

RELATÓRIO TÉCNICO Nº 02/2022

**Passo a Passo para implantação
do DSpace a partir do GIT**

PRESIDENTE DA REPÚBLICA

Jair Messias Bolsonaro

VICE-PRESIDENTE DA REPÚBLICA

Hamilton Mourão

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

Marcos Cesar Pontes

Ministro da Ciência, Tecnologia e Inovações

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

Cecilia Leite Oliveira

Diretora

Reginaldo de Araújo Silva

Coordenação de Administração – COADM

Gustavo Saldanha

Coordenação de Ensino e Pesquisa, Ciência e Tecnologia da Informação – COEPE

José Luis dos Santos Nascimento

Coordenação de Planejamento, Acompanhamento e Avaliação – COPAV

Anderson Itaborahy

Coordenação-Geral de Pesquisa e Desenvolvimento de Novos Produtos – CGNP

Bianca Amaro de Melo

Coordenação-Geral de Pesquisa e Manutenção de Produtos Consolidados – CGPC

Tiago Emmanuel Nunes Braga

Coordenação-Geral de Tecnologias de Informação e Informática – CGTI

Milton Shintaku

Coordenação de Tecnologias para Informação – COTEC



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
Instituto Brasileiro de Informação em Ciência e Tecnologia

RELATÓRIO TÉCNICO Nº 02/2022

Passo a Passo para implantação do DSpace a partir do GIT



Coordenação de Tecnologias
para Informação (COTEC)
Brasília
2022

EQUIPE TÉCNICA

Diretora do Instituto Brasileiro de Informação em Ciência e Tecnologia

Cecília Leite Oliveira

Coordenador-Geral de Tecnologias de Informação e Informática – CGTI

Tiago Emmanuel Nunes Braga

Coordenador do Projeto

Milton Shintaku

Normalização

Jaqueline Rodrigues de Jesus

Organizadores do Relatório

Lucas Ângelo da Silveira

Mirele Carolina Souza Ferreira Costa

Milton Shintaku

Revisão

Rafael Teixeira de Souza

Flávia Karla Ribeiro Santos

Pesquisadores

Deise Maria Antonio Sabbag

Fernanda Maciel Rufino

Fernando de Jesus Pereira

Jaqueline Rodrigues de Jesus

Lucas Ângelo da Silveira

Mirele Carolina Souza Ferreira Costa

Priscila Rodrigues dos Santos

Raíssa da Veiga de Meneses

Renata Monteiro Rodrigues

Rosilene Paiva Marinho de Sousa

Victor Ramos Silva

Capa e Projeto Gráfico

Victor Ramos Silva

Este Relatório Técnico é um produto do Projeto Estudos para implementação de Repositório Institucional na Biblioteca Terezinha Ayres Costa.

Ref. Processo SEI Nº 01302.000479/2018-80 (Processo de Contratação)

Ref. Processo SEI Nº 01302.000437/2020-63 (Processo de Execução)

Ref. FUNDEP - 28139

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto Brasileiro de Informação em Ciência e Tecnologia ou do Ministério da Ciência, Tecnologia, Inovações e Comunicações.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

Apresentação	06
1 Introdução	07
2 Implantação do DSpace	08
2.1 Restaurar <i>backup</i> do banco de dados	10
2.2 Restaurar <i>backup</i> da aplicação	11
2.3 Ativar e desativar o sistema	12
2.4 Teste da aplicação	12

Apresentação

Este guia é direcionado aos profissionais de informática e apresenta a instalação do software livre DSpace, versão 6.3 a partir dos códigos-fonte e do backup do banco de dados nos ambientes GIT e Drive do Ibict. O passo a passo especifica aspectos do sistema que apoiam o trabalho da equipe e tem o objetivo de descrever os comandos para a implantação da Biblioteca Digital da Agência Nacional de Vigilância Sanitária (Anvisa), considerando um ambiente de homologação que atenda aos requisitos de *hardware* e *software* descritos no Quadro 1:

Quadro 1 - Requisitos para a Biblioteca Digital.

Requisitos de *hardware*:

Sistema operacional Linux de 64 bits

Configuração mínima de 4 GB de memória RAM

Configuração mínima de 80 GB de disco rígido (HD)

Requisitos de *software*:

Java 8 com openJDK versão 1.8.0_171

Servidor web Apache Tomcat versão 8.5.69

Apache Solr

Software de apoio para baixar as dependências Apache Mavens versão 3.6.2

Software de apoio para baixar as dependências Apache Ant versão 1.10.7

Sistema de gerenciamento de banco de dados (SGBD) PostgreSQL versão 13.3

Fonte: Elaborado pelos autores (2022).

