UA CRIS

Elaboration phase 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

Índice

System Requirements

1. Requirements gathering and elicitation

At first, we held multiple meetings to evaluate the system's main goals and which of them would be the focus of our project. Through the insights given in meetings by multiple researchers we were able to narrow down the scope of the project and decide which API's and platforms we would need to use as a means of achieving our goals.

From all the available API's we decided that the ones whose use would benefit us the most and whose documentation required more in-depth studying were CienciaVitae, ORCID, SCOPUS, CrossRef and CERIF.

We decided that ORCID would be the best way to identify a researcher, that SCOPUS and CrossRef would be valuable sources of information since they are some of the largest abstract and citation databases of literature and that CERIF would be the ideal data model to follow.

Web of Science was also mentioned as a good source of publications and as something worth considering including in the future.

From previous existing works that already existed in this area the platforms that stood out to us were euroCRIS, PTCRISync and DSpace. We decided to study those platforms in order to better our understanding of the state of the art and technologies related to the system.

Since the main purpose of the project is to synchronize RIA with the databases whose API's were referred above, RIA was also part of the research.

In order to get a better understanding of all these platforms and API's we also conducted several meetings with researchers and engineers from UA that already had previous experience and knowledge regarding the topic at hand.

2. Context

Initially, after the researcher *log-in,* there should be a verification of the already available articles and papers of the respective user in RIA.

After this, a search through the platforms referred before should identify the missing documents and should be presented to the user so that they can be added to RIA according with the users' desire. There is always the case that the user already has every document available, if or when this happens the user should have the option to consult all his documents.

In a final stage of the project an interface to resolve the similar document problem should be developed. With the presentation of the documents side by side and the parts of the text that are similar between documents highlighted the user would choose which part he wants to be used in a final version, this would ultimately result in a unique document that would then be stored in RIA.

Since researchers are not the only stakeholders, both UA and IU can search for documents made by researchers associated with them.

3. State of The Art (SOA)

As mentioned in the Requirements Gathering section, CRIS is a solution that has already been successfully implemented in other universities and IUs.

In this section there will be some small texts regarding some CRIS related work as well as some important platforms that work with CRIS.

- **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.

- **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.

- **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.

- **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

- **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.

- **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.

- **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.

- **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.

- **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.

4. 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.

5. 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 save in RIA
- UC 6: The researches 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 consults information about UA CRIS.

