



Article

A Linked Data Application for Harmonizing Heterogeneous Biomedical Information

Nicola Capuano ^{1,*} , Pasquale Foggia ², Luca Greco ² and Pierluigi Ritrovato ² ¹ School of Engineering, University of Basilicata, 85100 Potenza, Italy² Department of Information and Electrical Engineering and Applied Mathematics, University of Salerno, 84084 Fisciano, Italy

* Correspondence: nicola.capuano@unibas.it

Abstract: In the biomedical field, there is an ever-increasing number of large, fragmented, and isolated data sources stored in databases and ontologies that use heterogeneous formats and poorly integrated schemes. Researchers and healthcare professionals find it extremely difficult to master this huge amount of data and extract relevant information. In this work, we propose a linked data approach, based on multilayer networks and semantic Web standards, capable of integrating and harmonizing several biomedical datasets with different schemas and semi-structured data through a multi-model database providing polyglot persistence. The domain chosen concerns the analysis and aggregation of available data on neuroendocrine neoplasms (NENs), a relatively rare type of neoplasm. Integrated information includes twelve public datasets available in heterogeneous schemas and formats including RDF, CSV, TSV, SQL, OWL, and OBO. The proposed integrated model consists of six interconnected layers representing, respectively, information on the disease, the related phenotypic alterations, the affected genes, the related biological processes, molecular functions, the involved human tissues, and drugs and compounds that show documented interactions with them. The defined scheme extends an existing three-layer model covering a subset of the mentioned aspects. A client–server application was also developed to browse and search for information on the integrated model. The main challenges of this work concern the complexity of the biomedical domain, the syntactic and semantic heterogeneity of the datasets, and the organization of the integrated model. Unlike related works, multilayer networks have been adopted to organize the model in a manageable and stratified structure, without the need to change the original datasets but by transforming their data “on the fly” to respond to user requests.

Keywords: linked data; biomedical ontologies; multilayer network analysis; neuroendocrine neoplasms; semantic information integration; polyglot persistence; multi-model database



Citation: Capuano, N.; Foggia, P.; Greco, L.; Ritrovato, P. A Linked Data Application for Harmonizing Heterogeneous Biomedical Information. *Appl. Sci.* **2022**, *12*, 9317. <https://doi.org/10.3390/app12189317>

Academic Editors: Antonio Sarasa Cabezuelo and Jordi Conesa Caralt

Received: 8 August 2022

Accepted: 14 September 2022

Published: 16 September 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The massive and continuous growth of biomedical data of a heterogeneous nature requires ever greater efforts aimed at their integration. Indeed, the proliferation of non-integrated and non-interoperable data greatly hinders their interpretation and prevents computer-assisted reasoning. Data integration and reproducibility are essential to biomedical studies; it becomes extremely important to observe the FAIR guiding principles where research works are recommended to be findable, accessible, interoperable, and reusable [1].

Generally, a key requirement for allowing data to be FAIR is the use of open approaches and standardized representation formalisms such as ontologies. Indeed, ontologies have proved crucial in supporting omics data integration [2]. However, since hundreds of biomedical ontologies have been designed and are currently available, a new problem has arisen, i.e., how to integrate different ontological schemes and make them interoperable [3]. To this end, in addition to syntactic issues related to the variety of models and formats, semantic problems arising from context-related different interpretations need to be encountered [4].

To get the most out of these data, different approaches are required for querying multiple ontologies and databases to provide researchers with biomedical information organized as interconnected entities in a semantic fashion [5]. The research described in this paper attempts to contribute toward this end within a specific domain, through an ontology-based *linked data* application. An approach to data aggregation and analysis in the field of rare *neuroendocrine neoplasms* (NENs) is proposed.

In previous research by the same authors [6] a novel linked data application was described for the same domain, integrating several existing biomedical ontologies including the *National Cancer Institute Thesaurus*, the *Mondo Disease Ontology*, the *MedGen* database, the *Disease Ontology*, the *Orphanet Rare Disease Ontology*, the *DisGeNet* database, and the *Gene Ontology*. Such data sources provided the relevant information to build a single knowledge model about NENs that is accessible via a client–server application.

In this work, an extension of the previous model through the integration of additional data sources is proposed. The knowledge model relies on the *multilayer network* formalism to semantically link heterogeneous data sources while grouping them into several “layers”, each corresponding to a specific “aspect” of the domain [7]. The original model consisted of three interconnected layers representing: information about diseases, affected genes, biological processes, and molecular functions of such genes and related gene products.

In addition to the information included in the original model, the extended version integrates information from new sources including the *Human Phenotype Ontology*, the *HPO-ORDO Ontological Module*, the *Human Protein Atlas*, the *Drug–Gene Interaction Database* as well as the *ChEMBL database*. Starting from the previous model, three additional layers have been designed and integrated to consider further aspects.

- The *first additional layer* provides details about disease-related *phenotypes* such as morphology, development, biochemical, physiological, and other features. It has been shown that a deep understanding of rare diseases together with the identification of prognostic and therapeutic implications can be accelerated by analyzing and correlating phenotypic and genomic data.
- The *second additional layer* provides an association between genes responsible for diseases and human tissue. This allows us to find commonalities between organs and illustrate the role of genes related to NENs.
- The *third additional layer* provides details about *drugs* with documented interactions with genes affected by NENs. The complexity in the extraction of relevant information about these tumors is motivated by the great heterogeneity in their biological features that can also determine heterogeneous responses to therapeutic agents. They often present subpopulations of cells with different angiogenic, invasive, and metastatic properties.

According to [8], biomedical ontologies pose several integration challenges due to both the complexity of the domain and the characteristics of the ontologies themselves which include thousands of classes. Furthermore, biomedical ontologies present profound organizational differences to the point of often being difficult to reconcile due to syntactic and semantic heterogeneity. Although properly integrated, such ontologies are often difficult for researchers to use due to the lack of user-friendly interfaces.

The developed model is aimed to meet these challenges by improving the way information is stored, enhancing interoperability while simplifying the work of scientists studying these rare diseases. Moreover, a user-friendly interface can guide healthcare professionals and researchers through the process of searching for relationships between pathologies (and their related genes) also highlighting the effectiveness of adopted drugs. The adopted multilayer network formalism is useful for organizing the model into a manageable and layered structure. In accordance with [4], a *workflow-based approach* has been adopted to integrate data from external sources “on the fly” on a frequently updated local copy, based on user requests, therefore without the need to modify the original datasets and without the risk of using obsolete data.

The paper is organized as follows: in Section 2 some background information on NENs is provided and the related work on biomedical data integration is summarized; in

Section 3 the starting point of this research is described, including the previous integration model as well the additional biomedical data sources; in Section 4 the structure of the new model and the related integration issues are presented; in Section 5 the extension of the developed application for browsing and querying the updated model is described. The last section summarizes the conclusions and outlines the ongoing work.

2. Background and Related Work

The research work described in this paper aims at supporting biologists, researchers, and specialists in the collection, organization, and analysis of existing biomedical data on *neuroendocrine neoplasms* (NENs) through the definition and development of a linked data application. During the last four decades, these neoplasms have shown a 6.4-times increasing age-adjusted annual incidence [9]. NENs have been observed in almost every tissue, either in the pure endocrine organs, the nerve structures, or in the diffuse neuroendocrine system.

The *World Health Organization* has defined two groups of NENs to enable consistent management of diseases regardless of their anatomical location [10]: *neuroendocrine carcinomas* (NECs) and *neuroendocrine tumors* (NETs). Although NETs appear as well-differentiated neoplasms and can be categorized into three levels—G1, G2, and G3 (low, intermediate, and high grade)—NECs are poorly differentiated neoplasms with only high grade (i.e., G3). Cell grade and differentiation often depend on different factors such as mitotic count and Ki-67 cell labeling index [11]. Moreover, NECs can be also classified into small- or large-cell-type NECs.

The integration of information about NENs is still quite difficult since variations depending on anatomical sites often lead to definitions different from the accepted and established ones. Heterogeneous data sources with different schemes and formats make the information retrieval and analysis process quite difficult to perform and often require the clinician to have advanced programming skills.

Some projects have been proposed to integrate and harmonize biological data sources. For example, in the *Gene Expression Data Warehouse* (GEDAW) project a data warehouse has been proposed to manage relevant information on liver gene expression data and related biomedical resources [12]. It provides the integration of gene information from several data sources, including *GenBank* and *BioMeKe*. *Bio2RDF* also has the purpose of transforming heterogeneous biomedical information into linked data using semantic Web technologies [13]. It currently consists of 11 billion triples and 35 connected datasets. *Bio2RDF* is part of the *Life Sciences Linked Open Data* (LSLOD), in the context of the *Linked Open Data* initiative [14].

