OPINION ARTICLE

REVISED ELIXIR and Toxicology: a community in development [version 2; peer review: 2 approved]

Marvin Martens[1], Rob Stierum[2], Emma L. Schymanski[3], Chris T. Evelo[1,4], Reza Aalizadeh[5], Hristo Aladjov[6], Kasia Arturi[7], Karine Audouze[8], Pavel Babica[9], Karel Berka[10], Jos Bessems[11], Ludek Blaha[9], Evan E. Bolton[12], Montserrat Cases[13], Dimitrios E. Damalas[5], Kirtan Dave[14], Marco Dilger[15], Thomas Exner[16], Daan P. Geerke[17], Roland Grafström[18,19], Alasdair Gray[20], John M. Hancock[21], Henner Hollert[22], Nina Jeliazkova[23], Danyel Jennen[24], Fabien Jourdan[25,26], Pascal Kahlem[27], Jana Klanova[9], Jos Kleinjans[24], Todor Kondic[3], Boï Kone[28], Iseult Lynch[29], Uko Maran[30], Sergio Martinez Cuesta[31], Hervé Ménager[32,33], Steffen Neumann[34], Penny Nymark[19], Herbert Oberacher[35], Noelia Ramirez[36], Sylvie Remy[11], Philippe Rocca-Serra[37], Reza M. Salek[38], Brett Sallach[39], Susanna-Assunta Sansone[37], Ferran Sanz[40], Haralambos Sarimveis[41], Sirarat Sarntivijai[21], Tobias Schulze[42], Jaroslav Slobodnik[43], Ola Spjuth[44], Jonathan Tedds[21], Nikolaos Thomaidis[5], Ralf J.M. Weber[45], Gerard J.P. van Westen[46], Craig E. Wheelock[47,48], Antony J. Williams[49], Hilda Witters[11], Barbara Zdrazil[50], Anže Županič[51], Egon L. Willighagen[1]

[1]Department of Bioinformatics - BiGCaT, Maastricht University, Maastricht, 6229 ER, The Netherlands
[2]Risk Analysis for Products In Development (RAPID), Netherlands Organisation for applied scientific research TNO, Utrecht, 3584 CB, The Netherlands
[3]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, 4367, Luxembourg
[4]Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, 6229 EN, The Netherlands
[5]Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Athens, 15771, Greece
[6]Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, 1113, Bulgaria
[7]Department Environmental Chemistry, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, 8600, Switzerland
[8]Université de Paris, Paris, F-75006, France
[9]RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic
[10]Department of Physical Chemistry, Palacky University Olomouc, Olomouc, 77146, Czech Republic
[11]Unit Health, VITO, Mol, 2400, Belgium
[12]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
[13]Chemotargets SL, Barcelona, 08028, Spain
[14]School of Science, GSFC University, Gujarat, 391750, India
[15]Forschungs- und Beratungsinstitut Gefahrstoffe (FoBiG) GmbH, Freiburg im Breisgau, 79106, Germany
[16]Seven Past Nine, Cerknica, 1380, Slovenia
[17]AIMMS Division of Molecular Toxicology, Vrije Universiteit, Amsterdam, 1081 HZ, The Netherlands
[18]Department of Toxicology, Misvik Biology, Turku, 20520, Finland
[19]Institute of Environmental Medicine, Karolinska Institute, Stockholm, 17177, Sweden

[20]Department of Computer Science, Heriot-Watt University, Edinburgh, UK
[21]ELIXIR Hub, Wellcome Genome Campus, Cambridge, CB10 1SD, UK
[22]Department Evolutionary Ecology & Environmental Toxicology (E3T), Goethe-University, Frankfurt, D-60438, Germany
[23]Ideaconsult Ltd., Sofia, Bulgaria
[24]Department of Toxicogenomics, Maastricht University, Maastricht, 6200 MD, The Netherlands
[25]MetaboHUB, French metabolomics infrastructure in Metabolomics and Fluxomics, Toulouse, France
[26]Toxalim (Research Centre in Food Toxicology), Université de Toulouse, Toulouse, France
[27]Scientific Network Management SL, Barcelona, 08015, Spain
[28]Faculty of Pharmacy, Malaria Research and Training Center, Bamako, BP:1805, Mali
[29]School of Geography, Earth and Environmental Sciences, University of Birmingham, UK, Birmingham, B15 2TT, UK
[30]Institute of Chemistry, University of Tartu, Tartu, 50411, Estonia
[31]Data Sciences and Quantitative Biology, Discovery Sciences, AstraZeneca, Cambridge, UK
[32]Institut Français de Bioinformatique, Evry, F-91000, France
[33]Bioinformatics and Biostatistics Hub, Institut Pasteur, Paris, F-75015, France
[34]Research group Bioinformatics and Scientific Data, Leibniz Institute of Plant Biochemistry, Halle, 06120, Germany
[35]Institute of Legal Medicine and Core Facility Metabolomics, Medical University of Innsbruck, Innsbruck, A-6020, Austria
[36]Institut d'Investigacio Sanitaria Pere Virgili-Universitat Rovira i Virgili, Tarragona, 43007, Spain
[37]Data Readiness Group, Department of Engineering Science, University of Oxford, Oxford, UK
[38]International Agency for Research on Cancer, World Health Organisation, Lyon, 69372, France
[39]Department of Environment and Geography, University of York, UK, York, YO10 5NG, UK
[40]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, 08003, Spain
[41]National Technical University of Athens, Athens, 15780, Greece
[42]Helmholtz Centre for Environmental Research - UFZ, Leipzig, 04318, Germany
[43]Environmental Institute, Koš, 97241, Slovakia
[44]Department of Pharmaceutical Biosciences and Science for Life Laboratory, Uppsala University, Uppsala, SE-75124, Sweden
[45]School of Biosciences, University of Birmingham, UK, Birmingham, B15 2TT, UK
[46]Division of Drug Discovery and Safety, Leiden Academic Center for Drug Research, Leiden, 2333 CC, The Netherlands
[47]Department of Respiratory Medicine and Allergy, Karolinska University Hospital, Stockholm SE-141-86, Sweden
[48]Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, 17177, Sweden
[49]Center for Computational Toxicology and Exposure, United States Environmental Protection Agency, Research Triangle Park, NC 27711, USA
[50]Department of Pharmaceutical Sciences, University of Vienna, Vienna, 1090, Austria
[51]Department Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, 1000, Slovenia

v2 **First published:** 08 Nov 2021, **10**(ELIXIR):1129
https://doi.org/10.12688/f1000research.74502.1
Latest published: 03 Oct 2023, **10**(ELIXIR):1129
https://doi.org/10.12688/f1000research.74502.2

Abstract

Toxicology has been an active research field for many decades, with academic, industrial and government involvement. Modern omics and computational approaches are changing the field, from merely disease-specific observational models into target-specific predictive models. Traditionally, toxicology has strong links with other fields such as biology, chemistry, pharmacology, and medicine. With the rise of synthetic and new engineered materials, alongside ongoing prioritisation needs in chemical risk assessment for existing chemicals, early predictive evaluations are becoming of utmost importance to both scientific and regulatory purposes. ELIXIR is an intergovernmental organisation that brings together life science resources from across Europe. To coordinate the linkage of various

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 03 Oct 2023	✓ view	
version 1 08 Nov 2021	? view	✓ view

1. **Christopher Southan**, University of Edinburgh, Edinburgh, UK

life science efforts around modern predictive toxicology, the establishment of a new ELIXIR Community is seen as instrumental. In the past few years, joint efforts, building on incidental overlap, have been piloted in the context of ELIXIR. For example, the EU-ToxRisk, diXa, HeCaToS, transQST, and the nanotoxicology community have worked with the ELIXIR TeSS, Bioschemas, and Compute Platforms and activities. In 2018, a core group of interested parties wrote a proposal, outlining a sketch of what this new ELIXIR Toxicology Community would look like. A recent workshop (held September 30th to October 1st, 2020) extended this into an ELIXIR Toxicology roadmap and a shortlist of limited investment-high gain collaborations to give body to this new community. This Whitepaper outlines the results of these efforts and defines our vision of the ELIXIR Toxicology Community and how it complements other ELIXIR activities.

Keywords
Toxicology, ELIXIR, interoperability, FAIR

Medicines Discovery Catapult, Macclesfield, UK

2. **Marta Eide**, University of Bergen, Bergen, Norway
Centre for Digital Life Norway, Bergen, Norway

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the ELIXIR gateway.

This article is included in the Nanoscience & Nanotechnology gateway.

This article is included in the Nanotoxicology collection.

This article is included in the ELIXIR Articles and Documents collection.

