The Use of Pre-Clinical Data Standardisation

In Silico Toxicology Network Meeting, 30th September, 2020

Who Are We?

- Consortium members of eTRANSAFE (IMI2) - the follow-on project from eTOX (IMI1)
- Specialist software development and consulting business active in regulatory toxicology since 1979
- Developers of nonclinical data collection systems and data preservation (data extraction) systems for long term electronic access (archiving) and research
- Developers of clinical coding browsing and auto-coding products (against SNOMED-CT, ICD-10, etc., code-sets).
- Working with data standards since 2012
- Member of PhUSE – headed-up Histopathology Visualization project (Nonclinical Topics Work Group)

What is SEND?

- What is the Standard for the Exchange of Nonclinical Data?
  - SEND is an implementation of the CDISC Study Data Tabulation Model (SDTM) for nonclinical toxicology studies
    - SEND is developed and maintained by the CDISC SEND team
  - Nonclinical studies refer to nonhuman studies that are conducted during drug development to address safety issues.
    - e.g. Tox studies to open clinical trials in humans
    - e.g. Carcinogenicity studies to support product labelling
  - Generally, these studies are reviewed by Pharm/Tox reviewers in CDER
- What does SEND do?
  - Provides a standardized presentation of toxicology study data in a electronic format.
  - Enables the development and use of visualization and analytical tools for these types of data.
  - Enables more effective and efficient review of nonclinical tox data.

