**SearchRx: An International Drug Label Application**

**Milestone 3**

**Team**

Bobby Cavezza

Cole Crawford

Alex Giannini

Yusheng Jiang

Peter Woo

Chen Lucy Zhang

**Customer**

David Edelen

**Teaching Staff**

Peter Henstock

Simili Abhilash

**Harvard University Extension School**

**Software Engineering Capstone**

**CSCI E-599 Spring 2023**

**Table of Contents**

[**SearchRx: An International Drug Label Comparison Tool** 3](#_Toc134436534)

[Abstract 3](#_Toc134436535)

[Introduction 3](#_Toc134436536)

[Review of Relevant Applications 4](#_Toc134436537)

[Methods 5](#_Toc134436538)

[Discussion 9](#_Toc134436539)

[Future Work 11](#_Toc134436540)

[System Design 13](#_Toc134436541)

[Infrastructure 13](#_Toc134436542)

[Database 15](#_Toc134436543)

[Elasticsearch 16](#_Toc134436544)

[Semantic Search 18](#_Toc134436545)

[API 19](#_Toc134436546)

[Front End 20](#_Toc134436547)

[Sub Section Comparison 20](#_Toc134436548)

[OpenFDA vs DailyMed 22](#_Toc134436549)

[Label Scraping/Parsing 22](#_Toc134436550)

[User Experience (UX) 25](#_Toc134436551)

[Tests 28](#_Toc134436552)

[Development Process & Lessons Learned 28](#_Toc134436553)

[Technologies Used 30](#_Toc134436554)

[Appendix A: Reqs 32](#_Toc134436555)

[Appendix B: Journal Publication Plans 38](#_Toc134436556)

[Appendix C: Team Strategy and Dynamics 40](#_Toc134436557)

[Appendix D: Team Contract 40](#_Toc134436558)

[Appendix E: Country Overview and Selection Process 40](#_Toc134436559)

[Appendix F: Initial Requirement List 48](#_Toc134436560)

[Appendix G : Important Links 57](#_Toc134436561)

## **SearchRx: An International Drug Label Comparison Tool**

### Abstract

SearchRX is an open-source web application that provides search and comparison of pharmaceutical drug labels across multiple countries. It is the only existing application to allow users to search drug labels from more than three countries and includes all drug labels approved for human use in the US, EU, Australia, and Canada. The application features search and comparison functions across individual label sections and uses vectors generated from an NLP model to provide predictive ‘semantic’ search. It also provides custom label upload so users can compare their own label to the existing database and faceted navigation, which lets users reduce their results with filters. SearchRx is primarily intended for researchers and pharmaceutical professionals but will also benefit patients and doctors.

**Keywords: Drug Labels, FDA, EMA, TGA, International Pharmaceutical Regulations, Regulatory Transparency, Pharmaceutical Research, Medicine Error**

### Introduction

Regulatory bodies around the world require pharmaceutical companies to create drug labels when bringing new drugs to market. These labels include details like dosage, side effects and trial information. Some governments have created publicly accessible databases of their labels, but none allow users to search and compare individual label sections (section search and comparison), a feature that would be useful for researchers and pharmaceutical professionals looking to understand the expectations of regulatory bodies. In addition, there are no existing platforms to facilitate search of international labels from countries other than the EMA (European Medicines Agency) and FDA (U.S. Food and Drug Administration) and HC (Health Canada). SearchRx’s primary goal is to provide a solution to this problem by providing search and comparison of label sections from the EMA, FDA, HC and TGA (Therapeutic Goods Administration) and including a label upload feature for users to upload and compare their own labels.

### Review of Relevant Applications

While SearchRx is the only known application with section search and comparison for multiple countries’ drug labels, some existing web-based applications contain similar features.

SearchRx is an update of an existing application called Drug Label Explorer (DLE) which only included data from the US and EU. SearchRx expands upon DLE by improving the section parsing process for the US and EU labels; including labels from Australia and Canada; improving the search experience through semantic search; improving the architecture using Infrastructure as Code best practices; and adding a data API for programmatic access.

Many countries provide drug labels online but none of them allow for easy comparison or allow the user to search for content within an individual section. In addition, no country provides custom label upload, version history or international comparison features, but they do provide general keyword search.

In the United States, the FDA and NIH both provide drug labels through FDALabel and DailyMed but neither site includes features like section search or keyword highlighting which allows users to more clearly understand the differences between regulatory agencies (National Center, Dailymed) . Similarly, the EMA and the TGA have websites that allow users to search for a specific drug according to associated keywords, but require users to download the full PDF to view any information which prevents users from searching for text outside of the set keywords (Ema, Therapeutic Goods Administration).

There are some commercial applications on the market that contain some features like SearchRx. DrugBank provides label search for the US, EU and Canada, but does not allow section search, semantic search or comparison (DrugBank Online). RXList only provides data for the US and does not allow section search, semantic search or comparison (Internet Drug index). Reed Tech Drug Label Navigator is a paid service from LexisNexis, which provides version history and some label comparison features, but only for labels in the EU and US (Reed Tech Navigator).

### Methods

To build the database of drug labels, SearchRx scrapes all human-approved drugs from all four countries' websites. The documents are then parsed and ingested into the database according to the headers present in the files.

Implementing section search and comparison across countries was determined a high priority because it could be used to understand each regulatory body’s expectations for each sections. However, because headers are not consistent across different countries, this presented a challenge. To resolve this, after ingestion, the sections across countries were compared using SciBert Embedding Cosine Similarity and then brought to a pharmaceutical industry expert to confirm its findings and adjust as needed. With this data, a table was created to group the sections across countries and provide a single unifying section name applicable to all relevant countries. The feature accomplishes its goal and can be seen in figure 1 below, which shows a comparison of the ‘indications and usage’ section for the drugs Zydelig from the EMA and Dovonex from the FDA. The content of the sections is very different and highlights the differing expectations of the two regulatory bodies.

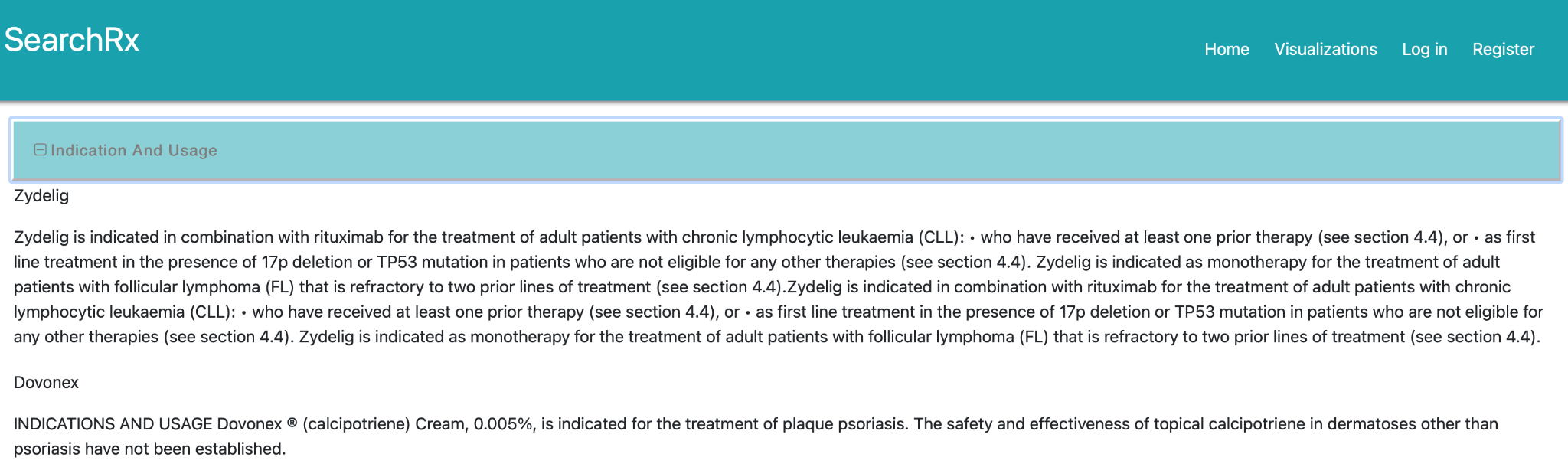


Figure 1: Section Comparison

In addition, allowing users to upload custom labels was determined to be a high priority because it would allow pharmaceutical professionals building labels the ability to compare their work to existing labels using section search and comparison. The application’s previous iteration, the DLE, included a login functionality which is used to allow users to upload and save custom labels to their account. Once logged in, the user can upload their label in XML format where it is parsed according to the sections created for other drugs. Once they are uploaded, they appear in the navigation bar and can be compared with other labels. This can be seen in figure 2 below.

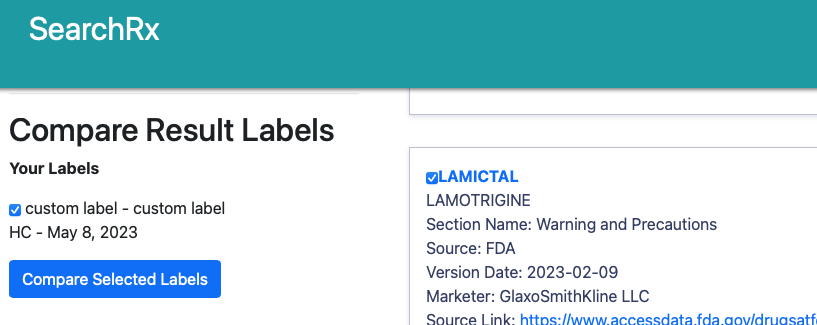


Figure 2: Custom Label Comparison

Enhancing the existing faceted navigation features was also determined to be a high priority. Faceted navigation is a filtering technique that allows users to reduce an existing search. The existing filters were improved upon by adding additional countries and the newly created section mapping. This can be seen in figure 3 below which demonstrates how the faceted navigation bar, on the left, can be used to display only HC labels created by MERCK CANADA .

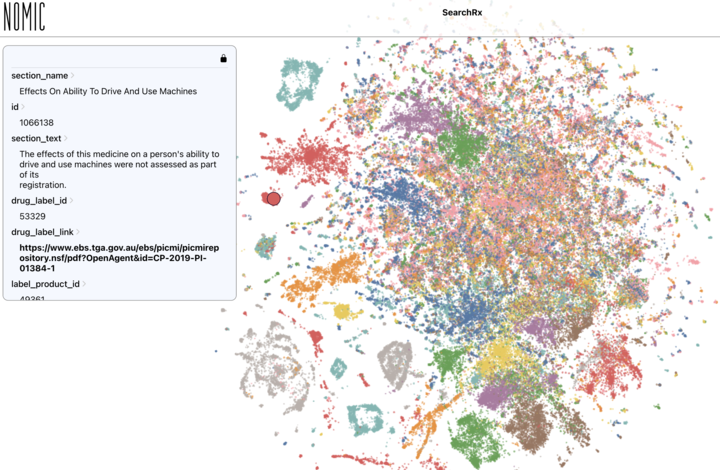
A screenshot of a computer

Description automatically generated with low confidence

Figure 3: Faceted Navigation

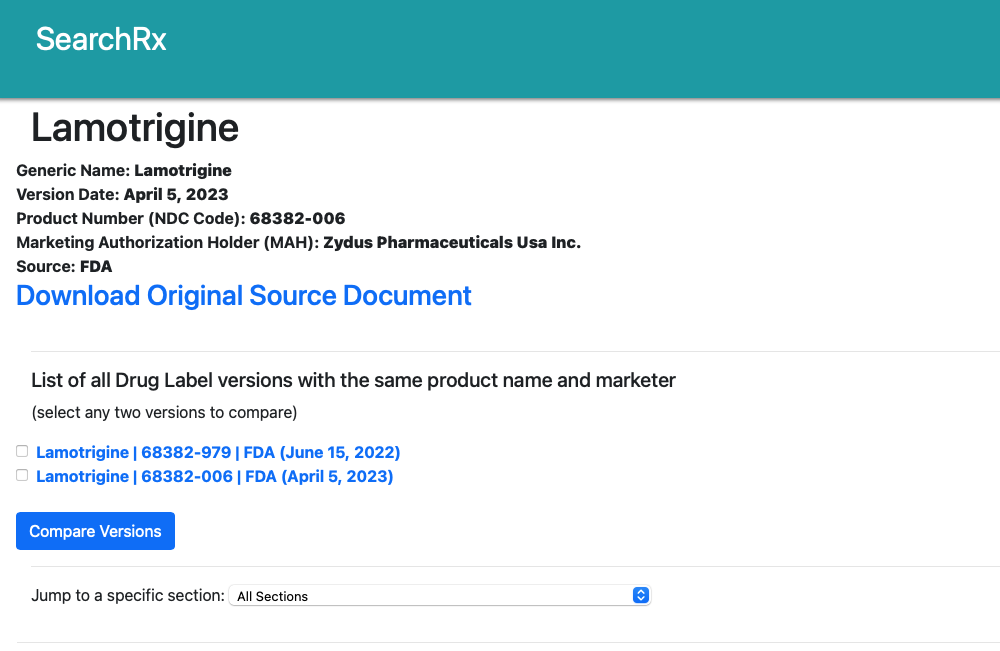
Building a semantic search feature was also determined to be a high priority. Providing predictive analysis creates enhanced user experience and allows users to find relevant labels more easily. With semantic search, the user’s query is transformed into a vector (array of integers) using a Natural Language Processing model called Bidirectional Encoder Representations from Transformers (BERT). This vector is then compared with existing data which allows SearchRx to map the query to relevant search terms. For example, if SearchRx is given the term “brain bleed,” it will return documents that contain “intracranial hematoma” because both reference the same underlying concept, even if they are not direct keyword matches.

