Chapter 1

ENTERPRISE APPLICATION INTEGRATION: APPROACHES AND PLATFORMS TO DESIGN AND IMPLEMENT SOLUTIONS IN THE CLOUD

Fernando Rosa-Sequeira *[1], Vítor Basto-Fernandes[2], and Rafael Z. Frantz[3]

[1]Instituto Politécnico de Leiria, Leiria, Portugal
[2]University Institute of Lisbon, Department of Information Science and Technology, Lisbon, Portugal
[3]Unijuí University, Department of Exact Sciences and Engineering, Ijuí, RS, Brazil

PACS 05.45-a, 52.35.Mw, 96.50.Fm. **Keywords:** Application Integration, Integration Platform-as-a-Service, Cloud Computing.

*Corresponding Author Email: 2120027@my.ipleiria.pt

Abstract

Nowadays companies have a software ecosystem composed of more than one application to support their business processes. On the Enterprise Application Integration (EAI) field can be found a set of methods, techniques, and tools to integrate them in a synchronous or asynchronous way. In this chapter, we review integration approaches and integration platforms available in the Cloud. We demonstrate the use of an integration platform by means of a case of study for a research outcomes and technological information management integration problem. The proposal addresses Portuguese and international science and research outcomes information management, and corresponding information systems. There are presented problems in interoperability between information systems. A business and technological perspective is provided, including the conceptual analysis and modelling, an integration solution based on 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.

1. Introduction

Organizations rely on information systems and software applications to support their business activities. 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) or 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 [18]. 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, techniques, 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 [15].

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. 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 todays enterprises. 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.

In a general manner, integration technologies from nowadays do not let to work at a high level of abstraction, e.g., the implementation of solutions demands for 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 modeled 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 [20].

The rest of this chapter is organized as follows: Section 2. discuss four common integration approaches; Section 3. introduces the cloud-based integration platforms ranked as “Leaders” in the Magic Quadrant of Gartner, Inc.; Section 4. illustrates the use of the Guaran Cloud integration platform to solve an integration problem in the context of research outcomes information management; Section 5. concludes this chapter.

2. Integration Approaches

Nevertheless applications integration may occur in the same machine, many times it does not; in fact, some machines may be thousands kilometers from each other, and so almost all integration solutions have to deal with a few fundamental challenges, being one of them networks unreliability and slowness. An integration solution has to transport data between computers across networks. Compared to a process running on a single computer, distributed computing has to be prepared to deal with a larger set of possible problems. Often, two systems to be integrated are in distinguished continents and data between them have to travel through phone-lines, LAN segments, public networks, and satellite links. On each of these steps delays or interruptions may occur. Sending data across a network is multiple orders of magnitude slower than doing a local method call. Designing a widely distributed solution in the same way it would approach a single application could result in disastrous performance implications. Integration solutions need to transmit information between systems that use different distinct programming languages, operating platforms and data formats. An integration solution needs to be able to interface with all these different technologies [18].

As stated by Gregor Hohpe and Bobby Woolf [18] application integration may be done in four different ways, namely: File Transfer, Shared Database, Remote Procedure Calling, and Messaging. In File Transfer each application may play the role of producer or consumer, producing files of shared data to others consume or consuming what others have produced. Shared Database is the integration approach where a common database schema is shared, in a single physical database; since there is not duplicate storage, there is not also any data transference between applications. In Remote Procedure Calling integration applications expose some of their functionalities in a way that other applications can access them as a remote procedure, and so the communication occurs in real-time and synchronously. Finally, in Messaging integration each application is connected to a common messaging system, where it publishes its own messages and it reads messages from other applications. Since applications may read messages from that channel in a later time after they have been published by another applications, the communication is asynchronous; applications just have to agree on a channel and a message format to be used [20]. In the following sections we provide more information on each integration approach.

2.1. File Transfer

File Transfer integration, shown in Figure 1 may be used in organizations having many independent applications, running in different languages and distinct platforms. While an application plays the role of "Producer", exporting data to files that other applications playing as "Consumers" will read. A cunning decision to take is which format shall 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 issue, with the support of an industry of readers, writers and transformation tools. Furthermore it should also be taken in mind how often data are updated, i.e., when must data be written or read [18].

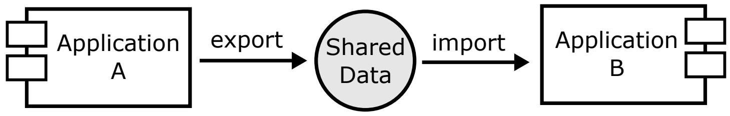

Figure 1. Integration by File Transfer.

