



## D9 – Report on tool evaluation and selection



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra



SWISS NATIONAL SCIENCE FOUNDATION

<i>Deliverable Lead:</i>	FORS, FFZG
<i>Related Work package:</i>	WP2
<i>Author(s):</i>	Alen Vodopijavec (FFZG) Marijana Glavica (FFZG) Bojana Tasic (FORS) Renate Kunz (FORS) Irena Vipavc Brvar (UL-ADP) Aneta Cekikj (ISPJR) Klime Babunski (ISPJR) Aleksandra Bradic Martinovic (IES) Ivana Vujović (CeMI) Artela Mitrushu (IDM) Nita Bylykbashi (CPC)
<i>Dissemination level:</i>	Public (PU)
<i>Submission date:</i>	31 <sup>st</sup> March, 2017
<i>Project Acronym:</i>	SEEDS
<i>Website:</i>	<a href="http://www.seedsproject.ch">http://www.seedsproject.ch</a>
<i>Call:</i>	Scientific cooperation between Eastern Europe and Switzerland (SCOPES 2013-2016)
<i>Start date of project:</i>	1 <sup>st</sup> May 2015
<i>Duration:</i>	24 months

## Version History

Version	Date	Changes	Modified by
1.0	February 28, 2017	Released version	FFZG
2.0	March 31, 2017	Revised version	FORS

## Acknowledgments

This report has been developed within the “South-Eastern European Data Services” (SEEDS) ([www.seedsproject.ch](http://www.seedsproject.ch)) project. The participant organisations of the SEEDS project are:

Name	Short Name	Country
Faculty of Humanities and Social Sciences University of Zagreb	FFZG	Croatia
Centre for Monitoring and Research, Podgorica	CeMI	Montenegro
Centre for Political Courage, Pristina	CPC	Kosovo
Institute for Democracy and Mediation, Tirana	IDM	Albania
Institute of Economic Sciences, Belgrade	IES	Serbia
Saints Cyril and Methodius University, Institute for Sociological, Political and Juridical Research, Skopje	ISPJR	Macedonia
Swiss Foundation for Research in Social Sciences, Lausanne	FORS	Switzerland
University of Ljubljana, Social Science Data Archive, Ljubljana	UL	Slovenia

## Table of Contents

1. Introduction .....	5
2. Shortlisting of available software solutions.....	5
2.1 FORSbase.....	6
2.2 Dataverse .....	7
2.3 DSpace.....	8
2.4 EPrints.....	8
3. List of criteria for evaluation of the archiving tools.....	9
3.1. General about software and software development .....	9
3.2. OAIS reference model coverage .....	10
3.3. Administration, user interface and access.....	11
3.4. Performance and security issues .....	13
4. Evaluation overview by country and conclusion .....	14
4.1. Albania.....	14
4.2. Croatia .....	15
4.3. Kosovo .....	16
4.4. Macedonia.....	17
4.5. Montenegro .....	18
4.6. Serbia.....	19
5. Annex: Lists of criteria for evaluation of the archiving tools.....	21
5.1 Albania.....	21
5.2 Croatia .....	28
5.3 Kosovo .....	35
5.4 Macedonia.....	42
5.5 Montenegro .....	50
5.6 Serbia.....	55

# 1. Introduction

In this task partner institutions were choosing the right tool to support their future archiving operations. This evaluation was done to point out specific demands of each partner as well as to identify their local resources concerning the technical infrastructure and expertise in the institution and/or possibilities of outsourcing. There was a list of criteria prepared with an ad-hoc methodology created specifically for this purpose. We have also applied some minor changes to the evaluation process during the assessment period. The initial plan that each partner will do the entire process of evaluation individually was changed and the description of available features of certain software was done by a smaller technical group (Switzerland and Croatia). In the second phase all partners made their specific evaluations, based on their previously identified priorities.

## 2. Shortlisting of available software solutions

There are lots of software products available and able to support data archiving, however each solution has its own vision of archiving and its own set of core implemented features that are used to shape processes and procedures. We had several sets of demands related to the software development process and the community around it, OAIS compliance, security considerations and performance.

While searching for the tools that would satisfy our basic demands, three categories of software products were identified:

- Products that are intended to build a general purpose repository, content agnostic and are able to support basic deposit + publish process;
- Products built specifically for data archiving and dissemination, either scientific research data or public sector open data;

- Products that aim to support data archiving practices specific for a particular scientific discipline.

Our final list of software products for evaluation was:

- Dataverse
- DSpace
- EASY
- EPrints
- FORSbase
- RODA

Several other possible solutions are available (Archivematica, Fedora Commons, Invenio, Islandora, Omeka), but we did not evaluate them further in detail, as they are too specific (i.e. focused on humanities) or too broad (frameworks), and they do not seem to fit the needs of the SEEDS project partners on the basis of the preliminary review.

After getting acquainted with software demo sites and available showcase repositories, RODA and EASY were considered too complex for installation and maintenance. We ended up with selecting FORSbase, Dataverse, DSpace and EPrints for the final assessment and comparison.

## 2.1 FORSbase<sup>1</sup>

FORSbase is a platform developed by FORS from Switzerland with the vision of supporting the whole research data lifecycle. It is a work environment for social scientists and FORS collaborators with integrated data storage for primary data and metadata, data preparation and exchange via standards (DDI etc.). The main characteristics of FORSbase are:

- Data can be downloaded immediately;
- Research projects enjoy greater visibility;
- Researchers can deposit their data themselves online;

---

<sup>1</sup> <https://forsbase.unil.ch/>

- Data get published more quickly;
- No registration required to access documentation and metadata;
- Studies are documented in more depth;
- Researchers, projects, and institutions are better linked;
- Getting in touch with researchers only takes one mouse click;
- Researchers can use their personal workspace to secure and preserve their files;
- Life as a data user/provider is easier!

FORSbase is fully OAIS compliant and the technologies behind it are JQuery and AJAX for frontend, Python/Django for middleware, MySQL database for storing meta information and Fedora Commons for storing the files and for the archival storage system. The main advantage of this architecture is the flexibility and modularity so the each layer could be easily replaced by some other technology in the future if needed. The software itself is not yet open-sourced but the plan is to publish the code under the open source licence in a near future. FORSbase is a relatively new product, currently used by FORS. One identical instance of the system is already used by other SEEDS partners. It is branded as “SEEDSbase” and implemented for the purpose of the “RRPP Data Rescue” project.

## 2.2 Dataverse<sup>2</sup>

The Dataverse software is being developed at Harvard's Institute for Quantitative Social Science (IQSS) with the help of many contributors and collaborators all over the world. This is a successor of the previous VDC - Virtual Data Center project (1999-2006). It is suitable for archiving the research data from any discipline and it covers basic elements of the OAIS reference model. It has the API, which enables fairly easy integration with other systems. The software is built using Java programming language and technologies around it and it uses PostgreSQL as database server. There are 22 Dataverse installations available online with over

---

<sup>2</sup> <http://dataverse.org/>

2000 dataverses and almost 50.000 datasets published<sup>3</sup>. This is far less when comparing to EPrints or DSpace but Dataverse is a special type of repository, much more research data oriented. It's also worth to mention that 50% of data providers come from the Social Sciences. The Dataverse application code is licensed under the Apache License, Version 2.0 which implies distribution of the source code and the right to reuse, change it and make derivative works as long as the source code is further distributed and required notices are included. This permissive license contains a patent license from the contributors of the code.

## 2.3 DSpace<sup>4</sup>

DSpace is a common and widely used repository software application. It is a Java based application with the set of tools and utilities that provide an asset store and associated metadata store. It supports the common interoperability standards (OAI-PMH, SWORD etc.) and it can be customized to match the implementer's demands. Today, DSpace is a part of a non-profit initiative DuraSpace whose main goal is to support development of technologies which will provide long-term, durable access and discovery of digital assets. The DuraSpace software ecosystem consists of DSpace, Fedora (modular file repository system) and VIVO (semantic discovery tool).

## 2.4 EPrints<sup>5</sup>

The first version of EPrints was released in November 2000 as a tool for supporting an Open Access initiative based on the idea of Steven Harnad and Les Carr. Many academic and research institutions are using it today as an institutional repository for depositing and long term preservation of their scientific output. EPrints is a 'Swiss knife' of a repository software, meaning you can use it in various ways and that you can adapt it to your needs. It does

---

<sup>3</sup> Data used is published and available through the Dataverse API. <http://dataverse.org/metrics>

<sup>4</sup> <http://dspace.org/>

<sup>5</sup> <http://eprints.org/>



however require a good expertise in Perl programming language. EPrints uses MySQL or any of its relatives (MariaDB, Percona) as a database server.

### 3. List of criteria for evaluation of the archiving tools

We have prepared a list of criteria that should be satisfied and each criterion items were divided into sections and categories. A priority tag was given to each criterion. Partners were deciding whether a certain functionality is a “must have” or a “nice to have” feature.

Additionally, we assigned scores for each shortlisted tool based on our assessment exercise:

- 0=not fulfilled/not present,
- 1=partly present,
- 2=fully present.

Commenting on each score was available, because there are many small differences in features’ implementation of various software products. Our goal was to cover as much facts and opinions as possible to make a good final decision. Finally, there was an estimate of the “development effort” for all the “must-haves” but “not fully present features” which ranged<sup>6</sup> from “none” to “high”. A complete list of criteria for the evaluation of the archiving tools for each partner can be found in the Annex (see chapter 5).

#### 3.1. General about software and software development

All of the selected and listed softwares are archiving tools capable of basic archiving actions. DSpace<sup>7</sup> and EPrints<sup>8</sup> have large and diverse user communities, while Dataverse<sup>9</sup> is currently

---

<sup>6</sup> All available options were: no estimate possible, none, low, medium and high.

<sup>7</sup> <https://github.com/DSpace>

<sup>8</sup> <https://github.com/eprints/eprints>

<sup>9</sup> <https://github.com/IQSS/dataverse>

used by a dozen of large US universities. FORSbase is used only at FORS, it is specifically developed for the social sciences and it covers all the archiving actions present in the OAIS model. FORSbase doesn't have a source code available yet but SEEDS partners can currently use the hosted instance of the software called SEEDSbase. This instance is hosted and maintained by FORS. The opening of FORSbase source code is planned for the near future. Other applications' source code is available on Github and the documentation is well maintained.

In terms of financial costs, neither solution requires a licence, but despite the fact that they are freely available, the costs of installation, implementation and future maintenance have to be planned. SEEDSbase is an exception again, because it's hosted at FORS and currently there are no additional costs<sup>10</sup> for SEEDS project partners. The costs for implementation, infrastructure resources and maintenance varies between project partners and partners will have to take into consideration their local expertise and resources.

### 3.2. OAIS reference model coverage

The OAIS environment is derived from the interaction of four entities: producers, consumers, management and the archive itself<sup>11</sup>. In our case, producers provide data, consumers discover, download and use data, and the archive curates and stores data, while managing the whole process.

Research data are represented as information package. Each information package is a conceptual container of two types of information: metadata and the data itself. Research data versions in that process are referred to as:

- SIP - submission information package,
- AIP - archival information package, and

---

<sup>10</sup> This will probably change in the future and repository hosting and support of future potential FORSbase instances, except SEEDSbase, might be charged by FORS.