1 Introdução

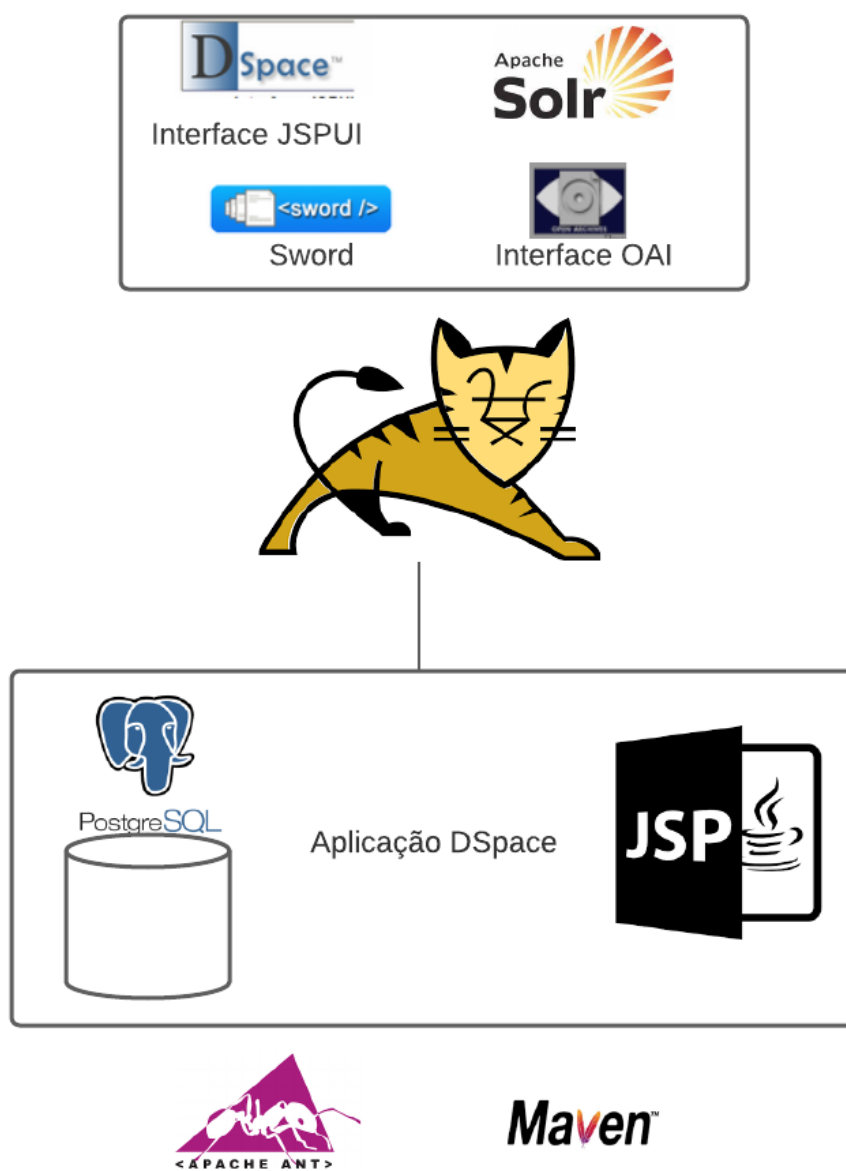
O DSpace é uma ferramenta de código aberto¹ originalmente desenvolvida pelo consórcio formado pela Biblioteca do Massachusetts Institute of Technology (MIT) e a Hewlett Packard (HP), tendo em vista a criação de repositórios institucionais ou temáticos no âmbito do Movimento de Acesso Aberto (Open Access), e mantida pela organização sem fins lucrativos DuraSpace². O *software* livre Dspace, mais utilizado no mundo para repositórios acadêmicos, também é muito requisitado no Brasil para a implementação de Bibliotecas Digitais em órgãos de governo de todas as esferas e poderes, em razão de sua capacidade de ofertar acesso ao texto completo e permitir a organização e descrição do acervo digital. Em vista disso, foi selecionado para a implementação da Biblioteca Digital da Anvisa.

