CrossMark

Scholarly Ontology: modelling scholarly practices

Vayianos Pertsas[1,2] · Panos Constantopoulos[1,2]

Abstract In this paper we present the Scholarly Ontology (SO), an ontology for modelling scholarly practices, inspired by business process modelling and Cultural-Historical Activity Theory. The SO is based on empirical research and earlier models and is designed so as to incorporate related works through a modular structure. The SO is an elaboration of the domain-independent core part of the NeDiMAH Methods Ontology addressing the scholarly ecosystem of Digital Humanities. It thus provides a basis for developing domain-specific scholarly work ontologies springing from a common root. We define the basic concepts of the model and their semantic relations through four complementary perspectives on scholarly work: activity, procedure, resource and agency. As a use case we present a modelling example and argue on the purpose of use of the model through the presentation of indicative SPRQL and SQWRL queries that highlight the benefits of its serialization in RDFS. The SO includes an explicit treatment of intentionality and its interplay with functionality, captured by different parts of the model. We discuss the role of types as the semantic bridge between those two parts and explore several patterns that can be exploited in designing reusable access structures and conformance rules. Related taxonomies and ontologies and their possible reuse within the framework of SO are reviewed.

✉ Vayianos Pertsas
vpertsas@aueb.gr

1 Department of Informatics, Athens University of Economics and Business, Athens, Greece

2 Digital Curation Unit, IMIS, Athena Research Centre, Athens, Greece

1 Introduction

The past two decades have witnessed significant changes in the way research is conducted by virtue of the use of information and communication technologies (ICT), changes encoded by terms like "e-science" and "d-science" [1]. This includes computationally intensive research, requiring interdisciplinary collaboration to tackle new kinds of problems in the study of complex systems, such as in science, engineering and medicine, as well as very large-scale data management and knowledge extraction. In parallel, Digital Humanities are emerging to incorporate these developments in the way scholarly work is conducted in the humanities. In all fields several European initiatives are currently developing large-scale digital research infrastructures that aim to bring together research resources, tools and services across Europe [2]. The emergence of such e-Infrastructures raises important questions as to how well their services fit the needs of the actual research life-cycle [3]. It is not sufficient to know which particular functionalities scholars want of digital infrastructures, or how they currently use digital tools and services, but rather, it is necessary to study these in the context of a broader user-centred perspective on scholarly research processes, so as to ensure that actual information needs are addressed [4] and working practices enhanced. In fact, the need to examine the humanities research process as a special kind of 'business process' that can be systematically analysed and recorded has been identified as early as the 1990s [5], where research is approached as a multi-stage iterative process involving series of tasks that typically take place in sequence but may also occur out of order. Furthermore, ontologies, with a pivotal role in e-Science, can provide the conceptual framework in which scientific processes and workflows can be structured, annotated and shared to become interoperable, inform sci-

entific reasoning and provide content and context for online dialogue in virtual communities [6–8].

In this context, we argue that an ontological framework for understanding and representing scholarly practices as a special kind of business processes can be instrumental to the success of a digital research infrastructure. Along this line we here introduce the Scholarly Ontology (SO), a conceptual framework intended to represent the domain of scholarly work in the digital age. The aim of the ontology is to provide a flexible framework for modelling scholarly practices while its scope covers the entire scholarly ecosystem. Specifically, the intuition behind SO is to address the aforementioned digital-era information needs by building on top of a solid, time-independent core ontological framework, properly mapped to upper foundational ontologies. This core framework can further be extended to cover fine-grained aspects of the entire scholarly domain, by either incorporating discipline-specific controlled vocabularies, through SO's type taxonomies, or by reusing other domain-specific ontologies that further specialize SO's classes/properties.

The paper proceeds as follows: Sect. 2 presents background-related literature that influenced the construction of SO; in Sect. 3 we present the ontology, commenting on its development method, explaining the rationale behind its structure as well as the core concepts along with their semantic relationships, and a use case demonstrating its potential usage; in Sect. 4 we illustrate the use of SO through a number of indicative queries that exploit its formalization in RDFS; in Sect. 5 we discuss several patterns and conformance rules, as well as possible extensions through the reuse of other related ontologies and we give a first account on preliminary user feedback and validation of the model; and in Sect. 6 we make concluding remarks.

2 Background

Understanding knowledge production and the research process has been the object of inquiry from several perspectives. In information science, works such as [9–11] identify scholarly activities and argue that, if viewed at an appropriate level of abstraction, scholarly research working practices tend to involve a finite, in fact rather small, set of fundamental processes, called scholarly primitives, common across disciplines. In social anthropology, Cultural-Historical Activity Theory (CHAT) [12], approaches the notion of activity as an intentional act of a subject that, using various physical or conceptual resources, interacts with other objects of the world to fulfill some motive or address a specific need. An activity system is then regarded as a hierarchy of activities, composed of conscious acts designed to meet hierarchically structured goals.

Goals, expressed as dependencies among actors or means-end and task-decomposition links, also play an important role in business process re-engineering and requirements analysis [13,14]. In addition, intentionality aspects of activities are addressed extensively in the digital libraries evaluation domain [15] with key aspects modelled as evaluation dimensions comprising, effectiveness, performance measurement, service quality, outcomes assessment and technical excellence along with their corresponding types based on the phases during which they take place.

In business process modelling, the notions of process and procedure expressing the 'what' and 'how' of actors' actions play a key role in modelling organizations [16,17]. Moreover, specialized ontologies from the enterprise domain capture the above notions through the use of special classes such as "commitment, authority, goals, agents", etc. [18] or express operational functions of an enterprise with subject areas such as "marketing, strategy, planning and organization" [19]. 'How to represent processes' is also a key intellectual challenge in management science, where notions of specialization and coordination are shown to provide significant leverage in developing and reasoning over process ontologies and process databases [20].

Processes as activities with temporal and spacial aspects are also crucial in the cultural domain, with the event-centric CIDOC CRM ontology [21] being the standard (ISO21127) for facilitating the integration, mediation and interchange of heterogeneous cultural heritage information. Carefully engineered and with extensive empirical grounding, the CIDOC CRM comprises a very substantial general ontology part which makes it a versatile reference ontology suitable to serve as the basis for developing various domain ontologies.

In the same perspective, the Unified Foundational Ontology (UFO) [22] formalizes the philosophical notions of Endurants (i.e. entities whose identity persists through time) and Perdurants (i.e. entities that are intrinsically temporal and whose identity does not persist through time), while including Agents (i.e. entities that cause or participate in Events) to fit the purpose of codifying the ontological foundations of business process.

In e-Science, scientific processes expressed as workflows play a fundamental role in modelling research activities and data flows [23], with systems such as Taverna[1] supporting scientists using Grid technology to conduct in-silico experiments in biology. The proliferation of such workflow systems has led to recent studies that attempt to categorize their features to assist end users according to their needs and practices [24].

Similar studies [3,25,26] analyse the working practices of researchers in the emerging field of Digital Humanities and determine the user requirements for digital infrastructures

---

[1] http://www.taverna.org.uk/.

such as the *European Holocaust Research Infrastructure* (EHRI) and the *Digital Research Infrastructure for the Arts and Humanities* (DARIAH). On the other hand, the *Network for Digital Methods in the Arts and Humanities*[2] (NeDiMAH) includes in its agenda an extensive charting of digital resources, methods, activities and tools, i.e. the environment and the processes of Digital Humanities.

A major part of this initiative is the development of the NeDiMAH Methods Ontology (NeMO) [27], in order to articulate the scholarly ecosystem of Digital Humanities and support an environment for documentation of research practices and methods of the field. NeMO provides a common methodological layer for arts and humanities researchers to develop, refine and share research methods that allow them to create and make best use of digital tools and collections. In addition, it promotes academic credibility for the area by supporting peer-reviewed scholarship and developing a commonly agreed nomenclature in the nascent field of Digital Humanities. In this context, NeMO can support maximizing the value of national and international e-research infrastructure initiatives as well as eliciting and prioritizing the functional requirements for planned digital infrastructures in the arts and humanities, following an evidence-based, user-centred approach.

NeMO was conceived as a layered structure comprising an upper layer including the most general concepts, a middle layer adding detail but still applicable across humanities domains and a lower layer intended to capture the details that differentiate domains and subjects of interest. This progressive detail is expressed by means of class and property specialization relations and controlled hierarchical vocabularies. It became apparent that the core concepts in NeMO, by virtue of their generality, may be applicable in modelling work in domains beyond the humanities as well and that it makes sense to pursue an elaboration of those core concepts. The outcome of this endeavour is the Scholarly Ontology (SO) presented here. A deductive framework can subsequently be applied, whereby NeMO and other domain-specific ontologies of scientific work can be derived as extensions of the SO backbone.

3 Scholarly Ontology

The Scholarly Ontology (SO) is inspired by the Cultural-Historical Activity Theory, grounded on evidence concerning the working practices and information behaviours of scholars, and views scholarly work as a special kind of 'business process' (see above). The ontology is event-centric since it is built around a central notion of activity and combines three perspectives: the agency perspective, concerning actors

and intentionality; the procedure perspective, concerning the intellectual framework and organization of work; and the resource perspective, concerning the material and immaterial objects consumed, used or produced in the course of activities.

