



# EMIF Deliverable 5.1: Establishing functional data warehousing and management

## Executive summary

---

### Executive Summary

The task is to establish functional data warehousing and management for 'omics data. The tranSMART data repository combines a data mart with access to federated sources of open and commercial databases. tranSMART provides a demonstrated schema and loading utilities for a variety of data types:

- **Phenotypic data**, such as demographics, clinical observations, clinical trial outcomes, and adverse events
- **High content biomarker data**, such as gene expression, genotyping, pharmacokinetic and pharmaco-dynamics markers, metabolomics data, and proteomics data
- **Unstructured text-data**, such as published journal articles, conference abstracts and proceedings, and internal studies and white papers
- **Reference data** from sources such as Medical Subject Headings (MeSH), Unified Medical Language System (UMLS), etc.
- **Metadata** providing context about datasets, allowing users to assess the relevance of results delivered by tranSMART.

On the short term, to meet the timelines, a drop-box solution will be used as a temporary solution while the tranSMART solution is further developed. A drop-box solution will allow the team to store and exchange files with an access and security procedure defining roles and related privileges.

### Contacts

EMIF-Metabolic: Dawn Waterworth and Ulf Smith

[Dawn.M.Waterworth@gsk.com](mailto:Dawn.M.Waterworth@gsk.com) / [ulf.smith@medic.gu.se](mailto:ulf.smith@medic.gu.se)