SOFIA UNIVERSITY

"ST. KLIMENT OHRIDSKI"

FACULTY OF MATHEMATICS AND INFORMATICS

RESEARCH INFORMATION SYSTEM SERVICES FOR THE OPEN SCIENCE CLOUD

DISSERTATION

Created for the for the purpose of acquiring the educational and science degree "Doctor"

in the field of 4.6 Informatics and Computer Sciences

in Doctoral Program "Information Technologies"- Computer Networking

Student:

Silvester Hasani

Supervisor:

Prof. Dr. Eliza Stefanova

Abstract

Open Science is the trend that aims to make scientific research and its dissemination available to all levels of the society, amateur or professional. It is considered the future of Scientific Research, and all the research institutions are transitioning towards Open Science.

In Europe, European Open Science Cloud is fostering the transition towards it by creating a set of rules and guidelines to be followed to make Research Data accessible by all EU researchers through interoperable FAIR data. From all the European Countries transitioning towards Open Science, Balkan countries are newly joining the transition. This is especially true for Albanian Institutions.

In this dissertation, Open Science is analyzed in detail. Starting from what is Open Science and moving on to Open Access, all key components of Open Science are explained. Open Data is also described and compared to FAIR data. Open Source is also described and the concept of Digital Repositories is detailed, as a key element in storing research data. Storing research data is an important task for Open Science.

Next, the European view on Open Science is introduced with the European Open Science Cloud (EOSC), moving to the most used Research Data Digital Repositories. EOSC defines the guidelines to be followed in Europe to be compliant for future collaboration in Open Science, thus a complete understanding of EOSC is necessary to continue the work. After that, the current state of Open Science in Balkan countries is captured, focusing on identifying the stages in the transition to Open Science and main problems faced by Balkan universities. To do so, a questionnaire is distributed to all relevant institutions and analyzed in detail. Key findings are found and the next steps are planned.

The careful analysis illustrates the need for the Albanian Open Science Cloud (AOSC). The goal of AOSC is to make Albanian Science open and to help Albania join the EOSC initiative. A prototype deployed in Albania is presented, following the European standards set by European Open Science Cloud. Albanian-CRIS is the repository that will help the build of Albanian Open Science

Cloud, to follow the transition to Open Science, and to join EOSC. The data structure of the repository is illustrated.

The analysis also indicates a need for an Open Science Policy to be implemented in Albanian Universities. The policy is presented and is taken into consideration by Albanian Research Institutions for implementation. The Open Access Mandate aspires to an Open Science transition and considers it as a critical component in enhancing the relevance of research on the Albanian community. The intention of Albania's first Open-Science Policy and Open Access Digital Repository is to make research data FAIR and make knowledge publicly open to all Albanian researchers.

In conclusion, this dissertation describes the detailed transition of Albanian Universities into Open Science and the next steps taken to foster the future of the research.

Acknowledgments

First of all, I would like to express my gratitude to my supervisor Prof. Dr. Eliza Stefanova for the extensive guidance and encouragement throughout the dissertation preparation and studies at Sofia University “St. Kliment Ohridski”. I am also thankful to Prof. Dr. Krassen Stefanov and Atanas Georgiev for guiding me through the dissertation as much as possible. I thank them for the valuable academic growth.

I owe a deep appreciation to people close to me for the inspiration they have given to me.

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

In this chapter, Open Science is introduced, followed by Open Access and Open Data, as the main elements of Open Science. General definitions are given for the most important topics related to Open Science and an in-depth overview of each element is discussed. A general understanding of Open Science and its related elements is given.

1.1. Open Science

The Open Science trend aims to "make scientific research (including publications, data, physical samples, and software) and its distribution freely accessible to everyone" [1]. Open Science is considered also Science that is transparent and accessible knowledge that is shared and developed through collaborative networks [2]. Another popular definition of Open Science is that "it encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives" [3]. Overall, the goal of Open Science is to provide free access to knowledge, by fulfilling its principles. Open Science consists of six principles which will be analyzed through the chapter: Open Data, Open Source, Open Methodology, Open Peer Review, Open Access, and Open Educational Resources. Through the years several advantages and disadvantages of Open Science have been apparent as shown in [4].

1.1.1. Advantages of Open Science

Advantages of Open Science include:

- Rigorous peer-review - since Open access publication of research papers and data is available, other scientists can provide much better peer-review related to the topic.
- Publicly funded science is publicly available – Open Science requires that all the research which are publicly funded should be open to the public.
- More transparent, reproducible and impactful science – Since data is transparent and everyone can access this information, scientific collaborations are much more possible.
- Researchers have better outreach towards science – Since access to information is open to everyone, researchers have access to scientific data regardless of their economic situation.
- Better social outreach – Since data is accessible to everyone, future citation possibilities are enhanced by other scientists accessing your research.

1.1.2. Disadvantages of Open Science

Disadvantages of Open Science include:

- Potential misuse – One good example is research that contains new security protocols. Since the protocol is going to be publicly available, there are higher chances of misuse.
- General misunderstanding by the public – People find it difficult to understand which are the steps needed to have an open-access research publication.
- Low-quality science - Since the major concern is just to have data open, sometimes this leads to low-quality science. In fact, this is the biggest disadvantage of Open Science which is analyzed in the next section related to quality challenges of Open Science.
- On some occasions, legal issues – Open science changes the way we understand ownership of research data, its privacy and the copyright. This leads to misunderstandings related to giving the copyright permissions to the publisher or not. This is analyzed in the next section related to the researcher's challenges in Open Science.

1.1.3. Quality Challenges of Open Science

As previously mentioned, the biggest disadvantage of Open Science is the lead to low-quality research data. When publishing in Open-Access journals, it is important to understand if the journal is of quality or not. Some Journals, present themselves as open-access ones, but in fact they are “predatory journals” which look after publishing costs to earn money. This has been a problem since the beginning of Open Science and has given a bad reputation in the past [5]. It is important for researchers to investigate the quality of the journals that they are trying to publish their papers in. Some points a researcher can investigate:

- Does the Journal clearly state which are the Authors Publishing Costs (APC) on their website?
- Do all the articles have a Digital Object Identifier (DOI)?
- Do all the articles correspond to the main subject that is stated on their website?
- Is the editorial board with researchers from the field of the subject?
- Does the journal clearly state the re-use Rights (i.e., Creative Commons)?
- Does the journal have an ISSN (International Standard Serial Number)?

These questions should clearly understand if the journal is a real Open-Access journal and not a “predatory journal”. Sometimes, researchers check the impact factor of the journal, but this might not be beneficial with Open-Access ones, since most of the time they are new journals and have not yet created a well-established name. However, the Directory of Open Access Journals (DOAJ) does quality assurance for Open Access Journals, and only lists the ones that guarantee a quality peer-review process in the articles [6].

1.1.4. Researcher’s Challenges in Open Science

When a researcher wants to publish in Open Science, there are a variety of challenges to be confronted with:

- There is a lack of awareness of the benefits of Open research data. Sometimes, researchers are happy with the current workflow of publishing their research and don’t want to change it. There is a lack of reward system towards Open Science for them, leading to even further reluctance to change.
- Repositories to support Open Science are still a work in progress for the majority of the research institutions. This means that the researchers don’t have a place where to store their research data.
- There is a need for policies implemented into Research Institutions to support and comply to Open Science. This ensures that a rewarding system for using Open Science would be in place, and researchers would be much more inclined in embracing Open Science.
- Researchers might have a financial and career challenge with APCs, for gold and hybrid Open access publishing. They are not rewarded for their work as much as they are usually rewarded in the traditional way of research publishing.

1.1.5. Open Science overview in Europe

In EOCD, Open Science is extended into opening all the research cycle as shown in Figure 1-a [7].

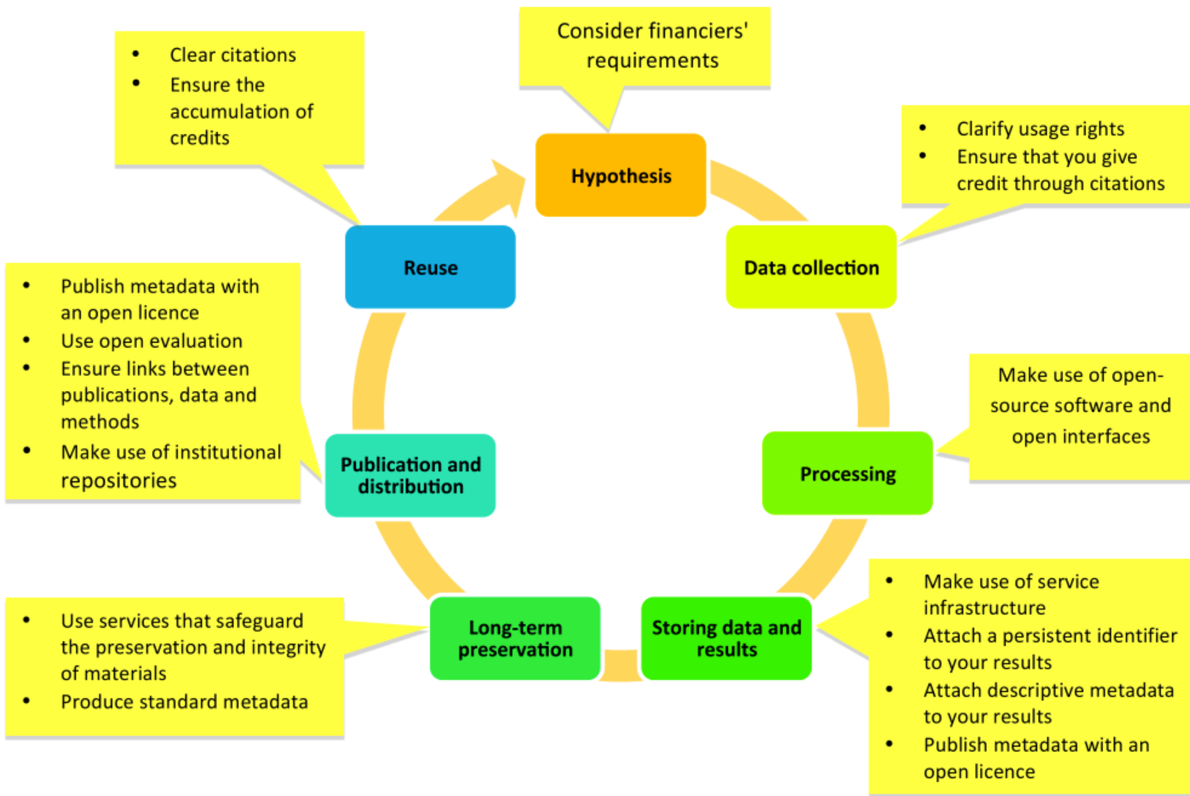

Figure 1-a: Open Science Research Cycle

OpenAIRE is a European project supporting Open Science [8]. It is a network of dedicated Open Science experts promoting and providing training on Open Science. OpenAIRE is a technical infrastructure harvesting research output from connected data providers. OpenAIRE aims to establish an open and sustainable scholarly communication infrastructure responsible for the overall management, analysis, manipulation, provision, monitoring, and cross-linking of all research outcomes. By implementing Open Science policies, OpenAIRE targets essential components and concerns of Open Science in Europe by supporting, accelerating, and monitoring the viability of Open Access to publications and research data. It contributes to the European

Open Science Cloud (EOSC) by providing a list of services that are interoperable with EOSC services, allowing for greater interaction with European Research Infrastructures.

This combination of knowledge and European Research Information platform enables services to researchers, research support organizations, funders, and content providers tools to perform Open Science such as:

- Integrated scientific information: links publication, project information, datasets and presents them in one place.
- Monitor and reporting on Open Science research outcomes for funders.
- Training sessions and support on all subjects related to Open Science and Open Access.
- Discovery of Open Science output per project, funder, or data provided.
- Exchange of metadata and content amongst data providers
- A general-purpose repository.
- An Open Science helpdesk.
- Monitoring and reporting mechanisms for research output per institution.
- Analyses massive collections of documents, related meta-data, and relational information.

As part of the Open Science movement, Open Access (OA) and Research Data Management (RDM) are also very important topics to understand. Open data, often known as open access to research data, is a vital step toward making science fully transparent and open. As research becomes more data-driven, scientific advancement becomes increasingly dependent on data availability. Researchers can use available information in new and complementary ways thanks to open data policies [4]. In the next sections, Open Access and Open Data are explained.

1.2. Open Access

As mentioned, Open Science relies upon Open Access. The Open Access definition is “a set of principles and practices for disseminating research findings online, free of charge or other restrictions”. Applying an open license for copyright eliminates restrictions to copying or reuse with open access [9]. It is also an international movement whose aim is to provide free and open online access to academic content including articles and research data. While publications can also be published differently, Open-Access publications are defined as publications that have no financial, legal, or technical barriers to accessing them. This means that anyone can read, download, copy, distribute, print, search for and search within the information, or use it in education or in any other way within the legal agreements.

1.2.1. Open Access Publications

When related to Research, Open Access is “a publishing model for scholarly communication that makes research information available to readers at no cost, as opposed to the traditional subscription model in which readers have access to scholarly information by paying a subscription” [5].

An open-access mandate is a policy adopted by a research institution, research funder, or government that requires all researchers part of the organization to make their published, peer-reviewed journal articles and conference papers open access by using two ways. The first way is by self-archiving their final, peer-reviewed drafts in a freely accessible institutional repository or disciplinary repository which is called the Green Open Access method and the second way is by publishing them in an open-access journal which is called the Gold Open Access way. Sometimes both ways are used. To have Open Access publications, four methods are used. They are called routes of publishing Open Access.

- Golden Route: This method of publication is with Full Open Access Journals, which means that publication is done via a publisher platform to a fully Open-Access journal. Most of the time, this route has a publication cost known as article processing charges (APC) which are either covered by the authors or by their institutions.
- Green Route: This method of publication is via a publicly accessible database, part of a trusted repository managed by a research organization. The full text of the academic publication is deposited on this repository.
- Diamond Route: This method of publication is via diamond journals, which are fully open access journals, but the difference from the golden route is the fact that no article processing charges are paid. These journals are funded by library subsidy models, institutions, or societies.
- Hybrid Route: This method of publication is via subscription journals that allow open access to publications of articles after paying the APCs.

Institutions that support Open Science and Open Access have an open-access mandate. This mandate is a policy adopted by the research institution to make their published, peer-reviewed journal articles and conference papers in an open access method by self-archiving their final, peer-reviewed drafts in a freely accessible institutional repository or disciplinary repository (Green Route) or in an open-access journal (Gold Route).

1.2.3. Advantages of Open-Access Publications

Some of the advantages of publishing research as Open access are:

- Faster and bigger dissemination of research data.
- People that cannot afford the subscription to journals, can read the research for free. Usually, if they find it interesting it increases the number of citations.

- New knowledge and ideas are disseminated faster. This is especially important in ICT, where technology is changing at the rate of months.
- Companies have more access to new ideas, which leads to more products that follow the latest standards and boosts the economy.
- Re-use possibility brings the knowledge to everyone as an “open-educational resource”.

1.2.4. Disadvantages of Open-Access Publications

Some of the disadvantages of publishing research as Open access are:

- Some fields, especially science, rate their research abilities by the impact factor of the journals which research data is published. However, Open access journals are relatively new which usually means their impact factor is lower than traditional journals.
- Some fields don’t have many Open access journals available for them to publish. This restricts their publishing options.
- Research institutions have not yet made payment regulations related to APCs. This means that researchers have a higher cost when publishing on Open access journals since they will pay the APC on their own.
- There are a lot of “predator” journals that spam researchers to publish their papers there. It takes some time and experience to distinguish genuine Open access journals.
- Researchers have extra work to supply research data to repositories. Besides that, there are sometimes some copyright issues risen.

1.2.5. Misconceptions of Open-Access Publications

There are some misconceptions related to Open access publishing. To clarify that, it is important for researchers to understand that as long as a genuine journal is chosen:

- The research publications are always going to be peer-reviewed.
- The research publications are going to be indexed.
- The research publications are going to have an impact factor.

There is also a clear distinction for researchers when choosing the publishing method. If the researcher decides to go with the Golden route, he has to pay the APC cost, which sometimes is very expensive. As explained before, the author retains the copyright, and the journal archives the research data into its repository before publishing it. For the moment most institutions don't offer funding, so the cost is totally covered by the author. If the researcher decides to go with the green route, he has to self-archive the article, and he doesn't pay the APC cost. However, self-archiving must be agreed upon with the publisher, and sometimes the publisher has some strict rules related to publishing the paper. On some occasions, the publisher also requests the copyright of the work, which might be inconvenient for fostering Open Science.

1.2.6. Open-Access Platforms

The European Commission set a contract for an Open Access platform for all the Horizon 2020 participants as a free service which started at early 2021 [10].

The system provides all Horizon 2020 research with a peer-reviewed publishing service with an open-access repository and no article costs. Participants are able to use the platform to fulfill the open access requirements of Horizon 2020 and its successor, Horizon Europe initiative. In the future, it will also offer open access publishing as the preferred method of research publications.

The system also adheres to the highest scientific and publishing standards, with by having a Scientific Advisory Board overseeing the publication of the research. It manages the whole publication process, including open peer review, post-publication archiving, and the preservation of the publication [10].

All the previous publications from Horizon 2020 and Horizon Europe-funded research, regardless of the field of research, are also eligible for publication on the platform. As in autumn 2020, the portal allows submissions for Horizon 2020-funded articles [10].

1.3. Open Data

Open Data is also essential to Open Science. “Data that can be freely used, updated, and shared by anyone for any purpose” is defined as open data. [11]. When this data is useful, usable, and used, Open Data is considered Open Knowledge. The main properties of Open Data are:

- The whole data must be available at a fair replication cost, preferably downloadable online. The information must also be in a usable and editable format.
- The data should allow reuse and redistribution, as well as processing with other datasets.
- Everyone should have access to the data, regardless of the research field or group.

There are many reasons for data to be Open however according to Open Knowledge Foundation the main reason are [12]:

- Transparency: Data should be open access for everyone, should be sharable to everyone, and should be reusable in any form they seek to.
- Have social and economic value: Opening data can aid in the development of new enterprises and services both socially and economically.
- Motivate engagement: Not only data should be transparent but also should motivate government, business, and organizations to open data to everyone. This way better decision-making can be made.

On the other hand, data should not only be open but also reusable which means that some standards should be used. An open data standard is a set of specifications and requirements for how some sets of data should be made publicly available. These standards should be open to everyone to contribute and are consisted of standards for:

- Schematic: Schematic standards define the structure of the data to be published. This includes the names, descriptions, and data types of data fields or columns. Schematic standards also may include how one dataset is related to another.
- Semantic: Semantic standards define the terminology or language in the data which is published.
- Atomic: Atomic standards define how basic elements of data must be represented when there is an opportunity for confusion. Atomic standards may represent individual data values or a combination of data values.

Open Data can be also used by the governments. Government can benefit by having open discussions with better decision-making, and the development of innovative new services just by opening the data to the general public. This is also called Open Government and is focused on allowing anybody to freely use, reuse, and redistribute government data [13]. It is essentially the same as Open Data, that Government Data should be freely accessible, used, and re-published by anyone. This perspective transforms into a mechanism that gives people the right to information about the changes into their government. It allows them to be aware of all their government decisions and prevents the governments from becoming data monopolies. To do so, people must have unrestricted access to government data and information, as well as the ability to analyze and share it with other citizens.

The doctrine of Open Government believes that citizens have the right to see papers and procedures so that effective overview of the general public may be carried out. It eliminates the possibility of government secrets. The beginnings of Open Government debates can be traced back to the European Age of Enlightenment, when philosophers discussed the best way to build a democratic society [13].

There are three main reasons why government data should be open [12]:

1. Transparency - Citizens must be aware of their government decisions. Having unrestricted access to this data, as well as the possibility to share it with others, makes the data much easier to analyze and visualize. Transparent data means that data can be properly shared, utilized and visualized.
2. Having social and economic value – As previously mentioned, having open data can aid the development of new enterprises and services both socially and economically. This is especially true with government data, which have a broader social impact when taking decisions.
3. • Motivate engagement – Normally when the government data is not open, citizens engage very little with the government, mainly during elections. Citizens may be considerably more directly informed and involved in decision-making by opening up data. This gives the citizens the ability to contribute much more to the governance processes.