3.1 Ontology development method

The design and development of the ontology was an iterative process with several repetitions of the following basic steps. It was carried out during the development of NeMO with additional rounds of steps 2, 3 and 5 for elaborating the SO core:

1. *Grounding*: The ground data supporting the validity of the ontology come from earlier empirical research using semi-structured interviews with scholars from across Europe that focused on analysing the research practices and capturing the information requirements of research infrastructures [3,25]. In addition we took into account earlier relevant models of scholarly research activity [28], as well as existing taxonomies from the interdisciplinary field of Digital Humanities (see Sect. 5.3).
2. *Domain Conceptualization*: Based on the analysis of the ground evidence, core concepts and relationships of the domain were identified by a team of analysts.
3. *Ontology Design*: Bearing in mind the existing related works (see Sect. 5.3) as well as reference ontologies [21], a first version of the ontology was constructed by a team of information and computer scientists and tested by a broader team of scholars from several disciplines, subsequently undergoing several rounds of elaboration. In this stage, modelling decisions regarding the layered architecture of the ontology were made.
4. *Controlled Vocabularies Construction*: With the first two layers of the ontology being relatively stable, the controlled vocabularies (CVs) of the third layer could be defined. In this stage, definitions in textual form as well as examples and mappings of terms of the ontology to and from terms of other taxonomies were developed.
5. *Ontology Formalization*: A machine-readable formalization was created in RDFS (RDF schema), to enable the use of the ontology in a wide range of applications accessing registries and knowledge bases. Furthermore the taxonomic parts of the ontology were designed in compliance with Simple Knowledge Organization System (SKOS).
6. *Ontology Validation and User Feedback*: Presentation of the ontology in several workshops and continuous elaboration and gathering of case studies along with relative questions that are asked by community members led to further refinement of the ontology concepts and proper-
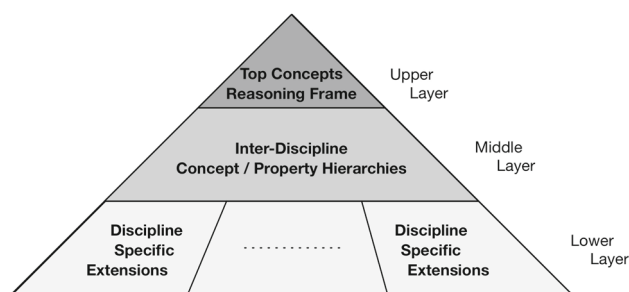

Fig. 1 Ontology structure

ties as well as the terms of the discipline-specific CVs (see Sect. 5.4).

3.2 Ontology structure

Architecturally, SO adopts a three-layered structure from abstract/general to concrete/special concepts to provide a flexible framework, adaptive to the multidisciplinary domain of scholarly work. As presented in Fig. 1 the ontology structure consists of the following:

– The *Upper Layer* that contains the most general concepts and properties acts as a frame of reference and provides the basis for compatibility with other reference ontologies [21,22].
– The *Middle Layer* that contains the hierarchical structures of more specific but still quite broad properties and concepts which are common across disciplines in the scholarly domain.
– The *Lower Layer* that contains the fine-grained aspects of research practices as well as various controlled vocabularies, specific to each aspect of scholarly work or scientific disciplines.

SO comprises the upper and middle layers. NeMO, on the other hand, consists of SO and a lower layer containing any domain-specific extensions of the middle layer concepts, as well as the relevant controlled vocabularies. Likewise, scholarly work ontologies for areas other than the humanities can be generated from SO by developing appropriate lower layer components. In the sequel we focus on the core concepts comprised in SO.

3.3 Concepts

All SO classes are considered subclasses of a top abstract class, *SO_Entity*, from which they inherit the basic properties *identifier, type* and *description*. In particular, the type property allows using arbitrary vocabularies and taxonomic schemes, thus enabling flexible characterization in parallel with a distinction of ontological classes, while the identi-

fier and description properties allow for identification by name or preferred identifiers and free text description, respectively. Further, meronymic decompositions generally apply. For instance, as seen below, *Groups* have *Persons* as members and may have internal hierarchical structures, *Activities* may consist of sub-activities and *Objects* may comprise other Objects as parts, subject to the constraint of not mixing material and immaterial ones. Figure 2 presents the hierarchy of SO Classes briefly explained below:

The most abstract/top entities of the ontology are *Actor, Event* and *Object*, expressing, respectively, the foundational ontological concepts of Agents, Perdurants and Endurants. Further specialization of Objects in the general subclasses of *ConceptualObject* and *PhysicalObject* classifies Endurants according to their nature. The above entities constitute the SO Upper Layer, comprising general concepts of foundational value, independent of domain. In fact, the concepts of this upper layer can also function as semantic links to foundational ontologies such as the Unified Foundational Ontology (UFO), concerning the domain of business process modelling [22] or CIDOC CRM [21], concerning the cultural domain. Specifically, SO:Event corresponds to UFO:Event and CRM:Event concepts; SO:Actor specialises the CRM:Actor and UFO:Agent classes, while SO:Object, the CRM:Thing and UFO:Endurant. Further specializations of Object follow the CRM classification with corresponding concepts the CRM:ConceptualObject and CRM:PhysicalObject respectively. Other abstract/top entities that are included in the hierarchies of [21,22] are not depicted in Fig. 2 for readability reasons. Moreover, the names of concepts in the second layer, although generic, adhere to the specifications of the scholarly domain since they belong to the SO name-space.

Actors are entities that can perform intentional acts for which they can be accounted or referenced. This distinguishes actors from tools and machines, which can only react to human intentions [29]. Actors can *participate* in activities, actively or passively, in one or more roles. *ActorRoles* are characterizations of the behaviour of an actor in a particular context [30]. The Actor class is further specialized in the subclasses *Person* and *Group* for, respectively, representing individual persons and collective entities in the Scholarly domain.

Activities are intentional acts carried out by instances of the Actor class; they have duration and occur at a specific time and place. They are real processes, as opposed to plans, or procedures for carrying out processes. *Projects* and *Courses* are two kinds of activity of particular interest in the scholarly domain that warrant specialized descriptions and are represented as subclasses of Activity.

Objects are discrete, identifiable, persistent items. They can be material, such as statues or computers, or immaterial,
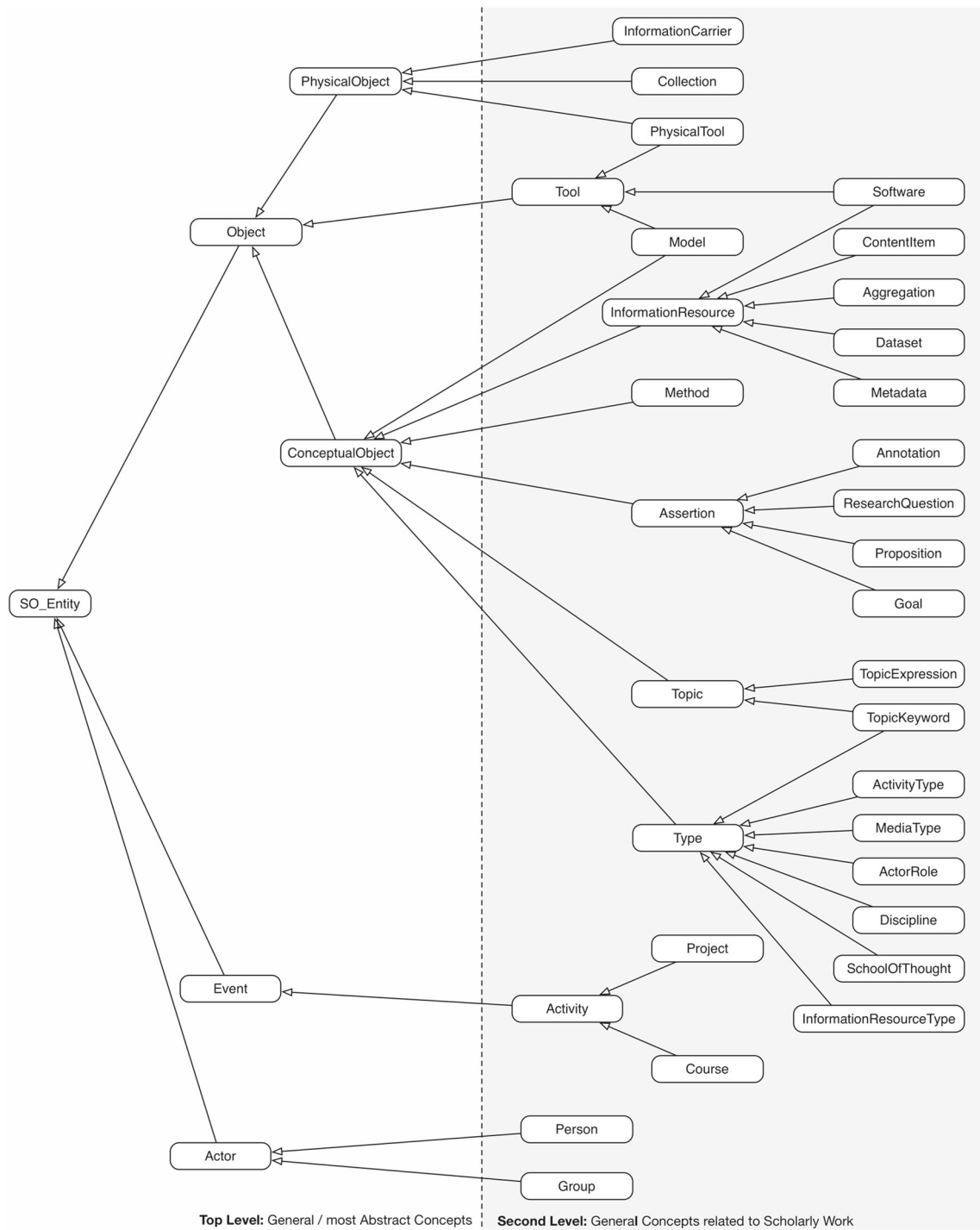
Fig. 2 SO class hierarchy. The arrows represent subclassOf relationships

