

# La Modellazione Diacronica di Risorse Termine-Ontologiche nell'Ambito delle Digital Humanities: Esperimenti su Clavius

Silvia Piccini, Andrea Bellandi, Giulia Benotto, Emiliano Giovannetti

Istituto di Linguistica Computazionale A. Zampolli

Via G. Moruzzi, 1 - CNR Pisa, Italy

name.surname@ilc.cnr.it

## Abstract

**English.** In this work, we present an experiment in the modeling of a diachronic termino-ontological resource named CLAVIUS through both the N-ary relations model and the 4D-fluents approach. Some of the salient differences of these two models are discussed. The overall objective of this research is to illustrate the main advantages and disadvantages in the adoption of a given model to build diachronic resources.

**Italiano.** In questo lavoro, si illustra un esperimento di modellazione di una risorsa termino-ontologica diacronica (CLAVIUS) secondo due approcci, quello N-ario e quello dei 4D-fluents. Le differenze salienti dei due approcci verranno presentate e discusse. L'obiettivo generale della ricerca qui introdotta è quello di mostrare i principali vantaggi e svantaggi che l'adozione di un determinato modello può comportare nella modellazione di risorse diacroniche.

## 1 Introduzione

*Panta rei* è la celebre espressione attribuita da Platone ad Eraclito. Tutto è sottoposto alla inesorabile legge del mutamento: la realtà, le categorie attraverso le quali la organizziamo e le parole che usiamo per parlare di essa.

Quali sono gli strumenti a disposizione dell'umanista digitale di oggi, che si trovi a dover rappresentare in modo esplicito e formale tale evoluzione diacronica dei concetti e dei termini in un determinato ambito, in modo che tale formalizzazione sia computabile ad un calcolatore?

In questi ultimi anni, ed in particolar modo nell'ambito delle *Digital Humanities*, si è sottolineata l'importanza di operare con tecnologie che siano alla base del *Semantic Web* e dei *Linked Open Data* per garantire interoperabilità e riuso delle risorse all'interno della comunità scientifica (Ciotti, 2014).

In questa ottica, le ontologie - e l'OWL, il loro linguaggio di rappresentazione standard - giocano un ruolo fondamentale. Tuttavia, il carattere fondamentalmente statico di questi ultimi e la necessità di modellare aspetti di evoluzione temporale sembrano a prima vista inconciliabili.

Le riflessioni che presentiamo in questo articolo nascono dalle esperienze condotte in seno al Progetto *Clavius on the Web*<sup>1</sup>. Tra gli obiettivi del Progetto, infatti, vi è anche quello di creare una risorsa termino-ontologica (RTO) che rappresenti l'evoluzione delle teorie matematico-astronomiche dall'antichità al XVI - XVII secolo, così come viene descritta da Clavius nei suoi *Euclidis Elementorum Libri XV. Accessit XVI e In sphaeram Ioannis de Sacro Bosco Commentarius*.

## 2 Il Contesto

Come sottolineato nell'Introduzione, il linguaggio OWL (e la sua estensione OWL2) è lo standard W3C per la creazione e condivisione di ontologie nel Semantic Web.

In particolare, OWL DL implementa la logica descrittiva  $SHOIN(D_n)$ , che garantisce una maggiore espressività rispetto a RDF e RDFS, senza compromettere la decidibilità e il meccanismo inferenziale.

Tuttavia, OWL è un linguaggio statico; in esso le proprietà e le relazioni tra entità sono fondamentalmente binarie, espresse sotto forma di triple  $\langle \text{Subject predicate Object} \rangle$ . Tale restrizione sintattica rende più complessa la rappresentazione

<sup>1</sup><http://claviusontheweb.it> (ultimo accesso: 13/10/2016)

di entità temporali, caratterizzate da relazioni ternarie.

