SECTION 1 - User community

Europe has steadily increasing demands on the assessment of chemicals, drugs, cosmetics ingredients and nanomaterials to lead to safer products, resulting in a strong toxicology research community with sub-communities *e.g.* in the drug development, environmental, and nanomaterial areas. Recent changes in European law for animal testing and new demands for testing of lower-volume chemicals and nanomaterials have triggered large-scale research into alternative testing approaches. These activities not only produce new biological and mechanistic insights, but also large amounts of new data, which have to be managed and shared for re-usage to avoid unnecessary duplication of experiments and, in this way, reduce animal testing. The goal is that the combination of data from integrated *in vitro* and *in silico* approaches will support (ultimately personalized) risk/benefit health analysis, faster drug innovation, a fact-based perception of chemical safety, safe-by-design nanomaterials and sustainable and safe economies.

The European Union supports toxicological and risk assessment (RA) projects with various funding programs. Recently, large collections of data have been released, resulting from research clusters, such as SEURAT-1[1], the EU NanoSafety Cluster (NSC) with NANoREG, EU-ToxRisk[2], the NORMAN network, and the EU Innovative Medicines Initiative (IMI) funded projects related to drug toxicology, including eTOX[3]. Data from these and other projects are becoming available, sometimes as Open Data (*e.g.* NANoREG) and sometimes as FAIR data. An example of the latter is European REACH data, which has recently been made FAIR by the Cefic-LRI-funded project AMBIT-LRI.

However, there are a few opportunities for data handling that need to be taken (see two recent roadmaps[4,5]). Recent studies show how powerful the combination of toxicology information and omics data is[6,7], but to be able to obtain the statistical significance to draw these conclusions, data from the US and Japan had to be combined. In contrast to large data sets like DrugMatrix, ToxCast/Tox21, and TG-GATEs from these countries, data from European projects is not often sufficiently integrated. There are also signs that the community is going in the right direction, e.g. the aforementioned data sets from diXa, NANoREG and REACH. With respect to the European Chemical Industry, TNO and others have been involved in various other Cefic-LRI activities related to data management (AIMT-3, AIMT-4).

In addition, the IMI-funded eTOX project[8] has established data integration approaches e.g. to enable the development of QSARs relating chemical structures to *in vivo* toxicopathological outcomes. As such, the project also delivered databases and approaches to ontology development, text mining approaches, and approaches for prediction of drug metabolism and pharmacokinetic features. Moreover, in 2017 the OECD launched an online survey underscoring the fact that data integration for safety is even of global concern for ultimate RA. The aim of such a knowledge base would be the integration of eCHEMportal, IUCLID, and OECD QSAR toolbox in light of the development of adverse outcome pathways (AOPs) and associated infrastructures (AOPWiki)[9]. The ‘data integration struggle’ from various perspectives (omics, computational chemistry and more ‘conventional’ toxicological data within REACH and pharma industry setting) will get even worse.

SECTION 2 - Roadmap

The above initiatives are mainly driven from user communities themselves (chemical industry, health research funding agencies, pharma, Member State organisations like OECD), but ELIXIR can contribute strongly to the existing infrastructure projects from a bioinformatics perspective, servicing the toxicology users.

For future risk assessment paradigms solely based on human-derived models, and in this way of higher relevance for adverse effects[10], various data types will need to be integrated across the conventional boundaries of RA. This involves external exposure assessment (*e.g.* via workplace or environmental modelling and detection), internal exposure characterisation (ADME-T *e.g.* via modelling and detection), metabolism, molecular toxicology, and cell and systems biology as more mechanistic data, are integrated into RA and regulatory toxicology. There will not only be more but also more diverse data as e.g. internal exposure data may be inferred from biomonitoring data and/or physiologically-based toxicokinetic modelling all the way up to the active sites involved in the molecular initiating events of AOPs.

It is also relevant to mention the newly emerging concept of the exposome here. To complicate matters even further, the exposome aims at characterising lifetime exposure (not only to chemicals in the narrowest sense, but also dietary components, lifestyle factors, environmental exposures, etc) or at least during vulnerable periods of life (infancy, childhood, and old age), in relation to health outcome and also integrates epidemiology. This is one clear demonstration of the trend that the previously distinct areas of toxicology, drug and product design and personalized/precision medicine are moving closer together and data sharing will be increasingly necessary across these and neighbouring disciplines.

