

ACESSO ABERTO A DADOS DE PESQUISA NO BRASIL

Dataverse documentação: identificadores persistentes



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

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: GABRIEL JUNIOR, Rene Faustino; AZAMBUJA, Luís Alberto Barbosa; MOURA, Ana Maria Mielniczuk; CAREGNATO, Sônia Elisa; PASSOS, Paula Caroline Schifino Jardim; PAVÃO, Caterina Groposo; BORGES, Eduardo Nunes; ROCHA, Rafael Port da; VANZ, Samile Andrea de Souza. Acesso aberto a dados de pesquisa no Brasil: identificadores persistentes para dados de pesquisa. 2020. Disponível em: <<http://hdl.handle.net/20.500.11959/1258>>

Sumário

1	INTRODUÇÃO.....	5
1.1	O problema.....	5
2	ESQUEMAS DE IDENTIFICADORES PERSISTENTES.....	6
2.1	<i>Uniform Resource Name (URN).....</i>	6
2.2	<i>Archival Resource Key (ARK).....</i>	7
2.3	<i>Persistent Uniform Resource Locator (PURL).....</i>	8
2.4	<i>Handle System.....</i>	9
2.5	<i>Digital Object Identifier (DOI).....</i>	10
2.5.1	Agências de registro do DOI.....	12
3	AQUISIÇÃO DO HANDLE.....	13
4	AQUISIÇÃO DO DOI.....	14
5	PERGUNTAS FREQUENTES SOBRE USO DO DOI.....	16
	REFERÊNCIAS.....	18

1 INTRODUÇÃO

O termo *identificador persistente* caracteriza um objeto digital acessível pela Web a fim de identificá-lo de forma única e independentemente da localização de sua representação. Este identificador possibilita que o recurso seja corretamente citado ao longo do tempo, permitindo também que várias cópias estejam disponíveis em locais diferentes.

Um identificador persistente (PI ou PID) é uma referência de longa duração a um documento, arquivo, página da Web ou outro objeto (ORCID, 2019). Mesmo que um recurso seja movido, ou sua propriedade seja transferida, os *links* permanecem efetivamente acionáveis, ou seja, podem ser acessados por meio de um navegador Web e serem direcionados para a fonte identificada. PURL, ISBN, ISSN, ISSI, DOI e *Handle* são exemplos de identificadores persistentes.

