SOFIA UNIVERSITY

"ST. KLIMENT OHRIDSKI"

FACULTY OF MATHEMATICS AND INFORMATICS

RESEARCH INFORMATION SYSTEM SERVICES

FOR THE OPEN SCIENCE CLOUD

ABSTRACT

of 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 a 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.

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]. Open Science consists of six principles: 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 [3]. Advantages of Open Science include: Rigorous peer-review since Open access publication of research reports and data, publicly funded science is publicly available, more transparent, reproducible, and impactful science. Disadvantages include: potential misuse, general misunderstanding by the public, low-quality science since the major concern is just to have it open and, on some occasions, even entrapment by platforms.

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 [3]. 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 [4]. 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. 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 a research institution, research funder or government that requires researchers, which usually are university faculty, research staff, or research grant recipients, 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).

1.3. 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. [6]. 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 [7]:

- 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.

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 [8]:

- Findable:
    a) F1. The (meta) data has a globally unique and persistent identifier (ID) [9].
    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 [9].
    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 [9].

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 [10]. 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 [11].

In Open Science data should not only be Open but it should also be FAIR.

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 [12]:

- 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. [13]
- 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.

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” [14]. 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 [15].

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 [16].

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” [17]. 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 [17]:

- 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.

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.

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” [18].

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 [19].

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 [20].

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 [21]:

- 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 [22]. 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.

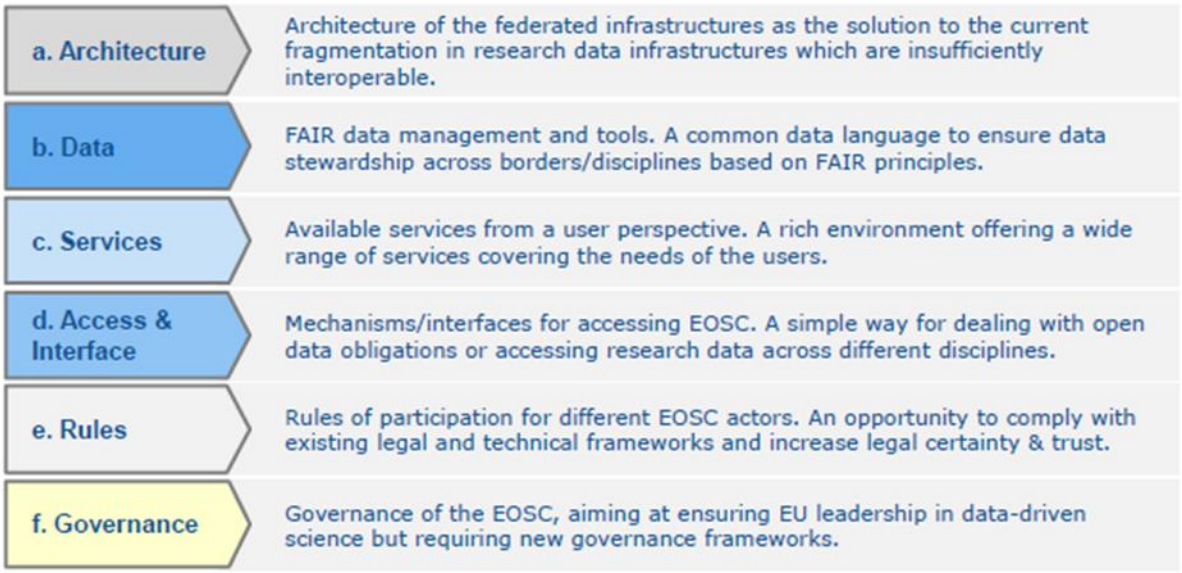

Figure 2-b: Actions of EOSC federated model

The architecture 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 [22].

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.

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.

2.2. Digital Libraries and Open AIRE

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 [23]:

- 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 [24]. 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 [24]. 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 [25]. 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” [26]. 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 [26] 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.

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 [27].

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” [28]. 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.

Figure 2-c shows the DSpace-CRIS interface which includes a demo at dspace-cris.4science.cloud [29].

