



9th International Symposium
on Leveraging Applications of Formal Methods, Verification
and Validation

-

Doctoral Symposium, 2021

Automating the referral pathways for
Multiple Myeloma through a Web Application and XMDD

Adam J Doherty

23 pages

Automating the referral pathways for Multiple Myeloma through a Web Application and XMDD

Adam J Doherty¹²

¹ Department of Computer Science and Information Systems
University of Limerick, Limerick, V94 T9PX, Ireland

Adam.J.Doherty@ul.ie

and

² Health Research Institute

HRI@ul.ie, <https://www.ul.ie/hri>

Abstract: Multiple Myeloma (MM), a type of bone marrow cancer, is diagnosed by measuring monoclonal proteins, paraproteins (PP), and serum-free light chains (SFLC) in the blood. These proteins can be detected in healthy individuals at a lower level. This condition is called Monoclonal Gammopathy of Uncertain Significance (MGUS). MGUS is associated with a risk of progression to MM at a rate of 1-2% per year. Early diagnosis of MM correlates with improved overall survival for patients, so early referral of suspect cases is important.

In collaboration with the Haematology experts at the University Hospital Limerick and the SCCE group in Computer Science, we designed and implemented a software application that improves and streamlines the current process. This (online) application is developed with modern XMDD technology, using the DIME low-code application development tool. The application faithfully maps the reference algorithm in an automated way and applies it to a consultation data set. The novelty consists in the adopted technologies, that improve the early validation and correctness of the software, and ease the human understanding and the modification turnaround of the application.

Keywords: Model Driven Development, Health Technologies, Scientific workflows, Risk stratification, DIME, Domain Specific Languages

1 Introduction

Multiple Myeloma is a type of bone marrow cancer that can be detected by measuring risk factors in the blood. These risk factors can be detected in healthy individuals, this condition is known as Monoclonal Gammopathy of Uncertain Significance (MGUS) [RR19]. In the Haematology Department of the University Hospital Limerick, a risk stratification process is used to detect patients with possible MGUS and refer them for further treatment or monitoring in dependence of the risk class. Based on the current risk factors in their test results, this risk stratification may also direct the General Practitioners (essentially, the family doctors) in managing the further care for a patient at lower risk. This risk stratification process is currently carried out manually within the Haematology Department. As staff is usually very busy, delays in this evaluation could possibly impact the patient referral timeline, and thus the speed of further care.

We aim to streamline this timeline by developing *MyMM*, a web-based application that will carry out the risk stratification process on a scheduled basis. By removing the human component from the referral process, MyMM aims to provide a fast and fail-safe risk stratification and referral for patients. The MyMM web-based application is designed and built using modern eXtreme Model Driven Development technologies [BNNS16]. A further goal of MyMM is to constitute a blueprint for other applications in the healthcare domain in the context of the Cancer research networks developing at the University of Limerick and potentially in Ireland. This research initially started as part of the UL Cancer Network (UICaN) Summer bursary scheme in 2020, resulting in the initial proof of concept for the application. This process was developed on the basis of a prior study conducted on the long term follow up care [KLT⁺18] and a patient risk stratification algorithm that was previously developed [RR19]. This risk stratification algorithm is currently used in the Haematology Department at the University Hospital Limerick to determine the further care for patients with possible MGUS. It is however implemented manually, i.e., a skilled healthcare staff member must know the algorithm and manually apply it to the patient's data sets that come in as a continuous stream from the laboratory. As there are several processes in the healthcare domain that require a human expert to process patient information along a well-defined algorithm, we are confident that this work will also serve as a template for other automation projects in the healthcare domain.

In this paper, Sect. 2 will outline the objectives of the MyMM application, then Sect. 3 will define the current manual process in place for patients to be screened and diagnosis for multiple myeloma. We then give a brief overview of the technologies used in developing the proposed solution in Sect. 4. Sect. 5 will introduce the MyMM application followed by the Model Driven Development process in Sect. 6. Lastly, in Sect. 7 and Sect. 8 we will discuss the results, conclusions, and the next steps for the MyMM application.

2 Objectives

This project has a number of goals:

Build on the current POC Application: It builds on the work already completed and bring the current proof of concept application to a fully functional web based application. This full scale application serves as a template for similar future automation projects.

Referral Optimisation: We aim to streamline and optimise the efficiency of patient referrals to the Haematology Department.

Provide a Reliable Process: We aim to automate the risk stratification of patients, and thereby ensure that each patient is processed quickly and correctly, with a top degree of accuracy for every patient. It gives certainty on the evaluation and referral timelines for each patient. As the diagnosis is time critical, it is a quality indicator to be able to process the data reliably and quickly, with a fast communication of the outcome to the requesting physician.

Remove human dependency: In the current process, each patient in the data set is manually processed through the risk stratification algorithm by a staff member of the Haematology

Department. This runs a risk of human error, with possible dire consequences, and for sure has an overall lack of visibility of the metrics and performance of the risk stratification. Another aspect of requiring human interaction is the time taken to complete the manual processing and the availability of the staff member to complete the task. By removing the human element from the process we intend to remove these possible risks.

GP Management of Low-risk MGUS: Patients with low-risk MGUS do not initially require a referral to the Haematology Department. Instead, an information segment is made available to requesting physicians, to provide them with guidance in regard to the specific patient's further needs. This information segment is intended to provide information to physicians and patients, with the aim of better educating both parties on the diagnosis and further care required for a patient with possible MGUS.

Automated Data Capture: As the risk stratification process is currently manual, there is no way of automating the data capture for statistics on the use of this process. As part of the proposed application, we aim to implement an automated data capture of the patients processed by the risk stratification. This data may be later used by the Haematology Department staff to further refine the risk stratification process.

Process Templating: The development of this application is intended to serve as a template for other projects in the healthcare domain. To achieve this, we analysed the specific requirements to create applications that process patient information in order to run a classification algorithm, and abstracted generalised functionalities that can be re-used throughout future projects of a similar pattern in the same domain.

3 The Multiple Myeloma Risk Stratification Algorithm

Multiple Myeloma is a form of bone marrow cancer with circulating monoclonal proteins. It is a very serious condition: with a survival rate of less than 60 % over 5 years[Ame], it is a form of cancer classified as a cancer of plasma cells[Woo15]. An early diagnosis of Multiple Myeloma is correlated with improved overall survival of patients, and early referral for suspected cases is required in order to improve the chances of a positive outcome. Multiple Myeloma is diagnosed on the basis of a set of parameters contained in the patient's blood. Even when the values of these proteins are low and there is no myeloma, low levels indicate a condition known as Monoclonal Gammopathy of Uncertain Significance (MGUS). Patients with MGUS risk progressing to Multiple Myeloma at a rate of 1-2% per year, so also for them it is important to have a follow-up, monitoring strategy.

3.1 Diagnosis

As the regional level three hospital for the Midwest region in Ireland, the University Hospital Limerick (UHL) is mandated to treat all the *Urgent – Serious conditions that require emergency intervention* for patients in its catchment. In its Haematology Department, a reference algorithm for the diagnosis of Multiple Myeloma and its predecessor MGUS are already in place, together with the related scoring system for patient referrals. They are currently systematically applied

to all the patients with the risk of MM [RR19]. Two main risk factors are used in the process of determining MGUS: a paraprotein level (PP) of above 15g/L and a Serum-free light chain ratio (SFLC) outside the normal range (0.26-1.65). If a patient's blood test result returns a PP value, then a SFLC ratio is automatically requested on the patient's sample. Once both a PP value and a SFLC ratio are available for a patient, then the patient's results are processed through the risk stratification algorithm shown in Fig. 1. The outcome of this process is to decide which one of seven pre-generated comments will be sent back to the requesting physician to manage the patients with low-risk MGUS and provides clear referral pathways for intermediate and high-risk MGUS patients. Effectively, this is a classification algorithm. The aim of this comment (textual advice) is to direct the General Practitioner to the proper management of the patient's care. The advice can range from suggesting follow-up checks with the patient at a set interval for patients with low-risk MGUS, to giving a clear pathway for referral to the hospital's Haematology Department for further assessment of patients with a medium to high-risk MGUS.

