



4º Plano de Ação Nacional em Governo

Compromisso 3 - Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil



4º Plano de Ação Nacional em Governo

Compromisso 3 - Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil

Marco 03 - Definição de diretrizes e princípios para políticas institucionais de apoio à Ciência Aberta

Instituições responsáveis:





Lista de ilustrações

Figura 1: Ciclo de Vida do Dado	13
Figura 2: Ciclo de Vida da Gestão do Dado	14
Figura 3: Ciclo de Vida da Curadoria do Dado	15



Lista de quadros

Quadro 1: Princípios e Diretrizes	8
--	---



Lista de abreviaturas e siglas

COAR	Confederation of Open Access Repositories
CSIP	Community Strategy and Implementation Plan
DCC	Data Curation Centre
FAIR	Findable Accessible Interoperable Reusable
JRC	Joint Research Centre
MP	Ministério do Planejamento, Orçamento e Gestão
OpenDOAR	Directory of Open Access Repositories
PGD	Plano de Gestão de Dados
RE3DATA	Registry of Research Data Repositories
ROAR	Registry of Open Access Repositories

Sumário

Apresentação	7
1. Quadro geral.....	8
2. Princípios	9
2.1 “Tão abertos quanto possível, tão fechados quanto necessário”, ou seja, disponibilização de dados abertos gratuitamente, porém atentos às questões éticas e legais.	9
2.2 Princípios FAIR	10
2.3 Capacitação da equipe e corpo técnico.....	10
3. Diretrizes para políticas de dados	11
4. Diretrizes para repositórios de dados	19
Referências.....	21
Anexo I – Orientações para os Princípios FAIR.....	24
Anexo II – Orientações PGD	25
Anexo III – Orientações vocabulários controlados.....	26
Anexo IV – Orientações para metadados.....	27
Anexo V – Orientações para visibilidade.....	28



Apresentação

1. Quadro geral

A seguir é apresentado um quadro com as Diretrizes e Princípios levantados e recomendados:

Quadro 1: Princípios e Diretrizes

Diretrizes	Princípios
Campo de aplicação da política	<p>“Tão abertos quanto possível, tão fechados quanto necessário”</p> <ul style="list-style-type: none"> • Abertura dos dados • Gratuidade de acesso • Questões éticas • Legalidade
Objetivo da política	<p>Princípios FAIR</p> <ul style="list-style-type: none"> • Encontrabilidade • Acessibilidade • Interoperável • Reutilizável
Responsabilidades e atribuições	<ul style="list-style-type: none"> • Capacitação da equipe e corpo técnico
<p>Orientações sobre o PGD</p> <ul style="list-style-type: none"> • Descrições dos dados • Formatos • Política e plano de preservação 	
<p>Orientações sobre gestão de dados</p> <ul style="list-style-type: none"> • Descrição do ciclo de vida do dado na instituição com orientações acerca das ferramentas 	
Aspectos éticos, morais e legais	
Diretrizes técnicas para repositórios	

Fonte: Elaboração própria.

2. Princípios

São três grandes princípios, que devem ser orientadores, ao tratar de uma política de dados abertos:

2.1 “Tão abertos quanto possível, tão fechados quanto necessário”, ou seja, disponibilização de dados abertos gratuitamente, porém atentos às questões éticas e legais.

Esse princípio abrange diversos conceitos, mas tem seu significado claro nos ideais e benefícios da Ciência Aberta em disponibilizar em acesso aberto o máximo possível, porém respeitando qualquer possível barreira ou impedimento legal, ético, moral ou de qualquer outro tipo. Para facilitar seu entendimento seguem os principais tópicos que aqui devem ser considerados:

- Abertura dos dados (questões sobre embargo)

Os dados por regra geral deverão ser abertos. Caso seja necessário, os dados terão períodos de embargo definidos.

- Gratuidade de acesso

A gratuidade de acesso é uma das características norteadoras do Acesso Aberto e vale ressaltar que essa gratuidade de acesso não exclui questões técnicas como processamento, maquinário e conexão com a rede. Cobrar qualquer taxa, mesmo que seja de manutenção, do acessante, já desconfigura este item.

- Questões éticas (sensibilidade dos dados)

Dados são de responsabilidade do mantenedor, ou seja, é responsabilidade da instituição garantir que não ocorram falhas de segurança ou publicação de dados pessoais ou identificadores que possam ser usados para prejudicar um indivíduo. Esta característica será mais explorada nas diretrizes.

