Hourglass-based interoperability through nanopublications in VODAN-A

André Cardoso Silva Ferreira [a,*] and Mirjam van Reisen [b,c]

[a] *Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands*
ORCID: *https://orcid.org/0000-0002-7838-0061*
[b] *Leiden University Medical Center, Leiden University, The Netherlands*
ORCID: *https://orcid.org/0000-0003-0627-8014*
[c] *Faculty of Humanities and Digital Sciences, Tilburg University, The Netherlands*

Abstract. This paper makes the case for the potential use case of nanopublications to expand the interoperability of the Virus Outbreak Data Analysis Network – Africa given their similarity to FAIR Digital Objects.

Keywords: Health records, interoperability, FAIR Digital Objects, nanopublications, VODAN-A

1. Introduction

This paper will explore the FAIR Hourglass Model [6,18] and how nanopublications could be useful and potentially implemented in the future as approximations of FAIR Digital Objects (FDOs) [19] in the Virus Outbreak Data Analysis Network – Africa (VODAN-A) to aid in its interoperability within the Internet of FAIR Data and Services (IFDS).

Firstly, the VODAN-A project will be introduced. Secondly, the IFDS will be set as context to lay out the importance of FDOs in guaranteeing interoperability according to the FAIR Hourglass Model. In the following sections nanopublications are explained in more detail and it is explored how they could be used in VODAN-A and their benefits.

2. VODAN-A

The VODAN-A project consists of a partnership between academic groups supporting 88 health facilities in Uganda, Kenya, Tanzania, Ethiopia, Somalia, Nigeria, Liberia, Zimbabwe and Tunisia [25] with greater coverage

---

*Corresponding author. E-mail: a.f.cardoso.silva.ferreira@umail.leidenuniv.nl.

expected across the continent expected within the next few years. The overarching goal of the initiative is to create a system where healthcare providers can collect and retain sovereignty over their patients records [23], where the data in every section of the project follows the Findable, Accessible, Interoperable and Reusable (FAIR) principles [17]. Significant effort has been poured into making VODAN-A adhere to the FAIR principles and the project has achieved success, particularly on findability and accessibility parts, as well as supporting data reuse [16]. This article deals with the question of how advancing interoperability could be considered in VODAN-A.

3. Internet of FAIR data and services and FAIR digital objects

The European Commission [5] and many other stakeholders in scientific fields [21] find that there is a need to build on the current Internet model to create an Internet of FAIR Data and Services referred to as the IFDS. In order to achieve more interoperability, the IFDS needs to replicate the growth of the current version of the Internet, whose success is often attributed to its hourglass model [2]. In order to do this, the FAIR Hourglass model [18], depicted in Fig. 1 was designed to foment adoption to the IFDS. Its shape showcases the high freedom of choices and implementation possibilities by users and developers on the top and bottom of the hourglass, the parts referring to the FAIRification processes responsible for converting raw data into FAIR-compliant units of information and the bulb related with high-level applications such as data integration, visiting and analytics, respectively. The narrow center conceptualizes the need to have a clearly defined FDO, a protocol to which all FAIR data should eventually comply to in order to be truly interoperable within the IFDS. The FDOs are the equivalent of the IP/TCP protocol in the Internet, but in this case spanning layers of the IFDS.

The notion of FDOs has already existed for some time [13], but has yet to be properly defined. While a working draft exists on the Fair Digital Objects Framework [4], this version is expected to be refined in the next year, following the 1st International Conference of FAIR Digital Objects (FDO2022) with the Leiden Declaration on FAIR Digital Objects [7].

4. Nanopublications

Some quasi-FDO solutions do already exist, including nanopublications which have been acknowledged to be good approximations for simple assertions [19,22]. These nanopublications are written in RDF and follow a clear
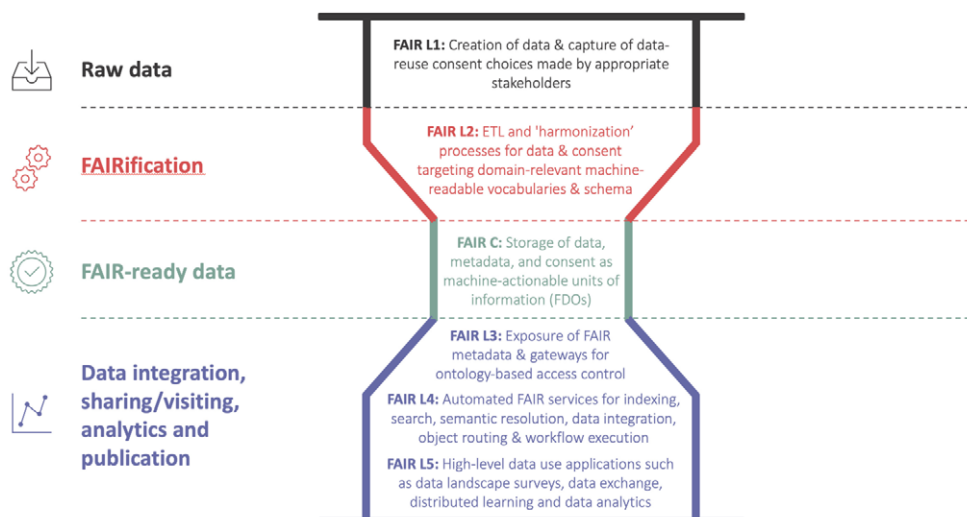