Because of the large amount of data, a visualization feature was determined to be a moderate priority. This was created to help understand the section mapping process and uses a third-party application called Nomic to group similar sections according to their text. The interactive visualization is available on the site and demonstrates the challenge of section mapping. This can be seen in the many colors shown in figure 4 below which analyzes the contents of the sections and groups them by color.



*Figure 4: Nomic visualization of SearchRx*

To effectively allow users to gain insight into how an approved drug label changes over time, a version history analysis feature was determined to be a high priority. This can be seen when a user navigates a specific drug label page. The picture below shows previous versions of the same product name appear along with a comparison option.



*Figure 5: SearchRx Single Page View with Version History*

Finally, SearchRx runs scheduled weekly updates to ensure that recently released data is always available to users. To do this, the application re-scrapes the agency websites and only pulls labels that are not present in its existing database.

### Discussion

SearchRx’s features provide users unequaled access to drug label information and the ability to gain unique insight into the expectations of regulatory bodies in each country. This will benefit (1) pharmaceutical professionals working to bring drugs into a new market, (2) researchers working to provide more regulatory transparency and (3) doctors in countries without drug label databases looking to reduce the possibility of medication error.

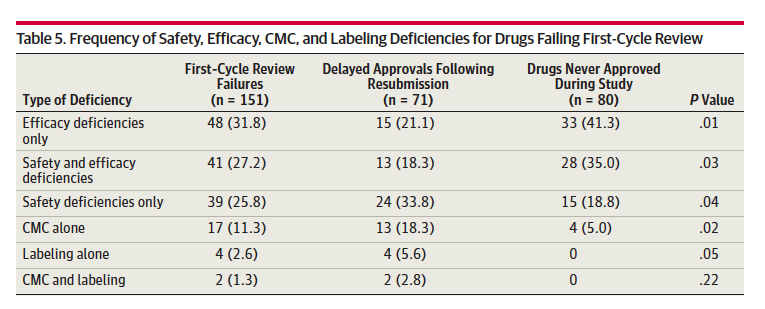
Pharmaceutical professionals can use searchRx’s custom label upload and section search and comparison features to understand the unique expectations of regulatory bodies, and therefore prevent costly delays. A 2014 study published by the American Medical Association found that labeling issues contributed to 3.9% of initial FDA rejections and 8.4% of delayed approvals following resubmission (Sacks 5). Furthermore, all drugs where labeling contributed were eventually approved showing that the errors were avoidable (Sacks 5). While the percent is relatively small, the implications are large. Drug approval is a costly process. Recent estimates put the average cost to approve a single drug at 935 million dollars (Wouters) and the mean delay time of 435 days when a drug is rejected (Sacks). SearchRx can be used to help prevent these labeling issues by providing custom label upload and section comparison. This will empower users to compare their own label with similar approved drugs from their intended regulatory body and make it easier to identify any lacking labeling requirements.

Figure 6: Highlights how labeling issues contribute to Label Rejection (Sacks 5)

In addition, researchers working toward transparency can use the SearchRx’s section search and comparison feature to more easily understand regulatory bodies. A 2014 study published in the New England Journal of Medicine describes policy changes that empowered EU citizens to request details of scientific studies used for a drug’s approval (Bonini). However, a 2022 study published in BMC Medicine details researchers attempts to re-create studies from this publicly available information and found that only ten of sixty-two EPARs contained enough information to reproduce studies (Siebert). SearchRx’s section comparison features will help these researchers more efficiently search through drug labels and possibly pinpoint where the EMAs regulatory criteria are lacking in comparison with other regulatory bodies. For example, if a drug were found lacking, they would be able to execute a search and comparison between different countries on similar approved drugs and instantly be able to determine if the EMA did not execute due diligence when compared to another country. Similarly, researchers in other countries could use it for the same purpose.

Finally, doctors and patients in countries that do not have searchable drug databases can use SearchRx’s international search capabilities to more easily find labels and therefore potentially reduce the chances for medication error. A medication error is defined by the FDA as a preventable situation where medication causes harm to a patient (Center for Drug Evaluation and Research) specifically lists Drug Labels as a way to combat Medication Error and affirms that these errors are very common events (Center for Drug Evaluation and Research). Therefore, in countries that do not have a regulatory approval process or a searchable drug label database there is a higher risk of medication error. In these countries, doctors have access to a wide range of drugs that may be approved in one or many of the countries in the SearchRx database. SearchRx can therefore help these doctors, and their patients, by allowing them to access approved human drug labels from multiple countries to find the right match and help them to reduce the potential for medicine error.

As a result, SearchRX’s functionality will serve an existing need to pharmaceutical companies, researchers, doctors and patients. Four countries of labels are currently included but can be readily modified to include additional countries. For this reason, SearchRX has been designed to make future development easier by using infrastructure as code to automate deployment on AWS services with low cost.

### Future Work

Additional work should include enhancing the scalability of the application. The current architecture is containerized, but horizontal and vertical auto-scaling could be configured to accommodate more users. In addition, the data loading and vectorizing currently takes over 24 hours to complete. This could be further optimized in numerous ways including containerizing the tasks. Finally, the inclusion of additional countries will enhance SearchRx. The UK could be added most easily due to the lack of a language barrier. However, it is possible that advancements in AI or the presence of translators with domain knowledge could lessen the difficulty presented with a language barrier in which case Japan, Brazil and China might be considered next.

**Conclusion**

The application in its current iteration accomplishes its primary goal of helping its users to search the contents drug labels and understand the requirements of regulatory bodies more easily. It provides users a free and open-source international database of drug labels with four countries and includes section search and comparison, semantic search, version history and custom label upload. As a result should be helpful to pharmaceutical professionals, researchers, doctors and patients and improved upon in future iterations.

**Bibliography**

“DailyMed - About.” *U.S. National Library of Medicine*, National Institutes of Health, <https://dailymed.nlm.nih.gov/dailymed/about-dailymed.cfm>.

“DrugBank Online: Database for Drug and Drug Target Info.” DrugBank Online | Database for Drug and Drug Target Info. Accessed April 1, 2023. <https://go.drugbank.com/>.

Ema. “Medicines.” *European Medicines Agency*, <https://www.ema.europa.eu/en/medicines>.

“The Internet Drug Index for Prescription Drug Information, Interactions, and Side Effects.” RxList. RxList. Accessed May 6, 2023. <https://www.rxlist.com/>.

Bonini, Eichler, H.-G., Wathion, N., & Rasi, G. (2014). Transparency and the European Medicines Agency — Sharing of Clinical Trial Data. The New England Journal of Medicine, 371(26), 2452–2455. <https://doi.org/10.1056/NEJMp1409464>

Brianna A da Silva and Mahesh Krishnamurthy, “The Alarming Reality of Medication Error: A Patient Case and Review of Pennsylvania and National Data,” Journal of community hospital internal medicine perspectives (U.S. National Library of Medicine, September 7, 2016), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5016741/>.

Center for Drug Evaluation and Research, “Working to Reduce Medication Errors,” U.S. Food and Drug Administration (FDA), accessed April 1, 2023, <https://www.fda.gov/drugs/information-consumers-and-patients-drugs/working-reduce-medication-errors>.

National Center for Toxicological Research. “FDALabel: Full-Text Search of Drug Labeling.” *U.S. Food and Drug Administration*, National Center for Toxicological Research, <https://www.fda.gov/science-research/bioinformatics-tools/fdalabel-full-text-search-drug-product-labeling#Additional%20Information>.

“Reed Tech Navigator™ for Drug Labels.” *Lexis Nexis | Reed Tech*, 7 June 2022, <https://www.reedtech.com/navigator-for-drug-labels/>.

Siebert, Gaba, J., Renault, A., Laviolle, B., Locher, C., Moher, D., & Naudet, F. (2022). Data-sharing and re-analysis for main studies assessed by the European Medicines Agency-a cross-sectional study on European Public Assessment Reports. BMC Medicine, 20(1), 177–177. <https://doi.org/10.1186/s12916-022-02377-2>

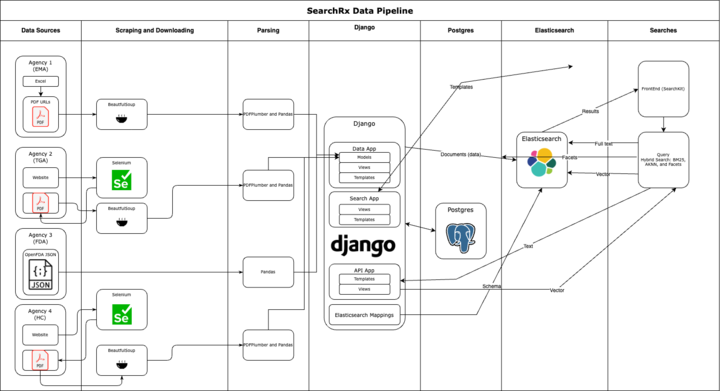
Sacks LV, Shamsuddin HH, Yasinskaya YI, Bouri K, Lanthier ML, Sherman RE. Scientific and Regulatory Reasons for Delay and Denial of FDA Approval of Initial Applications for New Drugs, 2000-2012. *JAMA.*2014;311(4):378–384. doi:10.1001/jama.2013.282542

Therapeutic Goods Administration (TGA). “Database of Adverse Event Notifications (Daen).” *Therapeutic Goods Administration (TGA)*, Therapeutic Goods Administration (TGA), 31 Mar. 2023, <https://www.tga.gov.au/safety/safety/safety-monitoring-daen-database-adverse-event-notifications/database-adverse-event-notifications-daen>.

Wouters, Olivier J et al. “Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018.” *JAMA*vol. 323,9 (2020): 844-853. doi:10.1001/jama.2020.1166

## System Design

SearchRx architecture and software decisions can be placed into seven categories: Infrastructure, Database, Elasticsearch, Comparison, Label Scraping/Parsing, UX and Frontend. This diagram provides a general overview of the SearchRx data processing pipeline, including the scraping, parsing, ingest, vectorization, and search steps.



### Infrastructure

The SearchRx team containerized the previous application, Drug Label Explorer (DLE), to facilitate local development. We used a Github organization and repo for version control, and wrote multiple Github Actions scripts to create continuous integration and deployment pipelines. The application’s prod environment is hosted on AWS and utilizes the following resources, which are provisioned with CloudFormation templates:

1 EC2 (Elastic Compute Cloud) instance, m6i.large with a 80 GB EBS volume

1 RDS (Relational Database Service) instance, t4g.medium

ECS (Elastic Container Service) cluster with 1 application task behind a Load Balancer

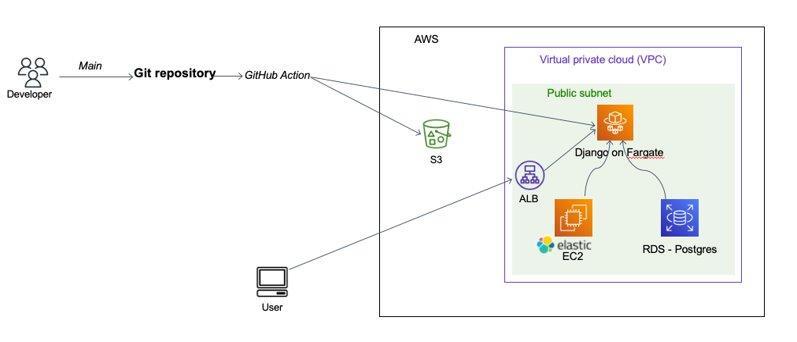
1 ECR (Elastic Container Registry) repository

1 Network Load Balancer with Target Groups

1 VPC with Public and Private Subnets

1 S3 Bucket

The diagram below shows the overall architecture design of the prod environment. Django is deployed on ECS and connects to Postgres (running on RDS database-as-a-service) and Elasticsearch (running on an EC2 virtual machine). Developers push code through a CI/CD pipeline triggered by GitHub Actions. When a pull request against main is created, the CI pipeline runs linting tests to ensure code formatting consistency; Pytest unit tests to test isolated, small units of functionality and prevent regressions; and Playwright end-to-end functional tests to test the API and frontend. If these tests pass, the pull request is marked for approval. After one or more manual approvals from a team member, the PR can be merged to the main branch. When a pull request which includes a change to the dle directory is merged to main, a deploy Github Action is triggered. The Action builds an image, pushes it to ECR, and updates the ECS task with the new image.



