Persons in Context. A Model to Represent Observations and Reconstructions of Historical Persons in Linked Data

By Pieter E. Woltjer, Ivo Zandhuis, Bob Coret, Mark Lindeman, Jeroen D. Balkenende, Richard L. Zijdeman and Rick J. Mourits

To cite this article: Woltjer, P. E., Zandhuis, I., Coret, B., Lindeman, M., Balkenende, J. D., Zijdeman, R. L., & Mourits, R. J. (2024). Persons in Context. A Model to Represent Observations and Reconstructions of Historical Persons in Linked Data. *Historical Life Course Studies, 14*, 105–125. https://doi.org/10.51964/hlcs19312

HISTORICAL LIFE COURSE STUDIES

VOLUME 14

2024

EHPS NETWORK

MISSION STATEMENT

HISTORICAL LIFE COURSE STUDIES

Historical Life Course Studies was established within *European Historical Population Samples Network* (EHPS-Net). The journal is the primary publishing outlet for research involved in the conversion of existing European and non-European large historical demographic databases into a common format, the Intermediate Data Structure, and for studies based on these databases. The journal publishes both methodological and substantive research articles.

Methodological Articles
This section includes methodological articles that describe all forms of data handling involving large historical databases, including extensive descriptions of new or existing databases, syntax, algorithms and extraction programs. Authors are encouraged to share their syntaxes, applications and other forms of software presented in their article, if pertinent.

Research articles
This section includes substantive articles reporting the results of comparative longitudinal studies that are demographic and historical in nature, and that are based on micro-data from large historical databases.

Historical Life Course Studies is a no-fee double-blind, peer-reviewed open-access journal supported by the European Science Foundation, the International Institute of Social History, the European Society of Historical Demography, Radboud University Press, Lund University and HiDO Scientific Research Network Historical Demography. Manuscripts are reviewed by the editors, members of the editorial and scientific boards, and by external reviewers. All journal content is freely available on the internet at hlcs.nl.

HISTORICAL LIFE COURSE STUDIES
VOLUME 14 (2024), 105–125, published 17-10-2024

Persons in Context

A Model to Represent Observations and Reconstructions of Historical Persons in Linked Data

Pieter E. Woltjer — CBG|Center for family history
Ivo Zandhuis — International Institute of Social History
Bob Coret
Mark Lindeman
Jeroen D. Balkenende — CBG|Center for family history
Richard L. Zijdeman — International Institute of Social History & University of Stirling
Rick J. Mourits — International Institute of Social History

ABSTRACT

Reconstruction of historical persons and family ties is the bread and butter of many researchers and genealogists. With the increasing digital availability of historical person records, the scope and depth of person reconstructions speaks to the imagination of researchers and genealogists. Yet, the lack of standardisation in the description of historical person data has hurt the interoperability and sustainability of both small and large databases. Persons in Context, or PiCo, presents a data model for historical person records within the Resource Description Framework (RDF). RDF or Linked Data is specifically designed for clear, unambiguous information exchange between multiple parties over the internet. We show how reuse of existing ontologies and concentric description are the building blocks of a flexible, straightforward, and stringent data model that emphasises provenance.

Keywords: Linked data, RDF, Family reconstitution, Person reconstruction, Data model, Historical database management, Standardization historical data, FAIR data, Ontology, Person observation

e-ISSN: 2352-6343
DOI article: https://doi.org/10.51964/hlcs19312

1 INTRODUCTION

For almost two centuries, civil registration data has inspired epidemiologists, social scientists, and statisticians (see, for example, Durkheim, 1897; Quetelet & Smits, 1832; Snow, 1855). The introduction of family reconstitution by Henry and Fleury (1956) ushered in a new era of research, as historians have been collecting person observations from archival records to reconstruct life courses and family relations ever since (Fauve-Chamoux et al., 2016). These efforts have resulted in today's databases varying from detailed local reconstitutions[1] to nation-wide reconstitutions with millions of records[2]. Each of these databases has proven its worth, but many have also been forgotten, as they are single, isolated sets of data. It is impossible for single scholars to have an overview of all these isolated databases and even overview studies only list the databases associated with larger reconstruction projects (Edvinsson et al., 2023; Mandemakers, 2023; Song & Campbell, 2017). The findability of these datasets can be improved by improving the interoperability of all historical databases as a whole (Hoekstra et al., 2016; Wilkinson et al., 2016).

Both big and small databases will benefit from improved interoperability. Large data sets obviously need standardisation before they can be matched, as the data is too complex (and labour intensive) for ad hoc standardisations by individual researchers. However, large data gathering efforts often exist in a closed system, which diminishes opportunities to add local data. Small, local datasets also known as "long tail of data", exist in obscurity or are at risk of disappearing into it. In linkage operations they are often "converted" to the database that is central to the study at hand, meaning others will need to repeat the work of "converting" the small database to their database. Unlike the large scale data infrastructures, these smaller sets receive less professional support and are hosted on personal websites or stored locally rather than in dedicated storage places. This compromises the findability and accessibility of datasets, as interoperability with professional standards is often limited and can cause datasets to disappear when they are no longer in use.

To deal with both the findability and interoperability of data, various solutions have been presented. The option is believed to lie in schemas, which are structured ways to describe data (Meroño-Peñuela et al., 2020; Mourits et al., 2024). Currently, there are six competing open source schemas[3] in the field of social history. Most of these have a limited scope and are specifically designed to tackle interoperability of census data (Minnesota Population Center, 2020; Roberts et al., 2003; Ruggles et al. 2011; Ruggles et al., 2021; Szołtysek & Gruber, 2016) or optimise person reconstructions for statistical analyses (Mourits et al., 2020). Only the Intermediate Data Structure (Alter & Mandemakers, 2014) is much broader in scope as it makes different types of data sources available for extraction. Efforts by IPUMS, MOSAIC, SwedPop, and the Utah Population Data Base among others (Smith et al., 2022; Swedpop, 2024; Szołtysek & Gruber, 2016) have shown how open source, and even proprietary, schemas can be used to combine and publish data within a project or institutional environment. Yet counterintuitively, these projects do not exist as one ecosystem, but are to a large extent non-interoperable islands that describe similar concepts in different terms. The lack of consensus within the field on how to describe common concepts hampers the discussion on how to describe specialised historical concepts. Moreover, it hinders the adaptation of best practices from other fields, for example, provenance (Mourits et al., 2024). Hence, current efforts to standardise data have only led to marginal gains in improved interoperability and concomitant improvements in findability and accessibility.

In the Netherlands, the Center of Family History, from here on CBG, has suggested a new way to standardise data in schemas. Computer scientists specialised in Linked Data worked together with archivists from the local, regional, and provincial archives and scholars from the International Institute of Social History to realign the wheel rather than reinvent it. Our goal was to describe historical data

1 For example, the Barcelona Historical Marriage Database (Pujadas-Mora et al., 2022), Keweenaw project (Trepal et al, 2021), Knodel's German village family reconstitutions (Knodel, 2002), Roteman Database (Geschwind & Fogelvik, 2000), Saquenay-Lac-St-Jean (Vézina & Bournival, 2020), Scania Economic Demographic Database (SwedPop, 2024).

2 For example, China Multi-Generational Panel Datasets (Lee & Campbell, 2016), LINKS, POPUM (SwedPop, 2024), Utah Population Database (Smith et al., 2022).

3 Intermediate Data Structure or IDS in brief (Alter & Mandemakers, 2014), IPUMS-USA (Ruggles et al., 2021), IPUMS-International (Minnesota Population Center, 2020), LINKS-gen (Mourits et al., 2020), MOSAIC (Szołtysek & Gruber, 2016), and North Atlantic Population Project or NAPP in brief (Roberts et al., 2003; Ruggles et al., 2011).

using existing terminology, as concepts such as gender, names, or (family) relation are not unique to history and are already described by other — more authoritative — institutions (see, for example, Davis & Galbraith, 2004; Gao et al., 2012; Guha et al., 2014). In this paper, we introduce a data model to make historical person data FAIRer[4] called *Persons in Context* or *PiCo*[5]. We argue that by reusing existing terminology, historians can adapt best practices from other disciplines, which gives them time and focus to discuss highly specialised vocabularies within their own field (Meroño-Peñuela et al., 2020; Mourits et al., 2024). PiCo does so by describing historical person data concentrically and by using general, domain-specific, and specialised schemas to describe general, domain-specific, and specialised concepts. In time, this will only make historical data more findable, accessible, and interoperable.

2 CONTEXT

In the Netherlands, a common schema for exchange of historical person data has been a practical issue for about a decade. Since 2010, archival institutions in the Netherlands have tried to tackle the exchange of historical person data via an XML standard called Archive to Archive (A2A).[6] Over time, A2A has become a well-adopted standard in the Netherlands, improving the exchange of historical person data from various record types between over 50 local, regional, and provincial archival institutions. Moreover, nearly 200 genealogical datasets have been made available in the A2A format[7] and A2A is also used by genealogical platforms for data harvesting, such as openarchieven.nl and wiewaswie.nl. However, the ever-increasing use of A2A and changing standards in computer science have highlighted some shortcomings in the original A2A schema.

