

OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

Deliverable Report D4.3

Report of the Service Integration with
OpenRiskNet (Final Report)



This project is funded by
the European Union

OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge
Integration and *in silico* Analysis and Modelling in Risk Assessment

Project Number 731075

www.openrisknet.org

Project identification

Grant Agreement	731075
Project Name	OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and <i>in silico</i> Analysis and Modelling in Risk Assessment
Project Acronym	OpenRiskNet
Project Coordinator	Edelweiss Connect GmbH
Start date	1 December 2016
End date	30 November 2019
Duration	36 Months
Project Partners	<p>P1 Edelweiss Connect GmbH Switzerland (EwC) P2 Johannes Gutenberg-Universität Mainz, Germany (JGU) P3 Fundacio Centre De Regulacio Genomica, Spain (CRG) P4 Universiteit Maastricht, Netherlands (UM) P5 The University Of Birmingham, United Kingdom (UoB) P6 National Technical University Of Athens, Greece (NTUA) P7 Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V., Germany (Fraunhofer) P8 Uppsala Universitet, Sweden (UU) P9 Medizinische Universität Innsbruck, Austria (MUI) P10 Informatics Matters Limited, United Kingdom (IM) P11 Institut National De L'environnement Et Des Risques, France (INERIS) P12 Vrije Universiteit Amsterdam, Netherlands (VU)</p>

Deliverable Report identification

Document ID and title	Deliverable 4.3 Report of the Service Integration with OpenRiskNet (Final Report)
Deliverable Type	Report
Dissemination Level	Public (PU)
Work Package	WP4
Task(s)	Task 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 and 4.7
Deliverable lead partner	UM
Author(s)	Danyel Jennen, Marvin Martens, Egon Willighagen, Chris Evelo (UM), Pantelis Karatzas, Philip Doganis, Haralambos Sarimveis (NTUA), Daan Geerke (VU), Ola Spjuth, Arvid Berg (UU), Denis Gebele, Atif Raza (JGU), Marc Jacobs (Fraunhofer), Lucian Farcas, Thomas Exner (EwC)
Status	Final
Version	V1.0
Document history	2019-11-19 Draft version 2019-11-29 Final version

Table of Contents

Summary	5
Introduction	6
Integration of services in case studies	7
DataCure	8
MetaP	9
TGX	11
ModelRX	13
SysGroup	15
AOPLink	16
RevK	17
Integration status of consortium-provided services	19
Services selected within the Implementation Challenge	27
Categorisation of services	29
Categorisation by service type	29
Categorisation by risk assessment area and applicability domain	33
Categorisation by target user	34
Categorisation by target industry	36
Conclusion	37
Glossary	38
References	39

Summary

This report describes the final status of the service integration including numbers of active services provided by the consortium, associated partners and other third parties. The work described in this report addresses all areas and tasks of WP4 (*i.e.* Toxicology, Chemical Properties and Bioassay Databases, Omics Databases, Knowledge Bases and Data Mining, Ontology Services, Processing and Analysis, Predictive Toxicology, Workflows, Visualisation and Reporting). Because of their importance for service integration, we also reference to work performed in WP1 on case studies and in WP2 on e-infrastructure interoperability and deployment.

The report includes examples of how the services provided by the consortium partners and associated partners are combined and integrated into workflows providing solutions to answer the specific case study questions. Further details are presented on how the services from the Implementation Challenge are strengthening the demonstration on the case studies.

A detailed, up-to-date description of all services integrated in OpenRiskNet including the extensive categorization of the integration status (a process based on a predefined set of eight operations) is available on the project's website. Here we will concentrate on statistics on the services integrated (*e.g.* categories, type, applicability domain and areas covered, targeted users or industries).

Introduction

For complete risk assessments and safe-by-design studies, data and tools from different areas have to be available. The work of WP4 is structured according to these areas and the different tasks deal with the integration of a set of prototype implementations later used as best-practice examples. The OpenRiskNet functionality is, thus, defined by, composed of and empowered by a variety of incorporated services (databases, knowledge bases, and tools for preprocessing, analysis and modelling). Integration of these tools started on day 1 and continued until the end of the project. The final report presented here describes the work achieved in the course of the project.

This work addresses the following tasks:

- Task 4.1 Toxicology, Chemical Properties and Bioassay Databases
- Task 4.2 Omics Databases
- Task 4.3 Knowledge Bases and Data Mining
- Task 4.4 Ontology Services
- Task 4.5 Processing and Analysis
- Task 4.6 Predictive Toxicology
- Task 4.7 Workflows, Visualisation and Reporting

Early OpenRiskNet services are adaptations of existing services developed by the consortium partners, prior to the start of the project or in other ongoing projects, or are a public version, modified and made available towards improved harmonisation and interoperability. Semantic annotation of the services has been further progressed in parallel to optimization of various aspects such as the technical specifications of the programming interfaces, deployment options, the security environment and the discovery services developed in WP2 and standards and recommendations regarding file formats, ontology usage and technical and scientific descriptions proposed by WP3. In addition to these consortium-provided services, more services were added via the Associated Partner Programme and especially via the Implementation Challenge¹.

While in the first period up to the Deliverable 4.1 report [1], the services were mainly selected and prioritised according to their readiness levels and to provide good showcases of the capabilities and benefits of the OpenRiskNet infrastructure to a larger audience; the second period up to the D4.2 report was driven by the needs of the case studies; and the final period embraced the services provided via the Implementation Challenge. In this final report we present for all case studies (DataCure, MetaP, TGX, ModelRX, SysGroup, AOPLink, RevK) (see also Deliverable 1.3 [2] and Deliverable 1.5 for more information on the case studies), how the tools provided by the consortium partners are combined and integrated into workflows providing solutions to answer the specific case study questions and how the winning services of the Implementation Challenges are included to strengthen the case. This is followed by detailed descriptions of the implementation state of the services, of Implementation Challenge winners and of the areas covered by the integrated services.

¹ <https://openrisknet.org/associated-partner-programme/implementation-challenge/>

Integration of services in case studies

Seven case studies have been defined concentrating on different aspects of the risk assessment process based on the feedback from the requirement analysis and existing risk assessment frameworks [2] (**Figure 1**).

- DataCure: Data curation and creation of pre-reasoned datasets and searching
- ModelRX: Modelling for Prediction or Read Across
- SysGroup: A systems biology approach for grouping compounds
- MetaP: Metabolism Prediction
- AOPLink: Identification and Linking of Data related to AOP-Wiki
- TGX: Toxicogenomics-based prediction and mechanism identification
- RevK: Reverse dosimetry and PBPK prediction

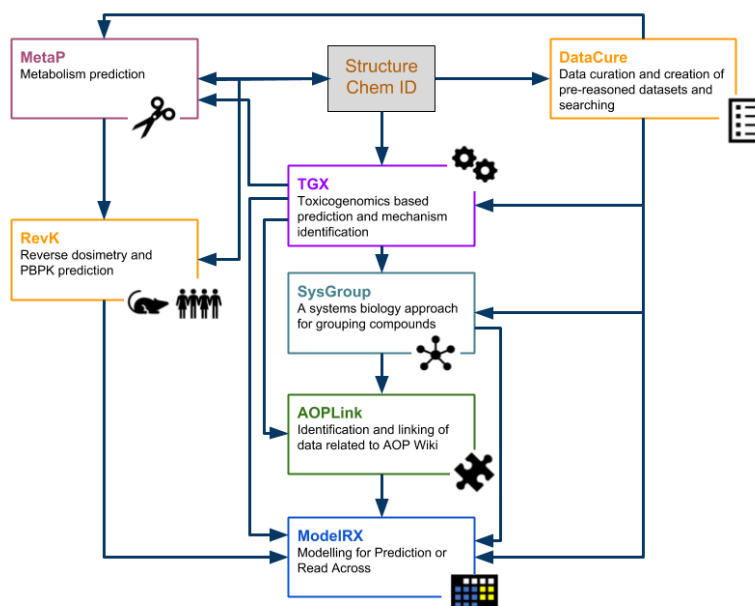


Figure 1. OpenRiskNet case studies²

For these case studies we will show how the integration of tools can lead to their innovative combination resulting in powerful workflows. The following descriptions include a status update report, involved partners and tools, workflows generated as well as the next steps if applicable. Tools fully integrated into OpenRiskNet and the corresponding consortium or associated partners are marked in bold below, to clearly identify these and separate them from tools needed to fill gaps not covered by OpenRiskNet yet.

² <https://openrisknet.org/e-infrastructure/development/case-studies/>

DataCure

In the Data curation and creation of pre-reasoned datasets and searching (DataCure³) case study, **EwC** (CS leader), **IM**, **Fraunhofer**, **UoB**, **UM** and **NTUA** work together on developing a system to allow users access to different data sources through the OpenRiskNet service. **EwC** and **UM** were representing data providers, **IM** developed data curation and quality control workflows, **Fraunhofer** and **UoB** extracted data from literature and generated pre-reasoned datasets while **UM** and **NTUA** brought in the user perspective in the need of data for predictive model generation, bioinformatics workflows and validation of key events of adverse outcome pathways.

This case study shows examples of practical implementations of data curation from various sources to be used by the remaining use cases. This includes 1) the identification of chemicals of concern, 2) the extraction of data (including merging of the data if coming from different sources), and 3) the collection of existing and supporting information (through text mining). All of the aforementioned steps are facilitated by the semantic annotation and API definition to be done for selected databases that are used in this case study and eventually for any new data to be accessed/analysed through the OpenRiskNet service.

The following data sources are used for this case and span the spectrum from biological interactions with chemicals to the actual physchem properties. They include toxicology sources of omics data like **TG-GATEs** and **DrugMatrix**, of high-throughput data like **ToxCast/Tox21** and ChEMBL, and of literature-curated data like Liver Toxicity Knowledge Base (LTKB), chemical information sources like ChEMBL and PubChem, and literature repositories such as PubMed and **ToxPlanet**. Specifically, after annotation and pre-processing (e.g. normalisation, filtering of transcriptomics data), these data are used for prediction analyses e.g. genotoxicity, carcinogenicity and/or improved grouping of compounds e.g. organ toxicity, drug-induced liver injury under the TGX and SysGroup case studies, respectively.