Corresponding authors: Marvin Martens (marvin.martens@maastrichtuniversity.nl), Egon L. Willighagen (egon.willighagen@maastrichtuniversity.nl)

Author roles: Martens M: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Stierum R**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Schymanski EL**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Evelo CT**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Aalizadeh R**: Writing – Review & Editing; **Aladjov H**: Writing – Review & Editing; **Arturi K**: Writing – Review & Editing; **Audouze K**: Writing – Review & Editing; **Babica P**: Writing – Review & Editing; **Berka K**: Writing – Review & Editing; **Bessems J**: Writing – Original Draft Preparation, Writing – Review & Editing; **Blaha L**: Writing – Review & Editing; **Bolton EE**: Writing – Review & Editing; **Cases M**: Writing – Review & Editing; **Damalas DE**: Writing – Review & Editing; **Dave K**: Writing – Review & Editing; **Dilger M**: Writing – Review & Editing; **Exner T**: Writing – Original Draft Preparation, Writing – Review & Editing; **Geerke DP**: Writing – Original Draft Preparation, Writing – Review & Editing; **Grafström R**: Writing – Original Draft Preparation, Writing – Review & Editing; **Gray A**: Writing – Review & Editing; **Hancock JM**: Writing – Review & Editing; **Hollert H**: Writing – Review & Editing; **Jeliazkova N**: Writing – Original Draft Preparation, Writing – Review & Editing; **Jennen D**: Writing – Original Draft Preparation, Writing – Review & Editing; **Jourdan F**: Writing – Review & Editing; **Kahlem P**: Writing – Review & Editing; **Klanova J**: Writing – Review & Editing; **Kleinjans J**: Writing – Original Draft Preparation, Writing – Review & Editing; **Kondic T**: Writing – Review & Editing; **Kone B**: Writing – Review & Editing; **Lynch I**: Writing – Original Draft Preparation, Writing – Review & Editing; **Maran U**: Writing – Review & Editing; **Martinez Cuesta S**: Writing – Review & Editing; **Ménager H**: Writing – Original Draft Preparation, Writing – Review & Editing; **Neumann S**: Writing – Review & Editing; **Nymark P**: Writing – Original Draft Preparation, Writing – Review & Editing; **Oberacher H**: Writing – Review & Editing; **Ramirez N**: Writing – Review & Editing; **Remy S**: Writing – Review & Editing; **Rocca-Serra P**: Writing – Review & Editing; **Salek RM**: Writing – Review & Editing; **Sallach B**: Writing – Review & Editing; **Sansone SA**: Writing – Review & Editing; **Sanz F**: Writing – Review & Editing; **Sarimveis H**: Writing – Original Draft Preparation, Writing – Review & Editing; **Sarntivijai S**: Writing – Review & Editing; **Schulze T**: Writing – Review & Editing; **Slobodnik J**: Writing – Review & Editing; **Spjuth O**: Writing – Original Draft Preparation, Writing – Review & Editing; **Tedds J**: Writing – Review & Editing; **Thomaidis N**: Writing – Review & Editing; **Weber RJM**: Writing – Review & Editing; **van Westen GJP**: Writing – Review & Editing; **Wheelock CE**: Writing – Review & Editing; **Williams AJ**: Writing – Review & Editing; **Witters H**: Writing – Review & Editing; **Zdrazil B**: Writing – Review & Editing; **Županič A**: Writing – Review & Editing; **Willighagen EL**: Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work received funding from the European Union's Horizon 2020 research infrastructure programme via the OpenRiskNet project under grant agreement No. 731075. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681002 (EU-ToxRisk), grant agreement No. 814572 (NanoSolveIT), grant agreement No. 825712 (OBERON), grant agreement No. 733032 (HBM4EU), grant agreement No. 953183 (HARMLESS), grant agreement No. 814401 (Gov4Nano), grant agreement No. 896141 (NTS-EXPOSURE), grant agreement No. 814425 (RiskGONE), grant agreement No. 731032 (NanoCommons), grant agreement No. 825489 (GOLIATH), and grant agreement No. 814426 (NanoInformaTIX). This work is supported by Dutch Governmental TNO Research Cooperation Funds. This work is supported by the European Union's Horizon 2020 programme under the Marie Skłodowska-Curie grant agreement No. 859891 (PRORISK). Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization. This work was supported by the Swedish Fund for Research Without Animal Experiments under grant number N2020-0005. This work was supported by the European Union,'s LIFE program LIFE-APEX, Grant agreement ID: LIFE17 ENV/SK/000355. This study was funded by the German Environment Agency within the PHION project (grant number 3718 674150). The Data Readiness Group is supported, in this ELIXIR Community, by the H2020 Precision Toxicology project (H2020-EU 965406), the Wellcome ISA-InterMine project (208381/A/17/Z), and the Wellcome FAIRsharing project (212930/Z/18/Z). This work was supported by CEFIC-LRI EEM9.3-IC/EEM9.4: Linking LRI Ambit chemoinformatic system with the IUCLID substance database to support read-across of substance endpoint data and category formation This work was supported by the French Ministry of Research and National Research Agency as part of the French MetaboHUB, the national metabolomics and fluxomics infrastructure (Grant ANR-INBS-0010). NFDI4Chem is supported by DFG under project number 441958208 SN acknowledges BMBF funding under grant number 031L0107. ELS acknowledges funding support from the Luxembourg National Research Fund (FNR) for project A18/BM/12341006. NR's research is funded by a Miguel Servet contract (CO19/00060) from Instituto de Salud Carlos III, cofinanced by the European Union. UM (Uni. of Tartu) is grateful for support to Ministry of Education and Research, Republic of Estonia through the Estonian Research Council (grant number PRG1509) and to European Union European Regional Development Fund through Foundation Archimedes (grant number TK143, Centre of Excellence in Molecular Cell Engineering). The work of EB was supported by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Martens M, Stierum R, Schymanski EL *et al.* **ELIXIR and Toxicology: a community in development [version 2; peer review: 2 approved]** F1000Research 2023, **10**(ELIXIR):1129 https://doi.org/10.12688/f1000research.74502.2

First published: 08 Nov 2021, **10**(ELIXIR):1129 https://doi.org/10.12688/f1000research.74502.1

REVISED **Amendments from Version 1**

We updated the manuscript based on reviewers' comments, and we made some minor textual changes and grammar corrections. A new section has been added on how to join the community, including a reference to the toxicology community webpage. Also, funding information was improved, putting grants from the same funders together instead of separate statements, and two references were added to the outputs of community joint activities.

Any further responses from the reviewers can be found at the end of the article

Introduction of the ELIXIR Toxicology Community

Toxicology as a field tries to understand the negative consequences that may arise from the interactions of chemicals with living organisms. This ELIXIR (European life-sciences Infrastructure for biological Information)[1] Community will concentrate on the focus areas of ELIXIR, including the protection of human, animal, and environmental health. There are several chemical and biological "interoperability" issues key to the toxicology field that translate into data interoperability issues. These include the connection between the action and activity of a particular chemical compound to its effective amount available at a biological target (the link between toxicodynamics and toxicokinetics). Typically, this is also a link between biological data analysis (including large-scale multi-omics) and kinetic modelling. Other examples include interactions between a compound and its target (a protein, nucleic sequence, or membrane structure for instance). This is primarily based on the interplay between chemistry (the chemical structure and, for example, its related properties in terms of functional groups, charge, shape, and related binding affinity) and biochemistry (like biomolecular 3D structures). Also, mixture toxicity needs to be considered as combinations of chemicals with synergistic or antagonistic behaviour, or a combination thereof. While chemicals with similar modes of action may act in terms of concentration addition, those with different modes may rather act according to independent action.[2,3] Substances with low toxicity may interact in concentration addition rather than as excess toxicity drivers of one compound. Often there is a need to translate the knowledge about one compound into knowledge about other compounds, where approaches like quantitative structure-activity relationships (QSARs) help to elucidate this knowledge and predict required property and toxicity, and in more general when "read-across approaches" come into play that are based on chemical data only, biological data only, or are hybrid.[4,5] These again need detailed information about the relationships between related chemical compounds, specific properties, and toxicological endpoints and, when considering chemical-biological interactions, details regarding both the chemical structures and adequate descriptions of the biological targets. Since toxicological endpoints can be represented in a myriad of ways, the toxicological effect data are often scattered over multiple repositories and databases hosting different types of data; *i.e.*, chemical structures, toxicity data (*in vivo* and *in vitro*), biological target details, and omics data. This is not a problem, per se, but the separation and segregation make the data difficult to find and connect. Currently, many of these deposition databases (where datasets can be archived) do not provide adequate descriptions regarding typical toxicological study designs and parameters, quality control, data acceptance criteria, or even clear identification of the compounds tested. This all adds to the need for an adequate FAIRification process, to make toxicology data more Findable, Accessible, Interoperable and Reusable (FAIR - see Go-FAIR).[6,7]

The field of toxicology would certainly benefit from clear, standardized guidelines for data capture, approaches to integrate and connect across multiple databases, clear data licensing for these data repositories, and tools that support accessing the data (as being developed by the ELIXIR Converge data brokerage work). Existing study capture tools can be extended with templates defined for toxicology, which end up in a central place (*e.g.*, Biosamples[8]) with links to omics and other data distributed over technology-specific deposition databases and BioStudies.[9] Relevant portals (*e.g.*, FAIRsharing.org[10,11]) should then be able to identify these studies, also linking to existing but scattered toxicology databases.

Risk assessment, consisting of hazard identification, characterisation, and risk evaluation in relation to exposure, involves expert opinions based on discussions and data interpretation. Streamlining this process is important also because of the huge number of chemicals known and produced in chemistry, biotechnology, or food production, with over 350,000 chemicals documented on the global market.[12] However, expert evaluation is and will remain crucial and calls for extensive data provenance both for the actual data (*e.g.*, how it was measured, where it was published, and whom to give credit for it), and for the risk assessments themselves. Interestingly this process stumbles on problems and solutions that have much in common with other fields, where part of the data needs to be hidden and other parts can be publicly accessed, as observed in pilot approaches in genetics, or patient data repositories.