In the early 2010s, XML was a logical vehicle to exchange historical person data, as it was the de facto standard for data transfer. However, as data is being exchanged on an ever-increasing scale, requirements on data exchange formats have become more stringent and new standards for data transmission have risen in prominence over the past decade. A2A was never designed to be used by the wide range of institutions on the myriad of historical person records that it is being used for today. As a result, A2A has become much less efficient today than it was a decade ago. Moreover, A2A was designed to model historical person observations on certificates, rather than person reconstructions based on all types of person observations, despite widespread academic and public interest in the latter. In order to circumvent this omission, ad hoc solutions have been implemented by different parties, which reduces the interoperability of the data. Clearly, the use of historical person data has grown beyond the original design of A2A and all systems based on A2A are crumbling down.

In recent years, various initiatives have arisen that use the Resource Description Framework (RDF) to describe large collections of data, also known as the Semantic Web or Linked Data. A key characteristic of RDF, is that everything described in a data set is unique and standardised with Uniform Resource Identifiers (URIs) represented as Uniform Resource Locators (URLs) or web addresses that retrieve and describe the resources, so that people can find this information. For example, Google describes persons in Schema.org with the URI, https://schema.org/Person as "A person (alive, dead, undead, or fictional)" that can have certain characteristics. Others can reuse this URI to describe persons in their own datasets.

The growing acceptance of RDF within computer science and the Dutch heritage sector makes it a logical choice describing historical person data. Various efforts have been undertaken to describe part of the information that is part of historical person records, mentions, or reconstructions. They provide more detailed information on persons in RDF and are established practices to make person data more findable and interoperable. For example, the Friend of a Friend (FOAF) vocabulary describes various person characteristics, mainly aimed at exchanging information on contact details. The BIO vocabulary lists a great many relations that persons can have, e.g. "sister", "brother-in-law", and makes it very helpful to describe (con)sanguineous relations. Moreover, Den Engelse and Van Wissen (2019) have shown that the RDF framework can describe historical person records if a strict distinction between person observations and person reconstructions is maintained. While each of these vocabularies have

4 See Wilkinson et al. (2016).
5 https://personsincontext.org
6 https://blogbob.coret.org/2015/02/een-stukje-geschiedenis-van-het-a2a-model.html
7 https://datasetregister.netwerkdigitaalerfgoed.nl/search.php?lang=en#eyJ0IjoiYTJhIn0=

their strong suits, none of them is encompassing enough to describe historical person data. Hence, historical person data needs to draw on a wide range of schemas to describe it, which is not easy as there is quite some overlap between existing vocabularies. Haphazardly combining insights from existing schemas leads to a myriad of terminology within and between different data sets. Therefore, studies on best practices within the humanities and historical demography suggest describing data "concentrically", that is, the use of general, domain-specific, and specialised schemas to describe general, domain-specific, and specialised concepts (Meroño-Peñuela et al, 2020; Mourits et al., 2024).

Taking elements from the pioneering efforts above, we present a data standard that allows for interoperability of both big and long tail data using RDF. The standard is applicable to the most common person observation documents in archives and allows for person reconstructions while at the same time being intuitive to users. We will first set out the requirements for such a data model for persons (Section 3), after which we present the model (Section 4), show applications of the model to various archival resources (Section 5), and reflect on the merits of the PiCo data model (Section 6).

3 REQUIREMENTS FOR A NEW DATA MODEL FOR PERSONS

From personal experience, a decade of exchanging historical person data has highlighted four key characteristics of an effective data model for historical person data. First, *flexibility* to apply our standard to a wide range of historical person data and to enable extending it to new types of historical person data in the future. Second, a *straightforward* model that can be easily implemented and used uniformly. Third, *stringency* in our definitions and modelling prevents ad hoc interpretations and enables efficient and effective automated processing. Fourth, *provenance* to capture the origin of historical documents, person observations, and person reconstructions. These four characteristics are met by the Resource Description Framework.

3.1 FLEXIBILITY

Flexibility is essential for any data model within the historical realm. Historical sources and concomitant person observations appear in many different forms, especially when variations over time and by space are taken into consideration. Without flexibility, a model breaks when a user wishes to incorporate a previously unused historical source. For instance, A2A was mainly designed to capture and use Dutch civil registries from the 19th and 20th centuries and had little consideration for other types of sources. As a result, capturing population registers required ambiguous use of the event-centred A2A model. Unlike birth, marriage and death certificates, the population registers are not based on life events, but continuous registration of the population. However, because A2A requires an event, data providers introduced "registration" as an event type. This new "event" should not be dated, as registrations in the population registers refer to an observation period instead of a specific date. Nevertheless, some data providers felt incentivised to add dates to registrations in the population register.

Similar ambiguities occur with other sources that observe persons actively over time rather than passively through events, indicating that a flexible data model should not be event- or source-based.[8] Therefore, we chose to make person observations the focal point of PiCo. PiCo was designed with genealogical research in mind. But as an extensible model, it is suitable for other types of research as well. The focus on genealogy emphasises the importance of identifying unique persons and the relations between them, which are a perfect base for other types of data, such as biographies of people involved in the Second World War or local time machine projects.[9] Anyone can extend the model with the required properties for every such domain. If we, in the future, want to use the same model with naming conventions from other languages, we must be able to add that to our model.

8 E.g., see these examples about F. Domela Nieuwenhuis, born on 31 December 1846 (https://en.wikipedia.org/wiki/Ferdinand_Domela_Nieuwenhuis): https://www.openarchieven.nlnha:f79b94f3-c978-2db8-a649-73f05336d50b; https://www.wiewaswie.nl/nl/detail/102815854; https://www.openarchieven.nl/saa:dde7f0a5-08a7-4249-9377-e66f51fd1a9b.

9 For example, https://www.oorlogsbronnen.nl/mensen or https://www.timemachine.eu/ltm-projects/.

We can create a flexible standard for data exchange by "using generic ontologies where possible, and domain ontologies where needed" (Meroño-Peñuela et al., 2020, p. 86). The reuse and combination of already developed standards is considered a best practice, although it has — to our knowledge — never been implemented as the basis for an entire data model within the fields of demography and genealogy. PiCo uses generic, domain-specific, and specialised ontologies to model historical person data. By describing the data concentrically in three layers, we ascertain that the model can be extended without breaking the mould. The generic ontologies are mostly fixed, as they are defined on a global level and out of our span of control. Domain-specific ontologies are still rigid, but can change or be extended if the domain deems it necessary. Most changes will occur in the use of specialised vocabularies that are directly affected by new research insights or developments by groups of scholars (Mourits et al., 2024).

Table 1 *Concentric description of a marriage certificate using PiCo*

Schema level	Schema name	Concept name
General	XSD[1]	date, int, string
	Schema.org[2]	spouse, parent, gender, familyName, givenName, ArchiveComponent, dateCreated, locationCreated
Domain-specific	BIO[3]	Marriage, date, partner
	PROV-O[4]	hadPrimarySource, wasDerivedfrom
Specialised	PiCo	PersonObservation, hasRole, huwelijkspartij, huwelijksakte, hasAge

Remarks: 1) Gao et al., 2012; 2) Guha et al., 2014; 3) Davis & Galbraith, 2004; 4) World Wide Web Consortium, 2013.

The PiCo community consists of different users. On the one end of the spectrum are archival institutions that provide data, whereas on the other end genealogists and researchers retrieve data for their studies. Each of the users can expand on the PiCo model to describe or generate data not currently included in the data model, and over time some of these practices will most likely become de facto standards. To streamline discussions and share best practices, the CBG chairs a committee with representatives from different user groups to monitor the use of ontologies and discuss possible extensions of the PiCo model.

3.2 STRAIGHTFORWARDNESS

Straightforwardness ensures that the data model is easy to use and can intuitively be used correctly. Archivists, genealogists, and programmers all have different objectives, but share a common interest in the proper description of data. For any standard to be used it should be self-evident how its use improves the representation and retrieval of data. Data models that are easy to use attract more users, have lower maintenance costs, and are easier to develop applications for. Over time this will create an ecosystem where it is evident why people should work with PiCo and both non-profit and commercial partners are inclined to introduce the standard in their software.

PiCo adopts existing best practices by reusing authoritative ontologies. Describing data structures concentrically helps to streamline discussions on how to represent historical person data. One of the most commonly used schemas to describe classes and properties is Schema.org, which makes it a logical choice to define generic concepts in historical person data. Schema.org has the advantage that it is broadly accepted and implemented which makes the data usable in contexts outside the genealogical and demographical domain (Meroño-Peñuela et al, 2020). Furthermore, Schema.org is easy to use. For example, a date property can be used to describe a person's date of birth directly, rather than via a separate "birth event", which would unnecessarily complicate the data model and concomitant queries. More specialised ontologies are required to describe concepts that are missing in Schema.org, such as provenance and events. Here, the World Wide Web Council (W3C) and BIO are obvious authorities, so that almost no tailor-made vocabularies are necessary.

Paradoxically, some complexity needed to be introduced to ease the implementation of PiCo and enable more advanced and sophisticated querying of the data. For purely practical reasons, PiCo has to be backwards compatible with the A2A standard, so that the existing collections of historical person data can be migrated to PiCo without considerable costs or time investments. This requires that XML-data in A2A-format can be directly converted to Linked Data in the PiCo format without manual edits or corrections. Secondly, we introduce a distinction between person *observations*, which are available in sources, and person *reconstructions*, which are constructed from one or more observations by an algorithm, genealogist, or researcher (den Engelse & van Wissen, 2019). This distinction is necessary to reconstruct the population from the different types of historical person records and clearly communicates to users which properties are used to indicate the characteristics and quality of a person reconstruction.

