



## Nanopublication Guidelines

Concept Web Alliance Working Draft 15 December 2013

### This version:

<http://www.nanopub.org/2013/WD-guidelines-20131215/>

### Latest published version:

<http://www.nanopub.org/guidelines/>

### Latest editor's draft:

[http://nanopub.org/guidelines/working\\_draft](http://nanopub.org/guidelines/working_draft)

### Previous version:

<http://www.nanopub.org/guidelines/1.8>

### Editors:

[Paul Groth](#), [VU University Amsterdam](#)

[Erik Schultes](#), [LUMC](#)

[Mark Thompson](#), [LUMC](#)

[Zuotian Tatum](#), [LUMC](#)

[Michel Dumontier](#), [Stanford University](#)

[Tobias Kuhn](#), [ETH Zurich](#)

### Authors:

[Alasdair J G Gray](#), [Heriot-Watt University](#)

[Christine Chichester](#), [Swiss Institute of Bioinformatics \(SIB\)](#)

[Kees Burger](#), [NBIC](#)

[Spyros Kotoulas](#), [VU University Amsterdam](#)

[Antonios Loizou](#), [VU University Amsterdam](#)

[Valery Tkachenko](#), [RSC](#)

[Andra Waagmeester](#), [Maastricht University](#)

[Sune Askjaer](#), [Lundbeck](#)

[Steve Pettifer](#), [University of Manchester](#)

[Lee Harland](#), [Pfizer/CD](#)

[Carina Haupt](#), [Uni Bonn / Fraunhofer](#)

[Colin Batchelor](#), [RSC](#)

[Miguel Vazquez](#), [CNIO](#)

[José María Fernández](#), [CNIO](#)

[Jahn Saito](#), [Maastricht University](#)

[Andrew Gibson](#), [LUMC](#)

[Louis Wich](#), [DTU](#)

[Jesse van Dam](#), [Wageningen UR](#)

[Mark Wilkinson](#), [Technical University of Madrid](#)

This document is licensed under a [Creative Commons ShareAlike Attribution 3.0 License](#).

## Abstract

**TODO:** Add an abstract!

## Status of This Document

This document is a specification by Concept Web Alliance. It has no official standing of any kind and does not represent the support or consensus of any standards organisation.

## Table of Contents

1. [Introduction](#)
2. [Conformance](#)
3. [Primer](#)
4. [Nanopublication Ontology](#)
5. [Well-formed Nanopublications](#)
6. [Query Template](#)
7. [Integrity Key](#)
8. [Further Information](#)
9. [References](#)
- A. [References](#)
  - A.1 [Normative references](#)
  - A.2 [Informative references](#)

## 1. Introduction

Good descriptions of data are essential to finding, understanding and ultimately reusing data. However, data is all too often published without adequate descriptions of its provenance.

Here, we describe nanopublications, a community-driven approach to representing structured data along with its provenance into a single

publishable and citable entity. A nanopublication minimally consists of an assertion, the provenance of the assertion, and the provenance of the nanopublication. The nanopublication is represented with and may be queried using Semantic Web technologies (RDF, OWL, SPARQL). Nanopublications also feature a mechanism to ensure the integrity of data and its provenance. Access to provenance enables users to assess the trustworthiness of data and provides a mechanism by which authors and institutions may be acknowledged for their contribution to the global knowledge graph. Nanopublications may be used to expose quantitative and qualitative data, as well as hypotheses, claims, and negative results that usually go unpublished. With nanopublications, it is possible to disseminate individual data as independent publications with or without an accompanying research article.

This document describes the structure of a nanopublication. It offers guidelines in their composition, implementation and use. It was produced by members of the Concept Web Alliance (CWA), an open collaborative community that is actively addressing the challenges associated with the production, management, interoperability and analysis of unprecedented volumes of data. We collaborate via [Github](#)

## 2. Conformance

As well as sections marked as non-normative, all authoring guidelines, diagrams, examples, and notes in this specification are non-normative. Everything else in this specification is normative.

The key words **MUST**, **MUST NOT**, **REQUIRED**, **SHOULD**, **SHOULD NOT**, **RECOMMENDED**, **MAY**, and **OPTIONAL** in this specification are to be interpreted as described in [\[RFC2119\]](#).

## 3. Primer

Nanopublications combines an assertion, the provenance of the assertion, and the provenance of the nanopublication into a single publishable and citable entity. Both the assertion and provenance are represented as RDF graphs.