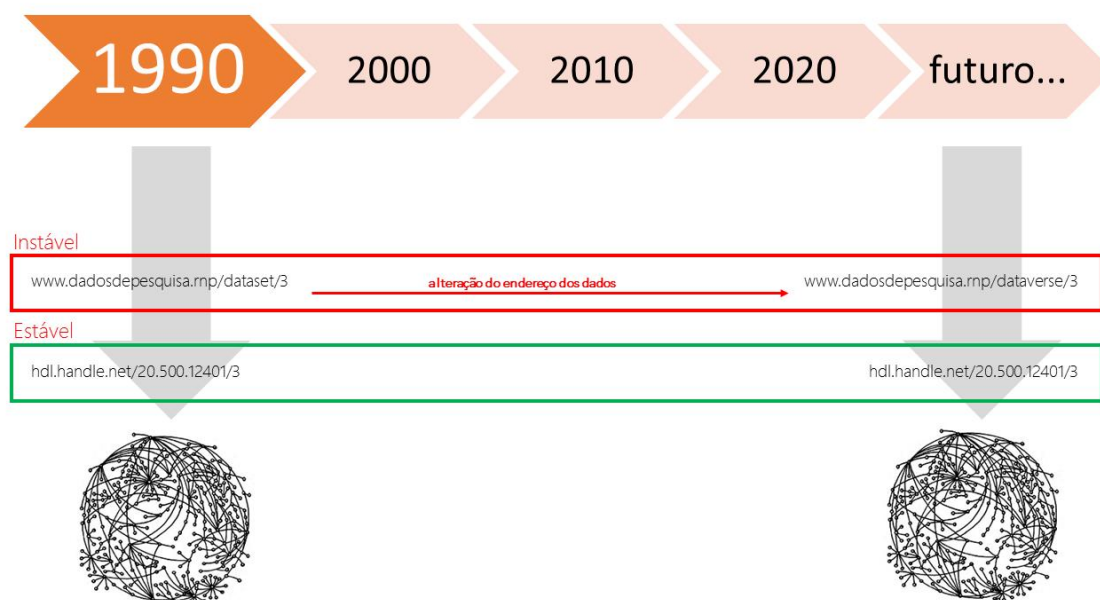
1.1 O problema

O *Uniform Resource Locator* (URL) é um identificador volátil, ou seja, sua utilização para descrever uma fonte de informação na Web torna-se um problema com o passar do tempo, pois depende de uma série de fatores. As razões mais comuns para que os recursos se tornem inacessíveis incluem (SAYAO, 2007):

- a) reorganização do sistema de arquivos em um servidor Web, visto que o URL associa o endereço deste servidor ao nome e caminho do recurso na mesma cadeia de caracteres;
- b) alteração da infraestrutura da instituição que mantém a fonte, tais como troca do nome de domínio, transferência do recurso para um novo URL;
- c) atualização das tecnologias, como sistemas de banco de dados e servidores de aplicações;
- d) desatualização do recurso ou perda de valor comercial, intelectual ou histórico e principalmente e
- e) falta de garantia de preservação dos dados.

Usar URL como chave de integridade referencial, ou seja, como link para um recurso Web em outros recursos, documentos, catálogos ou diretórios podem estabelecer falsas conexões. Já a utilização de indicadores persistentes para as fontes permite a manutenção do acesso por um longo período. A **Figura 1** apresenta um exemplo de identificador persistente para um recurso Web em que a URL foi alterada com o passar do tempo.

Figura 1 - Exemplo de identificador persistente evidenciando a alteração da URL com o passar do tempo.



2 ESQUEMAS DE IDENTIFICADORES PERSISTENTES

Os principais esquemas de identificadores persistentes atualmente em uso são detalhados nas próximas subseções.

2.1 *Uniform Resource Name* (URN)

URN ou Nome Uniforme do Recurso refere-se ao subconjunto do *Uniform Resource Identifier* (URI) com exceção do URL. É um identificador persistente e independente de localização, permitindo o mapeamento simples de *namespaces* em um único *namespace* URN. Por exemplo, são registrados os *namespaces* isbn e issn para citação de livros e periódicos.

A existência de um URN não implica a disponibilidade do recurso identificado, mas obriga a permanência global única e persistente, mesmo quando o recurso deixa de existir.

Atualmente, o termo URN está obsoleto, exceto no sentido muito restrito de um espaço de nomes formais para expressar um Identificador Uniforme de Recurso.