such as images, texts or organizational structures. Objects are involved in activities during which they are created, used, modified or destroyed, assuming specific roles. In SO we distinguish three broad categories of object involvement in activities: *input, output* and *tool*.

PhysicalObject comprises material objects, man-made or natural. Subclasses of *PhysicalObject* relevant in the present

context include: *Collection*—groups of physical objects collected by an actor for some purpose; *PhysicalTool*—physical objects that are used (but not consumed) in carrying out activities; *InformationCarrier*—physical, man-made objects designed to serve as carriers of Information Resources.

ConceptualObject comprises immaterial objects conceived in the human mind, which become objects of discourse.

They are borne by possibly multiple physical carriers, such as marks, paper, solid-state memory, or human memory and only cease to exist when the last carrier is destroyed. Subclasses of Conceptual Object of specific significance in the present context include *InformationResource, Type, Method, Model, Proposition* and *Topic*.

InformationResource comprises conceptual objects consisting of symbols and conveying propositions about things in a domain of discourse, e.g. data sets, texts, images, computer programs, vocabularies, sound or video recordings, mathematical expressions, etc. Information resources capture the discrete manifestations of conceptual objects on specific man-made carriers, have reproducible expressions and are borne by information carriers, yet they exist independently of those carriers. Groups of information resources can be modelled as *Aggregations*. Other subclasses consist of *Software* that comprises programs and machine readable code, *Dataset* for representing the contents of databases or matrices and *ContentItem* for the rest of the information resources that appear in various human-readable forms, e.g. images, sounds, texts, mathematical expressions, etc.

Model includes any kind of abstract representation, most notably information models. *Topic* comprises free expressions in natural language or keywords from controlled vocabularies, describing what the referred items are about and used as index terms. *Method* comprises specifications, procedures or recipes for carrying out activities, to be distinguished from those, as well as from activity types (see Sec. 3.5).

Type comprises conceptual objects used to characterize instances of entity classes and denoted by controlled terms, thus providing a powerful, flexible classification mechanism. Type subclasses of particular relevance in the context of SO are *ActivityType*—induces a taxonomic scheme for types of activities; *InformationResourceType*—characterizes information resources (see below); *MediaType*—lists the formats in which information resources are stored; *Discipline*—scholarly disciplines; *SchoolofThought*—different schools of thought that influence researchers especially in the humanities; *TopicKeyword*—various thematic keywords as found in conferences, scientific journals, etc.; *ActorRole*—different roles that characterize actors participating in activities.

Assertion includes all kinds of assertions in the scholarly domain and captures the intellectual essence of scholarly activity. *Annotation* is a subclass of Assertion including textual comments, ratings, classifications, comparisons and associations regarding specific objects of a domain. The basic distinction of annotations from other kinds of assertions lies in their existential dependence (the annotated object). On the other hand, their operational scope can be significantly
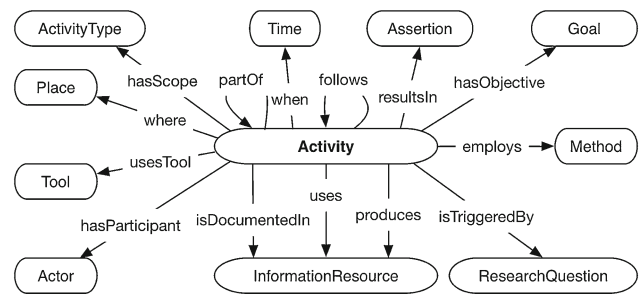

Fig. 3 Activity perspective

enhanced if supported by an intentional annotation model [31]. The *Goal* subclass of *Assertion* enables the formation of systematic collections of explicitly stated research goals. Finally, the *Proposition* and *ResearchQuestion* subclasses comprise assertions in affirmative or interrogative form respectively.

Tool aggregates objects such as physical tools, software and models that can be or have been used in carrying out activities. Its inclusion in the hierarchy admits a pragmatic rather than purely ontological justification.

We now present indicative semantic relations of the ontology through four complementary views: one centred on activity, and the other three corresponding to the perspectives of procedure, resource and agency. For readability reasons an overview of SO classes along with their corresponding properties is offered as Appendix. A complete documentation of all SO entities, along with taxonomies and mappings to other works, can be found in [27]. Furthermore, properties inherited from parent CIDOC CRM classes are generally not shown, unless they are of special significance or overriding applies.

3.4 Activity perspective

As shown in Fig. 3, the *partOf* property is used for modelling meronymic decompositions of activities. Causal ordering is expressed by the *follows* property. As mentioned in Sect. 3.3, the SO distinguishes three broad kinds of object involvement in an activity: input, output and tool. These are further specialized by the properties *produces, uses* and *isDocumentedIn* regarding the involvement of information resources and the properties *triggeredBy, hasObjective* and *resultIn* regarding the involvement of assertions or their specializations.

Other existing ontologies such as the Time Ontology[3] and BasicGeo[4] can be reused here to model the temporal and basic spatial aspects of activities. This is achieved by further
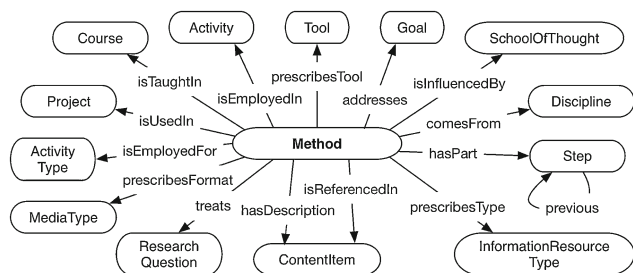
---

[3] http://www.w3.org/TR/owl-time.

[4] http://www.w3.org/2003/01/geo.

Fig. 4 Procedure perspective

Fig. 5 Resource perspective

Fig. 6 Agency perspective

specializing the *Time* and *Place* classes and the *when* and *where* properties, respectively.

The *ActivityType* class is essentially a hierarchically organized controlled vocabulary, spanning the generic research activity life-cycle as described in the information seeking behaviour literature [9–11] and is used in SO to express the general scope that a specific activity has, through the *hasScope* property. The entire ActivityType taxonomy—currently comprising 161 terms-together with their definitions and mappings from and to other relevant taxonomies can be found in [27].

3.5 Procedure perspective

Like in business process modelling, we distinguish between the procedure that prescribes how to perform a specific act and the act itself. In SO procedures or 'recipes' are captured by the *Method* class (see Fig. 4), while acts are captured by the *Activity* class. A method *hasDescription* explaining what it does, *isEmployedIn* an activity, *comesFrom* some discipline, may be *influencedBy* a school of thought, be *referencedIn* bibliography (a content item), or *taughtIn* some courses. A structured description of a method is enabled by the *hasPart* property yielding a recursive analysis into steps and the *previous* property establishing causal ordering of steps.

A method may *prescribe* specific information resource types for inputs and outputs, media types as formats and tools. Finally, methods are designed to *address* specific goals and *treat*—through their employment-specific research questions.

3.6 Resource perspective

Any conceptual object that has a concrete representation, borne by man-made carriers, is considered as a unit of information, independently of those carriers (paper, hard disk, etc.) and is treated as a resource that can be characterized by its topic, type, and format or be described by a set of metadata. In addition, information resources can be used as inputs or outputs of activities. *InformationResourceType*, and *Medi-*
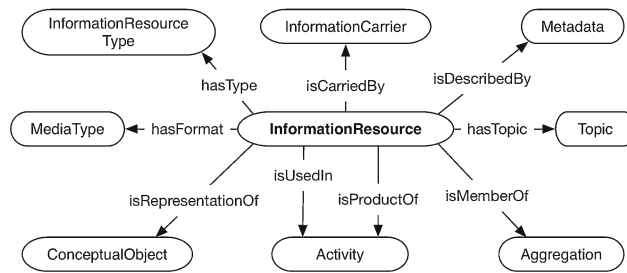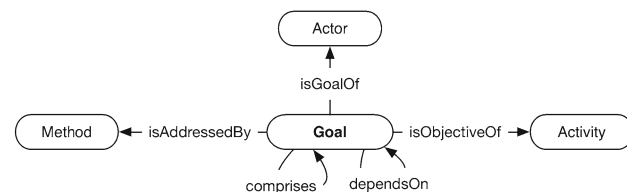
aType constitute controlled vocabularies describing the type and format of information resources and can be imported from authorities, such as the Marc21[5] bibliographic standard and the Internet Assigned Numbers Authority (IANA[6]). Instances of *InformationResource* can be grouped in aggregations and modelled through the use of an established data model, such as OAI-ORE.[7] The properties of *InformationResource* are shown in Fig. 5.

