OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

WP4 Service Integration

Danyel Jennen (UM)

General Assembly and 2nd annual meeting - 12-13 December 2018, Brussels (Belgium)

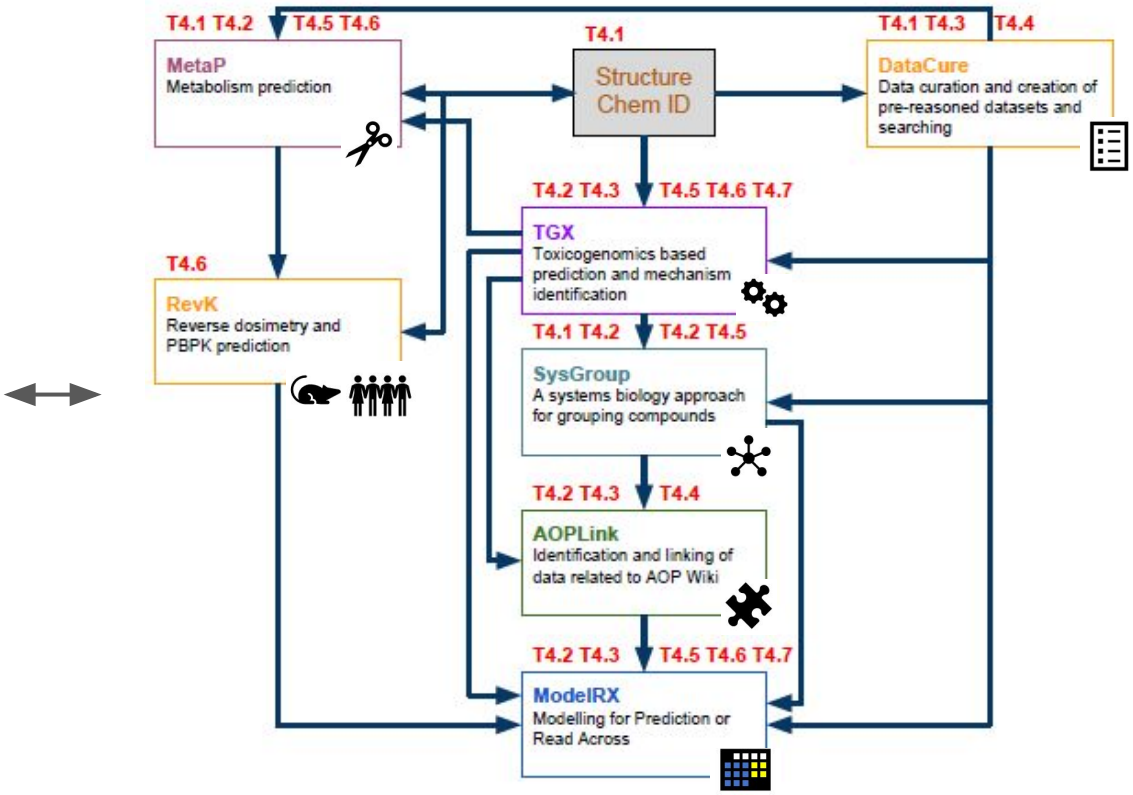
OpenRiskNet: Open e-Infrastructure to Support Data Sharing, Knowledge Integration and *in silico* Analysis and Modelling in Risk Assessment
Project Number 731075

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

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 - confidence predictor for logD		/	x	x	/			
	Modelling Web			x		x			
	CDK-Depict			x					
	Chemidconvert	x		x	x	x	x		x
	eNanoMapper - nanomaterial database	x							
	ToxRefDB	x		x					/
	ToxCast/Tox21 summary data	x		x	x				/
	Tox21 sample specific data	x							
	FDA Estrogenic Activity Database	x	/	x	x	x	x		/
4.2	TG-GATEs	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	x	/	x	x	x	x	/	/
	Data mining algorithms through JGU Weka	x	/	x	x	x	x		/
	SCAIView Scientific Literature Database	/	/	x	/				x

Task	Services integrated	1	2	3	4	5	6	7	8
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							
4.6	WEKA REST Service	x	/	x	x	x	x		/
	Lazar Toxicity Predictions	x	/	x	x	x	x		
	Jaqpot predictive modelling services	x	/	x	x	x	x	/	/
	Jaqpot PBPK modelling services	x	/	x	x	x	x	/	/
	Jaqpot applicability domain services	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

OpenRiskNet Services

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

OpenRiskNet Services

Category | Service type | User type | 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: Service, Data mining tool, Model, Model generation tool, Trained

Lazar Toxicity Predictions

INTEGRATED

Toxicity prediction

Lazar (Lazy Structure-Activity Relationships) takes a chemical structure as input and provides predictions for a variety of toxic properties. Lazar uses an automated and reproducible read across procedure to calculate ...

