\_\_x

These notes are only visible to privileged users, not regular users.

**type** array

**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

### **book\_series**

**MARC** 980\_\_a:BookSeries corresponds to true

Whether this “journal” is actually a serial whose volumes are books, i.e. a book series.

**type** boolean

### **control\_number**

ID of current record

**MARC** 001

Read-only field. Also called record ID or recid.

**type** integer

### **date\_ended**

**MARC** Not present.

Date of last publication of the journal.

**type** string

**format** date

**minLength** 1

### **date\_started**

**MARC** Not present.

Date of first publication of the journal.

**type** string

**format** date

**minLength** 1

### **deleted**

Whether this record has been deleted

**MARC** 980\_\_a/c:deleted

**type** boolean

### deleted\_records

MARC 981\_\_a

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *journal*

### doi\_prefixes

List of DOI prefixes for this journal

MARC 677\_\_d

This DOI prefix is the common start of DOIs in this journals, that all articles share.

---

**Note:** This is a list because journals can change publishers, and the new publisher will often assign new DOIs in its own prefix.

---

**type** array

**minItems** 1

Every element of **doi\_prefixes** is:

**type** string

**pattern** ^10\.\d+(\.\d+)?/.\*\$

**minLength** 1

### inspire\_categories

List of Inspire categories

**type** array

**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

### issns

List of ISSNs

MARC 022

**type** array

**minItems** 1

Every element of **issns** is:

**type** object

**Required** *value*

**Properties:** *medium, value*

### **medium**

Physical medium to which this ISSN refers

**MARC** 022\_\_b

**type** string

**minLength** 1

**Allowed values:**

- online
- print

### **value**

**MARC** 022\_\_a

**example** 0295-5075

**type** string

**pattern** ^\d{4}-\d{3}[\dX]\$

**minLength** 1

### **journal\_title**

**Reference** *title*

**MARC** 130

### **legacy\_creation\_date**

Date of record creation on legacy

**MARC** 961\_\_x

Only present if the record already existed on legacy Inspire.

**type** string

**format** date

**minLength** 1

### legacy\_version

Last record update

**MARC** 005

Used to indicate the version of the record on legacy.

**type** string

**minLength** 1

### license

**MARC** 540

**type** object

**Properties:** *license, url*

### license

License statement

**MARC** 540\_\_a

Either the short name of the license or the full license statement.

**example** CC-BY-4.0

**type** string

**minLength** 1

### url

URL of the license

**MARC** 540\_\_u

URL where the full license statement may be found, if only a short name is provided in `license`.

**type** string

**format** uri

**minLength** 1

### new\_record

Record replacing this one

**MARC** 970\_\_d

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *journal*



## proceedings

**MARC 690\_\_a:Proceedings** corresponds to `true`

Whether this journal publishes conference proceedings. If it publishes both conference proceedings and peer reviewed articles (depending on issue), both this field and *refereed* are `true`.

**type** `boolean`

## public\_notes

List of public notes

**MARC 500\_\_a, 640\_\_a, 680\_\_i**

Any notes about the document that do not fit into another field.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see *\_private\_notes*.

---

**type** `array`

**minItems** `1`

Every element of **public\_notes** is:

**Reference** *sourced\_value*

## publisher

List of publishers

**MARC 643\_\_b**

The first element of the list is the current publisher of the journal.

---

**Note:** This is a list because journals can change publishers.

---

**type** `array`

**minItems** `1`

Every element of **publisher** is:

**type** `string`

**minLength** `1`

## refereed

**MARC 690\_\_a:Peer review** corresponds to `true`, **690\_\_a:NON-PUBLISHED** to `false`

Whether this journal is considered to perform peer review. This assessment might differ from the journal's.

If the journal does not publish proceedings *proceedings*, all articles in it are flagged as *refereed*. Otherwise, it is only the case if the article is not a conference paper.

**type** `boolean`

### related\_records

List of related records

**MARC** 530

**type** array

**minItems** 1

Every element of **related\_records** is:

**Reference** *related\_record*

### self

**Reference** *journal*

### short\_title

**MARC** 711\_\_a

Normalized title of the journal

**example** Phys.Rev. D

**type** string

**minLength** 1

### title\_variants

List of journal name variants

**MARC** 730\_\_a

These name variants appear in references and are used to properly recognize citations.

**example** PHYS REVIEW

**type** array

**minItems** 1

Every element of **title\_variants** is:

**type** string

**minLength** 1

### urls

**type** array

**minItems** 1

Every element of **urls** is:

**Reference** *url*

### 3.1.9 seminars

Seminar

**type** object

**Required** *\_collections, end\_datetime, inspire\_categories, speakers, start\_datetime, timezone*

**Properties:** *\$schema, \_collections, \_private\_notes, abstract, acquisition\_source, address, captioned, contact\_details, control\_number, core, deleted, deleted\_records, end\_datetime, inspire\_categories, join\_urls, keywords, literature\_records, material\_urls, new\_record, public\_notes, self, series, speakers, start\_datetime, timezone, title, urls*

#### **\$schema**

**type** string

**format** uri

**minLength** 1

#### **\_collections**

Collections to which this record belongs

**type** array

**minItems** 1

Every element of **\_collections** is:

**type** string

**minLength** 1

**Allowed values:**

- Seminars

#### **\_private\_notes**

List of private notes

These notes are only visible to privileged users, not regular users.

**type** array

**minItems** 1

Every element of **\_private\_notes** is:

**Reference** *sourced\_value*

#### **abstract**

**Reference** *sourced\_value*

#### **acquisition\_source**

**Reference** *acquisition\_source*

### address

Location of the seminar

**Reference** *address*

### captioned

Whether this seminar contains captions

**type** boolean

### contact\_details

Contact information

**type** array

**minItems** 1

Every element of **contact\_details** is:

**Reference** *contact*

### control\_number

ID of current record

Read-only field. Also called `record ID` or `recid`.

**type** integer

### core

Whether this record is CORE

**type** boolean

### deleted

Whether this record has been deleted

**type** boolean

### deleted\_records

List of records that were deleted because they were replaced by this one. This typically happens when merging two records: one of them gets enriched with the information of the other one, which is then superfluous and gets deleted.

For the opposite concept, see *new\_record*.