3.7 Agency perspective

The agency perspective captures the 'who' and 'why' aspects of the domain by focusing on goals of actors and the intentional context of other relationships. Existing ontologies such as FOAF[8] and SiOC[9] can be reused in conjunction with SO to capture the social and community aspects of actors. This can be achieved by further specializing the corresponding SO concepts: *Person*, *Group*, *Project*, *Topic*, *ActorRole* and *InformationResource*. As depicted in Fig. 6, actors' goals constitute the *objectiveOf* activities and can be *addressedBy* instances of *Method*. In addition, they can be further decomposed into more refined goals or be dependent upon other goals as indicated by the *comprises* and *dependsOn* properties, respectively. Thus the notion of goal captures the successive refinement from high-level objectives down to narrower goals, as well as the manifestation of chains of dependency among goals. The notions of goal and topic together enable representing the research context.

The explicit specification of this class provides a straightforward mechanism for representing goals especially in cases involving instances of *Project* or *Course*, which frequently have predetermined goals. However, in cases where such an explicit specification is superfluous, terms from the *ActivityType* taxonomy can be exploited to capture the intentional context (see Sect. 5.1).

3.8 Use case example

By way of example we present a use case of SO in modelling a particular research activity for which a detailed textual account is available in published form [32]. Two researchers (the authors of [32]), used computational linguistic methods to analyse popular songs composed by Japanese female singer-songwriters. They gathered a sample of 116 song lyrics and they employed "Random Forests"—a machine learning method— to perform the classification experiments and extract important features regarding the distinctive lyrical characteristic of each singer-songwriter. Each activity used/produced information resources and resulted in several propositions that constitute the analysis on the subject and are represented through various tables, figures or text in the published paper.

The two researchers are modelled as instances of the Person class [Ac_1: Takafumi Suzuki] and [Ac_2: Mai Hosoya] with roles [R_1: Associate Professor] and [R_2: Researcher], respectively. Indicative instances of the Activity class are the general activity [A_1: Analysed popular songs] which are decomposed into its sub-activities: [A_2: Gathered 116 Songs] followed by [A_3: Applied Random Forests]. These can further be linked through the *hasScope* relationship with the *ActivityType* terms [Analysing], [Gathering] and [Classifying], respectively.

The method [M_1: Random forests]—as described in [32]—consists of 3 steps, modelled here as [$St_{1.1}$: Sample from i cases at random from the original text-feature matrix M[i,j]], followed by [$St_{1.2}$: Extract random subsets of [$root j$] variables from a bootstrap sample to make a sample for constructing an unpruned decision tree], which is followed by [$St_{1.3}$: Calculate the variable Importance (VIacu) for the classification experiments].

Other indicative elements of the model are [Software: MeCab, Uta-Map, Uta-Net], [Proposition: Pronouns, final particles, and auxiliary verbs are particularly important for discriminating the songs by ten Japanese female singer-songwriters], [ContentItem: Figs. 1 and 2, dataset of 116 song lyrics], [Goal: Gather a representative number of Songs as input for the Experiment], [Topic: Computational Stylistic Analysis of Popular Songs of Japanese Female Singer-songwriters], [Discipline: Computer Science], connected accordingly. Figure 7 presents a visualization of the above

as a graph. Several other use cases along with their graph visualizations can be found in [27].

4 Implementation and use

With a view to linked data usage and given the availability of various reasoned implementations [33,34], the formal expression of SO was done in RDFS. This formalization supports the development of an environment of inter-operable resources and services for discovering, understanding, selecting, linking and contributing content, tools and methods and also the use of SO as "semantic glue" between different existing taxonomies and controlled vocabularies in various disciplines. Further, SO is serialized in OWL/XML which provides for complex query answering, using appropriate query languages.

As mentioned in Sect. 3.1, during the validation stage of the ontology, a series of questions—representative of users' information needs—was gathered and properly formulated into queries, to validate the capabilities of the ontology, with respect to answering the different types of questions related to scholarly practices and research/scientific workflow. Elaboration and refinement based on those queries led to the current version of the model. Along this line, we present below indicative query examples in SPRQL v.1.1, as well as more complex query structures expressed in SQWRL, to illustrate the potential of SO in addressing the needs of representing knowledge and reasoning about scholarly practice.

4.1 Indicative SPRQL queries

Query 1 *Given a particular research question, retrieve all the related, existing information from the literature where it is described how this research question is addressed.* E.g. What are the individual characteristics of an Artist that are expressed in the lyrics without people noticing? ('RQ1').

```
SELECT DISTINCT ?Method ?Activity ?ContentItem
?Actor ?Proposition WHERE{
?Method so:isEmployedIn/so:isTriggeredBy |
so:isEmployedIn/so:isPartOf+/
so:isTriggeredBy |
so:isReferencedIn/
so:providesDocumentationFor/
so:isTriggeredBy so:RQ1 .
?Activity so:isTriggeredBy so:RQ1 .
OPTIONAL{?ContentItem
so:providesDocumentationFor/
so:isTriggeredBy | so:ProvidesReferenceFor/
so:isEmployedIn/so:isTriggeredBy so:RQ1.}
OPTIONAL{?Actor so:participatesIn/
so:isTriggeredBy so:RQ1 .}
OPTIONAL{?Proposition so:isResultOf/
so:isTriggeredBy |
so:isResultOf/SO:PartOf+/so:isTriggeredBy
so:RQ1 .} }
```
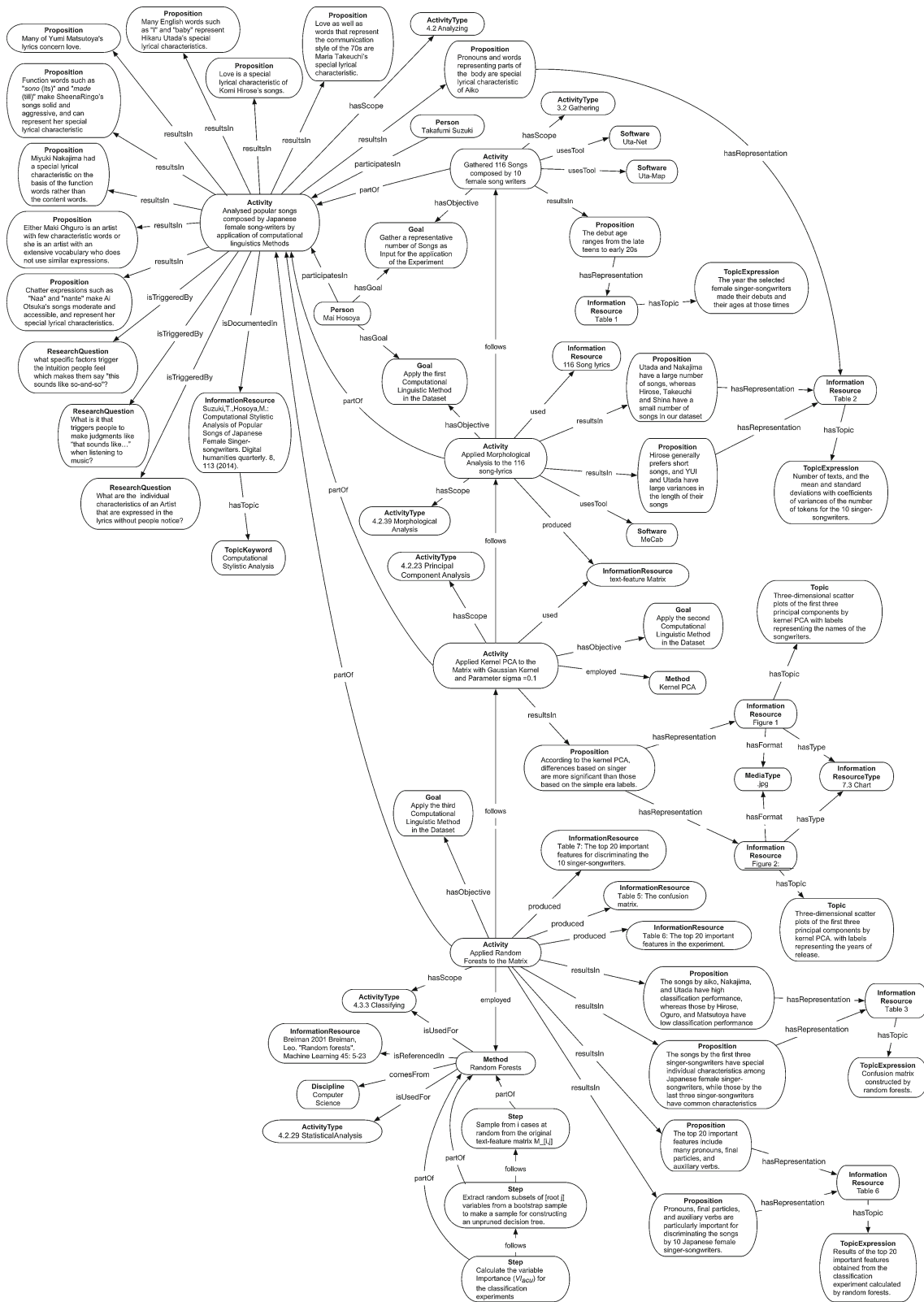
Fig. 7 Graph visualization of the modelling example