1.4. FAIR Data

FAIR is one of the most important aspects of Open Data. FAIR is data that is Findable, Accessible, Interoperable, and Reusable. All Open Data is FAIR Data since it is publicly available, editable and reusable. It is possible that some FAIR data is not open, hence not all FAIR data is open.

FAIR data adheres to the FAIR principles, which are outlined below [14]:

- Findable:
    a) F1. The (meta) data has a globally unique and persistent identifier (ID) [15].
    b) F2. Data is well described by metadata.
    c) F3. (Meta) data must be registered or indexed as a searchable resource.
    d) F4. The metadata clearly specifies the data identifier (ID).

- Accessible:
    a) A1. It is possible to obtain (meta) data by an identifier (ID) using a standardized communication protocol [15].
    b) A1.1 The protocol is open to the public, free of charge, and unlimited in implementation.
    c) A1.2 The protocol should be able to provide methods of authentication and authorization if required.
    d) A2. Metadata should be accessible even if the data becomes unavailable.

- Interoperable:
    a) I1. Use a well-defined, reachable, shared, and widely applicable description language for knowledge representation of (meta) data.
    b) I2. The (meta) data uses a vocabulary that follows the FAIR principle.
    c) I3. (Meta) data shall contain identifiable reference information to other (meta) data.
- Re-usable:
    a) R1. Meta (data) has a wealth of accurately related attributes [15].
    b) R1.1 (meta) data is published with a clear and accessible data usage license.
    c) R1.2 (meta) data is connected to its provenance.
    d) R1.3 (meta) data meets community standards for each discipline.

There are two concepts related to FAIR Data: Metadata and globally unique and persistent identifier. Metadata is defined as data that provides information about other data, so data about the data [16]. Metadata has many functions like helping with organizing data, providing digital identification, archiving, and preservation. It also helps with relevant discovery of similar sources but also to distinguish data that ant not the same.

A globally unique and persistent identifier is a long-lasting reference to a document, file, or other object that is globally unique and not bound to an institution. In Open Science, ORCID is the globally unique and persistent identifier used. ORCID is Open Researcher and Contributor ID, which is an alphanumeric code that uniquely identifies scientific and academic authors and contributors while following the Open Access principles by being a nonproprietary code [17].

In Open Science data should not only be Open but it should also be FAIR. That means that data that is not FAIR should be transformed to FAIR Data. This process is called FAIRification which consists of the steps shown in Figure 1-b [18].

The first step is to gain access and retrieve data that is not FAIR. After that, the data should be analyzed and inspected to provide a general view of the structure of the data and the relations between the data elements.

Next, a semantic model that is going to be used for such a dataset is selected, which is describing the meaning and relationships of all the entities in the dataset in an accurate way that includes a vocabulary that is globally recognized. A vocabulary is a computer-readable file that captures terms, their URIs, and descriptions. After that, the data is made linkable by applying the semantic model selected in the previous step.

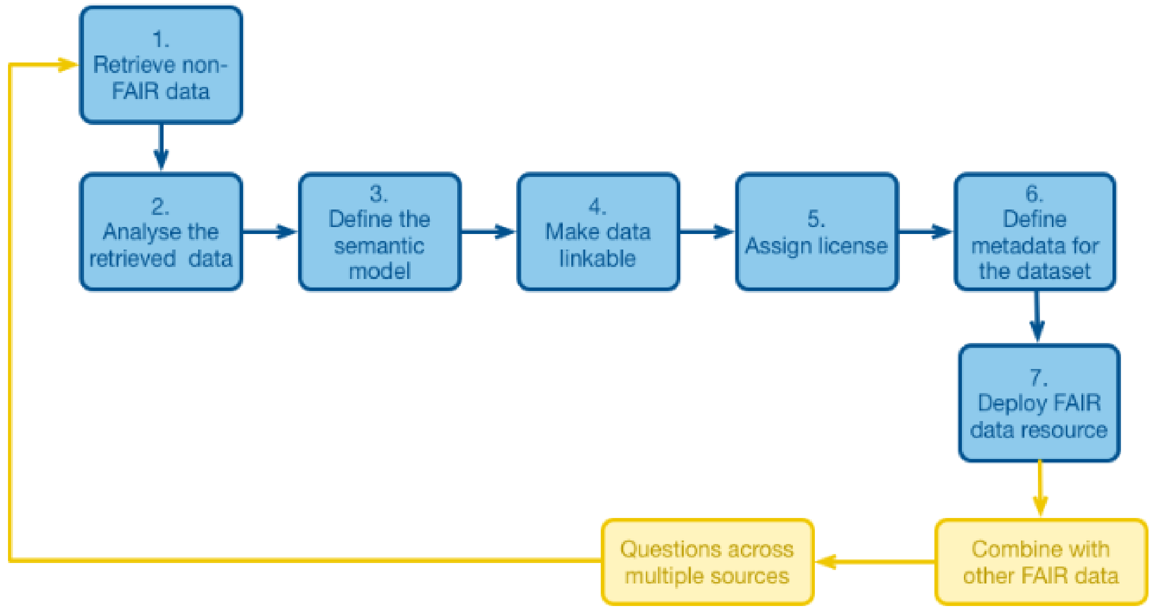

Figure 1-b: FAIRification process

A license is assigned to the metadata to allow reuse. Sometimes the absence of an explicit license may prevent others to reuse the data even if it was intended to be open access data. Finally, the metadata for the dataset is defined to support FAIR principles, and the data resource is deployed, including the relevant metadata and license to help the data get indexed in search engines.

1.5. Open Source

During this dissertation, the term Open Source is mentioned. Open source software is defined as software that not only provides access to the source code but also meets the criteria given below [19]:

- Free Redistribution – No one is prohibited from selling or giving away the software as part of a broader software distribution that includes programs from several sources.
- Source Code – The software should include the source code and allow for both source code and compiled form distribution.

- Derived Works – modifications and derived works must be allowed.
- The integrity of The Author's Source Code – It may be necessary for derived works to have a version number that differs from the original software.
- There should be no discrimination against individuals or groups in the software.
- These should be no discrimination against the field of the individuals in the software.
- Distribution of License - The rights associated to the software must apply to everyone to whom it is redistributed, without the requirement for such parties to sign an extra license. [20]
- The license must not be tied to a specific product.
- The license must not impose restrictions on other software.
- The license must not be technology-dependent.

Of course, there are some standards to be followed even for Open Science. The Open Science standards, or "open standards" the following criteria listed below should be met [21]:

1. No secrets related to the software: The standard must include every detail to have an interoperable service. Software bugs are always present, so all the information related to the software should be public to have a possibility to fix the problems during the implementation in different environments.
2. Availability: The standard must be freely and publicly available from a reliable source with nondiscriminatory features.
3. Patents: All patents required by the standard must be:
   a. Licensed to be used for free;
   b. Follow the standards of Open-Source software.
4. No Agreements: There must not be any requirement for execution of a license agreement, NDA, grant, click-through, or any other form of paperwork to deploy conforming implementations of the standard.

5. The standard's implementation must not need the use of any other technology that does not meet the standards previously mentioned.

Some tools are useful for Open Sources. Open-Source software differs from commercially licensed one that users have to pay to use it. Software is freely available to anyone with no fee to use it, and even the licensing is owned by a user community. A good example of Open-Source software are Apache Foundation software products [22].

For many years, the open-source movement has been a topic of debate. Many believe that open-source tools help the advancements of the field, however a lot of companies need to sell a product or license to survive. However, there are some companies that have made investments into open-source as well. Mozilla Firefox web browser is a good example of open-source software implemented by big a big company. Mozilla Firefox is a free download browser that appeals to the user community, rather than being a browser distributed while paying a license. Generally, open-source software is a collaboration of developers working together to aid the community instead of making profit [22].

Open Source has had an impact into businesses too. It enables them to reap the benefits of open-source software and methodologies. Both consumers and businesses can benefit from Open Source. The consumers benefit from a lower cost of ownership, for businesses it enables the possibility of collaboration. A lot of companies can work together to bring a new open-source product which benefit all of them; a product which none of the companies could be able to produce alone. Since consumers had a much lower cost of ownership, they are more inclined to accept bugs from open-source software, which in return can help businesses have a superior development of the product by having customer-identified bugs. The open-source model has a lot to offer the business world: both as consumers and contributors [23].

The open-source model also allows for increased security; because code is visible to the public, it will be subjected to heavy inspection, allowing for more vulnerabilities to be identified and addressed rather than being hidden until the unwanted person discovers them [23].

1.6. Digital repositories and Content Management

To store data, Digital repositories are needed. Digital repositories are “information systems that ingest, store, manage, preserve, and provide access to digital content” [24]. Digital repositories are also needed for storing research data. DSpace is a very famous Digital repository, originally designed at the MIT and HP Labs. DSpace is a “free, open-source software platform for building repositories of digital assets, with a focus on simple access to these assets, as well as their long-term preservation”. The goal of the repository is to have all the research materials important for a research institution under a single system, which can be used by all the academics [25].

Fedora is another open-source digital repository that is very prominent. It provides a group of tools that use open standards to have a service-oriented design. It provides authentication, access control, integrity checking and all the important tools needed to have a fully functional digital repository [26].

There are also standards for “trusted” digital repositories like TRAC (The Trustworthy Repositories Audit & Certification Checklist), superseded by the ISO 16363:2012—Audit and Certification of Trustworthy Digital Repositories international standard.

An open-access repository or open archive is “a digital platform that holds research output and provides free, immediate, and permanent access to research results for anyone to use, download and distribute” [27]. Such repositories must be interoperable according to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Search engines collect the content of open access repositories, creating a database of freely available research from around the world.

The advantages of using such repositories are numerous [27]:

- Making the institution's outputs available to a global audience.
- As a result, increasing the visibility and effect of these outputs.
- Introducing the institution to prospective employees, students, and other stakeholders.
- Better organization and evaluation of research projects.

- Facilitating and fostering multidisciplinary research approaches
- Assisting in the dissemination of digital resources, and
- Assisting students with their activities and providing a space for the establishment of e-portfolios.

Digital Commons, DSpace, and EPrints are the most popular open-access repositories, according to OpenDOAR. ArXiv, bioRxiv, Dryad, Figshare, and Zenodo are also used.

Digital repositories are closely related to Content management. Content management (CM) "is the process for collection, delivery, retrieval, governance, and overall management of information in any format" [28]. The definition is commonly used to describe the management of digital content, from production to permanent storage or deletion. CMS (content management systems) are also used to manage dynamic content. It is commonly adopted in multi-user scenarios with access permissions and workflow management. [29].

A content management system usually consists of three elements: output-layer, core-CMS, and data storage just like illustrated in Figure 1-c.

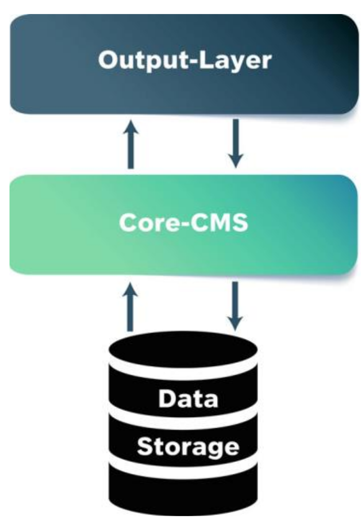

Figure 1-c: CMS architecture

The core of the CMS is used to collect data after the user authentication. The output layer formats and shows the data that was captured by the core. It is used to enable the standards to have a fully functional system for the end user. It deals with the throughput of the information and the transformation of it to a user readable format.

The data storage layer is used as an internal storage that only the core can access. Normally it holds all the information of the system, including the authentication and access data. Sometimes an external storage can be used to enhance the data storage, however the external storage has data formatted in a way that is usually user readable. This data must not be full data, to prevent possible leaking of the information.

The services provided by CMS are vast but indexing, search, retrieval, format management, revision control, and publishing are considered to be the essential. Other functions are:

- Easy-to-use indexing, search, and retrieval features index all data for quick access via search tools, and users can search by publication dates, keywords, or author [28].
- Format management makes it easier to convert scanned paper documents to PDF.
- Content can be edited after it has been published using revision features.
- Individuals can apply a template or a collection of templates approved by the institution, as well as other tools, to generate or alter materials.

A content management system (CMS) may also include tools for one-to-one marketing. One-to-one marketing refers to the ability to personalize content related to a specific user of the website, by creating content related to the specific characteristics of this user [28].

CMSs provide other popular services including:

- URLs that are preferable by search engines
- Online assistance integration.
- Permission-based access
- Customizable template support
- Simple installation procedures
- Support for multiple languages

- Unlimited depth and extent of content hierarchy
- Server specifications are simple.
- File managers that are incorporated
- Audit log that is implemented

Usually, instead of CMS, a content management framework (CMF) is “a system that facilitates the use of reusable components or customized software for managing Web content”. Aspects of a content management system and a Web application framework are combined in it.

Some Open-Source CMFs are:

- Apache Jackrabbit or Jakarta Slide – with Java language
- Open Semantic Framework - Drupal, OWL, PHP, and RDF
- RadPHP - MySQL, PHP 5.6+, PostgreSQL.

CMS software is usually commercial but there are also some Open-Source alternatives depending on the programming language used, listed below:

1. Java
    a. Apache Sling CMS
    b. DSpace
    c. Fedora Commons
2. ASP.NET
    a. mojoPortal
    b. Orchard Project
    c. C1 CMS
3. PHP
    a. CMSimple
    b. Drupal
    c. Magento

1.7. Objectives of the dissertation

As shown in the previous sections, the main goal of this dissertation is to analyze the situation and propose relevant measures to support the introduction of Open Science in Balkan universities. With this dissertation, I am trying to address the transition to Open Science, faced by such universities. The main focus will be on identifying main problems and proposing suitable solutions for introduction of Open Science in Albania, which has made almost no advancements to Open Science so far.

In particular, the following more specific objectives are targeted by this dissertation:

- Analyzing Open Science developments in Europe. Exploring the European view on Open Science: European Open Science Cloud.
- Identifying the stages in the transition to Open Science and main problems faced by Balkan universities. A detailed questionnaire will be developed and distributed to all relevant Balkan institutions. After a careful analysis, all main key findings will be identified and used to plan the next steps for the transition to Open Science of Balkan institutions.
- Developing a prototype of open science repository, following the European standards set by European Open Science Cloud. Deploying this prototype to support such transition to open science in Albania.
- Developing an Open-Science Policy to be implemented in Albania.
- Describing the next steps for implementing and deploying Open Science in Albania.

1.8. Publications related to the dissertation

Below is a list of publications that focus on Open Science in Europe and about the advancement of Open Science in Balkan institutions. These publications are listed in scientific journals and/or international conferences or scientific reports.

- Prompting an EOSC in practice - Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC), 2018. Directorate – General for Research and Innovation (European Commission) doi: 10.2777/112658. Available at: https://op.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1/language-en
- Relevance of the EOSC initiative and FAIR principles in the realm of Open Science and implementation phases of the EOSC, Michel Schouppe Jean-Claude Burgelman. Proceedings of the XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), Moscow, Russia, October 9-12, 2018. Available at: https://ec.europa.eu/research/openscience/pdf/eosc-fair_paper_schouppe-burgelman_2018.pdf#view=fit&pagemode=none
- Georgiev, Atanas; Stefanov, Krassen (2019) Bulgarian Open Science Digital Library - First Prototype, in Proc. of the 9th International Conference on Digital Presentation and Preservation of Cultural and Scientific Heritage (DiPP), Volume: 9, Pages: 251-258. WOS:000487853900025, ISSN: 1314-4006, eISSN: 2535-0366. Available at: http://dipp.math.bas.bg/images/2019/251-258_8_3.8_sDiPP2019-66_f_v.1.F_20190908.pdf
- Stanchev, Peter; Stefanov, Krassen (2019) Bulgarian Open Science Cloud, in Proc. of the 9th International Conference on Digital Presentation and Preservation of Cultural and Scientific Heritage (DiPP), Volume: 9, Pages: 259-264. WOS:000487853900026, ISSN: 1314-4006, eISSN: 2535-0366 Available at: http://dipp.math.bas.bg/images/2019/259-264_6_3.9_sDiPP2019-65_f_v.1.F_20190908.pdf

- Grigorov, A., Georgiev, A., Petrov, M., Stefanov, K., Varbanov, S. (2009) Building a Knowledge Repository for Life-long Competence Development, in IJCEELL V19 N4/5/6 2009, Special issue "Stimulating Personal Development and Knowledge Sharing", eds. R. Koper, K. Stefanov and D. Dicheva, pp.300-312. Available at: https://www.researchgate.net/publication/254324536_Building_a_Knowledge_Repository_for_Lifelong_Competence_Development

1.9. Structure of the dissertation

This dissertation is structured as below:

- The next chapter presents a literature review about Open Science in Europe. It analyzes Open Science developments in Europe. The main characteristics of the European Open Science Cloud are explored. Different Research Repositories are explained and described. A comparison between the Repositories is done.
- Chapter 3 identifies the stages in the transition to Open Science in Albania and Bulgaria. A questionnaire is distributed to capture the transition to Open Science in Balkan institutions, and a full evaluation of the current state of the transition is shown. It describes the decisions and the main element of the prototype of Open Access Digital Repository in Albania, as the first step towards Albanian Open Science Cloud. The architecture of such a prototype is developed.
- Chapter 4 presents the prototype and explains with details the main characteristics and details of the Albanian Repository. Then, the prerequisite software will be presented and the build of the digital repository using DSpace-CRIS. After that, the data model and the design of the Digital Library compliant to EOSC are detailed. Lastly, an Open Science Policy to be implemented is developed.
- Chapter 5 discusses conclusions and future works and in the last section, the bibliography is shown.
- Appendix A shows the questionnaire distributed.

Chapter 2. State of Open Science in Europe

In this chapter, the European view on Open Science is analyzed. Firstly, European Open Science Cloud is introduced as the model that Europe follows for Open Science. After that, digital repositories for research data management are illustrated and compared. Based on this analysis, a solution is proposed for the selection of a repository for the development of the Open Science Digital Repository.

2.1. European Open Science Cloud

The European Open Science Cloud (EOSC) is “a trusted digital platform for the scientific community, providing seamless access to data and interoperable services that address the whole research data cycle, from discovery and mining to storage, management, analysis, and re-use across borders and scientific disciplines” [30].

It was launched in Vienna in November 2018 to establish a European data infrastructure, integrating high-capacity cloud solutions and widening the services offered to Open Science [31].

EOSC is considered a platform that benefits all the researchers in the European Union by expanding the usage of data-driven science. It consists of a set of rules and standards to be followed, which incorporate FAIR data, interoperable services and metadata standards.

EOSC is part of the "European Cloud Initiative to build a competitive data and knowledge economy in Europe", launched in 2016. It has the goal to enhance Open Science and Science in general by three main characteristics, by having a well-built system to store and manage all the research data, tools to transport the research at high-bandwidth network, and powerful computing abilities to process the metadata [32].

In 2018, a governance was created to complete the objectives set by EOSC. It consisted of 3-tier architecture where each tier had its responsibilities:

- a Governance Board, which has people from the members of the European Commission. The role of the board is to decide the advancements of EOSC.
- an Executive Board, which consists of researchers and experts in the field of Open Science that advices newcomers on how to implement EOSC guidelines.
- a Stakeholders Forum, which consists of community members and the public sector to take part in the decision making of the future goals.

A declaration of the purpose of EOSC was created to assist in [33]:

- Recognizing the difficulties of implementing Open Science.
- Assuring that the objective of EOSC is to create a shared research information system compliant to Open Science for all the members of the European Commission.
- Assuring that EOSC is a process and not a project, and its implementation will be constantly adapting and changing to be interoperable.
- Endorse all the scientific advancements that endorse Open Science and EOSC implementation.

This enables a unified access point for all the researchers in EU by having a list of services to be offered:

- One access point to the system for all the research entities.
- Multi-disciplinary research cooperation by having all the data accessible through a single point.
- Implementation with Interoperable and FAIR data.
- Increased value of data science researches since all the data can be accessed easily.