2.2. Shared Database

Like in File Transfer approach, Shared Database integration, shown in Figure 2, may also be used in organizations having many independent applications, running in different languages and distinct platforms. However in this scenario, information must be shared rapidly and consistently. As suggested by its name, data from many applications are stored in the same database and so data consistency is ensured by database transaction management systems [18].

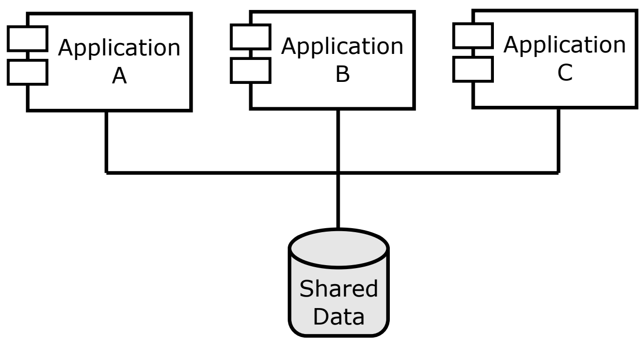

Figure 2. Integration by Shared Database.

2.3. Remote Procedure Calling

Remote Procedure Calling (RPC) integration, shown in Figure 3 may be used in situations where besides the need to share data between independent applications, running in different languages and distinct platforms, it is also needed to share functionalities 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 that interface. If one application needs to read or modify data of another application, that is done by making a call to the other application. In this situation each application maintains the integrity of the data owned by it. Being so each application can change its internal data or the way they are stored without having other applications affected. However, since this methodology works as 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. It should also be taken in mind that network issues may slow down or even cause fails in a part of the system that may affect the whole rest [18].

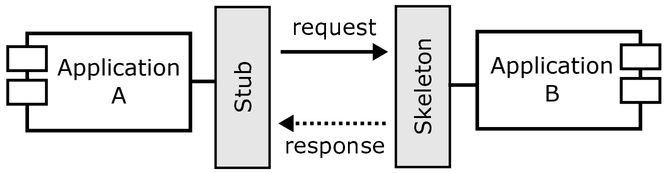

Figure 3. Integration by Remote Procedure Calling.

2.4. Messaging

Messaging Integration, shown in Figure 4, just like RPC might be used in organizations with independently built applications, running in different languages and distinct platforms, and where it is also needed to share functionalities in a responsive way to an event. However, unlike RPC integration, Messaging integration might be asynchronous. Just like a reaction to the common problems present in distributed systems (unavailability of systems, problems with network connections), Messaging systems enable transfer of data packets in a frequently, reliable, immediate way but also asynchronously, using customizable formats, y means of adapters. An adapter is a piece of code, independent of the application, which abstracts away the communication mechanism between the application and the message Having a system responsible for taking and delivering messages from one application to another one (or even more), allows the interoperability in situations where not all the systems are up-and-running at the same time. Nevertheless, it may occur that a sequence of messages may not be received in 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 [18].

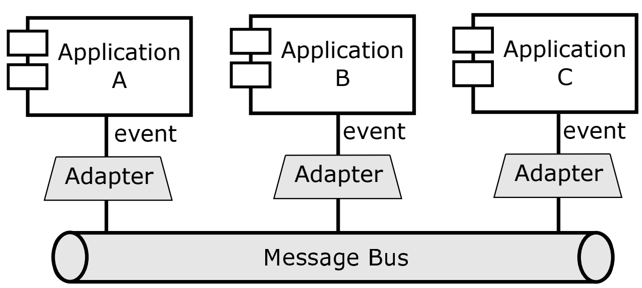

Figure 4. Integration by Messaging.

2.5. Chosen Approach

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

- systems to integrate are not 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.

Taking in mind the advantages and disadvantages of the integration approaches presented in previous sections, and the properties of this 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 [20].

