The PersonLink project: How to deal with historical events

Gerard Kuys, June 2021

For our PersonLink project, we intend to make our linked data express no less than the (nature of the) involvement of a particular historical person with a particular event at a particular place.

For reasons of economy and data availability, we shall limit the scope of this project to only those events, that have been described by A.J. van der Aa in his early-19th-century work. Even for a location-oriented *Geographic Dictionary* like Van der Aa's, such events may turn out to be complicated enough. Therefore, if we strive to consistently describe events, we have to define the minimal set of requirements to do so. These requirements are, that we can structure our event data so, that it conveys the *who*, *where*, *when* and *what* of a historical event as described by a particular source.

The latter requirement – the one about the source – is of crucial importance. If at some stage of our project we would choose to go beyond the work written by Van der Aa, we would need to describe versions of one and the same event by different authors. We might even want to provide in our data for differing versions of a single event – alternative versions by one and the same author in the course of time, or by several authors simultaneously or over time. In the end, we might also want to compare accounts of events alleged to be the same event. Since this can be done only by way of looking into the possible causes and correlations of events, we need to be able to express in our data assertions about the relation(s) between one event and the other – up to a point where such assertions may even contradict one another. So, what we set out to be a minimum requirement for expressing event descriptions in a dataset, has grown into a heavy workload. An event model that is able to cope with this kind of requirements, would it really exist?

As stated by Willem Robert van Hage e.a. [1], event models can be classified by four main design choices: domain (in)dependency, model orientation (focus either on RDF classes or on RDF properties), expressiveness (a deep model for a limited knowledge domain vs. a shallower model for all kinds of event descriptions), and, finally, the degree of formalisation. Of these criteria, not all are equally relevant for our case. The event model we are looking for must enable us to describe a (change in a) state of affairs in human society at a given time or time interval. Moreover, we need to be able to lay down both alternative accounts of the same event (if any) and variations in the way some event is being described. The requirements for our project fit perfectly well into the formula of Van Hage e.a.: the 'model needs to represent not only the description of who did what, when and where, but also the roles that each actor played, the time during which a role is valid and the authority according to whom this role is assigned' [2].

Such convergence does not necessarily mean, however, that the subject of van Hage's *Web Semantics* article, the Simple Event Model (SEM), should also be the guiding model for our PersonLink case. Designing the Simple Event Model was motivated chiefly by the desire to accommodate SEM to the wild variety of event descriptions on the World Wide Web. With SEM, hopefully even the most exotic event renderings could be recognised as such, and at least partially be made 'intelligible' for computer systems. If anyhow possible, SEM should enable users to detect

1 Willem Robert van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, Guus Schreiber, *Design and use of the Simple Event Model (SEM)*, in: Web Semantics: Science, Services and Agents on the World Wide Web 9 (2011), p. 134.

2 Van Hage e.a., o.c., p. 128.

the right actors and the right location and time frame for anything digital that comes close to an 'event'. In order to do just that, SEM has been designed to be very forgiving with regard to modelling variations, and to the absence of constraints. As a consequence, SEM cannot commit to a specific definition of an event. SEM would be happy to reuse any event typing vocabulary that fits the dataset currently processed. Instead of defining event patterns by itself, SEM provides placeholders for all types of – externally defined – event types, actor roles, locations, locations' roles and the like. SEM's designers must have realised, that the more interoperability you want, the less meaning you get. Probably that is why they have put their hopes on a Prolog API in order to to compensate by reasoning what's been left undecided in SEM's data structures [3].

For better or for worse, interoperability is not the main concern we have in our PersonLink project. Here, we have a strictly defined set of events – the locality-bound events as described by a single author – within the domain of the Netherlands' national history (including its colonies) well into the 1840's. In order to connect one strictly circumscribed data source to another, we need to be as specific as possible, so that we can tell – the example is just for the sake of argument – whether or not the siege of Grol in 1595 as described to by A.J. van der Aa could be classified to be in the same frame of events as the conquest of Groenlo in 1597 as depicted by Jan Wagenaar in Part IX of his 18th-century work of historiography [4]. Instead of by maximal interoperability, we in our project would be better served by reliable mappings between accounts of events as provided by unambiguously identifiable sources.

