

LODE: une ontologie pour représenter des événements dans le web de données

Raphaël Troncy¹, Ryan Shaw² et Lynda Hardman³

¹ EURECOM, Sophia Antipolis, France

² University of California, Berkeley, USA

³ CWI, Amsterdam, The Netherlands

Résumé : Nous faisons communément référence à la notion d'*événement* pour décrire une action ou quelque chose qui a lieu à un endroit particulier pendant une certaine période de temps. Ce concept est utile pour organiser et relier des faits individuels, et pour découvrir de nouvelles relations entre ceux-ci, en particulier sur le web de données. Dans cet article, nous proposons tout d'abord une comparaison des différents modèles permettant de représenter des événements, en mettant l'accent sur leur expressivité et les choix de modélisation. Nous décrivons ensuite l'ontologie LODE qui permet de représenter des événements dans le web sémantique. Nous présentons finalement un prototype de service fournissant des URIs stables pour les événements et retournant une description RDF pour chacune des dimensions les composants.

1 Introduction

Bien que leurs méthodes diffèrent de manière significative, les historiens et les journalistes ont l'habitude de narrer les enchaînements d'événements pour expliquer des phénomènes passés. L'initiative *Linked Data*¹ a comme objectif de publier et de connecter des jeux de données représentés en RDF dont les URIs identifient des documents web, des objets du monde réel, des relations les reliant et d'autres type d'informations. Pourtant, si certains vocabulaires sont devenus très populaires pour représenter les personnes, les lieux et d'une manière générale les entités nommées dans le web sémantique, aucun n'a véritablement émergé pour représenter la notion d'événement.

Le terme "événement" est lui-même polysémique. Il fait tout à la fois référence à des phénomènes passés (décrits dans des articles de presse ou expliqués par des historiens) et à des phénomènes planifiés dans le futur (notés dans un calendrier). Plusieurs formats et standards ont été proposés pour échanger des données structurées de calendrier (e.g. *iCal*). Dans cet article, nous nous concentrons sur la première catégorie : les phénomènes qui ont eu lieu dans le passé.

Notre contribution est double. Nous présentons tout d'abord une comparaison détaillée de plusieurs modèles ontologiques permettant de représenter des événements historiques (section 2). Ces modèles sont issus de communautés différentes et ont chacun

1. <http://linkeddata.org/>

leurs atouts. Notre but n'est donc pas de décrire un nouveau modèle, mais plutôt de résoudre un problème d'interopérabilité en proposant un ensemble d'axiomes logiques exprimant des correspondances entre ces ontologies (Shaw *et al.*, 2009) (section 3). Nous décrivons ensuite des outils pour nourrir ce modèle à partir de données semi-structurées telles que les frises chronologiques de Wikipédia. Nous présentons un prototype de service de type annuaire² fournissant des URIs stables pour les événements et retournant une description RDF pour chacune des dimensions les composant (section 4), avant de conclure et d'ouvrir quelques perspectives à ces travaux (section 5).

2 Comparaison d'ontologies centrées événement

Différentes ontologies OWL fournissant classes et propriétés pour décrire des événements et leurs relations ont été proposées (table 1). Dans cette section, nous présentons une

Ontologie	Namespace
CIDOC CRM	http://cidoc.ics.forth.gr/OWL/cidoc_v4.2.owl
Ontologie ABC	http://metadata.net/harmony/ABC/ABC.owl
Ontologie Event EO	http://purl.org/NET/c4dm/event.owl#
Ontologie SEM	http://semanticweb.cs.vu.nl/2009/04/event/
EventsML-G2	http://www.iptc.org/std/EventsML-G2/
DOLCE+DnS Ultralite	http://www.loa-cnr.it/ontologies/DUL.owl
F	http://events.semantic-multimedia.org/ontology/2008/12/15/model.owl
Ontologie OpenCYC	http://www.opencyc.org/

TABLE 1 – Ontologies pour représenter des événements

comparaison détaillée de ces modèles à partir de leurs dimensions principales : le type (section 2.2), les dimensions temporelle (section 2.3), et spatiale (section 2.4), la participation (section 2.5), la causalité (section 2.6) et la composition (section 2.7).

2.1 Aperçu des modèles

Bien que toutes les ontologies nommées ci-dessus contiennent des classes et des propriétés pour représenter le concept d'événement, elles ont été développées dans des buts différents. Ainsi, les ontologies CIDOC-CRM (Doerr, 2003) et ABC (Lagoze & Hunter, 2001) ont comme objectif de permettre l'interopérabilité des formats de métadonnées pour décrire des contenus multimédia possédés par les musées et les bibliothèques. Les événements visés sont donc historiques dans le sens large (e.g. guerre, naissance) ou liés aux objets décrits (e.g. changement de propriétaire, restauration).

