OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

WP4 Service Integration

Danyel Jennen (UM)

General Assembly and Final Consortium meeting
22 October 2019, Amsterdam (The Netherlands)

OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and *in silico* Analysis and Modelling in Risk Assessment
Project Number 731075

WP4 Objectives

- Integration of a representative number of services as best-practice examples with the corresponding step-by-step documentation
- Harmonisation of approaches to provide information on service status, capabilities and requirements over the interoperability layer
- Adoption of the data schemata and deployment as well as authentication & authorisation (optional) options proposed by WP 3 and 2, respectively
- Guarantee the error-free communication and integration of the services in the infrastructure and provide maintenance of the service throughout the duration of the project

WP4 Service Integration

ID	Title	Partners involved	Start month	End Month	Deliverables
T4.1	Toxicology, Chemical Properties and Bioassay Databases	UM, UoB, DC, NTUA	7	24	D4.1; D4.2; D4.3; D4.4
T4.2	Omics Databases	UM, UoB, CRG, DC, NTUA, CRG	7	24	D4.1; D4.2; D4.3; D4.4
T4.3	Knowledge Bases and Data Mining	UM, UoB, DC, NTUA, Fraunhofer	7	24	D4.1; D4.3; D4.4
T4.4	Ontology Services	UoB, UM	7	24	D4.1; D4.3; D4.4
T4.5	Processing and Analysis	DC, CRG, NTUA, UM, UoB, IM	7	24	D4.1; D4.3; D4.4
T4.6	Predictive Toxicology	JGU, NTUA, INERIS, UoB, UM, UU, IM	13	30	D4.3; D4.4
T4.7	Workflows, Visualisation and Reporting	IM, UU, UoB, DC, JGU	13	30	D4.3; D4.4

WP4 Deliverables

ID	Title	Due month	Lead partner	Type	Dissemination level
D4.1	Report of the Service Integration with OpenRiskNet (Initial Deployment)	12	UM	Report	Public
D4.2	Report of the Service Integration with OpenRiskNet (Intermediate Report)	24	UoB	Report	Public
D4.3	Report of the Service Integration with OpenRiskNet (Final Report)	36	UM	Report	Public
D4.4	Report on Re-Identification Risks and Private by Design Risk Management	12	DC	Report	Public

WP4 Milestones

ID	Title	Due month	Lead partner	Means of verification
MS7	**Adoption of the data schemata and deployment options by prototype services**	**18**	**UM**	**Well documented reference implementation of services**

WP4 Performance Metrics

Title
T4.1: Successful integration of >= **6** (until M18) and >= **10** (until M36) services
T4.2: Successful integration of >= **2** (until M18) and >= **4** (until M36) services
T4.3: Successful integration of >= **4** (until M18) and >= **10** (until M36) services
T4.4: Successful integration of >= **1** (until M18) and >= **2** (until M36) services
T4.5: Successful integration of >= **4** (until M18) and >= **6** (until M36) services
T4.6: Successful integration of >= **6** (until M18) and >= **10** (until M36) services
T4.7: Successful integration of >= **2** (until M18) and >= **3** (until M36) services

Services from ORN proposal a.k.a. Table 1.1

Service	Partner	Third Party	Service	Partner	Third Party
Toxicolog					
eChemPo					
ChEMBL					
ChEBI					
ChemSpid					
PubChem					
DrugBank	U				
ToxNet	U				

Service	Partner	Third Party	Service	Partner	Third Party
*omics Da					
diXa					
EGA					
Gene expr omnibus					

Service	Partner	Third Party	Service	Partner	Third Party
Knowledg					
AOP KB					
CTD					
Triple Ann Tool					
WikiPathways	U				
Pathway Annotation Tool	U				
Identifier Mapping Tool	U				
Chemistry	U				

Service	Partner	Third Party	Service	Partner	Third Party
Ontology					
Ontology Metadata					

Service	Partner	Third Party	Service	Partner	Third Party
Processir					
Toxygates					
Omics dat analysis to (TwinBoos					
Network m and analys (DiME/Co					
Image Analysis	N				
GO descriptor calculation	N				

Service	Partner	Third Party	Service	Partner	Third Party
Workflow					
Squonk					
Bioclipse					
Risk21 Ap					

Service	Partner	Third Party	Service	Partner
Predictive Toxicology				
Lazar	IST		BBRC	JGU
Nano-Lazar	IST		LAST-PM	JGU
PubChem Read Across	IST		FCDE	JGU
PBPK model	INERIS		PSCG	JGU

