Interactive Exploration and Collaborative Curation of an Industry-Scale Healthcare Knowledge Graph Using the WebProtégé Cloud-Based Editor

Maulik R. Kamdar, Ph.D.[1,*], Matthew Horridge, Ph.D.[2,*], Linda Wogulis, B.S.[1], Cailey Fitzgerald, M.S.[1], Josef Hardi, M.S.[2], Doug Anderson, M.A.[1], Katie Scranton, Ph.D.[1], Rafael S. Gonçalves, Ph.D.[2], Dru Henke, B.S.[1], Mevan Samarasinghe, M.S.[1], Mark A. Musen, M.D. Ph.D.[2]

[1]Elsevier Health Markets, Elsevier Inc., Philadelphia, PA

[2]Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA

[*]Both authors contributed equally to this research

ABSTRACT

Knowledge graphs are developed and used in academia and industry to tackle complex challenges in healthcare and biomedical research. Elsevier's Healthcare Knowledge Graph (HG) is developed by extracting and integrating medical knowledge from several heterogeneous sources such as clinical guidelines, medical textbooks, and legacy databases. The HG platform powers Elsevier's clinical search and decision support applications that are used by medical professionals for education, research, and patient care. Hence, the medical knowledge within HG must be trustworthy, updated, and manually vetted by medical subject matter experts (SMEs). However, the size and complexity of HG, as well as the steep learning requirements towards technologies and languages used for knowledge representation and querying pose several challenges for medical SMEs to curate HG. An academic–industry R&D collaboration explored the use of the WebProtégé cloud-based knowledge editing system to facilitate medical SMEs and other stakeholders to interactively explore and collaboratively curate HG. In this case report, we present our findings and experiences from this collaboration and showcase the approaches and methods developed to handle the size and complexity of HG within WebProtégé . We also demonstrate the novel features and modifications made to the editing, browsing, and search interfaces within WebProtégé to improve the usability and user experience of the system for medical SMEs. While an ideal 'one size fits all' knowledge editing system does not exist, establishing the R&D collaboration and deciding on a fixed set of requirements and priorities enabled us to quickly develop stakeholder-focused customizations within existing processes and WebProtégé .

1 INTRODUCTION

Clinicians always require access to accurate, updated, and trustworthy medical knowledge provided by renowned medical and scientific organizations. This medical knowledge is often disseminated through patient guidelines, medical textbooks, clinical journals, and synoptic overviews. This requirement manifests at several different stages for a clinician: *i)* during medical education, when a clinician learns about the different diseases, clinical findings, procedures, and treatments, *ii)* during research phases, when a clinician searches reference literature and scientific journals for novel information pertaining to a given disease or a drug, and *iii)* during the diagnosis, prognosis, and treatment of patients at the point of care. Access to trusted and accurate medical literature is also a requirement for other medical professionals (e.g., nurses, pharmacists) and even for patients who wish to learn about their conditions.

Medical professionals must search and synthesize information on focused, yet esoteric, questions from a broad set of literature sources (textbooks, guidelines, journal articles) during the course of a busy practice using search engines and information systems (e.g., UpToDate[1], Dynamed[2], ClinicalKey[3], PubMed[4]). However, medical knowledge continuously changes, evolves, and shifts as new discoveries are made. Whereas in 2010, medical knowledge was doubling only every 3.5 years, medical knowledge was projected to double every 73 days in 2020, and this doubling time is definitely decreasing as we advance in the 21^{st} century with the torrent of new medical knowledge published daily during the COVID-19 pandemic[5,6]. Clinicians need to stay updated on the latest FDA approved drugs, ongoing clinical trials, and practice guidelines from renowned medical associations. Moreover, the advent of affordable next generation sequencing technologies has resulted in the need for knowledge personalized to an individual or cohort of patients. The ever-growing volume of medical evidence, lack of awareness of which resource to search for specialty questions, skepticism and lack of trust regarding the quality of search results, and insufficient time are often cited as some of the main barriers to medical learning and point of care literature discovery[7,8].

Knowledge graphs are increasingly being developed and leveraged in academia and industry to tackle complex healthcare and biomedical challenges, such as drug discovery and safety, medical literature search, clinical decision support, and disease monitoring and management[9–14]. At Elsevier, we have developed Elsevier's Healthcare Knowledge Graph (HG) as a platform that enables enhanced content discovery for medical professionals through search, browsing, recommendation, and decision support services[14,15]. Since these services are often used by clinicians and other professionals for patient care, medical education, and research, the knowledge platform that is used to power these services must provide actionable, trusted and regularly updated medical knowledge. The curation and maintenance of the content in a large and complex knowledge graph would require both medical subject matter expertise and technical knowledge of the underlying graph technologies. In 2019–2020, an academic–industry research collaboration between the Elsevier Health Markets and the Stanford Center of Biomedical Informatics Research explored the use of the WebProtégé cloud-based knowledge editing system to facilitate medical subject matter experts (SMEs) to interactively explore and collaboratively curate Elsevier's Healthcare Knowledge Graph.

In the following sections, we provide background on Elsevier's Healthcare Knowledge Graph and the WebProtégé knowledge editor. We introduce the approaches developed to ensure that the WebProtégé knowledge editor can handle the scale and complexity of HG for editing and visualization purposes. Finally, we showcase the novel features developed and the lessons learnt through this collaboration to improve curation of industry-scale knowledge graphs through interactive and intuitive exploration interfaces.

2 THE NEED FOR KNOWLEDGE CURATION – A CASE STUDY ON ELSEVIER'S HEALTHCARE KNOWLEDGE GRAPH

2.1 Background on Elsevier's Healthcare Knowledge Graph

Elsevier's Healthcare Knowledge Graph (HG) consists of medical knowledge and data, integrated from heterogeneous healthcare sources such as clinical guidelines, scientific journals, medical textbooks, and legacy databases. HG uses popular linked data and semantic web technologies, such as the Resource Description Framework (RDF) and the JavaScript Object Notation for Linked Data (JSON-LD) for capture and representation of information, and the SPARQL Protocol and RDF Query Language (SPARQL) for querying and retrieval of information. There is a vast depth of literature that provide a preliminary understanding on these technologies[9,11–13,16–23].

Medical knowledge within HG is composed of medical concepts (e.g., RHEUMATOID ARTHRITIS, METHOTREXATE, KNEE PAIN) and medical term labels and synonyms for these concepts in different languages. Relations exist between concepts in the form of hierarchical relations (e.g., RHEUMATOID ARTHRITIS *has child* RHEUMATOID ARTHRITIS OF WRIST) and associative relations (e.g., RHEUMATOID ARTHRITIS *has symptom* KNEE PAIN). Concepts also have mappings to codes in external terminologies used in clinical data integration and electronic medical record systems, such as the International Classification of Diseases Version 10 (ICD-10) and Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) (**Figure 1A**).

The core of HG is built from the Elsevier Merged Medical Taxonomy (EMMeT), a polyhierarchical taxonomy, which was created by integrating and expanding on popular biomedical taxonomies. As a result, medical concepts in HG can have multiple parent concepts and multiple child concepts (**Figure 1B**). Medical concepts are also classified into different semantic types (e.g., DRUG BRAND NAMES) and semantic groups (e.g., DRUGS). Associative medical relations are classified into different relation types (e.g., *has symptom*) (**Figure 1C**). These relation types are derived after a careful investigation of the Ely taxonomy[24] questions asked by clinicians to computational systems at the point of care, such as *"What is the drug of choice for condition* X*?"* and *"What test is indicated in situation* X*?"*.