In addition, the infrastructure deployment itself is scripted using CloudFormation templates and a deploy pipeline (also in Github Actions) which can set up or modify the environment as necessary. Each element of the infrastructure is represented in the YAML templates except for the domain name registration and HTTPS setup. This can be triggered by making changes to individual CloudFormation templates or manually running the infra\_deploy GitHub workflow located in .github/workflows. Many individual configuration parameters and secrets are stored in AWS Parameter Store which can also be used to customize the deployment.

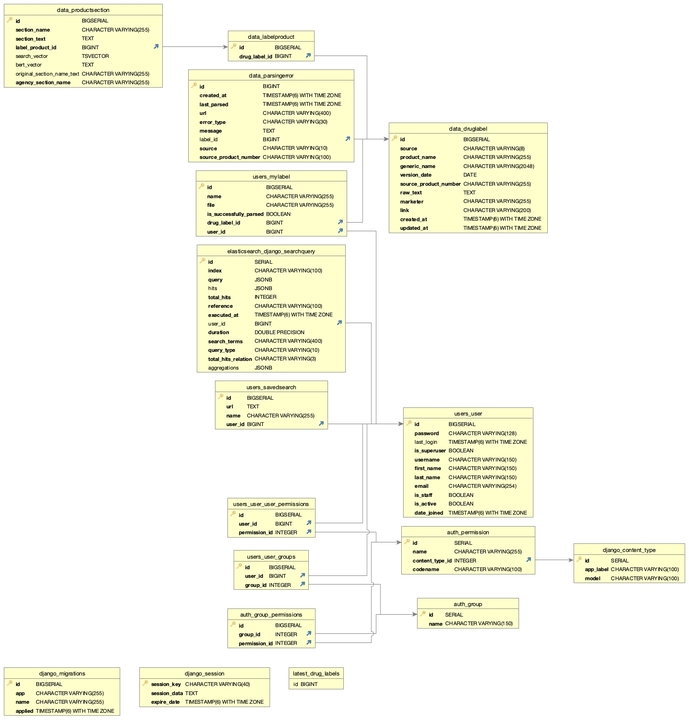
### Database

The previous DLE application used MariaDB for its sole, core database to power the Django web application. To improve the application’s performance and stability, the SearchRx team decided to migrate to Postgres, the recommended database for Django applications. Only minor changes were required to migrate from MariaDB to Postgres, specifically switching out a FULLTEXT field for Postgres’ tsvector field which is used by Django’s native SearchVectorField, and creating a custom trigger migration to update that vector when the source text changes. Even before Elasticsearch integration, changing out the database engine provided immediate improvements by reducing full text search query times to only a few seconds.

We kept the core DLE data model, but added several fields and a new model:

* data\_productsection.bert\_vector to store the PubMedBERT vector representation of the section full text
* data\_productsection.agency\_section\_name to store the agency name, repurposing data\_productsection.section\_name for the newly normalized metacategory
* data\_druglabel.created\_at and updated\_at to store when a drug label is added or updated, which we use to skip recently added labels if we need to kick off a re-ingest
* several new tables for elasticsearch\_django, a Python library we use to connect the web application to Elasticsearch
* a new model, ParsingError, which we use to capture errors encountered when parsing PDF, HTML, or JSON so that we can either skip labels with known issues, or revisit errors after improving the parsing logic

The UML diagram below shows the database design with its 17 tables.



### Elasticsearch

As the name implies, providing a robust drug label search experience is a major goal of SearchRx. While switching to Postgres for Django’s backend substantially improved the speed of searches on SearchRx, Postgres still limits how flexible and powerful those searches can be. Traditional keyword-based search algorithms such as TF-IDF (Term Frequency Inverse Document Frequency)[[1]](#footnote-1) or BM25[[2]](#footnote-2) measure how frequently a term appears in a document and normalizes that by the term’s frequency in the entire corpus, meaning uncommon words are more heavily weighted.

A priority for our client is implementing semantic search, a type of search which aims to understand the intent behind a query and get to the underlying meaning of the content being searched, rather than just matching for keyword hits. With this requirement in mind, Elasticsearch[[3]](#footnote-3) was included as a core component of the SearchRx tech stack specifically. SearchRx label data is still parsed and ingested into Postgres and is managed by Django, but it is also indexed into Elasticsearch for a better search experience. Elasticsearch added vector search capabilities in 2022 with the version 8.0 release which implements an approximate K-nearest neighbors’ algorithm to balance responsiveness with accuracy.[[4]](#footnote-4) [[5]](#footnote-5) Elasticsearch is robust and scalable; provides very fast responses (typically in milliseconds); supports full-text search and highlighting; and provides customizable search experiences through support for filtering, aggregations, and the Lucene query language.

Indexing existing data from Django into Elasticsearch was straightforward with the elasticsearch-django library. Because Elasticsearch is a NoSQL database, support for searching on relationships is limited, and it became evident that data needed to be denormalized during ingestion. Within Django, information about a single drug was split across multiple models, but Elasticsearch needed to retrieve and search from a single index. For this reason, SearchRx uses a single Elasticsearch index which includes a corpus of denormalized section documents combining the metadata from the drug\_label PSQL table with the vectors and section text from the productsection table.

After ingesting the data, we were able to achieve full text search query results across almost a million label sections in 10-20ms on a containerized version of Elasticsearch on a resource-bound development machine with search highlighting.

### Semantic Search

Semantic search was listed as a main requirement for searchRx and is meant to enhance the application by predicting a user’s intention when they search. For instance, a search for “brain bleeding,” in a semantic search engine would recognize that “cerebral hemorrhage” is a similar concept as “brain bleeding” and rank documents using that phrase highly. Semantic search typically tackles this problem by producing text embeddings, or high dimensional vector representations of the document. Semantic search engines can quickly perform vector math in a high dimensional space on indexed vectors normalized to unit length.

One difficulty encountered while implementing Semantic Search was selecting a model both trained on the correct domain data and tuned for text-embedding rather than other NLP tasks like fill-mask or text-prediction. Eventually, a sentence-transformers model pritamdeka/S-PubMedBert-MS-MARCO was used which “maps sentences & paragraphs to a 768 dimensional dense vector space and can be used for tasks like clustering or semantic search … in the medical/health text domain.”[[6]](#footnote-6) First, S-PubMedBERT-MS-MARCO was deployed via an NLP pipeline in Kibana, an Elasticsearch data analytics, visualization, and management tool. While the pipeline worked and the model loaded correctly, the vectorization job was extremely slow, taking about 25s per prediction even after allocating more resources and optimizing the Elasticsearch node configuration for an NLP pipeline. As a result, SearchRx pivoted to pre-computing the vectors based on the section text. These vectors are stored in Django as an JSONified string representing a list of floats and then indexed into the productsection documents into Elasticsearch, deserializing the vectors back into a list of floats along the way. Pre-computing vectors lowered prediction times to roughly 0.15s per section when computed in parallel. In total, vector-based searches take a few seconds, most of which is spent vectorizing the search query.

To implement semantic search, Elasticsearch requires an embedding to compare its index against, not just a search term. A vectorization service was added as the first component of the SearchRx data API so the text query could be vectorized and passed along as a vector to Elasticsearch. The initial implementation took roughly 35s to vectorize a query because the Django view was loading the S-PubMedBert-MS-MARCO model each time. Now, the model is pre-loaded during application startup and held in memory, which increases startup time but decreases vectorization time to ~2s, leading to a total search time of 2-3s. This is an increase over Elasticsearch’s native keyword search speeds, but still reasonably fast for semantic search, and both faster and more powerful than the original DLE application which did not have semantic search. More ways to improve search performance and reduce overall search time will be explored for Milestone 3.

### API

Adding an API for programmatic access of data was a business requirement for our client: technically-savvy analysts and scientists might want to export searches or specific labels from SearchRx for further analysis. The API has several endpoints: vectorize, search, search\_label, and searchkit/\_msearch. The searchkit/\_msearch (multi-search) endpoint is a proxy for Elasticsearch and is used by Searchkit for single page, AJAX-style search. The vectorize endpoint allows users to submit a search query and receive back a text embedding which was vectorized with the preloaded PubMedBERT model. The search endpoint allows users to programmatically search for drug label sections using a wrapper around Elasticsearch. Users can search any or all of the drug\_label\_generic\_name, drug\_label\_marketer, drug\_label\_product\_name, section\_name, and drug\_label\_source fields. The results are paginated, and users can page through results using the from and size URL parameters. The search endpoint also supports Elasticsearch’s simple\_query\_syntax[[7]](#footnote-7), which allowing for boolean operators such as +, -, |, or \*. The search\_label endpoint is relatively simple and allows users to directly find drug labels by searching against the drug\_label\_name field; this view searches against the Postgres database rather than Elasticsearch. Possible future improvements for the API include adding authentication via an API key tied to the user’s SearchRx user account; rate limiting based on the API key; and more advanced search filtering using semantic search or a hybrid search combining BM25 and semantic search.

### Front End

Initially, the team considered splitting the codebase into frontend and backend to use Django alongside a Javascript framework. However, it was decided that adding a frontend framework would introduce an un-necessary layer of complexity into the application. Ultimately, Django templates were decided on for the frontend which would allow the team to focus on making the single page view and the search portal more intuitive. Since SearchRx is a search engine at its heart, we decided to retool that page to use SearchKit,[[8]](#footnote-8) a Javascript widgets library specifically designed for Elasticsearch-based applications. SearchKit allowed us to quickly add complex filtering functionality as well as hybrid semantic search (keyword plus vector search) while still leaving us the flexibility to tune the searches through hooks. As noted above, we added a proxy to our API to protect Elasticsearch and remove the need to connect SearchKit to Elasticsearch via basic auth or API key.

