

Hospital del Mar Medical Research Institute, Barcelona Supercomputing Center, Lhasa Limited and PDS Consultants have contributed to a novel innovation recognised by the European Commission's Innovation Radar

An early innovation titled "[Pre-clinical text mining solution for treatment response](#)" undertaken by [Hospital del Mar Medical Research Institute \(IMIM\)](#) and [Barcelona Supercomputing Center \(BSC\)](#) in collaboration with [Lhasa Limited](#) and [PDS Consultants](#) within the [IMI2 eTRANSafe Project](#) has been recognised by the European Commission's Innovation Radar as being at the 'exploratory' stage for addressing the needs of existing markets.

What is this 'exploratory' innovation?

One of the goals of the eTRANSafe project is to address translational challenges in drug safety assessment. The need to capture expert conclusions of animal toxicology studies in a consistent, structured, machine-readable format was identified as a key aspect in utilising the shared pre-clinical safety data for the translational use cases of the project. Such expert conclusions contain observations for risk, severity and statistical significance qualified by the group, sex, day, specimen, and test to which they relate. However, these expert conclusions are not captured or available in the CDISC Standard for Exchange of Non-clinical Data (SEND), or in pre-clinical data originating from Laboratory Information Management Systems (LIMS), which form the core electronic pre-clinical data exchange formats employed within eTRANSafe.

To overcome these limitations the eTRANSafe consortium are developing a Pre-Clinical Text Mining Pipeline to extract expert conclusions for treatment-related findings from animal toxicology study reports into the eTRANSafe Pre-Clinical Database (figure 1). This pipeline has three elements:

1. A team of researchers from **IMIM** and **BSC** are developing text mining tools aimed at extracting the expert conclusions from animal toxicology study reports, such as the treatment related findings of the study. To enable this, a series of non-confidential toxicological reports from the pharmaceutical organizations involved in eTRANSafe have been securely shared with teams at IMIM and BSC and subsequently used by safety experts from the participating pharmaceutical organizations to identify the relevant terms. Sharing such data has been essential to assemble, evaluate and customize a text mining solution for automated supervised identification of toxicological relevant findings in unstructured data sources.
2. **PDS Consultants** have developed the concept of the "Study Report (SR)-Domain" as a proposed additional domain to SEND [[Drew et al., 2019](#)]. The SR-Domain forms the standardised target into which the text mining tools will extract the expert conclusions of an animal toxicology study. Additionally, PDS Consultants have developed an SR-Domain Editor application to facilitate the manual population or editing of an SR-Domain for an animal toxicology study. This application may be used by experts to validate the treatment related findings extracted by text mining methods or may be used by a Study Director to record significant study findings directly within an SR-Domain upon the conclusion of a study.
3. **Lhasa Limited**, the pre-clinical data honest broker in eTRANSafe, have integrated the SR-Domain within the cloud-hosted eTRANSafe Pre-Clinical Database Platform which they have designed and developed. This database is centred on the SEND format, however the novel SR-Domain has been included within the database schema to enable the expert toxicological findings of a pre-clinical study to be captured, stored and utilised by eTRANSafe partners.

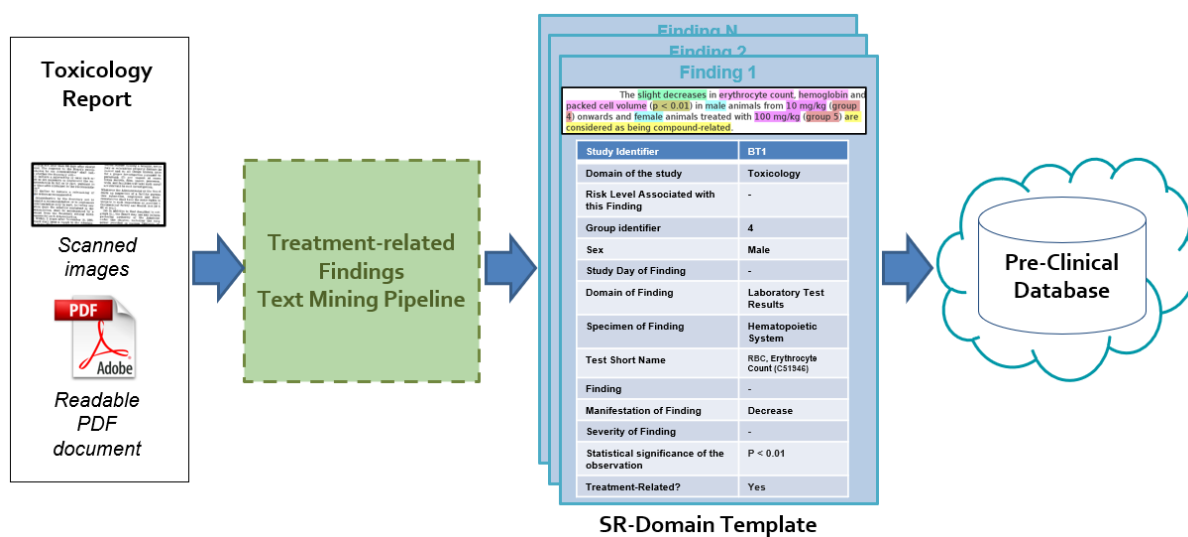


Figure 1. Pre-clinical text mining pipeline to extract treatment-related findings to the eTRANSafe Pre-Clinical Database. The text mining pipeline extracts relevant findings from toxicology study reports and populates the SR-Domain template which facilitates standardisation and import into the eTRANSafe Pre-Clinical Database.

On this innovation, William Drewe, the Leader of Lhasa’s eTRANSafe project team comments:

“The concept of the SR-Domain, which has been proposed by Phil Drew of PDS Consultants, is very exciting. This novel SEND-like domain has been well received by both the project partners of eTRANSafe as well as delegates of conferences where Phil has made presentations, such as the PhUSE Computational Science Symposium (CSS) which is attended by CDISC and the FDA. While the text mining methods being developed by IMIM and BSC are still in the early R&D stage, we hope that they will soon be able to automatically extract critical toxicological findings from study reports and standardise them within the SR-Domain template so that we at Lhasa can transfer these data into the eTRANSafe Pre-Clinical Database Platform we have developed. This should enable modellers and data users to understand the relevance of the SEND-based pre-clinical study data and enable these researchers to work towards some of the translational goals of eTRANSafe.”

This statement is complemented by Laura I. Furlong (IMIM) and Salvador Capella-Gutierrez (BSC):

“The development of a dedicated text mining solution for the extraction of relevant findings from toxicology reports has been exciting work as many components have been adapted from other knowledge domains to address this challenge. An important contribution of this effort is the generation of a manually annotated corpus of toxicologically relevant findings by experts from 10 of the pharmaceutical companies participating in eTRANSafe. This effort also required the donation of toxicology reports from the industry partners. This corpus constitutes the first resource of this kind and will enable the generation of text mining tools for the automated extraction of relevant information from legacy pre-clinical toxicology reports. These text mining tools will enable the rapid population of SR-Domain templates leaving the final validation of those annotations to human experts. These innovative text mining resources will therefore allow the optimization of resources and the extraction of hidden information from company archives.”

This work has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 777365 (“eTRANSAFE”). This Joint Undertaking receives financial support from the European Union’s Horizon 2020 research and innovation programme and from its EFPIA organisation members.