- Legalidade (licenças de uso e direitos autorais)

Sobre questões legais deve-se atentar à diversos dispositivos específicos que cada área pode dispor, porém recomenda-se sempre dedicar uma atenção especial à definição das licenças de uso e direitos de autorais. Dados em específico devem atentar a Lei Geral de Proteção aos Dados (LGPD) nº 13.709/2018 que também será abordada nas diretrizes. Sobre as licenças de uso deve-se adotar a que melhor se adapte à situação da organização, porém recomenda-se o uso da licença CC-BY que permite a reprodução, adaptação e derivação do trabalho, e o devido crédito ao seu autor, ou alguma outra licença Creative Commons.

2.2 Princípios FAIR

Em relação aos dados de pesquisa adota-se como princípio basilar os “Princípios FAIR” (*Findable, Accessible, Interoperable e Reusable* ou Encontrável, Acessível, Interoperável e Reutilizável).

2.3 Capacitação da equipe e corpo técnico

Este princípio representa a necessidade de se ter uma equipe de trabalho hábil, treinada, capacitada e educada, que possa enfrentar as necessidades e desafios que porventura se apresentem.

3. Diretrizes para políticas de dados

A partir dos princípios mencionados foram estabelecidas as seguintes diretrizes recomendadas que serão então explicadas a seguir, uma a uma:

- Campo de aplicação da política

Essa é uma diretriz geral para elaboração de normas e políticas, nela deve-se especificar a quais instâncias organizacionais a política em questão se aplica. Também deverão ser definidos quais cargos ou indivíduos devem ou não a seguir e em quais casos é ou não aplicada a diretriz.

- Objetivo da política

A diretriz geral para elaboração de normas e políticas deve abranger o objetivo do documento, o seu alcance e o que se almeja atingir com a política e suas premissas.

- Responsabilidades e atribuições

Devem ser descritas a estrutura organizacional e funcional das unidades organizacionais responsáveis pela coordenação, manutenção e desenvolvimento dos dispositivos tratados na política, bem como os atores envolvidos nesses processos e suas atribuições e fluxos de trabalho.

- Orientações sobre o Plano de Gestão de Dados (PGD)

O PGD deve um documento que descreve como os dados serão obtidos, documentados, organizados, armazenados, preservados e compartilhados durante a fase de pesquisa até o fim do projeto. É importante que constem no PGD no mínimo:

- Descrições dos dados (geral e metadados), bem como métodos e técnicas de coleta e documentação;
- Formatos e padrões adotados, e;
- Política e plano de preservação e disponibilização.

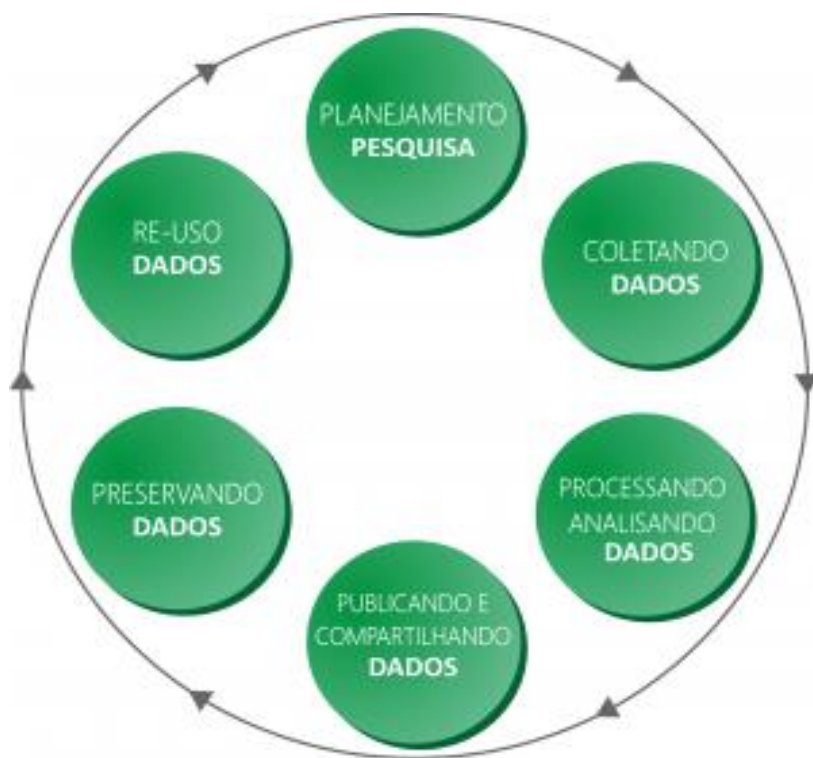
Cada instituição e agência financiadora pode exigir a elaboração de um PGD.