Note that the general question of Query 1 is analysed using the appropriate SO classes to address methods that either have been employed directly by the research activities triggered by the specific research question, or are referenced in the same literature where the above research activities are documented. Apart from the methods prescribing how activities corresponding to the specific research question are conducted, the relevant activities, describing the actual cases where this question was addressed can also be retrieved. Furthermore, additional information regarding the above, such as the actors that have been engaged or statements concerning the results of those activities can be presented, if possible, through the OPTIONAL clause. In case of transitive property such as *partOf*, using the + symbol tells the query processor to keep looking for activities that are part of other activities until it finds the requested input or runs out of entities interrelated with the specified property. Finally, resources such as texts, images, video, etc. that either document the bound research activities or provide reference for the related methods can also be presented optionally, enhancing the final results.

Query 2 *Find all relevant information regarding a specific topic of interest.* E.g.: Computational Stylistic Analysis ('TK1').

```
SELECT ?Method ?Activity ?ContentItem
?Proposition ?Actor ?Tool WHERE {
?Method so:regards |
so:isEmployedIn/so:hasParticipant/
so:hasInterest |
so:isEmployedIn/so:isPartOf+/
so:hasParticipant/so:hasInterest
so:TK1 .
?Activity so:hasParticipant/so:hasInterest |
so:employs/so:regards |
so:isDocumentedIn/so:hasTopic so:TK1 .
?ContentItem so:hasTopic so:TK1 .
OPTIONAL{?Proposition so:isResultOf
so:?Activity .}
OPTIONAL{?Actor so:hasInterest |
so:participatesIn/so:employs/
so:regards so:TK1 .}
OPTIONAL{?Tool
so:isUsedAsToolFor/so:hasParticipant/
so:hasInterest |
so:isUsedAsToolFor/
so:employs/so:regards so:TK1 .} }
```

In Query 2 the question is decomposed using SO classes, to partial questions concerning methods that regard a specific topic or are employed in the research activities that address that topic. The actual activities can also be retrieved in cases where either their participants took interest in that particular topic, or the methods they employed regard the specific topic, or the resources that documented them had as topic the indicated input. Optionally, any relevant material such as the propositions that express the results of the bound activities, the actors who share the same interest, or the tools that were

used or prescribed can also be retrieved, enhancing the final results.

Query 3 *Find all the researchers that share the same interest and retrieve any relevant information about them.* E.g.: Computational Stylistic Analysis ('TK1').

```
SELECT ?Person ?Activity ?Statement
?Method ?Tool ?ResearchQuestion WHERE {
?Person so:hasInterest |
so:participatesIn/so:employs/
so:regards |
so:participatesIn/so:isDocumentedIn/
so:hasTopic |
so:participatesIn/so:isPartOf+/
so:isDocumentedIn/so:hasTopic
so:TK1 .
?Activity so:hasParticipant ?Person .
OPTIONAL{?Statement so:isResultOf ?Activity .}
OPTIONAL{?Method so:isEmployedIn ?Activity .}
OPTIONAL{?Tool so:isUsedAsToolFor ?Activity .}
OPTIONAL{?ResearchQuestion so:Triggers
?Activity .} }
```

Note that in this case the query is relaxed since in addition to the direct relationship between an actor and the indicated topic keyword, alternate property paths are checked. Here too, the + sign is used to express the transitivity of the meronymic decomposition of activities. Furthermore, for the retrieved persons, additional information regarding the activities that he/she participated or—optionally—the employed methods, tools, research questions and statements of those activities, can also be retrieved.

4.2 Creating sets of SO entities with SQWRL

SPARQL has no native understanding of OWL semantics since it operates only on its RDF serialization [35]. On the other hand, SQWRL is built on the SWRL rule language [36] that is designed as an extension to OWL and incorporates all of its semantics. In addition it takes advantage of the built-ins of SWRL to define set operators that can be used for retrieval specifications. So built-in operators such as the sqwrl:makeSet or sqwrl:union, sqwrl:graterThan can be used to produce more complex queries taking advantage of the cardinality and set theoretic properties of owl constructs as in Queries 4 and 5 below:

Query 4 *List all the activities or methods that address a specific goal.* E.g.: Apply a computational linguistic method in the dataset ('Goal1').

```
Activity(?a)^hasObjective(?a, so:Goal1)^
Method(?m)^addresses(?m, so:Goal1) o
sqwrl:makeSet(?s1,?a)^sqwrl:makeSet(?s2,?m)^
sqwrl:union(?s3, ?s1, ?s2)->
sqwrl:select(?s3)
```

Query 5 *List all the activities that employ a specific method and consist of more than one sub-activities:*

```
Activity(?a)^employs(?a, m1)^hasPart(?a, ?s)^
swrlb:graterThan(?s,1)->sqwrl:select(?a)
```

Furthermore, the use of built-in operators such as *sqwrl: groupBy* in conjunction with the above can support some degree of closure without violating OWL's open world assumption, by partitioning OWL entities into sets under a group of arguments:

Query 6 *List the tools used in more than one activity employing methods which regard a particular research topic.* E.g. Computational Stylistic Analysis ('TK1'), and come from either Computer Science or Linguistics ('CS' or 'L'):

```
Tool(?t)^isUsedIn(?t, ?a)^employs(?a, ?m1)
regards(?m1, so:TK1)^comesFrom(?m2, so:CS)^
comesFrom(?m3,so:L) o sqwrl:makeSet(?s1, ?a)^
sqwrl:groupBy(?s1, ?t, ?m1)^sqwrl:size(?n, ?s1)^
sqwrl:graterThan(?n, 1)^sqwrl:makeSet(?s2, ?m1)^
sqwrl:makeSet(?s3, ?m2)^sqwrl:makeSet(?s4, ?m3)^
sqwrl:union(?s5, ?s3, ?s4)^
sqwrl:intersection(?s6, ?s2, ?s5)->
sqwrl:select(?t, ?s1, ?s6)
```

This case illustrates the use of counting, aggregation and disjunction. Sets of activities that employ methods regarding the specified research topic ('TK1') are created. Through the *sqwrl:groupBy* operator, these are grouped according to the tools used and the methods they employ. In a further step, the *sqwrl:size* and *sqwrl:greaterThan* operators contribute by filtering only the activities corresponding to tools that have been used in more than one activity. Finally, sets comprising all the methods that come from disciplines of Computer Science ('CS') or Linguistics ('L') are created and their union is used in further filtering the final result using the *sqwrl:intersection* operator. The output consists of the sets of requested tools together with their corresponding activities and methods.

5 Discussion

We have seen that the classes *Activity* and *Method* capture the distinction between describing how a deliberate act was actually carried out and describing a preconceived way for carrying out this type of activity. More generally, the 'how' and 'why' aspects of the scholarly domain, as captured by the *Method*, *ResearchQuestion* and *Goal* concepts, can be considered to represent a 'methodological level' concerning non-factual entities that prescribe how or explain why things are done. Conversely, the 'what' and 'who' aspects, as captured by concepts such as *Activity*, *InformationResource*, *Tool* and *Actor* represent factual entities of the scholarly domain and thus arguably belong to a 'factual level'.

Type, on the other hand, comprising the various kinds of controlled vocabularies employed for classification purposes, such as *ActivityType, InformationResourceType, MediaType, TopicKeyword*, can be regarded as a semantic bridge between the methodological and factual levels. Figure 8 illustrates

the above by displaying the majority of these interconnected concepts in a non-hierarchical manner. This indirect linking of concepts through the various types generates patterns that can be exploited in designing reusable access structures and conformance rules based on the interplay of intentionality and functionality properties. In the following two subsections we present a formalization of those patterns, which can be incorporated into the RDFS serialization of the model using SWARL or in the case of a knowledge base implementation, using a programming language such as Java.

