UA CRIS

Universidade de Aveiro
Departamento de Eletrónica, Telecomunicações e Informática

Projeto em Engenharia Informática
Technical Report

Alexandre Abreu, Dário Matos, Samuel Duarte, Pedro Almeida, Pedro Mateus, Renato Valente

Professor José Vieira

UA CRIS

Technical Report for Projeto em Engenharia Informática
Mestrado Integrado em Engenharia de Computadores e Telemática

Authorship of Alexandre Abreu (89290) alexandre.abreu@ua.pt,
Dário Matos (89288) dario.matos@ua.pt,
Samuel Duarte (89222) samuelduarte10@ua.pt,
Pedro Almeida (89205) pedro22@ua.pt,
Pedro Mateus (88858) @pedro.valente@ua.pt,
Renato Valente (89077) renatovalente5@ua.pt

Advised by Professor José Vieira

Keywords

Current research information system

Institutional repository

Information system

Research information management

System interoperability

Abstract

Managing the research work done in a university and its Research Units is often difficult and certainly a challenge faced by many universities, but it is also invaluable not only to the university itself but also to the funding agencies. Having a repository with all the scientific works produced is fundamental and makes the management of these Research Units' data a lot easier.

UA CRIS is a project developed at the University of Aveiro (UA) with the main purpose of aggregating all the research information pertaining the scholar content of UA by synchronizing all research work information in a single place. This system aims to provide an easy solution for researchers and Research Units to register and maintain their research information developed at UA by automatically retrieving the scholar's publications from existing citation databases and allowing future insertion of these publications on UA's Institutional Repository, RIA, an already existing operational and customised instance of DSpace, an open source software widely used for building open digital repositories.

The developed provides synchronization with the Open Researcher and Contributor ID (ORCID) using PTCRISync synchronization framework and its model follows the standards set by the Common European Research Information Format (CERIF) model maintained by euroCRIS.

This report presents a thorough description of this Current Research Information System (CRIS) from its architecture to its functionalities and focuses on the work developed by the authors: the design of the system's core and its main functionalities, database integrations, backend and frontend development, and its implementation. Also, the strategies and tools used to solve some of the problems and issues faced during the system's development life cycle are presented.

Acknowledgements

Developing a CRIS from scratch is a complex task and while our assignment was only to develop part of this system, as bachelor students we were overwhelmed when presented with this task, mostly due to the fact that we had no prior experience with projects of this magnitude, as well as with the complexity of managing research work. Above all, our inexperience with research and development was the first hurdle we had to overcome in the development of this project and as such we would like to express our deepest thanks to our advisor Professor José Vieira for believing in us and providing us the opportunity to participate in this project, for his hints and for all the support and connections provided by him.

Special thanks to Filipe Trancho for helping us throughout the whole project, for his invaluable technical support, for always being available to help us solve any problem we faced and for being the bridge between us and other engineers from the University of Aveiro IT services (sTIC) that contributed to this project.

We are grateful to Luís Carvalheiro and Ruben Silva for their continuous guidance and support through each stage of this project, for their hints and suggestions and for their willingness to give their time and energy for the development of this project.

We also acknowledge and appreciate the feedback, constructive criticism and encouragement offered by our teachers Osvaldo Pacheco, Rui Aguiar and José Rafael during our presentations for the university course Projeto em Engenharia Informática.

We would also like to thank all the other members of the UA CRIS team that participated in some of our weekly meetings and offered us precious advice during the planning of this project.

Thank you to all the sTIC engineers that helped us solve specific problems we faced when trying to implement some of the project's functionalities.

Thank you to all the members from different Research Units who took their time to give us helpful feedback on our prototype.

Last but not least we would like to thank our friends, colleagues and family for being a great source of support during these challenging times.

Table of Contents

Table of Figures

1 Introduction

A Current Research Information System (CRIS) is an information system designed to store and manage data related to research activity from a Research Unit. System interoperability is crucial for a CRIS to work since both the gathering and management of information and research activity heavily rely on the communication and information exchange of multiple different systems.

Currently, the information about the research work carried out at the University of Aveiro is spread out among various information systems, therefore one of the goals of UA CRIS is to aggregate all of this information in one single system that will be available for the existing Research Units to use and integrate in their own information systems.

Developing a complete and functional CRIS is a monumental task, thus the focus of the development of this project is on the core component of this system, the retrieval and synchronization of scientific publications.

UA CRIS is a platform that uses the researchers' ORCID iD to import and save the metadata of their publications so that they can later be submitted to RIA after the researcher curates the publications' metadata. It was built and developed with the objective of synchronizing various platforms with RIA and expediting research management.

This report goes over the work developed by the authors and tries to explain how the system was built and how it works. Taking this into account, the main focus of this report will be on the research done during the planning phase of this project, on the architecture of this system and on how the system and its functionalities were designed and implemented.

The outline of this report is as follows:

In this Chapter we present the motivations that led to the development of this project and its main goals and functionalities.

In Chapter 2 we present some of the research done on some systems similar to the one at hand and entities that are closely linked with the development of a CRIS.

Chapter 3 covers the overall systems' design and architecture outlining the APIs and citation databases used by our system and provides a simplification on how they interact with each other.

In Chapter 4 we go over the requirements analysis of UA CRIS. In this chapter we present the system's actors and use cases and focus on the functional and non-functional requirements of this system.

Chapter 5 describes the development and implementation of the system. First, we present the main tools used to develop this project and then explain in heavy detail the main functionalities developed for this system and how its database was built and implemented.

