



Cite this: *Nanoscale*, 2020, **12**, 4695

# Quality of physicochemical data on nanomaterials: an assessment of data completeness and variability†

Daniele Comandella,<sup>ib</sup> Stefania Gottardo, Iria Maria Rio-Echevarria‡ and Hubert Rauscher<sup>ib</sup>\*

Grouping and read-across has emerged as a reliable approach to generate safety-related data on nanomaterials (NMs). However, its successful implementation relies on the availability of detailed characterisation of NM physicochemical properties, which allows the definition of groups based on read-across similarity. To this end, this study assessed the availability and completeness of existing (meta)data on 11 experimentally determined physicochemical properties and 18 NMs. Data on representative NMs were mainly extracted from existing datasets stored in the eNanoMapper database, now available on the European Observatory on Nanomaterials website, while data on case-study NMs were provided by their industrial manufacturers. The extent of available (meta)data was assessed and data gaps were identified, thereby determining future testing needs. Data completeness was assessed by using the information checklists included in the templates for data logging developed by the EU-funded projects NANoREG and GRACIOUS. A completeness score (CS) between 0 and 1 was calculated for each (meta)data unit, template section, property, technique and NM. The results show a heterogeneous distribution of available (meta) data across materials and properties, with none of the selected NMs fully characterised. The average CS calculated for representative NMs (0.43) was considerably lower than for case-study NMs (0.68). The low CS was largely caused by missing information on sample preparation and standard operating procedures, and was attributed to a lack of harmonised data reporting and entry procedure. This study therefore suggests that a persistent use of well-defined and harmonised reporting schemes for experimental results is a useful tool to increase (meta)data completeness and ensure their integration and reuse.

Received 27th September 2019,  
Accepted 19th December 2019

DOI: 10.1039/c9nr08323e

[rsc.li/nanoscale](http://rsc.li/nanoscale)

## 1. Introduction

Nanotechnologies represent one of the four Key Enabling Technologies (KETs) (along with advanced materials, advanced manufacturing and processing and biotechnology) that have received funding from Horizon 2020, the European Union (EU) Research and Innovation Programme started in 2014.<sup>1</sup> Nanomaterials (NMs) have so far shown beneficial applications in various industrial processes and consumer products, ranging from food, packaging, cosmetics, and electronics to pharmaceuticals and medical devices. Innovation in nanotechnology is expected to bring even more new opportunities for economic growth and life improvement.<sup>2–5</sup>

At the same time, nanomaterials have been the subject of numerous scientific studies aimed to investigate their interactions with biological and environmental systems as well as develop experimental and predictive tools for characterising their properties and potential adverse effects on human health and ecological receptors.<sup>6,7</sup> The European Commission has contributed by funding research projects in this area since 2005. The resulting scientific advancements have triggered changes in the European legislation relevant for nanomaterials<sup>8,9</sup> and, most recently, in the Annexes to the REACH (Registration, Evaluation, Authorisation of Chemicals) Regulation,<sup>10</sup> which now contain a legal definition of the term ‘nanoform’ and specific information requirements that need to be fulfilled by companies when registering nanoforms of substances in a dossier. International bodies such as the Organisation for Economic Co-operation and Development Working Party on Manufactured Nanomaterials (OECD WPMN), the International Organization for Standardization (ISO) and the European Committee for Standardization (CEN) have followed up by publishing a series of test guidelines, gui-

European Commission, Joint Research Centre (JRC), 21027 Ispra, VA, Italy.

E-mail: [hubert.rauscher@ec.europa.eu](mailto:hubert.rauscher@ec.europa.eu)

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c9nr08323e

‡Present address: European Commission, Research Executive Agency, 1049 Brussels, Belgium.



dance documents and documentary standards that can be used to generate nanomaterial information with regulatory relevance.<sup>11</sup>

Despite these efforts, not all questions regarding nanomaterial safety have been fully answered.<sup>12</sup> Most of the available tools (exposure models, experimental protocols) have not undergone or accomplished a formal validation process yet and their data output has low potential for being accepted by regulators.<sup>7</sup> As a variety of new NMs and new applications of existing NMs are expected to be commercialised in the near-future, a comprehensive set of information demonstrating their safe use will need to be submitted to European authorities to access the market. Hazard data on a chemical substance is usually generated *via* animal testing, a combination of *in vitro* and *in vivo* methods or, if feasible and scientifically justified, using *in vitro* techniques and *in silico* models. Often, hazard data on a chemical substance is extrapolated *via* read-across from existing data on other chemicals that are considered to be sufficiently similar. The legislation in general requires that animal testing is minimised,<sup>13,14</sup> and scientists have recently called for a shift towards further reduction of animal testing in nanosafety.<sup>15</sup> The use of grouping and read-across, allowing data on a certain endpoint to be shared between nanomaterials whose physicochemical similarity is scientifically proven,<sup>16</sup> has been explored and specific approaches have been proposed in the literature.<sup>17,18</sup> In 2018, the GRACIOUS project<sup>19</sup> was launched with the aim of developing a comprehensive, science-based framework for grouping and read-across of nanomaterials, intended to facilitate the acquisition of data on nanomaterials for both regulatory purposes and safe-by-design.<sup>20</sup> The European Chemicals Agency (ECHA) has also contributed to the topic by publishing a guidance on how to apply the basic principles of grouping and read-across to nanoforms of the same substance in a REACH registration dossier.<sup>21</sup>

This intense scientific and regulatory activity has led to the generation of a large amount and variety of experimental data and information on NMs. Most of it can be manually retrieved from tables and graphs in articles published in peer-reviewed journals. In some cases, individual project-specific datasets were created; however, this was often done according to different reporting formats and without granting public access, thus preventing subsequent integration into a central repository and re-use of the data by other scientists or regulators for different purposes. Proper data management has thus become crucial for producing new knowledge on nanomaterials,<sup>22</sup> for example through computational modelling. Several initiatives have been taken in recent years to facilitate the move from field-specific datasets with limited access to a harmonised integrated infrastructure for nanomaterial data. In the EU, the eNanoMapper project<sup>23</sup> developed a computational infrastructure to import, store and share experimental data or calculated descriptors on physicochemical and (eco) toxicological properties of nanomaterials.<sup>24</sup> The system was also designed to record the details of how the data were generated (experimental conditions, methods) but it does not

include specific metrics to evaluate the quality of the data provided by the users.<sup>24</sup> Part of the data hosted in the eNanoMapper database is now accessible from the European Observatory on Nanomaterials (EUON)<sup>25</sup> and allows scientists and the general public to search and download data generated in specific EU-funded projects.

In 2011, the Nanoinformatics 2020 Roadmap<sup>26</sup> included (i) minimal information standards for data completeness and quality evaluation, (ii) proper data annotation and attribution, and (iii) standardised physicochemical characterisation of nanomaterials amongst the key priorities for sharing and integrating datasets. In the USA, the Nanomaterial Data Curation Initiative (NDCI) explored the critical aspect of data curation within the development of informatics approaches and concluded that establishing a minimum set of information to be reported and a standardised scheme to characterise both quality and completeness of the available datasets is fundamental to encourage integration in databases.<sup>22,27</sup> The recent Nanoinformatics 2030 Roadmap<sup>28</sup> recommends the use of a common data entry and transfer formats based on minimum information checklists, for example ISA-TAB-Nano,<sup>29</sup> and mentions the templates for data logging developed in the NANoREG project.<sup>30,31</sup> Moreover, the Roadmap endorses the FAIR (Findability, Accessibility, Interoperability, and Reusability) guiding principles for scientific data management and stewardship (including, collection, annotation, archival and long term care), which represent a minimal set of guiding principles and practices to more easily discover, access, integrate, re-use and cite the vast quantity of information being generated by contemporary data-intensive science.<sup>32</sup> One of the FAIR guiding principles explicitly states that (meta)data should be “richly described with a plurality of accurate and relevant attributes”, “associated with detailed provenance” and “meet domain-relevant community standards”.<sup>32</sup> Despite consistent investments and the large number of laboratories involved in the field of nanosafety, the size of the existing databases remains relatively small.<sup>33</sup> Tropsha and co-authors attributed the current situation to inefficient data collection and sharing caused by several factors, the most relevant being the time and effort required to report in a standardised template the detailed description of experimental conditions and to conduct the curation process, and the lack of a public database to which laboratory data on nanomaterials can be uploaded in an electronic format instead of being published in a journal format. As a result of this analysis, there has been a call on (i) researchers to further develop and accept minimal characterisation standards, (ii) data curators to evaluate data completeness and compliance, (iii) funding agencies to require data-upload to public databases as a condition for the award, and (iv) journals to require data-deposition into public databases as a prerequisite for accepting papers for publication.<sup>33</sup>