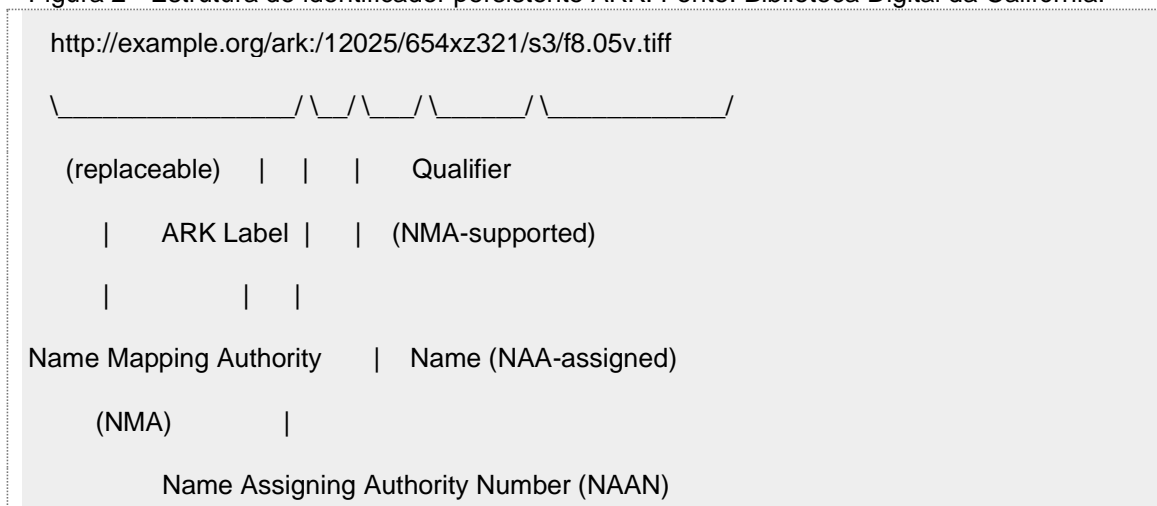
2.2 *Archival Resource Key (ARK)*

O ARK ou Chave de Recurso de Arquivamento é um esquema identificador concebido pela *California Digital Library*¹ (CDL), com o objetivo de identificar objetos de forma persistente. O esquema foi projetado com base em que a persistência "é puramente uma questão de serviço e não é inerente a um objeto nem conferida a ele por uma sintaxe de nomenclatura específica".

A Figura 2 apresenta a estrutura do ARK em detalhe. O identificador é representado por uma sequência de caracteres que contenha o rótulo "ark:". Quando está contido num URL, é precedido do protocolo e o nome do serviço que suporta o ARK, chamado *Name Mapping Authority (NMA)*. Este serviço é mutável e substituível. O identificador global único e imutável aparece após o rótulo, incluindo o nome o identificador *Name Assigning Authority Number (NAAN)* e o nome de acesso ao objeto seguido do qualificador.

¹ <http://www.cdlib.org/inside/diglib/ark/>

Figura 2 - Estrutura do identificador persistente ARK. Fonte: Biblioteca Digital da Califórnia.



2.3 Persistent Uniform Resource Locator (PURL)

PURL² ou Localizador Uniforme de Recursos Persistente é uma URL persistente, que provê um endereço para acessar um recurso Web usando códigos de status HTTP padrão. Um PURL é, portanto, um endereço da Web permanente que contém o comando para redirecionar para outra localização, que pode mudar com o tempo. Quando um autor precisa alterar a localização de um recurso, ele pode alterar o PURL apontando para a nova localização.

Este serviço foi desenvolvido pela *Online Computer Library Center (OCLC)*³ e desde 2016 é mantido pela *Internet Archive*⁴. O código-fonte é distribuído gratuitamente para as organizações que queiram estabelecer seus próprios serviços de resolução.

A estrutura do PURL é essencialmente a mesma do URL, sendo composta do protocolo, endereço do serviço de resolução de PURL e o nome do recurso⁵.

