

Annotating Evidence Based Clinical Guidelines

A Lightweight Ontology

Rinke Hoekstra^{1,2}, Anita de Waard³, and Richard Vdovjak⁴

¹ Dept. of Computer Science, VU University Amsterdam

² Leibniz Center for Law, University of Amsterdam

³ Elsevier Publishing

⁴ Philips Research

rinke.hoekstra@vu.nl, a.dewaard@elsevier.com, richard.vdovjak@philips.com

Abstract. This paper describes a lightweight ontology for representing annotations of declarative evidence based clinical guidelines. We present the motivation and requirements for this representation, based on an analysis of several guidelines. The ontology provides the means to connect clinical questions and associated recommendations to underlying evidence, and can capture strength and quality of recommendations and evidence, respectively. The ontology was applied in the conversion of manual annotations to RDF and used as part of a prototype clinical decision support system.

Keywords: clinical decision support, linked data, annotation, clinical guideline, evidence based, RDF

1 Introduction

Evidence based clinical guidelines follow the principle that every recommendation of the guideline should be supported by identifiable evidence in the form of published medical research. This is in contrast to the older variant, where guidelines are based on *consensus* within the scientific community. The strength of the evidence based approach is that such guidelines are more *adaptive* to new findings in medical research (most importantly *clinical trials*), and that they preserve *provenance* of recommendations in the form of citations to scientific literature.

Guidelines are part of a larger network of hypotheses, claims and pieces of evidence that span across multiple publications [1]. Presenting a clinician with the full text of a single guideline is therefore both *overwhelming* to the clinician, *incomplete* as it ignores the context in which the guideline was written, and *static*: it is frozen in time, a snapshot of the state of the art at time of publication. Furthermore, the form in which guidelines are currently published does not lend itself to patient centric presentation: guidelines are lengthy (digital) documents, with references (not hyperlinks) at the end, of which only a part may be relevant to a patient case.

To unlock the clinical knowledge contained in a guideline, we can choose from several languages to formalize the guideline and make it a “computer interpretable” guideline [2] (CIG). Specifying such executable models is a very knowledge intensive and laborious process [3]. Furthermore, only some of these languages offer means to link to evidence [4], and they generally are targeted to very concrete and *procedural* guidelines, akin to medical protocols. However, many evidence based guidelines exist that are much more *declarative* and are not readily implementable as a CIG. These declarative evidence based guidelines (DEG) nevertheless form a highly relevant information source for clinicians.

The main question underlying this work is: does access to declarative guidelines benefit from a lightweight modeling approach? In the context of clinical decision support systems: are lightweight models adaptive enough, and do they have sufficient power to link a patient to relevant guidelines and underlying evidence? This paper is the first step in answering our question: a lightweight ontology for annotating DEGs. Section 2 gives an analysis of the structure of DEGs. This provides us with the requirements for the ontology, which is described in section 3. The ontology describes DEGs at a meta-level; it builds on existing work on *annotation* languages. We ran a small pilot on a guideline on Febrile Neutropenia [5] and integrated the results in Hubble, a prototype CDS system [6]. Section 4 discusses the results and presents our ideas for future work.

2 Declarative Evidence Based Guidelines: Requirements

Evidence based guidelines⁵ are well structured documents that follow a clear recipe: introduction of *methodology* and definitions of recommendation *strength* and evidence *quality* and *level*, followed by a list of *clinical questions* and their *recommendations*, and finally a discussion of the *recommendation*, a *summary* of the underlying evidence and citations of medical publications. We briefly discuss each of these. Figure 1 shows an excerpt of a clinical guideline on Febrile Neutropenia containing a number of recommendations and a part of the associated evidence summary. For instance, the recommendation:

“At least 2 sets of blood cultures are recommended”

has an evidence summary containing:

“Recently, 2 retrospective studies found that 2 blood culture sets detect 80%–90% of bloodstream pathogens in critically ill patients, whereas ≥ 3 sets are required to achieve $>96\%$ detection”

that cites papers “49” and “50” as underlying evidence. It is this link between a recommendation, its summary and underlying evidence that we aim to capture. We use this excerpt as running example throughout the paper.

⁵ We studied a wide range of guidelines from the Department of Health (UK), AIMS (US), NABON (NL), NHRMC (AUS) and others.

