Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 106 (2017) 268 – 275

Procedia Computer Science

13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

On representing affiliations in the CERIF model

Alejandro Engelmann*, Christer Enkvist, Carl-Johan Syrén

SLU University Library, Swedish University of Agricultural Sciences, Sweden

Abstract

In most publications each author has one or many affiliations. This affiliation, or affiliations, is the one relevant for the publication at hand. In CERIF authors can be related to publications, and authors to affiliations thereby creating a relation between a publication and affiliation(s). However, if an author has multiple affiliations, CERIF cannot specify which of these that is relevant for a specific publication.

In the proprietary CRIS system Converis the affiliation problem is solved by the using entities called "business cards". The "business card" is an entity which connects to a person and an organisation unit and this entity can thereafter be connected to the publication providing a correct person-organisation-publication relationship.

We propose to model this entity as a properly classified cfPerson in the CERIF model. This entity can be referred to as business card, affiliated person, personality or avatar. The connection between the "real person" and the "affiliated person" is a cfPers_Pers entry representing the "has business card" relation. This model is used by SLU to transfer Converis data to a CERIF model without losing any information about the connection between publications, authors and organisations.

Result entities are related to the "business cards" in the standard CERIF way. Result entities are related to the organisation units through this "affiliated persons", but for backwards compatibility a direct relation can also be added to the model.

The use of "business cards" rather than "real persons" is in most cases more flexible as well as granular since it not only enables tracking of name changes and pseudonyms but can also be used when describing project members or even collaborations. However, when a reference should be associated with the unique real person, e.g. ORCID, then the "real" cfPerson should be used and not a "business card".

Peer-review under responsibility of the Organizing Committee of CRIS2016

Keywords: CERIF, Converis, Author affiliations, Business cards, Avatars

* Corresponding author. E-mail address: alejandro.engelmann@slu.se

Peer-review under responsibility of the Organizing Committee of CRIS2016
doi:10.1016/j.procs.2017.03.025

1. Introduction

When an author's name is given in a publication, it is almost always followed by an affiliation statement. An example is shown in Figure 1.

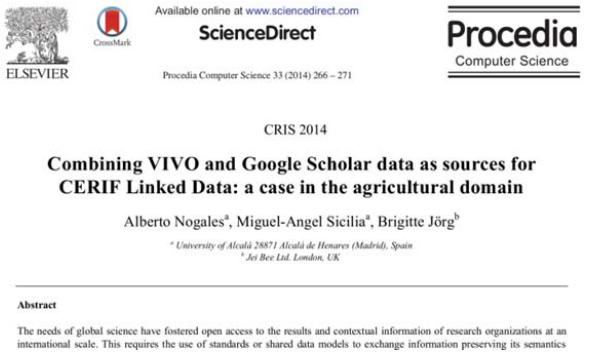

Figure 1. A typical header of a journal article showing how affiliations are reported.

Even the mention of creators and contributors to other works are very often linked to the organisation behind them. This can be seen as way of more precisely identifying the person, but also a recognition that the organisation behind the creator is in some way also responsible for the described work. When using a CRIS (Current Research Information System), it is important that this information is correctly recorded. This would allow to CRIS data to be used in bibliometric analysis as seen for example in the Leiden Manifesto[1].

In this paper we describe how the affiliation information is usually described in the CERIF[2,3] model mentioning its shortcomings, then we describe how the Converis[4] model does the same. In the following section we describe how the Converis concepts can be used in CERIF. Finally we show that the idea can be extended to conveniently record both intended and actual incomplete information about persons.

2. Author affiliations in CERIF

In the CERIF model many of the concepts mentioned are represented by first class entities. Author, creators or contributors are represented by cfPerson entities, organisations by cfOrganisationUnit entities, publications by cfResultPublication entities, and projects by cfProject entities, etc. In the model several other first and second class entities are described together with a semantic layer of classifiers and multi-language attributes. The relations between the base entities are represented by link entities classified by the semantic information and dated with start and end dates. Even a “fraction” representing the strength of the relation can be given for each link entity.

