Considerations about Uniqueness and Unalterability for the Encoding of Biographical Data in Ontologies

Thierry Declerck[1] and Rachele Sprugnoli[2]

[1]DFKI GmbH, Multilingual Technologies Lab
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
[2]Fondazione Bruno Kessler, Digital Humanities Group
Via Sommarive, 18, I-38123 Povo, Italy
declerck@dfki.de, sprugnoli@fbk.eu

Abstract

This paper results from observations that have been made while studying ontological and linked data-based approaches to the encoding of biographical data. Based on certain issues we discovered and which will be described here, we aim to call for a collaborative work towards guidelines for modelling biographical data in the standard Semantic Web representation languages. The need for guidelines became even more clear after reading an article, which described various types of errors in biographical data encoding that have been generated due to an unsuitable use of the `owl:sameAs` property when referring to the linked data-based description of the life of two literary authors. In this context, there is also a need to agree on the core element of which a biographical description constitutes. More specifically, we aim to determine the "biographical unit", which should be primarily modelled and to which all related information should be linked by using corresponding semantic properties. Apart from that, we will also discuss the need of the definition and use of synchronic versus diachronic properties associated with the modelled biographical unit. Regarding this point, we come to the conclusion that for the description of a biographical unit, there are probably no properties whose values remain unaltered over time. This is particularly true if the provenance information, that can provide contrasting values which, however, might be correct from different point of views, is taken into account.

Keywords: Ontologies, biographical units, linked data

1 Introduction

The issue of encoding changes occurring in a life has received some attention in the context of the formal representation of biographical data. In one study on this topic, Krieger and Declerck (2015) present considerations on synchronicity and diachronicity and how those aspects can be applied for defining properties in a formal ontology about biographical data. Building further on this study, one can eventually come to the conclusion that it is very difficult, if not impossible, to come up with unalterable property values that can be associated with an individual within a biographical description. This leads to the question whether there are any properties of human beings which are in fact immutable and which can therefore be used as the fixed pillar on whose base we can describe all other changeable aspects and characteristics of human beings.

At the actual state of our study, this seems not to be given. Let us take as an example the case of the soldier Manning, introduced in Wikipedia as "Chelsea Elizabeth Manning (born Bradley Edward Manning, December 17, 1987)".[1] In this case, the question arises whether this person remains the same after the change of sex, while we want to stick to the assumption that only one entry for this biographical unit should be kept.

We can also be confronted with uncertainty when looking at the birth date of a person, as this is an information that can still be modified or corrected in dependency of new data, also depending on the sources consulted. In addition to that, it is sometimes not even possible to state which source is the more reliable one, so that we have to encode biographical information mentioning its provenance, especially in the cases where we do not have a unique value. In the end, we only have the certainty that, biologically speaking, a person was born only once, but that various birth dates can be associated with this event, in dependency of perspectives and provenance of the information.[2]

Our intuition is that a very carefully designed ontology can offer support when dealing with a "biographical unit". This biographical unit might have no fixed characteristics, or properties, but on the basis of the large set of possibly divergent values of descriptors (classes and properties) and their organisation in one ontological space, it can be considered as one unique carrier of a life. This carrier should then be uniquely identified by a URI.[3]

In this paper, we concentrate thus on biographical data as giving an account of a person's life and achievements, not considering at this stage prosopography or what is sometimes referred to as "collective biography" (Davies and Gannon, 2006).

In the next sections we will first report on existing ontological modellisation initiatives for biographical data, before briefly describing the Linked Open Data cloud and

[1] `https://en.wikipedia.org/wiki/Chelsea_Manning`.

[2] The same remark applies for sure to the death date of a person, also biologically speaking, with the only difference that we can have biographies of living people, where a death date does not need to be specified, until the passing away of a person is being described in a biographical data set.

[3] This could also be an IRI (Internationalized Resource Identifier), but the URI stresses the "Uniqueness" of the resource identifier.

presenting the way biographical data is represented in this framework. This will be followed by a discussion of the paper by Brown and Simpson (2013), which describes how erroneous biographical data can be generated in the Linked Data framework due to the inappropriate use of the `owl:sameAs` property. Finally, we will present our ideas on how to overcome those issues, also calling for a collaborative work in order to generate guidelines to describe biographical data within the Linked Open Data cloud.