The main goal of EOSC is to allow universal access to data to all EU researches through an easy access point, that is cross-disciplinary access to all data, using services as data that are interoperable FAIR data. This means that EOSC helps to recognize Open Access and Open Science in all of Europe. To achieve this, several models were introduced as are shown in Figure 2-a.

	Current model	Federated	Centralized model
Resources	Fragmentation of resources and access to them	Integrated access to federated resources for ALL researchers	Access to a single data device (storage; single supercomputer?) for ALL researchers
Services	Varying quality of services currently	Service standards for all federated resources	Centrally decided standards
Interoperability	Varying levels of interoperability standards	Common standards for all federated resources	Centrally decided interoperability standards
Governance	Fragmented across 100+ institutions	Layered governance for EOSC participants – balanced stakeholders, MS, EC - with specific rules	Single body, centralized governance set up by EC
Costs & time to implement	Baseline	Marginally higher than baseline	Substantially higher than baseline
MS and stakeholder acceptance	Low	High	Extremely low

Figure 2-a: Different EOSC models proposed.

EOSC chose to integrate a federated model consisting of 6 main actions as shown in Figure 2-b [34]. The federated model aimed to provide integrated access to all of the services provided by EOSC. These services are following FAIR data. EOSC provides a list of FAIR data tools, standards, services, and catalogs. It also provides a list of services for all of the researchers including services to find, store, access, reuse, and distribute data generated by other researchers. EOSC provides an EOSC portal where all researchers can easily access all of the mentioned tools and information by having rules to participate and a governance board to decide the future of EOSC, Open Access, and Open Science in Europe.

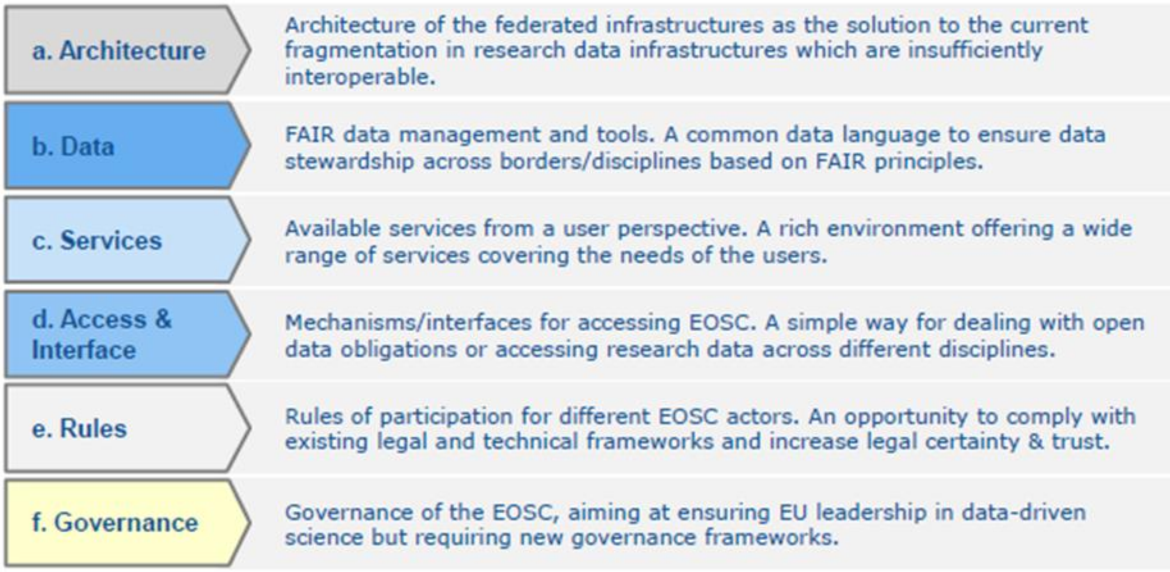

Figure 2-b: Actions of EOSC federated model

The architecture of the federated infrastructure which was chosen has the following characteristics [34]:

- A multidisciplinary data infrastructure accessible by all the members, which provides a list of tools to support all the services at national or international level.
- It will provide a single point of entry for all the research data systems that provide service as part of the EOSC.
- The core of the architecture will provide a Rule of Participation, which means that all the data infrastructures and services which the participants offer and use, will be compliant to standards set by the EOSC.
- The single point of entry will provide all the researchers tools to access information about Open Science, EOSC, and tools for data management and sharing, as well as computational power.

The architecture presented aids in the achievement of four objectives visioned when deciding EOSC. To begin, it aids in the organization of all the components of EOSC, including the single point of entry of the platform, the standards set for compliance, and tools to enable Open Science, monitor and regulate transactions across the system [34].

Secondly, it conducts a national and international survey to understand the willingness and the capabilities of scientific communities to participate in the EOSC development phase.

Thirdly, it establishes the Rules of Participation and the creation of FAIR certificates for all the infrastructures that are compliant with EOSC.

Finally, through the FAIR accreditation/certification scheme, it assists in the establishment of a list of all the data infrastructures that are compliant with EOSC, while monitoring if they are compliant during the advancements of the EOSC.

EOSC is an advocate of Open Science and Open Access in Europe. It is supported by cOAlition S, an international consortium of research funders. It is the origin of Plan S, which the main goal is "With effect from 2021, all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo" [35].

EOSC also helps to develop a data model that uses all the standards it sets to the researchers. So, information about Open Science, FAIR data, standards and services should be readily available to all users that access the infrastructure. It also has a FAIR Research Data framework which identifies and promotes FAIR data principles. The main goal is to allow all the researchers to share and reuse the research data available without having to transform the data set for their new research goal. Data that is interoperable and multidisciplinary provides a much better data-driven science and better data manipulation and understanding.

By having a FAIR Data certification scheme, researchers can identify if data they need to access or want to access can be reused by them and if certain research can be advanced. The goal of the FAIR Data Persistent Unique Identifier is to establish a common, cross-disciplinary, and cross-border policy for the management of EOSC persistent unique identifiers, ensuring long-term persistence and quality for all the research data objects [34].

Researchers that use EOSC, can:

- Use a single point of entry to identify and authenticate themselves to access of the resources of the system.
- Have a workspace where they can access and modify their preferences, work in progress and everything related to the system.
- Have access to all the information related to Open Science, EOSC, list of tools and services provided to them and also information about policies and compliance.
- Have access to FAIR data models, as a tool to understand which is the best metadata format they can use.
- Use services mentioned above to enhance Open Science and research data manipulation into their workspace.
- User services to make their existing data FAIR.

Most services are free, however when big data is involved, researchers might demand large computation power, so they need to request co-funding with another source. To ensure a smooth transition from legacy systems, the access model provides the researchers with a choice of alternative entry points for accessing EOSC services.

These several EOSC entry points all include [31]:

- An understandable front-end which is web-based where the researchers can freely navigate and understand how to access the information, they need from EOSC portal.
- a common platform as the back-end, so all the information they access is the same information found on the EOSC portal.

- Access to the EOSC portal which is considered a single point of entry for all the researchers regardless of their geographical location.

EOSC developed a platform that provides the following services:

- Establishes a common metadata format which can be used as an interoperable data set, that is multidisciplinary accepted and used. The dataset can be utilized by anyone that wants to be compliant with EOSC standards.
- Establishes a list of services that enable findability, interoperability for all the researchers using those services.
- Provides an overview of the EOSC platform prototype.
- Provides a single point of entry through the EOSC portal.

The portal provides a single point of entry for all the researchers to all the information related to Open Science, EOSC, list of tools and services provided to them and also information about policies and compliance. FAIR data models are also present as a tool to understand which is the best metadata format they can use. It also provides information about the vision and mission of EOSC including its roadmap.

2.2. Digital Libraries and OpenAIRE

Open Access digital libraries are also called Open Access Platforms. These platforms are scientific research repositories that are compliant to Open Access, they are digital libraries with three main attributes [36]:

- Scientific data is collected in digital form and all the knowledge of these publications are then indexed, accessed and searched from digital repositories.
- All data is organized in collection and hierarches by using different classification methods.

- All data is described by the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) metadata standard, so it can be used by any Open Access digital repository in the future. It follows all the standards of Open Access explained in the first chapter.

2.2.1. Interoperable Repositories

The Open Archives Initiative Protocol for Metadata Harvesting standard provides an application-independent interoperability framework based on metadata harvesting. There are two classes of participants in the OAI-PMH framework [37]. Data Providers administer systems that support the OAI-PMH as a means of exposing metadata and Service Providers that use metadata harvested via the OAI-PMH as a basis for building value-added services.

Scientific research repositories that are compliant with Open Access are also called Open-Access repositories. These kinds of repositories store the research output and provide free, immediate, and permanent access to all research results to everyone. They can download, use, distribute and share this data without having any sort of discrimination. These are digital libraries that have the ability to:

- Manage and create digital data
- Provide free access to scientific resources
- enable and encourage an interdisciplinary approach to research
- increase the quality of research publications
- manage teaching activities
- support students by providing access to theses and dissertations
- enable access to scientific results and push new collaborations at large-scale

The OAI-PMH provides an easy implementation for repositories. This is called a “minimal Repository Implementation” that describes the best practices for repository implementation that are interoperable [37]. It is important to emphasize that many optional concepts are not important to implement to provide the “minimal Repository Implementation”, but they provide desirable features which can be implemented when needed. The main characteristics are:

- Dublin Core (DC) as the basic metadata format. Of course, other metadata formats can be used, but DC is the minimum requirement for the repository.
- <about> Containers for self-referential metadata, which can be used for encoding rights.
- Sets as a method of exposing repository data to harvesters.
- Response Compression, which is used to have a performance boost. The repository data send to harvesters as a response, can be compressed.
- Flow Control, which means that a mechanism for setting a throughput from requests of the harvesters should be in place. This is the most complex part of OAI-PMH to implement.

2.2.2. OpenAIRE

In Europe, OpenAIRE is the key initiative that pushes open science digital libraries forwards [8]. OpenAIRE is a European project supporting Open Science. It is a network of dedicated Open Science experts promoting and providing training on Open Science. OpenAIRE is a technical infrastructure harvesting research output from connected data providers. OpenAIRE aims to establish an open and sustainable scholarly communication infrastructure responsible for the overall management, analysis, manipulation, provision, monitoring, and cross-linking of all research outcomes. By implementing Open Science policies, OpenAIRE targets important issues of Open Science in Europe and supports Open Access to papers and research Data. It has a key role in EOSC, by developing a list of services that are interoperable with EOSC services, which enables better integration with Research Infrastructures in Europe.

OpenAIRE has three layers of activities in EOSC: institutional, national, and international. It offers policies, training, and services. OpenAIRE helps to align specific policies for FAIR, Research Data management, and Open Science. It helps with training on Open Science topics, and with connection and integration with global research systems. The contributions include standards tools and services for Open Science implementation. That is, open access should be provided to all forms of scientific findings, including FAIR research data, open-source software, open education, open services, open protocols, and open methodologies to all European citizens.

EOSC has proposed a data model format to be used in CRIS. A current research information system (CRIS) is “a database or other information system to store, manage and exchange contextual metadata for the research activity funded by a research funder or conducted at a research-performing organization” [38]. The data model used provides a metadata representation of the research entities while enabling all the necessary tools to properly maintain the data that is going to be used by the system. The format of the data model used is CERIF. CERIF is the Common European Research Information Format [38] that has tools to provide:

- a concept for the research entities and their relationships with each other.
- a description of such research entities in the form of a model.
- a script to formalize the research entities.

A CRIS implementation can use full CERIF model, or parts of it, as long as all the relationships related to all research entities is correct. The main reasons for using the model are: it uses a neutral architecture which allows for different data models to be implemented; databases can be object-oriented or relational as long as they are configured correctly; advanced knowledge-based processes can be used into the system.

As we know, to have interoperability a structured metadata is necessary. That is why CERIF allows the implementation of a standalone CRIS but also the ability to define a metadata that works with other metadata formats, by having a wrapper around. This allows also backwards compatibility with systems that are already running.

In the year 2000, CERIF format was handed-over to euroCRIS, which made it possible to have multiple upgrades and extensions to the format. Now CERIF’s goal is a standardized European Research Area, which is very similar to the goal of EOSC.

2.3. Research Data Management Repositories

Research Data Management is a digital library that is used on research projects and all research data. Research data management (RDM) definition is a term that describes the organization, storage, preservation, and sharing of data collected and used in a research project [39].

There are a lot of Research Data management repositories but four are the most prevalent ones. Those are DSpace-CRIS, Fedora, and Zenodo (Invenio). The next sections will describe these repositories. It is important to note that all of these repositories are OpenAIRE compliant and open source. They are close to the European vision on Open Science and work closely with FAIR data principles.

2.3.1. DSpace-CRIS

DSpace-CRIS is "the first free open-source extension of DSpace for the Research Data and Information Management ever developed" [40]. DSpace is an open-source software system that provides:

- Tools to collect and describe digital assets using a specific workflow
- Tools to disseminate the digital assets that are collected
- Tools to preserve the digital assets

DSpace can be compared to full-stack web development, which has a front-end, back-end and database system. The data model used in the architecture is very configurable with the possibility to choose the metadata format used. By defaults, it used the Dublin Core metadata, also known as QDC (Qualified Dublin Core). However, the possibility to change the metadata is possible, and the implementation of other metadata formats is straightforward.

DSpace consists of a list of tools important for the management of digital assets, including the metadata tools. Even commercial plugins can be configured to work into the system. It also has the authentication and access control built in to allow only certain users to modify the configuration of the system. The administrator can also delegate specific tasks to other people using the built-in tools.

One important feature for Repositories, is to allow the export of backup files. These files can be easily exported by using AIP (Archival Information Packages) backup files.

DSpace complies with standard protocols and best practices that we mentioned for Open Science. It has tools to be interoperable with OAI-PMH, SWORD, OpenAIRE, and Driver.

One additional feature that DSpace-CRIS architecture brings is RESTful APIs. This means that the system can be interoperable with other elements of the modern web applications using REST responses/requests to handle the data from the database.

Oracle and PostgreSQL can be used with DSpace-CRIS and on upload a checksum is created to validate file integrity.

The core component of DSpace-CRIS is the institutional repository, providing visibility of all the stored data and object on the web. It is compliant with international standards important to Open Science: ORCID, OpenAIRE guidelines, PlanS, and FAIR principles.

Figure 2-c shows the DSpace-CRIS interface which includes a demo at dspace-cris.4science.cloud [41].

The dynamic data model of DSpace-CRIS allows you to collect and manage research data and information typical of a CRIS system, as well as define entities and attributes with their relationships [40]. A Current Research Information System (CRIS) is a database or other information system to store, manage and exchange contextual metadata for the research activity funded by a research funder or conducted at a research-performing organization.

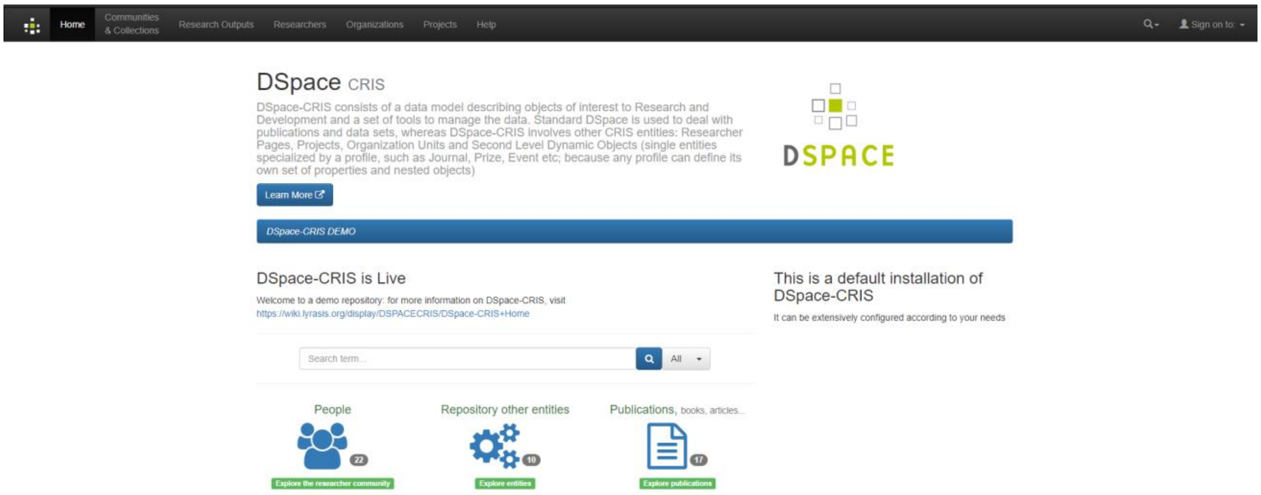

Figure 2-c: Demo of DSpace-CRIS interface

DSpace-CRIS can store the publications which are a very important part of the research, but also projects, patents, research profiles, grants, and organization units. It allows management of all metadata and data for each of the research entities.

DSpace-CRIS is comprised of a data model that describes all the Research Items related to the data and to the metadata of the research entities which are divided into Researcher Pages, Projects, Organization Units, Research Outputs and Communities. Each of the items has a relationship which the other elements using joins into the database.

The main features of DSpace-CRIS are:

- Management of all CRIS entities
- Global search and global indexing of entities
- Management of funding opportunities
- Management of researcher profiles and promotions
- Interoperability with external systems
- Statistical reports and analytics

2.3.2. Zenodo - Invenio

Zenodo and Invenio are always discussed together and there is a reason for that. Zenodo is a good example of a Research Data Management service that is hosted by CERN, but it is not a repository that you can run as your application. That is Invenio, which is a repository application that can help you have a similar experience to Zenodo [42][43]. Invenio is an open-source project that was initially developed by CERN. Zenodo is based on the Invenio digital library framework which is an open-source framework, hosted on CERN Data-center.

Zenodo characteristics related to Open Science are the following:

- Zenodo servers are controlled by a software configuration that ensures that the latest security updates are constantly installed. It manages huge pools of computing, storage, and networking resources throughout a datacenter using an OpenStack API or a dashboard. OpenStack is suited for heterogeneous infrastructure because it integrates with popular commercial and open-source technologies [43].
- All files uploaded to Zenodo are saved in an 18-petabyte disk cluster at CERN, by having two copies of every data in case of failures.
- Zenodo’s metadata and persistent identifiers are saved in a PostgreSQL with a 12-hourly backup cycle and a weekly backup.

An example of the Zenodo-Invenio interface is available at inspirebeta.net [25] shown in Figure 2-d. Inspire is the high-energy physics Information System that combines the successful SPIRES database content, curated at DESY, Fermilab, and SLAC for decades at CERN.

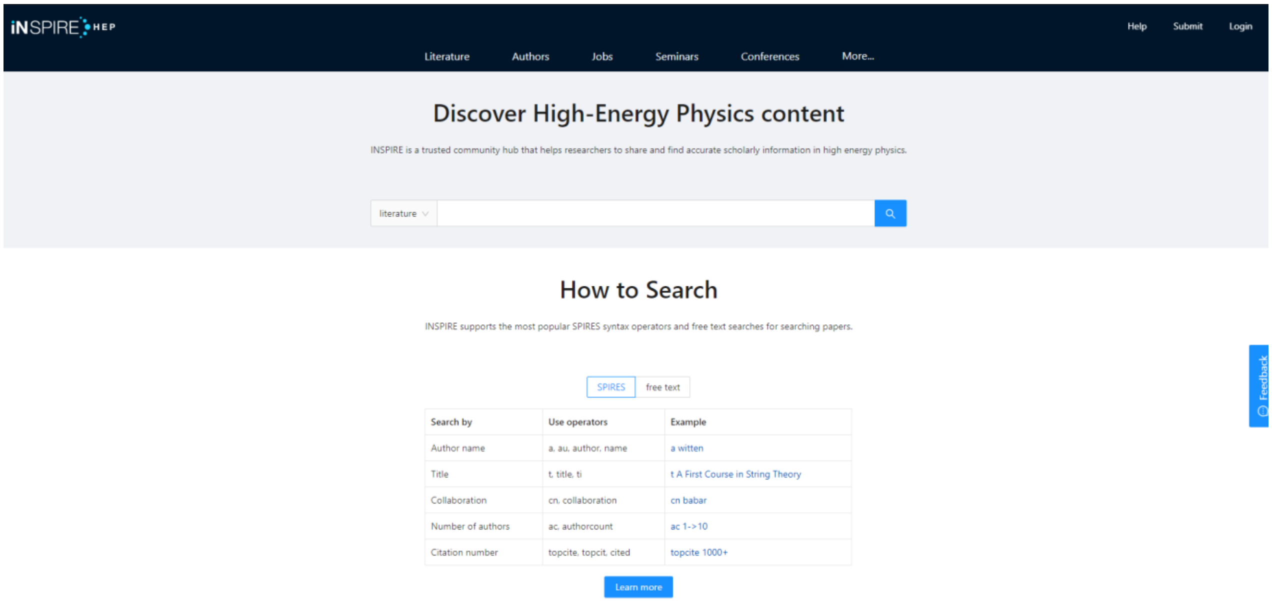

Figure 2-d: Invenio Interface at Inspirebeta.net

On the other end, Invenio is the foundation of Zenodo and now offers three types of services, including a research data management project that is open source called InvenioRDM. It comes with pre-configured repository profiles for institutional repositories and research data management systems.

