OpenRiskNet

RISK ASSESSMENT E-INFRASTRUCTURE

WP1 Requirement Analysis, Outreach and Case Studies - progress

Paul Jennings (VU) and Thomas Exner (EwC)

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

WP1 Objectives

Runs from month 1 to 36

- **Requirement analysis:** Insights from stakeholders end users (*researchers, risk assessors and regulators in the fields of chemistry, pharma, cosmetics, nanomaterials*) as well as data providers, tool developers and workflow integrators regarding current challenges. (***Task 1.1**, 1 to 6 M*)
- **Outreach Program:** Directed towards current and past EU consortia as well as other related activities and individuals involved in the development of solutions for animal free risk assessment. (***Task 1.2**, 1 to 36 M*)
- **Case studies:** Testing the functionality of the e-infrastructure within the context of real-world applications. (***Task 1.3**, 1 to 36 M*)
- **Usability testing** based on the proposed case studies in parallel to the ongoing (WP2) developments and (WP4) integration efforts (**no specific task** but should be included in report)

WP1 Requirement Analysis, Outreach and Case Studies

ID	Title	Partners involved	Start month	End Month	Deliverables
T1.1	Requirement Analysis	EwC, VU	1	6	D1.1
T1.2	Community Outreach	VU, EwC	1	36	D1.2; D1.4
T1.3	Case Studies	VU, EwC, JGU, CRG, UM, UoB, NTUA, Fraunhofer, UU, IM, INERIS	1	36	D1.3; D1.5