The *Knowledge Base of Biomedicine* (KaBOB) integrates 18 biomedical data sources using 14 ontologies from the Open Biomedical Ontologies (OBO) initiative [15]. Such a model integrates data sources by producing a single biomedical entity for each set of data source-specific equivalent identifiers. Similarly, the *Genomic and Proteomic Knowledge Base* (GPKB) links several biomedical data sources such as *Entrez Gene*, *UniProt*, *IntAct*, *Expasy Enzyme*, *GO*, *GOA*, *BioCyc*, *Kegg*, *Reactome*, and *OMIM* [16]. It provides a set of maintenance procedures to update the knowledge base depending on the evolutions of its sources and their consistency.

The *Software for Flexible Integration of Annotation* (SoFIA) aims at integrating omics information from several sources [17]. It relies on a minimal workflow that, given a starting goal indicated by the researchers, allows one to complete the task and return a relevant subset of information. More recently, the R package *Onassis* has been introduced to easily associate samples from large-scale biomedical repositories to ontology-based annotations [18]. *Onassis* leverages NLP techniques, biomedical ontologies, and the R statistical framework to identify, relate, and analyze datasets from public repositories.

In the domain of cancer research, *SysCancer* has the aim to provide an integrated system that combines different stages of cancer studies [19]. The data warehouse can allow a multidimensional analysis of collected and integrated data meant for public access. In [20] the authors developed a cancer staging ontology based on the guidelines of the *American*

Joint Commission on Cancer. The initial knowledge graph has been augmented by integrating additional open-source information about treatment and monitoring options depending on the inferred stage.

A network-based data integration framework for the semantic integration of clinical and omic data on breast cancer and neuroblastoma is presented in [21]. Here, a NoSQL database is used to combine heterogeneous raw data records and external knowledge sources. A cervical cancer ontology has been developed by [22] where the authors define 880 standardized concepts, 1182 common terms, 16 relations, and 6 attributes which are organized into 6 levels and 11 classes.

The first example of a linked data application, aimed at integrating and harmonizing existing information on NENs was reported in [6]. It connects data from several sources, providing a single access point to a detailed network of information, organized on three interconnected layers. The ontology developed in this latest paper and the related software prototype constitutes the starting point of the present study.

As anticipated in the introductory section, this work extends the existing ontological model by designing and integrating three additional layers, which refer to further aspects of the domain. On the other hand, the new model exhibits the innovations already proposed in the previous version compared to the existing literature, i.e., the use of multilayer networks to organize the model in a manageable structure, the use of a workflow-based approach to integrating external data “on the fly”, and the use of a user-friendly interface based on a multi-modal database with polyglot persistence.

3. Initial Model and New Information Sources

As anticipated in the previous section, the starting point of this work is the linked data application for NENs described in [6]. The next sub-section summarizes the content and the structure of the previous knowledge model whereas Section 3.2 provides a brief overview of the new biomedical information sources that, in the present study, have been integrated on top of the former model.

3.1. The Initial Integration Model

The former model already integrates several existing biomedical ontologies and databases that describe diseases, genes, gene products, biological processes, molecular functions, and the gene–disease and disease–disease associations. The entire list of the integrated data sources is reported in Table 1. A single knowledge model has been built by extracting and linking relevant information for research on NENs.

Table 1. List of the biomedical information sources integrated with the former model.

Information Source	Format	Website
National Cancer Institute Thesaurus (NCIT)	OWL, OBO	ncithesaurus.nci.nih.gov
Orphanet Rare Disease Ontology (ORDO)	OWL	www.ebi.ac.uk/ols/ontologies/ordo
Disease Ontology (DO)	OWL, OBO	disease-ontology.org
Mondo Disease Ontology (MONDO)	OWL, OBO	mondo.monarchinitiative.org
Gene Ontology (GO)	OWL, OBO, CSV	geneontology.org
MedGen database	CSV	www.ncbi.nlm.nih.gov/medgen
DisGeNet database	RDF, CSV	www.disgenet.org

Our model relies on the formalism of *multilayer networks* to address the heterogeneity of interconnected information; this allows us to represent complexity by generalizing a graph structure where the nodes and edges are distributed on different layers, each representing an “aspect” of the domain [23].

We define a multilayer network as a triple $M = (V, E, L)$ where the sets V and E represent, respectively, the nodes and edges of the network, whereas $L = \{L_1, \dots, L_d\}$ is the set of network layers [24]. In turn, each $L_i \in L$ is a subgraph $L_i = (V_i, E_i)$ composed by the nodes V_i and the edges E_i such that $V = \cup_{i=1}^d V_i$ and $E = \cup_{i=1}^d E_i$.

In Figure 1 a sketch of the former model is depicted. The three interconnected layers represent information on diseases, affected genes, and their functions. We define specific concepts as shared nodes that realize bridges between layers. Layer 1 collects information about NENs as available in NCIT, ORDO, and DO ontologies. The *disease* class (and its subclasses) acts as a bridge from layer 1 to layer 2 to highlight the variations in the human genome (available in the DisGeNet database) that lead to the NENs described at the first level as well as disease–disease associations based on their molecular causes.

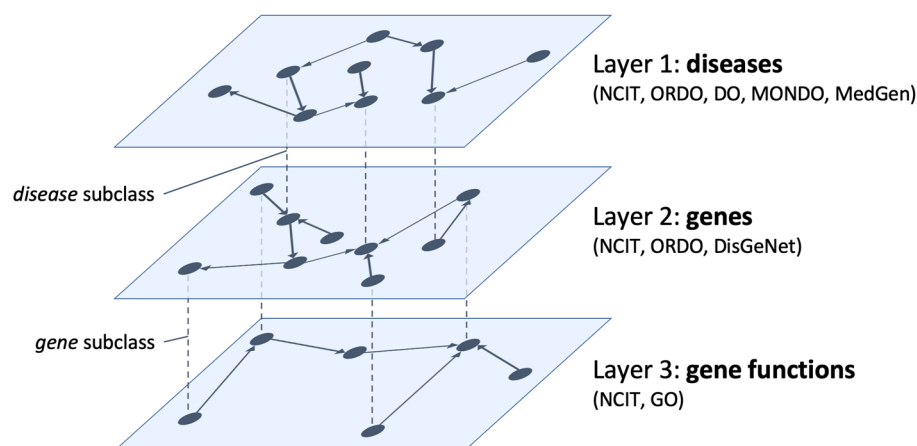


Figure 1. Visual representation of the initial integration model.

Then, the *gene* class (and its subclasses) acts as a bridge from layer 2 to layer 3 to highlight additional information on genes and gene products responsible for the onset of NENs (available in the *Gene Ontology*). This also includes genes' molecular functions (i.e., the elementary activities of a gene product at the molecular level, such as binding or catalysis) and biological processes (i.e., the operations or sets of events relevant to the operation of living units: cells, tissues, organs, and organisms).

We implement interlayer relations with the *equivalent-class* OWL statement. In the same way, newly defined ontologies are linked with the original ontological sources. We developed a client–server application to access, browse, and query the defined model.

3.2. The Additional Biomedical Information Sources

The list of additional data sources integrated with the updated knowledge model is reported in Table 2 which also includes a reference to the official website.

Table 2. List of additional biomedical information sources.

Information Source	Format	Website
Human Phenotype Ontology (HPO)	OWL, OBO	hpo.jax.org
HPO-ORDO Ontological Module (HOOM)	OWL	bioportal.bioontology.org/ontologies/HOOM
Human Protein Atlas (HPA)	TSV	www.proteinatlas.org
Drug–Gene Interaction Database (DGIdb)	TSV	www.dgldb.org
ChEMBL Database	RDF, SQL	www.ebi.ac.uk/chembl

The *Human Phenotype Ontology* (HPO) provides a standardized vocabulary of phenotypic abnormalities found in human diseases. It currently includes over 13,000 terms and more than 156,000 annotations describing phenotypic anomalies divided into five sub-ontologies that classify anomalies, link them to diseases, describe the mode of inheritance, the modifiers of clinical symptoms, the clinical course, and the frequency of specific clinical features. The ontological scheme is developed by the *Monarch Initiative*, using medical literature to improve biomedical research on rare diseases [25].