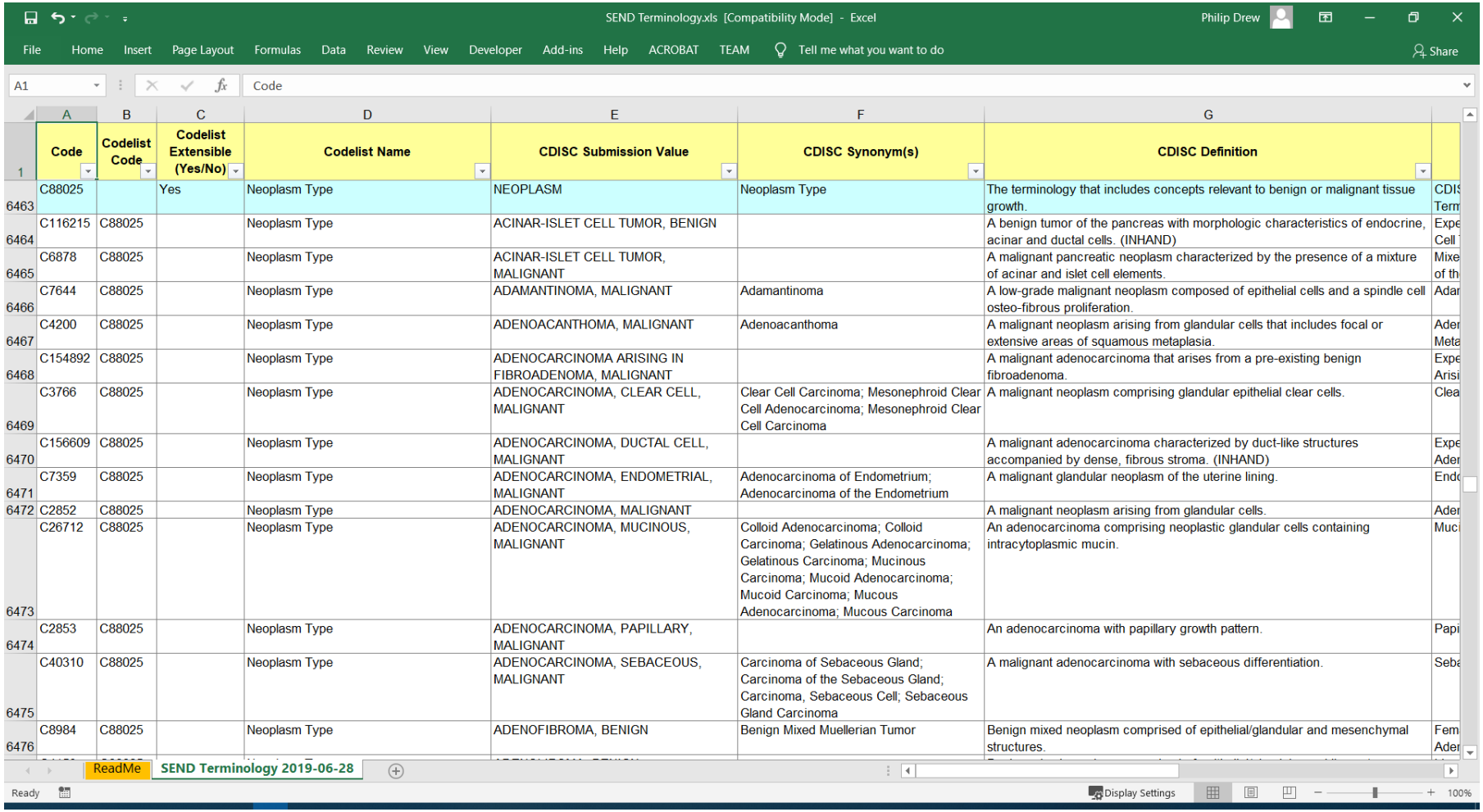
Features of SEND

- SEND compartmentalises data into *Domains* (e.g. Bodyweight, Food & Water, Clinical Signs, Gross Pathology, etc.)
- SEND uses a Controlled Terminology (SEND CT) for the harmonisation of species, specimens, laboratory parameters, measurement units, pathological terms, etc.
- SEND preserves original data (findings) ‘as is’
- SEND does not change data, nor does it impose new study requirements.
- SEND will not replace summary, interpretive, or other information in study reports. Only data tabulations.

Why Standards?

- Benefits: Aligned with CDER's goal of rapid acquisition, analysis, storage and reporting of regulatory data
  - Improve efficiency
    - Highly educated and experienced people are spending their time manually transcribing numbers into spreadsheets.
  - Improve review science
    - Pharmacologists and toxicologists can determine the nonclinical parameters that best predict adverse events in humans
  - Improve quality of reviews:
    - Improve information in written review to demonstrate basis for decisions
- SEND is now a CDER preferred, supported standard
  - CDER has processes and technology infrastructure for the receipt, processing, review, and archive of study data using SDTM/SEND

CDISC Controlled Terminology (CT)



	Code	Codelist Code	Codelist Extensible (Yes/No)	Codelist Name	CDISC Submission Value	CDISC Synonym(s)	CDISC Definition	
6463	C88025		Yes	Neoplasm Type	NEOPLASM	Neoplasm Type	The terminology that includes concepts relevant to benign or malignant tissue growth.	CDIS Term
6464	C116215	C88025		Neoplasm Type	ACINAR-ISLET CELL TUMOR, BENIGN		A benign tumor of the pancreas with morphologic characteristics of endocrine, acinar and ductal cells. (INHAND)	Expe Cell
6465	C6878	C88025		Neoplasm Type	ACINAR-ISLET CELL TUMOR, MALIGNANT		A malignant pancreatic neoplasm characterized by the presence of a mixture of acinar and islet cell elements.	Mixe of th
6466	C7644	C88025		Neoplasm Type	ADAMANTINOMA, MALIGNANT	Adamantinoma	A low-grade malignant neoplasm composed of epithelial cells and a spindle cell osteo-fibrous proliferation.	Adar
6467	C4200	C88025		Neoplasm Type	ADENOACANTHOMA, MALIGNANT	Adenoacanthoma	A malignant neoplasm arising from glandular cells that includes focal or extensive areas of squamous metaplasia.	Ader Meta
6468	C154892	C88025		Neoplasm Type	ADENOCARCINOMA ARISING IN FIBROADENOMA, MALIGNANT		A malignant adenocarcinoma that arises from a pre-existing benign fibroadenoma.	Expe Arisi
6469	C3766	C88025		Neoplasm Type	ADENOCARCINOMA, CLEAR CELL, MALIGNANT	Clear Cell Carcinoma; Mesonephroid Clear Cell Adenocarcinoma; Mesonephroid Clear Cell Carcinoma	A malignant neoplasm comprising glandular epithelial clear cells.	Clea
6470	C156609	C88025		Neoplasm Type	ADENOCARCINOMA, DUCTAL CELL, MALIGNANT		A malignant adenocarcinoma characterized by duct-like structures accompanied by dense, fibrous stroma. (INHAND)	Expe Ader
6471	C7359	C88025		Neoplasm Type	ADENOCARCINOMA, ENDOMETRIAL, MALIGNANT	Adenocarcinoma of Endometrium; Adenocarcinoma of the Endometrium	A malignant glandular neoplasm of the uterine lining.	Endo
6472	C2852	C88025		Neoplasm Type	ADENOCARCINOMA, MALIGNANT		A malignant neoplasm arising from glandular cells.	Ader
6473	C26712	C88025		Neoplasm Type	ADENOCARCINOMA, MUCINOUS, MALIGNANT	Colloid Adenocarcinoma; Colloid Carcinoma; Gelatinous Adenocarcinoma; Gelatinous Carcinoma; Mucinous Carcinoma; Mucoid Adenocarcinoma; Mucoid Carcinoma; Mucous Adenocarcinoma; Mucous Carcinoma	An adenocarcinoma comprising neoplastic glandular cells containing intracytoplasmic mucin.	Muci
6474	C2853	C88025		Neoplasm Type	ADENOCARCINOMA, PAPILLARY, MALIGNANT		An adenocarcinoma with papillary growth pattern.	Papi
6475	C40310	C88025		Neoplasm Type	ADENOCARCINOMA, SEBACEOUS, MALIGNANT	Carcinoma of Sebaceous Gland; Carcinoma of the Sebaceous Gland; Carcinoma, Sebaceous Cell; Sebaceous Gland Carcinoma	A malignant adenocarcinoma with sebaceous differentiation.	Seba
6476	C8984	C88025		Neoplasm Type	ADENOFIBROMA, BENIGN	Benign Mixed Muellerian Tumor	Benign mixed neoplasm comprised of epithelial/glandular and mesenchymal structures.	Fem Ader

