A Development and Integration Framework for Optimisation-Based Enterprise Solutions

Rodrigo Lankaites Pinheiro[1] and Dario Landa-Silva[1]

[1]*School of Computer Science, The University of Nottingham, Wollaton Road, Nottingham, NG8 1BB, United Kingdom*
{psxrp2, dario.landasilva}@nottingham.ac.uk

Keywords: optimisation development framework, decision support systems, academia-industry collaboration, workforce scheduling and routing

Abstract: The operations research literature includes some papers describing collaborative work between researchers and industry. However, not much literature exists that outlines methodologies to guide the development of a decision support module and its integration into an existing information management system. Here we describe a framework to aid the collaborative development of an optimisation solution by researchers and information system developers. The proposed framework also helps in the effective integration of the information management system and the decision support module. The framework is divided into three main components: a data model, a data extractor and validator, and a solution visualisation and auxiliary platform. We also describe our experience and positive results from applying the proposed development and integration framework to a project involving the development on an optimisation-based solution for workforce scheduling and optimisation problems. We hope that this contribution would be particularly useful for less experienced researchers and practitioners who embark on a collaborative development of a decision support module based on optimisation techniques.

1 INTRODUCTION

Decision support systems are pieces of software intended to aid humans in making decisions regarding specific problems in given domains. Over the past few decades, research and development of such systems have increased considerably (Power et al., 2011). Nonetheless, due to their complexity, decision support systems are mostly developed jointly by collaborative teams. Academics often enter into R&D projects in partnerships with practitioners to develop decision support modules – which may consist of one or more optimisation algorithms, simulation modules, inference engines, etc. – for solving a given problem (Fontana et al., 2006).

As with every software development environment, collaborative development of decision support systems presents its difficulties. The software engineering literature describes many methodologies for software development such as applying basic software paradigms (Pressman, 2010), or adhering to the Rational Unified Process (RUP) (Kruchten, 2004), Agile methodology (Martin, 2003) or Concurrent Engineering (CE) (Prasad, 1996). However, the general scope of these methodologies, often fails to address the unique challenges of R&D development.

The literature provides few works that propose general methodologies to aid the development of R&D projects (Barnes et al., 2002; Guarnaschelli et al., 2013). Some commercial tools exist to aid R&D projects development, such as the IBM ILOG ODM Enterprise (IBM, 2013). However, adhering to such tools implies adopting black-box algorithms (which may not be in the interest of researchers), limitations in technology (as compatibility issues arise) and often a considerable cost.

Here we present a data-centric development framework to assist in the collaborative development of optimisation-based enterprise solutions. The framework is a tool to facilitate the communication between development teams and to foster progress on the research of the optimisation problem being tackled. Moreover, it provides a layer between the information management system and the decision support module being developed. Hence, the practitioner does not need to force the use of specific technologies and methods and the researchers focus on the development of algorithms. We stress that the proposed framework is not intended to substitute software engineering methodologies and frameworks. Instead, it

is designed to provide both the researchers and practitioners with enough independence while still aiding them in the collaborative development. We also describe the application of the proposed framework to a real scenario involving the development of an optimisation-based enterprise solution.

The rest of this paper is structured as follows. Section 2 outlines the Workforce Scheduling and Routing Problems Project which is used to illustrate the application of the proposed framework. Section 2 also reviews previous related works. Section 3 presents the framework while section 4 presents the obtained results. Section 5 provides some insights and the conclusion of this work.

2 THE WSRP PROJECT AND RELATED WORKS

Part of the focus in our current research is the development of algorithmic solutions for workforce scheduling and routing problems such as home healthcare services. Workforce scheduling and routing problems (*WSRP*) are a class of problems in which workers (nurses, doctors, technicians, etc.) have to be allocated to tasks scattered in a geographical area, hence workers need to travel between locations. Workers have their own skills, availability, preferences, etc. while tasks have required skills, associated client preferences, etc. This class of problems combines features from scheduling problems and routing problems. For an overview of WSRP please refer to Castillo-Salazar et al. (2012).