Obviously, the Simple Event Model has clear advantages, such as roles not just for persons but also for objects. Also, it offers the feature of properties that are qualified by precise or imprecise indications of time. Unfortunately, however, at least for our PersonLink project, SEM refrains from the inclusion of participation modalities (like: the way some Role was exercised by an Agent participating in an Event), or a model of causality. Likewise, SEM leaves any typology of events in its event model to external classifiers. Beware, these are just the features we desperately need for our project. It is the outspoken intent of the designers of SEM 'to investigate, how other event models like GEM or F, can be used in combination with SEM to model other event properties like for example causality and correlation' [5]. It would be great to know, but just the promise of such an investigation does not help us find the most suitable event model for our project right now.

Leaving aside the event-related part of CIDOC-CRM – about which we agree with the designers of SEM that it is too much oriented towards the description of collectioners' artefacts – we now will have a look at the features of Event-Model-F [6]. 'F' is a model of events that is based on DOLCE+DnS Ultralite (DUL), and has been refactored to fit within the Ontology Design Patterns (ODP) extensions to DUL. This refactoring has the distinct advantage, that we could tailor specific historical subjects to specific areas of historical knowledge, while retaining the general framework of DUL. In this way, we get all the advantages of DUL as an upper ontology: its inherent connection of Agents and Objects with Time, its Quality feature meant to extend the description of properties, and the use of value spaces, for instance the one that contains one or more taxonomies of event types. Moreover, the DnS pattern, which offers a structure for distinguishing events and event descriptions, is central to DUL [7] – so that we can have diverging descriptions of the same set of events. The particular perspective the 'F' event vocabulary is meant to add to DUL, is the use of

3 Van Hage e.a., o.c., p. 129.

4 If all data were correct, such classification would depend on the granularity by which events are being defined: are we looking at the individual feat or at the general course of the Dutch Independence War/80-years War.

5 Van Hage e.a., o.c., p.135.

6 Ansgar Scherp, Thomas Franz, Carsten Saathoff, Steffen Staab, *F – A model of events based on the foundational ontology DOLCE+DnS Ultralite*, Conference Paper January 2009, with a 2014 update in: ResearchGate.net/publication/220916829.

different, contextualised views on events – thus opening the door to expressing important qualifications about events. Such qualifications may involve patterns expressing types of (time-indexed) participation, causality, mereology (especially inclusion relations), and, as stated, of diverging interpretations [8].

The nice thing about linked data/RDF is, that one approach about events need not exclude the other. Therefore, we decided to implement both SEM and 'F' in our linked-data version of our source text, in order to have the best perimeters for comparison. As we believe, the relative advantages of SEM versus DUL can only be explored against a considerable part of A.J. van der Aa's body of text. The results so far show that, whereas SEM is the more concise model, 'F' can be more expressive.

As a matter of fact, both models share the same framework for representing the basic features of an event. Like, for instance, the events reported by A.J. van der Aa related to the early development of the Twente textiles trade (Table 1).

Waar	***Wanneer***	***Wie***	***Wat voor geb.***	***Gebeurtenis***	***Rol in geb.***
Goor	1833	Thomas Ainsworth	Oprichting school	Thomas Ainsworth experimenteert in Goor met het weven van calicot en richt een weefschool op	Oprichter
Nijverdal	1836	Thomas Ainsworth	Oprichting bedrijf	Thomas Ainsworth richt een modelweverij op	Oprichter
Goor	1841	G.C. Arntzenius	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Kopende partij
Diepenheim	1841	G.C. Arntzenius	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Kopende partij
Hulten (NB)	1841	G.C. Arntzenius	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Kopende partij
Enter	1841	G.C. Arntzenius	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Kopende partij
Hulten (NB)	1841	Ned. Handel Mij.	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Verkopende partij
Diepenheim	1841	Ned. Handel Mij.	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Verkopende partij
Enter	1841	Ned. Handel Mij.	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Verkopende partij
Goor	1841	Ned. Handel Mij.	Transactie	G.C. Arntzenius koopt weefstoelen van de Nederlandsche Handel Mij.	Verkopende partij
Huis De Eversberg	13-02-1841	Thomas Ainsworth	Overlijden	Thomas Ainsworth sterft onverwacht	Overledene

