MSc Thesis in

Computer Engineering - Mobile Computing (MEI-CM)

Design and Implementation of an Enterprise Application Integration Solution for Science and Research Outcomes Information Management Using Guaraná Technology

Fernando José Rosa Sequeira

Leiria, September 2016

This page was intentionally left in blank.

MSc Thesis in

Computer Engineering - Mobile Computing (MEI-CM)

Design and Implementation of an Enterprise Application Integration Solution for Science and Research Outcomes Information Management Using Guaraná Technology

Fernando José Rosa Sequeira

Leiria, September 2016

This page was intentionally left in blank.

À Minha Família,
a base do que sou hoje, que esteve
sempre "lá".

This page was intentionally left in blank.

Agradecimentos

Começo por agradecer aos meus orientadores, professor Vitor Basto Fernandes e professor Rafael Zancan Frantz (o meu orientador do "outro lado do Atlântico"), pelo apoio e total disponibilidade que sempre demonstraram, paciência e motivação. Além do seu apoio em questões técnicas, devo-lhes a orientação que souberam dar a quem ela escasseia. Sem eles não teria sido possível concluir esta dissertação.

Um agradecimento também ao pessoal da i2Factory em Huelva, que me receberam durante uma semana para formação na ferramenta usada (Guaraná Cloud), e um agradecimento em particular aos membros da equipa de desenvolvimento com quem convivi mais, José Carlos Yerga, Borja Campina, Manuel Dominguez e Adrián Peña Castro.

This page was intentionally left in blank.

Resumo

Neste documento é apresentada uma proposta de Integração de Aplicações Empresariais para gestão da informação sobre produção científica e tecnológica. A proposta aborda a gestão de informação de produção científica e tecnológica nos contextos nacional, internacional e dos sistemas de informação correspondentes. Problemas inerentes à interoperabilidade entre sistemas de informação, assim como abordagens, tecnologias e ferramentas de integração são aqui apresentados(as) e aplicados(as). É apresentada uma perspetiva de negócio e tecnológica, incluindo a análise conceptual e a modelação, uma solução de integração baseada numa Linguagem de Domínio Específico e a plataforma de integração que executará a solução proposta. Para efeitos de ilustração e mais fácil compreensão da proposta e da sua implementação, o papel e as necessidades de um sistema de informação de suporte de uma unidade de investigação são assumidos como sendo o caso representativo ao longo do trabalho.

Palavras-chave: *Integração de Aplicações Empresariais, Domain-Specific Language, Gestão de Informação*

This page was intentionally left in blank.

Abstract

This document presents an Enterprise Application Integration based proposal for research outcomes and technological information management. The proposal addresses national and international science and research outcomes information management, and corresponding information systems. Information systems interoperability problems, approaches, technologies and integration tools are presented and applied to the research outcomes information management case. A business and technological perspective is provided, including the conceptual analysis and modelling, an integration solution based in a Domain-Specific Language (DSL) and the integration platform to execute the proposed solution. For illustrative purposes, the role and information system needs of a research unit is assumed as the representative case.

Key-Words: *Enterprise Application Integration, Domain-Specific Language, Information Management*

This page was intentionally left in blank.

List of Figures

This page was intentionally left in blank.

List of Tables

This page was intentionally left in blank.

Acronyms

A&I   *Abstracting and Indexing*

BPMS   *Business Process Management System*

CC0   *Creative Commons Zero*

CERN   *Conseil Européen pour la Recherche Nucléaire, nowadays European Organization for Nuclear Research*

CNPq   *Conselho Nacional de Desenvolvimento Científico e Tecnológico (until 1971, Conselho Nacional de Pesquisa)*

CRM   *Customer Relationship Manager*

DOI   *Digital Object Identifier*

DSL   *Domain-Specific Language*

EAI   *Enterprise Application Integration*

EID   *Electronic ID*

EIP   *Enterprise Integration Patterns*

EMBO   *European Molecular Biology Organization*

ERP   *Enterprise Resource Planning*

ESA   *European Space Agency*

ESB   *Enterprise Service Bus*

FCT   *Fundação para a Ciência e Tecnologia*

FCTSig   *Sistema de Informação da FCT*

GUI   *Graphical User Interface*

OECD   *Organization for Economic Co-operation and Development*

PDA   *Personal Digital Assistant*

PIM   *Personal Information Manager*

R&D   *Research and Development*

REST   *Representational State Transfer*

RPC   *Remote Procedure Call(ing)*

RPI   *Remote Procedure Invocation*

SCTN   *Sistema Científico e Tecnológico Nacional (Scientific and Technological National System)*

SOAP   *Simple Object Access Protocol*

Index

Introduction

Organizations rely on information systems and software applications to support their business activities (1). Interesting applications rarely live in isolation. Whether a sales application must interface with an inventory application, a procurement application must connect to an auction site, or a Personal Digital Assistant (PDA) Personal Information Manager (PIM) must synchronize with the corporate calendar server, it seems like any application can be made better by integrating it with other applications (2). Frequently, these applications are legacy systems, packages purchased from third parties, or developed internally to solve a particular problem. This usually results in heterogeneous software ecosystems, which are composed of applications that were not usually designed taking integration into account. Integration is necessary, chiefly because it allows to reuse two or more applications to support new business processes, or because the current business processes have to be optimized by interacting with other applications within the software ecosystem. Enterprise Application Integration (EAI) provides methodologies and tools to design and implement integration solutions. The goal of an EAI solution is to keep a number of applications' data in synchrony or to develop new functionality on top of them, so that applications do not have to be changed and are not disturbed by the integration solution (1).

1.1 Integration problems

Enterprises have typically hundreds of applications custom-built, acquired, part of a legacy system, or a combination, operating in multiple tiers of different operating systems and platforms. Some enterprises have dozens of Websites, more than one instance of SAP and countless departmental solutions (3). Creating a single, big application to run a complete business is next to impossible. Enterprise Resource Planning (ERP) have had some success at creating larger-than-ever business applications. The reality, though, is that even the heavyweights like SAP, Oracle, etc. only perform a fraction of the business functions required

in a typical enterprise. That can easily be seen by the fact that ERP systems are one of the most popular integration points in today's enterprises (3). Unfortunately, enterprise integration is no easy task. Software vendors offer EAI suites that provide cross-platform, cross-language integration as well as the ability to interface with many popular packaged business applications. However, this technical infrastructure presents only a small portion of the integration complexities. The true challenges of integration span far across business and technical issues (3).

Besides the Business point of view shown in the previous paragraph, it should also be taken in mind the Technology point of view. Problems with data representation, memory organization, data type representation, text coding (ASCII, Unicode, etc), binary (eventually proprietary) coding (databases, spreadsheets, text documents) should be expected to happen. Border (agnostic) formats, which are standards defining representations for common simple and structured data types (strings, arrays, records) have been developed for use in messages exchange and help on interoperability (CORBA CDR - Common Data Representation (4), JAVA serialization (5) , Web Services (6), etc.).

Lexical, syntactic and semantic mapping and transformation technologies were also developed for specific technology and business domains (EDI (7), XML (8), XSLT (9), ebXML (10), OWL (11), etc.).

Time and synchronisation, shared time references, cooperation and coordination semantics (message semantics, operations semantics, etc.) should also be taken in mind when dealing with information systems integration (3).

In a general manner, integration technologies from nowadays don't let to work at a high level of abstraction, for example, implementing solutions requires a knowledge of programming APIs. That is a limiting factor for the development and maintenance which turns the solution dependent on the integration platform. If solutions could be modelled in a platform independent language and the code needed for its implementation generated in an automatic manner, then it would be a cross-platform solution and the costs in implementation, maintenance and evolution would possibly be reduced.

2

1.2 Integration approaches

Nevertheless applications integration may occur in the same machine, usually it doesn't and sometimes machines are thousands kilometres from each other, and so almost all integration solutions have to deal with a few fundamental challenges, namely: networks are unreliable and slow. Integration solutions have to transport data from one computer to another across networks. Compared to a process running on a single computer, distributed computing has to be prepared to deal with a much larger set of possible problems. Often, two systems to be integrated are separated by continents and data between them have to travel through phone-lines, LAN segments, routers, switches, public networks, and satellite links. Each of these steps can cause delays or interruptions. Sending data across a network is multiple orders of magnitude slower than making a local method call. Designing a widely distributed solution the same way it would approach a single application could have disastrous performance implications. Any two applications are different. Integration solutions need to transmit information between systems that use different programming languages, operating platforms, and data formats. An integration solution needs to be able to interface with all these different technologies (2).

As stated by Gregor Hohpe and Bobby Woolf (2) application integration can be done in four different ways, namely: File Transfer, Shared Database, Remote Procedure Invocation, and Messaging. In File Transfer each application may produce files of shared data for others to consume, and may consume files that others have produced. The applications need to agree on the filename and location, the format of the file, the timing of when it will be written and read, and who will delete the file. Shared Database is the kind of integration where many applications share the same database schema, in a single physical database; since there is no duplicate storage, no data have to be transferred from one application to another. In Remote Procedure Invocation integration one application exposes some of its functionalities in such a way that other applications can access them as a remote procedure, thus the communication occurs in real-time and synchronously. Lastly, in Messaging integration each application connects to a common messaging system, publishing its own messages on the common message channel. Since an application can read messages from that common messaging system in a later time after they have been published by another application, the communication is asynchronous; applications only must agree on a channel and also on the message format.

Besides the previous classification, application integration might also be classified in another three different ways, namely: Point-to-point, Hub and Spoke, and Enterprise Service Bus (ESB). Point-to-point integration (or "Integration Spaghetti" as Gregor Hohpe did in "Hub and Spoke [or] Zen and the Art of Message Broker Maintenance" (12)), where one application connects directly into the application(s) on which there is something that is needed. The Hub and Spoke integration comes from a comparison to a bike wheal, where like on it, applications on the edge are connected by a spoke into a hub (12). Finally, on ESB integration applications are connected to a logical bus through smart connectors, which encapsulate system functionalities, providing an abstraction layer between bus and applications (13).

1.3 Problem to solve

The science and research outcomes information management process consists in the collection, structuring, processing and storage of information about researchers, publications, citations, projects, and other metadata about research activities and actors. In the Portuguese case, Foundation for Science and Technology (Fundação para a Ciência e Tecnologia - FCT) is the Portuguese national agency and authority for research promotion, funding, evaluation and national information outcomes management.

Although national research agencies and authorities are special observers of scientific and technical production at the national scale, other institutions also need this type of information for research planning, follow up, benchmarking, etc. Among these institutions are Higher Education Institutions (HEI), research institutes and other national and regional governmental agencies, and non-governmental industry and society actors.

Several initiatives have been developed to provide features and support for the needs of such information consumers. The most relevant initiatives worth to mention at the national level are the FCTSIG and DeGóis software platforms, representing the researchers national CV repository. While the former has a simple user interaction approach, the latter has highly structured data models and advanced features for researchers CV information management. Additionally, other initiatives at the national level took place targeting bibliometric data collection and science based indicator analysis. Having an exclusive bibliographic and

4

bibliometrics approach, these tools did not attract enough attention from the science and technology institutions, mainly due to their narrow scope for science and research outcomes analysis.

Several HEI have also developed internal systems for science and research outcomes management following their own data models, taxonomies and description syntaxes. National science and research institutions and corresponding information systems face nowadays the challenge of interoperating and exchange this type of information, in the scope of the science and technology national and international information ecosystem, for general and specific observation purposes. Among the international ecosystem components we can point out journals and conference publications repositories such as SCOPUS and Web of Science, international researchers information repositories such as ORCID, and several (less institutional) research oriented "social" networks such as Google Scholar, Research Gate and others. This global ecosystem is devoted to support research outcomes general information, lacks data harmonization and consistent identity management mechanisms, and raises severe difficulties for research outcomes analysis and evaluation at individual, institutional and national levels.

For the sake of simplicity and summarised description, without loss of generality, we assume in this work the role and perspective of a research unit actor. A research units needs, in a regular base (at least annually), to follow and assess its researchers activities and outcomes, whose CVs, activities, research outcomes are registered and updated in national funding agencies software platforms and international research production repositories.

In this document we present an EAI solution proposal for science and research outcomes information system integration, from the perspective of a research centre (or a similar research organization entity), responsible for pursuing, managing and reporting science and research activities (14).

1.4 Document structure

In the next chapter, Chapter 2, different integration approaches are introduced. File Transfer, Shared Database, Remote Procedure Invocation and Messaging are explained, and the reasons why the Messaging approach was selected for this work is presented.

Chapter 3 is focused on Messaging based integration technologies. Differences between Business Process Management System (BPMS) and Enterprise Application Integration (EAI) are highlighted, and a comparison study of messaging based integration technologies is presented. Guaraná is introduced here as the chosen technology to deal with the integration problem addressed in this dissertation.

Chapter 4 is devoted to the management aspects of information produced by science and research players such as researchers, research and development units, higher education institutions, national scientific and technological systems. This chapter also presents several platforms used to manage this type of information.

Chapter 5 is focused in Guaraná (an integration platform, just like Apache Camel or Spring Integration) and its Domain-Specific Language (DSL), development environment, etc. and corresponding use to build an integration solution for the integration problem presented in previous chapters.

Finally, in Chapter 6, conclusions are drawn and proposals for future work are described.

Approaches to Integration

Firstly, for enterprise integration, a significant shift in corporation technology policies might be required. Usually business applications are focused on a specific traditional area, like Customer Relationship Manager (CRM), Billing, Finance, etc. and many IT groups are organized in alignment with these functional areas. Therefore, for a well succeeded integration, besides establishing communication between multiple computer systems there should also be established communications between business units and IT departments (it should be taken in mind that "from now on" no group has the control of a specific application because each application is now part of an overall flow of integrated applications and services) (2).

Then, because an integration has a wide scope, the efforts for doing it typically have far-reaching implications on the business: a failure or a misbehaviour in an integration solution on which the most critical functions of the business have already been integrated may have high cost and impact in a company business (2).

Another very important constraint when developing integration solutions is the control that integration developers may have (or may not have!) over the applications to integrate. Most of the times applications to integrate are "legacy" systems or packaged applications that cannot be changed for being connected to an integration solution (2).

Despite the wide-spread need for integration solutions, only few standards have established themselves in this domain. The advent of XML, XSL and Web Services have marked the most significant advance of standards-based features in an integration solution. However, the hype around Web Services has also given grounds to new fragmentation of the marketplace, resulting in a flurry of new "extensions" and "interpretations" of the standards (2).

Besides this, the existence of XML Web Services standards address only a part of the integration challenges. The existence of a common representation - like XML - does not imply common semantics: for example, "account" may have many different semantics, connotations, constraints and assumptions in each participating system. Since business and technical decisions are involved in the resolution of semantic differences between systems, this process has been proven to be a very time-consuming and difficult task (2) (3).

All this show us that there are a lot of issues to consider when selecting an integration approach. There isn't also a "magic formula" or a "recipe" to follow with assured results.

We may easily concluded that there's more than one possible approach for applications integration and each approach addresses some of the integration criteria better than others. The various approaches might be summed up in four main integration styles, namely: File Transfer, Shared Database, Remote Procedure Invocation, and Messaging. In the File Transfer, applications may produce, share and consume files others produced. In Shared Database, applications store and query shared data in a common database. In Remote Procedure Invocation, each application exposes some of its procedures in order for being remotely invoked by another applications and exchange data or run some actions. In Messaging, each application connects to a common messaging system, using messages to exchange data and invoke business functions (2).