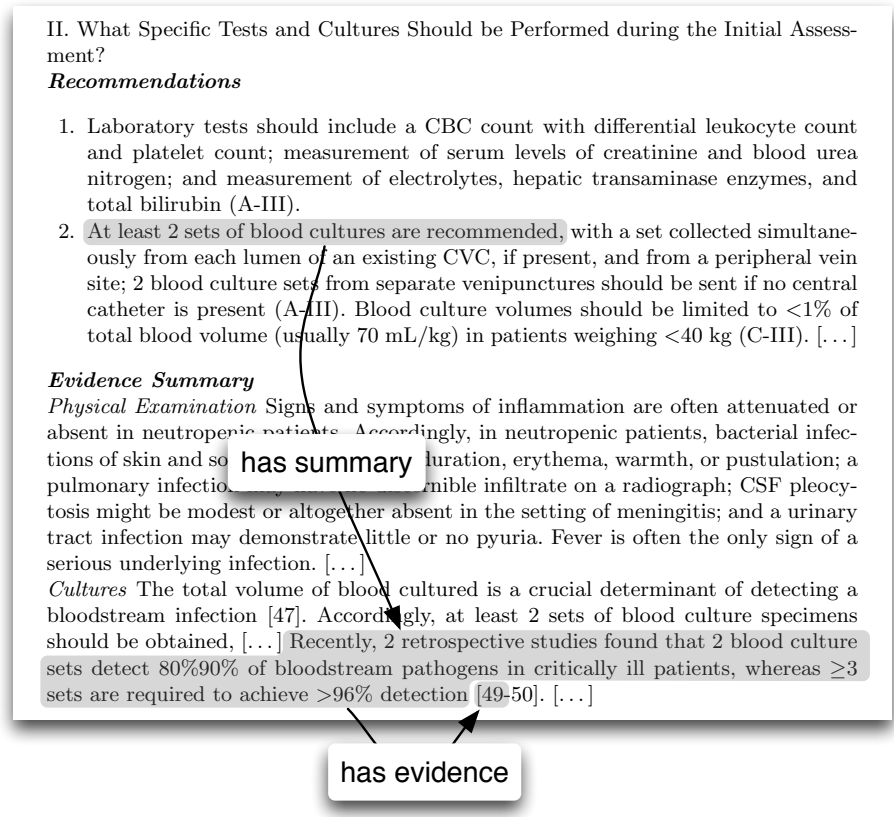


Fig. 1. Excerpt of [5], illustrating the structure of EBCGs.

Methodology and Definitions The **methodology** section (or introduction) motivates the existence of the guideline and details the procedure followed to identify clinical questions, and assess the evidence in clinical publications for determining requirements. The questions must be answerable by a systematic review; criteria must be formulated for including or excluding literature (to avoid bias); eligible studies must be categorized according to *quality* and *strength* of evidence; and this data must be synthesized and combined to a clear recommendation [7]. The *categorization* of evidence, and the domains on which the individual studies are evaluated, depends on the type of study: e.g. systematic reviews, randomized controlled trials, observational studies, and diagnostic test studies [8]. The *aggregated* strength of bodies of evidence is determined by the aggregate *quality* ratings of individual studies, the *quantity* (number of studies, sample size, etc.) and *consistency* of the body of evidence [8].

Although evidence categorization schemes are based on the same principles, the resulting categories (or codes) may vary significantly across different guide-

lines. Quality and strength may be represented independently or jointly; strength may be associated with the recommendation, rather than the evidence; and the *number* of categories may differ. For instance, in [5] (Figure 1), the recommendation strength is indicated with letters A-C (good-poor) and the evidence quality with numbers I-III (“one or more properly randomized, controlled trials”-“evidence from opinions...”), whereas e.g. the Australian NHRMC uses a combined system (I-IV) and the Dutch NABON adopts quality levels (1-4) and strength (A1,A2-D).⁶

Questions, Recommendations and Evidence Summary The core of the guideline is the list of **clinical questions** that motivated the selection of evidence. The questions are also the entry points for clinicians. Every clinical question is accompanied by a numbered list of **recommendations** (1. and 2. in in Figure 1). This suggests identifiability, but an item may actually contain multiple recommendations. In principle, every recommendation is backed by an **evidence summary** that motivates the recommendation by providing a synthesis of the underlying evidence studies. However, we have found that in many cases, recommendations exist without a discussion in the evidence summary. And conversely, the evidence summary may *implicitly* introduce new recommendations that are not listed as stand-alone recommendation. An example of this is the *Physical Examination* paragraph in the example above.