2 Overview of Existing Models for Biographical Data

Several repositories of biographies are available in digital format and specific schemata have been proposed to model life events to improve the analysis and understanding of these repositories.

BIO[4] describes a person's life seen as a series of interlinked events. Its vocabulary, expressed in OWL, has four core classes: `Person`, `Event`, `Relationship` and `Interval`. As for the `Event` class, BIO proposes a framework of 37 event types: some of these types apply to all people (e.g., `Birth`, `Death`), others are more specific (e.g., `Coronation`, `BarMitzvah`). Each event is characterised by four properties: `Date`, `Place`, `State` (i.e., territory involved in an event), and `Position` (i.e, employment position or public office). Other properties are used to relate an event to an agent (e.g., `Employer`, `Officiator`) or to temporally order an event with respect to another event (e.g. `Following Event`, `Preceding Event`). An extension of BIO has been proposed within the Shoah Ontology, a domain ontology that formally describes concepts and relationships characterizing the life and persecution of Jews in Italy between 1943 and 1945 (Brazzo and Mazzini, 2015). Here, the ontology class called `Persecution` is used to represent all main events related to the persecution of the victims (arrest, detention, deportation to a Nazi camp, transfer to another camp, liberation, death in a massacre). This class is connected to the `Person` class that is based on BIO extended with additional anagrafic/genealogical properties (e.g. `niece_nephewOf`).

The aim of the **Biography Light Ontology** is twofold (Ramos, 2009): i) encode life events following the 4W model, thus answering questions about what, where, when, who; ii) improve the interoperability among existing vocabularies such as LODE (Linking Open Descriptions of Events)[5] and BIO. Biography Light introduces the main class `BioEvent` with four subclasses that represent changes in the health of the biography's subject, his/her relations with other people, changes in location such as migrations, and inventions or discoveries made by the subject. Event properties are borrowed from LODE (e.g., `atPlace`) and from the Event Ontology (e.g., `isAgentIn`). This ontology has been developed within and adopted by the *Bringing Lives to Light: Biography in Context Project*, an initiative of the Electronic Cultural Atlas Initiative (ECAI)[6] at the University of California. The goal of the project was to design, develop and evaluate tools that can improve the understanding of biographical texts by connecting life events to contextual information, including their location, time of occurrence and related archival materials (Buckland and Ramos, 2010). Different datasets and sources of information were taken into account during the project: namely, the digital texts provided in the on-line Biographical Directory of the United States Congress,[7] the manually compiled chronology of Emma Goldman's lecture itinerary,[8] or the scanned page image of Irish texts.[9]

Bio CRM is a domain-specific extension of CIDOC CRM (Doerr, 2003): it provides a general model for representing biographical datasets that can be extended to meet the requirements of specific projects (Tuominen, 2016). This ontology makes a clear distinction between unary roles of actors, binary relations between actors and events in which actors participate having different roles. Events are described in terms of time, location, participants and other resources involved; moreover, they are organised in an hierarchy distinguishing, for example, ecclesiastical from educational events. Each event type has a corresponding class of permitted roles. Bio CRM has been developed by the Semantic Computing Research Group of Aalto University (Finland) within a set of experiments and projects focused on the linking, enrichment and visualisation of biographies with the aim of improving the reading experience of biographies by providing the users with a rich reading context. A first experiment, called *National Semantic Biography of Finland*, takes the short biographies published in the Finnish National Biography[10] as input data and works on a single type of event (i.e, achievements in the career of a person). An event extractor is used to identify snippets of texts containing words which express creation events, dates written in numbers, named entities of type location and a reference to the name of the subject person of the biography. Extracted information is then transformed in RDF following the Bio CRM model and linked to several external resources such as GeoNames and Wikipedia (Hyvönen et al., 2014). This approach has also been applied to the digitised historical register of the Finnish high school "Norssi", which includes information about the student lives of more than 10,000 alumni (Hyvönen et al., 2017).

Biography.owl is a lightweight ontology designed to represent biographical facts (Krieger and Declerck, 2015): its main feature is the tri-partite structure which entities are modelled with. More specifically, the most general class `Entity` has three subclasses, that is `Abstract` (describing concepts and roles), `Object` (describing physical things) and `Happening`. The latter includes both situations and events, the first being static and atomic, the second dynamic and decomposable. Happenings have properties related to their starting and ending date, the agents involved in them, and their location. Particular attention is devoted to pre- and post-conditions of a happening due to properties encoding causes and effects.

