HERO-Genomics: Bridging Genomic Data and Ontological Modelling

Laura Menotti[1], Mirco Cazzaro[1], Manuel Rueda[2], Ivo G. Gut[2] and Gianmaria Silvello[1]

[1]*Department of Information Engineering, University of Padua, Padua, Italy*
[2]*CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain*

Abstract
The Hereditary Ontology for Genomic Data (HERO-Genomics) supports the representation of genomic data specific to Amyotrophic Lateral Sclerosis (ALS) and Multiple Sclerosis (MS), enabling the documentation of gene mutations linked to these diseases. The current version includes a detailed framework for storing specific gene sequencing variations, such as Single Nucleotide Variants (SNVs). HERO-Genomics is part of the Hereditary Ontology (HERO), which is the backbone of the Semantic Data Integration platform of the HEREDITARY project, exploiting the Ontology-Based Data Access (OBDA) technology to query, aggregate, and join large heterogeneous data in a distributed manner using a unique query language,i.e. SPARQL.

Keywords
Ontology, Genomic Data, RDF, Data Modeling

1. Introduction

Genome sequencing is being used to aid diagnoses, particularly of rare diseases, to inform cancer treatment and progression, and to create predictive models for precision medicine [1]. The widespread availability of genomic data calls for a unified representation of genomes and trustworthy data discovery protocols. Global Alliance for Genomics and Health (GA4GH) developed two initiatives towards genomic data modeling: Phenopacket [2] and Beacon [3, 4].

Phenopacket [2] defines a machine-readable phenotypic description of a patient or a sample in the context of rare, common, or complex disease, or cancer. Phenopacket is a GA4GH standard for linking phenotypic descriptions, genetic information, diagnoses, and treatments to each patient. The Phenopacket schema is a data model represented using protocol buffers, a language-neutral, platform-neutral mechanism for serializing structured data. The data schema is described in .proto files and can be serialized with a hierarchical structure, e.g. JSON. Beacon

SWAT4HCLS '25: The 16th International SWAT4HCLS conference, February 24-27, 2025, Barcelona, Spain.
laura.menotti@unipd.it (L. Menotti); mirco.cazzaro@phd.unipd.it (M. Cazzaro); manuel.rueda@cnag.eu (M. Rueda); ivo.gut@cnag.eu (I. G. Gut); gianmaria.silvello@unipd.it (G. Silvello)
https://www.dei.unipd.it/~menottilau/ (L. Menotti); https://www.dei.unipd.it/~cazzaromir/ (M. Cazzaro); https://www.cnag.eu/manuel-rueda (M. Rueda); https://www.cnag.eu/ivo-g-gut (I. G. Gut); https://www.dei.unipd.it/~silvello/ (G. Silvello)
0000-0002-0676-682X (L. Menotti); 0009-0006-3856-7207 (M. Cazzaro); 0000-0001-9280-058X (M. Rueda); 0000-0001-7219-632X (I. G. Gut); 0000-0003-4970-4554 (G. Silvello)
© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

v2 [3, 4] is an Application Program Interface (API) specification for the federated discovery of both genomic and phenotypic data. The specification consists of two components, the framework and the models. The former defines the format for requests and responses, while the latter defines the biological data response structure. Beacon allows the translation of Variant Call Format (VCF) files into structured data, i.e. JSON, straightforwardly. This process is called *beaconization* and takes approximately 30 minutes per 1M variants. With parallel processing, a Whole Genome Sequence (WGS) for a cohort of 2,500 patients can be converted in a few hours.

