Data, Ontology and Harmonisation Needs for NSC & Projects

Following a series of consultation meetings with EU Nanosafey Cluster projects, and an open forum discussion in Brussels on Monday 25th January 2016, this short position paper reflects the current state of science with regards to nanosafety data and provides a needs assessment / call to action regarding the nanosafety research community's data, ontology and harmonization needs.

An expanded version, with the broader context is published in J. Nanomed Research, 2016, 3: 00070. DOI: 10.15406/jnmr.2016.03.00070 "Nanomaterial Ontologies for Nanosafety: A Rose by any Other Name…" by Eugenia Valsami-Jones, Iseult Lynch, and Costas A. Charitidis.

DATA/EXPERIMENTAL INTEGRITY/RELIABILITY

- The "*plethora of low-value results*" due to the lack of harmonized experimental protocols, problematic nanomaterials characterization and lack of reference materials is a problem that needs to be confronted.
- The "*plethora of low-value results*" and the published studies based on them, offer limited insight or can lead to distorted information with heavy repercussions to society.
- Need to move away from the mindset whereby it is "often better to be first and wrong than scooped and right" –Nature (2016) 529,256.doi:10.1038/529256a
- Experimental results are produced and stored in non-standardized forms.
- It is recognized that monitoring of Research Integrity cannot be properly done without a complete and tested method for inspecting data reliability, i.e. experimental reliability. Responsibility for data integrity needs to at the institutional and researcher level, and needs to be an integral part of data management.
- Experimental results in all scientific disciplines should not be just reported (and interpreted) with the *ad hoc* assumption of being free of errors or being unbiased.

In 2015 a research paper was published proving that 60% of publications in the field of Psychology cannot be reproduced (J. Bohannon, *Science* **349** (2015) 910-911). Also, in 2016 two publications revealed the disturbing fact of serious deficits in publications in the field of Biomedicine (S.A. Iqbal *et al*., *PLoS Biol*. **14** (2016) e1002333, C. Holman *et al*., *PLoS Biol*. **14** (2016) e1002331).

MEASURES

1. Databases

- The creation of databases has as a prerequisite the implementation of coherent experimental methods and materials used **through the implementation of unambiguous Standard Operation Procedures (SOPs)** and harmonized protocols, and linked via an agreed ontology in order for the results to be comparable.
- Open databases will also enhance the pace of research, since what is not useful for one researcher could be vital for someone else, as well as facilitating aggregation of data and

assessment of emerging patterns in the data e.g. across materials classes or between assay types.

- **Steps have been made for the systematic registry of nanomaterials characterization, environmental and health hazards assessment, high throughput and high content in a database infrastructure, with search capabilitie**s (eNanoMapper). Work is needed to add release, exposure and environmental fate aspects (with NanoFASE).
- Harmonization of existing databases is needed in order to join forces.
- Data must be stored in a standardized form.
- Databases should be supported by a commonly agreed domain ontology. This improves search quality and capabilities, but also enables integration of annotated data which are stored in different repositories.
- Tailored and user-friendly interfaces should be designed and implemented for different needs and usages. This includes explanations of data-related terminology (RDF, UDS) for experimentalists and intuitive flow processes for data flows from creation to curation and storage, written by a technical writer to avoid too much technicaljargon.
- User friendly tools should also be available for data preparation and upload, supporting many different import formats (such as ISA-Tab-Nano), custom spreadsheet templates and raw data files (such as microscopic images and high-throughput screening data)
- **Where possible, digital lab note-books should be integrated such that data management is directly linked to the data generation steps, rather than being an after-thought.**
- Integration and communication of a database with modelling and analysis tools allows exploitation of the data in the most efficient way, extraction of useful information and development of predictive mathematical models.

2. Open access data

- The EC Open Research Data Pilot initiative is a step towards the ability of the scientific community to validate results that appear in scientific publications and as a result a way to minimize sloppy science and inhibit research misconduct.
- The basic principle is to make research data open/visible, in order to facilitate validation of the results presented in scientific publications by any interested stakeholder.
- Access to raw experimental data (so called “underlying data”) must be given through electronic repositories (for example through OpenAIRE electronic infrastructure).
- Data storage in such repositories must be made in Discoverable, Accessible, Assessable-Intelligible, Usable beyond the original purpose for which it was collected and Interoperable.
- Such repositories must have state of the art equipment that ascertains the security of stored data for long periods (e.g. for 20 years or more), and allow the storage of several Tbytes for each user.