We are engaged in a R&D project in collaboration with an industrial partner in order to develop the optimisation engine for tackling large WSRP scenarios. The existing information system collects all the problem-related data and provides an interface to assist human decision makers in the process of assigning workers to visits. We are in charge of developing the decision support module that couples well with the information management system being developed and maintained by the industrial partner.

The use of methodologies, frameworks and guidelines aid the process of elaborate software development and can help to reduce risks and problems. In Section 4) we discuss in detail some of the risks and problems that the proposed framework has helped us to address in our project. The literature describes several methodologies and guidelines to help on the development of both decision support systems (Connors, 1992; Bui and Lee, 1999; Elgarah et al., 2002) and information management systems (Prasad, 1996; Martin, 2003; Kruchten, 2004). However, only

few works refer specifically to R&D projects and in our opinion, even fewer could be applicable to optimisation-based problems. We highlight the following ones:

- Several works focus on the management aspects of R&D projects. They discuss the advantages of a collaborative project between researchers and practitioners, how to assess such partnerships and how to evaluate the benefits. Among them, we can list the works of Balachandra and Friar (1997), Huchzermeier and Loch (2001), Santiago and Bifano (2005) and Fontana et al. (2006).

- There are some case studies of R&D projects that aim to identify characteristics of the software development process in such R&D projects to then propose a few general guidelines. Among these works we can highlight those by Pinto and Covin (1989), and Pillai et al. (2002).

- Barnes et al. (2002) presents a general good practice model for R&D projects, but not a methodology or framework. Hence, we recommend to adhere to their model while applying our framework, as both methodologies can be fully integrated. Their partner evaluation factors and university-industry issues practises can be addressed prior to the application of our framework. Our data modelling contributes to fulfill several of their project management topics, such as the clearly defined objectives and responsibilities, realistic aims, collaborative agreement, progress monitoring and effective communication. Besides,our data model helps in managing the cultural gap issue and our proposed solution visualisation and auxiliary platform aids with the outcome factors they defined.

- Guarnaschelli et al. (2013) proposed a general methodology for integrating the development of decision models with model driven software development. They divide the methodology in four parts, namely the business modelling, the decision problem definition, the service modelling and the reference modelling. The framework proposed in this paper can be integrated with their methodology since the elements we propose match the elements proposed in their suggested timeline. Our proposed data modelling starts after their suggested business modelling and progresses through the decision model identification and specification. Our data extractor and validator contributes to their decision problem realisation and the domain elements reference modelling. Finally our solution visualisation and auxiliary platform aids with the implementation step.
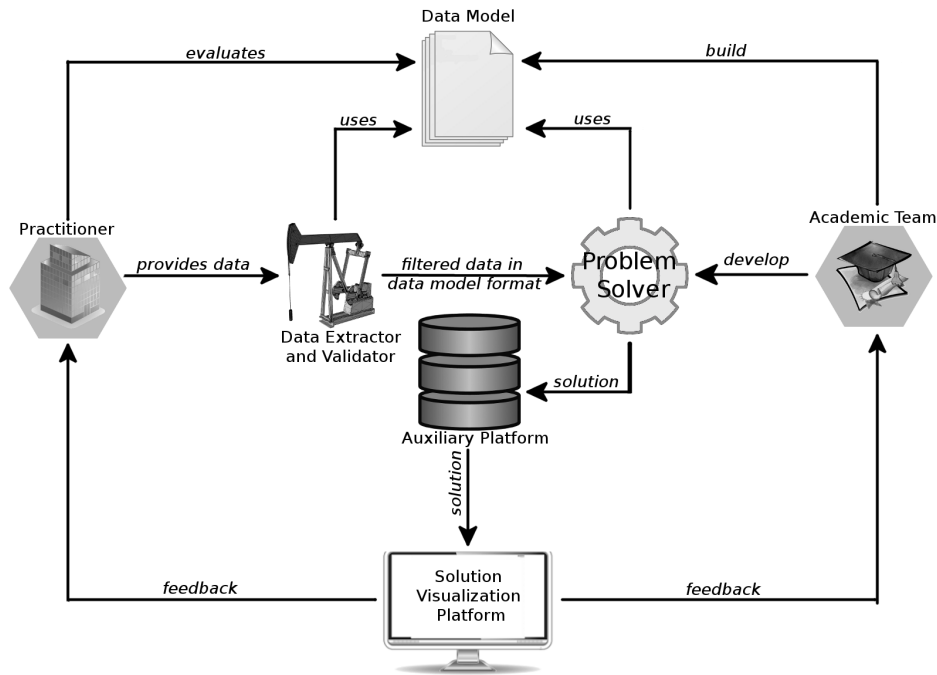
Figure 1: Overview of the development and integration framework.

