EPAD Deliverable 3.4

Data Discovery software for EPAD Register v1

Executive Summary

The EPAD project will recruit subjects from pre-existing Parent Cohorts across Europe to enter into a longitudinal cohort study and a clinical trial. This subject selection will employ a technology for data/subject 'discovery' – i.e., remote exploration of all participating Parent Cohort datasets, to find counts of subjects of interest without displaying subject data. When in place, this capability will constitute the 'EPAD Register'. It will be a virtual resource and not a traditional database, and it will be dynamic in nature in that subject lists and their searchable attributes will change with time.

The EPAD Register will use and adapt existing software, not least technologies devised partly or solely by EMIF. At the start of EPAD one of these, called 'Cafe Variome', was ready for use, and so it was adapted to meet the initial needs of the project. A complete description of the resulting platform is presented in Appendix A.

The development work so far undertaken includes: (i) establishing a pilot site (using DESCRIPA and EDAR content) to demonstrate the tool and stimulate discussions into how it should be adapted; (ii) defining a phased approach to enable increasingly deep data interrogation, including a common data model for initial Parent Cohort data discovery; (iii) establishing a 'harmonisation pipeline' to transform diverse source data into the common data model format; and (v) modifying Cafe Variome to support the required hub-and-spoke (single search portal) architecture and to offer a dedicated phenotype query builder interface.

The above work was completed and the 'EPAD Register V1.0 Demonstration site' was launched by July 2015. The system was initially populated with mock data, which is to be rapidly replaced by real Parent Cohort data as soon as it is made available. The platform will first operate as a centralised system, which will evolve into a federated architecture as Parent Cohorts begin installing the Café Variome and harmonisation pipeline software locally.

For more information: info@epad.org