**type** array

**minItems** 1

Every element of **deleted\_records** is:

**Reference** *seminar*

### end\_datetime

Seminar end datetime in UTC

**type** string  
**format** date-time  
**minLength** 1

### inspire\_categories

**type** array  
**minItems** 1

Every element of **inspire\_categories** is:

**Reference** *inspire\_field*

### join\_urls

List of URLs to attend the seminar remotely

**type** array  
**minItems** 1

Every element of **join\_urls** is:

**Reference** *url*

### keywords

List of keywords

Keywords give information about the specific contents of the record, much more precisely than *inspire\_categories*.

**type** array  
**minItems** 1

Every element of **keywords** is:

**type** object  
**Required** *value*

**Properties:** *schema, source, value*

### schema

Keyword vocabulary

Describes to which vocabulary the keyword in *value* belongs.

**INSPIRE** The keyword has been assigned by Inspire, and belongs to its vocabulary.

---

**Note:** If not present, the keyword is a free-form keyword, not necessarily part of any vocabulary.

---

**type** string

**minLength** 1

**Allowed values:**

- INSPIRE

### source

**Reference** *source*

### value

A keyword

It belongs to the vocabulary specified by *schema*.

**example** black hole: mass (for *schema* = INSPIRE)

**type** string

**minLength** 1

### literature\_records

List of literature records related to the seminar

**type** array

**minItems** 1

Every element of **literature\_records** is:

**type** object

**Required** *record*

**Properties:** *curated\_relation*, *record*

### curated\_relation

**type** boolean

### record

**Reference** *literature*

### material\_urls

Links to related material such as slides.

**type** array

**minItems** 1

Every element of **material\_urls** is:

**Reference** *url*

### **new\_record**

Record replacing this one

Contains a reference to the record replacing the current one, if it is marked as *deleted*.

**Reference** *seminar*

### **public\_notes**

List of public notes

Any notes about the document that do not fit into another field.

---

**Note:** These notes are publicly visible. For notes not shown to regular users, see *\_private\_notes*.

---

**type** array

**minItems** 1

Every element of **public\_notes** is:

**Reference** *sourced\_value*

### **self**

**Reference** *seminar*

### **series**

Seminar series that this is part of

**type** array

**minItems** 1

Every element of **series** is:

**type** object

**Required** *name*

**Properties:** *name, number*

### **name**

Name of the seminar series

**type** string

**minLength** 1

## number

Number of the seminar in the series

**type** integer

## speakers

List of speakers

**type** array

**minItems** 1

Every element of **speakers** is:

**type** object

**Required** *name*

**Properties:** *affiliations, curated\_relation, ids, name, record*

## affiliations

**type** array

**minItems** 1

Every element of **affiliations** is:

**type** object

**Required** *value*

**Properties:** *curated\_relation, record, value*

## curated\_relation

**type** boolean

## record

**Reference** *institution*

## value

ICN of affiliation

Currently, the old *legacy\_ICN* is used here. In the future, this will change and become the new *ICN*.

**type** string

**minLength** 1



**curated\_relation**

**type** boolean

**ids**

List of speaker's IDs

**type** array

**minItems** 1

Every element of **ids** is:

**Reference** *id*

**name**

The full name of the speaker, written in the standard last names, first names form.

**example** Smith, John P.S.

**type** string

**pattern** ^[\^, ]+(, [\^, ]+)?(, ?[\^, ]+)?\$

**minLength** 1

**record**

**Reference** *author*

**start\_datetime**

Seminar start date time in UTC

**type** string

**format** date-time

**minLength** 1

**timezone**

Timezone in which the seminar takes place

**type** string

**format** timezone

**minLength** 1

## title

**Reference** *title*

Title of the seminar

## urls

**type** array

**minItems** 1

Every element of **urls** is:

**Reference** *url*

## 3.1.10 elements

### acquisition\_source

Origin of the metadata in the record

**MARC** 541

Only the first source is stored: if the record later gets enriched with metadata coming from a second source, the *acquisition\_source* is not updated.

**type** object

**Properties:** *datetime, email, internal\_uid, method, orcid, source, submission\_number*

### datetime

Date on which the metadata was obtained

**MARC** 541\_\_d

This does not necessarily coincide with the creation date of the record, as there might be some delay between the moment the original information is obtained and a record is finally created in the system.

**type** string

**format** date-time

**minLength** 1

### email

Email address associated to the submitter's account

**MARC** 541\_\_b

This only gets populated when *method* is *submitter*.

**type** string

**format** email

**minLength** 1

## internal\_uid

Inspire user ID of the submitter

**MARC** 541\_\_a with `inspire:uid:` prefix.

This only gets populated when *method* is *submitter*.

**type** integer

## method

How the metadata was obtained

Possible values are:

**submitter** when obtained from a user submission. In this case, *orcid*, *internal\_uid* and *email* store identifiers of the submitter, and *submission\_number* an ID of the submission.

**oai** when obtained by OAI-PMH harvesting.

**batchuploader** when obtained through the batch uploader.

**hepcrawl** when obtained from a hepcrawl crawler. In this case, *source* stores the publisher or the repository that was crawled.

**beard** when created after author disambiguation

**type** string

**minLength** 1

**Allowed values:**

- submitter
- oai
- batchuploader
- hepcrawl
- beard

## orcid

ORCID of the submitter

**MARC** 541\_\_a with `orcid:` prefix

This only gets populated when *method* is *submitter*.

**type** string

**pattern** `^\d{4}-\d{4}-\d{4}-\d{3}[0-9X]$`

**minLength** 1

## source

**Reference** *source*

### submission\_number

Holding pen record ID of the submission

**MARC** 541\_\_e

This only gets populated when *method* is *submitter*.

**type** string

**minLength** 1

### address

Location associated to the record

**type** object

**Properties:** *cities*, *country\_code*, *latitude*, *longitude*, *place\_name*, *postal\_address*, *postal\_code*, *state*

### cities

List of cities

Sometimes conferences and institutes are officially located in a little-known place, but their location is widely known as being in a major neighboring city. In this way, we can store information about both the technically correct but little-known place and the well-known but technically incorrect place.

The cities should be ordered from most specific to least specific.

**example** ["Batavia", "Chicago"]

**type** array

**minItems** 1

Every element of **cities** is:

A city

**example** Geneva

**type** string

**minLength** 1

### country\_code

**Reference** *country\_code*

### latitude

Latitude of the location

**example** 50.7736

**type** number

## longitude

Longitude of the location

**example** 6.07532

**type** number

## place\_name

Name of the specific place where this is located.

- For a Conference record, this could be an institution or a hotel hosting the conference.

**MARC** 270\_\_b

**example** Dan Panorama hotel

- For an Institution record representing an independent institute and not a department, this could be a university where it is located.