For Chapter 6 we decided to place emphasis on the project's website that was built to communicate with the academic staff of the course 'Projeto em Engenharia Informática' and display information regarding the development of this project.

Chapter 7 focuses on our work methodology. It mentions the weekly meetings we had to discuss the development of the project, what our first steps were and how we decided to adopt the Scrum framework in order to improve our work methodology.

Chapter 8 provides a summary of the work done and what we were able to achieve in this project.

In Chapter 9 we go over how UA CRIS can still be improved by mentioning some of the functionalities that could be implemented in the future.

2 Current Research Information System – State of the Art

A CRIS system is the ideal bridge between the individual researchers and their research group management. Since most of the work is automated, the researcher can easily store their research work and enjoy the implemented functionalities of the CRIS system (such as an automated CV, etc.) and the research managers gains access to the whole research activity of their institution making it a lot easier to manage it, especially if the CRIS system also provides functionalities that focus on facilitating research management (such as automated reports of the research activity, etc.).

Since a CRIS system provides so many benefits relative to the effort required of its users there are many institutions around the world that desire to have a CRIS system. Due to the usefulness of having a fully functional CRIS system many institutions developed their own CRIS; some with the intention of using it as a commercial platform by providing an implemented CRIS solution to their user groups while others just use it as an in-house system designed to provide a solution to their own problems.

In this chapter we present and provide a small summary on CRIS related entities, citation databases and fully implemented CRIS systems we came across and studied in order to have a better understanding of this project.

2.1 CRIS Related Entities

This subchapter focuses mostly on entities that are directly related to the development of our CRIS system.

2.1.1 euroCRIS

euroCRIS is a European not-for-profit organization that promotes the cooperation of experts in the field of research information and research information systems. euroCRIS acts as a forum for discussion and resolution of all matters related to Current Research Information Systems and aims to be internationally recognized as the point of reference for all matters related to CRIS systems.

To ensure the interoperability of CRIS systems euroCRIS maintains the Common European Research Information Format (CERIF) standard, a data model for implementation of a CRIS. As an international standard for interoperation of a CRIS, CERIF is also an EU recommendation to its member states.

2.1.2 PTCRISync

In Portugal a solution called PTCRISync was developed in order to ensure the creation and sustained development of a national integrated information ecosystem, to support research management according to the best international standards and practices, it is an interesting framework to use in this project because it would remove the need to create two connectors (with ORCID and Ciência Vitae) simplifying the project's architecture.

"PTCRISync relies on ORCID as a central hub for information exchange between the various national systems and international systems." - from https://ptcris.pt/en/hub-ptcris-en/

PTCRISync is open source and is available on GitHub.

2.1.3 ORCID

Due to the names and other personal identifiers of researchers not being unique sometimes it becomes hard to know which authors contributed to which pieces of scientific literature. ORCID uses a non-proprietary unique alphanumeric code as a means of easily identifying scientific and other academic authors and contributors.

ORCID also aggregates and keeps records of the scholarly contributions made by researchers

2.1.4 SCOPUS

SCOPUS is one of the largest abstracts and citation database of peer-reviewed literature; through the SCOPUS API it is possible to retrieve data about published documents, their authors and their respective bibliographic data.

2.1.5 Crossref

Crossref is an organization that exists solely to help to put scholar content in context. By trying to gather and interlink everything in one spot and running an open infrastructure, Crossref vastly improves the system of scholarly communications. As a

DOI registration agency that is a member of the International DOI Foundation, Crossref not only allows its members to register their DOIs but also has available many other tools and services, the most important for the development of a CRIS being their API that allows the retrieval of metadata deposited by Crossref members.

2.1.6 RIA

RIA is the Institutional repository of the University of Aveiro, this information system stores and preserves the research outputs and intellectual property of the researchers at UA. One of the objectives of RIA is to spread the research works produced at UA in order to increase their visibility and impact, to do so they provide open access to these research works when authorized by the authors.

2.2 Existing implemented CRIS systems

Regarding existing CRIS systems, euroCRIS also maintains a Directory of Research Information Systems (DRIS), a directory with some of the existing CRIS systems around the world.

2.2.1 Pure

Pure is a CRIS that is currently used in some universities around the world, it is fully interoperable with ORCID and has already developed some of the features we hope to implement in our CRIS.

2.2.2 Authenticus

Authenticus is a platform developed and employed by the University of Porto that aggregates metadata of scientific publications from various Portuguese institutions. Authenticus is interoperable with other CRIS systems, provides synchronization with ORCID for import and export and has already developed some of the features we intend to implement in our CRIS.

2.2.3 Ciência Vitae

Ciência Vitae is the Portuguese system for science's curricular management. It acts as a repository to gather all the research information regarding Portuguese researchers. This platform was developed according to the international rules and best practices which makes it easily integrated with other systems, and that is why it is so useful.

Ciência Vitae also has an API available for free but restricted to institutional use, this API makes the bridge between the Ciência Vitae platform and external systems.