L'ontologie EO (Raimond *et al.*, 2007) a été développée pour être utilisée avec des ontologies musicales. Destinée principalement à la description d'interprétation artistique, rien de spécifique au domaine musical n'a été cependant inclus. Elle est aujourd'hui l'ontologie la plus utilisée dans la communauté *Linked Data*. L'ontologie SEM (van Hage *et al.*, 2009) permet elle de représenter des attaques maritimes dans des lieux parfois imprécis. EventsML-G2 est un modèle développé par l'IPTC pour échanger des informations structurées décrivant les événements entre les fournisseurs d'information

2. <http://www.linkedevents.org/> est une interface permettant de naviguer dans cet annuaire.

et leurs partenaires. Il permet de décrire les événements passés, instantanés et à venir, dans la presse.

DOLCE+DnS Ultralite (DUL) est une ontologie légère de haut niveau qui combine une axiomatisation simplifiée des concepts de DOLCE avec les patrons de représentation *Descriptions and Situations* (Gangemi & Mika, 2003). L'ontologie F Event est basée sur DUL dans la mesure où elle spécifie d'autres patrons pour représenter des relations (méréologique, participation, causale, corrélation) entre événements (Scherp *et al.*, 2009). OpenCYC est aussi une ontologie de haut niveau, mais suit une approche complètement opposée par rapport à DUL : plutôt que de fournir un ensemble très restreint de concepts, OpenCYC fournit des centaines de milliers de classes ayant pour but de modéliser "toute la réalité de l'homme".

2.2 Type fondamental : aspect et agentivité

Dépendamment de l'application visée, ces ontologies définissent différemment le concept d'événement (voir le tableau 2 pour la définition en prose en anglais). De

cidocE2.Temporal-Entity	"[E2.Temporal_Entity] comprises all phenomena, such as the instances of E4.Periods, E5.Events and states, which happen over a limited extent in time."
abceEvent	"An Event marks a transition between Situations."
eoEvent	"An arbitrary classification of a space/time region, by a cognitive agent."
semEvent	"Something significant that happens at a specified place and time."
eventsm1Event	"...something that happens and is subject to news coverage."
dulEvent	"Any physical, social, or mental process, event, or state."
fEvent	"...perduring entities (or perdurants or occurants) that unfold over time, <i>i.e.</i> , they take up time."
cyoSituation	"...a state or event consisting of one or more objects having certain properties or bearing certain relations to each other."

TABLE 2 – Définitions du concept *événement* dans plusieurs ontologies

plus, à l'exception de l'ontologie EO, elles permettent toutes de préciser le type fondamental de l'événement représenté. La manière de spécifier ce type varie cependant. Une première manière de distinguer le type d'événement est de considérer son *aspect*, c'est-à-dire si celui-ci correspond à un processus ou une activité en cours, ou si il s'agit plutôt d'une transition entre des états. Par exemple, OpenCYC définit le concept *Situation* et utilise cet aspect pour se spécialiser en *StaticSituation* et *Event*. Le premier dénote une situation dans laquelle l'état des choses persiste au cours du temps alors que le second fait référence à une situation où des changements sont observés.

CIDOC opère une distinction similaire mais conceptuellement moins claire entre les deux sous-concepts de *E2.Temporal_Entity* : *E3.Condition_State* et *E5.Event*. L'ambiguïté provient du fait que CIDOC dispose également du concept *E4.Period* pour désigner un type d'entité temporelle non statique mais qui n'implique pas nécessairement un changement d'état. Le concept *E3.Condition_State* permet de décrire "la condition physique d'un objet matériel" ce qui semble exclure,

par exemple, la description de l'état relatif de deux choses. `E3.Condition_State` est similaire au concept `Situation` de l'ontologie ABC qui permet de décrire l'état d'objets tangibles à un instant donné. Le concept `Event` est alors défini comme une transition entre deux instances de `Situation`. Ce choix de modélisation rend difficile la description d'un événement qui est caractérisé par un changement de la relation entre deux choses plutôt que par le changement d'état d'un seul objet.

Une autre distinction consiste à identifier si l'*agent* a produit l'événement. OpenCYC et DUL considèrent ainsi le concept `Action` comme un type particulier de `Event` tandis que pour CIDOC, `E7.Activity` spécialise `E5.Event`. Dans l'ontologie ABC, le concept d'`Action` est aussi défini comme quelque chose d'accompli par un agent, mais il est formellement disjoint du concept `Event` qui peut donc "contenir" des actions via la propriété `hasAction`. Ainsi, l'ontologie ABC suggère que les événements peuvent être complètement décrits comme un ensemble d'actions accomplies par un agent, ce qui peut être problématique pour modéliser un tremblement de terre.