**MARC** Not present.

**example** U. Libre Brussels

**type** string

**minLength** 1

## postal\_address

Full postal address in the original language, with one element for every line, but excluding the country.

**example** ["Notkestraße 85", "22607 Hamburg"]

---

**Note:** This is used for Institution records, but not for Conferences.

---

**type** array

**minItems** 1

Every element of **postal\_address** is:

**type** string

**minLength** 1

## postal\_code

Postal code of the location

**example** 1050

**type** string

**minLength** 1

## state

State or province of the location

For American states, standard 2-letter codes are used.

```
example TX
type string
minLength 1
```

## arxiv\_categories

A category that currently exists on arXiv. Note that some categories have been renamed and are not in this list. These are taken from the [arXiv API documentation](#).

```
example math.FA instead of its previous name, funct-an
type string
minLength 1
```

### Allowed values:

- astro-ph
- astro-ph.CO
- astro-ph.EP
- astro-ph.GA
- astro-ph.HE
- astro-ph.IM
- astro-ph.SR
- cond-mat
- cond-mat.dis-nn
- cond-mat.mes-hall
- cond-mat.mtrl-sci
- cond-mat.other
- cond-mat.quant-gas
- cond-mat.soft
- cond-mat.stat-mech
- cond-mat.str-el
- cond-mat.supr-con
- cs
- cs.AI
- cs.AR
- cs.CC
- cs.CE

- cs.CG
- cs.CL
- cs.CR
- cs.CV
- cs.CY
- cs.DB
- cs.DC
- cs.DL
- cs.DM
- cs.DS
- cs.ET
- cs.FL
- cs.GL
- cs.GR
- cs.GT
- cs.HC
- cs.IR
- cs.IT
- cs.LG
- cs.LO
- cs.MA
- cs.MM
- cs.MS
- cs.NA
- cs.NE
- cs.NI
- cs.OH
- cs.OS
- cs.PF
- cs.PL
- cs.RO
- cs.SC
- cs.SD
- cs.SE
- cs.SI
- cs.SY

- econ
- econ.EM
- econ.GN
- econ.TH
- eess
- eess.AS
- eess.IV
- eess.SP
- eess.SY
- gr-qc
- hep-ex
- hep-lat
- hep-ph
- hep-th
- math
- math-ph
- math.AC
- math.AG
- math.AP
- math.AT
- math.CA
- math.CO
- math.CT
- math.CV
- math.DG
- math.DS
- math.FA
- math.GM
- math.GN
- math.GR
- math.GT
- math.HO
- math.IT
- math.KT
- math.LO
- math.MG



- math.MP
- math.NA
- math.NT
- math.OA
- math.OC
- math.PR
- math.QA
- math.RA
- math.RT
- math.SG
- math.SP
- math.ST
- nlin
- nlin.AO
- nlin.CD
- nlin.CG
- nlin.PS
- nlin.SI
- nucl-ex
- nucl-th
- physics
- physics.acc-ph
- physics.ao-ph
- physics.app-ph
- physics.atm-clus
- physics.atom-ph
- physics.bio-ph
- physics.chem-ph
- physics.class-ph
- physics.comp-ph
- physics.data-an
- physics.ed-ph
- physics.flu-dyn
- physics.gen-ph
- physics.geo-ph
- physics.hist-ph

- physics.ins-det
- physics.med-ph
- physics.optics
- physics.plasm-ph
- physics.pop-ph
- physics.soc-ph
- physics.space-ph
- q-bio
- q-bio.BM
- q-bio.CB
- q-bio.GN
- q-bio.MN
- q-bio.NC
- q-bio.OT
- q-bio.PE
- q-bio.QM
- q-bio.SC
- q-bio.TO
- q-fin
- q-fin.CP
- q-fin.EC
- q-fin.GN
- q-fin.MF
- q-fin.PM
- q-fin.PR
- q-fin.RM
- q-fin.ST
- q-fin.TR
- quant-ph
- stat
- stat.AP
- stat.CO
- stat.ME
- stat.ML
- stat.OT
- stat.TH

**cnum**

CNUM identifier of the conference

The CNUM is based on the starting day of the conference, with an extra number appended to distinguish conferences starting on the first day.

**example** The first conference (in order of creation on Inspire) starting on December 25, 1987 has CNUM C87-12-25 whereas, say, the third one, has CNUM C87-12-25.2

**type** string

**pattern** ^C\d\d-\d\d-\d\d(\.\d+)?\$

**minLength** 1

**contact**

Contact person's data

**type** object

**Properties:** *curated\_relation, email, name, record*

**curated\_relation**

**type** boolean

**email**

Contact person's e-mail

**type** string

**format** email

**minLength** 1

**name**

Contact person's name

**type** string

**minLength** 1

**record**

**Reference** *author*

## country\_code

ISO 3166-1 or 3166-3 alpha 2 country code

**type** string

**maxLength** 2

**minLength** 2

### Allowed values:

- AI
- AN
- BQ
- BU
- BY
- CS
- CT
- DD
- DY
- FQ
- FX
- GE
- HV
- JT
- MI
- NH
- NQ
- NT
- PC
- PU
- PZ
- RH
- SK
- SU
- TP
- VD
- WK
- YD
- YU
- ZR

- AW
- AF
- AO
- AX
- AL
- AD
- AE
- AR
- AM
- AS
- AQ
- TF
- AG
- AU
- AT
- AZ
- BI
- BE
- BJ
- BF
- BD
- BG
- BH
- BS
- BA
- BL
- BZ
- BM
- BO
- BR
- BB
- BN
- BT
- BV
- BW
- CF

- CA
- CC
- CH
- CL
- CN
- CI
- CM
- CD
- CG
- CK
- CO
- KM
- CV
- CR
- CU
- CW
- CX
- KY
- CY
- CZ
- DE
- DJ
- DM
- DK
- DO
- DZ
- EC
- EG
- ER
- EH
- ES
- EE
- ET
- FI
- FJ
- FK

- FR
- FO
- FM
- GA
- GB
- GG
- GH
- GI
- GN
- GP
- GM
- GW
- GQ
- GR
- GD
- GL
- GT
- GF
- GU
- GY
- HK
- HM
- HN
- HR
- HT
- HU
- ID
- IM
- IN
- IO
- IE
- IR
- IQ
- IS
- IL
- IT

- JM
- JE
- JO
- JP
- KZ
- KE
- KG
- KH
- KI
- KN
- KR
- KW
- LA
- LB
- LR
- LY
- LC
- LI
- LK
- LS
- LT
- LU
- LV
- MO
- MF
- MA
- MC
- MD
- MG
- MV
- MX
- MH
- MK
- ML
- MT
- MM