3 THE PROPOSED FRAMEWORK

The framework consists of three components to help with the communication of teams, the understanding of the problem being tackled and the integration of the algorithmic solution. The first component is a data model understandable to both researchers and practitioners. The second component is a data extraction and validation mechanism. The third component is a solution visualisation and auxiliary platform. Figure 1 presents an overall diagram of the framework. Altogether, these elements create a layer between the practitioners and the researchers to assist them in achieving a common understanding of the problem and obtaining a higher communication level.

3.1 First Component: Data Model

The data model component applies data-centric development concepts (Bhattacharya et al., 2009; Cohn and Hull, 2009) to optimisation scenarios. The data model is a data abstraction in a file format capable of representing an instance of the optimisation problem. It serves as a layer between details of the problem domain and details of the algorithmic solution so that those building the model approach the problem from their own perspective. The purpose of the data

model building process is that both practitioners and researchers work in collaboration to create the file format and improve mutual understanding.

3.1.1 Modelling the Data

The data can be modelled into a set of files or a single file, depending on the characteristics of the optimisation problem. Ideally, the set of files contains all the information needed to run the solving algorithms. We recommend to have in each dataset all the information required to solve one problem instance instead of having data common to several instances in one file. This is because the time required for executing the optimisation algorithms is usually higher than the time required for data processing. More importantly, having all the data for a problem instance available in this way helps researchers to better understand the problem instance.

The WSRP scenario used to illustrate the framework in this paper contains a large dataset, hence we decided to split an instance into six files, each one containing a list of some sort. Figure 2 presents a simplified version of the dataset. Each file contains a list of the element that defines the file. Also, each file is linked to the mathematical model in order to show the correlation between the optimisation problem and the data files.