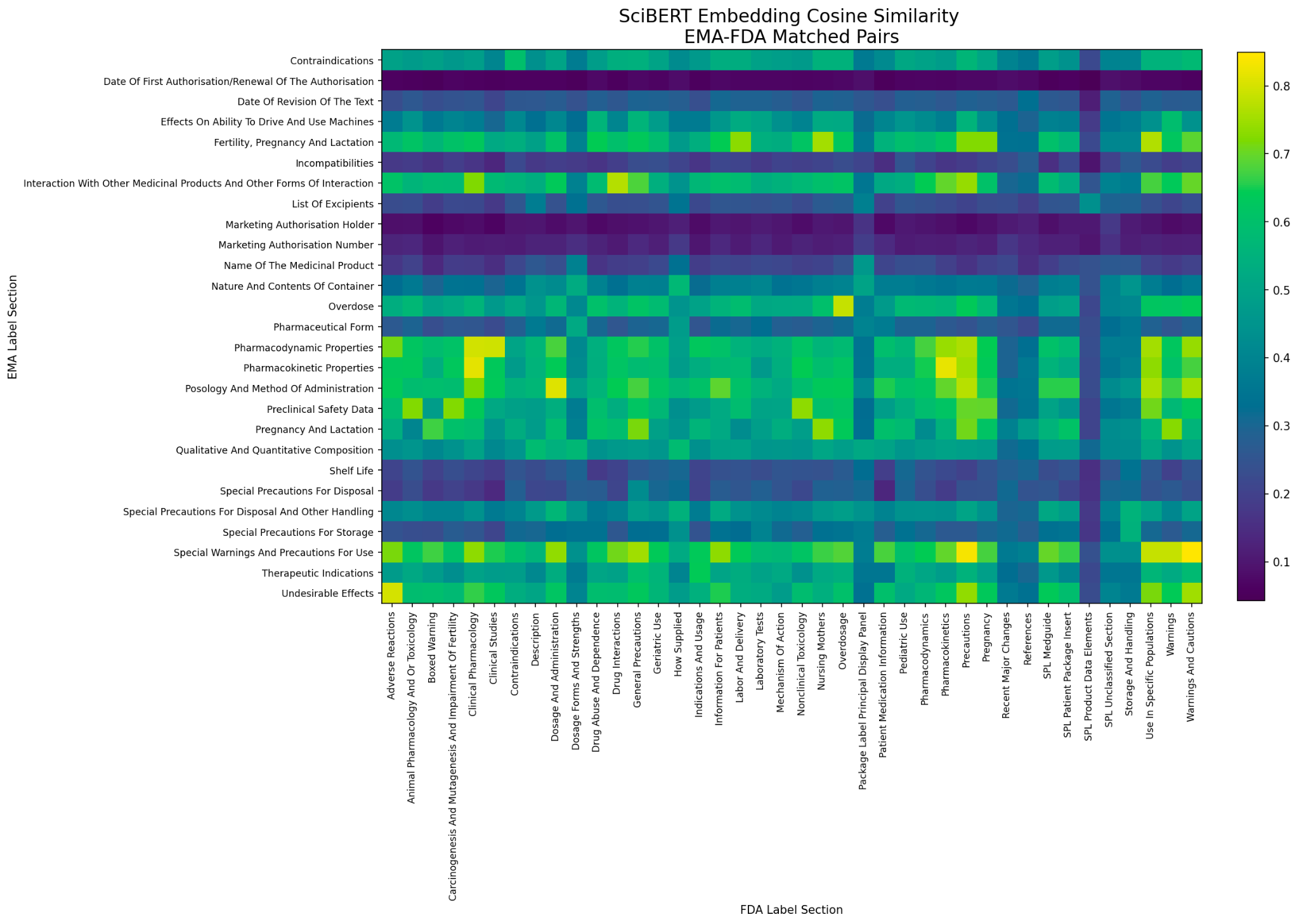
### Sub Section Comparison

Drug labels are organized into sections and subsections to provide specific information for safe and effective medication use. Comparing subsections requires matching drug labels from different countries, which can be challenging due to variations in metadata organization. Same information can be put under different section name. In addition, subsections from the same country can differ due to local regulations and guidelines. Therefore, comparing subsections in drug labels from different countries requires careful consideration.

From a front-end perspective, SearchRx solved this by providing a general search term that was mapped to multiple different subsections in the database. For example, the US, EU and Australia all have many different section names discussing side-effects (side effects, side-effects, adverse reactions etc…) but the user can search for all relevant subsections by selecting the side-effects tab.

To ensure the subsections are mapped accurately, the team set out to develop an automated solution and met with a Pfizer employee named Caleb Martin, who has been working on this problem. His code used a Jupyter Notebook to parse and analyze subsections from different countries and compare their similarity. He demoed the process to the team and was able to match nearly all of the EMA sections with FDA sections. He then shared his code and the SearchRx team adapted it to the data-loading section of the codebase.

Caleb’s code is able to compare all sections from two individual labels. The similarity of the matched section pair has been analyzed. The figure below shows that certain sections in EMA couldn't match with any FDA sections (e.g. Marking Authorisation Holder), while others (e.g. Pharmacodynamic Properties) showed a strong resemblance to multiple FDA sections.



Section comparison between two individual labels from EMA and FDA (From Caleb’s notebook). Section similarity is indicated by color, white suggests no matches.

### OpenFDA vs DailyMed

Drug labels in the United States are mainly maintained by FDA and NIH. They are stored in openFDA and DailyMed.

### Label Scraping/Parsing

**US (FDA)**

The previous year’s Drug Label Application used XML files from DailyMed to populate the database. However, it was determined that SearchRx should use openfda.gov because it has a more exhaustive list: OpenFDA contains over 200k labels while Daily Med has under 70k. In addition, OpenFDA encapsulates its data into JSON objects, while DailyMed keeps its data in XML files.

To load the data into the database, zip files were downloaded from the FDA database (<https://api.fda.gov/download.json>) and extracted. After that, the JSON objects were filtered based on the drug type and its approval status, then they were parsed and added to the database. Because OpenFDA uses the JSON format to organize the data, it makes it easier to parse with, compared to parsing PDF files.

**EU (EMA)**

Search RX populates its database with EPARs from the EMA (<https://www.ema.europa.eu/sites/default/files/Medicines_output_european_public_assessment_reports.xlsx>). The EMA maintains a searchable database of drugs on their website as well as a searchable spreadsheet with the pages for the individual drugs. Each page contains a link to the full EPAR document, in downloadable PDF form.

To load into the database, all 1300+ PDF EPARs of approved human drugs are downloaded using the excel spreadsheet. The total size of all the documents is under 1GB. The total time to download all documents was approximately 15 minutes. The average length of the PDFs were 59 pages. The EMA’s website standardizes the naming of these documents, so downloading them consisted of pulling the html page from each URL, isolating the link and downloading the PDF EPAR.

Once the PDF is downloaded, it is converted to a text file and parsed for headers using regular expressions that look for lines that start with section numbers (e.g., 1.1 or 2.3). This is an improvement from the previous DLE application which selects only 9 subsections. Taking this approach, allow all 20 plus subsections to be included in every EPAR. Finally, the Levenshtein distance is calculated to eliminate any invalid headers and the data is brought into the database.

**Australia (TGA)**

Scraping Australian labels is slightly different from parsing EU labels. There is no excel file that contains a list of URLs that point to individual drug labels. For AU labels, we needed to scrape the TGA website ([www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf](https://www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf/PICMI?OpenForm&t=pi&q=)) for the drug labels. One challenge that arose is the Access Terms page the application has to accept first. To solve this, Selenium is used to programmatically accept the terms by pressing the accept button. After the button is pressed, the cookies are stored and reused for downloading the rest of the labels.

Another difference between the AU labels and the EU labels is that not all AU labels are in the same format. For example, here are three different labels in different formats:

1. <https://www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf/pdf?OpenAgent&id=CP-2011-PI-01551-3>
2. <https://www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf/pdf?OpenAgent&id=CP-2009-PI-01102-3>
3. <https://www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf/pdf?OpenAgent&id=CP-2019-PI-01598-1>

The application’s parsing approach is that it will first try to parse the label like it’s in EU format (i.e., headers that clearly start with section numbers), and if it fails to parse with this approach, then it will fall back to a more rigid approach where it uses regular expressions to look for specific headers (i.e., DESCRIPTION or CONTRAINDICATIONS). Only a limited number of labels is in the third format above, if both approaches fail, then most likely the drug label is using the third format. The application is not able to handle that, so the drug information for that drug will be missed.

**Canada (HC)**

Similar to scraping Australian drug labels, we needed to scrape the HC website ([www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf](https://www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf/PICMI?OpenForm&t=pi&q=)) for the Canadian labels. Selenium is again used for selecting options (e.g., selecting status as “Approved”) and clicking the “Search” button. After the button is pressed, the cookies are stored and reused for downloading the rest of the labels.

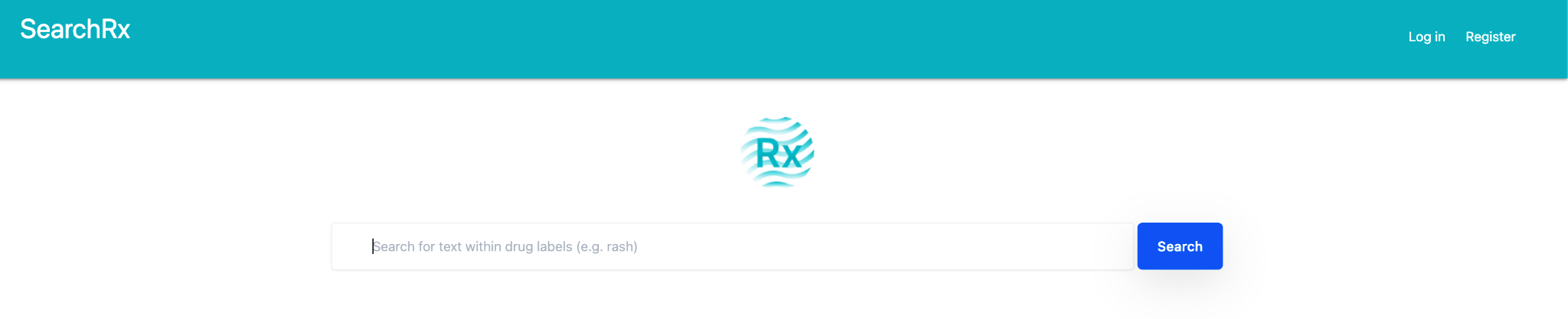
Parsing Canadian labels is more tricky and challenging than parsing the labels from other countries. Canadian labels are in at least three different formats. For example, here are three different labels in different formats:

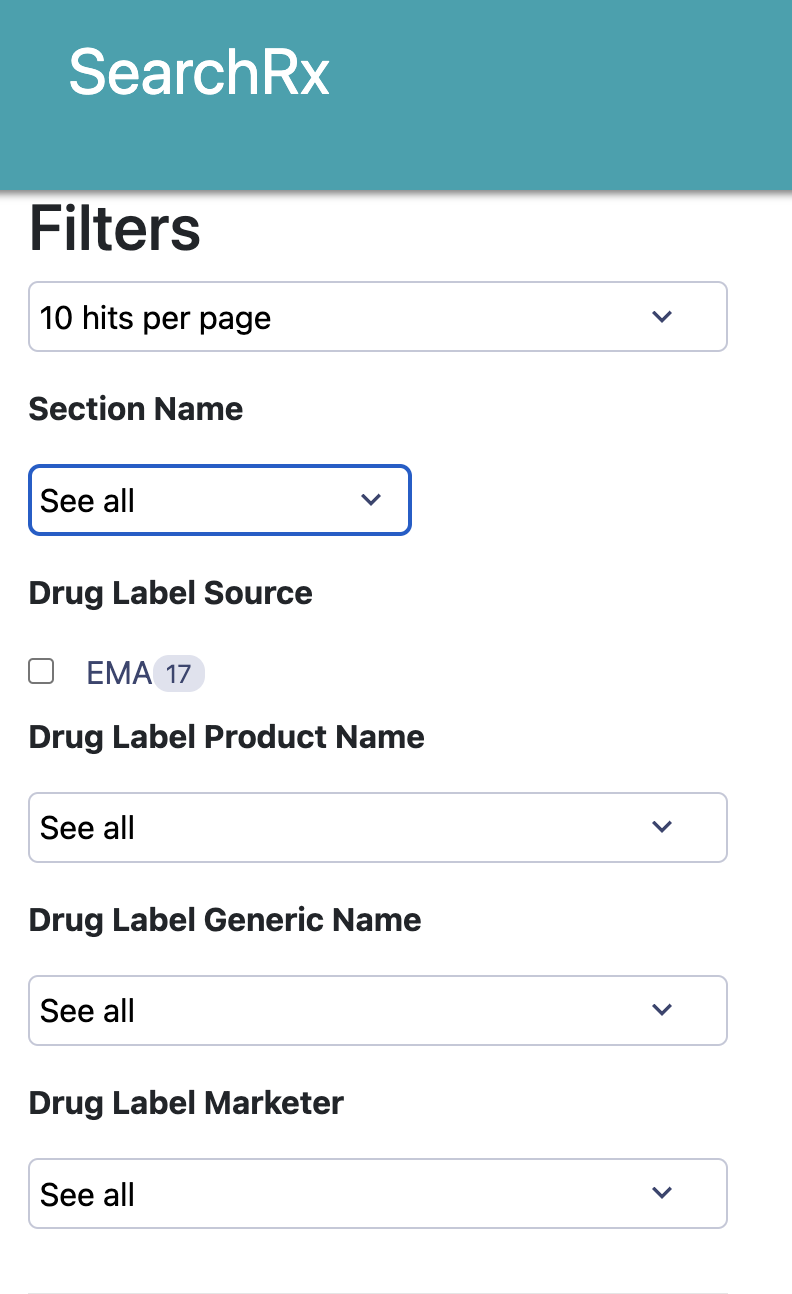
1. <https://pdf.hres.ca/dpd_pm/00062535.PDF>
2. <https://pdf.hres.ca/dpd_pm/00059995.PDF>
3. <https://pdf.hres.ca/dpd_pm/00003505.PDF>