- ME
- MN
- MP
- MZ
- MR
- MS
- MQ
- MU
- MW
- MY
- YT
- NA
- NC
- NE
- NF
- NG
- NI
- NU
- NL
- NO
- NP
- NR
- NZ
- OM
- PK
- PA
- PN
- PE
- PH
- PW
- PG
- PL
- PR
- KP
- PT
- PY

- PS
- PF
- QA
- RE
- RO
- RU
- RW
- SA
- SD
- SN
- SG
- GS
- SH
- SJ
- SB
- SL
- SV
- SM
- SO
- PM
- RS
- SS
- ST
- SR
- SI
- SE
- SZ
- SX
- SC
- SY
- TC
- TD
- TG
- TH
- TJ
- TK

- TM
- TL
- TO
- TT
- TN
- TR
- TV
- TW
- TZ
- UG
- UA
- UM
- UY
- US
- UZ
- VA
- VC
- VE
- VG
- VI
- VN
- VU
- WF
- WS
- YE
- ZA
- ZM
- ZW

### **degree\_type**

Academic degree type

The *other* value means that the degree type is not known or is not among the more specific values.

**type** string

**minLength** 1

**Allowed values:**

- other

- diploma
- bachelor
- laurea
- master
- phd
- habilitation

### document\_type

**type** string  
**default** article  
**minLength** 1

#### Allowed values:

- activity report
- article
- book
- book chapter
- conference paper
- note
- proceedings
- report
- thesis

### id

May satisfy *any* of the following definitions:

#### CERN institutional ID

Also called CCID (e.g. on legacy).

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string  
**minLength** 1

**Allowed values:**

- CERN

#### value

**example** CERN-12345

**type** string

**pattern** ^CERN-\d+\$

**minLength** 1

#### DESY identifier

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

#### schema

**type** string

**minLength** 1

**Allowed values:**

- DESY

#### value

**example** DESY-15456

**type** string

**pattern** ^DESY-\d+\$

**minLength** 1

#### Google Scholar profile identifier

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

#### schema

**type** string

**minLength** 1

**Allowed values:**

- GOOGLESCHOLAR

#### value

**example** AZUAeayXuibs  
**type** string  
**pattern** ^(\w|-){12}\$  
**minLength** 1

### Inspire BAI

The BAI (BibAuthorID) is based on the name of the author and is used as a human-friendly unique identifier. Its general form is a `firstnames.lastnames.number`, where `firstnames` is a list of first names (or initials) separated by dots, and similarly for `lastnames`. Only ASCII letters (without diacritics), numbers, dashes and apostrophes are allowed in the names. By default, only initials of first names are used, but this may be changed at a cataloguer's discretion in order to disambiguate otherwise homonymous authors.

---

**Note:** This identifier is not persistent, as it can be renamed to become more explicit (e.g. transforming an initial into a full name), or disappear when merging two different authors.

---

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

#### schema

**type** string  
**minLength** 1

#### Allowed values:

- INSPIRE BAI

#### value

**example** A.Einstein.1  
**type** string  
**pattern** ^((\w|\-|\'|)+\.)+\d+\$  
**minLength** 1

### Inspire ID

This identifier is assigned to any curated author record (HEPNAMES on legacy). It is mainly used by large collaborations (providing an *authors.xml* file) to uniquely identify the authors of their articles.

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

#### schema

**type** string

**minLength** 1

**Allowed values:**

- INSPIRE ID

#### value

**example** INSPIRE-12345678

**type** string

**pattern** ^INSPIRE-\d{8}\$

**minLength** 1

#### JACoW ID

Author identifier assigned by the Joint Accelerator Conferences Website (JACoW).

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

#### schema

**type** string

**minLength** 1

**Allowed values:**

- JACOW

#### value

**example** JACOW-12345678

**type** string

**pattern** ^JACoW-\d{8}\$

**minLength** 1

## KAKEN ID

Author identifier assigned by the Japanese [KAKEN](#) grants database.

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string

**minLength** 1

**Allowed values:**

- KAKEN

### value

**example** KAKEN-12345678

**type** string

**pattern** ^KAKEN-\d{8}\$

**minLength** 1

## LinkedIn profile

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string

**minLength** 1

**Allowed values:**

- LINKEDIN

### value

Identifier of LinkedIn profile, i.e. the part after `linkedin.com/in/` in the URL.

**example** kylecranmer

**type** string

**minLength** 1



## ORCID

**ORCID** provides an identifier for individuals to use with their name as they engage in research, scholarship, and innovation activities.

The ORCID identifier can be resolved by prepending `http://orcid.org` to the *value*, in order to get the ORCID record of the person.

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string

**minLength** 1

**Allowed values:**

- ORCID

### value

**example** 0000-0012-1234-5647

**type** string

**format** orcid

**pattern** `^\d{4}-\d{4}-\d{4}-\d{3}[0-9X]$`

**minLength** 1

## ResearcherID

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

### schema

**type** string

**minLength** 1

**Allowed values:**

- RESEARCHERID

## value

```
example A-1009-2008
type string
pattern ^[A-z]-\d{4}-\d{4}$
minLength 1
```

## SCOPUS author identifier

```
type object
Required schema, value
```

**Properties:** *schema, value*

## schema

```
type string
minLength 1
```

**Allowed values:**

- SCOPUS

## value

```
example 13245678901
type string
pattern ^\d{10,11}$
minLength 1
```

## SLAC identifier

```
type object
Required schema, value
```

**Properties:** *schema, value*

## schema

```
type string
minLength 1
```

**Allowed values:**

- SLAC

**value**

**example** SLAC-123456  
**type** string  
**pattern** ^SLAC-\d+\$  
**minLength** 1

**SPIRES HEPNAMES ID**

HEPNAMES entry in the old SPIRES system.

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

**schema**

**type** string  
**minLength** 1

**Allowed values:**

- SPIRES

**value**

**example** HEPNAMES-123456  
**type** string  
**pattern** ^HEPNAMES-\d+\$  
**minLength** 1

**Twitter handle**

**type** object  
**Required** *schema, value*

**Properties:** *schema, value*

**schema**

**type** string  
**minLength** 1

**Allowed values:**

- TWITTER

## value

A twitter handle, often preceded by @.

**example** Prof\_S\_Hawking

**type** string

**minLength** 1

## VIAF record ID

Record ID in the Virtual International Authority File (VIAF).