Toxicology, while established as a discipline, is also rapidly developing in areas, where, for example, new molecular methods to describe the adverse outcome of exposure to a toxic substance are not yet fully established. This offers opportunities for integration in systems biology approaches building on molecular pathway descriptions that benefit from

modern network biology approaches. Compound profiling, where, for example, using omics methods to characterize the effect of compounds on cells, is generally useful for categorizing compounds, or for the development of predictive pathway models. For such studies, the availability of high-quality annotations for compounds is of paramount importance to enable the use of omics profiles in toxicology.

Toxicology is also an applied field. There are important applications of toxicity tests in the regulatory field and large amounts of data are collected for that purpose, often for different (governmental) agencies. Making this data more "Findable" and "Reusable" is often seen as an important way to reduce both animal testing and the cost of registration of new compounds. If some data is not available to the public due to ownership by companies or other constraints, indexes should be developed to enable the use of this data in an aggregated form (such as the SPIN index of the Swedish Chemicals Agency (KEMI) and other Nordic chemical agencies). Typically, regulatory use requires very precise and rigid descriptions of protocols, including reporting methods. On the other hand, better insights into the toxicological mechanisms, including interactions between chemicals, benefit from more innovative research methods (*e.g.*, single-cell omics, induced stem cell applications, and imaging methods) and from the creative development of new analysis methods. While it is beneficial to combine results from both types of approaches (termed New Approach Methods (NAM) in the regulatory field), the corresponding interoperability issues are often quite different.

User community

Europe is steadily increasing demands on the risk assessment of chemicals, drugs, cosmetics ingredients and nanomaterials to lead to safer products, resulting in a strong toxicology research community with sub-communities in, for example, the drug development, environmental, nanomaterial and rare-disease 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 must 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, safer drug innovation with fewer needs to withdraw after registration, a fact-based perception of chemical safety, safe-by-design nanomaterials, and sustainable and safe economies.

The European Union supports toxicological and risk assessment projects with various funding programs. Recently, large collections of data have been released, resulting from research clusters, such as SEURAT-1,[13] the EU NanoSafety Cluster (NSC) with NANoREG, EU-ToxRisk,[14] the NORMAN Network, and the EU Innovative Medicines Initiative (IMI) funded projects related to drug toxicology, including eTOX,[15] eTRANSAFE,[16] and Eurion.[17] 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.[18] Furthermore, the FAIRplus project is collaborating with eTOX on making their data more FAIR, and the new Precision Toxicology will develop a data commons following the FAIR principles.

However, there are a few opportunities for data handling that need to be taken.[19,20] Recent studies show how powerful the combination of toxicology information and omics data is,[21,22] 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,[23] ToxCast/Tox21,[24] and TG-GATEs[25] from these countries, data from European projects is often not sufficiently integrated. Luckily, there are signs that the community is going in the right direction, *e.g.*, the aforementioned data integration by diXa,[26] NANoREG and REACH.[27] With respect to the European Chemical Industry, various authors have been involved in other Cefic-LRI activities related to data management (AIMT-3, AIMT-4).

In addition, the eTOX project[28] 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 the prediction of drug metabolism and pharmacokinetic features. Moreover, in 2017 the Organisation for Economic Co-operation and Development (OECD) performed an online survey underscoring the fact that data integration for safety is of global concern for ultimate risk assessment. The purpose of the resulting knowledge base is the integration of eCHEMportal (The Global Portal to Information on Chemical Substances[29]), IUCLID (International Uniform ChemicaL Information Database[30]), and OECD's QSAR Toolbox[31] supporting the development of Adverse Outcome Pathways and associated infrastructures (AOP-Wiki).[32] Despite these positive developments, the 'data integration struggle' from various perspectives (omics, computational chemistry and more 'conventional' toxicological data within REACH and pharmaceutical industry setting) remains a challenge.

Roadmap

The above initiatives are mainly driven by user communities themselves: the chemical industry, funding agencies, pharmaceutical companies, governmental agencies, and organisations such as Member State organisations, the OECD, and non-governmental organisations (NGOs). ELIXIR can contribute strongly to the existing infrastructure projects from a cheminformatics and bioinformatics perspective, providing tools and guidelines, linking and harmonizing the ongoing activities, and serving the toxicology users.

For future risk assessment paradigms solely based on human-derived models, and in this way of higher relevance for human adverse effects,[33] various data types will need to be integrated into and cross the conventional boundaries of risk assessment. This involves external exposure assessment (*e.g.*, via workplace or environmental modelling and measurements[34,35]), internal exposure characterisation (ADME-TK (absorption, distribution, metabolism, and excretion - toxicokinetics) such as via modelling and biomarker-based detection), toxicodynamics on a molecular level, and cell and systems biology. In this way, more data-driven mechanism-based evaluations and supporting data can be integrated into regulatory risk assessment. There will not only be more but also more diverse data such as internal exposure data, that may be inferred from biomonitoring data and/or physiologically-based toxicokinetic modelling to estimate target dose available at the active sites involved in the molecular initiating events of Adverse Outcome Pathways (AOPs).

Another relevant topic is the concept of the exposome,[36–38] which aims at characterising lifetime exposure (not only to chemicals in the narrowest sense, but also dietary components, lifestyle factors, environmental exposures, and more) in relation to health outcome. Often, vulnerable periods of life (infancy, childhood, and old age) are investigated, and the evaluation 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[39] but also environment and health and epidemiology are moving closer together. Data sharing will be increasingly necessary across these disciplines. High throughput data analysis in exposomics, for instance, shares many parallels with metabolomics and other higher-level omics analyses, with an added layer of chemical complexity on top.[38]

The toxicology community is large and well-established. The current list of proponents of an ELIXIR Toxicology Community only reflects a subset of a much larger community with a lot of commitments to, and activities around open collaboration. It has clear omics, knowledge management and data infrastructure needs to accommodate the increasing wish to predict toxicology without animal testing (*e.g.*, in SEURAT-1, EU-ToxRisk, eTOX, and OBERON[40]). 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 (*e.g.*, 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,[26] ToxBank[41] and eNanoMapper.[42]

To benefit the research community, small and medium-sized enterprises (SMEs) and larger industries, and to enable further future support to regulatory applications and upcoming calls (*e.g.*, Green Deal[43,44]), we need to reach an inclusive ecosystem of data, evaluation, and modelling tools. The current separate consortia from different toxicity-related and neighbouring disciplines already work towards data and knowledge that is FAIR.[6] After all, these aspects are essential to efficiently 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 need to be linked, such as on physiologically based toxicokinetic (PBTK) modelling, biological pathways describing affected metabolism and cell biology processes, metabolism, metabolic models and metabolism predictions, drug-response, omics (such as *biological identity*), chemical structures and associated metadata (use, hazard, transformations), QSARs, AOPs, REACH dossiers, and more. Simultaneously, an extension towards the human (preclinical toxicology) discipline should be initiated, in which exposure data are combined with internal exposure and early biomarkers of effect data (*e.g.*, from the European Human Biomonitoring Initiative (HBM4EU)[45] and environmental data from NORMAN[46,47]) towards pathways of toxicity. Interoperability with the standardization efforts of clinical research by CDISC is also important.

However, to reach such interoperable toxicology, resources need to be better integrated. Despite the work of many projects, their FAIR features can still be improved and applying newly developed FAIR metrics will help steer this.[48] Even though there is overlap in content with existing ELIXIR Communities (Table 1, key demands specifically fostering the integration for interoperable toxicology and risk assessment include the following roadmap[19,20]):

- chemical structure interoperability challenges (*e.g.*, links to ELIXIR Metabolomics Community[49])
- metadata, open standards (*e.g.*, links to ELIXIR Interoperability Platform, and TeSS[50])

Table 1. Overlap of the proposed ELIXIR Toxicology Community with ELIXIR platforms and communities.

	Overlap
Platforms	
Tools	Semantic annotation of services (*e.g.*, bio.tools[52]). Alignment with BioContainers[64] to make toxicology reproducible and redeployable (*e.g.*, OpenRiskNet)
Data	BioStudies was co-developed by the CHEMBL-EBI team, and builds on ArrayExpress,[65] an ELIXIR Core Data Resource. Better adoption of the core resources. Sharing toxicology workflows on WorkflowHub
Compute	ELIXIR Authentication and Authorisation Infrastructure (AAI) is used by OpenRiskNet. Better and more sustainable compute infrastructure reuse of core elements (ELIXIR AAI, modelling toolsets)
Interoperability	FAIR and Research Data Management (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 (also for ontology mapping and chemical (sub) structures), Common Workflow Language (CWL)[71] as interoperability for workflow, and standardized data exchange formats like the Investigation-Study-Assay (ISA) standard.[66]
Training	Bioschemas annotated tutorials. Several projects have training tasks that could be added to TeSS, already partially automated.
Communities	
3D-bioinfo & Intrinsically Disordered Proteins (IDP)	Structural information about molecular actions of toxic compounds, molecular initiating events in AOPs
Galaxy	Application Programming Interface (API) standards and ontological annotation of APIs
Marine Metagenomics	Ecotoxicology data reflects effects on populations of species
Plant Sciences	Ecotoxicology data reflects effects of toxicants on plants
Metabolomics	Toxicants Chemical identifiers, data standards, data repository, common resources
Proteomics	Data standards and repository for proteomics data
Microbial Biotechnology	Chemical transformations in biological systems of toxicants
Federated Human Data & Human Copy Number Variation Community	Federated search, phenotype, and genotype data
Rare Diseases	Shared biological pathways and interaction effects between rare diseases and exposure of toxicants