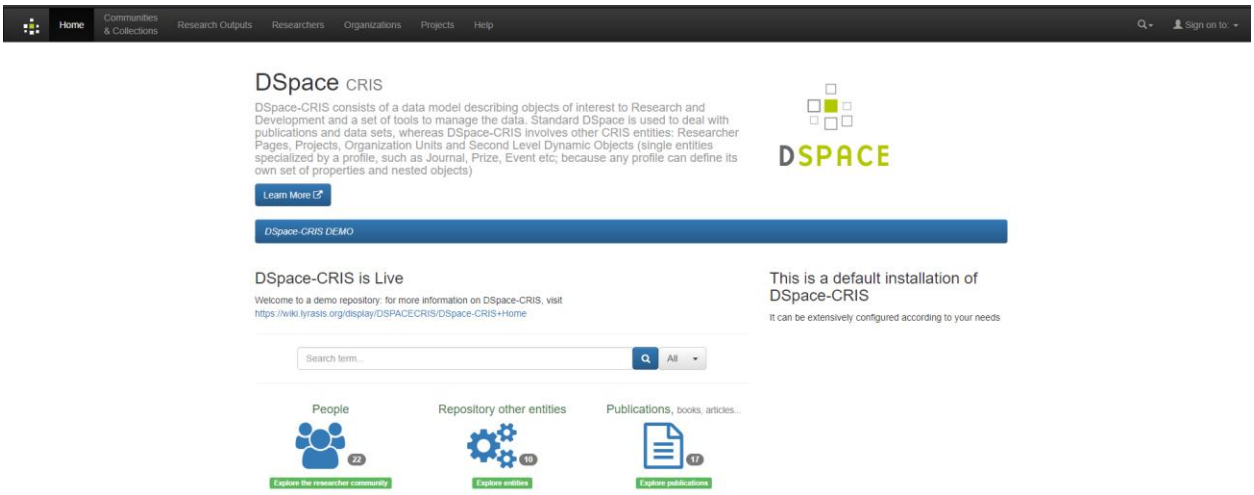

Figure 2-c: Demo of DSpace-CRIS interface

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 [28]. 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.

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.

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 [30][31]. 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.

An example of the Zenodo-Invenio interface is available at inspirebeta.net [31] 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.

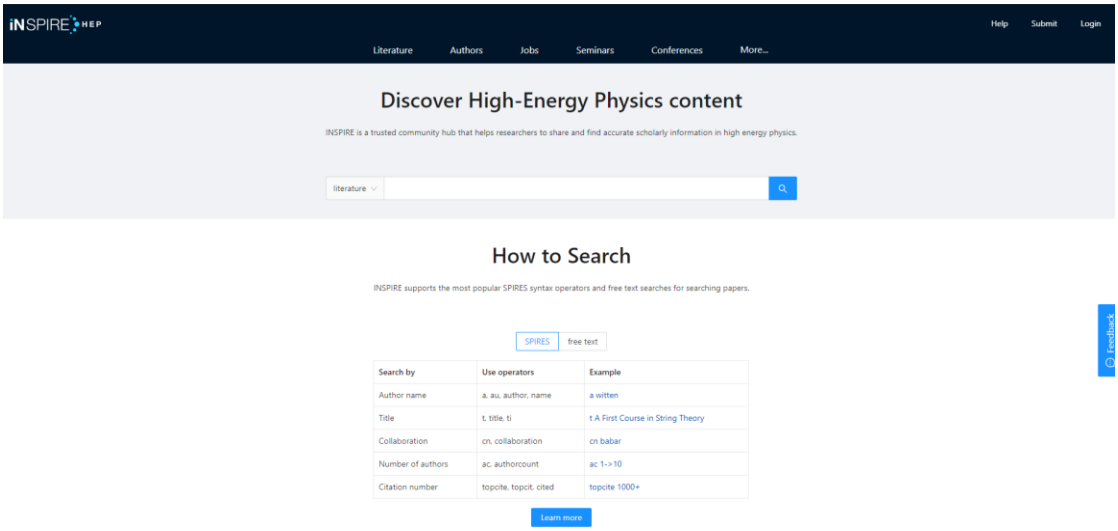

Figure 2-d: Invenio Interface at Inspirebeta.net

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 [32].

