



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-8 – Informação e Tecnologia

ESTUDO COMPARATIVO ENTRE O PLANO DE GESTÃO DE DADOS E A ISO 16363:2012: ÊNFASE NA GESTÃO DO OBJETO DIGITAL

COMPARATIVE STUDY BETWEEN MANAGEMENT DATA PLAN AND ISO 16363:2012: EMPHASIS ON DIGITAL OBJECT MANAGEMENT

Felipe Caldonazzo de Almeida Pereira - Universidade Estadual de Londrina (UEL)
Paula Regina Ventura Amorim Gonçalves - Universidade Federal do Espírito Santo (UFES)
Brígida Maria Nogueira Cervantes - Universidade Estadual de Londrina (UEL)

Modalidade: Resumo Expandido

Resumo: O movimento de Ciência Aberta apresenta propostas para uma nova prática no fazer científico, e o acesso livre aos dados de pesquisa é uma de suas características, assim, pode-se utilizar ferramentas que auxiliam os profissionais na disponibilização destes dados. Tendo como objetivo a comparação entre os documentos de Plano de Gestão de Dados e a seção 4 da ISO 16363:2012, visando identificar pontos em comuns e o alinhamento entre estes. O resultado foi a identificação de 83% dos campos categorizados em comum. Conclui-se que ambos os documentos estão em consonância com as práticas de Ciência Aberta e investigações futuras podem explorar outras questões, como o reuso de dados.

Palavras-Chave: ciência aberta; repositório de dados; gestão de dados de pesquisa; objetos digitais.

Abstract: Open Science movement shows a proposal for a new way of scientific activity, being open access one of its many features to research data, so, professionals may use tools to help making it available. The objective was to compare two documents, data management plan and ISO 16363:2012, seeking to identify common points and the align between these documents. The result was that 83% of the categories were in common. The conclusion was that both documents were in accordance to Open Science practices, and future investigations may explore other subjects, like the data reuse.

Keywords: open science; digital repositories; research data management; digital object.

1 INTRODUÇÃO

A Ciência Aberta surge como uma nova proposta para o fazer científico (FUNDAÇÃO OSWALDO CRUZ, [20--]), entre algumas características, estão o acesso aos dados de pesquisa de forma mais direta e organizada e transparência na coleta e facilidade de reuso (ALBAGLI; CLINIO; RAYCHTOCK, 2014; SAYÃO; SALES, 2016). Uma das principais fontes para se acessar os dados de pesquisa é o repositório de dados de pesquisa, um ambiente digital onde o pesquisador pode depositar os dados coletados ou gerados, permitindo que outros pesquisadores acessem e reutilizem esses dados de pesquisa, inclusive resultados inesperados ou dados negativos, apresentados como ciência invisível por Sayão e Sales (2020).

A maioria dos repositórios digitais possui um manual ou instruções que auxiliam o pesquisador a realizar o depósito de sua pesquisa, publicação ou dados de pesquisa. Já alguns repositórios de dados de pesquisa, por exigência de agências de fomento, indicam um documento onde são apresentados os campos que precisam ser preenchidos para o depósito dos dados.

Este documento segue recomendações para o Plano de Gestão de Dados (PGD) empregado (MONTEIRO; SANT'ANA, 2017), podendo, inclusive, instituir uma política de gestão de dados para diversas instituições de pesquisa poder compartilhar os dados de pesquisa em uma rede de repositórios de dados de pesquisa, como o da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, [2021]).

Segundo a FAPESP ([202-]), uma instituição não precisa possuir um único PGD, pois este documento pode variar de acordo com o domínio do conhecimento, mas indica que se deve responder a algumas questões, tais como: quais os dados gerados; como serão preservados e disponibilizados. *Digital Curation Centre* (DCC)(2013) oferece um *checklist* para que o pesquisador possa preencher, antes de se depositar os dados de pesquisa.

Repositórios de dados de pesquisa também seguem um modelo de ciclo de vida, como o do DCC, para a curadoria dos dados de pesquisa, preparando cada etapa, desde o desenvolvimento de uma política para o repositório, a coleta e o armazenamento, até a disponibilização dos dados para uma futura pesquisa.

Portanto, além do PGD, necessita-se que sejam instituídas diretrizes para que os repositórios sejam confiáveis, fazendo-se o uso da ISO 16363:2012, que, conforme afirmou Gonzalez (2017, p. 100), representa “[...] uma ferramenta que viabiliza a auditoria, a avaliação

e a certificação dos repositórios digitais”, por meio de documentação, metodologias entre outros instrumentos para a realização de auditoria para os repositórios digitais serem considerados confiáveis.