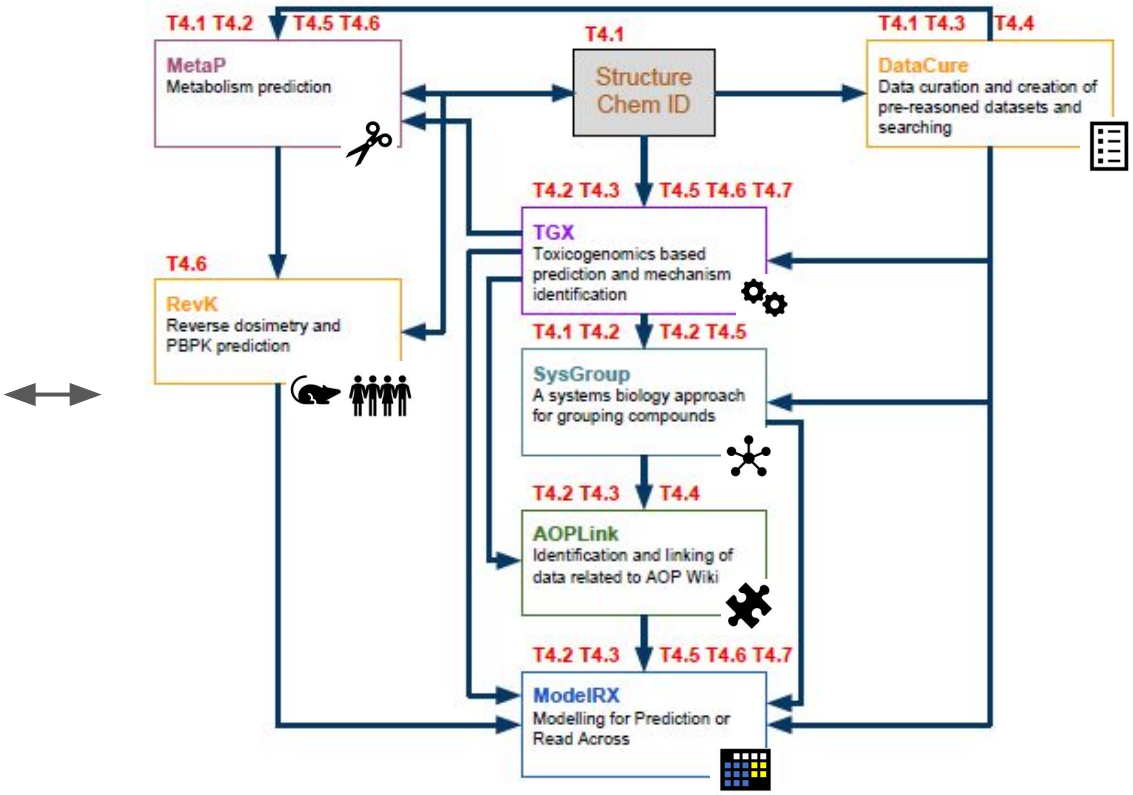
WP4 Performance Metrics (GA December 2018)

Title	Status M24	In progress*
T4.1: Successful integration of >= 6 (until M18) and >= 10 (until M36) services	5	*10*
T4.2: Successful integration of >= 2 (until M18) and >= 4 (until M36) services	1	*5*
T4.3: Successful integration of >= 4 (until M18) and >= 10 (until M36) services	4	*4*
T4.4: Successful integration of >= 1 (until M18) and >= 2 (until M36) services	-	*3*
T4.5: Successful integration of >= 4 (until M18) and >= 6 (until M36) services	2	*3*
T4.6: Successful integration of >= 6 (until M18) and >= 10 (until M36) services	6	*6*
T4.7: Successful integration of >= 2 (until M18) and >= 3 (until M36) services	3	*3*

*New services from implementation challenge not yet included

Tasks ↔ Case Studies

ID	Title
T4.1	Toxicology, Chemical Properties and Bioassay Databases
T4.2	Omics Databases
T4.3	Knowledge Bases and Data Mining
T4.4	Ontology Services
T4.5	Processing and Analysis
T4.6	Predictive Toxicology
T4.7	Workflows, Visualisation and Reporting



Service integration procedure → 8 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.

Task	Services integrated*	1	2	3	4	5	6	7	8
4.1	Squonk services for chemical property prediction	/		x		x			
	cpLogD		/	x	x	/			
	Modelling Web			x		x			
	CDK-Depict			x					
	Chemidconvert	x		x	x	x	x		x
	eNanoMapper - nanomaterial database	x							
	FDA Estrogenic Activity Database	x	/	x	x	x	x		/
	MetPred	x		x	x	x	x		
	LTKB			x		x			
	ToxRefDB	x		x					/
	ToxCast/Tox21 summary data	x		x	x				/
	Tox21 sample specific data	x							
4.2	TG-GATEs	x		x					
	Toxygates	x		x					
	diXa (via BioStudies)	/							
	Gene Expression Omnibus (GEO)	/							
	ArrayExpress	/							
	EGA Beacon							x	
	EGA Metadata API				x			x	x

Task	Services integrated*	1	2	3	4	5	6	7	8
4.3	BridgeDb	x		x		x		x	
	Data mining algorithms through Jaqpot	x	/	x	x	x	x	x	/
	Data mining algorithms through JGU Weka	x	x	x	x	x	x	*x*	/
	SCAIView Scientific Literature Database	/	/	x	/				x
	WikiPathways	/		/	x	/			
	AOP-Wiki	/		/		/			
	eNanoMapper - nanomaterial database	x		x	x	x	x	x	
4.4	Jenkins: ontology building and testing								
	Ontology Lookup Service (OLS)	/	/	x	/				
	Ontology Annotation Services (BELIEF Text Mining)	/	/	x	/				
4.5	Jaqpot processing and analysis services	x	/	x	x	x	x	x	/
	CDK descriptor calculation service	x	/	x	x	x	x	x	/
	PROAST and TCPL dose response modelling service	x							
	P450 SOM predictor	x	*x*	*x*	*x*		*x*		*x*
4.6	WEKA REST Service	x	x	x	x	x	x	x	/
	Lazar Toxicity Predictions	x	/	x	x	x	x		
	Jaqpot predictive modelling services	x	/	x	x	x	x	x	/
	Jaqpot applicability domain services	x	/	x	x	x	x	x	/
	Jaqpot PBPK modelling services	x	/	x	x	x	x	x	/
	httk package for PBPK modelling service	x	/	/	/	/	/		
4.7	Squonk Computational Notebook	/		x		x			
	Jupyter notebooks			x		x			
	Nextflow			x		x			