The **assertion graph** of the nanopublication contains an assertion that is comprised of one or more RDF triples (subject-predicate-object tuples). Examples of assertions include:

### Example

```
:assertion {
  ex:trastuzumab ex:is-indicated-for ex:breast-cancer .
}
```

### Example

```
:assertion {
  ex:BRCA1-gene ex:is-involved-in ex:breast-cancer .
  ex:BRCA1-gene ex:encodes ex:BRCA1-protein .
  ex:BRCA1-protein ex:is-expressed-in ex:breast .
}
```

The **provenance graph** of the nanopublication contains one or more RDF triples that provide information about the **assertion**. A nanopublication **MUST** have a provenance graph identifier linked to the assertion graph identifier. Provenance means, 'how this came to be', and may include any statement that discusses how the assertion was generated, who generated it, when was it generated, where was the assertion obtained from, and any other similar information. Examples of assertional provenance include:

### Example

```
:provenance {
  :assertion prov:generatedAtTime "2012-02-03T14:38:00Z"^^xsd:dateTime .
  :assertion prov:wasDerivedFrom :experiment .
  :assertion prov:wasAttributedTo :experimentScientist .
}
```

The **publicationInfo graph** contains one or more RDF triples that offer provenance information regarding the nanopublication itself. In this case, the subject of the triples in the publicationInfo graph **MUST** be the nanopublication URI and **SHOULD** contain attribution and timestamp. Examples of the nanopublication provenance include:

### Example

```
:pubInfo {
  ex:pub1 prov:wasAttributedTo ex:paul .
  ex:pub1 prov:generatedAtTime "2012-10-26T12:45:00Z"^^xsd:dateTime .
}
```

The nanopublication itself receives its own identifier and refers to its parts.

### Example

```
:head {
  ex:pub1 a np:Nanopublication .
  ex:pub1 np:hasAssertion :assertion .
  ex:pub1 np:hasProvenance :provenance .
  ex:pub1 np:hasPublicationInfo :pubInfo .
}
```

As far as applicable, provenance and publication info **SHOULD** use the provenance ontology [PROV-O](#) and/or ontologies that have a mapping to it, including Dublin Core and PAV.

## 4. Nanopublication Ontology

The structure of a nanopublication is defined by the following ontology using the Web Ontology Language (OWL). The namespace <http://www.nanopub.org/nschema>. Our namespace policy is that the current version of the ontology is always at this url. We suggest using [TriG](#) syntax for writing nanopublications.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix rdfg: <http://www.w3.org/2004/03/trix/rdfg-1/>.
@prefix owl: <http://www.w3.org/2002/07/owl#>.
@prefix np: <http://www.nanopub.org/nschema#>.
@prefix dc: <http://purl.org/dc/elements/1.1/>.
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.

<http://www.nanopub.org/nschema> a owl:Ontology ;
  dc:created "2012-10-26"^^xsd:date;
  dc:modified "2013-10-23"^^xsd:date;
  dc:description <http://www.nanopub.org/2013/WD-guidelines-20131215/> ;
  owl:priorVersion <http://www.nanopub.org/nschema-1.9> .

## A nanopublication should be associated with at most one assertion,
## provenance of assertion, and provenance of the nanopublication (publicationInfo)
## We use rdfg:Graph to denote a graph as used in the RDF 1.1. Trig syntax.
## Each sub class of rdfg:Graph should be separate Named Graph

np:Nanopublication rdf:type owl:Class.
np:Assertion rdfs:subClassOf rdfg:Graph.
np:Provenance rdfs:subClassOf rdfg:Graph.
np:PublicationInfo rdfs:subClassOf rdfg:Graph.

np:hasAssertion a owl:FunctionalProperty.
np:hasAssertion rdfs:domain np:Nanopublication.
np:hasAssertion rdfs:range np:Assertion.

np:hasProvenance a owl:FunctionalProperty.
np:hasProvenance rdfs:domain np:Nanopublication.
np:hasProvenance rdfs:range np:Provenance.