2.1 File Transfer

File Transfer integration, shown in Figure 1, might be used in organizations that have many applications independently built, running in different languages and platforms. An application plays the role of "Producer", exporting data to files that other applications, playing the role of "Consumers" will then read. A very important decision to take is on which format should be used. Since most of the times the output from an application is not understood by the other, integrators must go through a file processing task along it. Lately, XML format is being widely used for these issues, with the support of an industry of readers, writers and transformation tools. Besides the format for this data sharing, it should also be taken in mind how often data are updated, i.e., when must data be written and when must data be read (2).
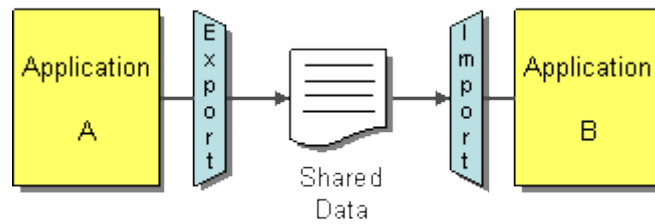
Figure 1 - Integration by File Transfer (2)

2.2 Shared Database

Shared Database integration, shown in Figure 2, might be used in organizations that have many applications independently built, running in different languages and platforms. A major difference from File Sharing integration, is that in this scenario information must be shared rapidly and consistently. As the name suggests, data from many applications are stored in the same database. Data consistency is ensured by database transaction management systems (2).
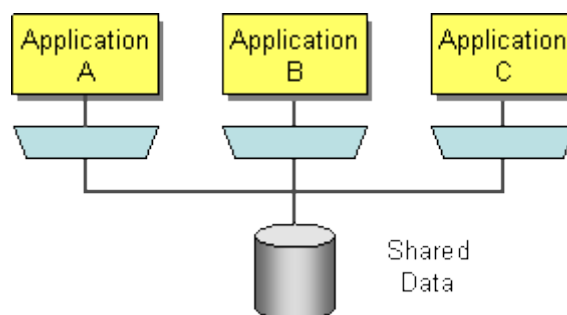
Figure 2 - Integration by Shared Database (2)

2.3 Remote Procedure Calling

Remote Procedure Calling (RPC) or Remote Procedure Invocation (RPI) integration, shown in Figure 3, might be used in organizations where besides the need to share data among applications independently built, running in different languages and/or platforms, it is also needed that functionalities may be shared in a responsive way. This may be achieved developing each application as a large-scale object or component with encapsulated data. For doing so, each application must provide an interface to allow other applications to interact with using this interface. If one application needs to read or even modify data of another

9

application, that is done by making a call to the other application. In this scenario each application maintains the integrity of the data it owns. As a consequence, each application can change its internal data, or the way they are stored, without having every other applications affected. A disadvantage on this methodology is that being this a synchronous system where applications are directly connected into each other, there is the risk to an application become overloaded and slow down the whole system. Also network issues may slow down or cause fails in part of the system that may affect the whole rest (2) (15).
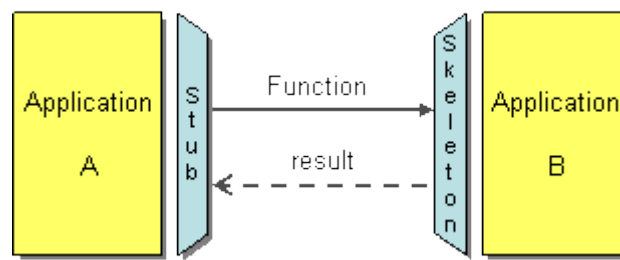
Figure 3 - Integration by Remote Procedure Calling (2)

2.4 Messaging

Messaging Integration, shown in Figure 4, just like RPC might be used in organizations that have many applications independently built, running in different languages and platforms where data and functionalities need to be shared in a responsive way. However, unlike RPC integration, Messaging integration might be asynchronous. As a reaction to the common problems present in distributed systems (unavailability of systems, problems with network connections) Messaging systems enable the transfer of data packets in a frequently, reliable, immediate way but also asynchronously, using customizable formats. Having a system responsible for taking and delivering messages from one application to another one (or more), allows interoperability in situations that not all the systems are up and running at the same time. However, it may occur that a sequence of messages may not be received by the same order that was sent (sometimes because a message fails or it took longer than another to be created) and the Message Bus will have to resend it again (2) (15).
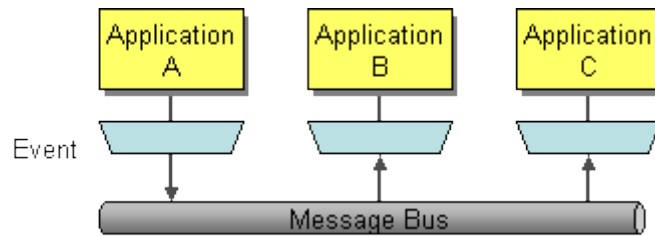
Figure 4 - Integration by Messaging (2)

2.5 Chosen approach

In order to select an integration approach to solve the integration problem addressed in this thesis, the following properties of the targeting scenario were considered:

- systems to integrate aren't all from the same institution;

- there is no connection between them;

- some systems provide data "as is", without any chance to request other data formats;

- outputs from one system have to be "worked on" before being used as inputs on other systems;

- some systems might suffer from temporary interruptions.

According to the advantages and disadvantages of the integration approaches presented in previous sections, and the properties of the integration scenario, the messaging integration approach was considered the most suited and promising for an integration solution with improved quality attributes such as reliability, scalability and availability (16).

This page was intentionally left in blank.

Integration technologies based in messaging systems

In the last years many proposals and tools have emerged to support the design and implementation of integration solutions. Hohpe and Woolf (2) documented several integration patterns found in the domain of Enterprise Application Integration (14).

Apache Camel, Spring Integration and Mule are open-source tools provided by Apache Software Foundation, VMWare Inc. and MuleSoft Inc., respectively, to support the design and implementation of Enterprise Application Integration solutions building on integration patterns. The most relevant concepts proposed in the architectures of these tools, inspired by the concept of enterprise service bus and Hohpe and Woolf integration patterns, are exchange or message, message channel, flow, endpoint, processor, and route. Mule, Camel and Spring Integration define slight variations and semantics for these patterns and combine them in specific ways to build integration solutions. The implementations, in all the mentioned tools, are carried out by means of a command-query API, and in the cases of Spring Integration and Mule can be assisted by an Eclipse-based IDE with a graphical editor (1).

Different from the previous technologies, Guaraná got inspiration from the Model-Driven Engineering discipline (17), shifting the focus from the source code to models. Models are abstractions that allow software engineers to focus on the relevant aspects of a software system while ignoring details that are irrelevant. Behind this discipline is the idea to raise the level of abstraction of the overall development process, to capture systems as a collection of reusable models, to separate business logic descriptions from a particular platform implementation, and to automate the implementation phase (14).

Additionally Guaraná provides both an ESB API assisted by an Eclipse-based IDE and a fully cloud based web graphical IDE. Guaraná Cloud was used in this thesis due to its

unique characteristics on what concerns to integration with a cloud infrastructure, graphical editor, and web based development environment.

In the next sections it will be presented the most relevant concepts proposed in the Guaraná platform.

3.1 Architecture of Guaraná

Guaraná DSL is a Domain-Specific Language, including an intuitive graphical notation, to design EAI solutions at a high-level of abstraction. In Guaraná, a message is an abstraction of a piece of information that is exchanged and transformed across an integration solution. It is composed of a header, a body, and one or more attachments. The header includes custom properties and frequently the following pre-defined properties: message identifier, correlation identifier, sequence size, sequence number, return address, expiration date, and message priority. The body holds the payload data, whose type is defined by the template parameter in the message class. Attachments allow messages to carry extra pieces of data associated with the payload, e.g., an image or an e-mail message (14). Figure 5 illustrates the graphical concrete syntax of the main constructors in Guaraná.
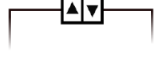
Notation	Concept	Notation	Concept
	Resource		Solicitor Port
	Integration Process		Responder Port
	Entry Port		Task
	Exit Port		Slot

Figure 5 - Guaraná Constructors (14)

Task: Represents an atomic operation that can be executed on messages, such as split, aggregate, translate, chop, filter, correlate, merge, resequence, replicate, dispatch, enrich, slim, promote, demote, and delay. Roughly speaking, a task may have one or more inputs from which it receives messages, and one or more outputs by means of which messages

14

depart. Depending on the kind of operation, a task may be stateless or stateful. In a stateless task, the completion of its operation does not depend on previous or future messages; contrarily, the operation of a stateful task depends on previous or future messages to be completed, such as the case of the aggregator task, which has to collect the different correlated inbound messages to produce a single outbound message. The vast majority of tasks in Guaraná technology are stateless (14).

Slot: A buffer connecting an output of one task to the input of another task aiming at messages to be processed asynchronously by tasks. A slot can follow different policies to serve messages to tasks, such as a priority-based output or a first-come, first-served. If a priority is defined in the message, slots follow the former policy; otherwise, the latter policy is adopted (14).

Port: Abstracts away from the details required to interact with resources within the software ecosystem. Roughly speaking, by means of a port it is possible to establish read, write, solicit, and respond communication operations with the resources being integrated (14).

Integration Process: Contains integration logic that executes transformation, routing, modification, and time-related operations over messages. An integration process is composed of ports that allow it to communicate with the resources being integrated, slots and a set of tasks to specify the integration logic (14).

Guaraná SDK provides a runtime environment for integration solutions in a multiplatform and multithreading technology developed in Java technology-based layers (framework, toolkits, adapters, tasks). Figure 6 presents a general overview of Guaraná Architecture.
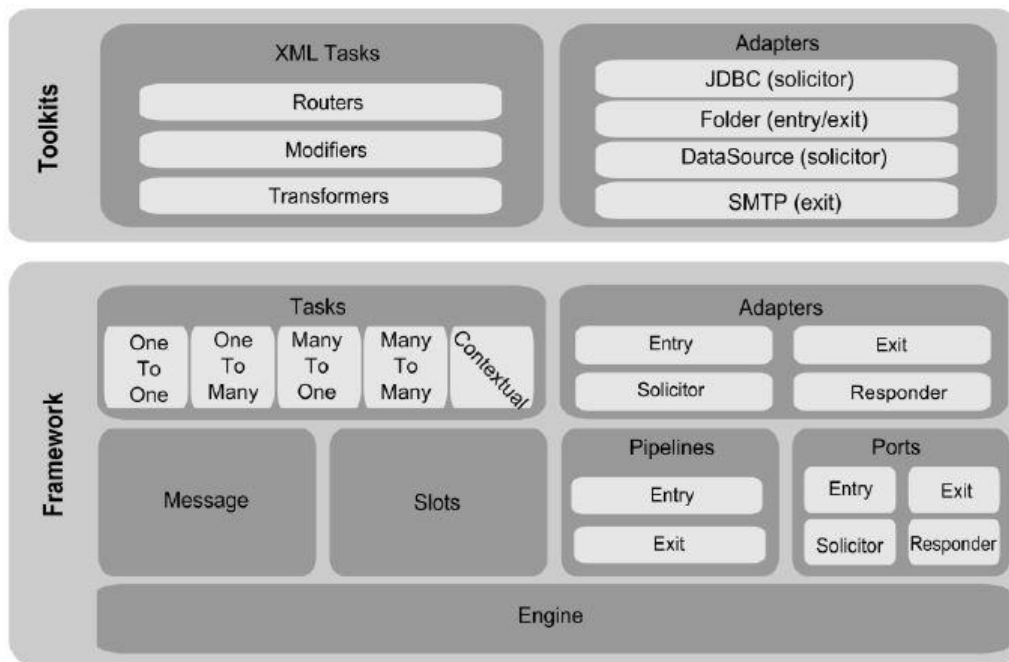
Figure 6 - Architecture of Guaraná Technology (18)

Conceptually, an integration solution aggregates one or more integration processes through which messages flow and are processed asynchronously. The integration flow is actually implemented as a Pipe and Filter architecture, in which the pipes are implemented by Slots and the filters are implemented by Tasks (14). Every task realizes an integration pattern, (2) and its execution depends on the availability of messages in all slots connected to its inputs. Slots are key constructors to enable asynchrony in an integration solution, thus messages are stored on them until they can be read by the next task in the integration flow. Messages do not appear in Figure 5, because they are not part of the conceptual model, they only exist and flow in the constructed and running integration solution (14).

3.2 Guaraná DSL

Guaraná DSL is a DSL to design EAI solutions at a high-level of abstraction. Guaraná's tasks are based on the Enterprise Integration Patterns (EIP) by Gregor Hohpe and Bobby Woolf (19).

Guaraná DSL tasks are grouped in the following major groups, according to their properties: Routers, Modifiers, Transformers, Stream Dealers, Mappers and Communicators. A Router type task is a task that does not change the messages it processes, but routes them

through a process; a Modifier type task is a task that adds or removes data to or from a message while preserving the message schema; a Transformer type task is a task that translates one or more messages into a new message with a different schema; a Stream Dealer type task is a task who deals with streams of bytes, used to zip or unzip, encrypt or decrypt, encode or decode it; a Mapper type task is a task used for changing the representation of the messages it processes (for example, turning a bytes stream into a XML document); a Communicator type task is a task used to encapsulate an adapter, serving two purposes: one, to allow adapters being exported to a registry so they can be accessed remotely and secondly, communicators might be setup to poll periodically a process or application using an adapter. Some actions may require a set of tasks and are done in a repeated manner along a solution. For example, whenever there is a query to an outer application or resource, it involves one Replicator, two Translator, one Correlator and one Enricher. This way, for not keep repeating this set of tasks, before a Solicitor Port to outer application or resource, this set of tasks can be abstracted using a composite task named "Enquirer". Thus, while helping for the clearness in the design of solutions, when the time for implementation is come it will be replaced for the corresponding set of tasks (1).

Next there will be shown the many tasks from Guaraná, grouped by type. Table 1 shows the notation and a description for Router tasks .

Notation	Task	Description
	Dispatcher	Dispatches a message to exactly one slot.
	Distributor	Distributes messages to one or more slots.
	Replicator	Replicates a message to all of the output slots.
	Correlator	Analyses inbound messages and outputs sets of correlated ones.
	Merger	Merges messages from different input slots into one output slot.
	Resequencer	Reorders messages into sequences with a pre-established order.
	IdempotentTransfer	Removes duplicated messages.
	Filter	Filters out unwanted messages.
	SemanticValidator	Validates the semantics of a message.
	CustomRouter	Allows for routing a message according to custom semantics.

Table 1 - Router tasks (1)

Table 2 shows the notation and description for Modifier tasks .

Notation	Task	Description
	Slimmer	Removes contents from the body of a message according to a static policy.
	ContextBasedSlimmer	Removes contents from the body of a message according to a dynamic policy that is provided by a context message.
	ContentEnricher	Adds static contents to the body of a message.
	ContextBasedContentEnricher	Adds dynamic contents from a context message to the body of a message.
	HeaderEnricher	Adds static contents to the header of a base message.
	ContextBasedHeaderEnricher	Adds dynamic contents from a context message to the header of a base message.
	HeaderPromoter	Promotes a part of the body of a message to its header.
	HeaderDemoter	Demotes a part of the header of a message to its body.
	CustomModifier	Allows to modify the header and body of a message according to custom semantics.

Table 2 - Modifier tasks (1)

Table 3 shows the notation and description for Transformer tasks .

Notation	Task	Description
■→▥	Translator	Transforms the body of a message from one schema into another.
■→▥▤	Splitter	Splits a message that contains repeating elements into several messages.
▤▥→■	Aggregator	Constructs a new message from several messages produced previously by a Splitter.
■→▥	Chopper	Breaks a message into two or more messages.
▥→■	Assembler	Constructs a new message from two or more messages.
▥→■	CrossBuilder	Constructs a new message that contains the Cartesian product of all inbound messages.
■→★▥	CustomTransformer	Allows for transformation of a message according to custom semantics.

Table 3 - Transformer tasks (1)

Table 4 shows the notation and description for Stream dealer tasks .

Notation	Task	Description
◇→◇	Zipper	Compresses a message.
◇→◇	Unzipper	Decompresses a message.
◇🔒◆	Encrypter	Encrypts a message.
◆🔓◇	Decrypter	Decrypts a message.
◇→◈	Encoder	Encodes a message.
◈→◇	Decoder	Decodes a message.

Table 4 - Stream dealer tasks (1)

Table 5 shows the notation and description for Mapper tasks .

Notation	Task	Description
■→◇	XML2Stream	Maps a XML message onto a stream of bytes.
◇→■	Stream2XML	Maps a stream of bytes onto a XML message.

Table 5 - Mapper tasks (1)

Finally, Table 6 shows the notation and description for Communicator tasks .

Notation	Task	Description
→◇	In Communicator	Used in ports to read messages.
◇→	Out Communicator	Used in ports to write messages.
→◇	In-Out Communicator	Used in responder ports to receive request messages and send reply messages.
→◇	Out-In Communicator	Used in solicitor ports to send request messages and receive response messages.

Table 6 - Communicator tasks (16)

Most relevant tasks introduced above and enumerated in Tables 1-6 are available in "Guaraná Cloud" platform, which is introduced in the next section.

3.3 Guaraná Cloud

Guaraná arises from the efforts of University of Seville researchers focused on the study of specific tools and notations to reduce design times and implementation of solutions in the field of integration computer systems and business information.

The aim of Guaraná Cloud is to provide software engineers the optimum technology to integrate traditional business resources (local applications, legacy systems, databases, files, web services, etc.), Internet applications (Software as a Service or SaaS) or Cloud platforms (Platform as a Service or PaaS) (18) (20).

