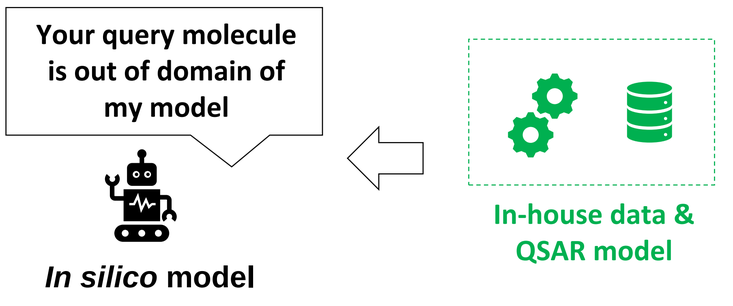
Using shared knowledge to improve predictions of adverse drug reactions and support decision making

J. Yung, A. Fowkes and I. Xanthis.

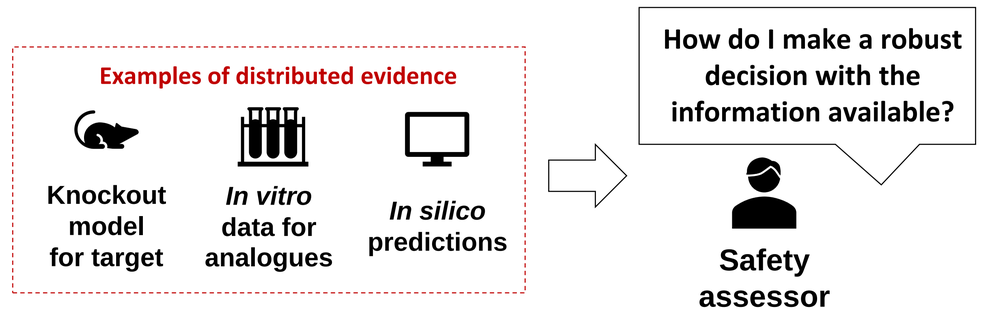
Lhasa Limited, 2 Canal Wharf, Leeds, LS11 5PS, United Kingdom

Current challenges facing toxicologists during safety assessments

In silico models haven't learnt from all high-quality data that is stored in proprietary data silos

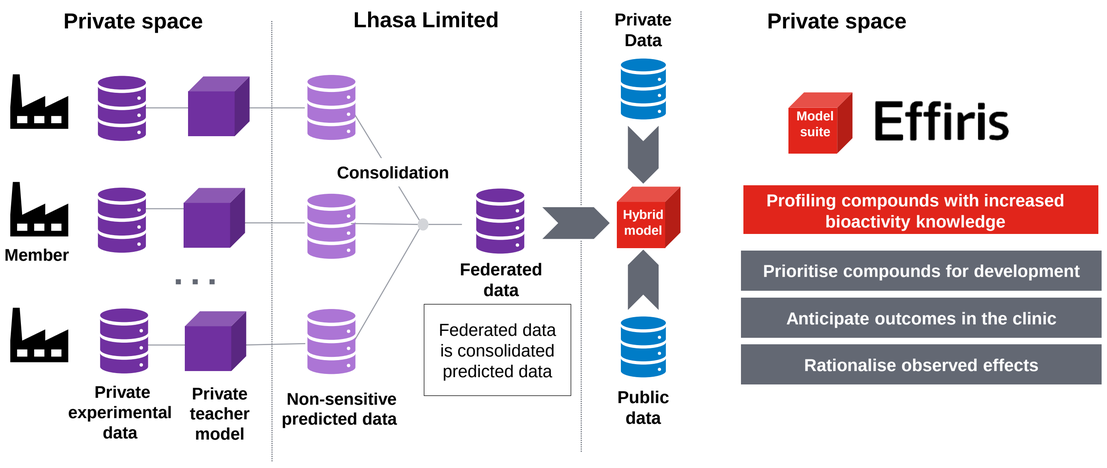


Supporting evidence for decision making is heterogenous, distributed and even discordant



