OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

Deliverable Report D4.2

Report of the Service Integration with OpenRiskNet (Intermediate 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 *in silico* Analysis and Modelling in Risk Assessment
Project Acronym	OpenRiskNet
Project Coordinator	Douglas Connect GmbH
Start date	1 December 2016
End date	30 November 2019
Duration	36 Months
Project Partners	P1 Douglas Connect GmbH Switzerland (DC) 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)

Deliverable Report identification

Document ID and title	Deliverable 4.2 Report of the Service Integration with OpenRiskNet (Intermediate 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	UoB
Author(s)	Danyel Jennen (UM), Jumamurat Bayjanov (UM), Lucian Farcal (DC), Noffisat Oki (DC), Daan Geerke (VU), Philip Doganis (NTUA), Marvin Martens (UM), Egon Willighagen (UM), Denis Gebele (JGU), Marc Jacobs (Fraunhofer), Tim Dudgeon (IM), Iseult Lynch (UoB), Thomas Exner (DC)
Status	Final
Version	V1.0
Document history	2019-01-23 First version 2019-02-21 Final version

Table of Contents

Summary

This report describes the 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). Due to 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.

This report is an update of Deliverable D4.1 “Report of the Service Integration with OpenRiskNet (Initial Deployment)” and is also directly linked to other deliverables (e.g. D1.3 Final definition of case studies, D2.2. Initial API version provided to providers of services and D2.3 Report on deployment of virtual infrastructures with service discovery and container orchestration).

The report includes examples on 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.

The extensive description of the integration status (a process based on a predefined set of eight operations) is completed by statistics on the services integrated into the OpenRiskNet e-infrastructure (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 preprocessing, analysis and modelling tools). Integration of these tools started on day 1 and will continue until the end of the project. The intermediate report presented here is an update of Deliverable D4.1 "Report of the Service Integration with OpenRiskNet (Initial Deployment)" [1] and describes the work in the second year 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 publicly 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, the 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 are coming in now via the Associated Partner Programme and especially the Implementation Challenge[1].

While in the first period up to the D4.1 report, 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 was driven by the needs of the case studies (see also Deliverable 1.3 [2] for more information on the case studies). We are presenting first four examples (DataCure, MetaP, TGX, and ModelRX), on 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 first Implementation Challenge are included to strengthen the case. This will be followed by detailed descriptions of the current 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 AOPWiki
- TGX: Toxicogenomics-based prediction and mechanism identification
- RevK: Reverse dosimetry and PBPK prediction

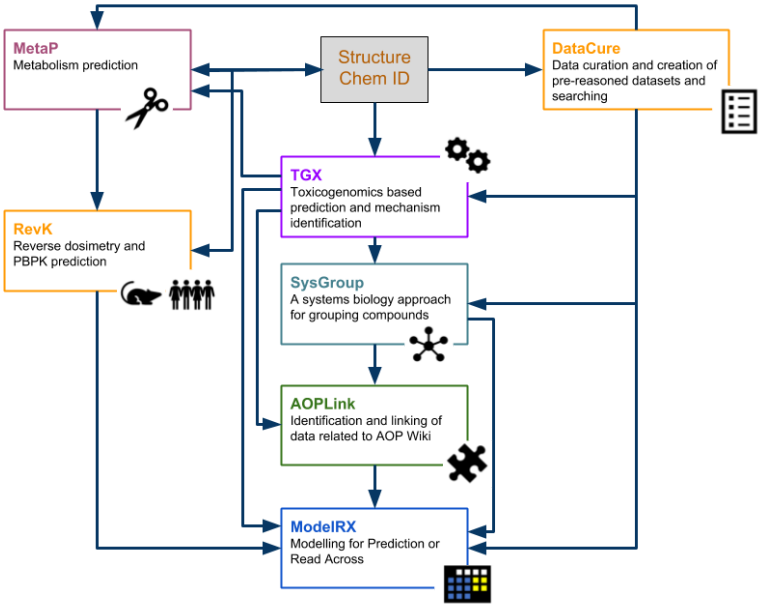

Figure 1. OpenRiskNet case studies[2]

We concentrate here on four of these as examples, DataCure, MetaP, TGX and ModelRX, showing 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.

[2] https://openrisknet.org/e-infrastructure/development/case-studies/

DataCure

In the Data curation and creation of pre-reasoned datasets and searching (DataCure[3]) case study, **DC** (CS leader), **IM**, **NTUA**, **Fraunhofer**, **UoB** and **UM** work together on developing a system to allow users access different data sources through the OpenRiskNet service. This case study shows examples of practical implementations of curation of data 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) and 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 sources of omics data like diXa data warehouse, ToxCast, LTKB, NCBI GEO, and EBI's ArrayExpress, while the chemical information data is currently being gathered from ChEMBL, and aggregated data repositories such as PubChem and ToxPlanet. Specifically, after annotation and pre-processing (e.g. normalisation, filtering of transcriptomics data), these data are used together with the chemical information for prediction analyses for e.g. genotoxicity, carcinogenicity and/or improved grouping of compounds for e.g. organ toxicity, drug-induced liver injury under the TGX and SysGroup case studies, respectively.

It is expected that in the process of collecting data from a variety of external sources and in different formats, text mining may be necessary to retrieve the useful information from the data collected before annotation can proceed. For this need, a text mining capability is also implemented to support the data curation process. This is done through the API of SCAIView (4.3) the semantic text search and retrieval engine. The API can be accessed from http://api.scaiview.com/swagger-ui.html. 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 4.4).

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' (eg. input: document-id or document-text, output: json, csv or triplestore) for use by other workflows in the OpenRiskNet Virtual Research Environment.