HG contains additional metadata such as prevalence statistics, cohort information (e.g., sex, ethnicity, age groups for which a given medical relation is valid), geographical information (e.g., countries or states where a given medical relation is valid), and provenance information (e.g., textbook snippets where a given medical relation is mentioned[14]). **Figure 1D** shows a schematic representation of cohort and provenance information associated with the relation — RHEUMATOID ARTHRITIS *has drug* METHOTREXATE. Medical knowledge is stored as structured *(subject, predicate, object)* RDF triples in HG, where the *subject* and *object* often refer to Uniform Resource Identifiers (URIs), which are unique and refer to different entities (i.e., medical concepts, documents, labels, cohorts, etc.).

HG's polyhierarchical taxonomy of medical concepts is manually curated by Elsevier's medical subject matter experts (SMEs). Medical relations between concepts can be manually curated by medical SMEs, retrieved from legacy databases, or extracted from unstructured text through automated machine learning and natural language processing (ML/NLP) pipelines[14, 15] (**Figure 2A**). A few statistics related to HG are presented in **Table 1**. Medical knowledge in HG is consumed by several clinical products in search, recommendation, and decision support, through various Application Programming Interface (API) services and projections[15].

Table 1: Size and complexity of Elsevier's Healthcare Knowledge Graph (HG). The first two columns indicate the number of medical concepts and manually-curated relations, categorized according to a few different concept and relation categories. The last column indicates the different types of additional information captured in HG.

Medical Concepts		**Manually-Curated Medical Relations**		**Other Information**	
DISEASES	75K+	*has clinical finding*	10K+	**Labels**	1.5M+
DRUGS	48K+	*has symptom*	27K+	**Cohorts**	10K+
PROCEDURES	63K+	*has cause*	20K+	**Mappings**	573K+
SYMPTOMS	90K+	*has procedure*	15K+	**Documents**	1.5M+
ORGANISMS	35K+	*has child concept*	50K+	**Excerpts**	217K+
...	**...**	**...**	**...**	**...**	**...**
Total Concepts	400K+	**Total Relations**	1M+		

2.2 Challenges to interactive exploration and collaborative curation

Healthcare knowledge graphs need continuous and collaborative curation to keep the medical knowledge constantly updated to reflect recent trends, events, and developments. SMEs who are proficient in medical knowledge discuss and work in coordination to search and distill medical facts (e.g., *'methotrexate can be used to treat rheumatoid arthritis in a certain population'*) from recent scientific and medical literature, and curate and refine extensive knowledge management systems using these facts[25]. While automated ML/NLP pipelines that use advanced models (e.g., BERT language models[26]) can assist in the identification and extraction of novel medical facts from unstructured scientific and medical literature[14,27], the facts need to be validated by SMEs if they are going to be used in clinician-facing computational systems at the point of care.

Generally, medical SMEs are not conversant in the technologies used behind healthcare knowledge graphs. They may have difficulty grasping the knowledge representation and querying technologies that have a steep learning curve for even experienced software engineers[11]. Medical SMEs need a knowledge editing system with advanced user interfaces to enable them to interactively explore and collaboratively curate medical knowledge in extensive knowledge graphs without worrying about the underlying graph complexity and knowledge representation techniques.

HG has myriad categories of other stakeholders such as solution and enterprise architects, ML/NLP scientists, data engineers, application developers, product owners, medical informaticists, business developers, and international stakeholders. These additional stakeholders may be interested in several different tasks associated with the knowledge graph (e.g., understanding how HG can improve their product, developing methods and applications that leverage or improve HG, aligning the business initiatives around knowledge management and technology infrastructure, etc.). There is a large variation in the roles, skills, the breadth and the depth of technical expertise and product knowledge across these stakeholders. The candidate knowledge editing system could ideally also be used by various stakeholders through different roles to interactively explore medical knowledge contained within HG.

Several key stakeholders at Elsevier conducted an extensive requirements analysis for a knowledge editing system through multiple day-long workshops to gather user requirements and expectations from medical SMEs and HG technology developers. These requirements were prioritized and categorized into broad buckets (e.g., search and view, data manipulation). A few of these requirements are listed in **Table 2** for an overview. While some of these requirements may seem trivial (e.g., Requirement **G1**), it was important to list and prioritize all requirements. It is extremely difficult for a single knowledge editing system to support all these requirements, especially for the scale of HG with billions of triples. It is also important to note that the satisfaction of some requirements is relative for different users. For example, different developers may have different expectations for a 'nominal' time to import the entirety of HG within a knowledge editor system (Requirement **G4**). Additionally, different SMEs may have different expectations for intuitiveness or interactivity for a user interface (Requirements **SV1** and **DM1**).

Elsevier evaluated several actively-maintained knowledge editing systems through practical experiments (e.g., importing the entirety of HG, UI personalization, integrity constraint checks, etc.). The goal of these experiments was to classify requirements listed in **Table 2** for each knowledge editing system as follows: *i)* the requirement is fully supported through built-in features of the system, *ii)* the requirement can be satisfied with minor modifications to the built-in features, and *iii)* the requirement is currently not satisfied and completely novel features or workarounds will need to be developed within the system. After several months of discussions and experiments in collaboration with the medical SMEs, Elsevier selected the WebProtégé knowledge editing system, a cloud-based ontology editor, developed and maintained at Stanford University's Center of Biomedical Informatics Research[28,29]. The academic–industry collaboration was established to further develop additional features and modifications to meet our requirements.

Table 2: A few requirements to be satisfied by a knowledge editor system for HG. The first three columns indicate the category of the requirement (**G** - General, **SV** - Search and View, **DM** - Data Manipulation, **IP** - Integrity and Provenance), the beneficiary stakeholder, and the description of the requirement, whereas the last column indicates how the requirement was satisfied by the WebProtégé knowledge editor system during our preliminary evaluation.

	Stakeholder	Requirement	WebProtégé Support
G1	SMEs and Developers	Support for collaborative editing through simultaneous sessions by SMEs, who may use diverse languages and browsers, and live in different regions	Built-in support for advanced browsers
G2	SMEs and Developers	Support for different user roles and allowed capabilities within the system	Built-in support
G3	Developers	Integrate the knowledge editor with the existing development workflows, infrastructure, and channels	Integration is easy due to use of open source software stack
G4	Developers	Easily import and export multiple versions of the entire industry-scale knowledge graph within the editor system under 'nominal' time	**Features needed**
G5	Developers	Ability to easily customize the search, viewing, and editing interfaces for different users	**Features needed**
SV1	SMEs	Intuitive and interactive interfaces to visualize different aspects of HG (e.g., HG polyhierarchical taxonomy, complex and granular relations, etc.)	**Features needed**
SV2	SMEs	Define personalized language and UI preferences within the system	Built-in support
SV3	SMEs	Search external biomedical vocabularies (e.g., SNOMED CT, ICD-10) for medical concepts	Built-in support with some modifications
SV4	SMEs	Perform advanced searches in the knowledge graph (e.g., mappings, fuzzy search)	Built-in support with some modifications
DM1	SMEs	Intuitive and interactive interfaces to create, edit, delete, or retire medical concepts and relations	**Features needed**
DM2	SMEs	Drag and drop hierarchical branches and translations for medical concepts within HG	Built-in support
DM3	SMEs	Perform bulk edits, additions, or deletions for medical concepts, relations, and labels within HG	**Features needed**
IP1	SMEs and Developers	Explore and/or validate large data sets (e.g., relations extracted from ML/NLP pipelines)	**Features needed**
IP2	SMEs and Developers	Collect and explore metadata around each edit or modification made by a user	Built-in support with some modifications
IP3	Developers	Incorporate constraints, quality assurance, and integrity checks around data input	**Features needed**

2.3 WebProtégé knowledge editing system

The WebProtégé knowledge editing system is a Web-based version of the Protégé desktop-based ontology editing tool, which has been widely used to edit and refine several popular medical ontologies and knowledge bases, including the Gene Ontology, National Cancer Institute Thesaurus, World Health Organization's International Classification of Diseases - Version 11. WebProtégé has been primarily used to edit ontologies and knowledge bases created using the Web Ontology Language (OWL). Recently, the use of WebProtégé has also been explored to curate industry-scale knowledge graphs, such as the Pinterest Taste Graph[30]. The publicly hosted version of WebProtégé (Hosted at `webprotege.stanford.edu`) currently has more than 68K user accounts and more than 99K projects.