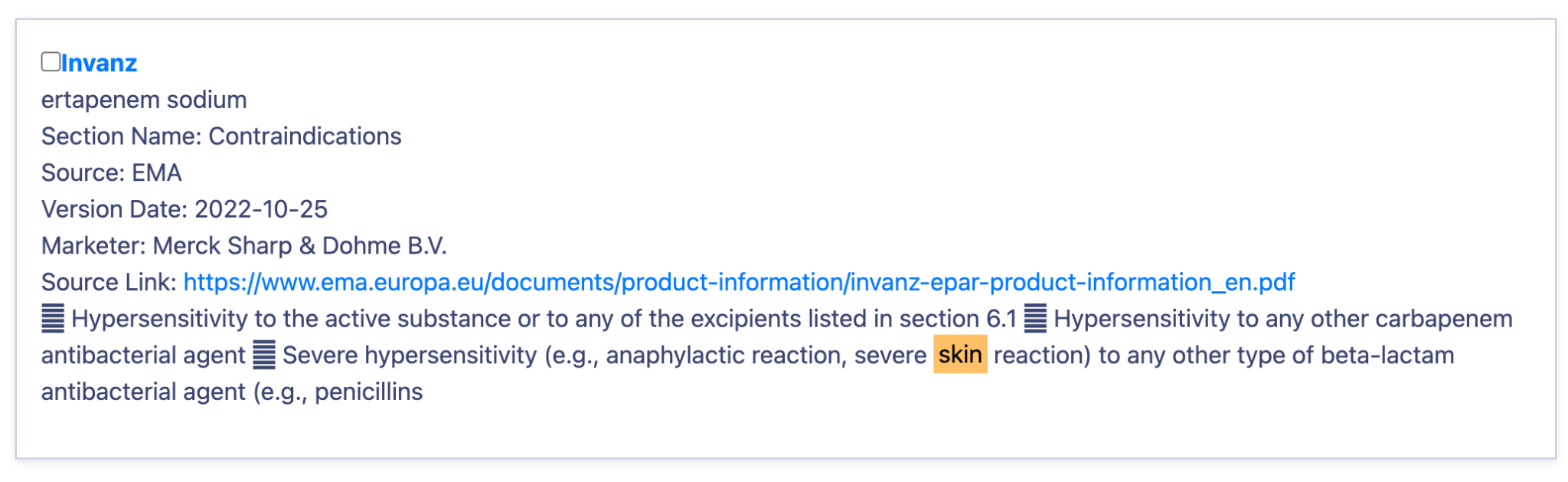
Because the Canadian labels have more variations, the parsing code requires more testing to account for all different cases. However, the parsing approach is similar to the approach for parsing AU labels. It will first try to parse the label like it’s in EU format, and if it fails to parse with this approach, then it will fall back to a more rigid approach where it uses regular expressions to look for specific headers.

### User Experience (UX)

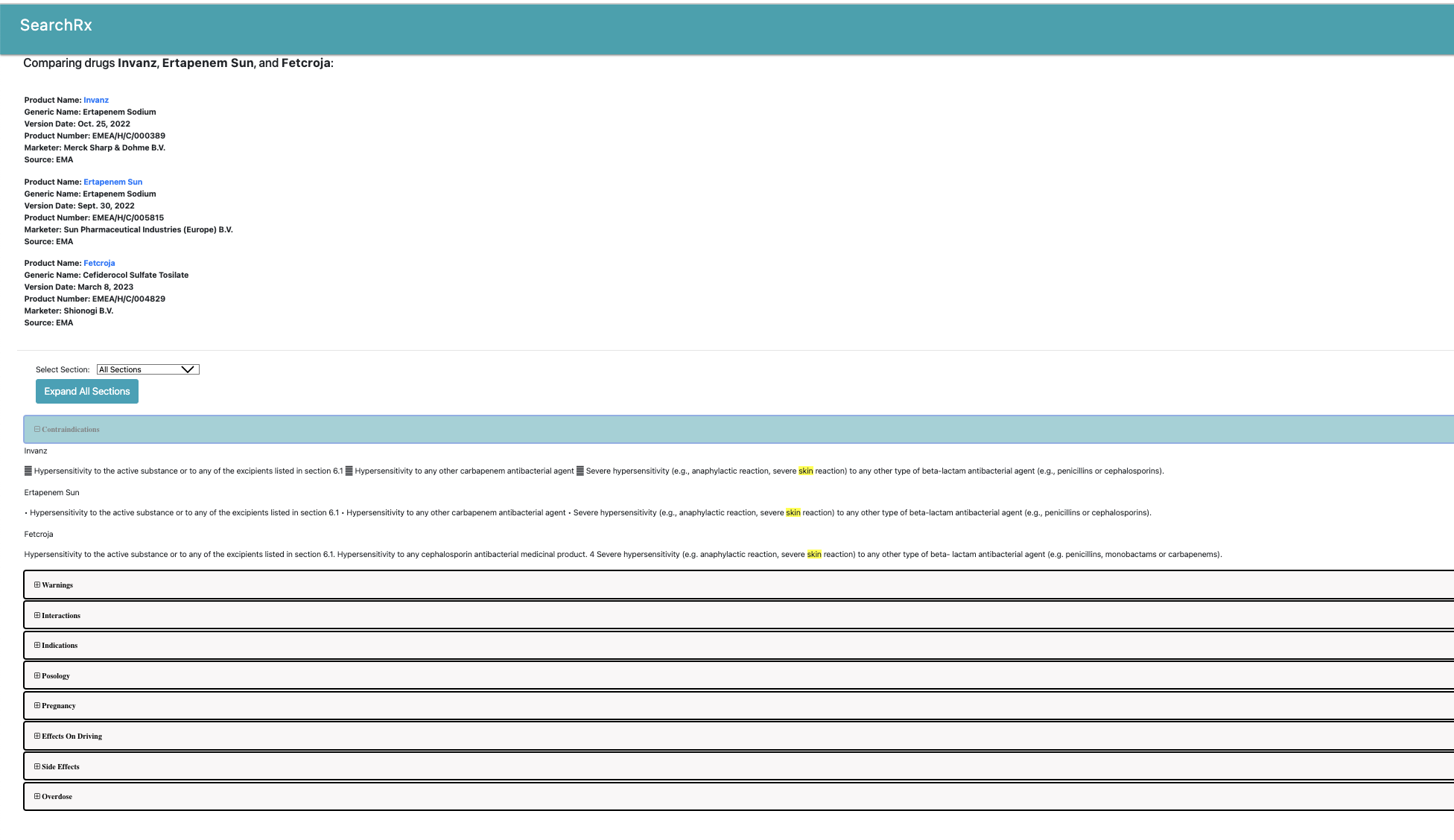
Improvements in search function UX aim to make it ease to use to find relevant results and refine their search queries. Some of the recent UX changes made to search functions include:



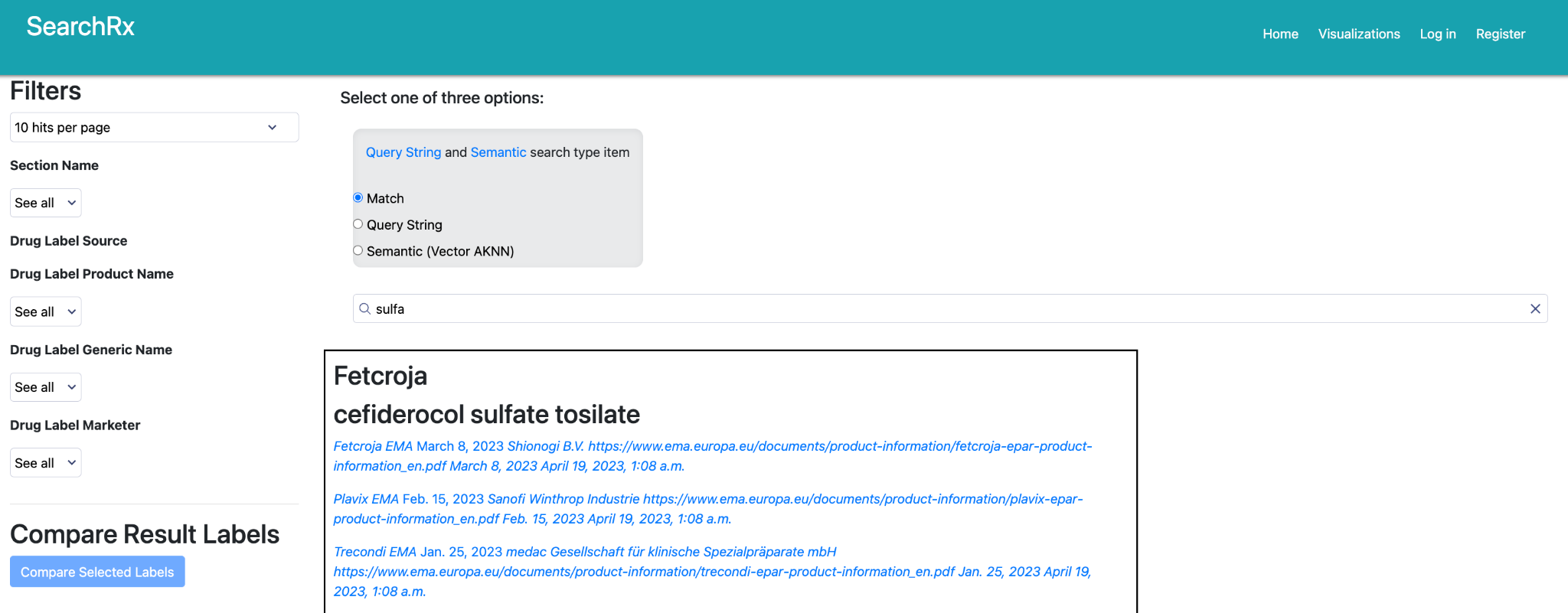
1. Simplified Navigation: The landing page has streamlined their navigation to make it easier for users to find what they need quickly. This includes reducing the number of dropdowns and grouping similar features together.
2. Customized Filters: Search function now offer filters and database options to help users refine their search queries and find specific types of results. A simple use case is to let users type their search queries. The customized filters and selection in dropdown helps users find relevant results faster and reduces the chances of spelling errors.
3. Rich snippets in card view(search result page): The current search result now displays rich snippets, which provide additional information about search results, such as links, . This helps users make more informed decisions before clicking on a result.



1. Personalization:For future development, search engines could use machine learning algorithms to personalize search results based on a user's search history and behavior. This helps users find relevant information more quickly.
2. Layout Optimization: With the rise of interface usage, the result page and compare page have optimized user experience for ease of use as well as visual clearness to easily identify the content user was looking for. This includes user-friendly design, faster load times, and simplified navigation. In addition, the comparison page was developed to allow users to compare labels more easily, the result page view is easy to use to see different results endlessly in vertical view instead of horizontally.



1. Accessibility: The landing page as well as result page have made efforts to improve accessibility for users with disabilities. This includes features such as color-blindness(drop usage of pure red and green), captions(tooltip), and color guided navigation.



To update the UX for SearchRx, several key changes were made. The team also met with the customer and discussed the various elements weekly to choose between vertical and horizontal view in the comparison page.

Further, the team worked to improve the landing result page, adding different components. An updated logo was created and a different style implemented.

The visual payout is customized to different types of user needs based on search type. The overall look and feel were changed to provide a more streamlined look.

Overall, these UX changes aim to make search functions more user-friendly and efficient, allowing users to find the information they need faster and with less effort.

## Tests

The DLE (Drug Label Explorer) included an existing test suite based on Python’s core unittest library. In addition to this, searchRx added a continuous integration pipeline using Github Actions to run the tests within ephemeral Docker containers for the Django application and Postgres. The pipeline also includes a linting test to ensure code consistency with the black, isort, and flake packages; team members use a Git pre-commit hook which lints their code locally before committing it to Github to reduce linting-only commits.

In addition, we migrated from Python's native unittest library to Pytest for writing and running unit tests. We also added Playwright, a powerful library for testing frontends and manipulating browser interactions in a manner similar to Cypress or Selenium. We used Playwright for end-to-end tests of the SearchRx frontend and data API. Further goals include increasing coverage for our end-to-end and unit tests, and updating the application's load tests.

## Development Process & Lessons Learned

Overall, the development process for SearchRX was a success. The client identified forty-five high priority requirements. Nineteen were completed during Miliestone 2 and twenty-four were completed during milestone three. In addition, several secondary requirements related to UX were fulfilled.

Over the course of the project, the requirements list experienced some changes. New requirements arose or were . One example is the Infrastructure pipeline created to automate the deployment of AWS resources which was requested to ensure the application would be easy to set up and maintain after the course concludes. Also, some requirements initially identified as high priority were lowered. One example of this is the user’s ability to include view the labels version history from the search results page.

To maximize productivity, the team used an Agile-like approach with daily check-ins on Slack and two scheduled meetings a week. At the Milestone 2 halfway mark, the team had an informal retrospective and discussed ways to improve. In addition, the team used JIRA to keep track of workflow, but general tasks were created instead of user stories. This more-strict Agile approach to help structure the work and ensure a more even distribution. It was implemented in Milestone 3 and included an estimation of individual user stories with story points and a weekly sprint cadence.