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

MetaP

In the Metabolism Prediction (MetaP[4]) case study, **VU** (CS leader), **UU**, University of Hamburg (**UHH**, associate partner and implementation challenge winner) and **JGU** work together on integrating tools for metabolite and site-of-metabolism (SOM) prediction. For a given (set of) query molecule(s), SOMs and/or metabolites will be predicted that can serve as input for other case studies. For that purpose, we incorporate and combine ligand-based metabolite predictors (e.g. MetPred, enviPath, FAME, SMARTCyp) as well as protein-structure and -dynamics based approaches to predict the sites-of-metabolism by Cytochrome P450 (CYP450s), which metabolise approximately 75% of the currently marketed drugs. To facilitate the combined use of the metabolite prediction approaches and their outcomes, we benefit from ongoing development in workflow management systems (Nextflow, Squonk, MDStudio) and we explore integration into and application of these platforms. Once integrated, the added value of multiple predictors will be the subject of a pilot study on consensus metabolite prediction.

Specific objectives of this CS are integration, comparison and combination of:

- Ligand-based Site-Of-Metabolism (SOM) prediction using reaction SMARTs, circular fingerprints and/or atomic reactivities
- QSBR (quantitative-structure biotransformation relationship) modelling of microbial biotransformation
- Protein-structure and -dynamics based prediction of CYP450 isoform specific binding and SOMs
- Predicting probabilities for specific reaction type events

During workflow development, model calibration and validation we take advantage of data from XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites, as available in ZINC, ChEMBL, DrugBank, EAWAG-BBD and/or the SMARTCyp and FAME suites.

UU has recently made a first API to their MetPred metabolite predictor available, which was available as web service already. MetPred predicts phase I metabolites by ranking the most probable SOMs and reaction type(s) for a given molecule based on similar atom environments and ReactionSMARTS in annotated datasets. Currently we focus on extending our toolbox towards enviPath (developed by JGU for prediction of microbial biotransformation pathways and products following rules represented by SMIRKS), SMARTCyp (SOM prediction for P450 metabolism based on fragment-mapping to pre-computed QM data and atomic accessibility), and protein-structure based predictors and plasticity models of partner VU for docking substrates into (flexible) P450 isoforms. As part of the implementation challenge, we also focus on integrating the FAME suite of associate partner **UHH**, which in its current version predicts CYP450 SOMs based on machine learning using few quantum and circular-environment based atomic descriptors. As a subsequent step we will perform a consensus study in which we will explore and quantify the added value of combined use of the different predictors.

[4] https://openrisknet.org/e-infrastructure/development/case-studies/case-study-metap/

TGX

In the toxicogenomics-based prediction and mechanism identification (TGX[5]) 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.

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 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 reproduced the results of that study. The workflow created using the Snakemake workflow manager is available from a GitLab software repository[6], where every step is clearly described in Snakefile to reproduce the approach described by Magkoufopoulou *et al.* (2012) [3] and is used as reference for the transfer 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 to a Nextflow-based workflow[7]. 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.

Currently, these workflows are being translated into more generic approaches so that they can be applied to other toxicogenomics studies. Here contributions from **ToxPlanet** (associate partner and implementation challenge winner) who provide access to hundreds of relevant databases/collections and from **Fraunhofer** who can provide well established text mining approaches, developed in collaboration with the DataCure case study, will provide the needed animal reference data.

[5] https://openrisknet.org/e-infrastructure/development/case-studies/case-study-tgx/
[6] https://gitlab.com/bayjan/openrisknet_magkoufopoulou
[7] https://github.com/openrisknet/nf-toxomix

ModelRX

In the Modelling for Prediction or Read Across (ModelRX[8]) case study, **NTUA** (CS leader), **JGU** and **UU** work together in the workflow that involves obtaining an initial (*training*) dataset within the OpenRiskNet environment, using tools to construct a model and applying the model to new *prediction* datasets in order to generate predictions.

The work of ModelRX contributes in two tiers of risk assessment framework, as described in Berggren et al., 2017 [4]:

- In Tier 0, providing computational methods to support suitability assessment of existing data and identification of analogues.
- In Tier 2, providing predictive modelling functionalities that are essential for final risk assessment.

The ModelRX case study built on the tools Jaqpot, Lazar 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
Jaqpot API	https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/#/
Lazar API	https://lazar.prod.openrisknet.org/
JGU WEKA Rest service	https://jguweka.prod.openrisknet.org/
Metpred cplogd ltkb	https://metpred.prod.openrisknet.org/ https://cplogd.prod.openrisknet.org/ http://ltkb-cpsign.prod.openrisknet.org/

APIs are more directed to developers, providing them with powers 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
Jaqpot GUI	https://ui-jaqpot.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
CPSign	https://squonk-notebook.prod.openrisknet.org/portal

[8] https://openrisknet.org/e-infrastructure/development/case-studies/case-study-modelrx/

We are working on the integration of our services through Jupyter notebooks that we are developing as a central point to access and use services from OpenRiskNet partners, from accessing data to developing models and generating predictions. For that purpose, we are gathering the work in a GitHub repository[9], so we can 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 support similarity identification to DataCure by providing tools for calculating theoretical descriptors of substances.

Additionally, work with associate partners assists the propagation of OpenRiskNet solutions and enhance the offering. We are working with two winners of the first implementation challenge: **NovaMechanics Ltd.** have developed the Enalos InSilicoNano platform and have experience in the risk assessment of nanomaterials and **BIGCHEM GmbH** offers OCHEM, a tool for descriptor calculations and model development. More information on both in the Implementation Challenge section below.

[9] https://github.com/OpenRiskNet/notebooks

Integration status of consortium-provided services

Since integration of services involves different steps from API development, packaging into containers and container orchestration, semantic annotation and listing in the OpenRiskNet and other service catalogues, becoming OpenRiskNet-complaint 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;
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.