[4] `http://vocab.org/bio/`

[5] `http://linkedevents.org/ontology/`

[6] `http://ecai.org/`.

[7] `http://bioguide.congress.gov/`

[8] See `http://metadata.berkeley.edu/emma/` for a prototype.

[9] For more details see `http://ecai.org/neh2007/`.

[10] `https://kansallisbiografia.fi/english`.

3 Discussing Existing Models

In this Section we discuss in more detail the approaches towards synchronic versus diachronic properties proposed by Krieger and Declerck (2015) and by the Biography Light Ontology, both briefly introduced in the preceding Section. Krieger and Declerck (2015) study how to classify relations (or properties) associated with classes of an ontology as being either synchronic or diachronic. The assumption behind this approach is that a date of birth is something that will not change over time ("a person is born only once"), while the profession exercised by a person can vary over time. While this study was mainly concerned with formalisation aspects, one of the results was that it is in fact very difficult, if not impossible to describe a property that will have only an unalterable value. We can assume that, biologically speaking, a person has indeed only one date of birth, but the statements about this event can be multiple, depending on the sources, or may be revised over time.

Furthermore, interesting statements about "changes" in the Biography Light Ontology could be found: "The Biography Light Ontology takes an event centric approach to the encoding of biographic texts. It is a lightweight framework for common biographic occurrences, such as changes in the health of a biographic subject, relationships between the subject and other people, social groups, or institutions, migration or the change of location of a subject, and biographic events pertaining to creations, inventions, or discoveries produced by the primary subject. The Biography Light model introduces the event type `bl:BioEvent` with four basic subclasses: `bl:ChangeOfHealth`, `bl:ChangeOfRelation`, `bl:ChangeOfLocation`, and `bl:Origination`" (Ramos, 2009). In other words, some properties of the Biography Light Ontology carry the name "Change". Using this ontology, we can extract event factoids modelled as instances of biographical event classes from a biography, as in the example below, adapted from (Ramos, 2009):

> Text: *Robert George Collier Proctor (1868-1903), bibliographer, was born in Budleigh Salterton, Devon, on 13 May 1868. He was educated at a preparatory school in Reading and at Marlborough College, before joining Bath College in 1881.*
> Event factoids:
>
> - `ChangeOfHealth`:
> - birth, 1868-05-13, Budleigh Salterton, Devon
> - `ChangeOfSocialRelation`:
> - studied at Marlborough College, before 1881
> - studied at Bath College in 1881

We can however assume that not only the properties identified in the Biography Light Ontology can change, but all the properties associated to a biographical entity are subjects to changes. The proposal would thus be to equip all properties with a time stamp (an instant or a duration). Technically, this can be done by either allowing n-tuples properties (Krieger, 2014) or by "reifying" a statement about a biographical unit. For example, `hasHealthStatus` can vary very often over time: we can thus "reify" the statement about the health status and encode it as a statement to be equipped with a time stamp in the object part of the resulting new triple. Wikidata[11] is using this method for marking the change of sex/gender by Bradley (later Chelsea) Manning,[12] but also for example for marking the number of inhabitants of a city (Hernández et al., 2015). In addition to that, the provenance of such information needs to be taken into account and to be encoded properly, so that the user can select between sources that seem to be more interesting or more reliable. At this point, we can take advantage of the work described by Ockeloen et al. (2013).[13]

4 The Linked Open Data Cloud

One of our goal of porting a model for biographical data into a Semantic Web compliant formal representation is to be able to publish those data as a specialised subset of the Linked Open Data (LOD) cloud. Figure 1 shows the shape of this cloud, as of 2018-04-30.[14]

Looking at this cloud in more detail, the reader can see the legend to the various colours used to mark the specialised subsets of the Linked Data infrastructure: Cross Domain, Geography, Government, Life Sciences, Linguistics, etc.

Biographical data is also present in LOD, but not (yet) in a specialised subspace. For example, there are a lot of biographical data encoded in the DBpedia node, which is classified as "Cross_domain".