In light of the observations exposed above, this paper describes a procedure for data collection and completeness evaluation concerning 11 experimentally measured physicochemical properties of 18 NMs selected in the GRACIOUS



project.<sup>19</sup> The project included representative test materials (carbon nanotubes, silver, barium sulphate, cerium dioxide and silica) and industrial materials as case-studies (unmodified and modified silica, uncoated and coated diketopyrrolopyrrole, Cu-phthalocyanine, hematite, cellulose nanofibrils and nanocrystals). Most of the data on the representative test materials was retrieved from existing datasets generated in the previous European research projects NanoTest,<sup>34</sup> MARINA<sup>35</sup> and NANoREG,<sup>31</sup> and made available *via* a database generated by the eNanoMapper project.<sup>24</sup> Data on industrial materials were provided by GRACIOUS project partners. The presented study aimed to evaluate the completeness of the collected (meta)data using as checklists the information items on sample preparation, method, instrument, results and uncertainty required by the templates for data logging developed in NANoREG<sup>31</sup> and then extended in GRACIOUS.<sup>36</sup> A completeness score between 0 and 1 was calculated for each data unit; individual scores were then averaged to obtain an overall assessment for each template section, property, technique and nanomaterial. The whole procedure was inspired by the compliance level calculated in the Nanomaterial Registry.<sup>37,38</sup> This paper reports and discusses both data collection and completeness evaluation. Variations in collected property values were also discussed through selected examples. GRACIOUS partners will use the results of this study to identify gaps and weaknesses in existing data and define the additional experimental characterisation of nanomaterial physicochemical properties within the project. The final dataset (including existing and newly generated data) will be used to test the performance of the GRACIOUS framework for grouping and read-across of nanoforms, currently under development.<sup>20</sup> The final dataset, curated and properly reported in the eNanoMapper database, will also be easier to re-use in later scientific studies as well as for regulatory purposes.

By means of practical examples, this study will show the need of assessing quality and completeness before archiving

data and making them accessible for future re-use, and proposes a transparent procedure on how to conduct this assessment using harmonised tools.

## 2. Methods

### 2.1 Materials and data selection

**Materials.** Two sets of manufactured NMs were considered in this study (Table 1). Group 1 includes five Representative Test Materials<sup>39</sup> from the JRC Nanomaterials Repository,<sup>40,41</sup> which over the last 15 years have been the subject of several studies promoted by the OECD Working Party on Manufactured Nanomaterials<sup>11,42</sup> and several EU-funded research projects (*e.g.* MARINA, NANoREG). For this reason, a considerable amount of physicochemical data on Group 1 NMs is expected to be available in the sources analysed in this study. Group 2 includes 13 NMs with growing industrial relevance provided by GRACIOUS partners and used as project case-studies.

**Data sources.** This study intended to assess the completeness of physicochemical data generated in previous research projects and made available to GRACIOUS *via* the eNanoMapper database,<sup>24,43</sup> with the aim of shaping the project testing strategy. In particular, GRACIOUS was granted use of experimental data generated in three EU-funded research projects: NanoTest, MARINA and NANoREG. The data providers are industrial or research bodies that were involved in the above-mentioned projects and agreed on incorporating the generated data into databases developed by eNanoMapper. Most of the experimental data generated in the NANoREG project were originally entered *via* a web-based entry tool through dedicated spreadsheets called templates,<sup>31</sup> or by using the templates directly (see section 2.2) in order to make data comparable and available for reuse after the project. In MARINA, characterisation results were also reported in specific

**Table 1** Nanomaterials used for this study. NM = nanomaterial; NP = nanoparticle

Name used in this study	Basic information	OECD code	JRC repository code
<b>Group 1: representative test materials</b>			
Ag NPs	Water-suspended, uncoated	NM-300K	JRCNM03000a
BaSO <sub>4</sub> NPs	Powder, uncoated	NM-220	JRCNM50001a
CeO <sub>2</sub> NPs	Powder, uncoated	NM-212	JRCNM02102a
MWCNTs	Multi-walled carbon nanotubes, powder, uncoated	NM-402	JRCNM04002a
SiO <sub>2</sub> NPs	Amorphous silica, powder, uncoated	NM-200	JRCNM02000a
<b>Group 2: GRACIOUS case study materials</b>			
Silica_unmodified	Water-suspended, unmodified colloidal silica	—	—
Silica_Al	Water-suspended, colloidal silica, Si partly substituted by Al onto the surface	—	—
Silica_silane	Water-suspended, silane-functionalised colloidal silica	—	—
DPP_nano	Organic pigment (diketopyrrolopyrrole), powder, transparent, red colour	—	—
DPP	Organic pigment (diketopyrrolopyrrole), powder, opaque, red colour	—	—
DPP_coated	Organic pigment (diketopyrrolopyrrole), powder, coated, opaque, red colour	—	—
CuPhthalo_blue	Powder, organic pigment blue 15, Cu-phthalocyanine	—	—
CuPhthalo_green	Powder, organic pigment green 7, Cu-phthalocyanine	—	—
Fe <sub>2</sub> O <sub>3</sub> _nano A	Powder, inorganic pigment, red colour	—	—
Fe <sub>2</sub> O <sub>3</sub> _nano B	Powder, inorganic pigment, red colour	—	—
CNF-50	Cellulose nanofibrils (nominal mean diameter: 50 nm)	—	—
CNF-80	Cellulose nanofibrils (nominal mean diameter: 80 nm)	—	—
CNC-25	Cellulose nanocrystals (nominal mean diameter: 25 nm)	—	—



spreadsheets. For all Group 1 NMs the GRACIOUS database served as the main source of data, which were extracted and then re-organised (section 2.3) for the purpose of the completeness assessment. For CeO<sub>2</sub> and BaSO<sub>4</sub> NPs, additional data were provided by project partners who carried out research activities on these materials before the start of GRACIOUS. For this reason, those data were not stored in the initial GRACIOUS database. For all Group 2 NMs, the data were provided by industrial project partners manufacturing those NMs. Most of these data were made available by partners in a structured way, by filling specific templates (section 2.2). Only in the case of cellulose nanofibres data were provided in an unstructured way by sharing the original documents reporting these data.

**Physicochemical properties and characterisation techniques.** In GRACIOUS, project partners agreed to collect or generate characterisation data based on 11 physicochemical properties (Fig. 1) that are considered to be relevant for regulatory purposes, particularly in the context of REACH. Composition, crystallinity, particle size, particle shape, chemical nature of the surface (henceforth “surface chemistry”), and specific surface area (SSA) are considered to be “priority properties”; they are essential for registration of nanoforms under

REACH<sup>10</sup> and considered to be potentially relevant for grouping and read-across by ECHA.<sup>21</sup> In addition, particle size is required for verifying compliance with the European Commission’s Recommendation on the definition of nanomaterial<sup>44</sup> and, under specific conditions, SSA by volume can also be used to identify nanomaterials.<sup>45</sup> Please refer to section S1 in ESI† for a detailed description of the properties’ selection process.

Each physicochemical property was associated with one or more measurement techniques (31 in total, Fig. 1), for which in-house expertise among project partners was available. Multiple techniques that are often employed to measure the same property were considered. For instance, two techniques that can measure the property “density” (VCM and Pycnometry) were selected.