x = full compliance, / = partly implemented

*New services from implementation challenge not yet included

OpenRiskNet Services

LOGIN INSTRUCTIONS → THE REFERENCE ENVIRONMENT → SETTING UP YOUR OWN VRE →

OpenRiskNet services

Submit a service

Category | Service type | User type | Targeted users | Filter | Reset

OCHEM models

IN PROGRESS

Prediction of chemicals

Prediction of different endpoints

Provided by: BigChem GmbH
Type: Trained model
Applicability domain: Predictive toxicology
Topic: Chemical properties, Risk assessment, Structure-activity relationship (SAR / QSAR), Predictive modelling
Biological area: NOAEL/LOAEL, Acute toxicity

AOP-DB SPARQL Endpoint

IN PROGRESS

This service is a Virtuoso SPARQL endpoint that is loaded with RDF of the Adverse Outcome Pathway Database (AOP-DB) from the US EPA, who won the implementation challenge.

Provided by: Maastricht University
Type: Database / data source
Applicability domain: Toxicology, Bioinformatics
Topic: Risk assessment, Information extraction

✓ For end-users

OpenRiskNet services

Submit a service

Category | Service type | User type | Targeted users | Filter | Reset

Jaqpot API

INTEGRATED

Generate, store and share predictive statistical and machine learning models

Jaqpot is a user friendly web-based e-infrastructure containing many data analysis and modelling microservices integrated under harmonised APIs. The Jaqpot infrastructure allows the user to build applications that preprocess data, compute ...

Provided by: National Technical University of Athens
Type: Analysis tool, Processing tool, Trained model, Model generation tool, Model, Data mining tool, Service
Applicability domain: Computational modelling, Predictive toxicology
Topic: Biokinetics, Predictive modelling

- For developers
- For end-users

DETAILS →

WikiPathways SPARQL Endpoint

IN PROGRESS

WikiPathways was established to facilitate the contribution and maintenance of pathway information by the biology community. WikiPathways is an open, collaborative platform dedicated to the curation of biological pathways. WikiPathways ...

Provided by: Maastricht University
Type: Database / data source
Applicability domain: Bioinformatics
Topic: Information extraction
Biological area: Acute toxicity, Carcinogenicity, Mutagenicity, Genotoxicity, Skin sensitisation, Omics, Transcriptomics

- For end-users
- For developers

DETAILS → VISIT SERVICE →

Jaqpot API

Generate, store and share predictive statistical and machine learning models

Jaqpot is a user-friendly web-based e-infrastructure containing many data analysis and modelling microservices integrated under harmonised APIs. The Jaqpot infrastructure allows the user to build applications that preprocess data, compute descriptors from raw data (such as electronic images), create, validate, store and share predictive machine learning models and generate reports in standard formats. Jaqpot has been developed by the Unit of Process Control and Informatics in the School of Chemical Engineering at the National Technical University of Athens.

API definition →

✓ For developers ✓ For end-users

Type: Service, Data mining tool, Model, Model generation tool, Trained model, Processing tool, Analysis tool
Categories: Knowledge bases, Processing and analysis, API Definitions for OpenRiskNet applications and data
Applicability domain: Computational modelling, Predictive toxicology
Topic: Predictive modelling, Biokinetics
Targeted industry: Chemicals, Nanotechnology
Targeted users: Risk assessors, Researchers, Students, Software Developers, Data managers
Relevant OpenRiskNet case studies:

- ModelRX - Modelling for Prediction or Read Across
- RevK - Reverse dosimetry and PBPK prediction

Support contact: https://github.com/KinkyDesign/jaqpot-web/issues
Documentation: https://github.com/KinkyDesign/jaqpot-web/
References and training materials:

- Chomenidis et al, 2017 (https://pubs.acs.org/doi/abs/10.1021/acs.jcim.7b00223)
- Video: https://www.youtube.com/channel/UC-j4T5s5li4iMm75AAOiJ7w

Provided by: National Technical University of Athens
Contact: hsarimv@central.ntua.gr
Licence: GNU Lesser General Public License 3 (LGPLv3.0)
Login required: Yes
Implementation status: API documentation available (Swagger-OpenAPI v2), Containerised, Available as web service, Application programming interface available
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

RevK Pharmacokinetics OpenRiskNet Case study using Jaqpot web modelling platform — Tutorial
Philip Doganis
15 Oct 2018
— Video

Model RX OpenRiskNet - Case study using Jaqpot web modelling platform — Tutorial
Philip Doganis
15 Oct 2018