<sup>11</sup> <http://www.oclc.org/research/publications/library/2000/lavoie-oais.html>

- DIP - dissemination information package.

SIP is the version provided by the data producer, and includes all kinds of checking, validating, anonymizing etc. AIP and DIP are curated sets of data with accompanied documentation that is ready for long term preservation (AIP) or for further dissemination and re-use. A DIP can be the same as an AIP (version and content), but it is also a common practice to disseminate only a subset of data, which is needed to validate results or is valuable for re-use. Some extra information, sensitive data etc. can be restricted even further in the archive and stored just on the AIP level.

FORSbase and Dataverse are compliant with the OAIS model and therefore better fit our whole workflow. It is possible to manage pre-SIP and SIP phase giving end users the opportunity to use the system as some sort of a virtual research environment with the features of safe data storage and file sharing platform, but there are also long term preservation subsystems in place, as well as a frontend for data publishing and dissemination.

DSpace and EPrints are lacking the features needed for the SIP phase and in addition, the default user roles are not sufficiently divided to enable granular permissions. Nevertheless, the basic self-archiving, archival and publishing features are well supported and are fulfilling most of our demands. Regarding long-term preservation, discoverability and accessibility it is also worth mentioning that DSpace, Dataverse and FORSbase provide DOI registration functionality by default, while EPrints has a plugin available.

### 3.3. Administration, user interface and access

Within Dataverse, a dataset or a collection of datasets can have many levels of access and there are plenty of different roles that the administrator can assign to users. Some of the defaults are:

- Dataverse admin - A person who has all permissions for dataverses, datasets, and files;
- Research team member - A person who can view both unpublished and published dataverses and datasets;
- Curator - only for datasets; a person who can edit License + Terms, edit Permissions, and publish datasets;
- Contributor - only for datasets; a person who can edit License + Terms, and then submit them for review;
- Etc.

FORSbase has 4 types of Roles:

- researcher basic (can just download the data),
- researcher full (can work on a study/dataset). To the researcher\_full one of 3 types of rights can be assigned: :
  - 1. can edit (can create/edit/delete a study/dataset/datafile),
  - 2. can manage (includes 1 plus can add/remove other researchers to the study),
  - 3. can submit (includes 1 and 2 plus can submit a study/dataset)
- archive collaborator (1, 2 and 3 plus manage persons and institution, create AIP and DIP, publish study/dataset, manage groups, manage circles, manage bibliographies, can manage contracts, etc.)
- admin (all the rights + some system rights)

In DSpace, a user can add a dataset to a specific collection, edit his own or others' items, while administrators can approve, edit or reject records. Users can also have full or limited access for accessing the content.

EPrints has a basic default set of roles:

- Registered user - who can add records and access all published content.
- Editor - approves, edits, rejects records.

- Administrator - a user with all systems components.

All evaluated applications can be customized in terms of user permissions. Dataverse has the most granular implementation of available user's actions and default roles can be modified or added.

### 3.4. Performance and security issues

Performance and security issues are related to the specific software itself but also to the system, used to run the service. All evaluated software runs best in Linux based environment. Software functions are optimized and function better with more CPU/RAM/SSD disks. Dataverse and DSpace are slightly more resource consuming, due to their Java environment. Our production sites show that we are facing different sizes of software installations, different numbers of users and different frequencies of visits. All of the revised software handles such diversity well.

In terms of security, none of the evaluated software had a critical security bug up to date, so we can say that all satisfy our demands. Anyway, all systems in general should be regularly updated (the operating systems itself as well as the other components (i.e. web server, Java environment etc.)).

The connection between the client and the server is encrypted and secured by using HTTPS - a protocol for secure communication over the internet. Its enoWeb server is configured for using the valid SSL certificates, which can be bought and obtained from a certificate authority or generated for free, using the Letsencrypt service.

## 4. Evaluation overview by country and conclusion

Dataverse is chosen as the appropriate solution by the Serbian team after their local test installation. While having experience with EPrints and investigating the features of other solutions, Dataverse is considered as the most complete solution that will enable a quick start and many valuable features, such as good access and permissions management, DDI support, versioning etc.

Other partners did not install any software on their own, neither do they have any prior experience with repository software applications, but they were provided with access to the Croatian Dataverse demo instance and other softwares' demo sites. They also checked documentation and frontend of a few live archives to get the feeling of what is feasible to achieve with each software application.

After the assessment, Kosovo and Montenegro decided to use Dataverse because in that way they can have their own installation and also use the possibility to get help from SEEDS project partners in the region.

Macedonia and Albania do not have enough resources in terms of expertise and available infrastructure. They will continue using the hosted SEEDSbase installation, as there are currently no additional costs for hosting. Later, once the code will be open-sourced, they will have the chance to take over their instance and maintain it by themselves, as well as get support, if needed.

### 4.1. Albania

Through the course of the SEEDS project, IDM was able to acquire some relevant knowledge related to data archiving. In addition, during the RRPP Data Rescue project, the Albanian team had the opportunity to put in practice this knowledge by archiving some studies of Albanian

researchers using SEEDSbase. This experience was very useful to understand some of the aspects of the tool evaluation.

However, despite the significant improvements there are still several important challenges that need to be faced in order for IDM to be capable of providing the relevant service.

The main issues to be addressed are the following:

- Lack of appropriate IT expertise within IDM, both in human resources and hardware and software equipment.
- Lack of funding, as part of the final decision for establishing the Archive, for regular daily activities.

Thus, given the above mentioned limitations IDM deemed appropriate to continue using SEEDSbase platform as the most optimal solution at this phase. However, when finalizing the development of the National archive in social sciences in Albania (ADAS), the technical infrastructure challenges will be addressed.

## 4.2. Croatia

The Croatian team has hands on experience with all of the evaluated software solutions. Furthermore, they are responsible for the maintenance and the development of several institutional repositories in production, based on EPrints. Despite the fact that there is no Java expertise available in the institution, Dataverse is evaluated as the most appropriate solution for their future needs.

During the SERSCIDA FP7 project Croatian team has developed a testing platform consisting of several software components. EPrints was used at the ingest phase as well as the archiving tool, Nesstar server was used as a dissemination platform. Experience in conducting this pilot was very useful for the current activities in the SEEDS project and evaluating the chosen software packages. First of all Nesstar was not on the list because it's a proprietary software and the

development was stalled. While EPrints is an all-in-one archiving solution, it really lacks some of the important features like building collections, hierarchy of collections with separated user/admin permissions, versioning data files inside the item etc. In the SEEDS project, the Croatian team has created a testing installation of Dataverse and explored publishing workflows, access control features, search and discovery features and metadata support. The possibility of building the hierarchy of collections and sub-collections (dataverses) and giving various permissions and access right inside Dataverse is one of the main advantages comparing it to other solutions. The flexible metadata framework (three levels of metadata) is also a great approach compliant with the Croatian plans of integrating Dataverse with the existing national repository infrastructure. The Croatian team is investigating possibilities of integrating Dataverse with the repository hosting service DABAR<sup>12</sup>, based on Islandora software. In this scenario Dataverse would be used in the pre-SIP and SIP phase, while the Islandora part would be used as a long term preservation storage system and/or a dissemination platform. Dataverse features also give the opportunity to enable self-archiving and a virtual research environment for research teams. This would be a win-win situation where researchers on the one hand would be provided with the secure online platform for storing and managing their data - sharing it with research teams and collaborators, and the Archive on the other hand would use it as the platform for pre-ingest getting the datasets with more or less complete documentation which could be easily supplemented by Archive staff or the providers themselves and thus prepare it for archiving.

### 4.3. Kosovo

For Kosovo, as it is the case with other Western Balkan countries (except Serbia and Bosnia and Herzegovina), who had no experience with data archives, the SEEDS project for the first time provided us with a whole new terminology in terms of technical infrastructure but most importantly with pushing the agenda for a real data archive in the future. However, this is just

---

<sup>12</sup> <http://dabar.srce.hr>



the beginning of a long journey which requires a lot of expertise and good will in order to establish a proper data archive in the future. While countries such as Croatia and Serbia have already fully or partially overcome such problems, the main difficulties in Kosovo as mentioned by other partners remain:

- Lack of appropriate IT department and expertise within our institution (CPC) in human resources as well as human and software equipment
- Lack of funding as part of the final decision for establishing the Archive for regular daily activities.

Nevertheless, after the testing done by our data specialist as well as the positive aspects mentioned above, we have come into conclusion that the Dataverse platform would be the best solution whenever the real deal with establishing a data archive comes to life. We are aware of the possible challenges, but we believe that having our own installation as well as using help from SEEDS partners in the region, who are going to use or who already use Dataverse, will be very helpful and appreciated.

#### 4.4. Macedonia

The SEEDS project provided a first, more complex experience with data archiving. Although the Macedonian team from the Institute for sociological, political and juridical research (ISPJR) gained knowledge and in-depth information during the SEEDS project, and also participated in practical activities connected with data archiving, some important conditions are still not fulfilled.

The situation is characterised by:

- Lack of an appropriate IT department within ISPJR in both aspects: human resources and hardware and software equipment
- Lack of funding, as part of the final decision for establishing the Archive, for regular daily activities

Having in mind the limited resources, we decided to continue using SEEDSbase as the best solution, due to the fact that in this way we overcome the limitations that at the moment are beyond our power.

However, we are fully aware that parallel to the dynamic of establishing the Macedonian Social Science Data Archive – MKDASS, a final decision regarding the technical infrastructure should be taken.

## 4.5. Montenegro

Montenegro has launched a national debate on data archiving within the SEEDS project.

Within the organisation and infrastructure building process of the project the CeMI team has strengthened its capacities in policies, and data management, as well as improving its knowledge on data archiving.

Within the first phase of the SEEDS project it could be concluded that there is a strong need within the social scientist community for data archiving, while those kinds of services are not in place. At the same time, there is a strong need for promoting open data, data archiving, but also standards in collecting, and managing data.

Through the process of analysing different software, using developed criteria for evaluation, the CeMI team has selected Dataverse as the repository solution. The main reasons behind this are the following: its availability, being an open-source software, being used by many well established universities, being already chosen by other repositories in the Region (which promotes collaboration and exchange of knowledge), having all necessary characteristics, documentation of the software is well written and maintained, the software seems to be well maintained, it also seems to be quite easy to use.

CeMI will continue to improve its capacities to provide quality of services in data preservation and sharing enabling, will promoting the establishment of an official national data repository for data from the social sciences (Montenegrin Social Science Data Archive Name/Crnogorski arhiv podataka iz istraživanja društvenih nauka), as well as its use among researchers.

## 4.6. Serbia

The development of a data archiving infrastructure in Serbia begun within the FP7 SERSCIDA project (2012-2014). One of the main results of this project was a prototype of the database which includes decisions about software requirements for a new archive. The following solutions were selected: EPrints for the Ingest phase, Metadata management, Archival storage and for maintaining the user database; Request Tracker (RT) for management of tasks, issues, and requests submitted by a community of users and Nesstar (Nesstar Publisher for preparing metadata descriptions for the SIP as well as for converting AIPs into DIPs and preparing the data for Access, Nesstar Server as a storage of distribution materials (parts of DIP) and Nesstar WebView for the presentation of data and metadata of a study and for on-line analysis and download). During and after the project, the Serbian team had a chance only to test this software. The main issues with EPrints were the lack of well-maintained documentation and the fact that the metadata fields are not DDI compliant. The main obstacle for using Nesstar was the complex structure of the tools and a license that must be renewed every two years. Additionally, the SERSCIDA project has enabled the purchase of hardware and today the Institute of Economic Sciences (IES) has a server DELL PowerEdge R720 which can support the development of archiving in social sciences in Serbia.