By accessing the URL http://guarana.solutions or the newest http://platform.i2factory.com , users are re-directed to https://platform.i2factory.com/login. After logging in (a login account must be provided by i2factory) users get into Guaraná Cloud Dashboard. As shown in Figure 7 and in Figure 8, the Dashboard or Control Panel consists in a main menu on the left (sometimes this menu hides itself, only showing the icons). Most of the area changes according to the context selection on the main menu. On the bottom it can be found a task bar which also changes according to the context. Depending on the user type ("User", "Super User" or "Administrator"), users have access to different features.
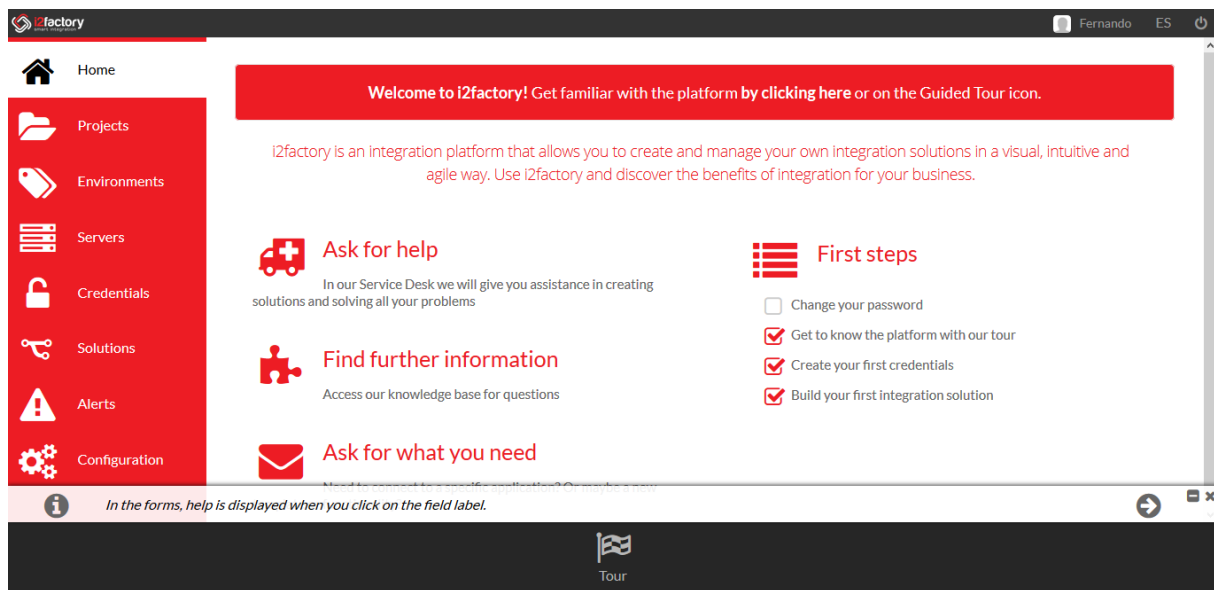
Figure 7 - Guaraná Cloud Platform Dashboard for "User" user type
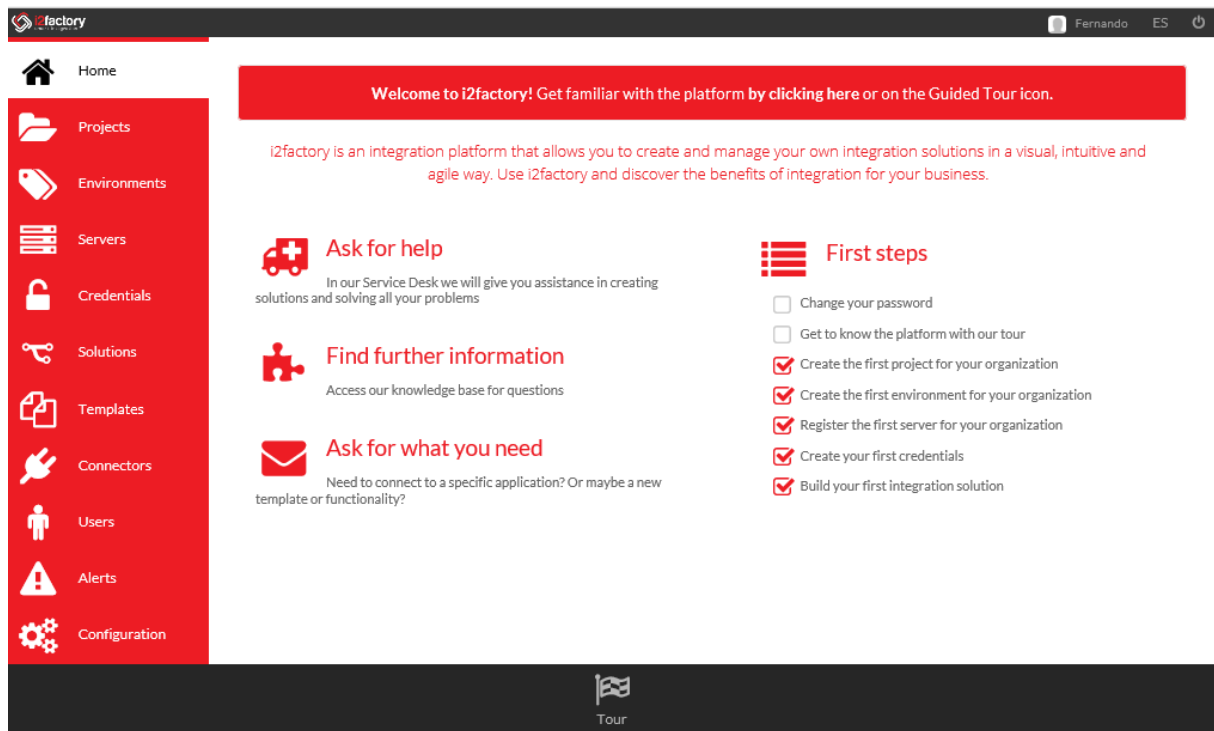
Figure 8 - Guaraná Cloud Platform Dashboard for "Super User" and "Administrator" user types

The set of common features or controls for all user types are Projects, Environments, Servers, Credentials and Solutions. In Projects users can create or edit existing projects. Projects are the main organizational unit, used to organize the remaining elements of a solution; Environments, Servers and Solutions must be created in a Project. Before creating anything else, a Project must be created (20) (21). Figure 9 shows the Graphical User Interface (GUI) when users select Projects.
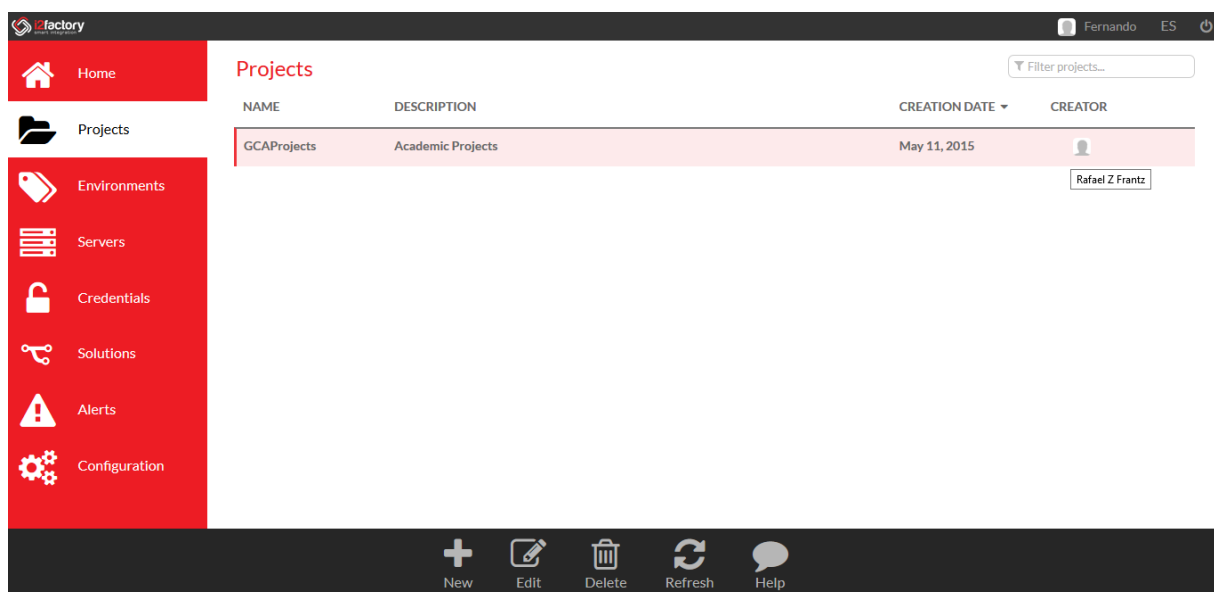
Figure 9 - "Projects" in Guaraná Cloud Dashboard

Environments are used to group and organize Servers, Credentials and Solutions. Before create a Server or a Solution, an Environment must be created (20) (21). Figure 10 shows the GUI when users select Environments.
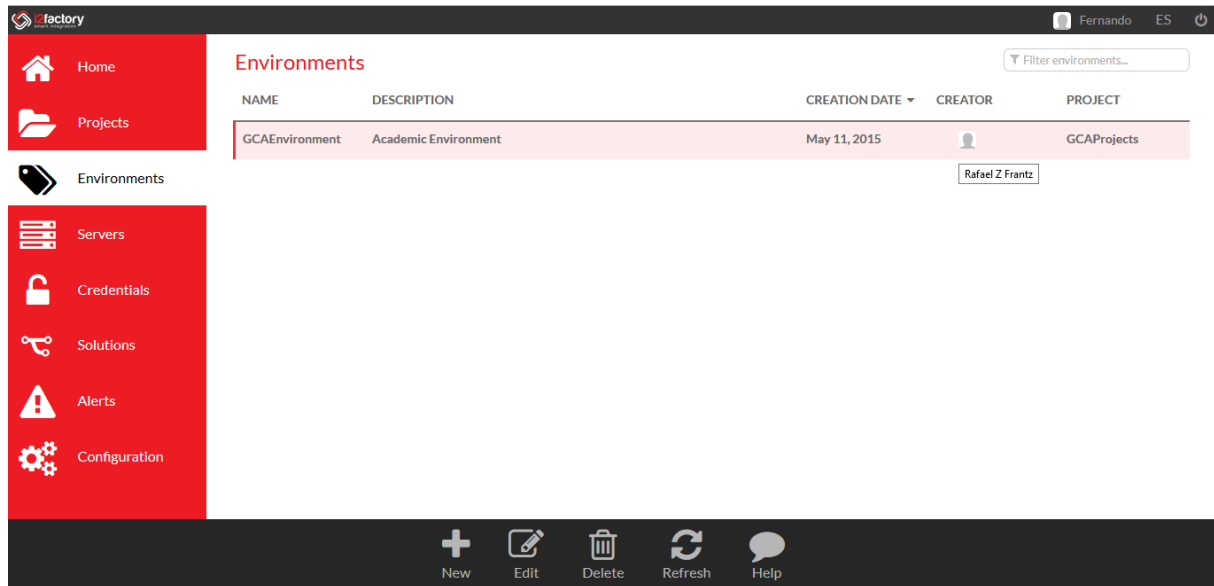
Figure 10 - "Environments" in Guaraná Cloud Dashboard

Server is the infrastructure on which Solutions deployed on Guaraná Cloud platform are hosted and ran. It needs a Project and an Environment to be created. A Server can be hosted in Guaraná Cloud or in a local infrastructure running an instance of Guaraná Cloud Server. While creating a Server, user may be allowed to select whether Debug is available or not, as well as pre-defined hardware settings, geographic location, a Project and an Environment (20) (21). Figure 11 shows the GUI when users select Servers.
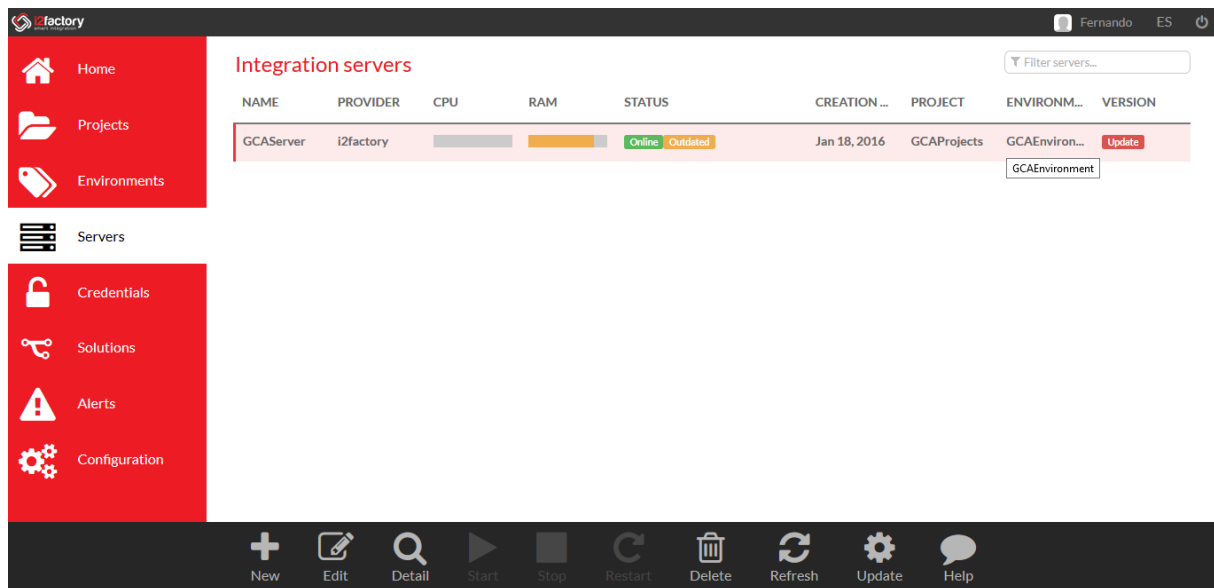
Figure 11 - "Servers" in Guaraná Cloud Dashboard

Figure 12 presents the detail view of a Server. This way it is possible to monitor the status and activity of the server where a solution has been deployed.
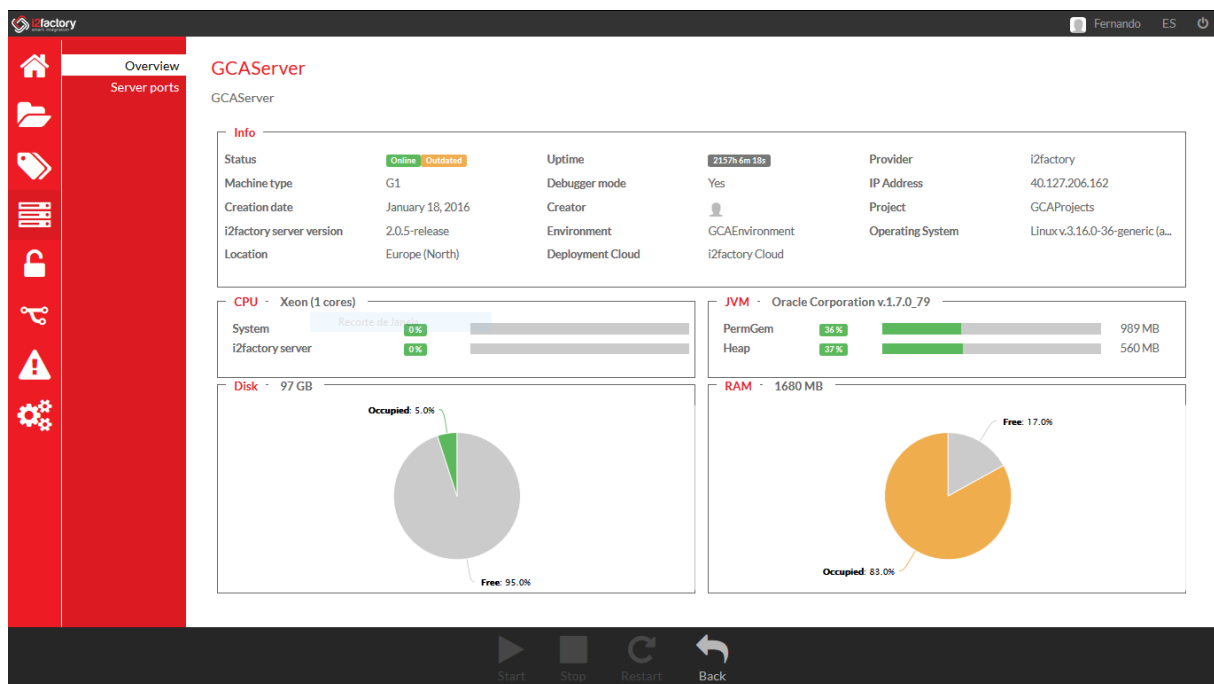
Figure 12 - "Server" detail in Guaraná Cloud Dashboard

