

Representing Biomedical Claims with Underspecified Statements in Nanopublications

Tobias Kuhn and Michael Krauthammer
Department of Pathology, Yale University School of Medicine
<http://www.tkuhn.ch> <http://krauthammerlab.med.yale.edu>

Nanopublications

Nanopublications (<http://nanopub.org>) have been proposed to ...

- ... make it easier to find, connect and curate core scientific statements
- ... determine the attribution, quality and provenance of scientific statements

Nanopublications use **Semantic Web** standards, concretely **RDF with named graphs**, to represent **scientific data and claims** in a machine-understandable form.

Problems:

- Existing RDF vocabularies are not sufficient for many scientific claims
- Correctly representing scientific claims in RDF is a difficult task

We propose to **extend nanopublications** by:

- Natural language representations for scientific claims
- Support for informal and underspecified claims

This is **work in progress** at an early design stage.

Scientific Claims as Sentences

We propose that authors should represent their **scientific claims as sentences**:

- Sentences in plain English (or another natural language)
- Independent sentences: no references (e.g. "this behavior") to the surrounding text

Simple example:

- Malaria is transmitted by mosquitoes.

Examples from existing abstracts:

Homelessness is a strong predictor of poor health outcomes and complicates the medical management of HIV. This housing intervention improved the health of HIV-positive homeless people. [PMID 19372524]

⇒

- Homelessness is a strong predictor of poor health outcomes.
- Homelessness complicates the medical management of HIV.
- Housing intervention can improve the health of HIV-positive homeless people.

The results of our study suggest that the MTHFR gene may influence the onset of menarche and natural menopause. This effect is probably due to the multiple SNP-SNP and SNP-environment interactions. [PMID 19593234]

⇒

- The MTHFR gene influences the onset of menarche and natural menopause.
- The effect of the MTHFR gene on the onset of menarche and natural menopause is due to multiple SNP-SNP and SNP-environment interactions.

Although we have found a slightly lower risk than other reports, the risk of developing neurodegenerative disease in idiopathic REM sleep behavior disorder is substantial, with the majority of patients developing Parkinson disease and Lewy body dementia. [PMID 19109537]

⇒

- The risk of developing neurodegenerative disease in idiopathic REM sleep behavior disorder is substantial.
- The majority of patients with idiopathic REM sleep behavior disorder who develop a neurodegenerative disease develop Parkinson disease and Lewy body dementia.

URIs that encode sentences as identifiers for scientific claims:

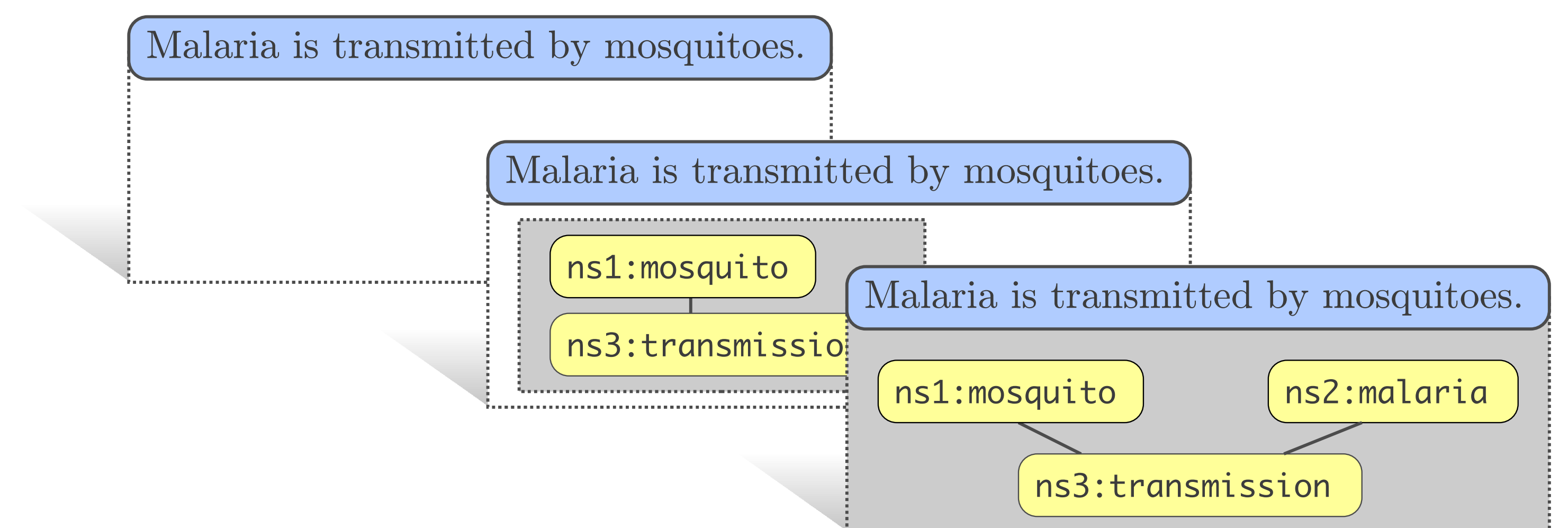
<http://statements.org/en/Malaria+is+transmitted+by+mosquitoes>