- continued ontology development (*e.g.*, links to ELIXIR Interoperability Platform and Ontology Lookup Service)
- interoperable computation (*e.g.*, links to the Galaxy[51] and bio.tools[52] communities)
- interactions with other ELIXIR Core Data Resources (*e.g.*, Ensembl and Europe PMC)
- interactions with other communities, including nanomedicine and health
- deployment of existing tools and modelling approaches on the ELIXIR Compute infrastructure (also allowing future growth towards risk estimations needing Monte Carlo approaches)
- integration of ELIXIR AAI
- InChI implementation for small molecule data
- Spectral database functionality (open implementations)

The concrete steps forward proposed in this roadmap include:

1. Leverage from open solutions (such as models, ontologies, educational material, and standards) developed by past and ongoing toxicology and ELIXIR projects

2. Connect more closely with the core ELIXIR resources (such as FAIR data and database interoperability), strengthen and connect the inclusive communities that have evolved over the past few years (OpenTox, eNanoMapper, diXa, OpenRiskNet, NORMAN) and older communities like the Federation of European Toxicologists & European Societies of Toxicology (EUROTOX), the Society of Environmental Toxicology and Chemistry (SETAC) and European Environmental Mutagenesis and Genomics Society (EEMGS, formerly known as EEMS)

3. Develop open community standards to support common interest (ontologies, APIs, data formats, deposition databases, and publication recommendations)

Specifically, we will continue to grow the list of involved toxicology research groups, projects, and ELIXIR Node activities. This Community has already held a joint meeting to select the key priorities and use cases (see Tables 2 and 3), resulting in this positioning paper. The ELIXIR Toxicology Community will continue to expand and search for contributors with relevant expertise as the community and activities mature. By bootstrapping from Open Science approaches developed in aforementioned projects (*e.g.*, Open PHACTS[53]), the new Community will focus on mutual benefit, an open and inclusive community, solving practical community problems. The goal is not to design domain-specific approaches, but a pragmatic approach that provides FAIR and open tools from the start, allowing all toxicology and neighbouring communities to benefit from these harmonized solutions. The inclusive community will involve the existing sub-communities in pharmaceuticals, *e.g.*, from eTOX, transQST,[54] and eTRANSAFE,[16] cosmetic ingredients (*e.g.*, from SEURAT-1), high and low-volume chemicals (*e.g.*, from HBM4EU and soon from Horizon Europe Partnership for the Assessment of Risk from Chemicals (PARC),[47] or from different Cefic-LRI (Lang-Range Research Initiative) projects), and nanomaterials (via the NanoSafety Cluster), thus building on shared needs and community solutions and strongly aligned to other ELIXIR communities. Open licensing and interfaces (such as ontologies, standards, and 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" would enable us to set priorities at a level above the individual projects. Existing components that will give this Community an initial boost include: software (*e.g.*, AMBIT with OpenTox API[55]), databases (*e.g.*, diXa, eNano-Mapper, AMBIT-LRI, NORMAN-SLE, and MassBank), ontologies (*e.g.*, the eNanoMapper ontology,[56] and AOP ontology[57]), interoperability concepts (*e.g.*, annotation of OpenAPIs, in collaboration with bio.tools, and semantic structural searches with IDSM), teaching/education material (*i.e.*, Bioschemas annotation of outreach activities,

Table 2. Examples of existing and anticipated collaboration.

	Existing collaboration/Reuse	Anticipated collaboration
Tools	Semantic annotation of services (*e.g.*, bio.tools)	Alignment with BioContainers of toxicology workflow efforts (*e.g.*, OpenRiskNet)
Data	diXa was co-developed by the CHEMBL-EBI team, and builds on ArrayExpress, an ELIXIR Core Data Resources	Better adoption of these Resources
Compute		Better and more sustainable compute infrastructure reuse of core elements (ELIXIR AAI, modelling toolsets)
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 (also for ontology mapping and chemical (sub) structures)
Training	Bioschemas annotated tutorials	Several projects have training tasks that could be added to TeSS, already partially automated

Table 3. The Toxicology Community roadmap is roughly defined by three themes (steps). Each step comes with a 10-year aim, further detailed with possible work that could be done in ELIXIR activities.

Roadmap step	10-year aim	Possible implementation study aim & general criteria
Disseminate existing open solutions (data, database software, models, ontologies, standards, or otherwise) developed among receptive toxicology projects, allowing the advantages of these solutions to become visible to the whole toxicology community	1. Integrating data types across toxicological exposure, biological, Adverse Outcome Pathways, chemical prioritization, hazard assessment, and risk assessment. Data spanning the toxicological domains, properties, exposure, kinetics and dynamics, that support hazard and risk assessment. Approaches include QSAR, QSARDB,[67] *in vitro* kinetics, kinetic modelling (toolboxes and parameter estimation), AOPs and Key Event measuring method data results, supporting read-across.	(a) Better adoption of the ELIXIR core and national node resources. (b) Share teaching material about data exchange between resources on TeSS. (c) Disseminate the need of linking database content, *e.g.*, with identifier mapping and solutions to match entries between databases (d) Disseminate existing open solutions through organising and participating in *e.g.*, toxicology-specific workshops, toxicology conferences, and through other targeted dissemination channels
	2. Interoperable Software and Predictive Models	(a) Explore technical connections between existing platforms (*e.g.*, OpenRiskNet) and the Compute Platform and make them more interoperable, working with ELIXIR Interoperability Platform. (b) Wider adoption of the EDAM ontology[68] and OpenAPI (c) Explore Common Workflow Language (CWL), port a risk assessment case study as CWL workflow
	3. Ontology and Standards	(a) Adoption of standards like InChI/InChIKey (b) Bioschemas adoption by web databases, and listing in Bioschemas Live Deploys (c) Semantic annotation of assays, cell lines, and Standard Operating Procedures (d) Encourage ontology reuse with dissemination of entering ontology terms as free text with autocorrect, autocomplete and autolookup (*e.g.*, Clinical Data Interchange Standards)

Table 3. *Continued*

Roadmap step	10-year aim	Possible implementation study aim & general criteria
Connect and stimulate cross-project collaborations by making all toxicology research output FAIR.	1. FAIRify existing community resources and solutions for findability and easier reuse	(a) Register all Toxicology resources in FAIRsharing[11,10] and/or the Registry of Research Data Repositories[69] to build an awareness and overview of ongoing activities and infrastructures (b) Link up with non-toxicology communities, such as GO FAIR and CDISC (c) Use LabLinks to link Toxicology resources to EuropePMC
	2. Increase FAIRness of relevant data	(a) Register/connect chemicals between PubChem, CompTox, ChEMBL, ChEBI, MolMeDB, IPCheM, IUCLID, HBM4EU, among others, and increase integration of toxicologically-relevant information in open resources (b) Develop citation standard for these solutions (c) Integrate kinetics databases and databases needed for kinetic modelling (d) develop and deploy tool boxes that make data registration in repositories easy (*e.g.*, analogous to OneDep in PDB)
	3. Inventorize Question-and-Answer (Q/A) platforms and sub-communities (*e.g.*, tags) on those platforms where people can learn about solutions	(a) Identify how/where toxicologists are asking questions and what their practices are (b) Define tags to be used cross-platform for toxicology and disseminate these among three Q/A platforms
	4. Develop journal editorial standards for minimal reporting standards for describing the chemical entities in deposition databases and compact identifiers in the main text	(a) Reach out to editors. Make an "easy to refer to" guideline available. Find enough journals to join/support this and point to these guidelines. (b) Define a simple format for reporting names and InChI (Keys) and/or SMILES. Consider the possible ontologies for study descriptions for other things often used in toxicology research and toxicity testing (like Salmonella Typhimurium strains, cell systems, and LD50 tests) (c) Advocate using compact identifiers (structured as "database:dbID" pairs, *e.g.*, uniprot: P1234) that are easy to use and human-readable and that can be resolved by adding links to identifiers.org or n2t.net.