The affiliation of an author would be represented by a connection of the cfPerson record, that is, the author “personal” record, to a cfOrganisitation record representing the author's institution. To model the authorship relation of this person to the publication, represented by a cfResultPublication record, a cfPerson_ResultPublication link record is used. It is possible to relate the publication directly to the institutions (or other organisation units) using cfOrganisationUnit_ResultPublication relations. A class diagram is shown in Figure 2, and in shown in Figure 3 an object diagram exemplifying how the model can be used for the publication in Figure 1.

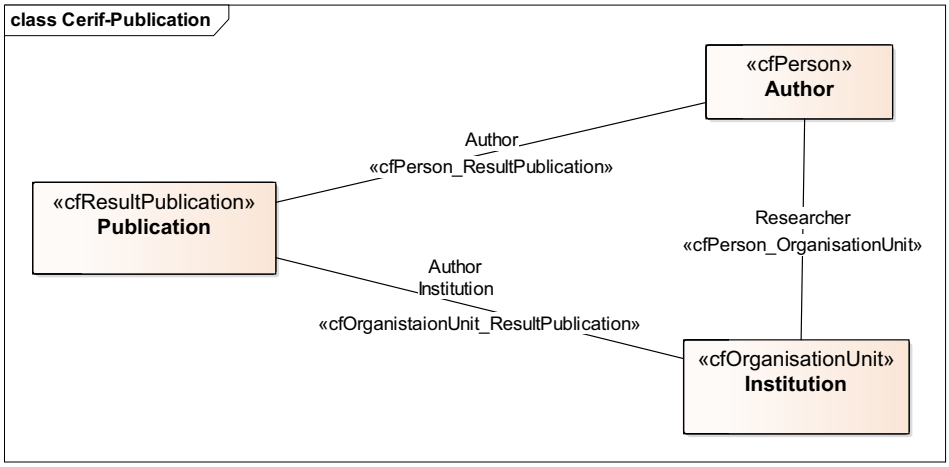

Figure 2. A class diagram showing the relations between Publications, Authors and their Institutions as they are described in standard CERIF.

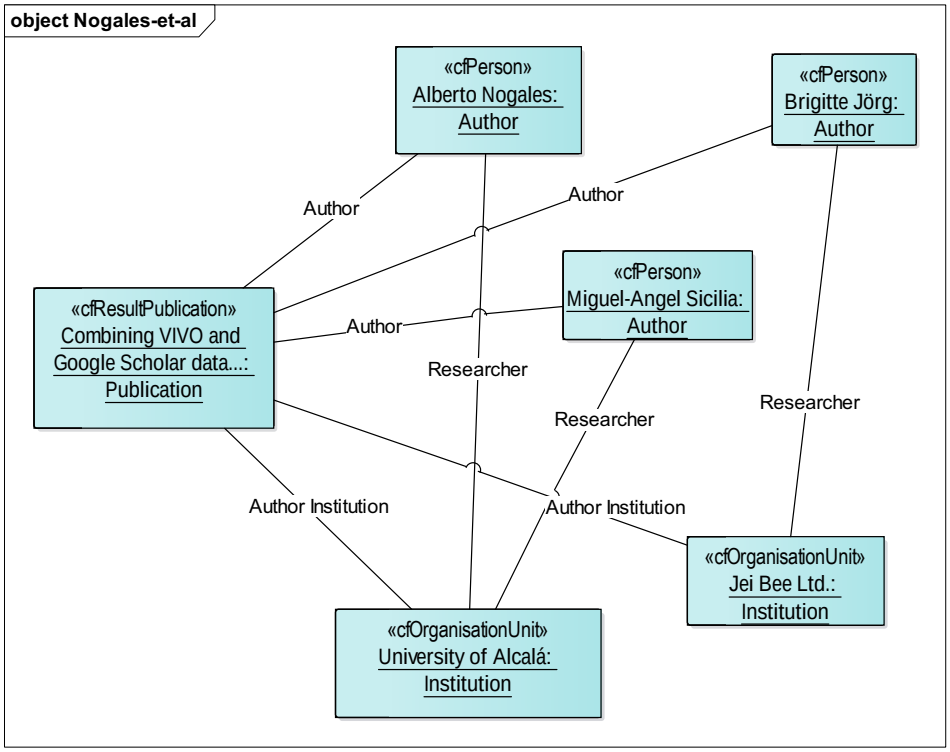

Figure 3. An object diagram showing the affiliation information in Figure 1 represented in standard CERIF.