Table 1: Basic modelling of events in the early Calicot textiles industry in Twente, according to A.J. van der Aa's *Geographic Dictionary*

The devil between SEM and 'F' is, however, in the details. For any relevant source text, we can, with the support of 'F', combine event chains into DUL Situations, and in this way also construct divergent accounts of events. Such accounts may even use the same event data, and only present varying sequences of these events, depending on the views of the author of an event description. The same thing can be done in SEM with *sem:accordingTo*, albeit that SEM has no notion of, and therefore no class for event clusters. That is why, in SEM, we are compelled to construct different

7 Valentina Presutti, Aldo Gangemi, *Dolce + DnS Ultralite and its main ontology design patterns*, in: Pascal Hitzler e.a., *Ontology Engineering with Ontology Design Patterns. Foundations and Applications*, Berlin-Amsterdam (IOS Press) 2016, p. 81-104.

8 Scherp e.a., o.c., p. 141. Thanks to Ansgar Scherp (University of Ulm) for sharing with us the OWL code of 'F'.

event views on the level of the single event in stead of connecting several events one to another. More often than not, this is beside the mark and cumbersome. For the ‘F’ model, on the other hand, we have a Situation class grouping events for us, together with a Description that may give a particular slant to the factual events within a Situation. Depending on whether we selected Van der Aa as a source or another account of the same event chain, queries on the ‘F’ model may produce different sequences of events, and of slightly varying sets of the actors involved. But also, we may encounter one and the same Event playing a role in various Situations.