Table 3. *Continued*

Roadmap step	10-year aim	Possible implementation study aim & general criteria
Design and ideally develop missing open community standards to support common interests (such as Open educational resources, ontologies, APIs, and data formats) in advancing toxicology research.	1. Engage with existing (ELIXIR) solutions, like Bioschemas and Bio.tools, ELIXIR AAI, Galaxy including these from connected organisations like RDA, Global Alliance for Genomics and Health (GA4GH), European Open Science Cloud (EOSC), and identify and communicate missing features.	(a) Continue semantic annotation of service APIs (*e.g.*, at bio.tools) and discuss limitations with service and solution providing communities (b) Semantic annotation of training materials, and databases, among others, with Bioschemas JSON-LD and discuss found limitations (c) ELIXIR AAI rollout in Toxicology platforms (already on OpenRiskNet platform as an example to build from) and report on experiences
	2. Sharing of standard operating procedures (SOPs)	(a) Explore existing solutions to share SOPs (*e.g.*, protocols.io) and discuss limitations
	3. Continued evaluation by the adoption of existing ontologies for annotation of toxicology research output	(a) Work with other communities on meta ontologies and mechanisms for (re) creating those from underlying ones and adding dedicated parts. With automated updates and feedback to integrate changes in the base ontologies, *e.g.*, Experimental Factor Ontology and the food & nutrition (Ontology of Nutritional Studies[70]).
	4. Communicate and where possible disseminate FAIR data entry and/or study capture with matching software and/or templates to bench toxicologists	(a) Engage with ELIXIR Training and explore solutions to repurpose existing solutions for data capturing better known in the ELIXIR Toxicology Community (*e.g.*, BioSamples, NanoSafety Cluster templates) (b) Make sure we can capture study designs based on templates can be used for toxicology studies (c) Adopt DataCite practices

in collaboration with TeSS), and virtual infrastructures (*e.g.*, OpenRiskNet Virtual Research Environment and 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 and National Resources. Various existing ELIXIR Communities need similar solutions, *e.g.*, for chemical structure handling, but also toxicology needs proteomics and metabolomics, toxicology involves human data, and ecotoxicology has a significant impact on crops and health.

The following projects have been adopting and integrating FAIR toxicology concepts but need integration with ELIXIR Platforms and Communities: eTOX, NanoCommons (NanoSafety Cluster), EU-ToxRisk, OpenRiskNet, OpenTox Foundation, Open PHACTS Foundation, and the diXa platform. Many other projects have a specific scientific focus but also need integration and some will work on the FAIR concepts. A non-exhaustive list is ACEnano, SmartNanoTox, HeCaToS, NewGeneris, EnviroGenoMarkers, EXPOsOMICS, HELIX, ASAT,[58] PATROLS, and HEALS, as well as new projects, including new Horizon 2020 projects RISK-HUNT3R and HARMLESS and the new Horizon Europe project PARC. Companies and organisations will profit from this Community either as users or as providers of services on top of the infrastructure, including ECHA (FI), Edelweiss Connect (CH), IdeaConsult Ltd. (BG), Misvik Biology (FI), TNO (NL), Seven Past Nine (SI), and the Swedish Academic Consortium for Chemical Safety (SwACCS, SE). Industries showed a strong interest in toxicology, demonstrated by their activities: Cosmetics Europe was participant in the SEURAT-1 cluster; chemical industries (Nanotechnology Industries Association) are a participant of the NSC; chemical branch organisation (Cefic) funds LRI projects around toxicology; and pharmaceutical industries funds toxicology research via IMI projects like eTOX, eTRANSAFE, and Open PHACTS. The Research Data Alliance (RDA) organized a workshop recently about the integration of toxicogenomics resources,[59] and collaboration with international organizations has already been established with, for example, the CompTox Chemicals Dashboard team of the US EPA[60] and PubChem from the US National Institutes of Health.[61] Joint activities include an ELIXIR BioHackathon Europe 2022 project around PubChem-compatible data exchange formats[62] and the use of Wikidata and Wikipedia for important chemicals.[63]

How to join

The Toxicology Community has a webpage at elixir-europe.org/communities/toxicology where it is possible to join the Community's mailing list and stay up-to-date with community activities such as the annual Face-to-Face or scheduled workshops. These are good opportunities to actively participate in the network.

Disclaimer

- Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

- The views expressed in this manuscript are solely those of the authors and do not represent the policies of the U.S. Environmental Protection Agency. Mention of trade names of commercial products should not be interpreted as an endorsement by the U.S. Environmental Protection Agency.

- The views expressed in this manuscript are solely those of the authors and do not represent the policies of and should not be interpreted as an endorsement by the U.S. National Institutes of Health.

Acknowledgements

We acknowledge the discussions with others in the toxicology community (see github.com/bigcaT-UM/ELIXIR-Tox) and ELIXIR for their support so far.

References

1. Crosswell LC, Thornton JM: **ELIXIR: a distributed infrastructure for European biological data.** *Trends Biotechnol.* may 2012; **30**(5): 241–242.
PubMed Abstract | Publisher Full Text

2. Kortenkamp A: **Low dose mixture effects of endocrine disrupters and their implications for regulatory thresholds in chemical risk assessment.** *Curr. Opin. Pharmacol.* 2014; **19**: 105–111.
PubMed Abstract | Publisher Full Text

3. Escher BI, Stapleton HM, Schymanski EL: **Tracking complex mixtures of chemicals in our changing environment.** *Science.* 2020; **367**(6476): 388–392.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Tropsha A: **Best practices for QSAR model development, validation, and exploitation.** *Molecular Informatics.* 2010; **29**(6-7): 476–488.
PubMed Abstract | Publisher Full Text

5. Patlewicz G, Ball N, Booth ED, *et al.*: **Use of category approaches, read-across and (q)sar: General considerations.** *Regul. Toxicol. Pharmacol.* 2013; **67**(1): 1–12.
PubMed Abstract | Publisher Full Text

6. Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **Comment: The FAIR Guiding Principles for scientific data management and stewardship.** *Scientific Data.* mar 2016; **3**(1): 160018–9.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Jacobsen A, de Miranda Azevedo R, Juty N, *et al.*: **FAIR Principles: Interpretations and Implementation Considerations.** *Data intelligence.* 2020; **2**(4): 10–29.
Publisher Full Text

8. Courtot M, Cherubin L, Faulconbridge A, *et al.*: **BioSamples database: an updated sample metadata hub.** *Nucleic Acids Res.* 11 2018; **47**(D1): D1172–D1178.
PubMed Abstract | Publisher Full Text | Free Full Text

9. Sarkans U, Füllgrabe A, Ali A, *et al.*: **From ArrayExpress to BioStudies.** *Nucleic Acids Res.* 11 2020; **49**(D1): D1502–D1506.
PubMed Abstract | Publisher Full Text | Free Full Text

10. The FAIRsharing Community, Sansone S-A, McQuilton P, *et al.*: **FAIRsharing as a community approach to standards, repositories and policies.** *Nat. Biotechnol.* April 2019; **37**(4): 358–367.
Publisher Full Text

11. The FAIRsharing Team: **FAIRsharing Website.** 2020.
Reference Source

12. Wang Z, Walker GW, Muir DCG, *et al.*: **Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories.** *Environ. Sci. Technol.* 2020; **54**(5): 2575–2584.
PubMed Abstract | Publisher Full Text

13. Gocht T, Berggren E, Ahr HJ, *et al.*: **The SEURAT-1 approach towards animal free human safety assessment.** *ALTEX.* 2015; **32**(1): 9–24.
PubMed Abstract | Publisher Full Text

14. Daneshian M, Kamp H, Hengstler J, *et al.*: **Highlight report: Launch of a large integrated European in vitro toxicology project: EU-ToxRisk.** *Arch. Toxicol.* may 2016; **90**(5): 1021–1024.
PubMed Abstract | Publisher Full Text | Free Full Text

15. Sanz F, Pognan F, Steger-Hartmann T, *et al.*: **Legacy data sharing to improve drug safety assessment: The eTOX project.** *Nat. Rev. Drug Discov.* nov 2017; **16**(12): 811–812.
Publisher Full Text

16. Pognan F, Steger-Hartmann T, Díaz C, *et al.*: **The eTRANSAFE Project on Translational Safety Assessment through Integrative Knowledge Management: Achievements and Perspectives.** *Pharmaceuticals.* 2021; **14**(3).
PubMed Abstract | Publisher Full Text | Free Full Text

17. The Eurion Consortium: **Eurion Project Website.** 2021.
Reference Source

18. Jeliazkova N, Koch V, Li Q, *et al.*: **Linking LRI AMBIT chemoinformatic system with the IUCLID substance database to support read-across of substance endpoint data and category formation.** *Toxicol. Lett.* 2016; **258**: S114–S115.
Publisher Full Text

19. EU-US Roadmap – Nanoinformatics 2030 – EU NanoSafety Cluster: 2020.
Reference Source

20. Karcher S, Willighagen EL, Rumble J, *et al.*: **Integration among databases and data sets to support productive nanotechnology: Challenges and recommendations.** *NanoImpact.* jan 2018; **9**: 85–101.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Kohonen P, Parkkinen JA, Willighagen EL, *et al.*: **A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury.** *Nat. Commun.* 8, jul 2017.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Luechtefeld T, Maertens A, Russo DP, *et al.*: **Global analysis of publicly available safety data for 9,801 substances registered under REACH from 2008-2014.** *ALTEX.* 2016; **33**(2): 95–109.
Publisher Full Text

23. Ganter B, Snyder RD, Halbert DN, *et al.*: **Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the drugmatrix® database.** *Pharmacogenomics.* 2006; **7**(7): 1025–1044.
PubMed Abstract | Publisher Full Text

24. Richard AM, Judson RS, Houck KA, *et al.*: **Toxcast chemical landscape: Paving the road to 21st century toxicology.** *Chem. Res. Toxicol.* 2016; **29**(8): 1225–1251.
PubMed Abstract | Publisher Full Text

25. Igarashi Y, Nakatsu N, Yamashita T, *et al.*: **Open TG-GATEs: a large-scale toxicogenomics database.** *Nucleic Acids Res.* 10 2014; **43**(D1): D921–D927.
Publisher Full Text