Provided by: in silico toxicology gmbh, Johannes Gutenberg Univertity
Type: Trained model, Application, Service
Applicability domain: Toxicology, Predictive toxicology

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 General Public License 3 (GPLv3)
Login required: Yes
Implementation status: API documentation available (Swagger-OpenAPI v2), Containerized, 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
— Video

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

Task	Services integrated	1	2	3	4	5	6	7	8
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								
	CDK descriptor calculation service								
	PROAST and TCPL dose response modelling service	x							
4.6	WEKA REST Service								
	Lazar Toxicity Predictions								
	Jaqpot predictive modelling services								
	Jaqpot PBPK modelling services								
	Jaqpot applicability domain services								
	httk package for PBPK modelling service	x	/	/	/	/	/		
4.7	Squonk Computational Notebook								
	Jupyter notebooks								
	Nextflow					x			

Squonk Comp Notebook

cpLogD

Data explorer

BridgeDb

Jaqpot GUI

JGU WEKA REST Service

SCAIView

Jaqpot GUI

Jaqpot API

JGU WEKA REST Service

Lazar Toxicity Predictions

Jaqpot GUI

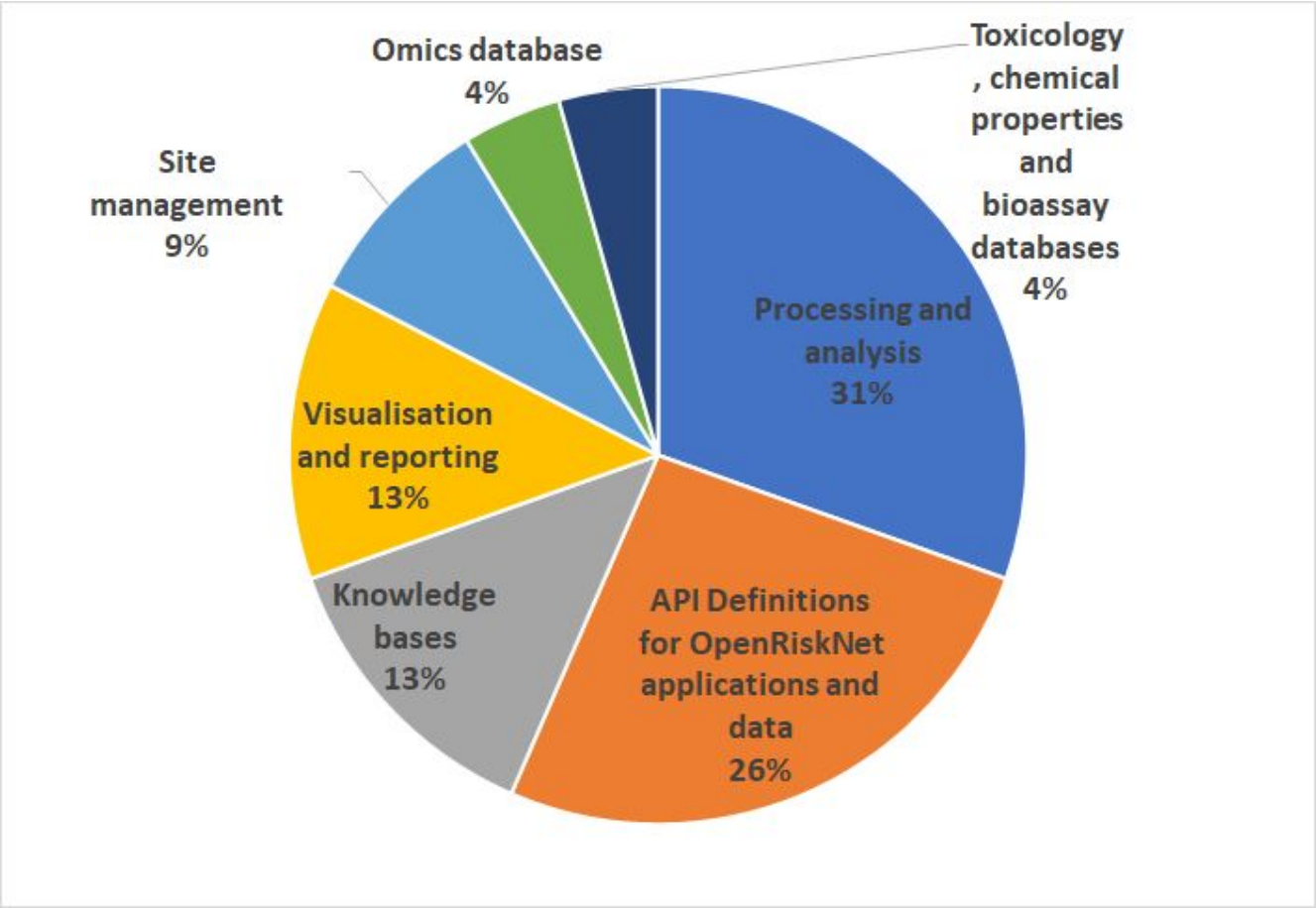
Jaqpot API

Squonk Comp Notebook

Jupyter Notebooks

Nextflow

x = full compliance, / = partly implemented


Omics database
4%
Toxicology
, chemical
properties
and
bioassay
databases
4%
Site
management
9%
Processing and
analysis
31%
Visualisation
and reporting
13%
Knowledge
bases
13%
API Definitions
for OpenRiskNet
applications and
data
26%

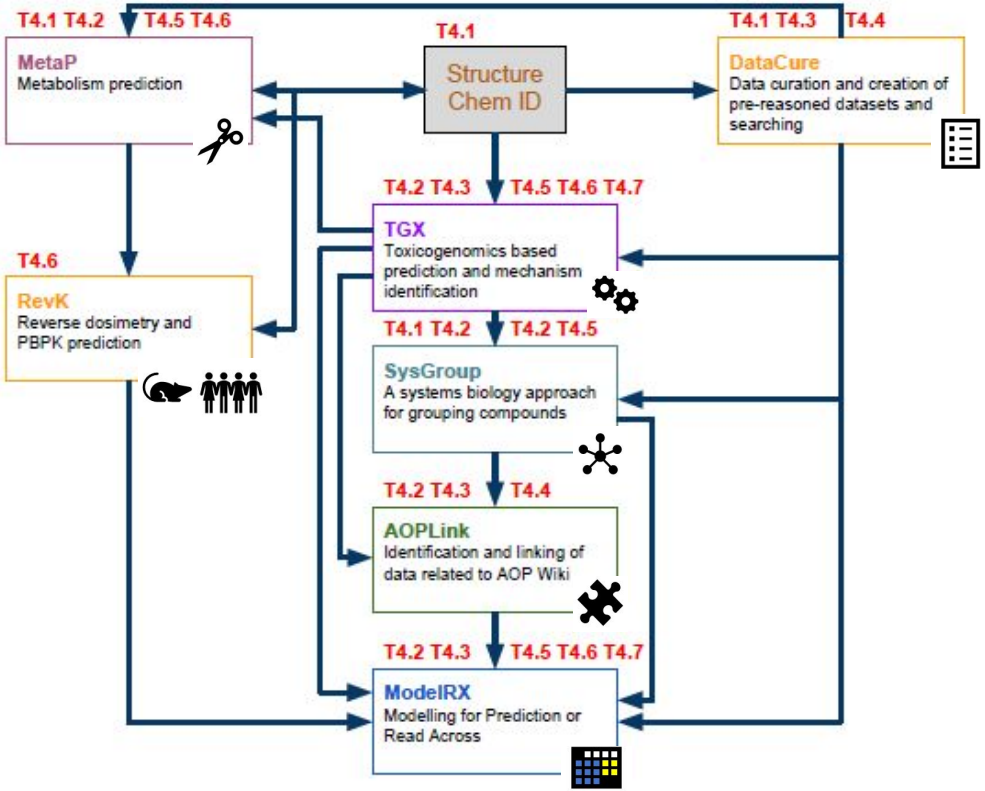
New services from the Implementation Challenge

							
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	

Service integration in Case Studies

Example MetaP → WP1

Example TGX



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 Magkoufopolou *et al* 2012 paper has been converted into NextFlow pipeline
- Second top-down case study on meta-analysis for genotoxicity prediction is under construction