A shortcoming in the CERIF model is that there is no possibility of specifying which attributes of the author are specifically relevant to the publication. The author could have used a special variant of their name, which indeed can be represented in the author's record, but it is impossible to indicate this in the model. In some cases the StartDate

and EndDate entities in the CERIF model can be used to choose the right name variant, but this is not always the case.

The same applies to the institution connection. An author can be connected to several institutions, sometimes at the same time, and this might make it impossible to pick the right institution to credit for a publication using the data contained in a CERIF model if the author has other publications attributed to the same institution as another author in the current publication.

3. Author affiliations in Converis

The proprietary system Converis model of author affiliation information is slightly different compared to CERIF. They have introduced a new entity called “Business Card”, which connects to both to the person entity and to the organisation entity. This entity is used to represent the authorship connecting it to the publication entity. A class diagram showing this is given in Figure 4.

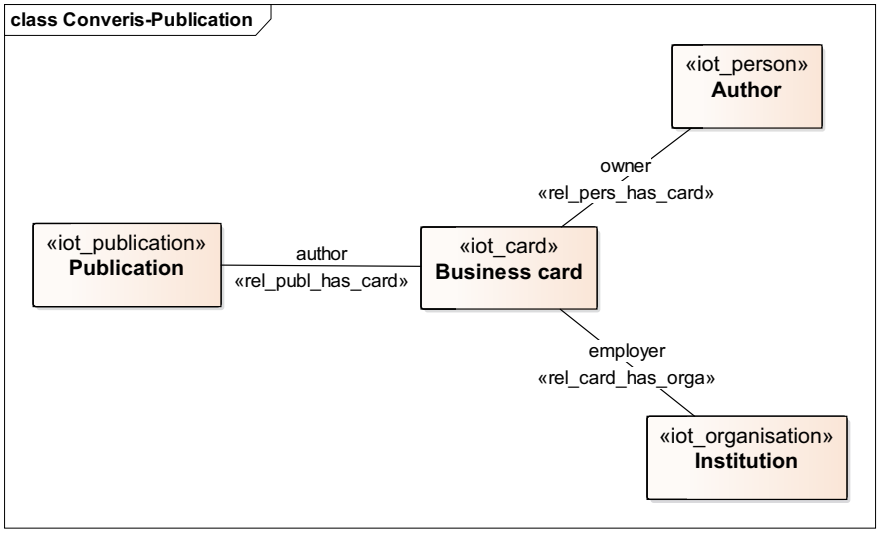

Figure 4. A class diagram depicting the relevant Converis entities and relations involved in the description of authorship and affiliations.

A person will get a Business Card for each unique name-organisation combination used in the person's publications. The same Card will be reused any time the person presents the same “credentials” when publishing. Any change for a person, or wish to change, to present them in another way would require a new Business Card.

4. Using the Converis affiliation concept in CERIF

The possibility of adding semantic information to CERIF entities enables us to modify the model to include the desired aspects of the Converis model.

In order to incorporate Business Cards in CERIF we have followed the same path as the one needed to model journals. In the Converis model, journal and articles are different entities but in the CERIF model, both are modeled with a cfResultPublication record semantically classified to model the difference.

We propose modeling the Business Card as a cfPerson record. A classification scheme is introduced to differentiate “Affiliated Persons”, the Business cards, from “Physical Persons”, the real persons. This is shown as a class diagram in Figure 5 and exemplified with previously used publication as an object diagram in Figure 6.

