Acesso Aberto Angola

Technical infrastructure requirements for the development of the National Repository of Angola

Work derived from the collaborative project among Redalyc UAEM, AmeliCA, MESCTI and U. Óscar Ribas funded by UNESCO for the development of an Open Access strategy in Angola

Project managers

Arianna Becerril-García, Eduardo Aguado-López, Alejandro Macedo-García, Eurico Wongo-Gungula

March, 2022
Vol. 28

unesco · Governo de Angola · mescti.gov.ao · UAEM · redalyc.org UAEM · AmeliCA · UÓR Universidade Óscar Ribas

Technical infrastructure requirements for the development of the National Repository of Angola. Work derived from the collaborative project among Redalyc UAEM, AmeliCA, MESCTI and U. Óscar Ribas funded by UNESCO for the development of an Open Access strategy in Angola. Vol. 28.

Project managers: Arianna Becerril-García, Eduardo Aguado-López, Alejandro Macedo-García, Eurico Wongo-Gungula.

March, 2022

Electronic distribution: License: CC-BY-NC-SA 4.0

Suggested citation: United Nations Educational, Scientific and Cultural Organization, Ministério do Ensino Superior, Ciência, Tecnologia e Inovação de Angola, Redalyc-Universidad Autónoma del Estado de México, AmeliCA & Universidade Óscar Ribas (2022). Technical infrastructure requirements for the development of the National Repository of Angola (Vol. 28). Zenodo. https://doi.org/10.5281/zenodo.6360882

This piece is the result of work carried out in 2021 as part of a collaborative project between UNESCO and the Redalyc Scientific Information System. The project had the collaboration of the Government of Angola through the Ministério do Ensino Superior, Ciência, Tecnologia e Inovação (MESCTI), the Universidad Autónoma del Estado de México (Mexico), AmeliCA and the Universidade Óscar Ribas (Angola), and its objective was to carry out a comprehensive diagnosis of scientific communication in Angola, a critical roadmap for the development of Open Access in the country and a series of technological developments in favor of non-commercial open scientific communication. The project resulted in a set of 38 works that are made public with the aim of contributing and influencing a technological, legal and knowledge infrastructure of non-commercial Open Access in Angola.

Project managers:

Arianna Becerril-García, Eduardo Aguado-López, Alejandro Macedo-García, Eurico Wongo-Gungula.

Working teams:

Mexico
Redalyc Scientific Information System
Universidad Autónoma del Estado de México

Research:
- Sheila Godínez-Larios
- Liliana González-Morales
- Marco Antonio Estrada-Medina
- Pedro Villegas-Hernández
- Brenda Uribe-Martínez

Software development:
- Alma Rosa Segundo Escobar
- Daniel Mejía Antolín
- Edgar Juárez Escamilla
- Thania del Carmen Colín Álvarez
- Jonatan E. Montes de Oca Ríos
- Jorge Juan Díaz Carbajal
- Domingo Anzaldo Bibiano

Art and design:
- Bernardo Bernal-Gómez
- Priscila Dávila-Morales
- Nayeli Lara-Martínez
- Abril Carmona Ochoa

Editorial Diagnosis:
- Lidia Abigail Villagómez Beltrán
- Ana Lilia Aladín Díaz
- Daniel Francisco Flores García

Communication:
- Pamela Amarillas Nava

Translation:
- Cristell Rubí Hernández Cruzaley
- Jessica Mireya Trujillo Zúñiga
- Martha Paulina Ibarra Quintana
- Nancy Verónica Derbéz Cruz
- Paola Andrea Carbajal García

Angola
Universidade Óscar Ribas

Research:
- Carla Santana
- Josefina Castillero
- Inês Portugal

Table of contents

I. ABOUT THE ANGOLA'S NATIONAL REPOSITORY HOSTING 5

ACQUISITION OF A SERVER 5

Virtual Server 5

Dedicated Server 7

II. ACQUISITION OF A HOSTING 10

DURASPACE 10

Mission 10

Services 10

About DSpaceDirect 11

Features 11

SUBSCRIPTION PLANS 14

DSpaceDirect Small 14

DSpaceDirect Medium 15

DSpaceDirect Large 15

DSpaceDirect Extra Large 15

III. STORAGE SPACE ESTIMATION 16

REFERENCES 16

List of figures

Figure 1 Virtual Server 5
Figure 2 Virtual Server Costs 6
Figure 3 Dedicated Server 7
Figure 4 Dedicated Linux Server Ultimate 8
Figure 5 Hosting Mexico 8
Figure 6 DS3-SSD 9
Figure 7 Sample 1 13
Figure 8 Sample 2 13

