NanoCommons

Nano-Knowledge Community

The European Nanotechnology Community Informatics Platform: Bridging data and disciplinary gaps for industry and regulators

This project has received funding from the European Union Horizon 2020 Programme (H2020) under grant agreement no. 731032

NanoCommons

Nano-Knowledge Community

When is a metadata set complete?

Thomas Exner, Edelweiss Connect, Switzerland

This project has received funding from the European Union Horizon 2020 Programme (H2020) under grant agreement no. 731032

Your most favoured piece of metadata

Go to www.menti.com and use the code 89 55 42 8

Relevant metadata

Mentimeter

ninchi

1

Completeness

NanoCommons
Nano-Knowledge Community

3.1 Data completeness key concept (reproduced and adapted from [3]):

The completeness of data and associated metadata may be considered a measure of the availability of the necessary, non-redundant (meta)data for a given entity. However, there is no definitive consensus regarding exactly how data completeness should be defined in the nanoscience, or wider scientific, community. Indeed, metadata availability may be considered an issue distinct from data completeness.

3.2 Data quality key concept (reproduced and adapted from [3]):

Data quality may be considered a measure of the potential usefulness, clarity, correctness and trustworthiness of data and datasets. However, there is no definitive consensus regarding exactly how data quality should be defined in the nanoscience, or wider scientific, community. Data quality may be considered dependent upon the degree to which the meaning of the data is "clear" and the extent to which the data are "plausible". In turn, this may be considered to incorporate (aspects of) data completeness. For example, data quality may be considered to be (partly) dependent upon the "reproducibility" of data and the extent to which data are reproducible and their reproducibility can be assessed.

3. Robinson, R.L.M., Lynch, I., Peijnenburg, W., Rumble, J., Klaessig, F., Marquardt, C., Rauscher, H., Puzyn, T., Purian, R., Åberg, C. and Karcher, S., 2016. How should the completeness and quality of curated nanomaterial data be evaluated? Nanoscale, 8(19), pp.9919-9943.

FAIRness

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

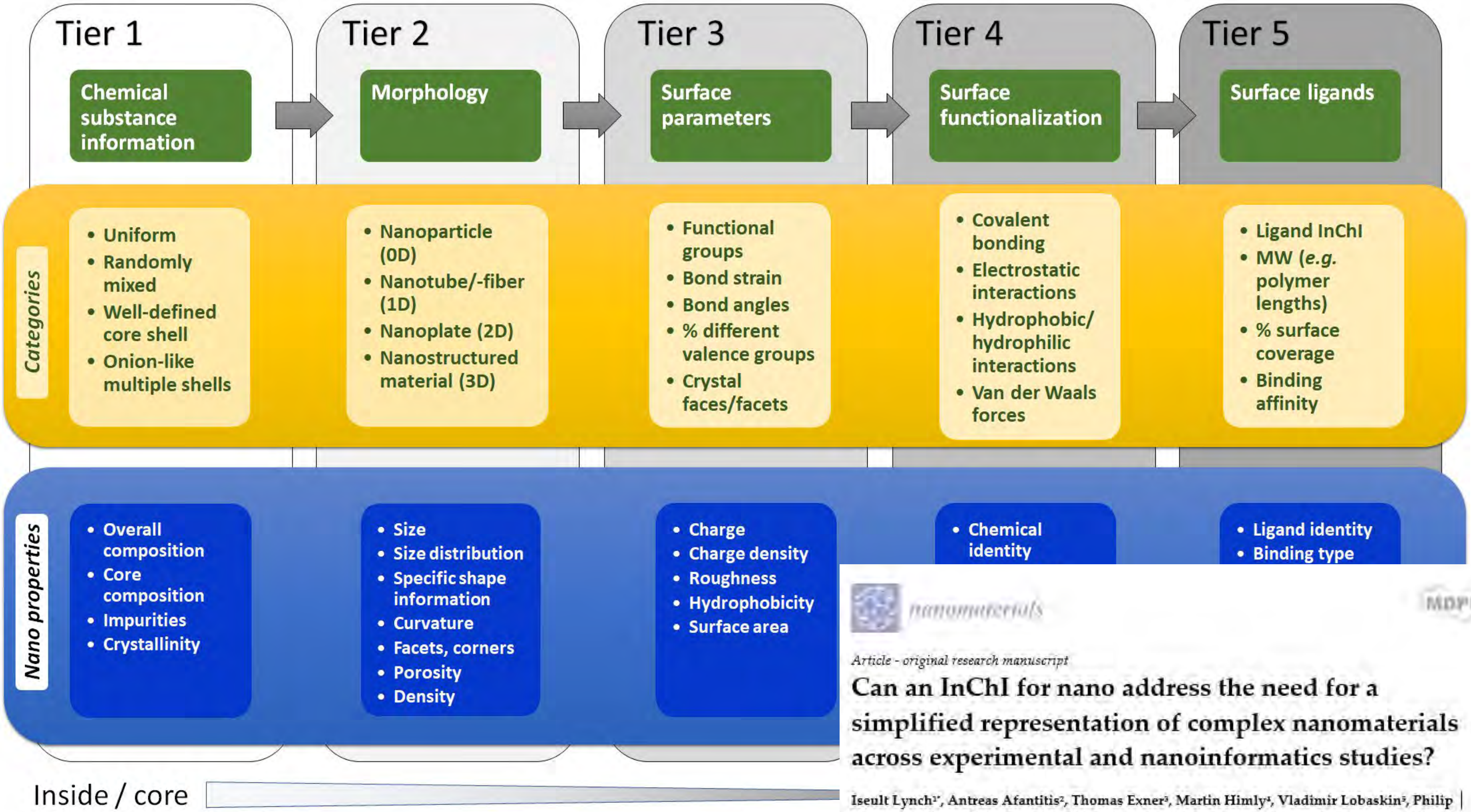
FAIRness

NanoCommons
Nano-Knowledge Community

- To be **Reusable**:
  - SR1: Do not limit the reported metadata to fulfil only the requirements of the study for which the data was produced. Sections 6 and 7 provide examples on the re-usage of data in a different computational context than the experimental data producer initially intended.
  - SR2: Establish a feedback loop between data creators, analysts and customers to continuously improve the metadata completeness and quality. Keep in mind that scientific progress can lead to new use cases and go beyond "standards" defined at a specific point of time.