Another observation is that in the guidelines we have seen, strength and quality codes are linked to a recommendation, but not to the text in the evidence summary, nor to the studies themselves. In other words, the reader is presented only the *outcome* of the meticulous weighing and categorizing of evidence, but not any of the intermediate results. This is a regrettable loss of information: insight in the assessed quality of individual studies can assist the clinician in weighing more nuanced cases. It is not always clear how or if the quality of evidence contributes to the strength of a recommendation. In our example, a strong ‘A’ recommendation may be backed by weak ‘III’ evidence.

Lifecycle Whenever a sufficient quantity, or sufficiently important new evidence arises, a guideline needs to be updated. Guidelines may thus undergo multiple *revisions* that are very similar in some respects (e.g. structure, questions) but very different in others (recommendations, evidence, authors). This is where guidelines differ from more regular publications in science, but it is similar to e.g. *working papers* in the social sciences and humanities. The relative dynamics of guidelines creates a significant maintenance burden for approaches that intend to capture the contents of a guideline in a formal model.

⁶ See <http://www.nhmrc.gov.au/guidelines/publications/cp30> and <http://www.oncoline.nl/mammacarcinoom> respectively.

3 A Lightweight Ontology

The analysis above identifies five core parts of evidence based guidelines: clinical questions, recommendations, evidence summaries, evidence studies and evidence quality scores. Our ontology should allow us to *identify* these parts, indicate their *type*, and *relate* them amongst each other. Given the variety in scoring schemes, we intend to express the strength and quality of recommendations and evidence in an explicit but lightweight model. The relations and types express an *interpretation* of a guideline: we do not purport to provide official representations, but rather use *annotations* to make explicit the structure of guidelines. This will allow for multiple, possibly competing, interpretations of the annotations. In this context, *provenance* information about the annotation process is very important.

Annotation of medical publications is certainly not new. Tools such as the BioPortal annotator [9] for automatic annotation against biomedical vocabularies and (manual) annotation environments such as Domeo [10], Utopia [11], brat [12] and Pundit [13] are very accessible and widely used. Recently, two initiatives for standardizing the representation of annotations in RDF – the Annotation Ontology⁷ and the Open Annotation Model (OA) – were merged into the latter.⁸ The lightweight guideline annotation ontology is an extension of the Open Annotation Model. In the following, we briefly introduce the core elements of the ontology, and illustrates them using recommendation “6.” in Figure 1. A more elaborate report, including extensive motivation and alternative representations is available from <http://bit.ly/AnnotatingGuidelines>. Over the past month, the Open Annotation model has seen a significant number of proposed changes. The work presented here is based on the specification of May 2012.⁹

The Ontology An annotation is a resource that relates to a *body* and a *target*, where the body is ‘somehow “about”’ (sic.) the target. The body of an annotation typically represents the *content* of the annotation, whereas the target of the annotation identifies the part of a document being annotated. Both body and target are typically represented by means of a URI, but they can also associate a *selector*, which can be used to identify a part of another resource (e.g. a particular string of characters). Although the OA defines properties for expressing provenance information such as authorship and generation time, these are quite restricted. We adopt the W3C PROV vocabulary instead.¹⁰

Unfortunately, the OA model cannot be used to *relate* (parts of) one or more documents.¹¹ In our scenario, we would like to be able to express e.g. that *this* particular bit of text in document *A* is a recommendation, that is supported

⁷ See <http://code.google.com/p/annotation-ontology/>

⁸ See <http://www.openannotation.org/>

⁹ See <http://www.openannotation.org/spec/core/20120501.html>

¹⁰ See <http://www.w3.org/TR/prov-o/>

¹¹ The vocabularies allow for multiple targets to an annotation, but this does not capture the *directedness* of the support relation.

