How can we unambiguously refer to materials and their corresponding data?

Thomas E. Exner[1], Gerd Blanke[2], Jaleesia Amos[3], Mark Wiesner[3], Joh Dokler[1], Maja Brajnik[1], and Iseult Lynch[4]

1. Introduction

Nanosafety research has clearly demonstrated that chemical composition is not enough to represent a nanomaterial and to differentiate it from similar materials, which can show very different properties and toxicities. This has led to the concept of nanoforms and sets of nanoforms, defined by a limited number of physicochemical properties to group materials, which can be handled together e.g., under REACH [1]. For the relatively limited number of commercially relevant nanomaterials, such a classification might be enough, but this changes already if complex nanomaterials are considered. Moving from nano to advanced and innovative materials, this much larger group should be handled from the start in a common materials data ecosystem, and provide data to many emerging modelling applications as applied e.g., in Safe-and-Sustainable-by-Design. Materials undergo multiple transformation steps during their production, which lead to distributions (e.g., particle size distributions in nanomaterials, distributions of polymer chain lengths) or complex, random bonding networks (e.g., coatings synthesised by sol-gel processes, cross-linked polymers) rather than homogeneous entities.

2. Material, sample and data provenance

The FAIR (Findable, Accessible, Interoperable and Re-Usable) principles stress that data can only be understood and therefore re-used if metadata is provided documenting the full provenance trail of the data, describing all manipulations from its creation to its current state including collection, processing, curation and storing. In the same way, documentated material and sample provenance trails are often the only possibility to fully understand what the object under consideration is, how it is related to other, similar materials, why it shows differences to these and even to identify batch-to-batch variations. In this talk, we will present multiple new data management ideas to document such material and sample provenance trails and integrate these as part of data provenance. Instance Maps [2] are used to structure and visualise the material manipulations including production and other life-cycle stages (e.g., use and end-of-life scenarios), and sample preparation and data collection steps (see Figure 1). In combination with existing metadata schemas for data provenance like PROV [3],

[1] Seven Past Nine d.o.o., Cerknica, Slovenia, thomas.exner@sevenpastnine.com
[2] StructurePendium Technologies GmbH, Germany
[3] Center for the Environmental Implications of Nano Technology (CEINT), Civil & Environmental Engineering, Duke University, Durham, North Carolina, USA
[4] School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK

improved structural representations like the InChI for nano (NInChI) [4] and unique identifiers such as the European Registry of Materials (ERM) [5], a new material representation system can be developed, which provides chemical information as precisely as possible but acknowledges the production-specific aspects of materials properties, providing this information for the current life-cycle stage and any past stages and transformations. Documentation in this way provides (i) a description of the real material or specific sample of it (if sample provenance is added) and the collection (distribution) of specific particles or substructure of the material; and (ii)enables all this information to be easily used to create input for data-driven or physics-based modelling, which need to use a specific structure or a small subset of these.