## 2.2 Templates as a tool for data processing

A series of spreadsheets, henceforth called “templates”, collected in a single Excel® workbook were chosen as the tool to structure existing data extracted from the eNanoMapper database and collect the data directly provided by partners. Each template is named after a given property and measurement technique (30 in total, as for Fig. 1) and is divided into two

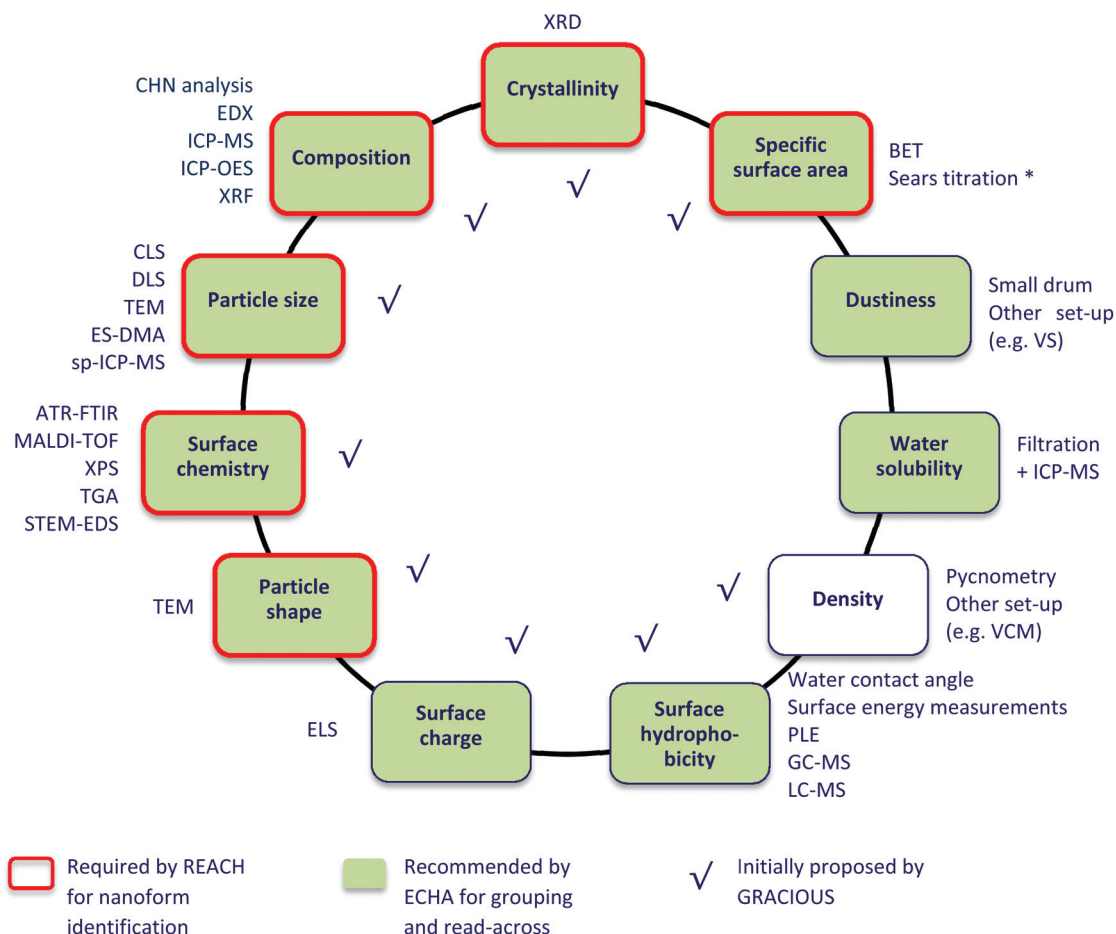


Fig. 1 Physicochemical properties selected in GRACIOUS and their associated measurement techniques. \* Only for silica-based nanoforms.





parts (Fig. S2 in ESI†). The top part shows a series of columns referring to a set of information items required to be recorded to sufficiently describe a characterisation experiment. The number of columns depends on the specific technique. From the left side to the right side, the columns are divided into 4 sections called Sample information, Method and instrument information, Results, and SOP (Standard Operating Procedure). The templates for DLS, TEM, and XRF also contain a section named Sample preparation where detailed information on how the sample was prepared for measurement is reported. The Results section includes one or more columns representing the desired outcome of the experiment and thus considered to be the “key” information items. The quantities recorded under these columns are called “property values”. The complete list of key information items is shown in section S3.† The bottom part of the template hosts the collected experimental (meta)data matching the required information items, including the property values. Each (meta)data row thus reports the (meta)data characterising an experiment performed to generate the property values. When the (meta)data from an experiment is transferred from the eNanoMapper database to the corresponding template, it is restructured into one (meta) data row including at least one property value.

Templates were first created in NANoREG<sup>31</sup> for a set of 12 physicochemical properties and 30 measurement techniques. They were developed for the purpose of harmonising data logging and intended to collect the experimental data generated by partners in a harmonised format, thus ensuring comparability, reproducibility and reuse after the project. In GRACIOUS, three original NANoREG templates were modified and six new templates developed.<sup>36</sup> In this study, the templates were not only used for data logging but also to assess the completeness of the whole dataset using the information items as criteria.

### 2.3 Procedure of data collection

In this article, the term “data record” is used to identify a unit of data as extracted from the eNanoMapper database. Each data record contains information on one specific quantity (or value) determined in an experiment using a specific measurement technique. If more quantities are determined, then more data records are associated with the same experiment. A “data row” is used to identify a unit of data as structured into the templates (section 2.2). It includes all information on an experiment and the quantities obtained from it as prescribed by the corresponding template.

The available data for Group 1 NMs were collected from the eNanoMapper database following a three-step procedure (more details in ESI, section S4†): (i) *Data extraction*. The database was first queried by each project (e.g. MARINA) and property (e.g. particle size). The resulting data records were extracted and manually screened for Group 1 NM names (as in Table 1) and selected measurement techniques (as in Fig. 1). Manual screening was necessary due to technical constraints in the query tool and to be able to detect any entry errors. (ii) *Data cleaning*. Only the (meta)data fulfilling the information items

requested by the templates were manually checked and, if needed, cleaned as explained below in section 2.4. (iii) *Transfer of extracted data to templates*. The cleaned data were then transferred to templates. In this step, the (meta)data resulting from an experiment (thus having the same identifier) and hosted in multiple data records was combined into a unique data row in the proper template. Data records that did not contain actual property values were discarded. For example, a data record referring to the technique *BET analysis* would be discarded if it did not report a value under the column *BET surface*, which was selected as key information item (see paragraph 2.2).

Industrial project partners provided the data for Group 2 NMs directly structured according to the templates. Only step (ii) *Data cleaning* was carried out. The same applied to the data for CeO<sub>2</sub> and BaSO<sub>4</sub> NPs provided by project partners.

### 2.4 Data cleaning actions

Data cleaning (step (ii)) of the collected dataset consisted of the following actions.

**Correction of entry errors.** Erroneously reported information in the original data was corrected. This included: correction of typos, attribution of the right technique or NM name and insertion of a missing measurement unit, if it was possible to infer it from the available (meta)data.

**Data conversion.** This action was necessary in order to comply with the information items requested by each template. Very often, for instance, measurement units reported in the database are different from those required by the templates and both value and unit needed to be converted.

**Duplicate elimination.** Data record duplicates and partial duplicates (records referring to the same measurement but storing complementary information about it) were identified thanks to their universally unique identifier (UUID) number.<sup>24</sup> Duplicates were removed and partial duplicates combined so to generate one individual data row in the templates.

### 2.5 Data evaluation

**Available data and gaps.** Data rows contained in every template were counted to quantify data availability and identify the gaps for each combination material/property. If at least one template for a certain property contained at least one data row about a given material, data on that combination property/material was considered available. If none of the templates referring to the same property contained any data rows for a given material, a data gap for that combination property/material was identified.