Diversi approcci sono stati proposti in letteratura per superare tale limite e rappresentare l'evoluzione diacronica dei concetti (Krieger, 2014; Welty et al., 2006). Tra questi vi sono: i) il versioning, che consiste nel creare una versione differente dell'ontologia, una per ciascun istante temporale (Grandi and Scalas, 2009), ii) l'estensione delle triple RDF con argomenti supplementari temporali (Krieger, 2012), iii) la reficazione, che trasforma le proprietà in classi (Manola and Miller, 2004), iv) il modello N-ario (Noy and Rector, 2006), v) il modello perdurantista o dei *4D-fluents* (Welty et al., 2006), vi) la codifica dell'estensione temporale attraverso nuove proprietà sintetiche (Gangemi, 2011). Allo stato dell'arte, i due principali approcci adottati nella modellazione di risorse diacroniche (solitamente ontologie) sono il modello delle relazioni N-arie e il modello dei perduranti. Nel primo, per ogni relazione temporale viene introdotto un nuovo oggetto, istanza della classe EVENT. La durata della relazione coincide con la durata dell'evento. In questo modello, pertanto, una proprietà n-aria è rappresentata come una classe, le cui istanze corrispondono alle istanze della proprietà. Nel modello *4D-fluents*, invece, le entità coinvolte in relazioni temporali sono rappresentate da oggetti 4-dimensionali detti *timeslices*. Ogni *timeslice* rappresenta una sorta di "fotogramma", in cui l'entità ad esso riferita ha natura immutabile (perdurante) in quello specifico istante o intervallo di tempo.

### 3 La risorsa CLAVIUS

Clavius costituisce una RTO. In accordo con le più recenti acquisizioni teoriche (Roche, 2007; Temmerman et al., 2005), in essa infatti sono formalizzate come indipendenti due componenti strettamente legate tra di loro: la componente terminologica nella quale vengono strutturati i termini presenti in un testo e la componente ontologica attraverso la quale vengono formalizzati i concetti evocati da quei termini. Il livello più alto dell'ontologia è rappresentato così dalle due classi OWL disgiunte CONCEPT e TERM. La superclasse CONCEPT assume due sottoclassi, ASTRONOMICAL CONCEPT e MATHEMATICAL CONCEPT, che a loro volta assumono rispettivamente tutti i concetti astro-

nomici e matematici. La superclasse TERM si articola, invece, in due sottoclassi ASTRONOMICAL TERM e MATHEMATICAL TERM. Quest'ultima assume le classi OWL disgiunte LATIN TERM e GREEK TERM, di cui i termini latini e greci costituiscono le istanze. Nella risorsa la parte concettuale è espressa in inglese mentre la parte terminologica è costituita dai termini latini, legati ai concetti che essi evocano attraverso la relazione *denotes* e la sua inversa *isDenotedBy*. L'ontologia è attualmente composta da 106 classi organizzate in quattro livelli gerarchici, un insieme di 10 *Data Properties* e 18 *Object properties*, che permettono di dare una rappresentazione precisa dei concetti e termini. Le relazioni sono suddivise in: relazioni lessicali, paradigmatiche tra i termini (iperonimia, iponimia, meronimia, olonimia, sinonimia e antonimia); relazioni interlivello tra termini e concetti; relazioni concettuali, molte delle quali sono state introdotte per dare una definizione più precisa del dominio matematico-astronomico (Piccini et al., 2016). Per quanto concerne la modellazione degli aspetti temporali è stata importata l'ontologia OWL-Time, che si basa sulle relazioni binarie tra intervalli introdotte in (Allen and Ferguson, 1997) ed è stata integrata una serie di regole *Semantic Web Rule Language* (SWRL), elaborate in (Batsakis et al., 2016). In questo modo possono essere rappresentate e trattate dai motori inferenziali sia informazioni temporali quantitative (informazioni temporali precise), sia informazioni temporali qualitative (delle quali, cioè, non si possono specificare con esattezza l'istante iniziale e finale). Queste ultime, in particolare, sono molto frequenti nel testo di Clavius, e più in generale, nella letteratura, dove non è sempre possibile definire con esattezza la data di inizio e di fine nella quale si è verificato o sussiste un determinato fatto<sup>2</sup>. Le relazioni tra le entità ontologiche sono state temporalizzate in base alla loro validità nel tempo, seguendo i due approcci, quello N-ario e quello perdurantista, come descritto nella Sezione 2.

### 4 Comparazione della risorsa nei due modelli

In questa Sezione sono definiti i criteri di comparazione dei due modelli e vengono presentati

<sup>2</sup>La rappresentazione di tali informazioni temporali è indipendente dai modelli ivi trattati e non verrà pertanto discussa.