3 System Modelling and Architecture

3.1 Overview of UA CRIS' architecture

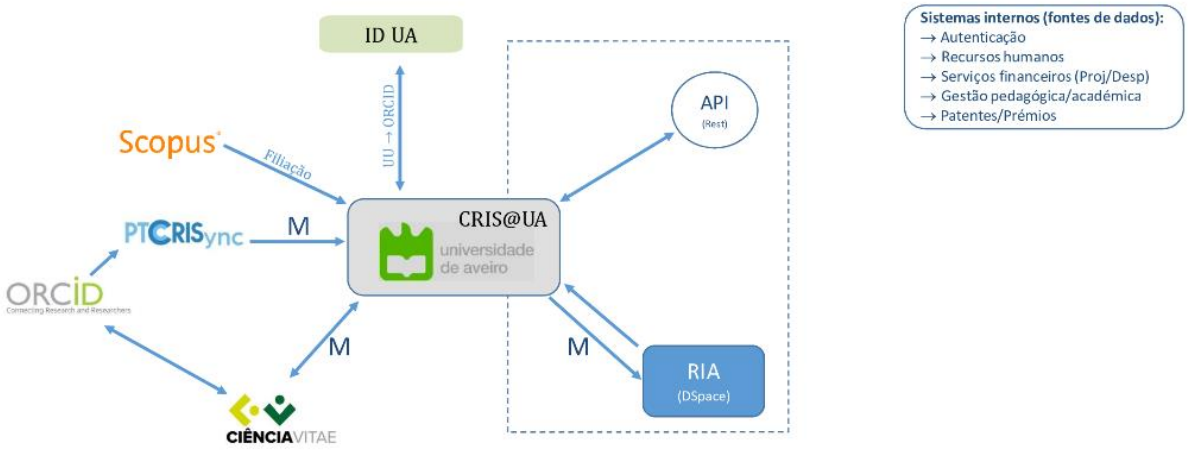

Figure 1 : UA CRIS' architecture

UA CRIS uses ORCID as one of the main sources of data through PTCRISync which also leads to having Ciência Vitae as an implicit source and an affiliation with Scopus.

The system retrieves general user information through the UA's Identity Provider (IDP) and relies on user input in order to get his ORCID iD.

RIA is the other main source of data as well as its destination.

In UA CRIS, the ORCID iD is used by PTCRISync in order to calculate the difference in publications between RIA and ORCID and retrieve the metadata from the resulting publications. As an extra goal there would be another API that would provide the available metadata to external users.

The "M" that can be viewed in some connections stands for manual intervention.

4 Requirement Analysis

4.1 Actors

- **The Researchers**, as one of the key actors of the project, will be one of the main users of the final product, since their work results in various documents to be published in the university repository but also on ORCID and other platforms. The researchers will be able to synchronize already published documents across different platforms while also being able to add more projects in production.
- **The CRIS** system is responsible for automated actions such as saving information to the database or retrieving the user's publications.
- **The Developers** will create, maintain and add features to UA CRIS.

4.2 Use-Cases

- UC 1: The researcher logs in with his UA's credentials.
- UC 2: The researcher inserts his ORCID iD in his profile.
- UC 3: The researcher checks his publications.
- UC 4: The researcher imports his publications that are not in RIA.
- UC 5: The researcher chooses which publications he wants to submit to RIA.
- UC 6: The researcher checks his saved publications.
- UC 7: The researcher checks the details of a publication.
- UC 8: The researcher fills the submission forms of a publication.
- UC 9: The researcher submits a publication, saving all the metadata from the form.
- UC 10: The researcher queries information about UA CRIS.

Figure 2: Researcher's Use Case Diagram

- UC 11: UA CRIS saves the user's ORCID iD.
- UC 12: UA CRIS saves the metadata of all the user's publications from RIA through the user's UA internal iD (IUPI).
- UC 13: UA CRIS calculates the difference of the publications through PTCRISync using the user's ORCID iD.
- UC 14: UA CRIS saves all the metadata of a publication retrieved from PTCRISync in its database.
- UC 15: UA CRIS saves the metadata that user inserts in each page of a form.

4.3 Functional Requirements

Since developing a fully functional CRIS is a very complex task, in the beginning stage our goal was to synchronize RIA with ORCID and create a simple user interface that allows researchers to easily select which articles they want to add to RIA. Ideally, the information system would be running in the background without bothering the user too much. However, that is extremely optimistic. Keeping in mind the goal of

automating the process of synchronization of scientific publications done in the UA, the system should be capable of showing the user's publications that are not yet synchronized with RIA, allowing individual publications to be chosen for synchronization.

In a more advanced stage of the project, synchronization with other platforms such as Ciência Vitae, SCOPUS, CrossRef and others would be ideal. After a successful synchronization of publications, the project could be extended to also synchronize patents, thesis and other scientific outputs.

In order to store all this information in accordance with the rest of the systems implemented in Europe, our database needs to follow the rules and standards set by euroCRIS.

4.4 Non-Functional Requirements

The list below has some properties and characteristics that should be considered when developing the system:

- **Reliability and resilience –** The system deals with pieces of scientific works; we should try to minimize the errors and handle them in a way where no data is lost or altered. Since CRISUA will be hosted on UA facilities' servers its consistency and performance are assured.

- **Security and data integrity –** Due to the nature and sensitivity of the publications we are handling we must assure that the system can't be breached, and that no data should be lost. As a module of the product, the database built specifically for UA CRIS will ensure the security of the users' data, not only protecting it while it is being used but also by releasing the information periodically. There are also additional security measures that will also be put in place the future but are currently out of the scope of our project.

- **Scalability -** The system should be easy to escalate. This means that there should be no conflicts when adding new IUs, researchers and even new platforms to synchronize with. To ensure that doing all of this is easy, the system is going to be developed in modules.

• **Efficiency -** During development there should be a large focus on efficiency and saving server resources simply due to the sheer amount of data that the system will be handling.