During the SERSCIDA project, the Serbian team tested the Nesstar platform, but the main obstacles were annual cost for the licence and the inability to store qualitative data sets. They also installed EPrints with Recollect plugin, and conclude that it was also sufficiently satisfactory for its incompatibility with DDI standard. During the RRPP Data Rescue project, they tested and used the SEEDSbase (FORSbase) platform and manage to archive 23 studies with quantitative

and qualitative datasets. This experience was very useful in the process of the tool evaluation, but they could not choose this platform because it has no open licence yet. Also, as a part of the CESSDA SaW project, they installed and tested a Dataverse instance. Finally, the Serbian team decided that Dataverse is the best solution for them because it is open licence, free of charge, has a good technical support and documentation of the software well written and maintained, has DDI metadata specification, and has all other relevant features that are important for good quality of data preservation.

The main problem they need to solve in the near future is the limited IT expertise. Possible solutions include the establishment of a strategic partnership with an organization that has a good IT support or outsourcing IT professionals that would satisfy the needs of the new Data Centre.

## 5. Annex: Lists of criteria for evaluation of the archiving tools

### 5.1 Albania

Req. #	Requirement description	Section	Category	Criteria	SEEDSBase			Dataverse			DSpace			EPrints	
					Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment
1.1.1	Open Source License. Everyone is able to download and install it.	I Community	General	must have	0	FORS will open the source code in the near future, most probably 2018	NA	2	Available on Github		2	Available on Github		2	Available on Github
1.1.2.1	Annual costs for license	I Community	General		0	Currently, no annual costs	NA	0	No costs		0	No costs		0	No costs
1.1.2.2	Annual costs for support	I Community	General		0	Currently there are no annual costs, but once the tool is opensource, there might be some costs for technical support	NA	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced
1.1.3	At least had been used by several institutions	I Community	General	must have	1	Used by FORS and SEEDS partners for RRRP project	NA	1	Dataverse is used by several large US Universities and in Europe DANS is using it as a part of their EASY system.		2	There is a strong user community worldwide. Many institutions use DSpace as their institutional repository		2	There is a plenty of institutions that are using EPrints. The user community is especially strong in the UK
1.1.5	Capability to be implemented / customized as archiving tool?	I Community	General	nice to have	2	It is an archiving tool	none	2	It is an archiving tool		2	It is an archiving tool		2	It is an archiving tool
1.2.1	Documentation of the software well written and maintained.	I Community	Software-Development	must have	1	There is documentation, but not very well maintained	none	2	Documentation is well maintained		2	Documentation is well maintained		2	Documentation not well maintained and there are occasional problems with the access to their wiki pages
1.2.2	Codes are maintained using version control system.	I Community	Software-Development	must have	2	Git is used as version control system	none	2	GIT		2	GIT		2	GIT

1.2.3	The dependency with other software components is tidily maintained.	I Community	Software-Development	must have	2	SEEDSbase is linked with FedoraCommons and the dependency is tidily maintained	none	2	Dataverse depends on many opensource components and dependancies are well maintained	2	Dataverse depends on many opensource components and dependancies are well maintained	1	Dataverse depends on many opensource components but dependancies are not very well maintained
1.2.4	If modules/features not open source, at least open APIs, and common programming language is used.	I Community	Software-Development	must have	2	Everything will be opensource in a near future. Programming language Python and framework Django were used for the development.	NA	2	Everything is opensource and the licences guarantee that this will persist in a future	2	Everything is opensource and the licences guarantee that this will persist in a future	2	Everything is opensource and the licences guarantee that this will persist in a future
1.2.5	Possibility to buy Plugins (Add-ons) for features not included in the open source version	I Community	Software-Development	nice to have	0	SEEDSbase is implemented in a way to easily integrate any additional functionality/module. There is no plan to sell specific plugins in a near future.	high	1	One cannot buy already developed plugins but it's possible to hire a developer or to contribute own solution	1	There is addons directory available and it's possible to contribute by developng or funding new features	1	EPrints has a plugin repository called Bazaar and it's possible to install plugins directly within the administrative web interface. Also, it's pretty easy to develop new plugins.
1.2.6	Internal IT Know-how	I Community	Software-Development	nice to have	0	The know-how is currently only at FORS	high	0	No, it's Java	0	No, it's Java	2	Yes, Perl
2.1.0	Multi-language support	II Functionality	General	must have	2	User interface is in 3 languages (EN, FR and DE) as for studies and datasets there can be used as many as needed languages. There will be some additional development effort in case you want to translate user interface in some other language.	low	0	It doesn't have multilang support and it's not on the official roadmap but we have some internal info that it's planned to start working on it.	2	Yes	2	Yes, metadata and interface could be internationalized
2.1.1	Ingest process	II Functionality	Ingest	must have	2	Fully covered	none	2	It depends on the way of using the tool. Submission is fully covered but one cannot	1	SIP cannot be saved in it's initial version	1	SIP cannot be saved in initial version

								save the state of SIP. Dataset can be either in draft/unpublished or published. Once published older states are lost but certain files could left as restricted for historical or other reasons.							
2.1.2	Virus check	II Functionality	Ingest	must have	2	Fully covered (antivirus check is currently disabled in SEEDSbase instance, but the functionality is in place)	none	1	Not by default		1	Not by default		1	Not by default
2.1.3	Check sum	II Functionality	Ingest	must have	2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.1.4	File type recognition	II Functionality	Ingest	must have	2	There is a list of desired filetypes that can be maintained by the archive. The system checks the uploaded filetype against this list, and if it's in a format that is not in the list the file will not be uploaded. In this way archive is making sure to receive right file formats as SIP.	none	2	MIME type recognition, tabular filetypes are converted on the fly, SPSS files are converted also and XML DDI is generated		2	MIME		2	MIME
2.1.5	Generate SIP	II Functionality	Ingest	must have	2	Once the submission process has started, the system generates SIP and freeze this version, so it can always be checked what was submitted at the first place and what were original documents.	none	1	Unpublished draft could be considered SIP but once published it becomes AIP		0	No		0	No
2.1.6	Generate Descriptive Information/Metadata	II Functionality	Ingest	must have	2	Fully covered. DDI compliant and in alignment with CESSDA recommended fields.	none	2	Yes, and DDI XML could be created from SPSS files		1	No, only what's filled in		1	No, only what's filled in



2.1.7	Coordinate Updates	II Functionality	Ingest	nice to have	2	All the updates made by the producer or archive are fully coordinated using status and versioning systems.	none	2	Actions are logged and files are versioned	2	Actions are logged and files are versioned	2	Actions are logged and files are versioned
2.2.1	Receive Data	II Functionality	Archival Storage	must have	2	Fully covered.	none	2	Covered	2	Covered	2	Covered
2.2.2	Long term preservation	II Functionality	Archival Storage	must have	2	All the data are stored in Fedora.	none	2	Preservation procedures are implemented	2	Preservation support is in place	1	There is a Preservation Toolkit available as plugin
2.2.3	Generate AIP	II Functionality	Archival Storage	nice to have	2	AIP is stored in Fedora	none	2	Publishing means creating AIP and DIP	2	Publishing means creating AIP and DIP	2	Publishing means creating AIP and DIP
2.2.4	Replace media / Convert file types	II Functionality	Archival Storage	nice to have	1	Currently the process of format migration is done manually. The people from the archive wanted to have better control here. No plans in near future to implement this as automatical conversion.	high	1	Automatic migration is not implemented but conversion is covered in preservation part	1	Automatic migration is not implemented but conversion is covered in preservation part	1	Preservation toolkit identifies archived formats and lists possible items that could be converted
2.2.5	Error Checking	II Functionality	Archival Storage	nice to have	2	Fully covered	none	2	Covered	2	Covered	2	Covered
2.2.6	Disaster Recovery	II Functionality	Archival Storage	nice to have	1	This is not done by the tool itself. It's rather on a server hosting side. How the server is maintained, how the data are backedup, etc.	NA	1	This is the part of backup/restore procedures not fully covered by the tool itself	1	This is the part of backup/restore procedures not fully covered by the tool itself	1	This is the part of backup/restore procedures not fully covered by the tool itself
2.3.1	Deposit contract	II Functionality	Administration	must have	2	SEEDSbase supports even different deposit contracts for different datasets.	none	2	It supports general terms of use and licences as well as special cases that can be applied to a collection or to individual dataset	2	General terms and licences, it's possible to add specific conditions	2	General terms and licences but it's also possible to add specific conditions
2.3.2	Database of Users	II Functionality	Administration	must have	2	Users, persons and institutions are captured	none	2	Yes	2	Yes	2	Yes

2.3.3	User management	II Functionality	Administration	must have	2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.3.4	Communication with users	II Functionality	Administration	nice to have	2	Fully covered	none	0	No		0	No		0	No
2.4.1	Download contract	II Functionality	Access	must have	2	SEEDSbase supports even different download contracts for different datasets.	none	2	Various possibilities		1	Similar to 2.3.1		1	Smilar to 2.3.1
2.4.2	Level of access	II Functionality	Access	must have	2	Fully covered	none	2	Fine grained system of permissions		2	User, Editor, SuperAdmin		2	User, Editor, SupeAdmin
2.4.3	Generate DIP	II Functionality	Access	must have	2	AIP is stored in Fedora	none	1	See 2.1.5		1	See 2.1.5		1	See 2.1.5
2.4.4	Data catalogue	II Functionality	Access	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.5	Simple and Advanced search facilities	II Functionality	Access	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.6	Browse metadata, data and associated documentation	II Functionality	Access	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered, but no faceted search
2.4.7	Download datasets in a variety of statistical formats	II Functionality	Access	nice to have	2	It can be decided what is included into DIP, this means we can include different file formats for the same dataset.	none	2	Different file formats are supported and some automatic conversions are implemented		1	Different file formats are supported		1	Different file formats are supported
2.5.1	Login, User Roles and Permissions	II Functionality	General	must have	2	There are 4 types of Roles: 1. researcher basic (can just download the data), 2. researcher full (can work on a study/dataset, and here we have 3 types of users: 1. can edit 2. can manage 3. can submit), 3. archive collaborator and 4.	none	2	Fine grained user control on the collection level as well as on the dataset and individual file level		2	Simple access controls		2	Simple access controls

						admin (all the rights + some system rights)									
2.5.2	Users are able to edit their own profiles	II Functionality	General	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.5.3	Users communicate with each other via platform internal messages	II Functionality	General	nice to have	2	They can communicate via internal messaging system and/or via email using the system. Also user can set if he wants to be notified by email every time he receives a message	none	0	No		0	No		0	No
2.5.4	The administrator of the platform is able to send E-Mails to the users	II Functionality	General	nice to have	2	He can send emails and/or private messages	none	0	No		0	No		0	No
3.1.1	Compatibility with operating systems	III Technical Basis	Platform	nice to have	2	SEEDSbase is fully compatible with Linux. It should work on any OS since it works on Apache and it's using Python.	none	1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported
3.2.1	Scalability e.g. with many users simultaneously	III Technical Basis	Performance	must have	2	We have currently 3300 active users in FORSbase and no performance issues reported. (If you noticed some performance issues in SEEDSbase it might be due to the power of the server, since the server we are using for SEEDSbase is less powerful).	none	2	It can be scaled by adding application and database servers		2	It works in different environments		2	It works in different environments
3.2.2	Usable through slow internet connections	III Technical Basis	Performance	nice to have	2	Fully covered	none	2	Covered		2	Covered		2	Covered

