

# Applicability of the Nano-publication Concept for Fostering Open Access in Developing and Transition Countries

Adam Sofronijević<sup>1</sup> and Aleksandra Pavlović<sup>2</sup>

<sup>1</sup>[sofronijevic@unilib.bg.ac.rs](mailto:sofronijevic@unilib.bg.ac.rs)

University of Belgrade, University library “Svetozar Marković”, Bulevar Kralja Aleksandra 71, 11000 Belgrade (Serbia)

<sup>2</sup>[pavlovic@unilib.bg.ac.rs](mailto:pavlovic@unilib.bg.ac.rs)

University of Belgrade, University library “Svetozar Marković”, Bulevar Kralja Aleksandra 71, 11000 Belgrade (Serbia)

## Abstract

The paper presents nano-publications as a publishing alternative primarily developed for exposing OA (Open Access) datasets. The possible role of nano-publications for fostering usage of OA in developing and transitional countries is analysed. The paper provides insight into the theoretical background of nano-publications and describes the steps necessary to develop this format of publication. In this framework we present a more detailed analysis of the possibilities of involving the scientific communities of developing and transitional countries in operations on OA datasets exposed via nano-publications. Incentives and ethical issues are discussed in order to provide the initial motivation for further research of this novel idea. The role of nano-publications in the area of humanities is briefly described, along with suggestions of the possibilities for involving humanity scientific communities of developing and transitional countries in nanopublishing. The advantages of fostering OA initiatives in developing and transitional countries by implementing activities regarding nanopublishing are presented in order to encourage contemplation on this format of publishing in the OA communities of these countries.

**Keywords:** nano-publication; Open Access; developing and transition countries; dataset publishing; ethical issues; humanities; metadata; scientific communication

## Introduction

A nano-publication is a very short declaration connecting two concepts by means of a third and providing metadata about this relation (conditions

under which the relation is viable, author, timestamp, etc.). An example of such a declaration would be: Trojans (concept 1) are transmitted by (concept 3) exe files (concept 2), with metadata describing the conditions (in PC computers), author, date of publication etc. Nano-publishing is a concept in the making that appeared as a response to the growing inadequacy of various aspects of existing tools for scientific communication. So far, the publishing industry and OA (Open Access) initiatives have not been very successful in using all the advantages of digital technology for the dissemination of scientific results, especially datasets. Therefore novel ideas and solutions are needed in order to expose the huge datasets that now exist, but are still invisible for those outside the labs producing them ([Bourne & de Waard, 2011](#)). Originating in life sciences, nano-publications seem to be envisioned and increasingly shaped as a tool for the efficient publishing of datasets. The example provided at the beginning of this section is general and not related to datasets, but an example of a nano-publication exposing small parts of datasets is given in the following section. The abundance of datasets is a relatively novel development in science. Not long ago quality datasets were strictly guarded and unavailable to outside researchers. Nowadays the gap between available datasets and the resources to even curate them let alone analyse them is widening each day. Contemporary tools for publishing datasets seem to be insufficient in comparison to the amount of data available and the possibilities that more effective data dissemination tools might offer. In this regard, one must observe that there are no sufficient incentives for researchers to publish datasets apart from those being used to support the publishing of datasets closely related to findings published in traditional scientific journals. The nano-publication concept has the potential to successfully face the challenge of providing a novel method of evaluating datasets and scientific work based on them, while at the same time preserving the values of the traditional means of scientific communication ([Mons et al., 2011](#)). When considering the ethical issues that need to be addressed in regards to nanopublications, it is good to keep in mind the ethical basis of the existing publishing paradigm ([Axton, 2011](#)). With nano-publications a new publishing paradigm is rising and all of its aspects need to be addressed – incentives to use it, operational details, but also the ethical issues. Nano-articles provide the basis for contemplation on the improvement of scientific communication in other areas such as

humanities. The nano-publication concept which we will discuss in the next section – though derived from life sciences – is nevertheless general enough to allow for contemplating the use of the concept in other scientific areas and for the purposes of publishing data originating in other scientific areas.