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.

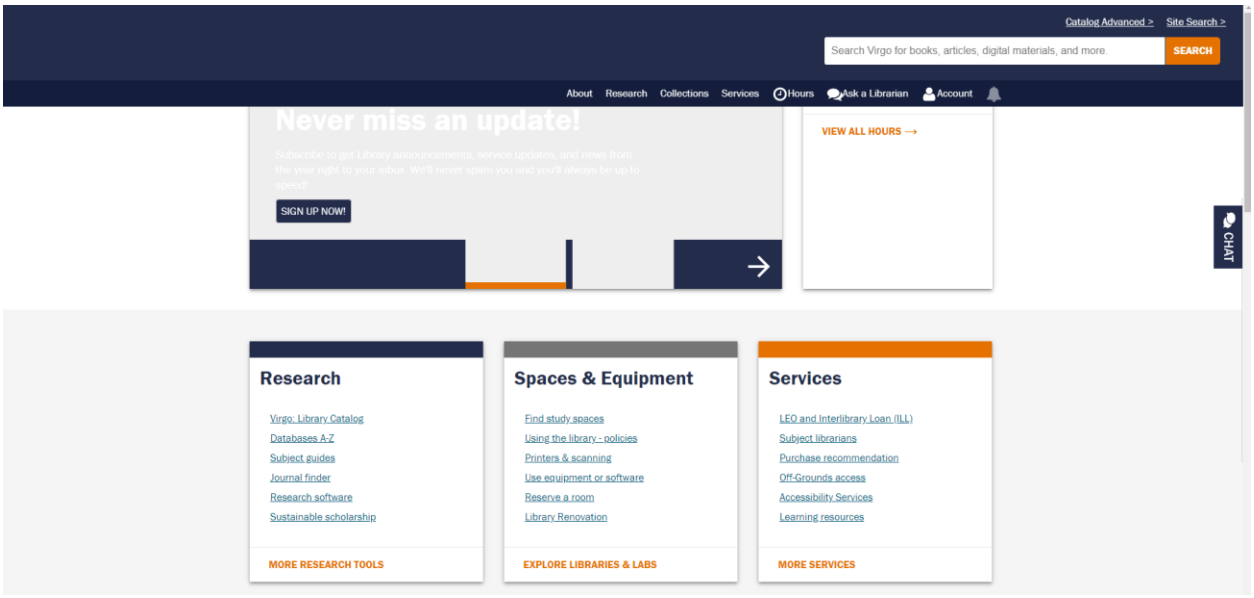

Figure 2-e: Fedora-based Interface at University of Virginia

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) [33] 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).[26] CERIF 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 [34]. It is disseminated in 15 Balkan countries including Albania and Bulgaria. However, both countries had low response participation of 1.91% in Albania and 7.48% in 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 [35]. 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" [36]. 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 [37].

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.