WebProtégé has several built-in features that instantly satisfied some of our requirements for the ideal knowledge editing system (**G1**, **G2**, **SV2**, **DM2**). These included collaborative editing with full change tracking across users, rollback and replay changes globally or per concept, user roles, discussion threads, chat and mail notifications, among others, all designed for collaborative curation of medical ontologies. The use of an open source software stack within WebProtégé (e.g., Tomcat web server, MongoDB document database) also satisfied an additional technology requirement. It should be noted that **Table 2** only lists some of the considered requirements, and only displays the results from our preliminary evaluation of the WebProtégé knowledge editing system.

Knowledge graphs and ontologies are stored as separate projects within the WebProtégé knowledge editing system and it is currently possible to store multiple versions of the same knowledge graph as different projects or make revisions to the same knowledge graph in the same project (part of requirement **G4**). As shown in **Figure 2C**, WebProtégé is equipped with the Bulk Edits API that enables developers to add or delete statements that correspond to groups of OWL axioms in bulk. The ontologies and edit histories are stored in a separate file system. Moreover, the Revisions API allows developers to export revisions or changes made by SMEs to a given project (e.g., addition of a new concept or a relation). Information on projects, users, user roles, preferences etc. is stored in a MongoDB document database. Additional customizations (e.g., UI design specifications like color and font, system settings such as the project dormant time) are stored in separate 'properties' or style files.

3 APPROACHES TO HANDLE SCALE AND COMPLEXITY OF HG WITHIN WEBPROTÉGÉ

In addition to the built-in features that satisfy some of Elsevier's requirements, there are several additional features and modifications we need to implement within Elsevier's Healthcare Knowledge Graph (HG) model and processes and in the WebProtégé editing system to enable medical SMEs to curate HG on a regular basis. We have developed additional methods and features to address the following requirements from **Table 2 — G4**, **G5**, **SV1**, **DM1**, **DM3**, **IP1**, **IP3**. We have grouped these additional methods and features in three main groups which we will explore further.

3.1 Handling complexity through model transformations

Since WebProtégé has been typically used for editing OWL ontologies in biomedicine, the display features in the editing system are geared toward the usage of OWL axioms and modeling patterns. However, HG has been developed using an RDF-based representation model. OWL and RDF can both be serialized to compatible formats and stored as graphs (e.g., triple-based formats such as Turtle). While we successfully imported the entirety of HG into the WebProtégé editor during our experiments to select a knowledge editing system, we were not able to visualize the rich polyhierarchical taxonomy, associated concept labels, or relations, in an intuitive way for the medical SMEs (e.g., indented tree layout visualization for taxonomies). To satisfy the requirements **SV1**, **DM1**, and **IP1**, we developed model transformation logic which converted the custom RDF-based representation model used in HG into an OWL-based representation model before importing into the WebProtégé knowledge editing system.

Through model transformation, all HG medical concepts are designated as OWL classes and the concept hierarchy is represented through the use of *rdfs:subClassOf* property. Properties that associate medical concepts to metadata (e.g., concepts are mapped to concept labels using (*hasMedicalName*, or concepts are mapped to codes in external terminologies using *hasExternalCode*) are represented as OWL annotation properties. All associated metadata are also represented as OWL named individuals to capture the additional information (e.g., concept labels have additional search keywords). An HG concept is associated with multiple OWL named individuals (e.g., multiple concept labels) by the OWL annotation properties (e.g., *hasMedicalName*) created through this modeling approach.

Currently, associative relations within HG are classified into different relation categories and generally have a subject concept and a object concept (e.g., in **Figure 1D**, RHEUMATOID ARTHRITIS is the subject, METHOTREXATE is the object, and *has drug* is the category of the relation). The same given relation can be derived from multiple sources (e.g., different medical textbooks can mention the relation between RHEUMATOID ARTHRITIS and METHOTREXATE) and can have associated granular information (e.g., cohort details, provenance information) that vary depending on the the source of the relation. For each unique relation between two HG concepts (i.e., each unique relation triple of subject, category, and object), we created an OWL class, which is a child class of the relation category. The new relation class has existential role restrictions with the associated medical concepts (i.e., RHEUTMATOID ARTHRITIS and METHOTREXATE). Moreover, each mention of this relation, as observed in different sources, is represented as an OWL named individual that belongs to the newly created relation class. The named individual can further have property assertions that associate additional relation metadata (e.g., cohort, provenance, strength of the relation) to the relation mention. This model transformation equips WebProtégé editing system to display the large number of NLP-extracted relations if the SMEs want to manually curate them in the future.

SPARQL CONSTRUCT query templates are used to extract triples from HG and transform the underlying model of those triples (**Figure 2B**). These templates are designed to transform different aspects of the knowledge graph (e.g., concept hierarchy, concept mappings). Revisions made daily by SMEs in the WebProtégé editing system are exported through the built-in Revisions API. However, these revisions are stored in a functional OWL format with the transformed model. Hence, we have developed reverse model transformation logic as well to parse the functional OWL revisions and convert them back to the representation model currently used in HG. A large number of projections and API services that are used to power various clinical applications and are dependent on the persistence of the HG data model. Through these model transformations between RDF and OWL, we ensure that the HG RDF-based representation model is not currently modified outside of the WebProtégé editing system.

3.2 Handling scale through improvements in processes, search, and automation

Traditionally users have uploaded their ontologies or knowledge graphs in the WebProtégé editing system through the user interface. A basic requirement (**G4**) is the ability to 'programmatically' import and export multiple versions of HG within the knowledge editing system through automated pipelines under 'nominal' time. While WebProtégé has a built-in Bulk Edits API to add or delete large numbers of triples, we were unable to import the entire HG before or after model transformation in one attempt. Common reasons for failure included network timeouts and issues parsing triples in a consistent manner under desirable time.

We developed a 'chunking' algorithm to split HG into smaller graph clusters with a set number of triples for upload within WebProtégé . Model transformations led to the creation of a large number of 'blank' nodes, especially through the use of different role restrictions for representing associative relations between HG concepts. Blank nodes are often used in RDF and OWL-based knowledge representations to associate additional granular information (e.g., **Figure 1D** has cohort and provenance information attached to a blank node that represents an HG relation) or to create equivalent classes in role restrictions (e.g., $\exists hasSubject.\mathbf{RheumatoidArthritis}$ is a class of relation individuals, where at least one linked subject belongs to $\mathbf{RheumatoidArthritis}$, and can be represented as a blank node in the OWL representation and a triple-based serialization). During chunking, we have to ensure that triples referring to a given blank node remain in the same cluster. To determine the optimal number of triples in each cluster, we experimented importing the entire HG into WebProtégé with different cluster sizes.