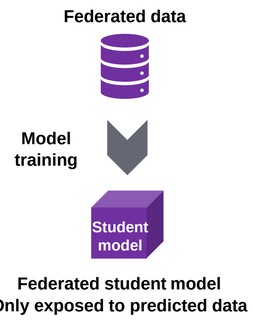
Solution: Federated data enables pre-competitive sharing of knowledge between organisations

Lhasa Limited has developed a federated learning platform called Effiris, which enables extraction of knowledge from multiple proprietary data silos without loss of confidential information. The approach labels a common set of public structures with endpoint predictions from teacher models trained on proprietary data. Acting as an honest broker, Lhasa Limited consolidates the predicted data from all the partners into a single robust dataset, whilst also performing quality checks. The consolidated data is then shared with members, who are then able to combine this new data with in-house and public data to generate hybrid models, which have learnt from multiple sources of knowledge.



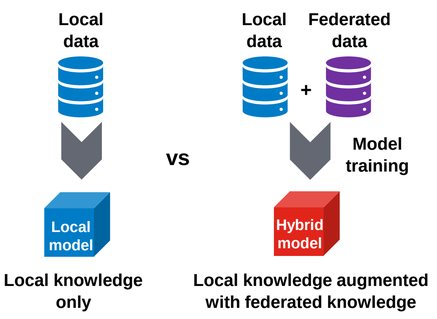
Shared bioactivity knowledge can be used to improve QSAR models

To examine if useful knowledge has been transferred successfully a range of experiments have been performed. By examining the knowledge in the federated data through testing of the student model, it was shown that for six out of eight priority endpoints the federated data contained knowledge as judged by exceeding a Matthews correlation coefficient (MCC) value of 0.2. The main factor that hampered knowledge transfer for the remaining two endpoints was a strong bias of the original private data across all the contributing members.



Endpoint	Federated student model (MCC)	Knowledge transferred? (MCC > 0.2)
COX-1	0.06	NO
COX-2	0.31	YES
GABA-A	0.19	NO
SERT	0.21	YES
DRD2	0.37	YES
CHRM1	0.42	YES
CHRM2	0.56	YES
hERG	0.36	YES

To examine if the shared knowledge for the six endpoints could be used to improve QSAR models, the federated data was combined with local public data to create hybrid classification models, which were then validated using a challenging cluster cross validation. The validation demonstrated that the approach resulted in improved QSAR models, which were able to make more correct predictions over a wider area of chemical space. These models will enable assessors to identify and act upon toxicity liabilities earlier in their workflows.

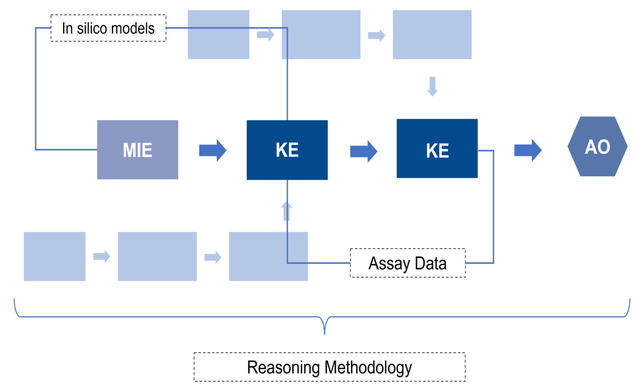


Average performance of models for the six endpoints where the federated data has demonstrated successful knowledge transfer

	Local QSARs	Hybrid QSARs	QSARs benefit from shared knowledge?
Performance (MCC)	0.50	0.54	YES
Coverage (%)	50	71	YES

Solution: AOPs provide a framework to organise evidence and support decision making

