

WestminsterResearch

http://www.wmin.ac.uk/westminsterresearch

Developing a database for automating regulatory affairs in the pharmaceutical industry.

Lindi Slevin Reza Shojanoori Radmila Juric

School of Informatics

Transactions of the SDPS / Journal of Integrated Design and Process Science, 9 (4). Slevin, Lindi, Shojanoori, Reza and Juric, Radmila, Developing a database for automating regulatory affairs in the pharmaceutical industry, 1-11., Copyright (2005), with permission from IOS Press.

The definitive version in the Journal of Integrated Design and Process Science is available online at:

http://iospress.metapress.com/openurl.asp?genre=article&issn=1092-0617&volume=9&issue=4&spage=1

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners. Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch. (<u>http://www.wmin.ac.uk/westminsterresearch</u>).

In case of abuse or copyright appearing without permission e-mail wattsn@wmin.ac.uk.

DEVELOPING A DATABASE FOR AUTOMATING REGULATORY AFFAIRS IN THE PHARMACEUTICAL INDUSTRY

Lindi Slevin Reza Shojanoori Radmila Juric Department of Information Systems, Cavendish School of Computer Science, University of Westminster, London, UK

The complex procedure of processing the marketing authorization of medicinal product licenses is a challenging task for government health authorities and the pharmaceutical industry across the world. Regulatory requirements have diverged significantly, and each country has its own regulations and procedures for marketing authorizations. The automation of and adequate software support for such procedures are critical factors that can improve the efficiency of regulatory authorities. In this paper we report on the design and implementation of a database whose role is to (a) support the automation of marketing authorization procedures, (b) address the interoperability of such procedures across the world, and (c) be reusable across a family of related applications. Our database is implemented in Oracle8i, and a distributed and component-based application has been built upon it using the J2EE technology.

Keywords: Marketing authorization, licensing application, submission and evaluation, software interoperability, EJB.

1. Introduction

A software solution, which automates marketing authorization of medicinal product licenses across the world, is a large-scale distributed data intensive application. (Please note that the 'marketing authorization of medicinal product licenses' is the procedure out of which a license may be granted. In our work we refer to this as 'marketing authorization'.) It requires sharing of data stored in databases and/or repositories and sharing of processes associated with various marketing authorization procedures and their prescribed regulations. To ensure

- (i) transparency of the results of marketing authorization applications,
- (ii) sharing of marketing authorization procedures across regulatory authorities, and
- (iii) interoperation of such procedures with existing healthcare systems,

we have implemented component-based software architecture, which automates marketing authorization procedures, as an EJB application (Juric et al., 2005). Its functionality is supported with various databases. In this paper we report on design and implementation of a database, whose role is to

- (a) support the automation of marketing authorization procedures,
- (b) address the interoperability of such procedures across the world, and
- (c) be reusable across a family of related applications.

Transactions of the SDPS

In section 2 we give a related background that highlights problems and procedures for marketing authorizations, which differ across various countries and regulatory authorities across the world. We also summarize our previous work in which a generic software architectural model for interoperable marketing authorization procedures, was implemented as an EJB application. In section 3 we formulate the aims of the paper and discuss related works that use databases in the same problem domain. In section 4 we describe our database design procedure, which is underpinned by discussions on our choice of entities, their attributes and their relationships. We also address the impact of our generic software architecture to the design of our database schemas. Section 5 discusses implementation issues and clarifies the role of submission/evaluation criteria in our database. We conclude and summarize our future work in section 6.

2. Related Background

Marketing authorization is one of the most important tasks undertaken by government health departments and their regulatory authorities, in every country in the world. The independent marketing authorizations are centered on regulations and guidelines for reporting and evaluating data on medicinal products' safety, quality and efficacy. These procedures are strictly defined to ensure that all standards on testing, manufacturing and controlling medicinal products are achieved. However, each country has its own system and procedures for marketing authorizations. These procedures differ, not only in vocabulary and definitions of medicinal products, but also in different organizational structures and practices of individual regulatory authorities. This represents a serious drawback for efficient local and worldwide licensing of medicinal products. The automation of such marketing authorization procedures and their adequate software support is a critical factor that can dramatically improve the efficiency of regulatory authorities and interoperation of regulatory systems across the world.