Potential medical knowledge in HG is also generated through additional methods (e.g., ML/NLP pipelines, ETL pipelines, or other bulk data deliveries). These updates need to be made to the HG version in WebProtégé for medical SMEs and other stakeholders to browse and validate the new knowledge (Requirement **DM3**). In these cases, we only upload the modifications (referred to henceforth as 'deltas') to WebProtégé , instead of re-importing the entire HG. There are several mechanisms to compute deltas between two versions of an RDF graph[31]. However, in many cases these mechanisms are not often computationally feasible and can become prohibitively time consuming, especially when dealing with the size of HG. We have developed domain-specific heuristics to compute deltas over different aspects of HG (e.g., concept hierarchy, concept mappings). Delta computation is only triggered when a given aspect of HG is updated. The computed deltas are uploaded into WebProtégé after model transformation and chunking.

A requirement around the search of medical concepts and relations within HG (Requirement **SV4**) led to a comprehensive avenue of improvements within the WebProtégé editing system. WebProtégé had a built-in search interface for simple searches of OWL classes using their labels, as well as a DLQuery (description logic query) interface for formulating advanced queries (e.g., over graphical paths or using regular expressions) that is shown in **Figure 3B**. However, after initial import of HG with more than 1.5M labels (including medical names, consumer names, synonyms, etc.), the original search was very slow to retrieve the relevant responses. We incorporated Apache Lucene within the WebProtégé editing system to improve the search performance in both the conventional search interface and the advanced query interface. The embedded Lucene search indexes are stored within the WebProtégé file system (**Figure 2C**). Moreover, UI customizations enabled developers to provide annotation properties (*rdfs:label*) and annotation property paths (e.g., *synonym* $\rightarrow$ *rdfs:label*) from where the labels will be indexed.

Elsevier uses automation to easily build and deploy new versions of the WebProtégé system from any branch in the publicly available GitHub repository (`https://github.com/protegeproject/webprotege`), as well as to import different versions of HG within the WebProtégé system. Elsevier extensively uses Docker and automation pipelines to build the WebProtégé web application archive (WAR), which is then stored and deployed on Elsevier's private cloud through the Tomcat web server. Similarly, the MongoDB document database is also built and deployed through such automation pipelines. It currently takes around 11 hours to import HG into WebProtégé through automated scripts and the Bulk Edits API, after the completion of model transformation and chunking HG into mini-graphs, with each mini-graph having approximately 500K triples (**Figure 2B**). Revisions are exported from the WebProtégé system on a nightly cadence through the Revisions API.

3.3 Improving user experience through novel WebProtégé UI features

WebProtégé has several built-in UI features and views that enable editors to explore, query, and edit knowledge graphs. Moreover, these features and views can be personalized as required by the users according to their language and display preferences. For example, as shown in **Figure 3A**, we have created a personalized layout of tabs, and different panels or 'views' in each tab (e.g., 'Editorial View' tab) that were of interest to our medical SMEs. However, additional user interface (UI) features were required to improve the usability and user experience of WebProtégé for Elsevier's medical SMEs to visualize and collaboratively edit HG (Requirements **G5**, **SV1**, and **DM1**).

The default views within WebProtégé are entity-centric. When a user is browsing a given entity (i.e., a class or an individual) all entities linked with the given entity through annotation properties and object properties at one hop away in the knowledge graph are shown in the WebProtégé browser. when viewing any given medical concept of HG within WebProtégé , all of these annotations and relations were displayed in a single list (through the built-in 'Class' view in WebProtégé). This forced medical SMEs who were viewing or editing large concepts with exhaustive medical knowledge (e.g., DIABETES MELLITUS or ASTHMA) to scroll an extremely long list and click multiple times to browse certain granular information (e.g., source or cohort of a given relation or mapping).

To improve this user experience we developed several sets of UI elements and features, grouped under a novel WebProtégé 'Forms' view (**Figure 3A**). It allows developers to create customizable display and editing interfaces which are more relevant and intuitive for medical SMEs to browse and edit HG within WebProtégé . Medical information is grouped into different sub-tabs, depending on the different aspects of HG (e.g., description, synonyms, mappings, relations). Each sub-tab further displays additional groupings and classifications of medical information depending on the context. For example, the 'Relations' sub-tab (**Figure 3A**), displays the different associative typed relations for RHEUMATOID ARTHRITIS, grouped according to the relation type (e.g., *has drug*). Within the sub-tab, relations are shown as a grid with different columns displaying the relation metadata (e.g., source of the relation, rank – strength of the relation). Through the use of advanced web-based features, such as pagination, tabs, and collapsible sections, the user can easily browse hundreds and thousands of nodes and edges, linked to any given medical concept, at different hops in the knowledge graph without being overwhelmed. The user can easily search or click to browse any other medical concept using the class hierarchy view displayed in the right of the 'Forms' view. When the medical SME wishes to edit medical information for any given concept, the same 'Forms' view can directly be transformed into an editorial interface by clicking on the 'Edit values' option. In certain cases, the editorial interface has interactive UI elements (e.g., dropdowns or radio buttons), depending on the expected type of user input.

HG developers collaborated with medical SMEs and decided on the most optimal and intuitive manner to display medical information in the 'Forms' view, and use another interactive administrator interface to customize the 'Forms' view. In this administrator interface, the developers can bind different elements in the knowledge graph (e.g., object property, instances, classes) to different UI elements for display (e.g., grid, text field, number field, etc.). This binding can be done recursively enabling graph traversal (e.g., the label of the source of a relation may be three or more hops away from the subject concept being browsed). These customized display and editorial forms are created for different aspects of HG (e.g., mappings), as well as for creating or for deprecating medical concepts in HG. The 'Forms' administrator interface provides the ability to incorporate simple constraints (e.g., only integer values in a given range for a given field) around the expected data input in the editorial forms (Requirement **IP3**).

During this project, we have developed several other minor, yet novel, UI features to improve the workflows and the user experience for medical SMEs while browsing and editing HG. For example, an initial version of the WebProtégé 'Entity Graph' view[30] visualized all outgoing edges from a given OWL class (hierarchical relations and property associations). In the case of medical concepts in HG that have a lot of associated medical knowledge (e.g., common diseases such as ASTHMA and DIABETES MELLITUS), this would become overwhelming to browse for a medical SME and would also affect the page loading time for the entire web application. We developed the ability to use edge filters with a default filter to only visualize hierarchical relations (i.e., *rdfs:subClassOf* edges) for the selected concept. This increasingly improved the user experience for less advanced and first time users and enabled medical SMEs to visualize the rich polyhierarchical taxonomy of HG (**Figure 1B**). Similar concept, edge, and language filters were also developed to improve search and display preferences within the UI. Through the use of these filters, medical SMEs were able to search for only ‘drug’ concepts or browse mappings or labels in a given language.

4 DISCUSSION

Manual curation and continuous refinement of industry-scale knowledge graphs, especially in healthcare and biomedical domains, are crucial to the successful use of those knowledge graphs. However, an ideal ‘one size fits all’ knowledge editing system does not exist. Different knowledge graphs and stakeholders have varying requirements and interpretations around how an ideal editing system should function. Elsevier’s preliminary experiments with different knowledge editing systems led us to select the WebProtégé system with their open source technology stack and establish a close R&D collaboration between the developers of the knowledge editing system and the developers of the knowledge graph, so that the required features can be developed internally.