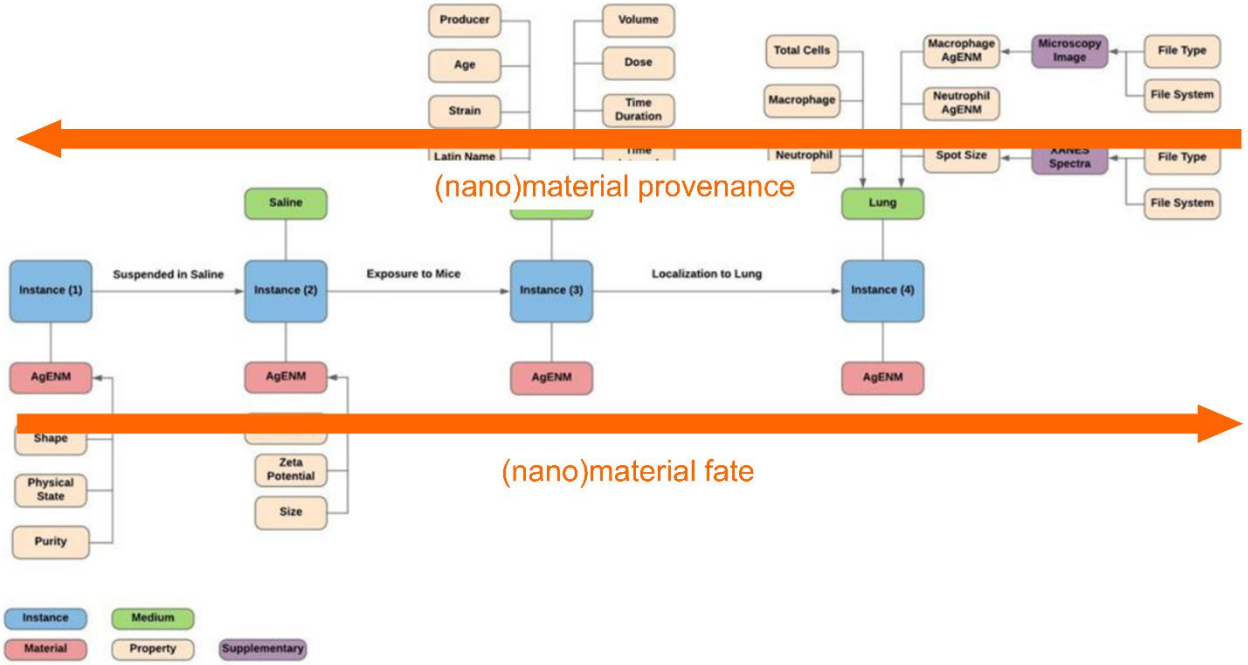

Figure 1: Instance map visualizing the fate of a nanomaterial in a mesocosm experiment looking at effects on mice lungs. Following the instance map backwards, the material provenance can be extracted giving a detailed description of the material and its previous life cycle stages (production and formulation for example).

3. Conclusions

Current identifiers cannot capture the complexity of materials research, leading to cases where (I) different materials or batches of materials lead to different experimental outcomes yet are referred to by the same identifiers (like CAS registry numbers) or (ii) different materials are shown to have the same basic physicochemical properties leading to misclassification as the same material. Thus, current materials characterisation as provided by many studies (and stored in databases) are not enough to prove sameness or similarity. To remove this ambiguity, we suggest using a life-cycle approach, which includes the history of previous stages as material and sample provenance trails. However, such a change of information management will require further developments for tools re-using the data starting with advanced non-structural similarity measures and input generation for modelling.

4. References

(1) European Chemicals Agency. *Appendix for Nanoforms to the Guidance on Registration and the Guidance on Substance Identification: Version 2.0 January 2022.*; Publications Office: LU, 2022

(2) Amos, J. D.; Zhang, Z.; Tian, Y.; Lowry, G. V.; Wiesner, M. R.; Hendren, C. O. *Sci Data* **2024**, *11*, 173. doi:10.1038/s41597-024-03006-8

(3) PROV Model Primer https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/ (accessed Apr 8, 2024)

(4) Lynch, I.; Afantitis, A.; Exner, T.; Himly, M.; Lobaskin, V.; Doganis, P.; Maier, D.; Sanabria, N.; Papadiamantis, A. G.; Rybinska-Fryca, A.; Gromelski, M.; Puzyn, T.; Willighagen, E.; Johnston, B. D.; Gulumian, M.; Matzke, M.; Green Etxabe, A.; Bossa, N.; Serra, A.; Liampa, I.; Harper, S.; Tämm, K.; Jensen, A. C.; Kohonen, P.; Slater, L.; Tsoumanis, A.; Greco, D.; Winkler, D. A.; Sarimveis, H.; Melagraki, G. *Nanomaterials* **2020**, *10*, 2493. doi:10.3390/nano10122493

(5) van Rijn, J.; Afantitis, A.; Culha, M.; Dusinska, M.; Exner, T. E.; Jeliazkova, N.; Longhin, E. M.; Lynch, I.; Melagraki, G.; Nymark, P.; Papadiamantis, A. G.; Winkler, D. A.; Yilmaz, H.; Willighagen, E. *Journal of Cheminformatics* **2022**, *14*, 57. doi:10.1186/s13321-022-00614-7

5. Acknowledgement

WorldFAIR (grant agreement No. 101058393), PINK (grant agreement No. 101137809), MACRAMÉ (grant agreement No. 101092686), and BIO-SUSHY (grant agreement No. 101091464) have received funding from the European Union's Horizon Europe Research and Innovation programme and for the UK Research and Innovation (UKRI) Horizon Europe Guarantee fund (Grant numbers 183197718 and 10097944, respectively).