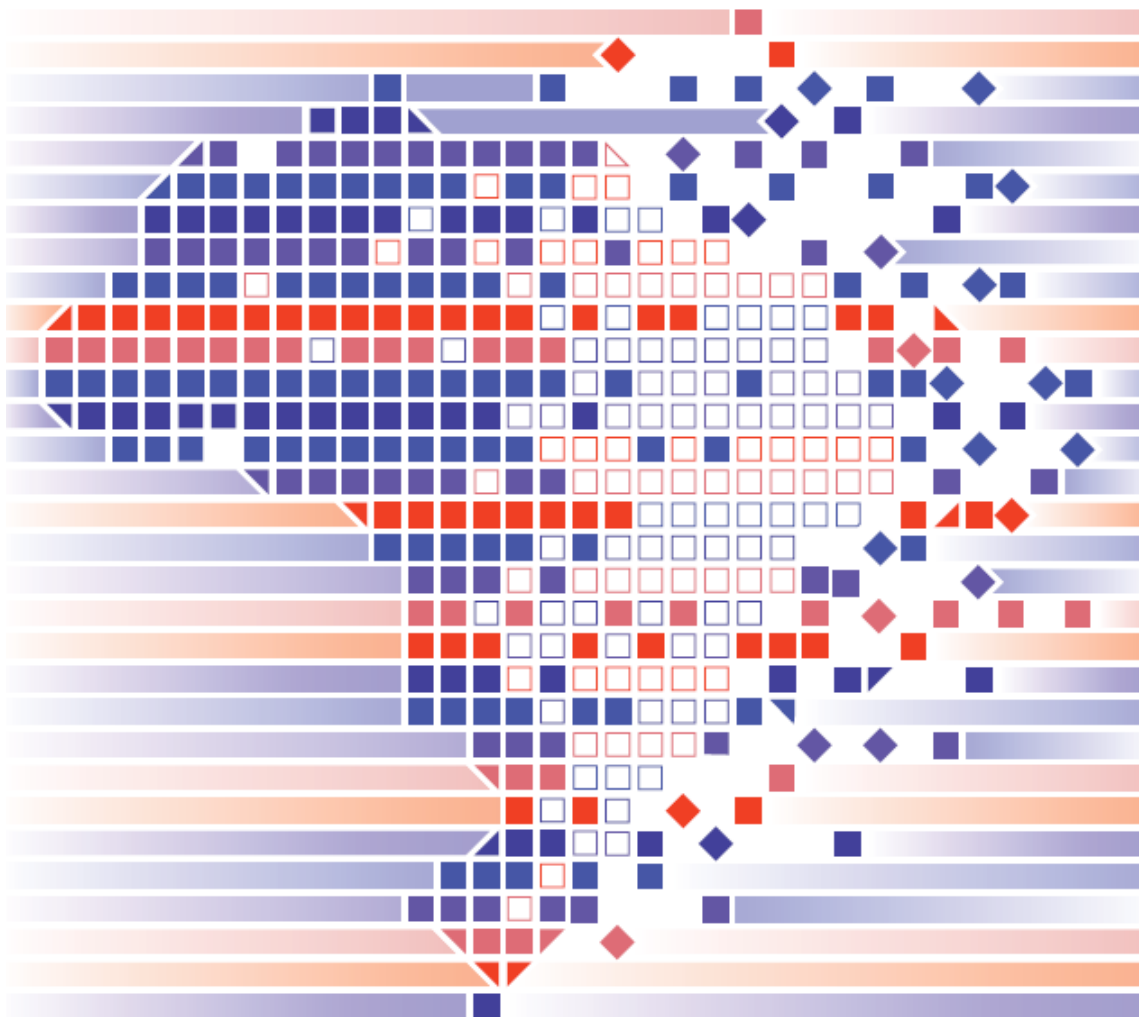


## Controle de Versionamento:

Versão	Autores	Data
1.1	<b>Universidade Federal do Rio Grande do Sul</b> <ul style="list-style-type: none"><li>● Rafael Port da Rocha</li><li>● Sônia Elisa Caregnato</li><li>● Rene Faustino Gabriel Junior</li><li>● Caterina Groposo Pavão</li><li>● Samile Andréa de Souza Vanz</li><li>● Paula Caroline Schifino Jardim Passos</li><li>● Ana Maria Mielniczuk de Moura</li></ul> <b>Universidade Federal de Rio Grande</b> <ul style="list-style-type: none"><li>● Eduardo Nunes Borges</li><li>● Adornete de Almeida Martins Junior</li></ul>	11/08/2020

# ACESSO ABERTO A DADOS DE PESQUISA NO BRASIL

Archivematica: documentação de instalação no CentOS



## **Coordenador**

Rafael Port da Rocha (Universidade Federal do Rio Grande do Sul)

## **Coordenadora adjunta**

Sônia Elisa Caregnato (Universidade Federal do Rio Grande do Sul)

## **Pesquisadores - Universidade Federal do Rio Grande do Sul**

Rene Faustino Gabriel Junior

Caterina Groposo Pavão

Samile Andréa de Souza Vanz

Paula Caroline Schifino Jardim Passos

Ana Maria Mielniczuk de Moura

## **Pesquisadores - Universidade Federal do Rio Grande**

Eduardo Nunes Borges

Adornete de Almeida Martins Júnior

## **Contato**

e-mail: [dadosdepesquisa@rnp.br](mailto:dadosdepesquisa@rnp.br) - site: <https://dadosdepesquisa.rnp.br>

Rua Ramiro Barcelos, 2777 - Campus Saúde Sala: 106 - Anexo 1 - Brasil - Porto Alegre - RS - CEP 90.035-007 - Telefone: +55(51)3308.5942



Como citar: MARTINS JUNIOR, Adornete de Almeida; BORGES, Eduardo Nunes; GABRIEL JUNIOR, Rene Faustino; ROCHA, Rafael Port da; CAREGNATO, Sônia Elisa; PAVÃO, Caterina Groposo; VANZ, Samile Andrea de Souza; PASSOS, Paula Caroline Schifino Jardim; MOURA, Ana Maria Mielniczuk de. Acesso aberto a dados de pesquisa no Brasil: Archivematica: documentação de instalação no CentOS.

## **Resumo**

Este relatório apresenta a implementação e a configuração do *Archivematica* e do *Archivematica Storage Service* para comunicação com o *Dataverse* e ingestão de dados publicados para armazenamento e preservação digital.

**Palavras-chave:** *Archivematica*, Ingestão, Preservação, Dados, *Dataverse*.

## **Sumário**

### **1. INTRODUÇÃO**

<b>2. PRÉ-REQUISITOS</b>	<b>5</b>
<b>2.1 REPOSITÓRIOS</b>	<b>5</b>
<b>2.2. ATUALIZAÇÃO DO SISTEMA</b>	<b>6</b>
<b>2.3. SELINUX</b>	<b>6</b>
<b>3. DEPENDÊNCIAS</b>	<b>7</b>
<b>4. ARCHIVEMATICA STORAGE SERVICE</b>	<b>8</b>
<b>5. ARCHIVEMATICA MCP SERVER E DASHBOARD</b>	<b>10</b>
<b>6. ARCHIVEMATICA MCP CLIENT</b>	<b>12</b>
<b>7. PÓS INSTALAÇÃO</b>	<b>14</b>
<b>7.1. FIREWALL</b>	<b>14</b>
<b>7.2. COMUNICAÇÃO DO STORAGE COM A DASHBOARD</b>	<b>15</b>
<b>7.3. COMUNICAÇÃO ARCHIVEMATICA STORAGE - DATAVERSE</b>	<b>17</b>
<b>7.3.1. USUÁRIO DO DATAVERSE</b>	<b>17</b>
<b>7.3.2. SPACES NO ARCHIVEMATICA STORAGE</b>	<b>18</b>
<b>7.3.3. LOCATIONS NO ARCHIVEMATICA STORAGE</b>	<b>20</b>
<b>7.4. DASHBOARD DO ARCHIVEMATICA</b>	<b>24</b>
<b>REFERÊNCIAS</b>	<b>27</b>

## 1. INTRODUÇÃO

Durante esta implementação, trabalhamos com um servidor virtual de configuração 2 vCPUs, 4 GB de memória principal e 50 GB de armazenamento em disco, foi utilizado o sistema operacional CentOS 7 e implantadas as versões do *Archivematica 1.11.2* e do *Archivematica Storage Service 0.16.1*, ambas no mesmo servidor. O Archivematica pode ser instalado em diferentes distribuições Linux, mas o CentOS é o mais recomendado em ambiente de produção, portanto considere estas configurações e versões ao replicar os experimentos.