This R&D collaboration also had a set goal of incorporating a technology system, developed and maintained by academic researchers, into an industry technology stack and workflow, with automated build and deployment processes, minimal downtime, production-level quality assurance, backups, and an interactive user experience for our medical SMEs. To achieve this, developers of the WebProtégé editing system and the developers of HG met on a regular cadence every alternate day to discuss progress and priorities. These meetings were often attended by medical SMEs, from different geographical regions, and infrastructure engineers to discuss their needs and requirements.

Previously, the WebProtégé team has established R&D collaborations with other companies to develop additional features within the system[30]. However, the uniqueness of this R&D collaboration with Elsevier Health Markets was that the WebProtégé team was not directly responsible for creating or extending HG itself due to the maturity of the graph. Moreover, the UI requirements of Elsevier’s stakeholders as well as the scale, multi-language support and the intricate structure of HG posed unique challenges for the WebProtégé system.

In some cases, where the collaboration fell short to meet the key requirements within the project period, alternate workarounds were outlined and developed to meet those requirements later. For example, to meet the requirement for quality assurance (QA) through complex model constraints and integrity checks around the exported revisions from the WebProtégé editing system, the developers later implemented a QA testing framework where the lead medical SME can accept or reject certain revisions made to HG. Automated QA tests also evaluate the graph revisions for model violations (**Figure 2B**). The rejected revisions are tagged within WebProtégé through a custom Tagging API service that internally uses the WebProtégé Bulk Edits API and the WebProtégé tagging framework[30].

In the previous section, we have given an overview of the different approaches and improvements made to our processes, as well as features and improvements made to the WebProtégé editing system to accommodate the requirements of different stakeholders. The novel features developed and the improvements made to WebProtégé (e.g., Lucene search, ‘Forms’ view) through this collaboration are generically applicable and customizable for different knowledge graphs in other domains as well. The novel features and modifications incorporated within the WebProtégé system are publicly available through the GitHub repository.

There are still several shortcomings and features that are required in the WebProtégé knowledge editing system to improve the workflow of medical SMEs and other stakeholders who search, browse, and edit HG. Two major avenues for novel feature development within the WebProtégé system are:

1. **Query Template Management:** The user can formulate advanced queries (e.g., regular expression query combined with graph traversal) using the DLQuery interface (**Figure 3B**). However, to formulate queries using this

interface, medical SMEs need to be aware of the exact modelling patterns (i.e., the data/object/annotation property paths), which is often difficult and non-intuitive. The steep learning requirements for formulating complex semantic queries is a common user experience problem, which has been well documented in prior research[11,32]. A query template management framework that enables HG developers and medical SMEs to save and share query templates with other users within the WebProtégé system is ideally required.

2. **User Role Management:** WebProtégé has some built-in user roles (e.g., *Viewer*, *Editor*, *Commenter*, *Manager*), and different users can perform different actions within the system, depending on the assigned role. However, user roles are not uniform across different projects and different knowledge graphs. For example, within HG, more granular user roles are required for users who can edit the graph with different capabilities (e.g., *Bulk Editor*, *Limited Editor*, *Knowledge Manager*). A user role management framework that enables HG developers to create custom user roles with different functionalities within the WebProtégé system is required.

5 CONCLUSION

In this case report, we have presented our adoption and further development of the WebProtégé cloud-based knowledge editing system to enable medical subject matter experts to regularly curate and refine Elsevier's Healthcare Knowledge Graph (HG). Elsevier's stakeholders selected WebProtégé from several candidate knowledge editing systems after a thorough analysis of requirements as well as after conducting several experiments and investigations on the handling of size and complexity of HG within the editing system. To incorporate the WebProtégé editing system firmly within Elsevier's industry technology stack, a close R&D collaboration was established between the WebProtégé and HG developers, along with medical SMEs and infrastructure engineers. Through this collaboration, several approaches, methods and processes were developed and improved iteratively to ensure that WebProtégé can accommodate the size and complexity of HG. We also collaboratively developed several novel features and modifications within WebProtégé to improve the usability and user experience of medical SMEs and other stakeholders, who browse, edit, and search HG. We believe that this case report will impress upon the readers the need for knowledge curation and refinement for industry-scale healthcare knowledge graphs and provide guidance on the methods and approaches that can be used to achieve the goal. Additionally, since features and modifications made to the WebProtégé knowledge editing system are released in the public version, we have provided the motivation and detail on the development of these features to facilitate their adoption by other academic and industry research groups in curating their knowledge graphs.

ACKNOWLEDGMENTS

The authors would like to acknowledge Craig E. Stanley Jr., David Conrad, Will Dowling, Sayema Islam, Dylan Montagu, Chris Wolf, Jessica Cox, Peter Willis, Danielle Walsh, Àlex Ausió, Andrés Carazo, Silvia Maifrede, Patti Novak, Leah Pilossoph, Michael Carroll, Sravanthi Tummala, David Childs, Dinakar Kulkarni, Mike Dunne, Steve Ross, Connor Skiro, Alex De Jong, Paul Snyder, and Chris Stoces, for discussions and support around the Elsevier–Stanford collaboration and on the implementation of the WebProtégé platform for Elsevier's Healthcare Knowledge Graph. Moreover, we acknowledge Tania Tudorache and Csongor Nyulas for constructive feedback around WebProtégé features and technology stack. This work was funded in part by a contract from Elsevier to Stanford University and by grant R01 GM121724 from the National Institute of General Medical Sciences.

AUTHOR CONTRIBUTIONS STATEMENT

MK and MH equally contributed to the technical development of the features, processes, and methods within HG and WebProtégé, as well as toward writing the manuscript. LW and CF contributed to developing processes within HG. JH contributed to feature development within WebProtégé. DA, KS, and RG provided valuable feedback to the manuscript as well as helped elaborate and evaluate the technical requirements. DH, MS, and MM established the goals of the collaboration. All authors have read and agreed to the manuscript.

COMPETING INTERESTS

MK, LW, CF, DA, KS, DH, and MS were employed by Elsevier Inc. during this collaboration and while writing the manuscript. This work was funded in part through a contract from Elsevier to Stanford University.

DATA AND CODE AVAILABILITY

Elsevier's Healthcare Knowledge Graph (HG) used and mentioned in this collaboration is a proprietary data source and can not be shared publicly. The WebProtégé editing system and all the novel features and modifications incorporated within WebProtégé system are publicly available through the GitHub repository (`https://github.com/protegeproject/webprotege`).