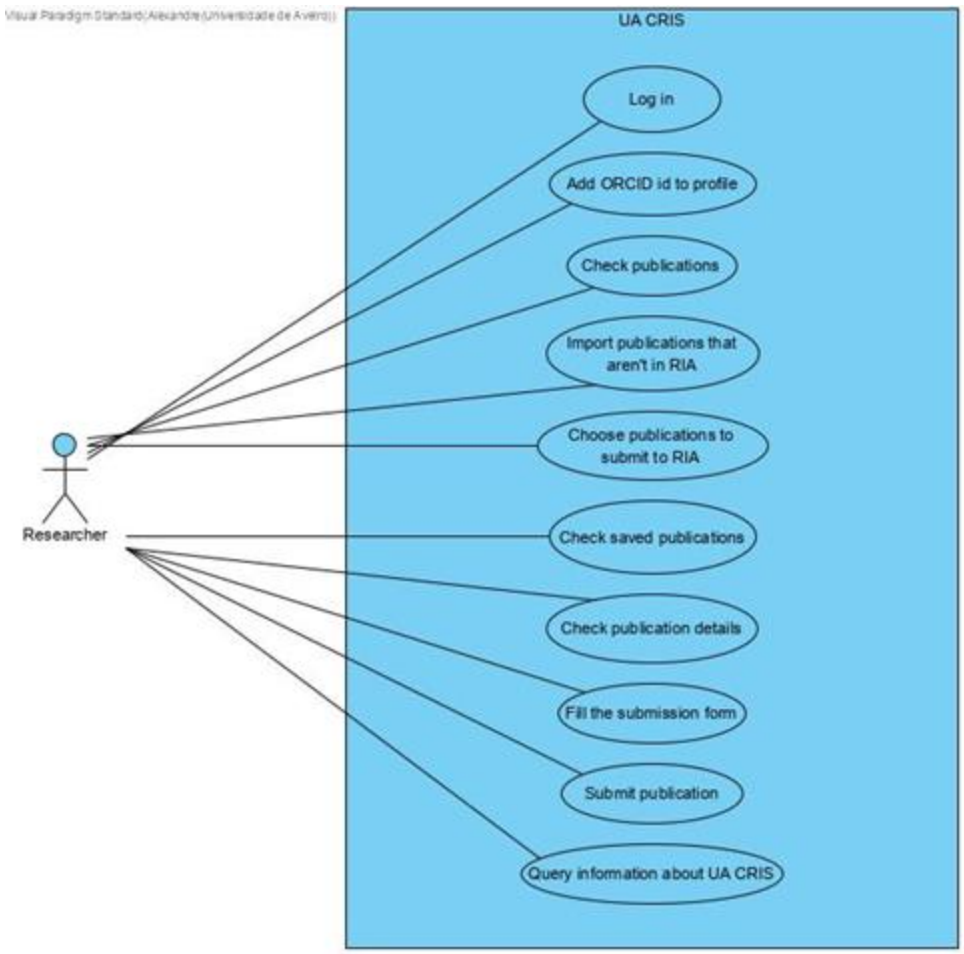

Figure 1: Researcher's Use Case Diagram

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

6. 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.

7. 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.

8. System architecture

This section presents a general overview of the system's domain, physical and technological model.