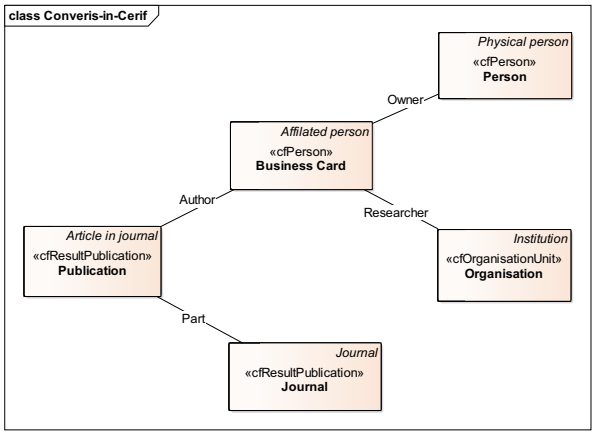

Figure 5. A class diagram showing the proposed extension of the standard way of modelling authorship and affiliations in CERIF.

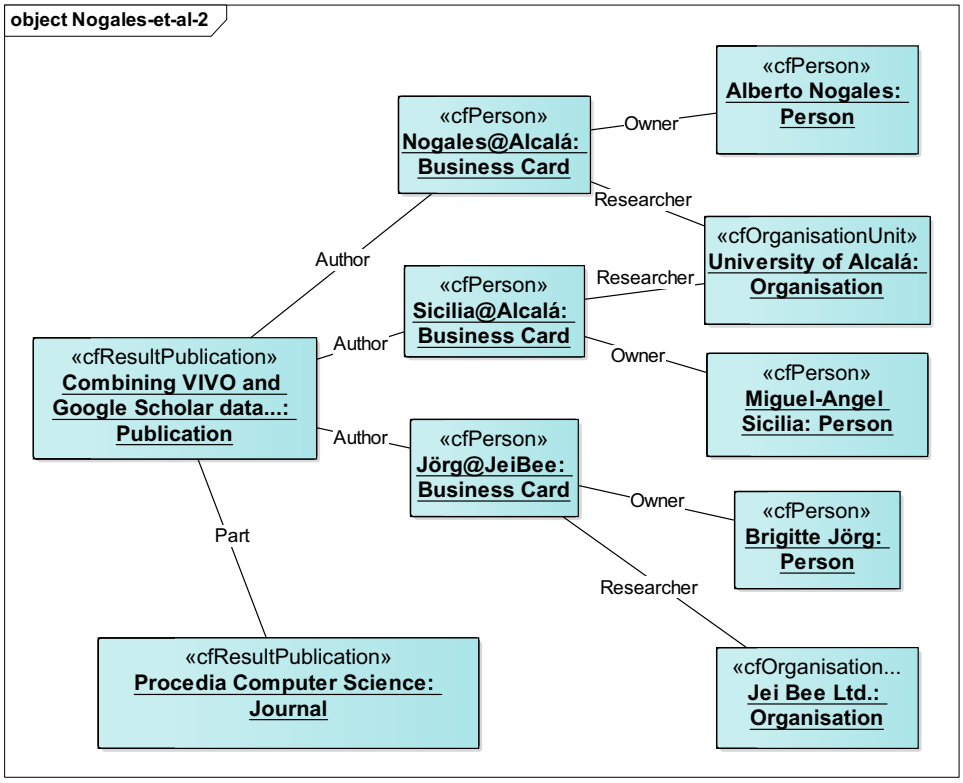

Figure 6. An object diagram modeling the publication in Figure 1 in the proposed schema.

The cfClassificationTerm Owner is well suited to describe the relation between a Business Card and a Person. To implement the Business Card into CERIF a new cfClassSchema is required for the identity type of the record and some classes for the actual types. An example can be found in Tables 1 and 2.

Table 1. The CERIF semantic entries for an Identity Type Classification Scheme.

CERIF Semantic Entry	Value
cfClassification SchemeId	9ebb4f55-97ab-4bac-91b1-e1f1b687de3d
cfClassification Scheme Name	Identity Type
cfClassification Scheme Description	This scheme contains CERIF vocabulary terms applicable in the cfPerson_Person link entity defining the person identity type.

Table 2. The CERIF semantic entries for the Source Classes of Business Cards and Physical Persons related to the Identity Type Classification Schema.

