

NSC & ECHA Data Harmonisation Workshop

The EU workshop on Data Harmonisation, co-organised with the European Chemicals Agency, took place on Thursday October 18th at the University of Birmingham Brussels Office.

The overall aim of the workshop was to bring together EU coordinators of recently completed, ongoing and newly starting projects and other interested parties, including ECHA, to discuss accessibility to, and curation of, nanosafety data. A key objective of the workshop was to increase the NanoSafety Cluster (NSC) community's ability to pool data, enhance data sharing and collaborations within the NSC, and facilitate future data import into ECHA's [European Union Observatory for Nanomaterials \(EUON\)](#).

[Report of Workshop](#)

[Consensus proposal \(view slides\)](#)

The day began with an introduction from Iseult Lynch of the consensus view from the previous day's NSC Meeting. This noted the need to identify actions by which the NSC could support EUON and respond to feedback from stakeholders. The ways to encourage and incentivise open data sharing was debated, and how to drive a cultural change in seeing the value of data sharing as a path to enriching data and to reach the goal of a connected, accessible and usable NSC database. The following points were discussed:

- Existing/older datasets should not be discarded, but considerable work would be needed to link these to more modern data and knowledge.
- Plans for data entry and sharing should be agreed by multiple members of the community, and expert workshops could be convened to work through and curate large datasets.
- There is a need to disseminate existing templates and tools to make everyone aware of what exists and what they need to align to. To this end, a video, webinar or shared document could be produced that presents and explains these aspects. This could be shared with Christine Hendren in relation to her experience reviewing the JRC templates
- The data entry and curation approach needs to be agreed from the beginning of a project. This could be problematic, however, as templates are not defined far enough to be standards. Nonetheless, the best examples could be gathered as a starting resource which can be continued to be built upon and evolved. The IOM templates could be shared as such an example.
- The data templates could be reviewed at the same time DMPs are reviewed, with such activities falling within WG F.
- Ensuring protocols are linked to the data when uploaded is vital. This could be made a mandatory requirement, i.e. No protocol, no upload.
- Data providers need to include an interpretation of the data, with the caveat that this may adapt over time as knowledge and understanding evolves.
- Making data public is very important, though this has the potential for data misuse if incorrectly interpreted or applied.

[The EUON: collaboration with the NSC \(view slides\)](#)

The discussion was followed by a presentation from Abdel Sumrein, ECHA, on areas of potential collaboration between the NSC and the EUON and what was expected and needed for the EUON. This linchpin presentation highlighted the general aims and future of the observatory, and how the EUON plans to use data from research projects. The aim for the collaboration would be to have access to information from existing and past projects, to agree on how data is collected, to have a common management system, and to increase the visibility of NSC projects. It would also use existing information from REACH etc. and national registries, and have a light integration with the eNM database. The presentation also looked at Nanomaterials registration and description in a regulatory context.

[Breakout Session 1](#)

Abdel's presentation provided a springboard for the subsequent breakout sessions, which were designed to discuss the main databases currently available and how their data can be harmonised across database capture systems (including consideration of SOPs), as well as database ontology harmonisation and requirements for transfer to the EUON. The session began with a couple of stimulus presentations from Andrea Haase ([view slides](#)), looking at the experiences and lessons learned from COST Modena and NanoReg2 disparate datasets, and Tassos Papadiamantis ([view slides](#)) on the ModNanoTox experience from a literature curated dataset.

[Breakout Session 2](#)

Christine Hendren ([view slides](#)) joined via Skype to present the NanoInformatics Knowledge Commons (NIKC) approach, and Nina Jeliaskova ([view slides](#)) then presented an overview of the eNM/NanoREG/NanoReg2 database. The subsequent roundtable group discussions focussed on data formats and opportunities and difficulties in data harmonisation. This highlighted that significant progress is still needed, specifically in terms of data curation, in ensuring ongoing data quality and completeness. Much of this is currently being done retrospectively by running projects wanting to re-use datasets from previous projects, however, long term this is an unsustainable approach.

