The GHGA Metadata Model: A Framework for National Data Sharing

Data sharing at both the national and international level benefits genomic medicine. Specifically, in diseases driven by genomic factors, such as cancer types or rare diseases, data sharing maximizes the utility and impact of cohorts, thereby aiding in translating research findings to therapies. The successful discovery of new findings, however, requires linking genomic data to health data and sharing both, which in turn necessitates national metadata standardization and harmonization along with a data protection framework.

The necessary infrastructure for FAIR (findable, accessible, interoperable, and reusable) data management, storage, and access on a national level will be provided by the German Human Genome-Phenome Archive (GHGA). In the work presented here, we introduce the underlying metadata model of GHGA. By exploring and building on several already existing models and in close discussions with stakeholders from genomic medicine, we defined a harmonized metadata model covering metadata elements pertaining to technical (experiment and analysis), individual (sample) and dataset data. Standardization of the model is achieved with the usage of several well-established ontologies and the definition of controlled vocabularies, making it self-describing, unambiguous, flexible, and expressive. The schematic model backbone is defined as YAML using the Linked Data Modeling Language (LinkML). As GHGA will be part of the federated European Genome Archive (EGA), our model is designed to be compatible with EGA.

Our model demonstrates how genomic and health data can be stored in accordance with General Data Protection Regulations and securely shared across German institutions. It balances the individual data subject's right to privacy while ensuring high-quality metadata to make genomic data in GHGA findable and reusable.

Please assign your poster to one of the following keywords.

Standards

Please assign yourself (presenting author) to one of the stakeholders.

Data Infrastructure Provider

Please specify "other" (stakeholder)

In addition please add keywords.

model, standardization, harmonization, ontologies, health

Primary author: Ms MAUER, Karoline (DZNE)

Co-authors: IYAPPAN, Anandhi; UNNI, Deepak; KRAUS, Florian; PARKER, Simon; TREMPER, Galina; SURUN, Bilge; MENGES, Paul; KIRLI, Koray; SCHULTZE, Joachim. L.; ULAS, Thomas; NAHNSEN, Sven

Presenter: Ms MAUER, Karoline (DZNE)

Session Classification: Postersession I

Track Classification: Postersession