3. Integration Platforms Based on Messaging

Gartner, Inc. is a world-wide enterprise that providers information regarding technology and its usage in the market. Its reports have been driving many technological and business decisions. The Magic Quadrant Report from 2017 [17] ranks many platforms of Enterprise Integration Platform as a Service (iPaaS) according to two parameters: ability to execute and completeness of vision. The former assess the ability of iPaaS providers to deliver platforms that respond to the expectations of software engineers and ensure their integration projects succeed. The later assess the capacity of iPaaS providers to support emerging requirements and lead the market and at the same time grow as a profitable and self-sustaining business. Platforms in this report are also classified into four possible profiles, namely: niche players, visionaires, challengers, and leaders. Niche players are start-ups or small companies in the market only in the past few years but with an excellent technology and very satisfied customers. Visionaires know the specific requirements of iPaaS market and inovate by their delivery models and market strategies. Some of them are in the iPaaS business as being part of a broader cloud strategy (SaaS-centric or PaaS-centric). Challengers have been in the market for several years, having a notable installed bases of thousands of clients. However they have a limited perspective on how iPaaS market will evolve, resulting in more narrowly focused offerings when compared to some of them competitors. Leaders have thousands of paying clients on their iPaaS offering, a solid reputation, a notable market presence and their platforms are well-proven and functionally rich, with regular releases in order to rapidly address this fast-evolving market. Lets introduce briefly the ones classified as "Leaders", Dell Boomi [3], Informatica [6], JitterBit [7], MuleSoft [8], Oracle [9], SnapLogic [12], as well Guaraná [15], which is the integration platform used by the authors in their research groups.

3.1. Dell Boomi

Founded in year 2000 in the U.S.A., Boomi became part of Dell's universe in 2010, operating as an independent unit of Dell. Dell Boomi is introduced as an iPaaS able to support integration application processes between cloud platforms, software-as-a-service applications and on-premise systems. Providing a visual designer with pre-built connectors, users built integration processes by point-and-click, drag-and-drop, without coding. After that they are deployed into a dynamic run-time engine. It provides a centralized management of all the integration solutions, no matter if deployed on cloud or on-premise. Dell Boomi provides a tool for test and watch the process rolling. It provides means to connect well know applications, like Dropbox, Jira and also standard connectors like FTP or HTTP.

3.2. Informatica

Informatica is a private company founded in 1993 in the U.S.A.. Informatica provides a visual designer by drag-and-drop web interface and self-service wizards, connectivity to applications on the cloud, as well as connectivity to on-premise applications and databases. It also provides wizards, pre-configured templates and out-of-the-box mappings. Developers use the design canvas to drag and drop data sources, targets, and advanced transformations. It allows to manage the state of orchestrations and business processes system-to-system interactions, be they synchronous, asynchronous, long or short-running. Informatica provides connectors for well known applications, like Dropbox, Google APIs, Jira, Microsoft Sharepoint, Salesforce, and standard protocols, like FTP, ODBC, REST and also connectors development.

3.3. JitterBit

Jitterbit was founded in 2003 on the U.S.A. JitterBit provides for the design of a graphical interface, allowing the re-use of existing code and business logic, point-and-click connectivity, drag-and-drop configuration, pre-built templates and also infusing any application with artificial intelligence. Deploy of solutions may be done 100% on cloud, on-premise, or hybrid. It is also possible the reuse of any application or code. JitterBit claims that its management data may be moved across applications, real-time analytics with consolidated data, real-time monitoring with alerts and the use of team permissions. It provides connections to many well Known applications, like Dropbox, Gmail, Jira, Microsoft SQL Server and also standard protocols, like FTP, HTTP, ODBC.

3.4. MuleSoft

Mulesoft was founded in 2006 on the U.S.A. With its Open Source tool, MuleSoft combines cloud-hosted and on-premises integration. It enables integration of software as a service and on-premises applications, APIs management, since its creation and publishing, all on a single platform. A repository for connectors, templates and APIs is available and might be enriched by users. Mulesoft provides connectors for well known applications, like Dropbox, Jira, Microsoft Sharepoint, Salesforce, and standard protocols, like FTP, HTTP, JDBC and also connectors development.

3.5. Oracle

Oracle was founded in 1977 on the U.S.A. Oracle integrations are developed by point and click in a browser based visual designer editor, allowing APIs publishing for external consumption. Users have access to connectors to all Oracle SaaS applications subscribed by them, native SaaS adaptors to integrate with other cloud applications, and integration with on-premise applications. It is possible to monitor transactions, key performance indicators, and also to detect and diagnose errors. Customers have access to pre-built integrations to use as-is or to customize and also to a Cloud Marketplace where pre-built adapters and integrations are traded. Developers have access to set of connectors to well-known applications, like GMail, Microsoft SQL Server, SAP R3 and also standard protocols, like FTP, REST.

