A Reference Terminology for Drugs

W.D. Solomon[1,2], C.J. Wroe[1,2], A.L. Rector[1,2],
J.E. Rogers[1,2], J.L. Fistein[1,2], P. Johnson[3]

[1]Medical Informatics Group, Department of Computer Science,
University of Manchester, UK

[2]The GALEN Organisation Limited, UK

[3]Sowerby Centre for Health Informatics at Newcastle, University of Newcastle, UK

Abstract

GALEN technology for re-usable terminologies using formal classification is being applied to the creation and maintenance of a reference terminology for drugs. GALEN's techniques are being used to address specific deficiencies of existing drug classifications that make it difficult to create and maintain guidelines to support prescribing in the care of patients with chronic diseases.

The reference terminology is in two parts; firstly, a re-usable and automatically-classified 'ontology' is built with GALEN technology; this describes generic drugs, their composition in terms of chemicals and chemical classes, their actions, indications and interactions. Secondly, a 'dictionary' of prescribable proprietary products is integrated with this ontology. The result is a drug resource designed to support both the traditional uses of a drug knowledge base (e.g. prescribing and messaging), and the specialised demands of guideline authoring and execution.

INTRODUCTION

"Despite the fact that the medications available are finite, countable and identifiable, no universally accepted standard exists for naming them"[1]

The Prodigy Project[2], funded by the UK's Department of Health, is currently extending an existing prescribing-support system for Primary Care, from a guideline (or protocol) system operational over a single consultation, to guideline and prescribing support for the long-term management of chronic diseases. As well as requiring a resource that contains details of prescribable products to a level suitable for creating and communicating prescriptions, Prodigy needs the ability to be able to include abstractions about drugs and drug classes. For example, a guideline may include the statement 'if the patient is on an inhaled steroid, then …'. This requires that the guideline author be able to create, use and re-use the abstraction 'inhaled steroid', and that a run-time execution engine be able to test current medications from an electronic patient record for inclusion in this abstraction.

Furthermore, there are a number of different abstractions – or axes of classification – that need to be supported, and these all need to be maintained with respect to each other, and with respect to the descriptions of individual prescribable products.

Existing Drug Knowledge Bases

Several proprietary drug knowledge bases (DKBs) exist, providing up-to-date and comprehensive cataloguing of products available for prescribing. However, each also provides different levels of added-value functionality and their information models are correspondingly different. All of these DKBs, together with other UK resources such as the NHS Clinical Terms and the British National Formulary (BNF), group drugs into classes in similar ways, but none has proved suitable for Prodigy.

Two specific problems with existing DKBs were encountered: firstly in identifying a specific subgroup or abstraction of drugs for a given guideline accurately and unambiguously; secondly in maintaining such groupings of drugs.

The classification of drugs presented, for example, to support navigation during the prescribing task did not lend itself well to the requirements of guideline authoring and execution. In addition, the implementation structures in which those abstractions were made could not be easily extended to support and maintain the number of different classifications that were required.

A prime source of these difficulties is the fact that existing drug classifications mix different axes without being explicit what those axes are: chemical structure, use, actions, or indications, or special properties such as whether a particular substance is water or lipid soluble.

Our fundamental hypothesis is that by using GALEN's formal classification technology, we will be able to provide a basis for a maintainable, multi-axial classification of drugs, suitable for guideline support and execution, as well as for supporting prescribing and messaging.

ARCHITECTURE

There are a number of different kinds of information about drugs that need to be represented:

- Detailed, rapidly changing information about products, proprietary names, pack sizes, and exact preparations: the 'drug dictionary'. This is the kind of information which changes rapidly, but it is well-understood, and there is much experience in delivering this kind of information in existing DKBs;
- A re-usable, less rapidly changing, set of information about generic drugs: the 'drug ontology'. This describes the chemical composition of drugs, their structure, physiological actions, therapeutic uses, and, where appropriate, information on the general form and route of administration;
- Application-specific information, for example describing specific prescribing regimens for particular clinical conditions, which may form part of a prescribing guideline.

