



Proposed schema changes -
have your say



@CrossrefOrg

Open schema development means

- opportunity for feedback
- no surprises (we hope)
- community investment

limited by:

- resources
- changing landscape
- conflicting needs of members

Priorities for upcoming update

- Expand identifier support
- Identifying and crediting contributors
- Improve data citation support
- Provenance



Don't panic!

Backwards compatibility - do we really want to mess with that?

- it's the best way to support changes
- it's not a big change
- supporting the old and new ways will be messy and confusing and not enforceable through the schema

BUT we try to stay backwards compatible for a reason, if we can't support current schema and new, we won't do it (for now...)

Expanding identifier support

Why: [identifiers are useful](#)

the plan:

- start with identifiers for people and organizations
- accept all identifiers registered with [identifiers.org](#)
- continue to develop services for open community led identifiers like ROR and ORCID

Contributors

We need to:

- extend identifier support for authors and institutions
- support CRediT
- make names less Western-centric
- refine affiliation support (ROR 🦁)
- expand support for organizational authors

Current contributor XML

```
<contributors>
  <person_name sequence="first"
contributor_role="chair">
  <given_name>Minerva</given_name>
  <surname>Housecat</surname>
  <affiliation>Crossref University</affiliation>
  <ORCID authenticated="true">
https://orcid.org/0000-0002-4011-3590</ORCID>
  </person_name>
  <organization sequence="first"
contributor_role="author">Crossref</organization>
</contributors>
```

person

@role (single)

@sequence (first or additional)

> given name and surname

> affiliation

> ORCID iD

organization

@sequence

@role

Current contributor XML

```
<contributors>  
  <person_name sequence="first" contributor_role="chair"> ← a single role  
    <given_name>Minerva</given_name>  
    <surname>Housecat</surname>  
    <affiliation>Crossref University</affiliation> ← a single string  
    <ORCID authenticated="true"> https://orcid.org/0000-0002-4011-3590</ORCID>  
  </person_name>  
  <organization sequence="first"  
    contributor_role="author">Crossref</organization> ← confusing?  
</contributors>
```


Person names

- **given_name** - no change
- **surname** - change to **family_name**
- **ORCID** - unchanged for now as a few services use this, may fold into identifier tags in the future
- **name-style** - replace with `alternate_name`
- add **alternate_name** and include `name-style` and `xml:lang` attribute
- require **given_name** *or* **surname** instead of just **surname** as not all names have surnames
- add [identifier support](#)

Alternate names

```
<alternate_name name_style="western" xml:lang="en">Minnie H</alternate_name>  
<alternate_name name_style="eastern" xml:lang="jp">ミネルバハウスキャット  
</alternate_name>
```

allows nicknames, aliases, and names in different character sets

Identifiers for names

```
<contrib_id contrib_id_type="isni">0000000121032683</contrib_id>  
<contrib_id contrib_id_type="orcid"  
authenticated="true">https://orcid.org/0000-0002-4011-3590</contrib_id>
```

allows multiple name identifiers and identifier type

Organization / corporate authors

```
<organization sequence="first" contributor_role="author">Crossref</organization>
```

- **Limited to just a name**
- **'organization' gets confused with 'affiliation'**

What will it look like?

```
<collab sequence="additional">
  <collab_name>Crossref</collab_name>
  <role role-type="data_curation"/>
  <contrib_id contrib_id_type="ror">https://ror.org/02twcfp32</contrib_id>
  <contrib_id contrib_id_type="isni">0000000405062673</contrib_id>
  <contrib_id contrib_id_type="wikidata">Q5188229</contrib_id>
</collab>
</contributors>
```

- **Change 'organization' to 'collab' (per JATS)***
- **add role support**
- **add contributor ID support (institutional identifiers)**

CRediT

What is it?

[CRediT](#) - Contributor Roles taxonomy; used to represent the roles typically played by contributors to scientific scholarly output.

Why is Crossref adopting it?

Current contributor types are very limited and CRediT is becoming widely used; many members are asking for it

Current roles:

- author
- editor
- chair
- translator

Roles for peer review:

- reviewer
- review-assistant
- stats-reviewer
- reviewer-external
- reader

CRedit roles

- Conceptualization
- Data curation
- Formal analysis
- Funding acquisition
- Investigation
- Methodology
- Project Administration
- Resources
- Software
- Supervision
- Validation
- Visualization
- Writing - original draft
- Writing - reviewing & editing

Changes:

- retire **@contributor_role** and replace with new repeatable **<role>** element - this allows multiple roles to be included
- add **@role-type** to capture role type.
- add **@vocab** to capture vocabulary

CRedit breaks backwards compatibility - we need to support multiple roles to accurately reflect contributions of authors

What will it look like?

```
<role role-type="conceptualization" vocab="credit"/>  
<role role-type="author" vocab="crossref"/>
```

or

```
<role role-type="data_curation" vocab="credit"/>Data Curation</role>  
<role role-type="author" vocab="credit"/>Author</role>  
<role role-type="other" vocab="credit"/>Sentence Diagrammer</role>
```

Affiliations

Current: `<affiliation>Crossref University</affiliation>`

Expand to include:

- **institution** - to capture info of the affiliated institution.
- **institution_name**
 - **institution_acronym**
 - **institution_place**
 - add **@country** attribute to **institution_place** to allow sorting of affiliations by country
 - **institution_department**
- add support for **institution_id** to support ROR and other identifiers