**Data completeness.** Data completeness may be defined as “a measure of the availability of the necessary, non-redundant data and associated (meta)data for a given entity”,<sup>27</sup> and describes the extent to which experimental details and results for an experiment are reported. Data completeness is related to data quality, which can be considered as a measure of the usefulness, clarity and correctness of data.<sup>27</sup>

The templates (section 2.2) are an ideal tool to determine data completeness, as the latter can be measured in terms of compliance with an information checklist.<sup>27,37</sup> The infor-



mation items required by the templates were therefore used in this study as criteria to assess the completeness of each data row. A “completeness score” (CS) was defined to numerically assess data completeness as the degree of compliance of the reported (meta)data with the information items requested by templates. For the row  $i$ ,  $CS_i$  is a number between 0 and 1 calculated by dividing the number of information items for which (meta)data are reported in a row by the total number of information items requested by the template (eqn (1)).

$$CS_i = \frac{\text{number of items reported in row } i}{\text{number of items required}} \quad (1)$$

Bearing in mind the preconditions for data selection (section 2.1), the  $CS_i$  is a measure of the completeness of the (meta)data stored in a row where the key property value(s) of the template are present. The CS associated to all the data collected in a specific template  $j$  (that is, for a certain combination NM/technique) was then defined as in eqn (2).

$$CS_j = \frac{\sum_{i=1}^n CS_i}{n}, \quad (2)$$

where  $n$  is the number of rows included in the template  $j$ . Similarly, the CS of a specific template section was determined by taking into account only the number of items belonging to the given section. Finally, CS for each nanomaterial and technique were calculated by averaging  $CS_j$ .

**Analysis of selected property values.** Other descriptors of data quality such as accuracy could not be comprehensively assessed due to the heterogeneity and small size of the collected dataset that was available when this study was conducted. However, the reported property values for a certain combination of NM and technique were discussed with the help of charts and used in combination with the completeness

scores to draw preliminary conclusions on the testing needs in GRACIOUS.

## 3. Results and discussion

### 3.1 Data collection

(i) *Data extraction* (only for Group 1 NMs). From the GRACIOUS database 15 408 data records were extracted and then reduced to 4050 after filtering for relevant NMs and further to 2704 after manual screening for relevant techniques. (ii) *Data cleaning*. Table 2 lists the performed data cleaning actions, the problems encountered and their occurrence. Corrections were carried out only when considered necessary and supported by the (meta)data available for that row, not with the intent of addressing every single inaccuracy found. For example, Ag NPs are referred to as “JRCNM03000a”, “Ag 16.7 nm”, “NM-300K”, “NM-300K (Ag)” or “dispersion medium” in the collected dataset under the columns “name” or “public name”. Only the action replacing “dispersion medium” with “Ag NPs” was applied because the name was recorded in other fields of that row, which clearly identified the material. The most frequent issue was data reported in the wrong field or format (460 times), followed by missing/not complying measurement units and information items described by inappropriate values, which sums up to a total amount of 1634 data cleaning actions for Group 1 NMs and 253 for Group 2 NMs. Apart from entry 2, the issues reported in Table 2 are common when free text is used to enter values in databases<sup>46</sup> and no steps of data curation based on scientific judgement are carried out. Since there was a lack of clear universally adopted criteria for evaluation of NM data, the insertion of data on NMs into eNanoMapper was carried out with the clear intention of “not to exclude automatically the unreliable data from further considerations”.<sup>24</sup>

**Table 2** Cleaning actions performed on the (meta)data extracted from the eNanoMapper database

Entry number	Problem encountered	Example	Cleaning action	Occurrence (Group 1)	Occurrence (Group 2)
1	Values not reported in right field and retrieved elsewhere	Concentration values reported as free text under “experimental code”	Data were moved to the correct column	460	
2	Measurement unit missing or not complying with template	Viscosity values expressed in <i>dyne</i> whereas the template requires them in <i>centipoise</i>	Unit was added or changed and values converted	396	
3	Information item described by inappropriate value	“n.a.”, “not recorded”, “not available” or “0” used to indicate absence of data	Inappropriate values were removed and cell left blank	375	253
4	Erroneous technique name	“Scanning electron microscopy” is reported as the technique, while the size values are expressed as zeta-average hydrodynamic diameter	The right technique was inserted	293	
5	Erroneous property name	“Water solubility” reported as property instead of “size” in records with sp-ICP-MS as technique	The right property was inserted	80	
6	Erroneous material name	“Dispersion medium” reported as NM name in records with NM-300K (meta)data	The right material name was inserted	30	
<b>Total occurrence →</b>				<b>1634</b>	<b>253</b>



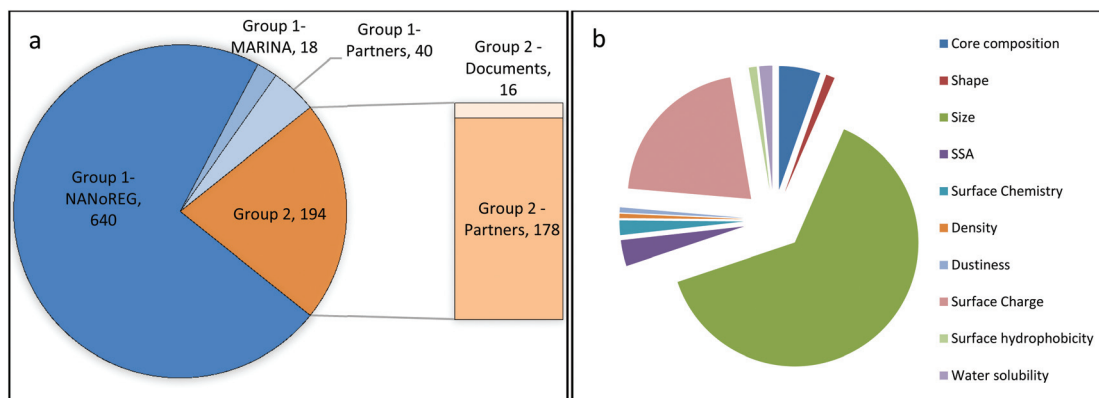


Fig. 2 (a) Pie chart showing the contribution of the various data sources to the collected dataset, in terms of number of data rows into GRACIOUS templates. (b) Pie chart showing the relative share of data rows for each physicochemical property.

(iii) *Transfer of data to templates* (only for Group 1 NMs). The restructuring of the cleaned data records according to the templates resulted in 777 data rows, which were reduced to 698 by discarding rows that did not contain the key property values. More than 90% of those rows contain (meta)data from NANoREG and about 3% from MARINA. The remaining rows are data provided by GRACIOUS partners (Fig. 2a). The database instance dedicated to data generated in NanoTest did not contain relevant records for Group 1 NMs but only records on NMs that were not selected for this study. For Group 2 NMs, collected data amounted to 194 data rows that were entirely provided by industrial project partners: 178 (more than 90%) were directly received in the format required by the templates, whereas the rest was manually inserted into the templates from the original documents. Each data row included the key property value(s). In conclusion, a total number of 892 rows were collected.

### 3.2 Available data and gaps

The distribution of the available data rows is very heterogeneous across all materials and properties, as shown by the green cells of Table 3. For Group 1 NMs, the largest number of rows was collected for SiO<sub>2</sub> NPs (231) and the smallest for MWCNTs (16). For Group 2 NMs, the number of rows ranges from a maximum of 22 for Fe<sub>2</sub>O<sub>3</sub>\_nano\_B, silica\_silane and silica\_unmodified to a minimum of five for CNC-25 and CNF-80. If both groups are considered, particle size is by far the property with the largest number of rows (564), followed by surface charge (Fig. 2b). All other properties are represented by less than 15 data rows each.

Available data and gaps for each combination material/property are shown in Table 3. None of the studied NMs can be considered completely characterised according to the 11 properties as at least one data gap (orange-coloured cells in Table 3) is present in each case. If only priority properties are considered, CeO<sub>2</sub> NPs is the only fully characterised nanomaterial. Group 1 NMs generally show fewer data gaps than Group 2 NMs. Particle size is the property with the fewest data gaps (only 3 out of 18 NMs do not report data on size). Table 3 can then be used to quickly identify the combinations of

material/property for which no data are available and where it is therefore necessary to set up experimental activities to characterise them and fill the gaps in the project. An extended version of Table 3 including available data for each combination technique/material could then help identify sub-gaps and shape the testing in detail. For example, data on a property for a certain material could be available only from some of the techniques chosen in the project. In this case, measurements with the missing techniques will have to be performed.