L'introduction de ce type de classification pour les événements dans une ontologie peut être problématique puisqu'il force l'ingénieur de la connaissance à adopter un point de vue particulier sur les faits qui se sont déroulés. Si ceci est plutôt souhaitable dans des domaines où la connaissance a besoin d'être représentée précisément, cela est moins justifié dans notre cas où l'objectif est d'améliorer la description et l'accès à des documents qui peuvent contenir des interprétations différentes d'un même événement. En effet, ce distinguo basé sur l'*aspect* ou l'*agentivité* n'est pas nécessairement intrinsèque aux faits passés, mais relève plutôt d'une interprétation. Qu'un événement historique ou rapporté dans la presse implique un changement identifiable, où que cette agentivité puisse être identifiée, fait le plus souvent objet d'un débat, et sa résolution ne devrait pas être un pré-requis pour pouvoir modéliser des faits passés à partir de concepts d'une ontologie. Cette séparation entre les événements et leurs interprétations est promue par DUL qui fournit le concept de `Situation` dont les instances peuvent décrire des points de vue différents attaché à une même instance de `Event`.

2.3 Événement et dimension temporelle

La dimension temporelle est une caractéristique majeure pour distinguer les événements. La relation entre un événement et une période chronologique est analogue à celle entre un lieu géographique et un système abstrait de coordonnées spatiales. Dans chaque cas, les instances du premier ont un sens communément et socialement accepté, tandis que dans le second cas, il s'agit d'un système arbitraire permettant de diviser un espace abstrait. Dans le langage OWL, un événement peut être lié à un intervalle temporel en utilisant une propriété *datatype* dont la valeur sera un littéral RDF représentant une date calendaire (e.g. en utilisant les types définis par XML Schema `xsd:date` ou `xsd:dateTime`). Une autre approche consiste à introduire une classe spécifique pour représenter les intervalles temporels et à utiliser ensuite une propriété *objet* pour lier l'instance d'événement à une instance de cette classe.

Les ontologies ABC, CIDOC et EO suivent toutes cette deuxième approche. ABC et CIDOC introduisent une classe particulière représentant un intervalle temporel tandis que EO réutilise le concept `TemporalEntity` défini dans l'ontologie OWL-Time (Hobbs & Pan, 2006). DUL dispose de deux propriétés : un événement peut être

directement daté en utilisant la propriété *datatype* `hasEventDate`, ou l'intervalle temporel peut être complètement explicité en instanciant le concept `TimeInterval` et en utilisant la propriété *objet* `isObservableAt`. L'avantage d'associer une date à un événement directement est la simplicité : moins de niveau d'abstraction à gérer et une comparaison ou un traitement des valeurs plus aisé à effectuer. Ainsi, la visualisation d'événements sur une frise chronologique est quasiment immédiate. Cette simplicité a cependant un coût : une incapacité à représenter certains intervalles temporels, par exemple, ceux pour lesquels on ne peut pas leur associer une datation précise, ou ceux qui ne sont pas bornés dans le temps, un cas fréquent pour les événements historiques. L'introduction de classes pour représenter les intervalles temporels permet d'effectuer des calculs et des raisonnements. L'inconvénient d'une telle approche est qu'elle introduit un certain nombre de concepts abstraits. Dans le cadre du web sémantique, le problème consiste alors à pouvoir identifier ces intervalles avec des URIs ou à gérer des noeuds anonymes dans des graphes RDF.

2.4 Événement et dimension spatiale

Tout comme pour la dimension temporelle, les événements peuvent être liés à des régions abstraites spatiales ou à des lieux qui font sens pour l'être humain. ABC, CIDOC et EO permettent ainsi d'associer un événement à une région spatiale. CIDOC définit le concept `E53.Place` comme "l'extension spatiale" auquel les événements font référence via la propriété `P7.took_place_at`. Les instances de ce concept peuvent avoir des noms (`E44.Place_Appellation`), mais il n'est pas possible de relier un événement à un lieu si celui-ci n'est pas défini dans l'espace. Le concept `Place` de ABC met également l'accent sur la définition géographique d'un lieu plutôt que son acceptation commune. La propriété `place` de EO a ainsi comme valeur une instance de `wgs84:SpatialThing` qui est aussi défini en terme de latitude et de longitude.

DUL est la seule ontologie à distinguer explicitement `Place` et `SpaceRegion`. Un événement peut ainsi être lié à `Place` via la propriété `hasLocation`, ou à `SpaceRegion` via la propriété `hasRegion`. Ce choix de modélisation est le plus flexible puisqu'il permet d'associer un événement à des lieux sans avoir à spécifier leur extension géographique. Par exemple, les historiens spécialistes de l'Antiquité travaillent avec des documents qui font indifféremment référence à des lieux réels ou mythiques. Le même problème se pose d'ailleurs avec des événements contemporains qui peuvent être liés à des lieux virtuels tels que ceux définis dans les jeux tels que *Second Life*. Enfin, le fait de distinguer clairement un lieu commun d'une extension spatiale permet de gérer le problème des lieux dont l'extension géographique change au cours du temps.

