

DATAVERSE 5



REPOSITÓRIO DE DADOS DE INVESTIGAÇÃO

**CARATERÍSTICAS E REQUISITOS
TÉCNICOS**

SOBRE O DOCUMENTO

Identificador	WP222062.5		
Aprovado por	Luís Miguel Ferros	Aprovado em	2023-05-11
Classificação	Público		
Distribuição	N/A		

REVISÕES

#	Data	Autor	Alterações
1	2022-11-17	José Carvalho	Documento inicial
2	2022-12-29	Márcia Fernandes	Revisão do documento
3	2022-01-20	Luís Miguel Ferros	Revisão do documento
4	2022-02-14	José Carvalho	Reformulação em três áreas
5	2023-05-11	Luís Miguel Ferros	Revisão final

SUMÁRIO EXECUTIVO

O software Dataverse tem como missão dar suporte a todas as áreas funcionais de um repositório de dados de investigação, desde a submissão de trabalhos até à sua publicação.

Este documento tem como objetivo descrever as principais características, funcionalidades e vantagens do software Dataverse, bem como os requisitos técnicos necessários à sua correta instalação.

**REPOSITÓRIO DE
DADOS DE
INVESTIGAÇÃO
PARA DIVULGAR E
PARTILHAR OS
DADOS
PRODUZIDOS
PELOS
INVESTIGADORES**

DATAVERSE

O software Dataverse tem como objetivo a gestão de todas as áreas funcionais de um repositório de dados de investigação, desde a submissão de trabalhos até à sua publicação.

O Dataverse possui um conjunto de mais-valias para os diferentes intervenientes de uma organização. Para os investigadores, permite-lhes ter total controlo sobre os seus dados, obter visibilidade na Web, crédito académico e aumentar as citações dos seus dados de investigação. Os grupos ou projetos de investigação podem criar e personalizar facilmente a sua área e disponibilizar um ponto de referência para os seus dados. Permite ainda ao investigador suportar os seus planos de gestão de dados necessários para candidaturas a projetos de financiamento.

Se possui uma revista científica pode promover o movimento *open data* através da submissão dos dados de investigação no repositório, os quais ficarão interligados com a publicação.

Este sistema permite disponibilizar um único ponto de acesso aos dados produzidos pela sua organização, permitindo a partilha, preservação, citação, exploração e análise por toda a comunidade internacional.

O Dataverse resulta da cooperação internacional da comunidade científica para garantir uma infraestrutura aberta para a ciência aberta, garantindo o cumprimento de boas práticas e a definição de planos de gestão de dados de acordo com as orientações dos projetos de financiamento e das políticas nacionais de partilha de dados.

Sem custos de licenciamento

Tratando-se de um software open-source, não existem custos de licenciamento associados ao produto Dataverse, o qual funciona sobre um sistema operativo open-source, o que permite, também, uma redução de custos ao nível da infraestrutura e software de suporte.

Princípios FAIR DATA

O software Dataverse suporta os 4 princípios FAIR DATA (Findable, Accessible, Interoperable and Reusable), garantindo, assim, o cumprimento dos requisitos basilares da Ciência Aberta.

Os seus dados, onde quiser

Pode optar por diferentes formas de alojamento dos seus dados, quer seja localmente ao nível dos serviços da sua organização ou de forma remota em infraestruturas externas dimensionadas para as características deste tipo de informação.

Citação de dados

Sendo as citações uma das motivações dos investigadores para a partilha dos seus dados de investigação, o Dataverse exhibe em cada conjunto de dados uma referência bibliográfica para inserir diretamente nos trabalhos que a referenciam, integrando ainda os serviços de gestão de referências bibliográficas como o Zotero ou Endnote.

Repositório versátil

Esta solução permite a definição de workflows específicos em diferentes áreas funcionais do repositório. No processo de submissão de dados pode optar por outras formas de integração sem ser o usual carregamento de ficheiros manual.

Pode também criar conjuntos de dados que estão fisicamente noutra infraestrutura, mas são registados, descritos e apresentados como se estivessem localmente disponíveis. Esta funcionalidade permite à organização criar um catálogo de dados próprio, mesmo com datasets que integram outros repositórios ou infraestruturas.

Adicionalmente, pode agregar para uma coleção do seu repositório, através de um interface OAI-PMH, conteúdos que estão disponíveis noutro repositório para a sua coleção de forma automática.

Finalmente, é possível a visualização de dados tabulares ou CSV diretamente no repositório, sem necessidade de descarregar e abrir localmente no seu dispositivo.