### 3.3 Data completeness

The completeness (or degree of compliance) of the available (meta)data rows with the information items requested by the templates varies greatly in the collected dataset. Table 4 shows the calculated CS for each combination material/technique for group 1 NMs. The overall CS is 0.43, which means that less than half of the information requested by the templates could be collected.

Looking at individual NMs, the average CS across all properties varies from 0.31 for MWCNTs to 0.54 for BaSO<sub>4</sub> NPs. For individual techniques, the average CS across all materials varies from 0.13 for composition measured by ICP-MS to 0.90 for surface charge determined *via* ELS (for IEP). This variability is partially due to the sources (Fig. S5a†): data rows provided by GRACIOUS partners who were directly asked from the start of the project to use the templates to record their existing data have the highest CS values. An example is surface charge of BaSO<sub>4</sub> NPs determined *via* ELS (for IEP) measurements, which has the highest CS because a larger amount of data has been provided by partners directly *via* templates. Data rows from NANoREG have a lower than expected average CS, equal to 0.43, despite most of the templates used in this study were the result of a large effort by NANoREG partners who aimed to log all their experimental data in a harmonised format that would facilitate reproducibility, comparability and reuse after the end of the project. The relatively low completeness of NANoREG data rows could be explained by a time lag between the finalisation of the templates and the availability of the data. Probably, the templates had been finalised at a later



**Table 3** Available data (green cells) and gaps (orange cells) for each combination material/property. The number of data rows available in the collected dataset for each combination is indicated in the green cells whereas the total number for each material and for each property is reported at the end of each row and column, respectively

	Composition	Crystallinity	Particle shape	Particle size	SSA	Surface chemistry	Density	Dustiness	Surface charge	Surface hydrophob.	Water solubility	Total
Group 1 NMs	Ag NPs	12	5	149	5	2	1		8		9	183
	BaSO <sub>4</sub> NPs	3		76	5				17	1		106
	CeO <sub>2</sub> NPs	3	2	114	6	1	1		33	1		162
	MWCNTs	6		7		1		2				16
Group 2 NMs	SiO <sub>2</sub> NPs		2	199	8	2			14		6	231
	Silica_unmodified	6		4	2				10			22
	Silica_Al	5		4					10			19
	Silica_silane	6		4	2				10			22
	DPP_nano			1	1	1			11	1		15
	DPP			1	1	1			11	1		15
	DPP_coated			1	1	1			11	1		15
	CuPhthalo_blue			1	1	1		1	11	1		17
	CuPhthalo_green				1	1		1	11	1		15
	Fe <sub>2</sub> O <sub>3</sub> _nano_A			1	1	1		1	11	1		17
	Fe <sub>2</sub> O <sub>3</sub> _nano_B			1	1	1		1	18	1		22
	GNF-50	1	1	1		2	1					6
	GNF-80	1	1	1		1	1					5
	CNC-25	1	1	1		1	1					5
	Total	44	4	564	30	17	5	6	186	9	15	892

stage of the project when some experimental activities were already carried out and results recorded by individual laboratories in a non-harmonised way. Data rows derived from original documents shared by partners and those generated in MARINA had the lowest average CS (0.31 and 0.21, respectively). In both cases, data were generated and recorded either without using a dedicated template or using a different one (e.g. MARINA) from that driving the completeness evaluation.

The overall average CS for Group 2 NMs is 0.68 (Fig. S5b†), much larger than for Group 1 NMs (0.43). Again, the greater completeness is linked to the direct request made by GRACIOUS to record the existing data in templates, providing as much (meta)data as possible.

The CS varies also greatly across template sections (Fig. S5a†). For Group 1 NMs, Results and Sample information are the most complete sections (CS = 0.52 and 0.45, respectively), followed by Method and instrument information (0.35), Sample preparation (0.24), and SOP (0.15). Therefore, there seems to have been a consistent tendency to preferentially report information on an experiment's outcome and sample identity rather than providing the complete SOP. As expected, the completeness of template sections for Group 2 NMs is greater due to greater contribution of data from partners who were explicitly asked to log their (meta)data as comprehensively as possible in the templates (CS = from 0.55 to 0.75).

### 3.4 Analysis of selected property values

Even if deemed important to fully define any testing strategy, the evaluation of data quality descriptors such as accuracy was not the object of this study. A preliminary analysis of the collected key property values was however carried out and the corresponding results are shown in this section. More results are illustrated in the ESI.†

**Amount of collected property values.** Very often there is only one available property value for a given NM and technique (see Table 2): this is the case for 25% and 70% of the measurement techniques used for Group 1 and Group 2 NMs, respectively. This means that in several cases the property was measured through one characterisation experiment performed by one laboratory. The quality of such experiments may vary a lot depending on the laboratory, for example laboratories with accredited GLP (good laboratory practice) are considered to be reliable and assumed to produce high quality data. Another aspect is the SOP that has been used to perform the measurement: following a national or international standard enhances the reliability of the results. Lack of data rows reporting results from different laboratories or missing information on the SOP used to perform the measurement may prevent other users from fully relying on and thus reusing the existing data. Additional characterisation by various laboratories could then be desirable to ensure measurement accuracy and potentially determine mean values or quality-related factors such as reproducibility or variability, especially when the method has not been standardised yet, which is often the case for NMs.<sup>47</sup> Only surface charge (for Group 1 and 2 NMs, section S6† and Fig. 3) and particle size (for Group 1 NMs, Fig. 4 and 5) show a rela-

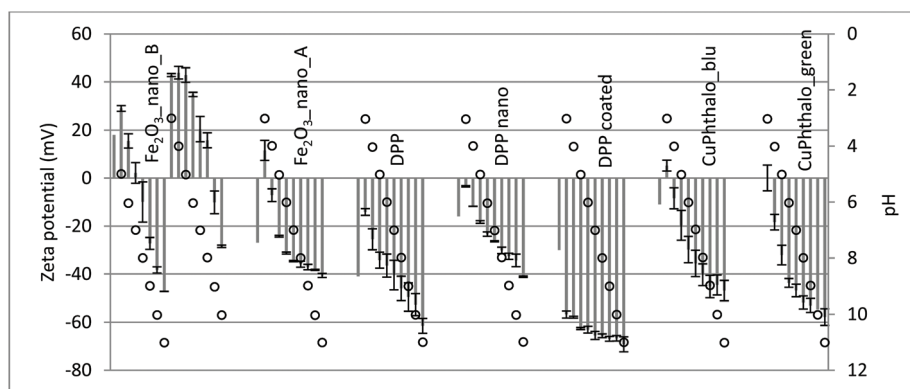




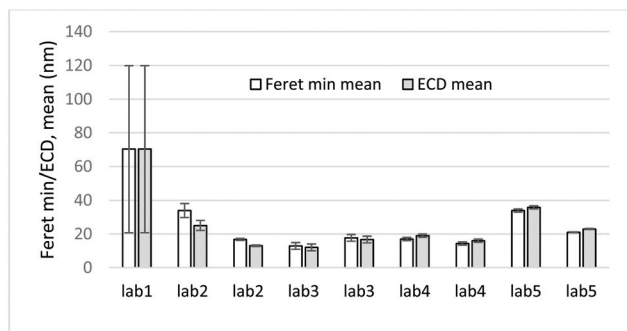
**Table 4** Completeness score (CS) for each combination material/technique (for Group 1 NMs). Green cells: data available. Orange cells: data gap. Only techniques with available data are reported