- In parallel to the above measures, the elaboration of Data Management Plans (DMP) must become an essential part of research. DMP is not mandatory yet; it applies only to the projects that have accepted to participate to Pilot Open Research Data initiative. According to this Pilot, DMP must be part of the deliverables at the beginning, midterm and end of Projects.
- **Once the NSC database is established, it should provide a menu of services (with associated costs) that future projects can buy-into, accompanied by standard texts to slot into the DMP at the proposal drafting stage, thereby ensuing projects build-in suitable costs for DM and providing a sustainable route to maintaining the database activities beyond the funding lifetime of the database project.**
- Data ownership issues will need to be addressed, including approaches such as data lisences, timed release of data to coincide with publications, etc.

3. Direct control

- Experimental data audits should be part of the *modus operanti* of all research laboratories, and institutional procedures and outcomes should accompany large datasets.
- Data audits are common for corporate biotechnology laboratories, but not for academic ones.
- Research entities should undergo independent audits of scientific data annually by certified public scientists, in much the same way as businesses and not-for-profit organizations are independently audited by certified financial accountants (J. L. Glick, *Ann. NY Acad. Sci.* **265**(1976)178–192).
- Quality Assurance audits must aim at eliminating: disorganized sample storage, inadequate data logging, variable experiments, unsecured data analysis, missed maintenance (M. Baker, *Nature* **529** (2016) 456-458).
- Research Integrity Offices, having the authority from the state, can be in charge of such data audits, with scientists of acknowledged record, experience and integrity.
- As first steps, data audits could be a mandatory part of the Quality System of laboratories or/and be a prerequisite for the completion of EU funded projects. Training to be provided to researchers at project initiation.
- Experimental instrument calibration (metrological services) should be performed regularly and supervised by an independent authority (e.g. National Metrological Institute). Calibrations should be made in a permanent base, by well-defined methods, i.e. traceable to National Metrological Standards (if applicable).
- **Ensure the production of high quality data of sufficient longevity and usefulness for research**. This will foster an integrated approach, e.g. in nanosafety assessment issues, characterized by the virtues of comparability, inclusivity and unanimity. Labs should be responsible for holding information on historical variability of cellular / biological assays within their organization.

- **Interlaboratory proficiency testing comparisons can be organized according to international standards (ISO, CEN)** to check competence of laboratories in performing measurements and proficiency in delivering accurate testing results, but also requires laboratories to know the historical variability of the assay in their facility.

4. Preventive measures against open access side effects (Against misconduct of misconduct elimination)

- Caution in the way that post publication peer review process is handled.
- Consider strengthening the role of the Code of Responsible research in nanoscience[1](currently includes "meaning, sustainability, precaution, inclusiveness, excellence, innovation and accountability") to include also QA and Open data aspects.
- This links also with recent debates (blogs[2], twitter etc.) regarding "data vultures" and "data parasites" following a poorly worded editorial on Data Sharing(Longo and Drazen, (2016) N Engl J Med 2016,374:276-277), which tries to address some of the issues by suggesting collaborative re-use of data but causes a controversy and significant backlash.
- "Don't let transparency damage science" (S. Lewandowsky and D. Bishop, *Nature* **529** (2016) 459-461)

FIVE DOUBLE-EDGED TOOLS

Legitimate tools of scholarly exchange can be weaponized.

Technique	**Use**	**Abuse**
Call for data	Permit the replication or inspection of analyses.	Impugn scientists' integrity (when data is already available); biased re-analyses.
Social-media posts	Highlight errors or questionable practices.	Stalk, libel, intimidate or harass.
Freedom-of-information requests	Reveal hidden conflicts of interest.	Launch a fishing expedition into private correspondence.
Call for retraction	Remove unethical or erroneous work from the literature.	Discredit inconvenient results.
Complaints to universities	Redress unethical conduct.	Damage reputation.

S. Lewandowsky and D. Bishop, *Nature***529** (2016) 459-461

[1]http://ec.europa.eu/research/science-society/document_library/pdf_06/nanocode-apr09_en.pdf
[2]http://www.thecogitoblog.com/blog/im-not-a-research-parasite-youre-a-data-vulture/