Credentials allow users to store authentication information for a given connector, for Cloud and Internet services. Being associated with a project and an environment, it can be re-used in as many solutions as needed as well to create multiple credentials for the same connector, in this case with different values. Guaraná Cloud provides quite a few Credentials,

amongst them Credentials to use with Gmail, Dropbox, Twitter (21). Figure 13 and Figure 14 shows the GUI when a Credential is selected.
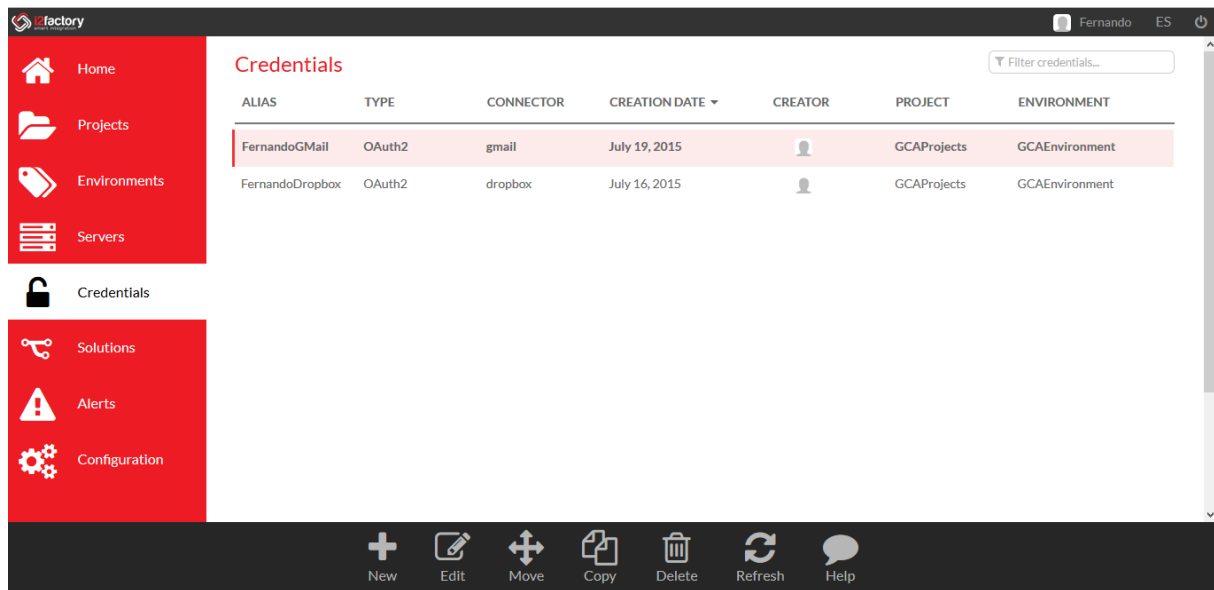
Figure 13 - "Credentials" detail in Guaraná Cloud Dashboard

Figure 14 - List of available Credentials in Guaraná Cloud

Solutions provides features for integration solutions management (creation, debugging, monitoring, deleting). Here the user access Guaraná Cloud Studio, where he can create new integration solutions from existing templates or starting from scratch. Even before entering Guaraná Cloud Studio there are useful information about the design status (if the integration solution design is valid or not; if the solution design has changed since the solution has been deployed; if the solution is running or not, or if is not even deployed in the server, etc.) (20) (21). After selecting the creation of a new solution or edit an existing one, users will have access to the Guaraná Cloud, a graphical IDE for solutions design. Figure 15 shows created solutions and corresponding context dependent information/options.
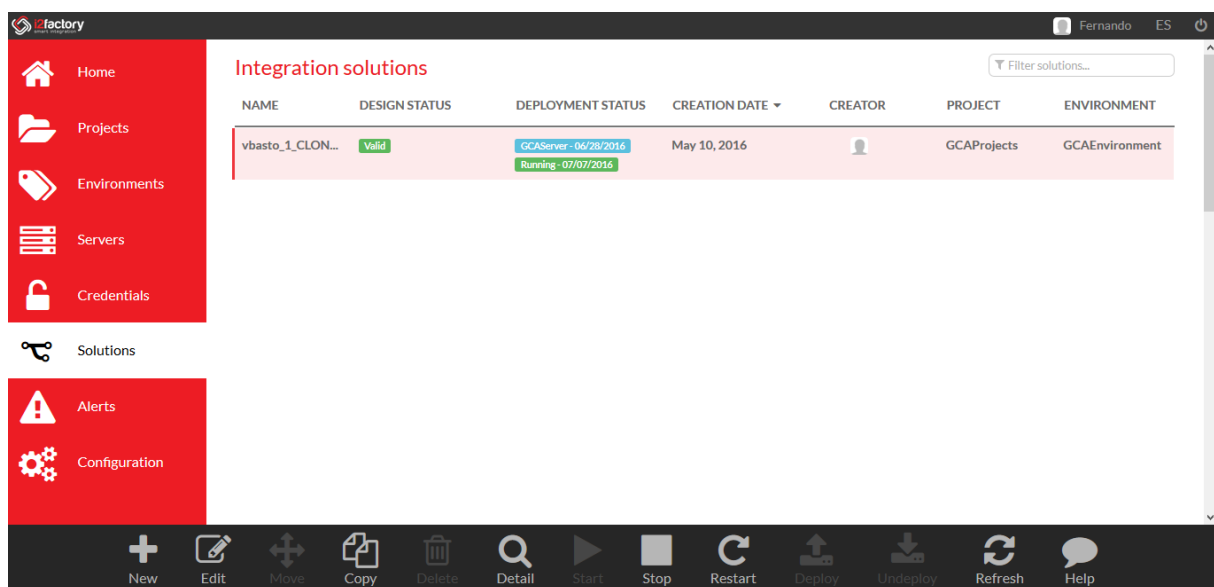
Figure 15 - "Solutions" detail in Guaraná Cloud Dashboard

As shown in Figure 16, Alerts section notifies users about errors or problems occurring while running the integration solution (20) (21).

Figure 16 - "Alerts" detail in Guaraná Cloud Dashboard

Configuration (Figure 17) allows for user information insertion/update.

Figure 17 - "Configuration" detail in Guaraná Cloud Dashboard

Templates (Figure 18) is only available to "Super User" and "Administrator" users type. Templates might be provided by Guaraná Cloud, other users or from solutions created by the user. Templates assure that solutions are well design, are easy to use, increase overall quality and productivity in solutions design.
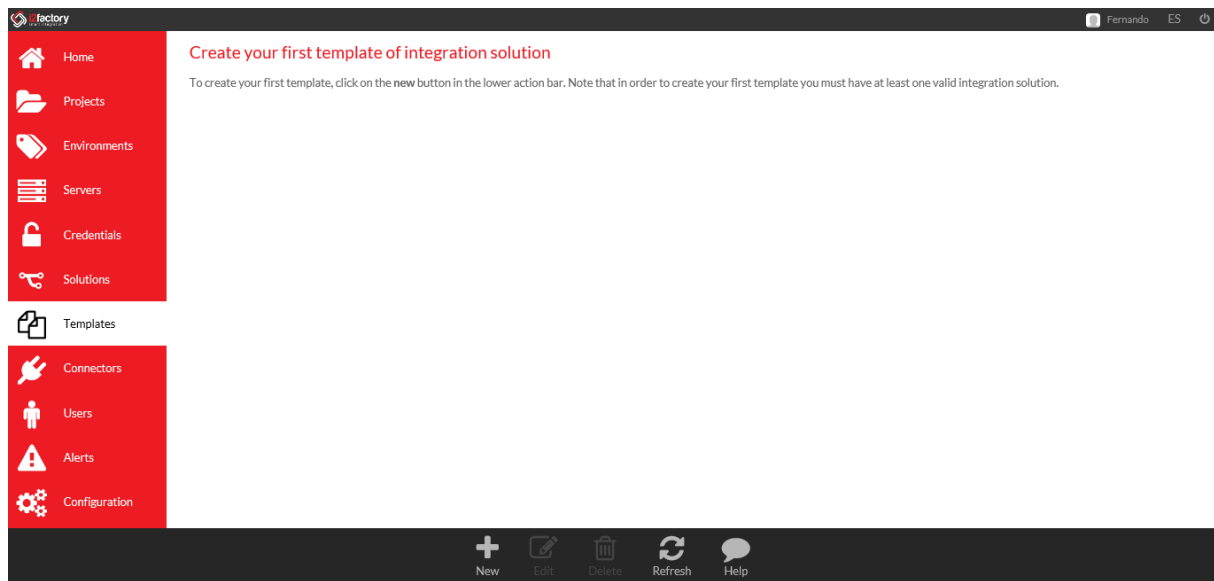
Figure 18 - "Templates" detail in Guaraná Cloud Dashboard

Connectors (Figure 19) is only available to "Super User" and "Administrator" users type. A set of commonly used connectors is available by default in the development environment, but new additional connectors may be purchased, developed, or removed from this set (21).
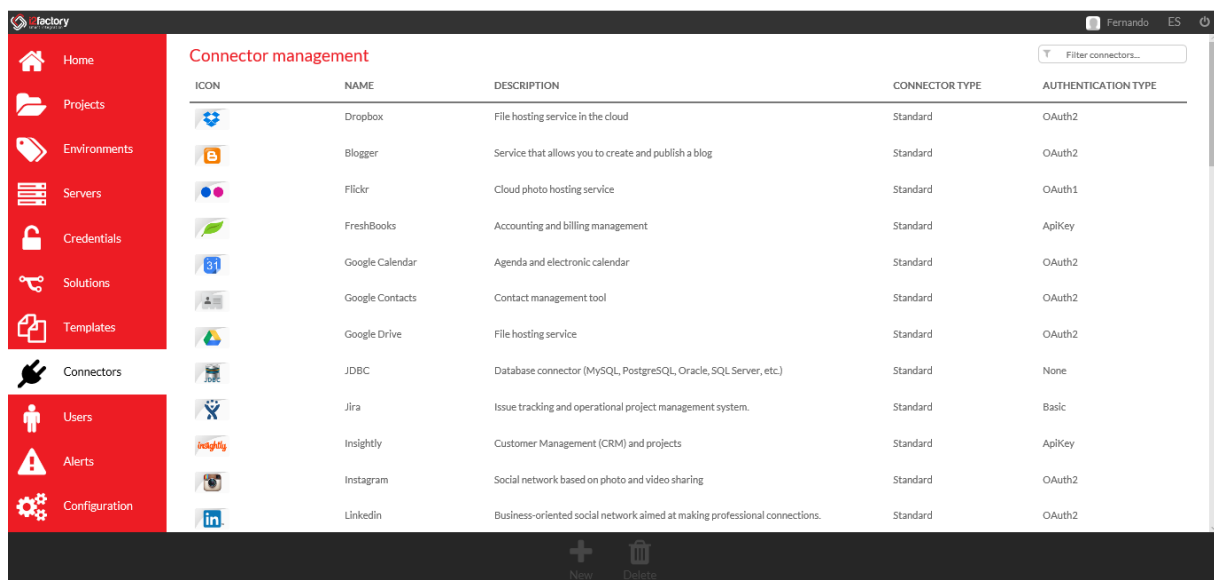
Figure 19 - "Connectors" detail in Guaraná Cloud Dashboard

Users (Figure 20) is only available to "Super User" and "Administrator" users type. It allows users to be added, edited, locked or deleted.

Figure 20 - "Users" detail in Guaraná Cloud Dashboard

After showing all the available controls in Guaraná Cloud Dashboard, Guaraná Cloud IDE graphical editor is introduced. Figure 21 and Figure 22 show us an IDE with four main areas: on the left hand side a List of Connectors (Figure 21) to be used as data input/output, or a List of Tasks (Figure 22) used to perform messages transformations; at the centre is the Development Area, where integration solutions are drawn by "Dragging and Dropping" connectors and tasks from the left panel and connecting them (connectors to tasks or tasks to tasks) with Slots (already introduced in section 3.1 Architecture); on the right hand side is the Properties Bar, which changes according to the selected item at the Drawing Area; at the bottom is the Action Bar with some of the most common actions found on other IDEs, such as Debug, Save, Cut, Copy, Paste, Delete, Undo, Redo, etc.
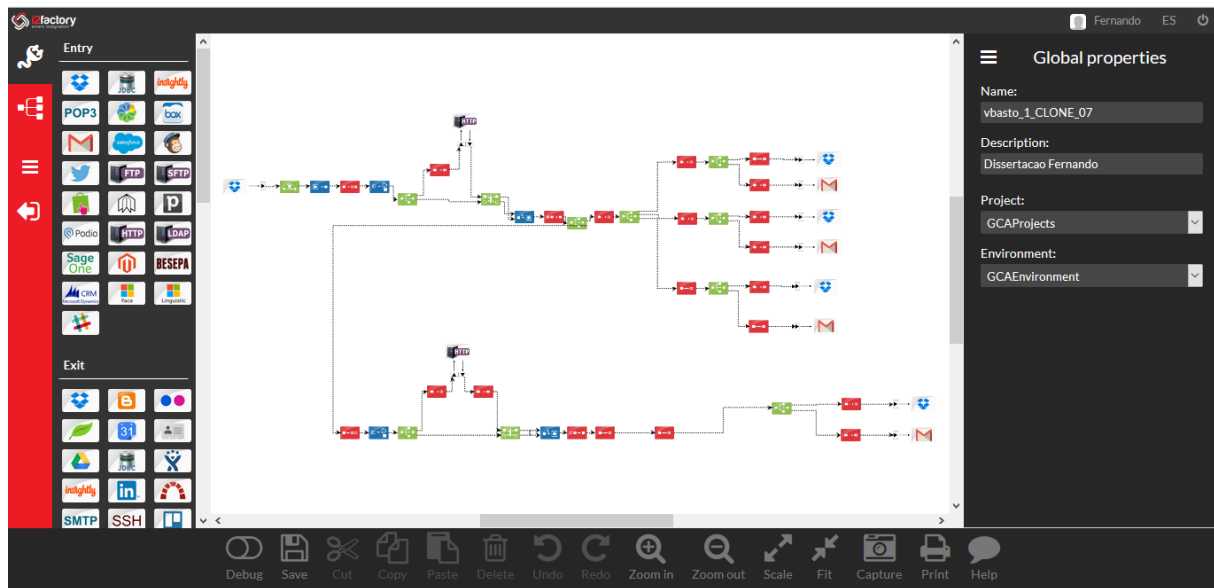
Figure 21 - Guaraná Cloud IDE - Connectors selected
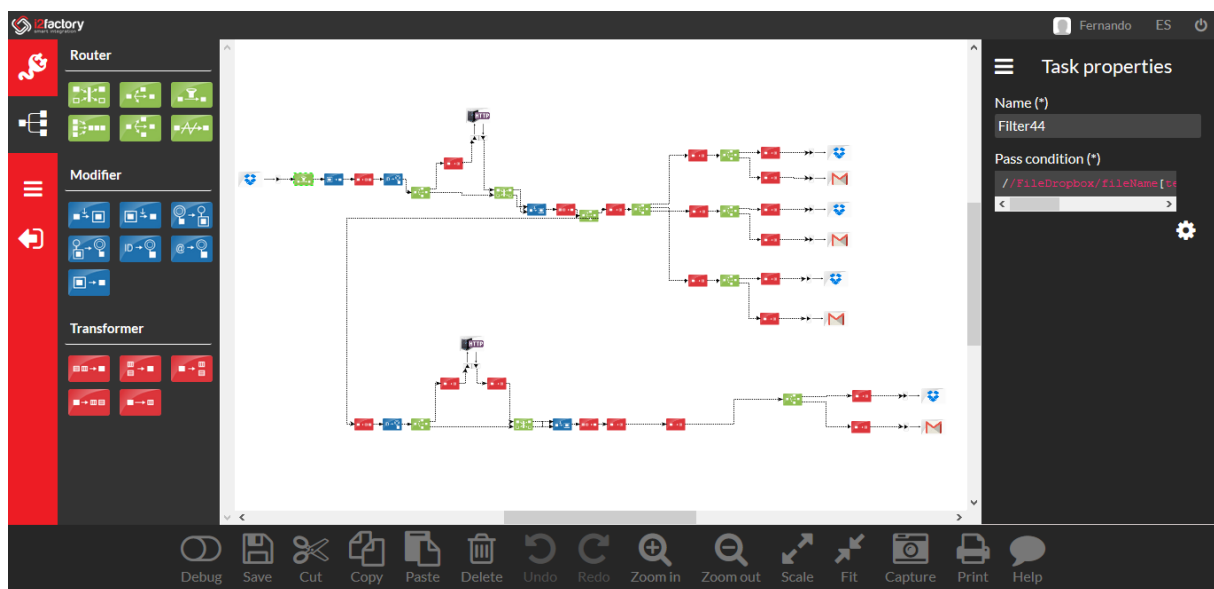
Figure 22 - Guaraná Cloud IDE - Tasks selected

Looking closer in "Connectors" available in Guaraná Cloud Platform (Figure 23), it is possible to see that they are grouped in four types, "Entry", "Exit", "Solicitor" and "Responder", the same already mentioned in Figure 5. This categorization also takes in mind the definitions in Table 6, from the Guaraná DSL definition.