2.5 Événement et participation

Une autre caractéristique commune à toutes ces ontologies est la possibilité de lier des événements à des agents (e.g. des personnes ou des organisations) et à des objets.

L'implication d'objets dans les événements. ABC définit deux types de propriétés pour lier un `Event` à une chose tangible (une `Actuality` dans le vocabulaire d'ABC). La propriété `involves` définit l'implication dans un sens très général tandis que la propriété `hasResult` spécifie la chose tangible qui résulte d'un événement. CI-

DOC définit la propriété `P12.occurred_in_the_presence_of`, qui, de la même manière que la propriété `involves` de ABC, permet de relier un `E5.Event` à un `E77.Persistent_Item` (i.e. un objet enduring) sans imposer un rôle particulier à cet objet outre sa seule implication dans le sens large du terme. Cette propriété est toutefois la racine d'une hiérarchie de sous-propriétés qui permet de réifier la nature de cette implication, par exemple, les sous-propriétés `P25.moved` et `P31.has_modified`. Contrairement au concept `Actuality` de ABC, le concept `E77.Persistent_Item` de CIDOC regroupe à la fois les objets tangibles mais également les concepts ou les idées. DUL définit la propriété `hasParticipant` pour relier un `Event` à un `Object`. Tout comme le concept CIDOC `E77.Persistent_Item`, le concept `Object` de DUL recoupe aussi bien les objets sociaux, mentaux et physiques. La propriété `factor` définie dans l'ontologie EO est tout aussi vague, n'ayant pas de co-domaine spécifié. EO définit finalement la propriété `product`, qui, comme la propriété `hasResult` de ABC, permet de relier un `Event` avec quelque chose qui résulte de cet événement.

La participation d'agents dans les événements. ABC définit la propriété `hasPresence` pour asserter qu'un agent est présent à un événement sans nécessairement impliquer que celui-ci y a un rôle actif. Cette propriété se spécialise d'ailleurs en `hasParticipant` pour justement indiquer que l'agent a un rôle causal ou actif pour l'événement. De la même manière, CIDOC définit les propriétés `P11.had_participant` (équivalent de `hasPresence` dans ABC) et `P14.carried_out_by` (équivalent de `hasParticipant` dans ABC). DUL spécifie la propriété `involvesAgent`, sous-propriété de `hasParticipant` pour lier un `Event` à un `Agent`. L'ontologie EO fournit la propriété générale `agent` qui a le même but. L'ontologie F a une manière unique pour modéliser la participation à un événement. Alors que DUL permet d'asserter qu'un objet ou un agent participe à un événement, F utilise le patron DnS (*Descriptions and Situations*) (Gangemi & Mika, 2003) pour mieux caractériser cette participation. La propriété `P14.1_in_the_role_of` de CIDOC permet également de réifier la nature de la participation d'un agent à un événement en précisant son rôle.

2.6 Événement, influence, but et causalité

La manière de modéliser des relations entre événements telles que la causalité, le but ou l'influence varie selon les ontologies. CIDOC et EO disposent toutes les deux de propriétés permettant de lier un événement à quelque chose de tangible ou pas. CIDOC définit par exemple la propriété `P15.was_influenced_by`, tandis que EO définit la propriété `factor`. EO ne distingue pas si cet objet a directement participé à un événement ou si il a eu une quelconque influence. De la même manière, la seule différence entre les propriétés `P12.occurred_in_the_presence_of` et `P15.was_influenced_by` de CIDOC semble être que l'objet soit physiquement présent pendant l'événement (et par implication, il s'agit donc d'une instance de `E77.Persistent_Item`). L'ontologie ABC ne dispose que de la propriété `hasResult` pour associer une vague causalité à l'événement. Dans le discours historique, on observe souvent un manque de consensus quant à la causalité, le but ou l'influence. Ainsi, les propriétés présentées ci-dessus ont peu de chance d'être vraiment utiles pour exprimer les relations complexes dont les historiens ont le plus souvent besoin. Le modèle F fournit ici un patron de description basé sur DnS beaucoup plus riche qui spécifie non

seulement la cause et l'effet d'un événement mais également la théorie justifiant cette assertion.

2.7 Événement, partie et composition

Il est souvent souhaitable de représenter qu'un événement *A* fait partie d'un autre événement *B*. La difficulté est de caractériser la sémantique à donner à cette relation méréologique. Il s'agit généralement d'une inclusion temporelle, mais celle-ci n'est pas suffisante. En effet, si quelqu'un se marie pendant les Jeux olympiques, cela ne signifie pas que ce mariage puisse être considéré comme un sous événement des Jeux olympiques. Les ontologies doivent donc distinguer ce qui relève d'une simple inclusion temporelle entre deux événements recouvrant une même période, d'une véritable relation entre un sous-événement faisant partie d'un événement plus large.