References

1. Wolters Kluwer Health. UpToDate. `https://www.uptodate.com/home`. Accessed: 2021-06-09.
2. EBSCO Health. Dynamed. `https://www.dynamed.com`. Accessed: 2021-06-09.
3. Elsevier. ClinicalKey Search Engine. `https://www.elsevier.com/solutions/clinicalkey`. Accessed: 2021-06-09.
4. US National Libraries of Medicine. PubMed. `https://www.ncbi.nlm.nih.gov/pubmed/`. Accessed: 2021-06-09.
5. Peter Densen. Challenges and opportunities facing medical education. *Transactions of the American Clinical and Climatological Association*, 122:48–58, 2010.
6. Holly Else. How a torrent of COVID science changed research publishing–in seven charts. *Nature*, 588(7839):553–554, 2020.
7. David A Cook, Kristi J Sorensen, John M Wilkinson, and Richard A Berger. Barriers and decisions when answering clinical questions at the point of care: a grounded theory study. *JAMA internal medicine*, 173(21):1962–1969, 2013.
8. Guilherme Del Fiol, T Elizabeth Workman, and Paul N Gorman. Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA internal medicine*, 174(5):710–718, 2014.
9. Prem Ramaswami. A remedy for your health-related questions: health info in the knowledge graph. *Google Official Blog*, 2018, 2015.
10. Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13):i457–i466, 2018.
11. Maulik R Kamdar, Javier D Fernández, Axel Polleres, Tania Tudorache, and Mark A Musen. Enabling web-scale data integration in biomedicine through linked open data. *NPJ digital medicine*, 2(1):1–14, 2019.
12. DCC AOCNP. Watson will see you now: a supercomputer to help clinicians make informed treatment decisions. *Clinical journal of oncology nursing*, 19(1):31, 2015.
13. Shima Dastgheib, Craig Webb, Qiaonan Duan, Rowan Copley, Gini Deshpande, and Asim Siddiqui. Accelerating drug discovery in rare and complex diseases. In *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018.
14. Maulik R Kamdar, Craig E Stanley, Michael Carroll, Linda Wogulis, William Dowling, Helena F Deus, and Mevan Samarasinghe. Text snippets to corroborate medical relations: an unsupervised approach using a knowledge graph and embeddings. *AMIA Summits on Translational Science Proceedings*, 2020:288, 2020.
15. Alex DeJong, Radmila Bord, Will Dowling, Rinke Hoekstra, Ryan Moquin, Charlie O, Mevan Samarasinghe, Paul Snyder, Craig Stanley, Anna Tordai, Michael Trefry, and Paul Groth. Elsevier's healthcare knowledge graph and the case for enterprise level linked data standards. In *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks co-located with 17th International Semantic Web Conference (ISWC 2018), Monterey, USA, October 8th - to - 12th, 2018.*, 2018.
16. Amrapali Zaveri and Gökhan Ertaylan. Linked data for life sciences. *Algorithms*, 10(4):126, 2017.
17. M Scott Marshall, Richard Boyce, Helena F Deus, Jun Zhao, Egon L Willighagen, Matthias Samwald, Elgar Pichler, Janos Hajagos, Eric Prud'hommeaux, and Susie Stephens. Emerging practices for mapping and linking life sciences data using RDF-A case series. *Web Semantics: Science, Services and Agents on the World Wide Web*, 14:2–13, 2012.

18. Olivier Bodenreider. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, 32(suppl 1):D267–D270, 2004.

19. Antony J Williams, Lee Harland, Paul Groth, Stephen Pettifer, Christine Chichester, Egon L Willighagen, Chris T Evelo, Niklas Blomberg, Gerhard Ecker, Carole Goble, et al. Open PHACTS: semantic interoperability for drug discovery. *Drug discovery today*, 17(21-22):1188–1198, 2012.

20. Graham Klyne and Jeremy J Carroll. Resource description framework (RDF): Concepts and abstract syntax. *W3C recommendation*, 2006.

21. Eric Prud'Hommeaux, Andy Seaborne, et al. SPARQL query language for RDF. *W3C recommendation*, 15, 2008.

22. Sean Bechhofer. OWL: Web ontology language. In *Encyclopedia of Database Systems*, pages 2008–2009. Springer, 2009.

23. Mark D Wilkinson, Benjamin Vandervalk, and Luke McCarthy. The Semantic Automated Discovery and Integration (SADI) web service design-pattern, API and reference implementation. *Journal of biomedical semantics*, 2(1):8, 2011.

24. John W Ely, Jerome A Osheroff, Paul N Gorman, Mark H Ebell, M Lee Chambliss, Eric A Pifer, and P Zoe Stavri. A taxonomy of generic clinical questions: classification study. *BMJ*, 321(7258):429–432, 2000.

25. Michelle Whirl-Carrillo, Ellen M McDonagh, JM Hebert, Li Gong, K Sangkuhl, CF Thorn, Russ B Altman, and Teri E Klein. Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics*, 92(4):414–417, 2012.

26. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

27. Bethany Percha and Russ B Altman. A global network of biomedical relationships derived from text. *Bioinformatics*, 34(15):2614–2624, 2018.

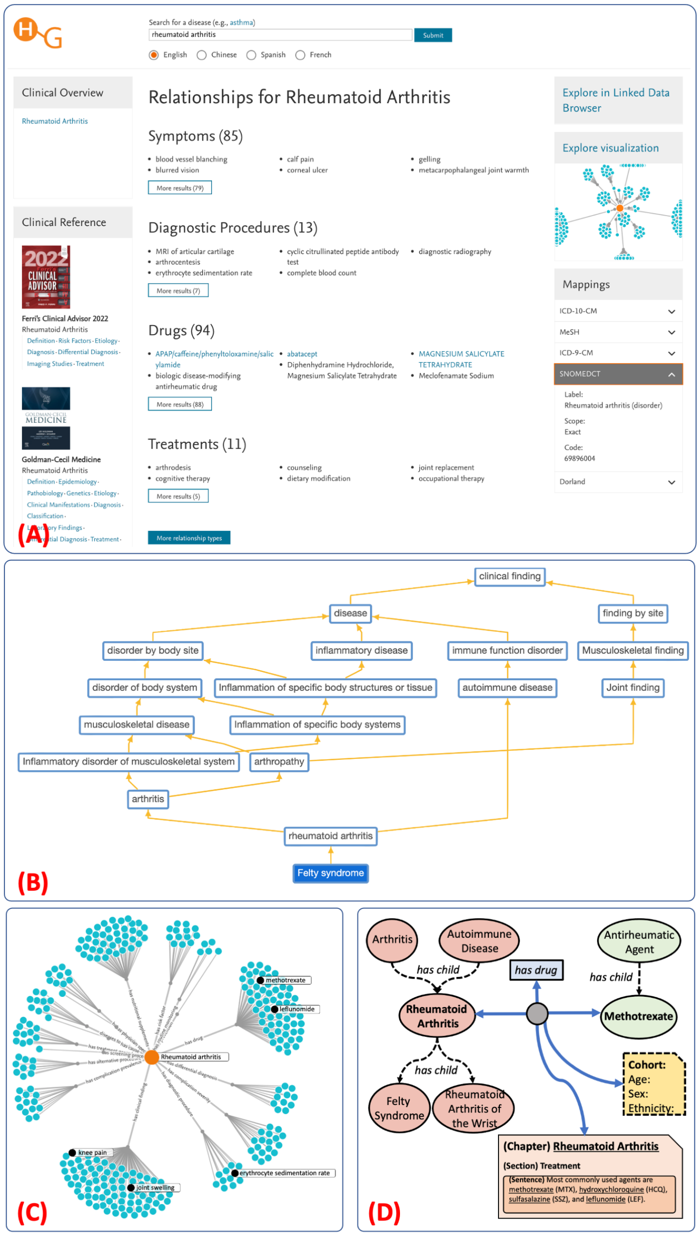
28. Matthew Horridge, Tania Tudorache, Csongor Nuylas, Jennifer Vendetti, Natalya F Noy, and Mark A Musen. WebProtégé: a collaborative Web-based platform for editing biomedical ontologies. *Bioinformatics*, 30(16):2384–2385, 2014.

29. Matthew Horridge, Rafael S Gonçalves, Csongor I Nyulas, Tania Tudorache, and Mark A Musen. Webprotégé: A cloud-based ontology editor. In *Companion Proceedings of The 2019 World Wide Web Conference*, pages 686–689, 2019.

30. Rafael S Gonçalves, Matthew Horridge, Rui Li, Yu Liu, Mark A Musen, Csongor I Nyulas, Evelyn Obamos, Dhananjay Shrouty, and David Temple. Use of OWL and Semantic Web Technologies at Pinterest. In *International Semantic Web Conference*, pages 418–435. Springer, 2019.