3.3 STRINGENCY

Stringency is a requirement for a model that aims to be both flexible and straightforward. At a first glance, these three concepts might seem to be incompatible. However, a clear and stringent model is by definition straightforward and gives freedom as it prevents confusion and alternative interpretations. Imprecise definitions lead to ambiguous data, which cannot be handled without interpretation or large-scale cleaning. As properties and classes in RDF have URIs containing descriptions that can be viewed by anyone, confusion about their definitions is kept to a minimum.

By being stringent, data models ascertain that the foundation of a database is solid and that third parties can start building upon it. The wrong use of event dates in population registers is a clear example of what can go wrong when a data model leaves too much room for interpretation. Representing this ambiguous data in the user interface of a genealogical website turns out to be cumbersome. Ambiguity also makes deductions based on data impossible, which lowers the possibilities and quality of software that can be written, such as automated processes for record linkage. The only way to solve these issues is to check all possibly affected cases by hand or enter the required information for a second time, which has become almost infeasible with the ever-growing amount of available historical person data.

Ten years of working with A2A has shown that data entry diverges most in the description of the *roles* of person observations and their *relations* to other person observations. In order to capture these two concepts precisely, specialised vocabularies are needed to model role types, such as "witness" or "testator", as well as relation types, such as "father", "stepfather", or "foster father". These vocabularies were not sufficiently enforced in A2A, which meant that the same roles and relations were coded differently, which made it unnecessarily cumbersome to write advanced queries on the historical person data.

An added advantage of a stringent model is that software can be used to ascertain data quality. Data providers can use technical solutions to standardise the data during data entry and check the quality of entered data, whereas users can check the data before consuming it. For RDF, the Shapes Constraint Language (SHACL) is developed for this purpose. In this language, rules — also known as shapes — are defined to validate whether the data complies with these rules. The CBG has the ambition to develop these shapes and make them available to data providers and users to help improve the quality of existing historical person data and prevent structural mistakes in the future.

3.4 PROVENANCE

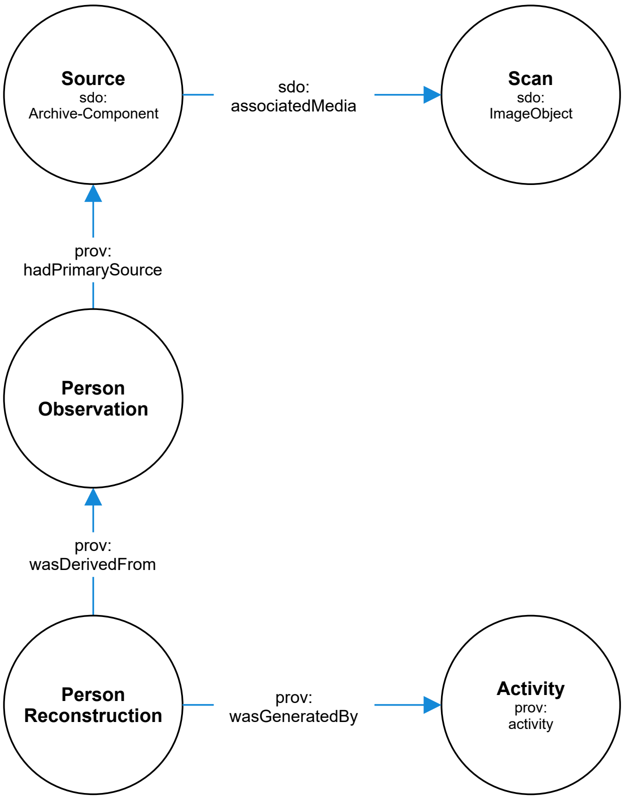
Finally, provenance allows researchers to keep track of how data was generated by stating the relationship between a historical argument and the primary and secondary sources it is based upon. Describing provenance requires a decentralised system, as the best moment to state when, where, and why data were created is when the data are being created. Hence, provenance data can only adequately be described by the person who generates it. Every institution can be asked to publish data with a standardised set of machine-readable information, which makes each statement clear and traceable, so that users can quickly gauge data quality.

In our case, the observations in the sources, such as censuses, the civil registry, militia registers, parish registers, personal cards, and population registers, eventually evolve into reconstructions of persons from information on historical records. Family trees made manually by genealogists are probably the most common example of person reconstructions. However, we expect that in most research programs the reconstructions will be created computationally (Idrissou et al., 2022; Raad et al., 2020). A user of the results of these computations should be able to know when, how and based on which sources a reconstruction was made. To be able to trace back the source from any reconstruction PiCo prescribes a strict distinction between observations and reconstructions, storing the relations between reconstructions and the underlying observations, and a link between the observation to the source where the observation is done. Combined, these three elements make that data presented in PiCo is always verifiable by the user.

4 THE MODEL

PiCo distinguishes three concepts in historical person data: sources, person observations within these sources, and person reconstructions based on one or multiple person observations. Combined, these three concepts describe historical sources as well as efforts by genealogists and historians to reconstruct families and life courses.

Figure 1 *The relation between person reconstructions, person observations, and sources in PiCo*



4.1 PERSON OBSERVATION

A person observation represents how a person is recorded. Therefore, they must always be related to a source. Person observations closely reflect data as it appears in the source, that is, with the age, occupation, and spelling of the person's name recorded as in the record. A notable exception are dates where the creator of the person observation is encouraged to write down the date according to the YMD ISO 8601 date standard (for example: "1802-07-29"), even though the source probably lists the date in a DMY or MDY format or another calendar entirely, for example "28 thermidor X" on the French Republican Calendar. Information on the original entry can be stored using an optional Triple, so that the database still shows the original date. The literal date and the ISO date can be distinguished by using xsd:type, for example:

afr:geboorteregister_1858_po_1 sdo:birthDate "1858-06-21"^^xsd:date .

afr:geboorteregister_1858_po_1 sdo:birthDate "21 Junij 1858"^^xsd:string .

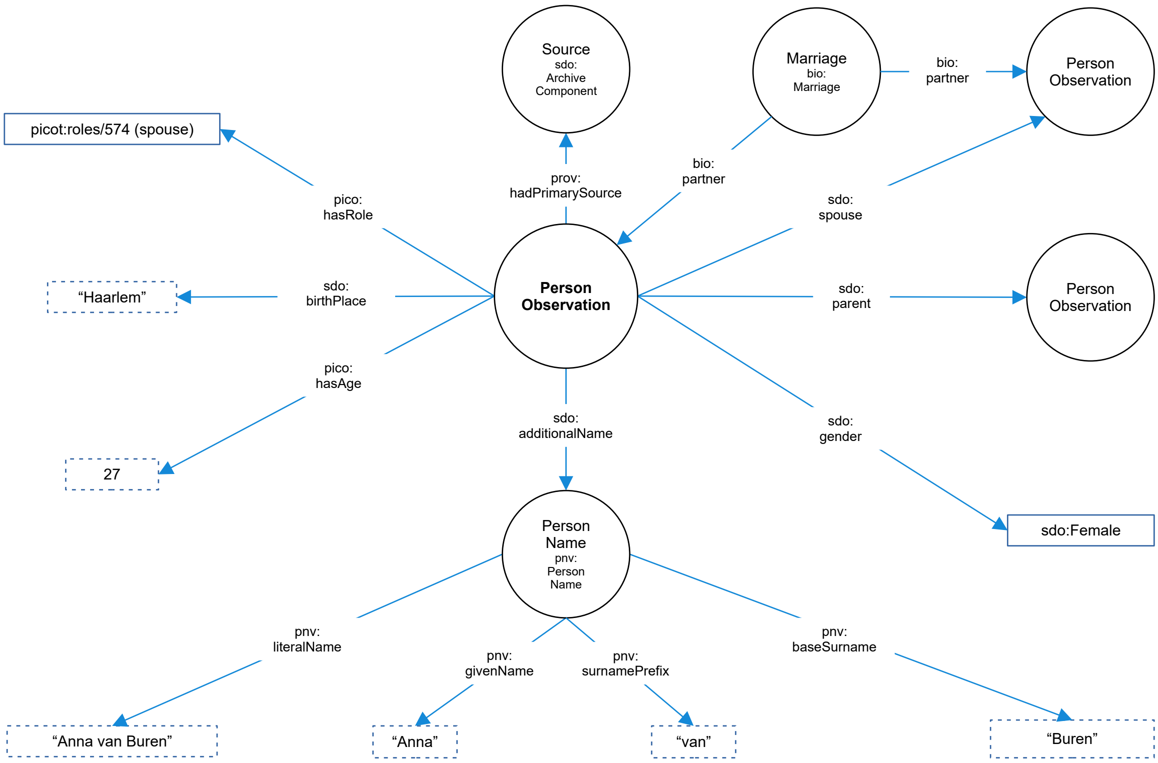
To describe person observations, PiCo uses the class pico:PersonObservation. Person observations are identified primarily by their name and date of birth. For these persons characteristics are listed, such as occupation, place of residence, religion, and place of birth. Most characteristics of person observations can be described using Schema.org. However, some properties that are typical for historical person records are unavailable in Schema.org. For example, some sources mention a person's age instead of their birth date. To deal with this omission in Schema.org, the property pico:hasAge is included in the PiCo model.