• **Maintainability -** The system should be developed while keeping in mind that there will be frequent maintenance in order to meet new requirements and correct some defects. Some more specific maintenance may also be required if the APIs used by system suffer some changes.

• **Usability -** Since the main goal is to automate the process of synchronizing with RIA in order to save time, it is also imperative for the system to be designed in a way that no time will be wasted due to usability problems. UA CRIS will be used by researchers and scholars with varying degrees of experience with this kind of system and as such its processes should be easy to comprehend.

5 Procedure/implementation

5.1 Project overview

The entire project was developed in Microsoft .NET Core (relying on C# as the main programming language) technology using the MVC (Model-View-Controller) software architecture. This technology was chosen by our advisor and his team before the start of this project.

For the development of this project we used Views and the Controllers functionalities of MVC. Both .NET Core and MVC were unknown to us, therefore before we could start the actual project, we had to study these components. At the start of the project we did not find a need to use Models and it was only in the final stages that we understood that Models would have facilitated some aspects of our work, so we implemented this functionality in some aspects we found more important, taking full use of all that MVC has to offer.

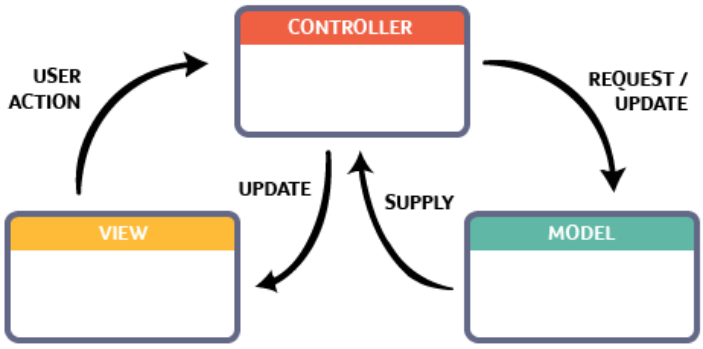

Figura 3 : MVC structure

The retrieval of information from ORCID (through PTCRISync) was the first step on the development of the project. To make this possible we quickly found the need to establish a product that should be available at any time for requests and retrievals, this product was an API to run the PTCRISync code.

As we did not know the resources we had at hand at the time we made the decision to allocate this API in a cloud platform called Heroku. Being dependent on third-party entities is never a good option therefore we acknowledge that this API should be transferred to sTIC.

For a project of this magnitude, a well-defined database, making use of the euroCRIS' CERIF model and some simplifications for our core purposes, was a major requirement. So, it was decided that one team member should have a deeper knowledge about the information stored here.

For a project of this magnitude, having a well-defined database, making use of euroCRIS' CERIF model with some simplifications for our core purpose, was a major requirement. Also, it was decided that one person should have a deeper knowledge about the information stored here. Since the course Base de Dados had begun simultaneously with Projeto em Engenharia Informática a bigger effort had to be done by the chosen person so that this part of project was not a liability. This database is hosted on a server at sTIC facilities.

Since this project was a product that was being made for the university, we were given a template with instructions and premade classes to help us achieve a product that had the university's design, this template can be found in:

http://static.ua.pt/css/semanticua/2.0b/demo/index.html

With the template presented before, previous knowledge in HTML5, JavaScript and CSS and new knowledge in integrating these three technologies with C# we developed the frontend part of the project, making it dynamic not only in the connection with the constructed database but also in the user interaction with the system.

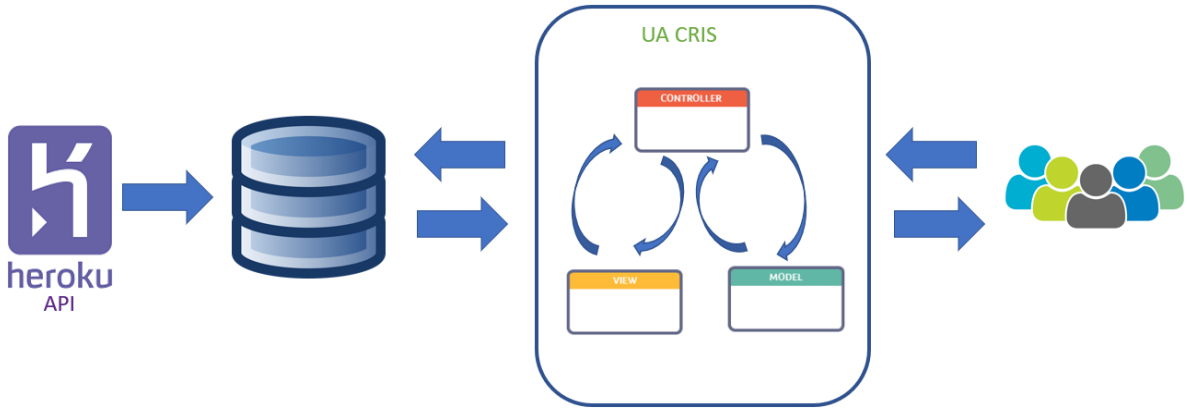

Figure 4 : Project Overview

5.2 Synchronization with ORCID

In order to retrieve publications from ORCID we used PTCRISync and as such an API to do so was built using Java and Maven. These tools were chosen for this because the library used from PTCRISync was also built using Java and Maven.