Fig. 1. Adapted from [18]. The FAIR Hourglass Model showcasing the raw data at the top which can then be made into FAIR data through a FAIRification process, culminating in machine-actionable units of information (FDOs) that can be used at a later stage by a wide variety of tools and services.

Fig. 2. Nanopublication example showcasing the interaction between two proteins. Structure from (left) [8] and nanopublication from (right) [10].

structure, enabling machine readability and, provided with the proper tools and descriptions, machine actionability. An example nanopublication in Fig. 2 highlights the key parts of the nanopublication are the assertion, the provenance and the publication info, which can be described as follows [14]:

1. The assertion: Essentially a statement containing the content of the nanopublication. Taking the example provided above: The product protein from gene ENSG00000103197 interacts through a particular pathway with a product of gene ENSG00000117020.
2. The provenance: An indication of the study/methodology/origin of the assertion. In the case of the example provided, the provenance is a study on thyroid hormone production and their peripheral downstream signaling effects regarding congenital hypothyroidism.
3. The publication info: A rich metadata of the nanopublication, containing information such as license, authors or institutions involved, which are key to determine reusability for other researchers. In the example provided in Fig. 2, we can see that the nanopublication was minted in 2020 as a part of a Maastricht University study under CC0 1.0 Universal (CC0 1.0) Public Domain Dedication.

As is evident by the example, some experiments have been attempted with nanopublications in the Biomedical field [12] [11] to optimize the output of data in a format that other researchers and systems in the community could interpret, enabling reusability and interoperability. The concept 'VODAN' is also published as a nanopublication [24] and can be seen in Fig. 3.

5. VODAN-A pipeline

The proposed pipeline in Fig. 4 follows the FAIR Hourglass Model and retains the same core principle of VODAN-A: to enable data sovereignty while stimulating FAIR cooperation. The process starts with each individual health clinic creating new records of patient admittance or disease evolution. To FAIRify the data, CEDAR templates are used with specific and pre-defined fields and ontologies. The CEDAR templates were used to create patient records from abstract registers at point of service. Existing data in csv format is also uploaded using a bulk-input facility created by the community. The data is stored in JSON-LD. Triple store also hosts RDF triples that can be visited. As a next step, nanopublications can be minted for each circumstance, from admitting a patient to the patient being diagnosed with a disease using a tool such as Nanobench [9].