In our previous works (Juric and Juric, 1999), (Juric and Juric, 2000) we have analyzed the local needs of various regulatory authorities and have extracted the common practices that exist across the world, which is essential if any interoperation between regulatory systems were to take place. In Fig. 1, we show the generic architectural model that allows automation of marketing authorizations across the world. The model is layered and component-based. Each regulatory authority may apply their own submission/evaluation procedures or any other that is available internationally (Juric and Juric, 2002). To illustrate the architecture we define the generic procedure for marketing authorizations. Its functionality is divided into two workflows:

- (i) submission of a licensing application for marketing authorization under local regulatory authority rules (\mathbf{R}_i) , and
- (ii) *evaluation* of a successfully submitted licensing application, under an evaluation procedure and its rules $D(E_i)$ available locally/internationally.

Reports resulting from such workflows are stored within a shared data repository, as in Fig. 1. The application layer provides a basic GUI functionality and controls interaction between users and any other layers within the system. This includes the appropriate choice of the R_i and $D(E_i)$ components involved in a particular licensing application submission for evaluation of a medicinal product. The domain layer consists of two families of components:

- 1. The R_i family of components contains a set of rules that should be followed by an applicant in order to have an automated licensing application submission, as in (i) above, within a particular regulatory authority. The R_i family may also include any future set of rules that originate within the International Conference on Harmonisation (ICH) (available at http://www.ich.org).
- 2. The $D(E_i)$ family of components (D denotes the domain of a specific regulatory authority) contains all available evaluation procedures and their rules, as in (ii) above, that originate from either different regulatory authorities or can be found within future harmonized activities from the ICH. Components from the domain layer use various data repositories and databases stored

within components of the persistence layer. The persistence layer contains data on licensing applications submitted for evaluation and the reports resulting from their evaluations. Our persistence and domain layers can be seen as a common repository of data and processes, where various applicants (such as pharmaceutical companies, regulatory authorities and hospitals) can share the data and services defined in our component-based architecture.



Fig. 1 The Generic Software Architecture for Interoperable Marketing Authorization Procedures

One example of a licensing application submission and its evaluation, placed within the software architecture from Fig. 1, is modeled as an EJB application and implemented within the J2EE (Juric et al., 2005). We have used Sun Studio Enterprise 7 and Oracle8i. Our application design has also generated COTS components (Juric and Williams, 2005), with a set of design patterns used throughout example components' modeling and deployment (Williams and Juric, 2005).

3. Aims of the Paper and Related Work

The aims of this paper are to:

- (1) develop a database to support the automation of marketing authorizations for medicinal product licenses,
- (2) develop a database schema, which serves a family of related applications that may be outside of this problem domain, and
- (3) address the interoperability of procedures for marketing authorizations across the world.

Currently, when a pharmaceutical company wants to apply for marketing authorization for a medicine that they have developed, they submit to a regulatory authority extremely large volumes of text, which are likely to be either structured or semi-structured. A regulatory authority and their evaluation agency must check if all the documentation is complete and in the correct format. It is only after this stage has been satisfied that the submitted documentation can be evaluated. Our aim (1) is to automate this procedure by allowing applicants to submit all their documentation online. We will also check if the submission rules, R_i , for such documentation have been satisfied. Once they have been satisfied, our software enables the submitted documentation to be evaluated through a specific evaluation procedure and its rules, $D(E_i)$. Only after the evaluation rules have been satisfied will an evaluator (employed by a regulatory authority) complete the evaluation of the documentation and issue

a license. Thus, the evaluation of licensing applications is a semi-automatic procedure. The database that supports submissions and evaluations, as in (i) and (ii) from section 2, should save a substantial number of man-hours per submission by storing submitted documentation in a structured or semi-structured format allowing semi-automatic evaluations, and storage of the results of evaluation in shared repositories.

The issue of automation of marketing authorization applications has been around for 20 years in America and 15 years in Europe. Quite a number of projects worldwide aim at reaching a partly or completely electronic data exchange between the pharmaceutical industry and the authorities involved (Franken, 2003). However, our work is the only one that aims to support the automation of procedures for marketing authorization of medicinal products and make them interoperable.