- Orientações sobre gestão de dados

Deve ser apresentada uma descrição do ciclo de vida do dado na instituição com orientações acerca das ferramentas. Existem diversos modelos de ciclo de vida de dados que podem ser utilizados de base para se descrever o próprio. Seguem os modelos para a descrições do ciclo de vida de dados:

- ❖ Para o ciclo de vida do **dado**: *Research Data Lyfe cycle* da *UK Data Archive* <<http://www.data-archive.ac.uk/create-manage>> que abrange: criação, tratamento, análise, preservação, acesso e reutilização dos dados.

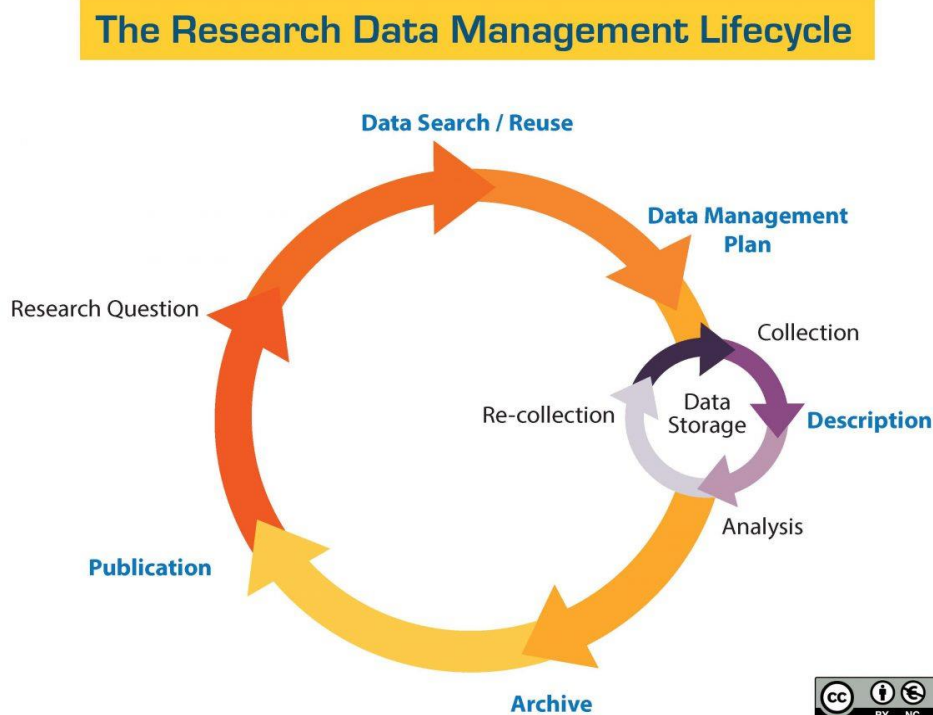
Figura 1: Ciclo de Vida do Dado



Fonte: Angues, 2018.

- ❖ Para o ciclo de vida da **gestão do dado**: O *Research Data Management Life Cycle* <<https://guides.library.ucsc.edu/datamanagement>> que abrange: Pesquisa ou reuso, PGD, Descrição (coleção, arquivo, análise e recoleta), Arquivamento e Publicação (Pergunta da pesquisa).

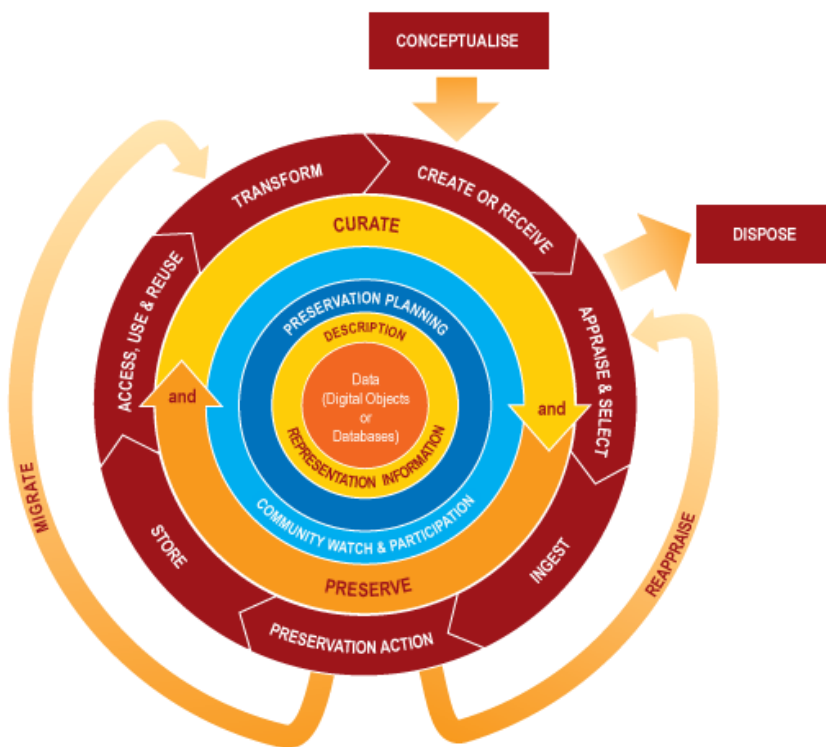
Figura 2: Ciclo de Vida da Gestão do Dado



Fonte: UC Santa Cruz, 2014.