Overall, the work was distributed evenly, though the team largely self-selected into particular areas of work (frontend, data parsing and ingest, search, cloud computing, UX). This limited the amount members were able to contribute to areas outside their focus.

Our initial estimation of the requirements were accurate with a few notable exceptions. The team estimated that creating a version history feature to store multiple versions of labels over time would take only two hours to implement. However, because of some of the design changes, this would have taken significantly longer. As a result, the customer downgraded two elements in the features, and only required version history be included in the Single Label View. In addition, the requirements surrounding the storing of personal labels was complicated by the inclusion of ElasticSearch. These requirements were initially estimated at 4 hours but took much longer as we built out additional features to combine personal labels stored in Django with section data retrieved from Elasticsearch and display both on a single search page. Finally, the requirement to include the EU and FDA labels was estimated at 2 hours because it was included in the previous application. However, it took much longer because it was determined that the data source and parsing method needed to be enhanced. All these changes are reflected in the Appendix A requirements list.

Risks to the project include the following:

**Navigating complexity of the underlying technologies**

This risk was mitigated during Milestone 2, but is still inherent in the project. The project uses many different technologies so any future teams should take this into consideration.

**Logistical Coordination of Team**

This is an inherent risk when coordinating a team of six people. This was mitigated with coordination on the slack channel.

**Reliance on Government Sources**

Because searchRx uses web scraping to pull data from government sources, it is dependent on their websites having consistent formatting. This was somewhat mitigate because the work occurred in a short amount of time, but Future teams should be aware of this in case any of these sites are re-designed.

### Technologies Used

|  |  |  |
| --- | --- | --- |
| **Technology** | **Area Used** | **Justification** |
| **Python** | Web server and data processing | Best language for NLP and Statistical Analysis |
| **Django** | Web Application/ORM | Gives a lot of value with little effort |
| **Postgres** | Database | Suggested Database for Elasticsearch |
| **Elasticsearch** | Search Engine | Provides faster search and enables semantic search development |
| **Searchkit** | Front End | Allows us to build front end search interfaces on top of Elasticsearch for great search experiences |
| **Kibana** | Data Backend | Provides a user-friendly web interface for visualizing exploring and analyzing data stored in Elasticsearch |
| **Docker** | Web server and data processing | Allows us to easily develop locally and scale as needed |
| **Html JS and CSS** | Front end web | Web standard languages |
| **Htmx** | Web Server/Front end | Lightweight library that allows us to create dynamic and interactive experiences with html css and javascript |
| **Selenium and Beautiful Soup** | Data Scraping | Allows us to crawl websites to pull relevant data for Database |
| **GitHub** | Version Control | GitHub is used to host our code |
| **GitHub Actions** | CI/CD Pipeline | Adds the ability to easily build and lint deployments |
| **CloudFormation** | Infrastructure As Code | Deploys AWS Resources in a consistent, reliable manner |
| **Pytest** | Testing | Improves upon unittest; integrates with ptytest-docker to make the test environment similar to prod |
| **Playwright** | Testing | End-to-end / frontend testing |

## Appendix A: Reqs

The following table includes all requirements for SearchRx that have been implemented. References to Iteration 1 refer to the DLE Application. A full list of all requirements can be found in appendix F.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Id** | **Name** | **Category** | **Milestone** | | **Hours** | | **Classification** | **Notes** |
| 1 | Secure Login Added | UX | 3 | 3 | | Must Have | | (if custom labels included) |
| 2 | Comparison to view two  labels. As a user scrolls through the page, both sides of the view should be  in sync. | Compare | 3 | 5 | | Must Have | | Stack labels vertically instead of side by side, should be able to  handle more than two labels |
| 3 | Search results are automatically highlighted in the vertical  comparison view of the drug labels | Compare | 3 | 5 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 4 | Semantic Search is  implemented for all available countries using Elasticsearch | Data | 3 | 40 | | Must Have | | Synonym, bert model |
| 5 | API returns results in JSON format | Data | 3 | 15 | | Must Have | | Django will be used to  render endpoint |
| 7 | Australia - TGA Labels  Added to Searchable Database and parse data | Data | 2 | 30-60 | | Must Have | | Assumes Database is searchable |
| 8 | System has access to FDA Drug Label data from OpenFDA to include  the latest versions for all approved prescription drug labels | Data | 2 | 1 | | Must Have | | Took longer than estimated to implement enhancements.  OpenFDA is new datasource. |
| 9 | System accesses data from  EU data source (PDFs), to include the latest version for all prescription  drug labels | Data | 2 | 10 | | Must Have | | Took longer than estimated to implement enhancements. All  subsections now included. |
| 10 | Drug label data can be accessed using MedDRA terms | Data | 3 | 1 | | Must Have | | May not be needed because  of Elastic Search |
| 11 | Data from all data  sources is standardized using the Findable, Accessible, Interoperable,  Reusable (FAIR) principles. This will be done with a uniform schema and  search tools designed to directly interface with such. | Data | 3 | 1 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 12 | Ability to upload labels conforming to a supported typeFDA/XML,  EU/PDF; labels are only available to the single user (by default) | MyLabels | 3 | 1 | | Must Have | | Re-estimate needed because of Elastic Considerations |
| 13 | Ability to share saved  labels with other registered users; after selecting a label and choosing an  email address, the system will send an email with a link to a page in the  system that shows the label | MyLabels | 3 | 1 | | Must Have | | Re-estimate needed because of Elastic Considerations |
| 14 | Sharing user-uploaded drug label, grants access to the registered  user with the recipients email address | MyLabels | 3 | 1 | | Must Have | | Re-estimate needed because of Elastic Considerations |
| 15 | User-uploaded drug labels  show up in the user’s search results along with other drug labels; only the  user who uploaded the label or other users with whom the label was shared  have access | MyLabels | 3 | 1 | | Could Have | | Re-estimate needed because of Elastic Considerations |
| 16 | Ability to Save searches | MyQueries | 3 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 17 | Application is Scalable  and Able to Handle 10 concurrent users | Non-functional | 2 | 20 | | Must Have | | Previous iteration did have difficulty with over 3 |
| 18 | There is a website is available on the public internet that allows people to run queries on drug labels | Non-functional | 3 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 19 | Website is protected by  industry standard TLS encryption | Non-functional | 3 | 1 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 20 | Website can handle a small number of concurrent users - tens | Non-functional | 2 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 21 | Database with security  measures to protect the data stored in the database including encryption at  rest and encryption in transit (via SSL) | Non-functional | 3 | 5 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 22 | Ability to Search by Product (Generic and/or Brand Name) for all  countries | SearchForm | 2 | 15 | | Must Have | | Identify generic name of  each drug, and display |
| 23 | Ability to Search by Application number, DEA schedule, NDC, UNI code, SET ID for all countries | SearchForm | 2 | 1 | | Must Have | | Implemented in Iteration 1 with only US and EU. Should be adapted  for new Countries |
| 24 | Ability to Search by drug Manufacturer for all countries | SearchForm | 2 | 8 | | Must Have | | Estimate assumes some  work was done in previous iteration |
| 25 | Ability to Search within  Label Section for all countries | SearchForm | 2 | 8 | | Must Have | | Was not implemented properly |
| 26 | Ability to perform wildcard search on drug label data when searching within drug label categories for all countries | SearchForm | 2 | 1 | | Must Have | | Implemented in Iteration  1 with only US and EU. Should be adapted for new Countries |
| 27 | Main page of the application has a SearchForm area that includes the functionality for searching the Drug Labels. In general this can include drop-downs, checkboxes, text fields, etc. | SearchForm | 2 | 1 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 28 | Ability to have multiple search criteria. Ability to apply up to 5 search criteria with AND operators. | SearchForm | 2 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 29 | After the search is  executed, the search results are displayed to the user. The search results view should display a list of the matching drug labels. The search results may be paginated when they exceed a specified number of drug labels. | SearchResults | 2 | 1 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 30 | In the SearchResults there is the ability to select two labels.  After selecting two labels, the user can then compare the labels. | SearchResults - Compare | 3 | 10 | | Must Have | | Stack labels vertically  instead of side by side, should be able to handle more than two labels |
| 31 | Ability to navigate to  the VersionHistoryView from the SearchResults view | SearchResults -  VersionHistory | 3 | 1 | | Could Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 32 | A details page for the drug label is shown after the user clicks on  an item from the search results. | SingleLabelView | 3 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 33 | The Search query  parameters used in the search are highlighted in the SingleLabelView | SingleLabelView | 3 | 1 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 34 | Ability to navigate to the VersionHistoryView from the  SingleLabelView | SingleLabelView - VersionHistory | 3 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 35 | Drug label search functionality is available to an unregistered / guest / null User | Users | 3 | 1 | | Must Have | | Completed in Iteration 1. Must launch Iteration 1 code base to  complete |
| 36 | Basic user authentication including sign-up, sign-in, and password  reset using email allows for additional features such as uploading labels,  saving queries, etc. | Users | 3 | 1 | | Must Have | | Completed in Iteration 1.  Must launch Iteration 1 code base to complete |
| 37 | Section (subsection)  Search is implemented across countries, sections are mapped to one another | UX | 3 | 30 | | Must Have | | Will include indication subsection |
| 38 | The users should be able to easily determine what product they are  viewing. The Generic Name, Product Name and Dosage/Type should be clearly indicated. | UX | 2 | 15 | | Must Have | | In the search or the  search results, we should know what name, dosage and strength. |
| 39 | App should be able to  update labels seamlessly. Weekly updates. | UX | 3 | 2 | | Must Have | | Should always show the most recent |
| 40 | Users can see the key words in the context of their keyword search.  Truncated section is shown. | UX | 3 | 5 | | Must Have | |  |
| 41 | Add PDF link on single  label view | UX | 2 | 2 | | Must Have | | Estimate assumes link is in database |
| 42 | A Version History View page is displayed showing changes to a drug  label over time | VersionHistory | 3 | 1 | | Could Have | |  |
| 43 | Search results  automatically highlighted in the version history page | VersionHistory | 3 | 1 | | Could Have | |  |
| 44 | Comparison page to Change From Horizontal to Vertical | UX | 3 | 10 | | Could Have | | DL-100 |
| 45 | The users should be able to easily determine what product they are viewing. The Generic Name, Product Name and Dosage/Type should be clearly indicated | UX+Front end | 3 | 5 | | Must Have | | DL-66 |
| 46 | Main page of the application has a SearchForm area that includes the functionality of searching the Drug Labels. In general this can include drop-downs, checkboxes, text fields | UX+Front end | 3 | 5 | | Could Have | | DL-55 |
| 47 | A details page for the drug label is shown after the user clicks on an item from the search results. | UX + front end | 3 | 5 | | Must Have | | DL-58 |
| 48 | dropdown menu UI  change the frontend of section filter close to the mock | UX+Front end | 3 | 10 | | Could Have | | DL-140  DL-63 |
| 49 | Compare search result vertically | UX+Front end | 3 | 5 | | Must Have | | DL-142 |
| 50 | Change differences color from very bright red to like a lighter pink | UX+Front end | 3 | 2 | | Could Have | | DL-199 |
| 51 | Direct match styling | UX | 3 | 3 | | Must Have | | DL 207 |
| 52 | Add a modal with search help information next to the search options | UX+Front end | 3 | 2 | | Could Have | | DL-215 |
| 53 | Style search type toggle | UX + Front end | 3 | 3 | | Must Have | | DL-212 |

## 

## Appendix B: Journal Publication Plans

We have identified several potential journals in which we could publish our research on collecting international drug labels and making them searchable and comparable.

*Pharma Analytica Acta,* published by Edelweiss Publications,is our top choice as it seems to be the most realistic publication possibility in terms of cost and prestige. It is a double-blind, peer reviewed, and Open Access publication that publishes a wide variety of pharmaceutical research, and specifically mentions “pharmaceutical technology” as a branch of pharmaceutical research in which its editors are accepting research. The journal has a broad definition of pharmaceutical technology, describing it as the “application of scientific knowledge or technology to pharmacy, pharmacology, and the pharmaceutical industry”[[9]](#footnote-9). The publication fee for *Pharma Analytica Acta* is $500.[[10]](#footnote-10)

Another choice would be to publish in a *J-STAGE* journal. *J-STAGE* is a platform for scholarly publications in Japan; because our database is adding Japanese PMDA labels, we could focus on a smaller regional journal. *Translational and Regulatory Sciences* is an international open access journal published by Catalyst Unit whose scope includes “compound and library management,” “data transparency and sharing,” “informatics,” and “new standards and approaches to facilitate sound and transparent regulatory decision-making regarding drugs and medical devices.”[[11]](#footnote-11) The submission charge and publication fee are “free as a temporary arrangement.”