The use of a data model at the centre of the development process helps to improve the communi-
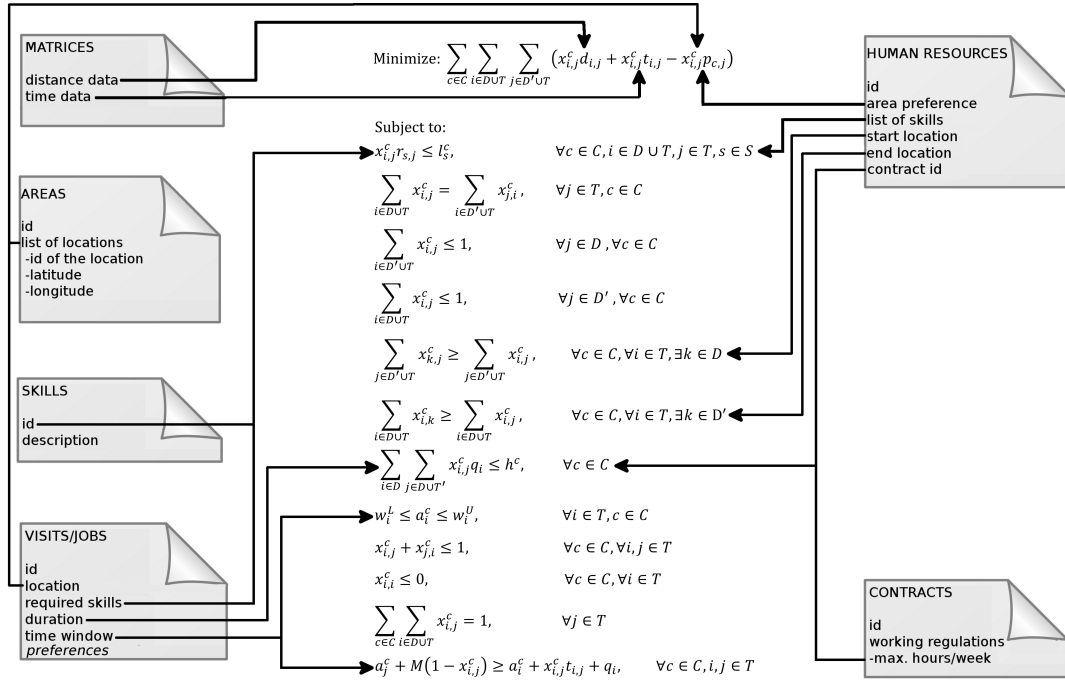
Figure 2: The files set for the WSRP project. Each file contains a list of the named elements and each element contains several fields related to the optimisation problem.

cation between researchers and practitioners (Cohn and Hull, 2009). The data files should preferably be human-readable and as self-explanatory as possible. This also allows researchers to easily alter the data in order to generate different test cases.

When developing the framework we opted to use eXtensible Markup Language (XML) (Harold and Means, 2009) as the file format for the model. We have chosen XML because it is easily readable, XML parsers are widely available, and XML offers a schema model. Such schema allows the creation of XML files that model other XML files and files attached to XML schemas are easily validated. Also, in the case of our WSRP project, both researchers and practitioners were already familiar with XML. However, an alternative is to use YAML (2013), which is clearer and more human-readable than XML. Besides, several YAML libraries are available for the most common programming languages.

Like in any other software development process, the development of a decision support system may be subject to many changes during the development. This must be considered when modelling the files, for example, by using required and optional fields. The required fields should be used for very basic data that must exist in order to validate the data for a problem instance. The optional fields are used to capture secondary constraints and objectives such as preferences.

In the modelling stage, the format used to represent a solution should help researchers to verify that the solution is valid, i.e. meets the problem constraints and additional requirements. This also helps practitioners and researchers to understand each other early in the process, on what constitutes a good-quality solution.

The final data model obtained is a layer between researchers and practitioners. This means that practitioners can work on the information management system using the data model to interact with the decision support module (even though it may not have been developed yet). Similarly, researchers can work on the algorithmic techniques without any specific knowledge of the information management system. Furthermore, this layer helps to ensure that any solution methodology developed by the academic team can be incorporated into the company's information system by only handling the corresponding model files. Therefore, the data model plays a crucial role when integrating the company's information system and the developed optimisation algorithms.

3.2 Second Component: Data Extraction and Validation

The sooner the academic team have access to data in the model format, the sooner research on the solv-

ing algorithms can begin. Then, sample datasets or a mechanism to efficiently assist in extracting the data would be useful. Figure 3 shows the schematic for the data extractor and validator developed as part of the framework. The figure also shows the interaction of this component with the database in order to generate the file set in the model format.

3.2.1 Data Extractor

The data extractor is the tool that retrieves the data from the information system and converts it to a set of files according to the data model. This data extractor plays an important role in the development of the decision support module. This is because a typical information system is likely to contain a large amount of data that is not needed for the optimisation problem and that should be filtered. Clearly the model itself helps to eliminate some of such data, but as expressed above, it is highly desirable that each dataset only contains the data required for a single problem instance. In our WSRP project, we only extract the visits/jobs for a given planning period. This helps to eliminate the processing of unnecessary data and improve the overall system's performance. Filtering the data at the retrieval stage is usually easier, more reliable and more efficient, specially in a SQL server environment. Figure 3 outlines this process of extracting and collecting raw data, sending it to the data validator, retrieving the validated data and finally generating the corresponding file set in the model format.