While provision of primary experimental data is a major goal of OpenRiskNet realised e.g. by the **EdelweissData system** by **EwC** and the case study shows how such data sources are brought into the OpenRiskNet ecosystem and further processed to combine different datasets to provide a large curated source for modelling or to provide all data to a weight-of-evidence risk assessment approach via **Jupyter** and **Squonk notebooks** by **IM**, to make use of legacy data, text mining can be applied to retrieve useful information from scientific papers via PubMed and regulatory reports via ToxPlanet. This is done through the API of the **SCAView**⁴ semantic text search and retrieval engine by **Fraunhofer**. The system can be queried for any compound of interest and a relevant context, e.g. 'benzene' AND 'carcinogen'. It will bring back all PubMed abstracts containing that piece of information. In the next step the retrieved documents can be further analysed via a text mining workflow (c.f. Ontology Annotation Services). An example for a text mining workflow is: "*find the relevant sentences, extract from them the relationships and properties of that chemical mentioned*". That workflow has to be predefined and started on the OpenRiskNet e-infrastructure. The workflow can then be accessed via a 'to be defined API' (e.g. input: document-id or document-text, output: json, csv or triplestore) for use by other workflows in the OpenRiskNet infrastructure.

³

<https://openrisknet.org/e-infrastructure/development/case-studies/case-study-datacure/>

⁴ <http://api.scaiview.com/swagger-ui.html>

MetaP

In the Metabolism Prediction (MetaP⁵) case study, **VU** (CS leader), **UU**, **HITeC/UHH**, (Associated Partner and Implementation Challenge winner), and **JGU** have worked together on integrating tools for metabolite and site-of-metabolism (SOM) prediction. The objective of this case study was to enable and facilitate metabolite prediction within the OpenRiskNet infrastructure and to evaluate and demonstrate the added value of these predictions. For a given (set of) query molecule(s), sites-of-metabolism (SOMs) and/or metabolites can be predicted that can serve as input for other case studies. For that purpose, we have incorporated and combined ligand-based metabolite predictors (e.g., **MetPred (UU)**, **FAME (HHU/HITeC)**, **SMARTCyp (external service integrated by VU)**) as well as protein-structure and -dynamics based approaches (**Plasticity tools (VU)**) to predict the SOMs by Cytochrome P450 (CYP450s), which metabolise approximately 75% of the currently marketed drugs. It is expected that this work will be continued after the end of the project to make services available for the prediction of microbial biotransformation pathways by integrating the enviPath data and software developed in part by **JGU**.

During method development, model calibration and validation, we used databases such as XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites. To enable and facilitate combined use of the metabolite prediction approaches and their outcomes, we benefited from ongoing development in workflow management systems and we made **Jupyter Notebooks** available to facilitate collection and visualization of results from the different available services. We illustrated the added value of having multiple predictors available and the corresponding Jupyter notebooks cover a pilot study on retrospective consensus prediction of known SOMs for drug compounds for which possible metabolite-associated toxicity was previously reported. We anticipate to present our results in an upcoming manuscript on tool integration, which will illustrate how using several predictors can provide additional value (compared to using individual tools) to (site-of-)metabolism prediction.

The following table gives an overview of metabolite prediction tools that are integrated and have been used in this case study. During method development, model calibration, and validation, advantage was taken of data from different databases for drugs, xenobiotics and their respective metabolites, as available in ZINC, ChEMBL, DrugBank, EAWAG-BBD and/or the **SMARTCyp** and **FAME** suites. Integration of enviPath (envipath.org) is still ongoing, which is a database and prediction system for microbial biotransformation of organic environmental contaminants.

Currently available MetaP tools

Tool	Input	Output	Method
MetPred (UU)	2D chemical structure of ligand	SOMs with Reaction Types for Phase I reactions	Preprocess Metabolite reaction database (>100K biotransformations) using MCS. For each query compound, look up similar atom environments based on circular fingerprints and use ReactionSMARTS to identify reaction types. See https://metpred.prod.openrisknet.org/

⁵ <https://openrisknet.org/e-infrastructure/development/case-studies/case-study-metap/>

FAME 3 (HHU/HITeC)	2D chemical structure of ligand	SOMs for Phase I, Phase II, or combined Phase I/II metabolism	Machine learning using 2D-circular-environment based atomic descriptors.
SMARTCyp 2.0 (external)	2D chemical structure of ligand	Rank atoms (SOMs) for P450-isoform specific reactions	Combining reactivity (from database on QM calculated transition state energies) with simple 2D molecular accessibility descriptors for SOM prediction. See smartcyp.sund.ku.dk/mol_to_som
Plasticity tools (VU)	3D Chemical structure of ligand	Prediction of most probable SOMs for P450-isoform specific reactions	Protein-structure and dynamics based prediction of substrate binding orientations and corresponding SOM in the active site of CYP isoforms (1A2, 2D6, 3A4).

As summarised in the table above, several services were made available in the MetaP case study. The listed services offer their functionality through RESTful APIs that are formalised according to OpenAPI specifications. The APIs are built using the Swagger toolchain and subsequently enable direct user interaction with the API endpoints using a browser-based User Interface (the Swagger UI). In addition, **MetPred (UU)** and **SMARTCyp (external service integrated by VU)** offer a custom browser-based interface to their service (see the links in the table above). The APIs enable access to the core features of the services as summarised above, and typically accept submissions of chemical structures in common file formats.

API endpoint input and output data exchange is standardised to a machine-readable JSON format. Together, the OpenAPI data type definitions and JSON-LD data annotations ensure seamless integration of the containerised services in the OpenRiskNet infrastructure and data exchange with other services.

Jupyter Notebooks that demonstrate the use of the listed services are freely available on GitHub⁶.

⁶ <https://github.com/OpenRiskNet/notebooks/>

TGX

In the toxicogenomics-based prediction and mechanism identification (TGX⁷) case study **UM** (CS leader), **VU**, and **CRG** are applying transcriptomics-based hazard prediction models for identification of specific molecular initiating events (MIE) based on (A) top-down and (B) bottom-up approaches. In this case study we focused on two top-down approaches.

For the first top-down approach, **UM** created a workflow from the earlier publication “A transcriptomics-based *in vitro* assay for predicting chemical genotoxicity *in vivo*” by Magkoufopoulou *et al.* (2012) [3], thereby reproducing their work as proof of principle. The workflow was created for one of the three approaches that were described in the study, using the available tools and methods. We followed guidelines described in the publication, but there were also some difficulties such as ambiguous descriptions of some techniques and a few omissions of software versions. Nevertheless, we were able to reproduce the results of that study. The workflow created using the Snakemake workflow manager is available from a GitLab software repository⁸, where every step is clearly described in Snakefile to reproduce the approach described by Magkoufopoulou *et al.* (2012) [3] and is used as a reference for transforming into an OpenRiskNet-based solution. The repository also includes required scripts as well as description of the steps necessary to reproduce the results. To make use of the harmonization and interoperability of OpenRiskNet, **CRG** converted the Snakemake-based workflow to a **Nextflow-based workflow**⁹. The Nextflow version uses containerised steps, thus making it easier to deploy on any cloud infrastructure, and applicable to OpenRiskNet Virtual Environments. Sharing the approaches in form of reusable workflows with the integrated services additionally improves the reproducibility of the research and will show that workflows written in different workflow-managers can inter-operate well. The Nextflow-based workflow has been translated into a more generic approach so that it can be applied to other toxicogenomics studies.

In the second top-down approach, **UM** collected transcriptomics data from the diXa data warehouse, NCBI GEO and EBI’s ArrayExpress on human, mouse and rat *in vitro* liver cell models, exposed to hundreds of compounds using the workflow from the first top-down approach. Merging of the obtained datasets per species could not be automated because of differences in the description of the datasets, e.g. differences in used ontologies, different metadata file formats, etc. Therefore, this step was carried out manually, after which the merged datasets were normalized and log₂ fold changes between exposure and matching controls were obtained. For all the compounds used in the experiments genotoxic and carcinogenic information was gathered from literature and several databases. Here contributions from **ToxPlanet** (Associated Partner and Implementation Challenge winner), whose access to hundreds of relevant databases/collections, and from **Fraunhofer**, whose well established text mining approaches via **SCAlview**, developed in collaboration with the DataCure case study, provided part of these toxicological reference data. The normalized transcriptomics data (per species) and genotoxicity/carcinogenicity information are ready to be fed into the prediction models of ModelRX. The transcriptomics data from the human, mouse, rat *in vitro* liver cell models and the toxicological information are available through OpenRiskNet¹⁰. The latter will also be

⁷ <https://openrisknet.org/e-infrastructure/development/case-studies/case-study-tgx/>

⁸ https://gitlab.com/bayjan/openrisknet_magkoufopoulou

⁹ <https://github.com/openrisknet/nf-toxomix>

¹⁰ https://gitlab.com/bayjan/openrisknet_meta_analysis_data

incorporated in **ToxicoDB** from **University Health Network** (Associated Partner and Implementation Challenge winner).

ModelRX

In the Modelling for Prediction or Read Across (ModelRX¹¹) case study, **NTUA** (CS leader), **JGU** and **UU** worked together in the workflow that involved (i) obtaining an initial (*training*) dataset within the OpenRiskNet environment, (ii) using tools to construct a predictive model, and (iii) applying the model to new datasets in order to generate predictions.

The ModelRX case study built on the tools **Jaqpote**, **Lazar** and **WEKA**, and **CPSign** by partners **NTUA**, **JGU** and **UU**, respectively. Work by partners led to the expansion of the offering to its current state. Functionality in the level of API (Application Programming Interface) is provided by the following tools:

Tool	URI
Jaqpote API	https://api-jaqpote.prod.openrisknet.org/jaqpote/swagger/#/
Lazar API	https://lazar.prod.openrisknet.org/
JGU WEKA Rest service	https://jguweka.prod.openrisknet.org/
MetPred	https://metpred.prod.openrisknet.org/
cplogd	https://cplogd.prod.openrisknet.org/
Ltkb models	http://ltkbnolelessvsmostcpsign.prod.openrisknet.org http://ltkbnolevslessmostcpsign.prod.openrisknet.org http://ltkbnolevsmostcpsign.prod.openrisknet.org
Blood-brain-barrier models	http://blood-brain-barrier-penetration-cvap-cpsign.prod.openrisknet.org http://blood-brain-barrier-penetration-cpsign.prod.openrisknet.org

APIs are more directed to developers, providing them with means for easy integration into their own services and use within external apps. In order to address the needs of the community more completely, OpenRiskNet members offer tools that can be used through a GUI (Graphical User Interface):

Tool	URI
Jaqpote GUI	https://ui-jaqpote.prod.openrisknet.org/
Lazar GUI	https://lazar.prod.openrisknet.org/prediction
Squonk	https://squonk-notebook.prod.openrisknet.org/portal
Jupyter	https://jupyterhub-jupyter.prod.openrisknet.org/hub/login

¹¹ <https://openrisknet.org/e-infrastructure/development/case-studies/case-study-modelrx/>

CPSign	http://modelingweb.prod.openrisknet.org
--------	---

We worked on the integration of our services through **Jupyter notebooks** that were developed as a central point to access and use the services from OpenRiskNet partners, from accessing data to developing models, and generating predictions. For that purpose, we gathered the work in a GitHub repository¹², so we could converge quickly to a solution that takes advantage of partners' tools within a common framework.