The risk stratification algorithm is shown in Fig. 1. In order to create a representation of the algorithm in our technology more faithfully as possible to the algorithm and also to its granularity and layout, we broke down the different components of the stratification and developed processes to match each section of the workflow. The initial scope we set ourselves was to focus on the case where IgG and/or IgA paraproteins are detected. This algorithm is reported in Fig. 1 (left). The other case on the right concerns the detection of IgM paraproteins. It is to a large extent identical, so we concentrated on the left branch, covering the IgG/IgA cases. We provide next a top to bottom description of the successive different screening criteria for the stratification, which map in our model to different value comparators.

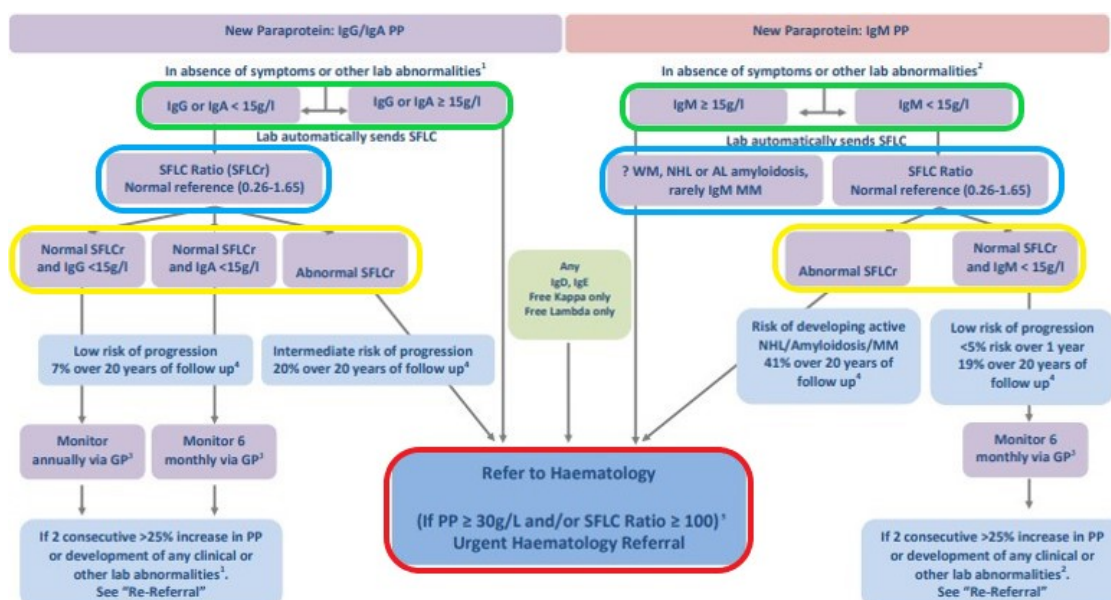


Figure 1: Excerpt from the UHL Guidance on Management of MGUS in Primary Care[RR19]: Risk stratification workflow

We call the successive steps *layers* as this suggests the successive specialization of the case by case evaluation in this decision tree.

Paraprotein Assertion (denoted in Green) With a threshold of 15g/L, we introduced a parametric comparator for decimal values. As the full range is considered (both less than and greater or equal to the threshold), we designed the comparator component in such a way that it can be utilised in both processes.

SFLC Ratio Thresholds (denoted in Blue) This layer checks whether the patient's SFLC ratio falls within the normal range, ($>0.26 - <1.65$). This layer of the stratification requires a dual threshold comparator for decimal values.

Outcome Layer (denoted in Yellow) This layer further refines the classification resp. to the outcome of the SFLC ratio check and the type of paraprotein present in the patient's data. At this layer, the normal/abnormal SFLC ratios determine the diagnostic outcome (low vs. intermediate risk of progression). This layer again uses Boolean comparators.

Referral Layer (denoted in Red) This layer of the workflow is used to assert the type of referral required. If a patient's initial PP value is above 15g/L or if their SFLC ratio is considered abnormal ($>0.26 - <1.65$) then the patient is referred to the Haematology Department. If the SFLC ratio is normal, different types of PP lead to generate a different response for the patient's further care through GP management: annually in case IgG was detected and every 6 months in case IgA was detected. This section of the workflow requires the use of the same Boolean operators and comparator types from the top layers, but with different threshold values.

We see that it is possible to abstract a small number of parameterisable components, that are generic enough to cover not only these MM/MGUS-specific decisions (on these data and thresholds), but general value-based classifications. Similarly, we can develop generic processes that we can utilise throughout this risk stratification algorithm and also in other cases.

Fig. 2 shows an excerpt of the risk stratification process model in the DIME development environment. Note that we have spent special care to keep in the Process model the same logic and the same steps as they are in the original algorithm in Fig. 1, in order to support the users' understanding through familiar layout and clear correspondences between the two versions.

We show in both Fig. 1 and in Fig. 2 the corresponding process blocks to highlight each layer of the process and how the corresponding DIME process model matches the similar layout and process flow. In Fig. 2 we see the initial starting point for the risk stratification in the Start block at the top of the figure that passes a patient's paraprotein value into the process model and holds this data object in context so that it can be then used throughout the rest of this process. The stratification process consists of a number of different, phase-specific process blocks that are appropriately connected. Each of these process blocks contains further finer granular processes, that carry out each comparison for each layer of the risk stratification seen in Fig. 1.

3.2 Manual Processing

The initial prompt for a patient's data to be processed through the risk stratification is sending a request for a blood sample to be processed labelled as Step 1 in Fig. 3 below. This can be a gen-

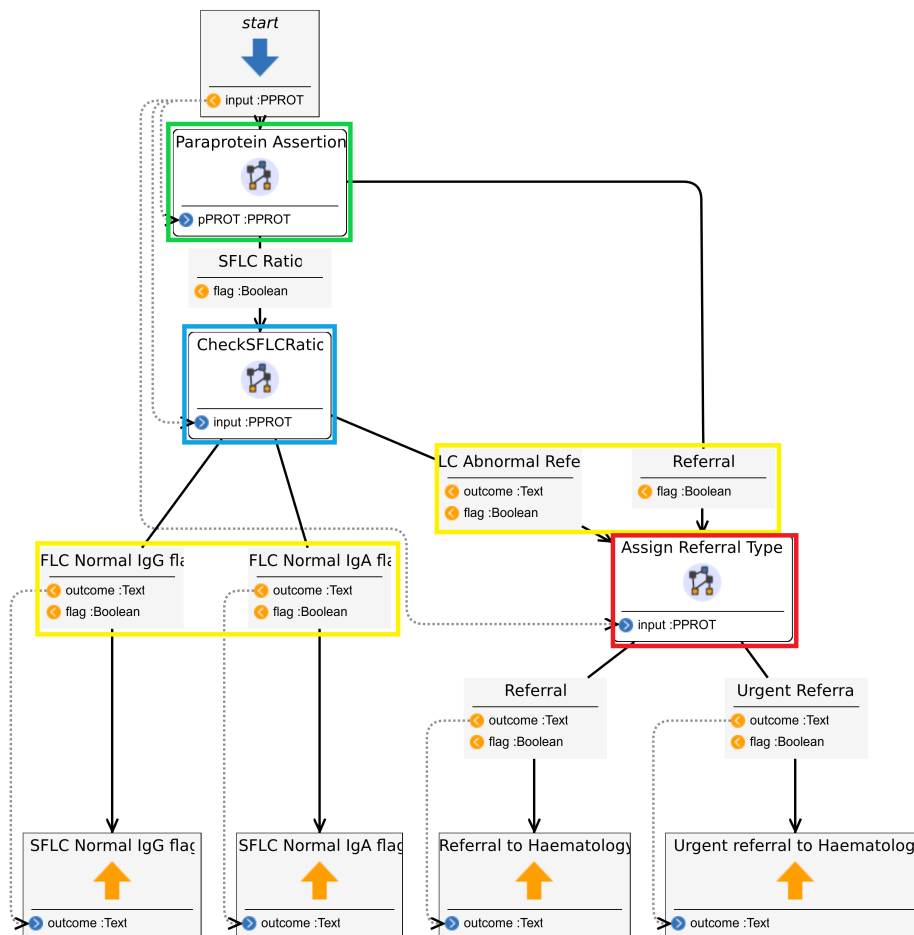


Figure 2: The risk stratification process model in DIME: excerpt of the top level control flow

