



SEI-SICITE 2021

Pesquisa e Extensão para um mundo em transformação

XI Seminário de Extensão e Inovação  
XXVI Seminário de Iniciação Científica e Tecnológica  
08 a 12 de Novembro - Guarapuava/PR



## Características para implantação de uma plataforma para facilitar o acesso aos dados do Repositório Brasileiro Livre para Dados Abertos do Solo

### Characteristics for implantation from a platform for easy access to data for Brazilian Free Repository for Open Soil Data

**Marcos Alexandre dos Anjos**

[marcosanjos@alunos.utfpr.edu.br](mailto:marcosanjos@alunos.utfpr.edu.br)

Discente do 7º período do Curso de Ciência da Computação. Universidade Tecnológica Federal do Paraná, Santa Helena, PR, Brasil.

**Alessandro Samuel-Rosa**

[alessandrrosa@utfpr.edu.br](mailto:alessandrrosa@utfpr.edu.br)

Docente do Curso de Agronomia, Universidade Tecnológica Federal do Paraná, Santa Helena, PR, Brasil.

#### RESUMO

Neste artigo apresentamos as principais características essenciais para a implantação do repositório de dados (RD) com o software DataVerse para o Repositório Brasileiro Livre para dados Abertos do Solo (FEBR). Os métodos e procedimentos constituiu numa pesquisa bibliográfica sobre princípios de repositório de dados, esquemas de metadados, identificadores persistentes, protocolos e certificações. No que diz respeito sobre o software de catalogação de metadados Dataverse, apresenta-se como uma solução completa seguindo padrões internacionais de protocolos. Assim concluímos que a recomendação apropriada para FEBR é adoção do software DataVerse juntamente com os protocolos e princípios para garantir a qualidade e organização dos dados. Sendo que isso inclui software para gerenciamento de metadados DataVerse, esquema de metadados, protocolo de interoperabilidade e padrão para identificador único persistente.

**PALAVRAS-CHAVE:** Metadados, DataVerse, Repositório de Dados.

#### ABSTRACT

In this article we present the main characteristics for implantation of a data repository with software DataVerse for the Brazilian Free Repository for Open Soil Data (FEBR). The methods and procedures consisted of a bibliographic search on research about data repositories, metadata scheme and persistent identifier and protocols and certifications. With regard to data and metadata to software Dataverse present as an alternative complete for repository data. We concluded then that the right recommendation for FEBR it's start using DataVerse software along the protocols and principles to ensure the quality and data organization. That includes DataVerse software for metadata management, metadata scheme, interoperability protocol and unique persistent identifier.

**KEYWORDS:** Metadata, DataVerse, Repository data.



SEI-SICITE 2021

Pesquisa e Extensão para um mundo em transformação

## INTRODUÇÃO

A Ciência do Solo apresenta um viés agrônomo histórico que justifica a denominação sobre uma visão do solo como substrato para o crescimento vegetal e sua importância ambiental (SILVA et al., 2019). Muitos estudos nesta área vêm sendo realizados na tentativa de compreender estas informações. Este volume de dados foi impulsionado pela internet que facilitou o acesso e disponibilização de troca de informações e dados (BIZER; HEATH; BERNERS-LEE, 2009). Os Objetivos do Desenvolvimento Sustentável (ODS) reconheceram a importância do uso sustentável do solo, quantidade e qualidade dos dados e informações do solo.

Os primeiros sinais de um movimento de dados científicos abertos vieram de Paul Ginsparg, em 1991, que criou um servidor de dados do laboratório de pesquisa para que os pesquisadores pudessem depositar os dados (SOUZA, 2012). Isso impulsionou o compartilhamento de dados de acesso aberto juntamente com este movimento iniciou-se o desenvolvimento de plataformas para compartilhamento de dados. Existem diversas plataformas para compartilhamento de dados mas o Registry of Data Repositories (RE3DATA, 2020) é um registro, que lista os repositórios de dados de pesquisa que abriga diversas áreas do conhecimento, sendo um registro global que oferece informações detalhadas sobre diversos repositórios de dados. Hoje o Brasil é referência em compartilhamento de dados e informações do solo em regiões tropicais (CAMARGO; ALVAREZ V.; BAVEYE, 2010).