The (meta)data is stored locally in triple store while adhering to each nation's relevant local privacy and medical laws [23]. Depending on the regulations in place, the analytics can then take place with the full (meta)data or an

```
@prefix this: <http://purl.org/np/RAdDKjIGPt_2mE9oJtB3YQX6wGGdCC8ZWpkxEIoHsxOjE> .
@prefix sub: <http://purl.org/np/RAdDKjIGPt_2mE9oJtB3YQX6wGGdCC8ZWpkxEIoHsxOjE#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix np: <http://www.nanopub.org/nschema#> .
@prefix orcid: <https://orcid.org/> .
@prefix nt: <https://w3id.org/np/o/ntemplate/> .
@prefix icc: <https://w3id.org/fair/icc/terms/> .
@prefix npx: <http://purl.org/nanopub/x/> .

sub:Head {
    this: np:hasAssertion sub:assertion ;
        np:hasProvenance sub:provenance ;
        np:hasPublicationInfo sub:pubinfo ;
        a np:Nanopublication .
}

sub:assertion {
    sub:community a icc:Community ;
        rdfs:comment "The primary objective of the VODAN Implementation Network is to showcase the creation and deployment of FAIR data related to COVID-19" ;
        rdfs:label "Virus Outbreak Data Network (VODAN) Implementation Network" ;
        rdfs:seeAlso <https://www.go-fair.org/implementation-networks/overview/vodan/> .
}

sub:provenance {
    sub:assertion prov:wasAttributedTo orcid:0000-0002-4182-7560 .
}

sub:pubinfo {
    sub:sig npx:hasAlgorithm "RSA" ;
        npx:hasPublicKey
"MIGfMA0GCSqGSIb3DQEBAQUAA4GNADCBiQKBgQDHYa4Qlb25/9DnqivTh2kxQMVpFQ0r/cXtOgDQM5dWNrGQVG9RVlXnyt9qowP+k+4bhzcHPQkYWd7YVFBX8FrEnBIReUuRLElYhFrn9jf7w2+YWN8WxKIyBaNZVtgjZpHwGrZOaLOIsl02kEK1G2Qc3FXYCZFWuRJDcafe58
;
        npx:hasSignature "aCUYgYKXkdnKcELn8aKgupTVAPUvpuAO5ZcfF5XLQ5V1KX03cMK4cBA5BKATUwiv2or9SN3b0s1rbxf10xhGccCFMkGf9Mlp1BqND50ep3GcDLzTUKqGkv9mblMPoWAkPgyPXkWa/viEqFILFA6T0GKscJ785wfX57OtAm8dyqg=" ;
        npx:hasSignatureTarget this: .
    this: dct:created "2020-05-01T10:05:25.575+02:00"^^xsd:dateTime ;
        dct:creator orcid:0000-0002-4182-7560 ;
        nt:wasCreatedFromTemplate <http://purl.org/np/RAtpy2wjYJtUUFDHB_qpSiqNR5ALN9dGHDwS97CWfuVe0> .
}
```

Fig. 3. The nanopublication with an assertion declaring the purpose of the VODAN-A network using FAIR data to tackle the COVID-19 pandemic.
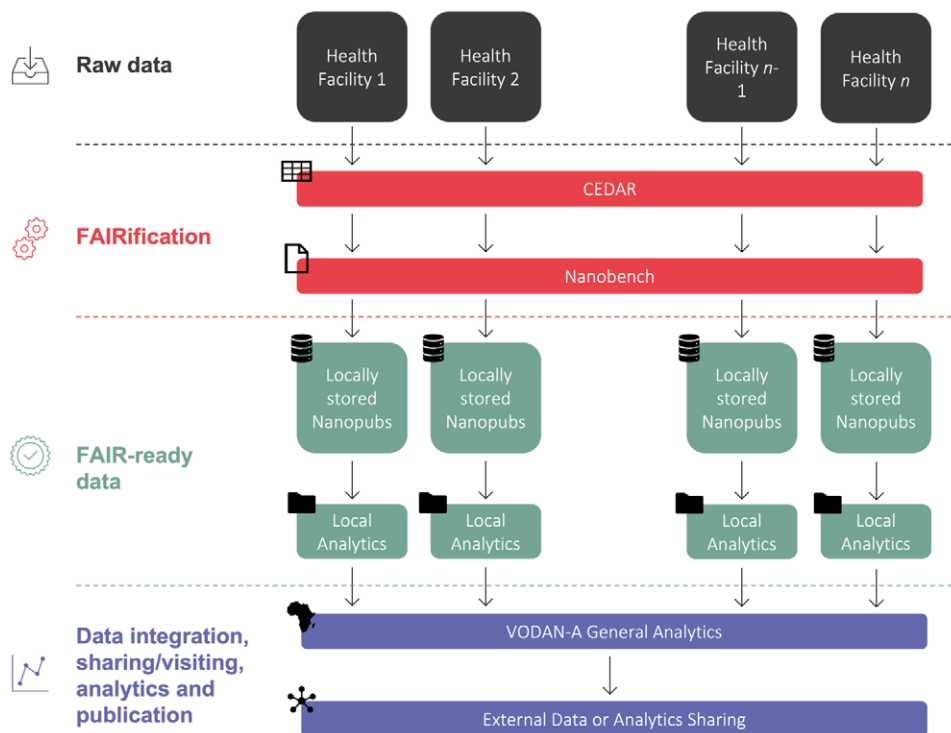
Fig. 4. Proposed high-level structure for VODAN-A using nanopublications as a minimum unit of usable data and taking the FAIR Hourglass Model into account. Here we can see showcased the journey starting in the top of the hourglass with raw data and FAIRification process, the center with FAIR-ready data in the form of nanopublications and the last part of data integration/sharing/visiting analytics and publication, corresponding to the bottom of the hourglass.

anonymized version at a local level, where the clinic can get insights into the patient intakes, diagnoses and outcomes [1]. To cooperate with other clinics on a higher level while retaining data ownership, sensitive data is not passed on towards extraterritorial entities. Given that nanopublications can act as linked data creating relationship graphs, the categories of the nodes and transitions could potentially be shared without ever revealing the (meta)data that generated it. With proper authorization and through a FAIR Data Point, the (meta)data stored locally could also be visited.