As pointed in the Deliverables 2.2 [5], 2.3 [6], 2.4 and 4.1 [1], many of these steps are themselves not one-time efforts but have to be continuously adapted to the progress of the project. Specifically, the bottom-up API concept described in Deliverable 2.2 lists specific steps, which have 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 will work technically. Due to the ongoing refinement, adaptation and optimisation of the concepts and the integrated services, we decided to not award the label of OpenRiskNet complicence to any of the integrated services for now, but instead to list separately the status of each service with respect to its compliance (to the current realisation of the OpenRiskNet concepts) for the eight operations specified above. The progress of the services previously prioritised (see Deliverable 4.1 [1]) was first included in the technical report of the first report period and is listed in an updated form in **Table 1**. Currently available services and their up-to-date integration status are available in the Service Catalogue[10] on the OpenRiskNet website. Examples of catalogue entries are shown in **Figure 2 and 3**, respectively.

[10] https://openrisknet.org/e-infrastructure/services/

Table 1. Integration status of OpenRiskNet services (x = full compliance, / = partly implemented)

Task	Services integrated	1	2	3	4	5	6	7	8
4.1	Squonk services for chemical property prediction	/		x	/	x			/
	cpLogD - confidence predictor for logD	x	/	x	x	/			
	Modelling Web			x		x			
	CDK-Depict			x		x			
	Chemidconvert	x		x	x	x	x		x
	eNanoMapper - nanomaterial database	x		x	x	x	x		
	ToxRefDB (Data Explorer)	x		x					/
	ToxCast/Tox21 summary data (Data Explorer)	x	/	x	x				/
	Tox21 sample specific data	x	/	x	/				/
	FDA Estrogenic Activity Database	x	/	x	x	x	x		/
	MetPred	x		x	x	x	x		
	P450 SOM predictor	/			x		x		
	LTKB								
4.2	TG-GATEs (Data Explorer)	x		x	/				/
	Toxygates	x		x	/				/
	diXa (via BioStudies)	/							
	Gene Expression Omnibus (GEO)	/							
	ArrayExpress	/							
4.3	BridgeDb	x		x		x		x	
	Data mining algorithms through Jaqpot (Jaqpot GUI)	x	/	x	x	x	x	x	/
	Data mining algorithms through JGU Weka	x	/	x	x	x	x		/
	WikiPathways SPARQL endpoint	/		/	x	/			
	AOP-Wiki SPARQL endpoint	/		/		/			
	eNanoMapper - nanomaterial database	x		x	x	x	x	x	
	SCAIView Scientific Literature Database	x	/	x	x		x		x

4.4	Jenkins: ontology building and testing								
	Ontology Lookup Service (OLS)	/	/	x	/				
	Ontology Annotation Services (BELIEF Text Mining)	x	/	x	/		/		/
4.5	**Jaqpot processing and analysis services** (Jaqpot GUI)	x	/	x	x	x	x	x	/
	CDK descriptor calculation service (Jaqpot API)	x	/	x	x	x	x	x	/
	PROAST and TCPL dose response modelling service	x						/	
4.6	**WEKA REST Service**	x	/	x	x	x	x		x
	Lazar Toxicity Predictions	x	/	x	x	x	x		x
	Jaqpot predictive modelling services (Jaqpot GUI)	x	/	x	x	x	x	x	/
	Jaqpot PBPK modelling services (Jaqpot API)	x	/	x	x	x	x	x	/
	Jaqpot applicability domain services (Jaqpot GUI)	x	/	x	x	x	x	/	/
	httk package for PBPK modelling service (Jaqpot API)	x	/	/	/	/	/	x	/
4.7	**Squonk Computational Notebook**	/		x	/	x			/
	Jupyter Notebook			x		x			
	Nextflow			x		x			

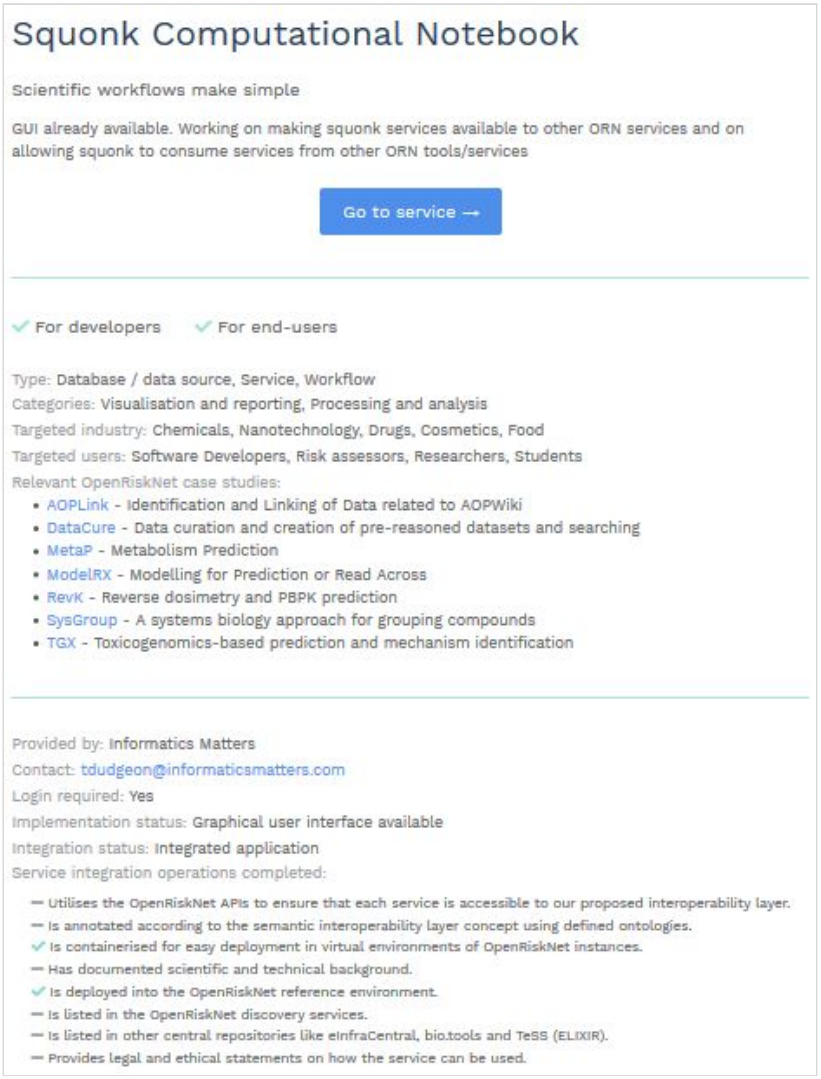

Figure 2. Screenshot of integrated Squonk services accessible on the OpenRiskNet Reference Virtual Environment