### **Defining nano-publications**

In order to define nano-publications clearly, an example available at ([Nanopublication Downloads section, 2012](#)) is provided, with only the most important parts presented for the sake of clarity:

```
<nanopublication id="0">
  <assertion>
    <subject>NG_000007.3:g.70628G>A</subject>
    <predicate>has variant frequency</predicate>
    <object>0.25%</object>
  </assertion>
  <condition>Sardinian</condition>
  <provenance>
    <dateofcreation>March 24, 2011</dateofcreation>
    ...
    <evidenceType>empirical</evidenceType>
    <authorID>Giardine et. al.</authorID>
    <curatorID>unresolved</curatorID>
    ...
  </provenance>
</nanopublication>
```

A nano-publication is identified uniquely by its id. The assertion part of the nano-publication offers the relation of the concepts and is followed by the condition under which the assertion holds and by the metadata describing the context of the relation. The nano-publication is presented by means of XML which enables machine readability.

Nano-publications will be presented in more detail by providing a short description of the steps necessary to perform in order to create a nano-publication as described in ([Mons & Velterop, 2009](#)).

The first step addresses the problem of ambiguity coming from the usage of different terms and identifiers that refer to a concept. A concept, which

---

is a unit of thought, may be related to an object in the physical world, but may also be related to a completely abstract object. The natural flow of any language, when used for communication by humans, is to describe the same concept by numerous different terms and identifiers. This creates ambiguity and causes problems for machine interpretation of such statements. Because of this, the first step in the creation of nano-publications is mapping the terms and identifiers to concepts, and by this the elimination of ambiguity.

The second step comprises the creation of statements from concepts. Statements are considered to be the smallest insights in exact science, consisting of three concepts. These three concepts are usually called subject, object and predicate, where the subject and object are associated by means of predicate. While the relationship between these three concepts is clear enough, the statement itself needs additional metadata in order to ensure clarity in regards to the conditions under which the statement holds, the authority, the creation date and other annotated metadata deemed necessary by the community using nano-publications. Adding these metadata is considered to be step three in nano-publication creation. We need to emphasise the simultaneous unity and difference of steps two and three. The completeness and consistency of the statement with the metadata describing it is imperative, but it is very important for us to notice that the process of the creation of a statement is not necessarily done by the same entity and in the same timeframe as the process of the annotation of that statement.

The third step is about making boundaries for the statements i.e. providing the range of validity, specifying authority and adding other important metadata about the statement so that it makes sense for the community using it. If conditional statements are used to describe the conditions under which the said statement is to hold, then it is imperative that the describing statement also comprises of unambiguous concepts. This is to ensure that operations on such a statement can be seamlessly performed along with operations on the original statement. In regards to authority, the importance of factors defined as status by ([Mons & Velterop, 2009](#)) is exceptional. Status may describe any of the following characteristics of the

annotated statement: peer-reviewed, curated, disputed, retracted, hypothetical, observational, repetitive, etc.

The fourth step brings annotated statements to the level of proper publications that can be cited and credited by treating them as nano-publications. By making this crucial step, the incentives to publish the short statements in the form of nano-publications will appear and their number should increase over time. ([Mons & Velterop, 2009](#)) distinguish three types of statements suitable for publishing in the format of nano-publications. These are curated statements, observational statements and hypothetical statements. Curated statements are the building blocks of ontologies and are intrinsically statements that are “as true as it gets in science”. They should be described by metadata containing data on the originator of the triple statement and its curator, in order to facilitate incentives for those two categories. Observational statements include factual statements originating in the literature or in datasets. The more observational statements become annotated and published in the format of nano-publications, the more knowledge will become efficiently accountable. Hypothetical statements represent triplets of yet undiscovered statements that by virtue of reasoning or data mining or inference from existing texts have a high probability of being true. The incentive for publishing such statements is that – when confirmed or “discovered ” by other statements that originate in experimental results – the authors of hypothetical statements may get credit for their prediction along with authors confirming it.