Property	Technique	Ag NPs	BaSO <sub>4</sub> NPs	CeO <sub>2</sub> NPs	MWCNTs	SiO <sub>2</sub> NPs	Average CS
Composition	CHN analysis	0.65	0.65				0.65
	EDX			0.54			0.54
	ICP-MS	0.09			0.18		0.13
	ICP-OES	0.3	0.6	0.57	0.26		0.43
	XRF		0.14	0.14			0.14
Crystallinity	XRD		0.28	0.28			0.28
Particle shape	TEM	0.45		0.42		0.45	0.44
Particle size	CLS					0.21	0.21
	DLS	0.24	0.26	0.26	0.23 <sup>a</sup>	0.26	0.25
	TEM	0.33		0.33		0.37	0.35
SSA	BET analysis		0.25	0.24		0.31	0.27
Surface chemistry	TGA		0.6		0.23	0.6	0.48
	XPS		0.72	0.72		0.19	0.55
Density	VCM		0.75	0.75			0.75
Dustiness	VS				0.64		0.64
Surface charge	ELS (for IEP)		0.90	0.90			0.90
	ELS (for Z-potential)	0.2	0.65	0.63		0.22	0.43
Surface hydrophobicity	Water contact angle		0.64	0.68			0.66
Water solubility	Filtration + ICP-MS	0.56				0.51	0.54
Average CS		0.35	0.54	0.5	0.31	0.35	0.43

<sup>a</sup> Even if data are available, DLS is not foreseen to be used to characterise MWCNTs.



**Fig. 3** Collected surface charge values (z-potential values, grey bars) and corresponding uncertainty bars for some Group 2 materials. Empty dots: pH values.



**Fig. 4** Mean Feret min and ECD (equivalent circle diameter) values and corresponding uncertainty bars from TEM measurements of CeO<sub>2</sub> NPs.

tively large amount of values. But this abundance is often only apparent: a closer look to the (meta)data associated to the property values reveals that measurements have been often carried out under different experimental conditions. For example, the surface charge values collected for each Group 2 NM (Fig. 3) are different because they have been determined at a different pH. This shows the importance of having access to and linking detailed information on applied method protocols and experimental conditions to a characterisation experiment (for example *via* the templates) to properly interpret, compare and reuse shared data.

**Data from complementary techniques.** Data on a given property can often be generated by using more than one technique. As the measurands for a given property are often method-defined,<sup>47</sup> the values available for a certain NM could describe the property in a different way. For example, the available





**Data from different techniques.** When the number of available values is sufficiently large, statistical quantities such as mean value and variability can be calculated. One can then

compare the distribution of values associated to the same property but generated by different techniques, and analyse potential differences. For the collected dataset, this is possible for size values of some Group 1 NMs. For CeO<sub>2</sub> NPs, the available size values generated by TEM (Fig. 4) and DLS (Fig. 5) show a very different dispersion. Collected TEM size values range from 12 nm to 73 nm, with an interquartile range (third quartile minus first quartile) of 15 nm and a mean value of 26 nm. Collected DLS size values are spread on a wider range (from 42 nm to 4175 nm) and show a larger interquartile range (115 nm) and mean value (342 nm). This outcome is in agreement with previous studies<sup>48</sup> and not surprising, as DLS



measurements are strongly affected by particle agglomeration and dispersibility in a given medium and cannot distinguish between constituent particles and agglomerates/aggregates. The higher DLS size values (1908, 4175, 1666, 1649 nm in Fig. 5, as reported in the data source) could have been caused by particle agglomerates, but the corresponding (meta)data do not contain information on sample preparation and dispersion protocol and it is thus not possible to clearly identify the reason. It is interesting to point out that, if only DLS values were available, it would not be possible to decide if the investigated CeO<sub>2</sub> is a nanomaterial according to the current definition by the European Commission<sup>44</sup> (which requires that 50% of the (constituent) particles in a material have at least one external dimension smaller than 100 nm to be identified as a nanomaterial). In addition, the median DLS values were not reported in the data sources. TEM is considered a more appropriate technique to measure the constituent particle size of materials within its application range.<sup>47</sup> In this case, median values were reported and could clearly identify CeO<sub>2</sub> NPs as a nanomaterial, as each median size value is less than 100 nm. Therefore, when only DLS data are available (as for MWCNTs and BaSO<sub>4</sub> NPs, Table 4), additional characterisation by electron microscopy, for example TEM, is recommended for regulatory purposes.<sup>49</sup> For MWCNTs, additional TEM characterisation is actually necessary, as DLS measurements are not suitable to determine the diameter of rod-like nanoforms.

**Lack of relevant information.** A property may need to be remeasured if available data rows do not include (meta)data for selected relevant information items. For example, although TEM can be used to determine both particle size and shape, TEM-derived shape values (such as aspect ratios) are less often reported than those for size. For CeO<sub>2</sub> NPs, only two shape values (Fig. 6) were reported from the same experiment that provided nine size values (Fig. 4). Another aspect is the lack of (meta)data describing the uncertainty of the property values. For instance, BET analysis from various laboratories generated comparable SSA values ( $26.9 \pm 2.0 \text{ m}^2 \text{ g}^{-1}$  for CeO<sub>2</sub> NPs, Fig. 6) but are reported without uncertainty despite the SSA template asking for it (as information item “associated error”).

## 4. Summary and conclusions

This study presented and applied a procedure for collection and completeness evaluation of existing physicochemical data for both representative NMs (Group 1) investigated in previous projects and NMs provided by GRACIOUS industrial partners and used as project case studies (Group 2). This assessment was primarily meant to shape the testing strategy needed to fill data gaps, with the goal of reusing as much as possible existing data and thus reducing testing efforts, so that only the identified data gaps would have to be filled in GRACIOUS.

Data completeness was assessed by assigning a completeness score, which compares the actual availability of (meta) data with the information required for a specific combination of physicochemical property and characterisation method by

the corresponding data logging template developed in NANoREG or GRACIOUS. The reported property values were discussed with the help of charts and used in combination with the completeness scores to draw preliminary conclusions on the testing needs in GRACIOUS.

This study identified numerous data gaps in the physicochemical data of Group 1 and Group 2 NMs. Therefore, additional characterisation will have to be carried out to fill those data gaps for the purpose of GRACIOUS. To this end, Table 3 could be used to quickly identify the combinations of material/property for which it is necessary to set up experimental activities in the project, for example data on particle size (for DPP\_coated, CuPhthalo\_green and Fe<sub>2</sub>O<sub>3</sub>\_nano\_B), and data on surface chemistry (for Ag NPs and silica-based Group 2 NMs).

Additional characterisation could be regarded as needed in cases where data are available but property values are few, too variable, or not well described (*i.e.* low CS). For example, as the amount of collected property values for some combinations of material/property is relatively small (*e.g.* 1 surface hydrophobicity value available per NM), it could be desirable to repeat the experiment to increase the number of available property values. This would allow partners to assess quality-related factors such as reproducibility or variability. At the same time, the experiment may need to be repeated even when many data exist, all generated with the same technique, but span over a large range of values. Moreover, data stored in the eNanoMapper database and made available for this study did score surprisingly low in terms of data completeness, with fundamental information often missing (*e.g.* about SOPs or uncertainty), which reduces the possibility of comparing and re-using the data in GRACIOUS and for other purposes. Here, additional characterisation could be performed to generate new values accompanied by a full description of experimental conditions, protocols, methods, *etc.*, as required by data logging templates developed in NANoREG and GRACIOUS.