2.3.3. Fedora

Fedora (or Flexible Extensible Digital Object Repository Architecture) is a modular architecture built on the principle that interoperability and extensibility are best achieved by the integration of data, interfaces, and mechanisms as clearly defined modules [44].

Fedora is a digital asset management architecture, upon which many types of Digital Libraries and Institutional Repository systems are built. Fedora is the underlying architecture for a Digital Repository and is not a complete system for the management, indexing, discovery, and delivery application. It is Java-based so potentially applicable to any platform. Fedora is FAIR compliant but it fails to provide a direct research data management repository out-of-box. Some research institutions and universities are using customized Fedora-based repository for data management like the University of Virginia (https://www.library.virginia.edu/) shown in Figure 2-e.

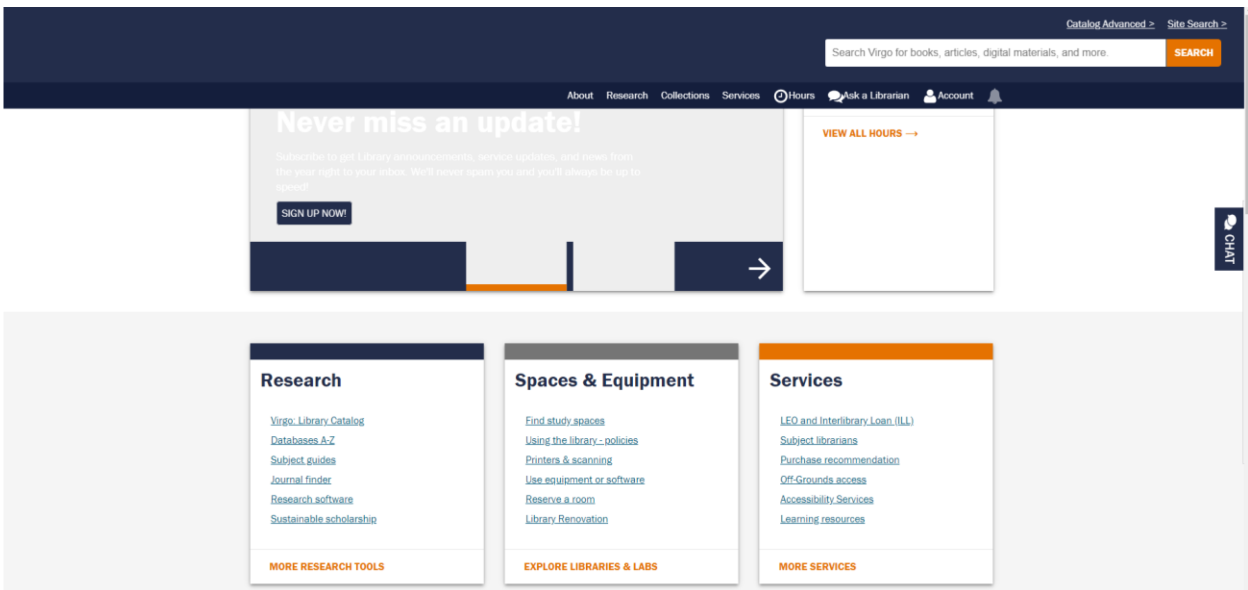

Figure 2-e: Fedora-based Interface at University of Virginia

2.4. Conclusions

Open Access Repositories are explained above with all of their characteristics. To summarize the three repositories, they are compared in Table 1 below.

Table 1: Comparison of Open-Source Repositories

	Fedora	DSpace-CRIS	InvenioRDM
User Interface provided	Yes	Yes	Yes
Dynamically Generated pages	Yes	Yes	No
Open-Source	Yes	Yes	Yes
Out-of-box configuration for RDM	No	Yes	Yes
Research publications entity management	Yes	Yes	Yes
Other research data entities management	No	Yes	Yes
OpenAIRE compliant	Yes	Yes	Yes
FAIR compliant	Yes	Yes	Yes
Flexible architecture	Yes	No	No
Multiple metadata format compatibility	Yes	Yes	Yes
Customizable metadata formats	Yes	No	Yes

All of the above-mentioned software's have a user interface, are open source, OpenAIRE, and FAIR compliant and support multiple metadata formats on their entities. However, when it comes to research entities only two provide out-of-box configuration. In conclusion, DSpace-CRIS and Invenio are more suitable for Research data management repositories in research institutions and universities since they are preconfigured for research entities. This doesn't mean that Fedora is not usable. As stated in the section above, some universities use Fedora, and the main reason for that is the fact that it is flexible. Its architecture allows integration with other systems present at their institutions and using Fedora they add the research data management practices to such systems. Choosing between DSpace-CRIS and Invenio is mostly convenient for a Research Information system.

For the dissertation, DSpace-CRIS will be used, and the main reason is the familiar user interface that researchers are used to. Besides that, it provides all the necessary research data management entities needed to have full management of research data in the repository and the pages are dynamically generated. A good example of this choice is the first prototype of Bulgarian Open Science Cloud(BOSC) [45] which uses DSpace-CRIS and is OpenAIRE and CRIS compliant. The data model chosen is relying upon the basic entities of the Common European Research Information Format model (CERIF).[38] CERIF (Common European Research Information Format) is the standard that the EU recommends to its member states for recording information about research activity. It is developed and maintained by EuroCRIS. EuroCRIS is created to address the problems of current research information systems worldwide with an emphasis on Europe. One advantage of InvenioRDM is the ability to work with customizable metadata formats, however, this advantage is not relevant in this dissertation since the format to be used will be CERIF compliant.

Following the model of BOSC, DSpace-CRIS is the open-source system used to create the first prototype of the open access repository in Albania.

Chapter 3. The Current State of Open Science in Balkan

In this chapter, the stages in the transition to Open Science and main problems are identified through a survey that was disseminated in Albania and Bulgaria. A comparison to understand the main differences between these countries to other European countries is described. The main characteristics of the first prototype towards Open Science for Albania are proposed at the end of the chapter.

3.1. State of Art

Other surveys that capture the current state of Open Science in Balkan countries. To start, one of them is the National Initiatives for Open Science in Europe (NI4OS-Europe) initiative which focuses on Open Science in South East European countries [46]. It is disseminated in 15 countries in the South East of Europe, including Albania and Bulgaria. However, it did not have a wide distribution in these countries by achieving only 1.91% of the total responses from Albania and only 7.48% from Bulgaria. Besides that, it discusses the funding criteria, rules, and policies on Open Science; awareness on EOSC, FAIR, and Open Data; it doesn't include questions related to research repositories or technical aspects of these systems.

An important initiative that captures the current state of Central European countries is the "National Initiatives" survey [47]. The survey is detailed and captures the transition of Central and North Europe to Open Science. However, it targets the crucial groups of EOSC and doesn't include Balkan Countries.

The most disseminated survey is the EUA "Research Assessment in the Transition to Open Science" [48]. It was distributed in 2019 to 32 countries with 260 valid responses. Neither Albania nor Bulgaria was part of the 2019 survey. Besides that, this survey doesn't discuss technical details but focuses more on Open Science awareness and its support in Institutions.

Other national initiatives can be mentioned, like the National science program "Information and Communication Technologies for a unified digital market in Science, Education, and Security" or ICTinSES, which is an initiative financially by the Ministry of Education and Science that focuses on widening the participation of the Bulgarian research community in the European Research Area by making strong relation in science, education, and society [49].

The survey disseminated captures the current state of Albania and Bulgaria towards Open Science.

3.2. Methodology

The survey presented in Appendix A tries to capture the transition to Open Science in Bulgaria and Albania. It is similar to EUA, which was not distributed in these countries. It also goes into some technical details which were not part of other surveys disseminated in Bulgaria and Albania. The survey tries to have more participation compared to NI4OS-Europe which is the most disseminated survey in Albania. This survey is used to evaluate the Albanian to Open Science and to plan relevant training and research activities in helping Albania to join the EOSC initiative. It captures the current state of Open Science in Albanian and compares it to Bulgaria and Europe.

The survey is structured with 32 questions. A combination of open questions, ranking, multiple-choice, and single-choice questions was included. The survey covers a wider variety of Open Science topics than the report of EUA, which consists of 20 questions. The topics include the current state of research assessment in Bulgarian and Albanian universities, as well as how institutions are reviewing their evaluation practices. Besides that, the survey is extended with technical questions about technical aspects such as Institutional Repositories (IRs) implementation, protocols, standards, persistent identifiers, and metrics used.

Questions on the current research assessment procedures are similar to the EUA report. That means that there were 3 sections covering research careers and research unit performance evaluation in the first section, open science approach and research funding allocation within the institution in the second section, and after that, the technical details are surveyed in the third section. The goal is to check the current research information system used on the universities is implemented, and which protocols, standards, and vocabularies are used. It provides an overview of other Institutional Repositories used at universities and finally is open access policies are implemented at an institutional level.

The respondents are from Academic and Research Staff, Lead positions at universities, library lead positions, and library staff. The report has 53 responses from 25 different institutions in Bulgaria and Albania. 13 responses are from 6 Albanian Institutions and 40 responses are from 19 Bulgarian Institutions. The sampling rate is close to 2 responses per institute. Albanian number of respondents is comparably low but all the participants are in key roles with the hope that their view fully reflects the state of transition to Open Science of Albanian Institutions. 11 of the respondents (85%) in lead Academic or Research positions except for 1 response from Research staff and 1 response from Library Lead Position.

The majority of respondents in Bulgaria are academic and research staff at 47%, and 27% are lead academic and research positions (Directors, Head of Departments), followed by 13% of leading positions at the universities (Deans, Rectors, Vice-Rectors), and 13% of the responses are from library lead positions and staff as shown in Figure 3-a.

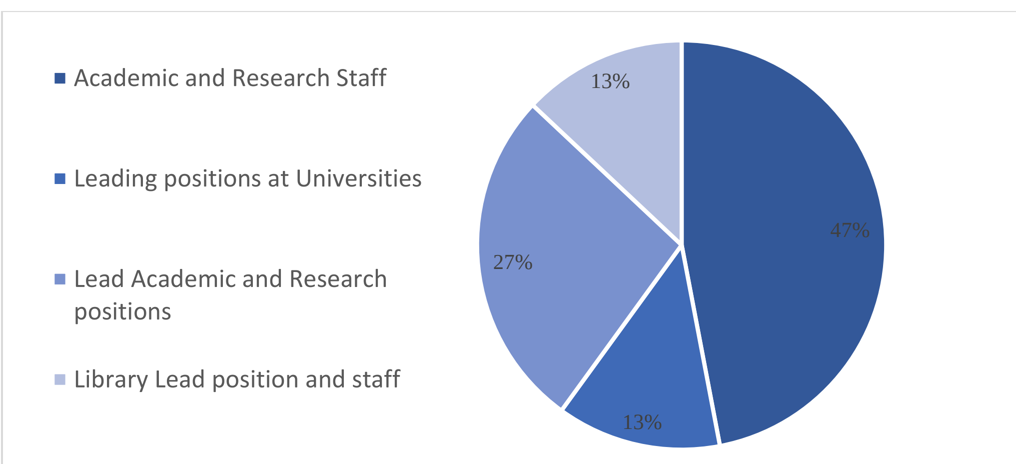

Figure 3-a: Respondent Profiles in Bulgaria.

3.3. Research Assessment results

In this section Research Assessment results from the survey are shown. The participants responded to how autonomous do they consider their institution in developing and implementing research assessment procedures for the evaluation of the performance of the research units. The results shown in Table 2 reveal completely different responses for Bulgaria and Albania.

Table 2: Autonomy to develop and implement research assessment procedures

Autonomy to develop and implement research assessment procedures	Bulgarian Institutions	Albanian Institutions	EUA report
High autonomy	11%	8%	44%
Mostly autonomous	60%	25%	39%
Some autonomy	24%	67%	14%
Low autonomy	5%	0%	3%

In Bulgaria, 60% of the respondents agreed that they are mostly autonomous; only 11% have high autonomy. The results are followed by 24% of respondents having some autonomy and 5% on having a low autonomy to develop and implement its procedures on research assessment. Compared to the EUA report, Bulgaria has a lower autonomy than other European countries to develop and implement Research Assessment Procedures. 44% of EUA report respondents are highly autonomous and 39% are mostly autonomous to develop and implement its research procedures.

In Albania, with 67% agreeing that they have low autonomy to develop such procedures; 25% of them having some autonomy, and only 8% having high autonomy.

Secondly, the responses of transparency of research assessment in universities are shown in Table 3. Bulgarian institutions are very similar, if not better to other European institutions according to the EUA report, while Albanian Institutions are far behind with more than half agreeing that such information is not available at all.

Table 3: Research Assessment Transparency

Research Assessment Transparency	**Bulgarian institutions**	**Albanian institutions**	**EUA report**
Publicly Available	75%	15%	63%
Internally Available	25%	31%	34%
Not Available	0%	54%	3%

Thirdly, each academic activity within the institutions is shown in Table 4. The results show that Bulgarian institutions focus mostly on research publications, research output data, and attracting external funding as the main elements for carriers in research. Bulgarian institutions value almost all the academic activities for careers in research more than the average European institution. In comparison, Albanian institutions' primary focus is teaching and mentoring activities; research publications are also important. Lastly, both countries focus more on Open Science and Open access for evaluating careers in research than the average European respondents of the EUA report.

Table 4: Most important academic activities within institutions

Research Assessment Transparency	**Bulgarian Institutions**	**Albanian institutions**	**EUA report**
Research publications	98%	69%	93%
Research Output Data	88%	38%	48%
Attracting External Funding	88%	31%	81%

Research Supervision Activities	78%	53%	63%
Research impact	78%	40%	64%
Research collaborations outside of academia	76%	46%	57%
Teaching Activities	68%	77%	62%
Research collaboration within academia	66%	46%	63%
Research Networking	63%	38%	57%
Social Outreach	53%	31%	45%
Open Science and Open Access	51%	46%	38%
Mentoring Activities	43%	61%	47%

Table 5 and table 6 show us the publication metrics and social outreach metrics used for evaluation within their institutions. Table 5 shows that 98% of the Bulgarian respondents' institutions and 92% of Albanian respondents' institutions use Journal Impact Factor (JIF) for the measurement of the research output of the researchers followed by 80% usage of SCImago Journal Rank (SJR) in Bulgaria and 23% in Albania. h-index is used in 72% of Bulgarian Institutions and 23% in Albanian Institutions. All the other publication metrics were not used besides CiteScore (4%) in Bulgaria. In comparison, EUA shows that 75% of European institutions use JIF, 70% use h-index and 31% use SCImago Journal Rank. Even though metric usage is very limited, both countries have a unified way of evaluating academic activities for research careers.

Table 5: Publication Metrics used

Publication metric	**Bulgarian institutions**	**Albanian institutions**	**EUA report**
Journal Impact Factor	98%	92%	75%
SCImago Journal Rank	80%	23%	70%
h-index	72%	23%	31%

Table 6 shows that in Bulgaria 78% use Data Citations, 28% use ResearchGate views, 13% use ImpactStory and 8% use Datacite. There were also 8% of respondents that don't know which is the altmetric used for the social outreach of research outcomes. In Albania, 77% of the respondents didn't know which was the altmetric used for social outreach of research outcomes. The other respondents show that 23% used Data Citations followed by 15% usage of ImpactStory. In the EUA report, 46% use Research Gate views, 35% use Data Citations, 13% ImpactStory, and 10% use Datacite.

Table 6: Altmetric used

Altmetric	Bulgarian institutions	Albanian institutions	EUA report
Data Citations	78%	23%	35%
ResearchGate Views	28%	0%	46%
ImpactStory	13%	15%	13%
Datacite	8%	0%	10%
Don't know	8%	77%	25%

3.3. Open Science assessment results

In this section, the Open Science assessment of responses from the questionnaire is shown. Table 7 states to us the opinion of researchers from Albania and Bulgaria to whom the science should be open. All (100%) agreed that science should be open to scientists from the same discipline. 87% of Bulgarian and 84% of Albanian respondents agree that science should be open to scientists of other disciplines. 55% of respondents of both countries agreed that science should be open to all citizens; civils and social organizations. 72% of Bulgarian respondents and 84% of Albanian respondents think science should be open to specially concerned groups. 68% of

Bulgarian respondents and 92% of Albanian respondents think science should be open to funders and policy makers. 85% of Bulgarian Respondents and 92% of Albanian respondents agreed that it should be open to industry and companies.

Table 7: To whom should science be open

Group	Bulgarian institutions	Albanian institutions
Scientists from the same discipline	100%	100%
Scientists from other disciplines	87%	84%
All Citizens	55%	55%
Civil and social organizations	55%	55%
Specially concerned groups	72%	84%
Funders and Policy makers	68%	92%
Industry and companies	85%	92%

Table 8 and 9 show us the reasons for Open Science and against it, in Bulgarian and Albanian institutions. 83% of the Bulgarian participants believe that efficiency, which means the sharing of data, procedures for optimizing science is a very important reason why should science be open. 78% of these participants believe that Rigour, which is open access, open data, and makes replicability easier, is a very important reason why science should be open; 83% believe that new and innovative economic possibilities are important for open science. 75% believe that equity (access to all scientific results methods or software) and ethics are very important reasons why should science be open; 72% believe that fairness is also important for open science; 70% believe that impact is a very good reason for open science; 48% believe that diversity is an important reason for open science. Albanian responses on the other side show that almost all reasons for open science are important. New economic possibilities, efficiency, and ethics are considered very important by all (100%) of the participants.

Table 8: Reasons for Open Science

Reasons for Open Science	**Bulgarian institutions**	**Albanian institutions**
Diversity	48%	77%
New Economic Possibilities	67%	100%
Efficiency	83%	100%
Equity	75%	92%
Ethics	75%	100%
Fairness	72%	85%
Impact	70%	77%
Rigour	78%	77%

Table 9: Reasons against Open Science

Reasons Against Open Science	**Bulgarian institutions**	**Albanian institutions**
Not a priority now	20%	15%
Public's lack of understanding	30%	15%
Public is not ready now	27%	15%
Risk to fundamental research	27%	23%
Interference with research integrity	32%	23%
Danger and potential misuse	32%	15%
Unfairness	32%	15%
Low quality	35%	15%

As shown in table 9, only 35% of Bulgarian respondents believe that low quality is an important reason against open science. 32% believe that potential misuse, unfairness, and interference with research integrity are very important reasons against open science. 30% agree that public's lack of understanding is also important; 27% agree that the public is not ready yet; 27% believe that open science is a risk to fundamental research. 20% believe open science is not a priority right now. On the other hand, 39% of Albanian respondents believe that low quality of research is an important reason against open science, 23% believe that interference with research integrity and risk to fundamental research are important reasons. All the other reasons are considered each 15% against open science.