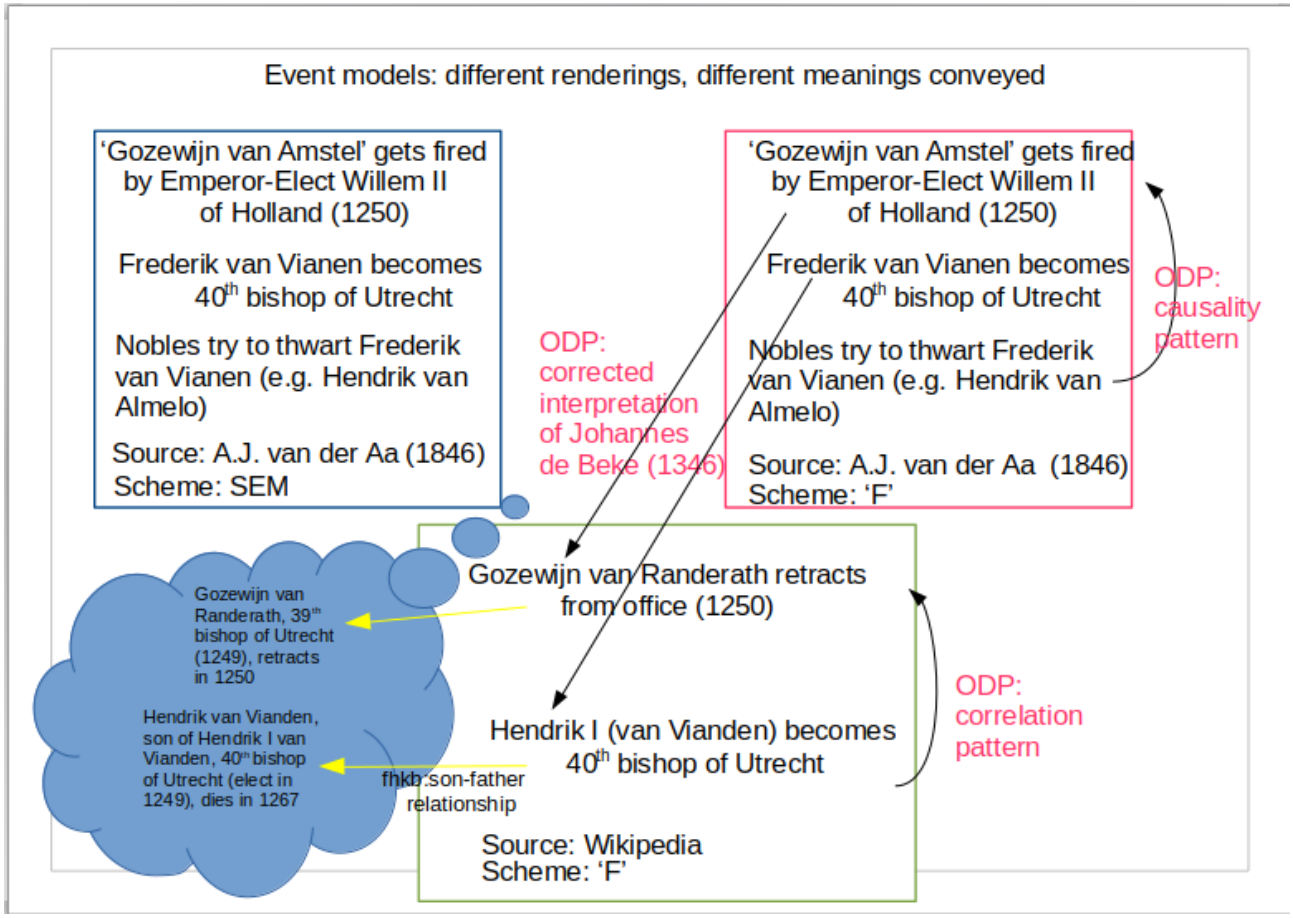

Diagram 1: Representing historical situations in sources, with SEM and 'F'

Equally, with ‘F’, we can construct causality chains, or compare interpretations, as visualised by Diagram 1 above. The rectangles indicate the rendering of an event according to a particular source expressed in a particular event vocabulary, while the cloud refers to the state of affairs according to some authority.

Van der Aa’s *Geographic Dictionary* is a work on locations, and, by consequence, it has to mention frequently the rights on landed property, both feudal and by transaction. Among the causes of shifting patterns of ownership encountered in Van der Aa’s text so far, we find hereditary succession, as well as the development policies pursued by the Nederlandsche Handel Mij. in the 1830’s. As we will work our way through Van der Aa’s text, we hope to get a more complete set of causes for events. These may be related to event types, and hopefully will turn out to be sufficient for a provisional classification – a taxonomy of event types as far as A.J. van der Aa’s domain is concerned.

Finally, a word about the evolution of historical objects in the course of time, and how to represent that evolution in data structures. In our earlier project outline document, we rejected the finite state models currently being used to express changes of objects in the course of time. Rather, we felt that ‘almost all entity types are subject to the impact of time, and have a life cycle of their own. And not just these entity types in isolation, but also the relations they have with their surrounding world, and the time-dependent typologies they refer to’. Vaguely we were thinking about some form of ‘morphing’ the central objects within our data models – where some essentialist core was thought to remain the same while most of its properties were subject to change.

However, when exploring the data as we found them in Van der Aa’s corpus of text, we gradually came to the conclusion that we would have to drop this image of ‘morphing’. First of all, as a work ordering rather than exploring knowledge, our primary source, Van der Aa’s *Geografisch Woordenboek,* hardly provides information about the inner development of places, persons or objects. Next, as far as we know, there is no linked data vocabulary handling the kind of ‘morphing’ we were looking for, least of all the finite state-oriented models like the Peircean approach we discussed in our project outline document.

But all this time, the solution had been staring us in the face. Already, we had taken notice of the advantages of the DOLCE + DnS (DUL) upper ontology and its ODP extensions, for instance like it has been used in the ‘F’ event model. When digging deeper, we found that this DUL upper ontology provides the framework for the representation of time that we were in need of. DUL is inherently time-oriented. While the old distinction between Endurants and Perdurants has been redefined and renamed into the ‘Lite’ness of DUL – in order to better connect to linked-data ways of thinking – the underlying concepts still are there. So, rather than delving ever deeper into the mystifying layers of historical objects to be ‘morphed’, we’d better apply as good as we can the ideas behind DUL. Already this approach proved to be beneficial when using to our benefit the ODP of the Time-Indexed Participations in the event model we adapted. Therefore we decided, at least for the duration of our PersonLink project, to go DUL and ODP all the way. If this decision will live up to its promise, we should be able to model events in two ways we could not model adequately before: 1) events which with every step grow ever more complex, and 2) events that seem or really are different in the eye of the beholder – facts and interpretations about having lived – or not – in the same reality. How far we got in doing so will be the subject of our next project document.