Work in ModelRX is in close connection with the DataCure case study, both in terms of ModelRX receiving and making use of the curated data from DataCure and in terms of ModelRX supporting similarity identification to DataCure by providing tools for calculating theoretical descriptors of substances.

Additionally, work with associate partners assisted the propagation of OpenRiskNet solutions and enhanced the offering. We worked with three winners of the first Implementation Challenge: **NovaMechanics Ltd.** have developed the **Enalos InSilicoNano platform** and have experience in the risk assessment of nanomaterials, **BIGCHEM GmbH** offers **OCHEM**, a tool for descriptor calculations and model development and the **Korea Institute of Technology** that made available a **NanoQSAR model** for prediction of cytotoxicity caused by metal and metal oxide nanoparticles. More information is provided in the Implementation Challenge section below.

¹² <https://github.com/OpenRiskNet/notebooks>

SysGroup

In the systems biology approach for grouping compounds (SysGroup¹³) case study, **UM** (CS leader) used an approach to identify groups of hepatotoxicant compounds based on their similarities in the mechanisms of action (omics-based), combined with chemical structure information (chemoinformatics).

In this case study, **TG-GATES** transcriptomics data from primary human hepatocytes exposed to 139 compounds for 24h and the highest dosage were used. This dataset was obtained from the OpenRiskNet service “Transcriptomics data from human, mouse, rat in vitro liver models”¹⁴. For the 139 compounds, PubChem¹⁵ was used to obtain 2D Tanimoto scores and PIDGIN3¹⁶ was used to retrieve ligand-based target predictions. Integration of the transcriptomics data with the chemoinformatics data was performed using iClusterPlus¹⁷, an integrative clustering framework developed to integrate diverse data types (i.e. binary, categorical, and continuous values) by a latent variable approach. In order to guarantee that each data type has the same weight in the analyses, scaled Euclidean distances were used for the transcriptomics profiles and ligand-based target predictions in the range of 0–1 (0 = most similar; 1 = most dissimilar), and for structural similarities the 2D Tanimoto scores were converted to the same range as the other data types, by subtracting the Tanimoto score from 1, i.e. they were used in the range 0–1 (0 = most similar; 1 = most dissimilar). For the iClusterPlus analysis, a fixed number of clusters was defined based on past performance from the DECO2 project (Cefic-LRI AIMT4¹⁸), i.e. 46 clusters (~1/3 of 139). The grouping of the chemicals in these clusters needs to be further investigated using toxicological data from the DataCure case study and various OpenRiskNet services, such as **ToxicoDB** and **ToxPlanet**. However, due to time constraints this final step was not achieved within the time frame of the project.

The workflow used in this case study was setup using GNU Make¹⁹ and is available through OpenRiskNet’s GitHub²⁰.

¹³

<https://openrisknet.org/e-infrastructure/development/case-studies/case-study-sysgroup/>

¹⁴ <https://openrisknet.org/e-infrastructure/services/164/>

¹⁵ https://pubchem.ncbi.nlm.nih.gov/score_matrix/score_matrix.cgi

¹⁶ <https://pidginv3.readthedocs.io/en/latest/>

¹⁷ <https://bioconductor.org/packages/release/bioc/html/iClusterPlus.html>

¹⁸ <http://cefic-lri.org/projects/aimt4-um-deco2-moving-from-deco-towards-oecd/>

¹⁹ <https://www.gnu.org/software/make/>

²⁰ https://github.com/OpenRiskNet/notebooks/tree/master/openrisknet_sysgroup

AOPLink

In the Identification and Linking of Data related to AOP-Wiki (AOPLink²¹) case study, **UM** (CS leader) aims to link experimental data to existing Adverse Outcome Pathways (AOPs) on AOP-Wiki. This was done with **Jupyter workflows** using the various annotations of chemicals, genes and processes that are present in the **AOP-Wiki** to find experimental data to support the biological plausibility of individual Key Events, or full AOPs. The other way around, experimental data of a stressor not yet described in the AOP-Wiki can be used to identify affected cellular processes within Key Events and AOPs, based on altered gene expression levels associated with the Key Event or from assays directly measuring the Key Event.

The central component in the AOPLink case study is the **AOP-Wiki**, the main repository for qualitative descriptions of AOPs. For this case study, an essential step was exposing the complete database in a query service as the existing repository does not provide such service. This led to the development of the AOP-Wiki Resource Description Framework (RDF) (paper in progress) from the existing XML data, followed by exposing the RDF in a Virtuoso SPARQL endpoint.

Another AOP-related repository is the **AOP-DB**, a repository developed at the **US EPA** that expands AOP knowledge with information about diseases, **ToxCast** assays, genes, etc. As a winner of the first round of the Implementation Challenges with the application by Holly Mortensen, parts of the SQL data of the AOP-DB have therefore also been converted into RDF as part of the Implementation Challenge. Just like for the AOP-Wiki RDF, the AOP-DB RDF has been exposed in a Virtuoso SPARQL endpoint service in the OpenRiskNet e-Infrastructure.

Because there is a focus on linking experimental data to AOPs and its components, the linkage of gene expression data was also done using **WikiPathways**, a molecular pathway database. This repository, exposed as RDF in a Virtuoso SPARQL endpoint just like AOP-Wiki and AOP-DB, can make the connection between the generalized descriptions of processes in the AOP-Wiki, and gene expression data such as transcriptomics. Gene expression data can be used to perform pathway analysis with this service as it includes detailed annotations of chemicals, genes and proteins for a large variety of species.

In order to integrate these different repositories and other tools to define a workflow and answer various research questions, **Jupyter notebooks** have been used in which different tools in AOPLink and DataCure were used to make the links between AOPs and experimental data. An essential tool for the integration of those services is facilitated by allowing the conversion of identifiers across databases for chemicals, genes and proteins. Therefore, a **BridgeDb** service was added in the OpenRiskNet e-Infrastructure.

To extend the applicability of this case study to the growing efforts in nanomaterial toxicity, the **eNanoMapper database** is added as a service that allows export and import of data related to nanomaterials.

Overall, the AOPLink case study is closely connected with the services developed in the DataCure case study, which provides access to a variety of experimental data sources such as **ToxCast** and **TG-GATES**, and this integration of services allows the searching for data related to AOPs, and *vice versa*.

²¹ <https://openrisknet.org/e-infrastructure/development/case-studies/case-study-aoplink/>

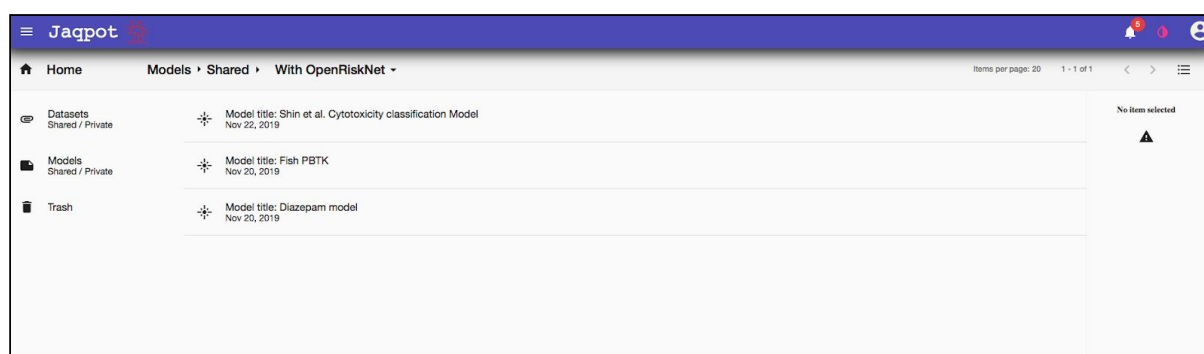
RevK

In the Reverse dosimetry and PBPK prediction (RevK²²) case study, **INERIS** (CS leader) and **NTUA** provide a web interface to physiologically-based pharmacokinetic models for forward and reverse dosimetry calculations. Forward calculations compute internal concentrations from given exposure doses, while Reverse calculations compute exposure doses from internal concentrations or measured biomarker levels.

The RevK case study builds on the **Jaqpot** web-based computational platform, providing access to relevant software used by the community or in-house client, developed within RevK, as shown in the following table:

Software	Scope	
httk https://cran.r-project.org/web/packages/httk/index.html	High-throughput toxicokinetic R package	API/GUI
In-house R PBPK client	Custom PBPK modelling workflow, developed by NTUA	GUI (R commands lead to model accessible in GUI)
PKSim http://www.systems-biology.com/products/pk-sim.html	Whole-body physiologically based pharmacokinetic modeling developed by BAYER AG, made Open Source in recent years.	API

Models made available over the graphical user interface of Jaqpot have been shared within Jaqpot through the OpenRiskNet organisation, as shown below (please note the two RevK models, “**Fish PBTK**” and “**Diazepam model**”):



Within RevK, INERIS and NTUA employed Jaqpot to provide access to the software used in each example and also demonstrate the usage with examples:

- **httk**: We present an example with Imazalil, developing the models as web services through the Jaqpot API or the Jaqpot GUI.
- **In-house R PBPK client**: Two examples are provided, a PBPK model for Diazepam in humans, and a PBPK model in fish that is not substance-specific.

²² <https://openrisknet.org/e-infrastructure/development/case-studies/case-study-revk/>

- **PKSim:** We describe the integration of a PBPK model for Theophylline, showing model development as a web service through the Jaqpot API.

As outlined in Figure 1, the work on RevK is closely connected to the ModelRX case study that can make use of the simulation results and insights gained from the effects of certain drugs and chemicals in the organisms being studied, in order to make decisions about optimal or new options. Also shown is the cooperation with the MetaP case study for the predictions regarding organisms' metabolism and the implications of compounds may have on them.