Where Schema.org is used for generic characteristics of a person observation, PiCo allows the use of specific properties where needed. For detailed annotations of (Dutch) person names, for example, the Person Name Vocabulary[10] is embedded in the model. This allows the data provider to distinguish between specific parts of a name, like the honorific title or the patronym. Users of the PiCo model from other countries are free to use their own specialised ontologies for specifying names. The use of the more generic Schema.org properties, like sdo:familyName and sdo:givenName, guarantees the interoperability of datasets using different ontologies for names.

Besides personal characteristics, historical records also document relationships between person observations. These relationships are consistently recorded in many of the different historical person records, which shows the importance of this information to the registrar. In PiCo a relationship describes the connection between two people, such as sdo:parent, sdo:spouse, or sdo:children. Generally, a person observation also has a specific function on the source, which PiCo calls a role. For example, one can say that a person has the role of "mother of the bride" on a marriage certificate. This implies there is also a connection with one of the other person observations on the marriage certificate, which PiCo makes explicit by linking the bride and her mother with the property sdo:parent. This distinction makes it easier to query the data and, for example, reconstruct persons.

A SKOS thesaurus for the roles of a person observation on a Source is published as part of the PiCo standard at https://terms.personsincontext.org/roles/.

Figure 2 *Linked data schema to describe historical person observations*



4.2 PERSON RECONSTRUCTIONS

A large number of genealogists and scholars have reconstructed historical persons by combining one or more person observations into a uniquely identifiable person reconstruction. This person reconstruction is very close to what most people would call a person. However, this person is reconstructed either automatically through specialised software or by hand via a more classical genealogical research. PiCo enables users to store and publish their person reconstructions in a standardised way. Each person reconstruction should include references to the person observations on which it is based. Since a person reconstruction always stems from one or more person observations, it must always have a relationship with at least one person record. The prov:wasDerivedFrom property is used for this purpose. The property prov:wasGeneratedBy is used, in combination with the class prov:Activity and prov:Agent, to indicate how and by whom the person reconstruction was created. When standardised person reconstructions become more widely available and used, this information allows genealogists

10 https://w3id.org/pnv

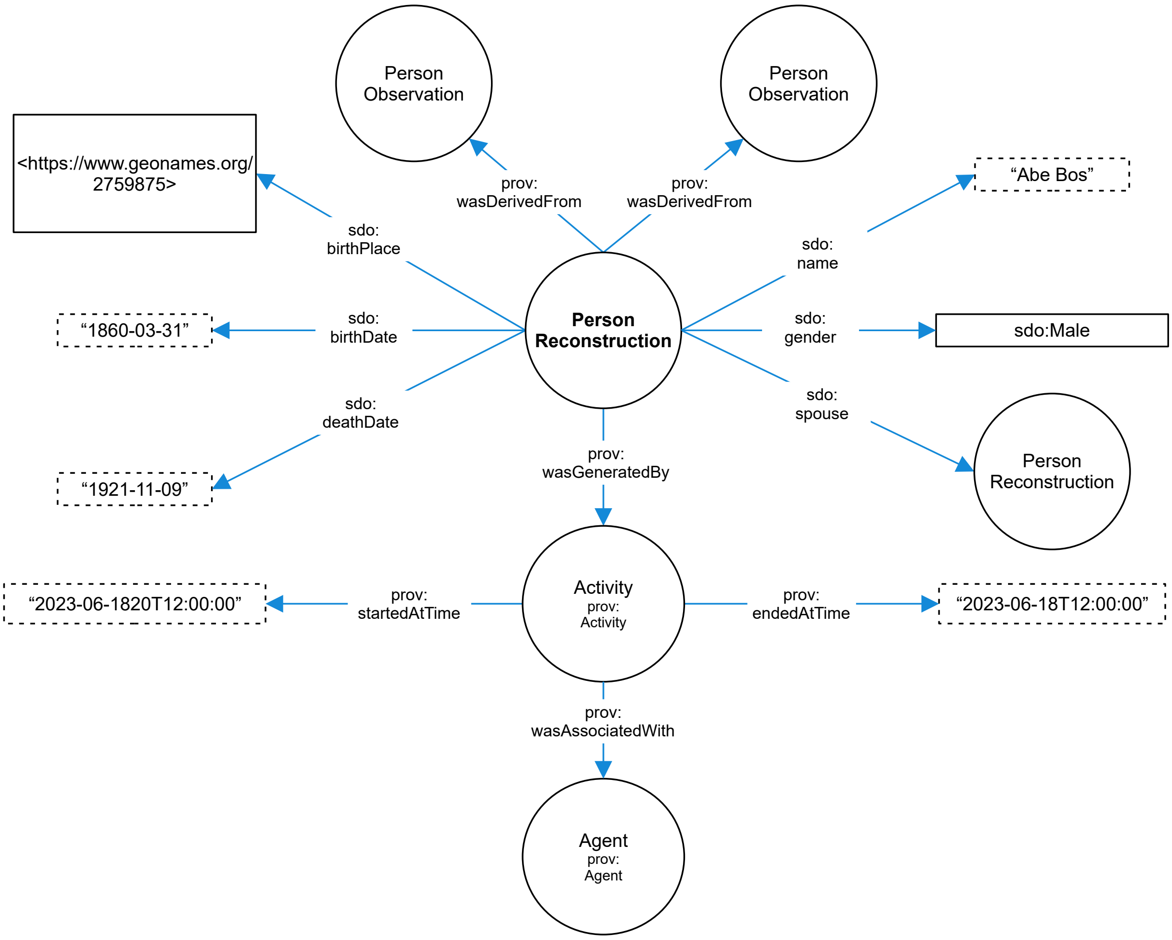
and scholars to understand differences between different reconstruction efforts.

Making person reconstructions necessitates reconciliation and standardisation of information on the matched person observations. We encourage data providers to use ontologies to describe information in person reconstructions. For example, there are dozens of town names in the Netherlands with an identical name and even more with very similar names. For a user searching for his great grandfather, it may be sufficient if he is registered as being born in "Alphen". For a computer algorithm trying to match millions of records, "Alphen" may be too ambiguous to create a certain match, as it may to refer to Alphen aan de Maas in Gelderland, Alphen aan den Rijn in Zuid-Holland, or Alphen-Chaam in Noord-Brabant. Adding a URI that uniquely identifies the misspelt variant of "Alfen" as Alphen aan den Rijn as defined by geonames increases the potential for querying and representing person reconstructions. The same principle applies to occupations and other characteristics that are available on the matched person observations.

ex:po_456 sdo:birthPlace "Alfen"^^xsd:string .

ex:po_456 sdo:birthPlace <https://www.geonames.org/2759875> .

Figure 3 *Linked data schema to describe historical person reconstructions*



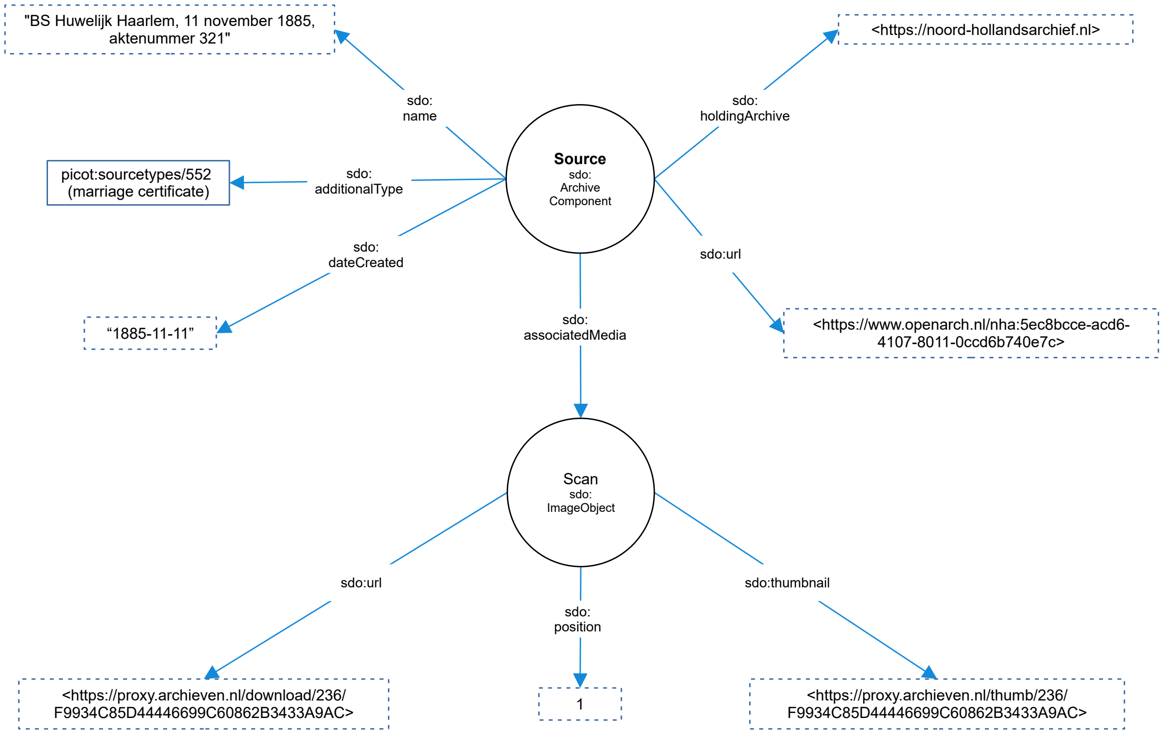
4.3 SOURCES

Person observations can appear on many types of historical documents. The PiCo ontology does not aim to describe these sources, as excellent data models, such as Records in Contexts and Dublin Core, already exist within the heritage domain. In order to model person observations, we require that institutions describe their sources with a single name property that combines identifying information about the source like its title and creation date. For example, a Marriage Certificate could be described as: "BS Marriage Haarlem, November 11, 1885, certificate number 321". As long as this description consists of various elements that identify the source, archival institutions are free to use any authoritative domain-specific standard to describe their historical person records.