3.2.3	User-friendly, clear and intuitive structure?	III Technical Basis	Performance	must have	1	It can be more userfriendly. It is planned to do the whole user interface redesign in 2018.	high	2	It could be optimized	high	2	It could be optimized	high	2	It could be optimized
3.3.1	How and where the information & data stored?	III Technical Basis	Security	must have	2	In FedoraCommons	none	2	Local database and filesystem		2	Local database and filesystem		2	Local database and filesystem
3.3.2	Is the connection between server and portal (and client) secure?	III Technical Basis	Security	must have	2	Fully covered	none	2	Covered, depends on web server		2	Covered, depends on web server		2	Covered, depends on web server

## 5.2 Croatia

Req. #	Requirement description	Section	Category	Priority	Development effort	SEEDSBase		Development effort	Dataverse			DSpace			EPrints	
						Score	Comment		Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment
1.1.1	Open Source License. Everyone is able to download and install it.	I Community	General	must have	NA	0	FORS will open the source code in the near future, most probably 2018	NA	2	Available on Github		2	Available on Github		2	Available on Github
1.1.2.1	Annual costs for licence*	I Community	General		NA	0	Currently, no annual costs	NA	0	No costs		0	No costs		0	No costs
1.1.2.2	Annual costs for support*	I Community	General		NA	0	Currently there are no annual costs, but once the tool is opensource, there might be some costs for technical support	NA	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human

										available for that purpose, this work has to be outsourced			resources available for that purpose, this work has to be outsourced			resources available for that purpose, this work has to be outsourced
1.1.3	At least had been used by several institutions	I Community	General	must have	NA	1	Used by FORS and SEEDS partners for RRPP project	NA	1	Dataverse is used by several large US Universities and in Europe DANS is using it as a part of ther EASY system.	2	There is a strog user community worldwide. Many institutions use DSpace as their institutional repository		2	There is a plenty of institutions that are using EPrints. The user community is expecially strong in the UK	
1.1.5	Capability to be implemented / customized as archiving tool?	I Community	General	nice to have	NA	2	It is an archiving tool	none	2	It is an archiving tool	2	It is an archiving tool		2	It is an archiving tool	
1.2.1	Documentation of the software well written and maintained.	I Community	Software-Development	must have	NA	1	There is documentation, but not very well maintained	none	2	Documentatio n is well maintained	2	Documenta tion is well maintained		2	Documentati on not well maintained and there are occasional problems with the access to their wiki pages	
1.2.2	Codes are maintained using version control system.	I Community	Software-Development	must have		2	Git is used as version control system	none	2	GIT	2	GIT		2	GIT	
1.2.3	The dependency with other software components is tidily maintained.	I Community	Software-Development	must have		2	SEEDSbase is linked with FedoraCommons and the dependency is tidily maintained	none	2	Dataverse depends on many opensource components and dependancies are well maintained	2	Dataverse depends on many opensource components and dependancies are well maintained		1	Dataverse depends on many opensource components but dependancies are not very well maintained	
1.2.4	If modules/features not open source, at least open APIs, and common programming language is used.	I Community	Software-Development	must have		2	Everything will be opensource in a near future. Programming language Python and framework Django were used for the development.	NA	2	Everything is opensource and the licences guarantee that this will persist in a future	2	Everything is opensource and the licences guarantee that this will persist in a future		2	Everything is opensource and the licences guarantee that this will persist in a future	

1.2.5	Possibility to buy Plugins (Add-ons) for features not included in the open source version	I Community	Software-Development	nice to have		0	SEEDSbase is implemented in a way to easily integrate any additional functionality/module. There is no plan to sell specific plugins in a near future.	high	1	One cannot buy already developed plugins but it's possible to hire a developer or to contribute own solution		1	There is addons directory available and it's possible to contribute by developng or funding new features		1	EPrints has a plugin repository called Bazaar and it's possible to install plugins directly within the administrative web interface. Also, it's pretty easy to develop new plugins.
1.2.6	Internal IT Know-how	I Community	Software-Development	nice to have		0	The know-how is currently only at FORS	high	0	No, it's Java		0	No, it's Java		2	Yes, Perl
2.1.0	Multi-language support	II Functionality	General	nice to have		2	User interface is in 3 languages (EN, FR and DE) as for studies and datasets there can be used as many as needed languages. There will be some additional development effort in case you want to translate user interface in some other language.	low	0	It doesn't have multilang support and it's not on the official roadmap but we have some internal info that it's planned to start working on it.		2	Yes		2	Yes, metadata and interface could be internationalized
2.1.1	Ingest process	II Functionality	Ingest	must have		2	Fully covered	none	2	It depends on the way of using the tool. Submission is fully covered but one cannot save the state of SIP. Dataset can be either in draft/unpublished or published. Once published older states are lost but certain files could left as		1	SIP cannot be saved in it's initial version		1	SIP cannot be saved in initial version

									restricted for historical or other reasons.						
2.1.2	Virus check	II Functionality	Ingest	nice to have	2	Fully covered (antivirus check is currently disabled in SEEDSbase instance, but the functionality is in place)	none	1	Not by default		1	Not by default		1	Not by default
2.1.3	Check sum	II Functionality	Ingest	must have	2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.1.4	File type recognition	II Functionality	Ingest	must have	2	There is a list of desired filetypes that can be maintained by the archive. The system checks the uploaded filetype against this list, and if it's in a format that is not in the list the file will not be uploaded. In this way archive is making sure to receive right file formats as SIP.	none	2	MIME type recognition, tabular filetypes are converted on the fly, SPSS files are converted also and XML DDI is generated		2	MIME		2	MIME
2.1.5	Generate SIP	II Functionality	Ingest		2	Once the submission process has started, the system generates SIP and freeze this version, so it can always be checked what was submitted at the first place and what were original documents.	none	1	Unpublished draft could be considered SIP but once published it becomes AIP		0	No		0	No
2.1.6	Generate Descriptive Information/Metadata	II Functionality	Ingest	must have	2	Fully covered. DDI compliant and in alignment with CEESDA recommended fields.	none	2	Yes, and DDI XML could be created from SPSS files		1	No, only what's filled in		1	No, only what's filled in
2.1.7	Coordinate Updates	II Functionality	Ingest	nice to have	2	All the updates made by the producer or	none	2	Actions are logged and files are		2	Actions are logged and files are		2	Actions are logged and files are

						archive are fully coordinated using status and versioning systems.	none		versioned		versioned		versioned
2.2.1	Receive Data	II Functionality	Archival Storage	must have	2	Fully covered.	none	2	Covered	2	Covered	2	Covered
2.2.2	Long term preservation	II Functionality	Archival Storage	must have	2	All the data are stored in Fedora.	none	2	Preservation procedures are implemented	2	Preservation support is in place	1	There is a Preservation Toolkit available as plugin
2.2.3	Generate AIP	II Functionality	Archival Storage	nice to have	2	AIP is stored in Fedora	none	2	Publishing means creating AIP and DIP	2	Publishing means creating AIP and DIP	2	Publishing means creating AIP and DIP
2.2.4	Replace media / Convert file types	II Functionality	Archival Storage	nice to have	1	Currently the process of format migration is done manually. The people from the archive wanted to have better control here. No plans in near future to implement this as automatical conversion.	high	1	Automatic migration is not implemented but conversion is covered in preservation part	1	Automatic migration is not implemented but conversion is covered in preservation part	1	Preservation toolkit identifies archived formats and lists possible items that could be converted
2.2.5	Error Checking	II Functionality	Archival Storage	nice to have	2	Fully covered	none	2	Covered	2	Covered	2	Covered
2.2.6	Disaster Recovery	II Functionality	Archival Storage	nice to have	1	This is not done by the tool itself. It's rather on a server hosting side. How the server is maintained, how the data are backedup, etc.	NA	1	This is the part of backup/restore procedures not fully covered by the tool itself	1	This is the part of backup/restore procedures not fully covered by the tool itself	1	This is the part of backup/restore procedures not fully covered by the tool itself
2.3.1	Deposit contract	II Functionality	Administration	must have	2	SEEDSbase supports even different deposit contracts for different datasets.	none	2	It supports general terms of use and licences as well as special cases that can be applied to a collection or to individual dataset	2	General terms and licences, it's possible to add specific conditions	2	General terms and licences but it's also possible to add specific conditions
2.3.2	Database of Users	II Functionality	Administration	must have	2	Users, persons and institutions are captured	none	2	Yes	2	Yes	2	Yes



2.3.3	User management	II Functionality	Administration	must have	2	Fully covered	none	2	Yes	2	Yes	2	Yes
2.3.4	Communication with users	II Functionality	Administration	nice to have	2	Fully covered	none	0	No	0	No	0	No
2.4.1	Download contract	II Functionality	Access	must have	2	SEEDSbase supports even different download contracts for different datasets.	none	2	Various possibilities	1	Similar to 2.3.1	1	Smilar to 2.3.1
2.4.2	Level of access	II Functionality	Access	must have	2	Fully covered	none	2	Fine grained system of permissions	2	User, Editor, SuperAdmin	2	User, Editor, SupeAdmin
2.4.3	Generate DIP	II Functionality	Access	must have	2	AIP is stored in Fedora	none	1	See 2.1.5	1	See 2.1.5	1	See 2.1.5
2.4.4	Data catalogue	II Functionality	Access	must have	2	Fully covered	none	2	Covered	2	Covered	2	Covered
2.4.5	Simple and Advanced search facilities	II Functionality	Access	must have	2	Fully covered	none	2	Covered	2	Covered	2	Covered
2.4.6	Browse metadata, data and associated documentation	II Functionality	Access	must have	2	Fully covered	none	2	Covered	2	Covered	2	Covered, but no faceted search
2.4.7	Download datasets in a variety of statistical formats	II Functionality	Access	nice to have	2	It can be decided what is included into DIP, this means we can include different file formats for the same dataset.	none	2	Different file formats are supported and some automatic conversions are implemented	1	Different file formats are supported	1	Different file formats are supported
2.5.1	Login, User Roles and Permissions	II Functionality	General	must have	2	There are 4 types of Roles: 1. researcher basic (can just download the data), 2. researcher full (can work on a study/dataset, and here we have 3 types of users: 1. can edit 2. can manage 3. can submit), 3. archive	none	2	Fine grained user control on the collection level as well as on the dataset and individual file level	2	Simple access controls	2	Simple access controls

						collaborator and 4. admin (all the rights + some system rights)									
2.5.2	Users are able to edit their own profiles	II Functionality	General	must have		2	Fully covered	none	2	Covered		2	Covered	2	Covered
2.5.3	Users communicate with each other via platform internal messages	II Functionality	General	nice to have		2	They can communicate via internal messaging system and/or via email using the system. Also user can set if he wants to be notified by email every time he receives a message	none	0	No		0	No	0	No
2.5.4	The administrator of the platform is able to send E-Mails to the users	II Functionality	General	nice to have		2	He can send emails and/or private messages	none	0	No		0	No	0	No
3.1.1	Compatibility with operating systems	III Technical Basis	Platform	nice to have		2	SEEDSbase is fully compatible with Linux. It should work on any OS since it works on Apache and it's using Python.	none	1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported	1	Linux by default - other OSs possible but not supported
3.2.1	Scalability e.g. with many users simultaneously	III Technical Basis	Performance	must have		2	We have currently 3300 active users in FORSbase and no performance issues reported. (If you noticed some performance issues in SEEDSbase it might be due to the power of the server, since the server we are using for SEEDSbase is less powerful).	none	2	It can be scaled by adding application and database servers		2	It works in different environments	2	It works in different environments
3.2.2	Usable through slow internet connections	III Technical Basis	Performance	nice to have		2	Fully covered	none	2	Covered		2	Covered	2	Covered

