

International Journal of Population Data Science

Journal Website: www.ijpds.org



Evolving the HDRUK Phenotype Library: Phenotype Creation and Editing

Daniel S Thayer¹, Jack Scanlon¹, Muhammad A Elmessary¹, Artur Zinnorov¹, Ieuan Scanlon¹, Alex Coldea¹, Hannah Davies¹, Carla Oliveira², Spiros Denaxas³, Emily Jefferson⁴, and Harry Hemingway³

¹SAIL Databank, Swansea University

²European Bioinformatics Institute

³Institute of Health Informatics, University College London

⁴Health Data Research UK

Objective

The HDRUK Phenotype Library (phenotypes.healthdatagateway.org) shares definitions used to measure concepts of interest (such as diagnoses or treatments) within health datasets. It already holds more than 1000 phenotypes, with researchers able to contribute their work via an API. We aimed to create a more user-friendly method of contributing to the Library.

Approach

We designed a phenotype creation workflow enabling users to create and submit new content via web interface. Goals included clarity and ease of use for a broad range of users.

Results

A home page shows researchers' own content. Researchers can create new phenotypes using a web form, entering metadata such as name, authors, description, and publications. Code lists are defined via one or more rules, including search terms or referring to another phenotype, or by CSV upload. Users can make their phenotypes accessible to a research group or to all authenticated users, as well as publish content on the web. Publication requests are reviewed to ensure content is complete and appropriate. Editing, with full history and version control, is also supported.

Conclusions

We implemented and released the new features. We are currently engaging researchers to get feedback and invite content submission.

Implications

The benefit of tools to support research transparency and repeatability is only realized when they are adopted. We hope that a GUI to support phenotype creation will broaden the Library's user base help it serve as an enabler of higher-quality, more efficient research across the worldwide health data research community.

