Repurposing of Drug Label Information to Create Actionable Intelligence with Metawise™

Paul Bradley and Jane Z. Reed
BioWisdom, Harston Mill, Cambridge, CB22 7GG, UK

Introduction

Improved toxicological prediction demands the best view of current and historic data. There are many barriers along the path to better data repurposing, pre-competitive sharing, and harmonisation – all key goals in the current climate of translational healthcare, including:

- **Document Structure** – legacy data is often in the form of unstructured text, with no tagging or mark up of important concepts
- **Linguistic Diversity** – domain experts use a varied vocabulary, which can confound conventional search approaches

About MetaWise

Metawise is a flexible and controllable system for search and identification of life science concepts from unstructured and other text sources. It uses a unique approach, based on term structure and semantics, to recognise the varied language used by domain specialists to reference important concepts, and can harmonise this language to any preferred standard. It features two major algorithmic components that deal with the following tasks:

- **Term Discovery** – rather than static dictionaries, Metawise utilises term *templates*, creating a more flexible system and providing a method for classifying novel concepts.
- **Term Translation** – Metawise allows terminology from a target lexicon (e.g. SNOMED CT) to be translated and mapped to descriptors in free text.

We present the application of Metawise to unstructured drug labels, creating actionable intelligence from this valuable data source for use in decision support by a US government agency.

Introducing Structure to Drug Package Inserts

As of May 2011, DailyMed featured labels for around 24,000 drug products submitted to the FDA. It constitutes a rich source of intelligence, with tagged sections including drug indications and usage, adverse reactions, contraindications, laboratory testing and boxed warnings. However, the text within each section is unstructured, with no mark-up of important concepts, and widespread use of synonyms such as "haemorrhage" and "bleeding". These features impede searching and navigation and limit the use of DailyMed in other applications.

Metawise allowed the rapid identification and mark-up of key medical language from the unstructured labels, and translated clinical terms to their SNOMED CT equivalents (Figure 1).

Figure 1: Mark-up of clinical language in a drug package insert from DailyMed (main image), and translation of "increase in average blood pressure" to the SNOMED CT concept "hypertension" (yellow box, inset).

Dealing with Legacy Text

Metawise uses an exhaustive matching method that is able to tolerate the natural variation that exists in free text, including stemming (plurals, US/UK variants and other inflected forms), synonyms (for whole phrases as well as individual words within phrases), and word insertions/deletions. Using this approach, terms from a target lexicon can be matched to document text with greater sensitivity than conventional dictionary look-ups (Table 1).

Table 1: SNOMED CT terms and matches from DailyMed labels. Matches include a combination of replaceable substrings, stemming and wording changes.

SNOMED CT Concept	DailyMed Matches
Anaphylactic-type reaction	Anaphylactoid reaction, anaphylactic reactions, anaphylactic shock, anaphylactoid and allergic reactions, anaphylactic reaction
Hypertension	elevation of blood pressure, increase in average blood pressure, high blood pressure, increased Blood Pressure

Delivering Actionable Healthcare Intelligence

The mark-up and harmonisation of medical language creates a powerful metadata index for document search and triage. However, this index is limited in terms of context, e.g. some drugs are ***indicated*** for the treatment of myalgia whereas others induce this type of muscular pain as an ***adverse effect***. This additional information was provided by the generation of ***assertional metadata***, which summarised the key facts in the form of subject-relationship-object triplets:

Diclofenac – HAS BOXED WARNING – GI Ulcer
Amiodarone – HAS ADVERSE EFFECT – AV Block

Assertional metadata creates a semantically normalised, actionable information layer over documents. Relationships create links between concepts (Figure 2), connecting drugs with adverse events, black box warnings and laboratory tests. This allows scientific hypotheses to be explored, and more direct questions to be posed than would be possible with concept mark-up alone, such as:

- *What adverse effects are observed in drugs that treat breast cancer?*
- *Which haematology tests should be considered when prescribing ACE inhibitors?*
- *Which black box warnings occur in medicines that list haemolytic anaemia as an adverse event?*

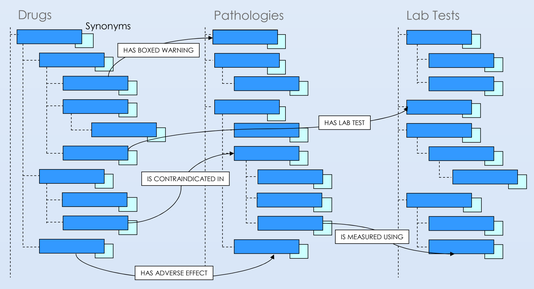

Figure 2: Assertional metadata (black arrows with relationships in text boxes), connecting taxonomies of drugs, pathologies and laboratory tests.

The Benefits of Organised Data

The highly organised assertional metadata layer created by Metawise is a powerful substrate for analysis, and allows different views to be taken of the data.

Diagnostic Drug Signatures

With label information converted to a highly structured form, it is possible to align equivalent features (adverse effects, indications, and so on) and note their frequency, or absence, across all labels. This generates a characteristic profile, or signature, for each drug that may share features with other drugs, e.g. those belonging to the same therapeutic class (Figure 3). Once expressed in this form the profiles are tractable to alignment algorithms, making it possible to ask questions such as:

- *Which drugs most closely resemble my drug of interest in terms of adverse event profile?*
- *My drug does not currently feature a boxed warning; which drugs with boxed warnings is it similar to and might this be a concern?*

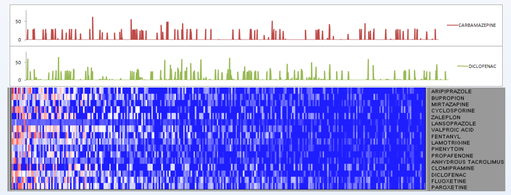

Figure 3: Creation of drug "signatures" from highly ordered assertional metadata, showing adverse event frequency plotted against pathology type (upper image) and clustering of profiles in OmniViz (TreeScape, lower image).

Interoperability with Other Data Sources

With terminology expressed in a consistent format, the content of DailyMed is made interoperable with other resources. For example, it is possible to seek corroborative evidence for DailyMed black box warnings in the FDA Adverse Event Reporting System (AERS, Table 2).

DailyMed Warning	In AERS?
Confusion	✓
Fever	✓
Hypotension	✓
Hypothermia	✓
Pneumonitis	✓
Pulmonary fibrosis	✓
Pulmonary toxicity	✓
Wheezing	✗

DailyMed Warning	In AERS?
Agranulocytosis	✓
Aplastic anemia	✓
Bone marrow suppression	✓
Fever	✓
Leucopenia	✓
Low platelet count	✗
Stevens-johnson syndrome	✓
Toxic epidermal necrolysis	✓

Table 2: Cross-checking of black box warnings for bleomycin (left) and carbamazepine (right) against FDA AERS, made easy with harmonised language.

Conclusions

Drug package inserts are a rich, but largely unstructured, source of pharmaceutical intelligence. The objective of this exercise was to repurpose this unstructured information for use in decision support. BioWisdom's Metawise allowed the rapid identification of clinical terms from DailyMed drug labels, and their harmonisation to SNOMED clinical terminology. The metadata index that this created provides a powerful substrate for analysis, revealing new trends, offering fresh insight and rendering DailyMed interoperable with other applications. Furthermore, the creation of assertional metadata forms an actionable information layer that can be used to answer complex scientific questions.