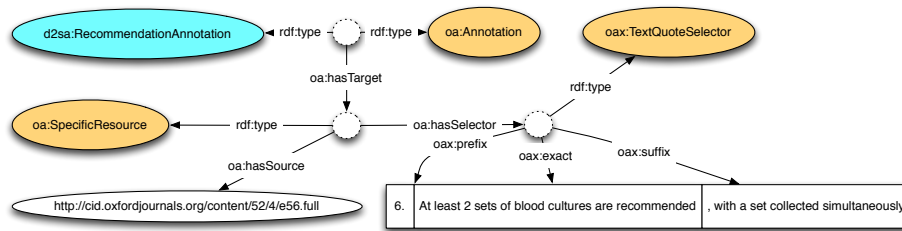


Fig. 2. A `d2sa:RecommendationAnnotation` with an `oax:TextQuoteSelector`.

by *that* particular bit of text in document *B*. The SWAN ontology¹² identifies discourse elements (research statements such as hypothesis and claim) and the relations between them, but can only be used to refer to the sources of those statements (actual publications) at the *document level*. There is no prescribed way to directly use SWAN relations together with the annotation vocabularies. SWAN relations hold between the research statements themselves, and not between annotations on the statements. Similarly, the Citation Typing Ontology (CITO)¹³ and the Bibliographic Ontology Specification (BIBO)¹⁴ are both catered more to traditional citation metadata.

3.1 Open Annotation-based Representation

The structure of OA-based annotations allows for three alternative representations for the link from a *recommendation*, its *evidence summary* to the underlying *evidence* in (external) publications (Figure 2):

1. as a *single annotation* with multiple targets,
2. as *two annotations*, distinguishing the recommendation from the evidence summary, and
3. using *three separate annotations* for recommendation, evidence summary and evidence.

Approaches 1 and 2 have some drawbacks. The first approach is concise, but makes it difficult to distinguish between recommendation, summary and evidence. The second approach confuses representation and annotation by saying that the summary *is an annotation*: who should be listed as the author of the annotation? Is that the author of the paper, or the creator of the annotation?

The third approach offers the most fine grained control over the link between the various parts of a guideline and express more detailed information about the relation between the summary and the evidence. We distinguish three types of annotations:¹⁵.

¹² See <http://code.google.com/p/swan-ontology/>

¹³ See <http://purl.org/spar/cito>

¹⁴ See <http://purl.org/ontology/bibo/>

¹⁵ The ‘d2sa’ namespace is defined as <http://aers.data2semantics.org/vocab/annotation/>.

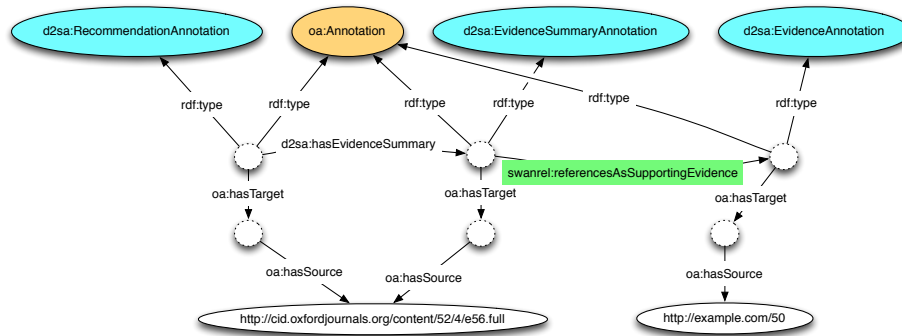


Fig. 3. Relations between a recommendation, a summary and an evidence.

- An instance of `d2sa:RecommendationAnnotation` is an annotation for the recommendation text in the guideline. Figure 2 shows an example using an OA `oax:TextSelector` to identify the recommendation in the guideline text (the target of the annotation). We will use a `d2sa:hasEvidenceSummary` property to relate it to the evidence summary.
- The `d2sa:EvidenceSummaryAnnotation` captures the relevant summaries for individual recommendations. A single recommendation may have multiple underlying evidences. The annotation uses a `swanrel:referencesAsSupportingEvidence` property to relate it to one or more evidence annotations.¹⁶
- Finally, the `d2sa:EvidenceAnnotation` is an annotation on the document that contains the evidence described in the evidence summary (not on the guideline). In its simplest form, it is an annotation (without body) on the evidence document as a whole. We can also make it more expressive by adding a selector for the actual text that provides the evidence.