CERIF Semantic Entry	Values	
cfClassfication ID	444aa6fb-b2a8-4cde-b41a-69d0fe69cdef	6ce00776-8f63-45de-afc0-9ab9d18961fc
cfClassification Term	Physical person	Business Card
cfClassificationDefinition	A physical person	A persona or group of personas.
cfDefinitionSource	https://en.wikipedia.org/wiki/Legal_personality	https://en.wikipedia.org/wiki/Persona
cfClassificationSchemeName	Identity Type	Identity Type
usage with CERIF	cfPers_Pers	cfPers_Pers

This extension of the use of the CERIF model allows us to keep track of the actual name and affiliation used by an author in each publication.

We suggests that the “Presented Name” is used to attach a name to the Business Card. One could imagine that it is the name printed in the card. On the other hand the “Physical Person” record would have the official names (“Passport Name”) and may even have links to the alternative names also appearing in the owned Business Cards.

The links of the Business Cards to the Institution, which as all relation in CERIF are classified with semantic information, indicate the position of the person within the organisation while the link of the Business Cards to the publications indicate the role of the person in the production of the publication, as shown in Figures 5 and 6.

When exporting the information to CERIF compliant systems not supporting the Business Card concept, this entity can be bypassed, losing some information, but keeping compatibility with other uses of the CERIF model.

5. An extension of the Business Card concept

Quite frequently, when harvesting metadata from publications, the (physical) person(s) referred to in the publication is not known. In cases like this, persons would be modeled with a cfPerson record which is not classified as a physical person. An extension of the Business Card concept would be to introduce the Avatar concept in the model. That is a holder for the information of a person which can be incomplete or artificial for any reason.

If and when the physical person behind this “Avatar” is identified a relation between the Avatar and physical person can be made keeping the original information in the “Avatar” record.

I our case, at the SLU Library, we have found that the incompletely identified author information harvested from an Eprints system is reasonably represented as Avatars in CERIF. A later analysis can connect these Avatars to identified person records in the same way as Business Cards are connected, by a cfPerson_Person link. This is shown in a class diagram in Figure 7. A proposal presented elsewhere5 allows us to track this kind of additions to the modeled data.

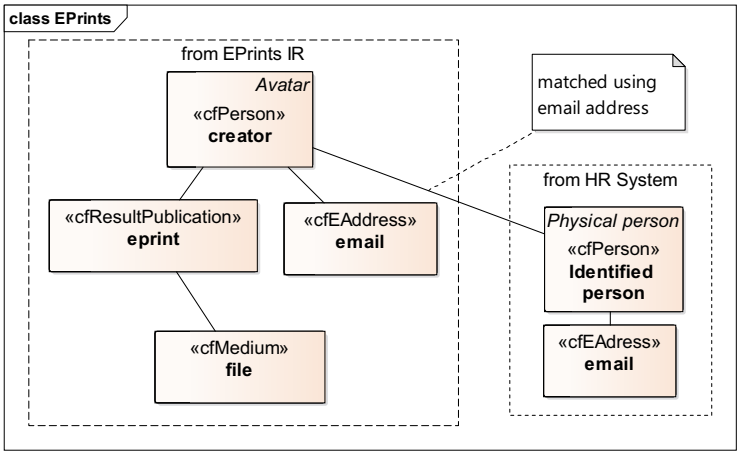

Figure 7. A class diagram showing how the incomplete information retrieved from an Eprints based IR (dashed boundary) could be improved using information from other sources, in this case the HR system (doted boundary), stored in the CERIF model.

Another generalization is the use of non-physical persons representing a group of person or even "affiliated" persons. This "group authors" appear sometimes in the list of authors just in the same way as other affiliated persons that we are modeling with business cards. Modeling group authors as an "Avatar", which is a cfPerson, instead of modeling groups as cfOrganisationUnits implies that it is not necessary to allow for other entities than cfPersons to link with "author" relations to publications.

The use of pseudonyms is of course modeled in the same way, allowing to state in the model that a certain publication was done using a certain name. In this case the pseudonym, or other variants of the "Passport Name", can be attached to the "Physical Person" as "Presented Name". However, in order to correctly describe the usage of the name in a certain publication the "Avatar" record is necessary.