Although the Group 1 NMs have been studied in various collaborative projects and with considerable public funding, a large part of the generated physicochemical data, even if published, is not available in a structured, user-friendly format: the data collection procedure from the eNanoMapper database required the correction of numerous errors, such as wrong names of NMs and techniques. Most of these issues originated from the data entry procedures used by specific projects or data transfer procedures into eNanoMapper, and do not represent a great hurdle to the understanding of the (meta)data when basic knowledge of physicochemical characterisation is available. However, a large part of the (meta)data concerning Group 1 NMs have been recently made publicly available *via* the eNanoMapper database on the EUON website, following an approach promoted by the European Council in its recent conclusions on chemicals (26 June 2019).<sup>§</sup> Under these new

<sup>§</sup> European Council, Council conclusions on chemicals, 26 June 2019, <https://www.consilium.europa.eu/en/press/press-releases/2019/06/26/council-conclusions-on-chemicals/>, (accessed 18 July 2019).



circumstances, it is desirable to call for an expert review of data on those NMs to improve the database's clarity and usability. It would be helpful to foresee such action in general for all data from any project that will be uploaded into eNanoMapper and then made publicly available on the EUON website in the future. Additionally, the database content could be improved by inserting information on data quality, such as the completeness evaluation as presented in this study. The CS could be implemented into the database as an additional field describing the data quality of each record (or experiment) in terms of completeness against the corresponding templates. It could also become an acceptance criterion when new data is uploaded into the database: in this case, data entering would only be allowed if a minimum CS value is met or if certain information items labelled as compulsory are provided. In this context, CS calculation may be customised by assigning higher weights to those information items that are deemed more important (e.g. the specification of the SOPs used). This option is being explored by eNanoMapper database developers, who are considering the possibility of implementing the completeness evaluation procedure presented in this paper in an automatic way in a future version of the database.

This study also shows that the (meta)data collected from sources which used the NANoREG and/or GRACIOUS templates as format for data logging have the highest CS values. Indeed, properties such as surface charge, particle size measured by ES-DMA and composition measured by ICP-OES reached CSs higher than 0.85 in all of the Group 2 NMs. This proves that the use of well-defined and harmonised reporting schemes for experimental results is a useful tool to increase (meta)data completeness and, ultimately, ensure their comparability, integration and re-use. It is recommended that the structure firstly designed in NANoREG and then extended to other physicochemical properties and methods in GRACIOUS is used in future projects as a blueprint to develop new templates covering other properties and methods for NM data logging.

The data analysed in this study and other data on nanomaterials are publicly available in databases (e.g. eNanoMapper) or will be made available soon. Based on the outcomes of this work, it is suggested that users who intend to carry out studies on safety (or any other) aspects of nanomaterials (i) critically analyse the completeness of the relevant existing data and (ii) carefully plan the generation, collection and reporting of new data and metadata with an emphasis on their completeness and comparability. The availability and accessibility of such datasets will foster their re-use for other scientific studies as well as for regulatory purposes. In particular, complete datasets are reproducible as all descriptors for the method, instrument settings, SOPs used, results and uncertainty are reported. In addition, they can be compared and properly integrated with other datasets.

Data completeness regarding 11 physicochemical properties was analysed in this work with the objective of enabling grouping and read-across for 18 nanomaterials in GRACIOUS. For this case study, a selection of well-studied Representative Test

Materials and novel industrial nanomaterials with growing commercial importance was chosen. Beyond that, the availability of complete datasets in a harmonised format would be beneficial for any substance when fulfilling regulatory information requirements, for instance for the registration of nanoforms and sets of similar nanoforms in a dossier under the European REACH Regulation. If it is likely that a specific substance will be produced with a large variety of different nanoforms, it is suggested to report the physicochemical properties of these nanoforms using harmonised templates such as those presented in this study. This would facilitate assessing the completeness of the (meta)data, interpreting the results and identifying any gaps. This way, it might be considerably easier to perform grouping and read-across between nanoforms, or even to define sets of similar nanoforms, which would ultimately reduce the costs of registration.

## Abbreviations

ATR-FTIR	Attenuated total reflection – Fourier transform infrared
BET	Brunauer–Emmett–Teller
CHN analysis	Carbon hydrogen nitrogen analysis
CLS	Centrifugal liquid sedimentation
CS	Completeness score
DLS	Dynamic light scattering
ECD	Equivalent circle diameter
ECHA	European Chemicals Agency
EDX	Energy-dispersive X-ray spectroscopy
ELS	Electrophoretic light scattering
ES-DMA	Electrospray-differential mobility analysis
GC-MS	Gas chromatography – mass spectrometry
ICP-MS	Inductively coupled plasma – mass spectrometry
ICP-OES	Inductively coupled plasma – optical emission spectrometry
IEP	Isoelectric point
JRC	Joint Research Centre
LC-MS	Liquid chromatography – mass spectrometry
MALDI-TOF	Matrix-assisted laser desorption ionization time of flight mass spectrometry
NM	Nanomaterial
NP	Nanoparticle
OECD	Organisation for Economic Co-operation and Development
PLE	Pressurized liquid extraction
REACH	Registration, Evaluation, Authorization and Restriction of Chemicals
sp-ICP-MS	Single particle – inductively coupled plasma – mass spectrometry
SOP	Standard operating procedure
SSA	Specific surface area
STEM-EDX	Scanning transmission electron microscope – energy-dispersive X-ray spectroscopy
TEM	Transmission electron microscopy





TGA	Thermogravimetric analysis
VCM	Volumetric centrifugation method
VS	Vortex shaker
VSSA	Volume-specific surface area
XPS	X-ray photoelectron spectroscopy
XRD	X-ray powder diffraction
XRF	X-ray fluorescence

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The GRACIOUS project was supported by the European Commission Horizon 2020 programme, Grant Agreement No. 760840.