- To be **Interoperable**:
  - SI1: Provide direct links to descriptions of the test methods, protocols and quality control measures to give the user the chance to evaluate data interoperability. In this way, additional information, which cannot all be covered by the metadata can be easily accessed.
  - SI2: Report protocol metadata in a structured and annotated way to allow harmonisation and interlinking of data. While duplication of information in the protocol and the metadata is sometimes needed or even preferred, guarantee consistency between both.

Nanomaterial identifier - InChI for nano



nanomaterials MDPI

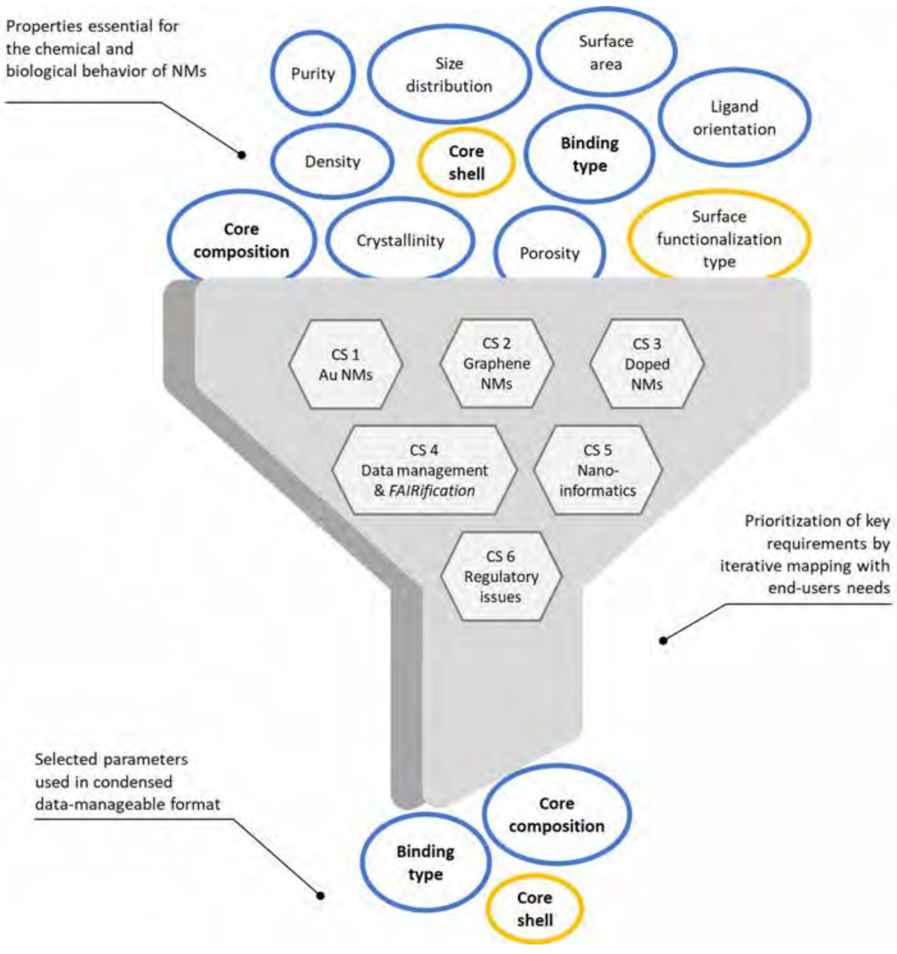
Article - original research manuscript

Can an InChI for nano address the need for a simplified representation of complex nanomaterials across experimental and nanoinformatics studies?

Iseult Lynch[1*], Antreas Afantitis[2], Thomas Exner[3], Martin Himly[4], Vladimir Lobaskin[5], Philip Doganis[6], Dieter Maier[7], Natasha Sanabria[8], Anastasios G. Papadiamantis[1], Anna Rybinska-Fryca[9], Maciej Gromelski[9], Egon Willighagen[10], Blair D. Johnston[11], Mary Gulumian[8], Marianne Matzke[12], Amaia Green Etxabe[12], Nathan Bossa[13], Angela Serra[14], Irene Liampa[6], Stacey Harper[15], Kaido Tämm[16], Alexander CØ Jensen[17], Pekka Kohonen[18], Luke Slater[1], Haralambos Sarimveis[6], Georgia Melagraki[2*]

Case studies

1. Library of Au NMs of different sizes and surface functionalities (ligands)
2. Library of carbon nanotubes
3. Complex chemistries and structures
4. NM-related data management and incorporation into the FAIR data landscape
5. NInChI in Nanoinformatics
6. Regulatory challenges



InChI for nanomaterials

Category 1: must have	Category 2a: nice to have	Category 2b: extrinsic properties	Category 3: out of scope
• Chemical composition • Size / size distribution • Shape • Crystal structure • Chirality • Ligand and ligand binding	• Structural defects • Density • Surface composition	• Surface charge • Corona • Agglomeration state • Dispersion	• Optical properties • Magnetic properties • Chemical state / oxidation state