3.6. SnapLogic

Snaplogic was founded in 2006 on the U.S.A. With its Open Source tool, SnapLogic provides a web-based user interface and a set of adapters, development of integration flows and a set of patterns to be used by integrators, via drag-and-drop. It integrates applications or data in the cloud, on-premise and hybrid. It guarantees data requests delivery and automatically monitors performance and data requests to ensure data delivery as well as compliance with Service Level Agreements (SLAs), companies policies and legal and regulatory requirements. Centralized object level, granular security and permissions enable integration to be extended through out customers organizations. It provides connectors to ERP, CRM, identity management, on-line storage, relational, columnar and key-value databases and standard technologies, like XML, REST, OAuth.

3.7. Guaraná

Guaraná technology arises from the efforts to provide specific tools and notations to reduce design times and implementation of solutions in the field of integration computer systems and business information. It got inspiration from the Model-Driven Engineering discipline [19], 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.

Guaraná Cloud aims 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). After logging in, users get in Guaraná Cloud Dashboard, and IDE where developers may create solutions from drafts or using templates, create templates, set up configuration values, access tutorials, watch for alerts. Solutions development is done in a "drag-and-drop" basis; users pick-up tasks and/or connectors, connect them and set up some values. There can be found connectors to well known applications from everyday work, like Dropbox, Gmail, Jira, Sage One, Salesforce, or generic connectors, like HTTP, FTP. Depending on the connectors and tasks, most of the times those settings are done just by clicking; only in few cases developers may have to type XSLT code or other parameter values. Even in the cases where XSLT code is required, developers may count with an internal tool to help on building the XSLT code. Developers may also count with a visual debugging tool to find out where errors occur or only to check if the solution is running as expected, or they can also look into logs, helpful when they cannot get enough information from the visual debugging tool. Furthermore, users also have access to statistics for better monitoring and log messages from solutions already running, not only on debug mode.

4. Study of a case

In this section we report on a study of a case to demonstrate the use of Guaraná integration platform. We introduce the integration problem to solve, the integration solution modeled using the domain-specific language of Guaraná and its implementation using Guaraná Cloud.

4.1. Problem to solve

Science and research outcomes information management process consists in collecting, structuring, processing and store 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 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 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.

FCT periodically launches contests for projects on R&D in all scientific domains, besides contests in specific scientific areas. Between years 2012 and 2016 there were about

2700 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 [5]. 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 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 [5]. In this chapter FCT DeGóis platform is taken as the reference digital repository and platform of research outcomes information. 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 [14]. 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 DeGóis, there is in Brazil Lattes Platform, which hosts a considerable amount of information about Portuguese researchers.

4.2.1. DeGóis Platform and Curriculum DeGóis

DeGóis platform [2] 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.

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.

4.2.2. ORCID

ORCID [10] 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. 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. 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.

4.2.3. Web of Science

According to their own words, Web of Science [13] 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.

Clarivate Analytics, 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.

4.2.4. Scopus

Scopus [11] 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. 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. 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.

4.2.5. Lattes Platform

The Lattes Platform [1] is the experience of the Brazilian National Council for Scientific and Technological Development (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.

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.

The Directory of Research Groups in Brazil [4] 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.

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

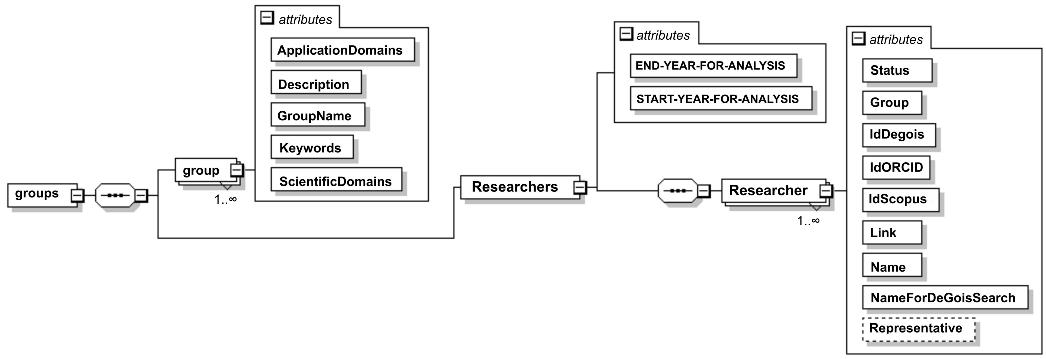

Figure 5. Researchers.xml Schema [20].