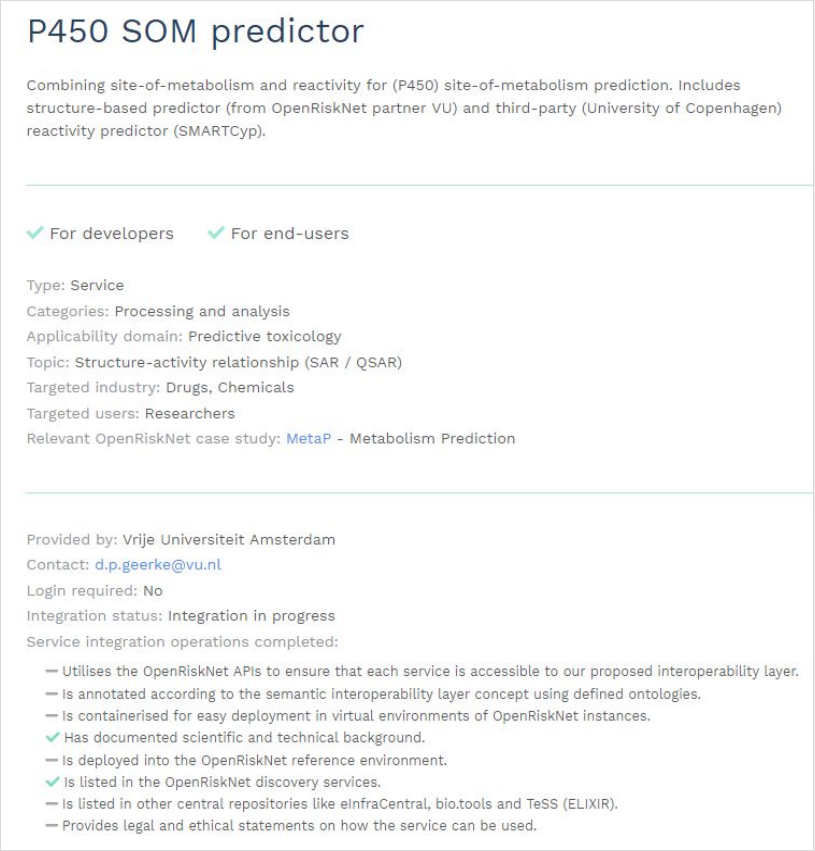

Figure 3. Screenshot of integration in progress of P450 SOM predictor service

Table 2 provides additional information on each service including the license type and the case studies, the service is already used in.

Table 2. Implementation status, licence information and link to case studies

Name	Implementation status	Licence type	Licence	Relevant case studies
ChemIdConvert	API documentation available (Swagger-OpenAPI v2), Application programming interface available	Open source	Attribution-ShareAlike	DataCure
BridgeDb identifier mapping service (Homo sapiens, Mus musculus and Rattus norvegicus)	Available as web service	Open source	Apache	AOPLink
BridgeDb identifier mapping service (variety of species)	Available as web service	Open source	Apache	AOPLink

cpLogD confidence predictor for logD	Available as web service, Application programming interface available	Open source (wrapper), Commercial (CPSign)	Apache	ModelRX
JGU WEKA REST Service	Available as web service, Containerised	Open source	GNU General Public License	ModelRX
MetPred	Available as web service, Application programming interface available	Open source (wrapper), Commercial) (database)	Apache	MetaP
AOP-Wiki SPARQL Endpoint		Open source (software), Proprietary data	GNU General Public License (software)	AOPLink
WikiPathways SPARQL Endpoint	Available as web service	Open source (software), Open data (data)	GNU General Public License (software), CCZero (data)	AOPLink
SCAIView	Available as web service, Application programming interface available, API documentation available (Swagger-OpenAPI v2), Containerised, Graphical user interface available	Fraunhofer license	free to academic use	DataCure, TGX
Jaqpot API	API documentation available (Swagger-OpenAPI v2), Containerised, Available as web service, Application programming interface available	Open source	GNU General Public License	ModelRX, RevK
Jaqpot GUI	Graphical user interface available	Open source	GNU General Public License	ModelRX
Lazar Toxicity Predictions	Containerised, Graphical user interface available, Available as web service	Open source	GNU General Public License	ModelRX
Nextflow	Containerised	Open source	Apache	TGX
P450 SOM predictor		Open source		MetaP
Jupyter Notebooks	Containerised, Graphical user interface available, Available as web	Open source		AOPLink, DataCure, MetaP, ModelRX, RevK, SysGroup, TGX

	service			
Data Explorer serving ToxCast, openrisknet.ToxRef DB and TG-GATEs data	API documentation available (Swagger-OpenAPI v2), Graphical user interface available, Available as web service	Open source	Attribution	DataCure
Squonk Computational Notebook	Graphical user interface available	Open source	Apache 2.0	AOPLink, DataCure, MetaP, ModelRX, RevK, SysGroup, TGX
eNanoMapper	Graphical user interface available, API documentation available (Swagger-OpenAPI v2), Containerised	Open source	LGPL v2.0	AOPLink