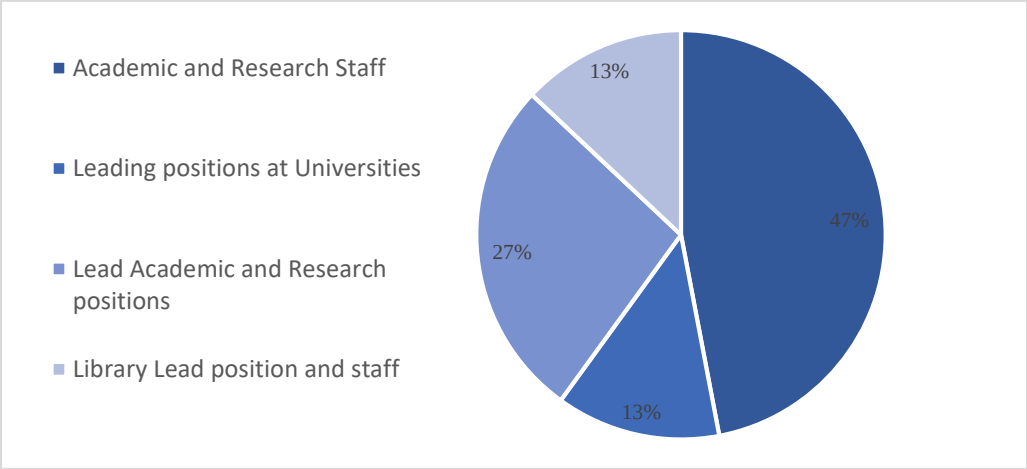

Figure 3-a: Respondent Profiles in Bulgaria.

3.3. 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/ [38].

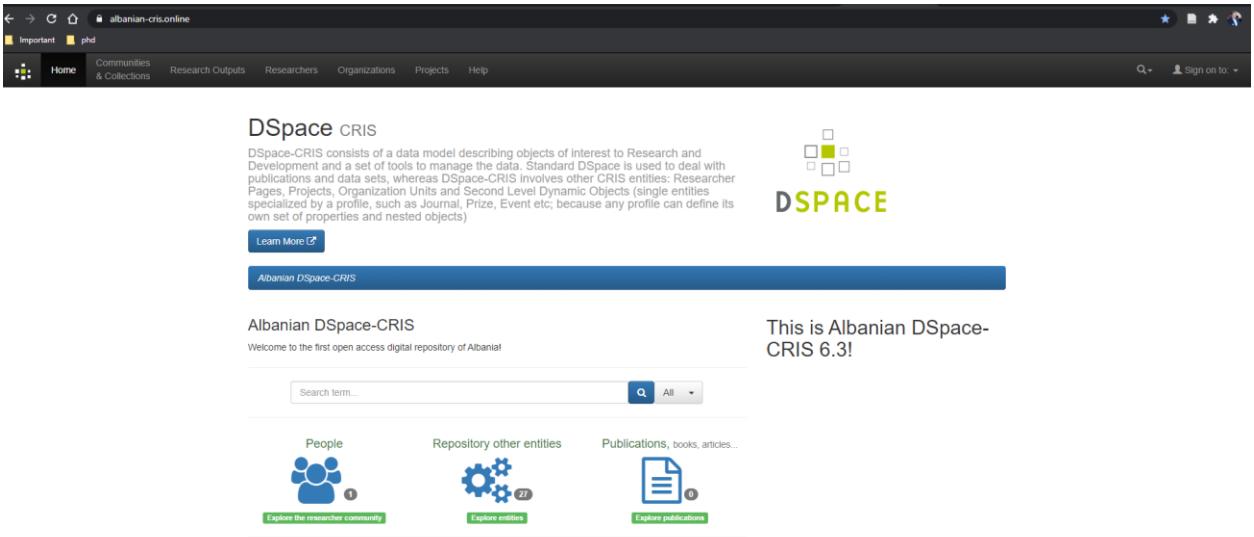

Figure 4-a: Albanian DSpace-CRIS

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 2 below with all the known data from such institutions. The information is checked with the Ministry of Education, Sports, and Youth in Albania.

Table 2: 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.