An overview of possible cases for the use of "Business Cards" or "Avatars" is shown in Figure 8.

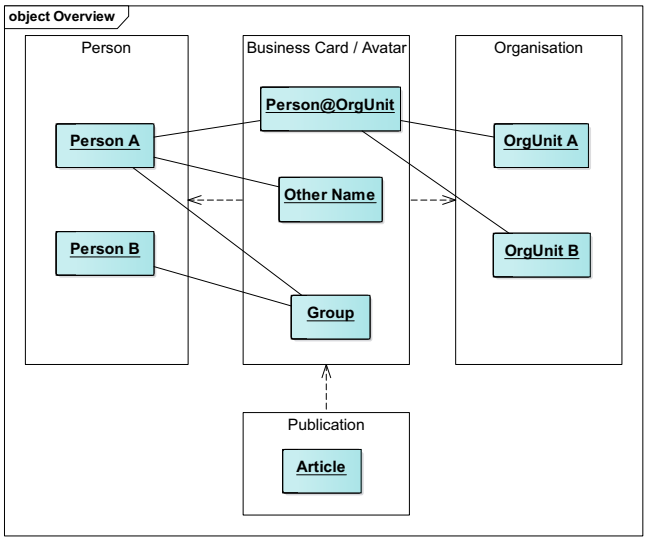

Figure 8. An Object diagram showing some cases where the use of "Business Cards" and "Avatars" are advantageous.

6. Identifiers for non physical persons

There are many identifiers that are used to single out authors and creators. One of the newest is ORCID[6,7] which is supposed to be an identifier of a physical person and probably should be attached to the "real" cfPerson.

Another identifier is ISNI[8] which is used by publisher that is supposed to identify the person. In many cases the same person gets different identifiers when their names are reported by different publishers. In those cases the identifiers could be attached to a "Business Card" representing the person in the publications by this publisher.

Even ResearcherID[9], the Thomson-Reuters author identifier can be issued several times to the same person and a "deduplication" might be difficult which suggests that this entity should be attached to the Business Card rather than the Physical Person.

7. Conclusions

We propose the addition of a new category of concepts to be modeled as cfPerson records. Adhering to the CONVERIS concept of Business Cards we propose to add it as a suitably classified cfPerson. This can be achieved by adding a new classification scheme called "Identity Types" to the semantic library of CERIF; containing classes as "Physical Person", used to tag real persons, and "Business Card" and "Avatar" to tag a dependent level of cfPersons. These secondary identities are attached to the real person by a cfPerson_Person relation classed as "Owner".

References

1 Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Nature **520**, 429–431 (23 April 2015) doi:10.1038/520429a

2. Jeffrey K, Houssos N, Jörg B, Asserson A Int. J. Metadata, Semantics and Ontologies, Vol. **9**, No. 1, 2014

3. Jörg B, Kerridge S, Cranner P Trowel S Ginty K Jeffery KG, Houssos N, Brasse V, Höllrigl T, Vestam T, Baker D, Foster S Dvořák J, Asserson A, Rasmussen H, Zendulkova D, Price A, Sicilia MA, Ruiz-Rube I, van Grootel G, Evans K, Strijbosch L, Cox M, Simons EJ. CERIF 15 Semantics: Research Vocabulary euroCRIS, 2013

4. CONVERIS, accessed 20 May 1016, < http://Converis.thomsonreuters.com/>

5 Alejandro Engelmann, Christer Enkvist, Carl-Johan Syrén. "Keeping track of provenance of CRIS entities and relations within the CERIF model".13th International Conference on Current Research Information Systems, CRIS2016, 9-11 June 2016, Scotland, UK

6 ORCID, accessed 20 May 1016, < http://orcid.org/>

7 Bohannon, J. Science, Vol. **351**(6270), p.213, 2016

8 International Standard Name Identifier, accessed 20 May 1016, <http://www.isni.org/>

9 Researcher ID, accessed 20 May 1016, <http://www.researcherid.com/>