CIDOC dispose des propriétés pour opérer une telle distinction : la propriété `P86` - `falls_within` permet d'asserter une relation d'inclusion temporelle entre deux intervalles, tandis que la propriété `P9` - `consists_of` permet d'exprimer une véritable relation méréologique entre événements. EO définit également la propriété `sub_event` et ABC la propriété `isSubEventOf` avec la même sémantique. De plus, ABC conceptualisant la notion d'événement comme un ensemble d'actions effectués par des agents, elle fournit également la propriété `hasAction` pour lier un événement aux actions qu'il contient.

DUL définit deux propriétés pour lier un événement à des sous-événements : `hasPart` et `hasConstituent`. La première propriété permet d'exprimer à la fois une inclusion temporelle (e.g. "l'année 1923 appartient au 20ème siècle") et une relation sémantique (e.g. "la 2ème guerre mondial inclut Pearl Harbour"). La deuxième propriété a comme objectif de modéliser que la connaissance du monde peut être représentée sur plusieurs couches selon leur niveau d'abstraction qui ne sont pas strictement inclus les uns aux autres. Par exemple, la société est constituée d'individus, même si ceux-ci ne sont pas des "parties" de la société puisque individus et société existent à différents niveaux d'abstraction. Cette distinction est aussi utile pour établir des relations entre événements. Ainsi, un événement complexe tel que la Révolution Française est composé de plusieurs sous-événements, mais ceux-ci ne font pas partie de l'événement principal de la même manière qu'un set fait partie d'un match de tennis.

En utilisant le patron DnS, l'ontologie F permet une fois de plus de décrire finement comment un événement peut être décomposé. Des situations spécifiques (un ensemble particulier d'événements) peuvent ensuite satisfaire cette description. Cette flexibilité permet par exemple de décrire dans quelle condition un événement peut être considéré comme faisant partie d'un autre et donc d'inférer des relations méréologiques plutôt que de les représenter explicitement à chaque fois.

3 LODE : une ontologie événement pour le web de données

Dans cette section, nous proposons une ontologie qui tire le meilleur parti des différents modèles étudiés ci-dessus. Notre but est de permettre la description interopérable des aspects "factuels" d'un événement, ce qui peut se caractériser en terme des "quatre Ws"

(*what, when, where, who*) : qu'est-ce qui s'est passé, où et quand cela s'est-il produit, qui était impliqué. Ces relations factuelles décrivant un événement ont comme objectif de représenter une réalité consensuelle et ne doivent donc pas être associées à une perspective ou une interprétation particulière. Notre modèle permet donc de représenter des informations pour lesquelles un consensus a été obtenu. Que celles-ci aient été obtenues empiriquement ou produites par la rhétorique est une question épistémologique. Notre modèle exclut donc les propriétés permettant de catégoriser les événements ou de les inter-liaisonner avec des relations causales ou métréologiques. Nous pensons que ces questions relèvent du domaine de l'interprétation et sont mieux gérées avec des approches telles que le patron DnS spécialisé dans l'ontologie F.

Le tableau 3 montre quelques unes des propriétés de notre modèle avec leurs correspondances dans les autres modèles discutés ci-dessus. L'ontologie contenant la définition formelle de ces axiomes est disponible à <http://linkedevents.org/model/>.

ABC	CIDOC	DUL	EO	LODE
atTime	P4.has_time-span P7.took_place_at	isObservableAt	time place	atTime inSpace atPlace
inPlace		hasLocation		atPlace
involves	P12.occurred_in_- the_presence_of	hasParticipant	factor	involved
hasPresence	P11.had_participant	involvesAgent	agent	involvedAgent

TABLE 3 – Exemple d'alignements entre propriétés de plusieurs ontologies événements

Agentivité. Notre modèle est agnostique quant à un jugement sur l'aspect ou l'agentivité d'un événement. Ainsi, il est possible de représenter un événement historique sans prendre position sur ce qui a changé ou qui en est responsable. Ce choix de modélisation a des conséquences quant à l'alignement de notre modèle vers les autres ontologies. Nous considérons ainsi que notre concept *Event* est exactement équivalent à celui défini dans DUL et EO puisque ceux-ci sont également agnostiques vis à vis de l'aspect ou de l'agentivité d'un événement. En revanche, notre concept n'est pas équivalent au concept CIDOC E5. *Event* puisque celui-ci exclut les processus, activités ou états en cours de la notion d'événement. Pour être compatible avec cette notion d'événement statique, nous définissons le concept *Event* comme une sous-classe du concept CIDOC E2. *TemporalEntity*, qui est lui même une super-classe de E5. *Event* (via E4. *Period*) et E3. *Condition_State*. Cet axiome de sous-classe par rapport à E2. *TemporalEntity* se justifie dans la mesure où ce dernier est défini comme "quelque chose qui se passe pendant une période de temps donnée" alors que notre définition, plus spécifique, indique en outre que cette chose a du être classifiée comme un événement par un agent, tel qu'un historien ou un journaliste.