eral practitioner or other healthcare providers. The next stage of this process is for this blood test to be transported to the Haematology Department in the University Hospital and for the sample to be processed, shown as Step 2 below. Step 3 below denotes the result of this sample being stored in a database called iLab and if the sample shows a paraprotein of any level, this will then trigger a SFLC ratio to be taken on the sample. Once both of these results are available, a senior member of the Haematology Department’s staff searches the patient’s result in the database and apply the risk stratification workflow in step 4. In the current process manual process, there is no way for the current system in place to notify the staff member once both results are available. This means that currently, the member of staff needs to continuously check if both results are available. Once completed, they then update the result with one of seven pre-generated comments. This comment is dictated by the risk stratification process and will determine the patient’s further care. Step 5 then denotes the system generating a HL7 message, which is then transmitted through a communication platform called HealthLink[Hea]. HealthLink is a secure communication platform with 5,665 clinicians registered [Hea] and is widely used in GP practices in Ireland. Lastly, Step 6

denotes the response message being communicated back to the requesting physician. Depending on the outcome of the risk stratification, this response will contain guidance on the patient’s further care in the case of a low risk MGUS and in the case of intermediate high risk, the message will include a referral for the patient to be seen by the Haematology consultant.

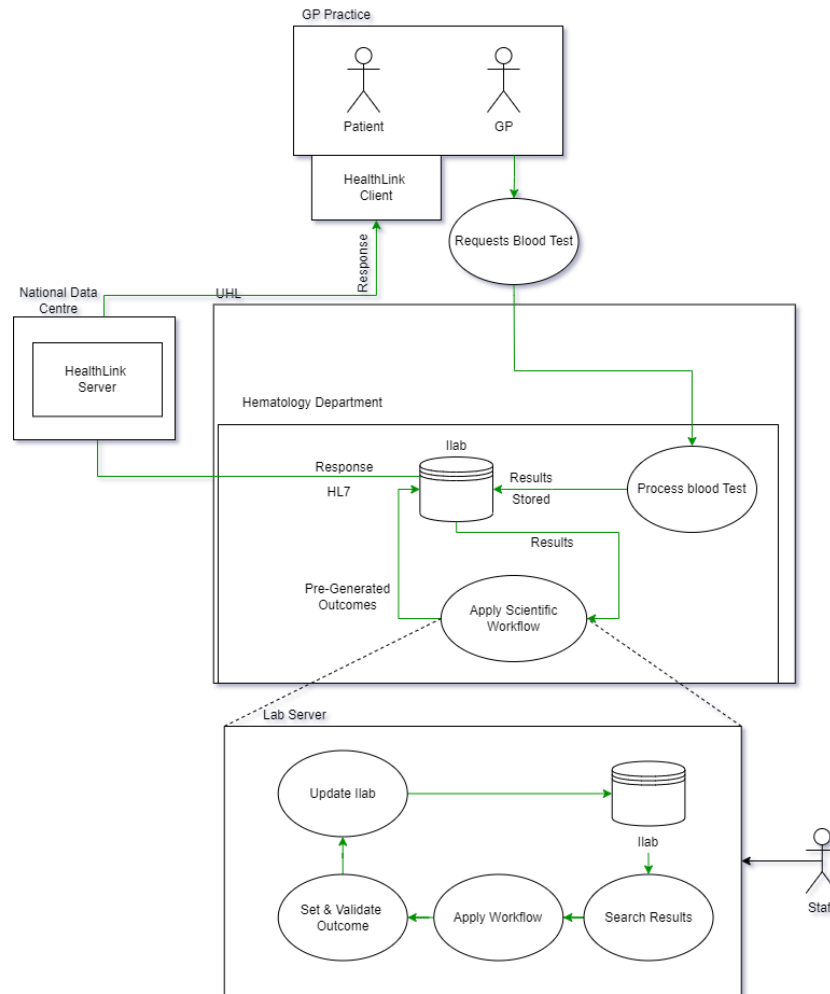


Figure 3: A depiction of the current manual process

4 The MyMM Technologies

In this section, we briefly describe the technologies in use in the MyMM web application and its environment: this encompasses the University Hospital, the GP network and in general the HSE, which is the national Health Service, mandating a number of specific technologies, protocols, and governance mechanisms.

4.1 XMDD development in DIME

We chose to develop the MyMM in DIME. DIME is a modern model driven low-code development framework for web applications [BNNS16], that is itself generated from higher level specifications in the Cinco [NLKS18] environment. Technically, it embraces the Language Driven Engineering (LDE) approach [SGNM19]. It is an evolution beyond the DyWA [SGNM19] environment, that had been already successfully applied to healthcare analyses, as in [MFG⁺14] and [WLM18]. A precursor had already been used to develop decision support systems [KM06] suitable for being transformed in a series of similar products, as we also aim in the context of the Health research Institute.

The DIME development environment lends itself to developing applications in an agile fashion while still upholding web-application development best practices. The development in DIME is not done through a manual code-base but instead through a graphical development environment and a library of generic or domain specific components. This makes DIME uniquely suited for use in mapping scientific workflows, like the risk stratification model in this project. DIME applications consist of 3 main, interlinked model types (data, processes and GUI) that are each used to define a different aspect of the resulting web application. They make use of libraries of atomic components called Service Independent Building Blocks (SIBs). There is another type called Native SIBs, which are the only ones manually implemented, or encapsulating services provided by external platforms. The advantage of this organisation is the ability to share encapsulated services and reuse them in different contexts. We take advantage of this capability by mapping the current risk assessment scientific workflow seen in Fig 1 to a maximally generic DIME process, that can be used as a template for other, future applications that map currently manual classification and the decision process to automated versions. The inclusion of the knowledge of the domain experts is critical when working in a healthcare domain. We find examples of XMDD Development using a graphical process modelling approach in the healthcare domain [BWLM17] has been seen to be effective. We aim to use the graphical environment to leverage this knowledge in the rapid development of the application. As the risk stratification is broken into different layers of assertions, we can break down these logical blocks into familiar processing [MBK13] done by the domain experts and the rapid deployment of web-based applications in DIME allows us to see reactive changes in the processing.

4.2 Health Level Seven

Health Level Seven (HL7) [HL7] is a global standard for electronic data exchange designed specifically for healthcare settings. At the beginning of the project, we aimed to use HL7 to return the processed result to the GP: the information guide or a referral for the patient, depending on their health risk status. While studying this technology, we found out that it includes the open-source library *Fast Healthcare Interoperability Resource* (FHIR) [OLY⁺20]. It is "a standard describing data formats and elements (known as "resources") and an application programming interface (API) for exchanging electronic health records (EHR)". Healthcare services use FHIR to create, transmit and digest HL7 messages. It is available in a number of different languages. As the FHIR solution has the ability to create an entire communication platform for communication through HL7, it served as an example of how we could generate and transmit

HL7 messages through a network. Due to the evolving infrastructure in the Haematology department throughout the development lifecycle of the MyMM application, another already existing path to communicate with GP's was found. As this method of communication was already integrated into the Haematology department, a decision was made during the project to use this communication method.