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.

For the sake of simplicity and summarized description, without loss of generality, it is assumed in this work the role and perspective of a research unit actor. A research unit 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.

4.3. Integration Solution

This section presents briefly the integration problem addressed in this chapter, 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 [16] 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 Center content management systems is shown.

4.3.1. Software ecosystem

As previously stated in this chapter, 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 Center at Polytechnic Institute of Leiria is taken as the reference case study. The integration solution developed in this chapter will replace the manual process of collecting, computing and updating research unit production, assure that researchers list of publications, participations on scientific events

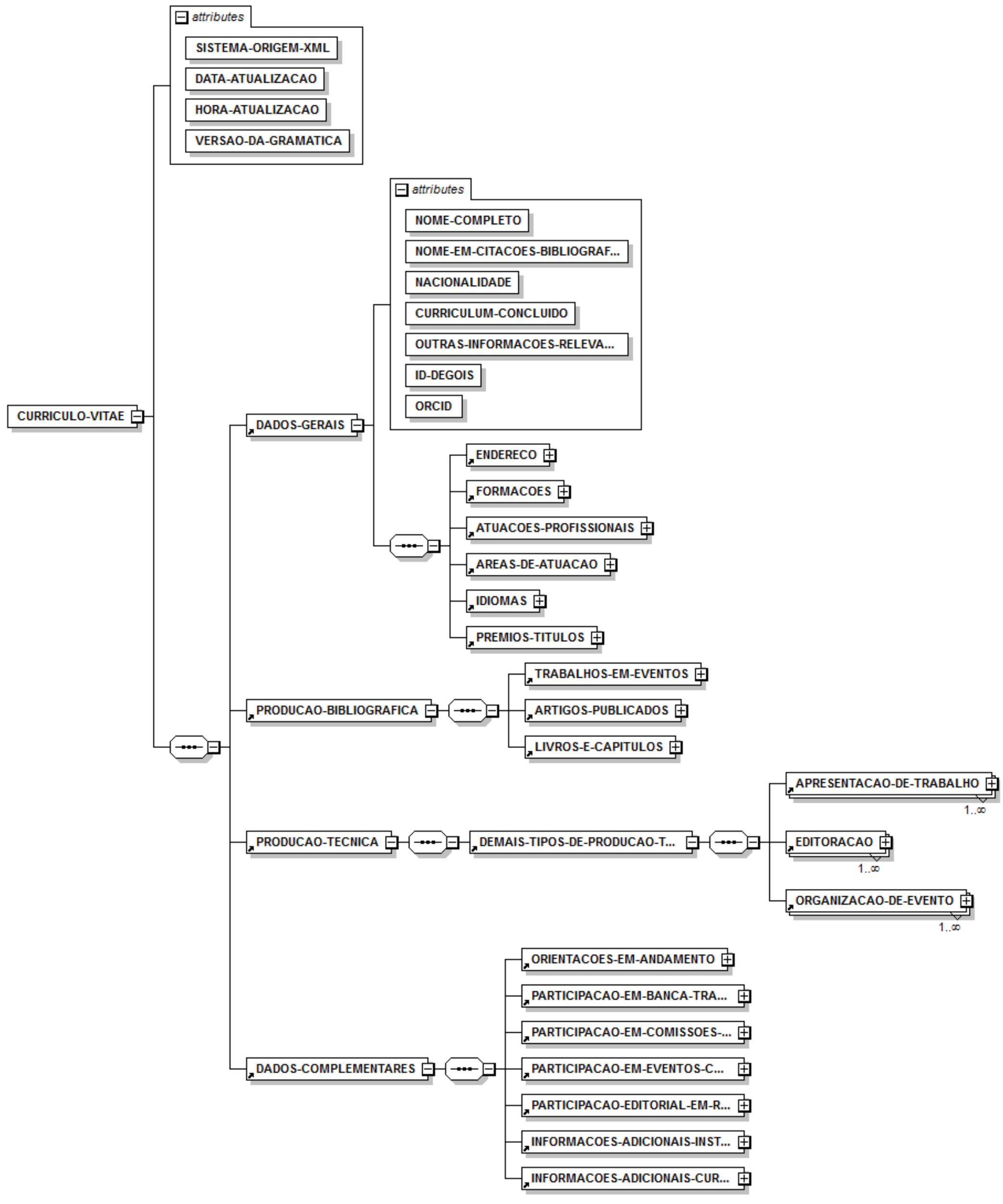

Figure 6. CV XML Schema used by “Plataforma DeGóis” [20].