Terminological Diversity Example

- BLDGLU
- Blood / serum glucose (Gluc) — Matrix info
- Blood Glucose — Matrix info
- Blood sugar — Matrix info
- Blood sugar (SUGAR) — Matrix info
- Glucose
- Fasted blood glucose concentration — Protocol info
- Fasting Glucose — Protocol info
- Fasting blood glucose (FBS) — Protocol info
- Fasting blood sugar (GLUCOSE) — Protocol info
- GLU
- GLU: Glucose
- GLUC
- GLUC.
- GLUC:G
- GLUCOSE
- Glicose — OCR error
- Glucose (GLUC)
- Glucose % of Control — Relative
- Glucose (GL) — Company acronyms
- Glucose (GLS) — Company acronyms
- Glucose (GLU) — Company acronyms
- Glucose (GLUC) — Company acronyms
- Glucose (GLUCOSE) — Company acronyms
- Glucose (Gluc.) — Company acronyms
- Glucose (NGLU)
- Glucose (UGLS)
- Glucose (UGLU)
- Glucose (sq)
- Glucose GLU
- Glucose concentration (Gluc)
- Glucose mg%
- Glucose per sample (USGL)
- Glucose, dipstick — Detection method
- Glucose, pool 1
- Glucose, pool 2
- Glucose/sample (USGL)
- Glucose:volume
- Maximal glucose — Distinct measurement types
- Mean glucose — Distinct measurement types
- Minimal glucose — Distinct measurement types
- P-GLUC
- Plasma glucose
- Plasma glucose (GLU)
- Plasma glucose (PGLU)
- S-GLUC
- Serum glucose
- Serum glucose
- Serum glucose (GLUC)
- Sugar — "Common" name
- Total glucose
- U-glucose (U-Gls)
- U-glucose, pool 1 (U-Gls)
- U-glucose, pool 2 (U-Gls)
- Urinary glucose (UGLU)
- Urinary glucose/volume (UGLC)
- Urine Glucose
- Urine glucose (Gls)
- Urine glucose (Glu-U)
- Urine glucose (UGLU)
- Urine glucose (UGLU) 6h — Distinct timing
- Urine glucose: volume

Why didn't we collect data using standard code lists?

⌘ **Code lists for most toxicology data domains are recent or not yet available**

- CDISC-SEND Controlled Terminologies (CTs) are helping standardize code lists
- Each company uses its own in-house standards/glossaries

⌘ **Complexity of the data is enormous**

- Minimum set of data collected for a study are: the protocol, in life observations, food/water consumption, **clinical pathology**, toxicokinetics, body weight, cause of death, organ weight, **anatomic pathology (micro/macro)**
- Other domains can also be present: ECGs, biomarkers, hematology, hemostasis, reprotox, palpable masses, genetics, genomics, ADME, etc.