Integration status of consortium-provided services

Since integration of services involves different steps, including API development, packaging into containers and container orchestration, semantic annotation and listing in the OpenRiskNet and other service catalogues, becoming OpenRiskNet-compliant is not a one-time activity but is a process of different stages including the following operations:

1. Utilising the OpenRiskNet APIs to ensure that each service is accessible to our proposed interoperability layer;
2. Annotating the services according to the semantic interoperability layer concept using defined ontologies;
3. Containerising the services for easy deployment in virtual environments of OpenRiskNet instances;
4. Documenting the scientific and technical background;
5. Deploying the service into the OpenRiskNet reference environment or a publicly available service linked into the OpenRiskNet service registry;
6. Listing the service in the OpenRiskNet discovery services;
7. Listing in other central repositories like eInfraCentral, bio.tools and TeSS (ELIXIR);
8. Providing legal and ethical statements on how the service can be used.

Many of these steps were themselves not one-time efforts but had to be continuously adapted to the progress of the project. Specifically, the bottom-up API concept described in Deliverable 2.2 [4] lists specific steps, which had to be performed to harmonise the existing APIs and annotate them semantically. Additionally, with the combination of the industry standard OpenAPI (formerly Swagger) and the semantic web standard JSON-LD, we have now defined the way the semantic interoperability layer is working technically. To monitor the progress of implementation of specific services, the status of each service with respect to its compliance to the current realisation of the OpenRiskNet concepts, described by the eight operations specified above, is listed in the Service Catalogue²³ on the OpenRiskNet website. Examples of catalogue entries are shown in **Figure 2** indicating two status types, i.e. integrated and in progress and in **Figure 3** showing more details for a fully integrated service. All services provided by the OpenRiskNet consortium are now compliant at least up to operation 6.

For operation 7, we are in contact with eInfraCentral and EOSC-hub (see also Deliverable 3.5 on dissemination and community activities) how to optimise the listings for larger sets of smaller services addressing a specific field and how these can then become part of the EOSC marketplace. However, a general listing of OpenRiskNet and some specific services combining a larger set of functionalities like Jaqpot are available in eInfraCentral and the EOSC service catalogue. For operation 8, all services have clear legal statements (licenses and terms of use). As described in detail in Deliverable 3.6, ethics statements need a harmonisation in a global, pan-European setting and the OpenRiskNet consortium is actively contributing and will continue to do so to establish ethics guidelines for data

²³ <https://openrisknet.org/e-infrastructure/services/>

management and sharing and ways how to document these with the data in data warehouses. Some of the services of the Implementation Challenge winners are still labeled as in development. This is caused by the fact that the Implementation Challenge project concentrated on the aspects with most beneficial for the winner allowing the service to be findable, accessible and combinable with other OpenRiskNet services. Based on the feedback from the winners acknowledging the benefits of being part of the OpenRiskNet infrastructure, it is guaranteed that work to achieve full integration will be continued beyond the OpenRiskNet project.

Figure 2. Screenshot of OpenRiskNet services accessible on OpenRiskNet Services. The status of the services, either integrated or in progress, is indicated by red circles. Service type and User type which can be selected from the drop down menus are directly indicated (see arrows).

ToxCast/Tox21 V3.2

ToxCast and Tox21 datasets (raw and summary) extracted from the MySQL database provided by US EPA

The dataset as provided by US EPA were transformed and are now available in the EdelweissData system for easy access via APIs. The most current version is 3.2. Data of version 3.1 is also available. ToxCast: Data for approximately 1,800 chemicals from a broad range of sources including industrial and consumer products, food additives, and potentially green chemicals that could be safer alternatives to existing chemicals is provided. These chemicals were screened in more than 700 high-throughput assay endpoints that cover a range of high-level cell responses. Tox21: Using a high-throughput robotic screening system housed at NCATS, researchers are testing 10,000 environmental chemicals (called the Tox21 10K library) for their potential to disrupt biological pathways that may result in toxicity. Screening results help the researchers prioritize chemicals for further in-depth investigation.

[Go to service →](#)

✓ For end-users ✓ For developers

Type: Database / data source

API Type: REST under OAS3 specification, REST, OpenAPI

Categories: Toxicology, chemical properties and bioassay databases

Applicability domain: Computational modelling, Toxicology, Predictive toxicology

Topic: Bioassay, Risk assessment, Information extraction

Biological area: NOAEL/LOAEL

Targeted industry: Other consumer products, Food, Cosmetics, Drugs, Chemicals

Targeted users: Regulators, Software Developers, Students, Researchers, Risk assessors

Relevant OpenRiskNet case study: [DataCure](#) - Data curation and creation of pre-reasoned datasets and searching

Documentation: <https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data>

Provided by: Edelweiss Connect GmbH

Contact: openrisknet@edelweissconnect.com

Licence: Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Login required: No

Implementation status: Graphical user interface available

Technology readiness level: TRL 8 – system complete and qualified

Integration status: Integrated application

Service integration operations completed:

- ✓ Utilises the OpenRiskNet APIs to ensure that each service is accessible to our proposed interoperability layer.
- ✓ Is annotated according to the semantic interoperability layer concept using defined ontologies.
- Is containerised for easy deployment in virtual environments of OpenRiskNet instances.
- ✓ Has documented scientific and technical background.
- Is deployed into the OpenRiskNet reference environment.
- ✓ Is listed in the OpenRiskNet discovery services.
- Is listed in other central repositories like einfraCentral, bio.tools and TeSS (ELIXIR).
- ✓ Provides legal and ethical statements on how the service can be used.

Resources & Training

Case Study description - Data curation and creation of pre-reasoned datasets and searching [DataCure]

Report

10 May 2019



Figure 3. Screenshot of ToxCast/Tox21 V3.2 service accessible on OpenRiskNet Services showing detailed information and links to the service, relevant case studies and resources & training. Completeness of service integration operations are also indicated.

Table 1 provides additional information on each service including the license type and the API version available together with the link to the service API specification on the reference environment. More information on the services especially their service category, applicability domain, user focus and use in case studies is given in the **Categorisation of services** section and in **Table 2** below.

Table 1. Implementation status, licence information and API information

Name	Implementation status	Licence type	License	API Type	API URL
cpLogD confidence predictor for logD (UU)	Available as web service, Application programming interface available	Open source (wrapper), Commercial (CPSign)	Apache	REST under OpenAPI2 specification, REST, OpenAPI	https://cplogd.prod.openrisknet.org/
Daphnia magna nanotoxicity database (KIT and EwC)	Graphical user interface available, OAS v3, Available as web service, Application programming interface available	Open source	Attribution 4.0 International (CC BY 4.0)	REST under OAS3 specification, REST, OpenAPI	https://api.staging.kit.cloud.douglasconnect.com/datasets/ab2bc1ee-99dc-4ddf-b1f9-9fdeb8a0f48c/versions/1
ToxCast/Tox21 V3.2 (EwC)	Graphical user interface available, OAS v3, Available as web service, Application programming interface available	Open source	Attribution-Share Alike 4.0 International (CC BY-SA 4.0)	REST under OAS3 specification, REST, OpenAPI	https://ui.staging.kit.cloud.douglasconnect.com/datasets
EdelweissData serving ToxCast, ToxRefDB and TG-GATEs data (EwC)	API documentation available (Swagger-OpenAPI v2), Graphical user interface available, Available as web service	Open source	Attribution	REST	https://toxcast-api.cloud.douglasconnect.com/beta/ui/
Jaqpote API (NTUA)	API documentation available (Swagger-OpenAPI v2), Containerised, Available as web service, Application programming interface available	Open source	GNU General Public License	REST under OpenAPI3 specification	https://api-jaqpote.prod.openrisknet.org/jaqpote/swagger/
Jaqpote GUI (NTUA)	Graphical user interface available	Open source	GNU General Public License	Based on Jaqpote API that uses REST under OpenAPI3	https://ui-jaqpote.prod.openrisknet.org/

				specification	
Lazar Toxicity Predictions (JGU)	Containerised, Graphical user interface available, Available as web service	Open source	GNU General Public License	REST, OpenAPI	https://lazar.prod.openrisknet.org/
TG-GATEs and DrugMatrix (EwC)	Graphical user interface available, OAS v3, Available as web service, Application programming interface available	Open source	Attribution-Share Alike 4.0 International (CC BY-SA 4.0)	REST under OAS3 specification, OAS3, REST, OpenAPI	https://ui.staging.kit.cloud.douglasconnect.com/dataset
Conformal Blood Brain Barrier Penetration Classification (UU)	Containerised, Graphical user interface available, Available as web service	Open source	-	REST under OAS3 specification, OAS3, REST, OpenAPI	http://blood-brain-barrier-penetration-cpsign.prod.openrisknet.org/
CVAP Blood Brain Barrier Penetration Classification (UU)	Containerised, Graphical user interface available, Available as web service	Open source	-	REST under OAS3 specification, OAS3, REST, OpenAPI	http://blood-brain-barrier-penetration-cvap-cpsign.prod.openrisknet.org/
Nanoinformatics Tool for the Virtual Screening of Metal Oxide Nanoparticles via Enalos Platform (NovaMechanics)	API documentation available (Swagger-OpenAPI v2), Application programming interface available	Open source	-	REST	https://model-nova-mech.prod.openrisknet.org/novamechanics/swagger-ui.html#!/rest--qnar--controller/ionOxideUploadFileUsingPOST
nanoQSAR model for the prediction of the cellular uptake and the virtual screening of nanoparticles via Enalos Platform (NovaMechanics)	API documentation available (Swagger-OpenAPI v2), Application programming interface available	Open source	-	REST	https://model-nova-mech.prod.openrisknet.org/novamechanics/swagger-ui.html#!/rest--qnar--controller/pacaUploadFileUsingPOST
Nano-QSAR to predict cytotoxicity of metal and metal oxide nanoparticles (KIT and NTUA)	API documentation available (Swagger-OpenAPI v2), OAS v3, Available as web service, Application programming interface available	Open source	-	REST, OpenAPI, REST under OAS3 specification, OAS3, Based on Jaqpot API that uses REST under OpenAPI3 specification	https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/
FAME 3 site-of-metabolism predictor (HITeC)	API documentation available (Swagger-OpenAPI v2), Containerised,	Open source	-	OpenAPI	http://fame3.dev.openrisknet.org/swagger-ui.html