3.2.2 Data Validator

After the data extraction, data should be validated to account for missing and inaccurate entries in the database plus other considerations. In the database for our WSRP project, this means for example, filtering out staff members that are not available in the planning period considered, or those whose skills are not appropriate for the given scenario, or those only available to work on different geographical locations. We also filter out data about locations and skills not used in the given scenario. Hence, the first step of the data validator is to trim data as much as possible and ensure that data references broken as result of this data trimming are also fixed.

Moreover the validator also checks for inconsistent data, such as mistyped postcodes, which if not detected can negatively affect the performance of the solving algorithm. For example, the data validator for the WSRP project checks for wrong postcodes by comparing them to the postcodes of other locations in the same geographical area. It also checks for any
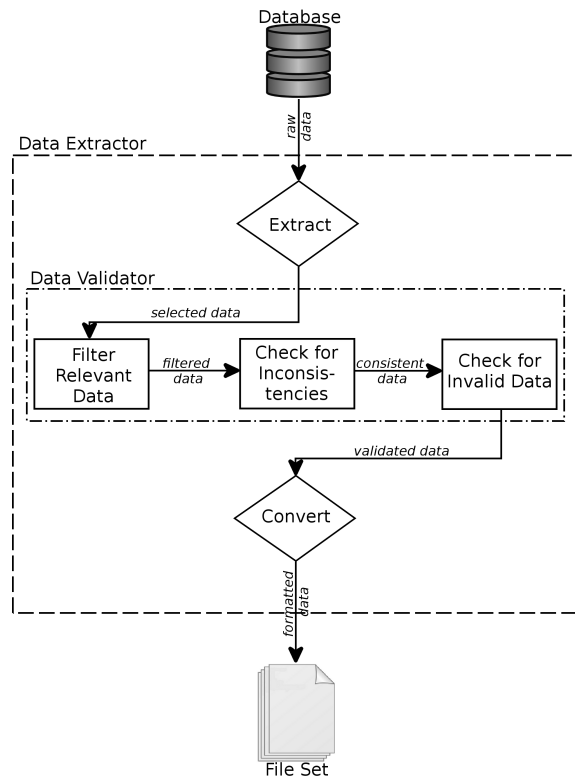

Figure 3: The data extractor and validator.

numerical value that deviates too much from the expected range of values. Figure 3 presents the proposed flow diagram for the data validator. The extractor and validator provides researchers with problem instances to assist the algorithm development, while also provides practitioners with a guiding mechanism to continue the data gathering and preparation.

3.3 Third Component: Solution Visualisation and Auxiliary Platform

The third component of the framework is a solution visualisation and auxiliary platform, a feature that is almost entirely designed to help researchers in the development. There are many information system development methodologies well consolidated in the software engineering literature. However, these are usally not fully applicable to the development of decision support systems, for which well established methodologies and frameworks are not easily identified in the literature.