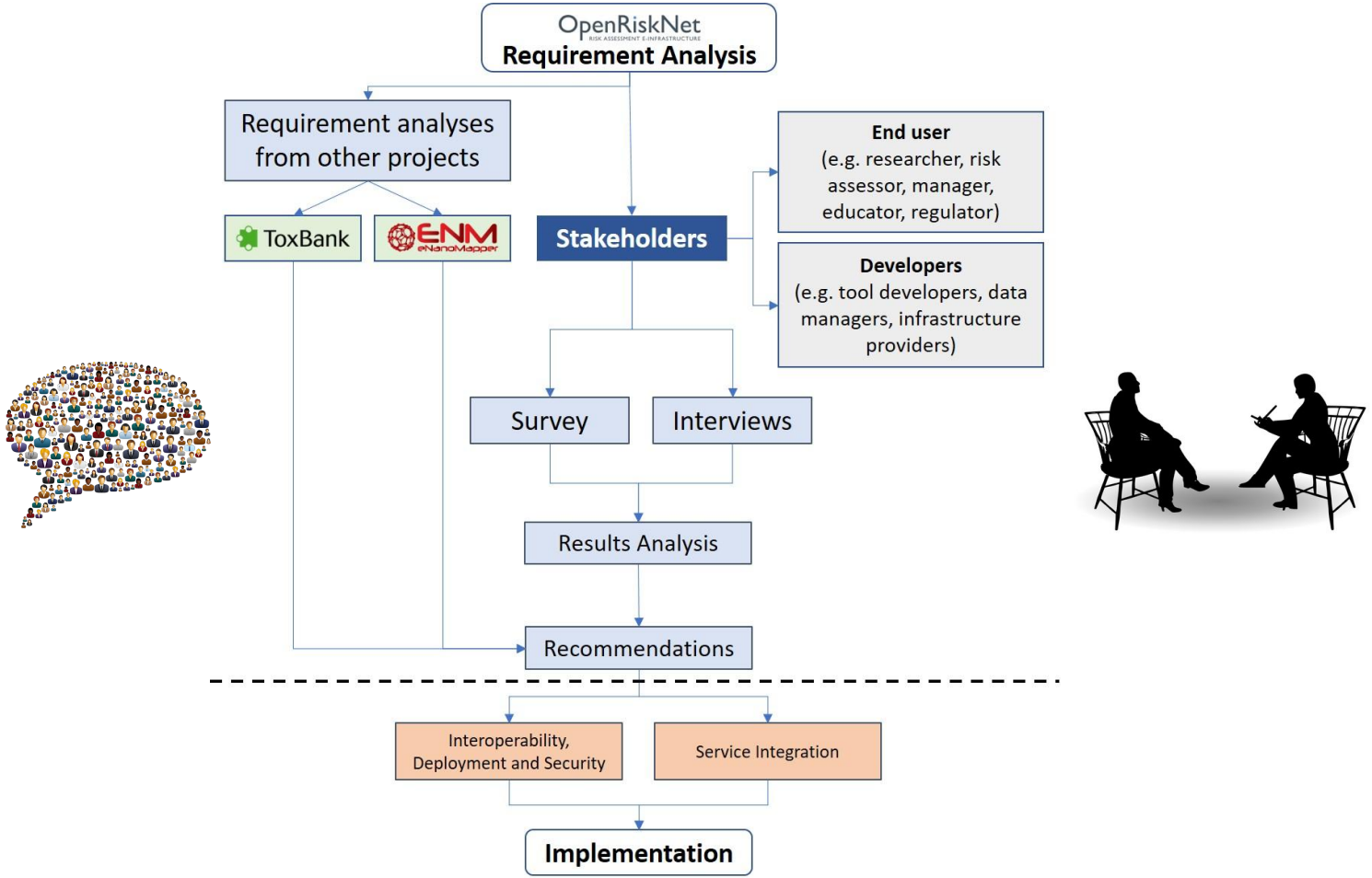
WP1 Deliverables

ID	Title	Due month	Due date	Lead partner	Type	Dissemination level
D1.1	**Report on requirement analysis and recommendations for WP2-4 including Data Management Plan - Version 2**	**6**	✓ **30 Nov 2018**	**EwC**	**Report**	**Public**
D1.2	**Associated partner program is established - Version 2**	**12**	✓ **30 Nov 2018**	**EwC**	**Other**	**Public**
D1.3	**Final definition of case studies**	**12**	✓ **30 Nov 2017**	**VU**	**Other**	**Public**
D1.4	Report on the results of the Implementation Challenge	36	30 November 2019	EwC	Report	Public
D1.5	Finalization of case studies and analysis of remaining weaknesses	36	30 November 2019	VU	Other	Public

WP No	Del Rel.	Del No	Title	Description	Lead I	Nature	Dissemination L	Est. Del. Date (an	Rev. Due Da	Receipt Date	Approval Date	Status
WP1	D1.1	D1	Report on requirement analysis and recommendations for W	The report will describe the results obtained f...	DC	Report	Public	31 May 2017		21 Jan 2019		Submitted
WP1	D1.2	D5	Associated partner program is established	Rules for the associated partner program will b...	DC	Other	Public	30 Nov 2017		21 Jan 2019		Submitted
WP1	D1.3	D6	Final definition of case studies	Case studies for testing the usability of the i...	VU	Other	Public	30 Nov 2017		18 Dec 2017	18 Dec 2018	Approved
WP1	D1.4	D17	Report on the results of the Implementation Challenge	This report will summarize the selected data so...	DC	Report	Public	30 Nov 2019				Pending
WP1	D1.5	D18	Finalization of case studies and analysis of remaining weakn	All case studies will have been finalized and r...	VU	Other	Public	30 Nov 2019				Pending

T1.1: Requirement Analysis

Deliverable 1.1


OpenRiskNet
RISK ASSESSMENT E-INFRASTRUCTURE
Requirement Analysis
Requirement analyses from other projects
ToxBank
eNanoMapper
Stakeholders
End user
(e.g. researcher, risk assessor, manager, educator, regulator)
Developers
(e.g. tool developers, data managers, infrastructure providers)
Survey
Interviews
Results Analysis
Recommendations
Interoperability, Deployment and Security
Service Integration
Implementation

Conclusions, Task 1 Requirement analysis

Collected invaluable up-to-date information on the tools, data sources, data annotation and protocoling requirements.

Optimize specific parts of the infrastructure concept improving the usability and the user-friendliness of the solution.

To do for final report

Everyone should collect events where she/he was talking to stakeholders about OpenRiskNet:

- After talks or during poster presentations
- During trade exposes (probably mainly Edelweiss Connect and Informatic Matters)
- Taking to associated partners and Implementation Challenge winners
- Ongoing project meeting (NanoSafety Cluster, NanoCommons, EU-ToxRisk) - Did we do any customization of OpenRiskNet services according to specifications coming from these interactions?
- New proposals

T1.2: Requirement Community Outreach

Deliverable 1.2: 1.4

Task 1.2 Community Outreach

Associated Partner Programme

To ensure the usability of the infrastructure, alignment with the community, as well as to pursue complete coverage of important data and tools for risk assessment

Officially launched at the OpenTox Euro 2017 conference (November 2017) >> Read the press release

- **Service providers** - integrate their databases and software tools into the OpenRiskNet infrastructure:
  - greater visibility of their tools by being listed in the OpenRiskNet discovery service;
  - infinite additional features by combining with other tools;
  - support for emerging techniques like API development and containerization/deployment.
- **Early adopters** - use the infrastructure for their predictive toxicology and risk assessment tasks:
  - easy access to a increasing number of tools using their preferred access route (web, workfl ow tools like knime, scripts) without the need of manually downloading of data and file conversion when moving from one tool to another;
  - harmonized access for comparison of different approaches.
- **Technology partners** - use their services and tools for the OpenRiskNet e-infrastructure development