The relation between a recommendation, its evidence summaries and underlying evidence is depicted in Figure 3. We believe that the separation between annotations on the guideline and annotations on the underlying evidence is a good thing. High granularity allows for expressive relations between evidence, summaries and recommendations. It has the advantage that evidence becomes a first class citizen, rather than just the target of an evidence summary annotation. For instance, we can reuse the evidence annotation as supporting some claim or hypothesis in yet another publication. More importantly, we have encountered examples where the evidence cited in an evidence summary may be a recommendation by itself, that in turn refers to some evidence.

A drawback is that increased expressiveness and verbosity go hand in hand: a single recommendation-summary-evidence chain (with evidence selector) is captured as 26 triples. In practice, every annotation will maintain a link to a timestamped, cached resource (a `oa:State`) on which the annotation was originally made. Adding an additional 4 triples per annotation.

¹⁶ For the SWAN ontology, see <http://code.google.com/p/swan-ontology/>.

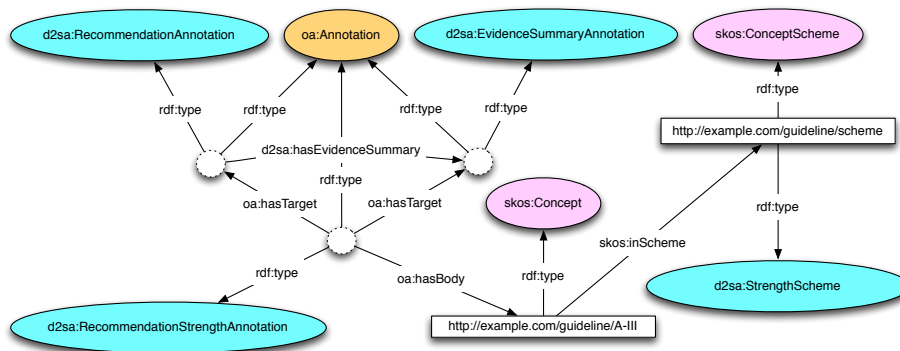


Fig. 4. The `d2sa:RecommendationStrengthAnnotation`

Evidence Quality The multitude of evidence rating schemes [7] requires a model that can accommodate different schemes, without committing to a particular standard vocabulary. We have seen that strength and quality make a statement about the *combination* of the recommendation and the evidence summary. It furthermore clear “quality of evidence” is based on a synthesis of the available evidence for a recommendation.

We represent the *strength of recommendation* as an annotation that has both the `d2sa:RecommendationAnnotation` and `d2sa:EvidenceSummaryAnnotation` as target (Figure 4). The *quality of evidence* when mentioned separately, can be an annotation to the evidence summary annotation (if it is an aggregated quality indication) or to a single `d2sa:EvidenceAnnotation`. The *body* of the annotation contains the strength or quality itself. We represent the strength and quality codes as instances of `skos:Concept`, belonging to a `skos:ConceptScheme` that represents the rating scheme being used. Using SKOS allows us to use lightweight relations between ratings without having to commit to expressive formal semantics.¹⁷ For instance, we may say that a rating A-II is `skos:broader` than A-I (for retrieving all A-II grade and stronger recommendations) without committing to a set-theoretic semantics that says that all A-I grade recommendations are necessarily also A-II grade.

4 Discussion

In the preceding sections, we analyze the structure of declarative evidence based guidelines, and introduce a lightweight ontology for making the structure of these guidelines explicit. The analysis shows that guidelines are often incomplete and even inconsistent, supporting the argument for an annotation-based approach. The ontology allows us to break down guidelines and evidence papers into their basic parts, and relating these parts to allow for a chain connecting

¹⁷ Simple Knowledge Organization System, <http://www.w3.org/2004/02/skos/>.