[LM:] Both initiatives provide a unified representation of genomic data; however, neither is RDF-compliant, which prevents the data from being represented as Linked (Open) Data in line with FAIR principles and limits interoperability with genomic ontologies. Thus, our work introduces the Hereditary Ontology for Genomic Data (HERO-Genomics), an ontology enabling the communication between Beacon and Resource Description Framework (RDF). This approach not only facilitates the FAIRification of genomic data but also supports the development of federated infrastructures that leverage OBDA systems. HERO-Genomics follows the Beacon Data Model, as it is well-maintained. Albeit the Genomic Data Modeling follows Beacon, many classes can be translated to the Phenopacket data model. [1] HERO-Genomics is part of the HERO, which is the backbone of the Semantic Data Integration platform of the HEREDITARY project, exploiting the OBDA technology to query, aggregate, and join large heterogeneous data in a distributed manner using a unique query language,i.e. SPARQL. [LM:] HERO has been developed in the context of the HEREDITARY project, whose goal is to harmonize and link heterogeneous sources of clinical, genomics, and environmental data on a large scale. The HEREDITARY framework exploits advanced federated analytics and learning workflows, to identify new risk factors and treatment responses focusing on brain-related disease, such as MS, ALS, and Parkinson's Disease, and diseases corning the gut-brain interplay.

The rest of this work is organized as follows: Section 2 describes the ontology design process, covering the domain requirements, the alignment with Open Biological and Biomedical Ontology Foundry (OBO) and FAIR principles, and the implementation choices. Section 3 outlines HERO-Genomics, offering an in-depth explanation of the representation of genomic variations and VCF files. Section 4 reports a use-case showing how HERO-Genomics can be used in practice to achieve Federated Data Access. Section 5 draws some final remarks and concludes the paper.

2. Methodology

The HERO-Genomics has been designed exploiting a co-design approach, collaborating with the medical partners and domain experts to embed their knowledge and, at the same time, to validate all the design choices. To this end, we operated iteratively, producing several (intermediate) versions of the ontology and discussing them with our domain experts.

The iterative discussion process with the medical partners ensures that these newly defined concepts correctly describe the corresponding real-world concepts and guarantees the semantic quality of the ontology.

[1] https://docs.genomebeacons.org/formats-standards/#phenopackets

Definition of the domain requirements In the context of the HEREDITARY project, genetic information is collected by means of VCF files comprising Single Nucleotide Polymorphisms (SNPs). A SNP is a one-base sequence where an individual's genome varies with respect to another sequence, usually called "Reference Genome". SNPs are SNVs present in a sufficiently large fraction, i.e. at least 1%, of a specific population. VCF is a text format widely used for storing genetic variations, such as SNVs, insertions, deletions, and structural variants, together with rich annotations [5]. The format was developed for the 1000 Genome Project [2] and has been adopted in several other projects, e.g. dbSNP [3]. Each VCF file comprises a header and a body. The former provides metadata describing the body of the file and keywords that optionally describe the fields used in the body. The latter is tap-separated and comprises 8 mandatory columns and an unlimited number of optional columns that may record additional information about the sample. When optional columns are used, the first of these columns describes the format of the additional columns. The mandatory columns are:

- CHROM, the name of the sequence, typically a chromosome, on which the variation is being called;
- POS, the one-based position of the variation;
- ID, the identifier of the variation, e.g. dbSNP rs identifier;
- REF, the reference base;
- ALT, the alternative allele;
- QUAL, the quality score associated with the inference of the given allele;
- FILTER, the flag indicating which filters have failed or PASS if all filters were successfully passed;
- INFO, the description of the variation. This column is highly variable and its content can vary across VCFs;
- FORMAT, an optional list of fields for describing the samples;
- SAMPLEs, optional values describing the samples.

The HERO-Genomics will answer the Beacon genomic queries provided and other queries of interest to the genomics community. The Beacon queries comprise:

- Sequence Queries for the existence of a specified sequence at a given genomic position;
- Range Queries for matching variants at least overlapping with a specified region;
- GeneId Queries for returning variants affecting a gene's coding region;
- Bracket Queries for matching variants falling in a start range and end range;
- Genomic Allele Queries for matching variants with the specified allele;
- Amino Acid Change Queries for matching variants with the specified amino acid change.

