@@ General

We need to understand whether we want to use spaces or “\_” to separate words in the keywords of key-value pairs. I changed it by using spaces since it is clearer for human readability.

# Research product

## local identifier

String (mandatory): Unique code identifying the research product in the SKG (if any, otherwise "stateless identifier").

Suggestion: use a URL as a string to make this resource dereferenceable on the Web.

## identifiers

List (recommended): A list of objects representing external identifiers for the entity. Each object is structured as follows.

* scheme String (mandatory): The scheme for the external identifier (e.g., doi, handle, url, pubmed, etc.).
* value String (mandatory): The external identifier.

Note: the current version of SKG-IF includes the following types of identifiers (to be specified as strings in the field “scheme”): doi, pmid, url …

## topics

List (optional): A list objects referring to Topics covered by the Research product. Each object in the list has the following properties:

* term String (mandatory): The identifier of a Topic relevant for the Research product.
* provenance List (recommended): A list of provenance information tracking the origin of the relation between a Topics and a Research product. Each topic provenance object has the following properties:
  + associated with String (mandatory): the local identifier of the agent responsible for the topic relation.
  + trust Float (mandatory): A numeric value associated to the trust given to the relation to a Topics. The float should be normalised in the range [0,1].

## contributions

List (optional): A list of objects that describe an Agent, its role, contribution, rank and declared affiliations to Organisations when working on a Research product. Each object is structured as follows:

* by String (mandatory): The identifier of an Agent contributing to the Research product.
* declared affiliations List (recommended): A list of Organisations that reflect the declared affiliations of an Agent for the Research product.
* rank Integer (recommended): The rank (i.e., order of appearance) of the Agent with a specific role (e.g. the order of an author in a list) of a Research product.
* role String (recommended): The role that an Agent had in the Research product, to choose among author, editor, and publisher.
* contribution List (recommended): The contributions that an Agent had in the Research product. Each element in the list is a String compliant with the CRediT taxonomy, i.e.:
  + conceptualization
  + data curation
  + formal analysis
  + funding acquisition
  + investigation
  + methodology
  + project administration
  + resources
  + software
  + supervision
  + validation
  + visualization
  + writing – original draft
  + writing – review & editing

## manifestations

List (optional): A list of objects representing multiple manifestations of the same Research product (e.g., a preprint, a postprint, etc.). Each manifestation object has the following structure:

* type String (mandatory): The type of the manifestation (e.g., preprint). The type has the following properties:
  + labels Object (mandatory): the labels describing the type (multiple for multilinguism). The object is a dictionary, the keys represent language codes following ISO 639-1; the special key none is reserved whenever the informtion about the language is not available or cannot be shared.
  + defined in String (mandatory): the URL of the schema of the manifestation type, e.g., a link to the vocabulary of allowed product types.
  + class String (suggested): The URL of the class identifying the entity (e.g., in an ontology) describing that type.
* dates Object (mandatory): Relevant dates for the manifestation. The object is a dictionary, the keys represent the type of date and the value is expressed as either a string or a list of string, where each string is compliant with the ISO 8601 datetime string. The possible dates that are specifiable are:
  + acceptance: The date of acceptance of an entity. Examples of entities to which a date accepted may be relevant are a thesis (accepted by a university examination board) or an article (accepted by a journal editor).
  + access: The date on which a particular digital item, such as a PDF or an HTML file, has been accessed by somebody.
  + collected: The date on which some item has been collected, for example the data gathered by means of questionnaires.
  + copyright: The date on which an entity has been copyrighted.
  + correction: The date on which something, for example a document, is corrected.
  + creation: The date on which an entity has been created.
  + decision: The date on which a particular endeavour, such as a grant application, has been or will be approved or rejected by somebody.
  + deposit: The date on which an entity has been deposited, for example in a library, repository, supplementary information archive, database or similar place of document or information storage.
  + distribution: The date on which something is distributed, for example the date on which a preprint of a document is e-mailed to colleagues and other academics by the author(s), or the date on which a printed announcement of forthcoming theatre events is mailed to those those on the theatre's mailing list.
  + embargo: The date before which an entity should not be published, or before which a press release should not be reported on. For open-access journal articles, the embargo date is the date before which availability of the open-access version of the article is restricted by the publisher, following subscription-access availability of the published work.
  + modified: The date on which an entity has been modified.
  + publication: The date of formal issuance of a resource (e.g. a publication or a patent).
  + received: The date on which some item is received, for example a document being received by a publisher.
  + request: The date on which an agent is requested to do something, for example a reviewer is requested to write a review of a paper submitted to a journal for publication, or an author is requested to supply a revised version of the paper in response to the reviews received.
  + retraction: The date on which something, for example a claim or a journal article, is retracted.
  + validity: Date of validity of a resource.
* identifiers List (recommended): A list of objects representing external identifiers for the manifestation. Each object is structured as follows.
  + scheme String (mandatory): The scheme for the external identifier (e.g., doi, handle, url, pubmed, etc.).
  + value String (mandatory): The external identifier.
* peer review Object (suggested): Whether the Research product has undergone a peer review. It must be specify only if information about peer reviewing exists, and has the following properties:
  + status String (mandatory): describe if the manifestation has been already reviewed (i.e. “peer reviewed”) or if it is currently under review (i.e. “under review”).
  + description String (suggested): describe the type of peer review that applies, to choose from “single-blind peer review”, “double-blind peer review”, “open peer review”.
* access rights Object (mandatory): The access right for the specific materialisation. It specifies the following properties:
  + status String (mandatory): describe if the manifestation is open access (“open”), closed access (“closed”), under embargo (“embargoed”), restricted access (“restricted”), or unavailable for some reason (“unavailable”).
  + description String (suggested): describe and qualify the specific status selected.
* licence String (recommended): The URL of the licence specific to the manifestation.
* version String (recommended): Version for a software or research data product.
* biblio Object (optional): An object containing bibliographic information about a manifestation. The object has the following properties:
  + issue String (optional): Issue number.
  + pages Object (optional): the pages where the manifestation in defined (within its venue). It includes the following information
    - first String (mandatory): The starting page.
    - last String (mandatory): The ending page.
  + volume String (optional): Volume number (for journals, books, conferences).
  + edition String (optional): The edition (for journals and books).
  + number String (optional): a number of the manifestation within the venue (e.g. chapter number)
  + in String (optional): A Venue identifier for the manifestation.
  + hosting data source String (optional): A Data source URL for the manifestation.`

## related products

Object (optional): A dictionary of objects representing related Research products, where the semantics of such relationships is specified as a key. It is structured as follows:

* cites List (optional): A list of Research products identifiers that are cited by the current Research product.
* is supplemented by List (optional): A list of Research products identifiers that are supplement of the current Research product.
* is documented by List (optional): A list of Research products identifiers that documents the current Research product.
* is new version of List (optional): A list of Research products identifiers that are prior versions of the current Research product.
* is part of List (optional): A list of Research products identifiers that contains the current Research product.

# agent

Agents

The Agent entity represents an individual (a person, an organisation, or another kind of entity being able to act) who is involved in the creation, publication, dissemination, etc. of Research products. An Agent can be an author, a reviewer, an editor, a publisher, a researcher, or any other stakeholder involved in the scholarly communication process.

This section describes the metadata fields for the Agent and all its subtypes (i.e. people and organisations).

## local identifier

String (mandatory): Unique code identifying the agent in the SKG (if any, otherwise "stateless identifier").

Suggestion: use a URL as a string to make this resource dereferenceable on the Web.

## identifiers

List (recommended): A list of objects representing external identifiers for the entity. Each object is structured as follows.

* scheme String (mandatory): The scheme for the external identifier (e.g., orcid, viaf, etc.).
* value String (mandatory): The external identifier.

Note: the current version of SKG-IF includes the following types of identifiers (to be specified as strings in the field “scheme”): orcid, viaf, …

## name

String (optional): The string containing whatever concatenation of a Agent’s name(s).

## entity type

String (mandatory): Field stating what kind of entity is being serialised. It can be “agent” (generic entity), “person”, or “organisation”.

@ person related attributes

## given name

String (optional): The given name of a Person.