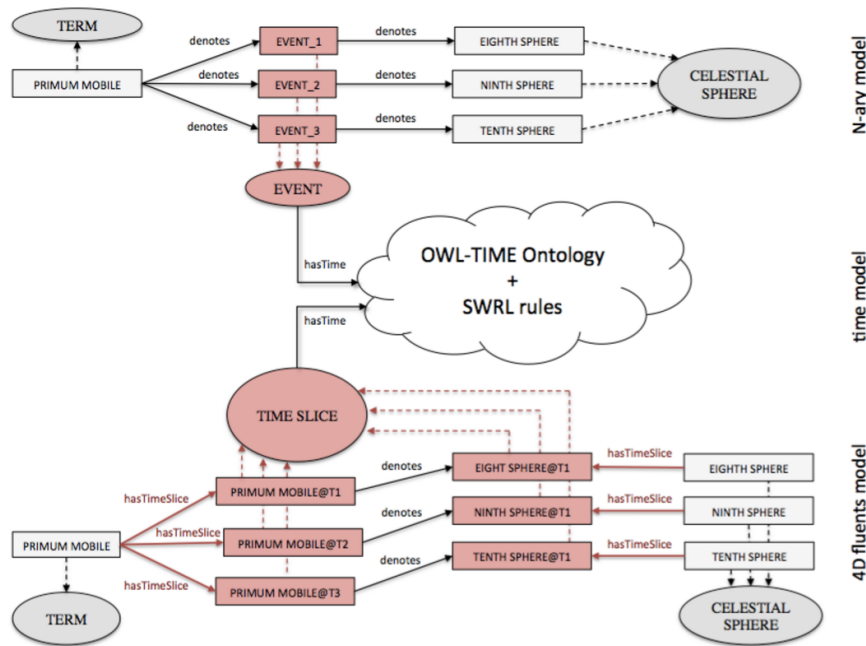


Figure 1: I due modelli a confronto. In rosso i nuovi oggetti da creare nei diversi modelli. Le linee tratteggiate rappresentano la relazione di *instance of*.

i vantaggi e le criticità di entrambi nella loro applicazione alla risorsa Clavius in OWL. Un primo criterio di comparazione riguarda il numero di assiomi logici necessari alla rappresentazione delle relazioni temporali. Come illustrato in precedenza, infatti, entrambi i modelli richiedono l'introduzione di nuove entità temporali (eventi o *timeslice*). A titolo d'esempio, presentiamo il mutamento nel tempo del concetto denotato dal termine *primum mobile*. Il numero delle sfere ed il loro relativo ordine costituisce un problema ampiamente dibattuto nell'astronomia del Seicento. Secondo la visione aristotelica del *kosmos* il movimento dei sette cieli attorno alla terra era dovuto ad una sfera, la più esterna, detta *primum mobile*. Mano a mano che nei secoli si vennero a scoprire maggiori dettagli sul moto dei pianeti, divenne necessario aggiungere altre sfere: così con Tolomeo il loro numero salì a 9, nelle tavole alfonsine a 10, quindi a 11 ed infine a 12 con Magini. Alla introduzione di una ulteriore sfera sul piano concettuale corrisponde un mutamento nel codominio della relazione interlivello *denotes* il cui dominio è rappresentato dal termine *primum mobile*: i) *primum\_mobile denotes eighth\_sphere* (dal 300 a.C. al 152 d.C.), ii) *primum\_mobile denotes ninth\_sphere* (dal 152 d.C. al 1252 d.C.), iii) *primum\_mobile denotes tenth\_sphere* (dal 1252 d.C. al 1589 d.C.). La Figura 1 mostra un esempio

di rappresentazione di questi *statements*.

Le classi, le proprietà e le istanze in rosso, rappresentano le entità ontologiche che debbono essere introdotte per tradurre in linguaggio formale tali *statements*. Si nota come il modello dei perduranti sia soggetto ad una maggiore proliferazione delle entità: accanto alla classe *TIMESLICE* sono stati introdotti una nuova *Object property* (*hasTimeSlice*), istanziata sei volte, e sei nuovi individui della classe *TIMESLICE* (le parti temporali degli individui coinvolti nella temporalizzazione).