Task	Services integrated*	1	2	3	4	5	6	7	8
4.1	Squonk services for chemical property predictio	/							
	cpLogD	/							
	Modelling Web			x		x			
	CDK-Depict			x					
	Chemidconvert	x				x	x		x
	eNanoMapper - nanomaterial database	x							
	FDA Estrogenic Activity Database	x	/	x	x	x	x		/
	MetPred	x		x	x	x	x		
	LTKB					x			
	ToxRefDB	x		x					/
	ToxCast/Tox21 summary data	x							/
	Tox21 sample specific data	x							
4.2	TG-GATEs	x		x					
	Toxygates	x		x					
	diXa (via BioStudies)	/							
	Gene Expression Omnibus (GEO)	/							
	ArrayExpress	/							
	EGA Beacon							x	
	EGA Metadata							x	x

Task	Services integrated*	1	2	3	4	5	6	7	8
4.3	BridgeDb	x							
	Data mining algorithms through Jaqpot	x			x	x	x	x	/
	Data mining algorithms through JGU Weka	x						x	/
	SCAIView Scientific Literature Database	/			/				x
	WikiPathways	/							
	AOP-Wiki	/							
	eNanoMapper - nanomaterial database	x			x		x	x	
4.4	Jenkins: ontology building and testing								
	Ontology Lookup Service (OLS)	/	/	x	/				
	Ontology Annotation Services (BELIEF Text Mining)	/	/	x	/				
4.5	Jaqpot processing and analysis services	x			x	x	x	x	/
	CDK descriptor calculation service	x			x	x	x	x	/
	PROAST and TCPL dose response modelling service	x							
	P450 SOM predictor	x				x			x
4.6	WEKA REST Service	x						x	/
	Lazar Toxicity Predictions	x							
	Jaqpot predictive modelling services	x	/	x	x	x	x	x	/
	Jaqpot applicability domain services	x			x	x	x	x	/
	Jaqpot PBPK modelling services	x	/	x	x	x	x	x	/
	httk package for PBPK modelling service	x	/	/	/	/	/		
4.7	Squonk Computational Notebook	/							
	Jupyter notebooks					x			
	Nextflow				x				

Squonk Comp Notebook

cpLogD confidence predictor for logD

ChemIdConvert

eNanoMapper database

MetPred

LTKB APIs

EdelWeissData (Data explorer)

Human, mouse & rat
in vitro liver data available

EGA Beacon

EGA Metadata API

BridgeDb identifier mapping service

Jaqpot GUI

JGU WEKA REST Service

SCAIView

WikiPathways SPARQL endpoint

AOP-Wiki SPARQL endpoint

eNanoMapper Database

Jaqpot GUI

Jaqpot API

P450 SOM predictor

JGU WEKA REST Service

Lazar Toxicity Predictions

Jaqpot GUI

Jaqpot API

Squonk Comp Notebook

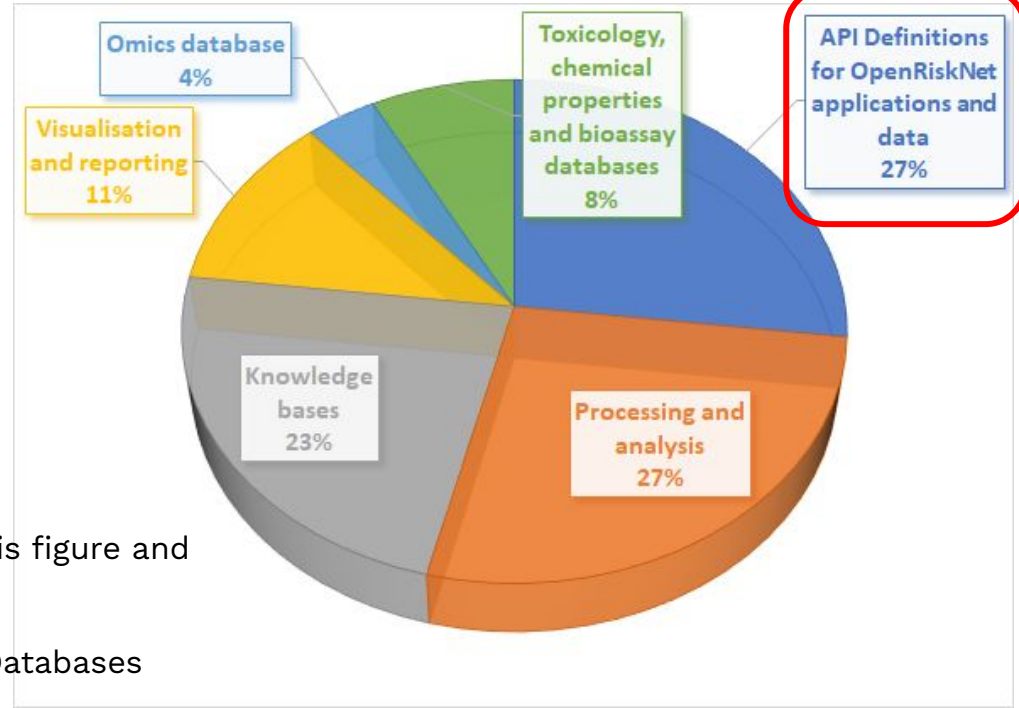
Jupyter Notebooks

Nextflow

x = full compliance, / = partly implemented

*New services from implementation challenge not yet included

Categories of services in ORN catalogue



There is an inconsistency between the categories in this figure and the ones in the tasks description