Services selected within the Implementation Challenge

The Implementation Challenge[11] 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 can apply for partial financially and strongly technically support by researchers and developers of OpenRiskNet partners. Similar to the internal services, the funds and the technical support will cover 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. At the end of the first selection period (31 October 2018), 6 winners have been selected by the scientific advisory board (November 2018) and integration of the following services has been started:

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: This model 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: University of Bergen 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: An online decision support tool for the design and virtual screening of nanoparticles
   Risk Assessment Tool for the Virtual Screening of Metal Oxide Nanoparticles through Enalos: Since experimental toxicity evaluation for the different types of NPs already available, is often expensive and time consuming, several computational approaches are proposed for the risk assessment of NPs. In this work, we have developed a predictive classification model for the toxicological assessment of iron oxide NPs with different core, coating and surface modification based on a number of different properties including size, relaxivities, zeta potential and type of coating.
5. Igor Tetko, BIGCHEM GmbH
   OCHEM models: OCHEM provides access to more than 100 models published on

[11] https://openrisknet.org/associated-partner-programme/implementation-challenge/

its web site for different endpoints. The implementation of this interface will allow other users execute predictions on the OCHEM website.
OCHEM descriptors: Provide access to several tens thousands descriptors as well as structural alerts and chemical functional groups calculated at the OCHEM platform.
OCHEM model development tool: OCHEM model development tools provide access to more than 20 modelling approaches which are based on original and published methods as well as frameworks. The recent developments include support GPU computing as well as multi-task modelling

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.

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 functionality being made available in ModelRX and AOP-DB into AOPLink case study.

Categorisation of services

Even if the case studies show the functionality of individual OpenRiskNet service and of the complete infrastructure, these services and general workflows are offered to the community to facilitate the generation 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 catalogue and can be used as search facets to locate suitable services via the web interface or, if already available for the services, 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, the different tasks for WP4 corresponding service types 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, which is indicated and used for structuring Table 1. 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 current status of the integrated services listed in the Service Catalogue. It can be seen that databases are somewhat underrepresented, which is caused by the uncertainty on how to deal with ethics and legal issues when re-distributing human and animal data. Since clear ethics guidelines and checklists are now available, this is expected to change in the near future. A more fine-grained categorisation of the type of services is additionally given in **Figure 5**. The coarse- and fine-grained categories are additionally listed in **Table 3** on a per-service base together with the API version available and the link to the service API specification on the reference environment.