The *HPO-ORDO Ontological Module* (HOOM) is an ontology that integrates ORDO information on rare diseases with HPO information on phenotypic anomalies. HOOM qualifies the annotations between clinical entities and phenotypic anomalies according to the frequency and integrates the notion of diagnostic criterion. HOOM is intended for researchers and pharmaceutical companies wishing to co-analyze associations of rare and common disease phenotypes. Being designed to integrate the information of two different models, it does not contain instances but only classes and relationships.

The *Human Protein Atlas* (HPA) maps human proteins in cells, tissues, and organs using the integration of several omic technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems biology [26]. It has contributed to thousands of publications in the field of human biology and disease and is recognized by the intergovernmental organization ELIXIR as a central European resource for the life science community. HPA consists of ten parts, each focusing on a particular aspect of the genome-wide analysis of human proteins.

In particular, the *HPA Tissue Section* has been used in this work. This section describes the expression profiles in human tissues of genes at both the *mRNA* and protein levels. The protein expression data of 44 normal human tissue types is derived from antibody-based protein profiling using immunohistochemistry. Protein data covers 15,323 genes (i.e., 76% of protein-coding genes) for which antibodies are available. The *mRNA* expression data is derived from deep RNA sequencing (RNA-seq) from 256 different types of normal tissue.

The *Drug–Gene Interaction Database* (DGIdb) is a Web resource that provides information on drug–gene interactions and druggable genes from publications, databases, and other Web sources [27]. Data on drugs, genes, and interactions are normalized and merged into conceptual groups. In the current version (4.0), DGIdb includes 100,273 interaction statements and 33,577 druggable gene category claims. In total, it includes 10,606 druggable genes and 54,591 drug–gene interactions, covering 41,102 genes and 14,449 drugs. DGIdb is accessible via a Web-based search interface, an application programming interface (API), and is downloadable as a collection of TSV archives.

The *ChEMBL* database is a large open-access drug discovery database managed by the *European Molecular Biology Laboratory* (EMBL). It is handled manually and has the purpose of capturing data and knowledge across the pharmaceutical research and development process. Information on molecules and their biological activity is extracted from full-text articles in several journals and supplemented with data on approved drugs and clinical development candidates, such as the mechanism of action and therapeutic indications [28]. It includes information on more than 2 million compounds and 14,000 drugs from more than 84,000 publications and about 200 datasets. It is accessible via a Web-based interface and can be downloaded as an SQL database or a collection of RDF files.

4. The Integration and Harmonization Process

To integrate the biomedical data sources described in Section 3.2, three additional layers were designed and harmonized with the initial model. Figure 2 shows the updated model which now consists of six interconnected layers. The additional layers (opaque in the figure) contain information on the phenotypic anomalies connected to the NENs (layer 1b), the involved human tissues (layer 4), and the pharmacological interactions with the connected genes (layer 5). As in the original model, we don't include explicit interlayer connections since they can be inferred from the projections of the same node in different layers. Furthermore, each pair of adjacent layers shares at least one node.

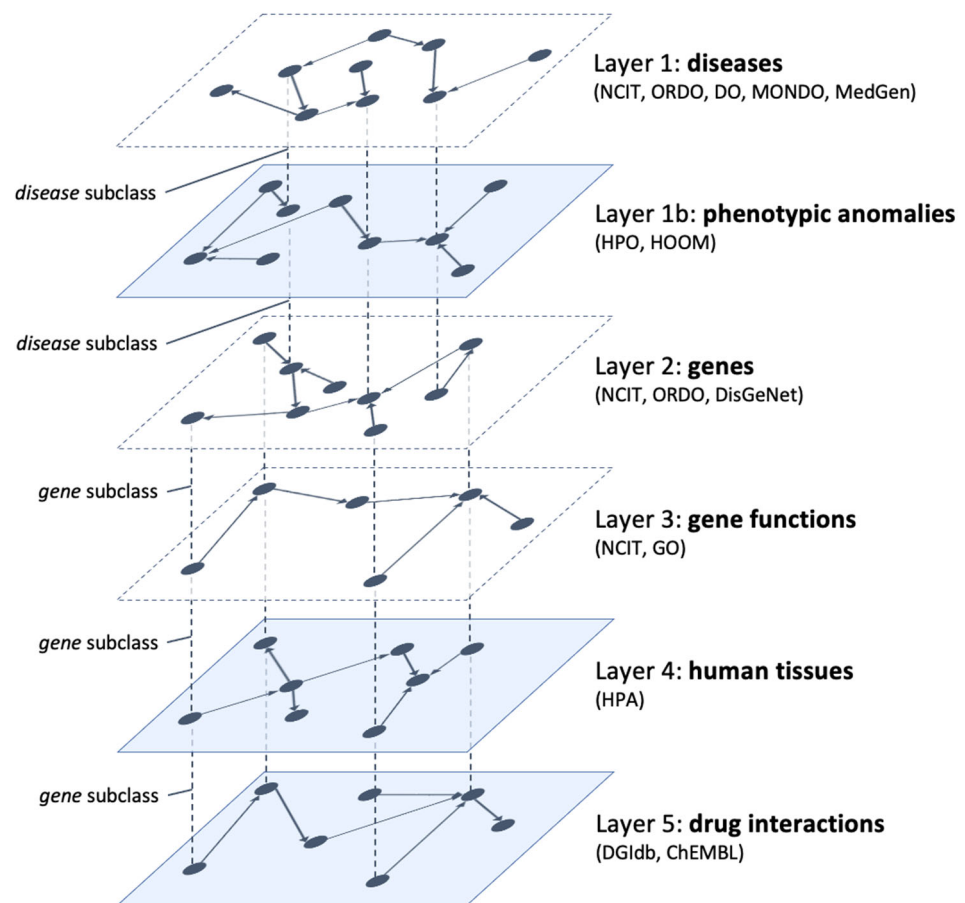


Figure 2. Visual representation of the updated integration model.

The new layer describing phenotypic anomalies, being closely related to diseases, is placed directly between layers 1 (diseases) and 2 (genes). In this way, the disease class (and its subclasses) allows a transition from layer 1 to layer 1b (to discover phenotypic anomalies related to diseases) and to layer 2 (to discover disease–gene connections). In turn, the gene class (and its subclasses) allows a transition from layer 2 to layer 3 (to discover gene functions), to layer 4 (to discover human tissues where gene products are expressed), and to layer 5 (to find drugs that have documented interactions with genes). In the next subsections, we describe in more detail the composition of each additional layer.

4.1. Layer 1b: Phenotypic Anomalies

As introduced in Section 2, NENs are rare diseases that include heterogeneous neoplasms such as high-grade NETs in the lung, mixed medullary and follicular cell carcinomas, intrathyroidal NENs with paraganglioma features, NENs of the breast, NETs in the kidney, NETs of the bladder, etc. [10]. Although a rare disease occurs in less than 1 in 2000 individuals, due to the high number of such diseases (about 8000 according to Orphanet), it is estimated that around 4% of the European population has a rare disease diagnosis [29].

According to [30], about 80% of rare diseases are of genetic origin. However, due to a lack of clinical and scientific knowledge, the molecular cause is unknown for about 40% of them. The second level of the proposed model already includes information on known genetic variants linked to NENs. However, despite the increasing number of identified genetic variants, their functional impact and, consequently, the connection with rare diseases is still largely unknown. Furthermore, even for diseases for which one or more causative genes have been identified, these often do not explain the totality of cases.

This lack of knowledge often prevents patients from receiving adequate and timely care. It is estimated that specific therapies are available for less than 10% of rare diseases,

including NENs. For this reason, having available detailed phenotypic data combined with ever-increasing amounts of genomic data is of enormous importance to accelerate the identification of clinically actionable prognostic or therapeutic implications and to improve the understanding of rare diseases [31]. Moreover, phenotype-based genomic analysis has also been shown to improve the diagnostic rate in patients with rare diseases [32].

Phenotypes are the observable traits of an organism. In medical contexts, however, the word phenotype is more often used to refer to some deviation from normal morphology, physiology, or behavior. A disease is commonly characterized by one or more phenotypic features which can affect all or only a subset of individuals with the disease as well as a time course over which the phenotypic features may have onset and evolve. The HPO ontology (see Section 3.2) describes a deep hierarchy of phenotypic abnormalities whereas the HOOM ontology (see Section 3.2) associates the phenotypic anomalies described in HPO with the clinical entities included in the ORDO ontology of rare diseases.