Another potential publication venue is the *Journal of Labeled Compounds and Radiopharmaceuticals*. This journal is published by Wiley Analytical Science with a 2-year impact factor of 1.949. The *Journal of Labeled Compounds and Radiopharmaceuticals* focuses on “all aspects of research dealing with labeled compounds,” though it primarily seems focused on the chemical and pharmacological methods used to produce such compounds as it is the official journal of the International Isotope Society.[[12]](#footnote-12) The cost of submitting is free, with an optional Open Access fee of $3450.

A final possible journal is one from the *Journal of the American Medical Association* network. *JAMA Health Forum* is an “international, peer-reviewed, online, open access journal that addresses health policy and strategies affecting medicine, health, and health care.”[[13]](#footnote-13) It receives roughly 1.8 million article views and downloads and almost 1 million online visits annually, and has an impact factor of 0.8. Article processing charges are $3000, and combined with the journal’s fairly competitive nature it is unlikely we will publish here.

## Appendix C: Team Strategy and Dynamics

The team will use Kanban based Agile methodology for task management and will have weekly meetings together and with the customer. The team will work to agree on the assigning of Jira tasks and slack will be used for daily communication. All team members have agreed to check slack and email daily. In addition, all major decisions will be voted on. In the event of a tie, a member of the teaching staff will be consulted. More details can be found in the team contract detailed in Appendix A.

## Appendix D: Team Contract

The international drug label team agrees to the following:

We expect everyone will do their best.

Disagreements are resolved by voting. In the scenario of a 3-3 tie, Simile is the tiebreaker.

We communicate asynchronously on slack and we meet at least once a week and have standing times on Sunday and Tuesday. Our expectations for asynchronous slack communication is that everyone checks the channel at least once per day.

Everyone will try their best to attend each meeting, but we understand that things happen. And if you can’t attend a meeting, it is expected that you send a virtual update.

Product decisions are documented as stories in jira and general team decisions are documented in Google Drive documents.

We have two standing meetings each week. One on Sunday evening and one on Tuesday evening. Our general plan is to commit to at least one meeting each week

## Appendix E: Country Overview and Selection Process

**European Union (included)**

The EU’s pharmaceutical governing body was established 1995 as the European Medicines Evaluation Agency (EMEA) and was renamed the EMA in 2009[[14]](#footnote-14). The organization’s goal is to streamline the drug approval process and reduce the time and effort needed for pharmaceutical companies to sell products in the EU[[15]](#footnote-15). The EMA equivalent to an FDA drug label is called a European Public Assessment Report Product Information Document (EPAR)[[16]](#footnote-16). EPARs include three sections and provides detailed information about past studies, drug characteristics and product labeling restrictions. The first section, ANNEX I, is titled the Summary of Product Characteristics (SmPC), and includes an in-depth discussion of dosage recommendations, negative side effects, drug interactions and Clinical Study Results. The second section ANNEX II is shorter and includes information about the manufacturer and regulations. Finally, the third section, ANNEX III, is titled Labeling and Package Leaflet (PL) and details how to represent the drug to the public.

The EMA maintains a searchable database of drugs on their website. When a user selects a drug in the database, they are brought to a summary page. On that page is a link to the full EPAR, in downloadable PDF form. While the EPARs are standardized, they frequently experience changes due to regular informational updates and EMA mandated format changes. This can be seen in a 2017 report produced by the EMA which recommends creating a new system to help make changes to EPAR formatting best practices, adding a ‘key information’ section to the PL and SmPC, digitizing the PL and the SmPC and providing more accommodation for patient input in the testing of the PL [[17]](#footnote-17). Therefore, the EPARs issued before and after this guidance will likely differ.

Like the previous version, DLE, it was determined that SearchRx should include all human approved EU labels. However, it also includes additional information. The EPAR subsections vary, but most include over twenty-five. The Drug Label Explorer only uses 9 subsections of the EPAR[[18]](#footnote-18) but SearchRx uses all 25. This enhancement should provide a more granular view of the labels and increase the effectiveness of the application.

**United States (included)**

The Food and Drug Administration regulates the drug approval process in the United States. There are many sources for US Labels including the NIH’s Daily Med and OpenFDA.gov. The previous DLE application used DailyMed a source for FDA labels[[19]](#footnote-19) which is a database managed by the National Library of Medicine and provides FDA labels in an .xml format[[20]](#footnote-20). Testing indicated under 70k labels available on DailyMed. This much less than openFDA.gov which is managed by the FDA and provides open source APIs to pull labels[[21]](#footnote-21). Testing indicated over 200k labels are available in PDF format.

As a result, SearchRx uses OpenFDA.gov labels. This required a significant overhaul of the import and updated functionality but was deemed worth the effort.

**Japan (not included)**

The primary Japanese government health authority that regulates pharmaceuticals is the Pharmaceuticals and Medical Devices Agency, or PMDA. The PMDA was established in 2004 and “conducts scientific reviews of marketing authorization applications of pharmaceuticals and medical devices” .[[22]](#footnote-22) The PMDA is “in charge of reviewing drugs and medical devices, overseeing post-market safety, and providing relief for adverse health effects[[23]](#footnote-23) and is in charge of regulating Japanese drug labels, which are referred to as package inserts, or “tempu bunsho.” These documents are prepared by local entities called Marketing Authorization Holders, or MAHs. Japan is the third largest pharmaceutical market in the world at $109 billion in 2020, its own pharmaceutical sector and drug discovery ecosystem is growing rapidly, and all major US pharmaceutical companies have Japanese subsidiaries.[[24]](#footnote-24) Japan is also a leader in drug eLabeling, and the Japanese PMDA database is potentially a rich store of drug label data.

While, these factors make Japan a desirable candidate for inclusion into SearchRX it is not included due to it’s lack of English labels. Automating translation is erroneous and could result in bad information being put forward. Considering bad information could be used to make healthcare decisions, it was not deemed too much of a risk. In addition. It is likely that e-labeling in Japan will increase the amount of English translations for drug labels available on the PMDA website, and the PMDA is actively trying to promote English translations for better international information sharing. Therefore, Japanese labels should be considered in the near future when automated translation gets better or the PMDA includes a greater percentage of English labels.

**China (not included)**

Drug labels in China are regulated by the National Medical Products Administration (NMPA), formerly known as China Food and Drug Administration (CFDA).[[25]](#footnote-25) This regulatory agency has a broad range of responsibilities including registering, licensing, and approving medical products, as well as monitoring, testing, and inspecting these products.[[26]](#footnote-26) The NMPA also establishes standards and regulations for the production, development, and distribution of medical products within China. To locate drug labels, individuals can use the search engine maintained by the NMPA's Center for Drug Evaluation at [http://www.cde.org.cn](http://www.cde.org.cn/). Additionally, drug labels are available for search on the China Medical Information Platform at [http://www.dayi.org.cn](http://www.dayi.org.cn/) or some third-party websites, for example DXY Drugs information at [http://drugs.dxy.cn](http://drugs.dxy.cn/). Another distinctive feature of drug labeling in China is the use of QR codes. Since 2017, drug manufacturers have been required by the NMPA to include QR codes on their product labels. These codes enable consumers to scan the label with their smartphones and obtain additional information about the drug, including its manufacturing and distribution records. This system aims to enhance transparency and deter the sale of counterfeit drugs.

While there are many benefits for including Chinese labels in SearchRX, the search results are primarily in Chinese and automating translation is erroneous and could result in bad information being put forward. As a result, it is too much of a risk to include but should be reconsidered if automated translation becomes more reliable or English labels become available.

**United Kingdom (not included)**

The UK had previously followed guidance and regulations on drugs and medical devices made by the European Medicines Agency (EMA) before the UK formally withdrew from the European Union on January 31, 2020. Because changing regulations involved lots of moving parts and stakeholders, as a transitory measure, the EU pharmaceutical laws remained in effect in the UK until December 31, 2020. Since January 1, 2021, the Medicines and Healthcare products Regulatory Agency (MHRA) has taken over the role of EMA and become the sole regulator on all medicines and medical devices in the UK, except for Northern Ireland.[[27]](#footnote-27) Northern Ireland remains an exception because it still follows some of the laws from the EU, based on the Northern Ireland Protocol.

MHRA is the agency that approves all the marketing authorizations (MAs). A pharmaceutical company needs to apply for a MA and get approved, before they can market the medicine in the UK.[[28]](#footnote-28) The application process requires the company to submit a document called summary of product characteristics (SmPC), which describes the medicine in great details. The SmPC needs to include all information of the medicine based on the product research and knowledge, which means including information like clinical parameters, dosage, warnings, side effects, shelf life and etc.[[29]](#footnote-29) Since SmPC contains extensive information about the medicine, it can be used as a good reference when there is any doubt about the drug (e.g., whether it is safe to use when pregnant).

MHRA maintains a database (<https://products.mhra.gov.uk/>), which contains the most up-to-date SmPCs for drugs sold in the UK, and the database is accessible by public for free. However, scraping the website presents some challenges, and while UKs inclusion in SearchRX would be beneficial it was deemed below Australia and Canada in priority. As a result, the UK is not included in the SearchRX Database.

**Brazil (not included)**

ANVISA is Brazil’s Health Regulatory Agency and was created in 1999. In 2016, ANVISA joined the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use - also known as ICH. This means that its standards will need to align with international standards more and more over time.

Brazilian drug labels are searchable online, but are not in English. As a result, it is not included in SearchRX. It should only be considered when automation get more reliable or ANVISA starts including labels in English.

**Canada (included)**

In Canada, all natural health products (NHPs) are regulated by Health Canada (HC) under the Food and Drugs Act and the Natural Health Product Regulations.[[30]](#footnote-30) Health Canada requires that all drug labels provide information about the name of the drug, its dosage form, and its strength, as well as the route of administration and address of the manufacturer/sponsor of the drug. In addition, drug labels must include information about the indications for use, recommended dosages, and any contraindications or warnings about potential adverse effects or drug interactions.[[31]](#footnote-31)

All authorized products undergo pre-market assessment for safety, efficacy, and quality and the degree of pre-market oversight varies depending on the risk of the product. The Canadian version of a drug label is a Product Monograph which provides detailed information about a drug's safety, efficacy, pharmacology, and clinical use. The NHP provides a searchable online database on Product Monograph’s and is included in SearchRx.