26. Hendrickx DM, Aerts HJWL, Caiment F, *et al.*: **diXa: a data infrastructure for chemical safety assessment.** *Bioinformatics.* 2015; **31**(9): 1505–1507.
PubMed Abstract | Publisher Full Text | Free Full Text

27. Čihák R: **Reach - an overview.** *Interdiscip. Toxicol.* 01 Jun. 2009; **2**(2): 42–44.
PubMed Abstract | Publisher Full Text | Free Full Text

28. Cases M, Briggs K, Steger-Hartmann T, *et al.*: **The eTOX data-sharing project to advance in Silico drug-induced toxicity prediction.** *Int. J. Mol. Sci.* nov 2014; **15**(11): 21136–21154.
PubMed Abstract | Publisher Full Text | Free Full Text

29. OECD: **eChemPortal.** 2020.
Reference Source

30. ECHA: **IUCLID.** 2020.
Reference Source

31. OECD: **QSAR Toolbox.** 2020.
Reference Source

32. Leist M, Ghallab A, Graepel R, *et al.*: **Adverse outcome pathways: opportunities, limitations and open questions.** *Arch. Toxicol.* nov 2017; **91**(11): 3477–3505.
PubMed Abstract | Publisher Full Text

33. Nymark P, Rieswijk L, Ehrhart F, *et al.*: **A Data Fusion Pipeline for Generating and Enriching Adverse Outcome Pathway Descriptions.** *Toxicol. Sci.* 2018; **162**(1): 264–275.
PubMed Abstract | Publisher Full Text

34. Ring CL, Arnot JA, Bennett DH, *et al.*: **Consensus modeling of median chemical intake for the u.s. population based on predictions of exposure pathways.** *Environ. Sci. Technol.* 2019; **53**(2): 719–732.
PubMed Abstract | Publisher Full Text | Free Full Text

35. Wambaugh JF, Wang A, Dionisio KL, *et al.*: **High throughput heuristics for prioritizing human exposure to environmental chemicals.** *Environ. Sci. Technol.* 2014; **48**(21): 12760–12767.
PubMed Abstract | Publisher Full Text

36. Wild CP: **Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology.** *Cancer Epidemiology and Prevention Biomarkers.* 2005; **14**(8): 1847–1850.
Publisher Full Text

37. Escher BI, Hackermüller J, Polte T, *et al.*: **From the exposome to mechanistic understanding of chemical-induced adverse effects.** *Environ. Int.* 2017; **99**: 97–106.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Vermeulen R, Schymanski EL, Barabási A-L, *et al.*: **The exposome and health: Where chemistry meets biology.** *Science.* 2020; **367**(6476): 392–396.
PubMed Abstract | Publisher Full Text | Free Full Text

39. Barouki R, Audouze K, Coumoul X, *et al.*: **Integration of the human exposome with the human genome to advance medicine.** *Biochimie.* 2018; **152**: 155–158.
PubMed Abstract | Publisher Full Text

40. Audouze K, Sarigiannis D, Alonso-Magdalena P, *et al.*: **Integrative strategy of testing systems for identification of endocrine disruptors inducing metabolic disorders—an introduction to the oberon project.** *Int. J. Mol. Sci.* 2020; **21**(8).
PubMed Abstract | Publisher Full Text | Free Full Text

41. Kohonen P, Benfenati E, Bower D, *et al.*: **The ToxBank data warehouse: Supporting the replacement of in vivo repeated dose systemic toxicity testing.** *Mol. Inform.* jan 2013; **32**(1): 47–63.
PubMed Abstract | Publisher Full Text

42. Jeliazkova N, Chomenidis C, Doganis P, *et al.*: **The eNanoMapper database for nanomaterial safety information.** *Beilstein J. Nanotechnol.* 2015; **6**(1): 1609–1634.
PubMed Abstract | Publisher Full Text

43. **H2020 Green Deal Call LC-GD-8-1-2020: Innovative, systemic zero-pollution solutions to protect health, environment and natural resources from persistent and mobile chemicals.** 2020
Reference Source

44. **H2020 Green Deal Call LC-GD-8-2-2020: Fostering regulatory science to address combined exposures to industrial chemicals and pharmaceuticals: from science to evidence-based policies.** 2020.
Reference Source

45. Ganzleben C, Antignac J-P, Barouki R, *et al.*: **Human biomonitoring as a tool to support chemicals regulation in the european union.** *Int. J. Hyg. Environ. Health.* 2017; **220**(2, Part A): 94–97. Special Issue:

Human Biomonitoring 2016.
Publisher Full Text

46. Dulio V, van Bavel B, Brorström-Lundén E, *et al.*: **Emerging pollutants in the EU: 10 years of NORMAN in support of environmental policies and regulations.** *Environ. Sci. Eur.* 5, December 2018; **30**(1): 5–4715.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Dulio V, Koschorreck J, van Bavel B, *et al.*: **The NORMAN Association and the European Partnership for Chemicals Risk Assessment (PARC): let's cooperate!.** *Environ. Sci. Eur.* December 2020; **32**(1): 100.
Publisher Full Text

48. Ammar A, Bonaretti S, Winckers L, *et al.*: **A semi-automated workflow for fair maturity indicators in the life sciences.** *Nanomaterials.* 2020; **10**(10): 1–14.
PubMed Abstract | Publisher Full Text | Free Full Text

49. van Rijswijk M, Beirnaert C, Caron C, *et al.*: **The future of metabolomics in ELIXIR [version 2; peer review: 3 approved].** *F1000Research.* 2017; **6**(1649).
Publisher Full Text

50. Beard N, Bacall F, Nenadic A, *et al.*: **TeSS: a platform for discovering life-science training opportunities.** *Bioinformatics.* 02 2020; **36**(10): 3290–3291.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Afgan E, Baker D, Batut B, *et al.*: **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update.** *Nucleic Acids Res.* jul 2018; **46**(W1): W537–W544.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Ison J, Ienasescu H, Chmura P, *et al.*: **The bio.tools registry of software tools and data resources for the life sciences.** *Genome Biol.* 2019; **20**(1): 164.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Williams AJ, Harland L, Groth P, *et al.*: **Open PHACTS: semantic interoperability for drug discovery.** *Drug Discov. Today.* 2012; **17**(21): 1188–1198.
PubMed Abstract | Publisher Full Text

54. Ferreira S, Fisher C, Furlong LI, *et al.*: **Quantitative systems toxicology modeling to address key safety questions in drug development: A focus of the transqst consortium.** *Chem. Res. Toxicol.* 2020; **33**(1): 7–9.
PubMed Abstract | Publisher Full Text

55. Jeliazkova N, Jeliazkov V: **AMBIT RESTful web services: An implementation of the OpenTox application programming interface.** *J. Cheminformatics.* 2011; **3**(1).
PubMed Abstract | Publisher Full Text | Free Full Text

56. Hastings J, Jeliazkova N, Owen G, *et al.*: **eNanoMapper: Harnessing ontologies to enable data integration for nanomaterial risk assessment.** *J. Biomed. Semant.* mar 2015; **6**(1): 10.
PubMed Abstract | Publisher Full Text | Free Full Text

57. Burgoon LD: **The AOPOntology: A semantic artificial intelligence tool for predictive toxicology.** *Appl. In Vitro Toxicol.* sep 2017; **3**(3): 278–281.
Publisher Full Text

58. Stierum R, Aarts J, Boorsma A, *et al.*: **Assuring safety without animal testing concept (ASAT). Integration of human disease data with in vitro data to improve toxicology testing.** *Toxicol. Lett.* 2014; **229**: S4.
Publisher Full Text

59. Hendrickx DM, Boyles RR, Kleinjans JCS, *et al.*: **Workshop report: Identifying opportunities for global integration of toxicogenomics databases, 26–27 June 2013, Research Triangle Park, NC, USA.** *Arch. Toxicol.* nov 2014; **88**(12): 2323–2332.
PubMed Abstract | Publisher Full Text | Free Full Text

60. Williams AJ, Grulke CM, Edwards J, *et al.*: **The CompTox Chemistry Dashboard: A community data resource for environmental chemistry.** *J. Cheminformatics.* nov 2017; **9**(1): 61.
PubMed Abstract | Publisher Full Text. | Free Full Text

61. Kim S, Chen J, Cheng T, *et al.*: **PubChem 2019 update: improved access to chemical data.** *Nucleic Acids Res.* 2019; **47**(D1): D1102–D1109.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Schymanski EL, Bolton EE: **FAIR chemical structures in the Journal of Cheminformatics.** *J. Cheminform.* 2021; **13**(1): 50.
PubMed Abstract | Publisher Full Text | Free Full Text

63. Sinclair G, Thillainadarajah I, Meyer B, *et al.*: **Wikipedia on the CompTox Chemicals Dashboard: Connecting Resources to Enrich Public Chemical Data.** *J. Chem. Inf. Model.* 2022; **62**(20): 4888–4905.
Publisher Full Text

64. Gruening B, Sallou O, Moreno P, *et al.*: **Recommendations for the packaging and containerizing of bioinformatics software [version 1; peer review: 2 approved with reservations].** *F1000Research.* 2018; **7**(742).
Publisher Full Text