This third component aims to be a communication layer between researchers and later between researchers and practitioners. Figure 4 presents a diagram to illustrate how the visualisation and auxil-
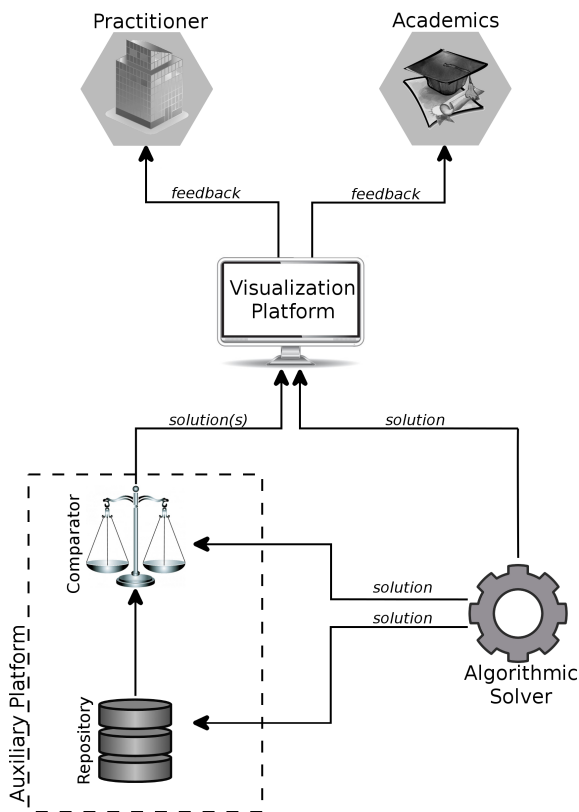
Figure 4: The visualisation and auxiliary platform.

iary platform interacts with the solver and the actors in the scenario, the company and the academic team. The solutions generated by the algorithmic solver can be stored into the repository, can be displayed in the visualisation platform and can be compared to other stored solutions. Through this visualisation platform, researchers and practitioners can have feedback from the solver and can evaluate the progress in the development of the solving algorithm.

3.3.1 Solution Visualiser

In an algorithm development team where each member works in a semi-independent manner with its own knowledge, tools and techniques, having a common solution visualiser can help to conduct a fair comparison between the different solution methods developed. Moreover, a mean to visualise solutions is a useful tool for algorithm developers that can help to save development time. Furthermore, having a common solution visualiser allows to present results obtained by different optimisation algorithms in a standardised format for practitioners to evaluate. Therefore, the visualiser is also a communication asset for the company and a tool by which researchers can present progress in their development.

Ideally the solution visualiser should be platform-free (able to run on any operating system), programming-free (not dependant of a specific programming language) and based just on the modelled format (not require extra files, configuration or parameters). To achieve these goals we can resort to some options which are a standalone application, a network-based application or a web-application. It is clear that a web-based environment presents many advantages like easy-access, good presentation and flexibility. The decision usually depends on the nature of the problem and preference of the development team. Furthermore, web-based and network applications are preferred since they make it easy to integrate with the auxiliary platform. In our scenario, we opted to develop a web-based visualisation platform.

3.3.2 Auxiliary Platform

The auxiliary platform is another tool specially conceived to aid on the research of the optimisation techniques. It consists of a repository of problem instances and solutions, plus an automated comparator of solutions. To compare solutions is fairly easy for single-objective problems in which the optimisation objective is a single scalar value. However, research on multi-objective approaches has expanded considerably and it is now common to apply such approach to real world problems, thus making the comparison between solutions more difficult. The specifics of the comparison are subject to the problem nature and to what the decision makers aim to achieve in a solution. The auxiliary platform should preferably provide charts showing strong and weak points of the solution as well as a detailed table containing useful information. In our WSRP project, the solution comparator is able to handle a pair of solutions or two sets of solutions. For the WSRP we display the distance travelled by the workers, the distance time, the travel cost, the payroll cost, and preferences. Also, an integration between the visualiser, the comparator and the repository is highly desirable since the usability of the tools escalates.

4 RESULTS

In this section we present the results obtained by applying the proposed framework to the WSRP project. We highlight the problems found during the development process and how applying the framework helped us to tackle these problems.

4.1 Better Collaboration Between Researchers and Practitioners

Communication Issues. Our experience showed that communication between the actors of the scenario can potentially be problematic. Information, requirements and expectations can be different from the perspective of researchers and practitioners, leading to misunderstandings and misconceptions. The data model building process played a major role in the researchers' understanding of the optimisation problem details. During the process we were able to identify gaps in our conceptualisation and fill them during a phase where changes caused minimal drawbacks. We were able to identify more constraints for the optimisation problem as well as alter some features to better suit the needs of the practitioner.