## 2. PRÉ-REQUISITOS

Primeiro passo é adicionar os repositórios ao sistema para instalarmos as dependências.

### 2.1 REPOSITÓRIOS

**Passo 1:** Adicionar repositório *Extra Packages for Enterprise Linux (EPEL)*.

```
sudo yum install -y epel-release
```

**Passo 2:** Adicionar repositório do Archivematica.

```
sudo -u root bash -c 'cat << EOF > /etc/yum.repos.d/archivematica.repo  
[archivematica]  
name=archivematica  
baseurl=https://packages.archivematica.org/1.11.x/centos  
gpgcheck=1  
gpgkey=https://packages.archivematica.org/1.11.x/key.asc  
enabled=1  
EOF'
```

**Passo 3:** Adicionar repositório extra do *Archivematica*.

```
sudo -u root bash -c 'cat << EOF > /etc/yum.repos.d/archivematica-extras.repo
```

```
[archivematica-extras]
name=archivematica-extras
baseurl=https://packages.archivematica.org/1.11.x/centos-extras
gpgcheck=1
gpgkey=https://packages.archivematica.org/1.11.x/key.asc
enabled=1
EOF'
```

**Passo 4:** Adicionar o repositório do *Elasticsearch*, o *Archivematica* pode ser instalado com ou sem ele, portanto é opcional.

```
sudo -u root rpm --import https://artifacts.elastic.co/GPG-KEY-elasticsearch
sudo -u root bash -c 'cat << EOF > /etc/yum.repos.d/elasticsearch.repo
[elasticsearch-6.x]
name=Elasticsearch repository for 6.x packages
baseurl=https://artifacts.elastic.co/packages/6.x/yum
gpgcheck=1
gpgkey=https://artifacts.elastic.co/GPG-KEY-elasticsearch
enabled=1
autorefresh=1
type=rpm-md
EOF'
```

## 2.2. ATUALIZAÇÃO DO SISTEMA

**Passo 1:** Atualizar pacotes já instalados.

```
sudo yum update -y
```

## 2.3. SELINUX

Para as dependências do *Archivematica* funcionarem corretamente precisamos ajustar as configurações do SELinux no CentOS caso esteja ativado.

**Passo 1:** Instalar utilitários de política do SELinux (utilitários Python).

```
sudo yum install -y policycoreutils-python
```

**Passo 2:** Permitir Nginx utilizar as portas 81 e 8001, na porta 81 será o proxy para o serviço *Archivematica Dashboard* e na porta 8001 o proxy para o serviço *Archivematica Storage*. É possível escolher outras portas para uso, troque de acordo com as suas configurações de ambiente e ajuste os comandos abaixo com a porta de preferência.

```
sudo semanage port -a -t http_port_t -p tcp 81
sudo semanage port -m -t http_port_t -p tcp 81
sudo semanage port -a -t http_port_t -p tcp 8001
sudo semanage port -m -t http_port_t -p tcp 8001
```

**Passo 3:** Confira se as portas foram liberadas.

```
sudo semanage port -l | grep http_port_t
```

**Passo 4:** Permitir Nginx se conectar ao MySQL e ao serviço Gunicorn.

```
sudo setsebool -P httpd_can_network_connect_db=1
sudo setsebool -P httpd_can_network_connect=1
```

**Passo 5:** Permitir o Nginx alterar os limites do sistema (*/etc/security/limits.conf*).

```
sudo setsebool -P httpd_setrlimit 1
```

### 3. DEPENDÊNCIAS

Com os repositórios adicionados, as dependências do *Archivematica* já podem ser instaladas.

**Passo 1:** Instalar os pacotes *Nano*, *Java*, *Python*, *Elasticsearch*, *MariaDB Server* e o *Gearmand*.

```
sudo -u root yum install -y nano java-1.8.0-openjdk-headless elasticsearch mariadb-server gearmand
```

**Passo 2:** Habilitar o *Elasticsearch* para começar com o sistema, iniciar o serviço e verificar o status (opcional).

```
sudo -u root systemctl enable elasticsearch
sudo -u root systemctl start elasticsearch
sudo -u root systemctl status elasticsearch
```

**Passo 3:** Habilitar o *MariaDB Server* para começar com o sistema, iniciar o serviço e verificar o status.

```
sudo -u root systemctl enable mariadb
sudo -u root systemctl start mariadb
sudo -u root systemctl status mariadb
```

**Passo 4:** Habilitar o *Gearmand* para começar com o sistema, iniciar o serviço e verificar o status.

```
sudo -u root systemctl enable gearmand
sudo -u root systemctl start gearmand
sudo -u root systemctl status gearmand
```

#### 4. ARCHIVEMATICA STORAGE SERVICE

O *Storage Service* é o gerenciador de pipelines, fontes de coleta e localizações em disco para a ingestão e armazenamento. Ele funciona como um backend de configuração do *Archivematica* e é instalado primeiro.

**Passo 1:** Instalar os pacotes do *Storage*. Dependendo da sua velocidade de conexão à internet, o *download* dos pacotes pelo repositório do *Archivematica* pode parar, então é preciso dar o comando de instalação novamente.

```
sudo -u root yum install -y python-pip archivematica-storage-service
```



**Passo 2:** Realizar as migrações no banco de dados SQLite. Essa tarefa deve ser executada como usuário "archivematica".

```
sudo -u archivematica bash -c "\
set -a -e -x
source /etc/sysconfig/archivematica-storage-service
cd /usr/lib/archivematica/storage-service
/usr/share/archivematica/virtualenvs/archivematica-storage-service/bin/python
manage.py migrate";
```

**Passo 3:** Iniciar o serviço do *Storage* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable archivematica-storage-service
sudo -u root systemctl start archivematica-storage-service
sudo -u root systemctl status archivematica-storage-service
```

**Passo 4:** Acessar o arquivo de configurações do Nginx para o *Storage Service* e alterar a porta 8001 para a porta escolhida na configuração do SELinux seção 2.3, passo 2. Caso tenha mantido a porta padrão 8001 pode pular essa etapa.

```
sudo -u root nano /etc/nginx/conf.d/archivematica-storage-service.conf
```

**Passo 5:** Iniciar o proxy Nginx e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable nginx
sudo -u root systemctl start nginx
sudo -u root systemctl status nginx
```

**Passo 6:** Iniciar o gerador de números aleatórios *rngd* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable rngd
sudo -u root systemctl start rngd
sudo -u root systemctl status rngd
```

**Passo 7:** Confira se o serviço do Archivematica Storage esta disponível em `http://<ip>:8001`. Criaremos o usuário de acesso nas próximas etapas.

## 5. ARCHIVEMATICA MCP SERVER E DASHBOARD

O *MCP Server* é o núcleo do *Archivematica* e controla os vários microserviços no sistema. As informações de configuração e processamento são mantidas no banco de dados. O usuário monitora e controla o status e o fluxo de trabalho por meio da *Dashboard*.

**Passo 1:** Instalar os pacotes da *Dashboard* e do *MCP Server*.

```
sudo -u root yum install -y archivematica-common archivematica-mcp-server
archivematica-dashboard
```

**Passo 2:** Após a instalação dos pacotes, crie o banco de dados MCP no MySQL.

```
sudo -H -u root mysql -hlocalhost -uroot -e "DROP DATABASE IF EXISTS MCP;
CREATE DATABASE MCP CHARACTER SET utf8 COLLATE utf8_unicode_ci;"
```

**Passo 3:** Crie o usuário *archivematica* no *MySQL* e não altere a senha do banco “demo” do comando abaixo. Ela é utilizada em várias migrações do código fonte do *Archivematica* e se alterada o serviço não vai iniciar corretamente.

```
sudo -H -u root mysql -hlocalhost -uroot -e "CREATE USER
'archivematica'@'localhost' IDENTIFIED BY 'demo';"
```

**Passo 4:** Dê permissões ao usuário “*archivematica*” para alterar o banco MCP.

```
sudo -H -u root mysql -hlocalhost -uroot -e "GRANT ALL ON MCP.* TO
'archivematica'@'localhost';"
```

**Passo 5:** Utilize o usuário “*archivematica*” para realizar as migrações no banco.