² <https://archive.org/services/purl/>

³ <http://www.oclc.org/>

⁴ <https://archive.org/>

⁵ <http://myPurlResolver.example/name>

2.4 Handle System

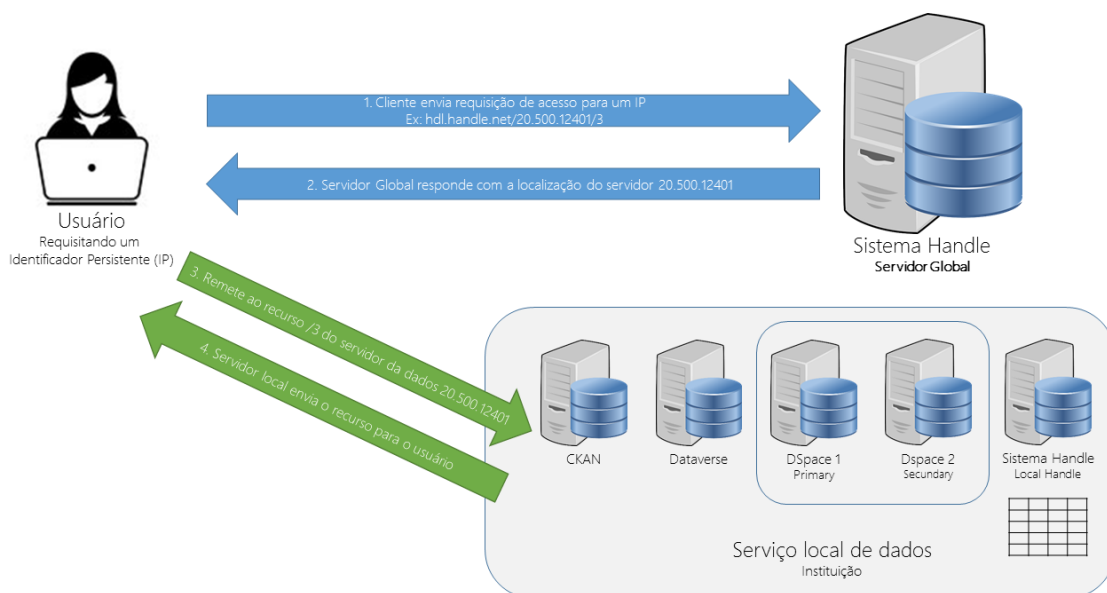
Handle⁶ é um sistema distribuído de grande alcance, voltado para identificação persistente de objetos digitais e outros recursos na *Web*. Este sistema independente de localização e foi desenvolvido para aplicações no domínio das bibliotecas digitais. Ele foi concebido para assinalar, administrar e resolver identificadores persistentes conhecidos como *handles*, que podem ser usados como *Uniform Resource Names* (URN) (SAYÃO, 2007).

Os *handles* são identificadores exclusivos e persistentes para recursos da Internet, com um registro central para resolver URLs para o local atual. Apesar de imutáveis, os *handles* permitem que as informações de localização e acesso aos objetos sejam alteradas. Cada Identificador está associado a um único recurso e a organização que criou ou mantém o recurso. O Sistema *Handle* também sustenta a infraestrutura técnica do DOI, que é um tipo especial de *handle*.

A Figura 3 apresenta um exemplo de requisição realizada no sistema *Handle*. Um cliente envia a requisição de acesso a um servidor global, que responde com a localização de um servidor local. O cliente então solicita ao servidor local o recurso de interesse, o qual envia o recurso para o usuário. Este servidor local pode ser uma instância de um repositório de dados de pesquisa implementada sobre softwares tais como CKAN, Dataverse ou DSpace, devidamente configurado com o sistema *Handle*.

⁶ <http://www.handle.net>

Figura 3 – Exemplo de requisição no Sistema *Handle*.



A estrutura do identificador persistente *Handle* é apresentada em detalhe na **Figura 4**. Além do serviço de resolução, o *Handle* é composto pelo identificador do provedor de dados e pelo identificador do recurso neste provedor.

Figura 4 - Estrutura do identificador persistente *Handle*.

SERVIÇO / <PREFIXO> / <SUFIXO>
 Handle:20.500.12401/31/3123

- **Serviço:** Serviço resolvedor o Identificador Persistente
- **Prefixo:** Prefixo identificador do provedor de dados
- **Sufixo:** Sufixo para identificação do recurso dentro do provedor de dados

2.5 Digital Object Identifier (DOI)

O DOI é um identificador digital de objetos (sejam natos digitais, digitalizados ou abstratos) que podem ser atribuídos por organizações associadas a uma das agências de registro do DOI. No Brasil, as duas agências mais conhecidas são

a CrossRef⁷, para artigos de periódicos e algumas outras publicações acadêmicas, e DataCite⁸ para uma ampla gama de objetos de dados.

O Sistema possui uma infraestrutura para garantir que uma URL seja resolvida para o local correto do objeto digital. O DOI.org é coordenado pela *International DOI Foundation* (IDF), uma organização sem fins lucrativos, fundada em 2000, que faz a governança e o gerenciamento da federação de agências de registro. Estas agências fornecem os serviços de registro do identificador de objeto digital para publicações, conforme norma ISO 26324 (2012) para o sistema DOI (DOI, 2019). O Sistema DOI fornece uma infraestrutura técnica e social para o registro e uso de identificadores interoperáveis persistentes, para uso em redes digitais. (DOI, 2019).