Estes casos de uso advêm das necessidades da comunidade e foram desenvolvidas pela mesma para suportar os seus processos na área de gestão de dados.

ARQUITETURA E MÓDULOS FUNCIONAIS

Este software é constituído por 5 módulos aplicativos que procuram ir ao encontro das necessidades de gestão do repositório de dados:

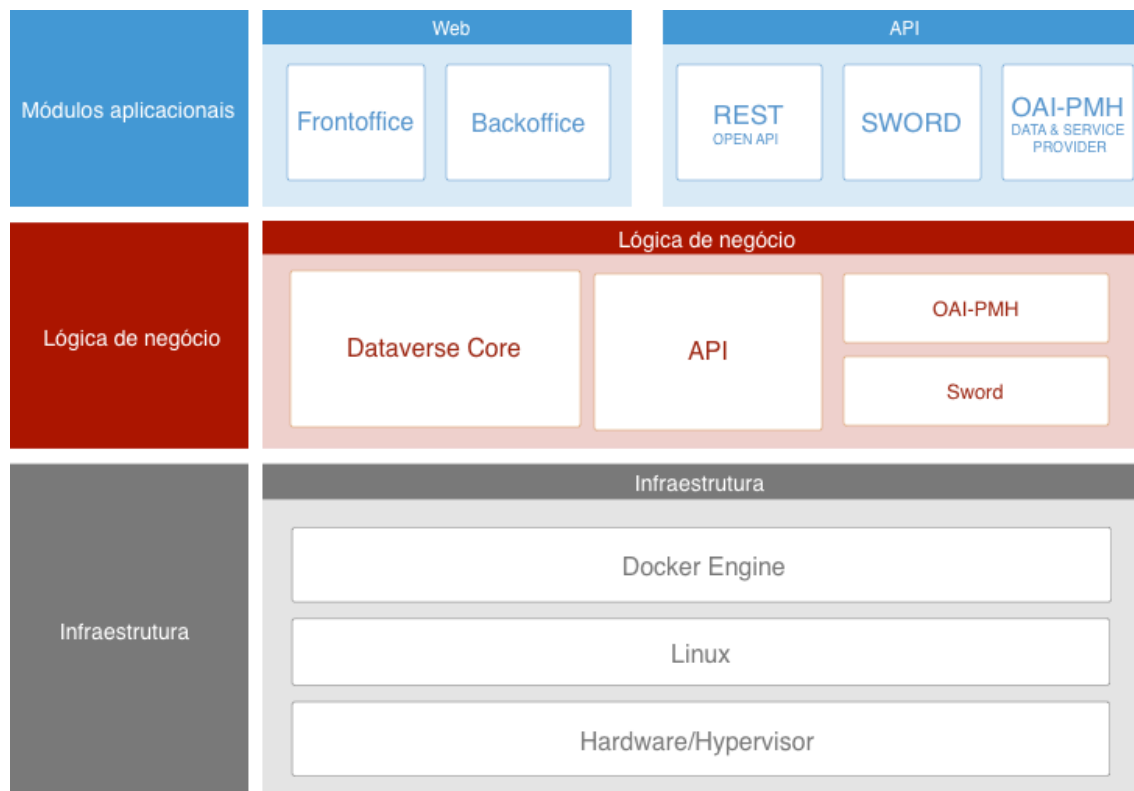


Figura 1 - Arquitetura do Dataverse

PORTAL PÚBLICO (FRONTOFFICE)

O interface público é responsável por disponibilizar toda a informação dos registos de dados do repositório para os utilizadores e serviços de indexação. Poderá pesquisar ou listar toda a informação para aceder aos conjuntos de dados, a sua descrição e visualizar ou descarregar os ficheiros. O acesso aos dados poderá ser totalmente aberto, fechado, ou através de mecanismos de registo de utilização, como o livro de visitas ou uma ligação privada para o acesso exclusivo.

O sistema fornece automaticamente o formato de citação para referenciar o conjunto de dados ou um ficheiro em particular, promovendo as boas práticas.

O acesso e registo de utilizadores pode ser efetuado através de contas locais ou de sistemas externos como o ORCID ou Shibboleth.

Identificadores persistentes

O Dataverse efetua uma reserva do DOI (Digital Object Identifier System) e regista toda a informação na DATACITE assim que os dados são publicados. O DOI fica disponível na referência bibliográfica para garantir a persistência a longo termo.

Estatísticas de uso

O Dataverse regista estatísticas de uso sobre os diferentes conjuntos de dados, permitindo medir o impacto quer do repositório no seu todo quer dos dados em particular. Integra a iniciativa internacional Make Data Count para partilhar de forma transparente dos dados do seu repositório.