Within Prodigy, these different kinds of information are maintained and implemented separately; the overall architecture is shown in Figure 1.
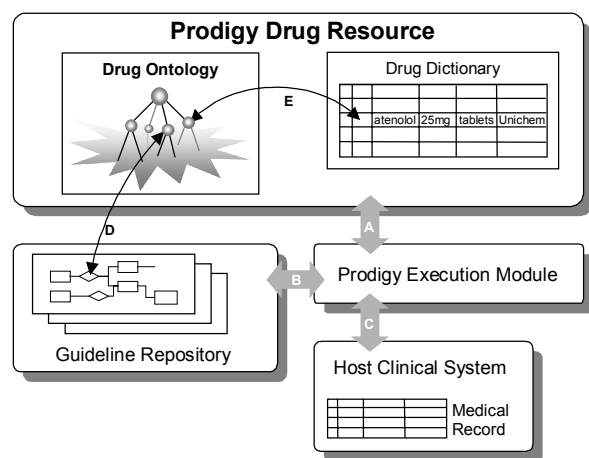

Figure 1: The Prodigy Architecture

The re-usable Prodigy Drug Resource is a combination of the 'drug ontology' and 'drug dictionary', whilst the application-specific information is maintained separately as part of a structured guideline in the Guideline Repository.

The Prodigy Execution Module (PEM) co-ordinates communication and data-flow between the Drug Resource, the Guideline, and the Host Clinical System (links shown as **A**, **B** & **C**).

Within this overall architecture, GALEN is taking primary responsibility for the Drug Ontology, which is the subject of the remainder of this paper.

Prodigy Requirements for the Ontology

The Drug Ontology is referenced by guidelines (link **D**), and by the Drug Dictionary (link **E**), and forms the focus for all generic information about drugs. It supports queries such as:

- "is this prescribable drug product a member of this drug class" – used for guideline support;
- "given a drug class, what are its prescribable products" – used for prescribing support;
- "is this drug contra-indicated given the presence of these conditions or other current medication" – used for prescribing support.

GALEN

The GALEN Programme has been applying description logic methods for representing a general clinical terminology[3,4]. Earlier work in the PEN&PAD project successfully demonstrated the value of GALEN prototypes in supporting the prescribing process[5]. Within Prodigy, we are using matured GALEN techniques for concept representation to build the Prodigy Drug Ontology, as an extension and development of the GALEN Common Reference Model of medicine.

The result is a hierarchy of drugs, automatically and simultaneously classified along many different axes. This is initially to support the specific requirements of Prodigy, but the flexible nature of the GALEN technology means that it is easy to add additional axes of interest at a later date, whilst still maintaining the functionality originally intended.

There are different steps required to use GALEN technology to build the drug ontology:

- designing and implementing a high-level ontological structure in GRAIL (the GALEN Representation and Integration Language)[6];
- designing and implementing an intermediate representation[7] to make it easy to populate the high-level ontological structure;
- populating and maintaining the structure subsequently using specialised methodologies, for which tools have been developed.

Subsequently, the Drug Ontology is integrated with the Drug Dictionary to produce the Prodigy Drug Resource, itself a part of the overall Prodigy Architecture.

IMPLEMENTATION

Principles of Ontology Development in GALEN

The key principle in developing an ontology using GALEN is to make all information explicit, and to record information on each axis independently, even at the cost of redundancy. For example, to record both that a drug has an action "bronchodilitation" and that it is used to treat "bronchoconstriction". The

goal is to avoid conflating axes even if they are, in general, parallel. This methodology is sometimes referred to as creating 'orthogonal taxonomies'[8,9].

The first task is therefore to establish the key axes for the high level structure. Each of these axes will become a semantic link in the resulting ontology, allowing drugs to be automatically classified along any, or along any combination, of these axes.

High level structure of the Prodigy Drug Ontology

Ingredients. Each drug preparation has at least one ingredient (and there may also be any number of other active ingredients or excipients).

It is not sufficient to group drugs by the ingredients they contain: the ingredients themselves must have an inherent classification based on structure. This allows a 'drug preparation containing prednisolone' to be automatically classified under a 'drug preparation containing steroid'.

Form & Route. These are intimately related and so are described together. Form can include 'tablet', 'liquid', 'powder' etc. Route can include 'inhaled', 'topical', 'intravenous', 'intramuscular' etc.