Il modello N-ario, d'altro canto, risulta più snello, richiedendo solo la creazione di tre eventi, uno per ogni *statement*. Tuttavia, tale modellazione modifica, da un punto di vista intuitivo, la semantica della relazione *denotes*, spezzando e duplicando la proprietà per legare l'evento ai due individui coinvolti nella temporalizzazione. In entrambi i modelli il dominio e il codominio di una proprietà temporale vengono modificati: nel modello N-ario si aggiunge ad essi la classe *EVENT*, nel modello dei perduranti si sostituiscono le classi originarie con la classe *TIMESLICE*. Precisiamo che la nostra analisi è limitata ai modelli delle relazioni da temporalizzare, ed è quindi indipendente sia rispetto al modo in cui gli intervalli e gli istanti temporali sono rappresentati (OWL-TIME in Figura 1), sia rispetto a come i costrutti e le regole di OWL ("SWRL rules" in Figura 1) vengono

Table 1: Confronto tra i modelli

	modello n-ario	modello 4D-fluents
<b>Metriche dell'ontologia</b>		
numero di classi	106	106
numero di individui	244	265
numero di istanze di object property	117	159
<b>Temporalizzazione del predicato</b>		
$pred : X \rightarrow Y$	$pred_{n-ary} : X \sqcup Event \rightarrow Y \sqcup Event$	$pred_{perd} : TimeSlice \rightarrow TimeSlice$
<b>Interrogazione</b>		
query SPARQL: lunghezza del path	2	3
query SPARQL: numero di variabili	2	3
<b>Tools</b>		
Availability	<i>yes</i>	<i>no*</i>

combinati per fornire una procedura di reasoning solida e completa. Per una trattazione accurata di queste problematiche rimandiamo a (Batsakis et al., 2016). In relazione ad una maggiore proliferazione di entità, inoltre, l'interrogazione della risorsa può diventare più complessa e quindi maggiormente prona ad errori. Per quanto riguarda l'interrogazione, abbiamo considerato una semplice query di interrogazione della risorsa per conoscere quali sono i concetti che il termine *primum\_mobile* ha denotato nel tempo. Le query sono state implementate tramite il linguaggio di interrogazione SPARQL nei due modelli. Per il modello N-ario abbiamo:

```
SELECT ?concept
WHERE {
  n-ary:primum_mobile a n-ary:TERM .
  n-ary:primum_mobile n-ary:denotes ?event .
  ?event a n-ary:EVENT .
  ?event n-ary:denotes ?concept }
```

Per il modello dei perduranti abbiamo:

```
SELECT ?concept
WHERE {
  ?ts@t1 4d:denotes ?ts@t2 .
  ?ts@t1 a 4d:TIMESLICE .
  ?ts@t2 a 4d:TIMESLICE .
  4d:primum_mobile a 4d:TERM .
  4d:primum_mobile 4d:hasTimeSlice ?ts@t1 .
  ?concept 4d:hasTimeSlice ?ts@t2 }
```

Da una analisi dei due codici si evince che per interrogare il secondo modello occorre un numero maggiore di variabili nella query e che il path dell'interrogazione del grafo è più lungo e coinvolge un maggior numero di entità ontologiche. Questo implica che, in termini di scalabilità ed efficienza nel processo di interrogazione, il modello N-ario potrebbe risultare migliore. Come ultimo criterio abbiamo considerato la disponibilità di strumenti che supportino l'umanista nella

creazione di RTO diacroniche senza richiedere la conoscenza dei modelli matematici sottostanti. Per quanto riguarda il modello N-ario è disponibile il plug-in di *Protégé CHRONOS-ED* (Preventis et al., 2012). Il plug-in per il modello *4D-fluents*, al contrario, risulta ad oggi non funzionante nelle ultime versioni di *Protégé* (vedi asterisco in Tabella 1). I risultati dell'analisi comparativa sono riassunti nella Tabella 1.