2.1. Alignment with FAIR Principles and OBO Standards

HERO-Genomics complies with the OBO [4] and Findable, Accessible, Interoperable, Reusable (FAIR) principles [6] [5], favoring its adoption in heterogeneous scenarios. The ontology is

[2] https://www.internationalgenome.org/
[3] https://www.ncbi.nlm.nih.gov/snp/
[4] https://obofoundry.org/principles/fp-000-summary.html
[5] https://www.go-fair.org/fair-principles/

defined according to the OWL 1.2 *Common Format*. The ontology is *open* and publicly available; its definition and detailed description can be found at http://hereditary.dei.unipd.it/ontology/genomics/. The proposed ontology relies on unique *URIs/Identifier Spaces* identified by the prefix https://w3id.org/hereditary/ontology/genomics/schema/. A description of the *Versioning* procedure is available as part of the documentation of the HERO-Genomics on the ontology web page. The *Scope* of HERO-Genomics is clearly defined: the ontology is meant to represent genomics data from different medical centers. Following the OBO principles, we associate *Textual Definitions* to each ontology class, also to favor its re-use in other scenarios. Before defining a new relation, *Relations* available on the Relations Ontology (RO) have been considered. None of HERO-Genomics relations presents the same meaning and could have been replaced with one of the RO – nevertheless, this possibility has always been scrutinized. For what concerns *Documented Plurality of Users* and *Commitment To Collaboration*, these aspects are intrinsic in developing and using HERO-Genomics. Indeed, HERO-Genomics has been developed in the context of the HEREDITARY Project, which includes partners from multiple countries (from the EU and USA). The co-design approach to devise the HEREDITARY ontology defines its *collaborative* nature. HERO-Genomics identifies its *Locus of Authority* into its developers, who are indicated on the web page of the ontology, and in the authors of this paper [LM:] , that comprises both medical experts and computer science specialists. HERO-Genomics follows strict *Naming Conventions* described in Section 2.2. Finally, the HEREDITARY consortium is actively working on the *Maintenance* and update of HERO-Genomics.

2.2. Design Choices

We follow some basic principles when defining classes and properties to provide consistency in the HERO-Genomics. These guidelines involve external referencing, annotation properties, and naming conventions.

External Referencing Reusing entities and properties already defined in other resources enforces collaboration and data consistency. External referencing is managed with annotation properties and using the Unique Resource Identifier (URI) of the term in the original thesaurus. In HERO-Genomics, external URIs are used when defining named individuals that refer to abstract concepts. On the contrary, when a new class is inserted in HERO-Genomics, it is defined within the HERO-Genomics namespace, and connected references are expressed using annotation properties.

Classes Definition and Annotation Properties All components of the HERO-Genomics have additional information in the form of annotation properties. All classes must have a label denoting the name and a comment, which provides a brief explanation – together with its source (e.g., other thesauri, websites, or textbooks). If the class has an equivalent in National Cancer Institute Thesaurus (NCIT) or other resources, the name and definition are inherited from the thesaurus. In this case, the class comprises another annotation property called `rdfs:isDefinedby` expressing the Internationalized Resource Identifier (IRI) corresponding to the resource term of reference. If there is an additional external reference, we use the property `rdfs:seeAlso`. Most biomedical vocabularies are mapped in the Unified Medical Language

System (UMLS)[6] with a unique identifier called Concept Unique Identifier (CUI) [7]. For each class that has a UMLS reference, the annotation property `dcterms:conformsTo` is instantiated with the URL of the corresponding concept.

Naming Conventions All components must have a label and a comment. About object properties, we use explanatory labels where the property range is included. In this case, the comment explains the relationship between the two classes involved. Concerning data properties, the label usually includes the name of the domain class so that its meaning is intuitive. A comment with the attribute description and, when available, the definition source are also included. Note that, all the HERO-Genomics components can comprise the `note` annotation property for additional remarks or business logic rules.