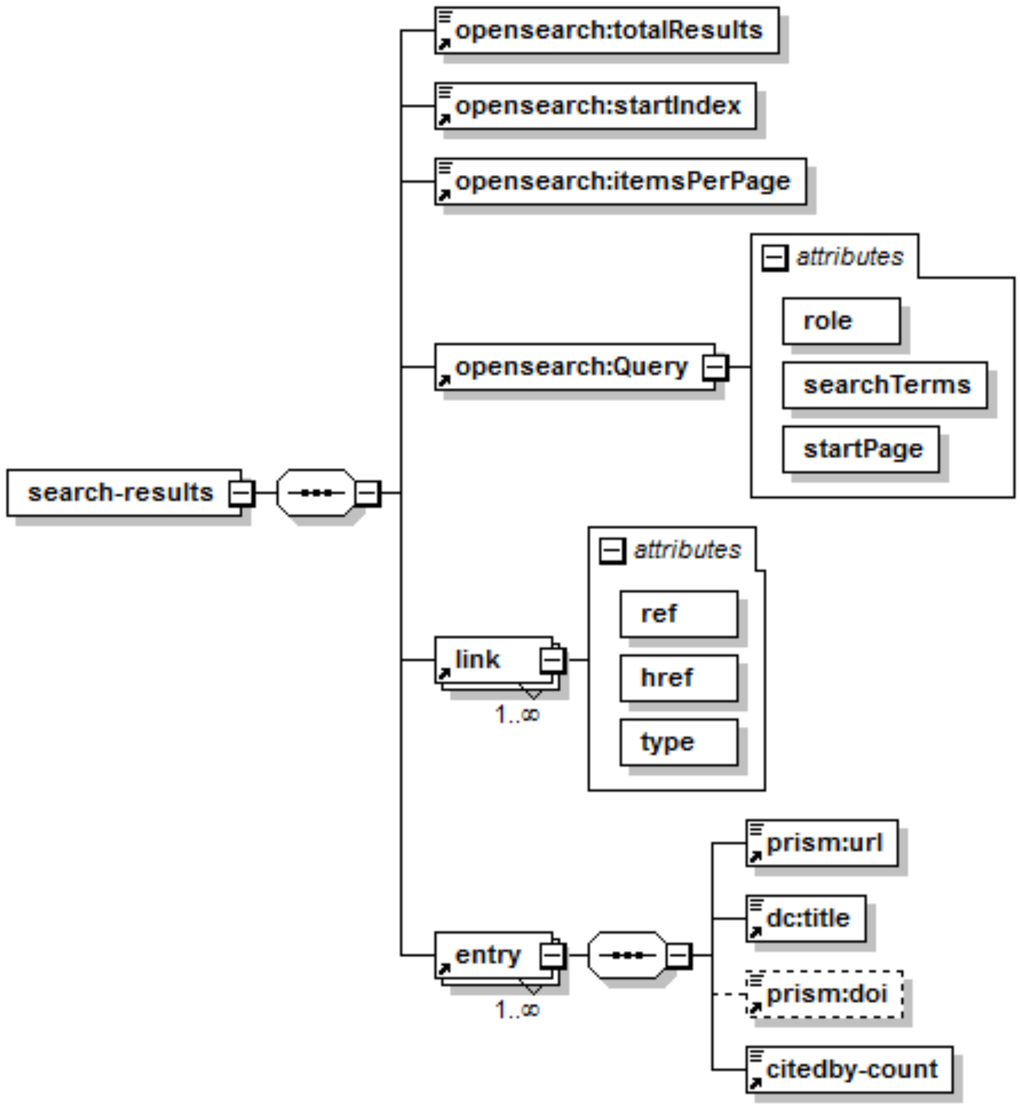

Figure 7. XML Schema used by "Scopus" [20].

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 chapter 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”. 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. Figure 5 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.

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”. Figure 6 shows the researchers CV XML schema used by “Plataforma DeGóis”. RESTful web services request/responses are exchanged between the integration solution and DeGóis platform to search/deliver a researcher CV identified by the `IdDegois` attribute.

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 7.