## 5 Discussione e lavori futuri

La modellazione della risorsa diacronica terminologica CLAVIUS attraverso il modello a relazioni N-arie e quello *4D-fluents* ha consentito di valutare, empiricamente, quelle che sono le differenze salienti tra i due paradigmi temporali. Per quanto riguarda l'aspetto di *authoring*, di primario interesse in questo lavoro in quanto funzionale alla necessità di un umanista digitale nella costruzione (tipicamente manuale) di una risorsa diacronica, i risultati hanno evidenziato come, ad oggi, l'approccio a relazioni N-arie risulti il più vantaggioso, in termini sia di numero di entità ontologiche coinvolte, sia di complessità delle query SPARQL sia di disponibilità di strumenti di supporto. Si intende approfondire questa ricerca prendendo in considerazione gli aspetti che, naturalmente, seguono quelli più essenziali di formalizzazione della terminologia e della conoscenza. In primis, verrà studiato l'impatto determinato dall'adozione di un certo modello sui meccanismi di *reasoning*, e quindi, sulla possibilità da parte di un motore inferenziale di derivare nuova conoscenza a partire da quella rappresentata esplicitamente nella risorsa.

## References

- James F. Allen and George Ferguson. 1997 *Actions and Events in Interval Temporal Logic*. In: Spatial and Temporal Reasoning. O. Stock, ed., Kluwer, Dordrecht, Netherlands, pp. 205-245.
- Sotiris Batsakis, Euripides Petrakis, Ilias Tachmazidis and Grigoris Antoniou 2016. Temporal Representation and Reasoning in OWL 2. *Semantic Web journal*, ISSN 1570-0844.
- Fabio Ciotti. 2014 *Digital Literary and Cultural Studies: State of the Art and Perspectives*. Between, vol. 4, no. 8.
- Noy N. Fridman, William Grosso, and Mark A. Musen 2000. *Knowledge-Acquisition Interfaces for Domain Experts: An Empirical Evaluation of Protege-2000*. SEKE2000 Proceedings. Chicago.
- Aldo Gangemi. 2011 *SuperDuper schema: an OWL2+RIF DnS pattern*. KCAP 2011 Deep Knowledge Representation Challenge Workshop.
- Fabio Grandi, and Maria Rita Scalas. 2009 *The Valid Ontology: A simple OWL temporal versioning framework*. Advances in Semantic Processing, SEMAPRO'09. Third International Conference IEEE, 2009. pp. 98-102.
- Hans U. Krieger. 2014 *A detailed comparison of seven approaches for the annotation of time-dependent factual knowledge in RDF and OWL*. Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation, p. 1.
- Hans U. Krieger. 2012 *A temporal extension of the Hayester Horst entailment rules and an alternative to W3C's n-ary relations*. Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS), pp. 323–336.
- Frank Manola and Eric Miller. 2004 *RDF primer*. Technical report, W3C.
- Natasha Noy and Alan Rector. 2006 *Defining N-ary Relations on the Semantic Web*. W3C Working Group Note 12, <https://www.w3.org/TR/swbp-n-aryRelations/>
- Piccini, S., Bellandi, A., and Benotto, G. 2016 *Formalizing and Querying a Diachronic Terminological Resource: the CLAVIUS Case Study*. Proceedings of From Digitization to Knowledge 2016 workshop (D2K), July 11, 2016, Krakow (Poland).
- Preventis A., Marki P., Petrakis E.G.M., Batsakis S. 2016 *Chronos: A Tool for Handling Temporal Ontologies in Protégé*. Proceedings of the 24th International Conference on Tools with Artificial Intelligence, Athens, Greece.
- Christophe Roche. 2007 *Le terme et le concept: fondements d'une ontoterminologie*. Actes de la première conférence TOTh, pp.1–22, Annecy.
- Evren Sirin, Bijan Parsia, Bernardo C. Grau, Aditya Kalyanpur, Yarden Katz 2007. *Pellet: A practical owl-dl reasoner*. Web Semantics: science, services and agents on the World Wide Web, 5.2 51-53.
- Rob Shearer, Boris Motik, and Ian Horrocks. 2008 *HermiT: A Highly-Efficient OWL Reasoner*. OWLED, Vol. 432.
- Rita Temmerman, Koen Kerremans, and Veerle Vandervoort 2005 *La termontographie en contexte(s)*. In Blampain, D., Thoiron, P., Van Campenhoudt, M. (eds). Mots, Termes et Contextes. Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie Terminologie Traduction. Bruxelles, 429- 439.
- Chris Welty, Richard Fikes, and Selene Makarios 2006 *A reusable ontology for fluents in OWL*. Proceedings of the 4th International Conference on Formal Ontology in Information Systems (FOIS), Vol. 150. pp. 226–336