O DOI de um objeto permanece fixo durante a vida útil do documento, enquanto sua localização e outros metadados podem ser alterados. Referir-se a um documento on-line pelo DOI deve fornecer um link mais estável do que simplesmente usar seu URL. Mas toda vez que um URL é alterado, o editor precisa atualizar os metadados do DOI para vinculá-lo ao novo URL. É de responsabilidade do editor atualizar o banco de dados do DOI.

A resolução de nomes DOI é fornecida por meio do *Handle System*, desenvolvido pela Corporação para Iniciativas de Pesquisa Nacional, e está disponível gratuitamente para qualquer usuário que encontre um nome de DOI. A resolução redireciona o usuário de um nome DOI para um ou mais dados digitados: URLs que representam instâncias do objeto, serviços como e-mail ou um ou mais itens de metadados.

Para o *Handle System*, um nome DOI é um identificador e, portanto, tem um conjunto de valores atribuídos a ele e pode ser considerado como um registro que consiste em um grupo de campos. Cada valor de identificador deve ter um

⁷ <https://www.crossref.org/>

⁸ <https://datacite.org/>

tipo de dados especificado em seu campo <type>, que define a sintaxe e semântica de seus dados.

Embora um DOI identifique de forma persistente e exclusiva o objeto ao qual está atribuído, a resolução do DOI pode não ser persistente, devido a problemas técnicos e administrativos. Para resolver um nome DOI, ele pode ser inserido em um resolvedor DOI⁹.

A estrutura do DOI é definida por <Código do Diretório>.<Código de Registro>/<Sequência Sufixo DOI>. Atualmente o código de diretório é 10. Os códigos de registro são distribuídos a grupos editoriais, publicadores e detentores de direitos autorais. O sufixo é determinado pelo gestor de conteúdo e identifica o objeto a ser referenciado.

2.5.1 Agências de registro do DOI

A principal função das agências de registro é fornecer serviços aos gestores de conteúdo, alocando prefixos de nomes de DOI, registrando nomes de DOI e fornecendo a infraestrutura necessária para permitir que os gestores declarem e mantenham metadados e dados de estado. Espera-se que este serviço adote medidas de garantia de qualidade, de modo que a integridade do sistema DOI como um todo seja mantida no nível mais alto possível, fornecendo resultados confiáveis e consistentes aos usuários. Isso inclui garantir que os dados de estado sejam precisos e atualizados e que os metadados sejam consistentes e estejam em conformidade com o núcleo do sistema DOI e os padrões de Perfil de Aplicativo apropriados (DOI, 2019).

A Figura 5 mostra exemplos de agências de registro DOI específicas para produções científicas, conjuntos de dados de pesquisa e mídias de entretenimento.

⁹ <http://doi.org>

Figura 5 - Agências de registro DOI



3 AQUISIÇÃO DO HANDLE

Não existe diferença na aquisição do *Handle* para repositórios de publicações ou para dados de pesquisa, pois a *DONA Foundation*¹⁰, coordenadora do Handle.net atribui um prefixo que redireciona para o provedor do cliente, sendo necessário este estar ativo para a resolução do identificador persistente.

O *Handle* pode ser adquirido diretamente no site da *Handle.Net Registry*¹¹ com o preenchimento do *Prefix Registration Form*. Para o primeiro registro, o custo é de US\$ 100,00 incluindo as taxas para um ano, e US\$ 50,00 a partir do segundo ano. O pagamento deve ser realizado por *PayPal*.

¹⁰ <https://www.dona.net/>

¹¹ <http://handle.net/payment.html>

4 AQUISIÇÃO DO DOI

Como descrito anteriormente, existem agências específicas para registro de diferentes tipos de DOI. O quadro abaixo apresenta detalhes sobre as diferentes agências para aquisição do DOI para dados de pesquisa.