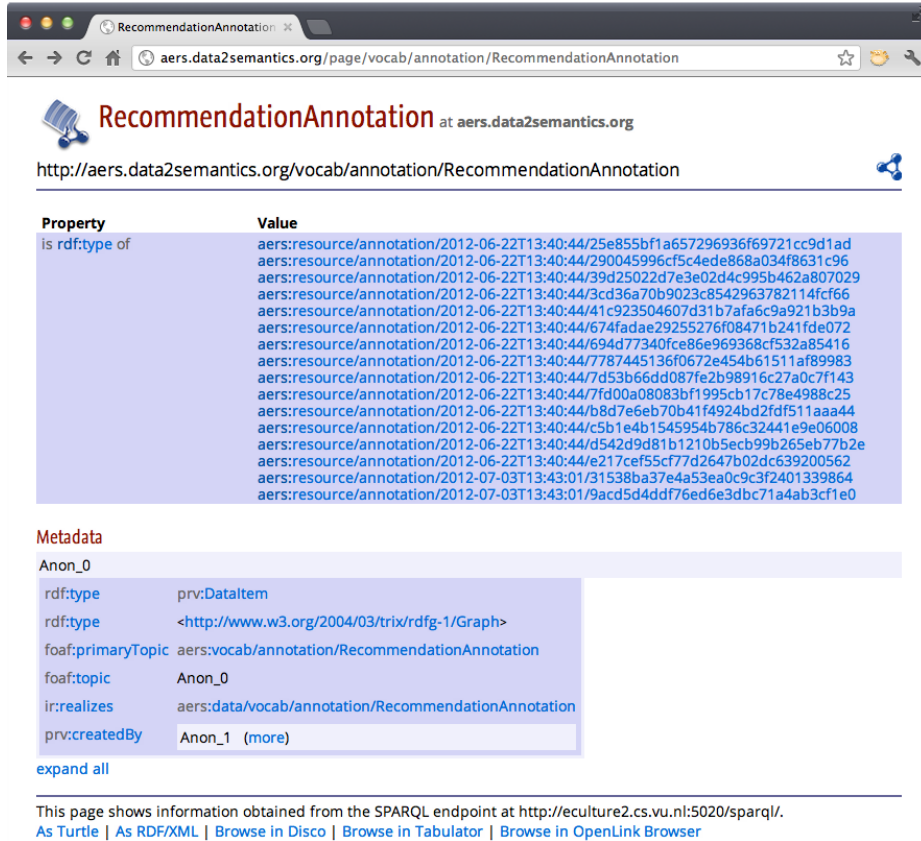


Fig. 5. Screenshot of the RecommendationAnnotation class in Pubby

every recommendation to its underlying evidence. We can provide *context* to a recommendation.

We used the ontology to add guideline and recommendation information to Hubble [6], a prototype CDS system that is hooked up to AERS-LD, a Linked Data version of AERS.¹⁸ Since the explicit connection between recommendations and evidence was lost in the writing of the guideline, we reverse engineer the relations between recommendations, underlying evidence and the patient characteristics to which these pertain. We manually analyzed two questions of a clinical guideline on neutropenic patients with cancer [5], identified recommendations and underlying evidence, using an Excel spreadsheet as intermediate format. This spreadsheet was automatically converted to RDF, and stored

¹⁸ Adverse Event Reporting System of the FDA. For AERS-LD, see <http://aers.data2semantics.org/>.

The screenshot displays the Hubble interface for a patient named John Doe. The patient's information includes a Patient ID, a comment stating 'over 60, matches all other criteria', and an age of 67. The diagnosis section lists Neutropenia, Diarrhoea, and Vomiting, each with 'Related Reports' and 'Browse' buttons. The drug administration section lists Cefprozil, Ambisome, Mercaptopurine, and Foradil, each with 'Molecule', 'Related Reports', and 'Browse' buttons.

The bottom section shows a table with three columns: Recommendations, Evidence Summary, and Supporting Evidence. The Recommendations column lists various clinical guidelines, such as 'Get Enough Blood In Two Bottles' and 'CSFs are not generally recommended for treatment of established...'. The Evidence Summary column provides a detailed summary of the evidence for one of the recommendations, discussing the use of myeloid CSFs. The Supporting Evidence column lists specific evidence sources, such as 'Filgrastim treatment used with antibiotics at the onset of febrile neutropenia...'. Arrows indicate the relationship between the recommendations and their supporting evidence.

Fig. 6. Screenshot of the recommendation browser in Hubble

in a 4Store triple store.¹⁹ An example recommendation represented using the lightweight ontology can be browsed from <http://bit.ly/RecommendationExample> (see also Figure 5).

The Hubble CDS is patient centric, that is, it presents relevant information based on a patient description (the top half of Figure 6). Since guidelines can be quite abstract, we used the BioPortal annotator to annotate all documents in our repository, and converted the annotations XML to the OA RDF format. UMLS CUI identifiers allowed us to link BioPortal output to LinkedLifeData.com and from there on to other Linked Data resources, including AERS-LD. The similarity between patients and scientific publications allows us to retrieve all recommendations grounded in a publication. These recommendations are listed in the bottom half of the Hubble UI, ordered by relevance. Following a drill-down model, selecting a recommendation shows the underlying evidence summaries, and selecting a summary shows the underlying evidences. The 'arrow' icons link to the RDF resource in the triple store, or, if applicable directly to the evidence paper.