CTAB-capped-gold nanoparticles, diameter=20 nm:

NInChI=1A

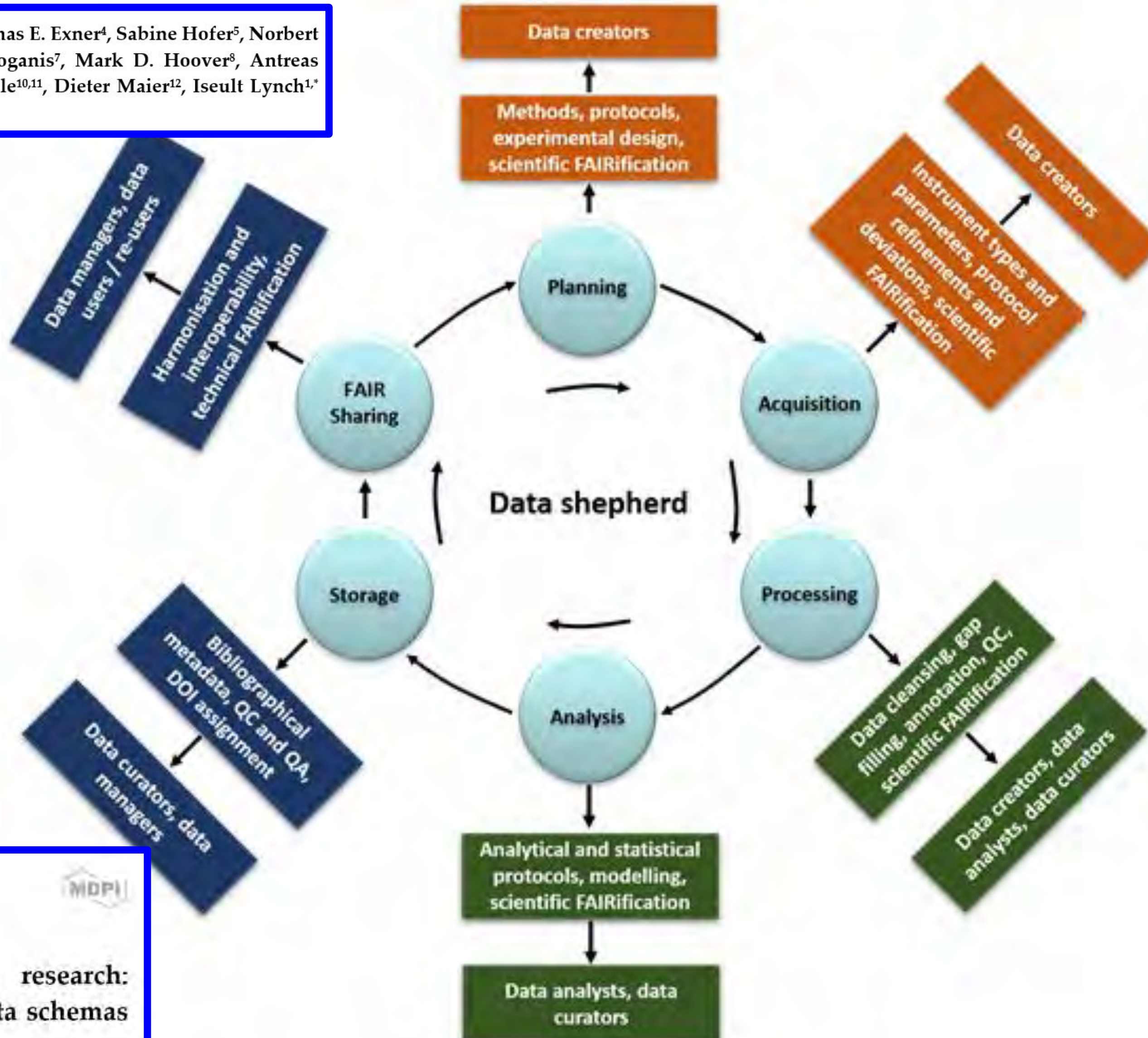
/Au**/msp/s20d-9**

!C19H42N.BrH/c1-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20(2,3)4;/h5-19H2,1-4H3;1H/q+1;/p-1

/y1&2

Data management, data lifecycle & metadata

Anastasios G. Papadiamantis[1,2,*], Frederick C. Klaessig[3], Thomas E. Exner[4], Sabine Hofer[5], Norbert Hofstaetter[5], Martin Himly[5], Marc A. Williams[6], Philip Doganis[7], Mark D. Hoover[8], Antreas Afantitis[2], Georgia Melagraki[2], Tracy S. Nolan[9], John Rumble[10,11], Dieter Maier[12], Iseult Lynch[1,*] and members of the NCI nanoWG



nanomaterials MDPI

Article

Metadata stewardship in nanosafety research: community-driven organisation of metadata schemas to support FAIR nanoscience data