https://openrisknet.org/associated-partner-programme/

Task 1.2 Community Outreach

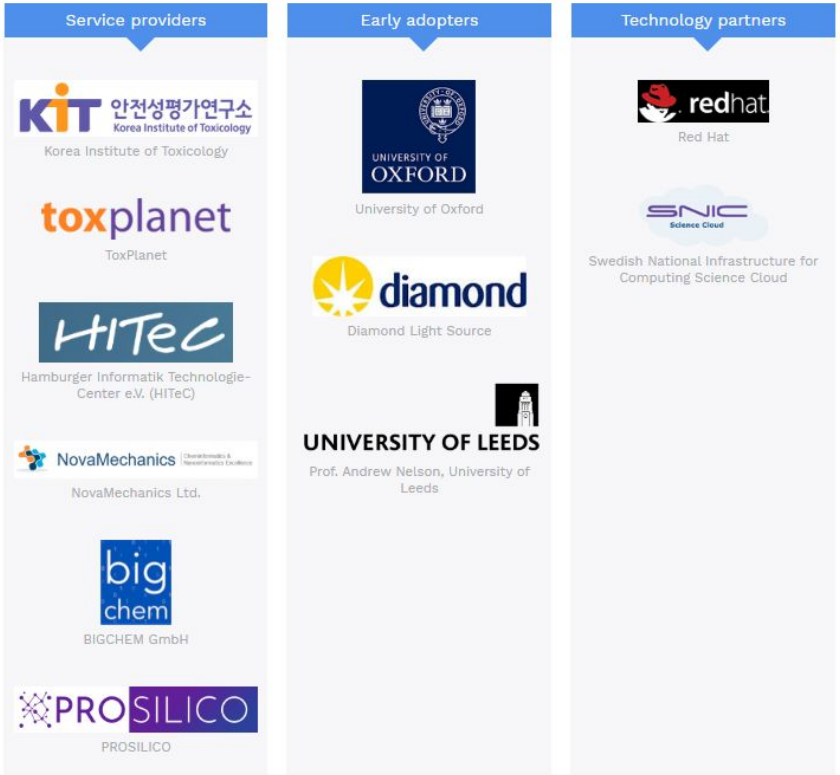
Programme Guideline ✅

Standard Agreement ✅

1. **Service providers**
2. **Early adopters**
3. **Technology partners**

Many others including major projects (NanoCommons, EU-ToxRisk, in3) but as gentlemen agreement

List of associated partners



Task 1.1 and 1.2 Contact partners

- Holly Mortensen, US EPA ➔ UM
- Hyun Kil Shin, Korea Institute of Toxicology ➔ EwC
- Matthias Timberlake, ToxPlanet ➔ UM/EwC
- Johannes Kirchmair, Universität Hamburg ➔ VU Amsterdam
- Antreas Afantitis, NovaMechanics Ltd ➔ NTUA
- Igor Tetko, BIGCHEM GmbH ➔ JGU
- Rachael Skyner, Diamond / Informatics Matters ➔ IM
- Urban Fagerholm, ProSilico ➔ UU
- Benjamin Haibe-Kains, University Health Network ➔ UM
- Katy Wolstencroft, Leiden University ➔ UM

Implementation Challenge - Agreements status

Issues / actions:

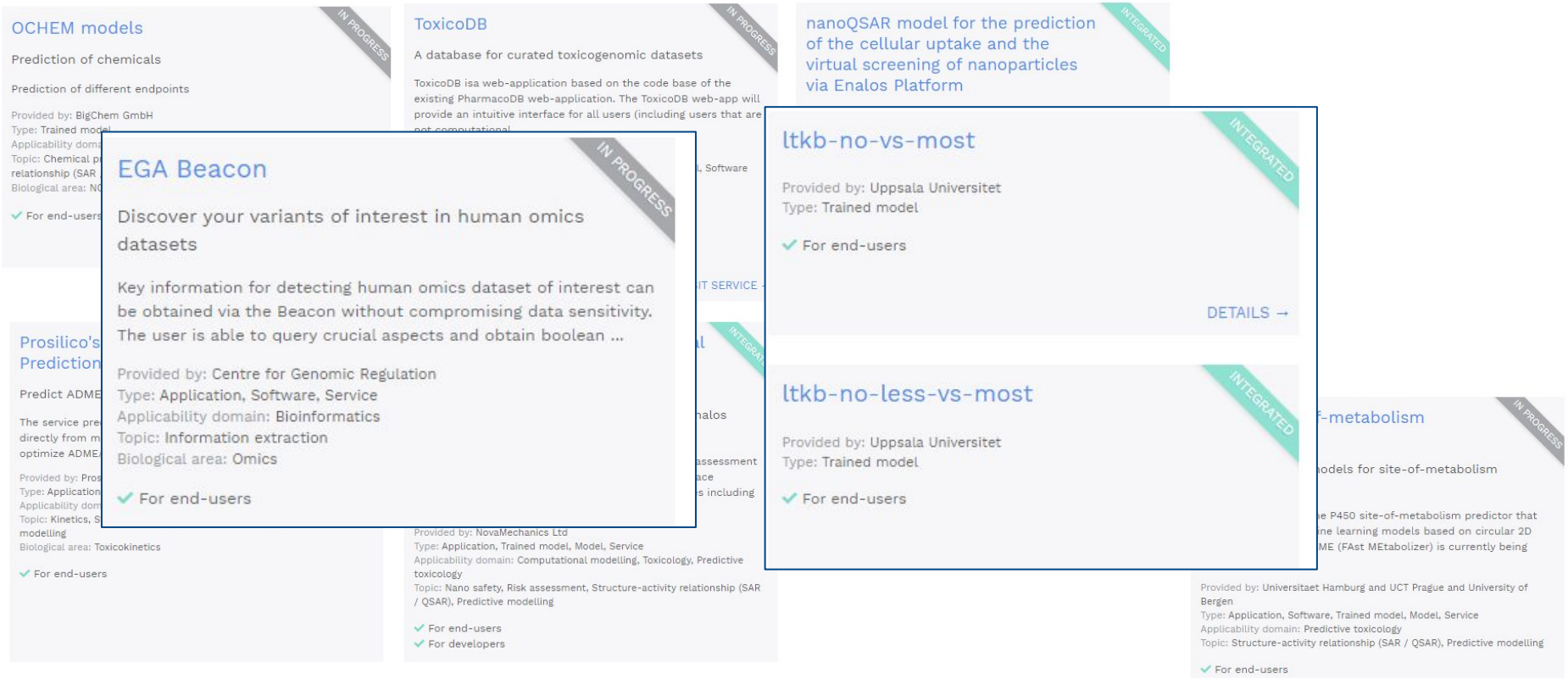
- Signing official associated partner agreement for complex especially between projects since these are not legal entities
- Standard contract most often not acceptable by associated partners even if only single sentences are changed
- After long consultations with advisors (Programme and Financial Officers, auditors), secondment agreements are the best (only) way
- Secondment agreements are not standard in many organizations

