

ACESSO ABERTO A DADOS DE PESQUISA NO BRASIL

Dataverse: Protocolo de Migração



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

Ana Maria Mielniczuk de Moura

Caterina Groposo Pavão

Paula Caroline Schifino Jardim Passos

Rene Faustino Gabriel Junior

Samile Andréa de Souza Vanz

Pesquisadores - Universidade Federal do Rio Grande

Eduardo Nunes Borges

Luís Alberto Barbosa Azambuja

Bolsistas - Universidade Federal do Rio Grande do Sul

Iván Andrés Fornos Angues

Victor Andrews Garcia Lima

Bolsistas - Universidade Federal do Rio Grande

Adornete de Almeida Martins Junior

Michael Lemos da Costa

Contato

e-mail: 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: AZAMBUJA, Luís Alberto Barbosa; BORGES, Eduardo Nunes; GABRIEL JUNIOR, Rene Faustino; MARTINS JUNIOR, Adornete de Almeida; COSTA, Michael Lemos da; ROCHA, Rafael Port da; CAREGNATO, Sônia Elisa; PASSOS, Paula Caroline Schifino Jardim; PAVÃO, Caterina Groposo; VANZ, Samile Andrea de Souza; MOURA, Ana Maria Mielniczuk de. Acesso aberto a dados de pesquisa no Brasil: Dataverse: protocolo de migração.

Resumo

Este relatório apresenta um protocolo de migração e atualização de software para o Dataverse. São apresentados conjuntos de procedimentos gerais para testar as modificações realizadas, implantar as atualizações necessárias e validar a migração de versão de *software*. Estas mudanças incluem a aplicação de correções, instalação de novos pacotes de idiomas, funcionalidades e recursos.

Palavras-chave: Dataverse. Dados abertos. Dados de pesquisa. Migração. Atualização. Segurança.

Sumário

1. Introdução	5
2. Protocolo para atualização do repositório de dados	5
2.1. Conjunto de procedimentos para testar as alterações	6
2.2. Conjunto de procedimentos para implantar as alterações	6
2.3. Conjunto de procedimentos para verificar o funcionamento do Dataverse	7
3. Atualizando versões do Dataverse	8
4. Considerações em relação a instalações sem o uso de VM	9

1. Introdução

O Dataverse é composto de vários componentes e serviços responsáveis por funcionalidades distintas no repositório de dados. Cada um destes componentes pode necessitar de atualizações, seja por melhorias, correções de falhas ou segurança. Essa é uma característica inerente a todo *software*, portanto a sua atualização deve ser prevista e planejada para poder ser realizada com o mínimo de risco e o máximo de eficiência.

Este roteiro pretende auxiliar o planejamento desta atualização, assim como de qualquer mudança que impacte ou altere significativamente um repositório baseado em Dataverse. Estas mudanças incluem a aplicação de correções, instalação de novos pacotes de idiomas, funcionalidades e recursos.

2. Protocolo para atualização do repositório de dados

O Dataverse pode ser instalado em diversos ambientes, tanto virtualizados quanto em servidores físicos dedicados, utilizando *containers* ou não. É possível também instalar em múltiplos sistemas operacionais. Essa diversidade traz uma grande liberdade aos profissionais de TI, mas também aumenta as preocupações quanto à segurança e ao funcionamento adequado.

Os procedimentos abaixo são indicados para instalações em ambientes virtualizados e devem ser usados como referência para locais que não tenham uma rotina própria de manutenção e atualização. O primeiro conjunto de procedimentos testa as alterações realizadas em uma máquina virtual independente. O segundo bloqueia a submissão e publicação de conjuntos de dados temporariamente na máquina em produção, realiza as alterações em uma máquina clonada, testa os resultados e por fim libera o acesso bloqueado.

2.1. Conjunto de procedimentos para testar as alterações

Clonar a máquina virtual (VM). Na VM clonada:

- Alterar a configuração de rede, atribuindo um IP válido com acesso externo e mesmas permissões de firewall da VM original. Este passo é necessário para poder acessar a rede externa, baixar atualizações e testar o acesso Web.
- Aplicar as correções e atualizações necessárias.
- Verificar as funcionalidades atualizadas.
- Verificar o funcionamento geral de todo o ambiente do repositório.