Figure 23 - Available Connectors in Guaraná Cloud Platform

Based in the tasks defined in Guaraná DSL (Table 1, Table 2, Table 3), Guaraná Cloud Platform has a set of tasks grouped by the same criteria as in the DSL definition (Figure 24). However, not all the tasks defined in the Guaraná DSL are available in Guaraná Cloud Platform.
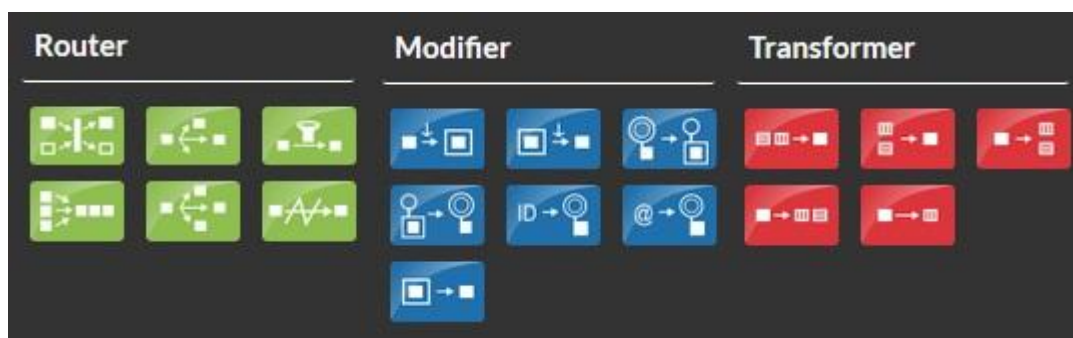
Figure 24 - Available Tasks in Guaraná Cloud Platform

Some other tasks have been implemented as being part of connectors, like the Table 5, that are part of Dropbox connectors. In this case (Figure 25), it's not only a task "Stream2XML" that is available together with the connector, but it also does the conversions or mappings to CSV, JSON, HTML, etc., meaning that is a connector (a "port", according to Guaraná DSL, described in Chapter 3.2) with a set of Mapper tasks (described in Table 5).

Figure 25 - Properties detail on Dropbox connector

Figure 26 shows a solution run in Debug Mode, which requires the server to be created with "Debugging" enabled.
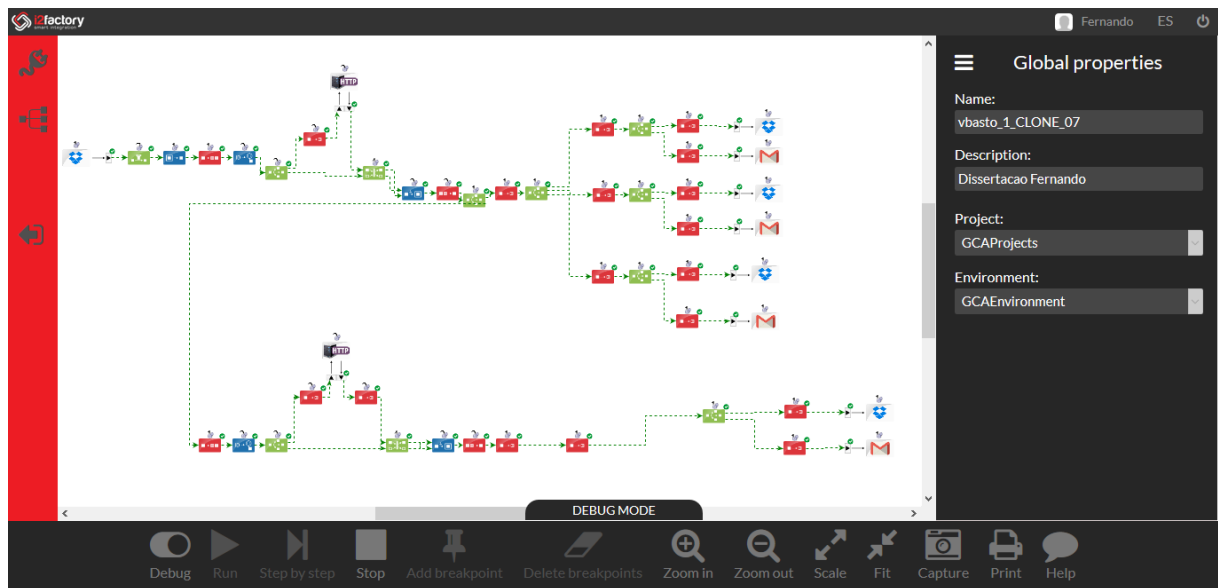
Figure 26 - Debug mode, showing a successful execution

When in "Running mode" it is possible to watch the Solution behaviour and track for errors. Users can have an overview of the amount of incoming and outgoing messages and other statistical data (Figure 27).
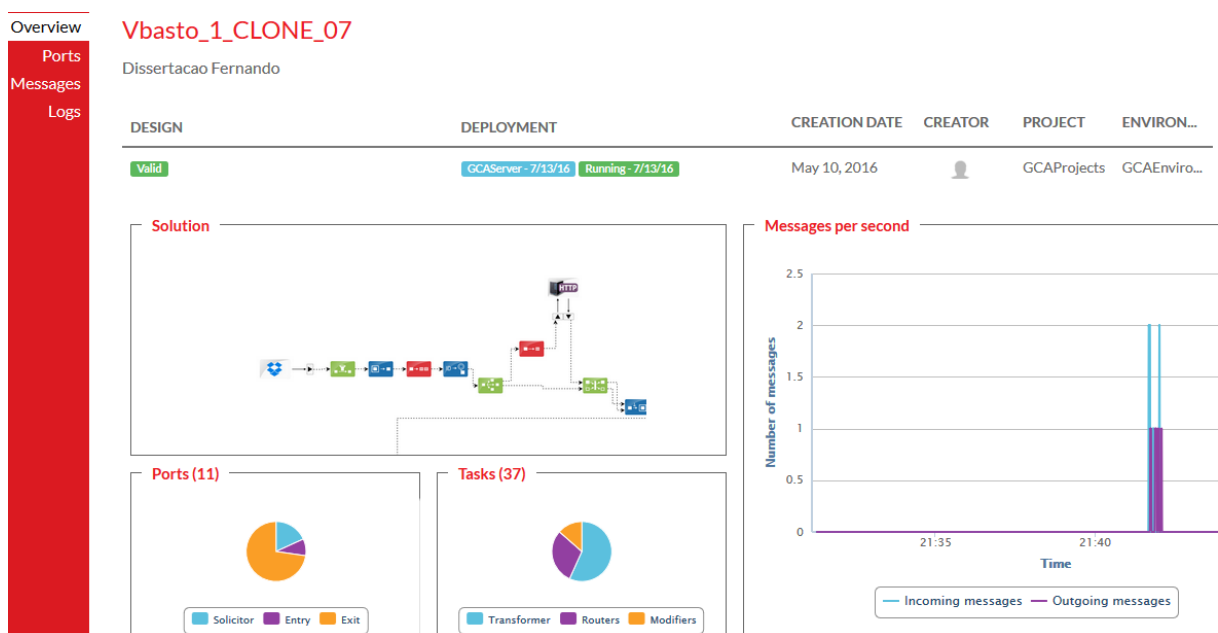
Figure 27 - Detail view from a running solution

Besides statistical data, it might also be useful to look into incoming or outgoing messages (Figure 28) or even messages between tasks (Figure 29) while looking for errors or unexpected results.

34
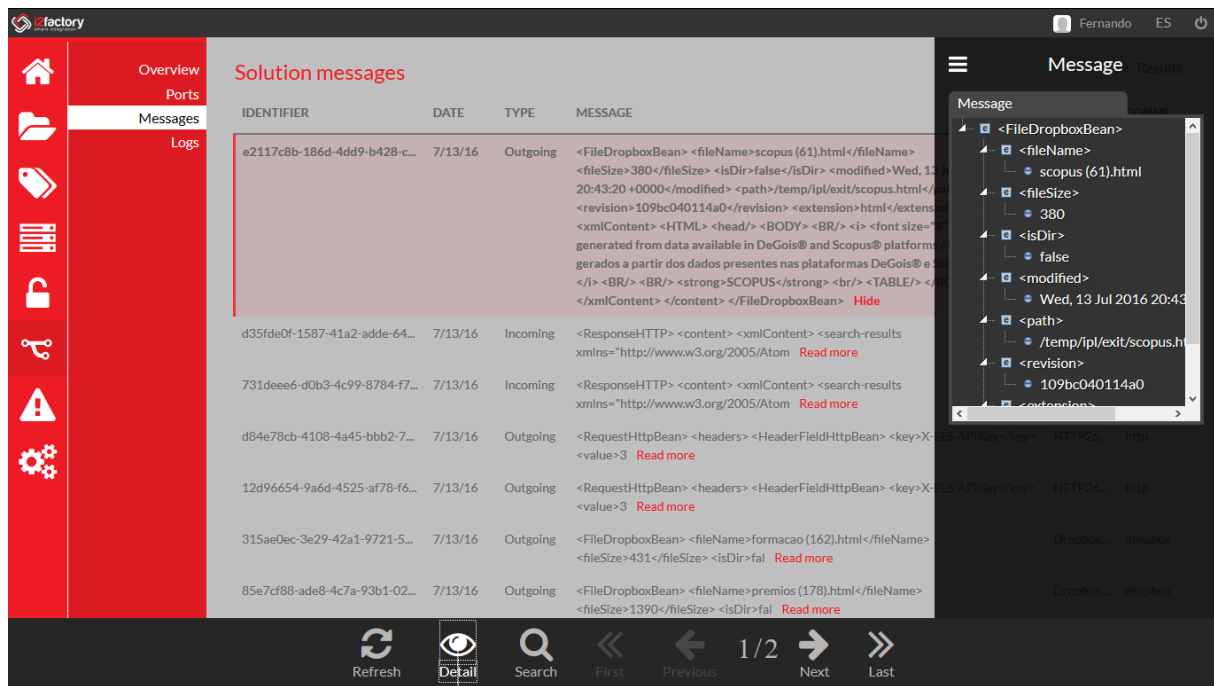
Figure 28 - Detail view of an outgoing message
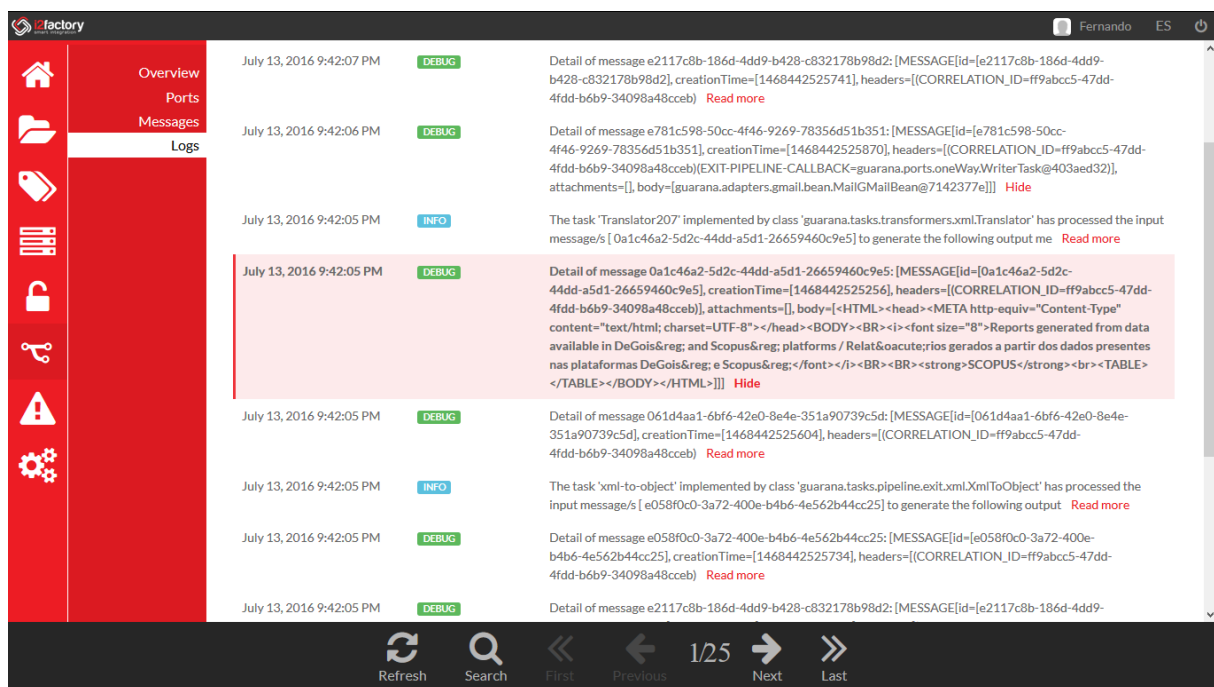
Figure 29 - Detail view of a message between tasks

The Guaraná Cloud Platform IDE presented in this section reveals the most relevant features used to develop the integration solution subject of this thesis. The next chapter presents in detail the integration scenario and corresponding properties, as well as the integration solution developed using Guaraná Cloud Platform.

This page was intentionally left in blank.

Research outcome from R&D units information management

FCT is responsible for promoting, funding, evaluating research and manages the national research outcomes information system. FCT is the major national research planning and management entity. Other players performing research and research management at regional, institutional and research units/groups level exist and may also be considered as research outcomes information managers, producers and consumers. In this chapter those players and their roles will be addressed, from the research outcomes information management and information systems integration perspectives.

4.1 The problem of scientific outcome information management

Research units represent an essential pillar in the consolidation of a modern and competitive scientific system. They promote creative environments where new ideas may come true and where researchers may find the proper conditions to realize their scientific projects and to their career development (22).

A major share of scientific research in Portugal is done in Research and Development Units (R&D) and Associate Laboratories, funded and evaluated by FCT. According to the latest data provided by FCT (23), in 2014 there were 292 R&D Units and 26 Associate Laboratories funded, across all research areas. Some are public and hosted by higher education institutions, other are private, non-profit organizations. In the period between 2011 and 2013 there were more than 22000 researchers affiliated to these institutions (24). Eighty four R&D Centres were rated as "Excellent" (Figure 30) in the last evaluation process (23). As stated by FCT, "R&D Centres of Excellence play a key role in advancing science in

Portugal, across all research areas." These Centres of Excellence are supported by FCT, establishing themselves internationally, by addressing issues of national and global relevance (23).
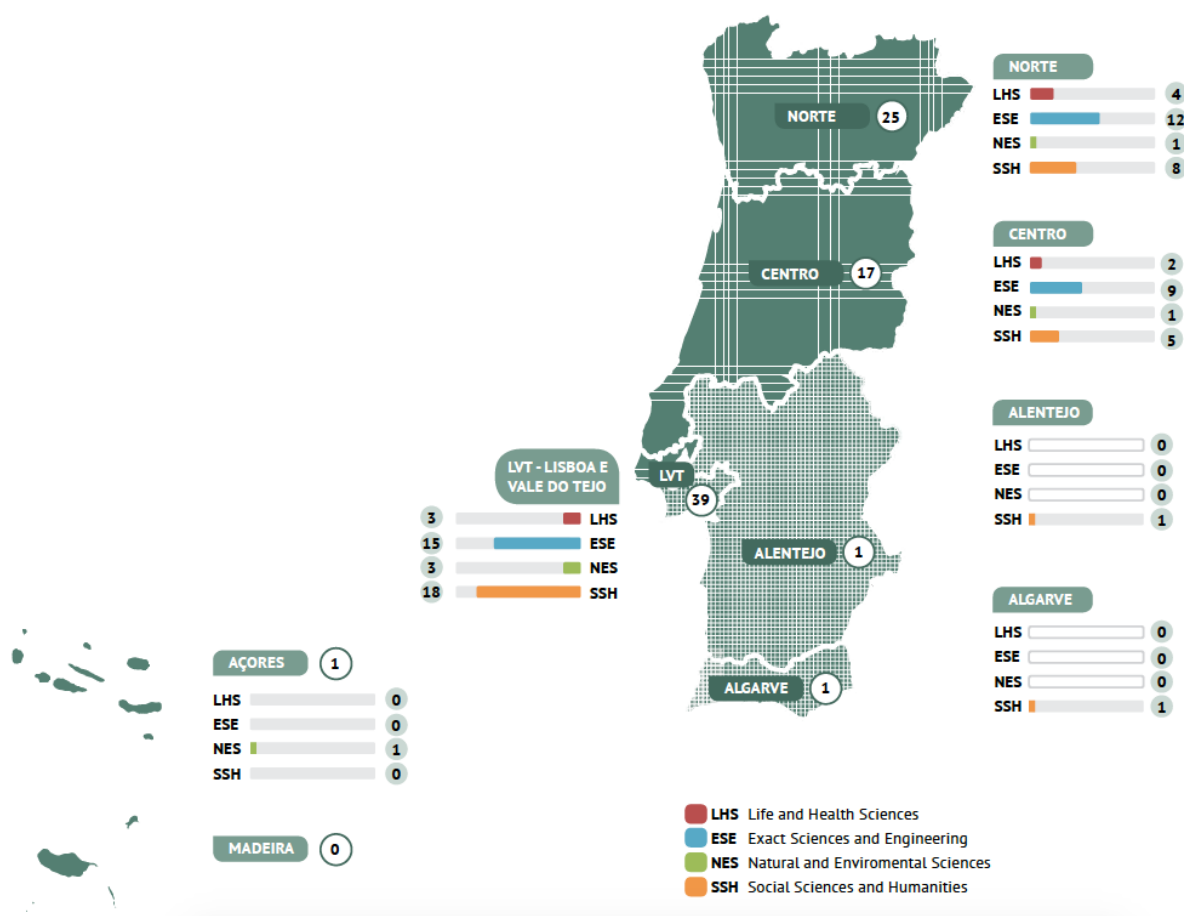
Figure 30 - Centres of Excellence (23)

FCT periodically launches contests for projects on R&D in all scientific domains, besides contests in specific scientific areas. Between 2011 and 2013 there were about 2300 projects supported by FCT. Besides this, FCT also ensures international partnerships with the U.S.A., the participation of Portuguese scientific community in bilateral and multilateral programs, contributions for international scientific organizations like CERN, ESA and EMBO (24). Another important responsibility of FCT is to collect, organize, compile, summarize, report and provide national research outcomes and activities information, by the means of electronic repositories and platforms.