It is strongly encouraged that a source includes a permanent link to the record at the archival institution. Ideally, this would be a web link to the inventory number that shows the source related entry and the page it is on. This link allows users to verify the source or to read it in person and extract more details that are not part of the person observation data, but are still relevant to the user.

A SKOS thesaurus for document types of Source is published as part of the PiCo standard at https://terms.personsincontext.org/sourcetypes/.

Figure 4 *Linked data schema to describe historical person records*



5 DESCRIBING HISTORICAL PERSON RECORDS

In this chapter, we will give some examples of person observations and reconstructions expressed in PiCo. For readability, all examples are provided in the Turtle syntax, but any of the other serialisation formats could have been used. The examples apply prefixes that substitute the base URL of an identifier. This is common use in Turtle. The property *sdo:familyName* should be read as https://schema.org/familyName. To keep triples intelligible, the prefix declarations are omitted from the examples. A complete list of prefixes used can be found in the Appendix.

5.1 AN EXAMPLE OF A PERSON OBSERVATION: DESCRIBING A BIRTH CERTIFICATE

The Dutch civil registry contains a wide range of person characteristics. By civil law, municipalities were obligated to register who was born, married, or died, to whom this person was related, who informed the local clerk about the event, and who bore witness to the event (Mourits et al., 2020). As a result, each birth certificate contains multiple person observations: an informant, the newborn child, one or two parents, two witnesses, and sometimes the local clerk. These person observations, the relationships between them, and the roles that these persons perform on the person record can be described in PiCo.

The example underneath describes the three most important people on a birth certificate: the parents and the newborn. Their names, dates of birth, and sex can all be described using Schema.org. The given name, family name, and complete name are all recorded as shown in the source. The mentioned birth date of the index person "21 Junij 1858" is converted into "1858-06-21", as Schema.org requires dates to be in ISO 8601 yyyy-mm-dd date format. Using the date format is obviously helpful for online searching and it enhances the possibility for matching person observations into reconstructions. However, we also chose to include the literal string along with the interpreted date to show how dates were originally written down in the source. Finally, sex was declared to be of the female gender.

Besides the basic demographic information we add information on age, relation, role, and provenance. The newborn obviously has no age yet, but for parents we need to describe their age. We do so with the specialised PiCo ontology, as it is not available in Schema.org. Next we describe the relations between the newborn and his parents using sdo:parent to state that person observations 2 and 3 are the parents of person observation 1, sdo:children to indicate that person observation 1 is the child of person observations 2 and 3, and sdo:spouse to note that person observation 2 and 3 are married. Finally, we note what roles the person observations have in the source. These are newborn for person observation 1 and informant for person observation 2. The role of the mother is mentioned on the birth certificate, but is redundant as it is already implicitly modelled via the relation sdo:parent. The role property is therefore omitted from the person observation of the mother.

In PiCo, this would look like this:

Code block 1 *Observations around Abe Bos on his birth certificate, in PiCo format*[11]

```
afr:geboorteregister_1858_po_1
        a                       pico:PersonObservation ;
        prov:hadPrimarySource   afr:geboorteregister_1858 ;
        sdo:name                "Abe Bos" ;
        sdo:familyName          "Bos" ;
        sdo:givenName           "Abe" ;
        sdo:birthDate           "21 Junij 1858"^^xsd:string ;
        sdo:birthDate           "1858-06-21"^^xsd:date ;
        sdo:birthPlace          "Joure" ;
        sdo:gender              sdo:Male ;
        pico:hasRole            picot:roles/575 ; # newborn
        sdo:parent              afr:geboorteregister_1858_po_2, afr:geboorteregister_1858_po_3 .

afr:geboorteregister_1858_po_2
        a                       pico:PersonObservation ;
        prov:hadPrimarySource   afr:geboorteregister_1858 ;
        sdo:name                "Sjouke Abes Bos" ;
        sdo:familyName          "Bos" ;
        sdo:givenName           "Sjouke" ;
        pico:hasAge             "30"^^xsd:decimal ;
        sdo:hasOccupation       "klokmakersknecht" ;
        sdo:address             "Joure" ;
        sdo:gender              sdo:Male ;
        sdo:spouse              afr:geboorteregister_1858_po_3 ;
        sdo:children            afr:geboorteregister_1858_po_1 ;
        pico:hasRole            picot:roles/489 . # informant

afr:geboorteregister_1858_po_3
        a                       pico:PersonObservation ;
        prov:hadPrimarySource   afr:geboorteregister_1858 ;
        sdo:name                "Geertruida van der Wijk" ;
        sdo:familyName          "van der Wijk" ;
        sdo:givenName           "Geertruida" ;
```

11 To improve readability of the examples, some non-existent prefixes and URIs are used. See the Appendix for an overview of prefixes used in the examples.

```
        sdo:additionalName      [ a                   pnv:PersonName ;
                                  pnv:literalName     "Geertruida van der Wijk" ;
                                  pnv:baseSurname     "Wijk" ;
                                  pnv:surnamePrefix   "van der" ; ] ;
        sdo:address             "Joure" ;
        sdo:spouse              afr:geboorteregister_1858_po_2 ;
        sdo:children            afr:geboorteregister_1858_po_1 ;
        sdo:gender              sdo:Female .

afr:geboorteregister_1858
        a                       sdo:ArchiveComponent ;
        sdo:additionalType      picot:sourcetypes/551 ;  # birth certificate
        sdo:name                "Geboorteregister 1858, archiefnummer 30-16, Burgerlijke Stand
                                Haskerland - Tresoar, inventarisnummer 1020, blad 051" ;
        sdo:holdingArchive      <https://www.tresoar.nl/> ;
        sdo:url                 <https://allefriezen.nl/zoeken/deeds/f7e82c08-47c5-de67-c636-
                                26aa1603f6c9> ;
        sdo:dateCreated         "1885"^^xsd:gYear ;
        sdo:locationCreated     "Haskerland" .
```

5.2 AN EXAMPLE OF A PERSON RECONSTRUCTION: COMBINING PERSON OBSERVATIONS

The child in the example above, is usually mentioned in other historical person records, for example his marriage certificate:

Code block 2 *Observations around Abe Bos on his marriage certificate, in PiCo format*

```
nha:huwelijksakte_1885_321_po_1
        a                       pico:PersonObservation ;
        prov:hadPrimarySource   nha:huwelijksakte_1885_321 ;
        sdo:name                "Abe Bos" ;
        sdo:familyName          "Bos" ;
        sdo:givenName           "Abe" ;
        pico:hasAge             27 ;
        sdo:birthPlace          "Joure" ;
        sdo:hasOccupation       "agent van politie" ;
        sdo:gender              sdo:Male ;
        pico:hasRole            picot:roles/574 ;  # spouse
        sdo:spouse              [ rdf:value           nha:huwelijksakte_1885_321_po_2 ;
                                  sdo:startDate       "1885-11-11"^^xsd:date ; ] ;
        sdo:parent              nha:huwelijksakte_1885_321_po_3 , nha:huwelijksakte_1885_321_po_4 .

nha:huwelijksakte_1885_321_po_3
        a                       pico:PersonObservation ;
        prov:hadPrimarySource   nha:huwelijksakte_1885_321 ;
        sdo:name                "Sjouke Abes Bos" ;
        sdo:familyName          "Bos" ;
        sdo:givenName           "Sjouke" ;
```

```
        sdo:additionalName      [ a                 pnv:PersonName ;
                                  pnv:literalName   "Sjouke Abes Bos" ;
                                  pnv:patronym      "Abes" ; ] ;
        sdo:gender              sdo:Male ;
        sdo:children            nha:huwelijksakte_1885_321_po_1 .

nha:huwelijksakte_1885_321_po_4
        a                       pico:PersonObservation ;
        prov:hadPrimarySource   nha:huwelijksakte_1885_321 ;
        sdo:name                "Geertruida van der Wijk" ;
        sdo:familyName          "van der Wijk" ;
        sdo:givenName           "Geertruida" ;
        sdo:additionalName      [ a                 pnv:PersonName ;
                                  pnv:literalName   "Geertruida van der Wijk" ;
                                  pnv:baseSurname   "Wijk" ;
                                  pnv:surnamePrefix "van der" ; ] ;
        sdo:gender              sdo:Female ;
        sdo:children            nha:huwelijksakte_1885_321_po_1 .
```

A genealogist or scholar can decide to combine both person observations into a single person reconstruction via either manual or automatic matching. The person reconstruction summarises information from the underlying person observations, such as the name, date of birth, date of marriage, and place of birth. There are no discrepancies between person observations to be reconciled, but the disambiguation could be removed from the place name by replacing the place name and occupational strings with a URI. Defining "Joure" as geonames:2753197 uniquely identifies it as a town in the municipality of Haskerland, located within the province of Friesland, the Netherlands. Similarly, "agent van politie" is defined as wikidata:Q384593 to make the title available in multiple languages and provide background information.