3.2.3	User-friendly, clear and intuitive structure?	III Technical Basis	Performance	must have	1	It can be more userfriendly. It is planned to do the whole user interface redesign in 2018.	high	2	It could be optimized	high	2	It could be optimized	high	2	It could be optimized
3.3.1	How and where the information & data stored?	III Technical Basis	Security	must have	2	In FedoraCommons	none	2	Local database and filesystem		2	Local database and filesystem		2	Local database and filesystem
3.3.2	Is the connection between server and portal (and client) secure?	III Technical Basis	Security	must have	2	Fully covered	none	2	Covered, depends on web server		2	Covered, depends on web server		2	Covered, depends on web server

## 5.3 Kosovo

Req. #	Requirement description	Section	Category	Criteria		SEEDSBase			Dataverse			DSpace			EPrints	
				Priority	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment
1.1.1	Open Source License. Everyone is able to download and install it.	I Community	General	must have	NA	0	FORS will open the source code in the near future, most probably 2018	NA	2	Available on Github		2	Available on Github		2	Available on Github
1.1.2.1	Annual costs for licence*	I Community	General		NA	0	Currently, no annual costs	NA	0	No costs		0	No costs		0	No costs
1.1.2.2	Annual costs for support*	I Community	General		NA	0	Currently there are no annual costs, but once the tool is opensource, there might be some costs for technical support	NA	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced

													outsourced		
1.1.3	At least had been used by several institutions	I Community	General	must have	NA	1	Used by FORS and SEEDS partners for RRPP project	NA	1	Dataverse is used by several large US Universities and in Europe DANS is using it as a part of their EASY system.		2	There is a strong user community worldwide. Many institutions use DSpace as their institutional repository	2	There is a plenty of institutions that are using EPrints. The user community is especially strong in the UK
1.1.5	Capability to be implemented / customized as archiving tool?	I Community	General	nice to have	NA	2	It is an archiving tool	none	2	It is an archiving tool		2	It is an archiving tool	2	It is an archiving tool
1.2.1	Documentation of the software well written and maintained.	I Community	Software-Development	must have	NA	1	There is documentation, but not very well maintained	none	2	Documentation is well maintained		2	Documentation is well maintained	2	Documentation not well maintained and there are occasional problems with the access to their wiki pages
1.2.2	Codes are maintained using version control system.	I Community	Software-Development	must have		2	Git is used as version control system	none	2	GIT		2	GIT	2	GIT
1.2.3	The dependency with other software components is tidily maintained.	I Community	Software-Development	must have		2	SEEDSbase is linked with FedoraCommons and the dependency is tidily maintained	none	2	Dataverse depends on many opensource components and dependancies are well maintained		2	Dataverse depends on many opensource components and dependancies are well maintained	1	Dataverse depends on many opensource components but dependancies are not very well maintained
1.2.4	If modules/features not open source, at least open APIs, and common programming language is used.	I Community	Software-Development	must have		2	Everything will be opensource in a near future. Programming language Python and framework Django were used for the development.	NA	2	Everything is opensource and the licences guarantee that this will persist in a future		2	Everything is opensource and the licences guarantee that this will persist in a future	2	Everything is opensource and the licences guarantee that this will persist in a future

1.2.5	Possibility to buy Plugins (Add-ons) for features not included in the open source version	I Community	Software-Development	nice to have		0	SEEDSbase is implemented in a way to easily integrate any additional functionality/module. There is no plan to sell specific plugins in a near future.	high	1	One cannot buy already developed plugins but it's possible to hire a developer or to contribute own solution		1	There is addons directory available and it's possible to contribute by developng or funding new features		1	EPrints has a plugin repository called Bazaar and it's possible to install plugins directly within the administrative web interface. Also, it's pretty easy to develop new plugins.
1.2.6	Internal IT Know-how	I Community	Software-Development	nice to have		0	The know-how is currently only at FORS	high	0	No, it's Java		0	No, it's Java		2	Yes, Perl
2.1.0	Multi-language support	II Functionality	General	must have		2	User interface is in 3 languages (EN, FR and DE) as for studies and datasets there can be used as many as needed languages. There will be some additional development effort in case you want to translate user interface in some other language.	low	0	It doesn't have multilang support and it's not on the official roadmap but we have some internal info that it's planned to start working on it.		2	Yes		2	Yes, metadata and interface could be internationalized
2.1.1	Ingest process	II Functionality	Ingest	must have		2	Fully covered	none	2	It depends on the way of using the tool. Submission is fully covered but one cannot save the state of SIP. Dataset can be either in draft/unpublished or published. Once published older states are lost but certain files		1	SIP cannot be saved in it's initial version		1	SIP cannot be saved in initial version

									could left as restricted for historical or other reasons.							
2.1.2	Virus check	II Functionality	Ingest	nice to have		2	Fully covered (antivirus check is currently disabled in SEEDSbase instance, but the functionality is in place)	none	1	Not by default		1	Not by default		1	Not by default
2.1.3	Check sum	II Functionality	Ingest	must have		2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.1.4	File type recognition	II Functionality	Ingest	must have		2	There is a list of desired filetypes that can be maintained by the archive. The system checks the uploaded filetype against this list, and if it's in a format that is not in the list the file will not be uploaded. In this way archive is making sure to receive right file formats as SIP.	none	2	MIME type recognition, tabular filetypes are converted on the fly, SPSS files are converted also and XML DDI is generated		2	MIME		2	MIME
2.1.5	Generate SIP	II Functionality	Ingest	nice to have		2	Once the submission process has started, the system generates SIP and freeze this version, so it can always be checked what was submitted at the first place and what were original documents.	none	1	Unpublished draft could be considered SIP but once published it becomes AIP		0	No		0	No
2.1.6	Generate Descriptive Information/Metadata	II Functionality	Ingest	must have		2	Fully covered. DDI compliant and in alignment with CEESDA recommended fields.	none	2	Yes, and DDI XML could be created from SPSS files		1	No, only what's filled in		1	No, only what's filled in

2.1.7	Coordinate Updates	II Functionality	Ingest	nice to have		2	All the updates made by the producer or archive are fully coordinated using status and versioning systems.	none	2	Actions are logged and files are versioned		2	Actions are logged and files are versioned		2	Actions are logged and files are versioned
2.2.1	Receive Data	II Functionality	Archival Storage	must have		2	Fully covered.	none	2	Covered		2	Covered		2	Covered
2.2.2	Long term preservation	II Functionality	Archival Storage	must have		2	All the data are stored in Fedora.	none	2	Preservation procedures are implemented		2	Preservation support is in place		1	There is a Preservation Toolkit available as plugin
2.2.3	Generate AIP	II Functionality	Archival Storage	nice to have		2	AIP is stored in Fedora	none	2	Publishing means creating AIP and DIP		2	Publishing means creating AIP and DIP		2	Publishing means creating AIP and DIP
2.2.4	Replace media / Convert file types	II Functionality	Archival Storage	nice to have		1	Currently the process of format migration is done manually. The people from the archive wanted to have better control here. No plans in near future to implement this as automatical conversion.	high	1	Automatic migration is not implemented but conversion is covered in preservation part		1	Automatic migration is not implemented but conversion is covered in preservation part		1	Preservation toolkit identifies archived formats and lists possible items that could be converted
2.2.5	Error Checking	II Functionality	Archival Storage	nice to have		2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.2.6	Disaster Recovery	II Functionality	Archival Storage	nice to have		1	This is not done by the tool itself. It's rather on a server hosting side. How the server is maintained, how the data are backedup, etc.	NA	1	This is the part of backup/restore procedures not fully covered by the tool itself		1	This is the part of backup/restore procedures not fully covered by the tool itself		1	This is the part of backup/restore procedures not fully covered by the tool itself
2.3.1	Deposit contract	II Functionality	Administration	must have		2	SEEDSbase supports even different deposit contracts for different datasets.	none	2	It supports general terms of use and licences as well as special cases that can be applied to a		2	General terms and licences, it's possible to add specific conditions		2	General terms and licences but it's also possible to add specific conditions

									collection or to individual dataset						
2.3.2	Database of Users	II Functionality	Administration	must have	2	Users, persons and institutions are captured	none	2	Yes		2	Yes		2	Yes
2.3.3	User management	II Functionality	Administration	must have	2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.3.4	Communication with users	II Functionality	Administration	nice to have	2	Fully covered	none	0	No		0	No		0	No
2.4.1	Download contract	II Functionality	Access	must have	2	SEEDSbase supports even different download contracts for different datasets.	none	2	Various possibilities		1	Similar to 2.3.1		1	Similar to 2.3.1
2.4.2	Level of access	II Functionality	Access	must have	2	Fully covered	none	2	Fine grained system of permissions		2	User, Editor, SuperAdmin		2	User, Editor, SuperAdmin
2.4.3	Generate DIP	II Functionality	Access	must have	2	AIP is stored in Fedora	none	1	See 2.1.5		1	See 2.1.5		1	See 2.1.5
2.4.4	Data catalogue	II Functionality	Access	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.5	Simple and Advanced search facilities	II Functionality	Access	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.6	Browse metadata, data and associated documentation	II Functionality	Access	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered, but no faceted search
2.4.7	Download datasets in a variety of statistical formats	II Functionality	Access	must have	2	It can be decided what is included into DIP, this means we can include different file formats for the same dataset.	none	2	Different file formats are supported and some automatic conversions are implemented		1	Different file formats are supported		1	Different file formats are supported
2.5.1	Login, User Roles and Permissions	II Functionality	General	must have	2	There are 4 types of Roles: 1. researcher basic (can just download the data), 2. researcher full (can work on a study/dataset,	none	2	Fine grained user control on the collection level as well as on the dataset and individual file level		2	Simple access controls		2	Simple access controls



						and here we have 3 types of users: 1. can edit 2. can manage 3. can submit), 3. archive collaborator and 4. admin (all the rights + some system rights)									
2.5.2	Users are able to edit their own profiles	II Functionality	General	must have	2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.5.3	Users communicate with each other via platform internal messages	II Functionality	General	nice to have	2	They can communicate via internal messaging system and/or via email using the system. Also user can set if he wants to be notified by email every time he receives a message	none	0	No		0	No		0	No
2.5.4	The administrator of the platform is able to send E-Mails to the users	II Functionality	General	nice to have	2	He can send emails and/or private messages	none	0	No		0	No		0	No
3.1.1	Compatibility with operating systems	III Technical Basis	Platform	must have	2	SEEDSbase is fully compatible with Linux. It should work on any OS since it works on Apache and it's using Python.	none	1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported
3.2.1	Scalability e.g. with many users simultaneously	III Technical Basis	Performance	must have	2	We have currently 3300 active users in FORSbase and no performance issues reported. (If you noticed some performance issues in SEEDSbase it might be due to the power of the server, since the server we are using for	none	2	It can be scaled by adding application and database servers		2	It works in different environments		2	It works in different environments