4.2 Data sources, data structures and scientific outcome information management platforms

FCT operates some websites and platforms with the purpose to publicly announce information about contests, national and institutional research results, evaluation reports, rankings, etc., about Portuguese R&D institutions (25). In this thesis FCT DeGóis platform is taken as the reference digital repository and platform of research outcomes information. Plataforma DeGóis was conceived having in mind the maximum flexibility for being used in different purposes, such as the publication of *Curricula* by entities from SCTN, by FCT or by researchers. A DeGóis *Curriculum* is more detailed than the *Curricula* available in other FCT platforms such as the FCTSig *Curriculum* and a direct consequence of that is that creating and updating a DeGóis *Curriculum* it's harder and longer than doing it in FCTSig. Adopting a DeGóis *Curriculum* may be part of a strategy to manage a researcher career in long term rather than using a FCTSig *Curriculum,* that might be an option when the goal is to quickly provide a *Curriculum* to FCT contests or other short term, temporary data requests about individual researchers activities (22). Besides FCTSig and Plataforma DeGóis, which are mostly concerned with research in Portugal or done by Portuguese, it should also be taken into account other international science and research platforms such as ORCID, Web of Science and Scopus. Similar to the Portuguese Plataforma DeGóis, there is in Brazil Plataforma Lattes, which hosts a considerable amount of information about Portuguese researchers.

4.2.1 *Plataforma DeGóis and Curriculum DeGóis*

Plataforma DeGóis is a tool owned by FCT for collecting, providing and analyzing the intellectual property production, scientific and *curriculum* information of the Portuguese researchers. Is a portal whose main features are the individual management of the *curriculum* by the user, query of science and research indicators and *curricula* search based on criteria related to *curriculum* content.

The *curricula* management system (*curriculum* DeGóis) allows registered users to create their *curricula*, to insert their personal data, personal and professional address, jobs, spoken languages, awards, titles gained and research paths, as well as all the kinds of scientific outcomes and a detailed description of the projects the researcher was or is involved. It also allows to register participation in evaluation boards, identify scientific areas in which researchers work, and relate the scientific outcome with Organization for Economic Co-operation and Development (OECD) international science fields identifiers that allow

comparison of the *curriculum* DeGóis with other models produced in other scientific communities (26).

DeGóis platform is owned by FCT, Ministério da Educação e Ciência (Ministry of Education and Science of Portugal) which, through a quadripartite agreement between the FCT, the Ministério da Ciência, Tecnologia e Inovação of Brazil (Ministry of Science, Technology and Innovation), the Gávea laboratory of the Department of Information Systems from University of Minho and the Stela group from the Federal University of Santa Catarina in Brazil, guarantees the maintenance of the basic principles of DeGóis platform, and establishes the legal and institutional way the project is developed (26).

4.2.2 ORCID

ORCID is an open, non-profit, community-driven effort to create and maintain a registry of unique researchers identifiers and a transparent method of linking research activities and outputs to these identifiers. ORCID is unique in its ability to reach across disciplines, research sectors and national boundaries. It is a hub that connects researchers and research through the embedding of ORCID identifiers in key workflows, such as research profile maintenance, manuscript submissions, grant applications, and patent applications (27).

Researchers may benefit from ORCID two core functions:
- a registry to obtain a unique identifier and then manage a record of activities;
- APIs that support system-to-system communication and authentication.

ORCID makes its code available under an open-source license and posts an annual public data file under a Creative Commons Zero (CC0) waiver for free download (27).

The ORCID Registry is available free of charge to individuals who may obtain an ORCID identifier, and then manage their record of activities and search for others in the Registry. Organizations may also use it, become members, link their records to ORCID identifiers, update ORCID records, receive updates from ORCID, register their employees and students with ORCID identifiers (27).

4.2.3 Web of Science

According to their own words, Web of Science has became the gold standard for research discovery and analytics as a consequence of their meticulously work indexing the most important literature in the world. Web of Science connects publications and researchers through citations and controlled indexing in curated databases spanning every discipline. Using Web of Science researchers may do a search for cited reference to track prior research and also to monitor current developments in over 100 year's worth of content that is fully indexed (28).

Thomson Reuters, the owner of Web of Science, claims to have the world's largest collection of research data, books, journals, proceedings, publications and patents:
- across regions, all disciplines and content types;
- connected through citations and
- for faculty, researchers and students.

Not being a publisher, it claims to offer unbiased metrics based on citation activity of the most impactful global and regional journals, books and proceedings for scholarly community, remaining free from proprietary involvement (29).

4.2.4 Scopus

Scopus claims to be the largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings. Scopus features smart tools to track, analyze and visualize research, delivering a comprehensive overview of the world's research output in the fields of science, technology, medicine, social sciences, and arts and humanities (30).

Scopus claims comprehensiveness, having twice as many titles and over 50% more publishers listed than any other Abstracting and Indexing (A&I) database, with interdisciplinary content that covers the research spectrum. Timely updates from thousands of peer-reviewed journals, preliminary findings from millions of conference papers, and the thorough analysis in an expanding collection of books ensure researchers have the most up-to-

date and highest quality interdisciplinary content available. Scopus claims to be the only leading database that is daily updated, rather than weekly. There can be found journals, books, open access journals, conference papers and patents (31).

Scopus supports data exportation to reference managers such as Mendeley, RefWorks and EndNote. Besides this, there is a set of APIs available to registered or non-registered users, being that the last ones have limited access to a basic metadata and basic search functionality (32) (33).

4.2.5 Plataforma Lattes

The Plataforma Lattes is the experience of CNPq in integrating *Curricula* databases, from research groups and institutions into a single Information System in Brazil. Its current dimension extends not only to the action of planning, management and operation of CNPq development, but also from other federal and state funding agencies, the state foundations that support science and technology, higher education institutions and also research institutes. Furthermore, it became strategic not only for planning and management activities, but also for the formulation of the Ministry of Science and Technology from Brazil policies and other governmental agencies in the area of science, technology and innovation (34).

The Curriculum Lattes has become a national standard in Brazil in the record of past and present life of students and researchers in the country, and is now adopted by all development agencies, universities and research institutes in the country. For its wealth of information and its increasing reliability and scope, has become indispensable and compulsory for the analysis of merit and competence of claims for funding in science and technology (34).

The Directory of Research Groups in Brazil is an inventory of active groups in the country. The constituents of human resources groups, research lines and the involved industry sectors, the specialties of knowledge, scientific, technological and artistic production and patterns of interaction with the productive sector are some of the information contained in the directory. The groups are located in higher education institutions, research institutes, etc. The individual information of the participants of the groups are obtained from their Curriculum Lattes (34).

42

The Directory of Institutions was designed to promote the organizations of the National System of Science, Technology and Innovation to the condition of users of the Lattes Platform. It records any and all organizations or entities which establish some kind of relationship with the CNPq (institutions in which students and researchers supported by CNPq develop their activities; institutions where research groups are housed; institutions that strive participate in these programs and services, etc). Public availability of the Platform data on the Internet gives greater transparency and reliability to the promotion activities of the CNPq and agencies that use it, strengthen exchanges between researchers and institutions, and is an inexhaustible source of information for studies and research. In the way that its information is recurrent and cumulative, also has an important role in preserving the memory of the research activity in Brazil (34).

This page was intentionally left in blank.

The Guaraná based integration solution

This chapter presents briefly the integration problem addressed in this thesis, followed by a technical description of the solution developed for the research outcomes information system integration, adopting the perspective of a research unit. The software ecosystem is introduced and a description of data sources and data structures are provided. The conceptual solution designed with Guaraná DSL and the solution implemented with Guaraná Cloud IDE are also described, and finally the web output generated by the solution that is published in the Computer Science and Communications Research Centre content management systems is shown.

5.1 Software ecosystem

As previously stated in this thesis, a major part of scientific research in Portugal is done in Research and Development (R&D) Units in Higher Education Institutions. Although any national R&D unit could be used here for the integration solution results analysis, Computer Science and Communications Research Centre at Polytechnic Institute of Leiria is taken as the reference case study. The integration solution developed in this thesis will replace the manual process of collecting, computing and updating research unit production, assure that researchers list of publications, participations on scientific events and other type of scientific outcomes are reported and valid, classify the publications according to quality ranks, aggregate, summarize and generate annual reports using software and communication tools such as e-mail, style sheets and text editors. Turning this into an automated process supported by integration software, only requiring a way to identify the researchers of a research unit, would reduce or eliminate the need for manual and/or ad-hoc procedures.

The EAI based integration solution proposed in this thesis involves the interaction with four main data sources/applications to collect or publish data about researchers and corresponding research outcomes: "Local Research Unit Characterisation", "Plataforma DeGóis", "Scopus" and "CMS Application". The first data source consists in a XML file stored in a file system accessible via TCP/IP protocols, containing basic data about

researchers, needed to feed the integration solution. Based on this data the integration solution creates and send requests of researchers CVs to "Plataforma DeGóis", which is the national repository of CVs owned and maintained by the science and research management national authority. The integration solution aggregates the researchers CVs into a research unit scale XML document, collects additional information related to the research outcomes referred in researchers CVs (e.g. a conference paper number of citations) available in "Scopus" platform, and finally, transforms the summarized data into a HTML document that is sent to the CIIC-IPLeiria Joomla Content Management System ("CMS Application"). All the integration tasks and interactions with the external applications are specified with Guaraná DSL and processes by the Guaraná integration engine introduced in previous chapters.

The data sources and data structures used in the integration solution are briefly and graphically presented next, and the corresponding full detailed specification are include in Appendix

.

Figure 31 shows a graphical representation of Researchers.xsd, a XML document schema defining the structure of data about researchers that feeds the unique input port of the entire integration solution. A sample file can be found at Appendix 1 - Researchers.xml.
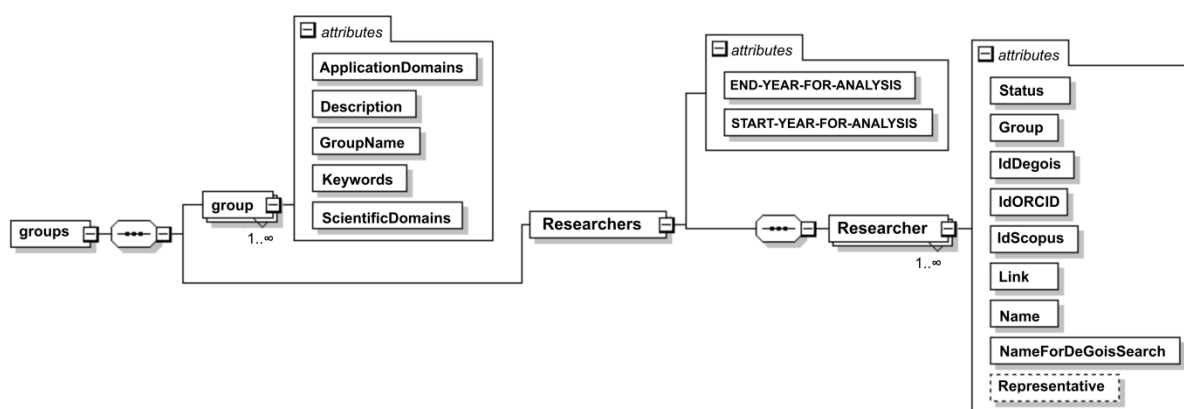
Figure 31- Researchers.xml Schema (14)

For the current integration solution only "Researcher" tags containing "Status" attribute equal to "Efetivo" are considered. "IdDegois" attribute must be previously and manually filled in the XML document with the corresponding researchers IdDegois, for the integration solution to look for their CV on "Plataforma DeGóis".

46

Figure 32 shows the researchers CV XML schema used by "Plataforma DeGóis"[1]. RESTful web services request/responses are exchanged between the integration solution and DeGois platform to search/deliver a researcher CV identified by the IdDegois attribute. For example, the HTTP request "GET http://193.136.2.109/XMLExportacao/CV/?idinst=ESTGIPL83&iddegois=02650000429 94737" --output-document=CV.xml" represents the RESTful web service request for the researcher CV with IdDegois 0265000042994737.
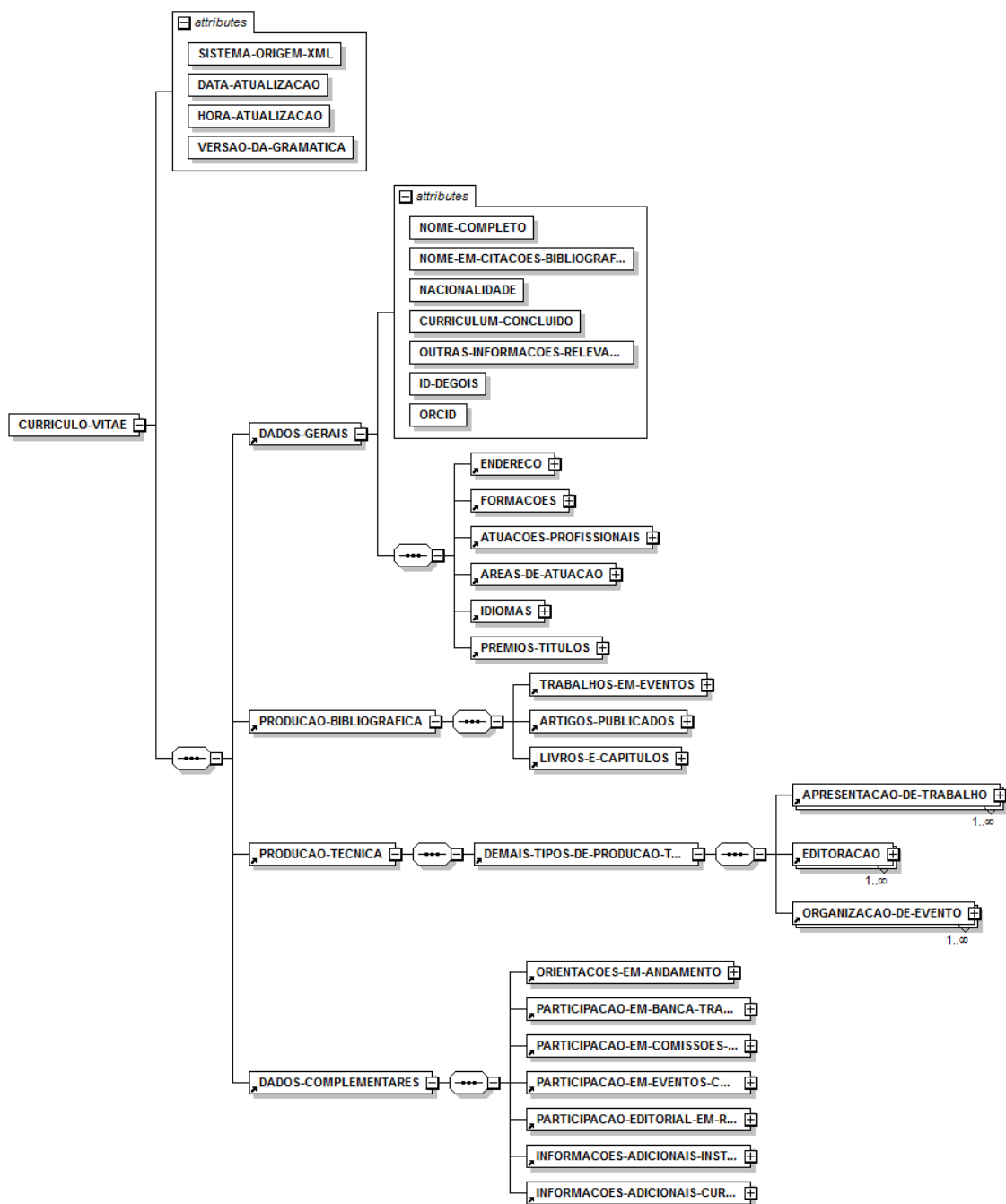
Figure 32 - CV XML Schema used by "Plataforma DeGóis"