Quadro1 – Descrição das agências fornecedoras de identificadores persistentes para registro de dados de pesquisa

Agência	Descrição
<i>DataCite</i>	<p>Organização internacional sem fins lucrativos que visa melhorar a citação de dados de pesquisa. Para fazer parte do DataCite, a instituição precisa ser membro e preencher um termo de compromisso com seu estatuto.</p> <p>Custo: Taxa de associado: US\$ 7.000,00 por ano (estimado)</p> <p>Tipos de parcerias:</p> <ul style="list-style-type: none">a) DataCentersb) Publicadoresc) Instituições de pesquisad) Financiadores <p>Informações:</p> <ul style="list-style-type: none">a) https://datacite.org/become.htmlb) support@datacite.org; patricia.cruse@datacite.org <p>Prós:</p> <ul style="list-style-type: none">a) Participação na rede de dados de pesquisa (DataCite)b) Atribuição ilimitada de DOIs para dados de pesquisa <p>Contras:</p> <p>Ter infraestrutura para gerenciamento local (país) das contas distribuídas</p>
<i>Global Dataverse Community Consortium</i>	<p>Consórcio da Comunidade Global Dataverse que tem como objetivo reunir esforços e fornecer um espaço colaborativo para as instituições, em apoio aos repositórios da Dataverse em todo o mundo.</p> <p>Custo:</p> <ul style="list-style-type: none">a) Taxa de associado: US\$ 500,00 por anob) Taxa de uso de registro: US\$ 900,00 por anoc) Taxa de excesso: US\$ 300,00 (acima de 10.000 DOIs) <p>Informações:</p> <ul style="list-style-type: none">a) https://dataverse.org/global-dataverse-community-consortium

	<p>b) Email: jonathan_Crabtree@unc.edu</p> <p>Prós:</p> <p>a) Relação custo/benefício para instituição b) Facilidade de registro</p> <p>Contras:</p> <p>a) Utilizar o software DataVerse para o repositório b) Conta única para registro</p>
CrossRef	<p>Organização sem fins lucrativos que existe para melhorar as comunicações acadêmicas. Responsável em marcar e compartilhar metadados; executar uma infraestrutura aberta; desenvolver ferramentas e serviços voltados ao conteúdo acadêmico no contexto.</p> <p>Custo:</p> <p>a) Taxa de associado: US\$ 275,00 a US\$ 50.000,00 por ano (depende da receita da instituição) b) Taxa de uso de registro DOI: US\$ 0,06 por registro corrente (<i>Datasets</i>) c) Taxa de uso de registro DOI: US\$ 0,03 por registro retrospectivo (<i>Datasets</i>)</p> <p>Informações: https://www.crossref.org/fees/</p> <p>Contra: Alguns softwares não registram automaticamente o DOI na Crossref, sendo necessário duplo trabalho.</p>
CrossRef via ABEC	<p>Associação dos Editores Científicos dos Brasil (ABEC)</p> <p>Custo:</p> <p>a) Taxa de associado: R\$ 500,00 por ano b) Taxa de uso de registro DOI: US\$ 1,00 por registro + impostos</p> <p>Faturado e convertidos em R\$ a cada três meses.</p> <p>Informações:</p> <p>a) https://www.abecbrasil.org.br/novo/2015/09/e-stabelecido-o-acordo-abeccrossrefibict-para-facilitar-a-atribuicao-de-dois/ b) doi@abecbrasil.org.br</p> <p>Prós: Facilidade para pequenas instituições com atribuição de poucos registros.</p> <p>Contra:</p> <p>a) Alguns softwares não registram automaticamente o DOI na Crossref, sendo necessário duplo trabalho. b) Restrito a editoras científica (sem fins comerciais)</p>

4 PERGUNTAS FREQUENTES SOBRE USO DO DOI

1 Preciso utilizar o DOI como indicador persistente para os dados de Pesquisa?

Não, pode-se utilizar outro indicador persistente, desde que o sistema de software de repositório suporte. Por exemplo, o Dataverse permite o uso do Handle. O problema não está na atribuição do indicador persistente, mas na garantia que a localização do objeto digital permanecerá ativo por um longo período de tempo e na qualidade dos metadados que descrevem os dados de pesquisa.