## family name

String (optional): The family name of a Person.

## affiliations

List (optional): A list of all the affiliations of a Person (à la ORCID). Each element of the list is structured as follows:

* affiliation String (mandatory): The identifier of the Organisation a Person is affiliated with.
* Role String (mandatory): The role that a Person had in the context of the Organisation. Needed for parsing purposes; fixed to “affiliate”.
* period Object (recommended): the time period where the Person was affiliated to the Organisation. It includes the following information:
  + start String (mandatory): The start datetime of the affiliation with the Organisation. The string should be compliant with the ISO 8601 datetime string.
  + end String (optional): The end datetime (if any) of the affiliation with the Organisation. The string should be compliant with the ISO 8601 datetime string.

@ organisation related attributes

## short name

String (optional): The short name/acronym for the Organisation.

## other names

List (optional): A list of other names, maybe in different languages, identifiying the Organisation.

## website

String (optional): The website URL for the Organisation.

## country

String (optional): The country code of the Organisation expressed as ISO 3166-1 alpha-2.

## type

List (optional): The types of the Organisation. One or more from the following values:

* archive
* company
* education
* facility
* government
* healthcare
* nonprofit
* funder
* research
* unspecified

# venue

## title

String (optional): The title of the Venue.

## type

String (optional): The type of the Venue. The String follows the vocabulary below:

* journal
* conference
* book
* repository
* unknown

## publisher → to be removed (now mapped via contributions)

## access rights

Object (optional): The access right for the specific journal. It specifies the following properties:

* status String (mandatory): describe if the journal is open access (“open”) or closed access (“closed”).
* description String (suggested): describe and qualify the specific status selected.

## creation date

String (optional): The date of creation of the Venue expressed as a ISO 8601 datetime string.

## contributions

List (optional) : A list of all the Agents that contributed to the Venue. Each element of the list is structured as follows:

* by String (mandatory): The id of an Agent.
* role List (mandatory): The roles of the Agent contributing to the Venue, to choose from the following list:
  + publisher
  + editor

# data sources

## policy

List (optional): A list of policies, documented in specific documents, that enable specific interactions / actions with the data source. Each item in the list is an Object with the following information:

* about String (mandatory): the type of policy to consider, to choose among the following possibilities:
  + submission: This policy provides a comprehensive framework for the contribution of research products. Criteria for submitting content to the repository as well as product preparation guidelines can be stated. Concepts for quality assurance may be provided.
  + preservation: This policy provides a comprehensive framework for the long-term preservation of the research products. Principles aims and responsibilities must be clarified. An important aspect is the description of preservation concepts to ensure the technical and conceptual utility of the content.
  + embargoed access (from COAR Access Rights 1.0): This policy provides the possibility of having / handling embargoed access refers to the resources in the data source.
  + metadata only access (from COAR Access Rights 1.0): This policy provides the possibility of having / handling metadata only access refers to the resources in the data source.
  + open access (from COAR Access Rights 1.0): This policy provides the possibility of having / handling data source resources that are immediately and permanently online.
  + restricted access (from COAR Access Rights 1.0): This policy provides the possibility of having / handling restricted access to the resources in the data source.
* target List (suggested): the types of resources to which the policy applies to. Each item in the list should be compliant with the following terms:
  + metadata
  + research data (from EOSC vocabulary [Research Product Type](https://wiki.eoscfuture.eu/display/PUBLIC/D.+v4.00+EOSC+Data+Source+Profile#D.v4.00EOSCDataSourceProfile-ResearchProductType))
  + literature (from EOSC vocabulary [Research Product Type](https://wiki.eoscfuture.eu/display/PUBLIC/D.+v4.00+EOSC+Data+Source+Profile#D.v4.00EOSCDataSourceProfile-ResearchProductType))
  + software (from EOSC vocabulary [Research Product Type](https://wiki.eoscfuture.eu/display/PUBLIC/D.+v4.00+EOSC+Data+Source+Profile#D.v4.00EOSCDataSourceProfile-ResearchProductType))
  + any
* documented at String (suggested): the URL of the document that describes the policy.
* description String (suggested): describe the type of policy, if necessary.