5.1 Modelling intentionality aspects of scholarly work

The *Goal* class supports the explicit representation of intentions in the form of autonomous goals. This is not always necessary or even relevant. Alternative representations of actors' intentions are supported by the *hasIntention* property taking values in the *ActivityType* class and the *hasInterest* property with values in *TopicKeyword*. Aspects of intentionality are likewise represented by the *hasScope* property of *Activity* and the *isUsedFor* property of *Tool*, both with values in *ActivityType*. These **intentionality properties** combined with **functionality properties**, such as the rest of the properties of *Activity* and *Actor* and the properties of *Tool*, *InformationResource* and *Method*, generate semantic paths that capture aspects of the intentional context of scholarly work. Specifically:

(I1) For each actor participating in an activity there must be at least one activity type in the scope of the activity, which is also within the intention of the actor (see Fig. 8):

$$\forall a : Activity, x : Actor | participatesIn(x, a)$$
$$\rightarrow (\exists t : ActivityType(hasScope(a, t) \land hasIntention(x, t))).$$

The particular activity type thus becomes the **pivotal** element in modelling the intentional context of the actor's participation in the activity. Similarly, activity types become pivotal in representing the intentional context of the use of tools and the employment of methods in an activity. Specifically:

(I2) Whenever a method is employed in an activity there must be at least one activity type in the scope of that activity, which the method is appropriate for:

$$\forall m : Method, a : Activity | employs(a, m)$$
$$\rightarrow (\exists t : ActivityType(hasScope(a, t) \land isEmployedFor(m, t))).$$

(I3) For each tool used in an activity, there must be at least one activity type in the scope of that activity, which the tool is appropriate for:

$$\forall l : Tool, a : Activity | usesTool(a, l)$$
$$\rightarrow (\exists t : ActivityType(hasScope(a, t) \land isUsedFor(l, t))).$$
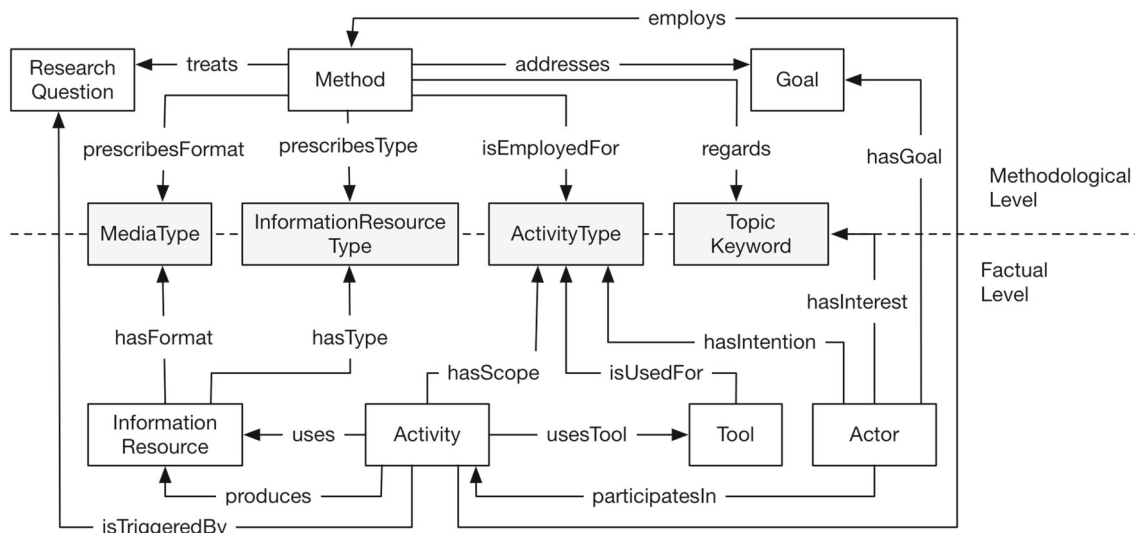
Fig. 8 Types as the semantic link between 'methodological' and 'factual' levels

5.2 Conformance rules

The functional choices made in carrying out scholarly work need to match the options relevant to the intended activity types or topics of interest. Such matching conditions are proposed here in the form of rules concerning resources, tools and actors.

Resource conformance (**RC1, RC2**): The information resources used in or produced by an activity must conform to the information resource types and formats prescribed by the method employed in the activity:

RC1:

$\forall m : Method, a : Activity, r : InformationResource,$
$rt : InformationResourceType, mt : MediaType$
$(employs(a, m) \wedge uses(a, r) \wedge hasType(r, rt) \wedge hasFormat(r, mt))$
$\rightarrow (prescribesType(m, rt) \wedge prescribesFormat(m, mt))$

RC2:

$\forall m : Method, a : Activity, r : InformationResource,$
$rt : InformationResourceType, mt : MediaType$
$(employs(a, m) \wedge produces(a, r) \wedge hasType(r, rt) \wedge$
$hasFormat(r, mt)) \rightarrow$
$\quad (prescribesType(m, rt) \wedge prescribesFormat(m, mt)).$

Tool conformance (**TC1, TC2**): When an activity is bound to employ a method, then the tools it uses must be among those prescribed by the method (for readability reasons the *prescribesTool* property is not depicted in Fig. 8). Also, when an activity is bound to use a tool, then it must employ a method that prescribes that tool:

TC1:

$\forall a : Activity, l : Tool, m : Method$
$(employs(a, m) \wedge usesTool(a, l)) \rightarrow prescribesTool(m, l)$

TC2:

$\forall a : Activity, l : Tool, m : Method$
$(usesTool(a, l) \wedge prescribesTool(m, l)) \rightarrow employs(a, m)$

Actor conformance (**AC1–AC6**): When an actor is involved in a certain type of activity (hasIntention ActivityType), then that actor must be using the tools and methods appropriate for the activity type or topic. Conversely, the tools and methods used by the actor should be enlisted as relevant to that activity type:

AC1:

$\forall c : Actor, t : ActivityType, a : Activity, l : Tool$
$(hasIntention(c, t) \wedge participatesIn(c, a) \wedge$
$\quad hasScope(a, t) \wedge usesTool(a, l)) \rightarrow isUsedFor(l, t)$

AC2:

$\forall c : Actor, t : ActivityType, a : Activity, l : Tool$
$(hasIntention(c, t) \wedge participatesIn(c, a) \wedge$
$hasScope(a, t) \wedge isUsedFor(l, t)) \rightarrow usesTool(a, l)$

AC3:

$\forall c : Actor, t : ActivityType, a : Activity, m : Method$
$(hasIntention(c, t) \wedge participatesIn(c, a) \wedge$
$hasScope(a, t) \wedge employs(a, m)) \rightarrow isEmployedFor(m, t)$

AC4:

$$\forall c : Actor, t : ActivityType, a : Activity, m : Method$$
$$(hasIntention(c,t) \wedge participatesIn(c,a) \wedge$$
$$hasScope(a,t) \wedge isEmployedFor(m,t)) \rightarrow employs(a,m).$$

Note that an actor using a tool is represented indirectly through an activity. Equivalently, when an actor is interested in a certain topic (hasInterest TopicKeyword), then that actor must be using the methods appropriate for that topic. Conversely, the methods used by the actor should be enlisted as relevant to that topic:

AC5:

$$\forall c : Actor, t : TopicKeyword, a : Activity, m : Method$$
$$(hasInterest(c,t) \wedge participatesIn(c,a) \wedge$$
$$employs(a,m)) \rightarrow regards(m,t)$$

AC6:

$$\forall c : Actor, t : ActivityType, a : Activity, m : Method$$
$$(hasInterest(c,t) \wedge participatesIn(c,a) \wedge$$
$$regards(m,t)) \rightarrow employs(a,m).$$

Note that an actor using a method is represented indirectly through an activity.

5.3 Related work and ontology reuse

Various conceptual models have been developed to describe scholarship and research. These can be distinguished into two broad categories:

First, models based exclusively on taxonomic categorizations of research activities, methods or tools such as the Taxonomy of Digital Research Activities in the Humanities[10] (TaDiRAH), the AHDS Taxonomy of Computational Methods[11] and the Oxford ICT Methods Taxonomy[12] from the field of Digital Humanities. The entity classes of SO constitute a superset of those offered by these models. With regard to relations, the SO not only provides subsumption, like the above models, but also a wide variety of semantic relations between the classes. Therefore, it is a richer representation of the domain of scholarly work, in which the above taxonomic models can be incorporated through broader term/narrower term mappings to the concept types of SO, specifically the activity types defined in NeMO [27]. Likewise, various existing SKOS vocabularies can be used in connection with relevant SO entity classes.

Second, ontological models capturing arbitrary semantic relations. These aim to represent different aspects of the scholarly domain: rhetorical and structural components of scientific discourse; bibliography and citations in the scholarly domain; scientific experiments and research activities; and social aspects in communities of practice.

More specifically, concerning rhetorical and structural aspects, the Argument Model Ontology[13] encodes arguments into a web of interrelated entities based on Tulmin's model of argumentation [37]. The SWAN[14] Ontology is a W3C recommendation for modelling scientific discourse, developed in the context of building a series of applications for biomedical researchers. In a similar way, the DoCO[15] Ontology provides a classification of document components. The above ontologies use a modular architecture while themselves foster reuse of other models (such as FOAF, DublinCore, SKOS etc.) and can provide specializations to SO's *ContentItem* as well as *Assertion* and *Proposition* classes.

