

Data Publishing Using Nanopublications

Mark
Thompson

Barend
Mons

Marco
Roos

Erik
Schultes

Leiden University Medical Center

ABSTRACT

Published data, whether in traditional publication formats such as research articles or in databases often lack a consensus structure (which slows search and reasoning) and provenance and citation models (which lowers incentive for publication [1]). Furthermore, in some disciplines the growing rate of data production exceeds the capacity of human comprehension. Together, these trends lead to the loss of valuable data from scientific discourse. Nanopublication is a data publication model built on top of existing Semantic Web technologies to counter these data dissemination and management trends [2]. A nanopublication represents the smallest unit of publishable information and consists of an (i) assertion and (ii) provenance [3]. The assertion takes the form of one or more semantic triples (subject-predicate-object combination). The provenance describes how the assertion ‘came to be’, and includes supporting information (e.g., context, parameter settings, a description of methods) and attribution to the authors (of content) and creators (of the nanopublication), institutions supporting the work, funding sources and other information like date and time stamps and certification. Creating a nanopublication requires a one-time effort to model the assertion and provenance as RDF named graphs [3]. After submission to an open, decentralized nanopublication store, the nanopublication will be available both to humans and automated inference and discovery engines. Nanopublications can be used to expose quantitative and qualitative data, experimental data as well as hypotheses, novel or legacy data and negative data that usually goes unpublished. Nanopublications are meant to augment traditional long-form narrative.

BODY

The nanopublication model incentivizes rapid, citable data dissemination, interoperability, semantic reasoning, and knowledge discovery.

References

- [1] B. Giardine, J. Borg, D. R. Higgs, et al. Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. *Nature Genetics*, 43(4):295–301, 04 2011. URL <http://dx.doi.org/10.1038/ng.785>.
- [2] B. Mons, H. van Haagen, C. Chichester, P.-B. t. Hoen, et al. The value of data. *Nature Genetics*, 43(4):281–283, Apr 2011. ISSN 1061-4036. doi: 10.1038/ng0411-281. URL <http://dx.doi.org/10.1038/ng0411-281>.
- [3] E. Schultes, P. Groth, A. Gibson, et al. *Nanopublication Guidelines*. The Open PHACTS RDF/Nanopublication Working Group, v1.81 edition, URL <http://www.nanopub.org>.

Volume 1 of Tiny Transactions on Computer Science

This content is released under the Creative Commons Attribution-NonCommercial ShareAlike License. Permission to make digital or hard copies of all or part of this work is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. CC BY-NC-SA 3.0: <http://creativecommons.org/licenses/by-nc-sa/3.0/>.