I. About the Angola's National Repository hosting

There are two possible options to host the Angola's National Repository:

Acquisition of a server

The first option was to acquire a server which focuses on allocating the national repository for Angola. An analysis of the physical resources that the server should have was made, for example: physical memory "RAM 32GB minimum), number of CPUs (8 CPUs minimum), storage on disk (500 GB were considered), speed of access to the red and the operating system.

The following proposals are made:

Virtual Server

https://nic.com/virtual-private-servers/

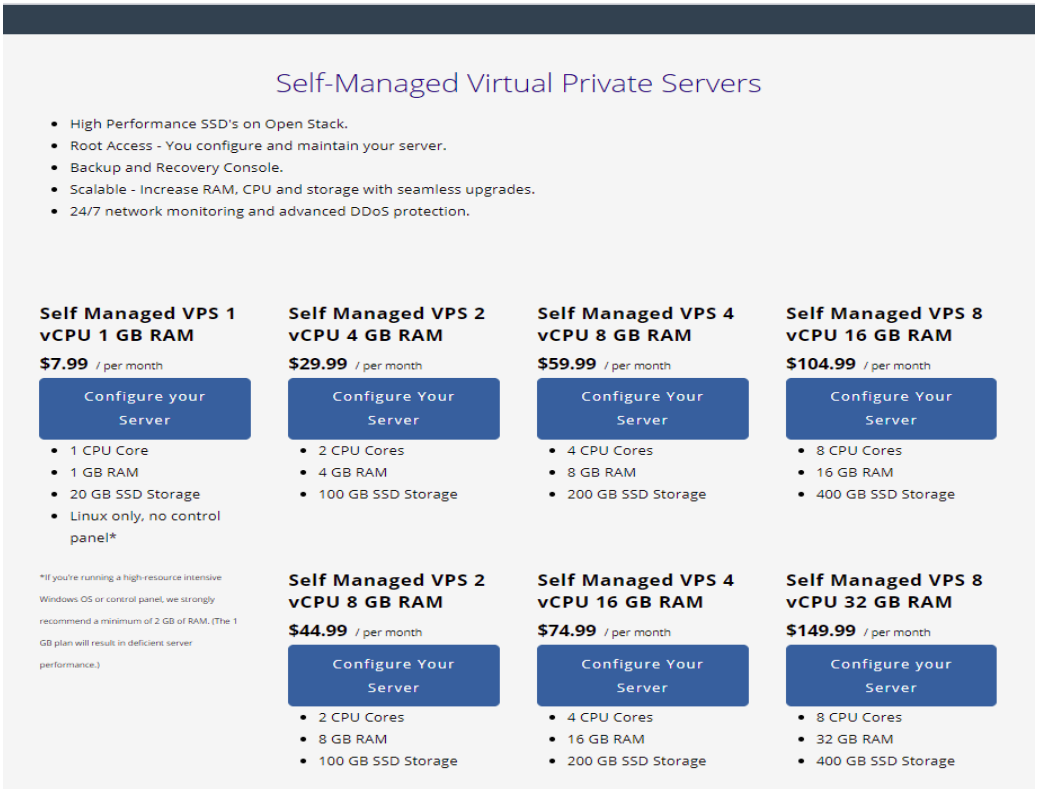

Figure 1 Virtual Server

Self Managed VPS8 VCPU 32 RAM $149.99/per month

Annual price: $1,799.88 => MX $36013.89

Operating Systems: CentOS or Ubuntu

Additional:
Add Deluxe Website Security

Block malware before it reaches your site and stop DDoS attacks with our advanced WAF. Plus, boost page load time 70% with our CDN.