DBpedia[15] started as an effort consisting in extracting structured data from Wikipedia (mainly its "infoboxes") and to encode this information in a Semantic Web compliant representation language. Nowadays, DBpedia is among the largest nodes in the LOD cloud. DBpedia organises its data on the basis of an ontology that was first developed starting from the Wikipedia category system, which can be found in the infoboxes and which evolved to a full ontology representing a directed acyclic graph. This ontology contains 4,233,000 instances (resources), among which 1,450,000 are about entities classified as "Person" and many about other topics that are inherently related to the description of a person, like places, organisations, work, etc.[16] The full ontology is browsable[17] and demonstrates that DBpedia is making use of a large set of (ontological) properties that can be used to describe a biography. Looking at the full ontology, one can also see the details on the information associated with a certain class, illustrating that, for exam-

[11] `https://www.wikidata.org/`

[12] At `https://www.wikidata.org/wiki/Q298423` the entity "Q298423" is marked as being "male" until 22 August 2013 and "transgender female" starting from 22 August 2013. The change of given name of the "Q298423" entity is marked in a similar way.

[13] "Provenance" is also a W3C recommendation, see `https://www.w3.org/TR/2013/REC-prov-dm-20130430/`.

[14] `http://lod-cloud.net/`.

[15] See `http://wiki.dbpedia.org/` for more details.

[16] For more details see `http://wiki.dbpedia.org/services-resources/ontology`.

[17] `http://mappings.dbpedia.org/server/ontology/classes/`.