	Available as web service, Application programming interface available				
ltkbnovsmost (UU)	API documentation available (Swagger-OpenAPI v2), Containerised, Available as web service	Open source	-	REST under OpenAPI2 specification	http://ltkbnovsmost-cpsign.prod.openrisknet.org/
ltkbnovslessmost (UU)	API documentation available (Swagger-OpenAPI v2), Containerised, Available as web service	Open source	-	REST under OpenAPI2 specification	http://ltkbnovslessmost-cpsign.prod.openrisknet.org/
ltkbnovslessmost (UU)	API documentation available (Swagger-OpenAPI v2), Containerised, Available as web service	Open source	-	REST under OpenAPI2 specification	http://ltkbnovslessmost-cpsign.prod.openrisknet.org/
ChemIdConvert (EWC)	API documentation available (Swagger-OpenAPI v2), Application programming interface available	Open source	Attribution-Share Alike	REST	http://orn-chemidconvert.prod.openrisknet.org/v1/ui/
OpenRiskNet service registry (EWC)	Available as web service	Open source (login required)	Apache	-	http://registry.prod.openrisknet.org/
JGU WEKA REST Service (JGU)	Available as web service, Containerised	Open source	GNU General Public License	REST, OpenAPI	https://jguweka.prod.openrisknet.org/
Jupyter Notebooks (external)	Containerised, Graphical user interface available, Available as web service	Open source (login required)	-	-	https://jupyter.prod.openrisknet.org/
MetPred(UU)	Available as web service, Application programming interface available	Open source (wrapper), Commercial (database)	Apache	REST under OpenAPI2 specification, REST, OpenAPI	https://metpred.prod.openrisknet.org/
OpenShift Console (external)	Graphical user interface available, Available as web service	Open source (login required)	-	Rest	https://docs.okd.io/3.7/rest_api/index.html
P450 SOM predictor (VU)	Containerised	Open source	Apache, GPL v2.0, and custom academic for docking tool	OpenAPI	http://mdstudio-smartcyp.prod.openrisknet.org/ui

			(www.tcd.uni-konstanz.de/plants_download/)		
SSO Console (external)	Graphical user interface available, Available as web service	Open source (login required)	-	Rest	https://www.keycloak.org/docs-api/8.0/rest-api/index.html
Squonk Computational Notebook (IM)	Graphical user interface available	Open source (login required)	Apache 2.0	REST	https://squonk-notebook.prod.openrisknet.org/portal
PySquonk (Diamond Light Source Ltd. and IM)	-	Open source	-	-	https://github.com/InformaticsMatters/pysquonk
Transcriptomics data from human, mouse, rat in vitro liver cell models (UM)	-	Open source	-	-	-
WikiPathways SPARQL Endpoint (UM)	Available as web service	Open source , Open data	GNU Lesser General Public License (LGPL) 2 (LGPLv2.0)	SPARQL	http://wikipathways.prod.openrisknet.org/sparql/
BridgeDb identifier mapping service (UM)	Available as web service	Open source	Apache	REST under OpenAPI2 specification	http://bridgedb.prod.openrisknet.org/swagger/
ToxPlanet (ToxPlanet)	API documentation available (Swagger-OpenAPI v2), OAS v3, Available as web service, Application programming interface available	Commercial	-	REST under OpenAPI2 specification, REST, OpenAPI, REST under OAS3 specification, OAS3	-
OCHEM models (BIGCHEM)	Application programming interface available	Open source	Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)	REST	https://ochem.eu/home/show.do
AOP-DB SPARQL endpoint (US EPA and UM)	Available as web service	Open source	GNU General Public License	SPARQL	http://aopdb-rdf.prod.openrisknet.org/sparql/
eNanoMapper database (UM)	Graphical user interface available, API documentation available (Swagger-OpenAPI v2), Containerised	Open source	LGPL v2.0	REST	http://nanomaterialdb-test.prod.openrisknet.org/ambient/
AOP-Wiki SPARQL Endpoint (UM)	Available as web service	Open source, Proprietary data	GNU General Public License	SPARQL	http://aopwiki-rdf.prod.openrisknet.org/sparql/
The Adverse Outcome Pathway Database (AOP-DB) (US EPA)	Graphical user interface available	Open source	-	-	-
Fragnet Search	API	Open source (login required)	Apache 2.0	REST under	http://fragnet-sea

(IM)	documentation available (Swagger-OpenAPI v2), Containerised, Available as web service	required)		OpenAPI2 specification, REST, OpenAPI	rch.fragnet-search.svc:8080/
ToxicoDB (University Health Network and UM)		Open source	GNU Lesser General Public License 3 (LGPLv3.0)	REST	http://toxicondb.ca/api
Prosilico's Human In Vivo ADME/PK-Prediction Studio (Prosilico)	Containerised, Available as web service	Open source (login required)	-	-	-
EGA Beacon (CRG)	Graphical user interface available, Application programming interface available	Open source	-	REST OpenAPI2	https://beacon-api.ega-archive.org/
EGA Metadata API (CRG)	Application programming interface available	Open source	-	REST	https://ega-archive.org/metadata/v2/
SCAView (Fraunhofer)	Available as a web service, Application programming interface available, API documentation available (Swagger-OpenAPI v2), Containerised, Graphical user interface available	Fraunhofer license	free to academic use	REST under OpenAPI2 specification	http://api.scaiview.com/swagger-ui.html
Nextflow (CRG)	Containerised	Open source (login required)	Apache		http://www.nextflow.io/
TeMOwl (Fraunhofer)	API documentation available (Swagger-OpenAPI v2), Containerised, Available as a web service, Application programming interface available	Open source (login required)	-	REST under OpenAPI2 specification	https://sam1.api.scaiview.com/swagger-ui.html
ToxTargetLinks (Leiden University and UM)	Graphical user interface available	Open source	-	-	https://cytargetlinker.github.io/pages/linksets

Services selected within the Implementation Challenge

The Implementation Challenge²⁴ was created to select external tools especially in areas of risk assessment not completely covered by the OpenRiskNet consortium to be prioritised for their integration into the e-infrastructure. Third parties were able to apply for partial financial and strongly technical support by researchers and developers of OpenRiskNet consortium partners. Similar to the internal services, the funds and the technical support covered the work associated with making the service OpenRiskNet compliant. This includes the adoption of the OpenRiskNet API concept including the interoperability layer, generation of the data schemata for in- and output as well as the containerization for deployment. More details on the Implementation Challenge is available in Deliverable D1.4 and we will only highlight here the services, which were integrated. After three selection periods, 10 winners have been selected by the scientific advisory board and integration of the following services has been achieved:

1. **Hyun Kil Shin, Korea institute of Toxicology**
Daphnia magna nanotoxicity database: Daphnia magna assay results were compiled according to the test guidelines
nano-QSAR to predict cytotoxicity of metal and metal oxide nanoparticles: The model made available, uses quantum mechanical descriptors based on two molecular structures: cluster and hydroxyl metal coordination complex. Prediction outcome is divided into three groups: toxic, uncertain, and innocuous
2. **Matthias Timberlake, ToxPlanet**
ToxPlanet: ToxPlanet aggregates chemical hazard and toxicology content (literature & technical reports) from hundreds of reputable sources and provides decision-support solutions used by chemical safety professionals around the world
3. **Johannes Kirchmair: HITEC and Universität Hamburg**
Fast MEtabolizer (FAME): FAME is a machine learning model for the prediction of sites of metabolism in xenobiotics. The available version covers cytochrome P450 metabolism. A new version covering also other metabolizing enzymes is in its final stage of development.
4. **Antreas Afantitis, NovaMechanics Ltd**
Enalos InSilicoNano platform - nanoQSAR model: The work on a nanoQSAR model for the cellular uptake prediction and virtual screening via Enalos InSilicoNano platform enables the deployment of a validated predictive model for nanoparticles cellular uptake, that is part of the Enalos infrastructure, into the OpenRiskNet reference environment. The model can be used for the predictions of cellular uptake for new structures that are imported to the server via sdf format.
Risk Assessment Tool for the Virtual Screening of Metal Oxide Nanoparticles through Enalos: In this work, a predictive classification model for the toxicological assessment of iron oxide NPs with different core, coating and

²⁴ <https://openrisknet.org/associated-partner-programme/implementation-challenge/>

surface modification has been developed, based on a number of different properties including size, relaxivities, zeta potential and type of coating. The model is part of the Enalos infrastructure and has been implemented into the OpenRiskNet reference environment.

5. **Igor Tetko, BIGCHEM GmbH**

OCHEM models: OCHEM provides access to more than 100 models published on its web site for different endpoints. The implementation of this interface will allow other users execute predictions on the OCHEM website.

6. **Holly Mortensen, US EPA**

AOP-DB (The Adverse Outcome Pathway Database): The AOP-DB serves to link molecular targets identified as molecular initiating events (MIEs) and key events (KEs) in the AOP-Wiki (<https://aopwiki.org>) to publically available data (e.g. gene-protein, pathway, species orthology, chemical, disease), in addition to ToxCast assay information.

7. **Urban Fagerholm, Prosilico**

Prosilico Human ADME/PK Studio: Prosilico Human ADME/PK Studio is based on Prosilicos's unique database of human clinical data provides the foundation to develop a wide scope of relevant prediction models for human ADME/PK.

8. **Rachael Skyner, Diamond Light Source Ltd.**

PySquonk: PySquonk, in its current form, is an early attempt at using the python requests module to create a python wrapper around Squonk RESTful services.

9. **Katy Wolstencroft, Leiden University** (on behalf of the European Joint Programme on Rare Diseases)

ToxTargetLinks: The service allows to connect toxicological molecular initiating events to rare disease molecular networks.

10. **Benjamin Haibe-Kains, University Health Network**

ToxicoDB: ToxicoDB is a web-application based on the code base of the existing PharmacoDB web-application. The ToxicoDB web-app will provide an intuitive interface for all users (including users that are not computational savvy) to mine the complex toxicogenomic data. This will increase the visibility of the data and the curation efforts of the OpenRiskNet and facilitate access to toxicogenomic data and analytical tool to the whole scientific community.

As already visible from the case study descriptions above, the services are deeply embedded into the case studies, e.g. FAME and ToxPlanet into MetaP and DataCure, respectively. Other examples are OCHEM and Enalosfunctionalities as well as the nano-QSAR model used as additional examples to validate the ModelRX concepts. Finally, AOP-DB enriched the knowledge available to theAOPLink case study.

Categorisation of services

Even if the case studies show the functionality of many OpenRiskNet services and of the complete infrastructure, these and additional services and more general workflows are also offered to the community to facilitate the execution and improve the quality of risk assessments. To help the user to find the best tools for a specific task, all services are categorised with respect to the target user group, target industry, risk assessment area and application domain. All these categories are included in the descriptions in the OpenRiskNet service catalogue²⁵ and can be used as search facets to locate suitable services via the web interface or via querying the semantic annotation.