Price: $19.99

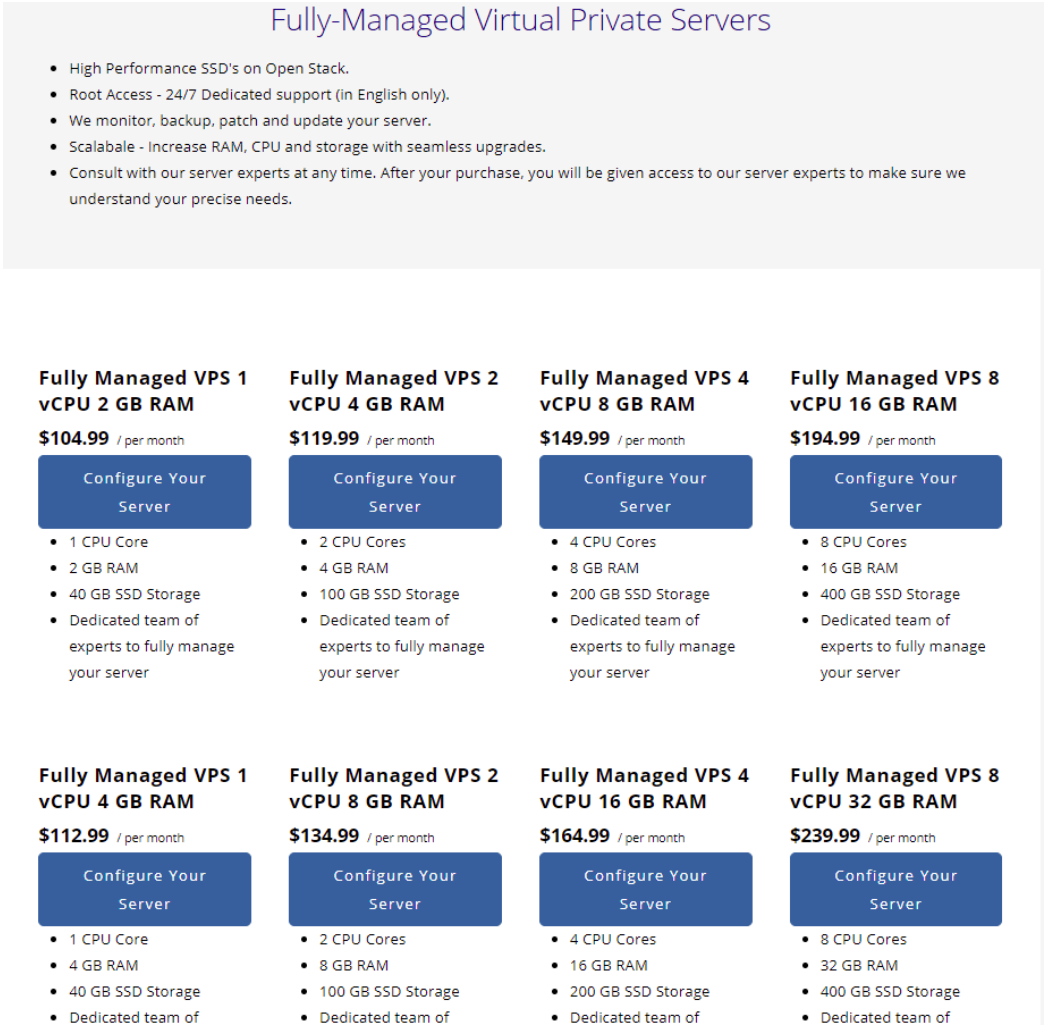

Figure 2 Virtual Server Costs

Fully Managed VPS8 VCPU 32 RAM $239.99/per month

Annual price: $2,879.88=> MX $57623.66

Operating Systems: Linux CentOS with cPanel/WHM

Additional:
Add Deluxe Website Security

Block malware before it reaches your site and stop DDoS attacks with our advanced WAF. Plus, boost page load time 70% with our CDN.