Temporalité. Notre modèle permet de relier un événement à un intervalle temporel. Tout comme pour l'ontologie EO, cet intervalle est modélisé avec la classe *TemporalEntity* de l'ontologie OWL-Time. Notre propriété *atTime* est donc équivalente à la propriété *time* de EO. De plus, *atTime* est définie comme une sous-propriété de *isObservableAt* définie dans DUL, dans la mesure où elle restreint son domaine pour n'inclure que des événements. De la même manière, *atTime* est aussi une sous-propriété de P4. *has_time-span* définie dans CIDOC puisque là encore, le domaine est restreint aux seuls événements plutôt que n'importe quel instance de

E2. *TemporalEntity*. Finalement, nous définissons *atTime* comme une propriété fonctionnelle, ce qui implique qu'un événement peut être associé à au plus un seul intervalle temporel. Si un désaccord subsiste quant à la valeur de cet intervalle pour un événement particulier, alors cette valeur doit correspondre au plus petit intervalle qui inclut les interprétations contradictoires ou être laissée non spécifiée.

Espace. Nous suivons la même approche que DUL pour distinguer une région abstraite définie dans l'espace d'un lieu qui fait sens pour l'être humain. Nous définissons la propriété *inSpace* pour relier subjectivement un événement à un espace borné. Tout comme pour la propriété *atTime*, *inSpace* est fonctionnelle. Elle est aussi définie comme une sous-propriété de *hasRegion* définie dans DUL puisque son domaine est restreint aux événements (et non pas n'importe quelle entité) et son co-domaine est restreint aux régions spatiales (et non pas n'importe quel espace dimensionnable). Tout comme pour EO, nous utilisons le concept *SpatialThing* défini dans l'ontologie basique de géographie WGS84. Ainsi, notre propriété *inSpace* est équivalente à la propriété *place* de EO. Notre concept d'événement étant plus large que celui de CIDOC, *inSpace* est aussi une super-propriété de *P7.took_place_at* défini dans CIDOC. En complément de cette propriété, nous définissons la propriété *atPlace* pour associer un événement à un lieu commun sans qu'il soit forcément nécessaire de définir l'extension géographique de ce lieu. Contrairement à *inSpace*, *atPlace* n'est pas défini comme une propriété fonctionnelle et plusieurs lieux peuvent donc être spécifiés. *atPlace* est aussi défini comme une sous-propriété de *hasLocation* de DUL, le domaine étant restreint aux seuls événements, et le co-domaine restreint aux lieux (plutôt que n'importe quelle entité).

Participation. Tout comme DUL, nous définissons une propriété pour lier un événement à une chose arbitraire (*involved*) que nous spécialisons lorsque cette chose est un agent (*involvedAgent*). Ces deux propriétés sont ainsi directement équivalentes à *hasParticipant* et *involvedAgent* définies dans DUL. Elles sont presque³ équivalentes aux propriétés *P12.occurred_in_the_presence_of* et *P11.had-participant* définies dans CIDOC. L'alignement avec EO est plus complexe. *involved* est plus spécifique que *factor* puisque le co-domaine de ce dernier est restreint aux seuls objets (ce qui exclut par exemple les "causes abstraites"). Mais elle est en même temps plus générale puisque contrairement à *factor*, cette propriété n'implique pas que l'objet a un rôle passif. Aucun axiome ne peut donc être établi entre ces deux propriétés. *involvedAgent* est une super-propriété de *agent* définie dans EO, puisque le rôle de l'agent peut être actif ou passif. Le fait de qualifier ce rôle est effectivement une interprétation subjective qui dépasse notre objectif de ne représenter que les aspects factuels d'un événement.

Causalité. Finalement, notre modèle ne contient pas de propriétés pour représenter l'influence, le but ou la causalité d'un événement. Il n'y a par conséquent pas de propriétés équivalentes à *P15.was_influenced_by* de CIDOC ou à *factor* de EO. De la même manière, nous ne pouvons exprimer de relation méréologique entre événements. Nous considérons que ce type d'information relève d'une interprétation dont la description peut être représentée par un patron complexe tel que *DnS* ou sa spécialisation dans l'ontologie F.