---

[1] Since the XSD file takes hundreds of pages and it's not included here, it is possible to find it under the URI https://drive.google.com/open?id=0B-jANqT6HWXFdTZuXzRSenNoUkU

Similar RESTful requests/responses are exchanged between the integration solution and the Scopus platform, in order to collect detailed data about researchers production items such as paper indexing ID, paper number citations, etc. The XML schema adopted by Scopus platform for research production items is shown in Figure 33. A sample file may be found at Appendix 2 - Scopus.xsd. For example, the HTTP request "GET http://api.elsevier.com/content/search/scopus?query=AU-ID(53363129900)&APIKey=39b3ce4ff467c92d5e07d04a13426ce8&httpAccept=application/xml&field=citedby-count,eid,prism:doi,dc:title" represents the RESTful web service request for the researcher CV with Scopus ID 53363129900.
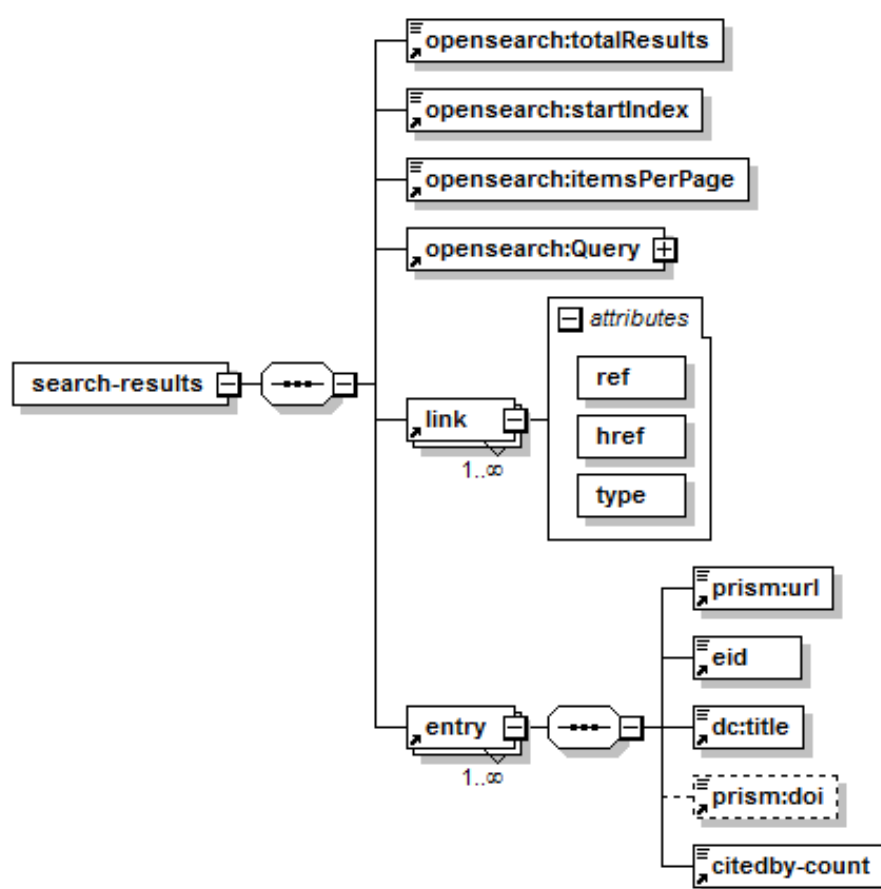
Figure 33 - XML Schema used by "Scopus"

5.2 Conceptual model in Guaraná DSL

The integration solution model specified with Guaraná DSL is shown in Figure 34 - Integration solution specification with Guaraná . An input XML file ("Researchers.xml") is stored in "Local Research Unit Characterization". That file contains the initial and main input

48

for the integration solution. There it will be found information about researchers (tag "investigadores"), namely their status (attribute "Estatuto"), the research group he or she belongs to (attribute "Grupo"), the researcher identification code in "Plataforma DeGóis" used for searches on that platform (attribute "IdDegois"), etc. The original flow of data (started with "Researchers.xml" contents) will be enriched with data provided from "Plataforma DeGóis" and from "SCOPUS". The expected output will be a set of HTML files that are sent to a "Joomla" CMS instance in the form of "Joomla" articles.
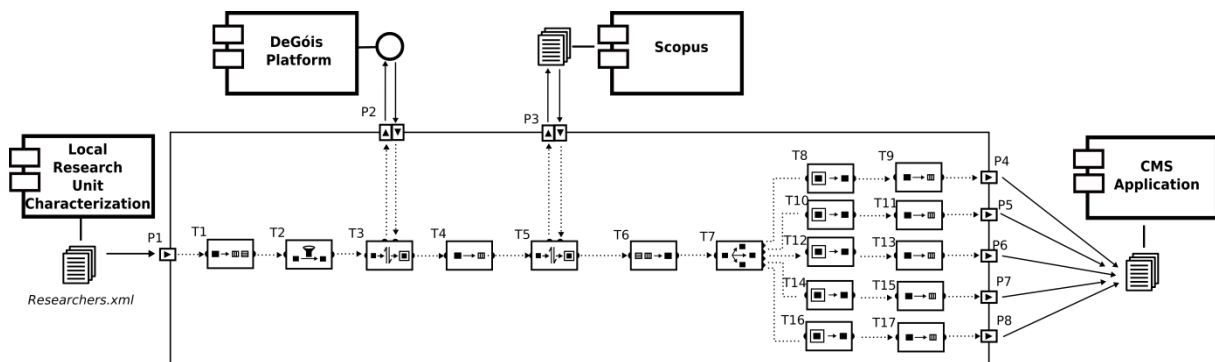
Figure 34 - Integration solution specification with Guaraná DSL

The workflow starts at entry port P1, which loads "Researchers.xml" contents and then periodically checks for changes on it. Task T1 splits the data obtained from P1 and each chunk corresponds to a researcher. From now on, each "Researcher" will be handled as a message.

Task T2 filters out researchers with attribute Status different from "Efetivo". Messages in the solution are then replicated at T3; one copy is used to build a "DeGóis" query, to be forwarded to "Plataforma DeGóis" by Solicitor Port P2. Solicitor Port P2 will then get a reply from "Plataforma DeGóis" ("Plataforma DeGóis" was queried by a researcher CV, query based on researcher's "IdDegois"). Still in task T3 messages will be merged, and the system will keep running with the same amount of messages it had right before starting task T3.

Task T4 changes message schema for the message that reaches task T5 to be able to hold new information coming from "Scopus". For example, XML attribute "ScopusId" is added to publication items XML elements to hold publications ScopusId retrieved from "Scopus" .

Task T5 retrieves information from "Scopus" to be associated with researchers CV information, as previously described.

From here onwards, information about researchers does not need to be treated individually. Task T6 re-unifies messages with information about each researcher into a single message, for research unit granularity processing. Task T7 replicates this unique message into five copies, which will be used to produce another five distinct HTML output documents, containing research unit scale indicators in a per research items type basis (projects, papers, organized events, awards, advanced training, news).

Tasks T8, T10, T12, T14 and T16 (Slimmer tasks) perform messages cleansing, preserving only information related to each of the specific research indicator to be calculated/processed. Finally, tasks T9, T11, T13, T15 and T17 perform messages transformation, more precisely, transformation of XML represented data into HTML documents, corresponding to the five different category of research item types. The output of these tasks (five HTML documents) is forwarded through Exit Ports P4-P8 to "CMS Application" in the form of an HTML CMS articles type, and immediately made accessible by the ("Joomla") CMS instance.

5.3 The solution in Guaraná Cloud

In this section we present the integration solution implemented for research outcomes information management. The solution was designed with the Guaraná DSL (presented in the previous section) and implemented in Guaraná Cloud platform.

Guaraná Cloud solution involves the collection, integration and transformation of data according to the following main workflows: read a XML file from Dropbox containing a list of researcher names that belong to the research unit; HTTP REST requests directed to "Plataforma DeGóis" to fetch XML representations of each researcher's CV; HTTP REST requests directed to Scopus to check if the publications contained in the researchers CVs are indexed by Scopus (and retrieve the corresponding Scopus IDs and citations in case they are indexed by Scopus); generation of a report (HTML document) per each type of research activity (projects, publications, news, awards, etc.); copy the generated HTML documents to a Dropbox folder, shared with the research unit Joomla CMS platform; send by email the

50

generated HTML reports to the research unit director.

Note that the integration solution and Joomla CMS shared Dropbox folder enables automatic updates of the research unit web site.

Figure 35 shows the Guaraná Cloud implementation of the integration solution described above, and Figure 36 shows an example of the generated HTML document output for the specific case of science dissemination activities (research unit activities announced in radio, newspapers, etc.).
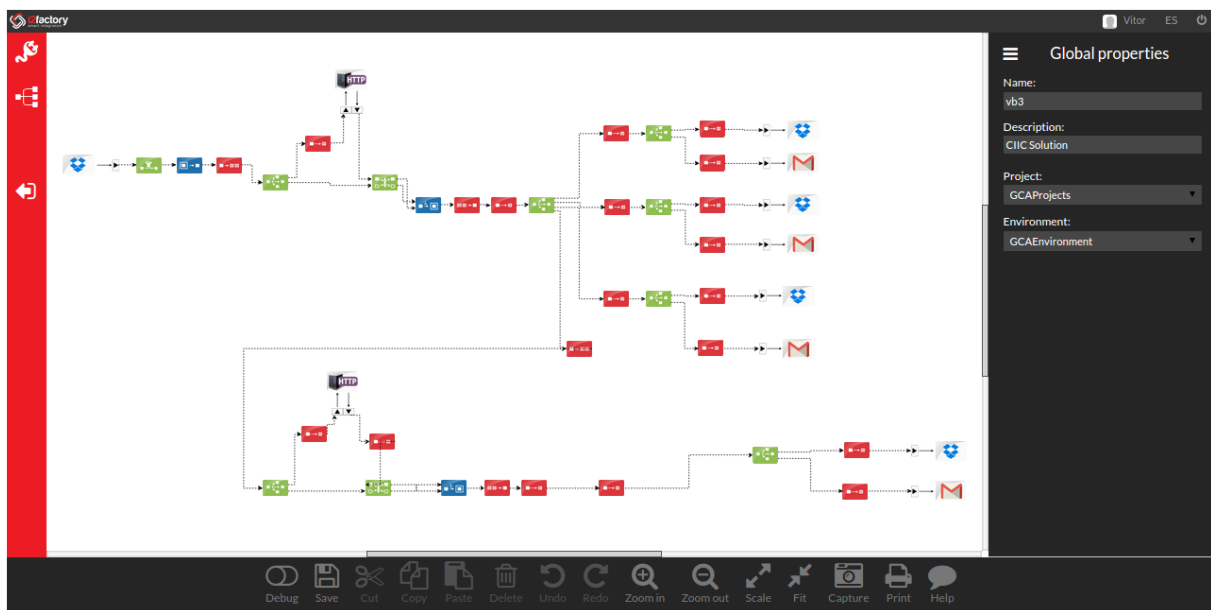
Figure 35 - Guaraná Cloud implementation of the integration solution

Figure 36 - Integration Solution generated report integrated with Joomla CMS

Conclusions and future work

Enterprise Application Integration (EAI) is a well-established research field, which provides methodologies, techniques and tools to design and implement integration solutions. Companies rely on EAI to reuse the applications that are available within their software ecosystems to support their business processes. There are currently several open-source integration platforms available for companies to assist the design and implementation of their integration solutions. The open-source integration platforms community got inspiration from the work of Hohpe and Woolf (2), which means they support the catalogue of integration patterns documented by these authors and follow the messaging based integration style. In this thesis Guaraná integration tools were studied and used to design and implement an innovative integration solution targeted for science and research outcomes information management. Guaraná was chosen due to its advantages with respect to some integration solutions quality attributes, with emphasis on platform independency.

Guaraná is divided into a domain-specific language and a set of tools from which stands out a cloud based editor and runtime system. Whereas Guaraná DSL can be used independently from engineering tools to design integration solutions and provide a full support to the integration patterns documented by Hohpe and Woolf (2), the Guaraná Cloud is an integrated development environment available on demand in the cloud and with a limited support to the DSL.

In this thesis, the research outcomes information management at research units, institutional and national levels were presented, as well as the overall research outcomes management ecosystems. Information producers, consumers, sources and platforms were addressed with focus on interoperability problems and information systems integration complexity. Firstly, Guaraná DSL was studied and used to model an integration solution to the science and research outcomes information management integration problem. Then,

Guaraná Cloud was studied and used to implement the model into an executable integration solution.

At the beginning, the lack of documentation was the Achilles' heel. The documentation available was only about Guaraná DSL. Since everything was new, it was not possible to find out any references about Guaraná Cloud. The only way to start deploying integration solutions using it was looking into Guaraná DSL definition, meaning, based in the name of components in Guaraná DSL and corresponding elements in Guaraná Cloud, and the descriptions from the components and expected behaviour in Guaraná DSL, and then, keep going by the try-fail methodology. Furthermore, some connectors available in Guaraná Cloud were not properly working, or at least, weren't working the way expected. For instance there was a connector to Local File System. However, it was a Cloud Environment without access to File System from where the server was running. After some e-mails it was clarified that that connector was a remain from former Guaraná versions. Another issue occurred with Google Drive Connector, where, once more because of the lack of documentation, only after some time and some e-mails it was realized that it was not possible to upload a file contents do Google Drive. At the first months of use there was no easy way for debugging. For that purpose it was necessary to deploy the solution, run the solution, wait for the end and then watch for messages tracked. It may look easy, but it turns into an Herculean task when the number of tasks or data entry increases.

Although Guaraná DSL is a simple and at the same time a rich modelling language, there is still a gap between the language and the tool support in the cloud to design the integration solutions. Guaraná Cloud has concentrated efforts to provide an extensible list of application adapters, which allow to communicate with the integrated applications, but has devoted less attention to support more building blocks of the DSL, such the different kinds of tasks. It makes more difficult the implementation of the model when the integration solution model has to be adapted due to a missing building block in the Guaraná Cloud.

Guaraná Cloud is a recent integration platform and is still under development. Considering this, it is important to: a) improve the DSL support by supporting new kinds of tasks; b) performing better testing and correct some bugs to improve reliability on the integrated development environment; c) the lack of documentation is an important problem, which must be solved by providing examples, tutorials, reports, and online help; d) access

control based in roles or something similar, avoiding or allowing a user to use another user credentials and access to solutions; e) copy or move solutions between servers, when a user is member of more than one group, having access to more than one server; f) keep an historical, version control or a mechanism that allows download and upload of full or part of an integration solution; g) allow users to create their own connectors according to their needs.

Regarding the integration solution developed in this thesis, it is important to highlight that it can be improved by aggregating other data sources/applications from which more scientific information could be extracted to enrich the web pages generated by the integration solution. Thus, in the future, the Brazilian WebQualis could provide the information regarding the ranking of each publication according to the Brazilian system; the ISI Web of Science could provide the information regarding the Journal Citation Report (JCR) impact factor for journal publications; other reports could be generated by the integration solution to enrich the analysis of CIIC-IPLeiria activity available on the web site.

From this thesis resulted a scientific paper titled "An EAI based integration solution for science and research outcomes information management" presented at CENTERIS 2015 Conference.

This page was intentionally left in blank.

Bibliographical references

1. **Frantz, Rafael Z.** *Enterprise Application Integration - An Easy-to-Maintain Model-Driven Engineering Approach.* 1ª ed. Sevilha : Universidad de Sevilha, 2012. p. 232.

2. **Hohpe, Gregor and Woolf, Bobby.** *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions.* s.l. : Addison-Wesley, 2003.

3. **Fernandes, Vitor Manuel Basto.** Information Systems Integration. *Unidade Curricular "Integração de Sistemas".* Leiria : ESTG - Instituto Politécnico de Leiria, 2014. Vols. Capítulo 1 - Introdução.

4. CORBA. [Online] Object Management Group, Inc. [Cited: 10 09 2015.] http://www.omg.org/spec/CORBA/.

5. Serializable Objects (The Java™ Tutorials > Java Naming and Directory Interface > Java Objects in the Directory):. [Online] Oracle and/or its affiliates. [Cited: 10 09 2015.] http://docs.oracle.com/javase/tutorial/jndi/objects/serial.html.

6. Web Services Architecture:. [Online] W3C®. [Cited: 10 09 2015.] http://www.w3.org/TR/ws-arch/.

7. What is EDI (Electronic Data Interchange)? | EDI Basics:. *EDI Basics:.* [Online] OpenText Corporation. [Cited: 10 09 2015.] http://www.edibasics.co.uk/what-is-edi/.