Categorisation by service type

To structure the work on service integration and to guarantee that all important areas are considered, service types corresponding to the different tasks for WP4 were proposed in the Description of Action:

- Task 4.1 Toxicology, Chemical Properties and Bioassay Databases
- Task 4.2 Omics Databases
- Task 4.3 Knowledge Bases and Data Mining
- Task 4.4 Ontology Services
- Task 4.5 Processing and Analysis
- Task 4.6 Predictive Toxicology
- Task 4.7 Workflows, Visualisation and Reporting

The services were categorised according to these tasks and linked to the case studies (see **Table 2**). Two services, EGA Beacon and EGA Metadata API, build the first links of the OpenRiskNet infrastructure to the clinical and human health area, demonstrating how these fields can be brought closer together with risk assessment and drug design. They do provide an important building block for working with sensitive genomic data in future safe-by-design approaches within virtual environments. However, due to this more global, community-overarching purpose of use, they were beyond the scope of what can be addressed by the OpenRiskNet case studies. Also the OpenRiskNet service registry, and the SSO and the OpenShift console have not been assigned to any case study since they are very general services to run the virtual environments.

The categories, in a slightly modified form to more reflect the needs of the users and not of the service providers (consortium partners), became one of the major search and browsing criteria in the OpenRiskNet Service Catalogue. **Figure 4** visualises the number of integrated services in each category listed in the Service Catalogue at the end of the project. A more fine-grained categorisation of the type of services is additionally given in **Figure 5**.

²⁵ <https://openrisknet.org/e-infrastructure/services/>

Table 2. Overview of services accessible on OpenRiskNet Services, the categories and case studies they belong to

Service	Category ^a								Case studies
	Task 4.1	Task 4.2	Task 4.3	Task 4.4	Task 4.5	Task 4.6	Task 4.7	Other ^b	
cpLogD confidence predictor for logD (UU)	✓							✓	ModelRX
Daphnia magna nanotoxicity database (KIT and EwC)	✓								DataCure
ToxCast/Tox21 V3.2 (EwC)	✓								DataCure, AOPLink
EdelweissData serving ToxCast, ToxRefDB and TG-GATEs data (EwC)	✓	✓							DataCure
Jaqpot API (NTUA)			✓		✓	✓		✓	ModelRX, RevK
Jaqpot GUI (NTUA)			✓		✓	✓	✓		ModelRX
Lazar Toxicity Predictions (JGU)	✓				✓	✓	✓		ModelRX
TG-GATEs and DrugMatrix (EwC)		✓							DataCure, TGX, AOPLink
Conformal Blood Brain Barrier Penetration Classification (UU)	✓					✓		✓	ModelRX
CVAP Blood Brain Barrier Penetration Classification (UU)	✓					✓		✓	ModelRX
Nanoinformatics Tool for the Virtual Screening of Metal Oxide Nanoparticles via Enalos Platform (NovaMechanics)						✓		✓	ModelRX
nanoQSAR model for the prediction of the cellular uptake and the virtual screening of nanoparticles via Enalos Platform (NovaMechanics)						✓			ModelRX
Nano-QSAR to predict cytotoxicity of metal and metal oxide nanoparticles (KIT and NTUA)						✓			ModelRX
FAME 3 site-of-metabolism predictor (HITeC)	✓				✓				MetaP
ltkbn-no-vs-most (UU)						✓		✓	ModelRX
ltkbn-no-vs-less-most (UU)						✓		✓	ModelRX
ltkbn-no-less-vs-most (UU)						✓		✓	ModelRX
ChemIdConvert (EwC)	✓							✓	DataCure
OpenRiskNet service registry (EwC)								✓	-
JGU WEKA REST Service (JGU)			✓		✓	✓		✓	MetaP, ModelRX
Jupyter Notebooks (external)					✓		✓		AOPLink, DataCure, MetaP, ModelRX, RevK, SysGroup, TGX
MetPred(UU)	✓							✓	MetaP
OpenShift Console (external)								✓	-
P450 SOM predictor (VU)					✓				MetaP
SSO Console (external)								✓	-
Squonk Computational Notebook (IM)					✓		✓		AOPLink, DataCure, MetaP, ModelRX, RevK, SysGroup, TGX
PySquonk (Diamond Light Source Ltd. and IM)					✓				DataCure, MetaP, ModelRX

Transcriptomics data from human, mouse, rat in vitro liver cell models (UM)	✓	✓							AOPLink, DataCure, ModelRX, RevK, SysGroup, TGX
WikiPathways SPARQL Endpoint (UM)			✓						AOPLink
BridgeDb identifier mapping service (UM)				✓				✓	AOPLink
ToxPlanet (ToxPlanet)	✓		✓						DataCure, TGX, SysGroup
OCHEM models (BIGCHEM)	✓								ModelRX
AOP-DB SPARQL endpoint (US EPA and UM)	✓		✓						AOPLink
eNanoMapper database (UM)	✓		✓						AOPLink
AOP-Wiki SPARQL Endpoint (UM)			✓						AOPLink
The Adverse Outcome Pathway Database (AOP-DB) (US EPA)	✓				✓				AOPLink
Fragnet Search (IM)					✓				DataCure, MetaP, ModelRX
ToxicoDB (University Health Network and UM)	✓	✓	✓		✓				TGX, SysGroup
Prosilico's Human In Vivo ADME/PK-Prediction Studio (Prosilico)	✓								MetaP, RevK
EGA Beacon (CRG)		✓						✓	-
EGA Metadata API (CRG)		✓						✓	-
SCAIView (Fraunhofer)			✓						DataCure
Nextflow (CRG)					✓		✓		AOPLink, DataCure, MetaP, ModelRX, RevK, SysGroup, TGX
TeMOwl (Fraunhofer)				✓					DataCure
ToxTargetlinks (Leiden University and UM)	✓								AOPLink, DataCure
Total	19	6	10	2	13	12	5	17	

^aSee main text for description of categories

^bOthers include “API Definitions for OpenRiskNet applications and data”, and “Site management”.

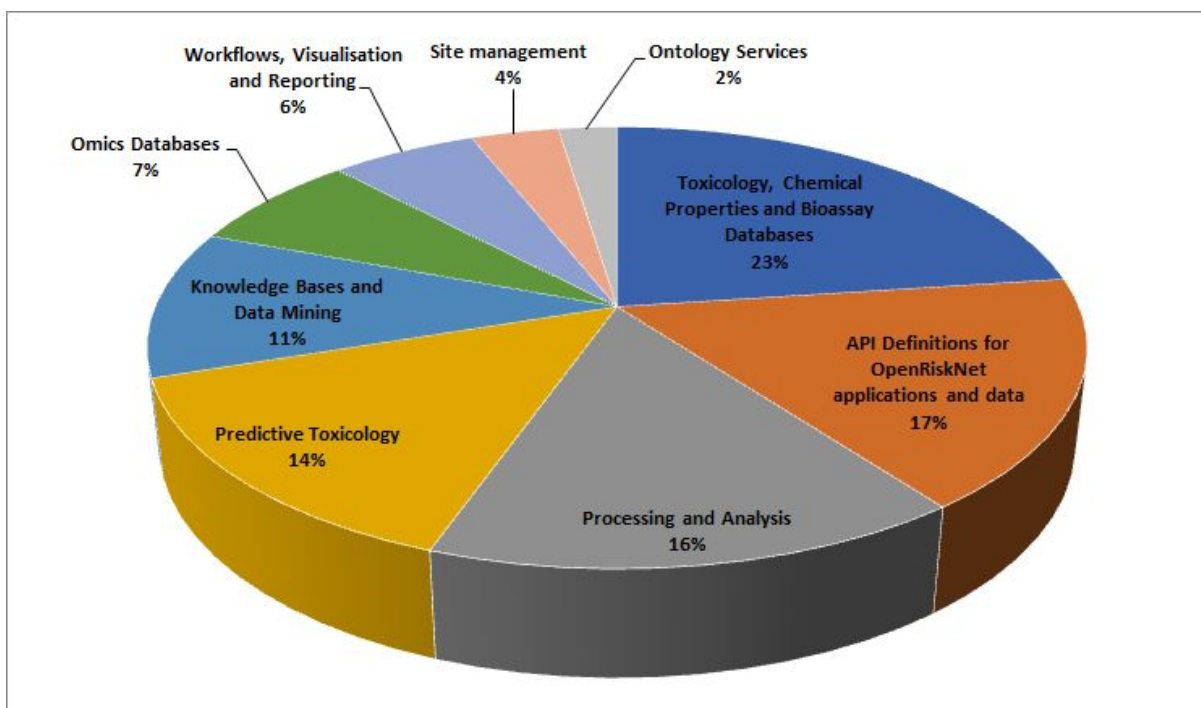


Figure 4. Categories of services included in OpenRiskNet catalogue

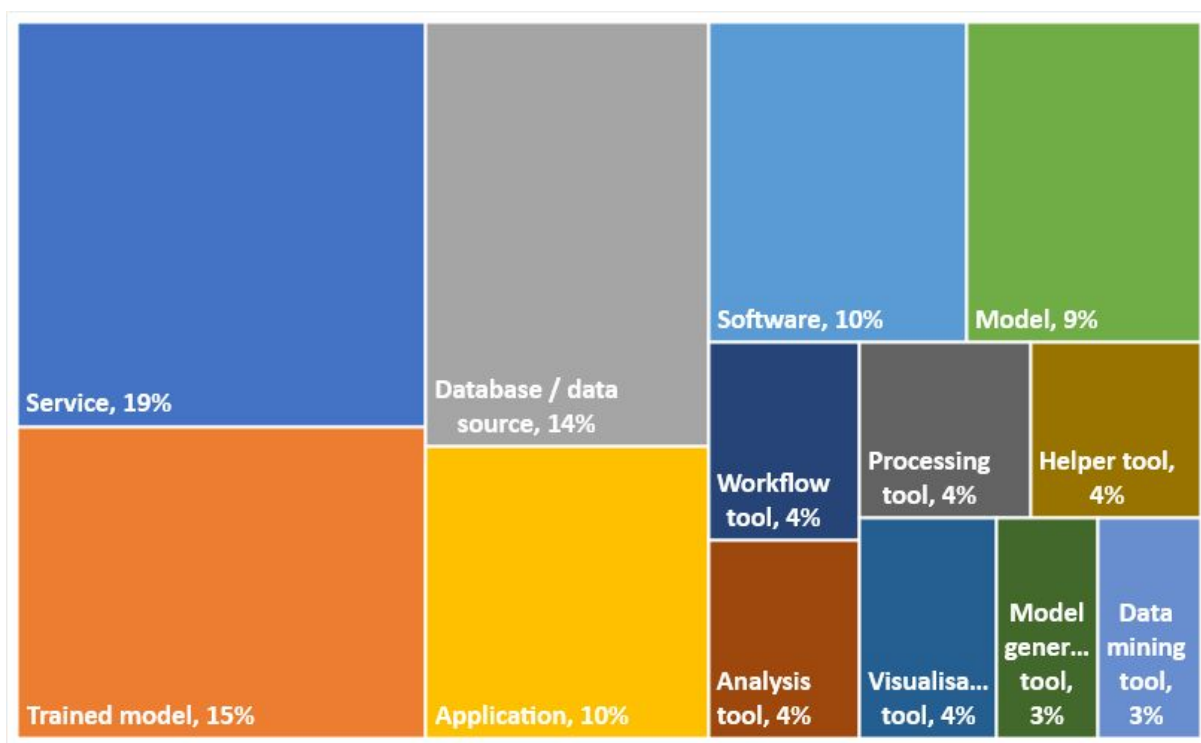


Figure 5. Types of services included in OpenRiskNet catalogue

Categorisation by risk assessment area and applicability domain

The applicability domains relevant to OpenRiskNet (bioinformatics, predictive toxicology, toxicology and computational modelling) are well covered by the current integrated services (**Figure 6**). Details for individual services can be accessed in the OpenRiskNet service catalogue.