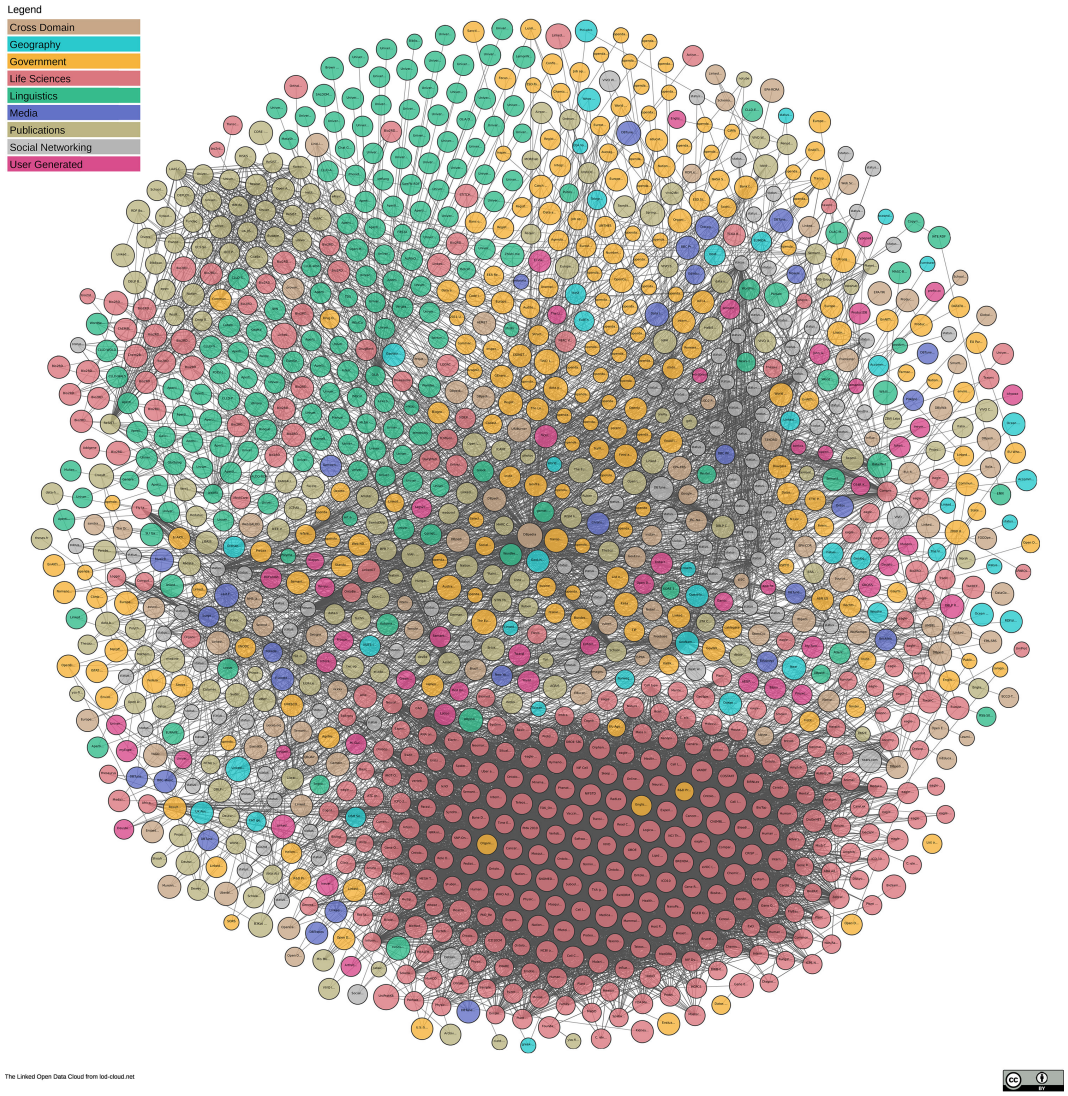

Figure 1: The shape of the Linked Open Data cloud, as of 2018−04−30.

ple, 257 properties are introduced for the class `Person`.[18] The information about domain and range of such properties is given and we can also recognise the type of each property, being either a data-type or an object-type property. Looking at the data-type property `birthDate`,[19] we can see that the class "Person" is defined as its domain and the `xsd:date` type as its range. This setting corresponds to our intuition that only a Person can have a birth date, but for example not a `Group` or even an `Agent` (a superclass of `Person` in the DBpedia ontology). However, it is important to note that the correct setting of domain and range of properties is just ensuring the flow of information to be inherited by the sub-classes of the class bearing the properties, but it is not a restriction on the instances of the classes that can be checked for avoiding inconsistencies.

DBpedia links its data to other knowledge sources using to this end OWL constructs such as `owl:sameAs`. This construct can cause problems and generate errors. An example of this issue is given by Brown and Simpson (2013), which will be described briefly in the following section.

5 Issues with "Michael Field"

Brown and Simpson (2013) describe problems with bibliographical entries in the Linked Data context. More concretely, it involves Katharine Harris Bradley and her niece Edith Emma Cooper. Both are authors of poetry and verse drama and formed a duo, for which they used the pseudonym "Michael Field". The use of pseudonyms is not seldom and has many reasons: in this specific case the choice of a pseudonym could have been motivated by the fact that the authors had an intimate relationship. "Hiding" themselves behind a pseudonym with a masculine name might have been a strategy to avoid social reprobation. In some knowledge sources the relation between each of the literary author and the pseudonym "Michael Field" is stated in such a way that the pseudonym is inheriting a birth/death date, and in the end even two birth/death

[18] `http://mappings.dbpedia.org/server/ontology/classes/Person`.

[19] `http://mappings.dbpedia.org/index.php/OntologyProperty:BirthDate`.

dates,[20] and meaning at the same time that each author is being associated with two birth/death dates, when the relation to the pseudonym is defined as a symmetric one. In this case the un-reflected use of the `owl:sameAs` property between one person and the pseudonym is enough for generating the wrong data, and associating the properties `birthDate` and `deathDate` to the pseudonym. Nevertheless, defining a restriction of the ontology, stating that only instances of the class `Person` can bear the properties `birthDate` and `deathDate` would suffice for avoiding this kind of problem.

Some data sets in the LOD, such as DBpedia, introduce "Michael Field" as an author.[21] All the explanation texts in the DBpedia page for "Michael Field" specify that the entry is about a pseudonym but at the formal ontological level it is introduced as a `Person`, which is wrong, as it should be an instance of a class `Pseudonym`. The same error can be observed in the Yago data set (Suchanek et al., 2007). In the Yago data, we can even see that the name of the pseudonym is segmented in a Given Name and a Family Name and that the pseudonym is bearing a `gender` property, with value "female".[22] We do think that this kind of information is not appropriate for a pseudonym.

The modelling in Wikidata seems to be more accurate, as it introduces "Edith Emma Cooper" as an instance of the class `human` and establishes a `part_of` relation to the "Michael Field" instance of the class `collective pseudonym`.[23]

We also noticed that DBpedia is making use of its property `dbo:wikiPageRedirects` in order to get to the page `http://dbpedia.org/page/Michael_Field\_(author)` when querying for "Edith_Emma_Cooper". While the property `dbo:wikiPageRedirects` is an extremely useful feature helping to normalize variants in names and then pointing them to the right DBpedia page, it is rather cumbersome in the case of "Edith_Emma_Cooper", as it would be better to land on the page describing her and not on a page that deals with the pseudonym she is sharing with another author.