In table 10 the challenges to performing Open Science are summarized. 78% of Bulgarians responded that budget or funding constraints are an important challenge to open science in their institution; 65% believe that authentic public engagement to open science is a challenge; 60% believe that fear about the value of open science is an important challenge; 52% believe that time constraint is very challenging as well; 50% believe that lack of clear steps to follow for open science is a challenge; 45% believe that the lack of proper infrastructure for open science is an important challenge. In Albanian institutions, the situation is almost the same, with an emphasis on budget and funding constraints as 69% of respondents think it is the most important challenge to open science in their institutions.

Table 10: Challenges to perform Open Science within institutions

Challenges to Open Science	Bulgarian institutions	Albanian institutions
Lack of proper infrastructure	45%	46%
Lack of clear steps to follow	50%	46%
Authentic public engagement	65%	46%
Budget and funding constraints	78%	69%
Time constraints	52%	54%
Fears and uncertainties for career development	60%	54%

In table 11, a summarization of how much training respondents would like to receive is shown. It is clear that all participants would like to receive more training for almost all the aspects of open science. This is especially true for Albania.

Table 11: Training needed in Open Science

Training in Open Science	Bulgarian institutions	Albanian institutions
Research and data management	70%	100%
Research integrity	63%	85%
Research publishing and dissemination	53%	77%
Collaborating and networking	48%	92%
Communicating science to the general public	68%	92%
Evaluation of research projects and researchers	58%	85%
Assessment of the impact of initiatives in public	85%	77%

3.4. Technical compliance assessment

In this section, technical details of the digital repositories and/or institutional repositories are discussed. Participants responded if their institution uses a Current Research Information System (CRIS). Only 10% of the Bulgarian institutions use CRIS and none is CERIF compliant. No CRIS is used by Albanian institutions.

When asked about the protocols, standards, and vocabularies used in their CRIS, ORCID was mentioned by Bulgarian respondents. Even though most of them do not use CRIS, 58% of Bulgarian institutions store institutional Research outputs such as publications, patents, or products in the Institutional Repository (IR); 13% of Bulgarian respondents didn't know if such a system was used in their institution; 29% answered they don't use IR at all. Figure 3-b shows the types of content stored in Bulgarian Institutional Repositories.

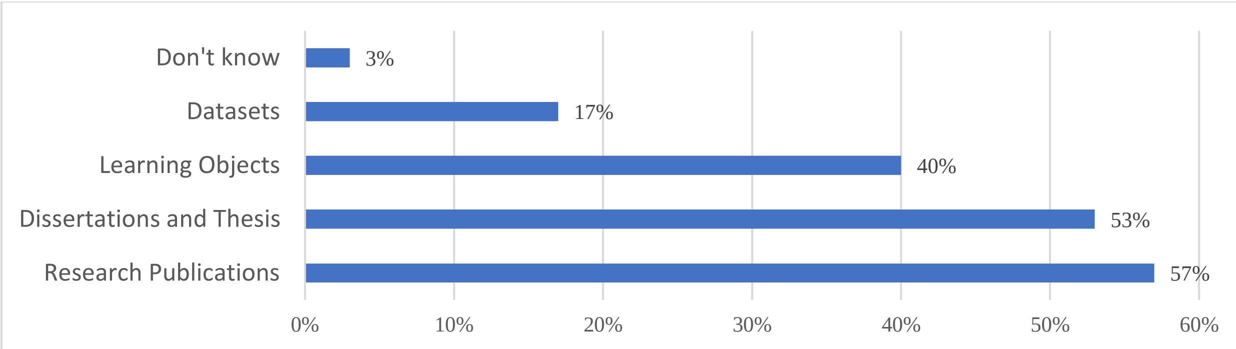

Figure 3-b: Types of content stored in Bulgarian Institutional Repositories

On the other hand, 38% of Albanian respondents answered that they use IRs; and 62% that their institution doesn't use such a system. When asked if their IR is OpenAIRE-compliant, only 5% of Bulgarian respondents answered positively and all of the Albanian respondents answered negatively.

When asked if their IR stored and registered researchers' persistent digital identifiers like ORCID, 32% of Bulgarian respondents didn't have such information while 42% answered positively. 92% of Albanian respondents answered negatively and 8% didn't have such information.

Finally, the participants were asked if an Open Access policy or mandate is used on their institution. Table 12 shows again different results between the countries. 62% of Albanian respondents didn't have it and 38% didn't know if such a mandate exists. In Bulgarian institutions, 45% have an open access policy/mandate and 32% didn't know if there is an open access policy/mandate to their institution.

Table 12:Open Access Policy or Mandate

Open Access Policy/Mandate used in institutions	**Bulgarian institutions**	**Albanian institutions**
Yes	45%	0%
No	23%	62%
Don't know	32%	38%

3.5. Conclusions

The survey provides a review of the current state of the transition to open science at Bulgarian and Albanian institutions. The conclusions of the survey show that Bulgarian and Albanian institutions support Open Science. Some key findings include:

- The majority of Bulgarian institutions have publicly available research assessment policy similar to other EU institutions, while Albania is far behind with more than half of the institutions without a research assessment policy.
- None of the Albanian universities have an Open Access mandate or the respondents don't know if such mandate exists.
- Albanian institutions don't have CRISs. Even in Bulgarian institutions, only 10% have such systems and none of them is CERIF compliant.

From the survey, it is clear that both Bulgarian and Albanian institutions are still at the early stages of transition towards Open Science except for small cases in Bulgaria. Bulgaria has already started to think and work towards an opening of Bulgaria science to the world. They have planned the Bulgarian Open Science Cloud (BOSC), based on the same principles, standards, and technologies of EOSC and fully compliant with it. They also have the first prototype of the Bulgarian Open Science Digital Repository.

The survey points us to the need for the Albanian Open Science Cloud (AOSC). The goal of AOSC is to make Albanian Science open and to help Albania join the EOSC initiative. In the next chapter, the deployment of the first prototype of AOSC is shown, created in a way to be fully compliant with the EOSC. The Albanian Open Science Cloud will provide the following principles:

- A unified research infrastructure to provide a better culture of research data management among scientists in Albania
- Data that is accessible, interoperable, and reusable (FAIR data).
- Tools to secure a reliable environment for FAIR data, FAIR data conversion services.

- An interface to work with data from different disciplines as well as an entry point with unique identification and authentication service.
- Information to open-source policies, information on FAIR data, guidelines for EOSC and OpenAIRE compliance.
- Services to find, access, reuse and analyze research data generated by other researchers with appropriate data sets compliant with EOSC.

The first step towards the Albanian Open Science Cloud is to provide an Open Access Research Digital Repository to preserve research outcomes and assets, compliant to EOSC. This is an Open-Source digital library that will be the first step to a fully-fledged Albanian Open Science Cloud. The Albanian Open Access Digital Repository has the following characteristics:

- Open-Source digital repository software DSpace-CRIS will be used to develop such an element.
- After further checking the data stored at Albanian universities, only the names of authors and publications can be extracted from the existing data. This means that the population of the repository is an important challenge on its own. On the other hand, EOSC and OpenAIRE compliant metadata can be used from the start without facing the difficulty of transforming the original data into FAIR data.
- The metadata to be used is CERIF data model that is compliant with OpenAIRE and EOSC. Only a subset of CERIF model is used since CERIF metadata model includes a wider range of entities that is not relevant to OpenAIRE.

For populating the digital library, we will start by checking the data on the Ministry of Education, Sports and Youth in Albania, and creating a list of all the Albanian Universities and researchers. Using this list, external bibliographic resources and dataset resources (like WoS, Scopus) are searched to provide a list of publications or metadata. The data imported from external bibliographic resources are then sent to universities and researchers to clean or provide additional metadata information. After that, additional information like altmetrics will be searched on the internet and added to different entities. Publication repositories are then accessed and downloaded as full-text information.

The next step is finding persistent digital identifies (like DOI, ORCID). This is performed by using regular expressions with CrossRef REST API to find DOI, and ORCID Public API to find ORCID identifiers for each of the researches on the DOI publication. Using DOI resolution services, additional information about the object (like URLs) is added to enrich the bibliographic metadata of such an object.

Finally, all the data is normalized to reflect the data model used in this repository to fully store the data following the EOSC and OpenAIRE guidelines for metadata.

Chapter 4. Albanian Open-Source Digital Library

In this chapter, the development of the Albanian Open-Source Digital Library fully compliant to EOSC is described. An analysis of the requirements to be compliant is done. Then, the prerequisite software will be presented and the build of the digital repository using DSpace-CRIS. After that, the data model and the design of the digital Library compliant to EOSC are illustrated. Lastly, an Open Science Policy to be implemented is developed.

4.1. Analysis of the requirements

The previous chapter showed us that Albanian researchers fully support Open Science. However, most of the institutions don't have a digital repository, let alone a compliant repository for Open Science. For that, an Open-Source Digital Repository is the next step towards Open Science in Albania. To build such a system to be fully compliant with EOSC, some decisions should be made.

Firstly, the system is built upon DSpace-CRIS. It is the first open-source Digital Repository for research institutions. The main reason why such a system is selected is the fact that it is compliant with international standards used in EOSC. These standards include facilitation for interoperability and data transfer. The standards adopted are ORCD, OpenAIRE guidelines for literature Repository management, PlanS by Coalition S, and FAIR data principles.

DSpace-CRIS allows different data models to be implemented, thus EOSC compliant metadata can be used. The main entities of DSpace-CRIS include Researchers, Projects, Communities and Collections, Organizations, and Research Outputs. Figure 4-a shows the current elements of DSpace-CRIS implemented in Albania. The system can be accessed at https://albanian-cris.info/[50].

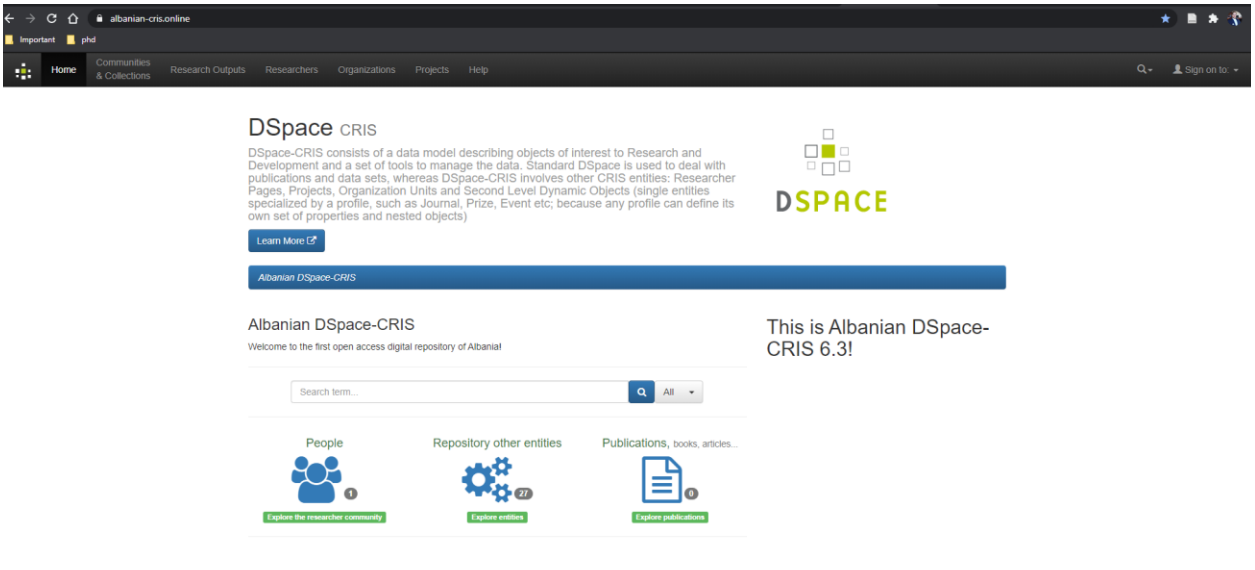

Figure 4-a: Albanian DSpace-CRIS

As stated before, no system is present in Albanian countries, which means that all data will have to be populated manually. The best option available is to create a list of all the educational institutions in Albania. Such a list is shown in Table 13 below with all the known data from such institutions. The information is checked with the Ministry of Education, Sports, and Youth in Albania.

Table 13: List of Institutions in Albania

Institution Name	Description	Type of Institution	Copyright
Academy of Albanological Studies	ASA	Public	© Akademia E Studimeve Albanologjike
Agricultural University of Tirana	UBT	Public	© Universiteti Bujqesor i Tiranes
Albanian University	albanianuniversity	Private	© Albanian University
Aldent University	UAL	Private	© Universiteti Aldent
Aleksandër Moisiu University of Durrës	UAMD	Public	© Universiteti Aleksandër Moisiu

Aleksandër Xhuvani University	UNIEL	Public	© Universiteti i Elbasanit "Aleksandër Xhuvani"
Armed Forces Academy	AFA	Public	© Akademia e Forcave të Armatosura
Bedër University	BU	Private	© Kolegji Universitar Bedër
Canadian Institute of Technology	CIT	Private	© CIT
Epoka University	EU	Private	© Epoka University
Eqrem Çabej University	UOGJ	Public	© UOGJ
European University of Tirana	UET	Private	© UET
Fan Noli University	UNKORCE	Public	© Universiteti "Fan S. Noli", Korçë
Luarasi University	LUARARI UNIV	Private	© Universiteti Luarasi
Marin Barleti University	UMB	Private	© Universiteti Barleti - Tiranë
Mediterranean University of Albania	UMSH	Private	© UMSH
Metropolitan University of Tirana	UMT	Private	© Metropolitan University of Tirana
Our Lady of Good Counsel University	UNIZKM	Private	© Our Lady of Good Counsel - Catholic University
Polis University	universitetipolis	Private	© UNIPOLIS
Polytechnic University of Tirana	UPT	Public	© UPT
Sports University of Tirana	UST	Public	© Universiteti I Sporteve Të Tiranës
University of Arts	UART	Public	© Universiteti i Arteve
University of Medicine, Tirana	UMT	Public	© UMT

University of New York Tirana	UNYT	Private	© University of New York Tirana
University of Shkodër "Luigj Gurakuqi"	UNISHK	Public	© University of Shkodër "Luigj Gurakuqi"
University of Tirana	UNITIR	Public	© UNITIR
University of Vlorë "Ismail Qemali"	UNIVLORA	Public	© Universiteti i Vlorës

The first step to populate the digital library is to start by checking the data on the Ministry of Education, Sports and Youth in Albania and creating a list of all the researchers in Albanian universities. Using this list, external bibliographic resources and dataset resources (like WoS, Scopus) are investigated to provide a list of publications or metadata. The data imported from external bibliographic resources are then sent to universities and researchers to provide additional metadata information. After that, additional information like altmetrics will be searched on the internet and added to different entities. Publication repositories are then accessed and downloaded as full-text information.

The next step is finding persistent digital identifies (like DOI, ORCID). This is performed by using regular expressions with CrossRef REST API to find DOI, and ORCID Public API to find ORCID identifiers for each of the researches on the DOI publication. Using DOI resolution services, additional information about the object (like URLs) is investigated and added when possible. This will enrich the bibliographic metadata of such an object.

Finally, all the data is normalized to reflect the data model used in this repository to fully store the data following the EOSC and OpenAIRE guidelines for metadata.

These Universities are stored and created as communities and also Research Organizations in the Open Access Digital Repository as shown in Figure 4-b. After that, for each University a list of Researchers is created. The list is then sent to each University for a double-check of the data.

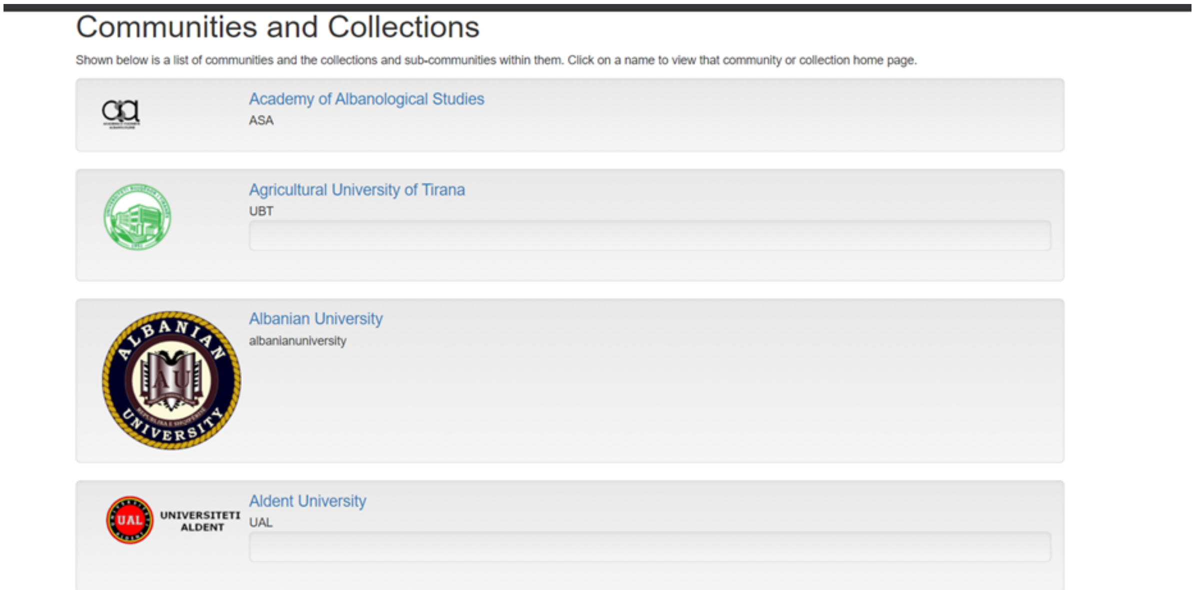

Figure 4-b: Sample of the List of Universities at Albanian Open Access Digital Repository

4.2. Prerequisite Software

Since DSpace-CRIS will be the system to be deployed, an environment is created to fully support the running of DSpace-CRIS. Starting with the operating system, an open-source Linux distribution was chosen. Ubuntu 18.04 LTS Server was installed which has support until 2023. Even though Ubuntu 20.04 LTS has been announced, it did not opt as a good choice since some of the dependencies needed for DSpace-CRIS were not fully functional at the moment of the deployment. The Operating System will be updated once the dependencies are bug-free. After installing the operating system into the server, the dependencies are installed and updated. The required dependencies are Java SE Development Kit 8 which is recommended for DSpace-CRIS installation as a development environment for building Java applications and components.

After that, PostgreSQL was installed which is needed for the deployment of the database. The version installed is 10.15. Ant and Maven were also installed to deploy the build of DSpace into the server. Lastly Tomcat 8 was installed to deploy the web apps to access the web interface of DSpace-CRIS.

Next, a user is added into Ubuntu called "dspace" to own the directory of the installation of DSpace-CRIS. A postgres user was also created, assigned as the user of the database to be created directly after the user. The database is needed for DSpace, thus the config file of PostgresSQL was modified to allow the user to access such database.

Following the next step, DSpace-CRIS is freshly installed and deployed. After the deployment of DSpace-CRIS, webapps needed were copied into the Tomcat, and Tomcat is initialized. After finishing all these steps, DSpace-CRIS can be accessed from http://localhost:8080/jspui.

The next step was to have a domain in use for the DSpace-CRIS. The domain was bought and configured to be http://albanian-cris.info:8080/jspui. To make the link more attractive, the Apache HTTPD server is installed and configured. The configuration is to make the redirection to the above link. Finally, an SSL certificate is needed to make the connection secure. The SSL certificate is configured for the system using Open digital certificates from Let's Encrypt (Let's Encrypt - Free SSL/TLS Certificates, no date). Let's Encrypt is a nonprofit Certificate Authority providing TLS certificates in support of Open Source. It is a free and automated CA that is open and runs for the public's benefit. It is provided by the Internet Security Research Group (Internet Security Research Group, no date) which allows digital certificates to be free and automatically updated. To perform such automatic renewal, a crontab is configured on the server to update the certificate every 3 months. Figure 4-c shows us the certificate details of the deployed system. The final access link to the deployed DSpace-CRIS system is finally https://albanian-cris.info/.