⌘ **Gold standard of toxicology is anatomic pathology**

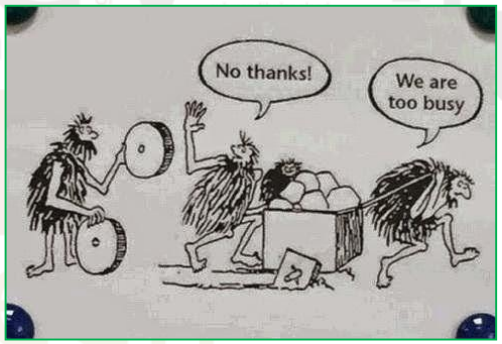
- No comprehensive standards exist (yet) for these data; it will be accelerated by the need for electronic submissions (FDA/SEND).
- Long term effort from toxicology societies to assemble guidance and code lists is now paying off. Code list assembled by INHAND is a case in point.
- Challenge: no relation/hierarchy between terms available in INHAND (manual curation) or CT.

Goals of data harmonization

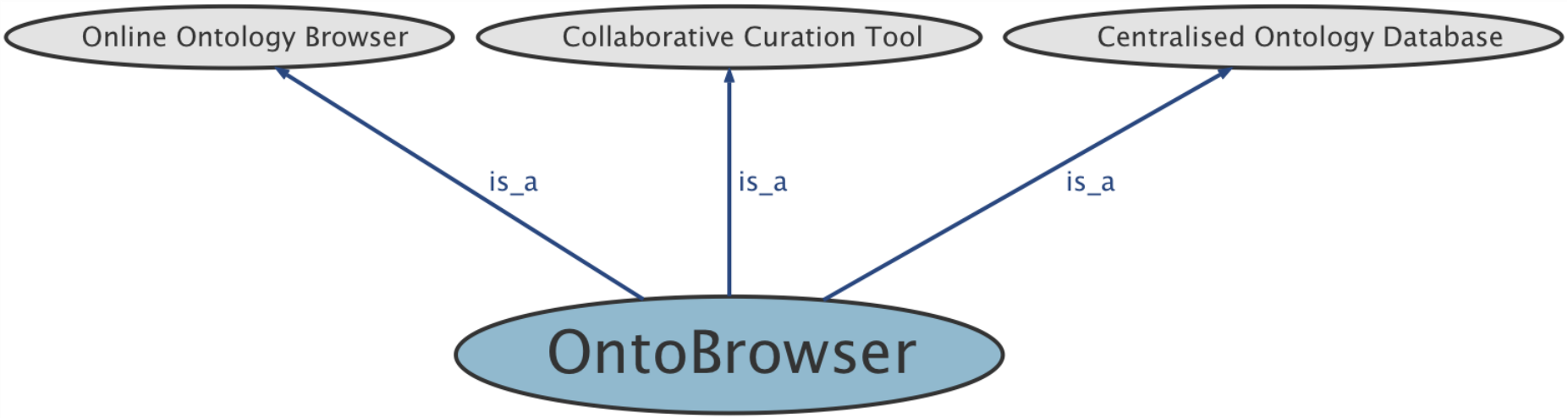
- Enable the work of modellers by reducing complexity of data
  - Code list/ontologies for descriptions
  - Harmonization of units/grades/etc.
- Get a single reference point for ontologies and code lists
  - List ontologies needed in eTOX/eTRANSAFE.
  - Re-use existing ontologies when suitable.
  - Complement/Edit reference ontologies when needed.
  - Create new ontologies when necessary.
- Provide a central point to browse and edit ontologies
  - Put in place teams and tools to create/edit/maintain ontologies.
  - Put in place relevant processes/governance regarding changes.
  - Expose concepts/terms to all eTOX/eTRANSAFE systems.

How did we move forward?

Verbatim terms curation is crucial. It is an expert job and, initially, no user friendly interface was available