Step five is about making nano-publications really useful. To achieve this, the redundancy in the present body of published scientific knowledge needs to be addressed. In order to facilitate understanding and following what is natural for language usage as a communication tool among humans, repetitions are very common in scientific publishing. This may be useful for us, but for machines this redundancy makes searching and retrieval inefficient. When converted to annotated triplets most of the data becomes redundant, thus making the remaining triplets really useful for researchers since they are now easily discoverable by machine aided search and retrieval.

As stated by ([“NanoPub: A beginner's guide to data publishing”, 2012](#)) nano-publications are based on open standards such as RDF and XML.

---

Adhering to standards should provide for machine readability of nano-publications, which has enormous importance for raising the efficiency of scientific communication. The lack of standards as of April 2012 should be noted, along with the existence of guidelines for the creation of nano-publications ([“Open PHACTS Nanopublication Guidelines v1.81”](#), 2012).

Some general principles of nanopublishing are presented at ([“NanoPub: A beginner's guide to data publishing”](#), 2012). The principle of added value addresses the novelty and quality of nano-publications by presenting examples of the origin of a predicate that associates an object with its subject (mathematical model, new experimental dataset, exposure of an existing dataset, etc.). The principle of transparency establishes the assurance of quality assessment of a nano-publication. The metadata describing statement should provide enough data on the predicate to ensure that others can assess its validity by author, date and time, spatial data and intrinsic quality of the predicate. The principle of ambiguity avoidance describes the need for all concepts in the statement and the metadata describing it to be unambiguously resolved to certain concepts. The principle of global reference addresses the issue of the referencing of entities in nano-publications. If an entity is established on the Web the accompanying URI should be used, and if not a UUID (Universal and Unique Identifier) standardized by the Open Software Foundation (OSF) should be derived from the available underlying database of concepts.

### **Nano-publications in developing and transition countries**

Nano-publications have the potential to be an important driver of scientific research in first world countries which are ample in resources that allow for the production of datasets in areas like life sciences. Could nano-publications also play a significant role in fostering communication and usage of OA resources in scientific communities of developing and transitional countries? In general these countries lack the resources and the infrastructure for conducting research that produces large datasets. Consequently entire scientific disciplines in these countries are left behind or become obsolete. In order to prevent this, the scientific communities of these countries should be focusing their efforts on the usage of freely available scientifically relevant datasets. This was not the case in the past,

because datasets were not freely exposed. All of this may change with the rise of nano-publications.

Scientific communities and librarians of developing and transitional countries can engage in several activities and operations based on freely available datasets that may start to emerge with the wider usage of nano-publications. In order to have the basis for participation in operations over datasets, they may engage in creating and editing concepts regarding the members of their own community in databases such as ConceptWiki. Furthermore they can add to the mapping efforts that lead to unambiguously resolving the terms and identifiers into unique concepts. The need for *a posteriori* mapping of terms and identifiers in already published texts is vast, and text and data mining combined with human curation input can provide significant results in this area. If we further follow the steps in the creation of nano-publications described in the previous section, steps two and three provide a rich source of activities that may be of interest to scientific communities of developing and transitional countries. The exposure of triplet statements from datasets that are freely available is one such activity. The annotation of these triplet statements that were either produced by those creating annotation or by someone else is the next possibility. Various statements regarding the context and provenance of the triplet provide many opportunities if an incentive system is created – primarily within scientific communities of developing and transitional countries, but also worldwide. The creation of hypothetical statements building on available datasets may be especially interesting. This comes from the high probability that a worldwide incentives system may emerge for these types of predictions. Redundancy removal described as step five in the previous section also offers some possibilities for the activities of members of scientific communities of developing and transitional countries. The analysis of statements and the removal of redundancies along with the clustering and linking of statements that are not redundant may present important work that should upgrade the existing semantic Web and provide for fair rewards for those involved with it.