By harmonizing the information included in HPO and HOOM with the classes and properties of the first three layers of the initial knowledge model, we were able to construct layer 1b, aimed at describing known associations between NENs and phenotypic anomalies. Figure 3 represents the main classes and relations of our model. Gray classes and bold relationships are introduced by the integration scheme of layer 1b except for the dotted classes which are projected from the previous and subsequent layers.

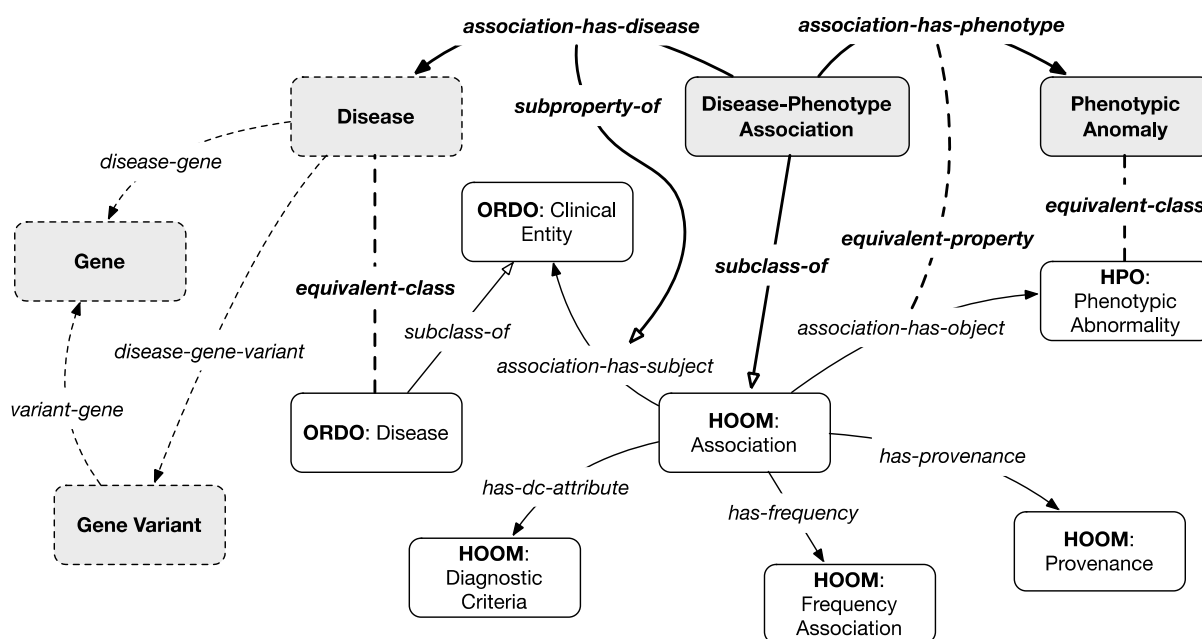


Figure 3. Relevant classes and relations of the additional layer 1b.

The main classes of this layer are *phenotypic anomaly*, equivalent to the *phenotypic abnormality* class from HPO (i.e., the ancestor class of all described phenotypic anomalies), and the *disease–phenotype association* class, a subclass of the *association* class from HOOM (describing known associations between phenotypes and clinical entities). Only associations related to diseases classified as NENs in the first layer have been considered. Furthermore, a set of properties, inherited from HOOM and described in Table 3, are linked to each association, to qualify it with information about the frequency and the diagnostic criteria, and the provenance (e.g., scientific articles or expert opinions).

Table 3. The *HOOM* sub-classes qualifying a disease–phenotype association.

Class	Sub classes	Description
Frequency Association	<i>Obligate</i>	The phenotypic abnormality is always present, and the diagnosis cannot be confirmed if it is absent
	<i>Very Frequent</i>	The phenotypic abnormality is present in 80% to 99% of cases
	<i>Frequent</i>	The phenotypic abnormality is present in 30% to 79% of cases
	<i>Rare</i>	The phenotypic abnormality is present in 5% to 29% of cases
	<i>Very Rare</i>	The phenotypic abnormality is present in 1% to 4% of cases
Diagnostic Criteria	<i>Criterion</i>	The phenotypic anomaly is used consensually to establish the clinical diagnosis
	<i>Exclusion</i>	The phenotypic anomaly allows to exclude the diagnosis
	<i>Pathognomonic</i>	The phenotypic anomaly is sufficient to undoubtedly establish the diagnosis
Provenance	...	Each subclass represents a set of scientific articles or expert advice qualifying the association

It should be noted that indirect connections between genes, gene variants, and phenotypic anomalies can be inferred from the model based on information from layer 2 that associates genes and gene variants with diseases.

4.2. Layer 4: Human Tissues

This layer enhances the model with information, inherent to NENs, on the human tissues associated with the genes that cause these diseases. This information is retrieved from the *Human Tissues* section of the HPA (see Section 3.2) which defines the distribution of gene products in the main tissues and organs of the human body. The collection and analysis of information relating to normal tissues are important and allow us to compare a pathological state with normality. On the other hand, inter-individual variations in the norm (e.g., age-related) can present a challenge in distinguishing a physiological condition from a pathological one.

The Human Tissues section of the HPA describes the level of expression of gene products in 44 different non-diseased tissues. These gene products and related genes play an important role in organ physiology and provide the basis for organ-specific research. By correlating genes and tissues it is possible to highlight genes that are simultaneously present in groups of tissues, compared to all other human tissues. Such information helps to find similar characteristics between different organs and allows us to elucidate the function of the genes associated with NENs.

Table 4 shows the main fields of the *HPA Normal Tissue Data Archive*. It is a tabular TSV file where each row represents the association between a gene and a human tissue. We filtered this extensive dataset to consider only the associations with a *medium* or *high* level of protein expression (*level* field), only for the subset of genes already included in layer 2 of the integrated model (therefore related to NENs). Then the value of the *reliability* field was considered. This value indicates the level of reliability of the analyzed protein expression pattern based on knowledge-based evaluation of available RNA-seq data, protein/gene characterization data, and immunohistochemical data from one or several antibodies designed toward non-overlapping sequences of the same gene. In our case, only the associations that are considered as *enhanced*, *supported*, or *approved* were retained whereas *uncertain* associations were discarded.

Table 4. The main fields of the *HPA Normal Tissue Data Archive*.

Field	Description	Example
gene	Gene identifier (<i>Ensembl</i> taxonomy)	ENSG00000000003
gene-name	Gene identifier (<i>HGNC</i> taxonomy)	TSPAN6
tissue	Name of a human tissue (controlled vocabulary of 44 terms)	breast
cell-type	Type of cells of the selected human tissue (controlled vocabulary)	glandular cells
level	Expression value of the gene within the tissue (<i>not detected, low, medium, high</i>)	high
reliability	Reliability of the expression value based on the evaluation of available data and literature (<i>enhanced, supported, approved, uncertain</i>)	approved

To harmonize the filtered information with the multi-layered knowledge model, an ontological representation of gene–tissue association is created as represented in Figure 4 where gray classes and bold relationships are introduced by the scheme of this layer except for the dotted classes which are projected from the previous layers. In particular, the classes *Tissues* and *Cell-Types* were introduced whose instances are taken from the *HPA-controlled* vocabularies.

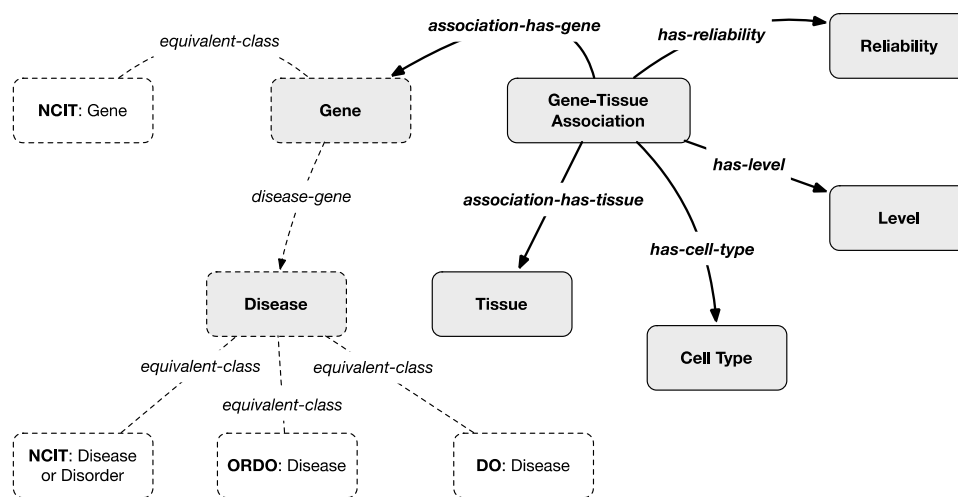


Figure 4. Relevant classes and relations of the additional layer 4.