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

## schema

**type** string

**minLength** 1

**Allowed values:**

- VIAF

## value

**example** 75121530

**type** string

**pattern** `^\d{7,9}$`

**minLength** 1

## Wikipedia Page\_name of the person

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

## schema

**type** string

**minLength** 1

**Allowed values:**

- WIKIPEDIA

**value**

A page name on the English Wikipedia, i.e. the Page\_name part in [http://en.wikipedia.org/wiki/Page\\_name](http://en.wikipedia.org/wiki/Page_name).

**example** Albert\_Einstein

Non-English Wikipedia pages can be referenced by prepending the language.

**example** it:Fabiola\_Gianotti

**type** string

**minLength** 1

**arXiv author identifier**

More info at arXiv.

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

**schema**

**type** string

**minLength** 1

**Allowed values:**

- ARXIV

**value**

**example** warner\_s\_1

**type** string

**pattern** `^\w+_(\w_)?\d+$`

**minLength** 1

**inspire\_experiment\_classification**

An INSPIRE experiment category. This is a hierarchical classification, with | used as a separator.

**type** string

**minLength** 1

**Allowed values:**

- Collider Experiments
- Collider Experiments|Hadrons
- Collider Experiments|Hadrons|p anti-p

- Collider Experiments|Hadrons|p p
- Collider Experiments|e+ e-
- Collider Experiments|e p
- Collider Experiments|Heavy Flavor Factory
- Collider Experiments|Heavy ion
- Collider Experiments|Detector development
- Fixed Target Experiments
- Fixed Target Experiments|High-momentum transfer
- Fixed Target Experiments|Hadron Spectroscopy
- Fixed Target Experiments|Deep inelastic scattering
- Fixed Target Experiments|Drell-Yan/Dilepton production
- Fixed Target Experiments|Flavor physics
- Fixed Target Experiments|Lepton precision experiments
- Fixed Target Experiments|Neutron/proton precision experiments
- Neutrino (flavor) experiments
- Neutrino (flavor) experiments|Accelerator
- Neutrino (flavor) experiments|Accelerator|short-baseline
- Neutrino (flavor) experiments|Accelerator|long-baseline
- Neutrino (flavor) experiments|Reactor
- Neutrino (flavor) experiments|Reactor|ultra-short-baseline
- Neutrino (flavor) experiments|Reactor|longer baselines
- Neutrino (flavor) experiments|Non terrestrial
- Neutrino (flavor) experiments|Non terrestrial|Atmospheric
- Neutrino (flavor) experiments|Non terrestrial|Solar
- Neutrino (flavor) experiments|Non terrestrial|Cosmic
- Neutrino (flavor) experiments|Neutrinoless double beta decay
- Neutrino (flavor) experiments|Neutrino mass
- Dark matter search experiments
- Dark matter search experiments|Non-accelerator
- Dark matter search experiments|Axion search experiments
- Dark matter search experiments|Dark Forces
- Cosmic ray/Gamma ray experiments
- Cosmic ray/Gamma ray experiments|Ground array
- Cosmic ray/Gamma ray experiments|Cerenkov array
- Cosmic ray/Gamma ray experiments|Satellite
- Cosmic ray/Gamma ray experiments|Balloon

- Other Rare-process/exotic experiments
- Other Rare-process/exotic experiments|Proton decay
- Other Rare-process/exotic experiments|Modified gravity and space-time
- Other Rare-process/exotic experiments|Magnetic monopoles
- Other Rare-process/exotic experiments|Fractionally charged particles
- Accelerator Test Facility Experiments
- Accelerator Test Facility Experiments|Electron and positron beams
- Accelerator Test Facility Experiments|Muon beams
- Accelerator Test Facility Experiments|Proton beams
- Accelerator Test Facility Experiments|Neutrino beams
- Astronomy experiments
- Astronomy experiments|CMB
- Astronomy experiments|Survey
- Astronomy experiments|Supernovae
- Astronomy experiments|Gravitational waves
- Astronomy experiments|Gravitational lensing/Dark matter
- Non-experimental
- Non-experimental|Data Analysis
- Non-experimental|Simulation tools
- Non-experimental|Simulation tools|Detector Simulation
- Non-experimental|Simulation tools|Event Simulation
- Non-experimental|Parton Distribution Fits
- Non-experimental|Lattice Gauge Theory
- Non-experimental|Neutrino Physics

## inspire\_field

Inspire category

**MARC** 65017 with 2 : INSPIRE

The Inspire category (also called field category) classifies the subject this record is related to. This classification schema is similar to the arXiv one (see *elements/arxiv\_categories.json*) and there is a way to map from arXiv categories to Inspire categories.

**type** object

**Required** *term*

**Properties:** *source, term*

### source

**type** string

**minLength** 1

#### Allowed values:

- arxiv
- curator
- cds
- magpie
- user

### term

**type** string

**minLength** 1

#### Allowed values:

- Accelerators
- Astrophysics
- Computing
- Condensed Matter
- Data Analysis and Statistics
- Experiment-HEP
- Experiment-Nucl
- General Physics
- Gravitation and Cosmology
- Instrumentation
- Lattice
- Math and Math Physics
- Other
- Phenomenology-HEP
- Quantum Physics
- Theory-HEP
- Theory-Nucl



**language\_code**

ISO 639-1 alpha 2 language code

Language code according to [ISO 639-1](#).

**example** el for Greek.

**type** string

**maxLength** 2

**minLength** 2

**Allowed values:**

- aa
- ab
- af
- ak
- am
- ar
- an
- as
- av
- ae
- ay
- az
- ba
- bm
- be
- bn
- bi
- bo
- bs
- br
- bg
- ca
- cs
- ch
- ce
- cu
- cv
- kw

- co
- cr
- cy
- da
- de
- dv
- dz
- el
- en
- eo
- et
- eu
- ee
- fo
- fa
- fj
- fi
- fr
- fy
- ff
- gd
- ga
- gl
- gv
- gn
- gu
- ht
- ha
- sh
- he
- hz
- hi
- ho
- hr
- hu
- hy

- ig
- io
- ii
- iu
- ie
- ia
- id
- ik
- is
- it
- jv
- ja
- kl
- kn
- ks
- ka
- kr
- kk
- km
- ki
- rw
- ky
- kv
- kg
- ko
- kj
- ku
- lo
- la
- lv
- li
- ln
- lt
- lb
- lu
- lg

- mh
- ml
- mr
- mk
- mg
- mt
- mn
- mi
- ms
- my
- na
- nv
- nr
- nd
- ng
- ne
- nl
- nn
- nb
- no
- ny
- oc
- oj
- or
- om
- os
- pa
- pi
- pl
- pt
- ps
- qu
- rm
- ro
- rn
- ru

- sg
- sa
- si
- sk
- sl
- se
- sm
- sn
- sd
- so
- st
- es
- sq
- sc
- sr
- ss
- su
- sw
- sv
- ty
- ta
- tt
- te
- tg
- tl
- th
- ti
- to
- tn
- ts
- tk
- tr
- tw
- ug
- uk
- ur

- uz
- ve
- vi
- vo
- wa
- wo
- xh
- yi
- yo
- za
- zh
- zu

## material

Material to which the field refers

The possible values are:

**addendum** for a document extending the main publication.

**additional material** for additional material attached to the main publication, that is neither the more specific `data` or `software`.