```
sudo -u archivematica bash -c "\
set -a -e -x
```

```
source /etc/sysconfig/archivematica-dashboard
cd /usr/share/archivematica/dashboard
/usr/share/archivematica/virtualenvs/archivematica-dashboard/bin/python
manage.py migrate
";
```

**Passo 6:** Iniciar o *MCP Server* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable archivematica-mcp-server
sudo -u root systemctl start archivematica-mcp-server
sudo -u root systemctl status archivematica-mcp-server
```

**Passo 7:** Iniciar a *Dashboard* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable archivematica-dashboard
sudo -u root systemctl start archivematica-dashboard
sudo -u root systemctl status archivematica-dashboard
```

**Passo 8:** Acessar o arquivo de configurações do Nginx para a *Dashboard* e alterar a porta 81 para a porta escolhida na configuração do SELinux seção 2.3, passo 2. Caso tenha mantido a porta padrão 81 podemos pular essa etapa.

```
sudo -u root nano /etc/nginx/conf.d/archivematica-dashboard.conf
```

**Passo 9:** Acessar o arquivo de configurações gerais do Nginx.

```
sudo -u root nano /etc/nginx/nginx.conf
```

**Passo 10:** Precisamos desabilitar o servidor default do Nginx que está configurado para utilizar a porta 80, para isso comentamos com o caractere “#” no início da linha toda a seção de texto abaixo.

```
# server {
#     listen    80 default_server;
#     listen    [::]:80 default_server;
```

```
# server_name _;
# root /usr/share/nginx/html;
#
# # Load configuration files for the default server block.
# include /etc/nginx/default.d/*.conf;
# location / {
# }
# error_page 404 /404.html;
#     location = /40x.html {
# }
# error_page 500 502 503 504 /50x.html;
#     location = /50x.html {
# }
# }
```

**Passo 11:** Reiniciar o proxy Nginx.

```
sudo -u root systemctl restart nginx
sudo -u root systemctl status nginx
```

## 6. ARCHIVEMATICA MCP CLIENT

O *Archivematica* possui uma ou mais instâncias *MCP Client* para executar o trabalho de ingestão dos pacotes. São implementações de micro serviços do *Gearman* que informam ao servidor quais tarefas eles podem executar e aguardam o servidor atribuir uma tarefa a eles.

**Passo 1:** Instalar os pacotes do *MCP Client*.

```
sudo -u root yum install -y archivematica-mcp-client
```

**Passo 2:** Após a instalação dos pacotes, crie um link simbólico para o 7-Zip.

```
sudo ln -sf /usr/bin/7za /usr/bin/7z
```

**Passo 3:** Ajustar a configuração do antivírus ClamAV.

```
sudo -u root sed -i 's/^#TCPSocket/TCP/g' /etc/clamd.d/scan.conf
sudo -u root sed -i 's/^Example//g' /etc/clamd.d/scan.conf
```

**Passo 4:** Caso deseje executar o *Archivematica* sem o *Elasticsearch*, desabilite as opções com os comandos abaixo ou altere “false” para “true” se o *Elasticsearch* foi instalado e iniciado (ver secção 2.1 passo 4 e secção 3 passo 2).

```
sudo sh -c 'echo "ARCHIVEMATICA_DASHBOARD_DASHBOARD_SEARCH_ENABLED=false" >> /etc/sysconfig/archivematica-dashboard'
sudo sh -c 'echo "ARCHIVEMATICA_MCPSERVER_MCPSERVER_SEARCH_ENABLED=false" >> /etc/sysconfig/archivematica-mcp-server'
sudo sh -c 'echo "ARCHIVEMATICA_MCPCLIENT_MCPCLIENT_SEARCH_ENABLED=false" >> /etc/sysconfig/archivematica-mcp-client'
```

**Passo 5:** Iniciar o *MCP Client* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable archivematica-mcp-client
sudo -u root systemctl start archivematica-mcp-client
sudo -u root systemctl status archivematica-mcp-client
```

**Passo 6:** Iniciar o serviço *File Information Tool Set (FITS)* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable fits-nailgun
sudo -u root systemctl start fits-nailgun
sudo -u root systemctl status fits-nailgun
```

**Passo 7:** Iniciar o antivírus *ClamAV* e habilitar para iniciar com o sistema.

```
sudo -u root systemctl enable clamd@scan
```

```
sudo -u root systemctl start clamd@scan
sudo -u root systemctl status clamd@scan
```

**Passo 8:** Reiniciar o *MCP Server* e a *Dashboard*.

```
sudo -u root systemctl restart archivematica-mcp-server
sudo -u root systemctl status archivematica-mcp-server
sudo -u root systemctl restart archivematica-dashboard
sudo -u root systemctl status archivematica-dashboard
```

**Passo 9:** Reiniciar o proxy Nginx.

```
sudo -u root systemctl restart nginx
sudo -u root systemctl status nginx
```

**Passo 10:** Confira se ambos serviços estão disponíveis em `http://<ip>:81` e `http://<ip>:8001`.

## 7. PÓS INSTALAÇÃO

### 7.1. FIREWALL

Recomenda-se a instalação de um serviço de firewall para proteção do sistema. Neste ambiente utilizamos o `firewalld`.

**Passo 1:** Instalando o *firewalld*.

```
sudo yum install -y firewalld
```

**Passo 2:** Configurar o *firewalld* liberando as portas do Nginx para acesso externo.

```
sudo firewall-cmd --add-port=81/tcp --permanent
sudo firewall-cmd --add-port=8001/tcp --permanent
```

**Passo 2:** Não esqueça de liberar a porta de conexão SSH utilizada ou perderá acesso a máquina. Aqui estamos utilizando a porta 2022.

```
sudo firewall-cmd --add-port=2022/tcp --permanent
```

**Passo 3:** Carregar as novas configurações para entrarem em vigor.

```
sudo firewall-cmd --reload
```

**Passo 4:** Habilitar *firewalld* para iniciar com o sistema.

```
sudo -u root systemctl enable firewalld
```

## 7.2. COMUNICAÇÃO DO STORAGE COM A DASHBOARD

Confira se o serviço do *Archivematica Dashboard* está disponível em <http://<ip>:81> e crie o usuário administrador. Na segunda etapa de configuração da *Dashboard*, é necessário configurar a comunicação com o *Storage Service*. Nessa etapa é solicitada a API Key do usuário administrador do *Storage Service*.

**Passo 1:** Criar o usuário administrador do *Archivematica Storage Service*.