4.3 HealthLink

HealthLink [Hea] is the National messaging broker used by GP practices in Ireland. It is an electronic way to send and receive information in a secure way. HealthLink is used to transport clinical patient information between Hospitals and other healthcare actors. We aim to leverage this ability as the communication platform between GPs and the hospital. This concerns receiving the stratification outcome, and the advice of referrals for the patients to the Haematology Department. Using HealthLink and the HL7 messaging standard we aim to achieve a safe and secure executable environment and transmission system for patient outcomes. The concrete HealthLink server is located offsite, in the national data centre in Dublin. It uses FTP to dispatch the HL7 formatted messages to the clinicians.

5 Screening with the MyMM web application

There are multiple different examples of automation approaches in the healthcare domain. The different areas and settings for healthcare automation require different technologies and depends on different forms of human interaction. Some examples include natural language processing for annotations provided by consultants in an emergency room setting [DBZ⁺13], to artificial intelligence to accurately diagnose through the use of chest X-rays [BMRS20]. Each of these different approaches to automation requires an element of human interaction in the completion of the process. These examples serve as a good template for approaching the automation problem based on their setting. This required us to look at all possible utilises available in our context. Our proposed solution is to integrate a web-based application to process each patient on a scheduled basis. This application has the ability to ingest a set of data from the Haematology Departments' database which is automatically generated with patients that have been tagged to be processed through the risk stratification. Each patient's results are processed through the risk stratification and one of the pre-generated outcomes is assigned to the patient in the system. The proposed connection for the application to ingest data consists of an API call to a laboratory server that returns a list of tagged patients to be processed.

Shown below in Fig. 4 is a depiction of the proposed solution. Step 1 denotes a clinician requesting a blood test for a patient. The sample is then transported to the Haematology department to be processed denoted as step 2. Once the sample has been processed, the result is then stored in iLab. If a paraprotein value is detected in the results at this point, a SFLC ratio will be automatically requested. At this stage of the process, the results are then sent to a database local to the MyMM application deployment environment seen in step 4 below. At predefined intervals, the MyMM application will ingest the patients results from this database, apply the risk stratification, set the pregenerated outcome for the patient, and insert this outcome to a table

in the local database. This outcome is then returned to the iLab database and Step 5 denotes this outcome then being converted to a HL7 formatted message, which is then sent to the HealthLink server. This is then sent back to the requesting physician to advise on the patients further care. Comparing the proposed solution to the manual process in Fig. 3, the additions we have added have no impact on the process of requesting a blood test or receiving the outcome for a requesting physician. The proposed connection for the application to ingest data consists of an API call to a laboratory server that returns a list of tagged patients to be processed. Initially, we designed this to communicate through JSON with the application. The output was initially thought as a validated HL7 formatted message, to be transferred via FTP to the HealthLink server. This required us to build a message factory for HL7 messages. We later made changes to these solutions to use an SQL database local to the deployment environment for the MyMM application for both the importing and exporting information for the application. One of the main reasons we altered this is that the SQL database could interact with iLab to store the outcome for the patient. Another reason for this orchestration is our inability to interact directly with the iLab database. iLab is a SQL database local to the Haematology Department which does not allow for external applications to make updates to its records. The iLab database is also connected to the national HealthLink server and to a number of validation services for the communication sent through HealthLink. The additional database that the MyMM application is integrated with, was set up by network administrators in the department so as to avail of these validation services which also allows for the outcome from the MyMM application to be recorded in iLab automatically. This allows for much better data completeness for the Haematology Department. The iLab database then generates HL7 messages to communicate the outcome with the HealthLink server and requesting physician. This led us to make the alterations to the proposed solution to include the SQL database into the applications design.

5.1 MyMM Application

The MyMM application is the realised web application created to automate the patient risk stratification. This application consists of multiple components to fit the requirements of the research. The main functional aspect of the application is to process patients through the risk stratification model. This process model has been built to generalise the operations needed to complete the required workflow. We have focused on accommodating records with IgG or IgA PP values. Taking this into account, the generalised comparators developed for the current scope of the risk stratification can be used in future iterations of the application to complete the full workflow depicted in the UHL Guidance on Management of MGUS in Primary Care[RR19]. One of the key features of the application is its ability for scheduled execution. This feature is instrumental to the objects of this research. This functionality allows us to set a timed process each day to update the current data model with new patient records, process all patients with a PP and a SFLC value available, and then to export the outcomes of this to the database table. This scheduled execution does not interfere with the normal operation of the application, allowing users to interact with the user interface of the application without issue. This kind of execution is not natively available in the DIME development environment and uses the Wildfly[WIL] application server functionality to achieve scheduled execution. The approach to the implementation of this functionality allows us to interact with complex data models in the data model in a concurrent

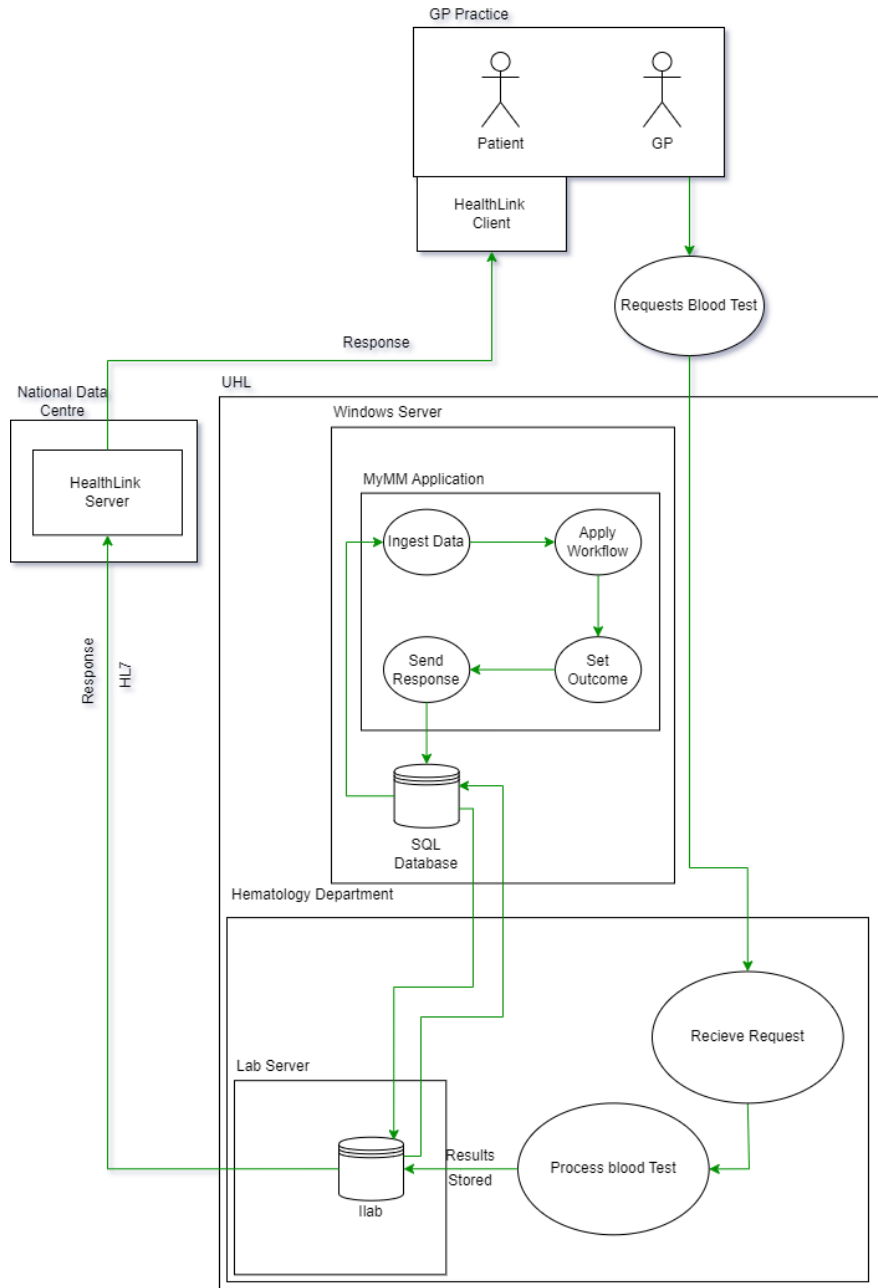


Figure 4: An overview of the proposed solution and its effect on the referral process