Price: $19.99

Dedicated Server

https://nic.com/dedicated-servers/

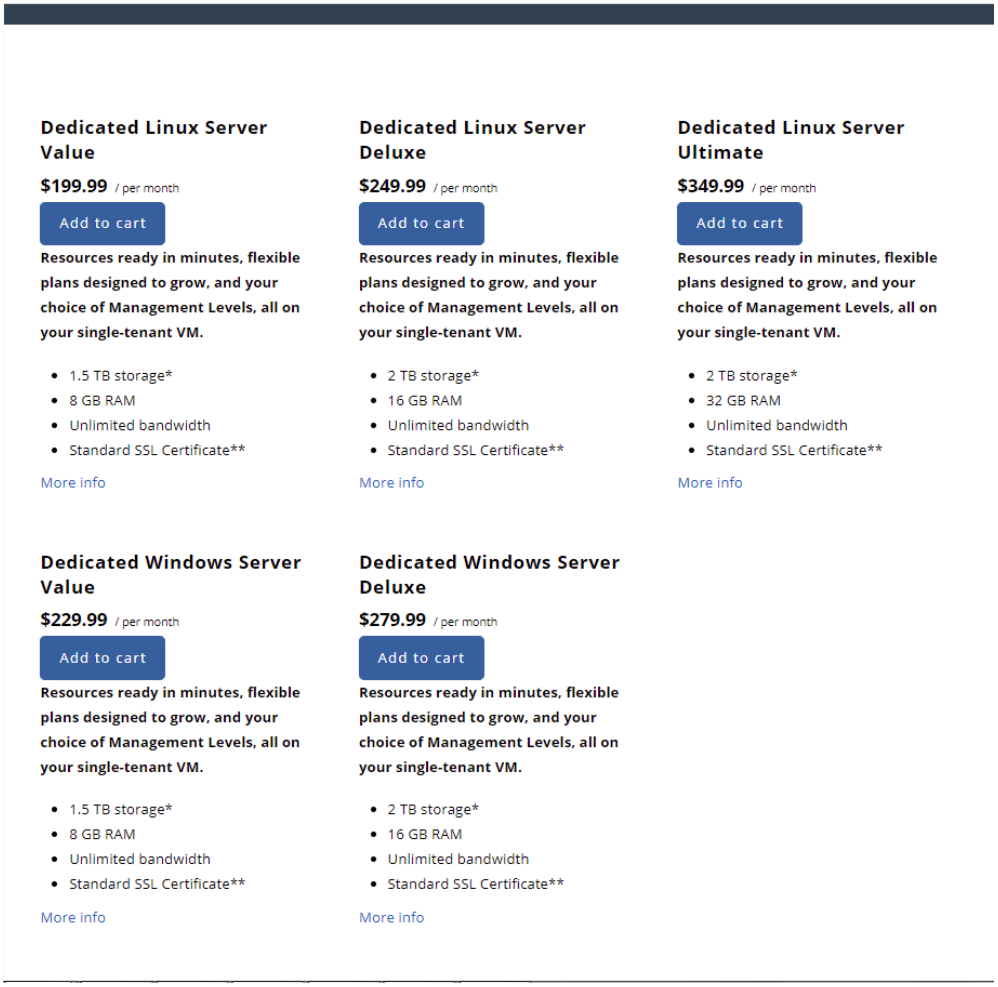

Figure 3 Dedicated Server

Dedicated Linux Server Ultimate $349.99/per month

Annual price: $5,999.88=> MX $120051.90

Operating Systems: CentOS or Ubuntu

https://www.digitalocean.com/pricing/

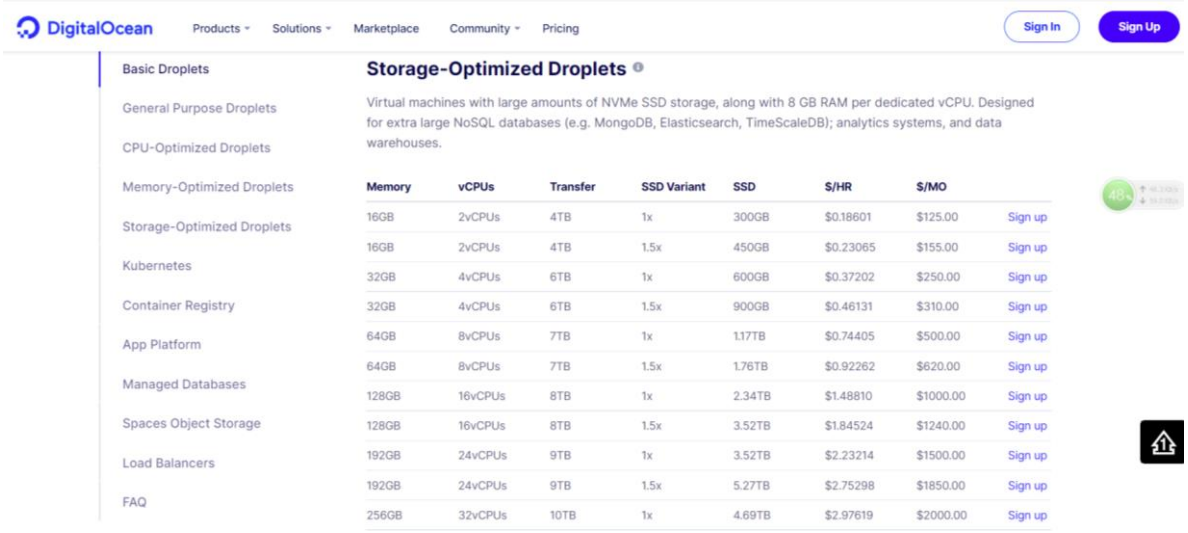

Figure 4 Dedicated Linux Server Ultimate

https://hosting-mexico.net/servidoresdedicados.htm

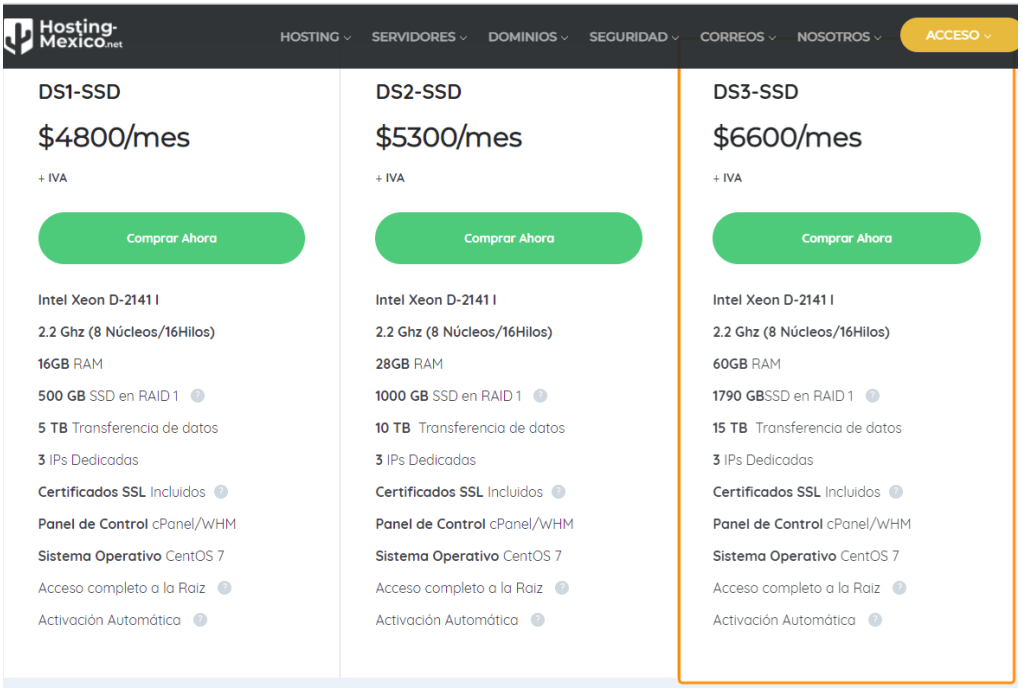

Figure 5 Hosting Mexico

DS3-SSD

Annual price: MX $6600/month=> MX $79200

Operating Systems: CentOS 7

https://mx.godaddy.com/hosting/dedicated-server#plans

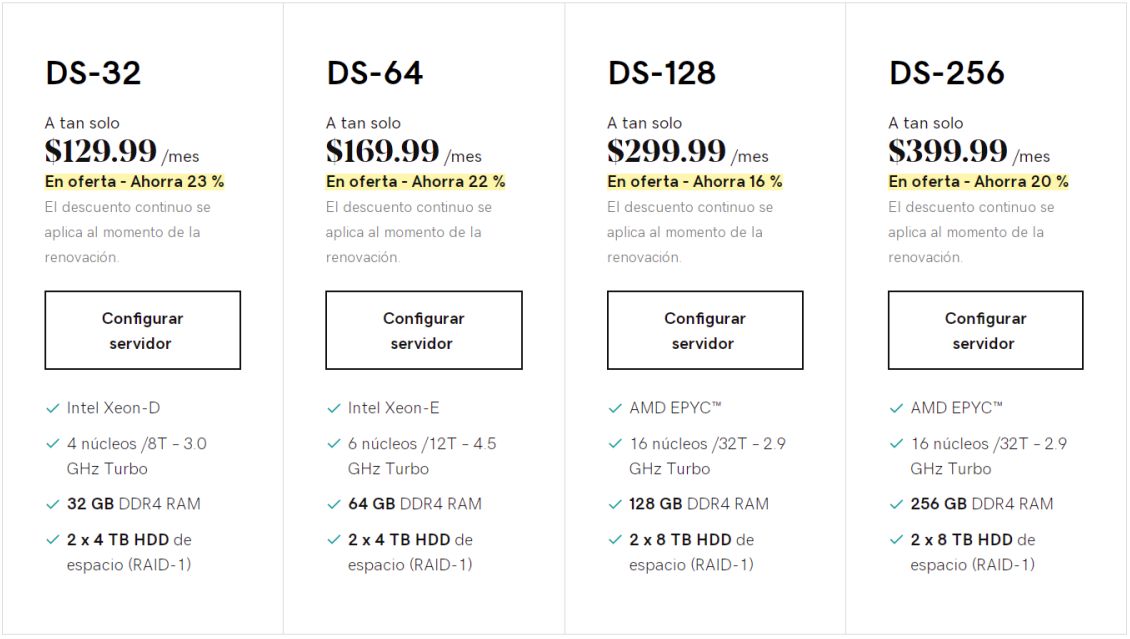

Figure 6 DS3-SSD

DS-256

Annual price: $5399.88 => MX $108046.47

Operating Systems: CentOS or Ubuntu