Usage of the Simple Knowledge Organization System (SKOS) Often we are interested in the abstract concept behind the medical term. In HERO-Genomics classism [8] is avoided for two important advantages: *i)* it dramatically reduces the number of required URIs, by not defining multiple named individuals; *ii)* it reduces the complexity of the queries. In HERO-Genomics, classification schemes that refer to abstract concepts already defined in other semantic resources, are modelled using the Simple Knowledge Organization System (SKOS) data model [7]. The SKOS data model allows for the storage of some particular information without instantiating one individual for each patient but by simply referring to the individual already instantiated as a concept. Note that this approach prevents us from describing the peculiarities of the specific entity. However, such a design principle is employed on components that do not have this requirement, i.e. for each class referring to a set of abstract terms without any associated data or object property.

3. The HEREDITARY Ontology for Genomic Data Modelling

The Genomic Data Modeling in the HEREDITARY ontology follows the Beacon Data Model, employing the "*Genomic Variation*" component as the core class of the ontology. HERO-Genomics allows the storage of genetic variations, e.g. SNVs, insertions, deletions, and structural variants, alongside their annotations, modeled with the "Case Level Variant" class. HERO-Genomics provides a direct mapping from VCFs files, or Beacon, to RDF format, thanks to classes "`VCF Record`", "`VCF File`", and "`Variant Quality`". The complete documentation of HERO-Genomics, including technical details, is available at: https://hereditary.dei.unipd.it/ontology/genomics/. The complete schema of HERO-Genomics can be found in the documentation website [8]. Figure 1 reports the schema of genomic variations and VCF files.

Genomic Variations HERO-Genomics records systemic variations, e.g. Copy Number Change, molecular variations, and legacy variations, e.g. SNVs. Annotations such as molecular effects and amino acid changes are represented by the class "`Molecular Attribute`". Class

[6] https://uts.nlm.nih.gov/uts/umls/home
[7] https://www.w3.org/TR/2009/REC-skos-reference-20090818/
[8] https://hereditary.dei.unipd.it/ontology/genomics/#figure1