To make person reconstructions reproducible, PiCo requires a description of who made the person reconstruction and from which person observations. Important properties are incorporated from the World Wide Web Consortium's (W3C) PROV ontology. prov:wasDerivedFrom states from which observations a reconstruction is derived. In this case certificates from the repositories from the Noord-Hollands Archief and Alle Friezen. prov:wasGeneratedBy in turn states that the person reconstruction was made by the CBG as part of reconstruction_activity_01. A reconstruction is incomplete without these two properties, as prov:wasDerivedFrom is required to check the quality of the provided reconstructions and prov:wasGeneratedBy to contact the provider.

The result would look like this:

Code block 3 *Abe Bos' person reconstruction in PiCo format*

```
cbg:person_reconstruction_1
        a                       pico:PersonReconstruction ;
        sdo:name                "Abe Bos" ;
        sdo:familyName          "Bos" ;
        sdo:givenName           "Abe" ;
        sdo:gender              sdo:Male ;
        sdo:hasOccupation       wd:Q384593 ;
        sdo:birthPlace          geonames:2753197 ;
        sdo:birthDate           "1858-06-21"^^xsd:date ;
        sdo:spouse              [ rdf:value         cbg:person_reconstruction_2 ;
```

```
                         sdo:startDate        "1885-11-11"^^xsd:date ; ] ;
     prov:wasDerivedFrom      nha:NL-HlmNHA_1503_537_IVa_63_po_1 ,
                              afr:geboorteregister_1858_po_1 ;
     prov:wasGeneratedBy      cbg:reconstruction_activity_01 .

cbg:reconstruction_activity_01
     a                        prov:Activity ;
     prov:wasAssociatedWith   cbg:reconstruction_agent_01 ;
     prov:startedAtTime       "2024-01-08T13:00:00"^^xsd:dateTime ;
     prov:endedAtTime         "2020-01-08T14:00:00"^^xsd:dateTime .

cbg:reconstruction_agent_01
     a                        prov:Agent, sdo:Person ;
     sdo:name                 "Ivo Zandhuis" ;
     sdo:sameAs               <https://orcid.org/0000-0003-0165-8341> .
```

5.3 OTHER USE CASE: DESCRIBING SLAVE REGISTERS

Pico is meant to describe the core demographic information from historical person records, as the endless variation in types of historical person records makes it next to impossible to include all the properties in one data model. A much more feasible approach is to extend PiCo with a secondary schema or ontology. The structure of Linked Data makes it easy to add the required predicates to describe data in a standardised way. As an example we show the entry in a Surinamese slave register from 1841 with the fictitious sr-ont schema, although, ideally, such information is described using shared terminology within the field of slavery studies.

The Surinam slave registers describe the lives of enslaved persons from 1830 until abolition in 1863. The slave registers contain both known properties of a person observation, such as the name, gender and primary source, but also new properties that are specific to the source (van Galen et al., 2023). These concepts are not described in Schema.org or bio and require a domain-specific ontology from the field of slavery studies. The slave registers contain the slave name of the enslaved person, their owner, and information on whether a person died, was born, bought, sold, escaped, or manumitted. There is a distinction between "slave names" and sdo:familyName, as enslaved persons were forbidden to have a family name, and only received them upon manumission. The researcher needs a domain-specific typology to define the properties of enslaved people.

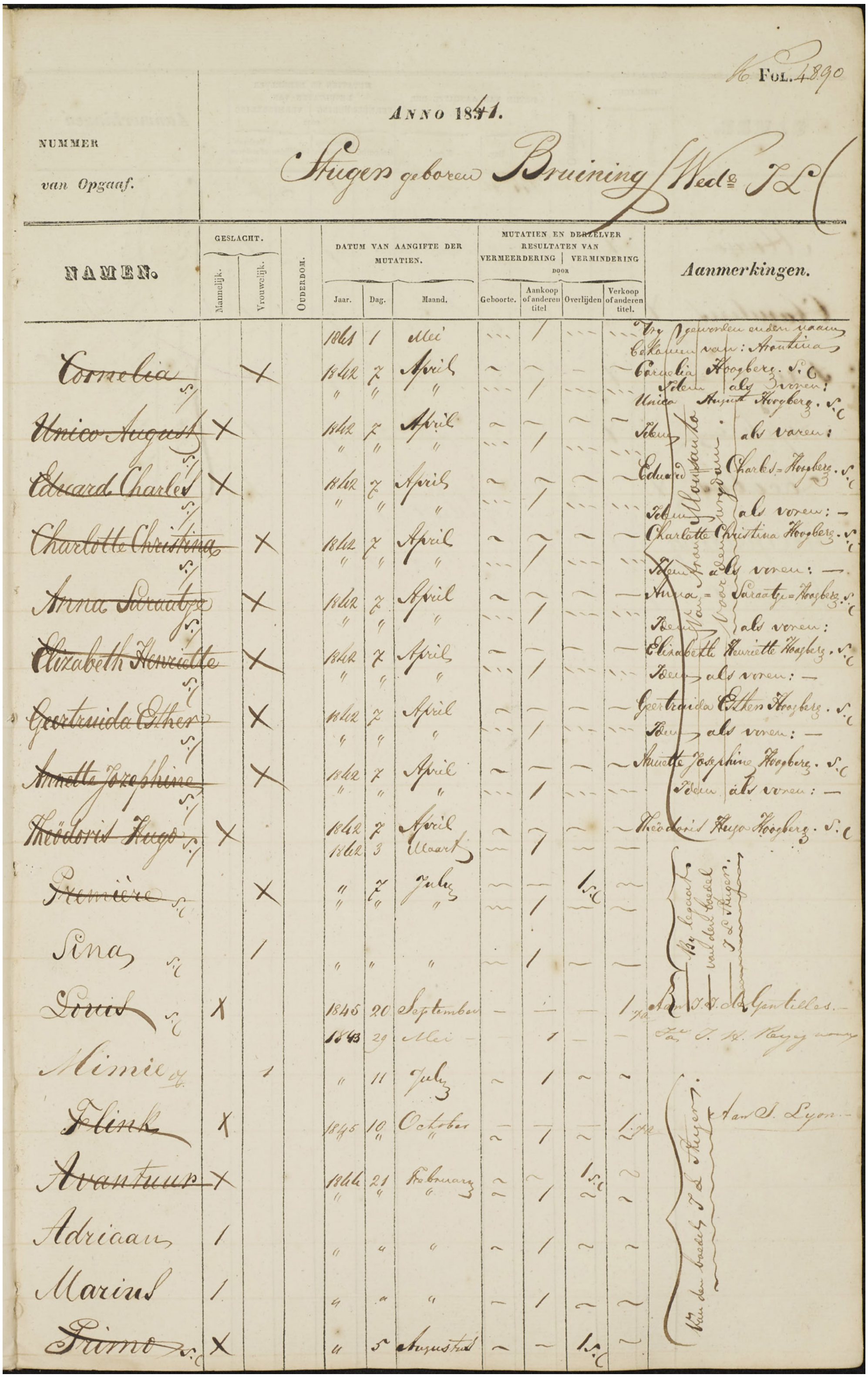
The example in code block 4 contains four person observations related to Unico August Hoogberg: the enslaved person himself, his owner, and his previous owner. Unico August Hoogberg was enslaved by the widow Stugen who bought Unico August on 1 May 1841 from Aron Monsanto and she manumitted Unico August almost a year later on 1842-04-07. We can describe Unico August Hoogberg's full name using sdo:name, sdo:familyName, and sdo:givenName. However, we need the domain-specific sr-ont to describe his enslaved name with sr-ont that we also use to describe his sr-ont:owners and sr-ont:manumission-date. In line with the PiCo standards, the date was standardised from "1842 | 7 | April" to ISO format: "1842-04-07". Finally, we specified that J.L. Stugen and his widow were married using sdo:spouse and note that mr Stugen is deceased with pico:deceased.