A *Gene–Tissue Association* class was also introduced whose instances are dynamically generated from the filtered version of the HPA Normal Tissue Data Archive. Each instance maps a gene with a tissue and a cell type by also specifying the related level (*medium* or *high*, which are instances of the *Level* class) and reliability (*enhanced, supported, or approved* which are instances of the *Reliability* class). Even in this case, indirect connections between diseases and tissues can be inferred from the model based on information from the previous layers that associate genes and diseases.

4.3. Layer 5: Drug Interactions

NENs are biologically heterogeneous and contain subpopulations of cells with different angiogenic, invasive, and metastatic properties. As their response to therapeutic agents is equally heterogeneous, their treatment still represents an important clinical problem [33]. Understanding drugs’ effects on NENs has been importantly investigated in the last years also using *in vitro* studies that have been essential to clarifying drug mechanism of action. Some innovative therapeutic options are also based on the study of the molecular pathways involved in the development and growth of NENs.

In this context, the research has recently focused on the so-called *druggable genome* that is, genes and gene products known or expected to interact with bioavailable compounds. In addition to the presence of a protein structure that can be powerfully bound by small molecules, good potential targets are proteins for which modulation of biological function could provide therapeutic benefits for the patient. Targeted therapy has proven to be a

successful strategy in oncology, with the introduction of new therapeutic agents, including monoclonal antibodies and small molecule kinase inhibitors [34].

Following this trend, whereas the previous levels of the model offer researchers the ability to find mutated or altered genes implicated in NENs, the last level is designed to provide them with information on compounds and drugs that show documented interactions with these genes. The main external information sources integrated into this level are *ChEMBL* and *DGIdb* (see Section 3.2) describing, respectively, drugs and drug–gene interactions.

Table 5 shows the main fields of the *DGIdb Interactions Archive*. It is a tabular TSV file where each row represents a documented interaction between a drug and a gene. We filtered this information (made of more than 85,000 associations) to consider only interactions with genes already included in the model. The standardized *HGNC* (Human Genome Organization Gene Nomenclature Committee) gene name was used to associate the correct *Gene* class from layer 2. Instead, the *drug-concept-id* field was used to link the right instance of the *Substance* class from the *ChEMBL* ontology.

Table 5. The main fields of the *DGIdb Interactions Archive*.

Field	Description	Example
gene-name	Gene identifier (<i>HGNC</i> taxonomy)	ITGB5
entrez-id	Gene identifier (<i>NCBI Entrez</i> taxonomy)	3693
drug-name	Name of the interacting drug	CILENGITIDE
drug-concept-id	Drug identifier (<i>ChEMBL</i> taxonomy)	CHEMBL429876
interaction-types	Types of gene–drug interaction (controlled vocabulary, see Table 6)	inhibitor
interaction-score	The strength of the interaction calculated by multiplying the number of sources that report the interaction with its specificity (if the gene or the drug interacts with many other drugs or genes, the interaction specificity is low, otherwise it is high)	9.55

Table 6. Types of drug–gene interactions supported by layer 5.

Interaction Type	Description
<i>activator</i>	A drug activates a biological response from a target, although the mechanism by which it does so may not be understood
<i>agonist</i>	A drug binds to a target receptor and activates the receptor to produce a biological response
<i>allosteric modulator</i>	Drugs exert their effects on their protein targets via a different binding site than the natural (orthosteric) ligand site
<i>antagonist</i>	A drug blocks or dampens agonist-mediated responses rather than provoking a biological response itself upon binding to a target receptor
<i>antibody</i>	An antibody drug specifically binds the target molecule
<i>antisense oligonucleotide</i>	A complementary RNA drug binds to an mRNA target to inhibit translation by physically obstructing the mRNA translation machinery
<i>inducer</i>	The drug increases the activity of its target enzyme
<i>inhibitor</i>	The drug binds to a target and decreases its expression or activity
<i>inhibitory allosteric modulator</i>	The drug will inhibit activity of its target enzyme
<i>inverse agonist</i>	A drug binds to the same target as an agonist but induces a pharmacological response opposite to that of the agonist
<i>modulator</i>	The drug regulates or changes the activity of its target, but it may not involve any direct binding to the target
<i>negative modulator</i>	The drug negatively regulates the amount or activity of its target, but it may not involve any direct binding to the target
<i>partial agonist</i>	A drug will elicit a reduced amplitude functional response at its target receptor, as compared to the response elicited by a full agonist
<i>positive modulator</i>	The drug increases activity of the target enzyme
<i>suppressor</i>	The drug directly or indirectly affects its target, suppressing a physiological process
<i>vaccine</i>	The drugs stimulate or restore an immune response to their target

To harmonize the filtered information with the multi-layered knowledge model, an ontological representation of drug–gene interaction is created as represented in Figure 5

where gray classes and bold relationships are introduced by the scheme of this layer except for the dotted classes which are projected from the previous layers. In particular, the new class *Drug* maps, through the *equivalent-class* relation, the external general *Substance* class from *ChEMBL*. The class *Gene–Drug Interaction* maps a row of the *DGIdb Interactions Archive* and connects the *Gene* class with the *Drug* class with the *interaction-has-gene* and the *interaction-has-drug* object properties, respectively.

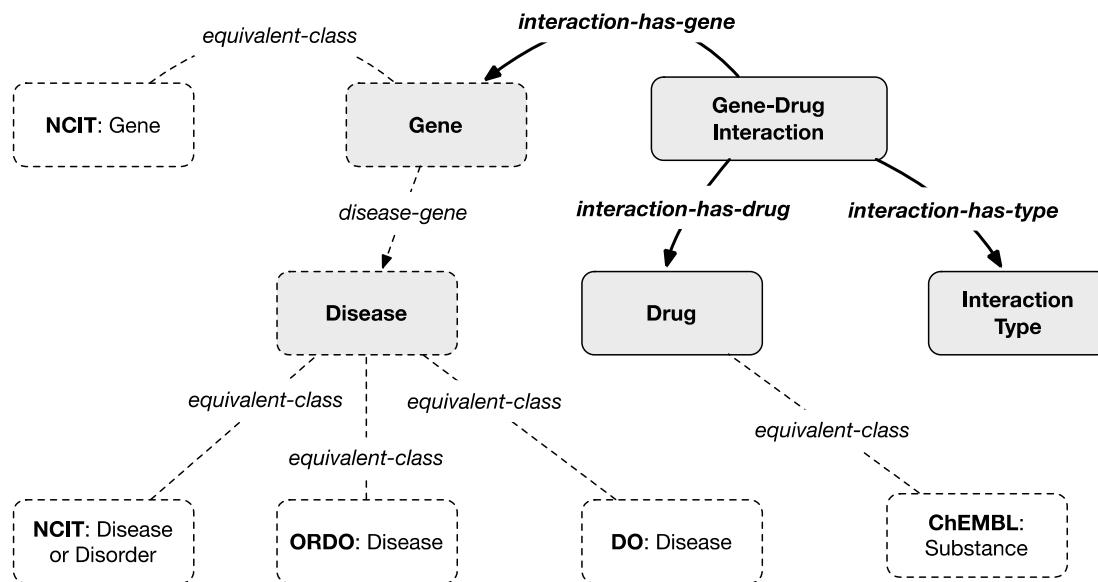


Figure 5. Relevant classes and relations of the additional layer 5.

Additional information, connected to the *Gene–Drug Interaction* class through the *interaction-has-type* property, is the *Interaction Type* that explains the way a drug or compound interacts with a gene according to a controlled vocabulary defined by *DGIdb* as reported in Table 6. Each term is represented within the model as an individual of the *Interaction Type* class. The meaning of each term is explained in the same table. Additional information like the interaction score (see Table 5) and a link to the *ChEMBL* Web page describing in detail each substance is included in the model through data properties.

5. Developed Prototype and Validation Results

A client–server application was developed as an extension of the one already presented in [6] to retrieve the relevant information from our integrated model. We selected *Virtuoso Universal Server* as the middleware to store the original biomedical ontologies and databases; such an open-source solution allows us to manage different data formats with several access protocols. We store the original datasets to the server, scheduling periodical updates starting from the original endpoints (not straightly used for performance issues). Our multilayer integration model is also hosted on the same server.