Responsibilities

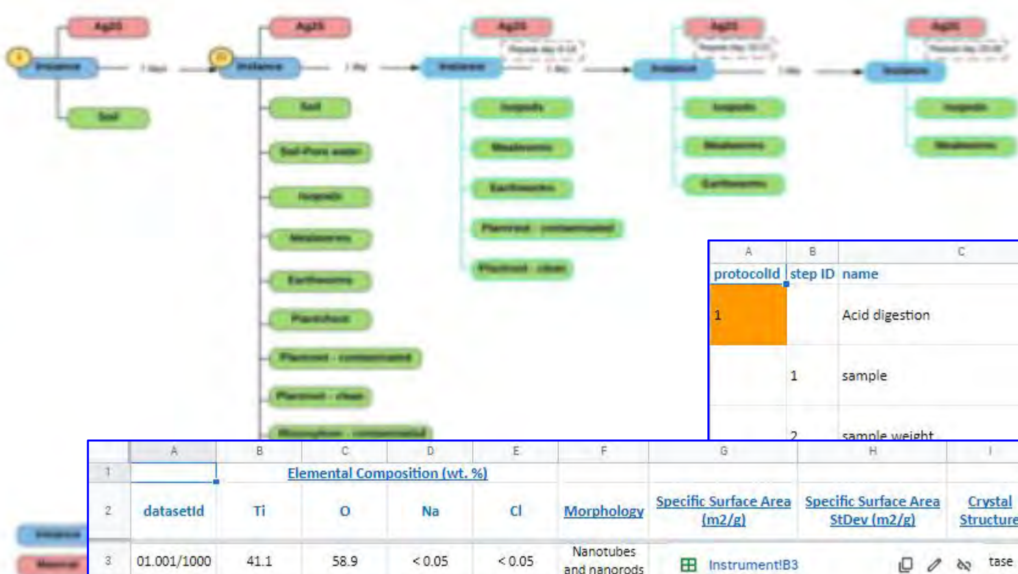
Table 1. Data roles, responsibilities and interactions. Adapted from Hoover et al. [30,41] and Woodall et

	Set objectives	Design Approach	Collect	Processing	Modelling/Analysis	Validate	Store	Sh
Creators	X	X	X	X		X		
Analysts		X		X	X	X		
Curators				X		X		
Managers							X	X
Customers	X							
Shepherds	X	X	X	X	X	X	X	X

Metadata templates - the NanoFase case

Instance Map

Terrestrial Mesocosm Experiment



protocolId	step ID	name	referencingId	durationValue/Unit	durationUnits	description	reference
1		Acid digestion				for fly ash samples	
	1	sample	1			dry powder	
	2	sample weight	1			20 mg	
						6 mL	
						1 mL	
						0.1 mL	

Protocol · Instrument · dictionary

	Elemental Composition (wt. %)								
datasetId	Ti	O	Na	Cl	Morphology	Specific Surface Area (m2/g)	Specific Surface Area StDev (m2/g)	Crystal Structure	TiO2 nanotubes concentration (ug/mL)
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods	Instrument!B3 BET		tase	0.00
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods			tase	6.25
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods	126.9	0.6	Anatase	12.50
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods	126.9	0.6	Anatase	25.00
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods	126.9	0.6	Anatase	50.00
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods	126.9	0.6	Anatase	100.00
01.001/1000	41.1	58.9	< 0.05	< 0.05	Nanotubes and nanorods	126.9	0.6	Anatase	200.00

Publication · Institution · People · Measurement · Protocol · Instrument · dictionary

NanoInformatics:

NanoCommons
Nano-Knowledge Community

MODA
Elements in materials modelling
NANODOME

THE SIMULATION			
1	USER CASE	Nanoparticle synthesis via gas phase condensation in industrial commercially-relevant processes. Prediction of the nanoparticle size distribution, morphology and internal composition via modelling of the gas phase condensation synthesis process, including homogeneous and heterogeneous nucleation, surface and internal chemical kinetics and composition, agglomeration, aggregation. Materials: Si, ZnO, Al_2O_3, Pt nanoparticles in $Ar/H_2/N_2/O_2$ atmospheres for synthesis processes in plasma, hot wall and flame reactors.	
2	CHAIN OF MODELS	MODEL 1	Electronic Density Functional Theory (Electronic)
		MODEL 2	Classical MD (Atomistic)
		MODEL 3	Coarse Grained Molecular Dynamics (Mesoscopic)
		MODEL 4	Fluid mechanics, Heat-Flow, Chemistry Reaction M Electromagnetism (Continuum)
3	PUBLICATION	N.A.	
4	ACCESS CONDITIONS	Electronic and Atomistic models are based on widely available or open-source licenses packages, such as Quantum ESPRESS ReaxFF, GROMACS, GARFFIELD. The mesoscopic model will be developed within the NanoDome under open-source license. Continuum models are based on the commercial package ANSY and on the open-source package OpenFOAM. Interfacing libraries and the material database for Si/Ar system open-source. Material database for reactive materials will be p under commercial license.	

Jaqpot

MODEL
Title: Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Description:
Linear Model for Predicting Solubility of C60 Fullerenes in Various Solvents

Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

The model is provided in the following publication: Farhad Gharagheizi & Reza Fareghi Alamdari (2008) A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents, Fullerenes, Nanotubes, and Carbon Nonstructures, 16:1, 40-57, DOI: 10.1080/15363830701779315

Full dataset is available in this link

Training dataset is available in this link

Test dataset is available in this link

A downloadable QMRF Report is available in this link

QMRF Report

1. QSAR Identifier

1.1.QSAR identifier (title):

Linear nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

The model has been presented in the publication "A Molecular-Based Model for Prediction of Solubility of C60 Fullerene in Various Solvents" Farhad Gharagheizi &Reza Fareghi Alamdari, Fullerenes, Nanotubes and Carbon Nanostructures, Volume 16, 2008 - Issue 1

1.2.Other related models:

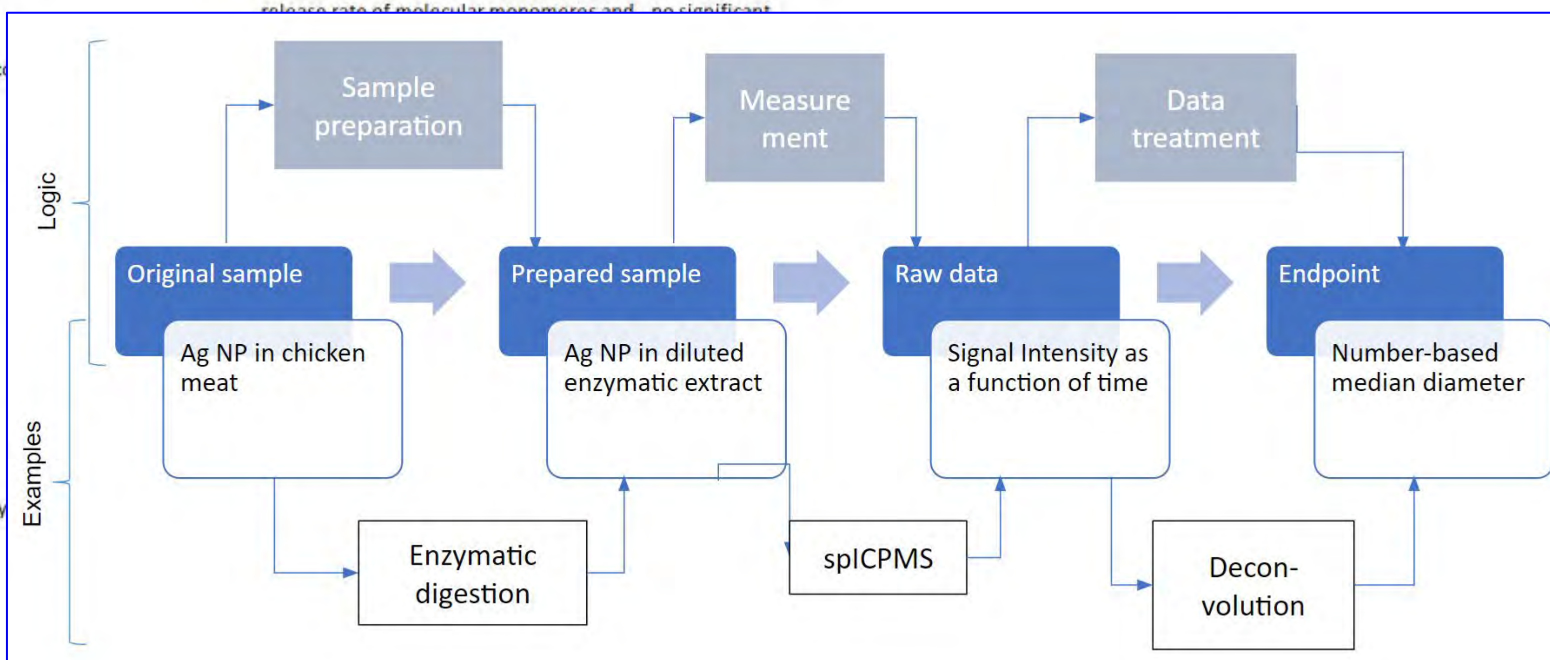
Neural Network nanoQSAR model predicting Solubility of C60 Fullerene in Various Solvents.

1.3.Software coding the model:

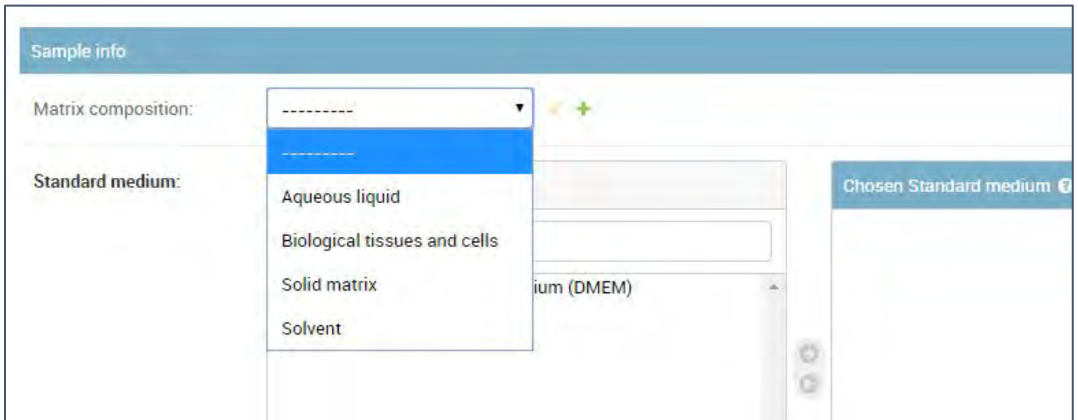
Jaqpot is a web platform that support development, validation and sharing of QSAR models apps.jaqpot.org