IDIOM Software have developed an XML-based publishing solution (available at http://www.idiominc.com/news/press-releases.asp?display= detail&id=80) which automates a different area of the medicinal industry: the submission of labeling and product information (required as a part of the marketing authorization) in the formats required by different regulatory bodies, and using different languages. Their software translates the text in product labels and product descriptions into XML format, which is then translated into a foreign language, thus simplifying the process of publishing multilingual, multicultural labeling and packaging content. This is an example of the usage of XML for data interchange, which does not involve any concept of submission or evaluation.

The EC allows the electronic submission of licensing applications for evaluation as advised on the EMEA's website (available at http://www.emea.eu.int/htms/human/presub/q24.htm). Applicants who submit electronically must use a "PCcompatible medium, e.g. CD-ROM or DVD together with 2 additional paper copies and must sign a letter in which they commit themselves to supplying a full paper copy within 48 hours upon request and confirm that the data on the CD-ROM/DVD supplied is identical to that in any written submission". However, there is no electronic repository that the EMEA has that might be included into our architectural solution and this is emphasized by the EMEA's statement that "the paper [copy] remains the formal submission".

The US Food and Drug Administration's (FDA) Centre for Drug Evaluation and Research (CDER) website (available at http://www.fda.gov/cder/regulatory/ersr/) provides guidance for submitting licensing applications in electronic format (in this case using a pdf file stored on a floppy disk, CD or digital tape).

The Computer Assisted New Drug Application (CANDA) was developed in 1986. It was viewed as a tool that would simplify the entire clinical information management process (see CANDA guidelines available at http://www.evolvingtech.com/etc/industry/submissions.html). Their product, Muse CANDA, provides a spreadsheet-like environment and is "one of the accepted commercial off the shelf software (COTS) products for CANDA Submissions". It is essentially a computerized extension of a paper-based data management process. Between 1991 and 1994, CANDA applications were about six months faster than traditional paper-based licensing applications. However, each CANDA is a one-of-a-kind production.

In 1995 the FDA launched its drug review and approval system through Submission Management and Review Tracking (SMART) (available at http://www.fda.gov/fdac/features/895_smart.html). They reengineered the drug approval process, and they also used CANDAs themselves. SMART was expected to transform their activities and let regulatory authorities retrieve their repositories, which would have allowed rapid comparison of new medicinal products with others of either the same type or for the same health problem.

Unfortunately, no more information has been published on SMART since 1995 and therefore it has been assumed that no other research which aims to provide the automation of the marketing authorization procedures is currently being carried out.

4. Designing the Database

In this section we describe the process of designing the database that will support the automation of marketing authorization. We identify the necessary data to be stored; we show a conceptual data model that represents the structure of the data; we discuss certain database design issues, which are specific to this problem domain. A complete Licensing Application data model, which supports functionality of a full-scale application, is available in Fig. 2.



Fig. 2 Complete Licensing Application Data Model

To ensure high cohesion and low coupling between our database schema and the application built upon it, we separate data models in (b) and (c) below according to the workflows described in (i) and (ii) in section 2. Thus, there are three data models:

- (a) the complete Licensing Application data model that represents a full scale application-specific data structure (available in Fig. 2),
- (b) the Submission data model (available in Fig. 3) that supports the submission workflow described in (i) of section 2, and
- (c) the Evaluation data model (available in Fig. 4) that covers the evaluation workflow described in (ii) of section 2.

All three data models from (a)-(c) above are based on the Marketing Authorisation Application (MAA) document (version 5.0) provided by the Medicines Control Agency (available at http://www.mca.gov.uk), which is a UK regulatory agency. However, our intension has been to make these models as generic as possible to fit the regulatory requirements of any other country in the world.



Fig. 3 Submission Data Model



Fig. 4 Evaluation Data Model

4.1. General Design Issues

There are many known approaches for designing database schemas, including top-down and bottom-up (DeMarco, 1978; Date, 2003). We have adopted the following approaches:

- (1) We used a combination of top-down and bottom-up approaches and exercised iterative development throughout the database design activities.
- (2) We primarily worked through the MAA document and identified potential attributes and entities.
- (3) We used our intuition as the main factor that influenced our first selection of attributes and entities.