We developed a lightweight Java desktop application to allow quick and easy user interaction. A visual interface is provided to specify input queries, which are in turn translated as SPARQL sequences and forwarded via HTTP to the server. Obtained results are shown graphically to the user. For RDF and SPARQL management, our client application relies on the Jena Framework. We also use *OWL API* for the client-side manipulation of OWL ontologies. Figure 6 summarizes our system architecture highlighting the main modules. Notice that the dashed dataset comes from the previous version.

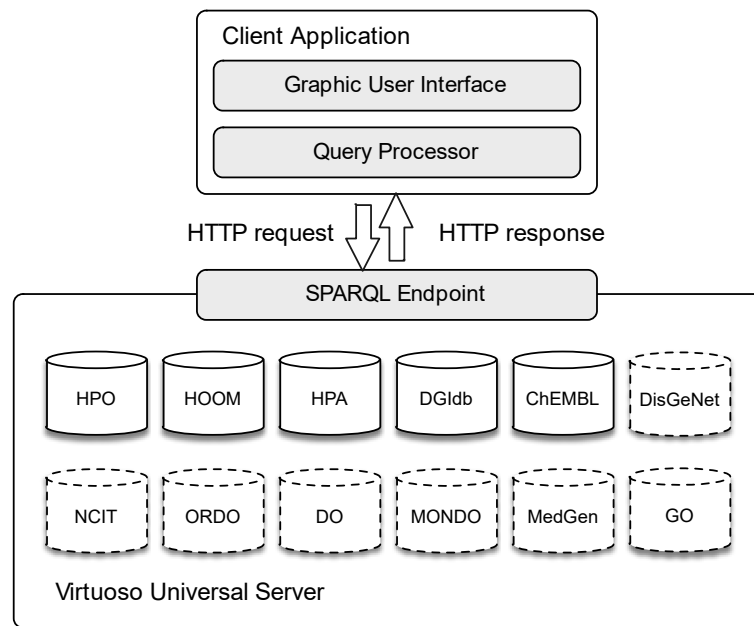


Figure 6. Client–server architecture of the developed prototype.

Figure 7 shows the “phenotypes” section of the client application. Once the user has selected a subset of NETs and/or NECs from the “diseases” area—see (Capuano, Foggia, Greco, and Ritrovato, 2022)—this section allows him to obtain the phenotypic anomalies associated with the selected diseases. The user can decide whether to carry out the analysis only on the previously selected diseases or not. The information obtained in this phase is extracted from layer 1b (see Section 4.1) starting from the rare diseases present in the ORDO ontology, and then extrapolating the associations with the anomalies in ORDO-HOOM and finally the information relating to the anomalies in HPO. The user can click on the *name* of the anomaly and on the *frequency* value to get more information.

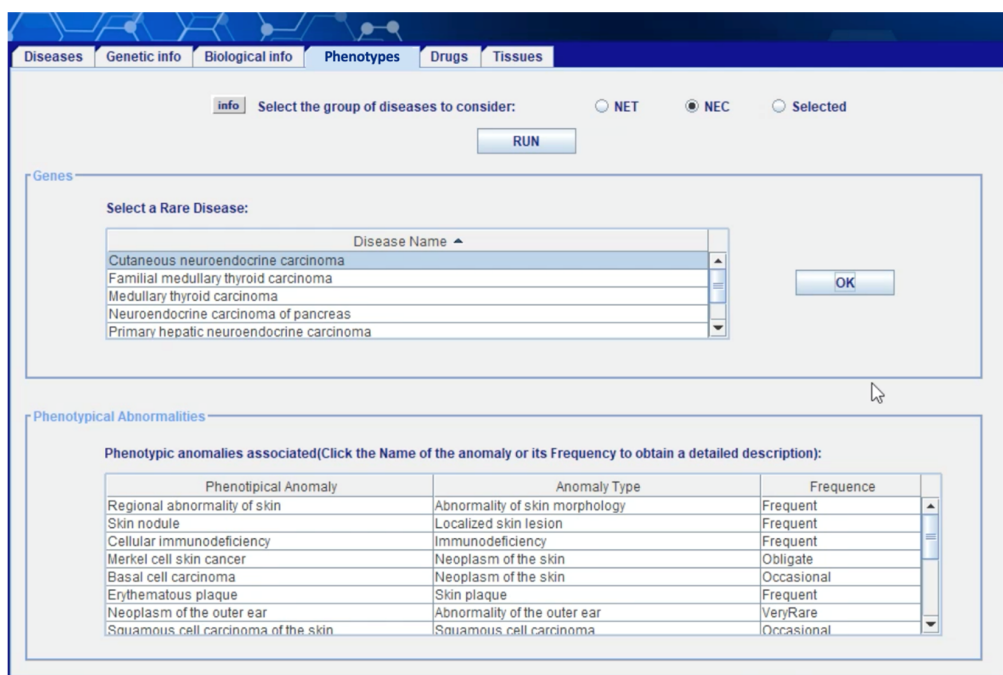


Figure 7. The “Phenotypes” section of the client application.

Figure 8 shows the “tissue” section of the client application. Information about the human tissues in which a neoplasm can occur is extracted from layer 4 (see Section 4.2) starting from the *Normal Tissue Data Archive* of HPA. Once the user has selected a subset of genes involved in NETs and/or NECs under investigation (including gene–disease, variation–disease, and disease–disease associations, cytogenetics anomalies, molecular anomalies, etc.)—through the “genetic information” tab described in (Capuano, Foggia, Greco, and Ritrovato, 2022)—he can obtain here the list of human tissues in which each gene has a medium or high level of expression. Displayed data on each tissue is associated with a link to the online version of HPA where the user can find additional information (see Figure 9).

The screenshot shows the 'Tissues' section of a web application. At the top, there are navigation tabs: Diseases, Genetic info, Biological info, Phenotypes, Drugs, and Tissues. The 'Tissues' tab is active. Below the tabs, there are three main sections:

- Load File:** A dropdown menu is set to 'hpa tissues.owl'. Below it, text reads 'Human Protein Atlas File in TSV Format (see https://www.proteinatlas.org/about/download)'. A 'Start' button is to the right.
- Genes:** A section with an 'info' icon and the text 'Select a group of diseases to consider:'. There are three radio buttons: 'NET' (selected), 'NEC', and 'Selected'. A 'RUN' button is to the right. Below this is a table titled 'Select a Gene:'.

Disease Name	Gene
Digestive System Neuroendocrine Tumor	GATA3
Digestive System Neuroendocrine Tumor G1	GATA4
Pancreatic Glucagonoma	GCG
Glucagonoma	GCG
Pancreatic Glucagonoma	GCGR

 Below the table is a dropdown menu for 'Select tissue information to get:' set to 'High Expression Tissues' and an 'OK' button.
- Tissues Information:** A section with the text 'Information related to Gene:'. Below it is a table:

Tissue	Clickable URI
pancreas	https://www.proteinatlas.org/learn/dictionary/normal/pancreas

Figure 8. The “Tissues” section of the client application.

Figure 10 shows the “drugs” section of the client application where the user can obtain information on compounds and drugs that show documented interactions with the genes selected in one of the preceding steps and, as consequence, can potentially impact the associated diseases. The information obtained in this phase is extracted from layer 5 (see Section 4.3) starting the integration of gene information included in GO with information on drugs included in ChEMBL and drug–gene interactions included in DGIdb. Displayed data on each drug is associated with a link to the online version of ChEMBL where the user can find additional information (see Figure 11).

A test installation of the server was conducted on a *Linux* machine with a 2.3 GHz quad-core Intel Core i7 processor and 16 Gb of RAM. With this hardware configuration, most of the queries are answered in a fraction of a second, and only the most complex ones (that combine information from semantic and non-semantic sources) require longer: in rare cases more than 2 s. These results are in line with recent RDF store benchmarks [35] which rank Virtuoso Universal Server as one of the fastest triple stores for both instant and analytical queries.

THE HUMAN PROTEIN ATLAS

MENU HELP NEWS

HUMAN PROTEIN ATLAS SECTIONS : TISSUE > PANCREAS

Protein expression of genes elevated in the pancreas

In-depth analysis of the elevated genes in pancreas using antibody-based protein profiling allowed us to visualize the expression patterns of these proteins in different compartments including islets of Langerhans, exocrine glands and ducts.

Proteins specifically expressed in islet cells of pancreas

The islet cells constitute 2% of the pancreas and are responsible for maintaining a steady blood glucose level by secreting hormones regulating uptake and release of glucose. Examples of proteins expressed in islet cells include *INS*, which is secreted following elevated blood glucose levels and stimulates glucose uptake upon binding insulin receptor, and, *GCG*, which elicits an opposite effect by raising blood glucose levels. Another important protein is *IAPP*, a hormone that regulates glucose metabolism and acts as a satiation signal.