- To streamline the interactions with winners, the complete in-kind budget is now managed by Edelweiss Connect (transfer from Maastricht University and University of Birmingham)
- Advantages: all knowledge is at one partner, direct transfer of money, harmonization of contracts, monitoring and reporting (if you see other advantages, please tell me since we have to argue this in the final report)
- **Outcome:** 4 signed contracts, 5 contracts signed by Edelweiss Connect waiting for co-signing of partners, 1 ongoing negotiation

Implementation Challenge - Agreements status

Challenge Session	Organisation	Contact Person	Contact in OpenRiskNet	Contract with	Contract signed	Budget	Service	Description in Service Catalogue
31 October 2018	National Center for Computational Toxicology, US Environmental Protection Agency	Holly Mortensen	UM	EwC		€7,345	The Adverse Outcome Pathway Database (AOP-DB)	
	Korea Institute of Toxicology	Hyun Kil Shin	EwC	EwC	✓	€9,500	Daphnia magna nanotoxicity database	
							Nano-QSAR to predict cytotoxicity of metal and metal oxide nanoparticles	
	ToxPlanet	Matthias Timberlake	UM/EwC	EwC	✓	€12,250	ToxPlanet database	
	BIGCHEM GmbH	Igor Tetko	JGU	EwC			OCHEM models	✓
							OCHEM descriptors	X
							OCHEM model development tool	X
15 July 2019	Hamburger Informatik Technologie-Center e.V. (HITeC)	Johannes Kirchmair	VU	EwC	✓	€11,000	FAME 2 site-of-metabolism predictor	✓
31 January 2019	Prosilico	Urban Fagerholm	UU	EwC	✓	€9,875	Prosilico Human Clinical ADME/PK-Studio	✓
	NovaMechanics Ltd	Antreas Afantitis	NTUA	EwC		€15,438	Nanoinformatics Tool for the Virtual Screening of Metal Oxide Nanoparticles via Enalos Platform	✓
							nanoQSAR model for the prediction of the cellular uptake and the virtual screening of nanoparticles via Enalos Platform	✓
	Diamond Light Source Ltd. / Informatics Matters Ltd.	Rachael Skyner	IM	EwC			BruteReg (with Squonk implementation)	X
		Frank von Delft					PySquonk	
30 April 2019	Princess Margaret Bioinformatics and Computational Genomics Laboratory, University Health Network, Toronto, Canada	Benjamin Haibe-Kains	UM	EwC		€12,000	ToxicoDB	✓
	Leiden University	Katy Wolstencroft	UM	EwC		€7,332	ToxTargetLinks	
					TOTAL	**€84,740**	**(Available budget: €134,000)**	

Partner services



T1.3: Case Studies

Deliverable D1.3; D1.5
Milestone 3

SEURAT-1

Towards the Replacement of in vivo Repeated Dose Systemic Toxicity Testing

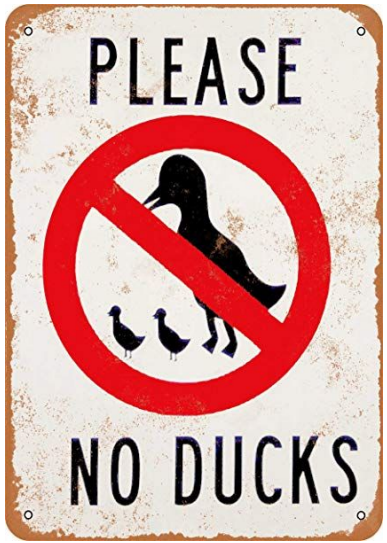
SPONSORED DOCUMENT FROM
COMPUTATIONAL TOXICOLOGY ELSEVIER FREE Full-Text Article