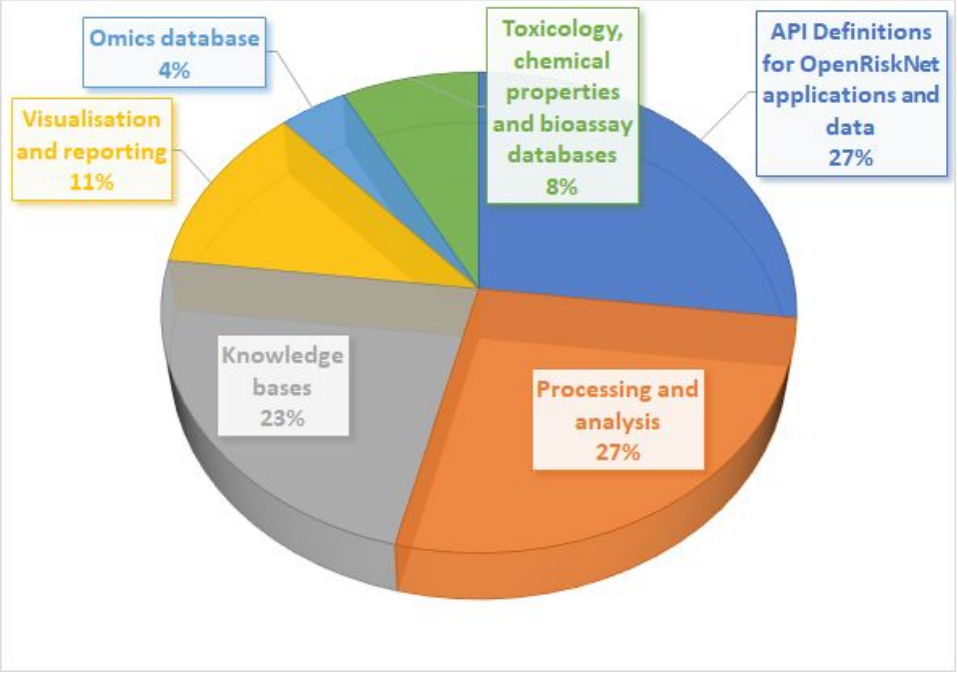

Figure 4. Categories of services included in OpenRiskNet catalogue

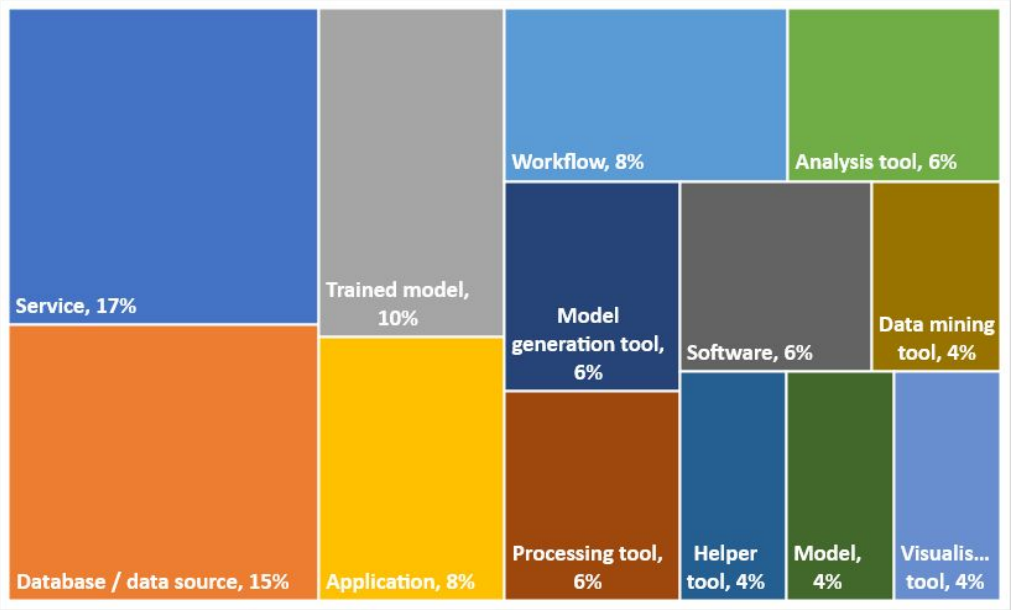

Figure 5. Types of services included in OpenRiskNet catalogue

Table 3. Categories and API types for the OpenRiskNet services

Name	Category	Service type	API Type	API URL
ChemIdConvert	API Definitions for OpenRiskNet applications and data	Helper tool	REST	http://orn-chemidconvert-chemidconvert.prod.openrisknet.org/v1/ui/
BridgeDb identifier mapping service (Homo sapiens, Mus musculus and Rattus norvegicus) https://www.bridgedb.org/	API Definitions for OpenRiskNet applications and data	Database / data source, Service	REST under OpenAPI2 specification	http://bridgedb.prod.openrisknet.org/swagger/
BridgeDb identifier mapping service (variety of species) https://www.bridgedb.org/	API Definitions for OpenRiskNet applications and data	Database / data source, Service	REST under OpenAPI2 specification	http://bridgedbfull-swagger.prod.openrisknet.org/swagger/
Conformal LogD Predictor	API Definitions for OpenRiskNet applications and data	Trained model		https://cplogd.prod.openrisknet.org/
JGU WEKA REST Service	API Definitions for OpenRiskNet applications and data	Model generation tool, Trained Model, Service	REST, OpenAPI	https://jguweka.prod.openrisknet.org/
Metabolic Site Predictor	API Definitions for OpenRiskNet applications and data	Trained model		https://metpred.prod.openrisknet.org/
AOP-Wiki SPARQL Endpoint https://aopwiki.org/	Knowledge bases	Database / data source	SPARQL	
WikiPathways SPARQL Endpoint https://www.wikipathways.org	Knowledge bases	Database / data source	SPARQL	http://sparql.wikipathways.org/
SCAIView http://academia.scaiview.com/academia/	Knowledge bases	Application, Service, Software	OpenAPI	http://api.scaiview.com/swagger-ui.html
Jaqpot API	Knowledge bases, Processing and analysis, API Definitions for OpenRiskNet applications and data	Service, Data mining tool, Model, Model generation tool, Trained model, Processing tool, Analysis tool	REST under OpenAPI2 specification	https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/
Jaqpot GUI http://www.jaqpot.org/	Knowledge bases, Processing and analysis, Visualisation and reporting	Visualisation tool, Application, Data mining tool, Model, Model generation tool, Trained model, Processing tool,	Based on Jaqpot API that uses REST under OpenAPI2 specification	https://ui-jaqpot.prod.openrisknet.org/

		Analysis tool, Workflow		
Lazar Toxicity Predictions https://lazar.prod.openrisknet.org/predict	Processing and analysis	Trained model, Application, Service	REST, OpenAPI	https://lazar.prod.openrisknet.org/
Nextflow http://www.nextflow.io	Processing and analysis	Service, Workflow, Software		
P450 SOM predictor	Processing and analysis	Service	OpenAPI	
Jupyter Notebooks https://jupyterhub-jupyter.prod.openrisknet.org/	Processing and analysis, Visualisation and reporting	Helper tool, Visualisation tool, Processing tool, Analysis tool, Software, Workflow		
Data Explorer serving ToxCast, ToxRefDB and TG-GATEs data https://data.dougla sconnect.com/	Toxicology, chemical properties and bioassay databases, Omics database	Database / data source, Application	REST	https://toxcast-api.cloud.douglasconnect.com/beta/ui/
Squonk Computational Notebook https://squonk-notebook.prod.openrisknet.org/portal	Visualisation and reporting, Processing and analysis	Database / data source, Service, Workflow	REST	
eNanoMapper database http://ambit.sourceforge.net/	Toxicology, chemical properties and bioassay database	Database	REST	http://nanomaterialdb-test.prod.openrisknet.org/ambit/

Categorisation by risk assessment area and applicability domain

The applicability domains relevant to OpenRiskNet (bioinformatics, predictive toxicology, toxicology and the computational modelling) are well covered by the current integrated services (**Figure 6**).