Agglomeration-related metadata and metadata questionnaires

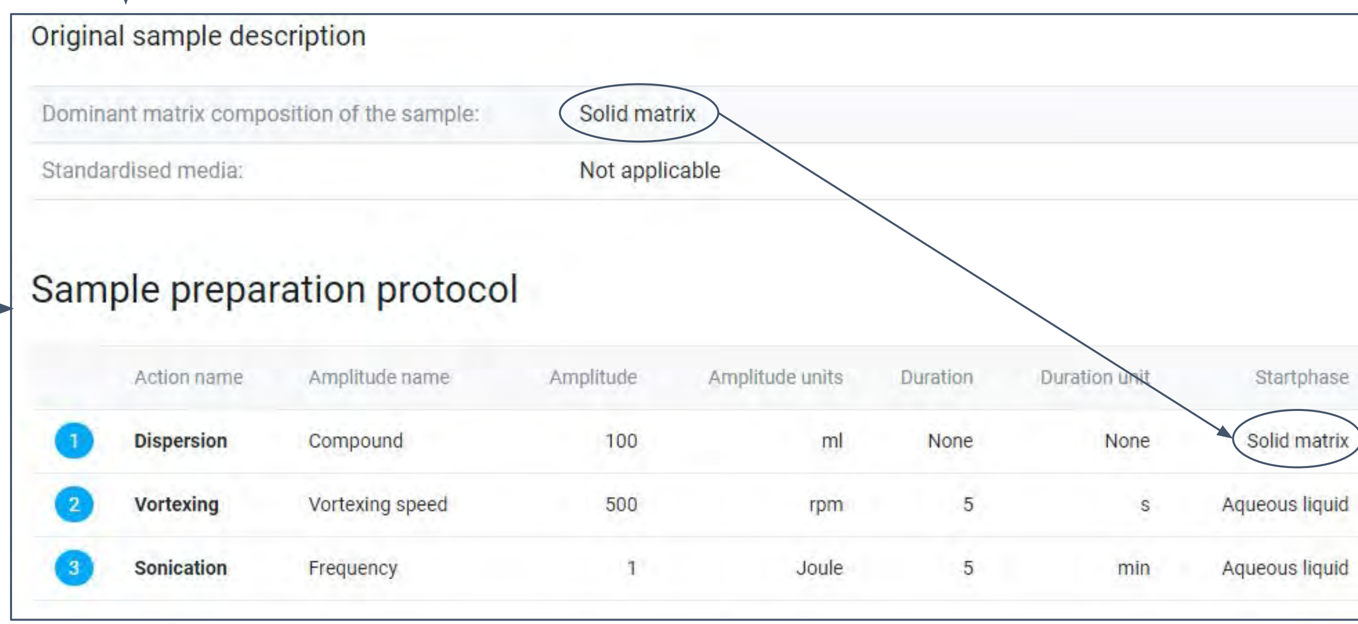
data object	NP descriptors	(meta)data	remarks / description	case study value
primary particle	size	diameter	diameter of primary particle	50.0 +/- 0 nm
		determination method	DLS, NTA, TEM, SEM, ...	NTA
		statistical measure	mean, mode, median, ...	mean + stdv
		size qualifier	hydrodynamic diameter, dried, ...	hydrodynamic diameter
	shape		shape of particle (spherical, rod, ...)	spherical
	aspect ratio		ratio of sizes in different dimensions	1
	density		density of primary particle	SiO_2: 2.2 g/cm^3, TiO_2: 4.24 g/cm^3
	surface charge		zeta potential of primary particle	-34 mV *)
	porosity		pore volume fraction	non porous
	polydispersity		polydispersity index, size distribution	monodisperse
	dissolution rate		release rate of molecular monomeres and	no significant
	synthesis protoco			
agglomerate	size			
	shape			
	aspect ratio			
	packing			
	effective density			
	polydispersity			
	stability			
experiment	method			
	model		for *in silico* only: which method is	*in vitro*, submerged,



ACEnano web interface



- Capture details on chemistry of the original sample and the sample after preparation
- Definition and example of standard medium
- Some actions are complex and could be described separately (e.g. dispersion protocol including detailed steps and equipment used)