What will it look like?

```
<affiliations>
  <institution>
    <institution_name>Crossref</institution_name>
    <institution_acronym>CR</institution_acronym>
    <institution_place country="us">Lynnfield, MA</institution_place>
    <institution_department>Feline Outreach</institution_department>
    <institution_id institution_id_type="ror">https://ror.org/02twcfp32</institution_id>
    <institution_id institution_id_type="isni">0000000405062673</institution_id>
    <institution_id institution_id_type="wikidata">Q5188229</institution_id>
  </institution>
  <institution>
    <institution_name>Institute of Sparse Metadata</institution_name>
</affiliations>
```

Data citations

```
<citation key="ref=1">
```

```
<unstructured_citation>Morinha F, Dávila JA, Estela B, Cabral JA, Frías Ó,  
González JL, Travassos P, Carvalho D, Milá B, Blanco G (2017) Data from: Extreme  
genetic structure in a social bird species despite high dispersal capacity.  
Dryad Digital Repository.
```

```
http://dx.doi.org/10.5061/dryad.684v0</unstructured_citation\>  
</citation>
```

```
<citation key="ref2">
```

```
<doi>10.5061/dryad.684v0</doi>  
<cYear>2017</cYear>  
<author>Morinha F</author>  
</citation>
```

- not a lot there
- can't tell it's a data citation without the DOI

Support types for citations

- journal
- journal_article
- book
- book_chapter
- conference_proceeding
- conference_paper
- standard
- dataset
- software
- report
- dissertation
- website
- preprint
- other

Also:

- add support for other identifiers
- add support for generic 'item_title'

What will it look like?

```
<citation cited_item_type="data" key="ref4">  
  <author>Morinha F</author>  
  <cYear>2017</cYear>  
  <item_title>Extreme genetic structure in a social bird species despite  
high dispersal capacity</data_title>  
  <institution>Dryad Digital Repository</institution>  
  <identifier  
type="uri">http://www.example.org/boohooidonthaveadoi</identifier>  
<doi>10.3201/nowihaveadoi</doi>  
</citation>
```

What will it look like?

```
<citation cited_item_type="data" key="ref4">  
  <unstructured_citation>Morinha F, Dávila JA, Estela B, Cabral JA, Frías  
  Ó, González JL, Travassos P, Carvalho D, Milá B, Blanco G (2017) Data from:  
  Extreme genetic structure in a social bird species despite high dispersal  
  capacity. Dryad Digital Repository.</unstructured_citation>  
</citation>
```

Provenance

Current: nothing for journal articles

For books, reports, standards, conference proceedings:

```
<publisher>  
  <publisher_name>Pumpkin Spice League of Hatred</publisher_name>  
  <publisher_place>Burlington, VT</publisher_place>  
</publisher>
```


What will it look like?

```
<publisher>  
  <publisher_name xml:lang="en">Pumpkin Spice League of Hatred</publisher_name>  
  <publisher_name xml:lang="jp">パンプキンスパイリーグオブ憎しみ </publisher_name>  
  <publisher_place country="us">Burlington, VT</publisher_place>  
</publisher>
```

Expand support for dates

- **citation date:** to capture member's preferred citation date (this can vary greatly between print and online across publishers)
- **submitted date:** per member and internal requests
- **copyright date:** to capture the specific copyright date as it may differ from a publication date

Other updates

- add e-location_id as an element
- add citations to peer reviews
- suggestions?

The future

new content types

versions

new relations

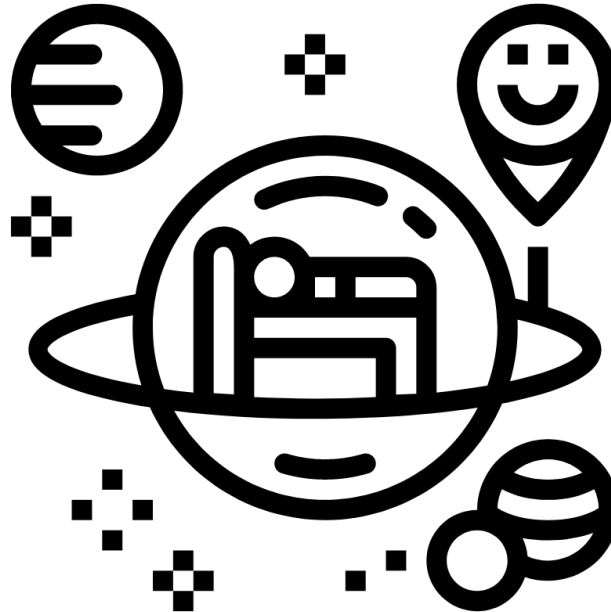
more funding
metadata?

article type taxonomy

Conference IDs

open access indicators

identifiers for ?



Next steps

- Feedback
- Open development
 - summarize feedback
 - share a schema and examples
 - share output plans (JSON, filters and facets for REST API)
 - share timelines for implementation
 - test
 - implement

Have feedback on this proposal?

Leave comments in the [proposal](#)

Send me an email (pfeeney@crossref.org)

Create a [GitLab](#) issue

★★ I'll be collecting feedback through January 15 ★★

Have your say!

Metadata Practitioners Interest Group

- will have monthly calls in 2020
- focus on sending Crossref metadata (for now)
- will be forming working groups for specific interests
 - JATS, preprints, new content types

Thank you!

Patricia Feeney

pfeeney@crossref.org

@schemaschemer