This PTCRISync library is available on GitHub and the API we built can either use it to fetch all of the publications from ORCID that are associated to a certain ORCID ID or to find the difference in the publications that are currently on ORCID but are not on a local repository. In this case our local repository is RIA, the institutional repository of the University of Aveiro, and this difference in publications is calculated by PTCRISync by comparing the publication's external IDs present in the metadata that was retrieved from ORCID such as their DOI, ISSN or other IDs that are related to the publication's source (AuthenticusID, Scopus EID, WOSUID).

In order to retrieve research works from RIA, an API was made available to us by sTIC. This API can be given the IUPI from a researcher at University of Aveiro in order to retrieve all of the publications stored at RIA associated with that IUPI. At the moment of writing this report the API does not retrieve all of the available information from each publication due to some problems, but it returns the crucial information necessary for this project and although this could not be provided until the deadline of this project, engineers at sTIC are working to try to fix this.

Now that the system has all of the publications from RIA associated with a certain researcher it needs to convert the data schema of these publications, that is currently the one used by RIA, into the data schema used by PTCRISync, that is the same one used by ORCID. This step is important due to the fact that PTCRISync will be comparing the publications retrieved from RIA with the publications associated with that researcher's ORCID ID in ORCID, without converting the data schema used in RIA into the one used in ORCID a comparison wouldn't be possible.

The next step of the system is to send the publications that were retrieved from RIA and had their data schema altered to the API we built so that it can use PTCRISync library to calculate the difference and find and return the publications that are associated with that researcher's ORCID ID on ORCID but are not currently being stored on RIA.

These publications are then displayed on UA CRIS to the researcher to choose which ones he wants to submit to RIA and after he curates the data on these publications, he can submit them to be stored in RIA. From then on these submissions are reviewed by the documental services of University of Aveiro (sbidm, *serviços de biblioteca,informação documental e museologia)* and after they validate these publications and insert them into RIA the process of synchronization with ORCID is over.

5.3 Database

In order to ensure system interoperability, we implemented our database using a subset of the conceptual model of CERIF, created by euroCRIS. Its hosting is provided by sTIC, and only some of their engineers and the UA CRIS group members have permission to access it. This resulted in an aggregate of 41 tables, 18 of them being dynamically updated throughout the development and usage of the web app.

In accordance with the MVC architecture, the interaction with the database was implemented through Entity Framework, a .NET tool that allows for a more object-oriented view of the database, and the usage of LINQ. This is a .NET component that allows for detailed querying and updating, easing the work of implementing SQL code in a C# environment. 18 C# classes representing database tables were created to ensure said interaction, grouped in the Data Models folder. All the queries, insert and update operations were defined via LINQ in separate methods to use in different situations. All these functions were developed in the Services folder, as the DatabaseServices file, to go along with the modular nature of the project.

Being the initial step of using the UA CRIS web app the login, the first change in the database is the addition of a Person, representing the user, linking it with a new person name, containing the full name of the user, and two identifiers, the user's IUPI and ORCID ID. The last one is manually input by the person who logs in, and the remaining mentioned data is provided by University of Aveiro's IDP.

Relatively to the projects' functionalities, the display of the user's publication metadata is a simple query selecting the user's works and their details present on the database, based on the ORCID ID and IUPI. For the result to not be null, the user must first import them. This process starts by adding the publication metadata exclusively contained in RIA, followed by the metadata available exclusively in ORCID.

Since these sources provide different data, the update/insert of the different fields is operated, respectively, as represented below:

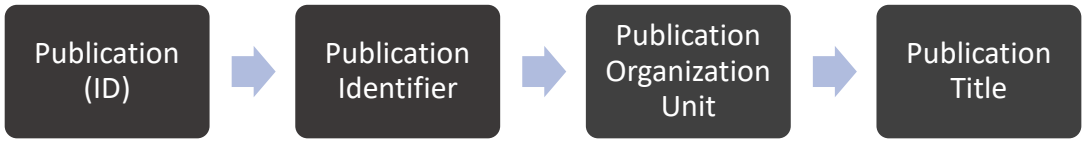

Figure 5 : Insertion of RIA metadata in the database

Figure 6 : Insertion of ORCID metadata in the database

In addition to the insertions, a method was developed to compare each publication's authors with the user who logged in, in order to not duplicate author information in the database caused by multiple name forms (Various ways a name can be displayed) and establish correct connections between authors and publication metadata. This comparison was based on similarity levels based on the presence of common words (names) and initials. The method's result is simulated in the diagram below, with fictional names.