Some preparations are intimately linked to their delivery device and so this too has to be included. For example: both nebulised and metered dose inhaled salbutamol posses the same form, route combination. The delivery device must be described to disambiguate the two preparations.

Pharmacological actions. These usually take the form of an action at the level of a receptor (agonism/ antagonism) or at the level of enzymes (inhibition).

Physiological actions. These usually take the form of increasing or decreasing features of a body system. For example, decreasing systemic blood pressure.

Indications. A list of reasonable indications for the drug preparation. Further features are added such as whether a given indication is specifically licensed.

It is not sufficient only to link the drug preparation to a condition. Some notion of a treatment goal must also be included. For example is the drug used to *treat* or to *prevent* a particular condition.

Side effects. These represent the less desirable effects of a drug. Additional features such as probability of occurrence can be added.

The line between wanted and unwanted drug effects may be blurred and may change for different contexts of use. However, for simplicity, the categorisation of the source text in the BNF, our experimental corpus, has been followed.

Interactions. Interactions have to describe both the drug that the interaction occurs with, together with the result of the interaction. Additional information can also be added on the relevant importance of the interaction information (for example whether it is hazardous or not).

Contraindications and cautions. These are used to provide links to conditions, the presence of which either totally preclude the use of a drug, or for which care should be taken. The degree and specific instructions of additional care can be added and the source text terminology includes examples such as 'avoid', 'care', 'reduce dose', 'monitor U&E'.

Pharmacodynamics. Information is included concerning absorption, half-life and excretion in a form-independent manner.

Figure 2 provides a representation of the structure that we have developed to satisfy the requirements of being able to classify along these axes.

Populating the Drug Ontology using an Intermediate Representation for drugs

The GALEN-IN-USE project[10] has demonstrated that given the need for a complex internal structure, an easy-to-use external format, or 'intermediate representation' is required for bulk authoring along with automated tools to translate from the external to the internal format[11]. These techniques were developed and applied in the area of surgical procedure classification, but the Intermediate Representation and tools developed for surgery were easily adapted and extended to make it possible to represent the structures required for drugs.

Integrating the Drug Ontology

Integration of the Drug Ontology with the Drug Dictionary. An existing DKB used for prescribing has been extended by linking individual DKB entries with concepts in the Drug Ontology that describe the structure, and hence the properties, of each such entry. This integration, shown as the link labelled **E** in Figure 1, results in the Prodigy Drug Resource.

Integration of the Prodigy Drug Resource with the Prodigy Architecture. Clinical system suppliers implement the Prodigy system as an add-on module. The GALEN Drug Ontology is an integral part of that module, and is delivered as part of it. In general, there are two ways in which GALEN technology can be delivered. The first is as a dynamic resource managed by a GALEN Terminology Server[3]. This is the most general form, and allows dynamic and arbitrary re-classification, and concept specialisation, to take place at run-time. This delivery mechanism is used, for example, by the Clinergy[TM][12] data entry system and the Classification Workbench[13].

The alternative, used in Prodigy, is applicable when there is no run-time requirement for such dynamic functionalities: the ontology is being used as a fixed semantic index. In this case, integration with Prodigy is via a static snapshot of the ontology, exported in a relational database format, with the separate

hierarchies for each axis made explicit. If the specific requirements of the application change, additional hierarchies can be generated automatically from the ontology as a batch process.

Additional mapping of concepts in the ontology to codes (either from standard schemes like SNOMED or NHS CTV3 (Read) Codes, or from existing DKBs) are included in the ontology export format.