Knowledge Background Issues. There is obviously difference between the background knowledge of the company developers and that of the academic team. Moreover, researchers in the team may have different expertise including maths, management science, computer science, statistics, etc. This variety could lead to some difficulties such as different interpretation of problem data, unfamiliarity with certain technology and different assessment of results. With the data model we created a structured set of files in which researchers were able to easily browse the problem data without knowing any specific technology employed on the information system. Besides, the visualisation platform proved to be a valuable asset to give feedback to the company. Thus, the framework aided the academics on minimising the knowledge-gap difficulties.

Integration Issues. Integrating a decision support module to an information management system can be difficult and tricky. The first issue is the problem data retrieval. The second issue is applying the developed algorithms to an existing information management system. Integrating the solver can be difficult, specially if it uses a specific tool, such as mathematical solvers or specific software libraries. The data model provided a layer between our developed algorithms and the information system and the researchers found themselves in a comfortable environment where they employed different technologies from the ones used by the company.

4.2 Better Research and Development Process

Teamwork. In our example project, the academic group was formed by several PhD students, each one using the WSRP as a background for its PhD research.

The application of the framework helped considerably to improve communication among academics. The optimisation problem was discussed and a common understanding achieved. Besides, building the tools in the proposed framework provided a good opportunity for teamwork and knowledge sharing.

Rework. When developing algorithmic solutions for a given optimisation problem, researchers tend to create their own versions of tools such as file models, data retrieval API's, solution viewers and representations, etc. It is rarely the case that researchers reuse these types of tools developed by other researcher. Instead, they tend to reuse algorithms in their quest to develop improved versions of those. In our WSRP project, the data extractor and the solution visualisation and auxiliary platform were employed by all academics. Although the deployment of these framework tools required some time, reusing these tools saved considerable time and effort overall by eliminating rework. Along with the data model we also developed a Java API to handle the XML files and access the data. This API further extended the usefulness of the model by providing the users with an easy way to access the problem data.

Benchmarking. In real world problems, a good benchmark dataset to evaluate the solutions achieved during the research stage, is often difficult to obtain. Sometimes it is possible to estimate some close enough targets, but as the complexity of the problem grows, such approaches become less practical. Hence, it is necessary to have a fair and clear way to benchmark solutions to the real world scenarios, obtained with different algorithmic techniques. In our WSRP project, the auxiliary platform provided the team with a valuable asset to compare solutions more fairly and improve the overall benchmarking process.

5 CONCLUSION

Research and development projects involving universities and companies are common in decision support systems development scenarios. Development teams around the world usually find their own solutions to such problems, based on experience, trial and error. However, despite the frequency of this situation, to the best of our knowledge few specific methodologies can be found in the literature to help researchers and company alike to efficiently develop decision support systems in such circumstances. In particular, no previous publications seem to exist proposing a methodological approach to develop optimisation-based decision-support systems.

In this paper we have presented a methodological framework to aid on the development and integration of optimisation-based decision support modules and information management systems. The proposed framework combines a collection of good practices along with a data-centric development. The proposed development and integration framework helps to: 1) promote communication between the research team and the company, 2) provide an interface for the integration of the solving algorithms and the information management system and, 3) provide tools to develop and evaluate different algorithmic solutions. Moreover, the framework aims to bestow the academic team full independence to use whatever methods and technologies they choose and an easy manner to integrate the designed algorithms to the management information system.

We also described our experience on applying the framework to an ongoing project for an optimisation-based decision support system involving workforce scheduling and routing problems (WSRP). By adhering to the proposed framework, we were able to identify several features related to the optimisation problem prior to the implementation of algorithmic techniques. We also were able to save development time by avoiding work being redone since the extractor, validator and visualiser tools were available for the development team. Furthermore, we were able to improve communication and collaboration between researchers and practitioners and among researchers.