6 On the "Biographical Unit"

As stated above, we did not find a property that can be considered as having a stable value in order to characterise a core element of an entry in a biographical dataset. By now, our intuition is that we just have to declare a class `Person`, being a "life carrier" and having a temporal span, to which all kind of relevant biographical properties can be assigned. Instances of this class are uniquely addressed by URI. This results in a highly abstract model.

[20] The birth and death dates of Katharine Harris Bradley and her niece Edith Emma Cooper are 27 October 1846/26 September 1914 and 12 January 1862/13 December 1913 respectively.

[21] `http://dbpedia.org/page/Michael_Field_(author)`.

[22] `https://gate.d5.mpi-inf.mpg.de/webyago3spotlxComp/SvgBrowser/`.

[23] See `https://www.wikidata.org/wiki/Q3719235` and `https://www.wikidata.org/wiki/Q839369`.

The core element of a biographical unit in such an ontology being a URI, we strongly discourage the use of the `owl:sameAs` property for linking to this unit. The very negative results of applying such a property to the description of entries in a biography in a linked data environment have been precisely and accurately documented in Brown and Simpson (2013), as we reported in Section 5.

Modelling the "Michael Field" data

We started our modelling experiment by encoding the biographical data described in Brown and Simpson (Brown and Simpson, 2013) in order to investigate how we could avoid the problems described in that paper. In particular, we developed OWL/RDF(s) code taking as a starting point the Biographical Ontology (Krieger and Declerck, 2015). Figure 2 depicts the basic class hierarchy we are using for modelling the "Michael Field" data also used by Brown and Simpson (2013). In this figure the small number of instances we have included are indicated in parentheses, which are basically the people named in (Brown and Simpson, 2013).

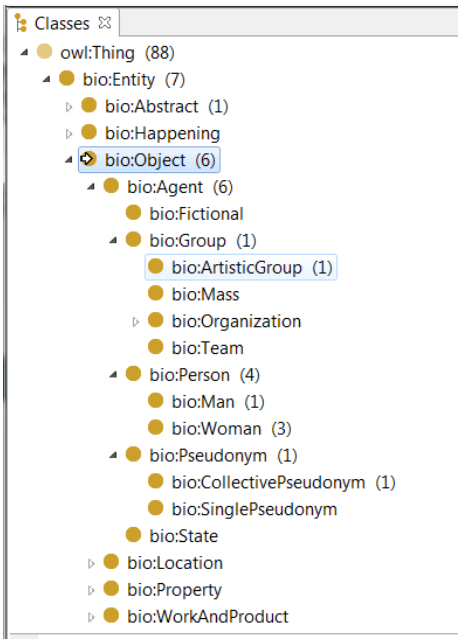

Figure 2: Overview of the Class Hierarchy used for modelling the "Michael Field" data.

The code in Listing 1 displays the way we apply a restriction to the class `bio:Person`, where we state that at least one date of birth has to be given, while we also state that the property `bio:dateOfDeath` is defined for this class.[24] The associated properties `bio:dateOfBirth` and `bio:dateOfDeath` not listed here are defined for domain `bio:Person` and range `xsd:date`, similar to the related properties in DBpedia or Wikidata.

[24] The definition of this class will for sure be updated to include information about provenance. We will also add a constraint stating that within a time period, to be counted from the birth date, a death date has to be given.

Listing 1: The class bio:Person

```
bio:Person
  rdf:type owl:Class ;
  rdfs:subClassOf bio:Agent ;
  rdfs:subClassOf [
     rdf:type owl:Restriction ;
     owl:minCardinality "1"^^xsd:date ;
     owl:onProperty <http://www.dfki.de/lt/
         onto/biography.owl#bio:dateOfBirth>
          ;
    ] ;
  rdfs:subClassOf [
     rdf:type owl:Restriction ;
     owl:onProperty <http://www.dfki.de/lt/
         onto/biography.owl#bio:dateOfDeath>
          ;
    ] ;
  owl:disjointWith bio:State ;
.
```

The code in Listing 2 introduces "MichaelField" as an instance of the Class `bio:ArtisticGroup`. This group consists of two instances of the class `Person`, which are described in the listings 4 and 5 below. It is important to note that it is the specific group, which is associated with the pseudonym `bio:Pseudonym_1` ("Michael Field"). None of the authors alone should be associated with the pseudonym, as it was the case in certain data sets in the LOD cloud.[25]

Listing 2: An instance of bio:ArtisticGroup

```
bio:MichaelField
  rdf:type bio:ArtisticGroup ;
  bio:hasActivity bio:Writer ;
  bio:hasMember bio:Woman_1 ;
  bio:hasMember bio:Woman_3 ;
  bio:hasPseudonym bio:Pseudonym_1 ;
  rdfs:label "\"Michael Field\""@en ;
.
```

The code in Listing 3 introduces "Michael Field" as an instance of the class `bio:CollectivePseudonym`.