**data** for data related to the main publication.

**erratum** for a document correcting errors in the content of the main publication. When the errors are not in the content but in the publication process, it is considered as an `editorial note` instead.

**editorial note** for a document correcting errors in the metadata of the publication but not its content. Otherwise, it is an `erratum`.

**preprint** for a document that is made public (typically on `arXiv`) but not yet published formally.

**publication** for the main publication. If it's not published, it's a `preprint`.

**reprint** for a reprint of the main publication.

**software** for software related to the main publication.

**translation** for a translation of the main publication.

**type** `string`

**minLength** `1`

**Allowed values:**

- addendum
- additional material
- data
- erratum
- editorial note

- preprint
- publication
- reprint
- software
- translation

## rank

Rank of academic position

**type** string

**minLength** 1

**Allowed values:**

- STAFF
- SENIOR
- JUNIOR
- VISITOR
- POSTDOC
- PHD
- MASTER
- UNDERGRADUATE
- OTHER

## records-files

File schema.

Describe information needed for files in records.

**type** object

**Properties:** *bucket, checksum, file\_id, filename, key, size, version\_id*

### bucket

**type** string

**minLength** 1

### checksum

**type** string

**minLength** 1

**file\_id**

**type** string  
**minLength** 1

**filename**

**type** string  
**minLength** 1

**key**

**type** string  
**minLength** 1

**size**

**type** integer

**version\_id**

**type** string  
**minLength** 1

**reference**

A reference to a cited document

This is structurally very similar to a Literature record, with unnecessary fields removed. For more information about the meaning of the different fields, see the corresponding field in *hep*.

**type** object

**Properties:** *arxiv\_eprint, authors, book\_series, collaborations, document\_type, dois, external\_system\_identifiers, imprint, isbn, label, misc, persistent\_identifiers, publication\_info, report\_numbers, texkey, title, urls*

**arxiv\_eprint**

arXiv eprint number

**MARC** 999C5r with arXiv: prefix

**example** 1703.07274

**type** string

**pattern** `^\d{4}.\d{4,5}|[\w.]+(-[\w.]+)?/\d+$`

**minLength** 1



## authors

List of authors

**MARC** 999C5h contains a comma-separated string with all authors, in `first_name last_name` order.

**type** array

**minItems** 1

Every element of **authors** is:

**type** object

**Properties:** *full\_name, inspire\_role*

## full\_name

Should be of the same form as *full\_name*.

**example** Smith, John G.D.

**type** string

**minLength** 1

## inspire\_role

**MARC** 999C5e for editor

**type** string

**minLength** 1

**Allowed values:**

- author
- supervisor
- editor

## book\_series

**type** object

**Properties:** *title, volume*

## title

**type** string

**minLength** 1

## volume

**type** string

**minLength** 1

## collaborations

**type** array

**minItems** 1

Every element of **collaborations** is:

**MARC** 999C5c

**type** string

**minLength** 1

## document\_type

**MARC** conceptually similar to 999C5d, but allowed values differ.

**Reference** *document\_type*

## dois

**type** array

**minItems** 1

Every element of **dois** is:

**MARC** 999C5a with prefix doi :

**type** string

**pattern** ^10\\.\\d+(\\.\\d+)?/\\.+\$

**minLength** 1

## external\_system\_identifiers

**MARC** 999C5u (as url)

List of identifiers of this document on external systems.

**type** array

**minItems** 1

Every element of **external\_system\_identifiers** is:

**type** object

**Required** *schema, value*

**Properties:** *schema, value*

**schema**

MARC 999C5u (as url)

**type** string

**pattern** `^\w+$`

**minLength** 1

**value**

External identifier

MARC 999C5u (as url)

**type** string

**pattern** `^\S+$`

**minLength** 1

**imprint**

**type** object

**Properties:** *date, place, publisher*

**date**

**type** string

**format** date

**minLength** 1

**place**

**type** string

**minLength** 1

**publisher**

MARC 999C5p

**type** string

**minLength** 1

## isbn

MARC 999C5i

ISBN if the cited document is a book.

---

**Note:** If the cited document is a book chapter, use *parent\_isbn* instead.

---

**type** string

**pattern** `^\d*[0-9X]$`

**minLength** 1

## label

MARC 999C5o

Label used to refer to the reference in the main text, and put in front of the reference in the reference list. Often, this is a number, found in square brackets. These brackets (or parentheses), if present, should not be put here.

**example** 42

**example** Jones1995c

**type** string

**minLength** 1

## misc

MARC 999C5m

Unrecognized information from the reference.

**type** array

**minItems** 1

Every element of **misc** is:

**type** string

**minLength** 1

## persistent\_identifiers

MARC 999C5a (with prefix hdl: or urn:)

**type** array

**minItems** 1

Every element of **persistent\_identifiers** is:

**type** object

**Properties:** *schema*, *value*

**schema**

**type** string

**minLength** 1

**Allowed values:**

- HDL
- URN

**value**

**type** string

**minLength** 1

**publication\_info**

**type** object

**Properties:** *artid*, *cnum*, *journal\_issue*, *journal\_record*, *journal\_title*, *journal\_volume*, *material*, *page\_end*, *page\_start*, *parent\_isbn*, *parent\_report\_number*, *parent\_title*, *year*

**artid**

**MARC** last part of 999C5s (if not a page)

**type** string

**minLength** 1

**cnum**

**MARC** 999C5b

**Reference** *cnum*

**journal\_issue**

**MARC** 999C5n

**type** string

**minLength** 1

**journal\_record**

**Reference** *journal*

### journal\_title

MARC first part of 999C5s

**type** string

**minLength** 1

### journal\_volume

MARC second part of 999C5s

**type** string

**minLength** 1

### material

Reference *material*

### page\_end

MARC Last part of 999C5s (if page-range)

**type** string

**minLength** 1

### page\_start

MARC Last part of 999C5s (if not article ID)

**type** string

**minLength** 1

### parent\_isbn

MARC 999C5i, but not distinguished from the *isbn* of the cited document.

