Background
One barrier to scalability and reproducibility is the phenotyping of cohorts (exposures, outcomes) of interest. The DARWIN EU® CC has created a repository of phenotypes deemed valid for regulatory research tested across the DARWIN EU Data Partner network named DECK (DARWIN EU Cohort Knowledge). We have also created a process that guides researchers through the steps needed for the use, generation, enhancement, and storage of phenotypes based on the OMOP CDM.

RESULTS
The process starts with the submission of a Phenotype Proposal Form with a semi-structured clinical description and an optimisation strategy that informs both concept set creation and the logics needed for cohort creation.

After assessing the feasibility of identifying such a clinical entity, the next step is to check whether the requested phenotype exists already in the phenotype library, or if needs to be created from scratch.

In case that a compatible phenotype already exists, the next step is to decide whether it is suitable for the proposed use, or if it needs to be modified, and how. Depending on the answer to these questions, a phenotype can be reused as is or it can be modified or adapted for the proposed new use. In case a phenotype needs modification, the resulting new version will be evaluated and (if approved) stored in DECK.

In case no compatible phenotype exists, a new one will be generated from scratch:

First, a search for potentially existing concept sets needed and evaluated or modified for the new use.

If no concept set exists, a new code list/concept set would be generated using the CodelistGenerator R package based on the specifications detailed in the Phenotype Proposal Form, and reviewed by clinical domain experts, and similarly evaluated for DARWIN EU® studies and Data Partners.

Once concept sets are available and deemed suitable, a cohort (or series of cohorts) will be created using software like ATLAS or programmatically using CapR, potentially including different flavours or optimisation strategies.

Following this step, the resulting cohorts will be evaluated using Cohortdiagnostics. Once approved, the cohorts and concept sets will be stored in DECK, with a clear registry of all the process steps, the request forms used, the clinical reviews, tracked changes, databases used for diagnostics, cohort evaluations, and the phenotyping and clinical experts involved.

CONCLUSION
A standardised procedure for creating, evaluating, and storing phenotypes has been created for DARWIN EU®, alongside of a phenotype library, i.e. DECK, that will be updated regularly as new phenotypes are added or revised. To aid in this process we are creating a web tool that follows the process and guarantees traceability, reproducibility, and reusability of concept sets and cohorts created and facilitates version control, user management.

Prototype of the phenotype tracker and library, the DECK