way. This scheduled execution is instrumental in developing automated applications and will be leveraged in multiple other applications in the same domain. The application has a minimalistic user interface seen in Fig. 5 below. This interface is used to view all patient records ingested by the application and the outcome of the risk stratification. This interface implementation has

the ability to trigger the process to update the current data set stored in the DIME data model, process each patient separately or all patients in the data model at once, and provide a user with a filtered view of all patients with a PP in their result set. The other aspects of the user interface include a process walk-through for staff members in the Haematology department to view a step by step walk-through as the MyMM application stratifies a patient's records. This aspect of the application will allow a user to select a patient's record and follow the application's processing of the risk stratification step by step through the workflow. Each step of the workflow is broken out into a different GUI model in DIME that will take the patient's record and provide an explanation for logic applied in each layer of the workflow. This educational segment is aimed for use by the staff members of the Haematology department to better educate them on how the process is applied, the different outcomes from the risk stratification. Both the scheduled processing of the Haematology departments data set and the walkthrough process are one in the same system. The web interface will be accessible by the staff to view the stratification process walkthrough while the scheduled processing doesn't rely on the user interface and can run without any requirement for a prompt from the web interface. In the following section, we describe the model driven approach taken to develop the main components to the MyMM application. This section describes the modelling process to map the Haematology Departments current data and the risk stratification process. This section will also define the model-driven approach taken to integration the MyMM application into a highly regulated network in the University Hospital for importing and exporting data into these environments.

6 The Model Driven Development

6.1 Data Model

In the initial stages of the application, we needed to map the DIME data layer to represent the data set that already existed inside the Haematology Department. Using the Data model in the DIME development Environment, we create representations for the different relationships between the data that we are trying to model and assign relationships between these complex objects to uphold the original data relations. During this process, we developed the data structure depicted in Fig. 6 to model a patient and their relationship with the rest of the data provided in the data source. The information stored in the database includes a record for each request sent to Haematology Department. A breakdown of each complex object:

ClinicianInfo Each record includes information on the requesting physician, which is used to communicate the outcome back to the requesting physician.

PatientData This complex object depicts the information that specific to the patient that is required to create a HL7 formatted message. This information is not directly required for the risk stratification and this has a one to one relationship with the patient.

Results The Results complex object contains a list of each different type of result that is possible to be returned from the processing of a patients blood test. This object has a relation to the PPROT object to store a list of paraproteins, if present. The Results object has a one to one relationship with a patient. This is to account for the process for a SFLC ratio to be

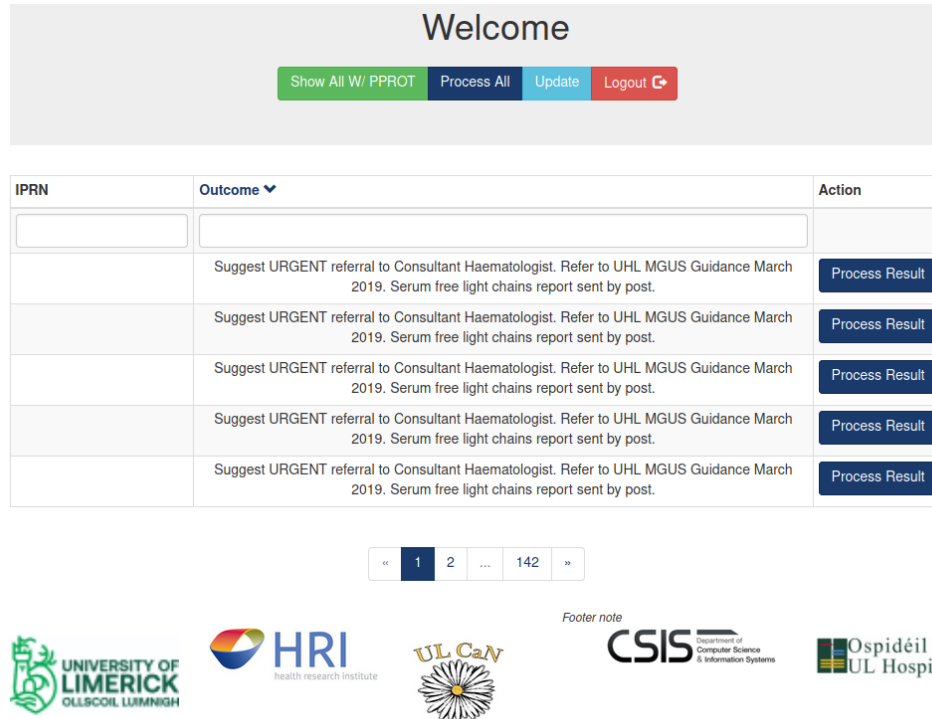


Figure 5: A view of the applications private home page

requested. As this happens after the result has been processed and a paraprotein detected, a partial result may be tagged to be processed without a SFLC ratio available.

PPROT This is modelled to hold the information required for paraproteins to be processed through the application. As a patient’s sample can possibly return multiple of paraproteins. A paraprotein has a many to one relationship with a Result object

As shown below, the system takes in far more information than just the risk factors for Multiple Myeloma. This is because the extended system aims to be able to send information based on the outcome of the assessment model back to the GP. For this purpose, we include the information required in separate complex objects associated with the patient model.

The initial scope of the application included taking the data from a web-based resource and populating the system at the application’s start. We developed a java native method that connects to the data source and ingests it into the Patient complex object shown in Fig. 6. Once the system was populated with the Patient information, we were able to view this information through the application to assert the correctness of the data ingestion.

We implemented a minimalist user interface for testing purposes, so that we can view all the entries in the data set to assure the data was ingested and updated correctly. The final application for processing the results and output will not require an interface as it will send the outcomes to the GP practice without any human intervention.

During this stage of the project, the PP value was provided as a comment on the lab’s results

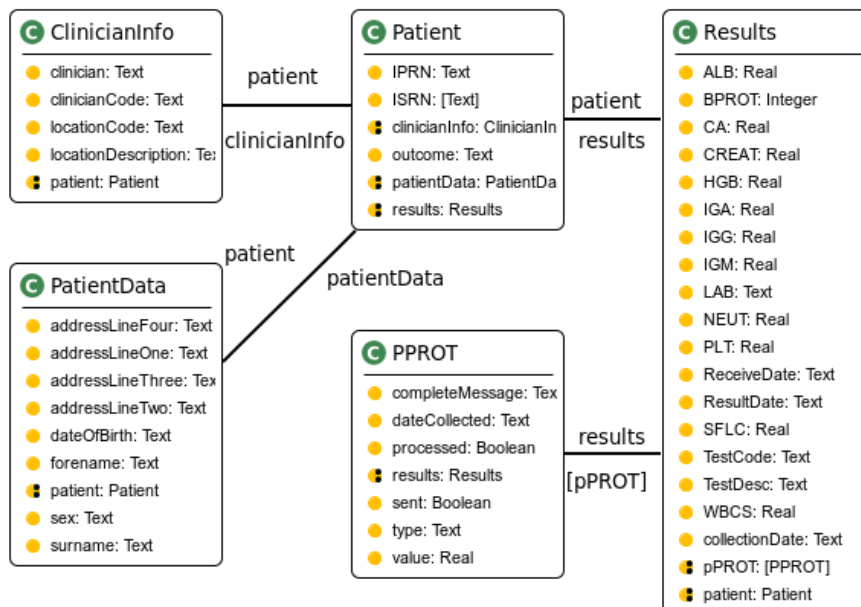


Figure 6: View of the Patient Model in the DIME data layer

as a freehand text. After attempting to extract the PP type and value from that freehand text, we decided that the natural language processing was outside the scope of this proof of concept and we proposed to the Haematology department to alter in the future the data structure for storing PP values and their corresponding types.