Listing 3: An instance of bio:CollectivePseudonym

```
bio:Pseudonym_1
  rdf:type bio:CollectivePseudonym ;
  bio:hasActivity bio:Writer ;
  bio:hasName "Michael Field" ;
.
```

The code in Listing 4 and Listing 5 below concerns the two authors involved in both the artistic duo with the associated pseudonym, but also related to each other by both a familiar and an intimate relation.

Listing 4: An instance of bio:Person

```
bio:Woman_1
  rdf:type bio:Woman ;
  bio:dateOfBirth "1846-10-27"^^xsd:date ;
  bio:dateOfDeath "1914-09-26"^^xsd:date ;
  bio:hasActivity bio:Writer ;
  bio:hasFirstName "Katherine Harris" ;
  bio:hasLastName "Bradley" ;
  bio:hasLover bio:Woman_3 ;
  bio:hasSister bio:Woman_2 ;
  bio:isMemberOf bio:MichaelField ;
.
```

Listing 5: An instance of bio:Person

```
bio:Woman_3
  rdf:type bio:Woman ;
  bio:dateOfBirth "1862-01-12"^^xsd:date ;
  bio:dateOfDeath "1913-12-13"^^xsd:date ;
  bio:hasActivity bio:Writer ;
  bio:hasFather bio:Man_1 ;
  bio:hasFirstName "Edith Emma" ;
  bio:hasLastName "Cooper" ;
  bio:hasLover bio:Woman_1 ;
  bio:hasMother bio:Woman_2 ;
  bio:isMemberOf bio:MichaelField ;
.
```

With this draft encoding our aim was to show how to avoid the issues described by Brown and Simpson (2013) who stress the need to have both a generic ontological framework for describing entities, but also a very specific encoding scheme for accurately modelling all aspects and subtleties of biographical data.

7 Towards a Sub-cloud of the LOD Dedicated to Biographical Data

Based on the observations we could make on the diverse efforts to encode biographies in a Semantic Web compliant format, which have been described in Section 2, Section 3 and Section 4, we see the need for reaching a wide consensus on this ontological design, exploring and possibly reusing existing biography vocabularies and ontologies.

In order to achieve this aim, we can build on the "Shared Data Model" initiative (Fokkens and ter Braake, 2018),[26] which was put in place at the DH Biographical Data Workshop held at the Digital Humanities 2016 conference.[27] We expect that generally accepted guidelines for the ontological encoding of biographical data can be derived from this moderated collection of data models.

In addition, we are advocating for a collaborative effort dedicated to establish a specialised sub-cloud of the LOD framework dedicated to data sets containing biographical data. In this way redundancies and inconsistencies in the modelling of biographical data could be avoided and the modelling of such data could also get a more salient position and an improved visibility in the LOD.

This community group could be organised in a similar manner to the W3C Community Group for the representation of language data in relation to ontologies and to the OKFN

[25] It is also to be mentioned that each member of this group also had an own pseudonym, which we do not display here, for reason of space.

[26] This is a moderated collaborative effort for sharing data models in the field of biography, resulting in a "Repository for Biographical Data Models" (Fokkens and ter Braake, 2018), which can be accessed at `https://github.com/cltl/BiographicalDataModels`.

[27] `http://www.biographynet.nl/dh-biographical-data-workshop/`.

Working Group on Linguistics[28] for building a domain specific subset of the Linked Data cloud, in this case the LLOD cloud.[29]

8 Conclusions

Based on our study of existing ontological models for biographical data, we came to the conclusion that it seems impossible to find one property of a human being that can remain stable within its lifespan. This has consequences on the modelling work, as we need to precisely define what constitutes the uniqueness of an entry in a biographical data set. We advocate for a solution, which consists in introducing a URI for each entry, which needs to be equipped fundamentally with two properties describing the dates of birth and of death. All values to be given to those (and other related) properties are mutable and can also vary in dependency of the provenance information, that also needs to be encoded in the biographical data set.