31. Dimitris Zeginis, Yannis Tzitzikas, and Vassilis Christophides. On the foundations of computing deltas between RDF models. In *The Semantic Web*, pages 637–651. Springer, 2007.

32. Maulik R Kamdar, Dimitris Zeginis, Ali Hasnain, Stefan Decker, and Helena F Deus. ReVeaLD: A user-driven domain-specific interactive search platform for biomedical research. *Journal of biomedical informatics*, 47:112–130, 2014.



Search for a disease (e.g., asthma)

rheumatoid arthritis

Submit

● English ○ Chinese ○ Spanish ○ French

Clinical Overview

Rheumatoid Arthritis

Clinical Reference

Ferri's Clinical Advisor 2022
Rheumatoid Arthritis
Definition · Risk Factors · Etiology · Diagnosis · Differential Diagnosis · Imaging Studies · Treatment

Goldman-Cecil Medicine
Rheumatoid Arthritis
Definition · Epidemiology · Pathobiology · Genetics · Etiology · Clinical Manifestations · Diagnosis · Classification · Laboratory Findings · Differential Diagnosis · Treatment ·

Relationships for Rheumatoid Arthritis

Symptoms (85)

- blood vessel blanching
- blurred vision
- calf pain
- corneal ulcer
- gelling
- metacarpophalangeal joint warmth

More results (79)

Diagnostic Procedures (13)

- MRI of articular cartilage
- arthrocentesis
- erythrocyte sedimentation rate
- cyclic citrullinated peptide antibody test
- complete blood count
- diagnostic radiography

More results (7)

Drugs (94)

- APAP/caffeine/phenyltoloxamine/salicylamide
- biologic disease-modifying antirheumatic drug
- abatacept
- Diphenhydramine Hydrochloride, Magnesium Salicylate Tetrahydrate
- MAGNESIUM SALICYLATE TETRAHYDRATE
- Meclofenamate Sodium

More results (88)

Treatments (11)

- arthrodesis
- cognitive therapy
- counseling
- dietary modification
- joint replacement
- occupational therapy

More results (5)

More relationship types

Explore in Linked Data Browser

Explore visualization

Mappings

ICD-10-CM

MeSH

ICD-9-CM

SNOMEDCT

Label:
Rheumatoid arthritis (disorder)

Scope:
Exact

Code:
69896004

Dorland

(Chapter) Rheumatoid Arthritis

(Section) Treatment

(Sentence) Most commonly used agents are methotrexate (MTX), hydroxychloroquine (HCQ), sulfasalazine (SSZ), and leflunomide (LEF).

Figure 1: Elsevier's Healthcare Knowledge Graph (HG): (A) A screenshot of HGViz, an in-house visualizer, provides a synoptic view of HG content. For the concept Rheumatoid Arthritis, HG has related content in different languages (e.g., Chinese, Spanish, French), linked content related to Rheumatoid Arthritis from existing journals and books (e.g., *Ferri's Cinical Advisor 2021*), other linked concepts in HG (e.g., drugs, symptoms, diagnostic procedures) through expert-curated semantic relations, and mappings to related concepts in other terminologies (e.g., ICD-10, SNOMED CT). **(B)** Rheumatoid Arthritis is organized in HG's rich polyhierarchical taxonomy with multiple parents (e.g., Arthritis), ancestors (e.g., Arthropathy), and children (e.g., Felty Syndrome). **(C)** A hairball visualization, centered on Rheumatoid Arthritis, showcases associative relations grouped according to different relation types (e.g., *has clinical finding*, *has drug*). **(D)** HG has additional granular information and metadata linked to each concept or relation, such as cohort information (e.g., age, sex, ethnicity, for which a given relation — Rheumatoid Arthritis *has drug* Methotrexate — is valid, or provenance information (e.g., text snippet from a medical textbook where the relation was extracted from).

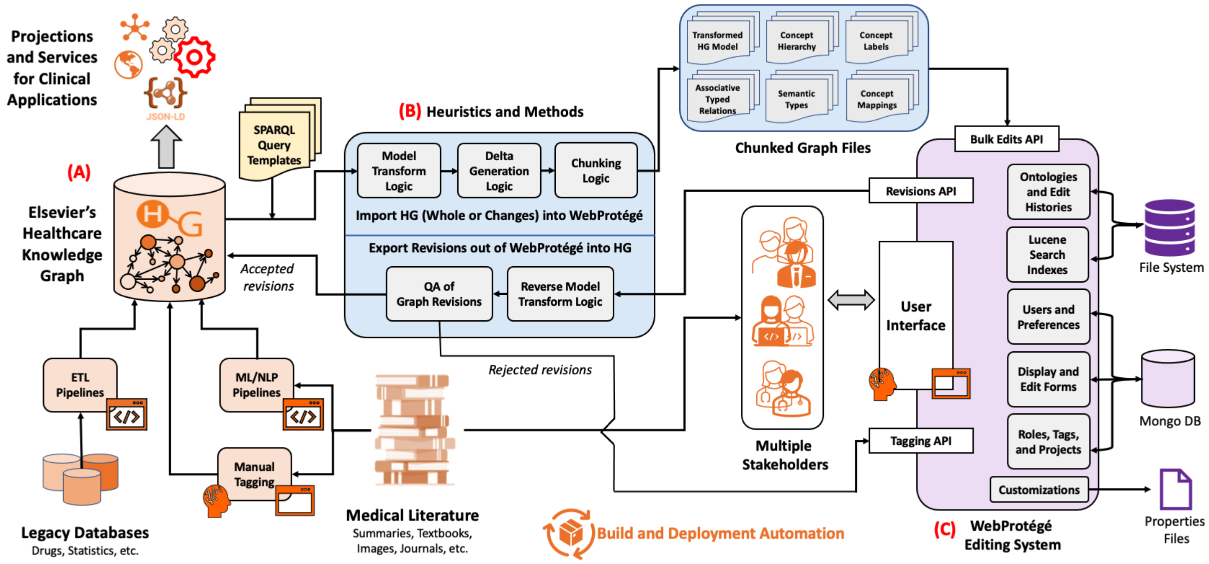
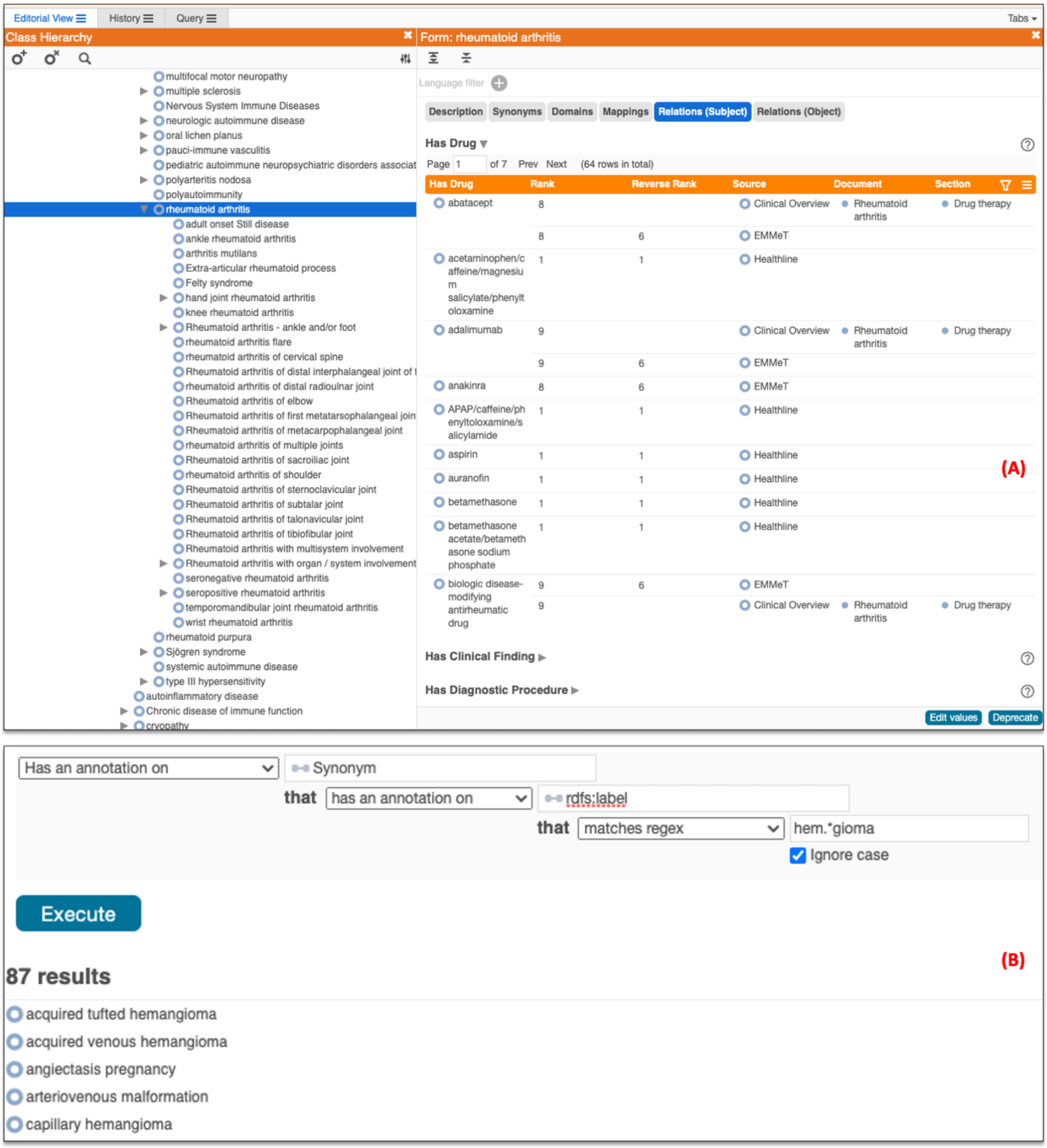