							SEEDSbase is less powerfull).									
3.2.2	Usable through slow internet connections	III Technical Basis	Performance	nice to have		2	Fully covered	none	2	Covered		2	Covered		2	Covered
3.2.3	User-friendly, clear and intuitive structure?	III Technical Basis	Performance	must have		1	It can be more userfriendly. It is planned to do the whole user interface redesign in 2018.	high	2	It could be optimized	high	2	It could be optimized	high	2	It could be optimized
3.3.1	How and where the information & data stored?	III Technical Basis	Security	must have		2	In FedoraCommons	none	2	Local database and filesystem		2	Local database and filesystem		2	Local database and filesystem
3.3.2	Is the connection between server and portal (and client) secure?	III Technical Basis	Security	must have		2	Fully covered	none	2	Covered, depends on web server		2	Covered, depends on web server		2	Covered, depends on web server

## 5.4 Macedonia

Req. #	Requirement description	Section	Category	Criteria		SEEDSBase			Dataverse			DSpace			EPrints	
				Priority	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment
1.1.1	Open Source License. Everyone is able to download and install it.	I Community	General	must have		0	FORS will open the source code in the near future, most probably 2018	NA	2	Available on Github		2	Available on Github		2	Available on Github
1.1.2.1	Annual costs for licence	I Community	General			0	Currently, no annual costs	NA	0	No costs		0	No costs		0	No costs

1.1.2.2	Annual costs for support	I Community	General		0	Currently there are no annual costs, but once the tool is opensource, there might be some costs for technical support	NA	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced
1.1.3	At least had been used by several institutions	I Community	General	must have	1	Used by FORS and SEEDS partners for RRRP project	NA	1	Dataverse is used by several large US Universities and in Europe DANS is using it as a part of their EASY system.	2	There is a strong user community worldwide. Many institutions use DSpace as their institutional repository	2	There is a plenty of institutions that are using EPrints. The user community is especially strong in the UK
1.1.5	Capability to be implemented / customized as archiving tool?	I Community	General	must have	2	It is an archiving tool	none	2	It is an archiving tool	2	It is an archiving tool	2	It is an archiving tool
1.2.1	Documentation of the software well written and maintained.	I Community	Software-Development	must have	1	There is documentation, but not very well maintained	none	2	Documentation is well maintained	2	Documentation is well maintained	2	Documentation not well maintained and there are occasional problems with the access to their wiki pages
1.2.2	Codes are maintained using version control system.	I Community	Software-Development	must have	2	Git is used as version control system	none	2	GIT	2	GIT	2	GIT
1.2.3	The dependency with other software components is tidily maintained.	I Community	Software-Development	must have	2	SEEDSbase is linked with FedoraCommons and the dependency is tidily maintained	none	2	Dataverse depends on many opensource components and dependance	2	Dataverse depends on many opensource components and dependanci	1	Dataverse depends on many opensource components but dependance

									s are well maintained			es are well maintained		s are not very well maintained	
1.2.4	If modules/features not open source, at least open APIs, and common programming language is used.	I Community	Software-Development	must have		2	Everything will be opensource in a near future. Programming language Python and framework Django were used for the development.	NA	2	Everything is opensource and the licences guarantee that this will persist in a future		2	Everything is opensource and the licences guarantee that this will persist in a future	2	Everything is opensource and the licences guarantee that this will persist in a future
1.2.5	Possibility to buy Plugins (Add-ons) for features not included in the open source version	I Community	Software-Development	nice to have		0	SEEDSbase is implemented in a way to easily integrate any additional functionality/module. There is no plan to sell specific plugins in a near future.	high	1	One cannot buy already developed plugins but it's possible to hire a developer or to contribute own solution		1	There is addons directory available and it's possible to contribute by developng or funding new features	1	EPrints has a plugin repository called Bazaar and it's possible to install plugins directly within the administrative web interface. Also, it's pretty easy to develop new plugins.
1.2.6	Internal IT Know-how	I Community	Software-Development	nice to have		0	The know-how is currently only at FORS	high	0	No, it's Java		0	No, it's Java	2	Yes, Perl
2.1.0	Multi-language support	II Functionality	General	must have		2	User interface is in 3 languages (EN, FR and DE) as for studies and datasets there can be used as many as needed languages. There will be some additional development effort in case you want to translate user interface in some other language.	low	0	It doesn't have multilang support and it's not on the official roadmap but we have some internal info that it's planned to start working on it.		2	Yes	2	Yes, metadata and interface could be internationalized

2.1.1	Ingest process	II Functionality	Ingest	must have		2	Fully covered	none	2	It depends on the way of using the tool. Submission is fully covered but one cannot save the state of SIP. Dataset can be either in draft/unpublished or published. Once published older states are lost but certain files could left as restricted for historical or other reasons.		1	SIP cannot be saved in it's initial version		1	SIP cannot be saved in initial version
2.1.2	Virus check	II Functionality	Ingest	must have		2	Fully covered (antivirus check is currently disabled in SEEDSbase instance, but the functionality is in place)	none	1	Not by default		1	Not by default		1	Not by default
2.1.3	Check sum	II Functionality	Ingest			2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.1.4	File type recognition	II Functionality	Ingest	must have		2	There is a list of desired filetypes that can be maintained by the archive. The system checks the uploaded filetype against this list, and if it's in a format that is not in the list the file will not be uploaded. In this way archive is making sure to receive right file formats as SIP.	none	2	MIME type recognition, tabular filetypes are converted on the fly, SPSS files are converted also and XML DDI is generated		2	MIME		2	MIME
2.1.5	Generate SIP	II Functionality	Ingest	must have		2	Once the submission process has	none	1	Unpublished draft could be		0	No		0	No

						started, the system generates SIP and freeze this version, so it can always be checked what was submitted at the first place and what were original documents.	none		considered SIP but once published it becomes AIP						
2.1.6	Generate Descriptive Information/Metadata	II Functionality	Ingest	must have		2	Fully covered. DDI compliant and in alignment with CEESDA recommended fields.	none	2	Yes, and DDI XML could be created from SPSS files	1	No, only what's filled in		1	No, only what's filled in
2.1.7	Coordinate Updates	II Functionality	Ingest	nice to have		2	All the updates made by the producer or archive are fully coordinated using status and versioning systems.	none	2	Actions are logged and files are versioned	2	Actions are logged and files are versioned		2	Actions are logged and files are versioned
2.2.1	Receive Data	II Functionality	Archival Storage	must have		2	Fully covered.	none	2	Covered	2	Covered		2	Covered
2.2.2	Long term preservation	II Functionality	Archival Storage	must have		2	All the data are stored in Fedora.	none	2	Preservation procedures are implemented	2	Presrvation support is in place		1	There is a Preservation Toolkit available as plugin
2.2.3	Generate AIP	II Functionality	Archival Storage	nice to have		2	AIP is stored in Fedora	none	2	Publishing means creating AIP and DIP	2	Publishing means creating AIP and DIP		2	Publishing means creating AIP and DIP
2.2.4	Replace media / Convert file types	II Functionality	Archival Storage	nice to have		1	Currently the process of format migration is done manually. The people from the archive wanted to have better control here. No plans in near future to implement this as automatical conversion.	high	1	Automatic migration is not implemented but conversion is covered in preservation part	1	Automatic migration is not implemented but conversion is covered in preservation part		1	Preservation toolkit identifies archived formats and lists possible items that could be converted
2.2.5	Error Checking	II Functionality	Archival Storage	nice to have		2	Fully covered	none	2	Covered	2	Covered		2	Covered

2.2.6	Disaster Recovery	II Functionality	Archival Storage	nice to have		1	This is not done by the tool itself. It's rather on a server hosting side. How the server is maintained, how the data are backedup, etc.	NA	1	This is the part of backup/restore procedures not fully covered by the tool itself		1	This is the part of backup/restore procedures not fully covered by the tool itself		1	This is the part of backup/restore procedures not fully covered by the tool itself
2.3.1	Deposit contract	II Functionality	Administration	must have		2	SEEDSbase supports even different deposit contracts for different datasets.	none	2	It supports general terms of use and licences as well as special cases that can be applied to a collection or to individual dataset		2	General terms and licences, it's possible to add specific conditions		2	General terms and licences but it's also possible to add specific conditions
2.3.2	Database of Users	II Functionality	Administration	must have		2	Users, persons and institutions are captured	none	2	Yes		2	Yes		2	Yes
2.3.3	User management	II Functionality	Administration	must have		2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.3.4	Communication with users	II Functionality	Administration	nice to have		2	Fully covered	none	0	No		0	No		0	No
2.4.1	Download contract	II Functionality	Access	must have		2	SEEDSbase supports even different download contracts for different datasets.	none	2	Various possibilities		1	Similar to 2.3.1		1	Smilar to 2.3.1
2.4.2	Level of access	II Functionality	Access	must have		2	Fully covered	none	2	Fine grained system of permissions		2	User, Editor, SuperAdmin		2	User, Editor, SupeAdmin
2.4.3	Generate DIP	II Functionality	Access	must have		2	AIP is stored in Fedora	none	1	See 2.1.5		1	See 2.1.5		1	See 2.1.5
2.4.4	Data catalogue	II Functionality	Access	must have		2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.5	Simple and Advanced search facilities	II Functionality	Access	must have		2	Fully covered	none	2	Covered		2	Covered		2	Covered

2.4.6	Browse metadata, data and associated documentation	II Functionality	Access	must have		2	Fully covered	none	2	Covered		2	Covered		2	Covered, but no faceted search
2.4.7	Download datasets in a variety of statistical formats	II Functionality	Access	nice to have		2	It can be decided what is included into DIP, this means we can include different file formats for the same dataset.	none	2	Different file formats are supported and some automatic conversions are implemented		1	Different file formats are supported		1	Different file formats are supported
2.5.1	Login, User Roles and Permissions	II Functionality	General	must have		2	There are 4 types of Roles: 1. researcher basic (can just download the data), 2. researcher full (can work on a study/dataset, and here we have 3 types of users: 1. can edit 2. can manage 3. can submit), 3. archive collaborator and 4. admin (all the rights + some system rights)	none	2	Fine grained user control on the collection level as well as on the dataset and individual file level		2	Simple access controls		2	Simple access controls
2.5.2	Users are able to edit their own profiles	II Functionality	General	must have		2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.5.3	Users communicate with each other via platform internal messages	II Functionality	General	nice to have		2	They can communicate via internal messaging system and/or via email using the system. Also user can set if he wants to be notified by email every time he receives a message	none	0	No		0	No		0	No