A Ciência do Solo, especificamente, apresenta um déficit na organização e padronização dos dados das pesquisas. Muitas vezes, os dados são compartilhados através de arquivos em formato PDF, tais arquivos dificultam a reusabilidade dos dados, aumentando significativamente o esforço de extração desses dados. Em 2016 foi criado o Repositório Brasileiro Livre de Dados Abertos do Solo (FEBR), que armazena os dados de pesquisa sobre os solos brasileiros. Seu diferencial é adoção de padrões para preservação, padronização e versionamento dos dados depositados, facilitando sua reusabilidade pela comunidade. Hoje o FEBR é o maior repositório de dados da pesquisa em ciência do solo no Brasil, com dados de mais de 20 mil locais em todo o Brasil. Dessa forma o FEBR precisa estar atualizado com desenvolvimento de padrões internacionais de interoperabilidade de dados.

Assim o FEBR tem como objetivo adotar um sistema para o gerenciamento dos metadados, mas que cumpre com algumas características que podem ser definidos padrões de metadados e interoperabilidade. Padrões de metadados descrevem como os dados foram produzidos, e o padrão de interoperabilidade possibilita a troca de informação entre a rede de repositórios global.

O objetivo tem como objetivo, a partir de uma profunda revisão bibliográfica e estudo de repositórios nacionais e internacionais de dados da pesquisa, apresentando pontos para aumentar a interoperabilidade e confiabilidade do FEBR.



## MATERIAIS E MÉTODOS

A pesquisa foi embasada em artigos científicos e documentações oficiais sobre alguns protocolos, na literatura foram encontrados poucos trabalhos relacionados ao tema de metadados. Avaliação dos softwares de gerenciamento de metadados foram realizadas uma consulta no re3data.org, onde que os três softwares mais utilizados são CKAN, DSpace e DataVerse. Também estamos acompanhando o webinar Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), sobre repositórios de dados de pesquisa.

O Comprehensive Knowledge Archive Network (CKAN) foi desenvolvido pela Open Knowledge Foundation com o objetivo de compartilhamento e visualização de dados. O DataVerse foi desenvolvido pela universidade de Harvard com o objetivo de compartilhamento, preservação e exploração de conjuntos de dados. O DSpace surgiu de uma demanda da biblioteca do Massachusetts Institute of Technology (MIT) e dos laboratórios da Hewlett-Packard com o propósito de desenvolvimento de repositórios e bibliotecas digitais.

A avaliação dos padrões de metadados e interoperabilidade foi realizada na forma de revisão bibliográfica. Os padrões garantem acesso e confiabilidade dos conjuntos de dados abertos, que estejam num lugar seguro e disponível. Sendo que uma boa prática deve ser escolhido o software para gerenciamento de metadados em sequência avaliar os protocolos que apresentam suporte. Para validar confiabilidade de um repositório de dados podemos submeter uma auditoria para *DSA-WDS Core Trustworthy Data Repositories Requirements (TRUST)*. É um órgão regulamentador de repositórios onde o repositório precisa cumprir 17 requisitos para obter o certificado.

## RESULTADOS E DISCUSSÕES

A pesquisa tem como primeira atividade analisar as características dos softwares de gerenciamento de dados de modo a escolher um software para o gerenciamento dos dados no repositório FEBR. Tendo em vista esse objetivo, foram apontadas algumas características principais, relacionadas ao propósito de cada software de gerenciamento de metadados, das quais julga serem relevantes para repositório FEBR, podem ser observadas no Quadro 1.

Quadro 1 – Objetivos sobre cada software de gerenciamento de metadados.

Software	Objetivo
CKAN	Repositório de Dados principal foco está na visualização dos dados, apresentando plugins e interface moderna. Seu uso predominante está relacionado à abertura dos governos.
DSpace	Repositório de Dados seu principal uso está relacionado a bibliotecas digitais, para publicações de revistas, teses, e produções acadêmicas.
DataVerse	Repositório de Dados usado está na preservação dos dados de pesquisa, conta com ferramentas de versionamento de dados.

Fonte: Autoria própria (2021).

Os softwares de gerenciamento de metadados apresentam algumas características semelhantes mas cada software apresentado foi construído para um propósito. Dessa forma, podemos avaliar que software Dataverse apresenta características as quais são as necessidades do FEBR. Pesquisadores da Rede Nacional de Pesquisa (RNP), Universidade Federal do Rio Grande do Sul (UFRGS) estão trabalhando com



construção de repositório para Acesso Aberto a Dados de Pesquisa (AADP). Onde esses pesquisadores apresentaram conclusões semelhantes na escolha do software de gerenciamento de metadados DataVerse.

Segunda parte da pesquisa foram avaliados princípios, protocolos para garantir a confiabilidade do repositório. No quadro 2, podemos observar os detalhes dos itens avaliados.

Quadro 2 – Avaliação itens para confiabilidade repositório de dados.