8.1. Domain Model

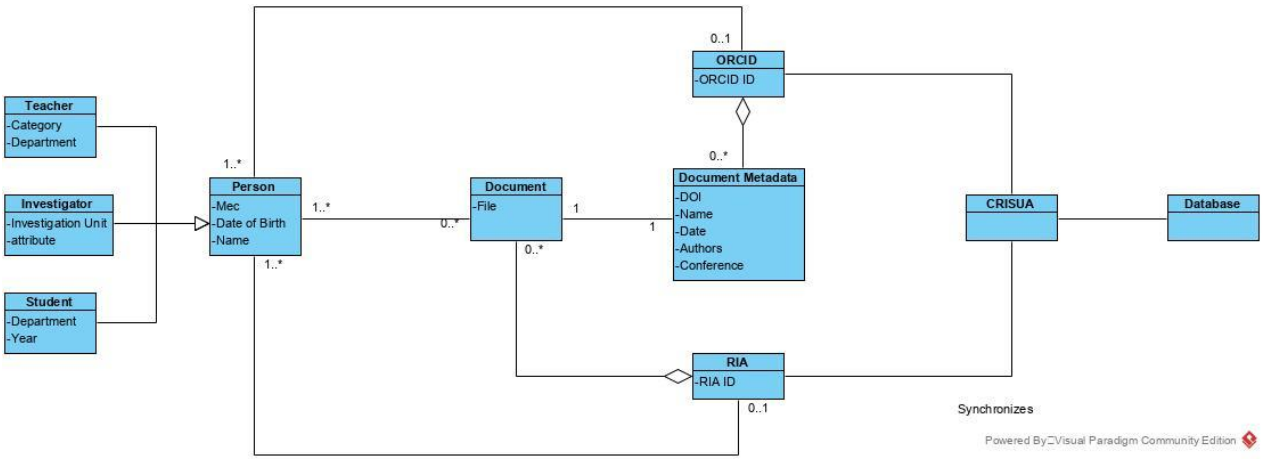

Figure 2: Domain Model

8.2. Physical Model

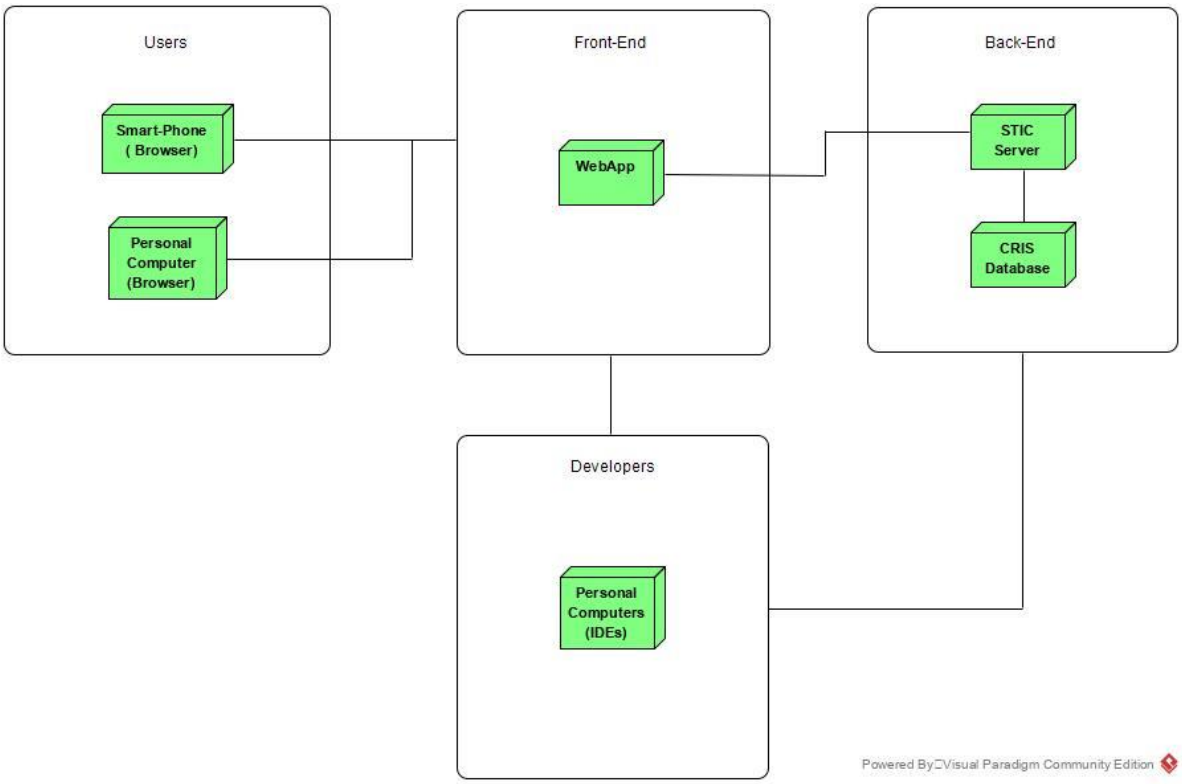

Figure 3: Physical Model

8.3. Technological Model

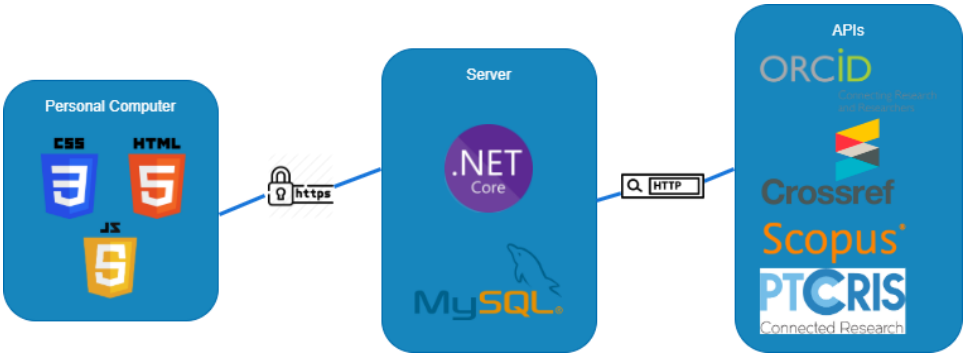

Figure 4: Technological Model

Glossary

- API – Application Programming Interface
- ORCID – Open Researcher and Contributor ID
- SCOPUS – Citation database
- CienciaVitae – Portuguese science curriculum manager
- CrossRef – DOI registration agency
- CRIS – Current Research Information System
- EuroCRIS – European CRIS
- CERIF – Common European Research Information Format
- PTCRISync – Synchronization framework
- DSpace – Used for implementing institutional repositories
- UA – Universidade de Aveiro
- IU – Investigation Unit
- RIA - Repositório Institucional da Universidade de Aveiro

Bibliography

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. Main features of CERIF.
https://www.eurocris.org/cerif/main-features-cerif
(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)