What this stage involved was that once the server was acquired, the following software should be installed:

· Operating system: Ubuntu Server 20.04.2 LTS (GNU/Linux 5.4.0-81-generic x86_64)

· Java: openjdk version "1.8.0_292"

· Apache Maven: version: "5.4.0-81-generic"

· Apache Ant: Apache Ant(TM) version 1.10.7

· Git: version 2.25.1

· PostgreSQL: version 10

· Apache Tomcat: version 7.0.109

- Aplicativo: DSPACE version 6.3
- Buy and configure a domain to have internet access.

To carry out all these activities, a technical personnel with knowledge on server administration is required. The first task of this person would be to install and configure all the previous tools and, mainly, DSPACE which is the necessary software to create the national repository for Angola.

Once installed, a person who administers the national repository for Angola is needed. This person would be in charge of creating users, communities, giving permissions, etc. in the repository.

II. Acquisition of a hosting

The second option is buying a repository server already installed and configured, for example: DSpaceDirect, which is a server provided by DuraSpace; which is acquired with a DSPACE in its last stable version, and nothing needs to be installed as the provider is in charge of these tasks which require a technical knowledge. In this server, the main task is to manage the repository; for example, create users, create communities, create collections, etc.

To know a little more about the provider, relevant information about this company is detailed, as well as the services, functionalities, benefits when acquiring DSpaceDirect and the plans it offers.