65. Athar A, Füllgrabe A, George N, *et al.*: **ArrayExpress update – from bulk to single-cell expression data.** *Nucleic Acids Res.* 2019; **47**(D1): D711–D715.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Sansone S-A, Rocca-Serra P, Field D, *et al.*: **Toward interoperable bioscience data.** *Nat. Genet.* Jan 2012; **44**(2): 121–126.
PubMed Abstract | Publisher Full Text | Free Full Text

67. Ruusmann V, Sild S, Maran U: **QSAR DataBank - an approach for the digital organization and archiving of QSAR model information.** *J. Cheminformatics.* 25, 2014; **6**(1).
PubMed Abstract | Publisher Full Text | Free Full Text

68. Ison J, Kalaš M, Jonassen I, *et al.*: **EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats.** *Bioinformatics.* 03 2013; **29**(10): 1325–1332.
PubMed Abstract | Publisher Full Text | Free Full Text

69. The RE3 Team: **Registry of Research Data Repositories Website.** 2020.
Reference Source

70. Vitali F, Lombardo R, Rivero D, *et al.*: **ONS: an ontology for a standardized description of interventions and observational studies in nutrition.** *Genes Nutr.* 12, 2018; **13**(1): 12.
PubMed Abstract | Publisher Full Text | Free Full Text

71. Amstutz P, Crusoe MR, Nebojša TN, *et al.*: **Common Workflow Language, v1.0. Specification, Common Workflow Language working group.** 2016.
Reference Source

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 11 October 2023

https://doi.org/10.5256/f1000research.156142.r211615

✔ **Christopher Southan**
[1] Deanery of Biomedical Sciences, University of Edinburgh, Edinburgh, UK
[2] Data Sciences,, Medicines Discovery Catapult, Macclesfield, UK

The authors have addressed my comments in this revision I thus hereby approve it.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics and cheminformatics (see LinkedIN)

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 08 April 2022

https://doi.org/10.5256/f1000research.78265.r124031

✔ **Marta Eide**
[1] Department of Biological Sciences, University of Bergen, Bergen, Norway
[2] Centre for Digital Life Norway, Bergen, Norway
[3] Department of Biological Sciences, University of Bergen, Bergen, Norway
[4] Centre for Digital Life Norway, Bergen, Norway

In this opinion article, the authors describe the establishment of a new ELIXIR Toxicology Community in order to enable the transition from traditional disease-specific observational models to target-specific predictive models. The article introduces the rapidly developing field of toxicology and the interoperability issues arising in combining chemical and biochemical analysis, biological data, and kinetic modelling. The authors argue for "*clear, standardized guidelines for data capture, approaches to integrate and connect across multiple datasets, clear data licensing for these data repositories, and tools that support accessing the data*". The ELIXIR Toxicology Community is suggested to 1) connect the different user communities (including industry, governmental agencies, NGOs, and more), 2) strongly contribute to link between, and harmonize the already existing relevant infrastructure projects, resources, and activities, and 3) provide useful tools, guidelines, and training.

Acknowledging the vast number of chemicals that are currently being produced, and the limited knowledge of the consequences of exposure to humans and the environment, this is a timely and important effort. From my impression, an ELIXIR Community is a good tool to drive and coordinate such an international initiative.

I find the article comprehensive and well-written, and their arguments well discussed. However, I have some points that I hope the authors will take into consideration when taking the initiative further:

1. It is early stated that "*in this ELIXIR Community, the focus will be primarily on the protection of human health*". I would argue that this is an unfavorable restriction for three reasons:
    - It disregards the 'OneHealth approach' recognized by WHO, FAO, OIE, and UNEP, which is defined as "*an integrated, unifying approach that aims to sustainably balance and optimize the health of people, animals and ecosystem. It recognizes the health of humans, domestic and wild animals, plants, and the wider environment (including ecosystems) are closely linked and inter-dependent*" (One Health High-Level Expert Panel Annual Report 2021). EFSA, ECHA, EMA *et al.* are also organizing a conference on the topic in June 2022 (https://www.one2022.eu). Seeing that animals and ecosystems are closely linked to human health, for example through foods, their responses to and effects from toxicant exposures are also highly important to study and include in risk assessments.

    - Live animals, as well as animal tissue, cells, and molecules, are often used as models for toxicological studies. As the toxicological defense systems are highly preserved through evolution, they can give a good indication of human outcomes. Thus, it is important to also ensure supporting infrastructure for integrating these as well.

    - Ontologies, standards, and data resources (like omics repositories, data analysis tools, etc.) are often already overlapping.

2. Although I welcome different stakeholders being identified as "user communities", I am concerned by the, from my understanding, lack of industry and governmental/administrative agencies in the author list of this article. I believe that if the aim of improving FAIR data integration for safety to enable ultimate risk assessments is to be met, it is eminent to include non-academic perspectives, experiences, and concerns into the initiative from early on. Experiences from the PARC project could for example be useful in how to approach this.

3. From my experience, the field of toxicology in general has limited experience with FAIR data management and is not familiar with the resources that are in place. I see that toxicologists' communication, dissemination, and training are mentioned in the proposed roadmap, but I am missing a more targeted approach to engaging with the communities. For example, such activities could include organizing toxicology-specific workshops, presenting at toxicology conferences, and targeted information through different channels. As the proposed ELIXIR Community seems to have a good overview of ongoing initiatives and infrastructures and a good anchoring in the scientific community, a mapping of their target audience and users in the field could be suggested as a starting point.

4. As mentioned in the paper, systems biology approaches are also getting implemented in toxicology, i.e. systems toxicology (described in Sturla *et al.,* 2014[1]. From my point of view, I am missing mentioning of tools for this, such as the FAIRDOM SEEK platform that integrates ISA-structured metadata and workflows with JWS-based biological systems modelling. However, there might already be other solutions for this that I don't recognize in the paper

References

1. Sturla SJ, Boobis AR, FitzGerald RE, Hoeng J, et al.: Systems toxicology: from basic research to risk assessment.*Chem Res Toxicol*. 2014; **27** (3): 314-29 PubMed Abstract | Publisher Full Text

Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Environmental toxicology, endocrine disruption, bioinformatics, systems toxicology, marine toxicology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 26 Sep 2023

Marvin Martens

Point by point replies:

- *In this opinion article, the authors describe the establishment of a new ELIXIR Toxicology Community in order to enable the transition from traditional disease-specific observational models to target-specific predictive models. The article introduces the rapidly developing field of toxicology and the interoperability issues arising in combining chemical and biochemical analysis, biological data, and kinetic modelling. The authors argue for "clear, standardized guidelines for data capture, approaches to integrate and connect across multiple datasets, clear data licensing for these data repositories, and tools that support accessing the data". The ELIXIR Toxicology Community is suggested to 1) connect the different user communities (including industry, governmental agencies, NGOs, and more), 2) strongly contribute to link between, and harmonize the already existing relevant infrastructure projects, resources, and activities, and 3) provide useful tools, guidelines, and training. Acknowledging the vast number of chemicals that are currently being produced, and the limited knowledge of the consequences of exposure to humans and the environment, this is a timely and important effort. From my impression, an ELIXIR Community is a good tool to drive and coordinate such an international initiative.*

We are happy to read that Dr. Eide agrees with our observation.

- *I find the article comprehensive and well-written, and their arguments well discussed. However, I have some points that I hope the authors will take into consideration when taking the initiative further:*
  It is early stated that "in this ELIXIR Community, the focus will be primarily on the protection of human health".
  I would argue that this is an unfavorable restriction for three reasons.
  It disregards the 'OneHealth approach' recognized by WHO, FAO, OIE, and UNEP, which is defined as "an integrated, unifying approach that aims to sustainably balance and optimize the health of people, animals and ecosystem. It recognizes the health of humans, domestic and wild animals, plants, and the wider environment (including ecosystems) are closely linked and inter-dependent" (One Health High-Level Expert Panel Annual Report 2021). EFSA, ECHA, EMA et al. are also organizing a conference on the topic in June 2022 (https://www.one2022.eu). Seeing that animals and ecosystems are closely linked to human health, for example through foods, their responses to and effects from toxicant exposures are also highly important to study and include in risk assessments.

This choice is based on the earlier focuses of ELIXIR itself. We agree that a wider idea of toxicology is important and we note that ELIXIR actually has broadened its focus too. We have updated the text to "*will be on the focus areas of ELIXIR, including the protection of human, animal, and environmental health.*"

- *Live animals, as well as animal tissue, cells, and molecules, are often used as models for toxicological studies. As the toxicological defense systems are highly preserved through evolution, they can give a good indication of human outcomes. Thus, it is important to also ensure supporting infrastructure for integrating these as well. Ontologies, standards, and data resources (like omics repositories, data analysis tools, etc.) are often already overlapping.*

The Toxicology Community is whenever possible using the ELIXIR Core Resources, and we note data Core and Node Services (https://elixir-europe.org/services) like Ensembl, deposition databases, and WikiPathways capture knowledge across species and often visualize homology where applicable. Indeed, the opportunity as an ELIXIR Community is that we can benefit from these overlaps, see the third roadmap theme in Table 3.