In all subsequent iterations a more logical structure emerged. We revisited the entities that had already been defined and finalized in the Licensing Application data model from (a) above, available in Fig. 2. Identifying and storing semantics was partially met by identification of entities and attributes as in (3) above (Codd, 1970). Establishing relationships along with their multiplicity and cardinality completed this effort. The final entities are all in Third Normal Form (Codd, 1972). The data model represents a generic model that can suit the UK and any other regulatory authority.

4.2. Defining Entities

4.2.1. Dealing with Multiple Marketing Authorizations

The MAA document is based around three different types of the entity, LicensingApplication:

- Mutual Recognition Procedure
- National Procedure
- Centralized Procedure

These form the hub of our design enabling all licensing applications to be classified under one of these categories. The {Mandatory, OR} notation in the Licensing Application data model (from Fig. 2) indicates that LicensingApplication is an abstract entity. This means that a licensing application has to be in the form of one of the subclasses that correspond to the three different types of the LicensingApplication entity. In other words, an instance of the superclass LicensingApplication must participate as a member of one and only one of its subclasses (Codd, 1979). (If an applicant wants to apply for more than one procedure for a particular product, they will need to submit separate licensing applications).

4.2.2. Dealing with a Choice of Marketing Authorizations

Each country or regulatory authority requires a separate marketing authorization. This means that two M:N relationships were needed between State (which equates to a country within the EU) and NationalProcedure, and between State and MutualRecognitionProcedure. The StateProcedure entity

had to be introduced into the data model (see Fig. 5) to cater for the M:N relationships. The exclusivity between the two 1:M relationships ensures that a licensing application can only be of one type.



Fig. 5 StateProcedure with its Exclusive Relationships

4.2.3. Dealing with a Generated Reports

We needed to decide how the data model would represent a report resulting from the submission or evaluation workflows (mentioned in (i) and (ii) from section 2). We had three options:

- (a) to include a link, in the LicensingApplication entity, to the feedback file,
- (b) to create a 1:M relationship, where each report was unique to a LicensingApplication, or
- (c) to create a M:N relationship where reports could be reused by different Licensing Applications, and one LicensingApplication could have more than one report (for multiple submission/evaluation attempts).

If we opted for (a) it would not be possible to retain a history of all reports generated for a particular LicensingApplication because each time a new report was generated, it would overwrite the last report in the database. If option (b) were implemented, it would enable one LicensingApplication to be linked with more than one report. For instance, a particular LicensingApplication may have reports for both a failed submission and a successful submission. (We wanted the reports for successful submissions and evaluations to contain standard text, but this option would not have allowed a standard report to be reused for other Licensing Applications).

As a result of the limitations that options (a) and (b) posed, we decided on using option (c). The implementation of an M:N relationship would require a Report entity, which would result in less redundancy and would enable the database to retain a history of reports (for each LicensingApplication) which the user could view. This has enabled us to use default feedback from the submission of valid Licensing Applications and their evaluations. We can also create tailored feedback for Licensing Applications that have failed their submission and/or evaluation.

The link entity, that evolved out of the M:N relationship between LicensingApplication and Report, is ApplicationReport (see Fig. 2). It contains the primary key of the LicensingApplication entity and the primary key of the Report entity. It also contains the date that the report was produced. The two foreign keys constitute the composite primary key for the ApplicationReport entity.

4.3. Defining Attributes

During the database design process only one entity has changed its status into an attribute. ActiveSubstance was initially thought to be a distinct entity. However, after consulting a medical consultant this proved to be unnecessary. Instead, an attribute was inserted into the Substance entity (see Fig. 2), to identify if a substance was active.

4.4. Recursive Relationships

All the relationships within our data models are standard 1:M relationships (see Fig. 2-4). There is only one example where we had to use a recursive relationship. If applicant 'x' is the marketing authorization holder for product 'x2005' in the UK and if they apply for marketing authorization for a new product (x2006) which is very similar to the old version, then they should provide the reference number for 'x2005' in their new licensing application. This has been represented in the data model by the recursive relationship for the Product entity, which is available in Fig. 2.