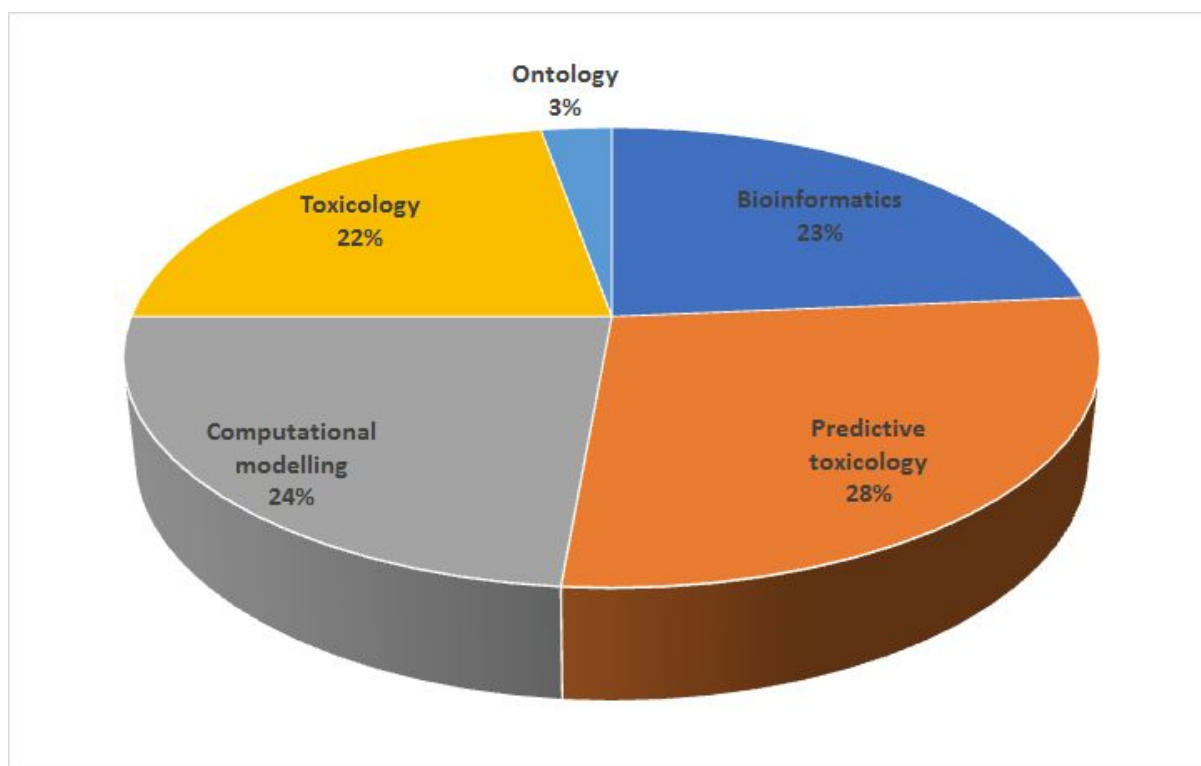


Figure 6. Applicability domains covered by OpenRiskNet services

Additionally, one or more descriptors on their specific topics or biological areas covered were assigned to each service (**Figure 7**) in order to understand the status of the infrastructure from this perspective, to additionally guide users to relevant services, but also to identify eventual areas of strength or gaps. From the statistics shown below a uniform distribution is observed. Please note that this information is based on the self-assessment done during the registration process completed by each service provider/owner.

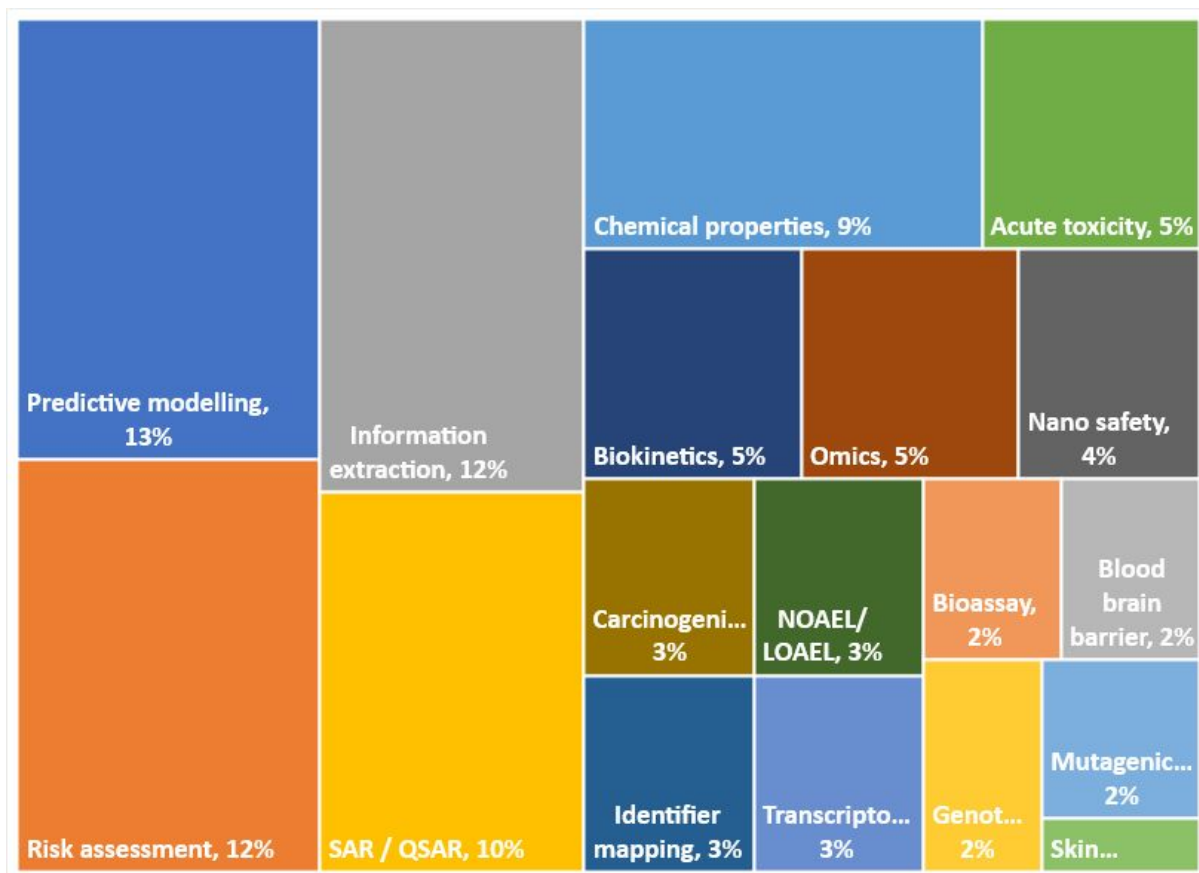


Figure 7. Specific topics and biological areas covered by OpenRiskNet services

Categorisation by target user

Services can also be categorised according to their target users. First, a very general categorisation into Developers and End-users was applied, which is also used to guide users through the website offerings. Advanced workflow tools like Squonk and graphical user interfaces are more suited for end-users while tools accessible only via APIs and jupyter notebook based workflows are provided mainly for developers and computational toxicologists, which integrate these in their own approaches. Both user categories are equally supported by the available services as can be seen in **Figure 8**. The more fine-grained user categories in **Figure 9** shows that also here all groups are represented appropriately. The only expectations are regulators but these are more recipients of the final results, which are included as evidence in regulatory applications, and not primary users of the services.