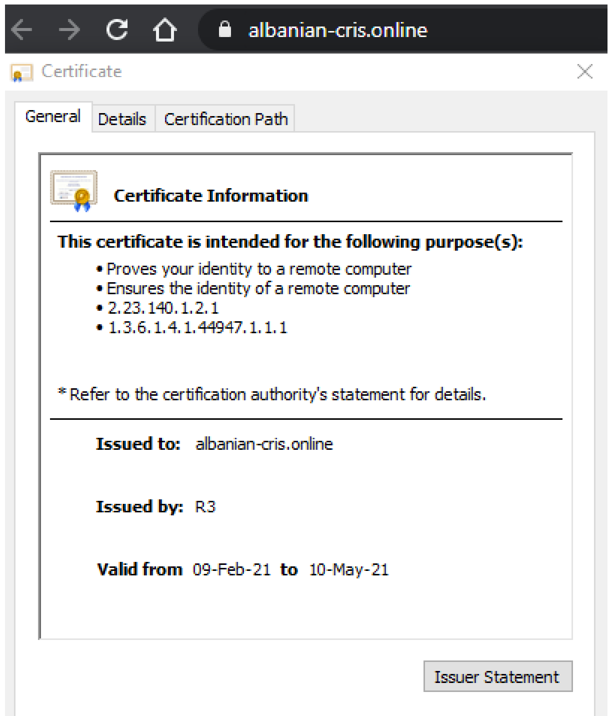

Figure 4-c: Certificate Information of https://albanian-cris.info

4.3. DSpace-CRIS deployment

DSpace-CRIS 6.3 is deployed into the server. The chosen DSpace-CRIS is the latest version at the moment of deployment. By default, DSpace-CRIS can be used to configure and maintain research data from the UI. The research data that can be input are the Organization Units, Researchers, Publications, Projects, and Communities. By default, even the possibility to enter equipment is possible from the User Interface. These are the preconfigured CRIS entities that can be customized and linked together.

Another feature that is by default is the possibility of assessment, reporting, and analytics of the whole repository system. That means that communities, organization units, researchers, or publications can show the number of downloads, the number of citations, and also Top/Global downloads. A global search and indexing are present too which helps with analytics.

Research Profile can be managed and promoted too. For example, the researcher can hide/show their profile and claim his profile. Figure 4-d shows an example of a Researcher Profile edit.

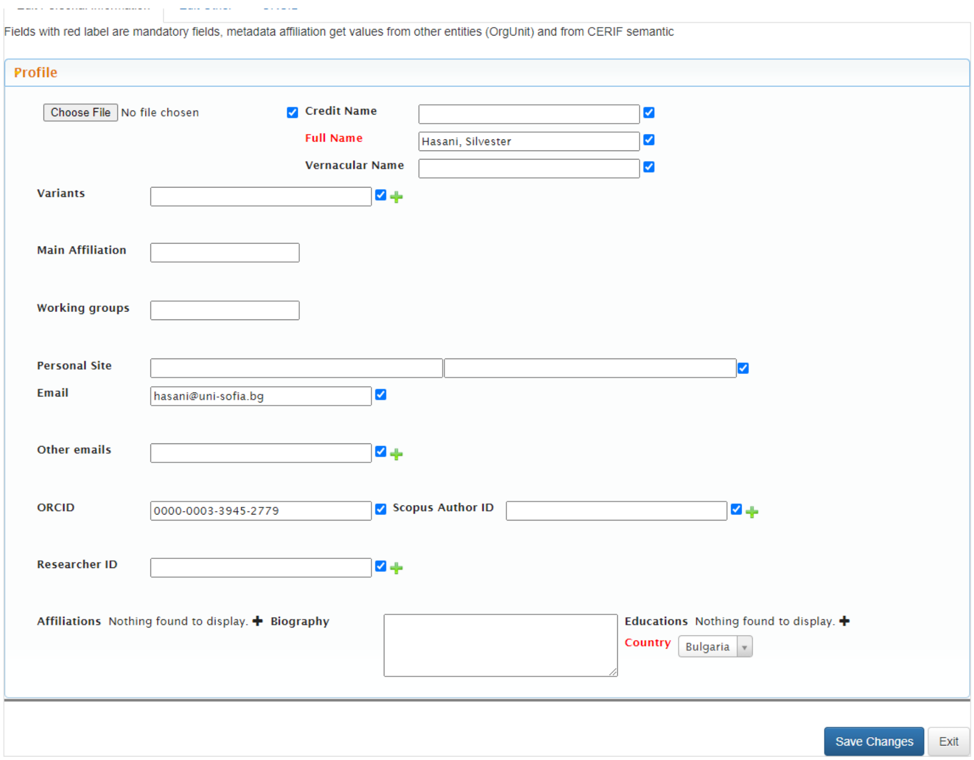

Figure 4-d: Edit Research Profile

Lastly, the DSpace-CRIS can offer interoperability with external systems and databases. Scripts are provided to query periodically bibliographic databases of Scopus and Web of Science. Also, ORCID integration for both public and member APIs. This can be shown in Figure 4-e where integration with ORCID is configured.

Figure 4-e: ORCID integration and login

It is possible to log in using an account created by the administrator of the system, or a new user registered into the system but also a Login using your ORCID digital identifier. Figure 4-f shows a sign-in by using the ORCID profile. Once you have logged in using the ORCID profile you can manage the researcher profile created at the repository. If you decide your data to be shown to the repository you have to set your researcher profile to public.

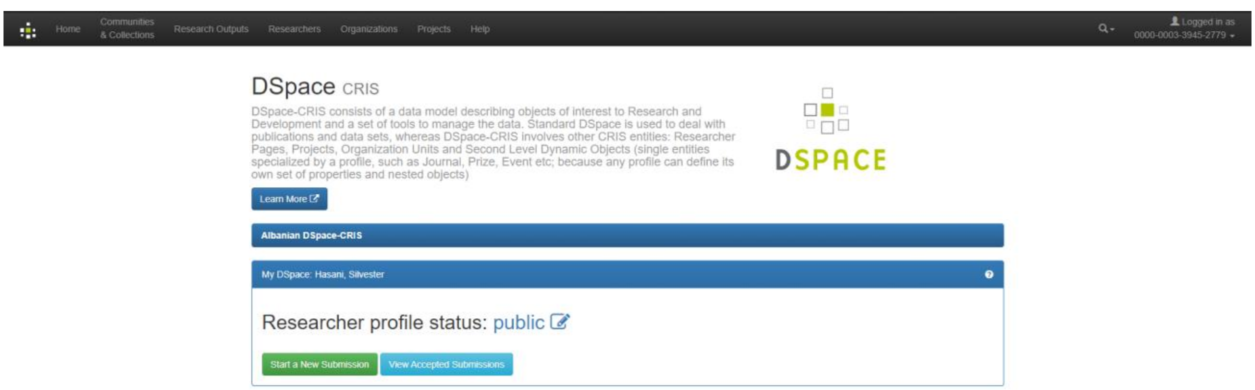

Figure 4-f: Login with ORCID

To have the repository fully functional, crontab jobs should be configured to retrieve information from external databases, compute and find statistical data, and also send subscription notifications. Table 14 shows all the crontab jobs running in the Albanian-CRIS configuration.

Table 14: crontab jobs at Albanian-CRIS

Crontab	Function	Time of Run
$DSPACE/bin/dspace generate-sitemaps > /dev/null	To make the content of the repository more findable in search engines.	Every 8 hours.
$DSPACE/bin/dspace oai import > /dev/null	Ensures that new content is available via OAI-PMH	Daily
$DSPACE/bin/dspace index-discovery > /dev/null	Ensures that deleted documents are cleaned from the repository search and index features	Daily
$DSPACE/bin/dspace index-authority > /dev/null	Ensure that Solr Authority is up to date	Daily
$DSPACE/bin/dspace stats-util -i	Removed known web spiders from the usage statistics	Daily
$DSPACE/bin/dspace sub-daily	Sends updates to users subscripted to daily notifications of different collections	Daily
$DSPACE/bin/dspace filter-media	Ensures that full text is available to newly added documents	Daily
$DSPACE/bin/dspace curate -q admin_ui	Ensures that any curation task set up by the administrator is executed	Daily
$DSPACE/bin/dspace checker -l -p	Verifies checksums of the files stored in the repository	Weekly

$DSPACE/bin/dspace checker -d 1h -p	Checks if the content is available to the repository	Weekly
$DSPACE/bin/dspace checker-emailer	Checks if the email config is running	Weekly
$DSPACE/bin/dspace cleanup > /dev/null	Permanently deletes all the bitstreams flagged Deleted	Monthly
$DSPACE/bin/dspace stats-util -s	Changing Solr indexing to improve performance	Yearly
sh $DSPACE /bin/tomcat.sh	Ensures that the web service is available for the users, if not it starts it.	Daily

Finally, as shown above, DSpace-CRIS digital repository system is configured with ORCID integration, where researchers can easily log in, modify and claim their research profile with the help of their ORCID. Next, integration to allow database import of bibliographic Records from Scopus and Web of Science is completed. Albanian-CRIS is configured such that real-time import of bibliographic records searching the external database or Web of Science and Scopus is done by the usage of identifiers, authors, or titles of the research entity. After that Periodic scanning of the external database to retrieve institutional publications to import. This is done by the scripts available out of the box from the DSpace-CRIS repository system available at the official website of the configuration of the system [51].

4.4. Data Model Deployed in Albanian-CRIS

The key feature of DSpace-CRIS is the fact that every entity data structure can be configured by a User Interface, by adding simple and complex fields. Relationships can be formed between new and old entities configured into the repository system. Once the data model has been configured, all entities have a public page, where some or every information may be shown, and they may be searched and browsed. Also, every relationship present in the data model of the repository system can be explored since they are connected together. By that, you can jump from Researcher Profile to organization units, projects, publications, and vice versa by using the relationship created into the repository system.

Figure 4-g shows the ER diagram of the DSpace-CRIS deployed into the Albanian-CRIS which is the default core of DSpace-CRIS. JDYNA_VALUES is used to manage all the relationships between all of the research entities configured into the system. CRIS_RPPAGE shows the data model and fields of the Researcher Profile, CRIS_PROJECT is the data model of the projects managed into the repository system, CRIS_ORGUNIT is the data model of organization units and CRIS_DO is the data model of Research Objects which include publications, equipment, and all the other research entities related to the dissemination of research information.

Figure 4-h illustrates the ER diagram of the tables related to Organization Units, Researcher Profile, Projects, and all the Research Entities. This shows the relationship of all the Research Entities and how you can jump from Researcher Profile to organization units, projects, publications, and vice versa.

Lastly, the configuration tables of the Organization Unit entity are shown in the ER diagram of figure 4-i.

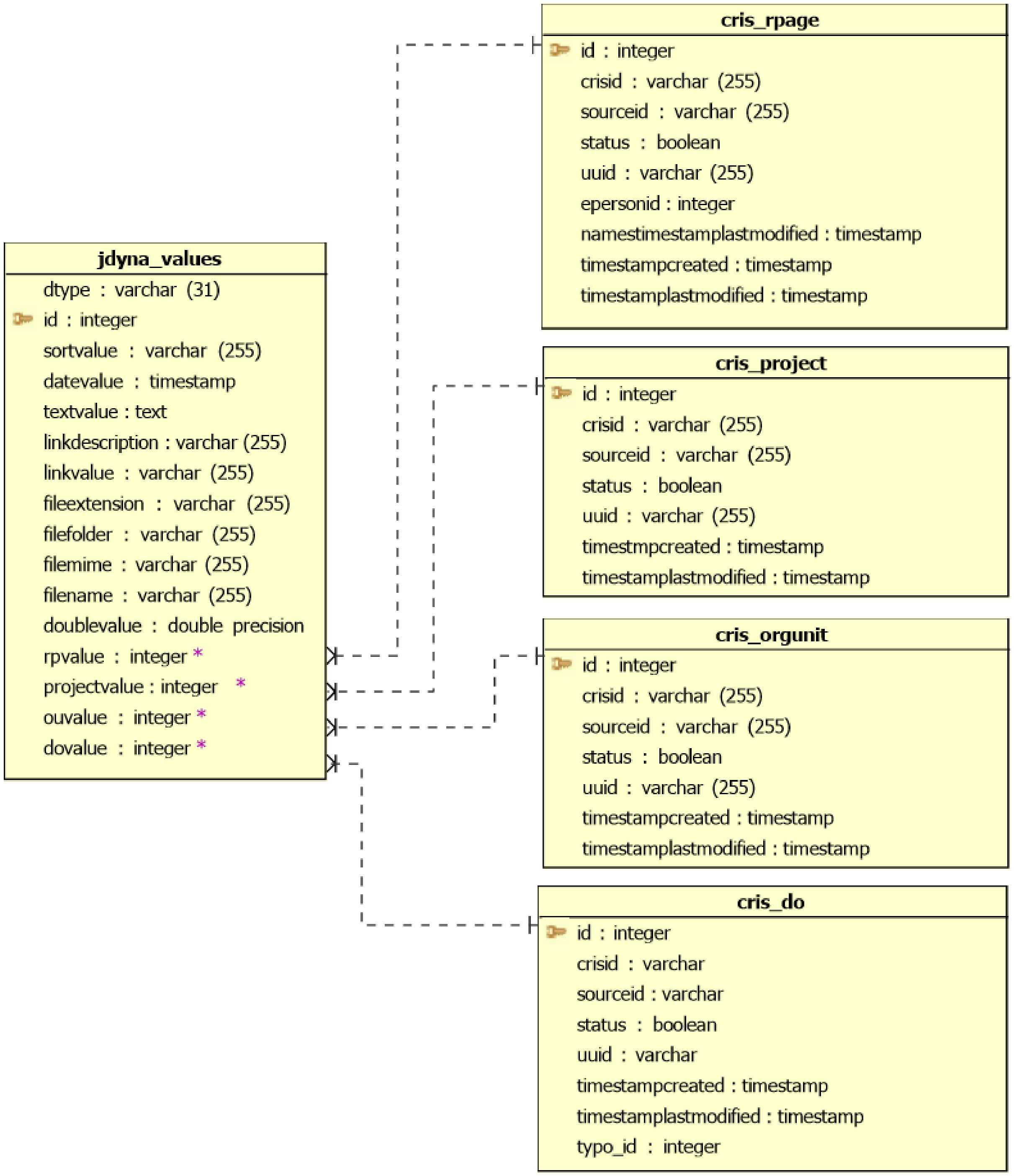

Figure 4-g: Albanian-CRIS core structure

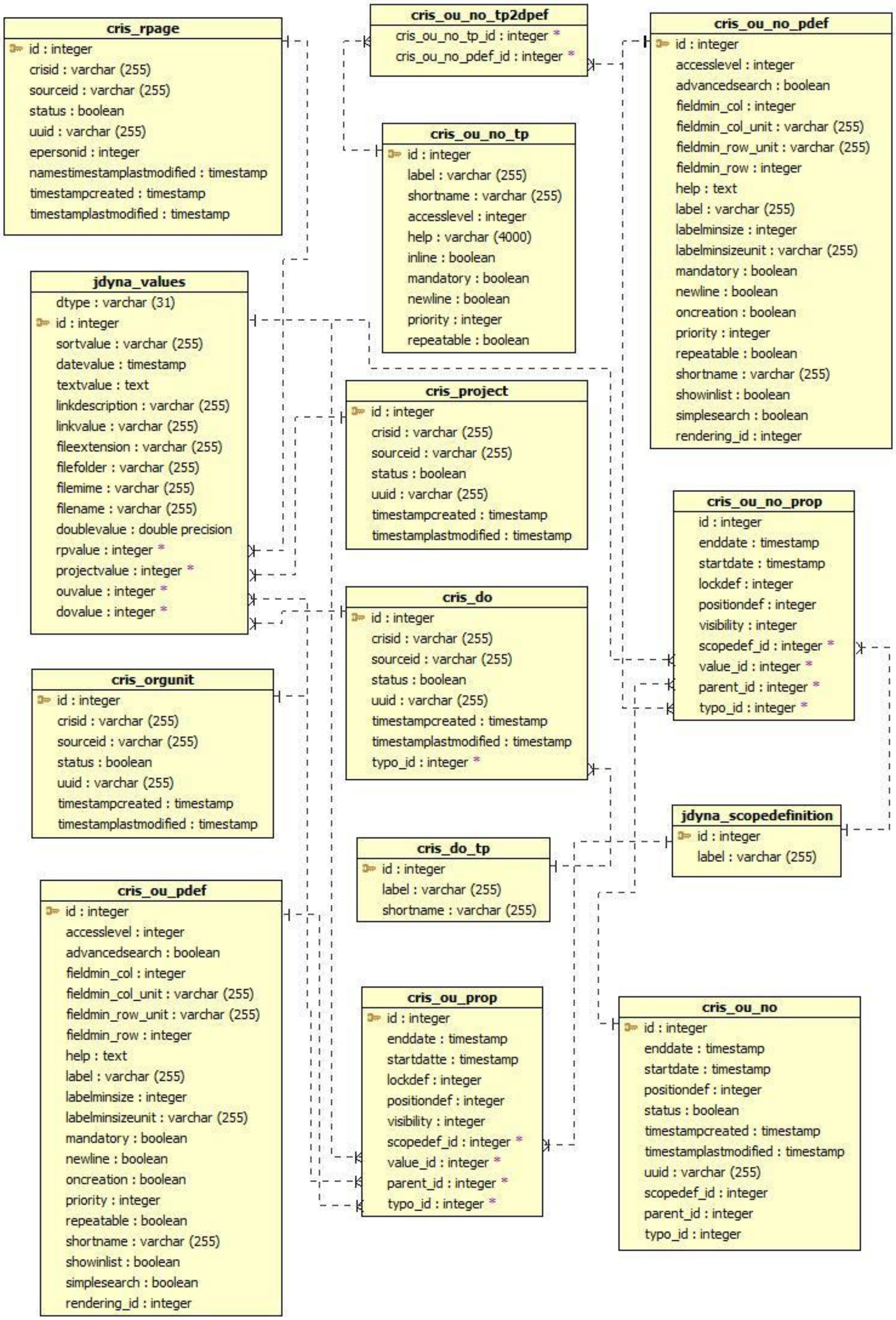

Figure 4-h: ER diagram of Albanian-CRIS

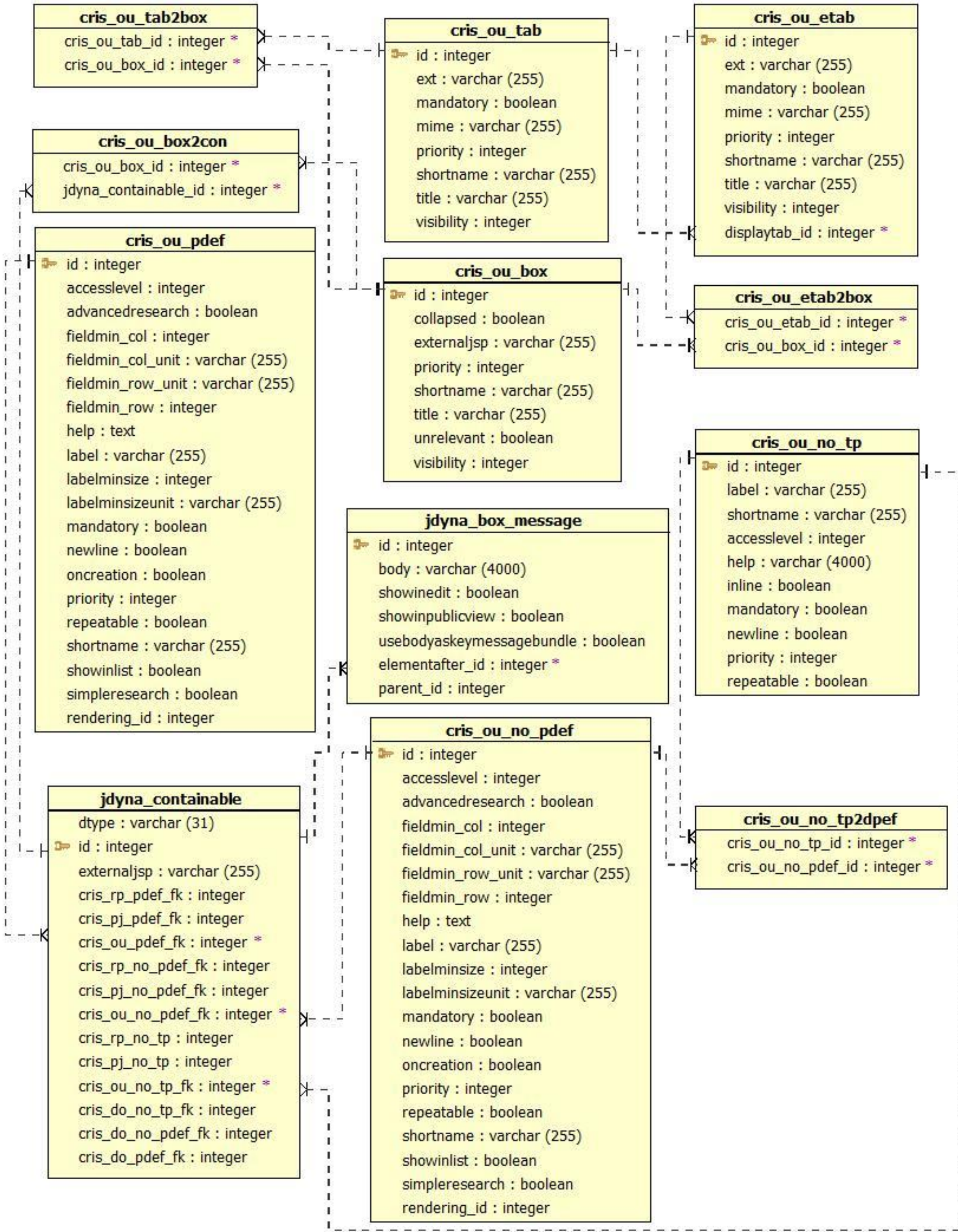

Figure 4-i: Configuration Tables of the Organization Unit entity

Tables 15-20 show all the fields of each research entity configured into the Albanian-CRIS. The fields and the general structure provide a simplified implementation of the key CERIF entities and concepts. For comparison, Researcher Profile is the People entity in CERIF mapping format, Organization Unit is OrgUnit, and Publications are the Publication Objects part of DSpace-CRIS.