DURASPACE

Mission

DuraSpace is an independent, non-profit organization that provides leadership and innovation for open technologies that promote durable and persistent access to digital data. It collaborates with academic, scientific, cultural, technological and research communities by supporting projects and promoting services to help ensure that current and future generations have access to collective digital heritage.

Services

DuraSpace is a supplier of innovative solutions to meet today's access, preservation and discovery challenges. Provides services to the digital preservation community that enable access to and long-term conservation of their digital assets.

Its services include DuraCloud, an easy and cost-effective way to archive, share and manage content in the cloud. One click creates multiple copies in the cloud in different locations with multiple providers, while ensuring content health through DuraCloud's automated content health check services.

DSpaceDirect is a low-cost, turnkey repository hosting service that can be used to safeguard and provide access to faculty and student documents, projects, and research through a DSpace interface.

ArchivesDirect is a hosting service offered by DuraSpace in collaboration with Artefactual Systems to create standards-based digital preservation content packages that are archived in secure long-term storage.

About DSpaceDirect

DSpaceDirect enables institutions of all sizes to store, organize and manage their repository content in the cloud. With a subscription to DSpaceDirect, organizations have an optimized DSpace repository ready to preserve and provide access to their content.

Features

DSpaceDirect offers the most current features available in the latest stable version of the open source DSpace software, including the following:

1. Provides a fully hosted software: DSpaceDirect comes fully installed and ready to go. System administration -server configuration, nightly server snapshots, server maintenance, and general system administration best practices- is fully managed, including free upgrades to the latest stable version of DSpace. You can also easily add more storage at any time or migrate to a locally installed DSpace in the future.

2. Content organization: DSpaceDirect offers the possibility to organize documents in a hierarchical way. Files or documents can be added to one or more Collections, and Collections can be grouped into larger Communities. Each Collection and Community can have its own logo and personalized home page, allowing the structure of an organization within DSpaceDirect to be replicated. Also, you can easily batch update all stored metadata content through the simple web interface.

3. Back up and content preservation: Subscription to DSpaceDirect includes DuraCloud, a cloud storage service from DuraSpace. This means that DSpaceDirect creates an additional backup of all content, and then replicates the copy to DuraCloud, where automatic content health checks are performed periodically. It allows users to easily

recover and restore individual files or entire collections that have been accidentally deleted or altered through the DSpaceDirect web interface.

4. Search and find content: All public content on DSpaceDirect is indexed by Google and other major search engines, helping users to make information available in the repository. Content can be easily referenced as all documents, collections and communities in DSpaceDirect are assigned a permanent external identifier. Content can be easily searched, and any search can be filtered and/or sorted to refine the results. DSpaceDirect offers options to search by Collection, Title, Authors, Subjects, or Date. You can jump to a specific item or sort your search results.

5. Enable Open Access: DSpaceDirect can easily provide open access to digital collections. DSpace supports various file formats and MIME types, including PDF, Word, JPEG, MPEG, and TIFF files. DSpaceDirect allows you to create an open repository to display the research and publications of students and teachers.

6. Customizing user access and workflows: DSpaceDirect has customizable user permissions, content repository licensing, and approval workflows. DSpaceDirect makes it easy to assign users specific roles with specific permissions. You can also decide whether new submissions to specific Collections are reviewed, so that individual submissions can be approved, edited or rejected. As needed, choose to allow users to limit access to specific content.

7. Select design: DSpaceDirect customizes the interface of the repository while keeping the site streamlined for easy updates. There are two layouts (sample 1, sample 2) and there is the option to add other branding like logos, favicon and custom color schemes.

Figure 7 Sample 1

Figure 8 Sample 2

8. Analyzing StatisticsCollapse usage statistics: DSpace provides its own internal usage statistics that track page views and downloads. If you have a Google Analytics account, you can also choose to add Google Analytics tracking to your DSpaceDirect account at no additional cost.

9. Access to customer service: DSpaceDirect includes 10 free support requests per year using a customer support ticket system.