The pancreas-specific proteome

- The pancreas transcriptome
- Protein expression of genes elevated in the pancreas
- Gene expression shared between pancreas and other tissues

Pancreas function

Pancreas histology

Background

Relevant links and publications

Dictionary
Pancreas

INS **GCG** **IAPP**

Figure 9. Additional online information from HPA linked to the “tissues” section.

Diseases Genetic info Biological info Phenotypes **Drugs** Tissues

info Select the group of diseases to consider: NET NEC Selected

RUN

Genes

Select a Gene to get Drugs:

Disease Name	Gene
lung small cell carcinoma	CHRNA4
lung small cell carcinoma	CNTNAP2
lung small cell carcinoma	COL11A2
lung small cell carcinoma	COL22A1
lung small cell carcinoma	COL4A2

OK

Drugs

Drugs associated to Gene (Click Synonyms of a Drug to obtain a detailed description):

Drug	Drug Type	Synonyms	Clickable URI
Nicotine	SmallMoleculeDrug	(â™)nicotine ; (S)-(â™)nicotine ; (S)-3-(1-met...	https://go.drugbank.com/drugs/DB00184
Fluoxetine	SmallMoleculeDrug	*(+)-N-Methyl-3-phenyl-3-(alpha;	https://go.drugbank.com/drugs/DB00472
Dextromethorphan	SmallMoleculeDrug	(+)-dextromethorphan ; D-methorphan ; delta...	https://go.drugbank.com/drugs/DB00514
Galantamine	SmallMoleculeDrug	(-)-Galantamine ; Galantamina ; Galantamin...	https://go.drugbank.com/drugs/DB00674
Ethanol	SmallMoleculeDrug	*1-Hydroxyethane ; Alcohol ; Alcohol (ethyl) ; A...	https://go.drugbank.com/drugs/DB00898
Pentolinium	SmallMoleculeDrug	Pentolineum ; Pentolonium ; Pentolinium;	https://go.drugbank.com/drugs/DB01090
Levacetylmethadol	SmallMoleculeDrug	*(-)-alpha-Acetylmethadol ; (1S;	https://go.drugbank.com/drugs/DB01227

Figure 10. The “Drugs” section of the client application.

ChEMBL Search in ChEMBL

Compound Report Card

Name And Classification

ID: CHEMBL52440
Name: DEXTROMETHORPHAN
Max Phase: 4 Approved
Molecular Formula: C18H25NO
Molecular Weight: 271.40

ChEMBL Synonyms: Dextromethorphan, DEXTROMETHORPHAN, Dimetane-DX, NSC-751452, C100003, C10003, C-10003, CTP-786, D6-DEXTROMETHORPHAN, d6-DM, DEUDEXTROMETHORPHAN, DEXTROMETHORPHAN D6, DEXTROMETHORPHAN HYDROBROMIDE

Synonyms From Alternate Forms: Dextromethorphan Hydrobromide Monohydrate, NSC-756723, Dextromethorphan polistirex, DEXTROMETHORPHAN POLISTIREX, d6-DM hydrobromide, D6-DM HYDROBROMIDE MONOHYDRATE, DEUDEXTROMETHORPHAN HYDROBROMIDE

Trade Names From Alternate Forms: BEECHAM COUGHCAPS, Benylin DM, BRONTYL, CONTAC COUGHCAPS, DELSYM, Dextromethorphan Hydrobromide OROS Tablets, Drixoral Cough, FRANOLYN DRY, FRANOLYN SED, PediaCare 1, Romilar, TUSSABRON, TUXIUM, Delsym, DEXTROMETHORPHAN POLISTIREX

Molecule Type: Small molecule

Structure Search

- Name And Classification
- Representations
- Sources
- Alternative Forms
- Molecule Features
- Drug Indications
- Drug Mechanisms
- Clinical Data
- Similar Compounds
- Metabolism
- Activity Charts
- Literature
- Target Predictions
- Calculated Properties
- Cross References
- UniChem Cross References
- UniChem Connectivity Layer Cross References

Figure 11. Additional online information from ChEMBL linked to the “drugs” section.

Our system validation was performed with the contribution of a domain expert that helped verify the consistency and correctness of the ontological knowledge as well as the quality of the alignment between the information sources. We adopted an iterative approach where the expert was asked to use the system and provide feedback; this allowed us to improve the level of alignment [36]. In the specific case, two validation iterations led to satisfactory results.

6. Conclusions and Further Work

In this paper, we have described the extension of previous research work aimed at designing and implementing a domain-specific linked data application for integrating relevant biomedical information on NENs. Additional biomedical aspects covered by the updated model include the phenotypic anomalies linked to these diseases, the involved human tissues, and the documented pharmacological interactions of existing drugs and compounds. Through the alignment and the integration of existing semantic and non-semantic biomedical sources, we were able to compose a knowledge base as a *multilayer network* managed through a multi-model database providing polyglot persistence. The model can be easily navigated and queried using a client application that provides a user-friendly graphical interface.

Several directions of extension of the proposed system can be envisaged. On the one hand, there is the possibility of integrating additional information sources within existing layers (integrating aspects already considered) as well as on additional layers (considering further aspects). On the other hand, it would be possible to apply the model to a connected biomedical domain, for example considering a different subset of rare diseases. The experimentation of the proposed system with researchers and professionals involved in the treatment of this type of neoplasms is also foreseen to collect feedback for system improvement as well as to assess the risks and critical success factors associated with the introduction of this kind of technology in real medical contexts [37].

Beyond the specific domain, the paper introduces and analyzes a way to integrate heterogeneous data sources, capable of being adapted to other contexts. Linked data makes the possible aggregation of information quite unlimited. Each information level can be enriched with further details so that the system becomes increasingly useful for

user support. On the other hand, the multi-layer organization would help to deal with the variety of information in a more organized and governable way.

Another promising research direction is the application of existing metrics, such as those defined in [38], to measure the quality of the integrated knowledge model in terms of *relationship richness*, *attribute richness*, *inheritance richness*, etc. Indeed, evaluating such metrics on the integrated model, which includes ontological and non-ontological information, could be an interesting but challenging task that would require revising the definition of such metrics to support hybrid models. Moreover, approaches to automatic ontology alignment could be investigated and incorporated into the proposed system as integrated schemas.