[Concluding remarks and decisions for way forward](#)

The process of harmonising the data produced by NSC projects is a very difficult task. Discussions across the 2 days had been very fruitful, but a clear way forwards and agreed consensus had not been reached and it is too early for EUON to take on some aspects. Preparation of the data would need to be done as a tiered approach; datasets would need to be prepared and updated, the data would then need to undergo a quality review process by which it is validated by data providers, reviewed at a NSC level by projects to agree the data quality criteria, and then curated. A single community database would also not be possible, but the same concept could be used across different databases to make them comparable. The possibility of a prize for the most FAIR and Open data was considered. This could involve a journal and be attached to the Young Scientist Forum, and act as an incentive to drive the cultural change in approach to data curation. Additionally, data training workshops on curation could be held, offering an accreditation upon completion.

The following points were also discussed:

- A group could be created at an NSC level that brings together all data curators to discuss these issues and agree a harmonised approach.
- The message would need to be that data curation is done as part of the data creation process, and not subsequently at the end of the project.
- Expert data curators could be brought in through NanoCommons, for example. However, this would require data to have been published and publication is always likely to come at the end of a project.
- Each project's dataset should undergo a process by which the accuracy of the data and the linked protocols are verified by an appointed member from that project. This, ideally, would be the content providers as this cannot be done by the EUON and would enable the curators to identify and work with the content providers.
- There needs to be further discussion on the common issues encountered when curating data, such as how to deal with outlying data and the need for these to be checked and verified. Thomas Exner could do a presentation on this topic at the next meeting.
- Different types of curation are required for different data uses, where the quality of the data depends on its purpose (e.g. regulators, risk assessment). Therefore, linking the data to such quality criteria could be something that is built into the database.
- A data curation Task Force could be set up with the deliverable to produce an agreed plan and explain the tiered approach. This may be more of an internal WG activity, but a defined timeline and deliverable would be necessary. Those involved could be drawn from the list of experts/expertise that will go on the NSC website.
- Managing expectations would be important. As data coming out of projects at the moment is not quite ready, the message should be clear on where the NSC is, what direction it is heading in, and the work it is doing to professionalise the data for the community.
 - e.g: a roadmap of the minimum steps required to get the data public. This would help to identify what is missing and how long it will take. Position this message correctly to focus on what can be achieved in the short term, to be honest about what is available, and the direction that is being taken.
 - Short Term Plans: Pick a popular, current topic of interest to the community and industry (e.g. reduction in animal testing) and summarise the state of the art and the NSC consensus in a digestible format that can go into a 2-pg summary for the newsletter. Suggestions for topics could also be received from other stakeholders and ECHA.
- The NSC coordination team and the Data Management WG will review and digest the outcome of the workshop and plan the next action in 2018.

Attendees

1	Eva Valsami-Jones	University of Birmingham (NSC Coordination Team)
2	Iseult Lynch	University of Birmingham (NSC Coordination Team)
3	Martine Bakker	RIVM
4	Anthony Bochon	ULB
5	Tom Carney	UoB
6	Costas Charitidis	NTUA
7	Damjana Drobne	University of Ljubljana
8	Maria Dusinska	NILU
9	Thomas Exner	Douglas Connect
10	Lucian Farcas	Douglas Connect
11	Wouter Fransman	TNO
12	Iris Garcia-Iglesias	UBU-ICCRAM
13	Camille de Garidel	CEREGE
14	Andreas Haase	BfR
15	Danail Hristozov	GD
16	Nina Jeliaskova	Idea Consult
17	Elias Koumoulos	NTUA
18	Vladimir Lobaskin	UCD
19	Marianne Matzke	NERC-CEH
20	Andrew Nelson	University of Leeds
21	Tassos Papadiamantis	University of Birmingham
22	Kai Paul	Blue Frog Scientific
23	Alexander Pogany	BMVIT
24	Dimiter Prodanov	IMEC
25	Peter Ritchie	IOM
26	Christa Schimpel	BioNanoNet
27	Tommaso Serchi	LIST
28	Claire Skentelbery	NIA
29	Neville Slack	CPI
30	Soco Vazquez Campos	LEITAT
31	Mar Viana	CSIC
32	Egon Willighagen	Maastricht University
33	Henrik Wolff	FIOH
34	Jorge Costa-David	European Commission
35	Jana Drbohlavova	European Commission
36	Georgios Katalagarianakis	European Commission
37	Erica Poot	European Commission
38	Abdelqader Sumrein	ECHA