Table 15: Fields used in Equipment Research Entity

Fieldname	Description	Mandatory
acronym	Acronym	No
identifier	Institution Unique Identifier	No
name	Name	No
ownerou	Owner (Organizations)	No
ownerrp	Owner (Persons)	No
description	Equipment Description	No

Table 16: Fields used in Events Research Entity

Fieldname	Description	Mandatory
acronym	Acronym	No
name	Event name	Yes
type	Event type	No
startdate	Start date	No
enddate	End date	No
location	Location	No
iso-country	Country	No
organizerou	Organizer of the Event	No
organizerpj	Organizer of the Event (project)	No

sponsorou	Sponsor of the Event	No
sponsorpj	Sponsor of the Event (project)	No
partnerou	Partner of the Event	No
partnerpj	Partner of the Event (project)	No
description	Event Description	No
keywords	Keyword(s)	No

Table 17: Fields used in Journals Research Entity

Fieldname	Description	Mandatory
picture	Picture of the Journal	No
name	Name	Yes
keywords	Subject Classifications	No
issn	ISSN	No
description	Description	No

Table 18: Fields used in Organization Unit Research Entity

Fieldname	Description	Mandatory
logo	Logo of the organization Unit	No
name	Organization name	Yes
director	Director	Yes
parentorgunit	Parent OrgUnit	No
date	Established	No
boards	Scientifics Board	No
city	City	Yes
crossrefid	Crossref Identifier	No
description	Description	No
iso-country	Country	Yes

Table 19: Fields used in Projects Research Entity

Fieldname	**Description**	**Mandatory**
acronym	Project Acronym	no
logo	Logo of the Project	no
title	Project title	yes
coordinator	Consortium Coordinator(s)	No
code	Internal project ID	No
partnerou	Partner Organization(s)	No
principalinvestigator	Project Coordinator	Yes
awardURL	Award URL	No
funder	Funder	No
fundingProgram	Funding Program	No
openaireid	OpenAIRE	No
contractorou	Contractor Organization(s)	No
memberou	Member Organization(s)	No
projectURL	Project URL	No
oamandate	OA Mandate	No
expdate	Expected Completion	No
status	Status	No
grantcurrency	Currency	No
grantamount	Amount	No
grantidentifier	Grant number	No
grantduration	Duration	No
granttype	Grant type	No
startdate	Start date	No
abstract	Abstract	No
coinvestigators	Co-Investigator(s)	No
keywords	Keyword(s)	No

Table 20: Fields Used in Researcher Profile Research Entity

Fieldname	Description	Mandatory
personalpicture	Personal picture of the researcher	No
preferredName	Credit Name	No
fullName	Full Name	Yes
policy	Policy	No
translatedName	Vernacular Name	No
variants	Variants	No
dept	Main Affiliation	No
system-orcid-profile-pref-biography	biography	No
workgroups	Working groups	No
personalsite	Personal Site	No
email	Email	No
otheremails	Other emails	No
interests	Interests	No
orcid	ORCID	No
scopusid	Scopus Author ID	No
authorid	Researcher ID	No
written	Written Languages	No
affiliation	Affiliations	No
biography	Biography	No
education	Educations	No
iso-country	Country	Yes
orcid-profile-pref-authorid	external-identifier-Researcher ID	No

orcid-profile-pref-iso-country	iso-3166-country	No
orcid-profile-pref-keywords	keywords	No
orcid-profile-pref-personalsite	researcher-urls	No
orcid-profile-pref-scopusid	external-identifier-Scopus Author ID	No
orcid-profile-pref-variants	other-names	No
orcid-projects-prefs	Projects imported by orcid	No
orcid-publications-prefs	Publications imported by orcid	No
orcid-push-crispj-activate-put	PUT mode enable for projects	No
orcid-push-crisrp-activate-put	PUT mode enable for profile	No
orcid-push-item-activate-put	PUT mode enable for publications	No
orcid-push-manual	Synchronization mode	No
qualifications	Qualifications	No
spoken	Spoken Languages	Yes
system-orcid-profile-pref-email	primary-email	No
system-orcid-profile-pref-fullName	name	No
system-orcid-profile-pref-otheremails	other-emails	No
system-orcid-profile-pref-preferredName	credit-name	No
system-orcid-token-activities-update	ORCID authorization scope (internally used)	No

system-orcid-token-authenticate	ORCID authorization scope (internally used)	No
system-orcid-token-person-update	ORCID authorization scope (internally used)	No
system-orcid-token-read-limited	ORCID authorization scope (internally used)	No
orcid-profile-pref-affiliation	affiliations-employment	No
orcid-profile-pref-education	affiliations-education	No

4.5. Open-Science Policy

In this section, we present the first Open-Science Policy which is going to be implemented in Albania. Survey results also stated that no Open-Access Policy or Mandate has been implemented in Albania. Following a collaboration with Albanian Universities, a policy to be implemented is created as below. The goal of this Open-Access Policy is to follow the trend of Open-Science in Albania. These universities are currently in the process of implementing such a Policy. The mandate aims the Open Science and sees it as a key element for increasing the impact of Research on the Albanian Community. It is going to make the research data FAIR and facilitate the access of knowledge.

It is divided into 5 subsections, which define the general principles of the policy and then moves to different aspects of Open Science. It starts with principles of publishing Open-Access, principles of FAIR data into the research, principles of Open Access software used, and finally it is focused on roles and responsibilities of Albanian researchers. During the policy a repository is mentioned which is the digital repository that was introduced in the previous section.

4.5.1. General Principles of the Policy

The policy is to make the data “as open as possible, as close as needed”. This means that:

- Research outputs of the University should be openly accessible. Ethical concerns, confidentiality, security, intellectual property rights, or publishers’ requirements might restrict how open the research data can be.
- All the requirements for research ethics from the Ministry of Education, Sports and Youth in Albania are met.
- Open Science activities are considered part of academic work and will be considered as a form of merit.
- Openness is going to be monitored for all the research output from the University.

4.5.2. Principles for Publishing Open-Access

The university aims to publish all publications according to the principles of Open Access and recommends publishing in any form of open access method.

If possible, all the peer-reviewed scientific articles and conference papers that do not conflict with the right of publishers or the authors should be published also in the Open-Access Digital Repository.

The theses of the students are published in electronic form in the Repository.

4.5.3. Principles of FAIR data into the research

Researchers of the university are encouraged to follow the FAIR principles for their research data. All the funded research data should be openly accessible unless there are legal or ethical obligations not to do so.

Data that is suitable for reuse should be made FAIR. All the metadata of the research should be documented properly and linked to the research publications. This data should be accessible even into the Open-Access Digital Repository.

4.5.4. Principles of Open-Access Methods and Software

Methods or software created by the researchers of the University should be openly shared, even when implemented with proprietary tools, to ensure the transparency of the research data. All the software created at the University should be shared as Open-Source Software to maximize the adoption by the research community of this research.

Derivate output from the researches such as tables and figures should be stored in the data research repository to allow future metadata analysis.

4.5.5. Roles and Responsibilities of the Researchers

All the researchers at the University who are going to publish the research data as open-access should ensure that all research data, software, and methods are properly documented and shared into the repository. If not possible, the metadata should be shared, and consider the research data as private. An additional explanation of why this data is restricted is advised when publishing the metadata.

All the researchers should familiarize themselves with the research data management and Open-Access publishing. They should also familiarize themselves with the Open-Access Digital Repository functionality. All the researchers should follow the ethical guidelines of the University for publishing research data.

When possible, always used ORCID researcher identifier, especially when submitting research output in Open-Access publishing. Always acknowledge the infrastructure and the affiliation with your department when publishing research outputs.

4.6. Conclusions

In this chapter, a full description of the Albanian Open-Source Digital Library is provided. Besides that, it is shown that the developed Open-Source Digital Library is fully compliant with the metadata standards of OpenAIRE and supported by EOSC. The repository is available at https://albanian-cris.info and currently, the population of the database is a work in progress.

Firstly, all the Albanian institutions to be input into the database of the repository are shown in detail and an analysis of the requirements is completed. After that, the deployment of the repository system is shown with all the prerequisite software, and configuration to make it fully functional is illustrated. Following that, the integration with external databases is explained, especially the ORCID integration which helps researchers to log in to the system easily.

Finally, the data model used relies on a set of basic entities as defined by the Common European Research Information Format CERIF model (Common European Research Information Format), and all of the research entities' relationships are shown. All of the fields included in the research entities are shown.

This Open-Source Digital Repository deployed in Albania is the first step to push Open Science in the country. Albania is part of Balkan institutions which are still the first steps towards Open Science. Albanian-CRIS is the repository that will help Albanian Open Science Cloud to follow the transition to Open Science and to join EOSC.

Lastly, an Open-Access Policy is developed and currently being adopted by Albanian Universities, stating the support towards Open Science.

Chapter 5. Conclusions

In this chapter, the contributions of the thesis are listed. Firstly, the stages in the transition to Open Science and main problems faced by Balkan universities are identified. After that, the next step for implementing and deploying Open Science in Albania is explained. Lastly, future works to be done in the field are discussed.

5.1. Main conclusions from dissertation work

During this dissertation, Open Science is discussed from a transition point of view as the future of the Research and Research cycle. The following main conclusions during this research are made:

1. Open Science is described and explained in detail. In the dissertation, all the key elements of Open Science as described starting from what is Open Science, moving on to Open Access. Open Data is also described and compared to FAIR data. Of course, all of the mentioned concepts are closely related to Open Source, thus even Open Source is described in detail. Finally, the concept of Digital Repositories is detailed, as a key element in storing research data. Storing research data is an important task for Open Science.
2. The European developments on Open Science is assessed and analyzed. Starting with the European Open Science Cloud, which has created the set of guidelines and rules to be followed to make Research Data accessible by all EU researchers through interoperable FAIR data. After that, the most used Research Data Digital Repositories are described and compared. Finally, the compliance to be followed for Open Science in Europe is shown.
3. The stages in the transition to Open Science by Balkan universities are investigated. A survey is distributed and detailed. The survey made it possible to understand that Albania

has no Current Research Information System. Researchers find Open Science important for the future of their country however they have no incentives.

4. By following the case of Bulgaria, the first prototype of Albanian Open Science Cloud (AOSC) is proposed and detailed. AOSC will provide Open Science for the Research Institutions in Albania.
5. The first Albanian Open Digital Repository is developed. DSpace-CRIS is used for the repository and ORCID id access is enabled, to allow all the researchers free and easy access to the system. This system is compliant with EOSC standards.
6. An Open Science Policy to be implemented in Albanian Universities is developed. The goal of this Open-Access Policy is to follow the trend of Open-Science in Albania. Some Albanian Universities are currently in the process of reviewing and implementing such Policies. The policy aims the Open Science and sees it as a key element for increasing the impact of Research on the Albanian Community. The final goal of the first Open-Science Policy in Albania is to make the research data FAIR and facilitate the access of knowledge for all Albanian researchers.

5.2. Contributions

In this section, contributions of this dissertation are listed.

5.2.1. Scientific contributions

1. A detailed explanation of Open Science and all its consisting elements was investigated. It serves as a fully comprehensive guide towards Open Science.
2. Proposed the architecture of the prototype of the Albanian Open Science Cloud which can help Albania join European Open Science Cloud faster. It can provide Open Science for all the Research Institutions in the country. This was presented at the ICERI2020 conference [52].

5.2.2. Applied contributions

1. Developed the first Albanian Open Digital Repository for Research data. It helps towards the Albanian Open Science Cloud by providing FAIR services to the researchers. This was presented as the first step in Albania towards Open Science at the ICAI 2021 conference [53].

5.2.3. Scientific and applied contributions

1. A survey was distributed and analyzed in Bulgaria and Albania, to understand the current state of Open Science in the region. The survey helped conclude that Albania is in the early stages of Open Science. The detailed results were presented at the ICAI2020 conference [54].
2. Developed the Albanian Open-Access Policy to embrace Open Science. The policy can help enhance the transition to Open Science for all the Research Institutions in the country. This was presented at the EDULEARN21 conference [55].

5.4. Scientific publications related to the dissertation work

1. **S. Hasani**, E. Stefanova, K. Stefanov, A. Georgiev, ARE WE READY FOR OPEN SCIENCE - THE ANSWER OF THE BALKAN UNIVERSITIES, ICERI2020 Proceedings, Publisher: IATED, 2020, pages: 1947-1953, ISSN (print): 2340-1095, ISBN: 978-84-09-24232-0, doi: 10.21125 / iceri.2020.0481, International 2020 [52].

2. **S. Hasani**, E. Stefanova, A. Georgiev, K. Stefanov, Current State of Open Science in Balkan Universities, Proceedings of the International Conference Automatics and Informatics (ICAI2020), Publisher: IEEE, 2020, pages: 1-6, ISBN: 978-1-7281-9309-0, doi: 10.1109 / ICAI50593.2020.9311337, Ref, IEEE Xplore, International, PhD [54].

3. **S. Hasani**, E. Stefanova, A. Georgiev, K. Stefanov, OPEN SCIENCE IN ALBANIA: FIRST OPEN-SCIENCE POLICY IMPLEMENTED IN ALBANIA, EDULEARN21 Proceedings, Conference name: 13th International Conference on Education and New Learning Technologies, Dates: 5-6 July, 2021, Pages: 12267-12271, ISBN: 978-84-09-31267-2, ISSN: 2340-1117, doi: 10.21125/edulearn.2021.2575 [55].

4. **S. Hasani**, E. Stefanova, A. Georgiev and K. Stefanov, "First Steps towards Open Science in Albania" 2021 International Conference Automatics and Informatics (ICAI), Varna, Bulgaria, 2021, pp. 235-238, doi: 10.1109/ICAI50593.2020.9311337, ISBN: 978-1-6654-2662-6, e-ISBN: 978-1-6654-2661-9, 10.1109/ICAI52893.2021.9639622 [53].

5.5. Future Works

Despite the contributions of this dissertation, some shortcomings can be continued as future works.

The most important one is the subject of the population of the Open Access Digital Repository. The process of the population is a work in progress and still needs a lot of time and dedication. Not only that but since the system will be expanding there is a need for many administrators to handle the research data and metadata.

Secondly, when the system becomes much more usable, migration to a better server is needed. Currently, the system is deployed in a dual-core processor with only 2GB of RAM. The moment the system needs a lot of simultaneous user access, a better system might be required to handle all the requests to the server.

Lastly, the deployed repository and Open-Science Policy are both under evaluation from different universities in Albania as personal initiatives. That means that in the future, a collaboration with the Ministry of Education, Sports and Youth in Albania will help the country join EOSC sooner.

Bibliography

[1] M. Woelfle, P. Olliaro, and M. H. Todd, “Open science is a research accelerator,” *Nature Chemistry*, vol. 3, no. 10, Art. no. 10, Oct. 2011, doi: 10.1038/nchem.1149.

[2] R. Vicente-Saez and C. Martinez-Fuentes, “Open Science now: A systematic literature review for an integrated definition,” *Journal of Business Research*, vol. 88, pp. 428–436, Jul. 2018, doi: 10.1016/j.jbusres.2017.12.043.

[3] OECD, “Open Science - OECD,” *Open Science - OECD*, 2018. https://www.oecd.org/sti/inno/open-science.htm (accessed Mar. 18, 2022).

[4] “What is Open Science?,” *ORION Open Science*, Sep. 27, 2017. https://www.orion-openscience.eu/resources/open-science (accessed May 05, 2020).

[5] “What is open access?,” May 04, 2020. https://www.openaccess.nl/en/what-is-open-access (accessed May 04, 2020).

[6] “Directory of Open Access Journals.” https://doaj.org/ (accessed Mar. 18, 2022).

[7] “Open Science Policies | FOSTER.” https://www.fosteropenscience.eu/foster-taxonomy/open-science-policies (accessed Apr. 04, 2021).

[8] K. Iatropoulou, “OpenAIRE,” *OpenAIRE*, Jun. 18, 2020. https://www.openaire.eu/ (accessed Jun. 18, 2020).

[9] Peter Suber, “Open Access Overview,” May 04, 2020. http://legacy.earlham.edu/~peters/fos/overview.htm (accessed May 04, 2020).

[10] “European Commission awards contract for setting up an open access publishing platform,” *European Commission - European Commission*. https://ec.europa.eu/info/news/european-commission-awards-contract-setting-open-access-publishing-platform-2020-mar-20_en (accessed Mar. 06, 2022).

[11] “The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge,” May 06, 2020. https://opendefinition.org/ (accessed May 06, 2020).

[12] “What is open?,” May 06, 2020. https://okfn.org (accessed May 06, 2020).

[13] Daniel Lathrop, *Open government*. O’Reilly, 2010. Accessed: May 21, 2020. [Online]. Available: http://archive.org/details/opengovernmentco00lath_0

[14] “FAIR Principles,” *GO FAIR*, Jun. 15, 2020. https://www.go-fair.org/fair-principles/ (accessed Jun. 15, 2020).

[15] M. D. Wilkinson *et al.*, “The FAIR Guiding Principles for scientific data management and stewardship,” *Sci Data*, vol. 3, no. 1, p. 160018, Dec. 2016, doi: 10.1038/sdata.2016.18.

[16] “Metadata | Definition of Metadata by Merriam-Webster,” Jun. 15, 2020. https://www.merriam-webster.com/dictionary/metadata (accessed Jun. 15, 2020).

[17] “ORCID,” Jun. 15, 2020. https://orcid.org/ (accessed Jun. 15, 2020).

[18] “FAIRification Process,” *GO FAIR*, May 10, 2020. https://www.go-fair.org/fair-principles/fairification-process/ (accessed May 10, 2020).

[19] “The Open Source Definition | Open Source Initiative,” May 04, 2020. https://opensource.org/osd (accessed May 04, 2020).

[20] “What is Open Source Software? Definition and FAQs | OmniSci.” https://www.heavy.ai/technical-glossary/open-source-software (accessed Mar. 06, 2022).

[21] “Open Standards Requirement for Software | Open Source Initiative,” May 04, 2020. https://opensource.org/osr (accessed May 04, 2020).