Concerning the bibliography and citation aspects of the scholarly environment, the SPAR[16] collection of ontologies codifies publishing aspects in compliance with upper ontologies, with modules such as the FaBIO and CiTO [38] offering good examples of potential specializations to SO's *InformationResource* class.

Concerning scientific experiments and research activities, the EXPO [39] Ontology formalizes the generic concepts of experimental design, methodology and results representation, while providing a controlled vocabulary for annotating domain specific activities from the disciplines of Physics or Biology. myExperiment [40] provides an ontological framework behind a workflow management and exchange system, thus offering the ability to share research objects (ROs) over a social research infrastructure. CRM-Sci [41], offers a CIDOC CRM-compatible formalization for integrating and exchanging metadata about scientific observation, measurements and processed data in descriptive and empirical sciences such as geology, geography, archaeology, biology, etc. Parts of the above ontologies can be reused as specializations to SO's *Activity*, *Method* and *InformationResource* classes. Especially in the case of CRM-Sci, the mutual compatibility with CIDOC CRM stands as a common backbone which reduces the labor of a potential alignment.

Concerning research communities SWRC [42] models entities typical of research communities of practice (such as proceedings, topics, etc.) as well as relations among them, while O'CoP [43] focuses on the members of a community and their roles. Parts of both ontologies could be

---

[10] http://tadirah.dariah.eu/vocab/index.php.

[11] http://www.ahds.ac.uk/about/projects/pmdb-extension/index.htm.

[12] http://digital.humanities.ox.ac.uk/methods/ict-methodology.aspx.

[13] http://www.essepuntato.it/2011/02/argumentmodel.

[14] http://www.w3.org/TR/hcls-swan/.

[15] http://www.essepuntato.it/lode/http://purl.org/spar/doco.

[16] http://sempublishing.sourceforge.net/.

reused/aligned with appropriate SO's classes (such as *Actor-Roles, Topic, Activity, Group*, etc.)

Clearly, the existing ontological models we have reviewed address specific aspects of scholarly practice while also attaining various degrees of general coverage. However, they do not offer the integrated perspective on scholarly practice which SO does. The SO offers the necessary semantic glue that will enable the reuse of those more specialized ontologies in a unified framework, while it also supports interoperability in wider contexts by virtue of its compatibility with foundational ontologies (as explained in Sect. 3.3). Influenced by CHAT and BPM, the SO offers a more refined treatment of intentionality and its interplay with functionality than the models reviewed above. Besides, it does not carry over the full complexity of modelling business processes as this would be beyond the scope of modelling scholarly practice.

5.4 Validation and user feedback

As mentioned in Sect. 2, SO is the core part of NeMO, an ontology specifically designed for the field of Digital Humanities. NeMO has been presented in various workshops where invited experts discussed its potential use, contributed case studies followed by questions that these can answer and provided feedback on the validity and functionality of the model. The outcome of these contributions is compiled in [27]. The design decisions underlying SO and NeMO, especially the fundamental distinction between activity, activity type and method, the treatment of intentionality and the use of the *Type* classes, were ascertained in those workshops. More specifically:

– The ontological distinction between activities representing acts that have actually been performed and methods representing prescriptions of "how to do things" and can be reused independently as needed, has contributed to the disambiguation of those concepts, often confused in modelling humanities working practices. This clarification was essential for establishing an ontological framework for modelling scholarly work.
– Modelling intentionality from various aspects and at various levels of strictness caters for the widely varying documentation needs depending on the discipline or the research subject, as observed in the workshops. This is addressed in SO through the ability to model intentions directly by explicitly instantiating the *Goal* class, or indirectly through the combination of appropriate properties (such as *hasIntention, isUsedFor*, etc.) and *Type* terms.
– The *Type* class allows using arbitrary vocabularies and taxonomic schemes (their terms interrelated through

SKOS[17] properties) for flexible characterization of items in parallel with classification to ontological classes. For example, an item can be declared as instance of a class (e.g. *Method*) by ontological criteria, but it can also be "tagged"—due to other features—with more than one terms from native or imported vocabularies (e.g. the *ActivityType* taxonomy), thus enhancing the expressivity of the model. In fact, this characterization mechanism is harmonized with the corresponding scheme used by the CIDOC CRM reference ontology [21].

The continuous refinement and elaboration on case studies led to the current version of SO. Its layered architecture supports disciplinary specializations in two ways: (a) by introducing domain-specific concepts at the third layer and (b) by introducing domain-specific classification schemes through the type classes. A sample of about 100 queries posed by humanities researchers in the course of their work was collected during the aforementioned workshops related to the case studies presented there [27]. Although this was not a specifically designed, exhaustive study of user queries, the sample is quite indicative of the kinds of enquiries made by scholars. In Sect. 4.1 we showed by way of example how these can be abstracted and encoded as SPARQL or SQWRL queries employing SO. This evidence suggests that SO (and NeMO, for that matter) can adequately address the types of questions related to scholarly practice and research/scientific workflow. However, counter-evidence may appear as further cases of scholarly practice are studied, which will trigger a step of evolution of SO (or NeMO). In an evolutionary perspective the proper criterion for judging the adequacy of SO (or NeMO, or any prospective descendant) is then the degree of stability: very stable upper layer, substantially stable middle layer and dynamically evolving lower layer.

6 Conclusion and future work

In this paper we presented SO, an ontology for modelling scholarly practices, based on notions from business process modelling and Cultural-Historical Activity Theory. SO constitutes the domain-independent core of the NeDiMAH Methods Ontology (NeMO), further elaborated to be capable of supporting the modelling and documentation of scholarly/scientific work in general. NeMO then fits as an extension of SO in the area of the humanities, and similar extensions could be generated for other areas inasmuch as SO captures the basic concepts of the scholarly ecosystem. We explained the rationale of the model, the core concepts and their semantic relationships through four complementary perspectives that characterize the research context. We demonstrated its representational capabilities through an example and, using an RDFS

---

[17] http://www.w3.org/2009/08/skos-reference/skos.html.

formalization of SO, we presented a set of queries that highlight potential uses. We also discussed certain aspects of intentionality in scholarly work, related model patterns and conformance rules. These patterns and rules can be useful in developing access structures and populating knowledge bases concerning scholarly work. Finally, we explored related work and its possible reuse in conjunction with SO.

Immediate plans of further work include the development of a collaborative environment for supporting the evolution of SO and the development of coordinated disciplinary knowledge bases documenting scholarly work. Exploration of and retrieval from those knowledge bases will be supported by appropriate queries derived from the case studies contributed in the aforementioned workshops and similar forthcoming activities for collecting user input. The outcomes will be contributed to the digital research infrastructures for the arts and humanities currently under development at the European (DARIAH-EU) and national (DARIAH-GR) levels. Moreover, further elaboration on reasoning mechanisms over the conceptual constructs and rules discussed in Sects. 5.1 and 5.2, as well as the creation of complete mappings between SO concepts and properties and corresponding concepts from upper ontologies such as [21,22], is currently under progress.

Appendix: Overview of SO classes and properties

Namespaces other than SO are declared as prefixes

ClassName	SubClassOf	Properties	Range
Activity	Event	employs	Method
		isTreggeredBy	ResearchQuestion
		produces	InformationResource
		uses	InformationResource
		isDocumentedIn	InformationResource
		hasParticipant	Actor
		usesTool	Tool
		where	Place
		when	Time
		hasScope	ActivityType
		partOf	Activity
		follows	Activity
		resultsIn	Assertion
		hasObjective	Goal
ActivityType	Type	isScopeOf	Activity
		isPurposeOfEmploymentOf	Method
		isIntentionOf	Actor
		isPurposeOfUseOf	Tool
Actor	UFO:Agent	participatesIn	Activity
	CRM:Actor	hasGoal	Goal
		hasIntention	ActivityType
		hasRole	ActorRole
ActorRole	Type	isRoleOf	ActorRole
Aggregation	InformationResource	hasMember	InformationResource
Annotation	Assertion		
Assertion	ConceptualObject	isResultOf	Activity
Collection	PhysicalObject		
ConceptualObject	CRM:ConceptualObject		
ContentItem	InformationResource	providesDescriptionFor	Mehtod
		providesReferenceFor	Mehtod
Course	Activity	teaches	Method
Dataset	InformationResource		
Discipline	Type	isOriginOf	Method
Event	UFO:Event		
	CRM:Event		
Goal	Assertion	isGoalOf	Actor
		dependsOn	Goal
		comprises	Goal
		isAddressedBy	Method
		isObjectiveOf	Activity
Group	Actor		
InformationCarrier	PhysicalObject		
InformationResource	ConceptualObject	isCarriedBy	InformationCarrier
		hasType	InformationResourceType
		hasFormat	MediaType
		isRepresentationOf	ConceptualObject
		isUsedIn	Activity
		isProductOf	Activity
		ProvidesDocumentationFor	Activity
		isMemberOf	Aggregation
		hasTopic	Topic
		isDescribedBy	Metadata
InformationResourceType	Type	isTypeOf	InformationResource
		isPrescribedAsTypeBy	Method
MediaType	Type	isPrescribedAsFormatBy	yMethod
		isFormatOf	InformationResource