- *Although I welcome different stakeholders being identified as "user communities", I am concerned by the, from my understanding, lack of industry and governmental/administrative agencies in the author list of this article. I believe that if the aim of improving FAIR data integration for safety to enable ultimate risk assessments is to be met, it is eminent to include non-academic perspectives, experiences, and concerns into the initiative from early on. Experiences from the PARC project could for example be useful in how to approach this.*

This concern is shared. The way ELIXIR is set up, industry cannot be a direct beneficiary of ELIXIR funding. This reflects on the Toxicology Community. We expect to overcome this in various ways: the user communities and community members themselves do include the industries, e.g. via the EU-funded projects or projects like PARC; second, ELIXIR has dedicated industry-oriented activities, which we plan to take advantage of too. At the time of writing, the latter had not yet materialized.

Our first Toxicology Community face-to-face meeting was/is actually co-located with a PARC meeting in Brussels, besides that multiple co-authors are members in PARC. This interaction with PARC also involves the Toxicology Community Implementation Study, which will involve focused workshops on various aspects that were presented in the Toxicology Community roadmap. One of the proposed workshops is aimed at leveraging FAIR solutions for Adverse Outcome Pathways, with a practical case study that is aligned with the PARC project.

- *From my experience, the field of toxicology in general has limited experience with FAIR data management and is not familiar with the resources that are in place. I see that toxicologists' communication, dissemination, and training are mentioned in the proposed roadmap, but I am missing a more targeted approach to engaging with the communities. For example, such activities could include organizing toxicology-specific workshops, presenting at toxicology conferences, and targeted information through different channels. As the proposed ELIXIR Community seems to have a good overview of ongoing initiatives and infrastructures and a good anchoring in the scientific community, a mapping of their target audience and users in the field could be suggested as a starting point.*

Indeed, projects like NanoCommons, VHP4Safety, RiskGONE, Gov4Nano have done research into how to "do" FAIR in Toxicology. The Toxicology Community will help channel this exchange of your gained insights. The use of the ELIXIR training platform TeSS in the whitepaper is a specific example activity in this area. We now also added an additional note in Table 3 to emphasize the need for dissemination of FAIR solutions by "organising and participating in e.g. toxicology-specific workshops, toxicology conferences, and through other targeted dissemination channels". In reality, we have already presented the Toxicology Community aims at various toxicology and computation biology meetings and conferences.

- *As mentioned in the paper, systems biology approaches are also getting implemented in toxicology, i.e. systems toxicology (described in Sturla et al., 20141. From my point of view, I am missing mentioning of tools for this, such as the FAIRDOM SEEK platform that integrates ISA-structured metadata and workflows with JWS-based biological systems modelling. However, there might already be other solutions for this that I don't recognize in the paper*

This paper is a whitepaper where we tried to find a balance with a high-level sketch and some practical examples based on already ongoing activities. There are a lot of possibilities,

more than ELIXIR will be able to support (financially or resource wise). We will have to continously look for activities between ELIXIR (and projects ELIXIR is involved in) and funded toxicology projects. For example, the first Toxicology Community face-to-face meeting is/was focused on seeking synergies and collaboration with PARC.

Therefore, where this whitepaper does not mention something, this does not imply the unmentioned approach (such as the FAIRDOM SEEK platform as Dr. Eide mentions) is excluded. In other words, this whitepaper is written with an “open world” assumption. We have added a ‘how to join’ section at the end of the article.

Competing Interests: No competing interests were disclosed.

Reviewer Report 21 March 2022

https://doi.org/10.5256/f1000research.78265.r124032

?

Christopher Southan
[1] Deanery of Biomedical Sciences, University of Edinburgh, Edinburgh, UK
[2] Data Sciences,, Medicines Discovery Catapult, Macclesfield, UK
[3] Deanery of Biomedical Sciences, University of Edinburgh, Edinburgh, UK
[4] Data Sciences,, Medicines Discovery Catapult, Macclesfield, UK

This wide-ranging review circumscribes an important data integration task with a major cat-herding dimension. Whilst also appearing herculean, this esteemed author collective fully understands what they are letting themselves in for and I wish them the best for this endeavor. It is, of course, early days, but I will make a few points (whether these might be addressed in a revision and blending in what other reviewers may come up with I will leave to the authors)

1. Harvesting tox data from the extant and future literature seems to neither be specifically addressed nor proposed via direct interactions with the pharmaceutical companies generating most of it. Standardised data from their large historical internal sets only surfaces sparsely and heterogeneously public databases. Companies such as LahsaVtic, Instem, and ToxPlanet claim to have large databases compiled from the literature. Might ChEMBL come into the frame here if they could strategically increase their toxicology data extraction from the literature, both prospectively and retrospectively? (So far with only 24 assays)

2. My impression is there are simply too many resources mentioned for realistic overarching harmonisation. Could these be cut back to a smaller “active membership” prioritised by the amount of data they have and will continue to generate?

3. Quote “standardized guidelines for data capture, approaches to integrate and connect across multiple databases”. As we know, this is going to be the really tough bit. It needs the major sources to share not only their data schema and metadata indexing but push for a level of convergence. It has to be said that ELIXIR does not have a good record on this. Unless I have missed it, none of the core data resources has made any significant changes in their individual (and often decades-old) data models to really enhance interoperability.

4. PubChem has 113,110 compounds from ~ 30 annotation sources (although most relatively small) indexed under “PubChem Compound TOC: Toxicological Information” and toxicology-related submitting sources cover over 0.5 million substances. In addition, the term “toxicology” matches 3,774 BioAssays covering 86,068 compounds. This makes PubChem a *de facto* toxicology integration hub. The ELIXIR effort should consolidate and expand this, including where an increase in toxicology curation by ChEMBL could feed into BioAssay and thus become interoperative with the 294,000 biological activity data points already in there.

Is the topic of the opinion article discussed accurately in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Are arguments sufficiently supported by evidence from the published literature?
Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bioinformatics and cheminformatics (see LinkedIN)

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 26 Sep 2023

Marvin Martens

Point by point replies:

- *This wide-ranging review circumscribes an important data integration task with a major cat-herding dimension. Whilst also appearing herculean, this esteemed author collective fully understands what they are letting themselves in for and I wish them the best for this endeavor. It is, of course, early days, but I will make a few points (whether these might be addressed in a revision and blending in what other reviewers may come up with I will leave to the authors)*

Thank you. Now that the first implementation study has been submitted, the first face-to-face is organized, we have a mailing list, things start to become more practical.

- *Harvesting tox data from the extant and future literature seems to neither be specifically addressed nor proposed via direct interactions with the pharmaceutical companies generating most of it. Standardised data from their large historical internal sets only surfaces sparsely and heterogeneously public databases. Companies such as LahsaVtic, Instem, and ToxPlanet claim to have large databases compiled from the literature. Might ChEMBL come into the frame here if they could strategically increase their toxicology data extraction from the literature, both prospectively and retrospectively? (So far with only 24 assays)*

Data collection has not been written down as a core activity for this community. We will support the adoption by toxicology projects of the ELIXIR resources. ChEMBL and ChEBI along with a few more databases are ELIXIR services, but most of them with a particular focus. Interoperability with large databases, like PubChem, is obviously envisioned, as also stated in the second theme Table 3. Currently, the Toxicology Community will support sharing data, but not expect deposition nor open licenses.

- *My impression is there are simply too many resources mentioned for realistic overarching harmonisation. Could these be cut back to a smaller "active membership" prioritised by the amount of data they have and will continue to generate?*

A broader ELIXIR Community does not exclude such smaller activities. See our relpy about the "open world" assumption to the other reviewer: let the Toxicology Community organically grow their community practices, extending, of course, on the core ELIXIR activities.

- *Quote "standardized guidelines for data capture, approaches to integrate and connect across multiple databases". As we know, this is going to be the really tough bit. It needs the major sources to share not only their data schema and metadata indexing but push for a level of convergence. It has to be said that ELIXIR does not have a good record on this. Unless I have missed it, none of the core data resources has made any significant changes in their individual (and often decades-old) data models to really enhance interoperability.*

We cannot comment here on the work of ELIXIR at large. So far, we have collaborated on interoperability with, for example, Bioschemas and do see convergence. The MolecularEntity and ChemicalSubstance types and profiles are adopted by a growing number of resources (e.g. ChEBI, Wikidata Scholia, MassBank) and EU projects (e.g. VHP4Safety, NanoCommons, NanoSolveIT).

- *PubChem has 113,110 compounds from ~ 30 annotation sources (although most relatively small) indexed under "PubChem Compound TOC: Toxicological Information" and toxicology-related submitting sources cover over 0.5 million substances. In addition, the term "toxicology" matches 3,774 BioAssays covering 86,068 compounds. This makes PubChem a de facto toxicology integration hub. The ELIXIR effort should consolidate and expand this, including where an increase in toxicology curation by ChEMBL could feed into BioAssay and thus become interoperative with the 294,000 biological activity data points already in there.*

In absense of a "European PubChem", the link with PubChem is indeed important. At the 2022 BioHackathon Europe, the Toxicology Community invited and worked with PubChem on interoperability.

We have updated the manuscript to reflect this. At the end of the paper, just above "how to

join", two examples are added on collaboration with PubChem and with the EPA. See also the collaboration with Mortensen et al. (2022)., doi: 10.3389/ftox.2022.803983.
Finally, there are other non-ELIXIR services of importance, like the international AOP-Wiki and the EPA CompTox Dashboard. Keeping our interactions with international organizations like the RDA, GO FAIR, and IUPAC continues to be important. Workshops, Face-to-Face meetings, Biohackathons, etc will formalize these interactions by making them practical.

Competing Interests: No competing interests were disclosed.