Code block 4 *Unico August Hoogberg's slave register entry in PiCo format*

```
nas:slavenregister_41_4890_po_1
     a                        pico:PersonObservation ;
     prov:hadPrimarySource    nas:slavenregister_41_4890 ;
     sdo:name                 "Unico August Hoogberg" ;
     sdo:familyName           "Hoogberg" ;
     sdo:givenName            "Unico August" ;
     sdo:gender               sdo:Male ;
```

```
        sr-ont:slave-name          "Unico August" ;
        sr-ont:wasEnslavedBy       [ rdf:value          nas:slavenregister_41_4890_po_4 ;
                                     sdo:endDate        "1841-05-01"^^xsd:date ; ] ;
        sr-ont:wasEnslavedBy       [ rdf:value          nas:slavenregister_41_4890_po_2 ;
                                     sdo:startDate      "1841-05-01"^^xsd:date
                                     sdo:endDate        "1842-04-07"^^xsd:date ; ] ;
        sr-ont:manumission-date    "1842-04-07"^^xsd:date .

nas:slavenregister_41_4890_po_2
        a                          pico:PersonObservation ;
        prov:hadPrimarySource      nas:slavenregister_41_4890 ;
        sdo:name                   "Wed. J.L. Stugen geboren Bruining" ;
        sdo:familyName             "Bruining" ;
        sdo:spouse                 nas:slavenregister_41_4890_po_3 ;
        sdo:gender                 sdo:Female .

nas:slavenregister_41_4890_po_3
        a                          pico:PersonObservation ;
        prov:hadPrimarySource      nas:slavenregister_41_4890 ;
        sdo:name                   "J.L. Stugen" ;
        sdo:familyName             "Stugen" ;
        sdo:givenName              "J.L." ;
        sdo:spouse                 nas:slavenregister_41_4890_po_2 ;
        pico:deceased              true ;
        sdo:gender                 sdo:Male .

nas:slavenregister_41_4890_po_4
        a                          pico:PersonObservation ;
        prov:hadPrimarySource      nas:slavenregister_41_4890 ;
        sdo:name                   "Aron Monsanto" ;
        sdo:familyName             "Monsanto" ;
        sdo:givenName              "Aron" ;
        sdo:gender                 sdo:Male .

nas:slavenregister_41_4890
        a                          sdo:ArchiveComponent, prov:Entity ;
        sdo:name                   "Nationaal Archief Suriname, Slavenregister, inv.nr. 41, fol. 4890" ;
        sdo:additionalType         "Slavenregistratie"@nl, "Slave Registration"@en ;
        sdo:locationCreated        "Suriname" ;
        sdo:dateCreated            "1841"^^xsd:gYear ;
        sdo:url                    "https://www.nationaalarchief.nl/onderzoeken/archief/Nadere
                                   Toegangen/invnr/NT00461.41/file/Slavenregister_InvNr_41_FolioNr_
                                   4890a" ;
        sdo:holdingArchive         <https://nationaalarchief.sr/> .
```

Figure 5 *Entry from the Surinamese slave registers*



16 Fol. 4890

ANNO 1841.

NUMMER van Opgaaf.

Stuger geboren Bruining / Wed.e J.L.

NAMEN.	Geslacht: Mannelijk	Geslacht: Vrouwelijk	Ouderdom	Datum van aangifte der mutatien: Jaar	Dag	Maand	Mutatien en derzelver resultaten van vermeerdering door: Geboorte	Aankoop of anderen titel	Vermindering door: Overlijden	Verkoop of anderen titel	Aanmerkingen
				1841	1	Mei		/			
Cornelia		×		1842	7	April					Cornelia Hoogberg. S.C
Unico August	×			1842	7	April					Unico August Hoogberg. S.C
Edward Charles	×			1842	7	April					Edward Charles Hoogberg. S.C
Charlotte Christina		×		1842	7	April					Charlotte Christina Hoogberg. S.C
Anna Saraatje		×		1842	7	April					Anna Saraatje Hoogberg. S.C
Elizabeth Henriette		×		1842	7	April					Elizabeth Henriette Hoogberg. S.C
Geertruida Esther		×		1842	7	April					Geertruida Esther Hoogberg. S.C
Annetta Josephine		×		1842	7	April					Annetta Josephine Hoogberg. S.C
Theodoris Hugo	×			1842 / 1842	7 / 3	April / Maart					Theodoris Hugo Hoogberg. S.C
Première		×		"	7	July		/			
Sina		/		"	"	"		/			
Louis	×			1846 / 1843	20 / 29	September / Mei					
Mimie		/		"	11	July		/			
Flink	×			1845	10	October		/			
Avantuur	×			1844	21	February		/			
Adriaan	/			"	"	"		/			
Marinus	/			"	"	"		/			
Primo	×			"	5	Augustus					

*Note: Entry in the slave register series 2 for the private slave owner widow J.L. Stuger née Bruining. At the top, the name of the slave owner and the page number (*folio number*) are mentioned, and the year this record started. Underneath from left to right, there are columns for names of enslaved, sex, date of mutation, the most common types of mutations (birth, purchase, death, sale) and remarks. People sold by the widow Stuger were struck off this entry and transferred to the entry of the new owner. Because information on enrollment and deregistration had to be entered on the same line in the register, the entries could easily become cluttered, as this example shows. For this reason, the registers were renewed every 4 to 10 years. This specific type of register was in use between 1838 and 1848 (NAS inventory number 41, folio 4890).*

6 DISCUSSION AND CONCLUSION

Historians have been working with a wide range of historical sources to reconstruct persons from historical person observations. Individual and collective efforts of historians and genealogists have given way to a plethora of reconstituted life courses and families. Yet, the lack of standardisation in the description of historical person data has hurt the interoperability and sustainability of both small and large databases. Without proper standardisation, it is next to impossible to take stock of all available historical person data, which obscures the visibility of all but the largest databases. Moreover, it makes it unnecessary labour-intensive to follow individuals between databases, either to follow migration between contexts or add detailed information to larger demographic databases.

PiCo introduces a standard for historical person data, which bridges different types of historical sources containing person data. Since its introduction in late 2023, PiCo has been implemented by archival institutions such as Het Utrechts Archief, is actively provided by genealogical platforms such as open archives initiative openarchieven.nl, recommended as a best practice by the Dutch Digital Heritage Network for GLAM institutions, and being used by historical expertise centres like WO2NET, the International Institute of Social History, and the Historical Database for Suriname and Curacao. Hence, the PiCo data model is rapidly becoming the de facto standard for exchanging historical person data within the Netherlands, as it allows for efficient communication between databases and stimulates reuse of software. Moreover, because historical person data is more standardised, it is easier to write software that extracts information from different types of sources, such as civil certificates, population registers, and slave registers. As such, PiCo moves beyond A2A, the previous Dutch national standard for describing historical person data, as well as existing schemas within historical demography, such as IDS, IPUMS, LINKS-gen, and NAPP.

By using internationally accepted Linked Data standards to describe person characteristics, historical person data become entangled in the "network of data". This makes it possible for genealogists and researchers to find historical person observations as well as reconstructions and their origin without a centralised website. Building a decentralised network of data ascertains that all historical data is only "one click away", as long as person observations and person reconstructions refer to one another. As a result, scans of the historical person records on enslaved persons hosted by the National Archives, person observations in a dataset hosted by the Radboud University, and person reconstructions at the CBG are interoperable.

Reusing existing standards also helps to focus discussions on data standardisation within historical demography and adjacent disciplines. PiCo discusses historical person data concentrically, which means that it reuses (parts of) global schemas to describe general concepts known by almost everyone, domain-specific schemas to describe concepts that are known to work in a domain like data science, and specialised vocabularies to describe concepts that are unique to historical person reconstructions. This guides discussions on how to describe things within historical demography towards the topics that have yet to be defined. The required general schemas and domain-specific schemas to describe historical person data are currently available, but some specialised vocabularies to describe historical person data are still missing. Here PiCo has developed its own terminology.

Not all variables on historical person documents are defined in PiCo, which is on purpose. The data model is designed to be as flexible and straightforward as possible and allows data providers to describe their core historical person data clearly and concisely. Rather than trying to be as comprehensive as possible, PiCo is easily extendable with other schemas. This gives data providers extensive flexibility in how they describe their data, as long as the core information on sources, person observations, and person reconstructions is (also) modelled according to the PiCo data model. For example, data providers can, for example, use specialised vocabularies to describe manumissions and transfers within slave registers. Over time, such extensions to PiCo will be encouraged as best-practices or even become part of PiCo schema if they are widely adopted.

PiCo contains two design choices to keep it straightforward in its use. The most obvious is the distinction between person observations and person reconstructions. Person observations describe unique entries in the sources, whereas person reconstructions state which person observations belong to the same individual. The strict distinction between observations and reconstructions allows us to add provenance on the actual action of grouping some observations: by whom, when and why. The second one helps to make historical person reconstructions by discerning between relations and roles to describe the relations between person observations. Relations refer to a standardised list of associations between person observations, most commonly child, parent or spouse, whereas roles state the link between a person observation and a historical person record, such as informant, witness, or father of the bride. This distinction makes it possible to model overlap between relations and roles, such as informants on birth certificates who are also the father, as well as important divergences, such as a father of the bride who is not the biological father.

Both distinctions are paramount for efficient querying and enable a new generation of software. The Linked Data technology that captures person observations can be used to derive additional links based on the properties. For instance, if a person observation states that person 1 and 2 siblings and in another observation we see that person 3 is the sibling of person 2, an additional relation can be inferred: 1 is the sibling of 3. This is called *knowledge inference* and gives way to automatic creation of person reconstructions. A similar process is used to automatically reconstruct life courses and families. Software to reconstruct historical persons can ingest any kind of historical person data that adheres to the PiCo data model. This will increase both the user-friendliness as well as the types of historical sources that can be processed with the matching program. Moreover, the necessary information to indicate the quality of person reconstructions is readily available, as PiCo states what metadata about the reconstruction process should be provided.

However, despite all the technical advantages made by PiCo, we should never forget that its success is dependent on community uptake. In the Netherlands alone, scores of volunteers have transcribed historical person observations, made person reconstructions, and attached those to secondary sources. Archival institutions organise persistent publication of this data and its relation to the original source. At the other end of the chain, scholars add value by means of aggregations and analysis. PiCo can serve as the basis for an ecosystem in which these persons and institutions interact and exchange information by having a clear design that makes communication more intuitive. Discerning between an observation in a source and the actual person behind it is not only straightforward and stringent, it most importantly fits how we as humans think and speak. This is paramount as it is the community that supports PiCo and not vice versa. Their combined effort generates data, maintains collections, and provides software for others to use. All in all, PiCo makes communication and processing of historical person data easier and we invite you to be part of this national and international community.