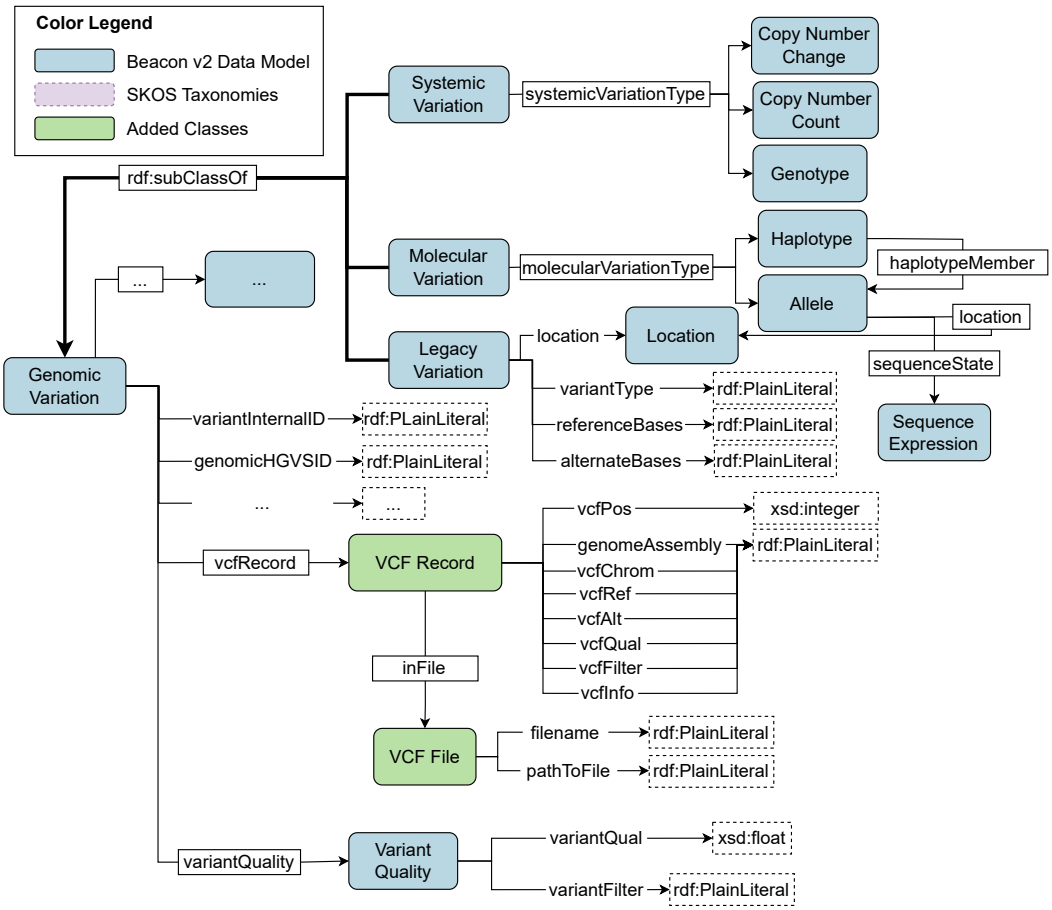

Figure 1: Genomic Variation Modeling in HERO-Genomics. HERO-Genomics records systemic variations, e.g. Copy Number Change, molecular variations, and legacy variations, e.g. SNPs. For each variation, one can store the corresponding VCF row and store additional annotations with classes "`Molecular Attribute`", "`Case Level Variant`", and "`Variant Level Data`". Classes in blue represent components from the Beacon Data Model, those in green were added to satisfy HEREDITARY's data requirements, and those in purple are classes that represent a SKOS taxonomy.

"`Case Level Data`" stores information about phenotypic effects, clinical interpretations, and the observed zygosity, a SKOS taxonomy from the Gene Ontology[9, 10] [9]. In addition, the class links the genomic variation to the corresponding biosample ("`Biosample`"), run ("`Run`"), analysis ("`Analysis`"), and patient ("`Patient`" from HERO-Clinical [10]). The location of the variation can stored with the class "`Location`", which provides a representation both for sequence location, to specify the sequence interval and reference genome, and chromosome location, to specify the cytoband interval of the chromosome where the variation occurs.

[9] https://geneontology.org/
[10] https://hereditary.dei.unipd.it/ontology/phenoclinical/

Systematic Variations comprise Copy Number Variation (CNV) and genotypes. CNV represents specific DNA segments that appear in a variable number of copies among individuals. In HERO-Genomics, CNVs are represented with two classes, namely "`Copy Number Change`" and "`Copy Number Count`". The former represents the change in the copy number of a sequence in a genome. This class records the location of the copy number change with the object property `location` that links the class to the "`Location`" class. One can also specify the type of variation copy change exploiting a taxonomy modeled using SKOS concepts, where values are from the Experimental Factor Ontology (EFO)[11], to be specific children of "`Copy Number Assessment`" [11]. On the other hand, "`Copy Number Count`" represents the integer number of copies of the DNA sequence in a genome. This class stores location information with the object property `location` and the integer number of copies of the sequence with data property `variationCopies`. Systematic Variations also comprises the class "`Genotype`", which stores the number of molecular variations with data property `molecularVariationCount` and the molecular variations as instances of class "`Genotype Member`".

Molecular Variations are variations on a contiguous molecule and can be classified into Haplotype and Allele. The former is a set of non-overlapping Allele members that co-occur on the same molecule. Thus, each haplotype is linked to its allele members with object property `haplotypeMember` ranging to class "`Allele`". The latter is a variant of the sequence of nucleotides at a particular location. Thus, class "`Allele`" has object property `location` to store the location of the allele and object property `sequenceState` to record the expression of the sequence state, represented by class "`Sequence Expression`". Sequence expressions can be "`Literal Sequence Expression`", i.e. an explicit sequence, "`Derived Sequence Expression`", i.e. a sequence that is derived from a sequence location, "`Repeated Sequence Expression`", or "`Composed Sequence Expression`".