10. Benefits: DSpaceDirect provides a streamlined, hosted DSpace repository managed by the nonprofit organization. Access, manage, and preserve any content or file type in a hosted environment that users can easily search for. DSpaceDirect will preserve and provide access to faculty and student documents, projects, and research. DSpaceDirect can also be used to store and share large image, video, and audio files.

    a. Quick start: A repository quickly and cost-effectively repository. No local servers or system administrators required. Duraspaces takes care of managing the software and servers, so the client can keep their costs low and focus on their users.
    b. Features to choose from: Possibility to select the characteristics that fit the needs. The customer can choose which features of DSpace to enable based on their local needs. Such as what the repository looks like, as well as customize the license text, etc.
    c. Free updates: Duraspace performs regular software updates, so the customer always has the latest stable version of DSpace.
    d. Content preservation: Content is automatically backed up, DSpaceDirect automatically integrates with DuraCloud, ensuring long-term access to your content.
    e. Permanent access to data: DSpaceDirect gives users the content they want, when they want it. DSpaceDirect is online and accessible from the web. In this way, you can add content and provide immediate access to your users, regardless of where they are.
    f. Data property: It is possible to move the repository and content to local servers at any time, since DSpace is an open source and has a global community of more than 1,500 institutions.

Subscription Plans

The following annual subscription plans are available for the DSpaceDirect service.

DSpaceDirect Small

The DSpaceDirect Small plan is suited for small organizations at the beginning stages of implementing a repository that have limited content and anticipate slow growth for the repository.

Annual Subscription Price: $3,940

Annual Storage: 75 GB

DSpaceDirect Medium

This plan is ideal for organizations at the beginning stages of implementing a repository that have a variety of content which needs to be stored and anticipate medium growth for the repository.

Annual Subscription Price: $5,780

Annual Storage: 150 GB

DSpaceDirect Large

This plan is designed for organizations at any stage of implementing a repository that have a large amount of content which needs to be stored and anticipate substantial growth for the repository.

Annual Subscription Price: $8,670

Annual Storage: 250 GB

DSpaceDirect Extra Large

This plan is designed for organizations at any stage of implementing a repository that have an extra-large amount of content which needs to be stored and anticipate substantial growth for the repository.

Annual Subscription Price: $9,750

Annual Storage: 500 GB

The plan that was acquired for the repository for Angola is DSpaceDirect Medium for 3 years. This plan is considered to estimate the space; for 5826 types of files it is required a 27 GB storage, for 29130 types of files it is required a space of 130GB. The storage that this plan offers is 150GB and it was considered as the most convenient option as it will be the initial phase of the project and will allow us to store almost 30,000 documents in the national repository. In case more space is needed, the provider will offer the possibility to dynamically increase the space.

For now, the repository for Angola was installed and configured in a Redalyc server to implement the repository while the mentioned plan is managed.

The server has the following features:

- ✓ Operating system AIX 7.2
- ✓ RAM memory 98 GB
- ✓ 4 Cores to 4.15 GHz

- ✓ 200 GB storage

Once the software is purchased to host the final Angola repository, Redalyc will migrate the data and files stored on the Redalyc server to the service purchased with Duraspace and finally hand over all administration to Angola staff.

III. Storage space estimation

For the storage space estimation, an Excel file (Personalized storage.xlsx) was implemented, in which all possible files that can be uploaded to the repository were specified, a size/weight was added per document to each file, this information was obtained from the analysis of various institutional repositories, including the institutional repository of UAEMex.

The size/weight of each document is multiplied by the number of files that are likely to be uploaded per year, resulting in the total size that is required in Gigabytes. This multiplication is applied for each document and, in the end, the sum of all the totals is made. With this, we can estimate the Gigabytes required for the repository for Angola.

As it is possible that these data may change due to different variables, 2 tentative stages are presented in the Excel file, but if they do not fit the real data required by the Angola project, there is a third option to customize the data according to the experience that during the implementation and use of the repository can be obtained to ensure a more realistic estimation.

References

Digital Ocean. (2021). *DigitalOcean – The developer cloud*. https://www.digitalocean.com/

Duraspace. (2021). *Duraspace*. Duraspace.Org. https://duraspace.org/

Network Information Center (NIC). (2021). *Network Information Center*. Network Information Center. https://nic.com/

Acesso Aberto
Angola