2.5.4	The administrator of the platform is able to send E-Mails to the users	II Functionality	General	nice to have		2	He can send emails and/or private messages	none	0	No		0	No		0	No
3.1.1	Compatibility with operating systems	III Technical Basis	Platform	nice to have		2	SEEDSbase is fully compatible with Linux. It should work on any OS since it works on Apache and it's using Python.	none	1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported		1	Linux by default - other OSs possible but not supported
3.2.1	Scalability e.g. with many users simultaneously	III Technical Basis	Performance	must have		2	We have currently 3300 active users in FORSbase and no performance issues reported. (If you noticed some performance issues in SEEDSbase it might be due to the power of the server, since the server we are using for SEEDSbase is less powerful).	none	2	It can be scaled by adding application and database servers		2	It works in different environments		2	It works in different environments
3.2.2	Usable through slow internet connections	III Technical Basis	Performance	nice to have		2	Fully covered	none	2	Covered		2	Covered		2	Covered
3.2.3	User-friendly, clear and intuitive structure?	III Technical Basis	Performance	must have		1	It can be more userfriendly. It is planned to do the whole user interface redesign in 2018.	high	2	It could be optimized	high	2	It could be optimized	high	2	It could be optimized
3.3.1	How and where the information & data stored?	III Technical Basis	Security	must have		2	In FedoraCommons	none	2	Local database and filesystem		2	Local database and filesystem		2	Local database and filesystem
3.3.2	Is the connection between server and portal (and client) secure?	III Technical Basis	Security	must have		2	Fully covered	none	2	Covered, depends on web server		2	Covered, depends on web server		2	Covered, depends on web server

## 5.5 Montenegro

Req. #	Requirement description	Section	Category	Criteria		SEEDSBase			Dataverse			Dspace			Eprints		
				Priority	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment	Development effort	Score	Comment	Development effort
1.1.1	Open Source License. Everyone is able to download and install it.	I Community	General	must have	NA	0	not still, possibly in 2 years	NA	2	Open Source Software. Written in Java		2	Open Source Software, written in Java. Developed at first at MIT and HP labs		2	Open Source Software, created in 2000, University of Southampton. Written in Perl	
1.1.2.1	Annual costs for license*	I Community	General	nic to have	NA	1		NA	2	free of charge		2	free of charge		2	free of charge	
1.1.2.2	Annual costs for support*	I Community	General	nic to have	NA	1		NA	2	free of charge		2	free of charge		2	free of charge	
1.1.3	At least had been used by several institutions	I Community	General	must have	NA	2			2	it is used by 2,023 teams, there are 22 institutions using this database		2	more than 500 organizations are using Dspace		2	UK based mainly	
1.1.5	Capability to be implemented / customized as archiving tool?	I Community	General	must have	NA	2			2	it is already used as an archiving tool by many institutions and teams		2	stated to be highly customized		2	it is archiving tool	
1.2.1	Documentation of the software well written and maintained	I Community	Software-Development	must have		2			2	it well maintained		2	very well		2	there is a dedicated team of experts	
1.2.2	Codes are maintained using version	I Community	Software-Development	must have		2			2	<a href="https://github.com/IQSS/dataverse">https://github.com/IQSS/dataverse</a>		2	<a href="https://github.com/DSpace/DSpace">https://github.com/DSpace/DSpace</a>		2	<a href="https://github.com/eprints/eprints">https://github.com/eprints/eprints</a>	

	control system.															
1.2.3	The dependency with other software components is tidily maintained.	I Community	Software-Development	must have	2		2	it well maintained		2	very well		2	plus there is a service of "building repositories that are configured to meet the particular requirements of organization. They work with clients to create repositories for research publications, open education resources, multimedia outputs and research data sets."		
1.2.4	If modules/features not open source, at least open APIs, and common programming language is used.	I Community	Software-Development	must have	2		2	it is open source, and within its strategy Dataverse has proclaimed promotion of open sources		2			2	it is open source		
1.2.5	Possibility to buy Plugins (Add-ons) for features not included in the open source version	I Community	Software-Development	niceto have	0		high	1	no Add-ons	high	2	<a href="https://wiki.duraspace.org/display/DSPACE/Plugin+Plugins">https://wiki.duraspace.org/display/DSPACE/Plugin+Plugins</a>	1	Version 3 of the software introduced a (Perl-based) plugin architecture for importing and exporting data, converting objects (for search engine indexing) and user interface widgets.		
1.2.6	Internal IT Know-how	I Community	Software-Development	niceto have	1		medium	1	it is always possible to hire additional consultant if needed (in case the question is on CeMI's know how)	medium	1	it is always possible to hire additional consultant if needed (in case the question is on CeMI's know how)	medium	2	it is always possible to hire additional consultant if needed (in case the question is on CeMI's know how)	
2.1.0	Multi-language support	II Functionality	General	niceto have	2			0	no	medium to high	2	<a href="http://dspace.org/sites/dspace.org/files/archive/1_5Documentation/configure.html#xmlui-multilingual">http://dspace.org/sites/dspace.org/files/archive/1_5Documentation/configure.html#xmlui-multilingual</a>	2			

2.1.1	Ingest process	II Functionality	Ingest	must have	2		2	well implemented, described and easy to follow	2	<a href="http://dspace.org/sites/dspace.org/files/archive/1_5_2Documentation/ch02.html#N1036E">http://dspace.org/sites/dspace.org/files/archive/1_5_2Documentation/ch02.html#N1036E</a>	2			
2.1.2	Virus check	II Functionality	Ingest	must have	2		1	it is not clear from demo version	medium to high	2	<a href="https://wiki.duraspace.org/display/DSPACE/Virus+Scan+Curation+Task">https://wiki.duraspace.org/display/DSPACE/Virus+Scan+Curation+Task</a>	1	<a href="http://www.eprints.org/tech.php/21315.html">confronting information, not clear... http://www.eprints.org/tech.php/21315.html</a>	
2.1.3	Check sum	II Functionality	Ingest	must have	2		2	provided		2	Dspace generates Checksums for all content files when they are ingested into the system. It also comes with a "Checksum Checker" which can be used to re-verify checksums at a later period in time.	2		
2.1.4	File type recognition	II Functionality	Ingest	must have	2		2	provided		2	<a href="http://www.darc.ntu.edu.tw/help/formats.html">http://www.darc.ntu.edu.tw/help/formats.html</a>	provided	2	provided
2.1.5	Generate SIP	II Functionality	Ingest		2		2			2	It is possible to choose to store a copy of the initial SIP	2		
2.1.6	Generate Descriptive Information/Metadata	II Functionality	Ingest	must have	2		2	automatically generated		2	Dspace stores all metadata internally in the database (in a format of slightly customized qualified Dublin core).	2		
2.1.7	Coordinate Updates	II Functionality	Ingest	niceto have	2		2	GIT		2	GIT	2	GIT	
2.2.1	Receive Data	II Functionality	Archival Storage	must have	2		2			2		2		
2.2.2	Long term preservation	II Functionality	Archival Storage	must have	2		2	Harvard originated program. There is a policy on long term preservation		2	guaranteed	2	<a href="http://preservation.eprints.org/">http://preservation.eprints.org/</a>	
2.2.3	Generate AIP	II Functionality	Archival Storage	niceto have	2		2			2	AIPs can be exported. AIP can represent either Items, Collections or Communities. However, these AIPs are external to the system.	2		
2.2.4	Replace media / Convert file types	II Functionality	Archival Storage	niceto have	1		high	2	provided possibility	2	yes	2		

2.2.5	Error Checking	II Functionality	Archival Storage	niceto have	2		1	automatically, only for ingest.	medium to high	2	yes		2	
2.2.6	Disaster Recovery	II Functionality	Archival Storage	must have	2		2			2	yes		1	not clear
2.3.1	Deposit contract	II Functionality	Administration	must have	2		2			2	yes		2	
2.3.2	Database of Users	II Functionality	Administration	must have	2		2			2	yes		2	
2.3.3	User management	II Functionality	Administration	must have	2		2			2	yes		2	
2.3.4	Communication with users	II Functionality	Administration	niceto have	2		0		high	0		high	0	
2.4.1	Download contract	II Functionality	Access	must have	2		2			1		high	1	
2.4.2	Level of access	II Functionality	Access	must have	2		2	different levels		2	different levels		2	different levels
2.4.3	Generate DIP	II Functionality	Access	must have	2		1		medium to high	2	The DIP, or Dissemination Information Package, contains an item as it is exported or disseminated from the DSpace archive.		2	<a href="http://www.eprints.org/tech.php/thread-14292.html">http://www.eprints.org/tech.php/thread-14292.html</a>
2.4.4	Data catalogue	II Functionality	Access	must have	2		2			2			2	
2.4.5	Simple and Advanced search facilities	II Functionality	Access	must have	2		2	intuitive and nicely done		2			2	
2.4.6	Browse metadata, data and associated documentation	II Functionality	Access	must have	2		2	intuitive and nicely done		2			2	

2.4.7	Download datasets in a variety of statistical formats	II Functionality	Access	nicely to have	2		2	intuitive and nicely done	2		2		
2.5.1	Login, User Roles and Permissions	II Functionality	General	must have	2		2	intuitive and nicely done	2		2		
2.5.2	Users are able to edit their own profiles	II Functionality	General	must have	2		2	yes	2		2		
2.5.3	Users communicate with each other via platform internal messages	II Functionality	General	nicely to have	2		0	no option high	0		high	0	high
2.5.4	The administrator of the platform is able to send E-Mails to the users	II Functionality	General	nicely to have	2		0	no option high	0		high	0	high
3.1.1	Compatibility with operating systems	III Technical Basis	Platform	nicely to have	2		2		2		2		
3.2.1	Scalability e.g. with many users simultaneously	III Technical Basis	Performance	must have	2		2		2	nicely done	2		
3.2.2	Usable through slow internet connections	III Technical Basis	Performance	must have	2		2		2		2		
3.2.3	User-friendly, clear and intuitive structure?	III Technical Basis	Performance	must have	1	NA	2	very much	2		2		
3.3.1	How and where the information & data stored?	III Technical Basis	Security	must have	2		2		2		2		

3.3.2	Is the connection between server and portal (and client) secure?	III Technical Basis	Security	must have		2			2	yes			2				
-------	--	---------------------	----------	-----------	--	---	--	--	---	-----	--	--	---	--	--	--	--

## 5.6 Serbia

Req. #	Requirement description	Section	Category	Priority	Development effort	SEEDSBase		Development effort	Dataverse		Development effort	DSpace		Development effort	EPrints	
						Score	Comment		Score	Comment		Score	Comment		Score	Comment
1.1.1	Open Source License. Everyone is able to download and install it.	I Community	General	must have		0	FORS will open the source code in the near future, most probably 2018	NA	2			2			2	
1.1.2.1	Annual costs for licence*	I Community	General	nic e to have		0	Currently, no annual costs	NA	0			0			0	
1.1.2.2	Annual costs for support*	I Community	General	nic e to have		0	Currently there are no annual costs, but once the tool is opensource, there might be some costs for technical support	NA	2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced		2	There are operating system costs and application maintenance costs that has to be planned. If the institution doesn't have human resources available for that purpose, this work has to be outsourced