Assim, este trabalho tem como objetivo comparar estes documentos, o PGD e a ISO 16363:2012, e identificar as semelhanças acerca da gestão dos dados de pesquisa, com o intuito de futuras pesquisas para o desenvolvimento de diretrizes nesse sentido. Para isso, será realizada uma pesquisa descritiva de natureza qualitativa, do tipo documental, com o uso da técnica de análise do conteúdo de Bardin (2002).

2 DESENVOLVIMENTO

Nesta seção serão apresentados o percurso metodológico, a análise e os resultados desta pesquisa.

2.1 Percurso metodológico

Esta pesquisa foi caracterizada como descritiva e por sua natureza quali-quantitativa, com a utilização de pesquisa do tipo documental, sendo realizada uma coleta de dados dos documentos selecionados do PGD da DCC e na ISO 16363:2012, com foco nas atividades de gestão da informação.

Com a utilização da técnica da Análise de Conteúdo (BARDIN, 2002), foi realizada a pré-análise, onde foram selecionados dois documentos, o PGD da DCC e a ISO 16363:2012, e houve também a delimitação do tema, a seleção dos campos que tratam a gestão da informação nos dois documentos.

Em seguida, explorou-se os materiais, onde selecionou-se todo o documento do PGD da DCC e a seção 4 da ISO 16363:2012, por ser a seção que trata especificamente da gestão de objetos digitais/gestão da informação. Por fim, os dados foram coletados em uma planilha do Excel, com as categorias fundamentadas nos oito campos gerais do PGD (dados administrativos, coleta dos dados, documentação e metadados, *compliance* ético e legal, armazenamento e *backup*, seleção e preservação, compartilhamento dos dados e responsabilidade e recursos) e comparada as ações exigidas com as recomendações da seção 4 da ISO 16363:2012.

2.2 Análise dos dados e resultados

A FAPESP ([202-]) recomenda que no PGD tenha as seguintes informações sobre os dados de pesquisa: a) descrição dos dados e metadados; b) questões éticas, legais, política de privacidade, segurança entre outros; c) política de preservação e compartilhamento; e d) descrição dos mecanismos, formatos e padrões para armazenamento.

Instituições, como a DCC, fornecem um documento para a realização de um *checklist* (DCC, 2013), onde o pesquisador segue um caminho para preencher as informações necessárias. Os campos mais gerais, tem-se: dados administrativos, coleta de dados, documentação e metadados, *compliance* éticos e legais, armazenamento e *backup*, seleção e preservação, compartilhamento, responsabilidades e recursos. Assim, tem-se:

- **Dados administrativos:** informações básicas sobre o projeto, como a identificação do projeto pela instituição/financiadora, nome do projeto, descrição do projeto, identificação e contato do pesquisador, data de criação e última versão, além de políticas relacionadas ao projeto (por exemplo: existência de um modelo de Gestão de Dados de Pesquisa).
- **Coleta dos dados:** informações sobre a coleta de dados, como quais dados serão coletados ou criados (tipo, formato, entre outros), existência de dados para reuso, forma da coleta de dados (metodologia empregada), organização das pastas e arquivos, organização das versões e garantia da qualidade dos dados.
- **Documentação e metadados:** uso de documentação para auxiliar na leitura e interpretação dos dados, como serão criados a documentação e os metadados, qual ou quais padrões de metadados serão utilizados.
- **Compliance ético e legal:** consentimento para preservação e compartilhamento dos dados, privacidade dos participantes da pesquisa, segurança de dados sensíveis, propriedade e direitos sobre os dados, permissão para o reuso.
- **Armazenamento e backup:** capacidade para armazenamento e *backup*, responsável (instituição ou organização) pelo armazenamento, segurança e recuperação, informações sobre o risco e o acesso aos dados seguros.

- **Seleção e preservação:** quais dados serão preservados, compartilhados e/ou descartados e por quanto tempo, isto inclui quem decide por este planejamento, como e onde será feito.
- **Compartilhamento dos dados:** como será compartilhado, como os usuários poderão encontrar os dados e em quais circunstâncias, qual repositório e se usará identificador persistente. Questões sobre dados restritos também deverão ser esclarecidos.
- **Responsabilidades e recursos:** apresenta quem serão os responsáveis pela coleta dos dados, produção de metadados, conferência e armazenamento, segundo as políticas indicadas. Também apresenta se tem a necessidade adicional de apoio técnico ou uso de *software/hardware* além dos já providenciados pelo repositório.

Além do PGD, a ISO 16363:2012 também apresenta informações e procedimentos que auxiliam os gestores dos repositórios nas atividades referentes ao gerenciamento da informação, mais especificamente na seção 4, sobre Gestão do Objeto Digital.

