"Please, sir, I want some more…" data

Guidelines for effective sharing of preclinical data

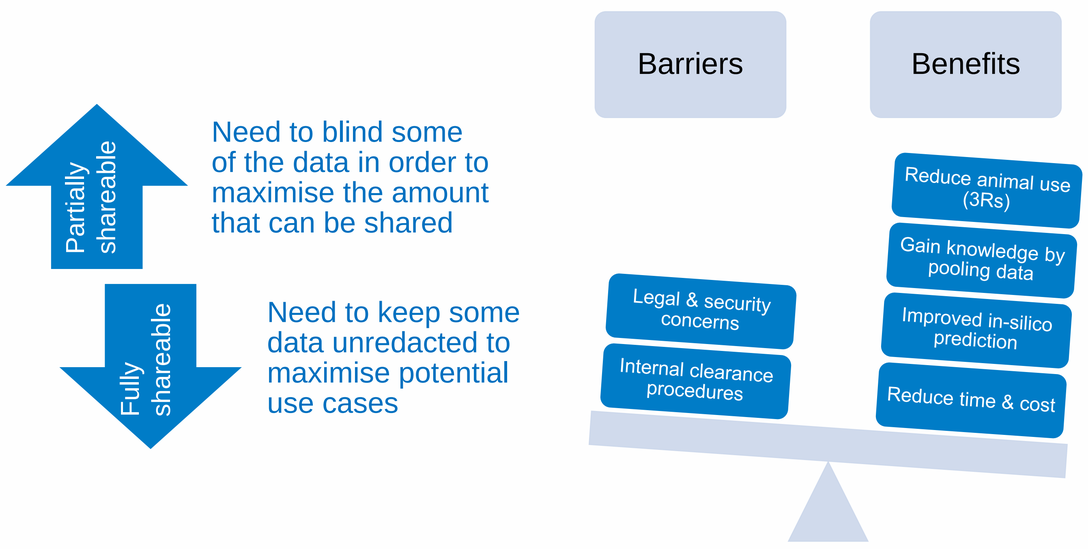
Katharine Briggs

Granary Wharf House, 2 Canal Wharf, Leeds, LS11 5PS

Why share data?

Pre-competitive data sharing can offer significant benefits in terms of reducing the time and costs involved in getting a new drug to market through more informed testing strategies and knowledge gained by pooling data. If sufficient data is shared and can be co-analysed, then it can also offer the potential for reduced animal usage and improvements in the in silico prediction of toxicological effects. However, our experience shows that obtaining internal clearance represents a significant hurdle to preclinical data being made more widely available. Organisations could address this by defining standard operating procedures for data sharing, simplifying the process and providing clear roles and responsibilities.

To investigate how barriers to data sharing might be overcome, a survey was conducted of pharmaceutical organisations participating in the Enhancing TRANslational SAFEty Assessment through Integrative Knowledge Management (eTRANSAFE) project (https://etransafe.eu/) [1][2].



When can it be shared?

Substantial agreement among data donors was obtained on the decision criteria for inclusion in internal clearance procedures to determine when data could be shared:

- Step 1 Whether data are already publicly available
- Step 2 Whether data are still of strategic interest
- Step 3 Whether there is a potential IP or legal conflict
- Step 4 Option to change the classification (to allow data to be shared at a later date)

Some partners were able to set up standard operating procedures on this basis enabling them to respond more effectively to data sharing requests. If these commonly agreed decision criteria could be codified into official guidelines, they could promote data sharing to the benefit of the scientific community.

What can be shared?

Substantial agreement among data donors was obtained on criteria for identifying data considered too sensitive to share for substances still of strategic interest:

- Chemical structure, chemical code, internal compound code, name or reference
- Pharmacological target, indication(s) and off-target in vitro panel
- Company name or identifier

Redaction of this data could allow more data to be shared and prompted the decision to use a tiered data classification scheme involving, unredacted, partially redacted and fully redacted data. Visibility of redacted data opens up the possibility for the setting up of one-to-one agreements for data access. In addition, some use cases, such as exploring virtual control groups, can still be met using partially redacted data [3]. Table 1 provides more details on what data would need to be redacted in the case of SEND formatted data in order to comply with these criteria.

How can it be shared?

Appropriate security procedures to prevent unauthorised access and unauthorised changes to data are vital to maintain the levels of trust needed. These procedures need to be proportionate to the risks involved and agreed with the data owners, encompassing physical security, network security, security of computer systems and files, as well as legal agreements and contracts.