If the cited document is a book chapter, this is the ISBN of the book.

**type** string

**pattern** `^\d*[0-9X]$`

**minLength** 1

### parent\_report\_number

MARC 999C5r but not distinguished from the cited document *report\_numbers*.

**type** string

**minLength** 1

**parent\_title****MARC** 999C5q

Title of the parent document, e.g. title of the book if the cited document is a book chapter.

**type** string**minLength** 1**year****MARC** 999C5y**type** integer**minimum** 1000**maximum** 2050**report\_numbers****MARC** 999C5r

---

**Note:** If the cited document is only part of a report, use *parent\_report\_number* instead.

---

**type** array**minItems** 1Every element of **report\_numbers** is:**type** string**minLength** 1**texkey****MARC** 999C5kKey that was used for this citation in the document. This may be extracted if the document is a PDF generated from a LaTeX, and possibly coincides with the one of the *texkeys* assigned to the cited document by Inspire and made visible in the TeX export formats.**example** Maldacena:1997re**type** string**minLength** 1**title****Reference** *title***MARC** 999C5t

## urls

**type** array

**minItems** 1

Every element of **urls** is:

URL of related document

**MARC** 999C5u

**type** object

**Required** *value*

**Properties:** *description, value*

## description

**type** string

**minLength** 1

## value

**type** string

**minLength** 1

## related\_record

Decorated link to a related record

**MARC** 78002, 78502, 78708`` and ``510 (depending on type of record and *relation*)

**type** object

**Required** *record*

**Properties:** *curated\_relation, record, relation, relation\_freetext*

## curated\_relation

**type** boolean

## record

URL of related record

**MARC** 510\_\_0 (for Institutions and Experiments), 530\_\_0 (for Journals), 78002/78502/78708w (for Literature).

**Reference** *any*



**relation**

Relation of the linked record to this one

The possible values are:

**predecessor**

**MARC** 510\_\_w:a (for Institutions and Experiments), 530\_\_w:a (for Journals) or the field comes from 78002 (for Literature).

The related record is a predecessor of the current one, i.e. the current record supersedes the related one.

- For a Literature record, this is typically the case if the related record is a preliminary note and this record is the final version.
- For a Journal record, this happens when the journal has changed name: the related record has the previous name and this one the new name.
- For an Institution record, this means that the related record represents an institution that does no longer exist and has become the one represented by this record.
- For a Project record, this happens when the related record is a project that finished and was superseded by the one in this record (e.g. two successive runs of an experiment).

**successor**

**MARC** 510\_\_w:b (for Institutions and Experiments), 530\_\_w:b (for Journals) or the field comes from 78502 (for Literature).

The related record is a successor of the current one, i.e. the current record is superseded by the related one.

- For a Literature record, this is typically the case if the current record is a preliminary note and the related record is the final version.
- For a Journal record, this happens when the journal has changed name: the current record has the previous name and the related one the new name.
- For an Institution record, this means that the current record represents an institution that does no longer exist and has become the one represented by the related record.
- For a Project record, this happens when the current record is a project that finished and was superseded by the related record (e.g. two successive runs of an experiment).

**parent**

**MARC** 510\_\_w:t (for an Institution record)

The related record is the parent of the current one, i.e. the current record is a child of the related one.

- For an Institution record, this means that the related record represents a parent institution (e.g. a university if this record is a department).

**commented**

**MARC** 78708i:commented (for a Literature record)

The related record has been commented by the current one, i.e. the current record is a comment on the related one.

---

**Note:** This only makes sense for a Literature record.

---

If none of the relations apply, a free textual relation can be provided in *relation\_freetext*.

---

**Note:** Only one directional links are currently allowed for technical reasons, but this should change in the future when Inspire can handle atomic updates of the two directions. For example, 78502 is currently discarded.

---

**type** string

**minLength** 1

**Allowed values:**

- predecessor
- successor
- parent
- commented

### relation\_freetext

**MARC** 510\_\_i (for Institutions and Experiments), 530\_\_i (for Journals) or 78708i (for Literature).

If none of the standard relations in *relation* fit, a textual relation can alternatively be provided here.

**type** string

**minLength** 1

### source

Source of the information in this field. As several records can be merged, this information allows us to remember where every bit of metadata came from and make decisions based on it.

**MARC** Often not present.

**type** string

**minLength** 1

### sourced\_value

This is used to add, besides the *value*, the *source* where this value came from.

**type** object

**Required** *value*

**Properties:** *source*, *value*

### source

**Reference** *source*

**value**

**type** string

**minLength** 1

**title**

**type** object

**Required** *title*

**Properties:** *source, subtitle, title*

**source**

**type** string

**minLength** 1

**subtitle**

**type** string

**minLength** 1

**title**

**type** string

**minLength** 1

**url**

URL of related document

**MARC** 8564

**type** object

**Required** *value*

**Properties:** *description, value*

**description**

**MARC** 8654\_\_y

**type** string

**minLength** 1

## value

**MARC** 8564\_u  
**type** string  
**format** uri  
**minLength** 1

### 3.1.11 json\_reference

#### any

Reference to another record

**type** object  
**Required** *\$ref*

**Properties:** *\$ref*

#### \$ref

URL to the referenced resource

**type** string  
**format** uri  
**minLength** 1

#### author

Reference to author record

**type** object  
**Required** *\$ref*

**Properties:** *\$ref*

#### \$ref

authors record

**type** string  
**format** uri  
**pattern** `.*\/api\/authors\/\d+ $\$$`   
**minLength** 1

### citeable

Reference to citeable (literature or data) record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

### **\$ref**

hep/data record

**type** string

**format** uri

**pattern** `.*\/api\/(literature|data)\/\d+ $\$$`

**minLength** 1

### conference

Reference to conference record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

### **\$ref**

conferences record

**type** string

**format** uri

**pattern** `.*\/api\/conferences\/\d+ $\$$`

**minLength** 1

### experiment

Reference to experiment record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

## **\$ref**

experiments record

**type** string

**format** uri

**pattern** .\*/api/experiments/\d+\$

**minLength** 1

## **institution**

Reference to institution record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

## **\$ref**

institutions record

**type** string

**format** uri

**pattern** .\*/api/institutions/\d+\$

**minLength** 1

## **job**

Reference to job record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

## **\$ref**

jobs record

**type** string

**format** uri

**pattern** .\*/api/jobs/\d+\$

**minLength** 1

## journal

Reference to journal record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

## **\$ref**

journals record

**type** string

**format** uri

**pattern** `.*\/api\/journals\/\d+ $\$$`

**minLength** 1

## literature

Reference to literature record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

## **\$ref**

hep record

**type** string

**format** uri

**pattern** `.*\/api\/literature\/\d+ $\$$`

**minLength** 1

## seminar

Reference to seminar record

**type** object

**Required** *\$ref*

**Properties:** *\$ref*

## \$ref

seminars record

**type** string

**format** uri

**pattern** .\*/api/seminars/\d+\$

**minLength** 1



Notes on how to contribute, legal information and changes are here for the interested.