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

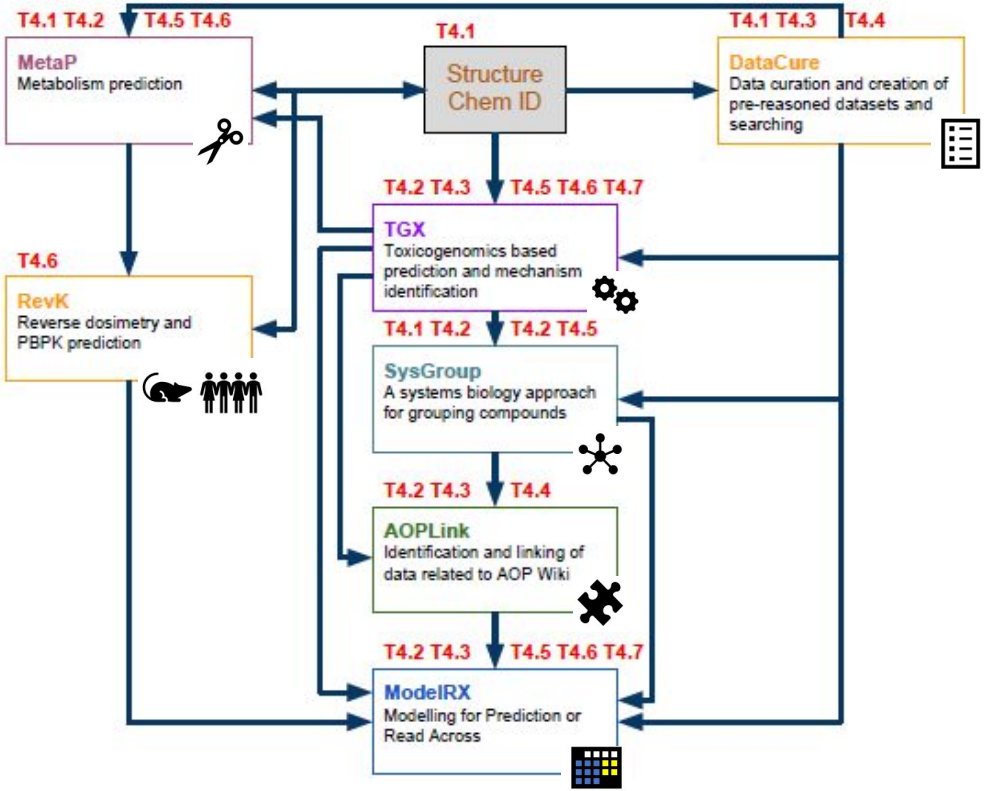
New services from the Implementation Challenge

							
Holly Mortensen, US EPA - “**AOP-DB** (The Adverse Outcome Pathway Database)”				x	x		
Hyun Kil Shin, Korea Institute of Toxicology - “**Daphnia magna nanotoxicity database**” and “**nano-QSAR to predict cytotoxicity of metal and metal oxide nanoparticles**”	x					x	
Matthias Timberlake, ToxPlanet - “**ToxPlanet database**”	x		x	x		x	x
Johannes Kirchmair, Universität Hamburg - “**FAst MEtabolizer (FAME)**”		x					
Antreas Afantitis, NovaMechanics Ltd - “**Enalos InSilicoNano platform: an online decision support tool for the design and virtual screening of nanoparticles**” and “**A Risk Assessment Tool for the Virtual Screening of Metal Oxide Nanoparticles through Enalos InSilicoNano Platform**”						x	
Igor Tetko, BIGCHEM GmbH - “**OCHEM models**”, “**OCHEM descriptors**” and “ **OCHEM model development tool**”	x					x	
Urban Fagerholm, Prosilico - **“Prosilico Human Clinical ADME/PK-Studio”**		x					x
Rachael Skyner, Diamond Light Source Ltd. - **“BruteReg”** and **“PySquonk”**	x	x	x	x	x	x	x
Benjamin Haibe-Kains, University Health Network - **“ToxicoDB”**	x		x	x	x		
Katy Wolstencroft, Leiden University (on behalf of EJP RD) - **“ToxTargetLinks”**					x		

Service integration in Case Studies

Example MetaP

Example TGX



Metabolism Prediction [MetaP]

CS leader: Daan Geerke (VU), **Involved:** UU, JGU, UHH

Risk Assessment Framework
Tier 0.1 (mol. structure), 1.5 (biokinetics), 1.6 (MoA)

AIM: Integration of tools for site-of-metabolism (SOM) prediction and metabolite prediction

Ligand-based metabolite predictors (e.g. MetPred) and incorporate protein-structure and -dynamics based approaches to predict the site of metabolism (SOM) by **Cytochrome P450 (CYPs)**, which metabolize ~75% of the currently marketed drugs.