The toxicology community is large and well established and the current list of proposers only reflects a subset of a much larger community with a lot of Open collaboration. It has clear omics and knowledge management needs to accommodate the increasing wish to predict toxicology without animal testing (SEURAT-1, EU-ToxRisk, eTOX). Foreseeing this need for better infrastructures, the community has previously contacted ELIXIR for collaboration. Various domain specific projects exist that service the toxicology community with computational and database knowledge (OpenRiskNet, NanoCommons) that can translate ELIXIR knowledge to the respective communities. These infrastructure projects are the successors of research projects focusing on data management, including diXa[11], ToxBank[12] and eNanoMapper[13].

To benefit the research community, SMEs and industry, and even enable further future support to regulatory applications, we need to reach an inclusive ecosystem of data. The currently separate consortia from different toxicity-related and neighboring disciplines already work towards data and knowledge that is findable, accessible, interoperable, and reusable (FAIR)[14]. After all, these aspects are essential to efficient assess the risk of new compounds and materials, as well as combined risks of current stressors (e.g. under the exposome concept). To further accelerate these activities, more toxicology-related data and knowledge needs to be linked, such as on PBPK, biological pathways, metabolism, drug-response, omics (*biological identity*), chemical structures, QSARs, AOPs, REACH dossiers, etc. Simultaneously, an extension towards the human

(preclinical toxicology) discipline should be initiated, in which exposure data are combined with internal exposure data (e.g. data from the European Biomonitoring Initiative (HBM4EU) and environmental data from NORMAN) towards pathways of toxicity.
However, to reach such interoperable toxicology, resources need to be integrated better. Despite the work of many projects, their FAIR features can still be improved. Example key needs that are unresolved, overlap with other ELIXIR communities but require innovative integration approaches for RA needs include (see the aforementioned roadmaps by Haase and Karcher[4,5]):

- chemical structure interoperability challenges (e.g. links to ELIXIR metabolomics community)
- metadata, open standards (e.g. links to ELIXIR interoperability platform and TeSS)
- continued ontology development (e.g. links to ELIXIR interoperability platform)
- interoperable computation (e.g. links to the Galaxy and bio.tools communities)
- interactions with other the ELIXIR core resource
- interactions with other communities, including nanomedicine and health

The concrete steps forward proposed in this roadmap include:

1. leverage from Open solutions (models, ontologies, educational material, standards etc) developed by past and ongoing toxicology and ELIXIR projects
2. connect more closely with the core ELIXIR resources (FAIR data, database interoperability, etc)
3. strengthen and connect the inclusive communities that have evolved over the past few years (OpenTox, eNanoMapper, diXa, OpenRiskNet, REACH, NORMAN)
4. develop open community standards to support common interest (ontologies, application programming interfaces, data formats)

Very specifically, we will continue to grow the list of involved toxicology research groups, projects, and Node activities. A joint meeting to select the key priorities and use cases (*e.g.* toxicogenomics, AOPs) is planned as well as a positioning paper.
By bootstrapping from Open Science approaches developed in aforementioned projects, the new community will focus on mutual benefit, an open and inclusive community, solving practical community problems. The goal is not to design a domain specific solution, but a pragmatic approach that provides FAIR and Open solutions from month one, allowing all toxicology and neighboring communities, to benefit from these harmonized solutions. The inclusive community will involve the existing sub-communities in pharmaceuticals, *e.g.* from eTOX, transQST, and eTRANSAFE, cosmetic ingredients, *e.g.* from SEURAT-1, high and low-volume chemicals, *e.g.* from different Cefic-LRI projects, and REACH supporters, and nanomaterials, via the NSC, building on shared needs and community solutions and strongly aligned to other ELIXIR communities. Open licensing and interfaces (ontologies, standards, formats) will encourage new solutions and collaborations, which will be accessible to any organisation and every project within and outside Europe. This will allow close interoperability with toxicology communities outside Europe that also use open approaches, while at the same time allowing compatibility with closed approaches too. This dual model has been demonstrated successfully in recent projects. A prioritized roadmap is essential; the label "ELIXIR Community" enables us to set priorities at a level above the individual projects.

Existing component that will give this community an initial boost, include, software (e.g. AMBIT with OpenTox API[15]), databases (*e.g.* diXa, eNanoMapper, AMBIT-LRI), ontologies (*e.g.* the eNanoMapper ontology[16], AOP ontology[17]), interoperability concepts (*e.g.* annotation of OpenAPIs, in collaboration with bio.tools), teaching/education material (Bioschemas annotation of outreach activities, in collaboration with TeSS), and virtual infrastructures (*e.g.* OpenRiskNet Virtual Research Environment, the NORMAN Digital Sample Freezing Platform). However, each of these approaches would benefit from integration in the ELIXIR Platforms (see examples in Table 1) and with Core Resources. Various existing ELIXIR Communities need similar solutions, *e.g.* for chemical structure handling, toxicology needs proteomics and metabolomics, toxicology involves human data, and ecotoxicology has significant impact on crops and health.