Comput Toxicol. 2017 Nov; 4: 31–44.
Published online 2017 Nov. doi: 10.1016/j.comtox.2017.10.001

PMCID: PMC5695905
PMID: 29214231

Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods

Elisabet Berggren,[a,*] Andrew White,[b] Gladys Ouedraogo,[c] Alicia Paini,[a] Andrea-Nicole Richarz,[a] Frederic Y. Bois,[d] Thomas Exner,[e] Sofia Leite,[f] Leo A. van Grunsven,[f] Andrew Worth,[a] and Catherine Mahony[g]



"From the beginning"

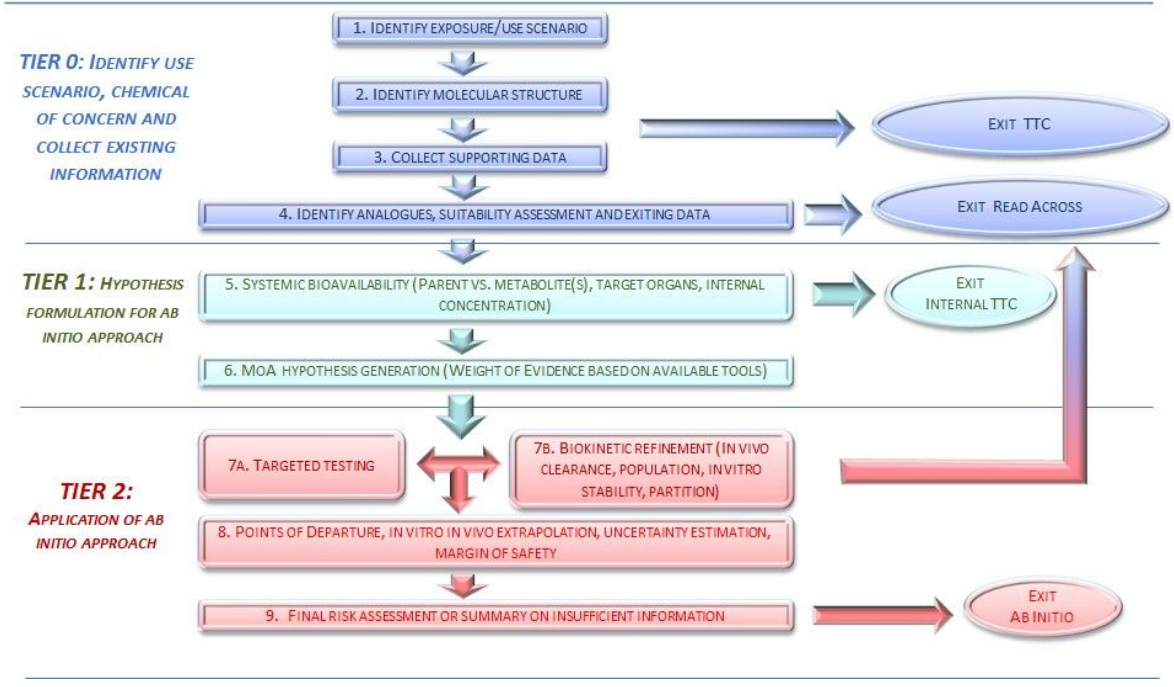
SEURAT-1

Towards the Replacement of in vivo Repeated Dose Systemic Toxicity Testing

Comput Toxicol. 2017 Nov; 4: 31–44.
Published online 2017 Nov. doi: 10.1016/j.comtox.2017.10.001
PMCID: PMC5695905
PMID: 29214231

Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods

Elisabet Berggren,[a,*] Andrew White,[b] Gladys Ouedraogo,[c] Alicia Paini,[a] Andrea-Nicole Richarz,[a] Frederic Y. Bois,[d] Thomas Exner,[e] Sofia Leite,[f] Leo A. van Grunsven,[f] Andrew Worth,[a] and Catherine Mahony[g]



CONCLUSION: This general "ab initio" workflow was developed as a means of structuring knowledge and data in a **logical sequence** for an integrated safety assessment applying non animal methods. **Workflow could be the basis for a full risk assessment and is aiming to provide a tool to guide the evaluation through the different steps to be considered and enable and gain confidence in decision making.**

The workflow is general enough to cover different types of chemicals, endpoints and exposure scenarios.

www.openrisknet.org

Case Studies: What is a case study ?

A particular instance of something used or analysed in order to illustrate a thesis or principle.