Caso problemas tenham sido identificados, identificar os erros e repetir todos os procedimentos.

2.2. Conjunto de procedimentos para implantar as alterações

Após ter certeza do correto funcionamento do Dataverse com as alterações testadas:

- Interromper a submissão, publicação e quaisquer outras alterações no ambiente do repositório, tais como a criação de usuários, atribuição de permissões, edição de datasets e dataverses.
- Clonar a VM e efetuar todas os procedimentos abaixo na VM clonada. Desta forma, a VM original ficará preservada e funcionando para consulta e recuperação de dados e informações durante o processo de atualização.
- Alterar a configuração de rede, atribuindo um IP válido com acesso externo e mesmas permissões de firewall da VM original. Este passo é necessário para poder acessar a rede externa, baixar atualizações e testar o acesso Web.
- Aplicar as correções e atualizações necessárias.
- Verificar as funcionalidades atualizadas.

- Verificar o funcionamento geral de todo o ambiente do repositório.
- Interromper o acesso às VMs original e clonada.
- Desligar e preservar a VM original.
- Alterar a configuração de rede (IP, host etc.) da VM clonada já atualizada, aplicando a configuração da VM original.
- Liberar submissão, publicação e alterações no repositório.

2.3. Conjunto de procedimentos para verificar o funcionamento do Dataverse

Toda atualização de *software*, seja para corrigir, melhorar ou acrescentar funcionalidades, mesmo as mais simples, podem afetar outras partes do sistema. Portanto, é imprescindível testar o ambiente do sistema para confirmar o perfeito funcionamento do Dataverse. Algumas correções e alterações envolvem mais de um componente de *software* e afetam mais de um recurso. Estas alterações precisam de um cuidado especial para garantir a integridade do sistema.

Um checklist completo dos procedimentos a serem testados precisa ser formalmente elaborado pela equipe técnica e ratificado pela curadoria e pela administração do repositório. Algumas das principais funções a serem testadas são listadas a seguir:

- Autenticação por todos modos disponíveis, tais como conta local, provedor de identidade institucional e ORCID.
- Criação de usuários.
- Alteração de senhas.
- Envio automático de e-mails, inclusive na troca de senha.
- Criação e exclusão de *roles*.
- Atribuição e exclusão de *roles* para usuários.
- Criação de dataverses (comunidades) e datasets.
- Submissão de conjuntos de dados.

- Publicação de conjuntos de dados.
- Pesquisa simples e avançada.
- Download de conjuntos de dados e de citação.
- Manipulação dos conjuntos de metadados.
- Tradução do sistema (todas os idiomas instalados).

3. Atualizando versões do Dataverse

Ao atualizar as versões do Dataverse é importante ler a página que descreve detalhes das releases¹. Ela contém a documentação dos novos recursos,

¹ <https://github.com/IQSS/dataverse/releases>

aprimoramentos e correções de erros de cada nova versão, além de incluir roteiros necessários para a atualização. A recomendação é sempre a de instalar todas versões de atualização em sequência. Por exemplo, se a intenção é atualizar da versão 4.16 para 4.19, deve ser instalada primeiro a *release* 4.17, então a 4.18, depois a 4.18.1 para só então instalar a 4.19.

4. Considerações em relação a instalações sem o uso de VM

Em instalações onde não esteja disponível o recurso de máquinas virtuais, e uma ou mais máquinas físicas dedicadas disponibilizem os serviços do repositório, é necessário um cuidado especial em preservar o ambiente antes de fazer qualquer mudança substancial no *software*. Isto pode ser feito com procedimentos automatizados ou manuais de *backup*, ou até mesmo com a clonagem completa dos discos rígidos.

Dependendo da atualização, além de preservar os conjuntos de dados armazenados no repositório, é necessário efetuar *backup* de configurações específicas do sistema operacional e do Sistema Gerenciador de Banco de Dados, e de vários componentes de *software* que integram o Dataverse. O inconveniente deste tipo de procedimento é a indisponibilidade do serviço de repositório já que certos *backups* exigem a parada dos sistemas.