<http://statements.org/en/The+majority+of+patients+with+idiopathic+REM+sleep+behavior+disorder+who+develop+a+neurodegenerative+disease+develop+Parkinson+disease+and+Lewy+body+dementia>

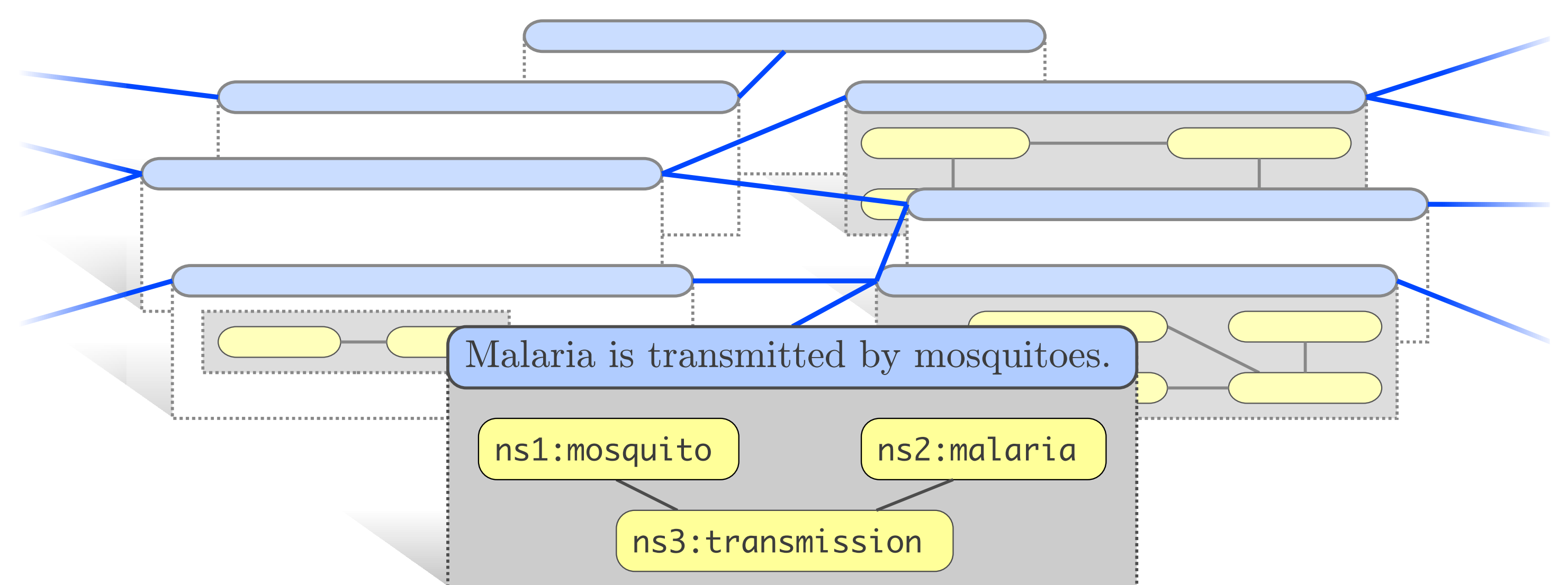
Linked Scientific Claims

Levels of Formalization:

- Informal (only a natural sentence)
- Underspecified (RDF representation for part of the sentence)
- Fully formal (RDF representation for the complete sentence)



Linking Scientific Claims:



Possible relations:

- [CLAIM] is equivalent to / contradicts / is similar to [CLAIM]
- [PERSON] agrees with / disagrees with / challenges [CLAIM]
- [STUDY] provides evidence for / provides counter-evidence against [CLAIM]
- ...

Implementation

In contrast to classical nanopublications, we divide assertions into a **head** and a **body** part (each being a named graph):

```
<>
Pub1 a Nanopublication .
Pub1 hasAssertion Pub1_Assertion .
Pub1_Assertion containsGraph Pub1_Assertion_Head .
Pub1_Assertion containsGraph Pub1_Assertion_Body .
...
```

The head refers to different (formal and natural) representations of the claim:

```
Pub1_Assertion_Head
Pub1_Assertion asSentence en/Malaria+is+transmitted+by+mosquitoes .
Pub1_Assertion asSentence de/Malaria+wird+durch+Muecken+uebertragen .
Pub1_Assertion asFormula Pub1_Assertion_Body .
Pub1_Assertion_Body containsGraph Pub1_Assertion_Partial .
```

The body represents the actual — possibly unknown — formal representation. The definition of the body can be included (fully formal claim) or left out (informal claim). As a third option, a partial formalization can be given (underspecified claim):

```
Pub1_Assertion_Partial
Trans1 a transmission .
Trans1 hasCarrier mosquito .
```