RESULTS

To act as a proof of concept, we have built and populated an ontology of drugs covering the first three chronic disease guidelines being developed in Prodigy: angina, hypertension and asthma.
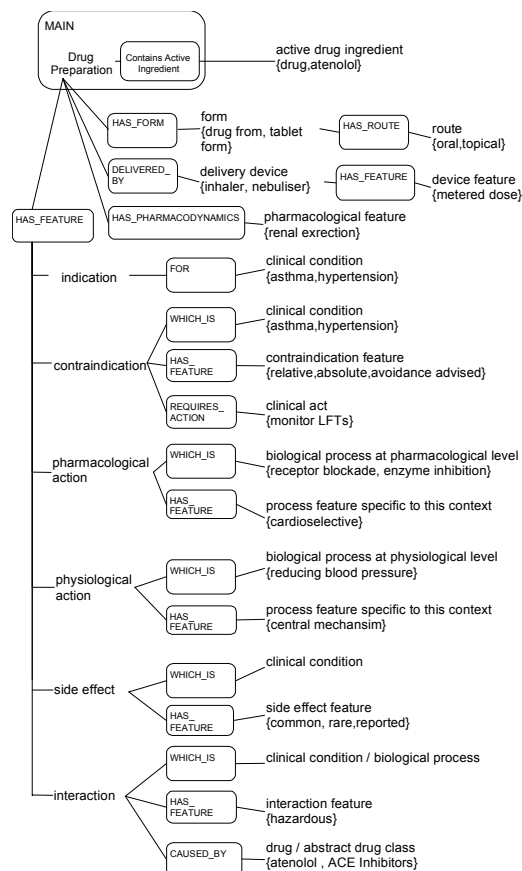

Figure 2: The High-level structure of the Drug Ontology

Initial scoping and experimentation took the form of examination and transcription of relevant sections of the cardiovascular and respiratory chapters of the BNF. This helped scope the task, and provided a corpus from which we developed the required high-level ontological structure (illustrated in Figure 2), and a suitable Intermediate Representation format for populating that structure.

One hundred drug preparation descriptions were created as a population of this structure. A total of 650 abstract drug class descriptions were formed,

reflecting the total number of distinct features applicable to the 100 drug preparations.

Seven hundred 'descriptors' (words or phrases used in the Intermediate Representation) were used together with 30 links. Of those 700 descriptors, 200 were disease terms, 50 signs or symptoms and 100 drug ingredients.

A number of abstract drug classes are enumerated automatically, producing a hierarchy of drugs classified along all of the required axes. These hierarchies are accessible individually, so as to support being able to ask questions along individual axes. Figures 3 & 4 show a sample of the single axis hierarchies produced showing atenolol classified in different ways (by indication, and by side effect).
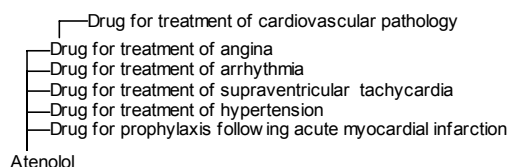

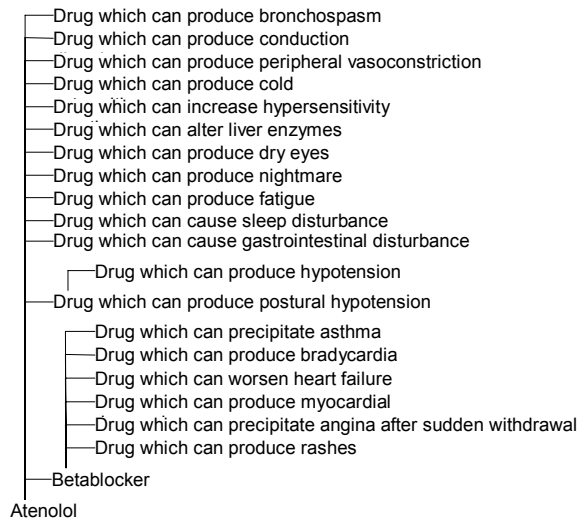**Figure 3: Atenolol classified by indication**

Figure 4: Atenolol classified by side-effect

DISCUSSION

The provision of high quality, timely information about drugs is a mission-critical part of health-care. The task is made more difficult by the complexity of the domain and the varying requirements – from messaging to prescribing to various forms of decision support – that exist. Within the Prodigy Project, we are applying GALEN technology with the aim of producing a generic, reusable resource, and one which is inherently more maintainable, as the classifier provides automatic classification.

The techniques and tools we have developed so far have proved successful in demonstrating the utility of

the technology. Specifically, using an existing corpus of generic information (the BNF), we have developed and populated a drug ontology to support the initial Prodigy guidelines. Work is now continuing to validate this ontology in the runtime execution module. Initial results are encouraging in that GALEN techniques have allowed us to produce, maintain and deliver a multi-axial drug classification.