6. On interoperability in VODAN-A using nanopublications

Dissecting the three interoperability principles [26], we can assess whether a nanopublication schematic can meet the requirements:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

The FAIR Hourglass Model conceptualizes that a minimum standard of machine-actionable (meta)data units is stored following a strict protocol or structure for this (meta)data. This is done in order for users and developers conducting research or developing tools on either the top or bottom of the hourglass to be able to design experiments using multiple sources of data as these come in very well defined packets with a clear structure that machines can be made to interpret and handle.

Nanopublications have a permanent identifier, are written in RDF and have a formal description in an assertion-provernance-publication information structure, designed to be unambiguously created and interpreted by machine actors [15]. These are approximations to FDOs and can be interoperable with other tools and services in other health clinics and external partners to the VODAN-A project, if sharing is authorized. This will enable cooperation between clinics and allow for VODAN-A stakeholders to position themselves as a go-to project for medical cooperation and research in the IFDS.

I2. (meta)data use vocabularies that follow FAIR principles

With the structure defined, the vocabularies to be used by the nanopublications can be those used by the relevant community or group domain. In the case of VODAN-A, customized ontologies could be created and published while also using those already made available in platforms such as BioPortal [3].

I3. (meta)data include qualified references to other (meta)data

While nanopublications are permanent and immutable, several nanopublications can be connected and reference each other, acting as versioned records. Additionally, records can also point to other nanopublications, creating a linked network of qualified and citable references.

7. Possibilities in VODAN-A for analytics using nanopublications

Using nanopublications as minimum units of information containing assertions about an object, allows for the creation of linked data units and, as such, visually representable graphs with a multitude of analysis possibilities [20].

The appearance of particular motifs in certain graphs can be of interest for analysis. As an example, a circular motif representing the patient being readmitted could be a signal that the initial treatment was not beneficial. In the case of services related to Antenatal Care (ANC) which is one of the data that is being collected, linking each visit and looking into the progress across visits until delivery. Furthermore, one can also link a mother's data if her record appears on Out Patient Register (OPD), which is a separate service. One can thus analyze and relate the treatment given to her and other diagnostic information at patient level. Centrality analysis could identify top nodes in a nanopublication-made network to determine, for instance, the most common symptoms or factors connected with a diagnosis.

Should VODAN-A expand to adopt more fields in the data entry stage, even if made by machines, such as medicines prescribed or patient evolution after medication, more patterns can be investigated. Using path analysis, the best course of treatment to treat a particular disease could be assessed by researching the shortest paths from a

particular stage to cure, which could be a powerful tool in novel outbreaks for which solid treatment plans do not yet exist, as was the case with COVID-19 in the earlier phases of the pandemic.

In order for these tools to be successful, however, the minting of nanopublications would have to follow certain VODAN-A-wide standards to guarantee that all facilities are outputting patient information using nanopublications in the same manner, creating similar paths.

8. Nanopublications and the Leiden declaration on FAIR digital objects

While nanopublications are a good first step towards increased interoperability within the IFDS, it is not guaranteed that they will prove to be a solution compliant with the upcoming Leiden Declaration on FAIR Digital Objects by multiple key stakeholders in the FAIR ecosystem to define the FDO framework.

The framework will allow for the FAIR Hourglass Model to start shaping the IFDS, as the center will be agreed on. Until then, it can be argued that nanopublications are but simply one of many available solutions in an attraction stage of the FDOs of the IFDS, with convergence on the matter expected soon.

Disclosures and conflict of interest

All authors report no conflict of interest.

References

[1] S. Amare, G. Taye, T. Gebreselassie and M. van Reisen, Realising health data-interoperability in low connectivity settings: The case of VODAN-Africa, 2022.

[2] M. Beck, On the hourglass model, *Communications of the ACM* **62** (2019), 48–57. doi:10.1145/3274770.

[3] BioPortal. Available from: https://bioportal.bioontology.org/ontologies/SIO.

[4] L.O.B. da Silva Santos, FAIR digital object framework documentation, 2022. Available from: FAIR digital object framework documentation.

[5] European-Commission, Realising the European open science cloud: First report and recommendations of the Commission high level expert group on the European open science cloud, 2016. Available from: https://data.europa.eu/doi/10.2777/940154.

[6] FAIR Wizard of Leiden, 2022. Available from: https://bit.ly/FWLnano.