Furthermore, we came across reports that detail errors in the encoding of biographical data in the Linked Data cloud and which were generated by the inappropriate use of ontological properties and vocabularies. This situation calls for the building of more collaborative work in the field of ontological modelling of biographical data and possibly also for a W3C Community Group dedicated to the creation of a biography specific sub-cloud in the LOD framework.

Acknowledgement

The DFKI contribution to this paper was partly supported by the H2020 project QT21 with agreement number 645452. We thank the anonymous reviewers of the first version of this paper for their very helpful comments. Our thanks go also to Eileen Schnur for proofreading and improving our text.

The paper is dedicated to the memory of Hans-Ulrich Krieger who unfortunately passed away in June 2017. He was the initiator of our efforts in this field and published the first version of the DFKI biography ontology.

9 References

Laura Brazzo and Silvia Mazzini. 2015. From the Holocaust Victims Names to the Description of the Persecution of the European Jews in Nazi Years: the Linked Data Approach and a New Domain Ontology. In *Book of abstract of DH 2015*.

S. Brown and J. Simpson. 2013. The curious identity of michael field and its implications for humanities research with the semantic web. In *2013 IEEE International Conference on Big Data*, pages 77–85, Oct.

Michael Buckland and Michele Renee Ramos. 2010. Events as a structuring device in biographical mark-up and metadata. *Bulletin of the Association for Information Science and Technology*, 36(2):26–29.

Bronwyn Davies and Susanne Gannon. 2006. *Doing collective biography: Investigating the production of subjectivity*. McGraw-Hill Education (UK).

Martin Doerr. 2003. The cidoc conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine*, 24(3):75.

Antske Fokkens and Serge ter Braake. 2018. Connecting people across borders: a repository for biographical data models. In *Proceedings of the 2nd conference on Biographies in a Digital World*.

Daniel Hernández, Aidan Hogan, and Markus Krötzsch. 2015. Reifying RDF: what works well with Wikidata? In *Proceedings of the 11th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS 2015)*, volume 1457 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Eero Hyvönen, Miika Alonen, Esko Ikkala, and Eetu Mäkelä. 2014. Life stories as event-based linked data: case semantic national biography. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track-Volume 1272*, pages 1–4. CEUR-WS. org.

Eero Hyvönen, Petri Leskinen, Erkki Heino, Jouni Tuominen, and Laura Sirola. 2017. Reassembling and enriching the life stories in printed biographical registers: Norssi high school alumni on the semantic web. In *International Conference on Language, Data and Knowledge*, pages 113–119. Springer.

Hans-Ulrich Krieger and Thierry Declerck. 2015. An owl ontology for biographical knowledge. representing time-dependent factual knowledge. In *Proceedings of the First Conference on Biographical Data in a Digital World 2015*. CEURS-WS.org, 7. Online-Proceedings: http://ceur-ws.org/Vol-1399/.

Hans-Ulrich Krieger. 2014. A detailed comparison of seven approaches for the annotation of time-dependent factual knowledge in rdf and owl. In *Proceedings of the 10th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (held in conjunction with LREC 2014)*. European Language Resources Association.

Niels Ockeloen, Antske Fokkens, Serge Ter Braake, Piek Vossen, Victor De Boer, Guus Schreiber, and Susan Legêne. 2013. Biographynet: Managing provenance at multiple levels and from different perspectives. In *Proceedings of the 3rd International Conference on Linked Science - Volume 1116*, LISC'13, pages 59–71, Aachen, Germany, Germany. CEUR-WS.org.

Michele R. Ramos. 2009. Biography Light Ontology: An Open Vocabulary For Encoding Biographic Texts. Technical report, Bringing Lives to Light: Biography in Context Project.

Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706. ACM.

Jouni Tuominen. 2016. Bio CRM: A Data Model for Representing Biographical Information for Prosopography. Version 2016-08-19. Technical report, Bringing Lives to Light: Biography in Context Project.

[28] See `https://www.w3.org/2016/05/ontolex/` and `https://blog.okfn.org/category/working-groups/wg-linguistics/`.

[29] `http://linguistic-lod.org/llod-cloud` for details on the Linguistic Linked Open Data cloud.