```
sudo -u archivematica bash -c "\
set -a -e -x
source /etc/default/archivematica-storage-service || \
source /etc/sysconfig/archivematica-storage-service \
|| (echo 'Environment file not found'; exit 1)
cd /usr/lib/archivematica/storage-service
/usr/share/archivematica/virtualenvs/archivematica-storage-service/bin/python
manage.py createsuperuser
";
```

**Passo 2:** Com o usuário criado, acessamos o *Storage Service* <http://<ip>:8001>, e seguindo os menus **Administration > Users > Edit** chegamos a API Key. Também há a possibilidade de ir diretamente através do endereço <http://<ip>:8001/administration/users/1/edit/>

The screenshot shows a web browser window with the URL `138.121.70.95:8000/administration/users/1/edit/`. The page title is "Archivematica Storage Service" and the navigation menu includes "Home", "Pipelines", "Spaces", "Locations", "Packages", "Administration", and "Log out". The main content area is titled "Edit User" and contains the following form fields and controls:

- Configuration** (link)
- Users** (link)
- Version** (link)
- Service callbacks** (link)
- Encryption keys** (link)
- Language** (link)
- Username:**  Required, 255 characters or fewer. Only letters, numbers, and @, ., +, -, or \_ characters.
- First name:**
- Last name:**
- Email address:**
- Administrator?**  Designates that this user has all permissions without explicitly assigning them.
- Edit User** (button)
- New password:**
- New password confirmation:**
- Note:** if you change your password, your API key will also be changed to a new, randomly generated one.
- Change Password** (button)
- API key:** `d279a0bf3049c7354ef7526926e9621c3a5a16c0`

**Passo 3:** Continuando a configuração, inserimos o domínio ou IP de acesso ao servidor do *Storage Service* e a porta, neste caso `http://localhost:8001` pois está na mesma máquina. Se os dois serviços estiverem instalados em servidores diferentes, deverá ser usado o endereço IP ou o nome de domínio completamente qualificado da sua instância do *Storage Service*, por exemplo, `http://storage.archivematica.org:8001`. Lembre-se de garantir que quaisquer regras de firewall (por exemplo, iptables, ufw, grupos de segurança da AWS etc.) estejam configuradas para permitir solicitações do IP da *Dashboard* para o IP do *Storage* na porta apropriada.

**Passo 4:** Inserimos o nome do usuário correspondente, caso tenha utilizado o padrão durante a configuração é "archivematica" e colocamos a API Key do *Storage* copiada no passo 2 para finalizar a configuração da *Dashboard*.



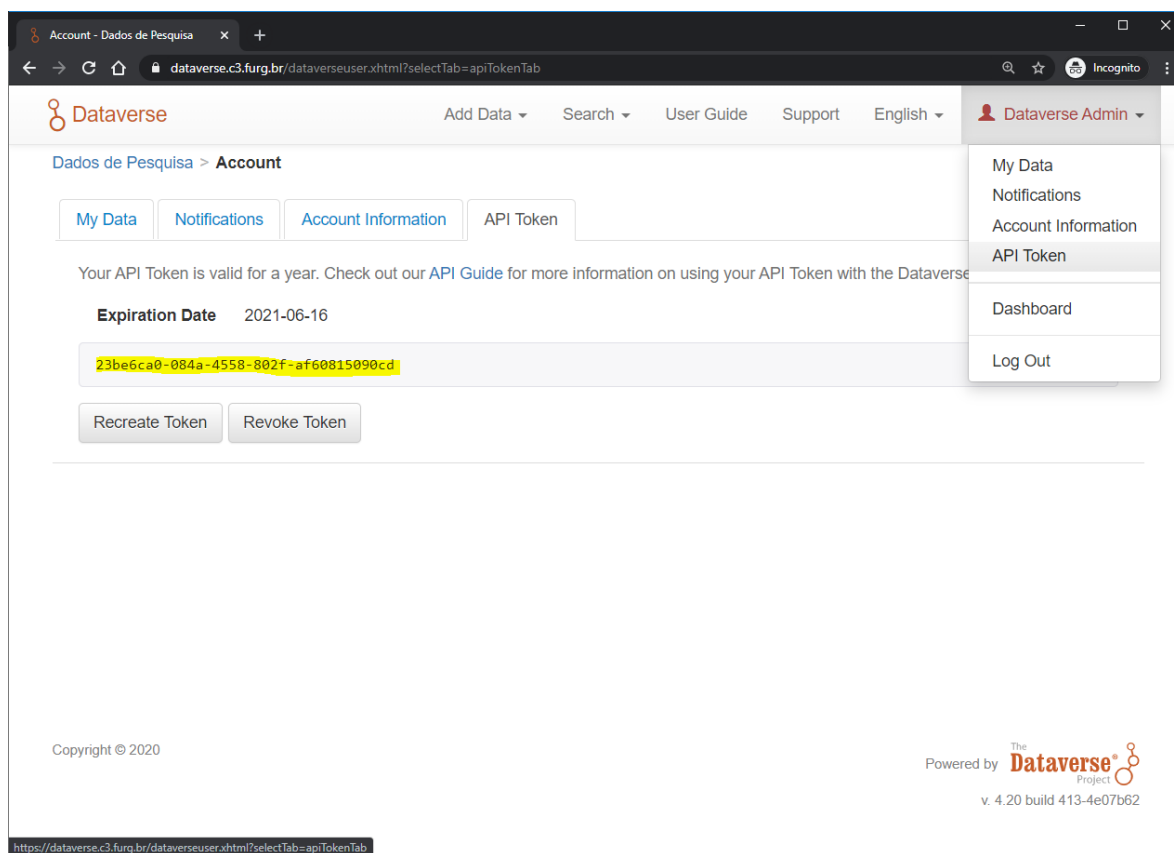
## 7.3. COMUNICAÇÃO ARCHIVEMATICA STORAGE - DATAVERSE

O *Archivematica* possibilita a ingestão de dados publicados em *Dataverse*, para isso é preciso realizar uma configuração no *Storage Service* criando essa conexão entre os servidores.

### 7.3.1. USUÁRIO DO DATAVERSE

Para que o *Storage Service* se comunique com o *Dataverse* é preciso atrelar um usuário existente durante a configuração.

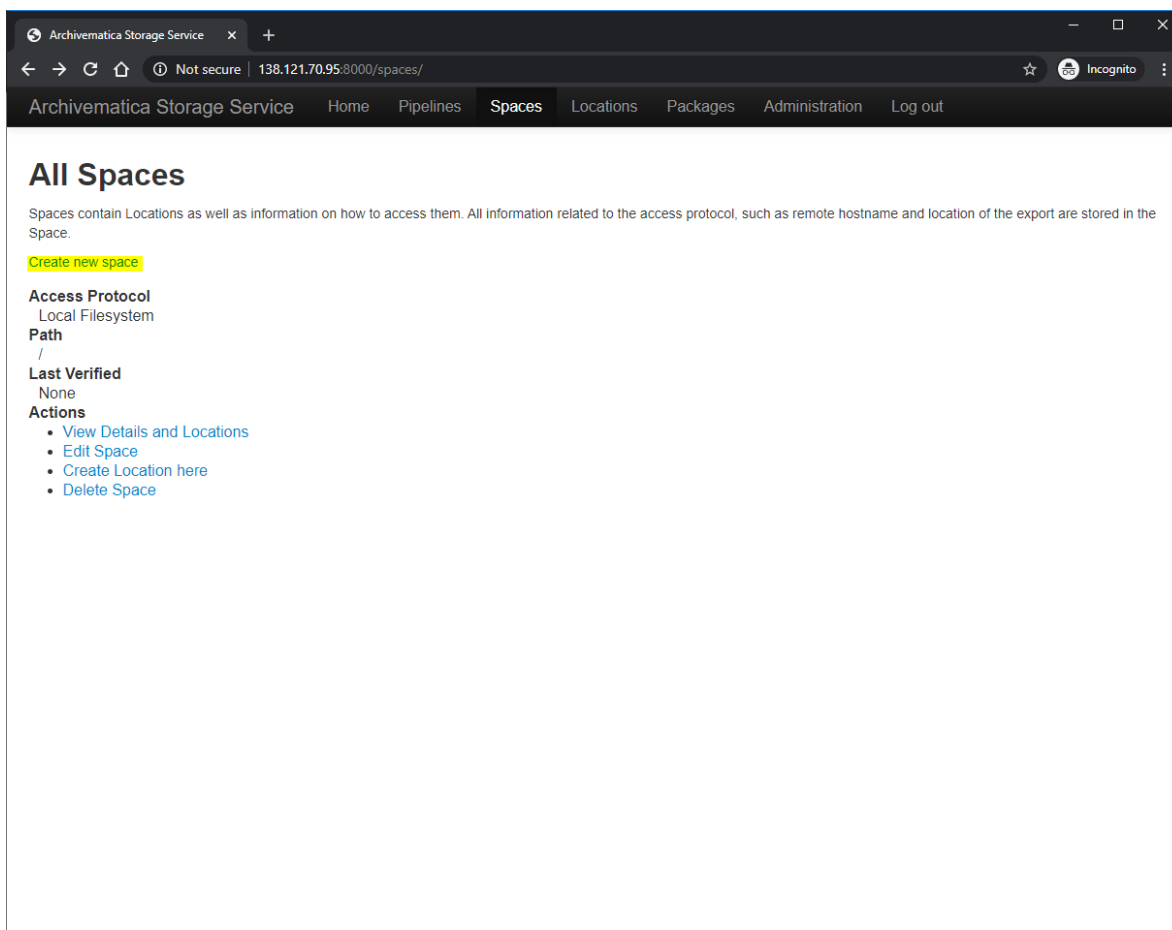
**Passo 1:** Para começar, acessamos o *Dataverse* com um usuário que tenha permissões de leitura aos datasets e copiamos a API Key.



The screenshot shows a web browser window displaying the Dataverse Admin interface. The browser's address bar shows the URL: `dataverse.c3.furg.br/dataverseuser.xhtml?selectTab=apiTokenTab`. The page title is "Account - Dados de Pesquisa". The navigation bar includes "Add Data", "Search", "User Guide", "Support", "English", and a user profile dropdown for "Dataverse Admin". The main content area is titled "Dados de Pesquisa > Account" and contains tabs for "My Data", "Notifications", "Account Information", and "API Token". The "API Token" tab is active, showing a message: "Your API Token is valid for a year. Check out our [API Guide](#) for more information on using your API Token with the Dataverse". Below this, the "Expiration Date" is listed as "2021-06-16". The API token itself is displayed as a long alphanumeric string: `23be6ca0-084a-4558-802f-af60815090cd`, which is highlighted in yellow. At the bottom of the token display area, there are two buttons: "Recreate Token" and "Revoke Token". The footer of the page includes "Copyright © 2020" and "Powered by The Dataverse Project v. 4.20 build 413-4e07b62".

## 7.3.2. SPACES NO ARCHIVEMATICA STORAGE

O menu *Spaces* no *Archivematica* é responsável pela criação e gerência de fontes de importação e armazenamento de dados. Para configurarmos um novo *Space* acessamos o *Storage Service* e clicamos em *Create New Space*.



