

THE METADATA CATALOGUE EXPLAINED

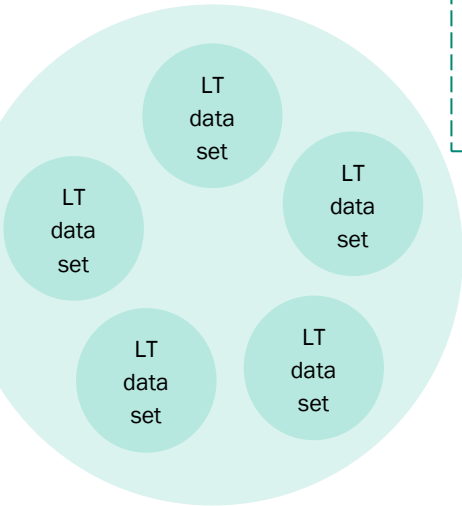
MOLGENIS is a free to use software which provides data on multiple datasets used by various projects including LIFECYCLE and ATHLETE

DataSHIELD enables secure and legally compliant access to data sets

Metadata – data about data!

Each data set may collect their data variables in a different way e.g. one cohort study may record height in centimetres, another in feet and inches. **Harmonisation** of the data variables within the Metadata Catalogue makes the data interoperable and enables researchers to interrogate and make use of more data when undertaking their research.

The actual data within the data sets is held locally and accessed by the researcher via **DataSHIELD**, subject to the necessary data access procedures. DataSHIELD only returns summary statistics to analysis questions from the researcher.



Data harmonisation – transforming multiple data sets into one cohesive, interoperable set e.g. ensuring all variables relating to height are in centimetres

LongITools (LT) data sets include:

- register-based cohorts
- prospective birth cohort studies
- biobanks
- longitudinal studies in adults
- randomised controlled trials

The Metadata Catalogue is open access and web-based using the Molgenis software platform. It follows and supports the **FAIR data** infrastructure principles – Findable, Accessible, Interoperable and Reusable.

Interoperable – ensuring multiple data sets can be used together and with other networks in the future

MOLGENIS

METADATA CATALOGUE

Data sets (e.g. cohorts, biobanks) – metadata detailing the LongITools data sets

Data dictionaries – the variables contained in the various data sets

Common data models – target variables for harmonisation and mapping from the data source variables

Institutions – contributors to the Metadata Catalogue searchable by type and country

EXPOSOME RESEARCH



In analysing data, LongITools researchers may determine the causal effects of disease. Using the Metadata Catalogue, they are able to assess all the data sets available from the LongITools consortium, and determine their suitability to answer specific exposome research questions.