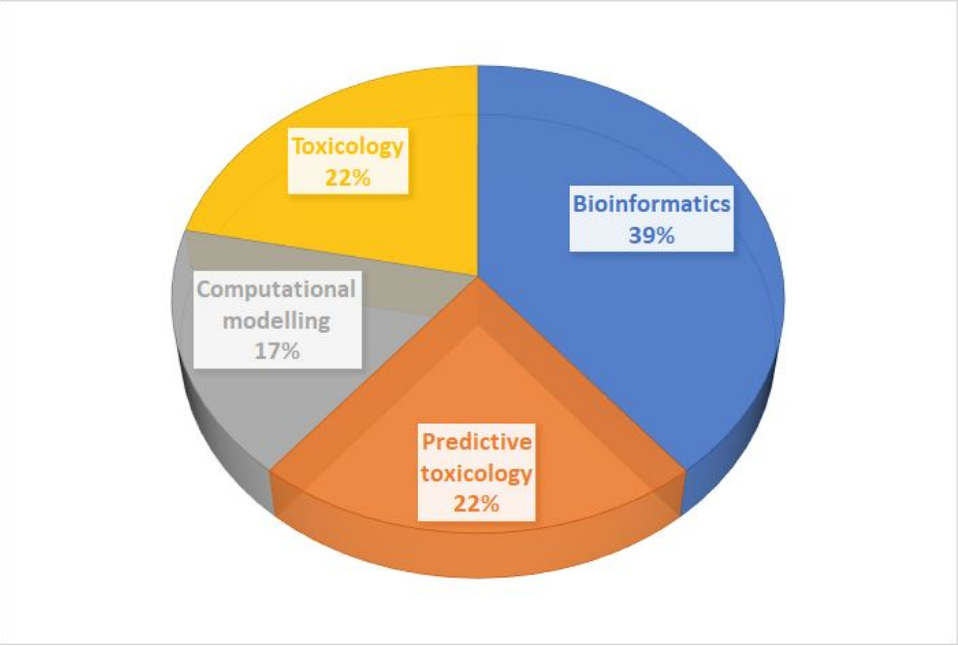

Figure 6. Applicability domains covered by OpenRiskNet services

Additionally, to each service one or more descriptors on their specific topics or biological areas covered were assigned (**Figure 7**) in order to understand the status of the infrastructure from this perspective, 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.

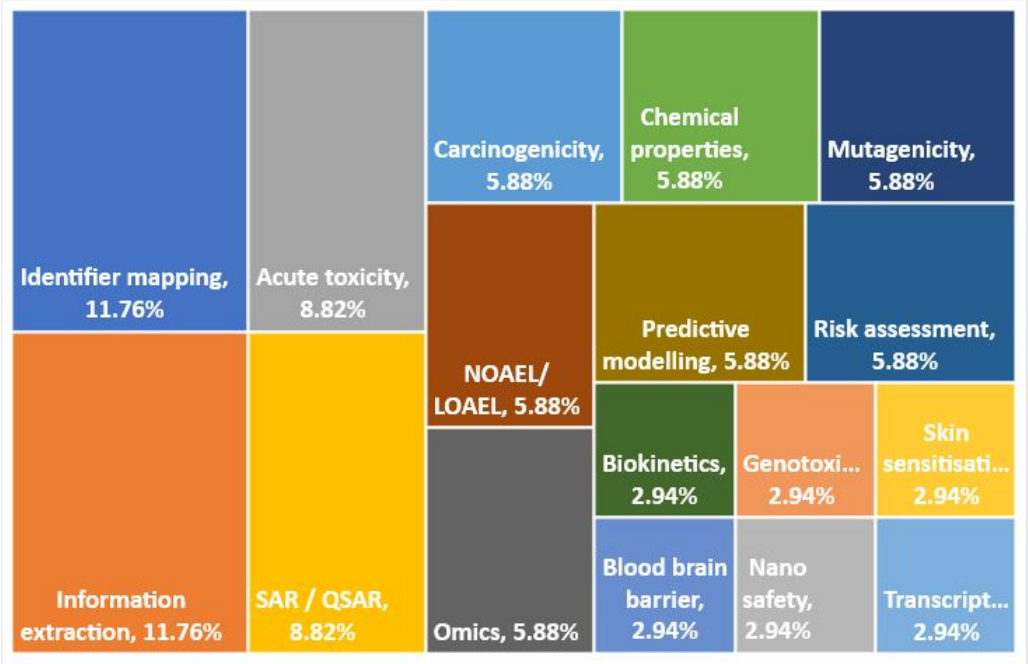

Figure 7. Specific topics and biological areas covered by OpenRiskNet services

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, 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.