4.3.2. Conceptual model in Guaraná DSL

The integration solution model specified with Guaran DSL is shown in Figure 8 - 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 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

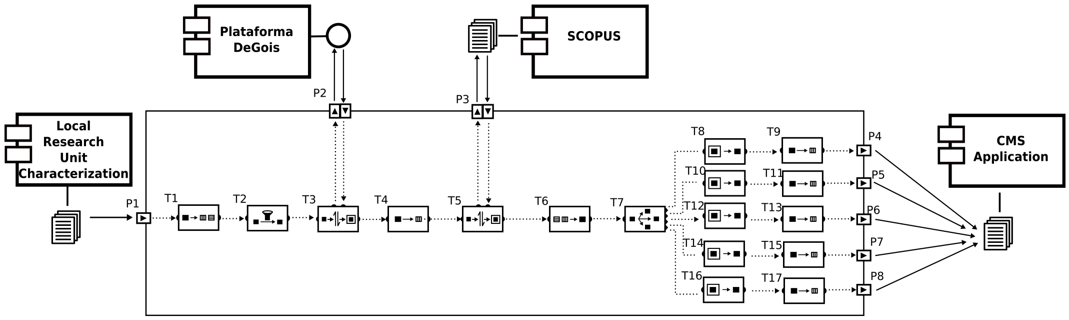

Figure 8. Integration solution specification with Guaraná DSL [20].

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.

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.

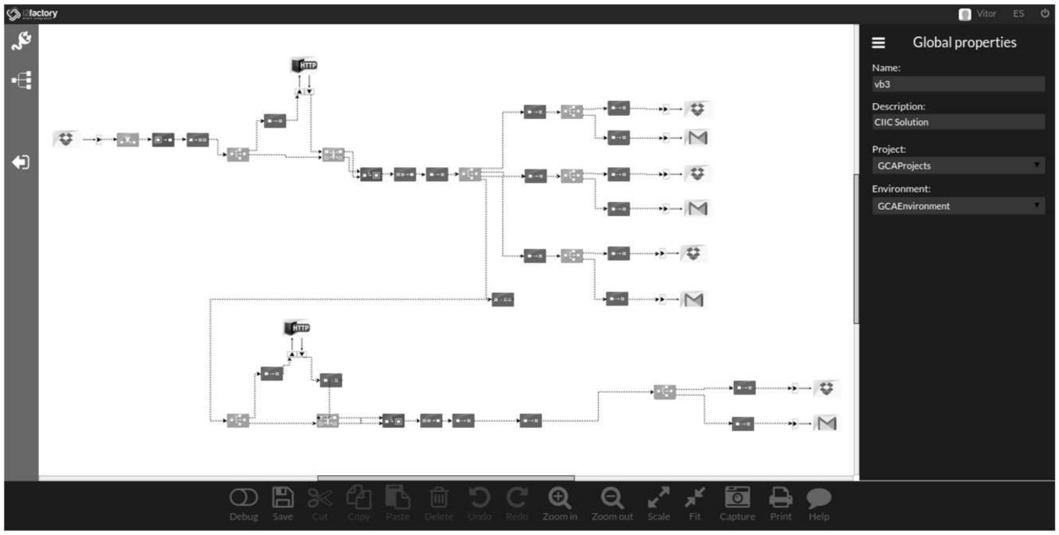

Figure 9. Guaraná Cloud implementation of the integration solution [20].

4.3.3. 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 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 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 9 shows the Guaraná Cloud implementation of the integration solution described above, and Figure 10 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.).