np:hasPublicationInfo a owl:FunctionalProperty.
np:hasPublicationInfo rdfs:domain np:Nanopublication.
np:hasPublicationInfo rdfs:range np:PublicationInfo.

```

## 5. Well-formed Nanopublications

The current ontology is loose in its specification to encourage adoption. We thus define an additional set of criteria to further constraint the usage of the ontology. These criteria may be adopted into further versions of the ontology.

A nanopublication **MUST** comply with all of the following criteria to be considered well-formed:

- A nanopublication consists of a set of RDF quads (i.e. subject-predicate-object + context)
- The context (i.e. graph) of each triple has to be specified as a valid URI (i.e. no null values)
- There is exactly one quad of the form '[N] rdf:type np:Nanopublication [H]', which identifies [N] as the nanopublication URI, and [H] as the head URI
- Given the nanopublication URI [N] and its head URI [H], there is exactly one quad of the form '[N] np:hasAssertion [A] [H]', which identifies [A] as the assertion URI
- Given the nanopublication URI [N] and its head URI [H], there is exactly one quad of the form '[N] np:hasProvenance [P] [H]', which identifies [P] as the provenance URI
- Given the nanopublication URI [N] and its head URI [H], there is exactly one quad of the form '[N] np:hasPublicationInfo [I] [H]', which identifies [I] as the publication information URI
- The URIs for [N], [H], [A], [P], [I] must all be different
- All triples must be placed in one of [H] or [A] or [P] or [I]
- Triples in [P] have at least one reference to [A]
- Triples in [I] have at least one reference to [N]

This is an example of a well-formed nanopublication in TriG notation:

### Example

```

@prefix : <http://example.org/pub1#> .
@prefix ex: <http://example.org/> .
@prefix np: <http://www.nanopub.org/nschema#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

:head {
  ex:pub1 a np:Nanopublication .
  ex:pub1 np:hasAssertion :assertion .
  ex:pub1 np:hasProvenance :provenance .
  ex:pub1 np:hasPublicationInfo :pubInfo .
}

:assertion {
  ex:trastuzumab ex:is-indicated-for ex:breast-cancer .
}

:provenance {
  :assertion prov:generatedAtTime "2012-02-03T14:38:00Z"^^xsd:dateTime .
  :assertion prov:wasDerivedFrom :experiment .
  :assertion prov:wasAttributedTo :experimentScientist .
}

:pubInfo {
  ex:pub1 prov:wasAttributedTo ex:paul .
  ex:pub1 prov:generatedAtTime "2012-10-26T12:45:00Z"^^xsd:dateTime .
}

```

And the same example in the N-Quads format:

### Example

```

<http://example.org/pub1> <http://www.nanopub.org/nschema#hasAssertion> <http://example.org/pub1#assertion> <http://example.org/pub1>
<http://example.org/pub1> <http://www.nanopub.org/nschema#hasProvenance> <http://example.org/pub1#provenance> <http://example.org/pub1>

```

```
<http://example.org/pub1> <http://www.nanopub.org/nschema#hasPublicationInfo> <http://example.org/pub1#pubInfo> <http://example.org/pub1> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.nanopub.org/nschema#Nanopublication>
<http://example.org/trastuzumab> <http://example.org/is-indicated-for> <http://example.org/breast-cancer> <http://example.org/pub1#assertion> <http://www.w3.org/ns/prov#generatedAtTime> "2012-02-03T14:38:00Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>
<http://example.org/pub1#assertion> <http://www.w3.org/ns/prov#wasAttributedTo> <http://example.org/pub1#experimentScientist> <http://example.org/pub1#assertion> <http://www.w3.org/ns/prov#wasDerivedFrom> <http://example.org/pub1#experiment> <http://example.org/pub1> <http://www.w3.org/ns/prov#generatedAtTime> "2012-10-26T12:45:00Z"^^<http://www.w3.org/2001/XMLSchema#dateTime>
<http://example.org/pub1> <http://www.w3.org/ns/prov#wasAttributedTo> <http://example.org/paul> <http://example.org/pub1#pubInfo>
```

## 6. Query Template

To extract an entire nanopublication from a triple store, the following SPARQL query template can be used:

### Example

```
prefix np: <http://www.nanopub.org/nschema#>
prefix : <...>
select ?G ?S ?P ?O where {
  {graph ?G { : a np:Nanopublication}} union
  {graph ?H { : a np:Nanopublication { : np:hasAssertion ?G} union { : np:hasProvenance ?G} union { : np:hasPublicationInfo ?G}}
  graph ?G { ?S ?P ?O }
}
```

## 7. Integrity Key

The goal of the integrity key is to establish an identifier that can be used to check if a nanopublication has changed, thus enforcing the immutability of nanopublications. Trusty URIs [10] are the recommended way of assigning integrity keys to nanopublications. This is the exemplary nanopublication from Section 5 after generating and attaching a trusty URI:

### Example

