

Figure S1. Data model for enrolling patients to studies.

Patients can be included in multiple studies and an informed consent is required for each study. In the crosstab 'individual_study' in the Trecode data model, it is recorded per patient - study combination which informed consent applies.

Figure S2. Description of DNA-methylation experiments in Trecode data model.

Array based DNA-methylation experiments are described similar to sequencing experiments. The preparation of the DNA sample (biomaterial) is described for array experiments in table 'assay'. As for sequencing experiments, the results (raw data) are linked to run. Database modeling and integration of the computational analyses is currently under active development.

Figure S3a. Automated data submission to EGA.

Ahead of sample metadata submission, queried sample data is EGA-pgp encrypted using Trecode's data encryption workflow (1). Checksums of the encrypted files, which are required for metadata submission, are automatically recorded in Trecode and encrypted files are transferred to EGA (2). A similar query is performed by the operator for metadata submission in which all required metadata for the queried samples and experiments is gathered by the Trecode platform and submitted to EGA's REST interface (3). The returned receipt is processed, and relevant information is passed to the operator and stored in the Trecode database (4).

Figure S3b. Data model entities related to EGA submission.

The Trecode platform keeps track on EGA submissions using the tables 'filehashes' for storing encrypted file checksums and 'submissions' for keeping track on which sample batches have been submitted and when. Table 'mappings' records the relation between the assigned identifiers by EGA and Trecode identifiers. Datasets for publications, which are stored in 'dataset', are composed from submitted runs and analyses in the Trecode web GUI. EGA datasets require a reference to a data policy ('policy') and a data access committee ('dac').