## persistent\_identity\_systems

List (optional): The persistent identifier systems that are used by the Data source to identify the ProductType it supports.

* product\_type String (mandatory): The Product type to which the persistent identifier is referring to, to choose among the following ones:
  + literature (from EOSC vocabulary Research Product Type)
  + research data (from EOSC vocabulary Research Product Type)
  + software (from EOSC vocabulary Research Product Type)
  + metadata
  + any
* pid\_schemes List (mandatory): the list of persistent identifier schemes used to refer to ProductTypes. Each elements must be drawn by the EOSC vocabulary Persistent Identity Scheme, i.e.:
  + doi
  + handle
  + ror
  + orcid
  + isni
  + arxiv
  + pmcid
  + ark

## audience

List (optional): The property defines the target audiences of the users of the data source. Each item specifies the following information:

* audience type String (mandatory): the type of the audience based on the vocabulary Jurisdiction, that may have the following values:
  + Global: intended for all users
  + National: intended for users of a country, e.g. national data repository
  + Regional: intended for users of a region (e.g. Europe)
  + Institution: intended for users of an institution, e.g. university publication repository
  + Research Infrastructure: intended for users of RIs and Clusters communities
  + e-Infrastructure: intended for users of horizontal e-infra services (e.g. EGI, GEANT, OpenAIRE, EUDAT, Fenix, gaia-x)

## data\_source\_classification

String (optional): The specific type of the Data source, based on the vocabulary Data Source Classification, which can be chosen among the following values:

* repository: services for the deposition, preservation, discovery, and access of research products metadata and files, e.g. PANGAEA, Zenodo, B2SHARE, EGI AppDB, BioTools
* aggregator: services for the aggregation of metadata about research products (and other entities) that are mainly collected (aka harvested) from data sources via APIs and then possibly curated/enriched by end-users, e.g. BASE, NARCIS, GESIS, B2FIND, OpenAIRE Research Graph
* scientific database: Scientific Databases intended to store structured information about scientific entities, e.g. PROTDB, ENA, ECRIN Studies Database
* journal archive: Data sources for the deposition (submission), preservation, discovery, and access of scientific journal articles (metadata and files), e.g. publishers sites, journal sites, etc.
* publisher archive: Scientific literature publisher, e.g. Elsevier, Copernicus, Frontiers, Springer, providing metadata and files of scientific publications relative to journals and conferences
* cris system: A current research information system (CRIS) is an information system to store, manage and exchange contextual metadata for the research products funded by a research funder or conducted at a research-performing organisation

## disciplines

List (optional): The disciplines for which the data source is dedicated. The disciplines must be specified using the Library of Congress Classification codes, available at <https://id.loc.gov/authorities/classification> (e.g. “PA3000-PA3049” for classical literature). In case the data source is discipline agnostic, the string “all” should be specified.

# grant

## funding agency

String (optional): The local identifier of an Organisation funding the Grant.

## duration

Object (optional): the duration of the Grant. It includes the following information:

* start String (mandatory): The start datetime of the Grant. The string should be compliant with the ISO 8601 datetime string.
* end String (optional): The end datetime (if any) of the Grant. The string should be compliant with the ISO 8601 datetime string.

## contributions

List (optional): A list of objects, each describing an Agent, its contribution, and declared affiliations to Organisations in the context of the Grant. Each object is structured as follows:

* by String (mandatory): The identifier of an Agent contributing to the Grant.
* declared affiliations List (recommended): A list of Organisations that reflect the declared affiliations of an Agent for the Grant.
* role List (recommended): The roles that an Agent had in the Grant. Each element in the list is a String compliant with the project roles in SCoRO (<https://sparontologies.github.io/scoro/current/scoro.html#http://purl.org/spar/scoro/ProjectRole>), i.e.:
  + co-applicant
  + lead applicant
  + project leader
  + project manager
  + project member
  + workpackage leader

# topic

## labels

Object (optional): the labels describing the topic (multiple for multilinguism). The object is a dictionary, the keys represent language codes following ISO 639-1; the special key none is reserved whenever the informtion about the language is not available or cannot be shared.