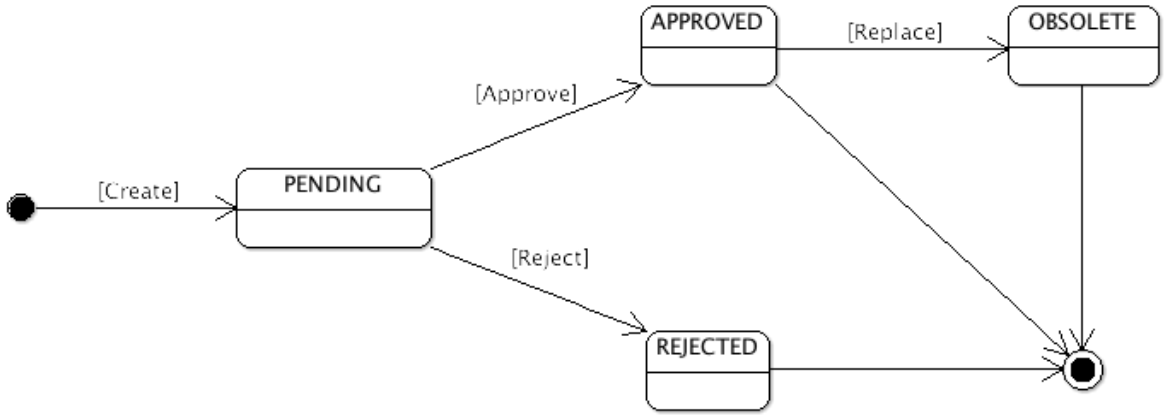
The Solution: OntoBrowser



OntoBrowser

Key Features

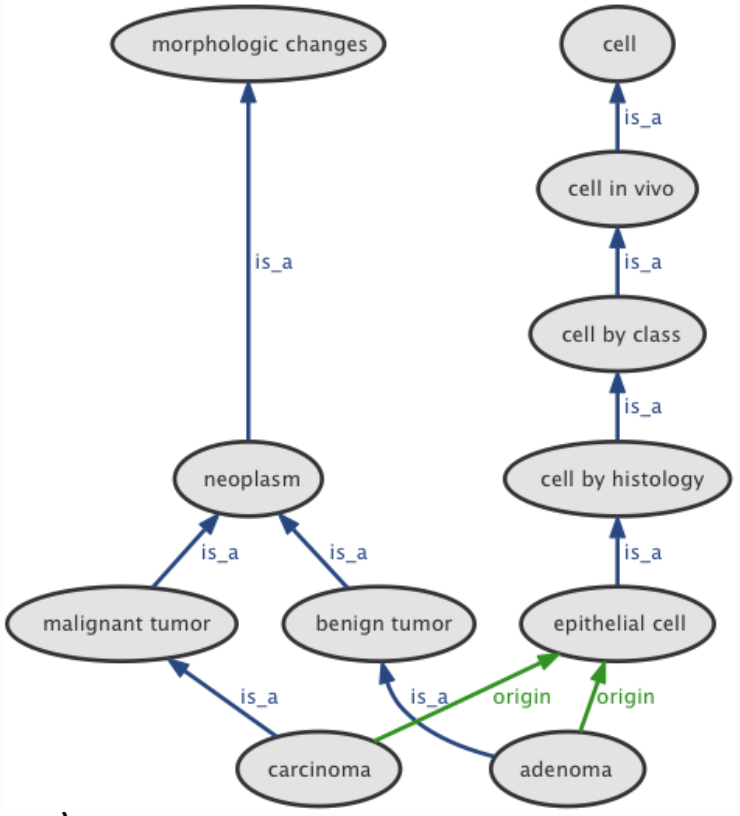
- Online collaborative ontology curation
- Browse-only mode for non-curators
- Interactive visualization
- Cross ontology searching
- Review/Approve workflow



OntoBrowser

Key Features (continued)

- Central database for all ontologies (model based on OBO file format)
- Direct access to terminology databases for vocabulary mapping
- Automated mapping of exact or similar matching synonyms
- Alerts for unmapped synonyms
- Versioning of ontologies
- Full curator history
- Cross ontology relationships:
  - Anatomy
  - Cell type
  - Histopathology
  - Protein ligand interaction
  - Toxicity events
- Open Source application (http://opensource.nibr.com)