6.2 Patient Assessment Process Model

The scientific workflow that assesses the health score that determines the patients further care is realized as a collection of process models. Fig. 1 shows the current algorithm used by the Haematology Department to evaluate a given patient’s pathway for further care. We identified the different elements that evaluate a patient’s current data with respect to thresholds and created the corresponding atomic and process SIBs. As our aim is the maximal reusability of the process components, we adopt a generic SIB approach for the comparators, that should not rely on or use any of the unique data types of Fig. 6. The process model in Fig. 2 follows the original logic in Fig. 1: each workflow node corresponds to a specific section of the patient assessment model. This allows users to map the actions of the system to the steps of the manual process. Many of these SIBs are in fact subprocesses so that we can manage the complexity by means of hierarchy. For example, in Fig. 2 the Paraprotein Assertion SIBs compares the patient’s current Paraprotein level to the threshold value: 15g/L. The referral pathway is followed according to the outcome, and the patient’s results will then be assessed to stratify the level of risk and the type of referral required.

In the initial design, the process SIBs that mimic the risk stratification were built directly along the original algorithm and thresholds, i.e. as non-reusable special purpose SIBs. Once the

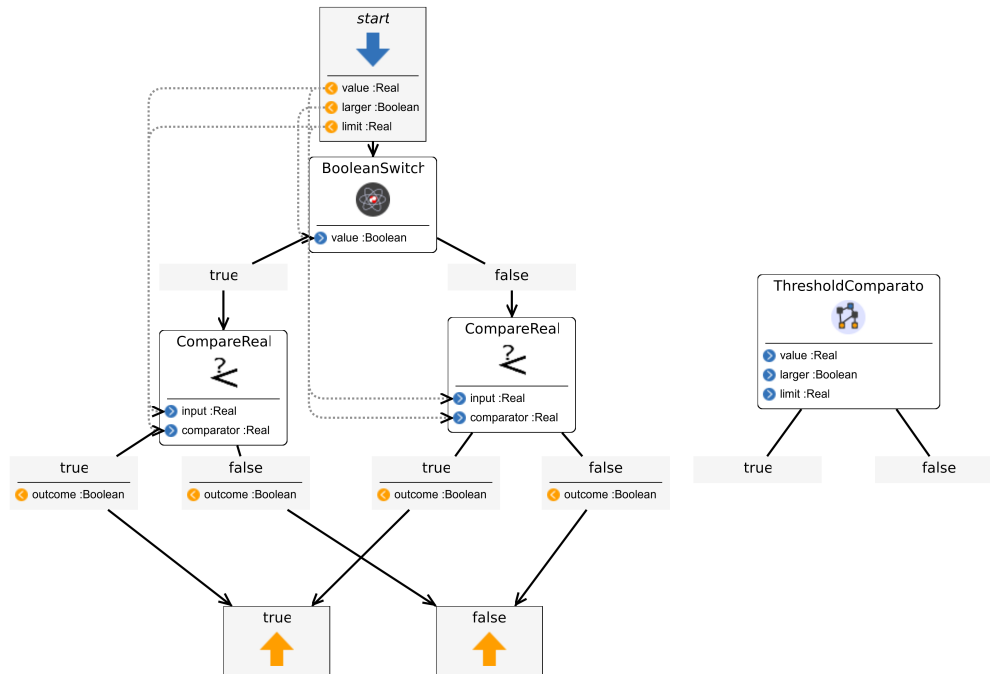


Figure 7: Abstract Comparator Processes

workflow was built into DIME process and the logic validated, we looked at how to abstract the comparisons in such a way that generic comparator SIBs could be reused in the future. Fig. 7 shows the body of one of the abstract comparators on the left, and its associated process SIB on the right. On the left, we can see that the model is parametric in the following three inputs: Value is the input value, of type real, Larger is a Boolean value indicating the comparison of choice: a *greater than* or *less than* comparison, Limit is the threshold, also of type real.

This generic comparator process uses internally the atomic SIB CompareReal, that was initially created for the top layer paraprotein value assertion of Fig. 1, in green. This abstraction provides a generic value comparator for real values that is independent of the project-specific complex data types.

Through these abstracted operators, we start to build a repository of different abstract operators for future projects in any application domain.

6.3 Automation

While so far patient records were analyzed one by one, by clicking on the MyMM GUI, our goal is to fully automate the process. We thus need a trigger process that is initialised at MyMM application start, then continuously monitors the arrival of new datasets and triggers their processing. This automation requires three mechanisms to be run autonomously: 1) a process to ingest new information for the Haematology Departments data source and update each pre-existing patient with the new result data, 2) a process that automatically processes the entire data set of patients, and 3) a process that formulates the risk stratification outcome and notifies the patients' GP

practices.

The process model for, essentially, batch processing all the results is realised in Fig. 8. This

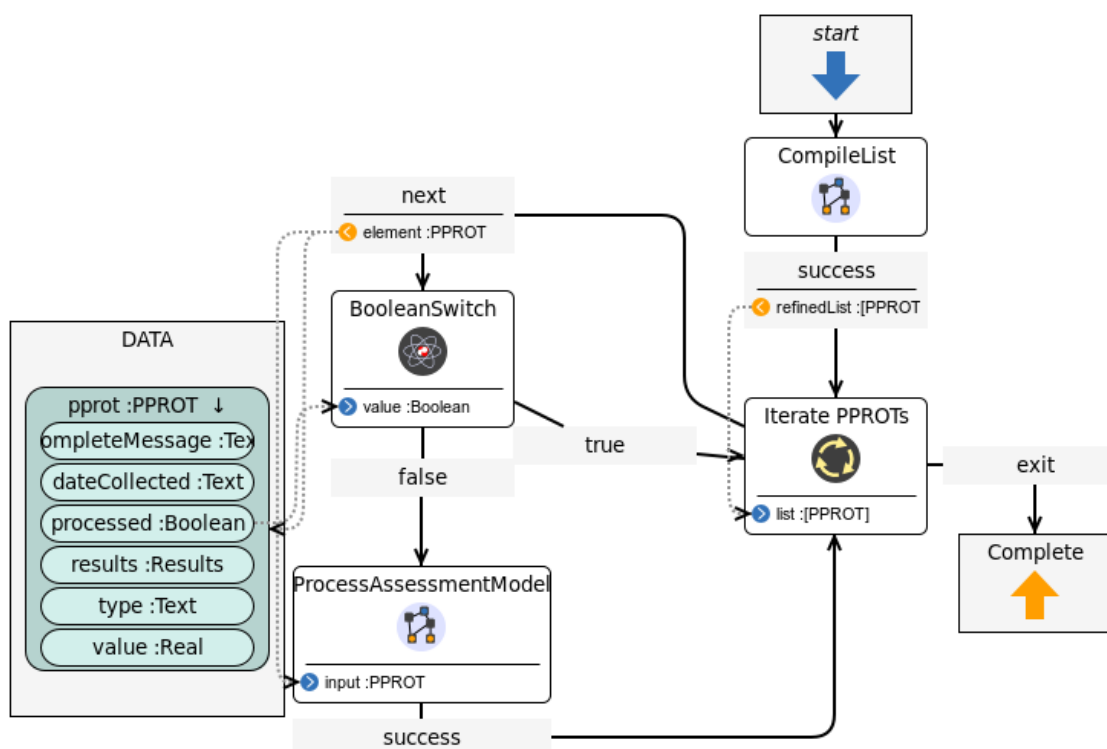


Figure 8: DIME process to iterate the individual record assessment over the entire dataset

solution is simple, but has a drawback: due to the single-thread nature of DIME processes, if this process is embedded in a time-based looping process, activated for example through a button in the MyMM application, the application’s interface will stop responding. The automation feature is now realized instead through the use of Timer Service [EJB] in Java EE, which allows us to set a scheduled execution of ingestion, processing and information export for the application. Specifically, we set the MyMM polling time to match the polling policy of the refresh of the ingestion database with new patient data. This way, we guarantee a quick stratification response and processing as soon as new patient data is provided.