```
@prefix this: <http://example.org/pub1.RAvVDzee5-fpWEFAvoa4Y3_7m9qIXJoKDTdBNbvWwnCiQ> .
@prefix sub: <http://example.org/pub1.RAvVDzee5-fpWEFAvoa4Y3_7m9qIXJoKDTdBNbvWwnCiQ#> .
@prefix ex: <http://example.org/> .
@prefix np: <http://www.nanopub.org/nschema#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

sub:head {
  this: np:hasAssertion sub:assertion ;
        np:hasProvenance sub:provenance ;
        np:hasPublicationInfo sub:pubInfo ;
        a np:Nanopublication .
}

sub:assertion {
  ex:trastuzumab ex:is-indicated-for ex:breast-cancer .
}

sub:provenance {
  sub:assertion prov:generatedAtTime "2012-02-03T14:38:00Z"^^xsd:dateTime ;
  prov:wasAttributedTo sub:experimentScientist ;
  prov:wasDerivedFrom sub:experiment .
}

sub:pubInfo {
  this: prov:generatedAtTime "2012-10-26T12:45:00Z"^^xsd:dateTime ;
  prov:wasAttributedTo ex:paul .
}
```

## 8. Further Information

Please check <http://www.nanopub.org> for further guides and community information.

## 9. References

- [1] Tobias Kuhn, Paolo Emilio Barbano, Mate Levente Nagy, and Michael Krauthammer. [Broadening the Scope of Nanopublications](#). In Proceedings of the 10th Extended Semantic Web Conference (ESWC). Springer, 2013.
- [2] Andrew Gibson, Jesse C.J. van Dam, Erik A. Schultes, Marco Roos, and Barend Mons (2012) Towards Computational Evaluation of Evidence for Scientific Assertions with Nanopublications and Cardinal Assertions (2012) CEUR Workshop Proceedings 952. [http://ceur-ws.org/Vol-952/paper\\_26.pdf](http://ceur-ws.org/Vol-952/paper_26.pdf)
- [3] Mark Thompson, Herman H.B.M. van Haagen, Barend Mons, and Erik A. Schultes (2012) Placing landMarks in the Knowledge Space: crowd-sourcing landmark publications for benchmarking text-mined predictions (2012) CEUR Workshop Proceedings 952. [http://ceur-ws.org/Vol-952/paper\\_14.pdf](http://ceur-ws.org/Vol-952/paper_14.pdf)
- [4] Patrinos et al (2012): Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain. Hum Mutat. 33(11):1503-12. <http://www.ncbi.nlm.nih.gov/pubmed/22736453>
- [5] Axton M (2011) Crowdsourcing human mutations. Nature Genetics 43: 279. <http://www.nature.com/ng/journal/v43/n4/full/ng0411-279.html>
- [6] Mons B et al (2011) The value of data. Commentary, Nature Genetics 43: 281-283. <http://www.nature.com/ng/journal/v43/n4/full/ng0411-281.html>
- [7] Groth P, Gibson A & Velterop J (2010) The Anatomy of a Nanopublication. Information Services & Use 30: 51-56. [http://www.w3.org/wiki/images/c/c0/HCLSIG\\$\\$\\$WANSIOCS\\$Actions\\$\\$\\$RhetoricalStructure\\$\\$\\$meetings\\$\\$\\$20100215\\$cwa-anatomy-nanopub-v3.pdf](http://www.w3.org/wiki/images/c/c0/HCLSIG$$$WANSIOCS$Actions$$$RhetoricalStructure$$$meetings$$$20100215$cwa-anatomy-nanopub-v3.pdf)
- [8] Barend Mons and Jan Velterop (2009) Nano-Publication in the e-science era. [http://www.nbic.nl/uploads/media/Nano-Publication\\_BarendMons-JanVelterop.pdf](http://www.nbic.nl/uploads/media/Nano-Publication_BarendMons-JanVelterop.pdf)
- [9] Linked Data: <http://www.w3.org/DesignIssues/LinkedData.html>
- [10] Tobias Kuhn and Michel Dumontier. [Trusty URIs: Verifiable, Immutable, and Permanent Digital Artifacts for Linked Data](#). In Proceedings of the 11th Extended Semantic Web Conference (ESWC). Springer, 2014.

## A. References

### A.1 Normative references

[RFC2119]

[Key words for use in RFCs to Indicate Requirement Levels](#)

### A.2 Informative references

No informative references.