2 Pode-se utilizar o DOI do CROSSREF para dados de pesquisa?

Sim, o Crossref permite a atribuição de DOI para dados de pesquisa (*datasets*), entretanto o conjunto de metadados não é suficiente para descrever um conjunto de dados. Além disso, devem-se considerar os custos, pois para cada registro no Crossref gera um custo de US\$ 1,00 para a instituição, e em muitos casos, existem várias versões de um mesmo conjunto de dados, todas referenciadas por diferentes DOIs. Portanto, não é recomendado.

3 Qual a diferença entre o Handle e o DOI nos registros de dados de pesquisa?

A principal diferença entre o Handle e o DOI está na infraestrutura. Enquanto no Handle ela é descentralizada, dependendo dos servidores dos provedores dos dados, no DOI a infraestrutura é centralizada, com o armazenamento nas agências de registros dos metadados de cada DOI. Em ambos os casos, se o servidor de armazenamento do provedor estiver “fora do ar”, o conteúdo dos dados não será acessível.

4 O repositório pode começar com Handle e depois contratar o DOI?

Sim, é possível iniciar com o Handle um repositório de dados e mais tarde alterar para o DOI, desde que se mantenha a estrutura do Handle para resolução de identificador persistente. Também é possível a contratação de um DOI de uma agência como a Crossref ou ABEC e depois se associar a DataCite. Em todos os casos, os DOI ou Handle já atribuídos não serão alterados. O repositório de Dados de Pesquisa pode utilizar métodos de registros DOI/Handle diferentes no decorrer do tempo, entretanto não concomitantes, ou seja, apenas um serviço estará ativo para publicação de novos *datasets* ou novas versões de *datasets* já publicados.

REFERÊNCIAS

DAVIDSON, J. Persistent Identifiers. DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre, 2006. Disponível em: <http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/persistent-identifiers>

DOI. The DOI system. Disponível em: <http://www.doi.org/index.html> Acesso em 30 jun. 2019.

DPC. Digital Preservation Handbook. Disponível em: <https://www.dpconline.org/handbook/technical-solutions-and-tools/persistent-identifiers>

IFLA. Persistent identifiers. Disponível em: <https://www.ifla.org/book/export/html/8789>

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Information and documentation -- Digital object identifier system. ISO Standard 26324, 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Space data and information transfer systems -- Open archival information system (OAIS) - Reference model. ISO Standard 14721 (CCSDS 650.0-M-2), 2012.

LANGSTONA, M.; TYLER, J. Linking to journal articles in an online teaching environment: The persistent link, DOI, and OpenURL. The Internet and Higher Education, v. 7, n. 1, 1st quarter, p. 51-58, 2004.
[DOI:10.1016/j.iheduc.2003.11.004](https://doi.org/10.1016/j.iheduc.2003.11.004)

ORCID. What are persistent identifiers (PIDs)? Disponível em: <https://support.orcid.org/hc/en-us/articles/360006971013-What-are-persistent-identifiers-PIDs> Acesso em 24/06/2019.

SANTOS, G. C. Estudo comparativo das vantagens da interoperabilidade entre o DOI e HANDLE para sistemas de informação. Boletim Técnico do PPEC, v. 4, 12 mar. 2019. Disponível em: <https://econtents.bc.unicamp.br/boletins/index.php/ppec/article/view/9169>

SAYAO, L. F. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes -URN, PURL, DOI, Handle System, CrossRef e OpenURL. Transinformação, Campinas, v. 19, n. 1, p. 65-82, abr. 2007 Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862007000100006&lng=en&nrm=iso Acesso em: 12 jun 2019.
DOI:10.1590/S0103-37862007000100006