Class "`Legacy Variation`" represents any other genomic variation. For instance, a SNP where base "A" is mutated to "T" can be represented as a Legacy Variation, by storing its location with object property `location`, the variant type "SNP" with data property `variantType`, data property `referenceBase` with value "A" and `alternateBase` with value "T".

Variant Call Format File One can store VCF files and each row with classes "`VCF File`" and "`VCF Record`". The former class comprises information about the file, such as its name and path, while the latter reports each column of the VCF row as a data property. Variant quality information, such as the quality tests performed and the variation score, can be reported using class "`Variant Quality`", which comprises properties `variantQual` for the variation quality score and `variantFilter` for the performed quality tests.

4. Ontology Deployment

[LM:] Genomic Queries

[11] http://www.ebi.ac.uk/efo/EFO_0030063

5. Conclusion

[LM:] This work presents the HERO-Genomics, an ontology enabling the communication between Beacon and RDF. HERO-Genomics follows the Beacon Data Model, as it is well-maintained and allows the translation of VCF files into structured data, i.e. JSON, straightforwardly. Future phases will consist of further refinements to HERO-Genomics for expanded genomics data capabilities. HERO-Genomics is part of the HERO, which is the backbone of the Semantic Data Integration platform of the HEREDITARY project, exploiting the OBDA technology to query, aggregate, and join large heterogeneous data in a distributed manner using a unique query language,i.e. SPARQL. HERO will be expanded to include aspects of the gut-brain interplay, with a focus on developing a gut microbiome ontology, an area currently with limited foundational work or related models. This approach will ensure that HERO remains robust and scalable, allowing it to capture complex biomedical data across a range of brain-related conditions and emerging research areas.

Acknowledgments

This project has received funding from the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074.

References

[1] K. J. Karczewski, M. P. Snyder, Integrative omics for health and disease, Nature Reviews Genetics 19 (2018) 299–310. URL: https://doi.org/10.1038/nrg.2018.4.

[2] M. S. Ladewig, J. O. B. Jacobsen, A. H. Wagner, D. Danis, B. El Kassaby, M. Gargano, T. Groza, M. Baudis, R. Steinhaus, D. Seelow, N. E. Bechrakis, C. J. Mungall, P. N. Schofield, O. Elemento, L. Smith, J. A. McMurry, M. Munoz-Torres, M. A. Haendel, P. N. Robinson, GA4GH Phenopackets: A Practical Introduction, Advanced Genetics 4 (2023) 2200016. URL: https://doi.org/10.1002/ggn2.202200016.

[3] M. Rueda, R. Ariosa, M. Moldes, J. Rambla, Beacon v2 Reference Implementation: a toolkit to enable federated sharing of genomic and phenotypic data, Bioinformatics 38 (2022) 4656–4657. URL: https://doi.org/10.1093/bioinformatics/btac568.

[4] J. Rambla, M. Baudis, R. Ariosa, T. Beck, L. A. Fromont, A. Navarro, R. Paloots, M. Rueda, G. Saunders, B. Singh, J. D. Spalding, J. Törnroos, C. Vasallo, C. D. Veal, A. J. Brookes, Beacon v2 and beacon networks: A "lingua franca" for federated data discovery in biomedical genomics, and beyond, Human Mutation 43 (2022) 791–799. URL: https://doi.org/10.1002/humu.24369.

[5] P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, . G. P. A. Group, The variant call format and VCFtools, Bioinformatics 27 (2011) 2156–2158. URL: https://doi.org/10.1093/bioinformatics/btr330.

[6] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak,

N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The fair guiding principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. URL: https://doi.org/10.1038/sdata.2016.18.

[7] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, Nucleic Acids Research 32 (2004) 267–270. URL: https://doi.org/10.1093/nar/gkh061.