What can be done:

https://www.ncbi.nlm.nih.gov/pubmed/27690270

Search…Retrieve…Model → 2016

↓ 2014

https://www.ncbi.nlm.nih.gov/pubmed/25405742

Int. J. Mol. Sci. **2014**, *15*(11), 21136-21154; doi:10.3390/ijms151121136

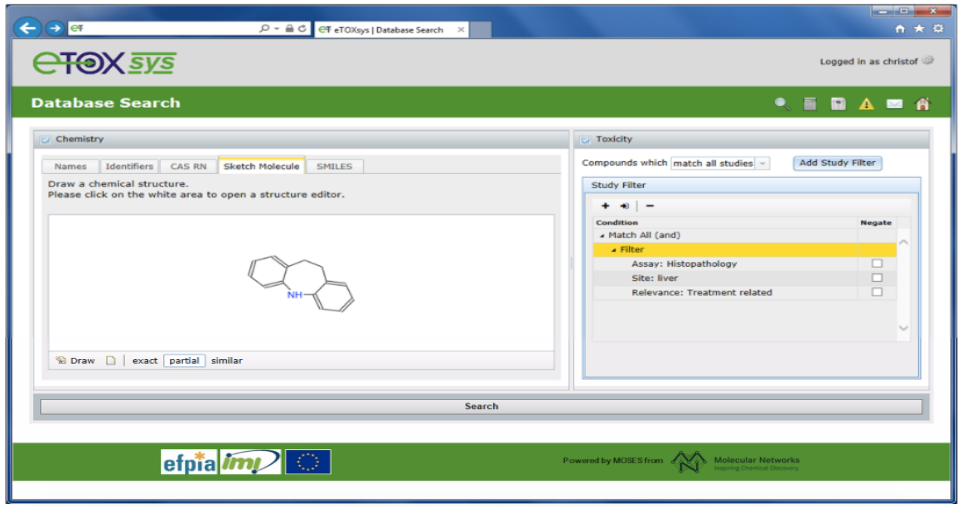
Open Access

Article

The eTOX Data-Sharing Project to Advance *in Silico* Drug-Induced Toxicity Prediction

Montserrat Cases [1,3], Katharine Briggs [2], Thomas Steger-Hartmann [3], François Pognan [4], Philippe Marc [4], Thomas Kleinöder [5], Christof H. Schwab [5], Manuel Pastor [1], Jörg Wichard [3] and Ferran Sanz [1,*]

Figure 4. Combined chemistry and toxicity database search.



Research Article

Hepatotoxicity prediction by systems biology modeling of disturbed metabolic pathways using gene expression data

Pablo Carbonell[1,2], Oriol Lopez[1], Alexander Amberg[3], Manuel Pastor[1] and Ferran Sanz[1]*

[1]Research Programme on Biomedical Informatics (GRIB), Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Dept. of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain; [2]Manchester Synthetic Biology Research Centre for Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, Manchester, UK; [3]Sanofi Aventis Deutschland GmbH, Preclinical Safety, Frankfurt am Main, Germany

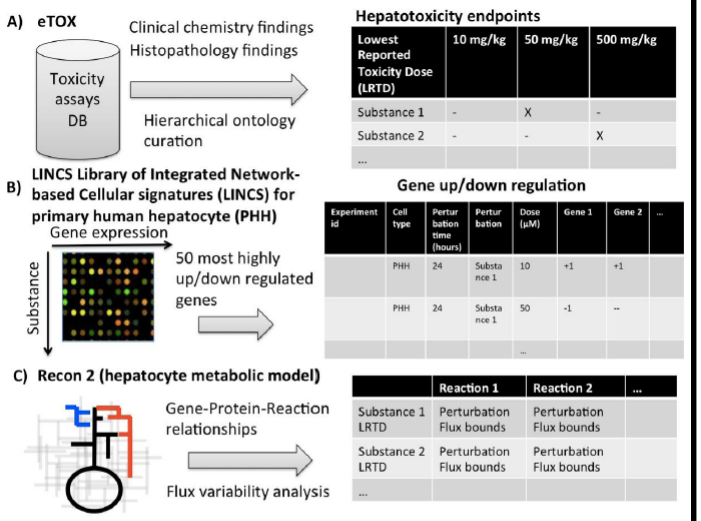


Fig. 1: General data extraction procedures developed in the study
a) Lowest reported toxic doses (LRTD) for hepatotoxicity endpoints were defined by toxicologist expert curation of the findings extracted from the clinical chemistry and histopathology tables of eTOX database and assigned LRTDs to one of the following categories: 10, 50, 100, and 500 mg/kg, depending if the LRTD value is lower than the category value; b) gene regulation for chemical substances in cells were compiled from the LINCS database for primary human hepatocytes. The contained information was perturbation time, dose and 50 top up/downregulated genes; c) a metabolic model for hepatocytes was used in order to determine upper/lower bound for reaction fluxes through flux variability analysis. Gene-protein-reaction associations in the model provided the information in order to link such information to gene expression data

Where can we go if we do it right?