3. L'équivalence n'est pas complète puisque notre concept d'événement est plus général

4 Applications

Dans un premier temps, nous avons implémenté une extraction automatique d'événements à partir des frises chronologiques de Wikipédia pour produire une description de ceux-ci dans le web de données. Dans un second temps, nous cherchons à intégrer des événements décrits selon des ontologies différentes. Il s'agit ici de vérifier la justesse de nos axiomes logiques et de voir si notre modèle résout les problèmes d'interopérabilité.

Extraire des événements à partir de Wikipédia. Les événements décrits dans les différentes frises chronologiques de Wikipédia varient en terme de domaine et de granularité ce qui en fait un bon test pour évaluer notre modèle. De plus, ces frises ne sont pas encore traitées par les projets DBpedia et Freebase. Notre contribution permet donc d'enrichir le web de données.

Les frises chronologiques de Wikipédia existent sous deux formes différentes :

- Celles dédiées à un sujet particulier (e.g. “la chronologie de l’histoire des techniques”) qui prennent généralement la forme d’une liste ou d’un tableau d’événements. En octobre 2008, Wikipédia avait environ 1000 chronologies de cette sorte. Cette liste d’événements est souvent divisée en groupes temporels (e.g. *Septembre 1939* or *12ème siècle*). Chaque événement correspond au minimum en une date et une description courte qui peut elle même être liée à d’autres articles de Wikipédia.
- Celles dédiées à une date particulière (e.g. “1996 en Irlande”) qui, outre la description des événements, incluent souvent le type d’événement décrit (e.g. naissance, mort, événement sportif). La forme la plus générale de ce type d’article est une année (e.g. “1976”), lui-même lié à des articles plus spécifiques pour les jours et les mois de l’année (e.g. “24 mai”).

Cette deuxième forme d’article a une structure plus homogène ce qui rend l’extraction d’événements plus facile à automatiser. Cependant, les différents événements ont rarement quelque chose en commun à l’exception du mois et de l’année auxquels ils sont associés. Notre but étant de lier les événements entre eux, à travers des thématiques, des lieux ou des participants communs, nous avons préféré traiter le premier type de chronologie.

Nous avons écrit des analyseurs spécifiques pour certain des articles de type chronologie les plus actifs de Wikipédia. Ces analyseurs identifient les événements contenus dans l’article et extraient pour chaque entrée leur date et description. Les titres des sous-sections de l’article sont également utiles puisque les dates fournies sont souvent relatives à ceux-ci. Nous modélisons ces dates en utilisant OWL-Time et nous utilisons la propriété `atTime` pour les lier à l’événement. Nous utilisons les termes de la description qui pointent vers d’autres articles sur Wikipédia pour déterminer quel type de relation créé entre l’événement et d’autres entités. Par exemple, si un événement contient la description “le Canada déclare la guerre à l’Allemagne” et que le terme “Canada” est lié à un autre article sur Wikipédia, nous effectuons une recherche de ce terme dans DBpedia (<http://dbpedia.org/resource/Canada>) afin d’identifier le type de cette ressource (<http://dbpedia.org/ontology/Place>). Dans ce cas, nous utiliserons la propriété `atPlace` pour décrire l’événement. Si DBpedia ne retourne aucun type que nous pouvons utiliser, nous qualifions par défaut la relation avec `involves`.

Nous avons extrait un premier ensemble d'événements à partir des quatre chronologies suivantes :

- “Chronologie de la Seconde Guerre mondiale” qui contient sept chronologies par année avec des événements impliquant des individus et une granularité journalière ;
- “Chronologie de l’Irlande” qui contient des événements pour un même lieu géographique recouvrant une très large période temporelle allant de l’âge de pierre à nos jours ;
- “Chronologie des attaques du 11 septembre” qui contient 147 événements sur une seule journée ;
- “Histoire évolutive du vivant” qui permet de tester notre modèle sur des événements grossiers et remontant à très loin dans le temps.

Interopérabilité avec des collections d'événements existantes. Pour évaluer la pertinence des alignements entre le modèle LODE et d'autres ontologies, nous avons intégré notre jeu de données composé des événements extraits à partir de Wikipédia avec deux autres collections composées d'événements décrits selon les ontologies EO et BIO⁴. Notre objectif est d'être capable d'interroger et de naviguer dans une base d'événements en utilisant des interfaces génériques visualisant le web de données telles que le serveur sémantique Cliopatria (Wielemaker *et al.*, 2008). Pour configurer ce dernier outil, nous avons simplement défini comme facette les différentes dimensions composant la description d'un événement.

Biographies du congrès. Le répertoire du congrès américain contient des articles bibliographiques courts, ainsi que l'ensemble des événements ayant trait à la vie du congrès, pour toutes les législatures des États Unis, de 1774 jusqu'à nos jours. La structure très homogène de ces événements rend facile leur extraction. Dans des travaux précédents, nous avons ainsi extrait 69228 événements décrits selon l'ontologie BIO.