1.1.3	At least had been used by several institutions	I Community	General	must have		1	Used by FORS and SEEDS partners for RRPP project	NA	1	Dataverse is used by several large US Universities and in Europe DANS is using it as a part of their EASY system.		2	There is a strong user community worldwide. Many institutions use DSpace as their institutional repository		2	There is a plenty of institutions that are using EPrints. The user community is especially strong in the UK
1.1.5	Capability to be implemented / customized as archiving tool?	I Community	General	must have		2	It is an archiving tool	none	2	It is an archiving tool		2	It is an archiving tool		2	It is an archiving tool
1.2.1	Documentation of the software well written and maintained.	I Community	Software-Development	must have		1	There is documentation, but not very well maintained	none	2	Documentation is well maintained		2	Documentation is well maintained		2	Documentation not well maintained and there are occasional problems with the access to their wiki pages
1.2.2	Codes are maintained using version control system.	I Community	Software-Development	nice to have		2	Git is used as version control system	none	2	GIT		2	GIT		2	GIT
1.2.3	The dependency with other software components is tidily maintained.	I Community	Software-Development	must have		2	SEEDSbase is linked with FedoraCommons and the dependency is tidily maintained	none	2	Dataverse depends on many opensource components and dependencies are well maintained		2	dSpace depends on many opensource components and dependencies are well maintained		2	Eprints depends on many opensource components but dependencies are not very well maintained
1.2.4	If modules/features not open source, at least open APIs, and common programming language is used.	I Community	Software-Development	nice to have		2	Everything will be opensource in a near future. Programming language Python and framework Django were used for the development.	NA	2	Everything is opensource and the licences guarantee that this will persist in a future		2	Everything is opensource and the licences guarantee that this will persist in a future		2	Everything is opensource and the licences guarantee that this will persist in a future
1.2.5	Possibility to buy Plugins (Add-ons) for features not	I Community	Software-Development	nice to have		0	SEEDSbase is implemented	high	1	One cannot buy already developed		1	There is addons directory		1	EPrints has a plugin repository



	included in the open source version			have							plugins but it's possible to hire a developer or to contribute own solution			available and it's possible to contribute by developng or funding new features			called Bazaar and it's possible to install plugins directly within the administrative web interface. Also, it's pretty easy to develop new plugins.
1.2.6	Internal IT Know-how	I Community	Software-Development	nic to have		0	The know-how is currently only at FORS	high	0	No, it's Java		0	No, it's Java			1	Yes, Perl
2.1.0	Multi-language support	II Functionality	General	must have		2	User interface is in 3 languages (EN, FR and DE) as for studies and datasets there can be used as many as needed languages. There will be some additional development effort in case you want to translate user interface in some other language.	low	0	It doesn't have multilang support and it's not on the official roadmap but we have some internal info that it's planned to start working on it.		2	Yes			2	Yes, metadata and interface could be internationalized
2.1.1	Ingest process	II Functionality	Ingest	must have		2	Fully covered	none	2	It depends on the way of using the tool. Submission is fully covered but one cannot save the state of SIP. Dataset		1	SIP cannot be saved in it's initial version			1	SIP cannot be saved in initial version

									can be either in draft/unpublished or published. Once published older states are lost but certain files could left as restricted for historical or other reasons.							
2.1.2	Virus check	II Functionality	Ingest	nic e to hav e		2	Fully covered (antivirus check is currently disabled in SEEDSbase instance, but the functionality is in place)	none	1	Not by default		1	Not by default		1	Not by default
2.1.3	Check sum	II Functionality	Ingest	mu st hav e		2	Fully covered	none	2	Yes		1	Yes		1	Yes
2.1.4	File type recognition	II Functionality	Ingest	nic e to hav e		2	There is a list of desired filetypes that can be maintained by the archive. The system checks the uploaded filetype against this list, and if it's in a format that is not in the list the file will not be uploaded. In this way archive is making sure to	none	2	MIME type recognition, tabular filetypes are converted on the fly, SPSS files are converted also and XML DDI is generated		2	MIME		2	MIME

						receive right file formats as SIP.										
2.1.5	Generate SIP	II Functionality	Ingest	must have		2	Once the submission process has started, the system generates SIP and freeze this version, so it can always be checked what was submitted at the first place and what were original documents.	none	2	Unpublished draft could be considered SIP but once published it becomes AIP		0	No		0	No
2.1.6	Generate Descriptive Information/Meta data	II Functionality	Ingest	must have		2	Fully covered. DDI compliant and in alignment with CESSDA recommended fields.	none	2	Yes, and DDI XML could be created from SPSS files		2	No, only what's filled in		2	No, only what's filled in
2.1.7	Coordinate Updates	II Functionality	Ingest	nic to have		2	All the updates made by the producer or archive are fully coordinated using status and versioning systems.	none	2	Actions are logged and files are versioned		2	Actions are logged and files are versioned		2	Actions are logged and files are versioned

2.2.1	Receive Data	II Functionality	Archival Storage	must have		2	Fully covered.	none	2	Covered		2	Covered		2	Covered
2.2.2	Long term preservation	II Functionality	Archival Storage	must have		2	All the data are stored in Fedora.	none	2	Preservation procedures are implemented		1	Presrvation support is in place		1	There is a Preservation Toolkit available as plugin
2.2.3	Generate AIP	II Functionality	Archival Storage	nic e to have		2	AIP is stored in Fedora	none	1	Publishing means creating AIP and DIP		0	Publishing means creating AIP and DIP		0	Publishing means creating AIP and DIP
2.2.4	Replace media / Convert file types	II Functionality	Archival Storage	nic e to have		1	Currently the process of format migration is done manually. The people from the archive wanted to have better control here. No plans in near future to implement this as automatic conversion.	high	1	Automatic migration is not implemented but conversion is covered in preservation part		1	Automatic migration is not implemented but conversion is covered in preservation part		1	Preservation toolkit identifies archived formats and lists possible items that could be converted
2.2.5	Error Checking	II Functionality	Archival Storage	nic e to have		2	Fully covered	none	1	Covered		1	Covered		1	Covered

2.2.6	Disaster Recovery	II Functionality	Archival Storage	mu st hav e		1	This is not done by the tool itself. It's rather on a server hosting side. How the server is maintained, how the data are backed up, etc.	NA	1	This is the part of backup/restore procedures not fully covered by the tool itself		1	This is the part of backup/restore procedures not fully covered by the tool itself		1	This is the part of backup/restore procedures not fully covered by the tool itself
2.3.1	Deposit contract	II Functionality	Administrati on	mu st hav e		2	SEEDSbase supports even different deposit contracts for different datasets.	none	2	It supports general terms of use and licences as well as special cases that can be applied to a collection or to individual dataset		2	General terms and licences, it's possible to add specific conditions		2	General terms and licences but it's also possible to add specific conditions
2.3.2	Database of Users	II Functionality	Administrati on	mu st hav e		2	Users, persons and institutions are captured	none	2	Yes		2	Yes		2	Yes
2.3.3	User management	II Functionality	Administrati on	mu st hav e		2	Fully covered	none	2	Yes		2	Yes		2	Yes
2.3.4	Communication with users	II Functionality	Administrati on	nic e to hav e		2	Fully covered	none	0	No		0	No		0	No
2.4.1	Download contract	II Functionality	Access	nic e to hav e		2	SEEDSbase supports even different download contracts for different datasets.	none	2	Various possibilities		0	Similar to 2.3.1		0	Smilar to 2.3.1
2.4.2	Level of access	II Functionality	Access	mu st hav e		2	Fully covered	none	2	Fine grained system of permissions		2	User, Editor, SuperAdmin		2	User, Editor, SupeAdmin

2.4.3	Generate DIP	II Functionality	Access	mu st hav e		2	AIP is stored in Fedora	none	1	See 2.1.5		1	See 2.1.5		1	See 2.1.5
2.4.4	Data catalogue	II Functionality	Access	mu st hav e		2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.5	Simple and Advanced search facilities	II Functionality	Access	mu st hav e		2	Fully covered	none	2	Covered		2	Covered		2	Covered
2.4.6	Browse metadata, data and associated documentation	II Functionality	Access	mu st hav e		2	Fully covered	none	2	Covered		2	Covered		2	Covered, but no faceted search
2.4.7	Download datasets in a variety of statistical formats	II Functionality	Access	nic e to hav e		2	It can be decided what is included into DIP, this means we can include different file formats for the same dataset.	none	0	Different file formats are supported and some automatic conversions are implemented		0	Different file formats are supported		0	Different file formats are supported

2.5.1	Login, User Roles and Permissions	II Functionality	General	must have	2	There are 4 types of Roles: 1. researcher basic (can just download the data), 2. researcher full (can work on a study/dataset, and here we have 3 types of users: 1. can edit 2. can manage 3. can submit), 3. archive collaborator and 4. admin (all the rights + some system rights)	none	2	Fine grained user control on the collection level as well as on the dataset and individual file level	2	Simple access controls	2	Simple access controls
2.5.2	Users are able to edit their own profiles	II Functionality	General	nic e to have	2	Fully covered	none	2	Covered	2	Covered	2	Covered
2.5.3	Users communicate with each other via platform internal messages	II Functionality	General	nic e to have	2	They can communicate via internal messaging system and/or via email using the system. Also user can set if he wants to be notified by email every time he receives a message	none	0	No	0	No	0	No

2.5.4	The administrator of the platform is able to send E-Mails to the users	II Functionality	General	nic e to hav e	2	He can send emails and/or private messages	none	0	No	0	No	0	No
3.1.1	Compatibility with operating systems	III Technical Basis	Platform	nic e to hav e	2	SEEDSbase is fully compatible with Linux. It should work on any OS since it works on Apache and it's using Python.	none	1	Linux by default - other OSs possible but not supported	1	Linux by default - other OSs possible but not supported	1	Linux by default - other OSs possible but not supported
3.2.1	Scalability e.g. with many users simultaneously	III Technical Basis	Performance	mu st hav e	2	We have currently 3300 active users in FORSbase and no performance issues reported. (If you noticed some performance issues in SEEDSbase it might be due to the power of the server, since the server we are using for SEEDSbase is less powerful)	none	2	It can be scaled by adding application and database servers	2	It works in different environments	1	It works in different environments
3.2.2	Usable through slow internet connections	III Technical Basis	Performance	mu st hav e	2	Fully covered	none	2	Covered	2	Covered	2	Covered



3.2.3	User-friendly, clear and intuitive structure?	III Technical Basis	Performance	must have		1	It can be more userfriendly. It is planned to do the whole user interface redesign in 2018.	high	2	It could be optimized		2	It could be optimized		2	It could be optimized
3.3.1	How and where the information & data stored?	III Technical Basis	Security	nic e to have		2	In FedoraCommons	none	2	Local database and filesystem		2	Local database and filesystem		2	Local database and filesystem
3.3.2	Is the connection between server and portal (and client) secure?	III Technical Basis	Security	must have		2	Fully covered	none	2	Covered, depends on web server		2	Covered, depends on web server		2	Covered, depends on web server