Is meant as:

- an in depth study of a particular situation
- method used to narrow down a very broad field of research into a researchable topic
- a guide to allow further elaboration and hypothesis creation on a subject
- an exercise to facilitate different disciplines to combine forces

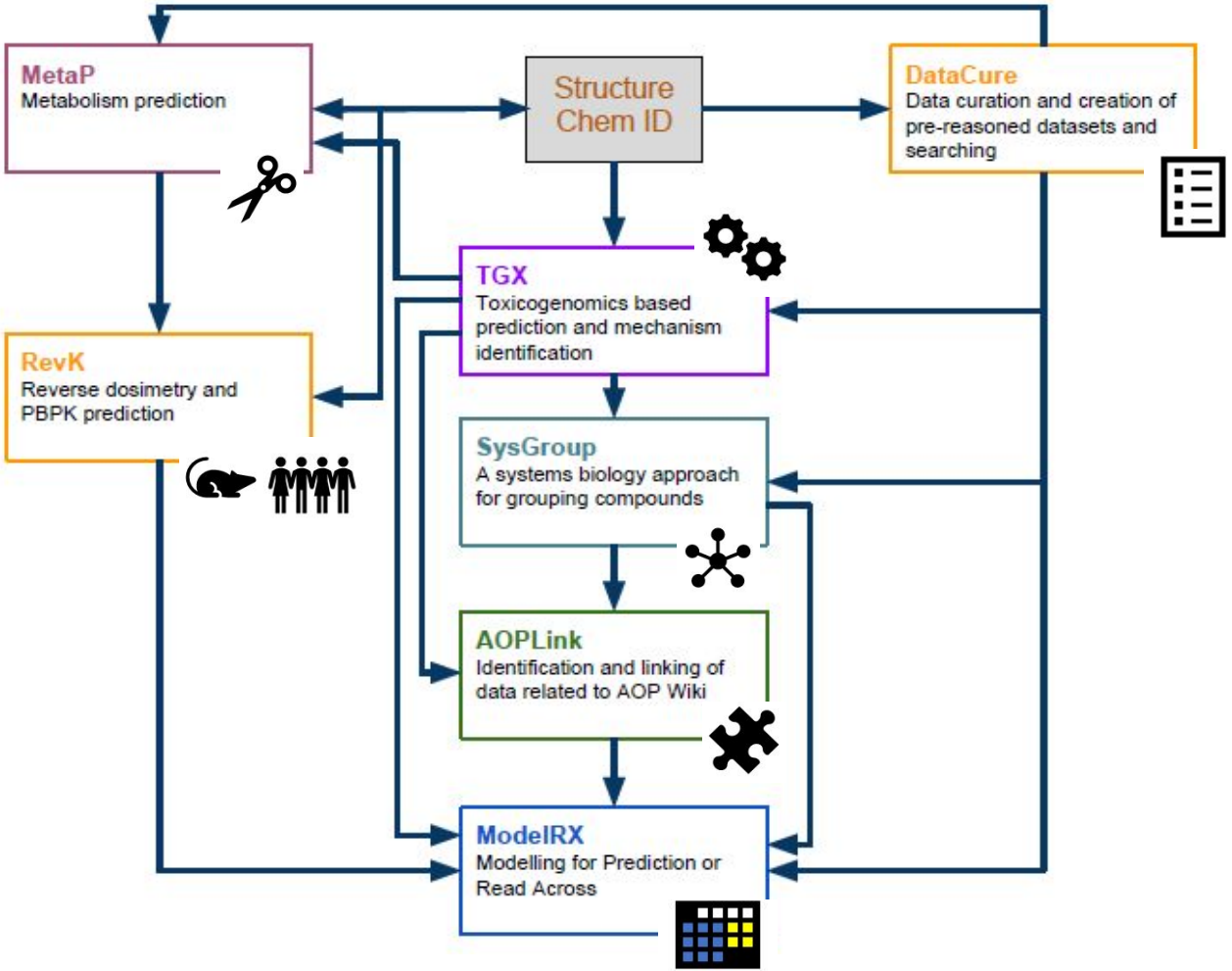
Is not meant as

- a sweeping statistical survey
- a complete answer to a particular question completely

Task 1.3 Case studies

D1.5	Finalization of case studies and analysis of remaining weaknesses	36	30 November 2019	VU	Other	Public

T1.3 Case Studies



[Deliverable 1.5]

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

CS leader: Noffisat Oki (EwC), **Involved:** EwC, IM, NTUA, Fraunhofer, UoB

Risk Assessment Framework
Tier 0.2-0.4 (data collection), 1.5 (biokinetics)

Outcome:

- Workflows to access data from different source
- Data curation workflow
- Semantic annotation of example datasets
- Text mining workflow

To do:

- Interaction with ToxPlanet
- ...

Databases

- Cheminformatic sources (PubChem)
- diXa (UM)
- ToxCast/Tox21 (EwC)
- FDA EADB (EwC, NTUA)

Tools / APIs

- ID converter service (EwC)
- Physchem, toxicological and omics databases: RDKit, CDK (NTUA), RDKit (IM), Data Explorer (EwC)
- Ontology/terminology/annotation: SCAIView / JProMiner / BELIEF (Fraunhofer), openBEL (Fraunhofer)

Metabolism Prediction [MetaP]

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

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

Outcome:

- Successful Integration of tools to predict (sites of) metabolism (also: third-party tools, implementation challenge)
- Diverse tools: ligand based (machine learning, QM based), protein-structure based (docking)
- Complementary tools: CYP450 metabolism, phase I, phase II, combined phase I/II
- Jupyter Notebooks to collect and visualize results from different services

To do:

- Update documentation, D1.5
- Output MetaP as input for other Case Studies?