La chronologie de Emma Goldman. Les éditeurs de l'anarchiste politique ont maintenu une chronologie au jour le jour des activités et de la localisation d'Emma Goldman et de ses associés. Cette chronologie sert en fait de document de référence pour permettre aux éditeurs d'établir où et quand ses oeuvres ont été produites et pour éventuellement détecter des inconsistances dans les livres d'histoire. Nous avons traité l'ensemble des années 1910 à 1916 et nous avons produit un jeu de données RDF en analysant ces dates, en géo-localisant les lieux et en disambiguant les personnes en utilisant DBpedia. En tout, 1041 événements de la vie d'Emma Goldman sont ainsi décrits en utilisant l'ontologie EO.

Problèmes rencontrés. Pour intégrer ces différents jeux de données, nous avons simplement utilisé les axiomes logiques définis dans LODE établissant un alignement entre les vocabulaires BIO et EO. Cliopatria ne permet cependant pas de visualiser immédiatement le résultat intégré, les inférences liées à l'utilisation des axiomes owl : `equivalentClass` et `owl:equivalentProperty` n'étant pas supportées. Nous avons donc simplement décomposé ces axiomes en deux axiomes utilisant `rdfs:subClass` (resp. `rdfs:subProperty`) pour obtenir le même résultat.

4. <http://vocab.org/bio/0.1/>

5 Conclusion et perspectives

Il y a une quantité phénoménale d'informations chronologiques sur le web. Les historiens et les journalistes sont de plus en plus intéressés pour présenter leurs travaux sous forme de données très structurées en complément d'une narration. Pourtant, sans aucun effort pour rassembler les différents modèles permettant de décrire des événements, il sera difficile de construire et de visualiser des réseaux denses d'information qui relient les événements. Nous avons proposé dans cet article le modèle LODE en se basant sur une analyse fine des ontologies existantes. Nos premières applications montrent que ce modèle est suffisant pour décrire des chronologies très variées et pour intégrer des données déjà existantes utilisant d'autres vocabulaires. Certains problèmes restent cependant à résoudre. Nous avons argumenté qu'une ontologie dédiée à la représentation des événements ne devait inclure que les propriétés pour lesquelles un certain consensus peut être établi et laisser à des ontologies spécifiques aux applications le soin de représenter les propriétés nécessitant une certaine interprétation. Mais avons-nous correctement identifié ces propriétés ? En particulier, la participation à un événement n'est-elle pas aussi subjective et donc une propriété qui devrait être exclue de ce modèle ?

Nous planifions de continuer à tester l'expressivité de notre modèle en convertissant de larges bases de données d'événements à venir telles que Eventful⁵ ou Upcoming⁶. En particulier, nous sommes en train d'imaginer et d'évaluer des interfaces centrées événements pour explorer, visualiser, comparer des contenus multimédia partageant des lieux, des personnes ou des thématiques communes.

Références

- DOERR M. (2003). The CIDOC Conceptual Reference Module : An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3), 75–92.
- GANGEMI A. & MIKA P. (2003). Understanding the Semantic Web through Descriptions and Situations. In *2nd International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE'03)*, p. 689–706, Catania, Italy.
- HOBBS J. & PAN F. (2006). Time Ontology in OWL. W3C Working Draft. <http://www.w3.org/TR/owl-time>.
- LAGOZE C. & HUNTER J. (2001). The ABC Ontology and Model. *Journal of Digital Information (JoDI)*, 2(2).
- RAIMOND Y., ABDALLAH S., SANDLER M. & GIASSON F. (2007). The Music Ontology. In *8th International Conference on Music Information Retrieval (ISMIR'07)*, Vienna, Austria.
- SCHERP A., FRANZ T., SAATHOFF C. & STAAB S. (2009). F—A Model of Events based on the Foundational Ontology DOLCE+ Ultra Light. In *5th International Conference on Knowledge Capture (K-CAP'09)*, Redondo Beach, California, USA.
- SHAW R., TRONCY R. & HARDMAN L. (2009). LODE : Linking Open Descriptions of Events. In *4th Asian Semantic Web Conference (ASWC'09)*, p. 153–167, Shanghai, China.
- VAN HAGE W., MALAISÉ V., DE VRIES G., SCHREIBER G. & VAN SOMEREN M. (2009). Combining Ship Trajectories and Semantics with the Simple Event Model (SEM). In *1st ACM International Workshop on Events in Multimedia (EiMM'09)*, Beijing, China.
- WIELEMAKER J., HILDEBRAND M., OSSENBRUGGEN J. v. & SCHREIBER G. (2008). Thesaurus-based search in large heterogeneous collections. In *7th International Semantic Web Conference (ISWC'08)*, p. 695–708, Karlsruhe, Germany.

5. <http://www.eventful.com/>

6. <http://www.upcoming.org/>