In general, data donors expected the level of security and privacy protection to match or exceed what was implemented in their own IT environment, e.g. compliance with ISO27001 22, GDPR (European Parliament and Council of the European Union, 2016) and HIPAA (United States, 1996). In addition, the requirements are expected to evolve in line with security best practices.

Conclusion

The data sharing guidelines developed within eTRANSAFE provide practical solutions for effective sharing of proprietary data which can help overcome potential company internal resistance to data sharing. We envisage that these guidelines could be reused by other proprietary data sharing initiatives and hope that they will obtain widespread adoption and recognition from regulatory bodies and international standard-setting organisations. The key recommendations of the eTRANSAFE data sharing guidelines are:

- A tiered data classification scheme to allow data to be protected as thoroughly as necessary, but also shared as widely as practical
- A structured company-internal clearance procedure in which decisions are centralised at each step in order to speed up & harmonise data sharing requests
- Appropriate security procedures to prevent unauthorised access & unauthorised changes
- Standardisation to facilitate data interoperability
- Quality assurance and quality control procedures to reduce errors
- Information on meta-data, traceability & provenance to facilitate reproducibility
- Sustainability plans to protect against obsolescence

Table 1: SEND domain variables scheduled to be redacted

Reason for redaction	Associated SEND variables
GDPR related identifiers	Comment (COVAL) Contributing Scientist (TSPARMCD=CNTRBSC) Evaluator (CLEVAL, COEVAL, DDEVAL, EGEVAL, MAEVAL, MIEVAL, PMEVAL, QEVAL, TFEVAL) Principal Investigator (TSPARMCD=PINV), Study Director (TSPARMCD=STDIR) & Sponsor's Monitor (TSPARMCD=STMON)
Company identifiers	Sponsoring Organization (TSPARMCD=SSPONSOR)
Pharmacology target	Pharmacologic Class (TSPARMCD=PCLASS) Planned Pharmacologic Target Common Name (TSPARMCD=PPTCNAM) & Mode of Action (TSPARMCD=PPTMDA) Entrez Gene Identifier (TSPARMCD=PPTEGID) & Entrez Gene Symbol (TSPARMCD=PPTEGSYM)
Compound identifiers	Description of Planned Arm (ARM) & Planned Arm Code (ARMCD & TXPARMCD=ARMCD) Conditions which cause an arm to Branch (TABRANCH) Group Identifier (PCGRPID & PPGRPID) & Group Label (TXPARMCD=GRPLBL) Sponsor-Defined Group Code (TXPARMCD=SPGRPCD) Set Code (SETCD) & Set Label (TXPARMCD=SETLBL) Description of Element (ELEMENT) & Element Code (ETCD) Unplanned Element (SEUPDES) Rule for Start of Element (TESTRL) & Rule for End of Element (TEENRL) Focus of Study-Specific Interest (FOCID) Name of Actual Treatment (EXTRT & TSPARMCD=TRT) & Lot Number (EXLOT), Primary Treatment CAS Registry Number (TSPARMCD=TRTCAS) Unique Ingredient ID (TSPARMCD=TRTUNII) Chemical Structure as SMILES (TSPARMCD=TRTSMILE) Parameter Category e.g. analyte the parameter is associated with (PPCAT), Pharmacokinetics Concentrations Test Name & Short Name (PCTEST & PCTESTCD) Study Title (TSPARMCD=STITLE) Test Article Physical Substance Classification (TSPARMCD=TAPHSCLS) Time Point Reference (CLTPTREF, CVTPTREF, EGTPTREF, EXTPTREF, LBTPTREF, PCTPTREF, PPTPTREF & VSTPTREF)
Other proprietary information	Study Identifier (TXPARMCD=STUDYID) IACUC Number (TSPARMCD=IACUC) & Project License Number (TSPARMCD=PPL) Treatment Vehicle (TSPARMCD/TXPARMCD=TRTV)

References

[1] Pognan, F, et al. 2021, 'The eTRANSAFE Project on Translational Safety Assessment through Integrative Knowledge Management: Achievements and Perspectives Guidelines for FAIR sharing of preclinical safety and off-target pharmacology data', Pharmaceuticals (Basel), 14, 237

[2] Briggs, K, et al. 2021, 'Guidelines for FAIR sharing of preclinical safety and off-target pharmacology data', ALTEX, 38, 187-197

[3] Steger-Hartmann, T, et al. 2020, 'Introducing the concept of virtual control groups into preclinical toxicology testing', ALTEX, 37, 343-349

Acknowledgement & Disclaimer

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777365 ("eTRANSAFE"). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

This communication reflects the author's view and neither IMI nor the European Union, EFPIA, or any Associated Partners are responsible for any use that may be made of the information contained therein.