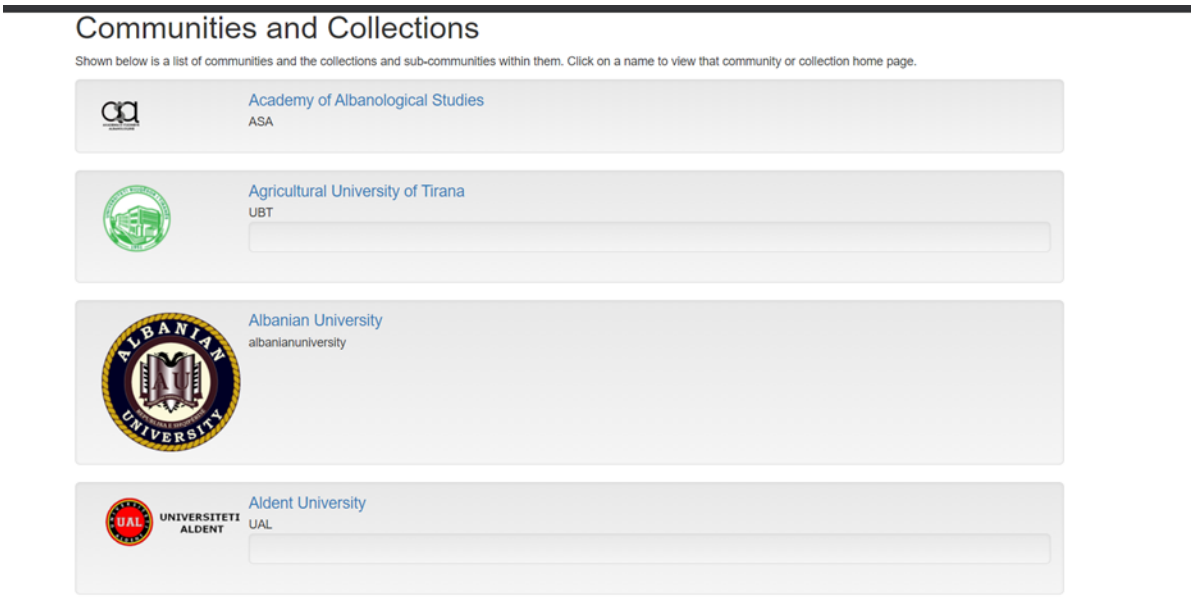

Figure 4-b: Sample of the List of Universities at Albanian Open Access Digital Repository

4.2. 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.

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-c where integration with ORCID is configured.

Figure 4-c: 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-d 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.

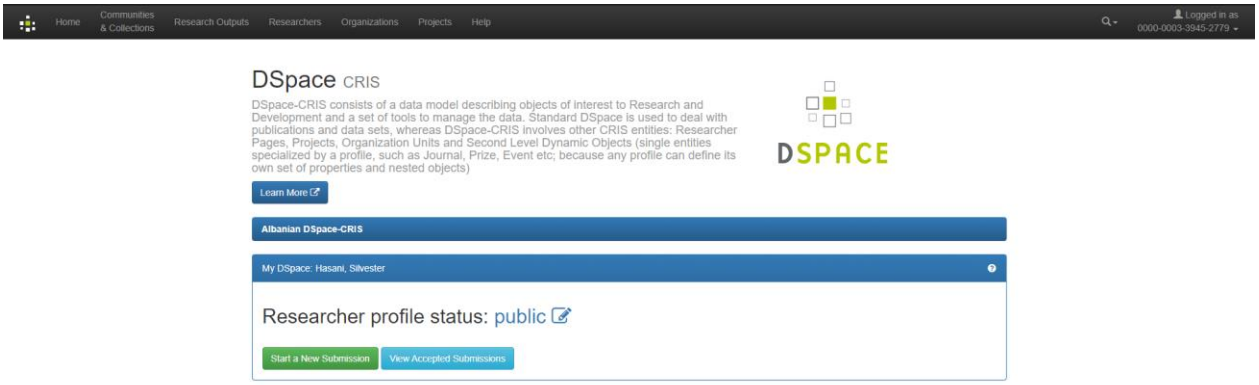

Figure 4-d: Login with ORCID

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 [39].

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-e 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.