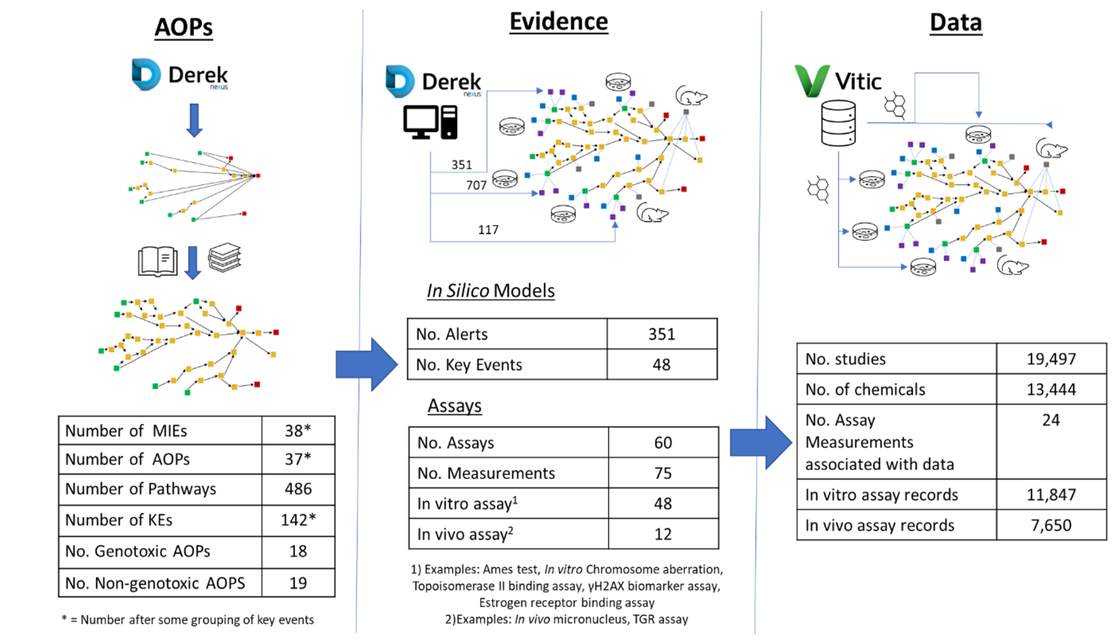
An AOP is a formalised approach to documenting a mechanism of toxicity. AOPs start with a molecular initiating event (MIE) and through additional key events (KEs), lead to an Adverse Outcome (AO). Each sequential KE (including the MIE and AO) are connected to each other through key event relationships (KERs). Formally, each KE should be measurable and therefore can be linked to a relevant assays and models.



Kaptis

- Scientifically-supported risk assessment decisions using AOPs
- Inform on what assay to run next
- Examine data for similar chemicals
- Identify adverse outcome liabilities

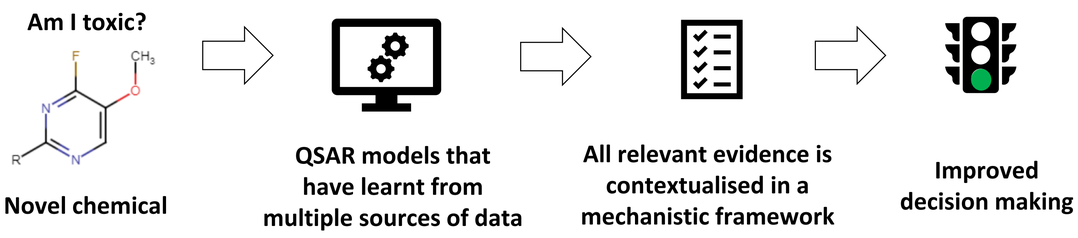
To populate an AOP framework for carcinogenicity with relevant evidence, knowledge held within the predictive system Derek Nexus was extracted, built upon, and arranged into a coherent network containing 37 AOPs. 60 Assays and 351 *in silico* alerts were then associated with KEs in this network and it was brought to life by associating data and contextualising evidence and predictions for over 13,400 compounds. This allows toxicity-relevant data to be queried from both the biological and chemical perspective and presented in a structured way to support decisions.



Conclusion

The development of *in silico* solutions which focus on sharing knowledge will enable safety assessors to better identify potential adverse drug reactions and make improved decisions through:

- Using models that have learnt from multiple sources of data, including proprietary ones
- Analysing all relevant evidence which is fully contextualised to support decisions



shared **knowledge** • shared **progress**