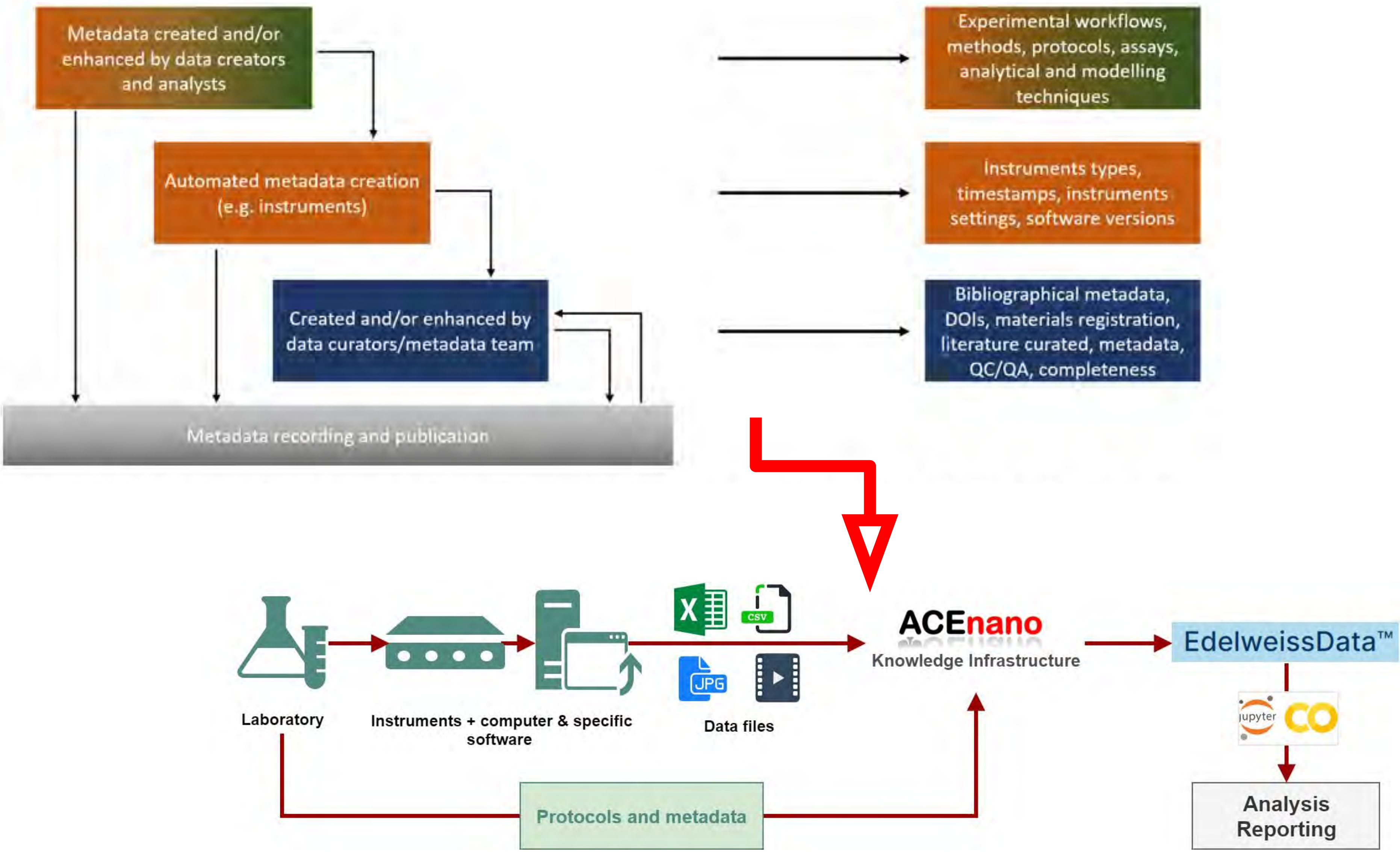

Original sample description

Dominant matrix composition of the sample:	Solid matrix
Standardised media:	Not applicable

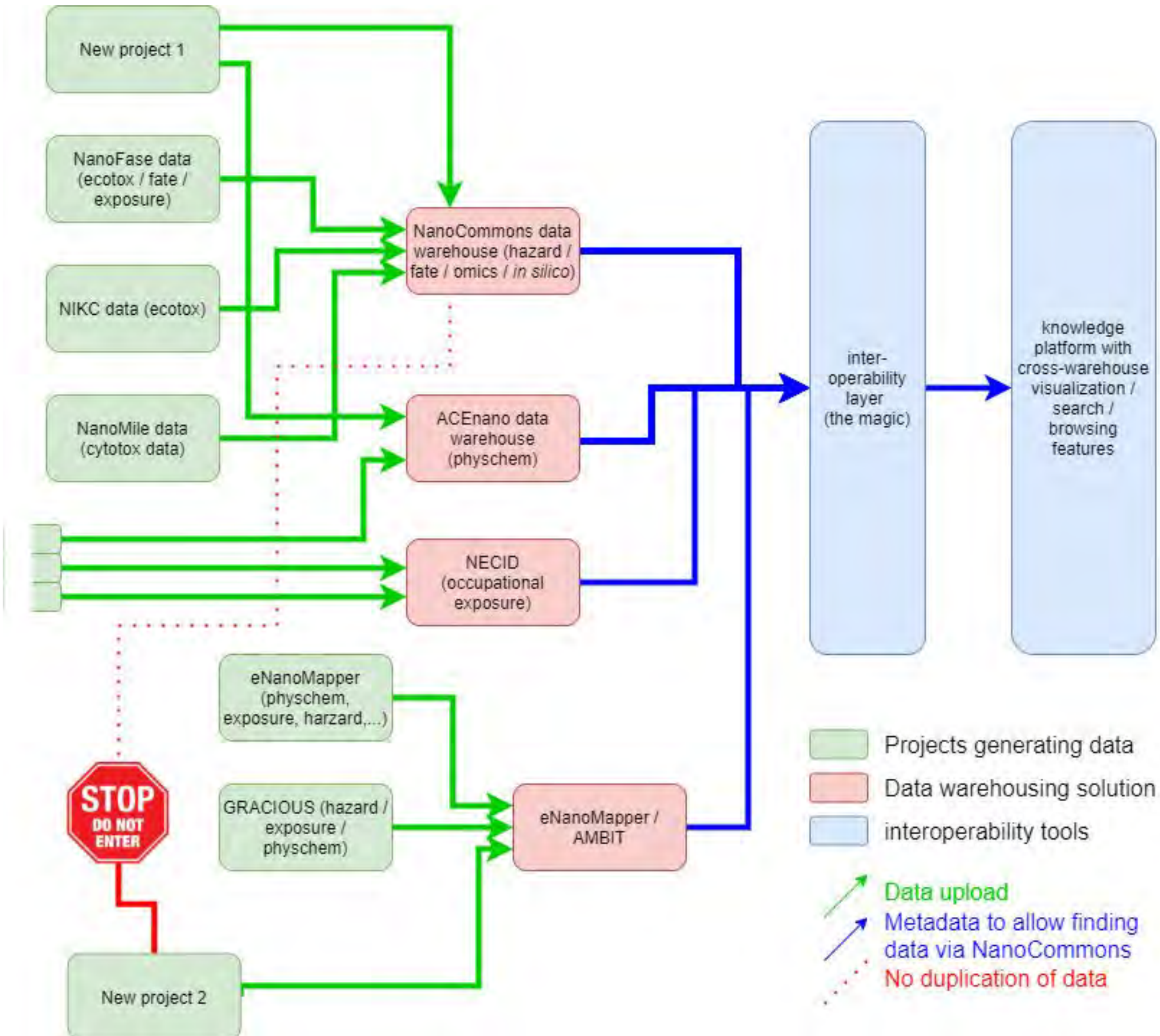
Sample preparation protocol

	Action name	Amplitude name	Amplitude	Amplitude units	Duration	Duration unit	Startphase
1	**Dispersion**	Compound	100	ml	None	None	Solid matrix
2	**Vortexing**	Vortexing speed	500	rpm	5	s	Aqueous liquid
3	**Sonication**	Frequency	1	Joule	5	min	Aqueous liquid