[7] GO-FAIR. 1st International FDO Conference, 2022. Available from: https://www.go-fair.org/events/1st-international-fdo-conference/.

[8] P. Groth, A. Gibson and J. Velterop, The anatomy of a nanopublication, *Information Services & Use* **30** (2010), 51–56. doi:10.3233/ISU-2010-0613.

[9] T. Kuhn, R. Taelman, V. Emonet, H. Antonatos, S. Soiland-Reyes and M. Dumontier, Semantic micro-contributions with decentralized nanopublication services, *PeerJ Computer Science* **7** (2021), e387. doi:10.7717/peerj-cs.387.

[10] R. Lahaije, E. Willighagen, E. Ehrhart, E. Weitz and L. Dupuis, Thyroid hormones production and peripheral downstream signaling effects (Homo sapiens) nanopublication, 2020. Available from: https://np.petapico.org/RAMc79WFgK30tHAeWj6PugKxfcqFsgErU2n8pLPJnwNrU.

[11] M. Lizio, J. Harshbarger, H. Shimoji, J. Severin, T. Kasukawa, S. Sahin et al., Gateways to the FANTOM5 promoter level mammalian expression atlas, *Genome Biology* **16** (2015), 22. doi:10.1186/s13059-014-0560-6.

[12] E. Mina, M. Thompson, R. Kaliyaperumal, J. Zhao, E. van der Horst, Z. Tatum et al., Nanopublications for exposing experimental data in the life-sciences: A Huntington's disease case study, *Journal of Biomedical Semantics* **6** (2015), 5. doi:10.1186/2041-1480-6-5.

[13] B. Mons, FAIR science for social machines: Let's share metadata knowlets in the Internet of FAIR data and services, *Data Intelligence* **1** (2019), 22–42. doi:10.1162/dint_a_00002.

[14] Nanopub website. Available from: https://nanopub.org/wordpress/.

[15] Nanopublications working draft, 2021. Available from: https://nanopub.org/guidelines/working_draft/.

[16] R. Plug, Y. Liang, M. Basajja, A. Aktau, P.H. Jati, S. Amare et al., FAIR and GDPR compliant population health data generation, processing and analytics, 2021, pp. 54–63. Available from: http://ceur-ws.org/Vol-3127/paper-7.pdf.

[17] M. Reisen, F. Oladipo, M. Stokmans, M. Mpezamihgo, S. Folorunso, E. Schultes et al., Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research, *Advanced Genetics* **6** (2021), 2. doi:10.1002/ggn2.10050.

[18] E. Schultes, The FAIR hourglass: A framework for FAIR implementation. *FAIR Connect* **1**(1) (2023), 13–17. doi:10.3233/FC-221514.

[19] E. Schultes, B. Magagna, T. Kuhn, M. Suchánek, L.B. da Silva Santos and B. Mons, The comparative anatomy of nanopublications and FAIR digital objects, *Research Ideas and Outcomes* **10** (2022), 8. doi:10.3897/rio.8.e94150.

[20] E. Schultes, M. Roos, L.O.B. da Silva Santos, G. Guizzardi, J. Bouwman, T. Hankemeier et al., FAIR digital twins for data-intensive research, *Frontiers in Big Data* **5** (2022). doi:10.3389/fdata.2022.883341.

[21] E.A. Schultes, G.O. Strawn and B. Mons, Ready, set, GO FAIR: Accelerating convergence to an Internet of FAIR data and services, 2018. Available from: https://www.semanticscholar.org/paper/Ready%2C-Set%2C-GO-FAIR%3A-Accelerating-Convergence-to-an-Schultes-Strawn/9a63e61ae2b530fe60df8d71fc1d76f45d713974.

[22] K.D. Smedt, D. Koureas and P. Wittenburg, FAIR digital objects for science: From data pieces to actionable knowledge units, *Publications* **8** (2020), 21. doi:10.3390/publications8020021.

[23] M. van Reisen, Connecting health data across jurisdictions through FAIR data-visiting: Ownership, localisation and regulatory compliance (OLR), 2022.

[24] VODAN-A Nanopublication. Available from: http://server.nanopubs.lod.labs.vu.nl/RAdDKjIGPt_2mE9oJtB3YQX6wGGdCC8ZWpkxEIoHsxOjE.

[25] VODAN-Africa webpage. Available from: https://www.vodan-totafrica.info/index.php.

[26] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak et al., The FAIR guiding principles for scientific data management and stewardship, *Scientific Data* **3** (2016), 160018. doi:10.1038/sdata.2016.18.