O Dspace foi desenvolvido com a linguagem de programação Java, sendo assim, sugere-se fortemente o uso do servidor de aplicação web Apache Tomcat, com a interface de usuário JSPUI, baseada em tecnologias JavaServer Pages (JSP). O SWORD e Interface OAI são nativos, mas não precisam ser ativados, pois esse sistema não precisa interoperar com outros. O Apache Solr é utilizado para indexação, e melhora o processo de busca. Pode-se utilizar o gerenciador de banco de dados PostgreSQL ou Oracle, mas, neste caso, usamos o PostgreSQL, por ser uma ferramenta livre. A Figura 1 apresenta a arquitetura de sistema utilizada na implementação do DSpace:

¹ Código aberto significa que os programas fontes (editáveis) ficam disponíveis para serem baixados e alterados se preciso, contrapondo-se aos softwares de código fechado, que disponibilizam apenas os programas executáveis.

² Disponível em: <https://duraspace.org/>. Acesso em: 04 fev. 2022.

Figura 1 - Estrutura tecnológica do DSpace.



Fonte: Elaborado pelos autores (2022).

2 Implantação do DSpace

Os arquivos necessários à migração da Biblioteca Digital da Anvisa encontram-se no GIT e no Drive do Ibict. Desse modo, realize o download dos arquivos como descrito no Quadro 2:

Quadro 2 - Endereço dos arquivos de backup da Biblioteca Digital da Anvisa.

Nome do arquivo	Endereço
jspui	Front-end da aplicação: https://bit.ly/3vRYuq9
dspace-fonte	Código fonte da aplicação: https://bit.ly/3vRYuq9
dspace.cfg	Configuração da aplicação: https://bit.ly/3vRYuq9
bancoDSpace	Banco de dados: https://drive.google.com/drive/u/1/folders/1nfT9kVU2cN7q8fdJdl7yJY5RAs3Q6req
assetstore.tar.gz	Pasta assetstore: https://drive.google.com/drive/u/1/folders/1nfT9kVU2cN7q8fdJdl7yJY5RAs3Q6req
statdata.tar.gz	Estatísticas da aplicação: https://drive.google.com/drive/u/1/folders/1nfT9kVU2cN7q8fdJdl7yJY5RAs3Q6req

Fonte: Elaborado pelos autores (2022).

A instalação completa do DSpace consiste em três árvores de diretórios separados:

1. Diretório de origem, assumido como “dspace-fonte”: onde os arquivos de configuração são usados apenas durante o processo de instalação inicial. Ele possui o diretório “dspace-fonte/dspace”, que contém todas as informações de compilação e configuração do DSpace, além do arquivo “dspace-fonte/local.cfg.EXAMPLE”, que é usado para armazenar quaisquer alterações de configuração locais para instalação.
2. Diretório de instalação, assumido como “DspaceDir”: onde o DSpace está instalado e em execução. É o local definido no “dspace.cfg” em “dspace.dir” e onde todos os arquivos de configuração do DSpace, scripts de linha de

comando, documentação e webapps serão instalados. Enquanto é executado, esse diretório é preenchido durante o processo de instalação e também pelo DSpace. Ele contém arquivos de configuração, ferramentas de linha de comando e as bibliotecas necessárias para executá-los. Após a construção e instalação inicial, as alterações nos arquivos de configuração devem ser feitas no diretório “DspaceDir”.

3. Diretório de implementação da web: é gerado pelo servidor na primeira vez em que encontra um arquivo dspace.war localizado em “tomcat/webapps/dspace”. Sua criação é totalmente automática.

Para migração, tomemos o usuário “dspace” como usuário mantenedor do banco de dados do software DSpace e o diretório de instalação do DSpace como “dspaceDir”. Para criar o diretório de instalação do DSpace, ou seja, “dspaceDir”, utilize um usuário root (ou um usuário com permissões apropriadas):

```
$ mkdir dspaceDir
```

Para realizar a implantação do sistema atual, siga os passos das seções seguintes.