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## 

## Appendix F: Initial Requirement List

The following table includes all requirements regardless of classification. Note: Iteration 1 refers to the previous project the Drug Label Explorer (DLE)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **Category** | **Milestone** | **Hours** | **Classification** | **Notes** |
| 1 | Secure Login Added | UX | 2 | 3 | Must Have | (if custom labels included) |
| 2 | Comparison to view two labels. As a user scrolls through the page, both sides of the view should be in sync. | Compare | 2 | 5 | Must Have | Stack labels vertically instead of side by side, should be able to handle more than two labels |
| 3 | Search results are automatically highlighted in the vertical comparison view of the drug labels | Compare | 2 | 5 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 4 | Semantic Search is implemented for all available countries using Elasticsearch | Data | 3 | 40 | Must Have | Synonym, bert model |
| 5 | API returns results in JSON format | Data | 2 | 15 | Must Have | Django would render the json endpoint. |
| 6 | Japan - PMDA Labels Added to Searchable Database and parse data | Data | 2 | 30-60 | Must Have | Assume Labels in English are possible. Must have unless blocker. |
| 7 | Australia - TGA Labels Added to Searchable Database and parse data | Data | 2 | 30-60 | Must Have | Assumes Database is searchable |
| 8 | System has access to FDA Drug Label data from DailyMed (SPL/XML), to include the latest versions for all approved prescription drug labels | Data | 2 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 9 | System accesses data from EU data source (PDFs), to include the latest version for all prescription drug labels | Data | 2 | 10 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 10 | Drug label data can be accessed using MedDRA terms | Data | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 11 | Data from all data sources is standardized using the Findable, Accessible, Interoperable, Reusable (FAIR) principles. This will be done with a uniform schema and search tools designed to directly interface with such. | Data | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 12 | Ability to upload labels conforming to a supported typeFDA/XML, EU/PDF; labels are only available to the single user (by default) | MyLabels | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 13 | Ability to share saved labels with other registered users; after selecting a label and choosing an email address, the system will send an email with a link to a page in the system that shows the label | MyLabels | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 14 | Sharing user-uploaded drug label, grants access to the registered user with the recipients email address | MyLabels | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 15 | User-uploaded drug labels show up in the user’s search results along with other drug labels; only the user who uploaded the label or other users with whom the label was shared have access | MyLabels | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 16 | Ability to Save searches | MyQueries | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 17 | Application is Scalable and Able to Handle 10 concurrent users | Non-functional | 3 | 10-Feb | Must Have | Previous iteration did have difficulty with over 3 |
| 18 | There is a website is available on the public internet that allows people to run queries on drug labels | Non-functional | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 19 | Website is protected by industry standard TLS encryption | Non-functional | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 20 | Website can handle a small number of concurrent users - tens | Non-functional | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 21 | Database with security measures to protect the data stored in the database including encryption at rest and encryption in transit (via SSL) | Non-functional | 1 | 5 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 22 | Ability to Search by Product (Generic and/or Brand Name) for all countries | SearchForm | 1 | 15 | Must Have | Identify generic name of each drug, and display |
| 23 | Ability to Search by Application number, DEA schedule, NDC, UNI code, SET ID for all countries | SearchForm | 1 | 1 | Must Have | Implemented in Iteration 1 with only US and EU. Should be adapted for new Countries |
| 24 | Ability to Search by drug Manufacturer for all countries | SearchForm | 1 | 8 | Must Have | Estimate assumes some work was done in previous iteration |
| 25 | Ability to Search within Label Section for all countries | SearchForm | 1 | 8 | Must Have | Was not implemented properly |
| 26 | Ability to perform wildcard search on drug label data when searching within drug label categories for all countries | SearchForm | 1 | 1 | Must Have | Implemented in Iteration 1 with only US and EU. Should be adapted for new Countries |
| 27 | Main page of the application has a SearchForm area that includes the functionality for searching the Drug Labels. In general this can include drop-downs, checkboxes, text fields, etc. | SearchForm | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 28 | Ability to have multiple search criteria. Ability to apply up to 5 search criteria with AND operators. | SearchForm | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 29 | After the search is executed, the search results are displayed to the user. The search results view should display a list of the matching drug labels. The search results may be paginated when they exceed a specified number of drug labels. | SearchResults | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 30 | In the SearchResults there is the ability to select two labels. After selecting two labels, the user can then compare the labels. | SearchResults - Compare | 1 | 10 | Must Have | Stack labels vertically instead of side by side, should be able to handle more than two labels |
| 31 | Ability to navigate to the VersionHistoryView from the SearchResults view | SearchResults - VersionHistory | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 32 | A details page for the drug label is shown after the user clicks on an item from the search results. | SingleLabelView | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 33 | The Search query parameters used in the search are highlighted in the SingleLabelView | SingleLabelView | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 34 | Ability to navigate to the VersionHistoryView from the SingleLabelView | SingleLabelView - VersionHistory | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 35 | Drug label search functionality is available to an unregistered / guest / null User | User | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 36 | Basic user authentication including sign-up, sign-in, and password reset using email allows for additional features such as uploading labels, saving queries, etc. | Users | 1 | 1 | Must Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 37 | Section (subsection) Search is implemented across countries, sections are mapped to one another | UX | 2 | 30 | Must Have | Will include indication subsection |
| 38 | The users should be able to easily determine what product they are viewing. The Generic Name, Product Name and Dosage/Type should be clearly indicated. | UX | 2 | 15 | Must Have | In the search or the search results, we should know what name, dosage and strength. |
| 39 | App should be able to update labels seamlessly. Weekly updates. | UX | 2 | 2 | Must Have | Should always show the most recent |
| 40 | Users can see the key words in the context of their keyword search. Truncated section is shown. | UX | 3 | 10 | Must Have | match keywords will shown in color |
| 41 | Add PDF link on single label view | UX | 1 | 2 | Must Have | Estimate assumes link is in database |
| 42 | A Version History View page is displayed showing changes to a drug label over time | VersionHistory | 1 | 1 | Must Have |  |
| 43 | Search results automatically highlighted in the version history page | VersionHistory | 1 | 1 | Must Have |  |
| 44 | Canada- Labels Added to Searchable Database and parse data | Data | 3 | 30-60 | Should Have | Assumes Database is searchable |
| 45 | Ability to export selected columns from multiple labels from the SearchResults in CSV | Export | 1 | 1 | Should Have |  |
| 46 | Ability to export Label Comparison to CSV | Export | 1 | 1 | Should Have |  |
| 47 | Basic Search Achieves less than 10 second response | Non-functional | 3 | 15 | Should Have | <https://www.nngroup.com/articles/response-times-3-important-limits/> |
| 48 | UK - MHRA Labels are Added to Searchable Database and parse data | UX | 3 | 30-60 | Should Have | Assumes Database is searchable |
| 49 | If prior version exist, allow the user to navigate to them, from the search results (or landing page)… | VersionHistory | 1 | 5 | Should Have | Once a drug is selected, previous versions in their entirety can be viewed in the application |
| 50 | A person w/computer science background could put it into production within a few days | Architecture | 3 | 30 | Could Have | Dockerize, Infrastructure as Code, and lowest possible cost |
| 51 | Survey implemented to gather data about users. | Data | 2 | 3 | Could Have | Would need to define what to do with the data afterwards. |
| 52 | System has access to FDA Drug Label data from DailyMed (SPL/XML), to include ALL historical versions for all prescription drug labels for the previous 3, 5, or 7 years (TBD) | Data (History) | 1 | 5 | Could Have | Completed in Iteration 1. Must launch Iteration 1 code base to complete |
| 53 | Ability to export the Version History View to CSV | Export | 1 | 1 | Could Have |  |
| 54 | All export CSV pages include highlighting of the search parameters, when present | Export | 1 | 1 | Could Have |  |
| 55 | Ability to Search by Product Characteristics (color, imprint, shape, size, scoring, etc) for all countries | SearchForm | 1 | 40 | Could Have |  |
| 56 | Ability to Filter Search Results - Pharmacologic class for all countries | SearchForm | 1 | 1 | Could Have |  |
| 57 | Ability to Filter Search Results - marketing categories for all countries | SearchForm | 1 | 1 | Could Have |  |
| 58 | Application is hardened to meet security standards | Security | 3 | 5 | Could Have | Comply with security benchmarks. |
| 59 | Provide users with customized results | UX | 2 | 20 | Could Have | by adding different dropdown menu filter, the result is filtered the way the user wants |
| 60 | Search by MedRA Terms given the tags instances of terms in labels | UX | 3 | 15 | Could Have | If tags are provided, we can add to labels and then provide search. |
| 61 | Ability to Filter Search Results by Manufacturer for all countries | UX | 1 | 4 | Could Have | Left Hand Column Display |
| 62 | Ability to Filter Search Results by Country for all countries | UX | 1 | 4 | Could Have | Left Hand Column Display |
| 63 | Ability to Filter Search Results by Marketing (i.e. Application Type) for all countries | UX | 1 | 4 | Could Have | Left Hand Column Display |
| 64 | Ability to Filter Search Results in an assess Level Of Effort To Change From Horizontal to Vertical | UX | 3 | 4 | Could Have | Left Hand Column Display |
| 65 | Ability to Filter Search Results by Country for all countries | UX | 3 | 4 | Could Have | Left Hand Column Display |

# Appendix G : Important Links

Application URL: <https://searchrx.org>

GitHub URL: <https://github.com/searchrx>

1. (2011). TF–IDF. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi-org.ezp-prod1.hul.harvard.edu/10.1007/978-0-387-30164-8\_832 [↑](#footnote-ref-1)
2. https://en.wikipedia.org/wiki/Okapi\_BM25 [↑](#footnote-ref-2)
3. https://www.elastic.co/ [↑](#footnote-ref-3)
4. https://www.elastic.co/guide/en/elasticsearch/reference/current/release-notes-8.0.0.html [↑](#footnote-ref-4)
5. https://www.elastic.co/guide/en/elasticsearch/reference/current/knn-search.html [↑](#footnote-ref-5)
6. https://huggingface.co/pritamdeka/S-PubMedBert-MS-MARCO [↑](#footnote-ref-6)
7. https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-simple-query-string-query.html [↑](#footnote-ref-7)
8. https://www.searchkit.co/ [↑](#footnote-ref-8)
9. *Aim and Scope | Edelweiss Pharma Analytica Acta*. (n.d.). Retrieved March 1, 2023, from<https://edelweisspublications.com/aim-and-scope.php?journal_id=45> [↑](#footnote-ref-9)
10. *For Authors | Edelweiss Publications*. (n.d.). Retrieved March 1, 2023, from<https://edelweisspublications.com/price.php> [↑](#footnote-ref-10)
11. *Instructions for authors – Translational and regulatory sciences*. (n.d.). Translational and Regulatory Sciences. Retrieved March 1, 2023, from<https://www.cutrs.jp/submission-instructions_t/> [↑](#footnote-ref-11)
12. *Journal of Labelled Compounds and Radiopharmaceuticals Author Guidelines*. (n.d.). Analytical Science Journals.<https://doi.org/10.1002/(ISSN)1099-1344> [↑](#footnote-ref-12)
13. *For Authors | JAMA Health Forum | JAMA Network*. (n.d.). JAMA Network. Retrieved March 1, 2023, from<https://jamanetwork.com/journals/jama-health-forum/pages/for-authors#fa-about> [↑](#footnote-ref-13)
14. “EMEA Becomes Ema.” *PMLive*, PMGroup Worldwide Limited, 14 Dec. 2009, <https://www.pmlive.com/pharma_news/emea_becomes_ema_197492>. [↑](#footnote-ref-14)
15. Van Norman, Gail A. “Drugs and Devices: Comparison of European and U.S. Approval Processes.” JACC. Basic to Translational Science, vol. 1, no. 5, 2016, pp. 399–412, <https://doi.org/10.1016/j.jacbts.2016.06.003> [↑](#footnote-ref-15)
16. Ema. (n.d.). *Product information*. European Medicines Agency. Retrieved February 15, 2023, from <https://www.ema.europa.eu/en/glossary/product-information> [↑](#footnote-ref-16)
17. *European Medicines Agency Action Plan Related to the Commission ...* <https://www.ema.europa.eu/en/documents/other/european-medicines-agency-action-plan-related-european-commissions-recommendations-product_en.pdf>. [↑](#footnote-ref-17)
18. DrugLabelExplorer. “Dle/load\_ema\_data.Py at Main · Druglabelexplorer/DLE.” GitHub. Accessed February 17, 2023. <https://github.com/DrugLabelExplorer/dle/blob/main/dle/data/management/commands/load_ema_data.py>. lines 64-105 [↑](#footnote-ref-18)
19. DrugLabelExplorer, “Dle/load\_fda\_data.Py at Main · Druglabelexplorer/DLE,” GitHub, accessed April 1, 2023, https://github.com/DrugLabelExplorer/dle/blob/main/dle/data/management/commands/load\_fda\_data.py. [↑](#footnote-ref-19)
20. “DailyMed,” U.S. National Library of Medicine (National Institutes of Health), accessed April 1, 2023, https://dailymed.nlm.nih.gov/dailymed/. [↑](#footnote-ref-20)
21. Open.fda.gov, accessed April 1, 2023, https://open.fda.gov/. [↑](#footnote-ref-21)
22. *History | Pharmaceuticals and Medical Devices Agency*. (n.d.). Retrieved February 16, 2023, from<https://www.pmda.go.jp/english/about-pmda/outline/0002.html> [↑](#footnote-ref-22)
23. Japan Health Authority (PMDA) Regulations. (n.d.). *Global Regulatory Partners, Inc.* Retrieved February 16, 2023, from<https://globalregulatorypartners.com/resource-center/regulatory-intelligence-platform/japan-regulatory-intelligence/japan-health-authority-pmda-regulations/> [↑](#footnote-ref-23)
24. *Japan—Pharmaceuticals*. (2022, November 9). International Trade Administration.<https://www.trade.gov/country-commercial-guides/japan-pharmaceuticals> [↑](#footnote-ref-24)
25. <https://en.wikipedia.org/wiki/National_Medical_Products_Administration> [↑](#footnote-ref-25)
26. <http://english.nmpa.gov.cn/2019-07/25/c_390614.htm> [↑](#footnote-ref-26)
27. <https://www.raps.org/news-and-articles/news-articles/2021/5/the-uk-regulatory-landscape-post-brexit> [↑](#footnote-ref-27)
28. <https://www.gov.uk/guidance/apply-for-a-licence-to-market-a-medicine-in-the-uk> [↑](#footnote-ref-28)
29. <https://www.rpharms.com/development/trainee-pharmacists/product-characteristics-summary> [↑](#footnote-ref-29)
30. Smith, Jogalekar, S., & Gibson, A. (2014). Regulation of natural health products in Canada. Journal of Ethnopharmacology, 158, 507–510. https://doi.org/10.1016/j.jep.2014.06.037 [↑](#footnote-ref-30)
31. Health Canada, 2023. <https://www.canada.ca/en/health-canada.html> [↑](#footnote-ref-31)