REFERENCES

Alter, G., & Mandemakers, K. (2014). The Intermediate Data Structure (IDS) for longitudinal historical microdata, version 4. *Historical Life Course Studies, 1*(1), 1–26. https://doi.org/10.51964/hlcs9290

Davis, I., & Galbraith, D. (2004). *BIO: A vocabulary for biographical information*. vocab.org. https://vocab.org/bio

den Engelse, M., & van Wissen, L. (2019). *Reconstructions and observations in archival resources: Ontology specification*. https://w3id.org/roar

Durkheim, É. (1897). *Le suicide: Étude de sociologie* [Suicide: A study in sociology]. Félix Alcan.

Edvinsson, S., Mandemakers, K., Smith, K. R., & Puschmann, P. (2023). *Harvesting: The result and impact of research based on historical longitudinal databases*. Radboud University Press. https://doi.org/10.54195/HYLR8777

Fauve-Chamoux, A., Bolovan, I., & Sogner, S. (2016). *A global history of historical demography: Half a century of interdisciplinarity*. Peter Lang.

Gao, S. S., Sperberg-McQueen, C. M., & Thompson, H. (2012). *W3C XML schema definition language (XSD) 1.1 Part 1: Structures.* W3.org. https://www.w3.org/TR/2012/REC-xmlschema11-1-20120405/

Geschwind, A., & Fogelvik, S. (2000). The Stockholm historical database. In P. K. Hall, R. McCaa & G. Thorvaldsen (Eds.), *Handbook of international historical microdata for population research* (pp. 207–230). Minnesota Population Center. https://international.ipums.org/international/resources/microdata_handbook/1_12_sweden_stockholm_ch13.pdf

Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema.org: Evolution of structured data on the web. *Communications of the ACM, 59*(2), 44–51. https://doi.org/10.1145/2844544

Henry, L., & Fleury, M. (1956). *Des registres paroissiaux à l'histoire de la population. Manuel de dépouillement et d'exploitation de l'état civil ancien* [From parish registers to population history. Manual for processing and analyzing the old vital records]. Institut national d'études démographiques.

Hoekstra, R., Meroño-Peñuela, A., Zijdeman, R. L., Rijpma, A., Dentler, K., & Zandhuis, I. (2016). An ecosystem for linked humanities data. In A. Adamou, E. Daga & L. Isaksen (Eds.), *1st Workshop on Humanities in the Semantic Web (WHiSe)* (pp. 85–96). CEUR Workshop Proceedings. https://doi.org/10.1007/978-3-319-47602-5_54

Idrissou, A., van Wissen, L., & Zamborlini, V. (2022). The lenticular lens: Addressing various aspects of entity disambiguation in the semantic web. *Graphs and Networks in the Humanities*.

Knodel, J. E. (2002). *Demographic behavior in the past: A study of fourteen German village populations in the eighteenth and nineteenth centuries*. Cambridge University Press.

Lebo, T., Sahoo, S., McGuinness, D., Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. (2013). *PROV-O: The PROV ontology* (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/prov-o/

Lee, J. Z., & Campbell, C. D. (2016). *China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN), 1749–1909* [Data set]. Inter-university consortium for political and social research. https://doi.org/10.3886/ICPSR27063.v10

Mandemakers, K. (2023). "*You really got me*". *Ontwikkeling en toekomst van historische databestanden met microdata* [Development and future of historical databases with microdata]. Erasmus University Rotterdam. https://doi.org/10.25397/eur.23256467

Meroño-Peñuela, A., de Boer, V., van Erp, M., Zijdeman, R. L., Mourits, R. J., Melder, W., Rijpma, A., & Schalk, R. (2020). CLARIAH: Enabling interoperability between humanities disciplines with ontologies. In G. Cota, M. Daquino & G. L. Pozzato (Eds.), *Applications and practices in ontology design, extraction, and reasoning* (pp. 73–90). https://doi.org/10.3233/SSW200036

Minnesota Population Center. (2020). *Integrated public use microdata series, International* (Version 7.3) [Data set]. IPUMS. https://doi.org/10.18128/D020.V7.3

Mourits, R. J., van Dijk, I. K., & Mandemakers, K. (2020). From matched certificates to related persons. *Historical Life Course Studies, 9*, 49–68. https://doi.org/10.51964/hlcs9310

Mourits, R. J., Riswick, T., & Stapel, R. J. (2024). Common language for accessibility, interoperability, and reusability in historical demography. *AISoLA Conference Proceedings*. https://aisola.org/papers/mourits-riswick-stapel.pdf

Pujadas-Mora, J. M., Fornés, A., Ramos Terrades, O., Lladós, J., Chen, J., Valls-Fígols, M., & Cabré, A. (2022). The Barcelona Historical Marriage Database and the Baix Llobregat Demographic Database: From algorithms for handwriting recognition to individual-level demographic and socioeconomic data. *Historical Life Course Studies, 12*, 99–132. https://doi.org/10.51964/hlcs11971

Quetelet, A., & Smits, E. (1832). *Recherches sur la reproduction et la mortalité de l'homme aux différents âges et sur la population de la Belgique* [Inquiries upon human reproduction and mortality at different ages, and on the Belgian population]. Hauman.

Raad, J., Mourits, R., Rijpma, A., Schalk, R., Zijdeman, R., Mandemakers, K., & Meroño-Peñuela, A. (2020). Linking Dutch civil certificates. In A. Adamou, E. Daga & A. Meroño-Peñuela, *WHiSe 2020 Workshop on Humanities in the Semantic Web 2020* (pp. 47–58). CEUR-WS. https://ceur-ws.org/Vol-2695/paper6.pdf

Roberts, E., Ruggles, S., Dillon, L. Y., Gardarsdottir, Ó., Oldervoll, J., Thorvaldsen, G., & Woollard, M. (2003). The North Atlantic Population Project: An overview. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 36*(2), 80–88. https://doi.org/10.1080/01615440309601217

Ruggles, S., Fitch, C. A., Goeken, R., Hacker, J. D., Nelson, M. A., Roberts, E., Schouweiler, M., & Sobek, M. (2021). *IPUMS ancestry full count data* (Version 3.0) [Data set]. IPUMS. https://doi.org/10.18128/D014.V3.0

Ruggles, S., Roberts, E., Sarkar, S., & Sobek, M. (2011). The North Atlantic Population Project: Progress and prospects. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 44*(1), 1–6. https://doi.org/10.1080/01615440309601217

Smith, K. R., Fraser, A., Reed, D. L., Barlow, J., Hanson, H. A., West, J., Knight, S., Forsythe, N., & Mineau, G. P. (2022). The Utah Population Database: A model for linking medical and genealogical records for population health research. *Historical Life Course Studies, 12*, 58–77. https://doi.org/10.51964/hlcs11681

Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). John Churchill.

Song, X., & Campbell, C. D. (2017). Genealogical microdata and their significance for social science. *Annual Review of Sociology, 43*, 75–99. https://doi.org/10.1146/annurev-soc-073014-112157

SwedPop. (2024). *Version 3 of harmonised data in SwedPop* [Data set]. SwedPop. https://doi.org/10.48524/SWEDPOP-V3-0

Szołtysek, M., & Gruber, S. (2016). Mosaic: Recovering surviving census records and reconstructing the familial history of Europe. *The History of the Family, 21*(1), 38–60. https://doi.org/10.1080/1081602X.2015.1006655

Trepal, D., Lafreniere, D., & Stone, T. (2021). Mapping historical archaeology and industrial heritage: The historical spatial data infrastructure. *Journal of Computer Applications in Archaeology, 4*(1), 202–213. https://doi.org/10.5334/jcaa.77

van Galen, C. W., Mourits, R. J., Rosenbaum-Feldbrügge, M., A. B., M., Janssen, J., Quanjer, B., van Oort, T., & Kok, J. (2023). Slavery in Suriname: A reconstruction of life courses, 1830–1863. *Historical Life Course Studies, 13*, 191–211. https://doi.org/10.51964/hlcs15619

Vézina, H., & Bournival, J.-S. (2020). An overview of the BALSAC Population Database: Past developments, current state and future prospects. *Historical Life Course Studies, 9*, 114–129. https://doi.org/10.51964/hlcs9299

Wilkinson, M. D., Dumontier, M., Aalbersberg, IJ. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3, 160018.* https://doi.org/10.1038/sdata.2016.18

APPENDIX LIST OF PREFIXES USED IN THE EXAMPLES

PREFIXES OF EXISTING URLs

@prefix geonames: <https://geonames.org/> .

@prefix pico: <https://personsincontext.org/model#> .

@prefix picot: <https://terms.personsincontext.org/> .

@prefix pnv: <https://w3id.org/pnv#> .

@prefix prov: <http://www.w3.org/ns/prov#> .

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

@prefix sdo: <https://schema.org/> .

@prefix xsd: <https://www.w3.org/2001/XMLSchema#> .

@prefix wd: <https://wikidata.org/wiki/> .

PREFIXES USED FOR FICTIONAL URLs

@prefix afr: <https://allefriezen.nl/zoeken/> .

@prefix cbg: <https://data.cbg.nl/> .

@prefix nas: <https://data.nationaalarchief.sr/> .

@prefix nha: <https://www.noord-hollandsarchief.nl/> .

@prefix sr-ont: <https://ontology.nationaalarchief.sr/> .