Item	Descrição
FAIR	Os princípios FAIR que significa <i>Findable, Accessible, Interoperable</i> e <i>Reusable</i> . FAIR aderiu uma solução para construção de backbone, ou seja uma rede de comunicação para uso e distribuição dos dados pela <i>World Wide Web</i> através da rede de computadores (DATA FAIRPORT, 2014).
TRUST	Modelo de auditoria DSA-WDS Core Trustworthy Data Repositories Requirements (TRUST), constituído por 16 requisitos, podendo o repositório obter um certificado nível básico caso consiga cumprir todos os requisitos (MÉNDEZ RODRÍGUEZ, 2002).
Esquema Metadados	Os esquemas de metadados para domínio na Web representam as características e atributos de um objeto real com intuito de identificar um modelo para posteriormente realizar a recuperação (ALVES, [s.d.]). A representação esquema metadados é codificada com linguagem de marcação <i>eXtensible Markup Language</i> (XML).
Identificador Regular Persistente	O modelo de padrão dos metadados necessita apresentar um identificador único para suprir as necessidades dos protocolos de comunicação (“Open Archives Forum. Implementing oai-pmh”, [s.d.]) . Dessa forma, cada conjunto de dados no repositório deve conter um identificador único usado para requisição dos métodos de uma informação. A padronização dos metadados devem ser persistentes e identificáveis, sendo assim, um item importante dos princípios do FAIR.
Protocolo de Interoperabilidade	Podemos entender que interoperabilidade é a capacidade de sistemas da tecnologia da informação (TIC) e todos seus processos se comunicarem, auxiliando a troca de informações (ARAÚJO, 2012). Open Archives Initiative (OAI) um protocolo de comunicação que surgiu a partir de uma necessidade da comunicação entre bases de repositórios estabelece um modelo padrão de comunicação via redes de computadores (GARCIA; SUNYE, 2003)

Fonte: Autoria própria (2021).



SEI-SICITE 2021

Pesquisa e Extensão para um mundo em transformação

## CONCLUSÃO

Este estudo demonstrou que existem soluções tecnológicas que atendem às necessidades do FEBR como por exemplo o DataVerse. E temos protocolos que conversam com essas tecnologias garantindo a interoperabilidade entre os repositórios. Dessa forma, seguindo as demandas de cada protocolo, podemos garantir a confiabilidade do repositório. Assim concluímos que a recomendação mais apropriada para FEBR é adoção do DataVerse juntamente com adesão dos protocolos e princípios para garantir a qualidade e organização dos dados. Sendo que isso inclui software para gerenciamento de metadados DataVerse, esquema de metadados DataCite ou OpenAire, protocolo de interoperabilidade OAI-PMH e padrão para identificador único persistente baseado na WEB no padrão URL/URI.

## AGRADECIMENTOS

Este trabalho foi financiado pela da UTFPR/PROREC como financiadora do PIBEX, na forma de bolsa de projeto de extensão (Edital 01/2020 – PROREC/UTFPR).

## REFERÊNCIAS

- ALVES, R. C. V. Metadados como elementos do processo de catalogação. p. 134, [s.d.].
- ARAÚJO, C. J. A. DE. **Um modelo para interoperabilidade entre instituições heterogêneas.** text—[s.l.] Universidade de São Paulo, 7 dez. 2012.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data – the story so far. International journal on semantic web and information systems. v. 5, p. 3, 22 jan. 2009.
- CAMARGO, F. A. DE O.; ALVAREZ V., V. H.; BAVEYE, P. C. Brazilian soil science: from its inception to the future, and beyond. **Revista Brasileira de Ciência do Solo**, v. 34, n. 3, p. 589–599, jun. 2010.
- DATA FAIRPORT. **Jointly designing a data FAIRPORT.** 2014. Disponível em: <<https://www.lorenzcenter.nl/lc/web/2014/602/info.php3?wsid=602>>
- GARCIA, P; SUNYE, M. O protocolo OAI-PMH para interoperabilidade em bibliotecas digitais. In: CONGRESSO DE TECNOLOGIA PARA GESTÃO DE DADOS E METADADOS DO CONE SUL. v. 1, 2003.
- MÉNDEZ RODRÍGUEZ, E. Metadados y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales. p. 429, 2002.
- Open Archives Forum. Implementing oai-pmh.** , [s.d.]. Disponível em: <<http://www.oaforum.org/tutorial/english/page4.htm>>. Acesso em: 20 dez. 2020
- SILVA, L. C. L. DA et al. Pedodiversidade no estado de Minas Gerais - Brasil. **Caderno de Geografia**, v. 28, n. 1, p. 18–38, 25 mar. 2019.