So far we have explained in some detail the usage and purpose of nano-publications in life sciences. In developing and transitional countries humanities and social sciences scholars are still numerous when compared

to statistics of the first world countries. Besides this, humanities and social sciences in developing and transitional countries are especially hard-hit by the widening gap in skills and knowledge when compared to their peers in first world countries, and also by budget cuts due to the ongoing financial crisis. Therefore, in order to discuss the role of nano-publications in developing and transitional countries we need to touch upon their possible role in humanities. Very few mentions of this role are available as of April 2012. ([Mons & Velterop, 2009](#)) mention in passing the applicability of nano-publications in humanities and conclude that it may have an important role in this area as well, and that most of the characteristics described on the basis of examples originating in life sciences may also be applicable for humanities. ([Chambers, 2011](#)) provides a more detailed example of a nano-publication originating from humanities. Based on this example it seems that nanopublishing should be embraced by the humanities scientific community, especially in regards to recent developments regarding the increasing interest in digital humanities. This may be an important field of activity for scientific communities in the humanities from developing and transitional countries, since resources needed to employ nanopublishing are mostly available to them thanks to the fact that they are mostly open sourced. Basic ICT equipment should suffice to employ these open sourced technologies and it is up to the mentioned scientific communities to demonstrate their innovativeness and employ the idea originating from other scientific fields to their own with merit and ethical stance. The creation of a database comprising concepts in humanities may be followed by nanopublishing initiatives similar to the one presented in ([Chambers, 2011](#)).

All mentioned activities are very promising if fair use is applied i.e. if ethical standards of scientific conduct are respected. In order to ensure this, especially in scientific communities of developing and transitional countries that may still be struggling to achieve proper levels of scientific ethics, a system of regulations and control need to be established. On the other hand, imposing too much control and regulation in an area that builds on OA principles and high ideals of international scientific community collaboration may damage the overall effort and stall the process to a standstill. Therefore we propose a balanced approach and employment of self-regulatory mechanisms. One such example may be Concept-

Wiki, where quality is intrinsically ensured by the properties of Wiki technology that enables seamless internal control of those involved in the process of creation and the editing of concepts. In general we propose usage of Web 2.0 technologies for the dissemination of results in the area of nanopublishing because by their own virtue these technologies provide a transparent environment that is likely to expose any misbehaviour. Mandatory blogging of results for all those involved in this line of scientific work is one example of activities that add to the transparency of the nanopublishing process. The importance of the ethical issues regarding scientific work in the area of nano-publications is underlined in ([Groth, Gibson & Velterop, 2010](#)), with an emphasis on citation and the *asserted by* mechanism that allows for the evaluation of the source grounding the base for nano-publications. Existing guidelines for the creation of nano-publications ([“Open PHACTS Nanopublication Guidelines v1.81”, 2012](#)) do encompass an Integrity key that should ensure the authority of nano-publications and a nano-publication ID that should ensure its unique identification among other nano-publications. As of April 2012 these two parameters are not yet technically described in these guidelines, but their mere presence suggests the importance of ethical issues in nanopublishing.

Finally, does nano-publishing have the potential of fostering OA in developing and transitional countries? We have shown that strong incentives may exist for the usage of nano-publications in developing countries. Since all concepts and technologies involved with nano-publications are open sourced and compliant with the OA initiative, the sheer usage of nano-publications should be considered as an important part of OA usage. Further, by becoming familiar with different open source technologies and ideas that are intrinsic for nano-publications, they may become a specific driver of OA in developing and transitional countries. Therefore we propose special support for nano-publication initiatives from the OA community in order to foster the usage of OA materials in conjunction with nanopublishing. Also, the sheer usage of nano-publications in developing and transitional countries will ensure the dissemination of novel technologies and ideas such as RDF, XML, URI, UUID, ConceptWiki, etc. that ground a basis for the usage and flourishing of OA initiatives.

## **Conclusions**

Nano-publications have always been related to OA (Open Access). This mostly comes from the opposition of authors contemplating it to the socially inefficient system prevailing the contemporary publishing industry. Nano-publications were related to OA either directly – by referencing the concept as such by authors – or indirectly – by means of relating the background of authors contemplating it. In all cases nano-publications were considered as OA publications only. Nanopublishing may also be considered as an important driver of OA promotion and usage in general. This especially stands for developing and transitional countries, where the specific circumstances of scientific activities predominantly determined by lack of resources promotes the usage of existing datasets via nano-publishing. A possibility for nano-publishing usage in the area of humanities is also an important aspect of their overall applicability in these countries.