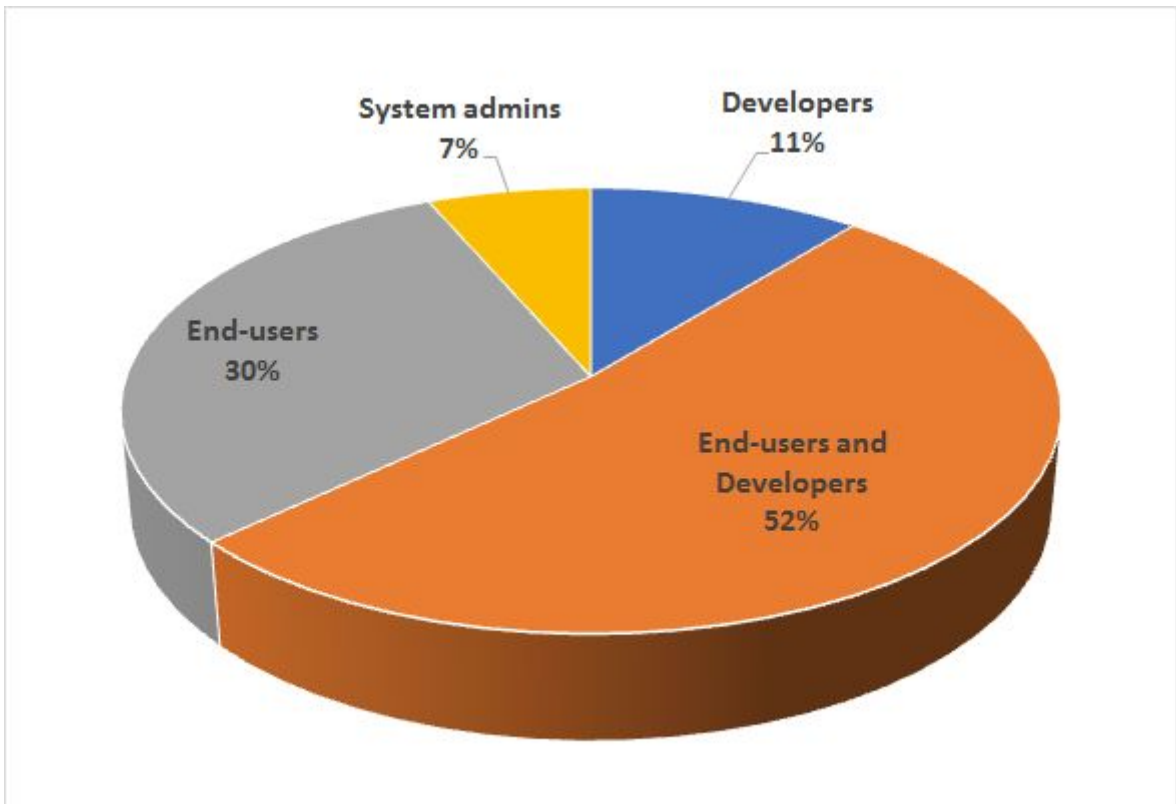


Figure 8. Overall view on OpenRiskNet services based on the user types

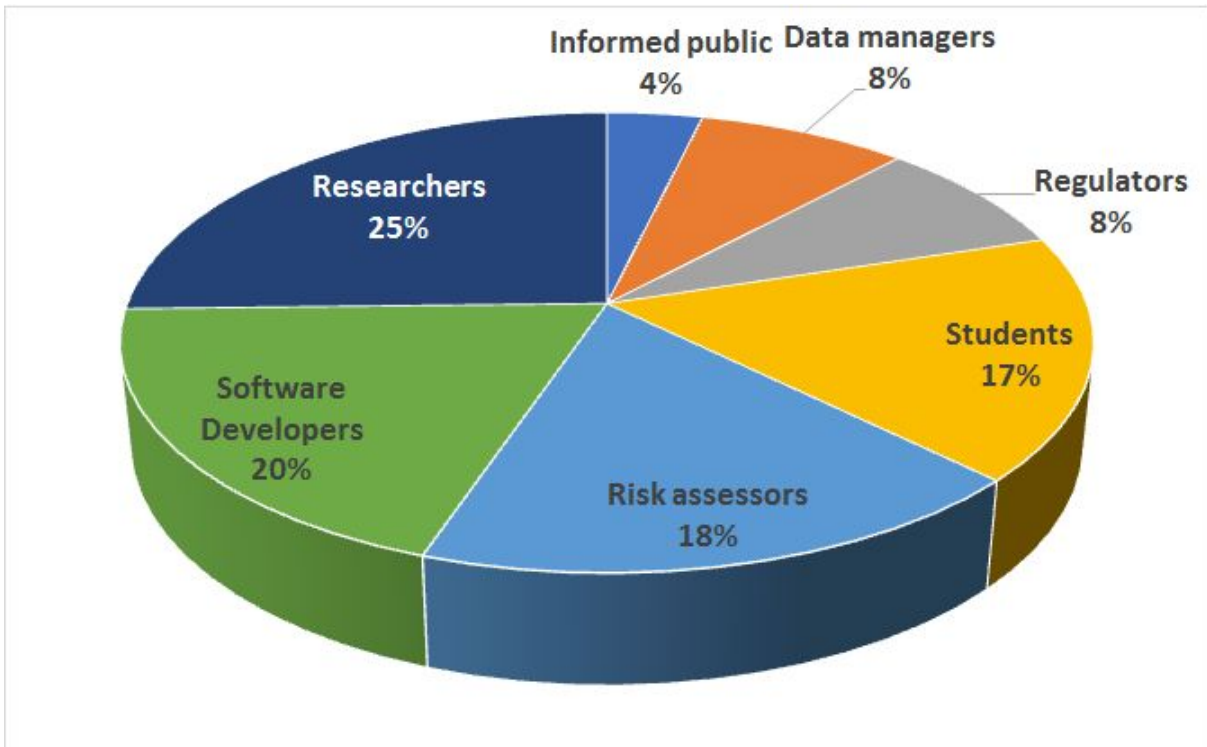


Figure 9. Type of users targeted by OpenRiskNet services

Categorisation by target industry

Finally, OpenRiskNet services address the needs of different industries that require risk assessment. The services cover mainly the chemicals and drug industries, but also sectors like the nanotechnologies, cosmetics, food or other consumer products (**Figure 10**). However, especially nanotechnology and nano-enabled products will be further strengthened by the services provided by the Implementation Challenge winners and the NanoCommons project funded under the call INFRAIA-02-2017: Integrating Activities for Starting Communities, which has completely adopted the OpenRiskNet e-infrastructure.

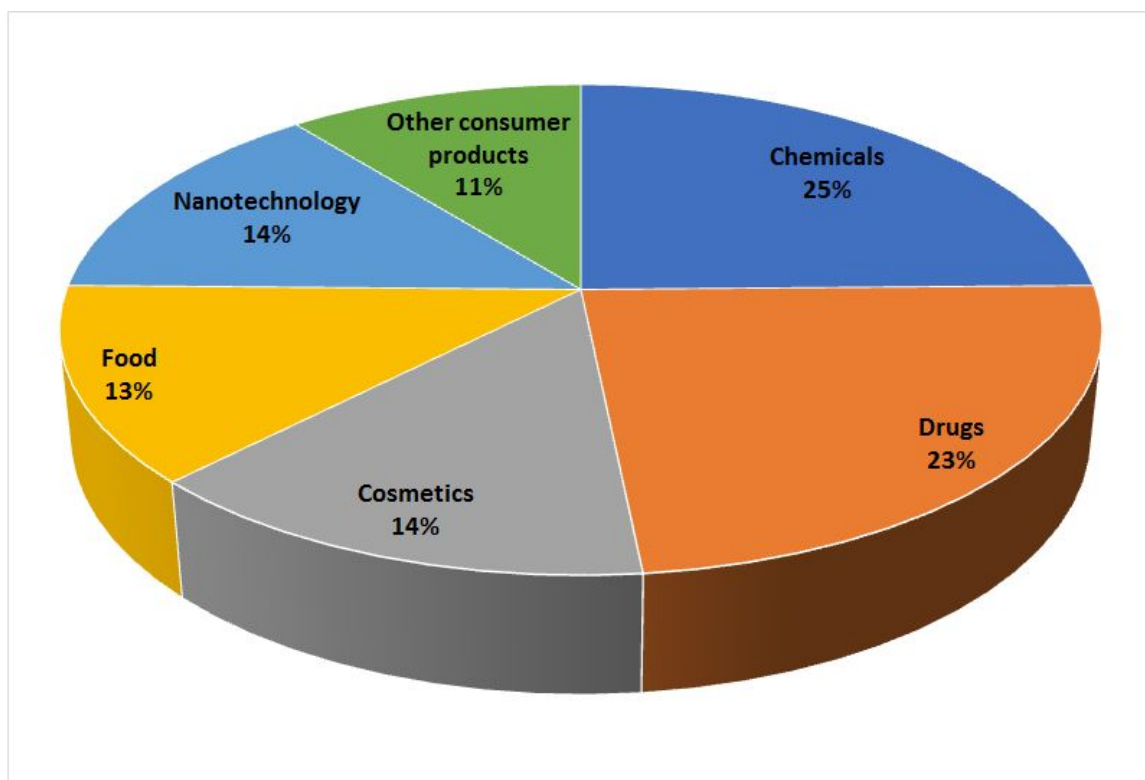


Figure 10. Industries targeted by OpenRiskNet services

Conclusion

The integration of services was finalized at the end of the OpenRiskNet project. In the past 3 years, 45 services were integrated, including 12 services from the Implementation Challenge winners, that cover the 6 categories indicated at the beginning of the project (Tasks 4.1-4.6) and two additional ones, i.e. “API Definitions for OpenRiskNet applications and data”, and “Site management”. Please note that some of the services are very powerful offering solutions for multiple risk assessment areas, large data and knowledge collections or general data management solutions. For example, Jaqpot is providing many hazard prediction and biokinetics models and is also a general tool for developing new models. This is not represented by the total number of services since each service is counted equally independent of the amount of available functionalities. Table 2 gives a better representation of the completeness of coverage of the risk assessment areas by all service and how the applicability of a specific service in general is. Moreover, 60% of the services can be applied in at least two different fields and the most general ones in up to four. This does, however not mean that the more specific services are less important since they provide highly optimised and unique solutions in this field.

The services integration also included the development of workflows to support the case study work by automating complex tasks only achievable by the combination of multiple services and, more generally, to showcase the interoperability of the services and the additional functionality and flexibility accessible when using the APIs instead of the graphical user interfaces. Categorisation of the services, also used to guide the user to relevant services in the OpenRiskNet Service catalogue²⁶, showed that the OpenRiskNet goals to provide services for all application domains in risk assessment and for different user groups could be reached at the end of the project.

²⁶ <https://openrisknet.org/e-infrastructure/services/>

Glossary

The list of terms or abbreviations with the definitions, used in the context of OpenRiskNet project and the e-infrastructure development is available:

<https://github.com/OpenRiskNet/home/wiki/Glossary>

References

1. Jennen D, Dudgeon T, Sarimveis H, Doganis P, Karatzas P, Rautenberg M, et al. Report of the Service Integration with OpenRiskNet - Initial Deployment (Deliverable 4.1). 2018. doi:10.5281/zenodo.1484309
2. Jennings P, Exner T, Farcal L, Oki N, Sarimveis H, Doganis P, et al. Final definition of case studies (Deliverable 1.3). 2018. doi:10.5281/zenodo.1479127
3. Magkoufopoulou C, Claessen SMH, Tsamou M, Jennen DGJ, Kleinjans JCS, van Delft JHM. A transcriptomics-based in vitro assay for predicting chemical genotoxicity in vivo. *Carcinogenesis*. 2012;33: 1421–1429.
4. Rautenberg M, Karwath A, Kramer S, Dudgeon T, Spjuth O, Bachler D, et al. Initial API version provided to providers of services (Deliverable 2.2). 2018. doi:10.5281/zenodo.1479444