Table 1: Examples of existing and anticipated collaboration.

	Existing collaboration / Reuse	**Anticipated collaboration**
Tools	Semantic annotation of services (e.g. bio.tools)	Alignment met BioContainers of toxicology workflow efforts (e.g. OpenRiskNet)
Data	diXa was co-developed by the CHEMBL-EBI team	Better adoption of the core resources
Compute		Aligning handling of human data in computation
Interoperability	FAIR and RDM standards have already been adopted by various projects	Registries of toxicology tools need integration with FAIRsharing; There is a huge identifier mapping service need
Training	Bioschemas annotated tutorials	Several projects have training tasks, listing in TeSS would be great

The following projects have been adopting and integrating FAIR toxicology concepts, but need integration with ELIXIR Platforms and Communities: eTOX, NanoCommons (NSC), EU-ToxRisk (69 partners), OpenRiskNet (11 partners), OpenTox Foundation, Open PHACTS Foundation, and the diXa platform. Many other projects have a specific scientific focus but also need integration. A non-exhaustive list is ACEnano, SmartNanoTox, HeCaToS, NewGeneris, EnviroGenoMarkers, Exposomics, HELIX, ASAT, PATROLS, and HEALS.

Companies and organisations that will profit from this Community either as users or as providers of services on top of the infrastructure, include ECHA (FI), Douglas Connect (CH), IdeaConsult Ltd (BG), Misvik Biology (FI), TNO (NL), and SweTox (SE). Industries showed a strong interest in toxicology, demonstrated by their activities: Cosmetics Europe was participant in the SEURAT-1 cluster; chemical industries (NIA) are participant of the NSC; chemical branch organisation (Cefic) funds LRI projects around toxicology; and, pharmaceutical industries funds toxicology research via IMI projects like eTOX and Open PHACTS. The Research Data Alliance organized a workshop recently about integration of toxicogenomics resources[18], and collaboration with international organizations has already been established with, for example, the CompTox Chemistry Dashboard team of the U.S.A. EPA[19].

References

1. Gocht, T., 2014. The SEURAT-1 approach towards animal free human safety assessment. ALTEX.
2. Daneshian, M., Kamp, H., Hengstler, J., Leist, M., van de Water, B., 2016. Highlight report: Launch of a large integrated European in vitro toxicology project: EU-ToxRisk. Archives of Toxicology 90, 1021–1024.
3. Sanz, F., Pognan, F., Steger-Hartmann, T., Díaz, C., Cases, M., Pastor, M., Marc, P., Wichard, J., Briggs, K., Watson, D.K., Kleinöder, T., Yang, C., Amberg, A., Beaumont, M., Brookes, A.J., Brunak, S., Cronin, M.T.D., Ecker, G.F., Escher, S., Greene, N., Guzmán, A., Hersey, A., Jacques, P., Lammens, L., Mestres, J., Muster, W., Northeved, H., Pinches, M., Saiz, J., Sajot, N., Valencia, A., van der Lei, J., Vermeulen, N.P.E., Vock, E., Wolber, G., Zamora, I., 2017. Legacy data sharing to improve drug safety assessment: the eTOX project. Nature Reviews Drug Discovery 16, 811–812.
4. Haase, A., 2017. EU US Roadmap Nanoinformatics 2030, https://www.nanosafetycluster.eu/Nanoinformatics2030.html.
5. Karcher, S., Willighagen, E.L., Rumble, J., Ehrhart, F., Evelo, C.T., Fritts, M., Gaheen, S., Harper, S.L., Hoover, M.D., Jeliazkova, N., Lewinski, N., Marchese Robinson, R.L., Mills, K.C., Mustad, A.P., Thomas, D.G., Tsiliki, G., Hendren, C.O., 2017. Integration among databases and data sets to support productive nanotechnology: Challenges and recommendations. NanoImpact.
6. Kohonen, P., Parkkinen, J.A., Willighagen, E.L., Ceder, R., Wennerberg, K., Kaski, S., Grafström, R.C., 2017. A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. Nature Communications 8, 15932.
7. Luechtefeld, T., 2016. Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008-2014. ALTEX.
8. Cases, M., Briggs, K., Steger-Hartmann, T., Pognan, F., Marc, P., Kleinöder, T., Schwab, C., Pastor, M., Wichard, J., Sanz, F., 2014. The eTOX Data-Sharing Project to Advance in Silico Drug-Induced Toxicity Prediction. International Journal of Molecular Sciences 15, 21136–21154.
9. Leist, M., Ghallab, A., Graepel, R., Marchan, R., Hassan, R., Bennekou, S.H., Limonciel, A., Vinken, M., Schildknecht, S., Waldmann, T., Danen, E., van Ravenzwaay, B., Kamp, H., Gardner, I., Godoy, P., Bois, F.Y., Braeuning, A., Reif, R., Oesch, F., Drasdo, D., Höhme, S., Schwarz, M., Hartung, T., Braunbeck, T., Beltman, J., Vrieling, H., Sanz, F., Forsby, A., Gadaleta, D., Fisher, C., Kelm, J., Fluri, D., Ecker, G., Zdrazil, B., Terron, A., Jennings, P., van der Burg, B., Dooley, S., Meijer, A.H., Willighagen, E., Martens, M., Evelo, C., Mombelli, E., Taboureau, O., Mantovani, A., Hardy, B., Koch, B., Escher, S., van Thriel, C., Cadenas, C., Kroese, D., van de Water, B., Hengstler, J.G., 2017. Adverse outcome pathways: opportunities, limitations and open questions. Archives of Toxicology.