Nano-publication theory and practice are both in a very early phase. A lot of additional contemplation and implementation experience is needed in order to further develop this idea and achieve its wider usage. In spite of this, it is obvious even in this early phase that nano-publications are inherently efficient in scientific knowledge dissemination. This important fact may be used as a basis for further contemplation and practical implementation, but one also has to have in mind that this too could lead to the commandeering of this publication form by the for-profit publishing industry. Nano-publications, especially some of their elements like the open availability of data (full text / raw data) and the drop of emphasis on the “publish or perish” attitude ([Laika's Medliblog, 2010](#)), have been perceived in a broader context as a new form of scientific dissemination and access provisioning that can contribute in reducing the tensions caused by the rigidity of the publishing industry. Because of this, theoretical and practical provisions may be needed in order to enable equally qualitative usage in different business models employing this novel concept.

## **References**

- Axton, M. (2011). Crowdsourcing human mutations. *Nature Genetics*, 43(4), 279-279. [doi:10.1038/ng0411-279](https://doi.org/10.1038/ng0411-279).
-

- Bourne, P., & de Waard, A. (2011). Notes from the breakout sessions. In: *Workshop Beyond the PDF* held on January 19-21, 2011 at University of California San Diego. Retrieved from <http://sites.google.com/site/beyondthepdf/home/program-draft/notes-from-the-breakout>, 2012 Jan 25.
- Chambers, S. (2011). Nano-publications in the arts and humanities. Retrieved from <http://www.slideshare.net/schambers3/nanopublications-in-the-arts-and-humanities>, 2012 Feb 15.
- Groth, P., Gibson, A., & Velterop, J. (2010). The anatomy of a nanopublication. *Information Services and Use*, 30(1), 51-56. Retrieved from [http://www.w3.org/wiki/images/c/c0/HCLSIG\\$\\$SWANSIOC\\$\\$Actions\\$\\$RhetoricalStructure\\$\\$meetings\\$\\$20100215\\$cwa-anatomy-nanopub-v3.pdf](http://www.w3.org/wiki/images/c/c0/HCLSIG$$SWANSIOC$$Actions$$RhetoricalStructure$$meetings$$20100215$cwa-anatomy-nanopub-v3.pdf)
- Laika's, M. (2010). Will Nano-Publications & Triplets Replace The Classic Journal Articles? Retrieved from <http://laikaspoetnik.wordpress.com/2010/06/23/will-nano-publications-triplets-replace-the-classic-journal-articles/>, 2012 Jan 24.
- Mons, B., & Velterop, J. (2009). Nano-Publication in the e-Science Era. In: *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*. Retrieved from [http://www.w3.org/wiki/images/4/4a/HCLSIG\\$\\$ISWC2009\\$\\$Workshop\\$\\$Mons.pdf](http://www.w3.org/wiki/images/4/4a/HCLSIG$$ISWC2009$$Workshop$$Mons.pdf)
- Mons, B., Van, H.H., Chichester, C., Hoen, P., Den, D.J.T., Van, O.G., et al. (2011). The value of data. *Nature Genetics*, 43(4), 281-283. [pmid:21445068](https://pubmed.ncbi.nlm.nih.gov/21445068/), [doi: 10.1038/ng0411-281](https://doi.org/10.1038/ng0411-281).
- NanoPub: A beginner's guide to data publishing (2012). Retrieved from <http://nanopub.org/original/index.html>, 2012 Apr 10.
- Nanopublication Downloads section, Human Hemoglobin Genetic Variation (2012). Retrieved from [http://nanopub.org/original/hemoglobin\\_nano-pub\\_download.html#](http://nanopub.org/original/hemoglobin_nano-pub_download.html#), 2012 Apr 10.
- Open PHACTS Nanopublication Guidelines v1. 81 (2012). Retrieved from [http://www.openphacts.org/documents/publications/OpenPHACTS\\_Nanopublication\\_Guidelines\\_v1.8.1.pdf](http://www.openphacts.org/documents/publications/OpenPHACTS_Nanopublication_Guidelines_v1.8.1.pdf), 2012 Apr 10.