Upload workflows



And don't forget about semantics



The ontology development universe

NanoCommons
Nano-Knowledge Community

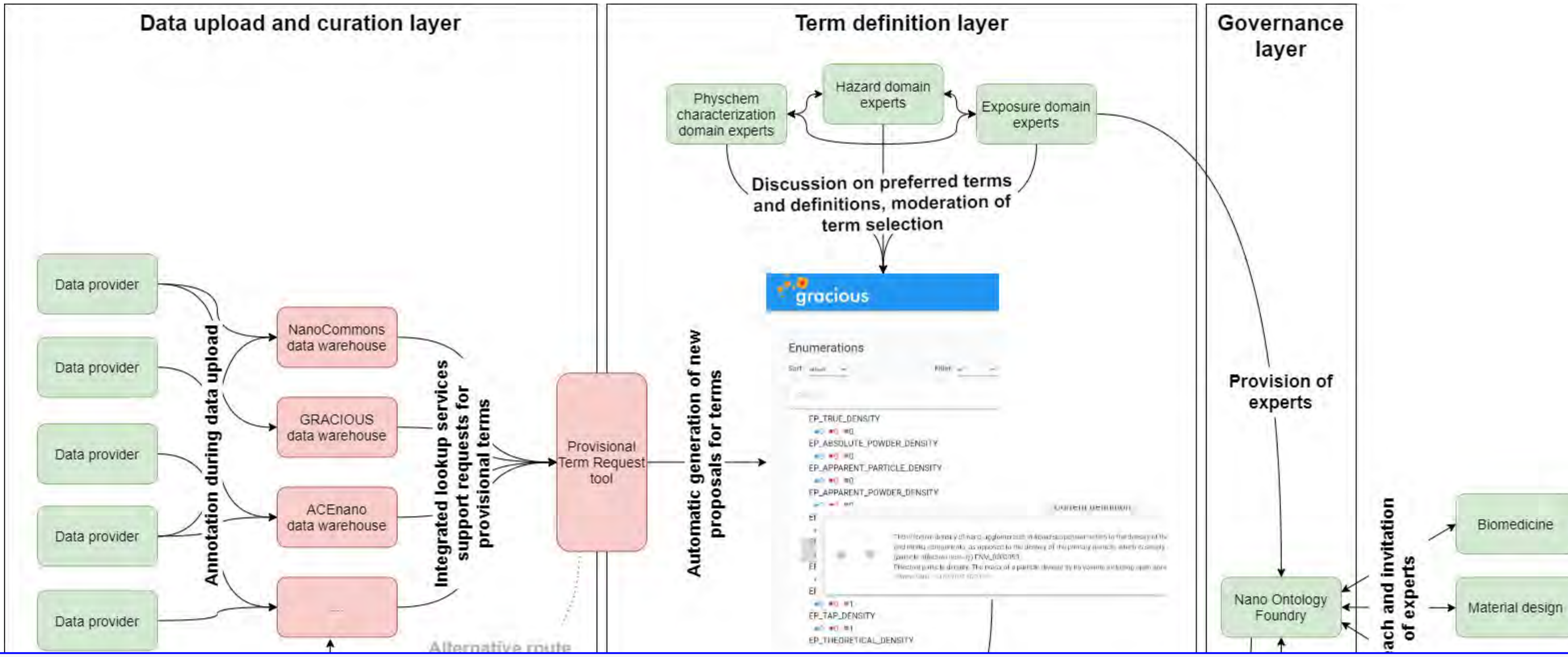


Table 2. Definition of the term "instance" for data creators, analysts, curators, managers, and customers.

Role	Example Term	Definition
Data creator	Experimental instance	A specific part of an assay or method
Data analyst	Training and test instances	A set of specific data entries used for training, testing and validating a predictive model
Data curator	NIKC instance	The reported nanomaterial in a system at a specific moment in time
Data manager	Database instance	A set of the background processes and memory structure needed by the database software to access the data
Data customer	All of the above depending on the specific use case	

When is a metadata set complete?

Never!?

But at least (minimal) standards should be defined and respected to cover most applications.

And these should improve over time.

Also based on new technology acceptable by all users.

Input from all sides are needed to achieve this!

Table 1. Data roles, responsibilities and interactions. Adapted from Hoover et al. [30,41] and Woodall et al. [42]. Adapted wi

	Set objectives	Design Approach	Collect	Processing	Modelling/Analysis	Validate	Store	Share	Quality Control	Ann	unication ducation
Creators	X	X	X	X		X			X		X
Analysts		X		X	X	X			X		X
Curators				X		X			X		X
Managers							X	X	X		X
Customers	X								X		X
Shepherds	X	X	X	X	X	X	X	X	X		X

Whatever your role is,

in the NSC WG F!

https://www.nanocommons.eu/ta-access/
https://ssl.biomax.de/nanocommons/cgi/login_bioxm_portal.cgi

Your most favoured piece of metadata

NanoCommons
Nano-Knowledge Community

Go to www.menti.com and use the code 89 55 42 8

Relevant metadata

Mentimeter

ninchi

1

Thank you *for your attention!*

NanoCommons

Nano-Knowledge Community

Iseult Lynch
Tassos Papadiamantis
Dieter Maier
Egon Willighagen
Martin Himly
Danail Hristozov
Alex Zabeo
and many more

Joh Dokler
Lucian Farcal
Maja Brajnik
Barry Hardy

Whatever your role is,

in the NSC WG F!

NSC Education Day @ NANOSAFE 2020,
16 November 2020