Nesta seção, os tópicos abordados se referem à ingestão (*ingest*) do objeto digital, a criação de um Pacote de Informação de Arquivo (*Archival Information Package* - AIP), ao plano de preservação, à preservação do AIP, à gestão da informação e ao seu acesso. Um detalhe importante sobre a ISO 16363:2012, é a escolha de algumas nomenclaturas, como por exemplo: o uso dos termos *shall* e *must* (deve), para uma ação necessária; o *should* (deveria), para uma ação opcional, mas desejada; o *may* (pode), para uma ação só opcional; e o *is, are and will*, para afirmações de um fato.

- **Ingestão (*ingest*) do objeto digital:** corresponde à subseção 4.1, que determina os procedimentos para a aquisição do conteúdo, sendo composta por 10 recomendações: que vão desde a identificação das informações de conteúdo e as propriedades destas informações que o repositório preservará (subitem 4.1.1) até a guarda dos registros atuais das ações tomadas ao se depositar o objeto digital, para medidas de auditoria (subitem 4.1.8);
- **Criação de um AIP:** corresponde à subseção 4.2, que determina as recomendações para a criação do Pacote de Informação para Arquivamento, ou AIP em inglês. São 28 recomendações, de ações mais gerais a ações mais

específicas, visando a criação de um conjunto de informações que auxiliem na preservação de longo prazo dos objetos digitais;

- **Plano de preservação:** são recomendações, divididas em seis subitens, que determinam desde a criação de uma estratégia documentada de preservação (subitem 4.3.1 até o fornecimento de evidência da efetividade das atividades de preservação (subitem 4.3.4);
- **Preservação do AIP:** são seis subitens que recomendam a preservação do pacote de informação de arquivamento visando a garantia de sua integridade e sua recuperação em longo prazo;
- **Gestão da informação:** dividido em quatro subitens, recomendam ações para a garantia da descrição dos recursos digitais, visando a encontrabilidade destes recursos, a associação entre as descrições e os pacotes de informação de arquivamento;
- **Gestão do acesso:** são recomendações, divididas em quatro subitens, que visam a utilização de políticas de acesso, o registro de erros ou anomalias, a garantia de rastreamento ao objeto digital autêntico, confirmando a confiabilidade do repositório.

A partir da categorização dos campos do PGD da DCC, foi possível identificar a relação destes campos e as recomendações da ISO 16363:2012, sendo que dos 31 campos categorizados da PGD, 83% tinha alguma indicação explícita na ISO 16363:2012, e mais da metade dos campos abrangia mais de uma recomendação da ISO.

Quadro 1 – PGD e ISO 16363:2012

DCC PGD	ISO 16363:2012
DADOS ADMINISTRATIVOS	RECOMENDAÇÕES
identificação do projeto pela instituição/financiadora	4.1.1; 4.1.2; 4.1.4
nome do projeto	4.1.1; 4.1.2; 4.1.4
descrição do projeto	4.1.1
identificação e contato do pesquisador	4.1.1; 4.1.2; 4.1.4
data de criação e última versão	4.1.1, 4.1.5
outras políticas relacionadas ao projeto	4.1.1; 4.1.4

DADOS DE COLETA	
qual(is) tipo(s) de dados será(ão) coletados(s)/coletados(s)	4.1.1
os dados coletados utilizam formatos ou softwares que permitem compartilhamento e acesso de longo prazo	4.1.1; 4.1.1.1; 4.3; 4.4
existência de dados para reuso	
forma de coleta/criação (metodologia empregada)	4.1.1
organização das pastas e arquivos	
organização das versões dos dados coletados	4.1.1;
garantia da qualidade dos dados	4.1.1; 4.1.3; 4.1.5; 4.1.8; 4.2; 4.4; 4.6.2
DOCUMENTAÇÃO E METADADOS	
informação para auxiliar leitura e interpretação dos dados	4.1.2; 4.2.5; 4.2.7
como serão criados a documentação e os metadados	
qual(is) padrão(ões) de metadados será(ão) utilizado(s)	4.1.1; 4.1.5
COMPLIANCE ÉTICO E LEGAL	
permissão para preservação e compartilhamento dos dados	4.1.2; 4.1.6; 4.2; 4.3; 4.6.1
privacidade dos participantes da pesquisa ou segurança dos dados sensíveis, propriedade e direitos sobre os dados	
detalhes sobre permissão para o reuso dos dados	
ARMAZENAMENTO E BACKUP	
capacidade para o armazenamento dos dados	
como e quem será responsável pelo backup e recuperação	4.1.6; 4.4
informações sobre riscos e permissão de acesso aos dados seguros	4.3; 4.6.1
SELEÇÃO E PRESERVAÇÃO	
quais dados serão preservados, compartilhados e/ou descartados	4.3
informações sobre preservação de longo tempo dos dados	4.1.8; 4.3; 4.4
COMPARTILHAMENTO DOS DADOS	
como os dados serão compartilhados ou encontrados pelos usuários	4.2.4; 4.3.2.1; 4.5; 4.6.1
informações sobre o repositório utilizado	4.2.4.2
uso de identificadores persistentes	4.2.4.1.1

informações sobre dados restritos	4.1.2; 4.6.1
RESPONSABILIDADES E RECURSOS	
responsáveis pela coleta dos dados	4.1.6
responsáveis pela produção de metadados, checagem e armazenamento (gestão dos dados)	4.1.3; 4.1.5; 4.1.6; 4.1.8; 4.2.5
necessidade de apoio técnico adicional, uso de software/hardware além dos disponíveis no repositório	4.2.5