**Passo 1:** Na configuração de um novo *Space*, definimos o protocolo de acesso (ex.: *Filesystem*, *Dataverse*, *Dspace*, *S3*) a um recurso que queremos utilizar pelo *Archivematica Storage*. Ajustamos o espaço e o caminho em disco para armazenar esses recursos temporariamente e a longo prazo. Neste ambiente utilizamos a pasta padrão do *Archivematica*, por exemplo “*/var/archivematica/dataverseName/staging*” para armazenamento temporário e “*/var/archivematica/dataverseName/storage*” para longo prazo. É importante lembrar o que o usuário “*archivematica*” do sistema precisa ter acesso de leitura e escrita nestas localizações do disco.

The screenshot shows a web browser window with the URL `138.121.70.95:8000/spaces/create/`. The page title is "Archivematica Storage Service" and the navigation menu includes "Home", "Pipelines", "Spaces", "Locations", "Packages", "Administration", and "Log out". The main heading is "Create Space".

**Access protocol:** A dropdown menu is set to "Dataverse". Below it is the text "How the space can be accessed."

**Size:** A text input field contains "0". Below it is the text "Size in bytes (optional)".

**Path:** A text area contains `/var/archivematica/dataverseName/storage`. Below it is the text "Absolute path to the space on the storage service machine."

**Staging path:** A text area contains `/var/archivematica/dataverseName/staging`. Below it is the text "Absolute path to a staging area. Must be UNIX filesystem compatible, preferably on the same filesystem as the path."

**Passo 2:** Continuando a configuração para acesso a dados de um *Dataverse*, ainda precisamos definirmos o protocolo, o domínio e a porta da API de administração. É importante definir o protocolo HTTP devido ao possível redirecionamento para HTTPS efetuado pelo proxy. Lembre-se também de liberar acesso a porta da API no firewall dos servidores a fim de realizar as determinadas conexões. Neste ambiente, estamos acessando o piloto experimental do Repositório Institucional de Dados de Pesquisa da RNP (<https://dadosabertos.rnp.br>) e vamos configurar da seguinte forma:

**HOST:** <http://dadosabertos.rnp.br:8080>

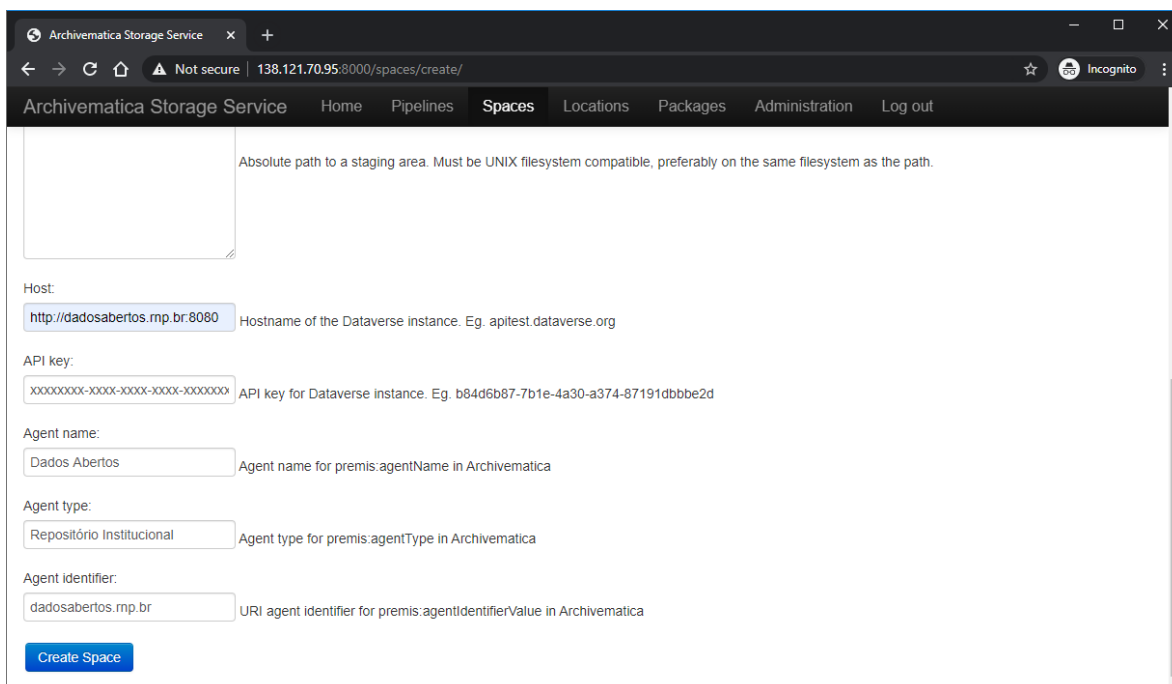
**API Key:** xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxx

**Passo 3:** Definimos nomes de identificação para os pacotes ingeridos pelo *Storage Service*. Estas definições serão importantes para poder identificar de onde vem os pacotes que serão preservados a longo prazo no *Archivematica*.

**Agent Name:** Dados Abertos

**Agent Type:** Repositório Institucional

**Agente Identifier:** dadosabertos.rnp.br



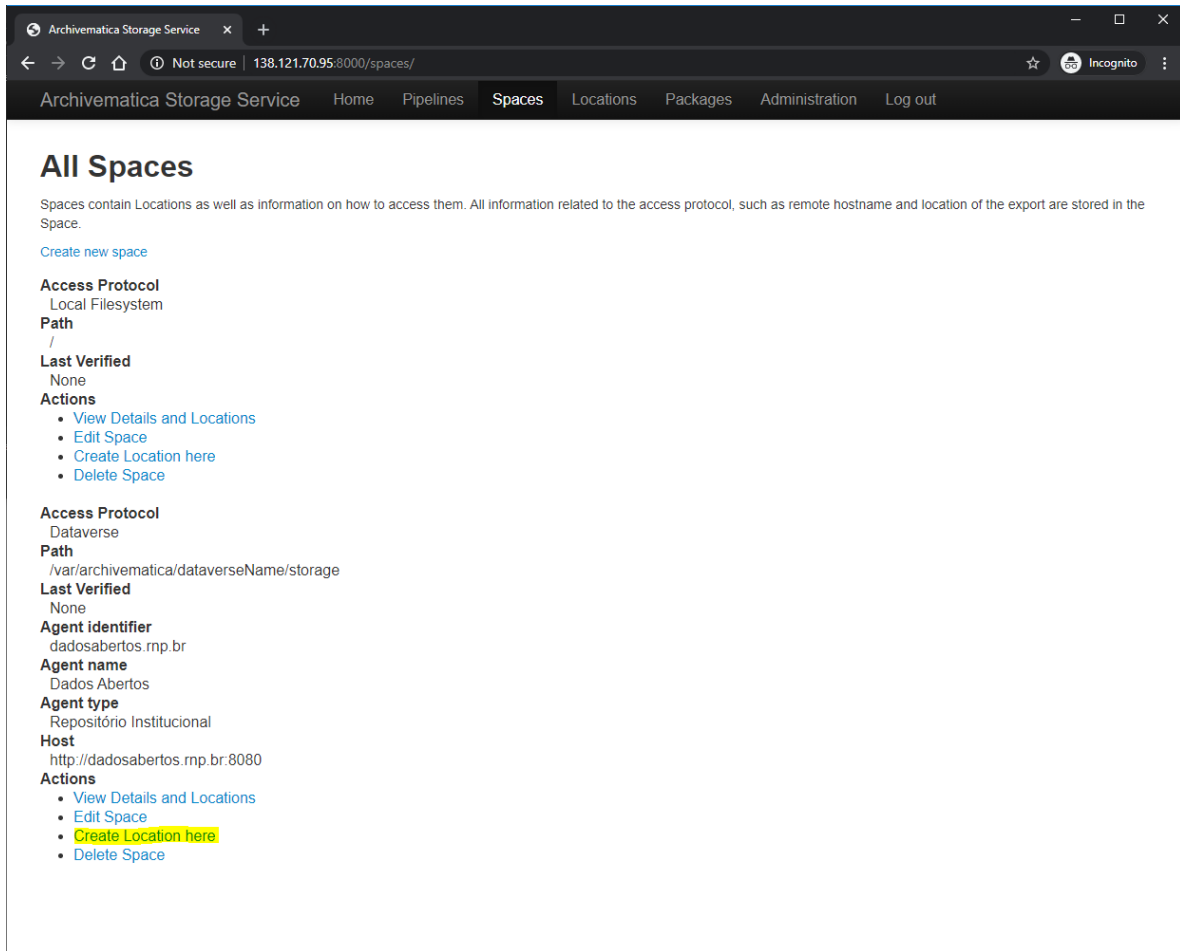
The screenshot shows a web browser window with the URL `138.121.70.95:8000/spaces/create/`. The page title is "Archivematica Storage Service" and the navigation menu includes "Home", "Pipelines", "Spaces", "Locations", "Packages", "Administration", and "Log out". The main content area contains a form for creating a space. At the top, there is a text input field for an absolute path to a staging area, with a note: "Absolute path to a staging area. Must be UNIX filesystem compatible, preferably on the same filesystem as the path." Below this are several labeled input fields: "Host" (with value `http://dadosabertos.rnp.br:8080`), "API key" (with a masked value), "Agent name" (with value "Dados Abertos"), "Agent type" (with value "Repositório Institucional"), and "Agent identifier" (with value `dadosabertos.rnp.br`). A blue "Create Space" button is located at the bottom of the form.

**Passo 4:** Finalize a criação clicando em *Create Space*.

### 7.3.3. LOCATIONS NO ARCHIVEMATICA STORAGE

*Location* podem indicar uma localização interna ou externa dependendo do protocolo de acesso do *Space*. Configurando um *Space* com protocolo *Filesystem* definiremos uma localização interna e para isso basta definir uma pasta no Disco Local. Já no caso de um *Space* com protocolo *Dataverse*, atrelamos um usuário utilizando a API Key (secção 7.3.2 passo 2) e então temos uma localização externa de onde os dados estão e poderemos acessar de acordo com o nível de autorização deste usuário.

**Passo 1:** Para começar, logo após salvar o *Space* criado na secção anterior. Clique em *Create Location Here*. Também é possível criar uma *Location* clicando no menu *Spaces* onde será listado todos cadastrados no *Storage Service*.

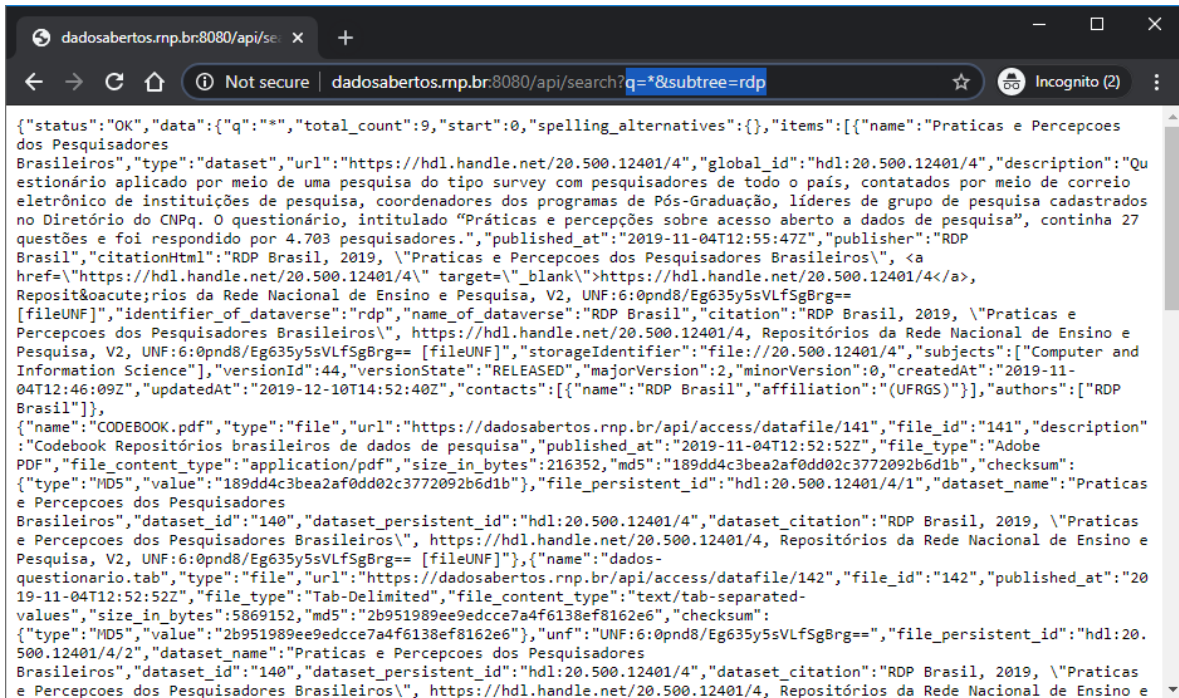


**Passo 2:** *Purpose* é onde definimos o propósito da comunicação, neste caso utilizaremos *Transfer Source* pois queremos transferir os arquivos da fonte *Dataverse* para o armazenamento no *Storage Service*.

**Passo 3:** *Pipeline* é o canal de comunicação da *Dashboard* como *Storage Service* que terá acesso a esta *Location*. Fizemos a configuração deste *pipeline* na secção 7.2. Como temos apenas uma *Dashboard*, por enquanto utilizaremos ela. Mas há a possibilidade de ter múltiplas instâncias de *Dashboard* que utilizam o mesmo *Storage Service*.

**Passo 4:** *Relative Path* é a parte de configuração da API de administração do *Dataverse* e o conteúdo é que define acesso ao datasets listados para ingestão. *Query* é o texto que iremos procurar e *Subtree* representa o *Dataverse* lógico.

**Exemplo 1:** Listar todos (\*) resultados do Dataverse lógico “rdp” hospedado no Repositório Dados Abertos (<http://dadosabertos.rnp.br>):