6.4 Data Ingestion and Output

When addressing how the system transfers information into and out of the application, there are certain constraints to be considered. All the test results conducted in the Haematology Department are stored on their local iLab server, and we do not have the ability to directly interface our application with it in an automated fashion. Therefore, reports are pulled automatically from

iLab and stored in a network location local to our application. This is ingested via a new SIB that pulls the report from an API in a JSON format compatible to the MyMM data model. This was the initial approach to the integration of the MyMM application into the Haematology Department but this approach evolved to use a currently already existing SQL database: our application can ingest the consultant's data set from it and write to a table with patients that require referrals.

As this new solution utilises a path for referral that already exists in the hospital network, this removes the requirement for our system to create, validate and transmission of HL7 messages. Testing the HL7 messaging for this would have required access to the hospital's network, on a live HealthLink system, potentially incurring unforeseen issues during the integration stage of the development life-cycle. Another concern the new solution addresses is the interaction with the hospital's network, only testable by deploying onsite in the University Hospital. This was a concern throughout the development as we were unable to test on-site for connectivity and other levels of authentication in the hospital's highly regulated network. Lastly, by using the pre-existing path for sending referrals, we avail of the validation server in the hospital's local network, that validates messages sent to the HealthLink server. For this alteration, we modified



Figure 9: Initial Populate SIB created to populate the DIME data layer

the initialPopulate SIB in Fig. 9 to read from a database instead of the expected JSON formatted web resource. This led to a larger rework on the ingestion feature of the project, because the database is structured differently from the original synthetic data set we were provided, and follows a different naming convention.

7 Results

As a result of the development of this MyMM application, we have been able to create multiple different reusable components. These components can be viewed as building blocks to a larger goal creating a healthcare domain specific language. We aim to continue to expand this DSL through the development of further web-based applications, e.g. a patient-oriented information portal in collaboration with ULCaN Pillar 1 (Dr. Dervla Kelly and Dr. Des Leddin). This portal is not directly related to automation but still can be used to advance the healthcare DSL and allows for expansion into different areas of healthcare, outside the realm of automation.

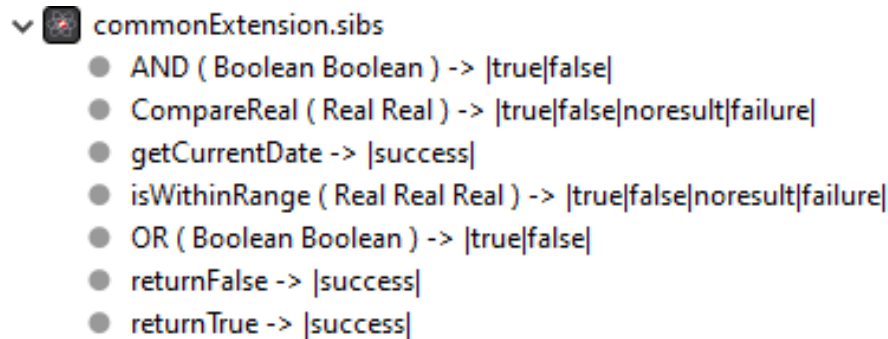


Figure 10: The CommonExtension Service Library

7.1 Reusable SIBs and Processes

Throughout the development cycle of this application, we have had the goal to use the development to create a template for other projects in the health domain. During the development process, we have created multiple Service Independent Building Blocks (SIBs) in the DIME Development Environment. When starting the development of the MyMM application, there weren't other examples of using graphical development frameworks to realise these types of automation service libraries in the healthcare domain. Following the creation of the MyMM application and its service libraries, we created a number of other applications in the healthcare domain that reuse and add to these services libraries. These SIBs are not native to the DIME development environment and have been developed for the purpose of being used in future projects. We have broken out our blueprinting into different kinds of Domain Specific Languages (DSL).

7.1.1 CommonExtension Service Library

Currently, DIME comes with a common service library, a collection of Java native methods to aid in developing DIME applications. These native java methods are then described in the CommonExtension.SIBs to adapt them from a code-based solution to DIME's graphical development environment. During the development of this application, the need for value comparators and other Boolean operations, lead to the development of the CommonExtension service library shown in Fig. 10 below.

Certain functionality like the `getCurrentDate` seen above in Fig. 10 was developed for a specific purpose in the application but after refactoring, this method was abstracted away from its dependency on this application's specific resources. The aim for this step was to contribute these SIBs to the current common SIBs collection, that come with the DIME distribution and is provided to all the developers.

7.1.2 Healthcare Service Library

The reason we created a healthcare DSL is to continuously extend it throughout the development of several projects in the healthcare domain, that share it. As we can see in Fig. 11, we have for example the functionality to create a connection to the consultant's data set. This was originally


- ▼  healthCareRelated.sibs
 - CombineList (DataType DataType) -> |success|
 - DatabaseConnect -> |success|
 - hl7ROMMessageCreate -> |success|
 - isResultsNull (DataType) -> |true|false|noresult|failure|

Figure 11: The Healthcare Service Library

designed with the intent of being able to connect to a single data set but with few alterations, this was abstracted to work in general for any database connection for use in other projects. In most healthcare settings, there will be a requirement to ingest a data set from/to a pre-existing database. Patient data information is highly regulated and therefore outside the scope of being collected in the DIME data models in this application. The initial requirements for this application foresaw the system to directly communicate with the national HealthLink server, to send the outcome of the risk stratification. This required the application to generate an HL7 formatted message with the patient's outcome, to be sent through the hospital network. After this feature was partially developed, an alternative pre-existing pathway was preferred, using a database that has the ability to parse HL7 messages. The alteration from one way of communicating with the HealthLink server to an already existing pathway addressed certain issues with the development and testing of the HL7 messages on the hospital network. This new pathway has other steps for validating the HL7 messages and communication with the HealthLink server already established. As the HL7 message generator was already partially developed, we continued with the small alteration required to complete the scope of the feature. This led to the design of the hl7ROMMessageCreate seen in Fig. 11.

7.1.3 MyMM specific Service Library

Lastly, we created a MyMM specific service library to handle bespoke processing requirements for this application. Each of the functionalities in Fig. 12 has been developed to serve a specific


- ▼  myelomaRelated.sibs
 - backgroundServices -> |success|
 - initialPopulate -> |success|failure|
 - isSFLCEmpty (DataType) -> |true|false|
 - pushResults -> |success|
 - refreshData -> |success|
 - updateData -> |success|

Figure 12: The MyMM DSL: the functionalities required specifically for this application

purpose in this application, covering bespoke requirements. In Fig. 12, the SIBs isSFLCEmpty

checks whether a SFLC ratio is available for this patient. This SIBs is designed using a datatype that is specific to this DIME application, i.e., these SIBs are dependent on the data structure for this application. This means that for another application to reuse these SIBs it would need to have the same complex data structure used to model a patient in this application. As another example of this specific application dependency, the `pushResults` functionality seen above collects a specific set of patients in the data layer, then adds them to a database table so that they are then relayed to the requesting clinician. This requires the data structure and the endpoint to be the same as for this application, which in this case is dependent on the infrastructure where this application is deployed.