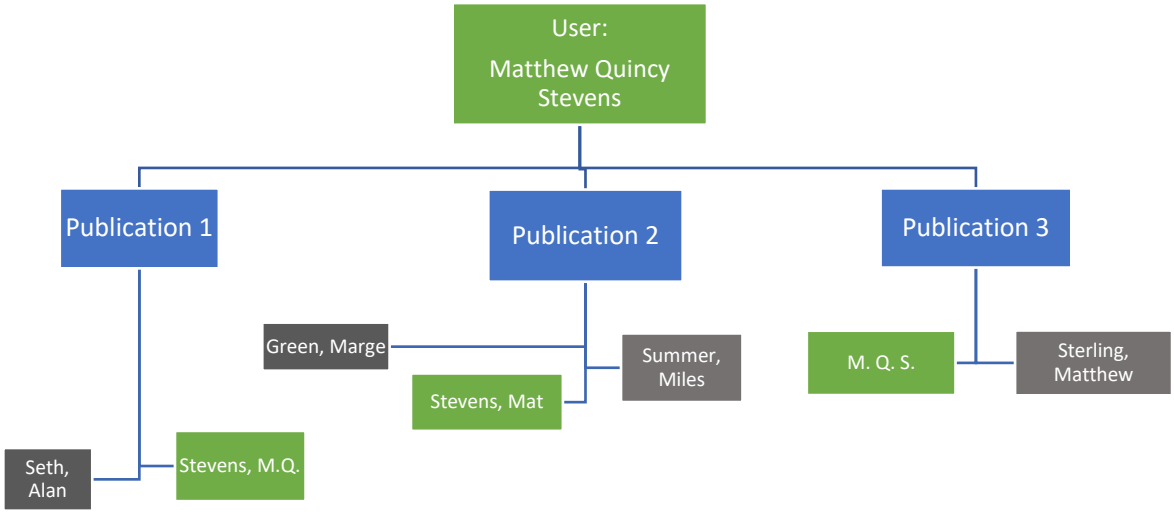

Figure 7 : Author name merging method

Other functionalities of the web app, such as altering a publication's state and metadata, result in multiple updates of the database. The first one occurs when the user decides to start the process of importation to RIA. Next, every page of checking and altering data from a certain publication implies an update in the database to store the changes made by the user, to safeguard the accuracy of the metadata.

5.4 Flexible handling of Publications (States)

When handling the several publications that are passed on from page to page, with the metadata also possibly suffering some changes made by the user, and the display of all the information available there was a need to create different states for our publications. This would simplify not only the user's perception of how he should proceed regarding a certain publication but also our view, as developers.

With this settled we then proceeded to define the states, after some discussion while always being aware that a publication's final state in the scope of this project wouldn't be its actual final state we decided upon these states:

1. **Publication Imported –** The first state of the publication, automatically given by CRIS as soon as it is retrieved from PTCRISync, publications in this state will have a page where they can be selected to move on to the next state.

2. **Publication Saved –** The publication is moved to this state on command from the user. The purpose of this state is to distinguish the publications that the user wants to submit for insertion to RIA, from the ones he doesn't. Publications in this state use the same page as the ones from the previous state.
3. **In review by the user –** Publications in this state have had, at least, the first page of RIA's forms consulted. The purpose of this state is so that the user can tell which publications were visited or edited. Publications in this state had their submission interrupted. These publications have a page in which they can be reviewed.
4. **Publication ready for submission to RIA –** Final state of a publication in the scope of this project, if a publication has reached this state it means the user has filled all the forms that were presented to him, now the publication is no longer editable and is ready to be submitted in RIA. These publications share the page with the publications in the previous state.
5. **Publication submitted to RIA**
6. **Publication validated by the library**

The last two states (States 5 and 6), although not used in the scope of this project, are important and should be taken into account when developing a fully functional CRIS. We still felt the need to delineate them for future work on this project as they will not lead to significant changes on the functionalities already developed.

It should also be noticed that we do not allow for state regression.

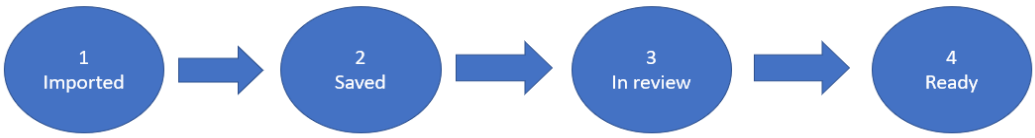

Figure 8 : State Diagram

5.5 Built-in Forms for Publication insertion in RIA

As explained above, all the metadata from the publications retrieved is stored in our database. However, at this point, the publications might not be ready to be submitted to RIA as some metadata might be missing. To solve this issue, the original forms from RIA are displayed so that the user can fill the missing metadata. It is worth

to note that the forms are already loaded with the metadata that is already stored in the database, so the user only has to fill the missing data.

The reason some of the fields of the forms are disabled is that, because we are using a subset of the conceptual model of euroCRIS' CERIF database, there is no way to store that information.

Every time the user advances or goes back in the form's pages, the information in the fields of the form is retrieved and saved to the database. This is done to allow the user to leave the process of submitting a publication to RIA at any moment and not losing the progress.

5.6 UX/UI and Application Flow

To develop the Front-End of the application we started off with the template that was given to us and adapted it to our project. After joining opinions from our advisor and from the team that helped us develop this project, we had a well-formed idea of what we wanted our CRIS to look like and what pages it should have. From this, we developed a Low Fidelity Protype that was then presented not only to the academic staff from the course but also to a selection of researchers (future users) from different Research Units so that they could give guidance and constructive criticism for us to improve the application.

We decided upon the following pages:

- **'Perfil' –** Profile page in which the user can insert his ORCID iD.
- **'Minhas Publicações' -** Here the user can check his publications that are on RIA and the ones we got from ORCID (Pubs in all states).
- **'Importar Publicações' -** Here the user can import from ORCID to CRIS and save the publications he wants (Pubs in state 1 and to state 2).
- **'Publicações Guardadas' –** Here the user can visualize and submit/edit his saved publications (Pubs in state 2,3 and 4).
- **'Forms' -** RIA's forms with pre-loaded metadata that is already in UA CRIS, here the user can edit and submit the metadata of his publications.
- **'Sobre' –** Contextual information on the project.