Databases For the purpose of method development and model calibration and/or verification, data from XMetDB and other open-access databases for drugs, xenobiotics and their respective metabolites can be used

XMetDB, SMARTCyp, ZINC, ChEMBL, EAWAG-BBD

Tools / APIs

- MetPred (UU)
- FAME (UHH/HITeC; implementation challenge)
- enviPath, UM-PPS (JGU)
- SMARTCyp (external service, integrated by VU)
- Plasticity tools (VU)

Service integration

To facilitate combining metabolite prediction approaches or using MetaP outcomes as input for other predictors, we take advantage of available workflow management systems and we have explored integration of Jupyter Notebooks into the platform. This aids in investigating the added value of having multiple predictors available, which has been subject of a pilot study on metabolite 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)

First top-down case study

Outcome:

- 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.

To do:

- Containerize workflow
- Interaction with ToxPlanet & ToxicoDB (UHN)

Second top-down case study

Outcome:

- In vitro liver transcriptomics data from multiple human, rat & mouse cell models is available via Gitlab

To do:

- Finalize prediction analyses
- Containerize workflow
- Interaction with ToxicoDB (UHN)

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
- Second top-down case study on meta-analysis for in vivo genotoxicity prediction in human, rat and mouse in vitro cell models

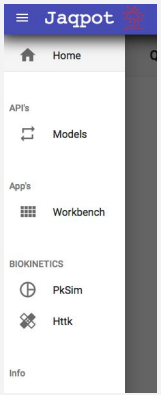
https://openrisknet.org/e-infrastructure/development/case-studies/case-study-tgx/

Reverse dosimetry and PBPK prediction [RevK]

CS leader: Frederic Bois (INERIS) **Involved:** NTUA

Outcome:
PBPK modelling is offered through the integration of 2 modelling environments:

- ***PKSIM*** (over API): open source from Bayer Pharmaceuticals, fixed model type
- ***httk*** (over API and Jaqpot 5 UI): R package from US EPA, more capabilities and greater modelling freedom
- ***JaqpotforR***: deploy PBPK model directly from R



To do:

- Add more simulation options in ***JaqpotforR***
- Efforts by to integrate a second httk model from INERIS using ***JaqpotforR***

Risk Assessment Framework
Tier 1.5 (biokinetics)

Databases Open systems Pharmacology, httk

Tools / APIs

- PKSim through Jaqpot (NTUA) https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/#/biokinetics
- httk through Jaqpot (NTUA) https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger/#/biokinetics
- R httk client (in progress, NTUA, INERIS)

Service integration

Two PK (pharmacokinetics) modelling environments have been integrated in OpenRiskNet e-infrastructure through the NTUA *Jaqpot* modelling platform: The extensive, industry strength, *PKSim* PBPK open-source software, and the US EPA developed R package *httk*. With *httk* the user can request concentration-time profiles at any dose and for any length of time following exposure to about 500 chemicals, in rats, rabbits, dogs, mice or humans. A client is being developed by NTUA and INERIS to present the user with an interactive interface automatizing the requests and presentation of the results.

https://openrisknet.org/development/case-studies/case-study-revk/

Identification and Linking of Data related to AOP-Wiki [AOPLink]

CS leader: Marvin Martens, Egon WIllighagen, Chris Evelo (UM)
Involved: EwC, UoB, CRG

Outcome:

- Discoverable annotated BridgeDb API
- Development of the AOPLink RDF (AOP-Wiki + WikiPathways + AOP-DB), and loaded and exposed as Virtuoso SPARQL endpoints
- Implementation challenge service: AOP-DB RDF
- Workflows utilizing the AOPLink RDF, linking knowledge repositories and experimental data to AOPs.

To do:

- Improved linking of AOPs with WikiPathways via KE genes.
- Development of pathway analysis Jupyter notebook with gene expression data related to AOP stressor chemicals.
- Knowledge base linking nanomaterials to MIEs

Risk Assessment Framework
Tier 0.3, 0.4 (Support Data), 1.6 (MOA)

Databases

- AOP-Wiki, AOP-DB: AOP knowledgebase (AOP-KB);
- WikiPathways, Reactome: biological pathway database;
- eNanoMapper, EPA Chemistry Dashboard, NORMAN Network: experimental data.