8. Extensible Markup Language (XML):. *Extensible Markup Language (XML):.* [Online] W3C®. [Cited: 10 09 2015.] http://www.w3.org/XML/.

9. XSL Transformations (XSLT):. [Online] W3C®. [Cited: 10 09 2015.] http://www.w3.org/TR/xslt.

10. ebXML - Enabling A Global Electronic Market:. [Online] OASIS Open. [Cited: 10 09 2015.] http://www.ebxml.org/geninfo.htm.

11. OWL Web Ontology Language Overview:. [Online] W3C®. [Cited: 10 09 2015.] http://www.w3.org/TR/owl-features/.

12. **Hohpe, Gregor.** Hub and Spoke [or] Zen and the Art of Message Broker Maintenance - Enterprise Integration Patterns:. *Enterprise Integration Patterns.* [Online] [Cited: 19 9 2015.] http://www.enterpriseintegrationpatterns.com/ramblings/03_hubandspoke.html.

13. **Bakker, Loek.** Goodbye Hub-and-Spoke, Hello ESB? Integration Architecture With BizTalk 2004 | Microsoft Cloud:. *.Net Developers Journal.* [Online] [Cited: 19 09 2015.] http://dotnet.sys-con.com/node/121831.

14. *An EAI based integration solution for science and research outcomes information management.* **Sequeira, Fernando Rosa, et al.** s.l. : Elsevier, 2015, Procedia Computer Science, Vol. 64, pp. 894-901. doi:10.1016/j.procs.2015.08.604.

15. **Haugg, Igor Gamste and Frantz, Rafael Z.** *Otimização Aplicada a Engenharia de Software para Auxiliar na Escolha de Provedores e Implantação de Soluções de Integração de Aplicações Empresariais em Nuvem.* Ijuí : Programa Institucional de Bolsas de Iniciação Tecnológica e Inovação PIBITI/CNPq - 2013-2014, 2014.

16. **Frantz, Rafael Z., Reina-Quintero, A. M. and Corchuelo, Rafael.** A Domain-Specific Language to Design Enterprise Application Integration Solutions. *International Journal of Cooperative Information Systems (IJCIS).* 2011, Vol. 20, pp. 143-176.

17. **Schmidt, Douglas C.** Guest Editor's Introduction: Model-Driven Engineering. *IEEE Computer.* 2006, Vol. 39, 2, pp. 25-31.

18. **Menor, Enrique.** *Dossier Técnico Guaraná 2.0.* s.l. : Inteligent Integration Factory S.L., 2011.

19. *On the Design of a Domain Specific Language for Enterprise Application Integration Solutions.* **Frantz, Rafael Z., Molina-Jimenez, Carlos and Corchuelo, Rafael.** Malaga : MoSE, 2010. ISSN 1613-0073.

20. **i2factory.** Visión global. *i2factory Academy.* [Online] i2factory, 5 6 2016. [Cited: 5 7 2016.] https://confluence.i2factory.com/pages/viewpage.action?pageId=884776.

21. *Enterprise Application Integration: A Domain-Specific Language to Design Platform-Independent Models.* **Frantz, Rafael Z., et al.** s.l. : International Journal on Software and Systems Modeling (SoSyM), 2016.

22. **Filipe, Ronaldo Carvalho.** *Desenvolvimento de uma plataforma de gestão das atividades de investigação e produção científica para unidades de investigação.* Leiria : Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, 2013.

23. **FCT - Fundação para a Ciência e Tecnologia.** FCT — Sobre a FCT:. *FCT - Fundação para a Ciência e a Tecnologia:.* [Online] 2015. [Cited: 23 09 2015.] http://www.fct.pt/documentos/Brochura_FCT_web.pdf.

24. **FCT - Fundação para a Ciência e a Tecnologia.** FCT — Principais linhas de Ação:. *FCT - Fundação para a Ciência e a Tecnologia:.* [Online] 2015. [Cited: 24 09 2015.] http://www.fct.pt/linhasatividadefct.

25. —. FCT — Portais FCT:. *FCT - Fundação para a Ciência e a Tecnologia:.* [Online] [Cited: 24 09 2015.] http://www.fct.pt/portaisfct.

26. **Plataforma DeGóis® - FCT.** Plataforma DeGóis - Plataforma Nacional de Ciência e Tecnologia:. *Plataforma DeGóis:.* [Online] [Cited: 24 09 2015.] http://degois.pt/index.jsp?id=1.

27. What is ORCID? | ORCID:. *ORCID | Connecting Research and Researchers:.* [Online] ORCID. [Cited: 26 09 2015.] http://orcid.org/node/47.

28. **Thomson Reuters.** Web of Science - IP & Science - Thomson Reuters:. [Online] [Cited: 24 09 2015.] http://wokinfo.com/.

29. —. The Citaton Connection - Real Facts - IP & Science - Thomson Reuters:. *Web of Science.* [Online] [Cited: 2015 09 25.] http://wokinfo.com/citationconnection/.

30. **Elsevier.** Scopus | The largest database of peer-reviewed literature | Elsevier:. [Online] [Cited: 25 09 2015.] https://www.elsevier.com/solutions/scopus.

31. —. Content:. *Scopus | The largest database of peer-reviewed literature | Elsevier:.* [Online] [Cited: 25 09 2015.] https://www.elsevier.com/solutions/scopus/content.

32. —. Features:. *Scopus | The largest database of peer-reviewed literature | Elsevier:.* [Online] [Cited: 25 09 2015.] https://www.elsevier.com/solutions/scopus/features.

33. —. Api:. *Scopus | The largest database of peer-reviewed literature | Elsevier:.* [Online] [Cited: 25 09 2015.] https://www.elsevier.com/solutions/scopus/features/api.

34. **CNPq.** Sobre a plataforma - Portal CNPq:. *Plataforma Lattes:.* [Online] [Cited: 25 09 2015.] http://www.cnpq.br/web/portal-lattes/sobre-a-plataforma.

This page was intentionally left in blank.

Appendix

This page was intentionally left in blank.

8.1 Appendix 1 - Researchers.xml

The following is sample file from the XML file used to start the integration solution. It's the Researchers.xml file located at the Local Research Unit Characterization.

```
<?xml version="1.0" encoding="UTF-8"?><?xml-stylesheet type="text/xsl" href="merge.xsl"?>
<grupos>
        <grupo ApplicationDomains="e-Business, e-Commerce, internet of things, e-Security,
auditing and certification" AreasAplicacao="Negócio eletrónico, comércio eletrónico, internet
das coisas, segurança em redes, auditoria e certificação" AreasCientificas="Mobilidade IP,
sistemas de deteção/prevenção de intrusões, prevenção de perda de dados, autenticação e
controlo de acessos, redes definidas por software, qualidade de serviço e qualidade da
experiência, segurança em redes e computação em nuvem" Descricao="Comunicações e Telemática"
Description="Communications and Telematics" Keywords="Computer, devices and network
communications; Network centric computing; Network and information security; Software defined
networks" NomeGrupo="CT" PalavrasChave="Computadores, dispositivos e redes de comunicação;
Computação centrada na rede; Segurança da informação e das redes; Redes definidas por
software" ScientificDomains="IP mobility, intrusion detection/prevention systems, data loss
prevention, authentication and access control, software defined networks, quality of service
and quality of experience, security in cloud computing networks"/>
        <grupo ApplicationDomains="Scientific discovery by unknown solution spaces exploration,
quality improvement of business processes and services, innovation in product and services
design, decision support systems and business intelligence" AreasAplicacao="Descoberta
científica baseada em análise de dados, melhoria de qualidade e otimização de processos e
serviços organizacionais, suporte à inovação no desenho de produtos e serviços, sistemas de
apoio à decisão e descoberta de conhecimento" AreasCientificas="Paradigmas computacionais
inspirados na natureza, teoria dos jogos e métodos exatos. Desenho, desenvolvimento e
adaptação de algoritmos para otimização multimodal, dinâmica, discreta, estrutural e
multiobjetivo" Descricao="Inteligência Computacional e Otimização" Description="Computational
Inteligence and Optimization" Keywords="Natural computing; Adaptive and complex systems;
Optimization; Business intelligence" NomeGrupo="CIO" PalavrasChave="Computação Natural;
Sistemas Adaptativos e Complexos; Otimização; Extração de Conhecimento"
ScientificDomains="Computational paradigms gleaned from nature, game theory and exact methods.
Design, development and adaptation of algorithms for multimodal, dynamic, discrete, structural
and multi-objective optimization"/>
        <grupo ApplicationDomains="Digital contents production and multimedia development (in
tourism, cultural heritage protection, history and science dissemination), creative industries
(digital video and sound based creativity), educational and entertainment multimedia"
AreasAplicacao="Produção e desenvolvimento de conteúdos multimédia (no turismo, proteção do
património cultural, disseminação histórica e científica), industrias criativas (criatividade
em vídeo e som digital), educação e entretenimento multimédia" AreasCientificas="Mundos
virtuais, ambientes virtuais, realidade aumentada, multimédia tridimensional e estereoscopia,
grande alcance dinâmico, modelação procedimental, gramáticas de formas"
Descricao="Investigação em Computação Gráfica e Som" Description="Computer Graphics and Sound
Research" Keywords="Multimedia; Digital games; Computer graphics; Sound and music research"
NomeGrupo="CGSR" PalavrasChave="Multimédia; Jogos digitais; Computação gráfica; Som e música"
ScientificDomains="Virtual worlds, virtual environments, augmented reality, tridimensional
multimedia and stereoscopy, high dynamic range, procedural modeling, shape grammars"/>
        <grupo ApplicationDomains="eTherapy, eHealth support services, clinical information and
```

management integrated systems, collaborative, education and management tools" AreasAplicacao="Tecnologias para o bem estar, personalização e desmaterialização de processos de saúde com recurso a tecnologias de informação e comunicação, terapia eletrónica e saúde eletrónica, gestão e integração de sistemas de informação clínicos, sistemas educacionais e colaborativos" AreasCientificas="Envelhecimento saudável, gerontotecnologia, previsão e prescrição, inclusão e acessibilidade digital" Descricao="Informática Aplicada à Saúde" Description="Health Informatics" Keywords="Ambient assisted living; Gerontotechnology; Mobile health; Healthcare information systems" NomeGrupo="HI" PalavrasChave="Desenvolvimento de tecnologias para o bem estar personalizado; Gerontotecnologia; e-Saúde; Sistemas de informação para a saúde" ScientificDomains="Healthy aging, gerontotechnology, predictive and prescriptive analysis, digital inclusion and accessibility"/>
        <investigadores ANO-FIM-PARA-ANALISE="2014" ANO-INICIO-PARA-ANALISE="2008">
                <investigador Estatuto="Efetivo" Grupo="CIO" IdDegois="0265000042994737" IdORCID="0000-0003-4269-5114" IdScopus="53363129900" Link="http://www.degois.pt/visualizador/curriculum.jsp?key=0265000042994737" Nome="Vitor Manuel Basto Fernandes" NomeParaBuscaDeGois="Vitor Manuel Basto Fernandes"/>
                <investigador Estatuto="Efetivo" Grupo="CGSR" IdDegois="2672961131904375" IdORCID="0000-0002-5966-3218" IdScopus="35092342500" Link="http://www.degois.pt/visualizador/curriculum.jsp?key=2672961131904375" Nome="Alexandrino José Marques Gonçalves" NomeParaBuscaDeGois="Alexandrino Jose Marques Goncalves"/>
                <investigador Estatuto="Associado" Grupo="CT" IdDegois="" IdORCID="" IdScopus="" Link="http://ww2.estg.ipleiria.pt/website/index.php?id=102002&pessoa_id=278593" Nome="Nuno Miguel Afonso Veiga" NomeParaBuscaDeGois="Nuno Miguel Afonso Veiga"/>

                <investigador Estatuto="Doutoramento" Grupo="CIO" IdDegois="" IdORCID="" IdScopus="" Link="http://ww2.estg.ipleiria.pt/website/index.php?id=102002&pessoa_id=278587" Nome="Marco Paulo Monteiro Ferreira" NomeParaBuscaDeGois="Marco Paulo Monteiro Ferreira"/>

                <investigador Estatuto="Colaborador" Grupo="HI" IdDegois="" IdORCID="" IdScopus="" Link="http://ww2.estg.ipleiria.pt/website/index.php?id=102002&pessoa_id=278601" Nome="Pedro Miguel Cardoso Gago" NomeParaBuscaDeGois="Pedro Miguel Cardoso Gago"/>

                <investigador Estatuto="ColaboradorGrupos" Grupo="HI" IdDegois="" IdORCID="" IdScopus="" Link="" Nome="Andreia Maria Ferreira Viveiros Tavares dos Reis" NomeParaBuscaDeGois="Andreia Maria Ferreira Viveiros Tavares dos Reis"/>

        </investigadores>
</grupos>

8.2 Appendix 2 - Scopus.xsd

The following is a sample file from the XML Schema from a generated output XML file from Scopus:

```xml
<?xml version="1.0" encoding="UTF-8"?>

<!-- W3C Schema generated by XMLSpy v2016 rel. 2 sp1 (x64) (http://www.altova.com) -->
<xs:schema  targetNamespace="http://www.w3.org/2005/Atom"  xmlns="http://www.w3.org/2005/Atom"
xmlns:dc="http://purl.org/dc/elements/1.1/"        xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/"
xmlns:prism="http://prismstandard.org/namespaces/basic/2.0/">
        <xs:import                               namespace="http://purl.org/dc/elements/1.1/"
schemaLocation="scopus_20160907_023.xsd"/>
        <xs:import                      namespace="http://a9.com/-/spec/opensearch/1.1/"
schemaLocation="scopus_20160907_021.xsd"/>
        <xs:import            namespace="http://prismstandard.org/namespaces/basic/2.0/"
schemaLocation="scopus_20160907_022.xsd"/>
        <xs:element name="eid">
                <xs:simpleType>
                        <xs:restriction base="xs:string">
                                <xs:enumeration value="2-s2.0-33646445529"/>
                                <xs:enumeration value="2-s2.0-33746497332"/>
                                <xs:enumeration value="2-s2.0-58049189140"/>
                                <xs:enumeration value="2-s2.0-80054061102"/>
                                <xs:enumeration value="2-s2.0-84864331565"/>
                                <xs:enumeration value="2-s2.0-84866025024"/>
                                <xs:enumeration value="2-s2.0-84875373076"/>
                                <xs:enumeration value="2-s2.0-84877900731"/>
                                <xs:enumeration value="2-s2.0-84935052337"/>
                                <xs:enumeration value="2-s2.0-84939518774"/>
                                <xs:enumeration value="2-s2.0-84962866428"/>
                                <xs:enumeration value="2-s2.0-84962921692"/>
                                <xs:enumeration value="2-s2.0-84974802649"/>
                                <xs:enumeration value="2-s2.0-84978764270"/>
                        </xs:restriction>
                </xs:simpleType>
        </xs:element>
        <xs:element name="link">
                <xs:complexType>
                        <xs:attribute name="ref" use="required">
                                <xs:simpleType>
                                        <xs:restriction base="xs:string">
                                                <xs:enumeration value="first"/>
                                                <xs:enumeration value="self"/>
                                        </xs:restriction>
                                </xs:simpleType>
                        </xs:attribute>
                        <xs:attribute name="href" use="required" type="xs:anyURI"/>
```

```xml
                    <xs:attribute name="type" use="required">
                            <xs:simpleType>
                                    <xs:restriction base="xs:string">
                                            <xs:enumeration value="application/xml"/>
                                    </xs:restriction>
                            </xs:simpleType>
                    </xs:attribute>
            </xs:complexType>
    </xs:element>
    <xs:element name="entry">
            <xs:complexType>
                    <xs:sequence>
                            <xs:element ref="prism:url"/>
                            <xs:element ref="eid"/>
                            <xs:element ref="dc:title"/>
                            <xs:element ref="prism:doi" minOccurs="0"/>
                            <xs:element ref="citedby-count"/>
                    </xs:sequence>
            </xs:complexType>
    </xs:element>
    <xs:element name="citedby-count">
            <xs:simpleType>
                    <xs:restriction base="xs:byte">
                            <xs:enumeration value="0"/>
                            <xs:enumeration value="1"/>
                            <xs:enumeration value="12"/>
                            <xs:enumeration value="3"/>
                    </xs:restriction>
            </xs:simpleType>
    </xs:element>
    <xs:element name="search-results">
            <xs:complexType>
                    <xs:sequence>
                            <xs:element ref="opensearch:totalResults"/>
                            <xs:element ref="opensearch:startIndex"/>
                            <xs:element ref="opensearch:itemsPerPage"/>
                            <xs:element ref="opensearch:Query"/>
                            <xs:element ref="link" maxOccurs="unbounded"/>
                            <xs:element ref="entry" maxOccurs="unbounded"/>
                    </xs:sequence>
            </xs:complexType>
    </xs:element>
</xs:schema>
```