ClassName	SubClassOf	Properties	Range
Metadata	InformationResource	describes	InformationResource
Method	ConceptualObject	prescribesTool	Tool
		isEmployedIn	Activity
		isTaughtIn	Course
		isUsedIn	Project
		isEmployedFor	ActivityType
		prescribesFormat	MediaType
		treats	ResearchQuestion
		hasDescription	ContentItem
		isReferencedIn	ContentItem
		prescibesType	InformationResourceType
		hasPart	Step
		comesFrom	Discipline
		isInfluencedBy	SchoolOfThought
		addresses	Goal
Model	Tool ConceptualObject		
Object	CRM:Thing UFO:Endurant	isInvolvedIn	Activity
Person	Actor		
PhysicalObject	CRM:PhysicalObject		
PhysicalTool	PhysicalObject Tool		
Project	Activity	usesMethod	Method
Proposition	Assertion		
ResearchQuestion	Assertion	Triggers	Activity
		istreatedBy	Method
SchoolOfThought	Type	influences	Method
Software	InformationResource		
Tool	Object	isUsedAsToolIn	Activity
		isPrescribedAsToolBy	Method
		isUsedAsToolFor	ActivityType
Topic	ConceptualObject	isTopicOf	InformationResource
TopicExpression	Topic		
TopicKeyword	Topic Type		
Type	ConceptualObject		

References

1. Berman, F., Fox, G., Hey, A.J.G.: Grid Computing. Wiley, New York (2003)
2. Research Infrastructures in the Digital Humanities. Science Policy Briefing, vol. 42. European Science Foundation, ISBN: 978-2-918428-50-3 (2011). http://www.esf.org/fileadmin/Public_documents/Publications/spb42_RI_DigitalHumanities.pdf
3. Benardou, A., Constantopoulos, P., Dallas, C.: An approach to analyzing working practices of research communities in the humanities. Int. J. Humanit. Arts Comput. 7, 105–127 (2013)
4. Case, D.O.: Looking for information: a survey of research on information seeking, needs, and behavior. Academic Press, San Diego, CA (2002)
5. Bearman, D.: Overview and discussion points. In: Research agenda for networked cultural heritage. Getty AHIP, pp. 7–22 . Santa Monica, CA (1996)
6. Brodaric, B., Gahegan, M.: Ontology use for semantic e-science. Semant. Web (2010). doi:10.3233/SW-2010-0021
7. Thanos, C.: The future of digital scholarship. Procedia Comput. Sci. 38, 22–27 (2014)
8. Thanos, C.: Mediation: the technological foundation of modern science. Data Sci. J. 13, 88–105 (2014)
9. Meho, L.I., Tibbo, H.R.: Modeling the information-seeking behaviour of social scientists: Ellis's study revisited. J. Am. Soc. Inf. Sci. Tech. 54, 570–587 (2003)
10. Palmer, C.L., Cragin, M.H.: Scholarship and disciplinary practices. Annu. Rev. Inf. Sci. (2008)
11. Unsworth, J.: Scholarly primitives: what methods do humanities researchers have in common, and how might our tools reflect this? King's College, London (2000). http://www3.isrl.illinois.edu/~unsworth/Kings.5-00/primitives.html. Accessed 4 Aug 2009
12. Victor Kaptelinin, B.A.N.: Acting with Technology: Activity Theory and Interaction Design, p. 1347 (2006)
13. Yu, E., Giorgini, P., Maiden, N., Mylopoulos, J., Fickas, S.: Modelling strategic relationships for process reengineering. In: Social Modelling for Requirements Engineering, p. 11152
14. Sun, J., Loucopoulos, P., Zhao, L.: Representing and elaborating quality requirements: the QRA approach. Lecture Notes in Computer Science, vol. 8217, pp. 446–453. Springer, Heidelberg (2013). doi:10.1007/978-3-642-41924-9_37
15. Tsakonas, G., Papatheodorou, C.: An ontological representation of the digital library evaluation domain. J. Am. Soc. Inf. Sci. Technol. 62(8), 1577–1593 (2011)

16. Dietz, J.L.G.: Enterprise Ontology: Theory and Methodology, Enterprise Ontology (2006)

17. Weske, M.: Business Process Management. Springer, Berlin (2012). doi:10.1007/978-3-642-28616-2

18. Fox, M.S.: The TOVE project towards a common-sense model of the enterprise. In: Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, vol. 604, p. 2534. Springer, Berlin (1992). doi:10.1007/BFb0024952

19. Uschold, M., King, M., Moralee, S.: The enterprise ontology. Knowl. Eng. Rev. **13**(1), 31–89 (1998)

20. Malone, T.W., Crowston, K., Herman, G.A.G.A1.: Organizing Business Knowledge: The MIT Process Handbook. MIT Press, Cambridge (2003)

21. Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M. (eds.): Definition of the CIDOC conceptual reference model (version 5.0.1). ICOM/CIDOC CRM Special Interest Group (2009). http://cidoc.ics.forth.gr/official_release_cidoc.html

22. Guizzardi, G., Wagner, G.: A unified foundational ontology and some applications of it in business modelling. In: CAiSE Workshops, pp. 129–143 (2004)

23. Workflows for e-Science: Scientific workflows for grids (2006)

24. Deelman, E., Gannon, D., Shields, M., Taylor, I.: Workflows and e-Science: an overview of workflow system features and capabilities. Futur. Gener. Comput. Syst. **25**, 528–540 (2009)

25. Benardou, A., Constantopoulos, P., Dallas, C., Gavrilis, D.: Understanding the information requirements of arts and humanities scholarship. IJDC **5**, 18–33 (2010)

26. Constantopoulos, P., Dallas, C., Doorn, P., Gavrilis, D., Gros, A., Stylianou, G.: Preparing DARIAH. In: Proceedings of the International Conference on Virtual Systems and MultiMedia (VSMM08). Nicosia, Cyprus (2008). http://www.dcu.gr/dcu/Documents/documents/preparing-dariah/en/attachment

27. NeDiMAH Methods Ontology (NeMO). http://nemo.dcu.gr

28. Benardou, A., et al.: A conceptual model for scholarly research activity. In: Unsworth, J., Rosenbaum, H., Fisher, K.E. (eds.) iConference 2010 Proceedings (Urbana–Champaign, Ill., 2010), p. 2632 (2010)

29. Doerr, M., Kritsotaki, A., Christophides, V., Kotzinos, D.: Reference Ontology for Knowledge Creation Processes. Collaborative Knowledge Creation, pp. 31–52. Sense Publishers, Rotterdam (2012). doi:10.1007/978-94-6209-004-0_3

30. Francesconi, F., Dalpiaz, F., Mylopoulos, J.: TBIM: A Language for Modelling and Reasoning about Business Plans. Conceptual Modelling. Springer, Berlin (2013)

31. Doerr, M., Tzobanakis, M.: On information organization in annotation systems. In: Lecture Notes in Computer Science, p. 189200. Springer, Berlin (2005)

32. Suzuki, T., Hosoya, M.: Computational stylistic analysis of popular songs of japanese female singer-songwriters. Digit. Humanit. Q. **8** (2014). http://www.digitalhumanities.org/dhq/vol/8/1/000170/000170.html

33. Stocker, M., Smith, M.: Owlgres: a scalable OWL reasoner. Owled (2008)

34. Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., Katz, Y.: Pellet: A practical OWL-DL reasoner. **5**(2), 5153 (2007). doi:10.1016/j.websem.2007.03.004

35. O'Connor, M.J., Das, A.K.: SQWRL: a query language for OWL. OWLED (2009)

36. SWRL: a semantic web rule language combining OWL and RuleML. http://www.w3.org/Submission/SWRL/

37. TOULMIN, S.E.: The Uses of Argument, 2nd edn. Cambridge University Press, Cambridge (2003)

38. FaBiO and CiTO: ontologies for describing bibliographic resources and citations. **17**, 3343 (2012). doi:10.1016/j.websem.2012.08.001

39. Soldatova, L.N., King, R.D.: An ontology of scientific experiments. J. R. Soc. Interface **3**(11), 795–803 (2006)

40. Newman, D., Bechhofer, S., De Roure, D.: myExperiment: an ontology for e-research (2009)

41. Doerr, M., Rousakis, Y., Hiebel, G.: CRMsci: the scientific observation model. 135 (2014)

42. Sure, Y., Bloehdorn, S., Haase, P., Hartmann, J., Oberle, D.: The SWRC ontology semantic web for research communities. In: Presented at the EPIA'05: Proceedings of the 12th Portuguese Conference on Progress in Artificial Intelligence, Berlin (2005)

43. Tifous, A., Ghali, El, A., Dieng-Kuntz, R., Giboin, A., Christina, C., Vidou, G., : An Ontology for Supporting Communities of Practice. ACM, New York (2007)