```
{
  "status": "OK",
  "data": {
    "q": "*",
    "total_count": 9,
    "start": 0,
    "spelling_alternatives": {},
    "items": [
      {
        "name": "Praticas e Percepcoes dos Pesquisadores Brasileiros",
        "type": "dataset",
        "url": "https://hdl.handle.net/20.500.12401/4",
        "global_id": "hdl:20.500.12401/4",
        "description": "Question\u00e1rio aplicado por meio de uma pesquisa do tipo survey com pesquisadores de todo o pa\u00eds, contatados por meio de correio eletr\u00f4nico de institui\u00e7\u00f5es de pesquisa, coordenadores dos programas de P\u00f3s-Gradua\u00e7\u00e3o, l\u00edderes de grupo de pesquisa cadastrados no Diret\u00f3rio do CNPq. O question\u00e1rio, intitulado \"Pr\u00e1ticas e percep\u00e7\u00f5es sobre acesso aberto a dados de pesquisa\", continha 27 quest\u00f5es e foi respondido por 4.703 pesquisadores.",
        "published_at": "2019-11-04T12:55:47Z",
        "publisher": "RDP Brasil",
        "citationHtml": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", <a href='\"https://hdl.handle.net/20.500.12401/4\" target='\"_blank\">https://hdl.handle.net/20.500.12401/4</a>, Reposit\u00f3rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBr== [fileUNF]",
        "identifi\u00e7\u00e3o_of_dataverse": "rdp",
        "name_of_dataverse": "RDP Brasil",
        "citation": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", https://hdl.handle.net/20.500.12401/4, Reposit\u00f3rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBr== [fileUNF]",
        "storageIdentifier": "file://20.500.12401/4",
        "subjects": ["Computer and Information Science"],
        "versionId": 44,
        "versionState": "RELEASED",
        "majorVersion": 2,
        "minorVersion": 0,
        "createdAt": "2019-11-04T12:46:09Z",
        "updatedAt": "2019-12-10T14:52:40Z",
        "contacts": [{"name": "RDP Brasil", "affiliation": "(UFRGS)"}],
        "authors": ["RDP Brasil"]
      },
      {
        "name": "CODEBOOK.pdf",
        "type": "file",
        "url": "https://dadosabertos.rnp.br/api/access/datafile/141",
        "file_id": "141",
        "description": "Codebook Reposit\u00f3rios brasileiros de dados de pesquisa",
        "published_at": "2019-11-04T12:52:52Z",
        "file_type": "Adobe PDF",
        "file_content_type": "application/pdf",
        "size_in_bytes": 216352,
        "md5": "189dd4c3bea2af0dd02c3772092b6d1b",
        "checksum": {"type": "MD5", "value": "189dd4c3bea2af0dd02c3772092b6d1b"},
        "file_persistent_id": "hdl:20.500.12401/4/1",
        "dataset_name": "Praticas e Percepcoes dos Pesquisadores Brasileiros",
        "dataset_id": "140",
        "dataset_persistent_id": "hdl:20.500.12401/4",
        "dataset_citation": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", https://hdl.handle.net/20.500.12401/4, Reposit\u00f3rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBr== [fileUNF]",
        "name": "dados-questionario.tab",
        "type": "file",
        "url": "https://dadosabertos.rnp.br/api/access/datafile/142",
        "file_id": "142",
        "published_at": "2019-11-04T12:52:52Z",
        "file_type": "Tab-Delimited",
        "file_content_type": "text/tab-separated-values",
        "size_in_bytes": 5869152,
        "md5": "2b951989ee9edc7a4f6138ef8162e6",
        "checksum": {"type": "MD5", "value": "2b951989ee9edc7a4f6138ef8162e6"},
        "unf": "UNF:6:0pnd8/Eg635y5sVLFsGBr==",
        "file_persistent_id": "hdl:20.500.12401/4/2",
        "dataset_name": "Praticas e Percepcoes dos Pesquisadores Brasileiros",
        "dataset_id": "140",
        "dataset_persistent_id": "hdl:20.500.12401/4",
        "dataset_citation": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", https://hdl.handle.net/20.500.12401/4, Reposit\u00f3rios da Rede Nacional de Ensino e"
      }
    ]
  }
}
```

Podemos ver que a Query retornou 9 resultados, logo o *Relative Path* ficaria:

```
Query: *
Subtree: rdp
```

**Exemplo 2:** Listar todos (\*) resultados da “root” do Reposit\u00f3rio Dados Abertos (<http://dadosabertos.rnp.br>):

```
dadosabertos.rnp.br:8080/api/se...
Not secure | dadosabertos.rnp.br:8080/api/search?q=*%&subtree=root
{"status": "OK", "data": {"q": "*", "total_count": 137, "start": 0, "spelling_alternatives": {}, "items": [{"name": "Campus Inteligentes", "type": "dataverse", "url": "https://dadosabertos.rnp.br/dataverse/campusinteligentes", "identifier": "campusinteligentes", "published_at": "2019-11-29T19:00:16Z"}, {"name": "Praticas e Percepcoes dos Pesquisadores Brasileiros", "type": "dataset", "url": "https://hdl.handle.net/20.500.12401/4", "global_id": "hdl:20.500.12401/4", "description": "Qu estionário aplicado por meio de uma pesquisa do tipo survey com pesquisadores de todo o país, contatados por meio de correio eletrônico de instituições de pesquisa, coordenadores dos programas de Pós-Graduação, líderes de grupo de pesquisa cadastrados no Diretório do CNPq. O questionário, intitulado \"Práticas e percepções sobre acesso aberto a dados de pesquisa\", continha 27 questões e foi respondido por 4.703 pesquisadores.", "published_at": "2019-11-04T12:55:47Z", "publisher": "RDP Brasil", "citationHtml": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", <a href=\"https://hdl.handle.net/20.500.12401/4\" target=\"_blank\">https://hdl.handle.net/20.500.12401/4</a>, Reposit&ocirc;rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBrG== [fileUNF]", "identifier_of_dataverse": "rdp", "name_of_dataverse": "RDP Brasil", "citation": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", https://hdl.handle.net/20.500.12401/4, Reposit&ocirc;rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBrG== [fileUNF]", "storageIdentifier": "file://20.500.12401/4", "subjects": ["Computer and Information Science"], "versionId": 44, "versionState": "RELEASED", "majorVersion": 2, "minorVersion": 0, "createdAt": "2019-11-04T12:46:09Z", "updatedAt": "2019-12-10T14:52:40Z", "contacts": [{"name": "RDP Brasil", "affiliation": "(UFRGS)"}], "authors": ["RDP Brasil"]}, {"name": "CODEBOOK.pdf", "type": "file", "url": "https://dadosabertos.rnp.br/api/access/datafile/141", "file_id": "141", "description": "Codebook Reposit&ocirc;rios brasileiros de dados de pesquisa", "published_at": "2019-11-04T12:52:52Z", "file_type": "Adobe PDF", "file_content_type": "application/pdf", "size_in_bytes": 216352, "md5": "189dd4c3bea2af0dd02c3772092b6d1b", "checksum": {"type": "MD5", "value": "189dd4c3bea2af0dd02c3772092b6d1b"}, "file_persistent_id": "hdl:20.500.12401/4/1", "dataset_name": "Praticas e Percepcoes dos Pesquisadores Brasileiros", "dataset_id": "140", "dataset_persistent_id": "hdl:20.500.12401/4", "dataset_citation": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", https://hdl.handle.net/20.500.12401/4, Reposit&ocirc;rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBrG== [fileUNF]"}, {"name": "dados-questionario.tab", "type": "file", "url": "https://dadosabertos.rnp.br/api/access/datafile/142", "file_id": "142", "published_at": "2019-11-04T12:52:52Z", "file_type": "Tab-Delimited", "file_content_type": "text/tab-separated-values", "size_in_bytes": 5869152, "md5": "2b951989ee9edc7a4f6138ef8162e6", "checksum": {"type": "MD5", "value": "2b951989ee9edc7a4f6138ef8162e6"}, "unf": "UNF:6:0pnd8/Eg635y5sVLFsGBrG==", "file_persistent_id": "hdl:20.500.12401/4/2", "dataset_name": "Praticas e Percepcoes dos Pesquisadores Brasileiros", "dataset_id": "140", "dataset_persistent_id": "hdl:20.500.12401/4", "dataset_citation": "RDP Brasil, 2019, \"Praticas e Percepcoes dos Pesquisadores Brasileiros\", https://hdl.handle.net/20.500.12401/4, Reposit&ocirc;rios da Rede Nacional de Ensino e Pesquisa, V2, UNF:6:0pnd8/Eg635y5sVLFsGBrG== [fileUNF]"}]}
```

Podemos ver que agora a *Query* retornou 137 resultados, logo todos datasets dentro de todos *Dataverse* lógicos, e o *Relative Path* ficaria:

```
Query: *
Subtree: root
```

**Passo 5:** *Description* é onde definimos uma descrição identificável para essa *Location*. Ela será importante quando listarmos todas *Locations* disponíveis no *Dashboard* de transferência (ver secção 7.4, passo 2).

**Passo 6:** *Quota* é onde definimos o máximo de espaço em disco que iremos disponibilizar para essa *Location*. Deixe “0” se não quiser delimitar.

The screenshot shows the 'Create Location' form in the Archivematica Storage Service web interface. The browser address bar shows the URL: `138.121.70.95:8000/spaces/b0ba5345-504e-4b05-807c-289d225eef8e/location_create/`. The navigation menu includes 'Home', 'Pipelines', 'Spaces', 'Locations', 'Packages', 'Administration', and 'Log out'. The form fields are as follows:

- Space:** `b0ba5345-504e-4b05-807c-289d225eef8e: /var/archivematica/dataverseName/storage (Dataverse)`
- Purpose:** A dropdown menu set to 'Transfer Source'. Description: 'Purpose of the space. Eg. AIP storage, Transfer source.'
- Pipeline:** A dropdown menu set to 'Archivematica on 6e0fd84a1c3'. Description: 'UUID of the Archivematica instance using this location.'
- Relative Path:** A text area containing 'Query: \*' and 'Subtree: root'. A 'Browse' button is next to it. Description: 'Path to location, relative to the storage space's path. Optional. Filter Dataverse by matches and/or Dataverse subtree.'
- Description:** A text input field containing 'Dados Abertos'. Description: 'Human-readable description.'
- Quota:** A text input field containing '0'. Description: 'Size, in bytes (optional)'
- Enabled:** A checked checkbox. Description: 'True if space can be accessed.'
- Set as global default location for its purpose:** An unchecked checkbox.
- Create Location:** A blue button at the bottom of the form.

The browser's address bar at the bottom shows the URL: `138.121.70.95:8000/spaces/.../location_create/`.

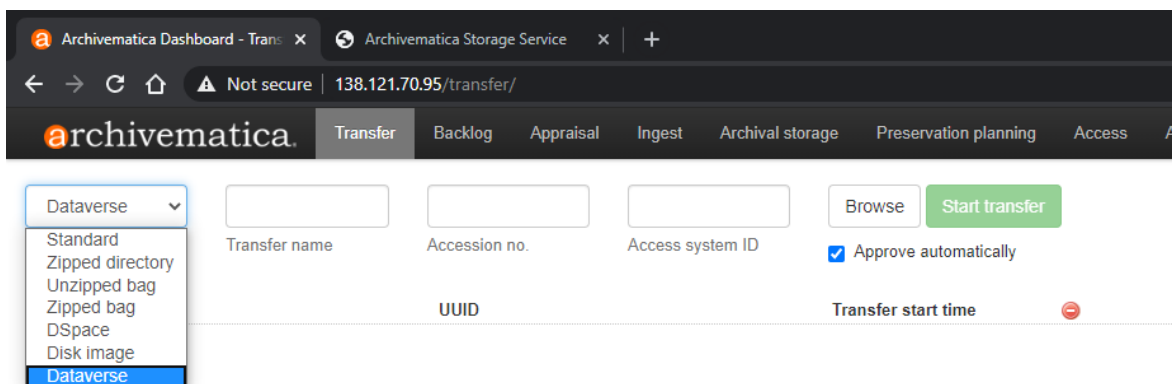
**Passo 7:** Finalize a criação da *Location* clicando em *Create Location*. Após essa etapa, a configuração do *Storage Storage* estará finalizada. Acesse a *Dashboard* (<http://<MY-IP>:81>) e tente realizar a ingestão dos datasets do *Dataverse* cadastrado.

#### 7.4. DASHBOARD DO ARCHIVEMATICA

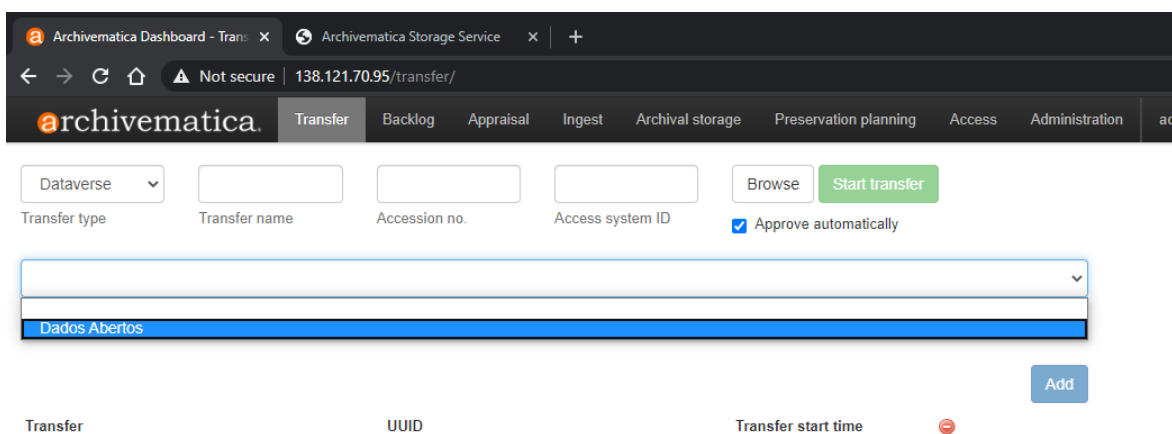
Da perspectiva do usuário final, todas funções do *Archivematica* ocorrem a partir de um painel web. Na *Dashboard* é exibido todo o andamento dos micro serviços de processamento dos arquivos digitais conforme forem progredindo no pipeline de ingestão e preservação. No *Archivematica Dashboard*, *Transfer* é o processo de mover qualquer conjunto de objetos digitais para o *Storage Service* e transformar os materiais em um pacote de informações de envio (SIP).



**Passo 1:** Para realizar a ingestão de um *Dataset*, acesse o menu *Transfer* da *Dashboard* e selecione a fonte de ingestão *Dataverse* no primeiro seletor de opções. Após clique em *Browse* para listar os *Dataverses* disponíveis configurados através do *Storage Service*.



**Passo 2:** Escolha o *Dataverse* e será listado os *Datasets*.



**Passo 3:** Selecione um *Dataset* e clique em *Add*.

Archivematica Dashboard - Transfer | Archivematica Storage Service | 138.121.70.95/transfer/

archivematica | Transfer | Backlog | Appraisal | Ingest | Archival storage | Preservation planning | Access | Administration

Dataverse [v] [ ] [ ] [ ] [Browse] [Start transfer]

Transfer type | Transfer name | Accession no. | Access system ID |  Approve automatically

Dados Abertos [v]

- 153 (Dados de medições da RedeIPE - 2012)
- 57 (Dados de medições da RedeIPE - 2013)
- 70 (Dados de medições da RedeIPE - 2014)
- 71 (Dados de medições da RedeIPE - 2015)
- 72 (Dados de medições da RedeIPE - 2016)
- 73 (Dados de medições da RedeIPE - 2017)
- 74 (Dados de medições da RedeIPE - 2018)
- 75 (Dados de medições da RedeIPE - 2019)
- 818 (Dados de medições da RedeIPE - 2020)
- 810 (Documentação de instalação e configurações - Ubuntu)**
- 168 (Pesquisa Campus Inteligente)
- 140 (Práticas e Percepções dos Pesquisadores Brasileiros)
- 824 (Repositórios que utilizam o Dataverse para dados de pesquisa)

[Add]

Transfer	UUID	Transfer start time
----------	------	---------------------

**Passo 4:** Preencha os campos de identificação do pacote após a ingestão. *Transfer Name* é o nome da transferência, este se tornará o nome do pacote de informações de arquivamento (AIP) resultante. Este é o único campo obrigatório. *Accession no.* é o número único dado a uma nova ingestão inserida no catálogo, a inserção de um número de acesso para sua transferência resultará na cópia do número de acesso no arquivo AIP METS como um evento de registro. Não é usado para identificar ou procurar o AIP no Archivematica. *Access System ID*, a inserção de um campo de ID do sistema de acesso ao configurar sua transferência permite automatizar o processo de upload de um DIP para o AtoM ou o Binder. O Archivematica automaticamente captura esse valor quando atinge o microserviço DIP de *upload*.

The screenshot shows the Archivematica Dashboard interface. At the top, there are browser tabs for 'Archivematica Dashboard - Trans' and 'Archivematica Storage Service'. The address bar shows '138.121.70.95/transfer/'. The navigation menu includes 'Transfer', 'Backlog', 'Appraisal', 'Ingest', 'Archival storage', 'Preservation planning', 'Access', and 'Administration'. Below the navigation, there are input fields for 'Transfer type' (set to 'Standard'), 'Transfer name', 'Accession no.', and 'Access system ID'. There are 'Browse' and 'Start transfer' buttons. A checkbox for 'Approve automatically' is checked. Below this, a table lists transfer details for 'dadosabertos-root-810' with UUID '2deea79d-301f-4bb7-a430-6cc681f00795' and start time '2020-06-23 15:41'. A list of microservices follows, with 'Job: Move to select file ID tool' highlighted in green and marked as 'Completed successfully'. Other jobs include 'Job: Do you want to perform file format identification?' which is 'Awaiting decision'.

**Passo 5:** Com as etapas anteriores completas, clique em começar a transferência e acompanhe os microsserviços realizarem o pipeline de ingestão. Algumas etapas precisam de interação para confirmar uma tarefa ou não, enquanto isso o processo de transferência ficará parado aguardando o usuário decidir.

## REFERÊNCIAS

Artefactual Systems Inc. Documentação de instalação do Archivematica no CentOS/RedHat. 2020. Disponível em: <https://www.archivematica.org/en/docs/archivematica-1.11/admin-manual/installation-setup/installation/install-centos/#install-pkg-centos>. Acesso em: 23 de Julho de 2020.

Artefactual Systems Inc. Documentação do Archivematica. 2020. Disponível em: <https://www.archivematica.org/en/docs/archivematica-1.11/#home>. Acesso em: 23 de Julho de 2020.

Artefactual Systems Inc. Documentação de Administração do Storage Service. 2020. Disponível em: <https://www.archivematica.org/en/docs/storage-service-0.16/administrators/#administrators>. Acesso em: 23 de Julho de 2020.

Artefactual Systems Inc. Documentação de Transferência do Archivemática Dashboard. 2020. Disponível em:

<https://www.archivemática.org/en/docs/archivemática-1.11/user-manual/transfer/transfer/#transfer>. Acesso em: 23 de Julho de 2020.