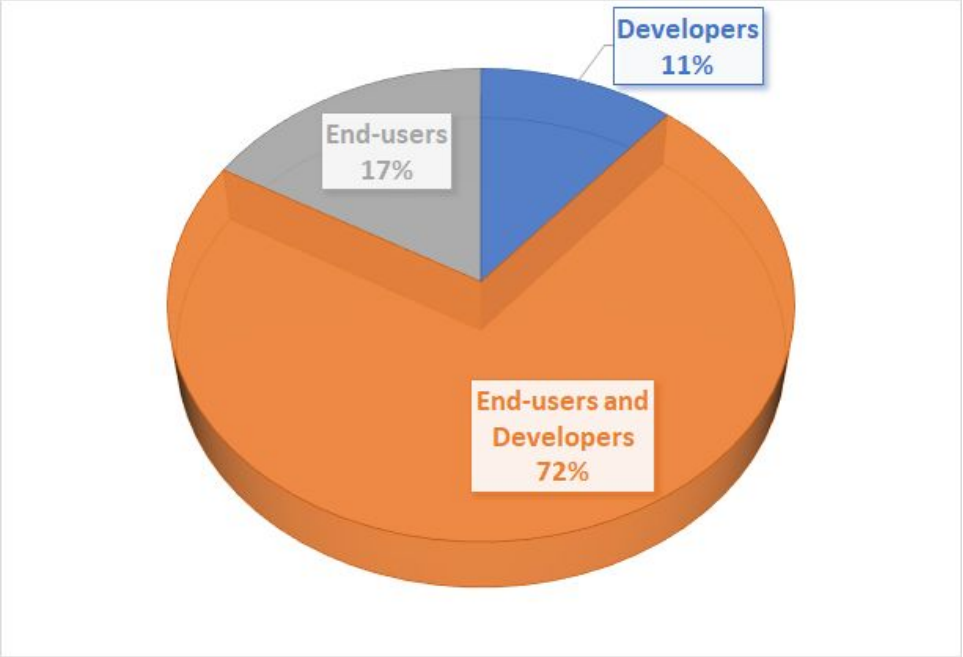

Figure 8. Overall view on OpenRiskNet services based on the user types

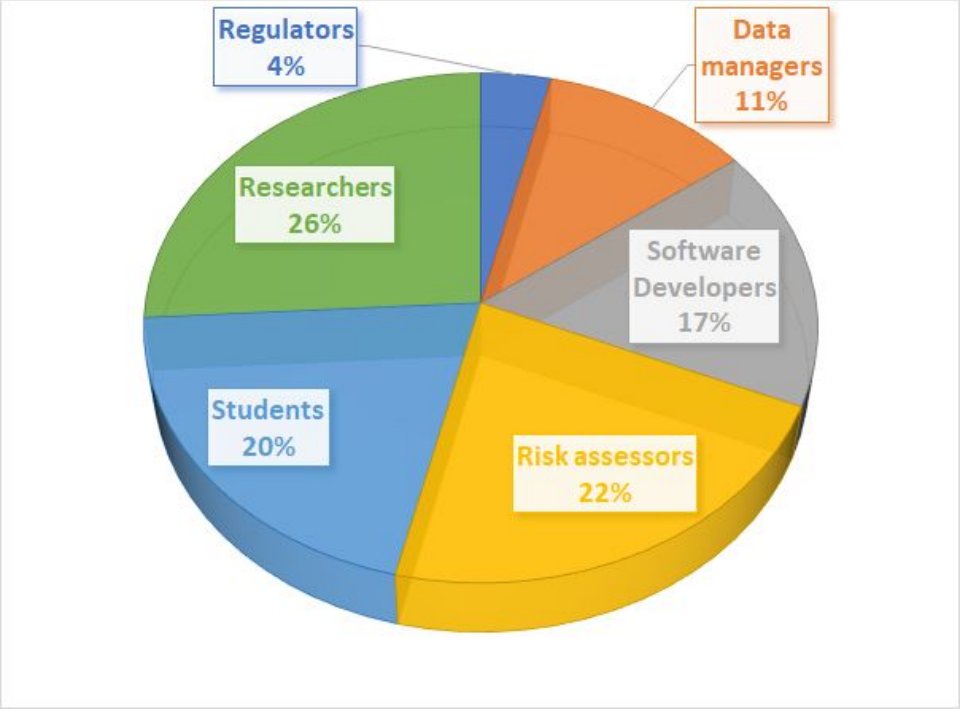

Figure 9. Type of users targeted by OpenRiskNet services

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.

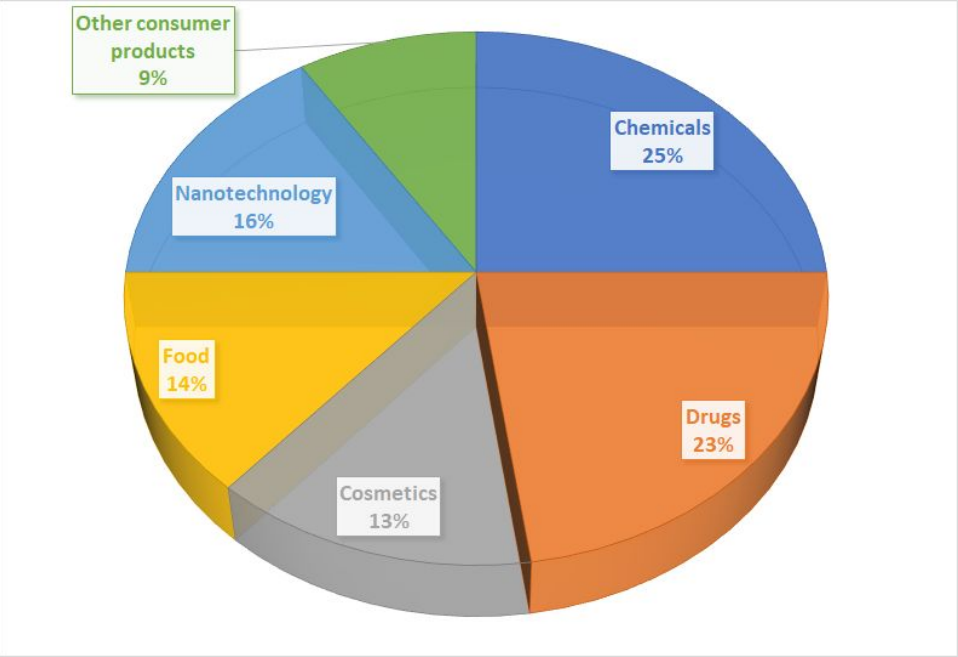

Figure 10. Industries targeted by OpenRiskNet services

Conclusion

The integration of services was continued in year 2 of the project and, in this way, expanded the portfolio of possible applications of the OpenRiskNet infrastructure. Besides improving the implementation status of services integrated in year 1 by finalising work needed to fulfill additional criteria, new implementations were started especially in collaboration with the Implementation Challenge winners. 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. Categorization of the services, also used to guide the user to relevant services in the OpenRiskNet Service catalogue, show that the OpenRiskNet goals to provide services for all applicability domains in risk assessment and for different user groups can be reached at the end of the project. For this, the service-integration work will be continued in year 3 again strengthening the collaboration with third parties in the Associated Partner Programme and the Implementation Challenge as well as the support of the case studies.

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. Oxford University Press; 2012;33: 1421–1429.

4. Berggren E, White A, Ouedraogo G, Paini A, Richarz A-N, Bois FY, et al. Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods. Comput Toxicol. 2017;4: 31–44.

5. 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

6. Spjuth O, Dudgeon T, Bachler D, Gebele D, Rautenberg M, Alvarsson J, et al. Report on deployment of virtual infrastructures with service discovery and container orchestration (Deliverable 2.3). 2018; doi:10.5281/zenodo.1479475