## 4.1 Contributing

Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given.

### 4.1.1 Types of Contributions

#### Report Bugs

Report bugs at <https://github.com/inspirehep/inspire-schemas/issues>.

If you are reporting a bug, please include:

- Your operating system name and version.
- Any details about your local setup that might be helpful in troubleshooting.
- Detailed steps to reproduce the bug.

#### Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with “bug” is open to whoever wants to implement it.

#### Implement Features

Look through the GitHub issues for features. Anything tagged with “feature” is open to whoever wants to implement it.

### Write Documentation

inspire-schemas could always use more documentation, whether as part of the official inspire-schemas docs, in docstrings, or even on the web in blog posts, articles, and such.

### Submit Feedback

The best way to send feedback is to file an issue at <https://github.com/inspirehep/inspire-schemas/issues>.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome :)

### 4.1.2 Get Started!

Ready to contribute? Here's how to set up *inspire-schemas* for local development.

1. Fork the *invenio* repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/inspire-schemas.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
$ mkvirtualenv inspire-schemas
$ cd inspire-schemas/
$ pip install -e .[all]
```

4. Create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

Now you can make your changes locally.

5. When you're done making changes, check that your changes pass tests:

```
$ ./run-tests.sh
```

The tests will provide you with test coverage and also check PEP8 (code style), PEP257 (documentation), flake8 as well as build the Sphinx documentation and run doctests.

6. Commit your changes and push your branch to GitHub:

```
$ git add .
$ git commit -s \
  -m "component: title" \
  -m "* Adds your new feature." \
  -m "* Fixes an existing issue." \
  -m "* Improves and existing feature." \
  -m "" \
  -m "Sem-Ver: new feature" \
```

(continues on next page)

(continued from previous page)

```
-m "Sem-Ver: breaks api"  
$ git push origin name-of-your-bugfix-or-feature
```

7. Submit a pull request through the GitHub website.

### 4.1.3 Pull Request Guidelines

Before you submit a pull request, check that it meets these guidelines:

1. The pull request should include tests and must not decrease test coverage.
2. If the pull request adds functionality, the docs should be updated. Put your new functionality into a function with a docstring.
3. The pull request should work for all the currently supported Python versions, as of writing that is 2.7, and 3.5. Check [https://travis-ci.org/inspirehep/inspire-schemas/pull\\_requests](https://travis-ci.org/inspirehep/inspire-schemas/pull_requests) and make sure that the tests pass for all supported Python versions.

## 4.2 License

Inspire-schemas is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

Inspire-schemas is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with Inspire-schemas; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307, USA.

In applying this license, CERN does not waive the privileges and immunities granted to it by virtue of its status as an Intergovernmental Organization or submit itself to any jurisdiction.



## CHAPTER 5

---

### Changelog

---

Here you can find the [full changelog](#) for this version



i

inspire\_schemas, 3  
inspire\_schemas.api, 5  
inspire\_schemas.utils, 5





- 
- A**  
author\_id\_normalize\_and\_schema() (in module *inspire\_schemas.utils*), 5
- B**  
build\_pubnote() (in module *inspire\_schemas.utils*), 6
- C**  
classify\_field() (in module *inspire\_schemas.utils*), 6  
convert\_new\_publication\_info\_to\_old() (in module *inspire\_schemas.utils*), 6  
convert\_old\_publication\_info\_to\_new() (in module *inspire\_schemas.utils*), 6  
country\_code\_to\_name() (in module *inspire\_schemas.utils*), 7  
country\_name\_to\_code() (in module *inspire\_schemas.utils*), 7
- F**  
filter\_empty\_parameters() (in module *inspire\_schemas.utils*), 7  
fix\_reference\_url() (in module *inspire\_schemas.utils*), 7  
fix\_url\_add\_http\_if\_missing() (in module *inspire\_schemas.utils*), 7  
fix\_url\_bars\_instead\_of\_slashes() (in module *inspire\_schemas.utils*), 7  
fix\_url\_replace\_tilde() (in module *inspire\_schemas.utils*), 7
- G**  
get\_license\_from\_url() (in module *inspire\_schemas.utils*), 7  
get\_paths() (in module *inspire\_schemas.utils*), 7  
get\_refs\_to\_schemas() (in module *inspire\_schemas.utils*), 7  
get\_schema\_path() (in module *inspire\_schemas.utils*), 7  
get\_validation\_errors() (in module *inspire\_schemas.utils*), 7
- I**  
inspire\_schemas (module), 3  
inspire\_schemas.api (module), 5  
inspire\_schemas.utils (module), 5  
is\_arxiv() (in module *inspire\_schemas.utils*), 8
- L**  
load\_schema() (in module *inspire\_schemas.utils*), 8  
LocalRefResolver (class in *inspire\_schemas.utils*), 5
- N**  
normalize\_arxiv() (in module *inspire\_schemas.utils*), 8  
normalize\_arxiv\_category() (in module *inspire\_schemas.utils*), 8  
normalize\_collaboration() (in module *inspire\_schemas.utils*), 9  
normalize\_collaboration\_name() (in module *inspire\_schemas.utils*), 9  
normalize\_isbn() (in module *inspire\_schemas.utils*), 9
- R**  
resolve\_remote() (in module *inspire\_schemas.utils.LocalRefResolver* method), 5
- S**  
sanitize\_html() (in module *inspire\_schemas.utils*), 9  
split\_page\_artid() (in module *inspire\_schemas.utils*), 9  
split\_pubnote() (in module *inspire\_schemas.utils*), 9
-

## V

`valid_arxiv_categories()` (*in module inspire\_schemas.utils*), 9

`validate()` (*in module inspire\_schemas.utils*), 9