## References

- European Commission, What is Horizon 2020?, <https://ec.europa.eu/programmes/horizon2020/what-horizon-2020>, (accessed 17 April 2019).
- M. C. Roco, C. A. Mirkin and M. C. Hersam, *J. Nanopart. Res.*, 2011, **13**, 897–919.
- C. W. Noorlander, M. W. Kooi, A. G. Oomen, M. V. Park, R. J. Vandebriel and R. E. Geertsma, *Nanomedicine*, 2015, **10**, 1599–1608.
- L. F. Fraceto, R. Grillo, G. A. de Medeiros, V. Scognamiglio, G. Rea and C. Bartolucci, *Front. Environ. Sci.*, 2016, **4**, 20.
- R. J. Peters, H. Bouwmeester, S. Gottardo, V. Amenta, M. Arena, P. Brandhoff, H. J. Marvin, A. Mech, F. B. Moniz and L. Q. Pesudo, *Trends Food Sci. Technol.*, 2016, **54**, 155–164.
- A. Duschl and G. Windgasse, *J. Nanopart. Res.*, 2018, **20**, 335.
- A. P. K. Jantunen, S. Gottardo, K. Rasmussen and H. P. Crutzen, *NanoImpact*, 2018, **12**, 18–28.
- H. Rauscher, K. Rasmussen and B. Sokull-Klüttgen, *Chem. Ing. Tech.*, 2017, **89**, 224–231.
- K. Rasmussen, H. Rauscher, S. Gottardo, E. Hoekstra, R. Schoonjans, R. Peters and K. Aschberger, in *Nanomaterials for Food Applications*, Elsevier, 2019, pp. 381–410.
- European Commission, *Off. J. Eur. Union*, 2018, **L 308**, 1–20.
- K. Rasmussen, H. Rauscher, P. Kearns, M. González and J. R. Sintes, *Regul. Toxicol. Pharmacol.*, 2019, **104**, 74–83.
- H. F. Krug, *Angew. Chem., Int. Ed.*, 2014, **53**, 12304–12319.
- European Parliament and Council, *Off. J. Eur. Union*, 2010, **L 276**, 33–79.
- European Parliament and Council, *Off. J. Eur. Union*, 2006, **L 396(1)**, 1–849.
- N. Burden, K. Aschberger, Q. Chaudhry, M. J. Clift, S. H. Doak, P. Fowler, H. Johnston, R. Landsiedel, J. Rowland and V. Stone, *Nano Today*, 2017, **12**, 10–13.
- A. Mech, K. Rasmussen, P. Jantunen, L. Aicher, M. Alessandrelli, U. Bernauer, E. Bleeker, J. Bouillard, P. Di Prospero Fanghella and R. Draisci, *Nanotoxicology*, 2019, **13**, 119–141.
- J. H. Arts, M. Hadi, M.-A. Irfan, A. M. Keene, R. Kreiling, D. Lyon, M. Maier, K. Michel, T. Petry and U. G. Sauer, *Regul. Toxicol. Pharmacol.*, 2015, **71**, S1–S27.
- A. Oomen, E. Bleeker, P. Bos, F. van Broekhuizen, S. Gottardo, M. Groenewold, D. Hristozov, K. Hund-Rinke, M.-A. Irfan and A. Marcomini, *Int. J. Environ. Res. Public Health*, 2015, **12**, 13415–13434.
- European Commission, CORDIS, EU research results, GRACIOUS: Grouping, read-across, characterisation and classification framework for regulatory risk assessment of manufactured nanomaterials and safer design of nano-enabled products, <https://cordis.europa.eu/project/rcn/212339/factsheet/en>, (accessed 11 July 2019).
- V. Stone, S. Gottardo, E. A. J. Bleeker, H. Braakhuis, S. Dekkers, T. Fernandes, A. Haase, N. Hunt, D. Hristozov, P. Jantunen, N. Jeliaskova, H. Johnston, L. Lamon, F. Murphy, K. Rasmussen, H. Rauscher, A. Sánchez Jiménez, C. Svendsen, D. Spurgeon, S. Vázquez-Campos, W. Wohlleben and A. G. Oomen, 2019, submitted to Nature Nanotechnology.
- ECHA, Guidance on information requirements and chemical safety assessment – Appendix R.6-1 for nanomaterials applicable to the Guidance on QSARs and Grouping of Chemicals, [https://echa.europa.eu/documents/10162/23036412/appendix\\_r6\\_nanomaterials\\_en.pdf/71ad76f0-ab4c-fb04-acba-074cf045eaaa](https://echa.europa.eu/documents/10162/23036412/appendix_r6_nanomaterials_en.pdf/71ad76f0-ab4c-fb04-acba-074cf045eaaa).
- S. Karcher, E. L. Willighagen, J. Rumble, F. Ehrhart, C. T. Evelo, M. Fritts, S. Gaheen, S. L. Harper, M. D. Hoover and N. Jeliaskova, *NanoImpact*, 2018, **9**, 85–101.
- European Commission, CORDIS, EU research results, eNanoMapper: A database and ontology framework for nanomaterials design and safety assessment, <https://cordis.europa.eu/project/rcn/110961/factsheet/en>, (accessed 11 July 2019).
- N. Jeliaskova, C. Chomenidis, P. Doganis, B. Fadeel, R. Grafström, B. Hardy, J. Hastings, M. Hegi, V. Jeliaskov and N. Kochev, *Beilstein J. Nanotechnol.*, 2015, **6**, 1609–1634.
- ECHA, European Observatory on Nanomaterials (EUON), eNanoMapper, <https://euon.echa.europa.eu/enanomapper>, (accessed 12 July 2019).
- D. de la Iglesia, S. Harper, M. D. Hoover, F. Klaessig, P. Lippell, B. Maddux, J. Morse, A. Nel, K. Rajan and R. Reznik-Zellen, Nanoinformatics 2020 roadmap, <https://greennano.org/content/nanoinformatics-2020-roadmap>, (accessed 11 July 2019, 2019).
- R. L. Marchese Robinson, I. Lynch, W. Peijnenburg, J. Rumble, F. Klaessig, C. Marquardt, H. Rauscher,



- T. Puzyn, R. Purian and C. Åberg, *Nanoscale*, 2016, **8**, 9919–9943.
- 28 A. Haase and F. Klaessig, EU US roadmap nanoinformatics 2030, <https://zenodo.org/record/1486012#.XYjehNFL1s>, (accessed 11 July 2019).
- 29 D. G. Thomas, S. Gaheen, S. L. Harper, M. Fritts, F. Klaessig, E. Hahn-Dantona, D. Paik, S. Pan, G. A. Stafford and E. T. Freund, *BMC Biotechnol.*, 2013, **13**, 2.
- 30 European Commission, CORDIS, EU research results, NANoREG: A common European approach to the regulatory testing of nanomaterials, <https://cordis.europa.eu/project/rcn/107159/factsheet/en>, (accessed 11 July 2019).
- 31 S. Totaro, H. Crutzen and J. Riego Sintes, *Data logging templates for the environmental, health and safety assessment of nanomaterials*, Publications Office of the European Union, Luxembourg, 2017, DOI: 10.2787/505397.
- 32 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, *Sci. Data*, 2016, **3**, 160018.
- 33 A. Tropsha, K. C. Mills and A. J. Hickey, *Nat. Nanotechnol.*, 2017, **12**, 1111.
- 34 European Commission, CORDIS, EU research results, NanoTest: Development of methodology for alternative testing strategies for the assessment of the toxicological profile of nanoparticles used in medical diagnostics, <https://cordis.europa.eu/project/rcn/86691/factsheet/en>, (accessed 12 July 2019).
- 35 European Commission, CORDIS, EU research results, MARINA: Managing risks of nanoparticles, <https://cordis.europa.eu/project/rcn/101207/factsheet/en>, (accessed 12 July 2019).
- 36 S. Gottardo, G. Ceccone, H. Freiburger, N. Gibson, M. Kellermeier, E. Ruggiero, B. Stolpe, W. Wacker and H. Rauscher, *GRACIOUS data logging templates for the environmental, health and safety assessment of nanomaterials*, Publications Office of the European Union, Luxembourg, 2019, DOI: 10.2760/142959.
- 37 M. L. Ostraat, K. C. Mills, K. A. Guzan and D. Murry, *Int. J. Nanomed.*, 2013, **8**, 7.
- 38 K. C. Mills, D. Murry, K. A. Guzan and M. L. Ostraat, *J. Nanopart. Res.*, 2014, **16**, 2219.
- 39 G. Roebben, K. Rasmussen, V. Kestens, T. Linsinger, H. Rauscher, H. Emons and H. Stamm, *J. Nanopart. Res.*, 2013, **15**, 1455.
- 40 S. Totaro, G. Cotogno, K. Rasmussen, F. Pianella, M. Roncaglia, H. Olsson, J. M. R. Sintes and H. P. Crutzen, *Regul. Toxicol. Pharmacol.*, 2016, **81**, 334–340.
- 41 European Commission, JRC Nanomaterials Repository, <https://ec.europa.eu/jrc/en/scientific-tool/jrc-nanomaterials-repository>, (accessed 10 July 2019).
- 42 OECD, Testing Programme of Manufactured Nanomaterials, <https://www.oecd.org/chemicalsafety/nanosafety/testing-programme-manufactured-nanomaterials.htm>, (accessed 10 July 2019).
- 43 ENanoMapper database, <https://data.enanmapper.net/>, (accessed October 2018).
- 44 European Commission, *Off. J. Eur. Union*, 2011, **L 275**, 38–40.
- 45 H. Rauscher, G. Roebben, A. Mech, N. Gibson, V. Kestens, T. P. J. Linsinger and J. R. Sintes, *An overview of concepts and terms used in the European Commission's definition of nanomaterial*, Publications Office of the European Union, Luxembourg, 2019, DOI: 10.2760/459136.
- 46 K. Guzan, K. Mills, V. Gupta, D. Murry, C. Scheier, D. Willis and M. Ostraat, *Comput. Sci. Discovery*, 2013, **6**, 014007.
- 47 K. Rasmussen, H. Rauscher, A. Mech, J. Riego Sintes, D. Gilliland, M. González, P. Kearns, K. Moss, M. Visser, M. Groenewold and E. A. J. Bleeker, *Regul. Toxicol. Pharmacol.*, 2018, **92**, 8–28.
- 48 C. Singh, S. Friedrichs, G. Ceccone, N. Gibson, K. A. Jensen, M. Levin, H. Goenaga Infante, D. Carlander and K. Rasmussen, *Cerium Dioxide, NM-211, NM-212, NM-213. Characterization and test item preparation*, Publications Office of the European Union, Luxembourg, 2014, DOI: 10.2788/80203.
- 49 EFSA Scientific Committee, A. Hardy, D. Benford, T. Halldorsson, *et al.*, *EFSA J.*, 2018, **16**, 5327.

