



# EMIF Deliverable 1.2: Private Remote Research database environment with pooled data from research cohorts

## Executive summary

---

### Executive Summary

In order to perform analyses across the cohorts available in EMIF-AD we need to pool these cohorts. In this report, we describe the technical solution for pooling of data across cohorts for a selected number of variables.

Specific obtained results:

- 4 AD cohorts were used and a list of priority variables was defined to pool data in TransSMART.
- Webprotégé was used to harmonize data of the selected variables and served as a communication tool between the EMIF-AD and EMIF-platform teams. Via this tool it could be defined how the selected variables map to measurements in the local data sources.
- The pooled data in TransSMART is available as a virtual cohort, which means that data of a cohort only needs to be uploaded once.
- All data analysis functionalities in TransSMART can be used to analyse a pooled dataset. Thus, possibilities for cross trial analyses are the same as those for single cohort analyses.
- Future challenges include pooling data of high dimensional data and optimizing the security framework of TransSMART so that also partial data access to cross trial data is possible.
- Next steps include the further development of the knowledge object library as a sustainable solution for data harmonisation and the upload of new cohorts for data pooling in TransSMART (e.g. for the WP3 1000 subjects study)

### Contacts

EMIF-AD: Pieter Jelle Visser and Johannes Streffer

[pj.visser@maastrichtuniversity.nl](mailto:pj.visser@maastrichtuniversity.nl) / [JSTREFFE@its.jnj.com](mailto:JSTREFFE@its.jnj.com)