[8] D. Allemang, J. Hendler, F. Gandon, Good and bad modeling practices, in: Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL, Association for Computing Machinery, New York, NY, USA, 2020, pp. 436–440. URL: https://doi.org/10.1145/3382097.3382113.

[9] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology, Nature Genetics 25 (2000) 25–29. URL: https://doi.org/10.1038/75556.

[10] T. G. O. Consortium, S. A. Aleksander, J. Balhoff, S. Carbon, J. M. Cherry, H. J. Drabkin, D. Ebert, M. Feuermann, P. Gaudet, N. L. Harris, D. P. Hill, R. Lee, H. Mi, S. Moxon, C. J. Mungall, A. Muruganugan, T. Mushayahama, P. W. Sternberg, P. D. Thomas, K. Van Auken, J. Ramsey, D. A. Siegele, R. L. Chisholm, P. Fey, M. C. Aspromonte, M. V. Nugnes, F. Quaglia, S. Tosatto, M. Giglio, S. Nadendla, G. Antonazzo, H. Attrill, G. dos Santos, S. Marygold, V. Strelets, C. J. Tabone, J. Thurmond, P. Zhou, S. H. Ahmed, P. Asanitthong, D. Luna Buitrago, M. N. Erdol, M. C. Gage, M. Ali Kadhum, K. Y. C. Li, M. Long, A. Michalak, A. Pesala, A. Pritazahra, S. C. C. Saverimuttu, R. Su, K. E. Thurlow, R. C. Lovering, C. Logie, S. Oliferenko, J. Blake, K. Christie, L. Corbani, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, C. Smith, A. Cuzick, J. Seager, L. Cooper, J. Elser, P. Jaiswal, P. Gupta, P. Jaiswal, S. Naithani, M. Lera-Ramirez, K. Rutherford, V. Wood, J. L. De Pons, M. R. Dwinell, G. T. Hayman, M. L. Kaldunski, A. E. Kwitek, S. J. F. Laulederkind, M. A. Tutaj, M. Vedi, S.-J. Wang, P. D'Eustachio, L. Aimo, K. Axelsen, A. Bridge, N. Hyka-Nouspikel, A. Morgat, S. A. Aleksander, J. M. Cherry, S. R. Engel, K. Karra, S. R. Miyasato, R. S. Nash, M. S. Skrzypek, S. Weng, E. D. Wong, E. Bakker, T. Z. Berardini, L. Reiser, A. Auchincloss, K. Axelsen, G. Argoud-Puy, M.-C. Blatter, E. Boutet, L. Breuza, A. Bridge, C. Casals-Casas, E. Coudert, A. Estreicher, M. Livia Famiglietti, M. Feuermann, A. Gos, N. Gruaz-Gumowski, C. Hulo, N. Hyka-Nouspikel, F. Jungo, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, I. Pedruzzi, L. Pourcel, S. Poux, C. Rivoire, S. Sundaram, A. Bateman, E. Bowler-Barnett, H. Bye-A-Jee, P. Denny, A. Ignatchenko, R. Ishtiaq, A. Lock, Y. Lussi, M. Magrane, M. J. Martin, S. Orchard, P. Raposo, E. Speretta, N. Tyagi, K. Warner, R. Zaru, A. D. Diehl, R. Lee, J. Chan, S. Diamantakis, D. Raciti, M. Zarowiecki, M. Fisher, C. James-Zorn, V. Ponferrada, A. Zorn, S. Ramachandran, L. Ruzicka, M. Westerfield, The Gene Ontology knowledgebase

in 2023, Genetics 224 (2023) iyad031. URL: https://doi.org/10.1093/genetics/iyad031.
[11] J. Malone, E. Holloway, T. Adamusiak, M. Kapushesky, J. Zheng, N. Kolesnikov, A. Zhukova, A. Brazma, H. Parkinson, Modeling sample variables with an Experimental Factor Ontology, Bioinformatics 26 (2010) 1112–1118. URL: https://doi.org/10.1093/bioinformatics/btq099.