Fonte: elaborado pelos autores, baseado em DCC (2013) e ISO 16363:2012

Os campos que não possuem recomendação explícita da ISO se referem às questões sobre o reuso dos dados (mas que pode estar implicitamente ligado a partir das recomendações para o acesso à longo prazo), à forma de nomeação dos arquivos, à privacidade dos participantes da pesquisa e capacidade do armazenamento no repositório, o que indica que a formulação de um PGD está em consonância com os requisitos para a confiabilidade dos repositórios digitais no quesito gestão da informação.

Tanto o uso da PGD quando o uso da ISO 16363:2012 indicam que os gestores dos repositórios possuem uma preocupação com a descrição dos conjuntos de dados e que essas informações possam auxiliar os usuários a localizarem os dados e que estes sejam confiáveis, com o uso de identificadores persistentes e únicos, o uso de metadados descritivos, além da correta identificação dos dados depositados e da recuperação ao longo do tempo.

3 CONSIDERAÇÕES FINAIS

Os dados de pesquisa deixam de ser subproduto do fazer científico para se tornarem uns dos principais ativos produzidos por uma pesquisa (SAYÃO; SALES, 2016), além de serem um dos elementos que auxiliam na transparência da pesquisa científica. Outrossim, como foi percebido pela análise dos documentos, a forma como serão apresentadas as informações sobre os dados de pesquisa, como foram coletados, descritos (quais instrumentos, ferramentas ou esquema de metadados), organizados, disponibilizados para recuperação e descoberta, a monitoração de erros e o devido planejamento para preservação indicam preocupação tanto com as atividades do próprio repositório quanto às necessidades dos usuários.

Estas ações demonstram a importância do uso do PGD e da ISO 16363:2012 para a atividade do pesquisador e dos gestores de repositórios digitais, e em especial para o

profissional da informação, que tem função importante para o funcionamento do que se conhece como Ciência Aberta, além da participação em algumas propostas para aprimoramento da comunicação científica e do uso de repositórios digitais (LEITE; COSTA, 2016).

REFERÊNCIAS

ALBAGLI, S.; CLINIO, A.; RAYCHTOCK, S. Ciência Aberta: correntes interpretativas e tipos de ação. **Liinc em Revista**, Rio de Janeiro, v. 10, n. 2, p. 434-450, nov. 2014.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2002.

DIGITAL CURATION CENTRE. **Checklist for a Data Management Plan**. 2013. Disponível em: https://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf. Acesso em: 06 jun. 2021.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Gestão de Dados**. [202-]. Disponível em: <https://fapesp.br/gestaodedados>. Acesso em: 06 jun. 2021.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Metabusador de dados de pesquisa**. [2021]. Disponível em: <https://metabusador.uspdigital.usp.br/>. Acesso em: 28 jun. 2021.

FUNDAÇÃO OSWALDO CRUZ. **Glossário da ciência aberta**. [20--]. Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 24 ago. 2021.

GONÇALEZ, P. R. V. A. **Repositórios arquivísticos digitais confiáveis**: identificação de requisitos com ênfase no acesso à informação. 2017. 166 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 16363:2012**: Space data information transfer systems -- Audit and certification of trustworthy digital repositories. Genebra, 2012.

LEITE, F. C. L.; COSTA, S. M. S. Modelo genérico de gestão da informação científica digital para instituições de pesquisa na perspectiva da comunicação científica e do acesso aberto. **Investigación Bibliotecológica**: Archivonomía, Bibliotecología e Información, v. 30, n. 69, p. 43-74, maio/ago. 2016. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0187358X16300600>. Acesso em: 06 jun. 2021.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Plano de gerenciamento de dados no contexto dos repositórios de dados das universidades. In: **WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA**, 1., 2017, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2017. Não paginado.

Disponível em: <https://repositorio.ufsc.br/xmlui/handle/123456789/180293>. Acesso em: 18 maio 2021.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90–115, maio/ago. 2016.

Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/27939>.

Acesso em: 12 maio 2021.

SAYÃO, L. F.; SALES, L. F. A ciência invisível: por que os pesquisadores não publicam seus resultados negativos? **Informação & Informação**, Londrina, v. 25, n. 4, p. 98-116, out./dez. 2020. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/40016>. Acesso em: 24 ago. 2021.