Autenticação federada

O repositório Dataverse permite a configuração de várias formas de autenticação, como diretórios locais usando Shibboleth (SSO), serviços externos como Google, ORCID ou Github (OAuth2) e ainda através de serviços OpenID.

Proteção de dados

O Dataverse permite ir ao encontro das orientações e normativos da proteção de dados, nomeadamente do Regulamento Geral sobre Proteção de Dados (RGPD).

Adaptável a qualquer dispositivo e à identidade da organização

O interface gráfico, disponível em vários idiomas, é configurável e personalizável, sendo ainda responsivo, o que permite consultar os dados a partir de qualquer dispositivo móvel.

GESTÃO (BACKOFFICE)

A área de gestão está restrita a utilizadores com permissões para poderem adicionar novos conjuntos de dados ou para a gestão de Dataverses. Os Dataverses são coleções que podem ter uma estrutura hierárquica para a organização do seu repositório em, por exemplo, faculdades ou projetos de investigação. Cada Dataverse é especificamente configurável, através de uma descrição, design gráfico personalizado, assim como a respetiva gestão de permissões.

Para suportar o ciclo de vida e a curadoria dos conjuntos de dados, o Dataverse permite manter um registo em rascunho até que estejam reunidas todas as informações necessárias para ser publicado. Neste processo, o conjunto de dados pode ser validado por vários elementos, revisto, e definidos os termos de acesso através de licenças e livro de visitas. O livro de visitas permite obter uma opinião do utilizador antes deste ter acesso ao conjunto de dados, permitindo por exemplo obter o contacto ou a finalidade de utilização dos dados.

Integração no ciclo de vida dos dados

O repositório permite a integração no ciclo de vida dos dados, parametrizando os workflows, versões, descrições e toda a definição de licenças e tipo de acesso necessários à sua publicação.

Estatísticas de uso

O Dataverse regista estatísticas de uso sobre os diferentes conjuntos de dados, permitindo medir o impacto quer do repositório no seu todo quer dos dados em particular. Integra a iniciativa internacional Make Data Count para partilhar de forma transparente a relevância dos dados do seu repositório.

Versionamento

Garanta o registo das alterações efetuadas ao longo do tempo sobre os seus conjuntos de dados, através da criação de versões totalmente identificadas, registadas e descritas para quem consulta o seu repositório de dados.

Controlo total

Tenha total controlo dos acessos, através das funcionalidades disponibilizadas pelo Dataverse. Pode optar por partilhar de forma privada os conjuntos de dados, usar "guest books" para saber quem usou os seus dados ou simplesmente bloquear ou definir um embargo.

Descrição pormenorizada dos dados de investigação

Opte por usar esquemas de metadados normalizados, predefinidos de acordo com as práticas da comunidade internacional, ou adote os seus próprios esquemas para maximizar a descrição dos conjuntos de dados e facilitar a sua organização e recuperação.

ADMINISTRAÇÃO

A administração permite a gestão dos acessos e permissões dos utilizadores, definindo o papel que cada utilizador pode desempenhar no sistema. Poderá ainda mover conjuntos de dados entre diferentes Dataverse para manter o seu repositório devidamente organizado. O Dataverse possui simultaneamente um servidor OAI-PMH e a possibilidade de agregar conteúdos de clientes externos através deste protocolo.

O servidor OAI-PMH do Dataverse permite expor toda a informação disponível publicamente no seu repositório para outros serviços externos que poderão agregar periodicamente os metadados. Além da interface geral que partilha todos os dados, podem ser criados conjuntos de dados (*sets*) totalmente personalizados, como por exemplo apenas os conjuntos de dados de um determinado autor, departamento, projeto, etc.

Para poder integrar novos dados alojados externamente, pode configurar diferentes fontes de dados que sejam compatíveis com o protocolo OAI-PMH. Cada novo cliente de OAI-PMH pode ser atualizado manualmente, diariamente ou semanalmente. Cada um deles pode ter formatos de metadados distintos ou agregar apenas uma parte dessa fonte de dados (*set*).

Disseminação

Com esta solução é possível expor esquemas de metadados através do protocolo OAI-PMH de acordo com formatos standard como Dublin Core, Data Document Initiative (DDI), OpenAIRE, e outros. Possui ainda várias API's para pesquisa, submissão, acesso, migração ou métricas. Suporta JSON-LD para integrar, por exemplo, com o Google Dataset Search, entre outros serviços.