[22] “What are Open-Source Tools? - Definition from Techopedia,” *Techopedia.com*, May 04, 2020. https://www.techopedia.com/definition/3295/open-source-tools (accessed May 04, 2020).

[23] “Open Source for Business | Open Source Initiative,” May 04, 2020. https://opensource.org/osforbusiness (accessed May 04, 2020).

[24] “Digital Repository - an overview | ScienceDirect Topics,” May 05, 2020. https://www.sciencedirect.com/topics/computer-science/digital-repository (accessed May 05, 2020).

[25] M. Smith *et al.*, “DSpace: An Open Source Dynamic Digital Repository,” *D-Lib Magazine*, vol. 9, no. 1, Art. no. 1, Jan. 2003, doi: 10.1045/january2003-smith.

[26] T. Staples, R. Wayland, and S. Payette, “The Fedora Project: An Open-source Digital Object Repository Management System,” *D-Lib Magazine*, vol. 9, no. 4, Apr. 2003, doi: 10.1045/april2003-staples.

[27] "Open Access - 1st Edition." https://www.elsevier.com/books/open-access/jacobs/978-1-84334-203-8 (accessed Mar. 06, 2022).

[28] "What is a Content Management System (CMS)? Definition from WhatIs.com," *SearchContentManagement*, May 21, 2020. https://searchcontentmanagement.techtarget.com/definition/content-management-system-CMS (accessed May 21, 2020).

[29] A. Klahold and M. Fathi, *Computer Aided Writing*. 2020. Accessed: May 21, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-27439-9

[30] "EOSC," *EOSC Portal*, Jul. 04, 2018. https://eosc-portal.eu/about/eosc (accessed Mar. 14, 2022).

[31] "EOSC Launch," May 04, 2020. https://eosc-launch.eu/home/ (accessed May 04, 2020).

[32] mattste, "European Open Science Cloud," *Shaping Europe's digital future - European Commission*, Oct. 28, 2016. https://ec.europa.eu/digital-single-market/en/european-open-science-cloud (accessed Jun. 18, 2020).

[33] "EOSC Vienna Declaration," May 04, 2020. https://eosc-launch.eu/declaration/ (accessed May 04, 2020).

[34] European Commission, "EOSC Strategic Implementation Roadmap," *EOSC Strategic Implementation Roadmap*, May 2018. https://ketlib.lib.unipi.gr/xmlui/bitstream/handle/ket/1170/eosc_strategic_implementation_roadmap_large.pdf?isAllowed=y&sequence=1 (accessed Dec. 28, 2020).

[35] European Science Foundation, "'Plan S' and 'cOAlition S' – Accelerating the transition to full and immediate Open Access to scientific publications," *Plan S Overview*, Jun. 18, 2020. https://www.coalition-s.org/ (accessed Jun. 18, 2020).

[36] A. Georgiev and K. Stefanov, "Bulgarian Open Science Digital Library - First Prototype," in *Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings*, Burgas, Bulgaria, Sep. 2019, p. 8. [Online]. Available: https://dipp.math.bas.bg/images/2019/251-258_8_3.8_sDiPP2019-66_f_v.1.F_20190908.pdf

[37] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, “Open Archives Initiative - Protocol for Metadata Harvesting - v.2.0,” Jun. 14, 2002. https://www.openarchives.org/OAI/openarchivesprotocol.html (accessed Jun. 18, 2020).

[38] “Main features of CERIF | euroCRIS,” Jun. 20, 2020. https://www.eurocris.org/cerif/main-features-cerif (accessed Jun. 20, 2020).

[39] J. Coombs, “LibGuides: Research Data Management: RDM Explained,” Jun. 19, 2020. https://library.dmu.ac.uk/rdmguide/rdmexplained (accessed Jun. 19, 2020).

[40] “DSpace-CRIS Home - DSpace-CRIS - LYRASIS Wiki,” Jun. 19, 2020. https://wiki.lyrasis.org/display/DSPACECRIS/ (accessed Jun. 19, 2020).

[41] “Home | DSpace-CRIS DEMO,” May 04, 2020. https://dspace-cris.4science.it/ (accessed May 04, 2020).

[42] “InvenioRDM — inveniosoftware.org,” Jun. 19, 2020. https://inveniosoftware.org/products/rdm/#faq (accessed Jun. 19, 2020).

[43] “Zenodo - Research. Shared.,” Jun. 19, 2020. https://about.zenodo.org/infrastructure/ (accessed Jun. 19, 2020).

[44] “Fedora - The Flexible, Modular, Open-Source Repository Platform,” *Duraspace.org*, Jun. 19, 2020. https://duraspace.org/fedora/ (accessed Jun. 19, 2020).

[45] P. Stanchev and K. Stefanov, “Bulgarian Open Science Cloud,” in *Digital Presentation and Preservation of Cultural and Scientific Heritage. Conference Proceedings*, Burgas, Bulgaria, 2019, vol. 9, p. 6. [Online]. Available: https://dipp.math.bas.bg/images/2019/259-264_6_3.9_sDiPP2019-65_f_v.1.F_20190908.pdf

[46] “NI4OS- Europe – National Initiatives for Open Science in Europe,” Jul. 08, 2020. https://ni4os.eu/ (accessed Jul. 08, 2020).

[47] A. Bodlos *et al.*, “EOSC Pillar ‘National Initiatives’ Survey (SUF edition).” AUSSDA Dataverse, 2019. doi: 10.11587/VOSVGK.

[48] B. Saenen, R. Morais, V. Gaillard, L. Borrell-Damián, and F. Tobon, “Data for Research Assessment in the Transition to Open Science. 2019 EUA Open Science and Access Survey Results.” Zenodo, Jan. 07, 2020. doi: 10.5281/ZENODO.3600122.

[49] “Home | ICT,” Jul. 08, 2020. https://npict.bg/ (accessed Jul. 08, 2020).

[50] “Home | Albanian DSpace-CRIS.” https://albanian-cris.info/ (accessed Oct. 06, 2021).

[51] “Integration with external databases (Scopus, Web of Science, etc.) - DSpace-CRIS - LYRASIS Wiki.” https://wiki.lyrasis.org/pages/viewpage.action?pageId=78163330 (accessed Mar. 05, 2021).

[52] S. Hasani, E. Stefanova, K. Stefanov, and A. Georgiev, “ARE WE READY FOR OPEN SCIENCE – THE ANSWER OF THE BALKAN UNIVERSITIES,” in *ICERI2020 Proceedings*, Online Conference, Nov. 2020, pp. 1947–1953. doi: 10.21125/iceri.2020.0481.

[53] S. Hasani, E. Stefanova, A. Georgiev, and K. Stefanov, “First steps towards Open Science in Albania,” in *2021 International Conference Automatics and Informatics (ICAI)*, Sep. 2021, pp. 235–238. doi: 10.1109/ICAI52893.2021.9639622.

[54] S. Hasani, E. Stefanova, A. Georgiev, and K. Stefanov, “Current State of Open Science in Balkan Universities,” in *2020 International Conference Automatics and Informatics (ICAI)*, Varna, Bulgaria, Oct. 2020, pp. 1–6. doi: 10.1109/ICAI50593.2020.9311337.

[55] S. Hasani, E. Stefanova, K. Stefanov, and A. Georgiev, “OPEN SCIENCE IN ALBANIA: FIRST OPEN-SCIENCE POLICY IMPLEMENTED IN ALBANIA,” *EDULEARN21 Proceedings*, pp. 12267–12271, 2021.

Appendix A

Survey disseminated in Bulgaria and Albania for the current state of Open Science.

Please tick the box to show your consent to be part of the research: ☐

Your Name: Click or tap here to enter text.

Your Institution name: Click or tap here to enter text.

Your Position in the Institution: Click or tap here to enter text.

Your Professional experience in the institution: Click or tap here to enter text.

1. **Is your institution performing research assessment for careers in research, performance evaluation and/or allocation of research funding?** (Select one)
   a. Yes. (Continue to question 2) ☐
   b. No, but it is being developed (continue to question 6) ☐
   c. No. (continue to question 8) ☐

2. **Information on the methods used in your institution's assessment procedure for careers in research is:** (Select one)
   a. Publicly available ☐
   b. Internally available ☐
   c. Not available ☐

3. **Rate the following aspects of academic work within your institution's research assessment approach for careers in research?** (From 1 to 5, 1=Unimportant and 5=Very Important)
   a. Research Publications 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
   b. Data Research Output 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

c. Research Collaborations within Academia 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

d. Research Collaborations outside academia 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

e. Research Impact 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

f. Social Outreach 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

g. Research Networking 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

h. Attracting external research funding 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

i. Open Science and Open Access 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

j. Research supervision Activities 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

k. Teaching Activities 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

l. Mentoring Activities 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

4. **Which of the following does your institution use to measure the research output of researchers?** (Multiple Choice)

a. Journal Impact Factor (JIF) ☐

b. h-index ☐

c. EigenFactor ☐

d. SCImago Journal Rank (SJR) ☐

e. Source Normalized Impact per Paper (SNIP) ☐

f. CiteScore ☐

g. Don't know ☐

h. Others Click or tap here to enter text.

5. **Which of the altmetrics does your institution use to measure the societal outreach of research outcomes?** (Multiple Choice)

a. Altmetric.com ☐

b. ImpactStory ☐

c. PLUMx ☐

d. Datacite ☐

e. Bookmetrix ☐

f. F1000Prime ☐

g. Data Citations ☐

h. ResearchGate views ☐

i. Don't know ☐

6. **Which existing models does your institution use for developing its own research assessment?** (Multiple Choice)

a. Principles and guidelines developed by other universities ☐

b. San Francisco DORA (Declaration on Research Assessment) ☐

c. The Leiden Manifesto ☐

d. The Metric Tide ☐

e. Doesn't use existing principles and guidelines ☐

f. Don't know ☐

g. Other principles and guidelines Click or tap here to enter text.

7. **How autonomous do you consider your institution on developing and implementing its research assessment procedures for evaluating the performance of research units?** (Single choice)

a. Highly autonomous ☐

b. Mostly autonomous ☐

c. Some autonomy to develop and implement its procedures ☐

d. Low autonomy to develop and implement its procedures ☐

e. Don't know ☐

"Open access science is the dissemination of scientific knowledge that is as wide as possible, free of charge to all users, and accessible online." (Open Access)

8. **Have you been involved in Open Access publishing as an author, a journal editor or in any other role?** (Single Choice)
    a. Yes ☐
    b. No ☐

9. **Do you share your publication in an Open Access Repository?** (Single Choice)
    a. Yes ☐
    b. No ☐

10. **Do you think researches should be Open Access?** (Single Choice)
    a. Yes ☐
    b. No ☐

11. **Do you see any advantages or disadvantages of publishing one's research results in Open Access?** (Open Question)

    Click or tap here to enter text.

12. **Would it differ with regard to research fields, geographic or economic position, personal situation, etc.?** (Open Question)

    Click or tap here to enter text.

13. **Do you think Open Access Research will bring changes to your own research practices?** (Open Question)

    Click or tap here to enter text.

14. **Do you think there is any other important aspect in order to understand Open Access and related issues better?** (Open Question)

Click or tap here to enter text.

15. **In your opinion, to whom should science be open?** (Rank from 1 to 5, 1=should not be opened, 5=should be very open)

    a. Scientists from the same discipline 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    b. Scientist from other discipline 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    c. All citizens 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    d. Civil and social organizations 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    e. Specially concerned groups 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    f. Funders and Policy makers 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    g. Industry and companies 1.☐ 2.☐ 3.☐ 4.☐ 5.☐

16. **In your opinion, how open do you think the different aspects of the scientific process should be to the scientific community (all scientists)?** (Rank from 1 to 5, 1=should not be opened, 5=should be very open)

    a. The research priorities 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    (what topics, how much funding…)
    b. The design of the research 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    (what methodologies, what ethical considerations…)
    c. The research process 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    (data gathering, data management, replicability …)
    d. The research results 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    (knowledge, publications, patents…)
    e. The research outcomes 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
    (design of final products for end users…)

17. In your opinion, how open do you think the different aspects of the scientific process should be to the society (all citizens)? (Rank from 1 to 5, 1=should not be opened, 5=should be very open)

- **a.** The research priorities 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (what topics, how much funding…)
- **b.** The design of the research 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (what methodologies, what ethical considerations…)
- c. The research process 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (data gathering, data management, replicability …)
- d. The research results 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (knowledge, publications, patents…)
- e. The research outcomes 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (design of final products for end users…)

18. In your opinion, how open do you think the different aspects of the scientific process should be to funders and policy makers? (Rank from 1 to 5, 1=should not be opened, 5=should be very open)

- **a.** The research priorities 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (what topics, how much funding…)
- **b.** The design of the research 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (what methodologies, what ethical considerations…)
- c. The research process 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (data gathering, data management, replicability …)
- d. The research results 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (knowledge, publications, patents…)
- e. The research outcomes 1.☐ 2.☐ 3.☐ 4.☐ 5.☐
  (design of final products for end users…)

19. **In your opinion, why should science be open?**

	Not a reason For Open Science	**A relatively important reason**	**An important reason**	**The most important reason for Open Science**	**I don't know it / I don't have enough information**
Diversity: incorporation of underrepresented groups in science (gender, races, cultures, etc.).	☐	☐	☐	☐	☐
New and innovative economic possibilities: crowdfunding, new types of founders, etc	☐	☐	☐	☐	☐
Efficiency: sharing of data, procedures and/or to optimize science.	☐	☐	☐	☐	☐
Equity: access for all to scientific results, methods, software, etc., regardless of economic capacity or institutional affiliation.	☐	☐	☐	☐	☐
Ethics: Open Science is aligned with principles of research integrity.	☐	☐	☐	☐	☐
Fairness: Science is often funded by society, so all results from the research	☐	☐	☐	☐	☐

should be available to society					
Impact: To outperform traditional metrics for scientific impact: larger audience, higher engagement, etc.	☐	☐	☐	☐	☐
Rigour: Open access, open data and/or open replicability make science easier to review.	☐	☒	☐	☐	☐

20. **In your opinion, why should science NOT be open?**

	Not a reason Against For Open Science	**A relatively important reason against**	**An important reason against**	**The most important reason against Open Science**	**I don't know it / I don't have enough information**
Not a priority now. Currently, there are higher priorities in the scientific community	☐	☐	☐	☐	☐
Public's lack of understanding. Society cannot make decisions or have a useful input without an understanding of	☐	☐	☐	☐	☐

science/the scientific process.					
Public is not ready now. Society is not ready for participation in science (lack of skills, tools, etc).	☐	☐	☐	☐	☐
Risk to fundamental research. Open Science would only benefit applied science and be detrimental to fundamental research.	☐	☐	☐	☐	☐
Low quality. By releasing publications prior to classical peer-review, the veracity of papers will be difficult to assess by individual researchers Quality is not guaranteed by preprint servers.	☐	☐	☐	☐	☐
Danger and potential misuse. Open Science may interfere with research integrity (ex: release of medical personal data). It could also facilitate misuse of research results (e.g. biological weapons).	☐	☐	☐	☐	☐
Lack of incentives. Open data / publication runs	☐	☐	☐	☐	☐

counter “meritocracy” and individual effort, and they are not captured and rewarded through traditional metrics.					
Unfairness. If a research group generates knowledge with own resources, it could be unfair if others use this knowledge to get economic benefits for themselves	☐	☐	☐	☐	☐

21. Imagine in your everyday work at your institution you decide to embrace (or you already have embraced) an Open Science perspective. What do you think (or know) are the most important barriers you will be facing?

	Very important barrier	**Important barrier**	**Low barrier**	**Not a barrier at all**	**I don’t know it / I don’t have enough information**
Lack of proper infrastructure. How/where do I store open data?	☐	☐	☐	☐	☐
Lack of clear steps to follow. How do I begin? How do I proceed?	☐	☐	☐	☐	☐

Authentic public engagement. How do I achieve representative samples of citizens (motivate people other than already concerned groups)?	☐	☐	☐	☐	☐
Budget and funding constraints. How do I find the funding to ensure open data and open publications?	☐	☐	☐	☐	☐
Time constraints. I don´t have time to practice Open Science, it is too time-consuming.	☐	☐	☐	☐	☐
Fears and uncertainties for career development. Will my Open Science practice be valued at institutional level or during my career? Does it mean I will receive more funding or merit?	☐	☐	☐	☐	☐

22. Do you receive training from your institution related to Open Science?

	I receive adequate training	**I need more training**	**These topics are not relevant for my specific professional tasks**	**I don't know it / I don't have enough information**

Research and data management (Data storage, sharing, FAIR - “Findable, Accessible, Interoperable, and Reusable” - approaches)	☐	☐	☐	☐
Research integrity (Animal Research, data analysis and interpretation, research with human samples, good practice in the lab, etc.)	☐	☐	☐	☐
Research publishing and dissemination (Open Access, pre-prints, peer review)	☐	☐	☐	☐
Collaborating and networking (How to improve collaboration through Open Science)	☐	☐	☐	☐
Communicating science to the general public (Different audiences, practical guides to getting started, online and offline options)	☐	☐	☐	☐
Involving the general public in research (Citizen	☐	☐	☐	☐

science: data gathering, data analysis, use of results)				
Evaluation of research projects and researchers	☐	☐	☐	☐
Assessment of the impact of initiatives in public	☐	☐	☐	☐

23. Do you receive support or incentives from your institution related to Open Science?

	I receive adequate support or incentives from my institution	**I would like to receive more support or incentives enough**	**These kind of support or incentives are not relevant for my specific professional tasks**	**I don't know it /I don't have enough informatio n**
Written guidelines (webpage/leaflet/videos), policies, recommendations	☐	☐	☐	☐
Technical infrastructure (templates, software, storage, databases, publication and/or data repositories, etc.)	☐	☐	☐	☐
Specialist support (experts on different aspects of Open	☐	☐	☐	☐

Science, research data committees, courses, workshops, etc.)				
Financial support and rewards	☐	☐	☐	☐
Career's perspectives and recognition	☐	☐	☐	☐

24. **Is your institution using a Current Research Information System (CRIS)? If so, provide the name of the CRIS, acronym, URL of the CRIS?**

    Click here to enter text.

25. **Select the protocols, standards and vocabularies used in the CRIS.** (Multiple Choice)

    a. OAI-PMH ☐
    b. Shibbloleth ☐
    c. CERIF ☐
    d. CORDIS ☐
    e. FOS ☐
    f. CASRAI ☐
    g. ORCID ☐

26. **If the CRIS is compliant with CERIF, please indicate the version of CERIF.**

    Click here to enter text.

27. **Do you store institutional Research Outputs such as publications, patents or products in the IR (Institutional Repository)?** (Single Choice)

   a. Yes ☐

   b. No ☐

   c. I don't know ☐

28. Does your Institution have an open access policy/mandate? (Single Choice)

   a. Yes ☐

   b. No ☐

   c. I don't know ☐

29. What type of content does your Institutional Repository store? (Multiple Choice)

   a. Research Publications ☐

   b. Datasets ☐

   c. Dissertations and Thesis ☐

   d. Learning Objects ☐

   e. Other Click here to enter text.

30. **Which other institutional services does your repository shares information with?** (Multiple Choice)

   a. Library Management System ☐

   b. Integrated Search Systems ☐

   c. Learning Management System ☐

   d. Project Management System ☐

   e. Human Resources Management System ☐

   f. Financial System ☐

   g. Researchers Webpages ☐

   h. Curricula Systems ☐

i. Institutional or Author Evaluation ☐

j. Research and Development Units ☐

k. Other Click here to enter text.

31. Is your Institutional Repository OpenAIRE-compliant? (Single Choice)

a. Yes ☐

b. No ☐

c. I don't know ☐

32. **Does your Institutional Repository register researchers' persistent digital identifiers ORCID? (If any please describe steps taken to integrate IR with CRIS)** (Single Choice)

a. Yes ☐

b. No ☐

c. I don't know ☐

Click here to enter text.

Declaration of Originality

I declare that the thesis “Research Information System services for the Open Science Cloud” is the result of my own original work that was done during my PhD study at Sofia University between years 2019 and 2022.