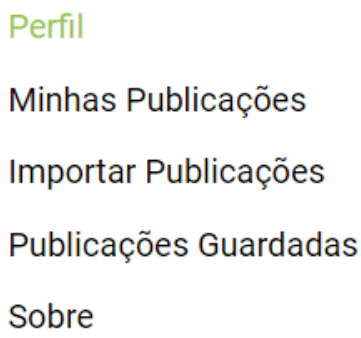

Figure 9 : CRIS' pages

From the advice that we received we implemented ways to filter the information that we present to the user, such as filters by year or state.

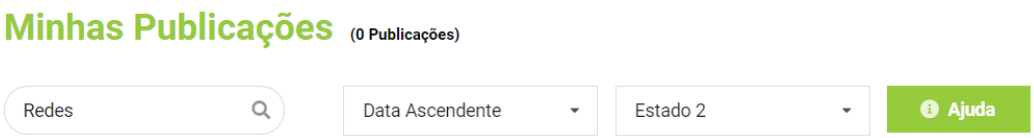

Figure 10 : Information Filtering

Through a simple table the user can select which publication he wants to save in UA CRIS. The importation of publications that are not yet in CRIS can be done in the click of a button.

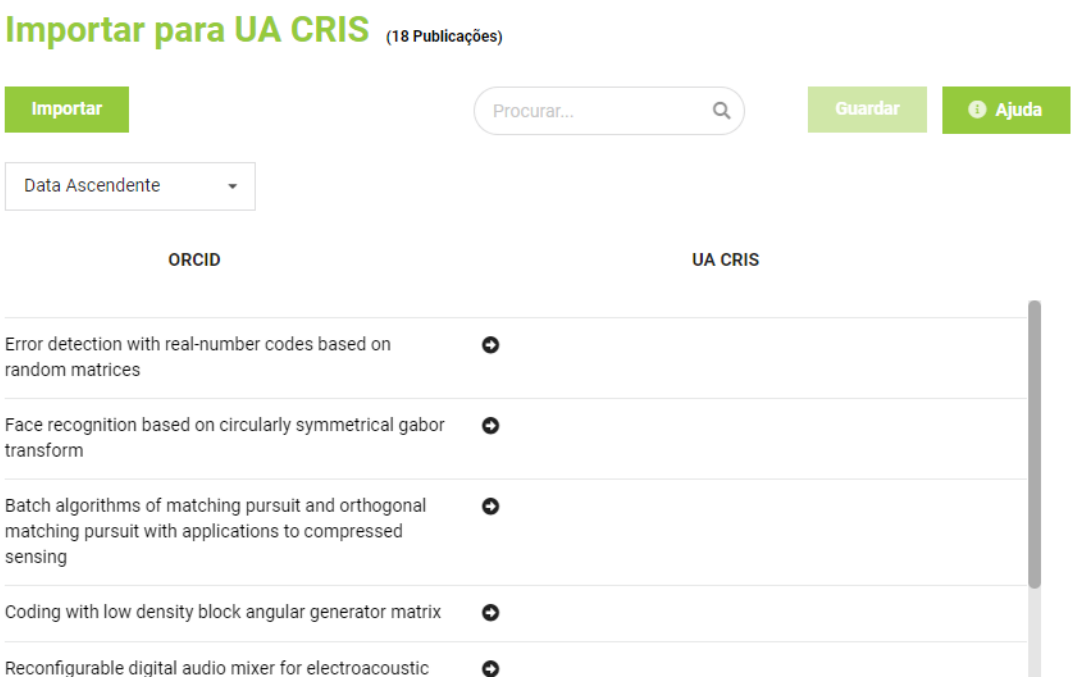

Figure 11 : Page "Importar Publicações"

6 Communication with the Academic Staff

Since this project was developed in the scope of an academic course, it was necessary to have constant communication with the academic staff in order to keep them updated and informed regarding the development of the project. Although we had weekly classes and milestone deliveries, they were not enough to transpire the real state of the project and as such we were asked to develop a simple static website to display information pertaining to updates on this project.

This website features the architectures and deliverables that we developed for the various milestones; some overviews from when some of the entities referred before were studied; a video of a low fidelity prototype that ended up being the overall design for our end product; a proof of concept demonstrating that the API implementing PTCRISync was working and last but not least, a Backlog that substitutes the Gantt Diagram developed in the first milestone; in this backlog it is possible to analyse the task completion from each Sprint since the moment we started implementing a SCRUM methodology.

We acknowledge that this website could have been richer in terms of information for its real purpose, but we hope it has made the work we developed more transparent.

7 Work methodology

In order for this project to be successful it was necessary to have weekly meetings right from the get-go, firstly to understand and narrow down the problem at hand, secondly to plan the design of the system and thirdly to discuss the implementation of the system and eventual problems that arose during this phase of the system's development.

At first we started off with information gathering and every member was given their own task, from studying APIs and citation databases that could be useful, to studying already implemented solutions to this problem (State of the Art), only then were we able to find what we consider the best way to design this system and begin the actual first steps of the implementation.

We did not do this work alone, various researchers gave us their opinion on this problem, using their experience on the topic to guide us regarding what we had to keep in mind when developing this project. One particular example was a meeting we had with the Vice-Rector of University of Porto in which we were shown *Authenticus,* even though this happened in the final stages of the project it gave us a wider view of what a fully functional CRIS would be and definitely showed us that we were on the right track.

During the semester we noticed that our work methodology could be improved so we attended a presentation from an engineer at sTIC about the SCRUM methodology. This topic was not totally new to us but having an introduction to how it is applied in the professional world opened new doors for us regarding how this methodology could be best applied to the development of our project. We started implementing it by dividing the team by roles such as Product Owner, Scrum Master and Development team while having a clear view of our users and by utilizing a Sprint system with tasks that we then used to facilitate the communication with the Academic Staff as stated before, all of this was extremely beneficial to our development as students and future engineers and these are valuable tools that we will take to our future on this field.