Integrações externas

O sistema Dataverse foi desenvolvido a pensar na integração de outras ferramentas específicas utilizadas na gestão de dados de investigação. Como exemplo, pode integrar o repositório de dados com Dropbox, Github, Jupyter Notebooks, vocabulários controlados (API/SKOSMOS), R. Rsync, entre outros.

Organize à sua maneira

Defina uma estrutura personalizada através de comunidades e coleções com a hierarquia necessária para fazer face às suas necessidades. Além da estrutura, pode ainda personalizar cada comunidade para ir ao encontro das suas necessidades, como colocar o logótipo do projeto ou do grupo de investigação.

ACESSIBILIDADE DE CONTEÚDOS WEB

As Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0 são um conjunto de recomendações emanadas pela W3C que visam tornar o conteúdo Web mais acessível. O cumprimento destas diretrizes faz com que o conteúdo publicado na Web se torne acessível a pessoas com incapacidade, nomeadamente cegueira e baixa visão, surdez e baixa audição, dificuldades de aprendizagem, limitações cognitivas, limitações de movimentos, incapacidade de fala, fotossensibilidade, entre outros. Seguir estas diretrizes permite também que o conteúdo Web se torne mais usável por utilizadores em geral e por dispositivos móveis como *smartphones*, *tablets* ou relógios de pulso.

Dada a importância desta temática, foi criada legislação que visa promover a adoção destas diretrizes por todo o Estado Português. O Regulamento Nacional de Interoperabilidade Digital (RCM n.º 91/2012 de 8 de novembro) esclarece a obrigatoriedade do cumprimento das diretrizes WCAG 2.0 nível AA por todos os sítios Web do Estado que disponibilizem serviços online¹. O artigo 2º da lei n.º 36/2011 determina que a lei se aplica a:

- Órgãos de soberania;
- Serviços da administração pública central, incluindo institutos públicos e serviços desconcentrados do Estado;
- Serviços da administração pública regional;
- Sector empresarial do Estado.

A KEEP SOLUTIONS é solidária com esta iniciativa e assegura que todos os seus produtos estão em plena conformidade com o nível AA+ das Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0.

¹ Ver também Lei n.º. 36/2011, de 21 de junho. Estabelece a adoção de normas abertas nos sistemas informáticos do Estado (PDF, 188KB); RCM n.º 91/2012 de 8 de novembro. Regulamento Nacional de Interoperabilidade Digital (PDF, 16,1MB); DGAEP. (Fevereiro de 2013). Estrutura orgânica da AP Portuguesa pós-PREMAC (PDF, 1,8MB); DGTf. (30/09/2012). Definição de Sector empresarial do Estado (PDF, 102KB).

REQUISITOS TÉCNICOS

O servidor será responsável por alojar a base de dados que contém a informação produzida pelos utilizadores, bem como o servidor Web que garantirá o acesso à informação aos utilizadores externos ao repositório.

A secção que se segue descreve os requisitos mínimos necessários para o correto funcionamento da aplicação.

SERVIDOR DE BASE DE DADOS E APLICACIONAL

RAM	8 GB 16 GB recomendados
CPU	Intel Quad-Core ou superior
Disco	500 GB em LVS Depende do volume de dados e da sua taxa de crescimento
Sistema operativo	Linux LTS (Debian, Ubuntu)
Software	Docker Engine
Rede	100 Mbit/s ou superior 1 Gbit/s é recomendável em contexto de incorporação de grandes volumes de dados

 www.keep.pt
 +351 253 066 735
 info@keep.pt
 sales@keep.pt
 KEEP SOLUTIONS, LDA.
Rua Rosalvo de Almeida, n° 5,
4710-429 Braga
Portugal

KEEP SOLUTIONS

A KEEP SOLUTIONS é uma empresa que tem como missão o fornecimento de soluções avançadas para gestão e preservação de informação.

A nossa abordagem assenta no fornecimento de sistemas de informação e na prestação de serviços com vista a permitir aos nossos clientes realizar uma gestão mais eficiente dos seus ativos de informação.

A empresa iniciou a sua atividade em 2008, tendo adquirido o estatuto de spin-off académica da Universidade do Minho, por se tratar de uma iniciativa empresarial com fortes laços de cooperação com centros de investigação e departamentos desta instituição.

Os nossos clientes encontram-se principalmente no setor público, nos domínios cultural, educacional, patrimonial e científico, mais concretamente nas áreas de arquivo, biblioteca e museu.

Apostamos no desenvolvimento contínuo de soluções inovadoras. Para isso, permanecemos ativos na produção de conhecimento científico, participando ativamente em projetos de I&D em cooperação com instituições nacionais e internacionais.