Objectives: Integration, comparison and combination of tools for metabolism prediction

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

openrisknet.org/e-infrastructure/development/case-studies/case-study-metap

Databases During method development, model calibration and validation we will use data from XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites.

XMetDB, SMARTcyp, ZINC, ChEMBL, EAWAG-BBD

Tools / APIs

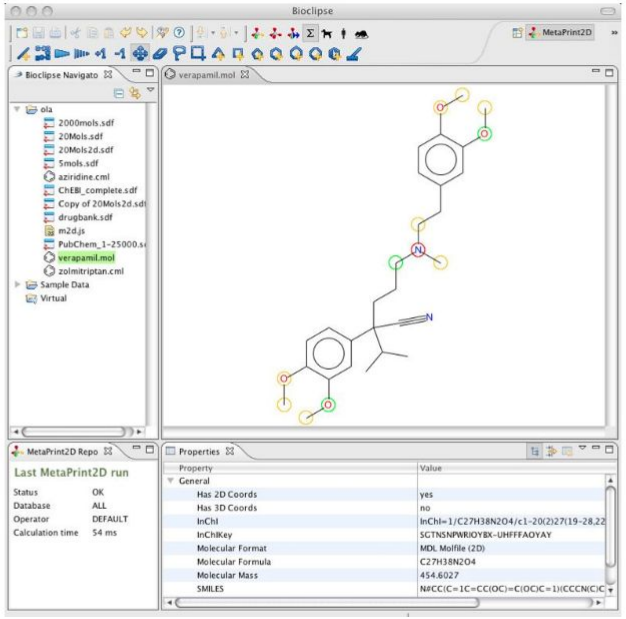
- MetPred (UU)
- Metaprint2D & MetVap (UU)
- UM-PPS (JGU)
- enviPath (JGU)
- SMARTCyp (external service, integration by VU)
- Plasticity tools (VU)
- FAME (UHH; implementation challenge)

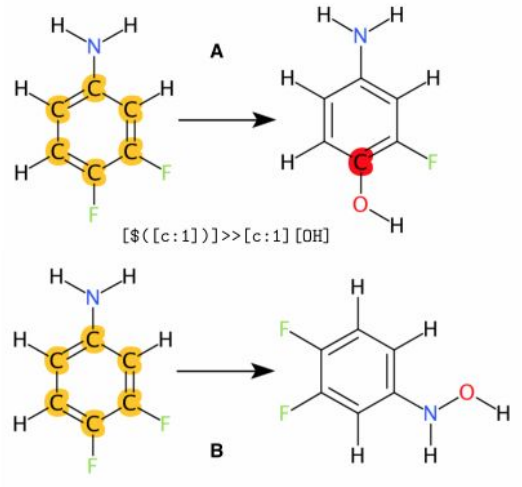
Service integration

To facilitate combining metabolite prediction approaches and using MetaP outcomes as input for other predictors, we will take advantage of ongoing development in workflow management systems (Nextflow, Squonk, MDStudio) and we will explore integration into/with and use of these platforms. Once integrated the added value of multiple predictors will be subject of a pilot study on metabolite prediction..

MetPred

Predicts phase I metabolites: MetPred ranks most probable sites-of-metabolism (SOMs) and reaction types based on similar atom environments and ReactionSMARTS in annotated dataset [webservice; API available]

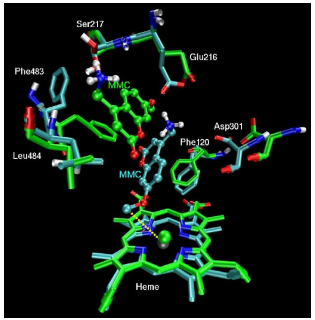


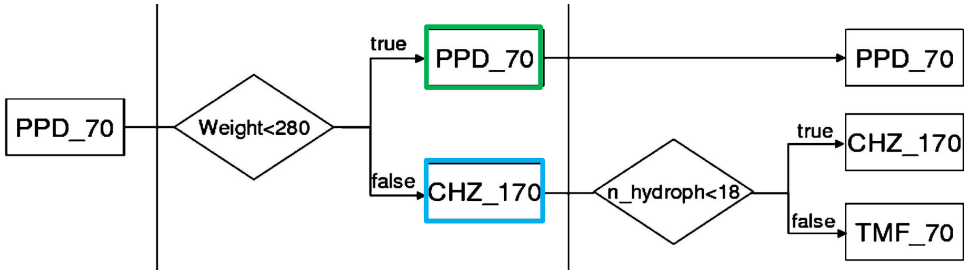


Carlsson et al. BMC Bioinformatics 2010
Arvidsson et al. Proc. Machine Learn. Res. 2017

Last years expansion of the toolbox

- **FAME 2.0**: SOM prediction (also for phase I, phase II or enzyme/isoform specific metabolism) from machine learning using (<)15 quantum and circular-environment based atomic descriptors [implementation challenge]
- **enviPath**: prediction of microbial biotransformation pathways and products using rules represented by SMIRKS
- **SMARTCyp**: SOM prediction for P450 metabolism based on fragment-mapping to pre-computed high-level QM data and atomic accessibility, extended with simple ligand-based pharmacophore rules for specific isoforms
- **Protein-structure based predictors**: plasticity models for docking into (flexible) P450 isoforms





Adapted from Hritz et al. J Med Chem 2008

Complementary tools, and we will explore added value of combined use for e.g. **consensus prediction**

Toxicogenomics based prediction and mechanism identification [TGX]

CS leader: Danyel Jennen (UM), **Involved:** UM, VU, CRG

Risk Assessment Framework
Tier 0.3-0.4 (data collection), 1.6 (MOA)

AIM: To provide a transcriptomics-based hazard prediction model for identification of specific molecular initiating events (MIE)

The foreseen transcriptomics-based hazard prediction model will be applied based on:

(A) top-down Creation of prediction models based on differentially regulated genes

(B) bottom-up Using knowledge of stress response pathways to integrate data sets for their activation or inhibition (bottom-up approach).