8 Conclusions and Next Steps

As advances in healthcare technologies are made, already existing infrastructure needs to be augmented with bleeding-edge technological advancements, in the same IT ecosystems. The main purpose of the MyMM application is to improve the early validation and risk stratification of multiple myeloma patients but it also serves a number of other purposes in the healthcare domain. Using the graphical development environment of DIME, we start to bridge the gaps in understanding between healthcare experts and information technology experts, and to accelerate the advancements in the domain. The use of DIME as a medium to communicate understanding with healthcare experts has been invaluable throughout the development life-cycle of the MyMM application. The application development supports the approach taken previous examples work[[MBK13](#)]. Through the use of traditional code-based application development, we would have no realistic expectation of different domain experts being able to participate in the development of the MyMM application. By automating the risk stratification of the patients with the MyMM application, we establish a kind of technology adoption through adaptation of new technology to legacy systems. We show that it is possible, and effective in enabling reliable forms of validation of processes that were previously completed manually. The development life-cycle of this application has not been without its challenges: for example, key requirements such as the scheduled execution are not included in DIME's native development environment. However, with the support of the DIME developers community, this was easily addressed. The MyMM serves as the first application developed with the aim of creating a healthcare domain specific language, and as a template for other web-based applications projects in the same domain.

The next stage of this work is to assert the correctness of the risk stratification process modelled in the application. We aim to use the initial synthetic data set provided by the Haematology Department to assert the correctness of the outcomes provided by the application: the outcomes for the synthetic data will be compared against the expected outcomes for the data set. This validation will serve as a reference to the correctness of the application before applying the risk stratification to actual patient data in a safe executable environment. Prior works like [[SCA⁺21](#)] define an approach of validating automated healthcare systems that we use as a starting point for our approach to systems validation.

The subsequent stage will be integrating the application into the Hospital network, taking into consideration that it is a highly regulated environment. This network will contain multiple levels of security to store and transmit patient information through the network. This needs to be done

in close collaboration with the UCH technical experts and responsables.

Acknowledgements: This work is funded by the HRI (Health Research Institute) through the UL Cancer Network (ULCaN) Summer bursary program 2020 and the ULCaN Pillar 4 project 2021-22. I would like to extend my deepest gratitude to my Supervisor, Professor Tiziana Margaria for their feedback and insight throughout my research. I would like to extend my sincere thanks to Dr. Ruth Clifford for their generous knowledge and expertise on this project. I am also grateful to Oliver Power and Antonio Rechemartinez for the support and in-depth knowledge.

Bibliography

- [Ame] Survival rates for multiple myeloma. <https://www.cancer.org/cancer/multiple-myeloma/detection-diagnosis-staging/survival-rates.html>
- [BGP08] H. Brenner, A. Gondas, D. Pulte. Recent major improvement in long-term survival of younger patients with multiple myeloma. *Blood, The Journal of the American Society of Hematology* 111(5):2521–2526, 2008.
- [BMRS20] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Computer Methods and Programs in Biomedicine* 196:105608, 2020.
- [BNNS16] S. Boßelmann, J. Neubauer, S. Naujokat, B. Steffen. Model-driven design of secure high assurance systems: an introduction to the open platform from the user perspective. In *The 2016 International Conference on Security and Management (SAM 2016). Special Track “End-to-end Security and Cybersecurity: from the Hardware to Application*. Pp. 145–151. 2016.
- [BWLM17] S. Boßelmann, A. Wickert, A.-L. Lamprecht, T. Margaria. Modeling directly executable processes for healthcare professionals with xmdd. In *Service Business Model Innovation in Healthcare and Hospital Management*. Pp. 213–232. Springer, 2017.
- [DBZ⁺13] L. Deleger, H. Brodzinski, H. Zhai, Q. Li, T. Lingren, E. S. Kirkendall, E. Alessandrini, I. Solti. Developing and evaluating an automated appendicitis risk stratification algorithm for pediatric patients in the emergency department. *Journal of the American Medical Informatics Association* 20(e2):e212–e220, 2013.
- [EJB] Understanding EJB Timer Services. https://docs.oracle.com/cd/E16439_01/doc.1013/e13981/undejdev012.htm#CCHIGFEC. (Accessed on 09/13/2021).
- [Hea] Healthlink. <https://www.healthlink.ie/healthlinkhome/>. (Accessed on 09/13/2021).
- [HL7] HL7-Definition V2. <https://hl7-definition.caristix.com/v2/HL7v2.5.1>. (Accessed on 09/13/2021).

- [KLT⁺18] R. A. Kyle, D. R. Larson, T. M. Therneau, A. Dispenzieri, S. Kumar, J. R. Cerhan, S. V. Rajkumar. Long-term follow-up of monoclonal gammopathy of undetermined significance. *New England journal of medicine* 378(3):241–249, 2018.
- [KM06] M. Karusseit, T. Margaria. Feature-based modelling of a complex, online-reconfigurable decision support service. *Electronic Notes in Theoretical Computer Science* 157(2):101–118, 2006.
- [MBK13] T. Margaria, S. Boßelmann, B. Kujath. Simple modeling of executable role-based workflows: An application in the healthcare domain. *Journal of Integrated Design and Process Science* 17(3):25–45, 2013.
- [MFG⁺14] T. Margaria, B. D. Floyd, R. Gonzalez Camargo, A.-L. Lamprecht, J. Neubauer, M. Seelaender. Simple management of high assurance data in long-lived interdisciplinary healthcare research: a proposal. In *International Symposium On Leveraging Applications of Formal Methods, Verification and Validation*. Pp. 526–544. 2014.
- [NLKS18] S. Naujokat, M. Lybecait, D. Kopetzki, B. Steffen. CINCO: a simplicity-driven approach to full generation of domain-specific graphical modeling tools. *International Journal on Software Tools for Technology Transfer* 20(3):327–354, 2018.
- [OLY⁺20] A. Y. Odisho, H. Lui, R. Yerramsetty, F. Bautista, N. Gleason, E. Martin, J. J. Young, M. Blum, A. B. Neinstein. Design and development of referrals automation, a SMART on FHIR solution to improve patient access to specialty care. *JAMIA open* 3(3):405–412, 2020.
- [RR19] O. B. R Clifford, A. Reche. UHL Guidance on Management of MGUS in Primary Care. 2019.
<https://healthservice.hse.ie/filelibrary/ulh/uhl-guidance-management-of-mgus-in-primary-care.pdf>
- [SCA⁺21] A. Shah, W. Clarida, R. Amelon, M. C. Hernaez-Ortega, A. Navea, J. Morales-Olivas, R. Dolz-Marco, F. Verbraak, P. P. Jorda, A. A. van der Heijden et al. Validation of automated screening for referable diabetic retinopathy with an autonomous diagnostic artificial intelligence system in a Spanish population. *Journal of diabetes science and technology* 15(3):655–663, 2021.
- [SGNM19] B. Steffen, F. Gossen, S. Naujokat, T. Margaria. Language-driven engineering: from general-purpose to purpose-specific languages. In *Computing and Software Science*. Pp. 311–344. Springer, 2019.
- [WIL] Wildfly documentation. <https://docs.wildfly.org/>. (Accessed on 09/13/2021).
- [WLM18] A. Wickert, A.-L. Lamprecht, T. Margaria. Domain-specific design of patient classification in cancer-related cachexia research. In *Proceedings of the 6th Conference on Formal Methods in Software Engineering*. Pp. 60–63. 2018.

- [Woo15] D. E. Wood. National Comprehensive Cancer Network (NCCN) clinical practice guidelines for lung cancer screening. *Thoracic surgery clinics* 25(2):185–197, 2015.