Figure 2: Integration of WebProtégé editing system in the Elsevier's Healthcare Knowledge Graph technology ecosystem: **(A)** Elsevier's Healthcare Knowledge Graph extensively uses ETL pipelines to extract, transform, and load medical information from legacy databases and ML/NLP pipelines, as well as manual tagging interfaces, to identify and extract medical relations from several sources of medical literature. Medical knowledge is represented as a knowledge graph and made available to several clinical applications through projections and sources. **(B)** Through several heuristics and methods, medical knowledge in HG is transformed and broken down into smaller chunks, representing different aspects of HG (e.g., concept hierarchy, concept labels, associative relations), for import into the WebProtégé knowledge editing system using the built-in Bulk Edits API. Similarly, changes made by medical subject matter experts (SMEs) are exported out of the WebProtégé system using the Revisions API and transformed back to HG's model. After QA of the graph revisions, approved edits are incorporated in HG and rejected revisions are tagged in WebProtégé for further edits. **(C)** The WebProtégé knowledge editing system uses open source technologies, such as the Tomcat web server to run the application, the MongoDB document database to store users, forms, preferences, roles, projects, etc., and a file system to store ontologies, edit histories, and search indexes. Moreover, WebProtégé allows several different stakeholders, including medical SMEs, to explore HG through different roles and permissions.



- multifocal motor neuropathy
- multiple sclerosis
- Nervous System Immune Diseases
- neurologic autoimmune disease
- oral lichen planus
- pauci-immune vasculitis
- pediatric autoimmune neuropsychiatric disorders associat
- polyarteritis nodosa
- polyautoimmunity
- rheumatoid arthritis
  - adult onset Still disease
  - ankle rheumatoid arthritis
  - arthritis mutilans
  - Extra-articular rheumatoid process
  - Felty syndrome
  - hand joint rheumatoid arthritis
  - knee rheumatoid arthritis
  - Rheumatoid arthritis - ankle and/or foot
  - rheumatoid arthritis flare
  - rheumatoid arthritis of cervical spine
  - Rheumatoid arthritis of distal interphalangeal joint of f
  - rheumatoid arthritis of distal radioulnar joint
  - Rheumatoid arthritis of elbow
  - Rheumatoid arthritis of first metatarsophalangeal join
  - Rheumatoid arthritis of metacarpophalangeal joint
  - rheumatoid arthritis of multiple joints
  - Rheumatoid arthritis of sacroiliac joint
  - rheumatoid arthritis of shoulder
  - Rheumatoid arthritis of sternoclavicular joint
  - Rheumatoid arthritis of subtalar joint
  - Rheumatoid arthritis of talonavicular joint
  - Rheumatoid arthritis of tibiofibular joint
  - Rheumatoid arthritis with multisystem involvement
  - Rheumatoid arthritis with organ / system involvement
  - seronegative rheumatoid arthritis
  - seropositive rheumatoid arthritis
  - temporomandibular joint rheumatoid arthritis
  - wrist rheumatoid arthritis
- rheumatoid purpura
- Sjögren syndrome
- systemic autoimmune disease
- type III hypersensitivity
- autoinflammatory disease
- Chronic disease of immune function
- cryopathy

Has Drug ▾

Page 1 of 7 Prev Next (64 rows in total)

Has Drug	Rank	Reverse Rank	Source	Document	Section
abatacept	8		Clinical Overview	Rheumatoid arthritis	Drug therapy
	8	6	EMMeT		
acetaminophen/caffeine/magnesium salicylate/phenyltoloxamine	1	1	Healthline		
adalimumab	9		Clinical Overview	Rheumatoid arthritis	Drug therapy
	9	6	EMMeT		
anakinra	8	6	EMMeT		
APAP/caffeine/phenyltoloxamine/salicylamide	1	1	Healthline		
aspirin	1	1	Healthline		
auranofin	1	1	Healthline		
betamethasone	1	1	Healthline		
betamethasone acetate/betamethasone sodium phosphate	1	1	Healthline		
biologic disease-modifying antirheumatic drug	9	6	EMMeT		
	9		Clinical Overview	Rheumatoid arthritis	Drug therapy

Has Clinical Finding ▸

Has Diagnostic Procedure ▸

Has an annotation on: Synonym
that has an annotation on: rdfs:label
that matches regex: hem.*gioma
☑ Ignore case

87 results

- acquired tufted hemangioma
- acquired venous hemangioma
- angiectasis pregnancy
- arteriovenous malformation
- capillary hemangioma

Figure 3: User Interfaces of Elsevier's Healthcare Knowledge Graph within the WebProtégé Editing System: **(A)** The 'Editorial View' tab is a fully-customized display for the medical SME to browse and visualize HG. The left panel shows an indented tree visualization of HG's concept hierarchy, whereas the right panel shows the 'Forms' display of information associated for the selected concept (e.g., RHEUMATOID ARTHRITIS) organized in different sub-tabs. The 'Relations (Subject)' sub-tab displays the associative relations where the medical concept is a subject, with the focus on *has drug* relation type. **(B)** A complex path query executed within the WebProtégé editing system to search for medical concepts in HG whose synonyms match the regular expression pattern '*hem.*gioma*'.