REFERENCES

Balachandra, R. and Friar, J. (1997). Factors for success in r amp;d projects and new product innovation: a contextual framework. *Engineering Management, IEEE Transactions on*, 44(3):276 – 287.

Barnes, T., Pashby, I., and Gibbons, A. (2002). Effective university industry interaction:: A multi-case evaluation of collaborative r&d projects. *European Management Journal*, 20:272 – 285.

Bhattacharya, K., Hull, R., and Su, J. (2009). A data-centric design methodology for business processes. In *Handbook of Research on Business Process Modeling, chapter 23*, pages 503 – 531.

Bui, T. and Lee, J. (1999). An agent-based framework for building decision support systems. *Decision Support Systems*, 25(3):225 – 237.

Castillo-Salazar, J. A., Landa-Silva, D., and Qu, R. (2012). A survey on workforce scheduling and routing problems. *Practice and Theory of Automated Timetabling*, pages 29 – 31.

Cohn, D. and Hull, R. (2009). Business artifacts: A data-centric approach to modelling business operations and processes. In *IEEE Data Engineering Bulletin 32*.

Connors, D. T. (1992). Software development methodologies and traditional and modern information systems. *SIGSOFT Softw. Eng. Notes*, 17(2):43 – 49.

Elgarah, W., Haynes, J., and Courtney, J. (2002). A dialectical methodology for decision support systems design. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*.

Fontana, R., Geuna, A., and Matt, M. (2006). Factors affecting universityindustry r&d projects: The importance of searching, screening and signalling. *Research Policy*, 35(2):309 – 323.

Guarnaschelli, A., Chiotti, O., and Salomone, H. E. (2013). Decision models as software artifacts - bridging the business-software gap for decision support systems. In *ICORES 2013 - Proceedings of the 2nd International Conference on Operations Research and Enterprise Systems, Barcelona, Spain, 16-18 February*, pages 86 – 95.

Harold, E. and Means, W. (2009). *XML in a Nutshell*. In a nutshell. O'Reilly Media.

Huchzermeier, A. and Loch, C. H. (2001). Project management under risk: Using the real options approach to evaluate flexibility in r&d. *Manage. Sci.*, 47:85–101.

IBM (2013). IBM ILOG ODM Enterprise. http://www-03.ibm.com/software/products/en/ibmilogodmente/.

Kruchten, P. (2004). *The Rational Unified Process: An Introduction*. The Addison-Wesley object technology series. Addison-Wesley.

Martin, R. (2003). *Agile Software Development: Principles, Patterns, and Practices*. Alan Apt Series. Prentice Hall/Pearson Education.

Pillai, A., Joshi, A., and Rao, K. (2002). Performance measurement of r&d projects in a multi-project, concurrent engineering environment. *International Journal of Project Management*, 20(2):165 – 177.

Pinto, J. K. and Covin, J. G. (1989). Critical factors in project implementation: a comparison of construction and r&d projects. *Technovation*, 9:49 – 62.

Power, D., Burstein, F., and Sharda, R. (2011). Reflections on the past and future of decision support systems: Perspective of eleven pioneers. In Schuff, D., Paradice, D., Burstein, F., Power, D. J., and Sharda, R., editors, *Decision Support*, volume 14 of *Annals of Information Systems*, pages 25–48. Springer New York.

Prasad, B. (1996). *Concurrent Engineering Fundamentals: Integrated product and process organization*. Prentice-Hall international series in industrial and systems engineering. Prentice Hall PTR.

Pressman, R. (2010). *Software Engineering: A Practitioner's Approach*. McGraw-Hill series in computer science. McGraw-Hill Higher Education.

Santiago, L. and Bifano, T. (2005). Management of r d projects under uncertainty: a multidimensional approach to managerial flexibility. *Engineering Management, IEEE Transactions on*, 52(2):269–280.

YAML (2013). YAML Aint Markup Language. http://www.yaml.org.