2.1 Restaurar backup do banco de dados

Acesse o Postgres e crie o banco de dados utilizando o usuário que irá manipular a base de dados do DSpace. Para isso, execute os comandos a seguir:

```
$ su postgres
$ sudo su
$ psql
postgres=# CREATE DATABASE novobanco ENCODING 'UTF-8' OWNER
dspace;
postgres=# \q
$ psql novobanco
postgres=# CREATE EXTENSION pgcrypto;
postgres=# \q
```

Para restaurar o backup do banco de dados, utilize o arquivo baixado no drive (neste, relatório usamos o nome “bancoDSpace”) e faça o *restore* do backup para o banco de dados criado. Para isso, execute o comando a seguir:

```
$ pg_restore -d novobanco bancoDspace
```

Em seguida, verifique se a restauração do banco de dados foi realizada corretamente, executando os seguintes comandos para listar as tabelas do banco de dados do DSpace:

```
$ su postgres  
$ sudo su  
$ psql novobanco  
postgres=# \d
```

2.2 Restaurar backup da aplicação

Para restaurar a pasta “assetstore”, que armazena o conteúdo físico, isto é, os documentos inseridos na biblioteca digital, mova o arquivo “assetstore.tar.gz” com o backup da pasta “assetstore” para dentro do diretório de instalação (neste, relatório usamos o nome “dspaceDir”) e descompacte o arquivo:

```
$ mv assetstore.tar.gz dspaceDir/
```

```
$ tar -vzxf assetstore.tar.gz
```

As estatísticas produzidas pelo DSpace referentes a acessos, número de documentos inseridos, downloads de documentos, entre outras informações, são armazenadas fora do banco de dados. Para restaurar o conteúdo de estatísticas, mova o arquivo “statdata.tar.gz” com o backup das estatísticas para o diretório “statistics”, localizado em “dspaceDir/solr/statistics”, e descompacte o arquivo com o seguinte comando:

```
mv statdata.tar.gz dspaceDir/solr/statistics  
$ tar -vzxf statdata.tar.gz
```

Para restaurar as configurações do DSpace, mova o arquivo “dspace.cfg” para dentro do diretório de instalação (neste, relatório usamos o nome “dspaceDir”).

```
mv dspace.cfg dspaceDir/config/
```

Para reindexar os itens no DSpace, execute os comandos abaixo no diretório de instalação (neste, relatório usamos o nome “dspaceDir”):

```
$ cd dspaceDir/bin
$ ./dspace index-discovery -f
$ ./dspace solr-reindex-statistics
$ ./dspace oai clean-cache
$ ./dspace oai import -c -v
```

Por fim, reinicie o servidor de aplicação web tomcat com o comando:

```
$ service tomcat restart
```

Com relação ao front-end, a nova instalação do DSpace será adequada em conformidade com o backup da interface de usuário “jspui”. Assim, mova o conteúdo da pasta “jspui” para “DspaceDir/webapp/jspui”. É importante ressaltar que esse processo deve exigir cuidados, visto ser preciso ir substituindo os arquivos e testando para não quebrar o sistema. Após seguir essas etapas, a Biblioteca Digital da Anvisa estará instalada na url indicada pelo servidor web e funcionando corretamente.

2.3 Ativar e desativar o sistema

Para ativar ou desativar o sistema, basta reiniciar o servidor web Apache Tomcat que torna disponível o acesso ao DSpace. Os comandos abaixo servem, respectivamente, para ativar, desativar ou verificar o status de execução do servidor tomcat:

```
$ sudo service tomcat start
```

```
$ sudo service tomcat stop
```

```
$ sudo service tomcat status
```

2.4 Teste da aplicação

Antes de testar a implantação do DSpace, verifique as diretivas abaixo. A título de exemplo, o nome de usuário e senha do banco de dados restaurado estão configuradas corretamente no arquivo de configuração e customização “dspace-fonte/dspace/config/local.cfg”:

db.username = nome de usuário do banco de dados, utilizamos “dspace”
db.password = senha do banco de dados, utilizamos “123”
dspace.dir = definição do diretório de instalação, utilizamos “caminho/dspaceDir”
dspace.baseUrl = URL completa da página inicial do DSpace do servidor (incluindo porta)
solr.server = URL completa do servidor Solr
db.url = URL JDBC completa para o banco de dados (exemplos são fornecidos no arquivo local.cfg.EXAMPLE)
db.dialect = Qual dialeto de banco de dados usar, utilizamos o PostgreSQL
db.schema = esquema de banco de dados a ser usado (os exemplos são fornecidos no local.cfg.EXAMPLE)

Realizados todos os passos anteriores, tem-se o DSpace pronto e disponível na URL indicada pelo servidor web. Para testar com o servidor web Tomcat disponível na porta 8080, navegue até o diretório “jspui”. Exemplo: <http://localhost:8080/jspui/.0> site da Biblioteca Digital da Anvisa será indexado conforme mostrado na Figura 2:

Figura 2 - Tela do site da Biblioteca Digital da Anvisa.



Fonte: Elaborado pelos autores (2022).