Tools / APIs

- BridgeDb, ChemIdConverter: identifier mapping;
- PathVisioRPC: pathway analysis;
- eNanoMapper database test instance

Results

- BridgeDb service, AOP-Wiki and AOP-DB SPARQL endpoints, operational in VRE
- Report on AOPWiki<>WikiPathways linking options

Activities

- Continued development identifier mapping databases
- Semantification of AOPWiki
- AOP Portal (http://aop.wikipathways.org)
- Exploration of APIs around semantic web technologies

https://openrisknet.org/development/case-studies/case-study-aoplink/

A systems biology approach for grouping compounds [SysGroup]

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

Risk Assessment Framework
Tier 0.2-0.4 (data collection, read across)

Outcome:

- Flowchart of required steps finalized

To do:

- Develop workflow
- Containerize workflow
- Interaction with ToxPlanet & ToxicoDB (UHN)

Databases

- diXa /BioStudies (UM)
- PubChem
- ChEMBL

Tools / APIs

- PubChem
- ChEMBL or PIDGIN
- (pre)processing tools for gene expression data (e.g. microarray data) → e.g. arrayanalysis.org (UM)
- iClusterPlus

Required steps

- Chemical similarity calculated by 2D or 3D Tanimoto coefficient (PubChem)
- Protein target prediction (ChEMBL/PIDGIN)
- Interface to diXa for obtaining gene expression data
- Integration of the multiple data sources and grouping by iClusterPlus

Service integration

- Integration with other case studies is needed. SysGroup acquires information from the DataCure case study and can feed into AOPLink and ModelRX.

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

Modelling for Prediction or Read Across [ModelRX]

Risk Assessment Framework
Tier 0.4 (read-across), 1.5, 1.6

CS leader: Harry Sarimveis (NTUA) **Involved:** JGU, UU, EwC

Outcome:

- Given a dataset with a defined end-point to be predicted, develop a **predictive model**, validate it using OECD principles and integrate it in risk assessment workflows for **filling gaps** in available experimental information.
- A training data set can be obtained from an ORN data source. Predictive models are trained with ORN modelling tools and are deployed as web services, documented and ontologically annotated. Model validation using OECD guidelines. Finally, models can be used to get predictions.
- Consensus modelling for all ORN-generated predictions have been set up in Jupyter notebooks.

To do:

- Finalise Jupyter notebooks library https://github.com/OpenRiskNet/notebooks/tree/master/ModelRX

Databases ChEMBL-eNanoMapper-Gene Ontology (GO) - Any ORN data providing service

Tools / APIs

- Jaqpot (NTUA):(API) https://api-jaqpot.prod.openrisknet.org/jaqpot/swagger, (GUI) https://ui-jaqpot.prod.openrisknet.org
- Lazar (JGU):(API/GUI) https://lazar.prod.openrisknet.org/
- WEKA Rest Service (JGU): (API) https://jguweka.prod.openrisknet.org
- Jupyter Notebooks (IM) https://jupyterhub-jupyter.prod.openrisknet.org/hub/logi
- Squonk Computational Notebook (IM) https://squonk-notebook.prod.openrisknet.org/portal
- CPlogd(UU) https://cplogd.prod.openrisknet.org/draw/ https://cplogd.prod.openrisknet.org Metpred (UU) https://metpred.prod.openrisknet.org/ LTKB(UU) http://ltkb-no-vs-most-cpsign.prod.openrisknet.org/, http://ltkb-no-vs-less-most-cpsign.prod.openrisknet.org, http://ltkb-no-less-vs-most-cpsign.prod.openrisknet.org, BBB (UU) http://blood-brain-barrier-penetration-cvap-cpsign.prod.openrisknet.org, http://blood-brain-barrier-penetration-cpsign.prod.openrisknet.org,

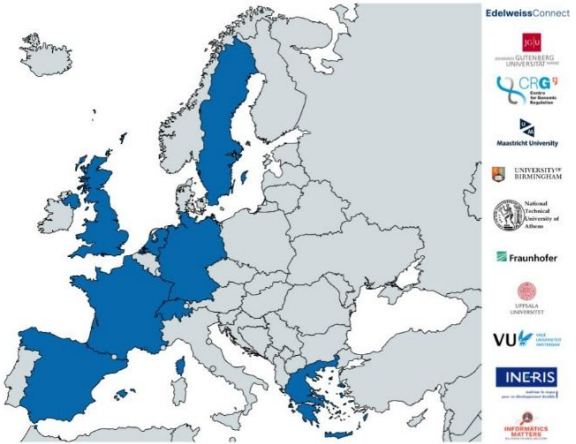
Service integration: **Jaqpot** has integrated WEKA, Scikit-Learn and R Caret as well as modelling and analysis services which include, scaling/normalisation, variable selection, model validation, algorithms for defining the DoA and standard reporting formats (QMRF). **Lazar** is a modular framework for predictive toxicology. Similar to the *read across* procedure in toxicological risk assessment, lazar creates local QSAR models for each compound to be predicted. The **WEKA** modelling web-services by JGU provide an OpenRiskNet compliant REST interface to the machine learning algorithms from the WEKA Java Library. IM made available **Jupyter notebooks** infrastructure and the **Squonk Computational notebook**. UU provides APIs for **CPlogD**, **metpred** and **LTKB** predictions. All these services can be accessed through Jupyter notebooks.

Deliverable 1.5

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)