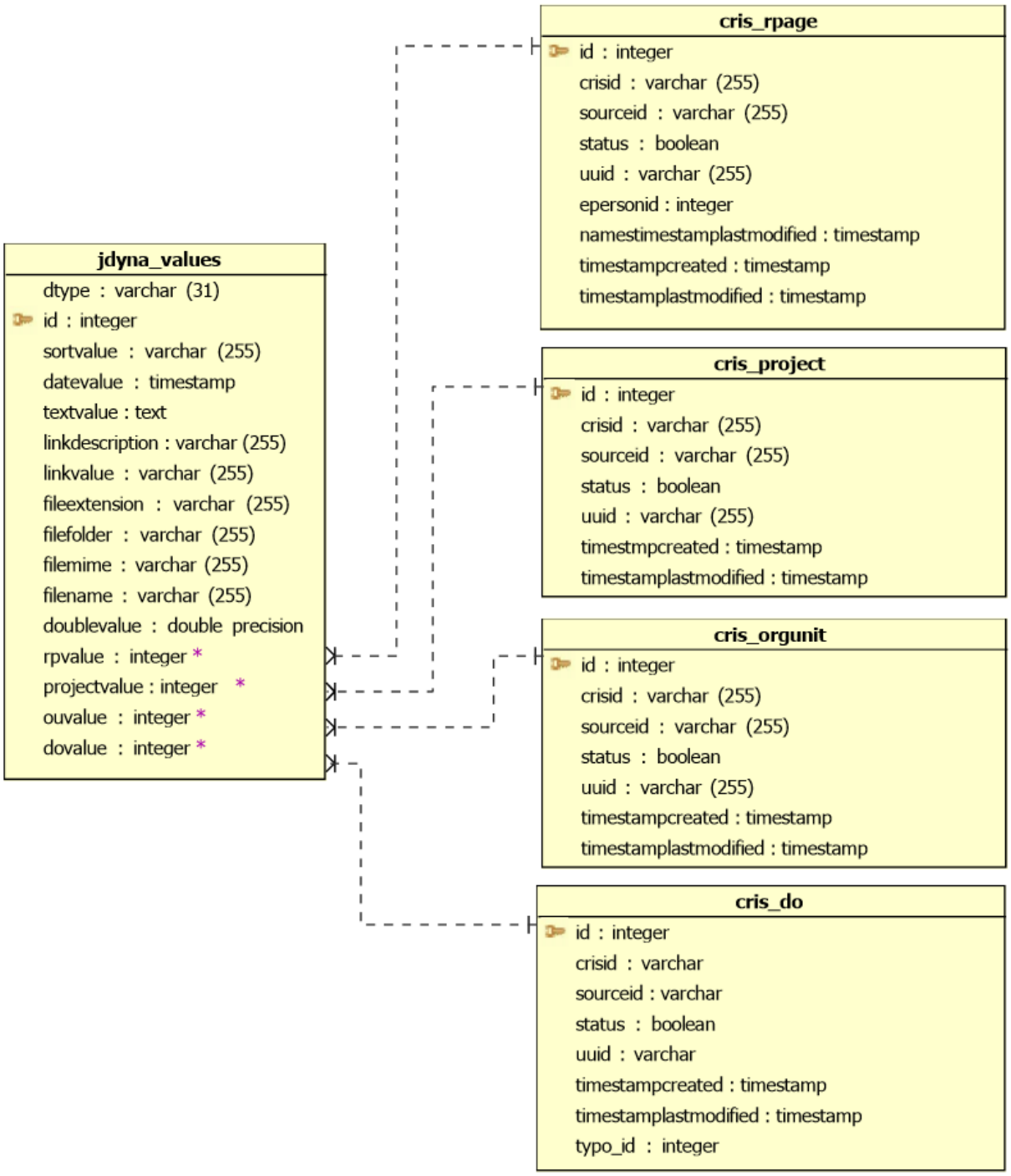

Figure 4-e: Albanian-CRIS core structure

All the fields of each research entity 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.

4.5. Open-Science Policy

In this paper it is also presented 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.

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.

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.

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 and Future Works

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 [40].

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 [41].

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 [42].
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 [43].

5.3. 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 [40].

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 [42].

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 [43].

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 [41].

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

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

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

[6] “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).

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

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

[9] 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.

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

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

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

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

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

[15] 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.

[16] 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.

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

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

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

[20] 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).

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

[22] 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).

[23] 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

[24] 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).

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

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

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

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

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

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

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

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

[33] 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

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

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

[36] 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.

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

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

[39] “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).

[40] 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.

[41] 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.

[42] 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.

[43] 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.

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.