5. Conclusions

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

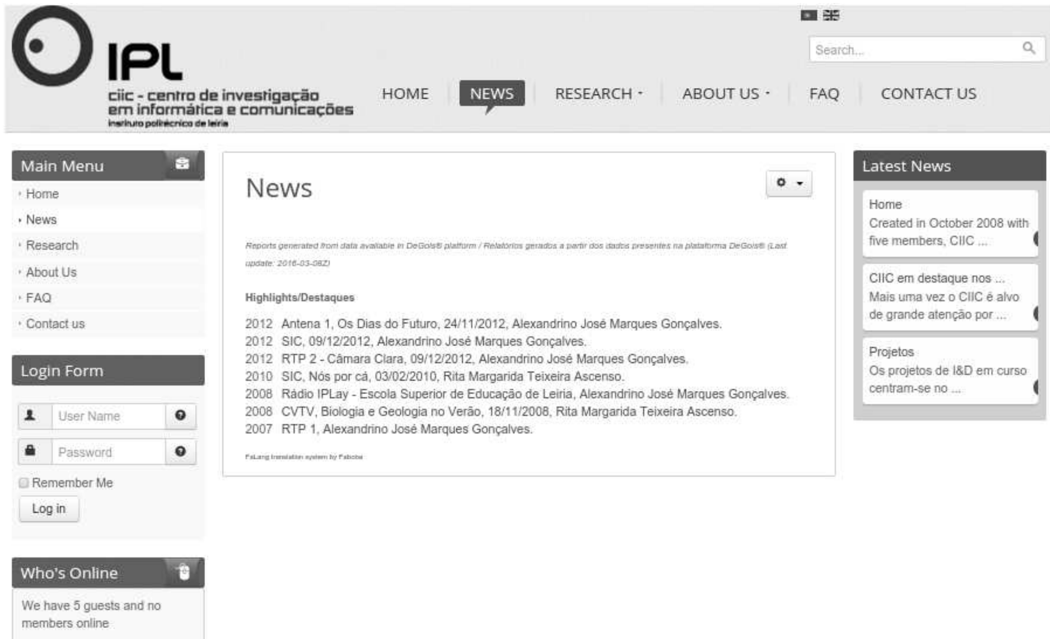

Figure 10. Integration Solution generated report integrated with Joomla CMS [20].

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 [18], which means they support the catalogue of integration patterns documented by these authors and follow the messaging based integration style. In this chapter 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 independence.

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 [18], the Guaraná Cloud is an integrated development environment available on demand in the cloud and with a limited support to the DSL.

In this chapter, 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.

Although Guaraná DSL is a simple and at the same time a rich modeling 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 chapter, 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.

Acknowledgements

Fernandos Rosa Sequeira and Vitor Basto-Fernandes' work was supported by the Portuguese republic national funds through the FCT - Portuguese Foundation for Science and Technology, I.P., under the project UID/CEC/4524/2016. Rafael Z. Frantz's work was supported by Capes postdoctoral programme, grant 88881.119518/2016-01.

References

[1] *Brazilian Lattes Curriculum Platform.* http://lattes.cnpq.br, 2017.

[2] *DeGóis CV Platform.* http://degois.pt, 2017.

[3] *Dell Boomi Home.* https://boomi.com, 2017.

[4] *Directory of Research Groups in Brazil.* http://lattes.cnpq.br/web/dgp, 2017.

[5] *Foundation for Science and Technology Home.* http://www.fct.pt, 2017.

[6] *Informatica Home.* https://www.informatica.com/products/cloud-integration, 2017.

[7] *Jitterbit Home.* http://www.jitterbit.com, 2017.

[8] *Mulesoft Home.* https://www.mulesoft.com/, 2017.

[9] *Oracle Cloud Home.* https://cloud.oracle.com/integration, 2017.

[10] *ORCID Home.* https://orcid.org/about/what-is-orcid/mission, 2017.

[11] *Scopus Database.* https://www.elsevier.com/solutions/scopus, 2017.

[12] *SnapLogic Home.* https://www.snaplogic.com, 2017.

[13] *Web of Science.* http://wokinfo.com/, 2017.

[14] R. C. Filipe. *Development of a platform for the management of research activities and scientific production for research units.* Master's thesis, Polytechnic Institute of Leiria, 2013

[15] R. Z. Frantz and R. Corchuelo. *On the design of a maintainable software development kit to implement integration solutions. The Journal of Systems and Software*, 111(1): 89–104, 2016

[16] R. Z. Frantz, A. M. R. Quintero, and R. Corchuelo. *A domain-specific language to design enterprise application integration solutions. International Journal of Cooperative Information System*, 20(1):143–176, 2011

[17] K. Guttridge, M. Pezzini, E. Golluscio, E. Thoo, K. Iijima, and M. Wilcox. *Magic quadrant for enterprise integration platform as a service.* Technical report, Gartner, 2017

[18] G. Hohpe and B. Woolf. *Enterprise integration patterns: Designing, building, and deploying messaging solutions.* Addison-Wesley, 2003

[19] D. C. Schmidt. *Guest editor's introduction: Model-driven engineering. IEEE Computer*, 39:25–31, 2006

[20] F. R. Sequeira, R. Z. Frantz, I. Yevseyeva, and M. Emmerich. *An EAI based integration solution for science and research outcomes information management. Procedia Computer Science*, 64(1):894–901, 2015