The MIEs can include, but are not limited to:

(1) Genotoxicity (p53 activation), (2) Oxidative stress (Nrf2 activation), (3) Endoplasmic Reticulum Stress (unfolded protein response), (4) Dioxin-like activity (AhR receptor activation), (5) HIF1 alpha activation and (6) Nuclear receptor activation (e.g. for endocrine disruption).

Databases

- diXa / BioStudies (UM)
- TG-GATEs
- EU-ToxRisk (nascent)
- HeCaToS (nascent)
- ArrayExpress / GEO

Tools / APIs

- top-down: Data normalisation tools, prediction tools such as Caret;
- bottom-up: ToxPi

Service integration

- Service integration will be needed for the omics databases; knowledge bases and data mining; processing and analysis.

Activities

- First top-down case study based on Magkoufopoulou *et al* 2012 paper
- Second top-down case study on meta-analysis for genotoxicity prediction in human, rat and mouse in vitro cell models

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

Example TGX (1) → top-down approach

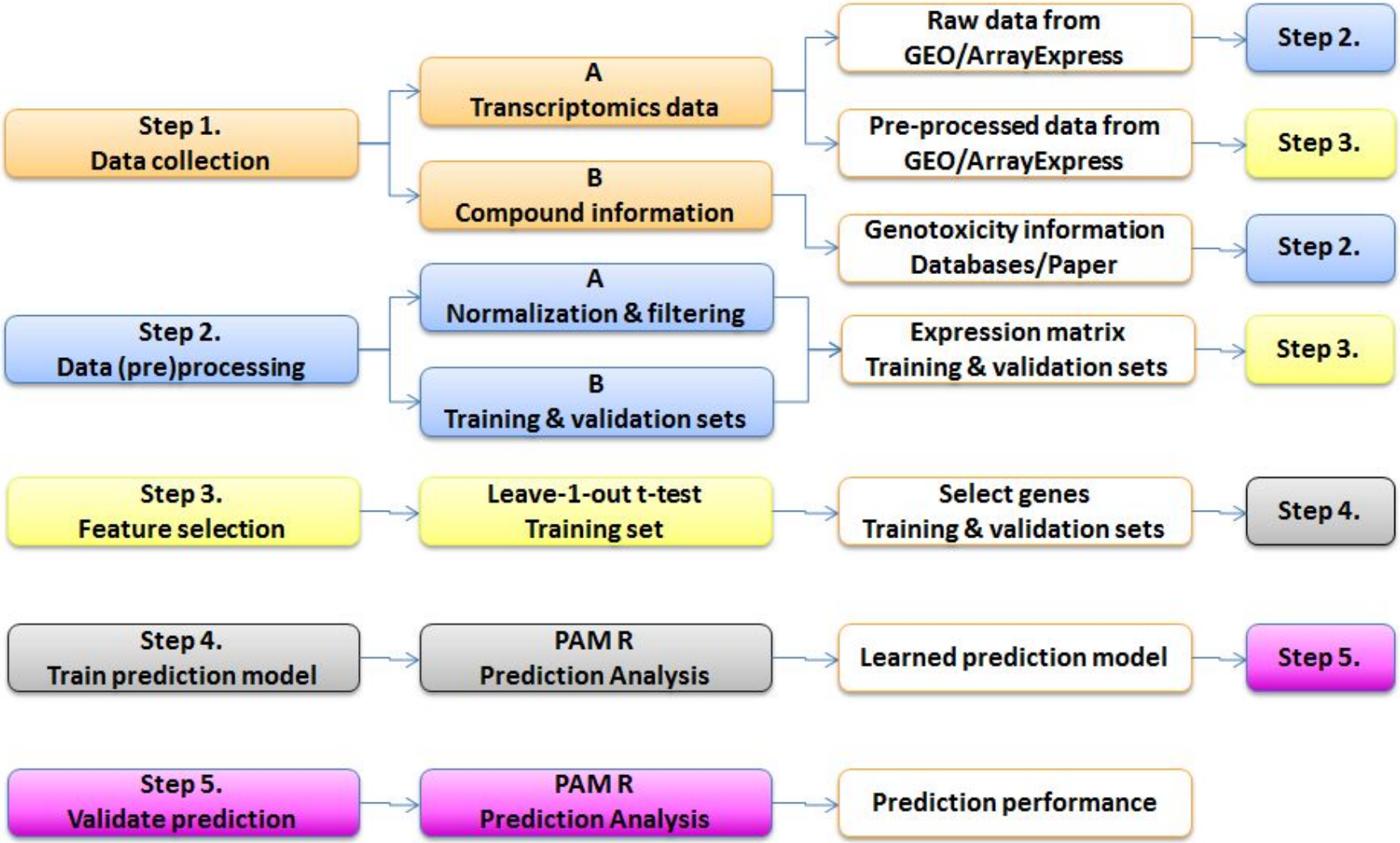
Case study 1:

Create workflow based on Magkoufopoulou *et al.* 2012

Carcinogenesis vol.33 no.7 pp.1421–1429, 2012
doi:10.1093/carcin/bgs182
Advance Access Publication May 23, 2012

A transcriptomics-based *in vitro* assay for predicting chemical genotoxicity *in vivo*

C.Magkoufopoulou[1,2], S.M.H.Claessen[1], M.Tsamou[1], D.G. J.Jennen[1,2], J.C.S.Kleinjans[1,2], J.H.M.van Delft[1,2,*]


Step 1.
Data collection
A
Transcriptomics data
B
Compound information
Raw data from
GEO/ArrayExpress
Pre-processed data from
GEO/ArrayExpress
Genotoxicity information
Databases/Paper
Step 2.
Step 3.
Step 2.
Step 2.
Data (pre)processing
A
Normalization & filtering
B
Training & validation sets
Expression matrix
Training & validation sets
Step 3.
Step 3.
Feature selection
Leave-1-out t-test
Training set
Select genes
Training & validation sets
Step 4.
Step 4.
Train prediction model
PAM R
Prediction Analysis
Learned prediction model
Step 5.
Step 5.
Validate prediction
PAM R
Prediction Analysis
Prediction performance

Example TGX (2)

Current status

Workflow has been established in Snakemake and is available via Gitlab

Workflow is converted into NextFlow pipeline;

Converted into a generic workflow applicable to other datasets.

Next steps

Workflow will be containerized;

(Publication of approach)

Example TGX (3) → Generic workflow

- Data collection from GEO/ArrayExpress → API available
- Retrieve (geno)toxicity information
  - from databases
    - possible, but depends of required data
    - expert knowledge needed
  - from paper
    - difficult as presentation of information may vary
    - online, pdf → difficult, because of format;
      suppl. data as Word, Excel, txt files → possible
- All other steps comprise of R-scripts → easy to adapt

toxplanet

TOXICODB

Fraunhofer SCAI

→ data collection relies on available metadata, used format & ontologies

Collaboration with implementation challenge winners

toxplanet

Find the chemical hazard and toxicology literature you need with our information portal

Our Federated Search engine searches the content from 500+ websites and quickly delivers relevant literature.

UHN

TOXICODB

Investigate the pathways triggered by exposure to toxic substances (coming soon)

PDF: Table detection and -segmentation

Fraunhofer SCAI

Example TGX (4) → top-down approach

Case study 2:

Meta-analysis for genotoxicity prediction in human, rat and mouse *in vitro* cell models

In this case study transcriptomics data obtained from multiple data sources will be used to build a prediction model for *in vivo* genotoxicity.

A similar approach will be applied as in case study 1.

Preliminary results on the rat data have been presented at the ICCA-LRI workshop 2018, Ottawa, Canada
Preliminary results on the human data have been presented at EUROTOX 2018, Brussels, Belgium

Example TGX (5) → top-down approach

The generic workflow will be applicable in this case study as well.

- Additional databases have been used, e.g. the diXa Data Warehouse
  → included in EBI's BioStudies → an API is available
- In vitro liver transcriptomics data from multiple human, rat & mouse cell models is available via Gitlab
- (Geno)toxicity data on all compounds has been collected
  → valuable source for ToxicoDB

Other services from the implementation challenge can be incorporated.
→ ToxPlanet databases potentially can provide genotoxicity information

Deliverable 4.3

To do:

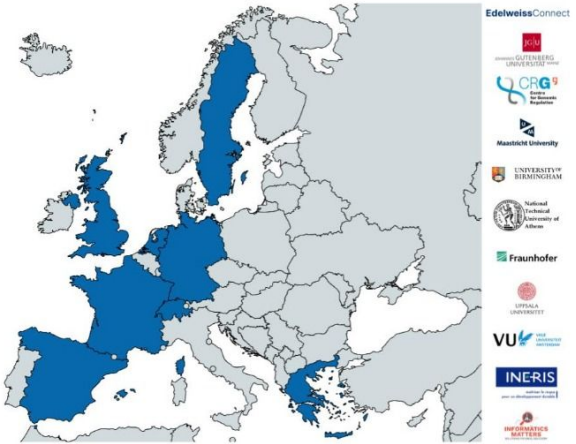
Update Deliverable 4.2 with the latest numbers:

- Add new services also from implementation services to table on slide 11
- Adjust categories of services included in OpenRiskNet catalogue to the ones described in the tasks (see slide 16)
- Integrate (new) services with case studies

Acknowledgements

OpenRiskNet (Grant Agreement 731075) is a project funded by the European Commission within Horizon 2020 Programme

Project partners:



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 INERIS, France (INERIS)
P12 Vrije Universiteit Amsterdam, Netherlands (VU)