4.5. Impact of the Application and the Domain Specific Layers

The EJB application and its components, built upon our database (which is available from Juric et al., 2005), have had some impact on the way in which our database schema has evolved. Components from the application and domain specific layers that control user interfaces and application functionality have changed some of our initial database design decisions, for example, 'retrieving all available evaluation procedures after valid submissions' and 'generating reports for a failed evaluation'. We give two specific examples in sections 4.5.1 and 4.5.2.

4.5.1. Dealing with 'Status' in the Licensing Application data model

It was important to incorporate the concept of a Licensing Application's "status", as every Licensing Application needs its status recorded throughout the process of applying for marketing authorization. Therefore, the "status" was represented as an attribute. This has made searching simple and will enable an applicant to track the progress of their Licensing Application. Possible values for this attribute are:

valid submission		passed evaluation
invalid submission	1	failed evaluation

4.5.2. Dealing with File Links

Another way in which the deployment and implementation of example components from the domain specific layers (in Fig. 1) influenced the physical design of the data model is in the modeling of 'reports'. The Report attribute within the ApplicationReport (discussed in section 4.2.3) has been incorporated using a link to a word file, instead of using a large text field as an attribute. This would make the application layer more efficient when generating reports as it would extract the text from the word file and display it to the user.

5. Database Implementation Issues

Although the database design incorporates all three types of LicensingApplication: Mutual Recognition Procedure, National Procedure and Centralized Procedure (as shown in Fig. 2) only the UK National Procedure has been implemented in the prototype. As a result, only the shaded entities in the Licensing Application data model (given in Fig. 2) have been implemented.

5.1. Disjoint Types of LicensingApplication

The three subclasses of the LicensingApplication entity are disjoint and as a result, they had to be mapped onto three separate entities that have common attributes listed in the LicensingApplication entity (see Fig. 6). This would have resulted in three separate tables being created for each ApplicationManufacturer. However, as our implementation decision focused on the NationalProcedure only, the highlighted entities from Fig. 2 are the only ones that had to be created as tables. Thus, the

design issue of exclusivity is irrelevant in our current implementation, as there is no longer any relationship between MutualRecognitionProcedure and StateProcedure.



Fig. 6 Generalised LicensingApplication Entity with its Sub-Types

Had we implemented the full-scale software solution, the user would have been able to choose between types of Licensing Applications. We would have had two foreign keys in the StateProcedure table referencing the MutualRecognitionProcedure and the NationalProcedure tables. These two foreign keys would be able to accept null values. A database trigger would have been required to enforce the rule that "at any given time, an instance of the StateProcedure table references one and only one instance of the MutualRecognitionProcedure or the NationalProcedure".

5.2. Submission Criteria and Evaluation Criteria Tables

Our workflows from (i) and (ii) in section 2 deal with submission of Licensing Applications and their evaluations. This means that the submission workflow uses submission rules R_i from the domain specific layer to check if a Licensing Application satisfies the criteria for successful submission. The same applies to evaluation rules E_i , which are applied to Licensing Applications that have been submitted. However, all these criteria are kept, as they are needed for applying R_i and E_i rules, within the domain layer, for instance, in the SubmissionCriteria entity (available in Fig. 3) and the EvaluationCriteria entity (available in Fig. 4). This means that rules R_i and E_i are defined upon the criteria set in these two tables. For example, when an evaluation is requested, the components from the application and domain specific layers would then compare the criteria stated in the EvaluationCriteria table within the submitted Licensing Application. If the document passes the evaluation criteria controlled by E_i , a default report would be displayed to the user, stating that the submitted Licensing Application has passed the evaluation procedure.

6. Conclusions

In this paper we report on the design and implementation of a database which supports the automation of procedures for marketing authorizations. This work is part of our ongoing research into a generic software solution for the automation of such procedures, focusing on their interoperability across various regulatory authorities in the world.

We have met all the aims listed in section 3. Our generic database schema (in Fig. 2) can assist in marketing authorization procedures in any country in the world. Such procedures are interoperable. This means that each country can choose which procedures are applied to licensing applications that are submitted to them (see sections 4.2.1; 4.2.2; 5.2). A family of related software applications can reuse our database schemas (given in Fig. 2 - 6) if they involve a similar workflow of application 'submissions' and their 'evaluations'. Thus, our data models can be reused for submitting and evaluating visa applications for the UK Home Office and any similar procedures.