¹⁹ See <http://4store.org>.

Future work includes the addition of annotations for clinical questions as these form the primary entry point for clinicians. Secondly, we aim to support automatic annotation of guidelines using this ontology, either through a pattern-based approach [3] or using a combination of NLP and machine learning. The core difficulties here are detecting what summary belongs to which recommendation, and generalizing the approach across different guidelines. Thirdly, given a sufficiently large body of annotations, presenting relevant results to the user (ranking) will become an issue. We are considering both expert-based ranking (e.g. using karma) and contextual ranking mechanisms. Finally, we are exploring the possibilities of merging our Open Annotation based ontology with the named-graph based approach of Nanopublications [14]. This allows us to add more content to the *body* of annotations, enabling e.g. a hybrid approach that combines our ontology with a more formal representation of the guideline.

Acknowledgements This work was funded under the Dutch COMMIT program as part of Data2Semantics. Special thanks goes to Adianto Wibisono, who worked on the converter from BioPortal to the Open Annotation model.

References

1. de Waard, A., Shum, S.B., Carusi, A., Park, J., Samwald, M., Sándor, Á.: Hypotheses, evidence and relationships: The hyper approach for representing scientific knowledge claims. In: Proceedings of the 8th ISWC, Workshop on Semantic Web Applications in Scientific Discourse, Berlin, Springer (October 2009)
2. Annette ten Teije, Silvia Miksch, P.L., ed.: Computer-based Medical Guidelines and Protocols: A Primer and Current Trends. Volume 139 of Technology and Informatics. (2008)
3. Seyfang, A., et al.: Maintaining formal models of living guidelines efficiently. In Bellazzi, R., Abu-Hanna, A., Hunter, J., eds.: Artificial Intelligence in Medicine. Volume 4594 of LNCS. Springer (2007) 441–445
4. Peleg, M., et al.: Comparing computer-interpretable guideline models: A case-study approach. *J. Am. Med Inform Assoc.* **10**(1) (2003) 52–68
5. Freifeld, A.G., et al.: Clinical practice guideline for the use of antimicrobial agents in neutropenic patients with cancer: 2010 update by the infectious diseases society of america. *Clinical Infectious Diseases* **52**(4) (2010) 56–93
6. Hoekstra, R., Magliacane, S., Rietveld, L., de Vries, G., Wibisono, A.: Hubble: Linked data hub for clinical decision support. In: Post-Conference Proceedings of the ESWC 2012. (2012)
7. Lohr, K.N.: Rating the strength of scientific evidence: relevance for quality improvement programs. *International Journal for Quality in Health Care* **16**(1) (2003) 9–18
8. West, S., King, V., Cary, T., Lohr, K., McKoy, N., Sutton, S., Lux, L.: Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment 47, Research Triangle Institute - University of North Carolina, North Carolina (April 2002)
9. Shah, N., Bhatia, N., Jonquet, C., Rubin, D., Chiang, A., Musen, M.: Comparison of concept recognizers for building the open biomedical annotator. *BMC Bioinformatics* (S14) (2009)

10. Ciccarese, P., Ocana, M., Clark, T.: DOMEQ: a web-based tool for semantic annotation of online documents. In: Proceedings of Bio-Ontologies, Vienna, Austria (2011)
11. Attwood, T.K., Kell, D.B., McDermott, P., Marsh, J., Pettifer, S.R., Thorne, D.: Calling international rescue: knowledge lost in literature and data landslide! *Biochemical Journal* **424**(3) (Dec 2009) 317–333
12. Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S., Tsujii, J.: brat: a web-based tool for nlp-assisted text annotation. In: Proceedings of the Demonstrations Session at EACL 2012. (2012)
13. Grassi, M., Morbidoni, C., Nucci, M., Fonda, S., Ledda, G.: Pundit: Semantically structured annotations for web contents and digital libraries
14. Groth, P., Gibson, A., Velterop, J.: The anatomy of a nanopublication. *Information Services and Use* **30**(1-2) (2010) 51–56