Author Contributions: Conceptualization, P.R. and P.F.; methodology, N.C. and L.G.; software, N.C. and L.G.; validation, P.R. and P.F.; formal analysis, N.C. and L.G.; writing-original draft preparation, N.C. and L.G.; writing-review and editing, N.C. and P.R.; supervision, P.F.; funding acquisition, P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the *RarePlatNet* project on *diagnostic and therapeutic innovations for neuroendocrine and endocrine tumors and glioblastoma through an integrated technological platform of clinical, genomic, ICT, pharmacological, and pharmaceutical skills*, funded by the Campania region, Italy under the grant on POR CAMPANIA FESR 2014/2020, axis 1, objective 1.2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wilkinson, M.; Dumontier, M.; Aalbersberg, I.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.; de Santos, L.B.; Bourne, P.E.; et al. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]
2. Hancock, J. Editorial: Biological ontologies and semantic biology. *Front. Genet.* **2014**, *5*, 18. [[CrossRef](#)] [[PubMed](#)]
3. Wang, Z.; He, Y. Precision omics data integration and analysis with interoperable ontologies and their application for COVID-19 research. *Brief Funct. Genom.* **2021**, *20*, 235–248. [[CrossRef](#)] [[PubMed](#)]
4. Messaoudi, C.; Fissoune, R.; Hassan, B. A Survey of Semantic Integration Approaches in Bioinformatics. *Int. J. Comput. Inf. Eng.* **2016**, *10*, 2058–2063.
5. Kamdar, M. Mining the Web of Life Sciences Linked Open Data for Mechanism-Based Pharmacovigilance. In Proceedings of the WWW'18: Companion Proceedings of the The Web Conference; ACM: New York, NY, USA, 2018; pp. 861–865.
6. Capuano, N.; Foggia, P.; Greco, L.; Ritrovato, P. A semantic framework supporting multilayer networks analysis for rare diseases. *Int. J. Semant. Web Inf. Syst.* **2022**, in press. [[CrossRef](#)]
7. Hammoud, Z.; Kramer, F. Multilayer networks: Aspects, implementations, and application in biomedicine. *Big Data Anal.* **2020**, *5*, 2. [[CrossRef](#)]
8. Faria, D.; Pesquita, C.; Mott, I.; Martins, C.; Couto, F.; Cruz, I. Tackling the challenges of matching biomedical ontologies. *J. Biomed. Semant.* **2018**, *9*, 1–19. [[CrossRef](#)]
9. Effraimidis, G.; Knigge, U.; Rossing, M.; Oturai, P.; Rasmussen, A.; Feldt-Rasmussen, U. Multiple endocrine neoplasia type 1 (MEN-1) and neuroendocrine neoplasms (NENs). *Semin. Cancer Biol.* **2021**, *79*, 141–162. [[CrossRef](#)]
10. Rindi, G.; Klimstra, D.; Abedi-Ardekani, B.; Asa, S.; Bosman, F.; Brambilla, E.; Scarpa, A.; Scoazec, J.; Travis, W.D.; Tallini, G.; et al. A common classification framework for neuroendocrine neoplasms: An International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. *Mod. Pathol.* **2018**, *31*, 1770–1786. [[CrossRef](#)]
11. Nagtegaal, I.; Odze, R.; Klimstra, D.; Paradis, V.; Rugge, R.; Schirmacher, P.; Washington, K.M.; Carneiro, F.; Cree, I.; The WHO Classification of Tumours Editorial Board. The 2019 WHO classification of tumours of the digestive system. *Histopathology* **2019**, *76*, 182–188. [[CrossRef](#)]
12. Guérin, E.; Marquet, G.; Burgun, A.; Loréal, O.; Berti-Equille, L.; Leser, U.; Moussouni, F. Integrating and Warehousing Liver Gene Expression Data and Related Biomedical Resources in GEDAW. In Proceedings of the 2nd International Workshop on Data Integration in the Life Sciences, San Diego, CA, USA, 20–22 July 2005; pp. 158–174.
13. Belleau, F.; Nolin, M.; Tourigny, N.; Rigault, P.; Morissette, J. Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *J. Biomed. Inform.* **2008**, 706–716. [[CrossRef](#)] [[PubMed](#)]
14. Bizer, C.; Heath, T.; Berners-Lee, T. Linked Data—The Story So Far. *Int. J. Semant. Web Inf. Syst.* **2009**, *5*, 1–22. [[CrossRef](#)]
15. Livingston, K.; Bada, M.; Baumgartner, W.; Hunter, L. KaBOB: Ontology-based semantic integration of biomedical databases. *BMC Bioinform.* **2015**, *16*, 126. [[CrossRef](#)] [[PubMed](#)]

16. Masseroli, M.; Canakoglu, A.; Ceri, S. Integration and querying of genomic and proteomic semantic annotations for biomedical knowledge extraction. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2016**, *13*, 209–219. [[CrossRef](#)]
17. Childs, L.; Mamlouk, S.; Brandt, J.; Sers, C.; Leser, U. SoFIA: A data integration framework for annotating high-throughput datasets. *Bioinformatics* **2016**, *37*, 2590–2597. [[CrossRef](#)]
18. Galeota, E.; Kishore, K.; Pelizzola, M. Ontology-driven integrative analysis of omics data through Onassis. *Sci. Rep.* **2020**, *10*, 703. [[CrossRef](#)]
19. Bensch, W.; Borys, D.; Fujarewicz, K.; Herok, K.; Jaksik, R.; Krasucki, M.; Kurczyk, A.; Matusik, K.; Mrozek, D.; Ochab, M.; et al. Integrated system supporting research on environment related cancers. In *Recent Developments in Intelligent Information and Database Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 339–409.
20. Seneviratne, O.; Rashid, S.; Chari, S.; McCusker, J.; Bennett, K.; Hendler, J.; McGuinness, D. Knowledge Integration for Disease Characterization: A Breast Cancer Example. In Proceedings of the 17th International Semantic Web Conference, Monterey, CA, USA, 8–12 October 2018; pp. 223–238.
21. Zhang, H.; Guo, Y.; Prospero, M.; Bian, J. An ontology-based documentation of data discovery and integration process in cancer outcomes research. *BMC Med Inform. Decis. Mak.* **2020**, *20*, 292. [[CrossRef](#)]
22. Hong, N.; Chang, F.; Ou, Z.; Wang, Y.; Yang, Y.; Guo, Q.; Ma, J.; Zhao, D. Construction of the cervical cancer common terminology for promoting semantic interoperability and utilization of Chinese clinical data. *BMC Med Inform. Decis. Mak.* **2021**, *21*, 309. [[CrossRef](#)]
23. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.; Moreno, Y.; Porter, M. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271. [[CrossRef](#)]
24. Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C.; Gómez-Gardeñes, J.; Romance, M.; Sendiña-Nadal, J.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122. [[CrossRef](#)]
25. Robinson, P.; Köhler, S.; Bauer, S.; Seelow, D.; Horn, D.; Mundlos, S. The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease. *Am. J. Hum. Genet.* **2008**, *83*, 610–615. [[CrossRef](#)] [[PubMed](#)]
26. Thul, P.; Lindskog, C. The human protein atlas: A spatial map of the human proteome. *Tools Protein Sci.* **2018**, *27*, 233–244. [[CrossRef](#)] [[PubMed](#)]
27. Freshour, S.; Kiwala, S.; Cotto, K.; Coffman, A.; McMichael, J.; Song, J.; Griffith, M.; Griffith, O.L.; Wagner, A.H.; Wagner, A.; et al. Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts. *Nucleic Acids Res.* **2021**, *49*, D1144–D1151. [[CrossRef](#)] [[PubMed](#)]
28. Mendez, D.; Gaulton, A.; Bento, A.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *45*, D930–D940. [[CrossRef](#)] [[PubMed](#)]
29. Maiella, S.; Olry, A.; Hanauer, M.; Lanneau, V.; Loughi, H.; Donadille, B.; Rodwel, C.; Köhler, S.; Seelow, D.; Jupp, S.; et al. Harmonising phenomics information for a better interoperability in the rare disease field. *Eur. J. Med Genet.* **2018**, *61*, 706–714. [[CrossRef](#)]
30. Aitken, M.; Kleinrock, M.; Muñoz, E.; Porwal, U. *Orphan Drugs in the United States, Rare Disease Innovation and Cost Trends through 2019*; IQVIA Institute for Human Data Science: Parsippany, NJ, USA, 2020.
31. Boycott, K.; Rath, A.; Chong, J.; Hartley, T.; Alkuraya, F.; Baynam, G.; Brookes, A.J.; Brudno, M.; Carracedo, A.; den Dunnen, J.T.; et al. International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* **2017**, *100*, 695–705. [[CrossRef](#)]
32. Wright, C.; Fitzgerald, T.; Jones, W.; Clayton, S.; McRae, J.; van Kogelenberg, M.; King, D.A.; Ambridge, K.; Barrett, D.M. Genetic diagnosis of developmental disorders in the DDD study: A scalable analysis of genome-wide research data. *Lancet* **2015**, *385*, 1305–1314. [[CrossRef](#)]
33. Fidler, I.; Ellis, L. Chemotherapeutic drugs—More really is not better. *Nat. Med.* **2000**, *6*, 500–502. [[CrossRef](#)]
34. Dupont, C.; Riegel, K.; Pompaiah, M.; Juhl, H.; Rajalingam, K. Druggable genome and precision medicine in cancer: Current challenges. *FEBS J.* **2021**, *288*, 6142–6158. [[CrossRef](#)]
35. Atemez, G.; Amardeilh, F. Benchmarking commercial RDF stores with publications office dataset. In Proceedings of the European Semantic Web Conference; Springer: Cham, Switzerland, 2018; pp. 379–394.
36. Dragisic, Z.; Ivanova, V.; Lambrix, P.; Faria, D.; Jiménez-Ruiz, E.; Pesquita, C. User validation in ontology alignment. In Proceedings of the International Semantic Web Conference; Springer: Cham, Switzerland, 2016; pp. 200–217.
37. Vicedo, P.; Gil-Gómez, H.; Oltra-Badenes, R.; Guerola-Navarro, V. A bibliometric overview of how critical success factors influence on enterprise resource planning implementations. *J. Intell. Fuzzy Syst.* **2020**, *38*, 5475–5487. [[CrossRef](#)]
38. Tartir, S.; Arpinar, I. Ontology evaluation and ranking using OntoQA. In Proceedings of the International Conference on Semantic Computing (ICSC 2007), Irvine, CA, USA, 17–19 September 2007; pp. 185–192.