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

Example TGX (1) → top-down approach

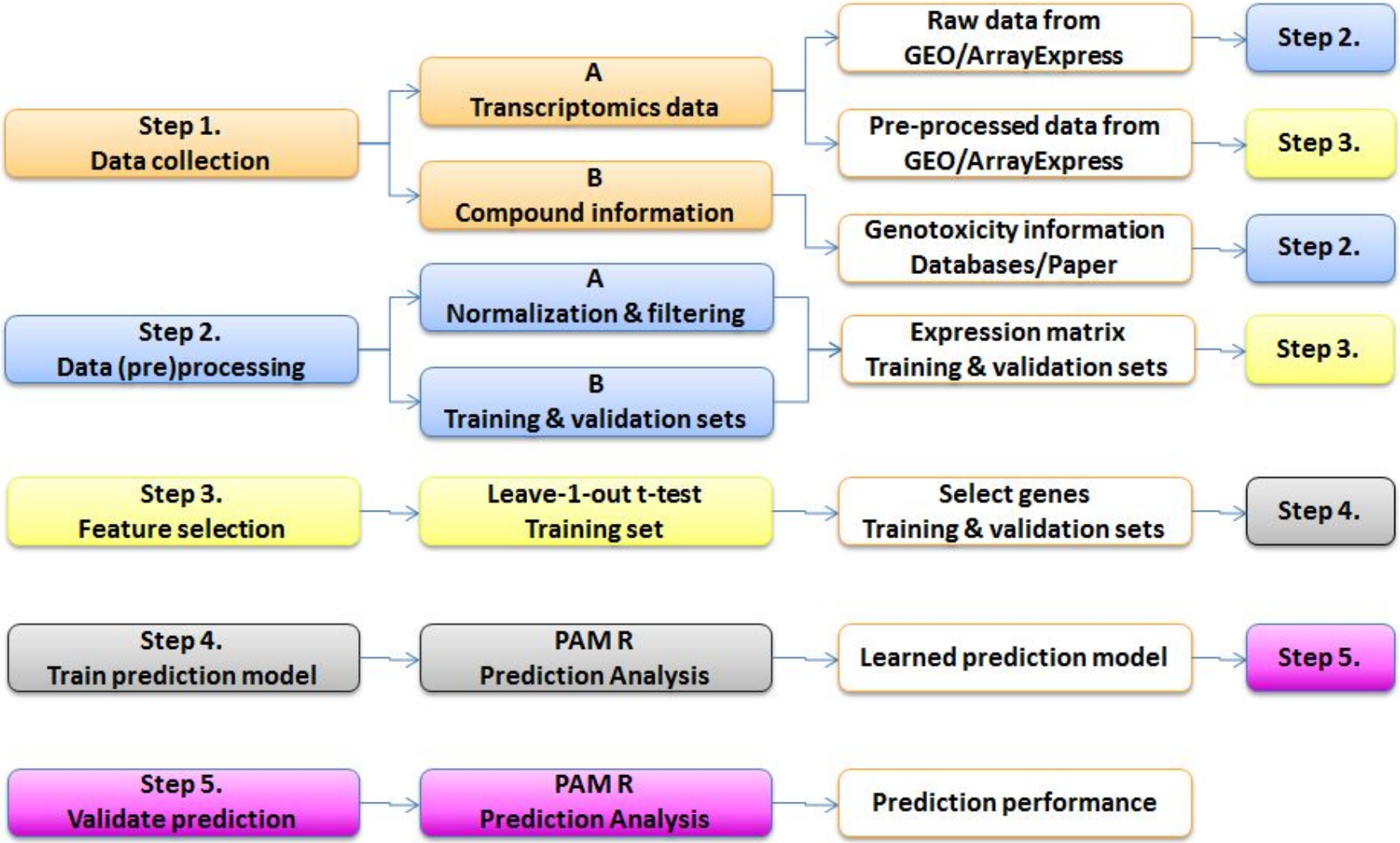
Case study 1:

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.

Workflow is converted into NextFlow pipeline;

Next steps

Workflow will be containerized;

Conversion into a generic workflow applicable to other datasets.

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

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

Example TGX (4) → top-down approach

Case study 2:

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

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

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.

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

Additional databases will be used, e.g. the diXa Data Warehouse
→ an API needs to be developed

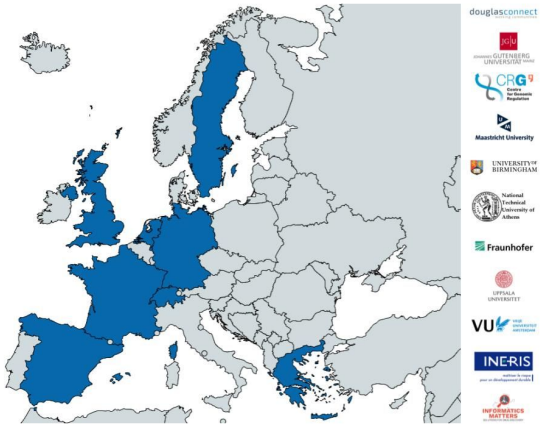
New services from the implementation challenge can be incorporated.

→ the ToxPlanet database potentially can provide genotoxicity information

Acknowledgements

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

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