There are numerous opportunities for future works. We will:

- work on a full-scale implementation that will further evaluate our database schema from Fig. 2,
- include more complex licensing application submissions and their evaluation procedures and address the implementation of submission/evaluation rules and their criteria within our architecture (Williams et al. 2005),
- facilitate the elimination of redundant information that applicants have to provide on different pages of their licensing application, and
- analyze Health Level Seven (HL7) (available at http://www.hl7.org), an ANSI accredited US health industry communication messaging standard, that has extended the protocol for exchange of healthcare information to include data repositories that are important for marketing authorizations. We would like to see if HL7 can be used as a support in the communication between different component layers of our architecture in Fig. 1. This includes communication between different applications and communication between applications and underlying infrastructures. The outcome of such work may give us more insight into the designing of databases and applications for marketing authorizations.

Our automation of licensing application submissions could restructure the format and decrease the volume of submitted data/information when applying for a medicinal product license. It would be important to see if this could affect our database design.

7. Acknowledgements

This work could have not been delivered without the advice of Dr Maurice Slevin, Consultant Medical Oncologist from St Bartholomew's Hospital, in interpreting the MAA document. We wish to thank our colleague Stephen Williams for his constructive comments and proof-reading of the paper.

8. References

Codd, E. F., 1970, "A Relational Model of Data for Large Shared Data Banks", Communications of the ACM, Vol. 13, No. 6, pp. 377-387.

Codd, E. F., 1972, "Further Normalisation of the Data Base Relational Model", Data Base Systems, Courant Computer Science Symposia Series 6, Prentice-Hall, pp.33-64.

Codd, E. F., 1979, "Extending the Database Relational Model to Capture More Meaning," ACM Transactions of Database Systems, Vol. 4, No. 4, pp.397-434.

Date, C. J., 2003, An Introduction to Database Systems, Addison-Wesley Publishing Company, Wokingham, England.

DeMarco, T., 1978, "Structured Analysis and System Specification," Yourdon Press, New York, USA.

Journal of Integrated Design and Process Science SEPTEMBER 2005, Vol. 7, No. 4, pp. 10

Franken, A., 2003, "ESubmission Management in the Field of Regulations Concerning the Marketing Authorisation Process - Differences and Similarities in Europe and the USA," Pharmazeutische Industrie, Vol. 65 No. 5A, pp 463-472.

Juric, R. and Juric, J., 1999, "The Application of the UML to the Modelling of Automated Support of Evaluating Medicinal Products across Different Regulatory Requirements," 1999 International Conference on Information Technology Interfaces Proceedings, Pula, Croatia, June 15-18, pp. 283-291.

Juric R. and Juric, J., 2000, "Applying The UML Modelling Elements in Complex Business Environment," 2000 Conference on Integrated Design and Process Technology Proceedings, June 4-8, Texas, USA, ISSN 1090 – 9389.

Juric R. and Juric, J., 2002, "Applying Component Based Modelling in the Process of Evaluation of Medicinal Products," 2002 Conference on Integrated Design and Process Technology Proceedings, California, USA, June 23-28, ISSN 1090 – 9389.

Juric, R., Williams, S., Slevin, L., Shojanoori R., Courtenage, S., and Milligan, P., 2005, "Component Based Automation of Regulatory Affairs in the Pharmaceutical Industry," submitted to the Journal of Healthcare Informatics, 2005.

Juric, R., Williams, S., and Milligan, P., 2005, "Experiences of Generating COTS Components when Automating Medicinal Product Evaluations," 2005 International Conference on Software Engineering and Knowledge Engineering Proceedings, Taipei, Taiwan, July 14-16.

Williams, S., and Juric, R., 2005, "Design Patterns for Distributed Applications in Healthcare Systems," 2005 International Conference On Information Technology Interfaces Proceedings, Dubrovnik, Croatia, June 20-23.

Williams, S., Juric, R., Shojanoori, R., and Slevin, L., 2005, "Submission and Evaluation Rules and Criteria for Marketing Authorisation of Medical Product Licences," submitted to the Journal of Software Engineering and Knowledge Engineering, 2005.