8 Conclusion

UA CRIS was designed with the objectives of aggregating all the research work information developed at UA in one system while also providing an easy way for researchers and Research Units to register and maintain this information. By doing this it could also easily reach the goal of simplifying and expediting research management. In order to do so it has been designed and implemented with a focus on efficiency and scalability in order to be able to handle a large amount of concurrent users, flexibility and interoperability to ensure that there would be no problems on the side of Investigation Units when trying to utilize and incorporate this system in conjunction with their own information systems.

Most of the objectives presented at the start of this project were achieved when building this system, the system is able to get all of the publications from a researcher's ORCID ID through an API using PTCRISync and is able to get all of the publications in the local repository, RIA, through an API given to us by sTIC. After this the system can once again use PTCRISync to determine the difference in the retrieved research works and obtain the publications that were retrieved from ORCID but were not found in RIA.

From then on, the system has a list of all the publications that the researcher might be interested in submitting for insertion to RIA and tries to provide not only an easy way for the researcher to choose which publications he wants to submit for insertion but also for the researcher to check and review those publications in order to ensure that the system automatically filled the information fields necessary for submission with the correct information from the available metadata.

The use of the CERIF model was also crucial for the design and implementation of this CRIS and ensures the interoperability of the system.

9 Future Work

Developing a CRIS is a never-ending work, there is always something that can be improved in the system and new functionalities that can be implemented. Some improvements that we were not able to develop but believe that should be the next steps going forward in the development of this system were:

- Creating specific functionalities for the Investigation Unit managers and coordinators, such as being able to access all the publications from said Investigation Unit and also have some statistics regarding those publications and their corresponding researchers. To do this we believe that we should use Elsevier's SciVal since we already have access to its member API, and it is an excellent tool to assess and visualize research performance.
- Gain access to SCOPUS to obtain ORCID iDs from other authors and Investigation Units that are in ORCID and not RIA.
- We are only retrieving information on the publications from PTCRISync and not from RIA, we would like to retrieve information from both sources for a richer CRIS.
- Create an API that allows for external users to retrieve information from CRIS. This was an extra goal of our project, we started developing this API, but ended up running into some problems, so we decided not to present it as part of the project.

10 References

ORCID. Structure of the ORCID Identifier.

https://support.orcid.org/hc/en-us/articles/360006897674

(Last visited 04/06/2020)

Elsevier. Pure

https://www.elsevier.com/solutions/pure

(Last visited 05/06/2020)

ORCID. Crossref.

https://orcid.org/members/001G000001C8dNEIAZ-crossref

(Last visited 05/06/2020)

Crossref. About us.

https://www.crossref.org/about/

(Last visited 05/06/2020)

ORCID. What is ORCID.

https://orcid.org/about/what-is-orcid/mission

(Last visited 05/06/2020)

PTCRISync. PTCRISync.

https://ptcris.pt/en/hub-ptcris-en/

(Last visited 06/06/2020)

FCCN. PTCRISync Overview.

https://github.com/fccn/PTCRISync/wiki/Overview

(Last visited 06/06/2020)

RIA UA. Welcome to RIA.

https://ria.ua.pt/

(Last visited 06/06/2020)

RCAAP. RIA - Repositório Institucional da Universidade de Aveiro.

https://diretorio.rcaap.pt/handle/2/1540

(Last visited 06/06/2020)

Elsevier. SCOPUS.

https://www.elsevier.com/solutions/scopus

(Last visited 06/06/2020)

Ciência Vitae. O que é o CIÊNCIAVITAE?

https://cienciavitae.pt/perguntas-frequentes/

(Last visited 06/06/2020)

Ciência Vitae. API

https://www.cienciavitae.pt/api/

(Last visited 06/06/2020)

Fábio André Domingues. Authenticus: Architecture and Mechanisms to Support a National Repository of Scientific Publications

https://hdl.handle.net/10216/83466

(Last visited 07/06/2020)

euroCRIS. Why does one need a CRIS?

https://www.eurocris.org/why-does-one-need-cris

(Last visited 08/06/2020)

Elsevier. How using a Current Research Information System can support science implementation.

https://www.elsevier.com/connect/how-using-a-current-research-information-system-cris-can-support-open-science-implementation

(Last visited 08/06/2020)

RCAAP. O Sistema DSPACE.

http://projecto.rcaap.pt/formar/mod1/contents/o_sistema_dspace.html

(Last visited 08/06/2020)

euroCRIS. Main features of CERIF.

https://www.eurocris.org/cerif/main-features-cerif

(Last visited 08/06/2020)

euroCRIS. What is euroCRIS.

https://www.eurocris.org/what-eurocris

(Last visited 08/06/2020)

euroCRIS. CERIF in brief.

https://www.eurocris.org/eurocris_archive/cerifsupport.org/cerif-in-brief/index.html

(Last visited 08/06/2020)

euroCRIS. DRIS

https://dspacecris.eurocris.org/cris/explore/dris

(Last visited 08/06/2020)

Elsevier. SciVal – features.

https://www.elsevier.com/solutions/scival

(Last visited 09/06/2020)

Appendix A - Project Website

Check the website we used for the course of “Projeto em Engenharia Informática” for additional information - http://crisua.pt/