10. Nymark, P., Rieswijk, L., Ehrhart, F., Jeliazkova, N., Tsiliki, G., Sarimveis, H., Evelo, C.T., Hongisto, V., Kohonen, P., Willighagen, E., Grafström, R.C., 2017. A data fusion pipeline for generating and enriching Adverse Outcome Pathway descriptions. Toxicological Sciences.
11. Hendrickx, D.M., Aerts, H.J.W.L., Caiment, F., Clark, D., Ebbels, T.M.D., Evelo, C.T., Gmuender, H., Hebels, D.G.A.J., Herwig, R., Hescheler, J., Jennen, D.G.J., Jetten, M.J.A., Kanterakis, S., Keun, H.C., Matser, V., Overington, J.P., Pilicheva, E., Sarkans, U., Segura-Lepe, M.P., Sotiriadou, I., Wittenberger, T., Wittwehr, C., Zanzi, A., Kleinjans, J.C.S., 2014. diXa: a data infrastructure for chemical safety assessment. Bioinformatics 31, 1505–1507.
12. Kohonen, P., Benfenati, E., Bower, D., Ceder, R., Crump, M., Cross, K., Grafström, R.C., Healy, L., Helma, C., Jeliazkova, N., Jeliazkov, V., Maggioni, S., Miller, S., Myatt, G., Rautenberg, M., Stacey, G., Willighagen, E., Wiseman, J., Hardy, B., 2013. The ToxBank Data Warehouse: Supporting the Replacement of In Vivo Repeated Dose Systemic Toxicity Testing. Molecular Informatics 32, 47–63.
13. Jeliazkova, N., Chomenidis, C., Doganis, P., Fadeel, B., Grafström, R., Hardy, B., Hastings, J., Hegi, M., Jeliazkov, V., Kochev, N., Kohonen, P., Munteanu, C.R., Sarimveis, H., Smeets, B., Sopasakis, P., Tsiliki, G., Vorgrimmler, D., Willighagen, E., 2015. The eNanoMapper database for nanomaterial safety information. Beilstein Journal of Nanotechnology 6, 1609–1634.
14. Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L.B. da S., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3, sdata201618+.
15. Jeliazkova, N., Jeliazkov, V., 2011. AMBIT RESTful web services: an implementation of the OpenTox application programming interface. Journal of Cheminformatics 3, 18+.
16. Hastings, J., Jeliazkova, N., Owen, G., Tsiliki, G., Munteanu, C.R., Steinbeck, C., Willighagen, E., 2015. eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment. Journal of Biomedical Semantics 6.
17. Burgoon, L.D., 2017. The AOPOntology: A Semantic Artificial Intelligence Tool for Predictive Toxicology. Applied In Vitro Toxicology.
18. Hendrickx, D.M., Boyles, R.R., Kleinjans, J.C.S., Dearry, A., 2014. Workshop report: Identifying opportunities for global integration of toxicogenomics databases, 26–27 June 2013, Research Triangle Park, NC, USA. Archives of Toxicology 88, 2323–2332.
19. Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S., Richard, A.M., 2017. The CompTox

Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics 9.