### Overview

In medical research, extensive data collections are collected at great personal and financial expense, which are subject to a wide range of regulatory requirements due to their nature: person-related, sensitive subject data. Although the idea of data sharing is now supported by many clinical researchers, there is a lack of mechanisms for the structured processing, permanent storage and organized release of interoperable data to external scientists.

[Immagine che contiene testo

Descrizione generata automaticamente](https://www.health-atlas.de/)

The Leipzig Health Atlas (LHA) serves as a research data management repository for archiving and sharing data(sets) from scientific projects in the biomedical and medical informatics fields. It includes data and publications from research studies such as clinical trials, epidemiological cohorts or registries.

The LHA is guided by the FAIR principles (Findable, Accessible, Interoperable, Reusable). Currently, more than 18 research consortia contribute data from the domains of lymphoma, glioma, sepsis, hereditary colorectal and breast cancer, among others. The target audience includes clinicians, epidemiologists, geneticists, pathologists, biostatisticians, and digital patient model developers. The LHA is publicly available - some content subject to legal access restrictions - at <https://www.health-atlas.de>.

### Technical background

The architecture of the LHA portal is based on the freely available software [SEEK](https://fair-dom.org/platform/seek/). Additional tools for hypothesis generation ([i2b2](https://www.i2b2.org/)), visualization ([tranSMART](https://transmartfoundation.org/)), interactive modeling for risk assessment (based on [R-Studio Shiny](https://shiny.rstudio.com/)) or deployment ([Docker](https://www.docker.com/)/[Kubernetes](https://kubernetes.io/)) are offered on a case-by-case basis. Before making datasets available, it is necessary to curate the data and metadata, obtain permissions from the data owners, consider the required privacy criteria, and assign semantic tags for better discoverability. In the course of data extraction, contractual and technical access procedures must be established and implemented.

### Example clinical trial ([HNSCC](https://www.health-atlas.de/studies/33) study)

The HNSCC study is an observational cohort study that characterizes head and neck squamous cell carcinomas (HNSCC) with different HPV16 DNA and RNA (E6\*I) status. The database used in the publication consists of clinical data from the hospital information system and genomic data from tumor sequencing. In order to support the reproducibility of research and to promote the reuse of data, the clinical data are offered in various data files compliant to different standards, including for example simple comma separated data or complex structures according to the [CDISC ODM standard](https://www.cdisc.org/standards/data-exchange/odm).

|  |
| --- |
| Immagine che contiene testo  Descrizione generata automaticamente |
| URL: <https://www.health-atlas.de/studies/33> |

### Options for utilizing HL7 FHIR

Interoperability and traceability are key issues for a research data repository that wants to make data available to third parties for reuse in secondary research projects. Experience shows that the existing processes for requesting and extracting data are not sufficiently descriptive, so that, for example, questions about the semantics of individual data elements or the provenance of the data sources and transformation processes must be clarified for each individual requester.

An investigation has already been conducted regarding the compliance of the LHA to the [indicators developed by the RDA FAIR Data Maturity Model Working Group](https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0). In principle, the results were satisfactory, but there were major weaknesses in the area of reusability.

The use of HL7 FHIR allows a clearly detailed description of contextual attributes of data and data sets. Currently, plans are underway for the programmatic implementation of a semi-automatic generation of a data export in FHIR under the recommendations of the implementation guide presented here.

### Outlook and future utilization

The Leipzig Health Atlas is to be expanded in the long term to become the U Leipzig medical faculty's research data management platform. More importantly, it will also serve as a basis for the realization of distributed Local Data Access Points within the framework of the [National Research Data Infrastructure for Personal Health Data (NFDI4Health)](https://www.nfdi4health.de/en/). Furthermore, the concept will be established in the SMITH consortium of the Medical Informatics Initiative for making research data available.

**Acknowledgement**: The LHA project was funded by the [BMBF](https://www.bmbf.de/bmbf/en/) program i:DSem (Integrative Data Semantics for Systems Medicine, funding code 031L0026).