Future work will need to focus on a number of outstanding issues.

Validation and Quality Assurance

Drug information must be authored with care. Extensive internal and external validation and QA are required before general release. At present, visual inspection of the hierarchies we have produced confirms that valid and meaningful classification occurs, and we anticipate further feedback once the Prodigy system is further developed. Tools and methods are currently being developed for on-going maintenance and systematic checking.

Scaling up

Our experience with similar processes in surgical procedures within the GALEN-IN-USE Project[10] provide a basis of techniques and methodologies for developing and maintaining the Drug Ontology as it scales. The task of completing the Drug Ontology from existing resources, and the level to which this can be automated, needs to be formalised and supported with software tools. The proof-of-concept that we have now completed has exposed the detailed requirements for those tools, and development is continuing.

Standards activity

We hope that the Drug Ontology may form a generic and reusable part of various standardisation exercises currently underway. In the US, the HL7 Vocabulary Special Interest Group is working in this area, and in Europe, CEN TC251 PT31 is developing a prescription messaging standard[14], and work is underway to ensure harmonisation between that work and the Drug Ontology work described here.

References

1. Cimino JJ, Huff S, Broverman C, McNamara T, Nelson SJ. Development of a Standard Terminology to Support Medication Messages. Journal of the American Medical Informatics Association 1998;Fall Symposium Special Issue.

2. Purves I. Prodigy interim report. Journal of Informatics in Primary Care 1996:2-8.

3. Rector A, Solomon WD, Nowlan W, Rush T. A Terminology Server for Medical Language and Medical Information Systems. Methods of Information in Medicine 1995;34:147-157.

4. Rogers J, Rector A. The GALEN ontology. In: Brender J, Christensen J, Scherrer J-R, McNair P, eds. Medical Informatics Europe (MIE 96). Copenhagen: IOS Press, 1996:174-178.

5. Solomon WD, Heathfield H. Conceptual modelling used to represent drug interactions. In: Barahona P, Veloso M, Bryant J, eds. Twelfth International Congress of the European Federation for Medical Informatics, MIE-94. Lisbon, Portugal, 1994:186-190.

6. Rector A, Bechhofer S, Goble C, Horrocks I, Nowlan W, Solomon WD. The GRAIL concept modelling language for medical terminology. Artificial Intelligence in Medicine 1997;9:139-171.

7. Rogers J, Solomon WD, Rector A, Zanstra P. From rubrics to dissections to GRAIL to classifications. Medical Informatics Europe (MIE-97). Thesalonika, Greece: IOS Press, 1997:241-245.

8. Rector A. Coordinating taxonomies: Key to re-usable concept representations. In: Barahona P, Stefanelli M, Wyatt J, eds. Fifth conference on Artificial Intelligence in Medicine Europe (AIME '95). Pavia, Italy: Springer, 1995:17-28.

9. Rector AL, Zanstra PE, Solomon WD, et al. Reconciling Users' Needs and Formal Requirements: Issues in developing a Re-Usable Ontology for Medicine. IEEE Transactions on Information Technology in BioMedicine 1999;2(4):229-242.

10. Rector A, Rossi Mori A, Consorti F, Zanstra P. Practical development of re-usable terminologies: GALEN-IN-USE and the GALEN Organisation. International Journal of Medical Informatics 1998;48(1-3):71-84.

11. Rector AL, Baud R, Ceusters W, et al. A comprehensive approach to developing and integrating multilingual classifications: GALEN's Classification Workbench. Journal of the American Medical Informatics Association 1998;Fall Symposium Special Issue:1115.

12. Kirby J, Cope N, Souza Ad, Fowler H, Gain R. The PEN&PAD Data Entry System. In: Brender J, Christensen J, Scherrer J-R, McNair P, eds. Medical Informatics Europe (MIE-96). Copenhagen: IOS Press, 1996:430-434.

13. Zanstra PE, Haring EJvd, Flier F, Rogers JE, Solomon WD. Using the GRAIL language for Classification Management. Medical Informatics Europe. Thessaloniki, Greece, 1997.

14. CENTC251. PT31 Prescription Messaging. 1998.