- ❖ Para o ciclo de vida da **curadoria do dado**: O *DCC Curation Lifecycle Model* <<http://www.dcc.ac.uk/resources/curation-lifecycle-model/lifecycle-model-faqs>>, abrangendo diversas etapas e envolvendo várias camadas de atuação, indo desde a conceituação à transformação final.

Figura 3: Ciclo de Vida da Curadoria do Dado



Fonte: *Data Curation Center*.

- Aspectos éticos, morais e legais

Alguns aspectos legais devem ser seguidos, além dos específicos para cada área do conhecimento que pode dispor de dispositivos e normas próprias, existem alguns gerais que também devem ser observados, são eles:

- Lei n. 13.709/2018 – Lei Geral de Proteção de Dados - LGPD (Brasil, 2018) estabelece normas específicas para o tratamento de informações pessoais
<http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>, seus principais pontos são:



- Ela estabelece os direitos dos titulares dos dados pessoais;
 - Aplica-se a qualquer operação de tratamento de dados pessoais;
 - O tratamento de dados pessoais somente poderá ser realizado mediante o fornecimento de consentimento pelo titular;
 - O controlador deverá comunicar à autoridade nacional e ao titular a ocorrência de incidente de segurança.
- Decreto n. 8.777/2016 - Política de Dados Abertos do Governo Federal <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8777.htm> Instituiu a Política de Dados Abertos do Poder Executivo federal ao:
- Promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal sob a forma de dados abertos;
 - Aprimorar a cultura de transparência pública;
 - Franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal;
 - Fomentar o controle social e o desenvolvimento de novas tecnologias.
- Lei nº 12.527/2011 - Lei de Acesso à Informação (LAI) <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm> Regula o acesso a informações previsto na Constituição Federal ao:
- Pode-se ter acesso a qualquer informação pública produzida ou custodiada pelos órgãos e entidades da Administração;
 - A busca e fornecimento da informação é gratuita;



- 20 (vinte) dias para atender ao pedido;
- Obrigação de publicação prévia de um rol mínimo de informações nos sites dos órgãos;
- Obrigatória a instalação do Serviço de Informações ao Cidadão - SIC.

Sobre os aspectos éticos e morais encontramos principalmente questões relacionadas à privacidade dos dados, recomendamos então a utilização da seguinte categorização e técnicas para tratar o assunto:

- Categorização de dados segundo sua sensibilidade:
 - Identificadores: possibilitam identificar unicamente indivíduos (ex: CPF)
 - Semi-identificadores: quando associados às informações externas possibilitam a identificação (ex: data de nascimento)
 - Sensíveis: dados considerados confidenciais (ex: registros médicos)
 - Não sensíveis: dados que não causam qualquer ameaça se revelados

A partir da categorização pode-se definir estratégias caso necessário para garantir a confidencialidade dos dados através de técnicas de anonimização e pseudonimização.

- Na anonimização deve ser retirado do dado pessoal o seu elemento identificativo, para tanto deve-se tratar a privacidade sob duas perspectivas, a privacidade sintática e privacidade semântica:
 - Privacidade sintática: abrangem apenas dados numéricos e as técnicas recomendadas são:
 - Supressão: remoção de parte dos dados.



- Generalização: consiste em substituir os valores por outros menos específicos, mas mantendo a representação.
- Privacidade semântica: abrange dados com significado semântico e a técnica recomendada é:
 - Adição de ruídos: transforma ou perturba o conteúdo original dos datasets.
- Na pseudonimização deve haver uma separação entre os referidos atributos e a identificação do titular dos dados.

4. Diretrizes para repositórios de dados

- Vocabulários controlados

Deve-se adotar o uso de vocabulários controlados em qualquer fase do projeto além de evitar falsos cognatos e ambiguidades também colaboram para o alinhamento à outros projetos tanto internos quanto externos.

- Padrões de metadados

Deve ser escolhido e aplicado um padrão de metadados para descrição dos dados, isso é de extrema importância para garantir a qualidade da divulgação e interoperabilidade. Ele deve levar em consideração o tipo de recurso descrito, no caso dados de pesquisa, seu campo do conhecimento e público destinado.

- Diretrizes OpenAIRE

As OpenAIRE Guidelines for Literature Repository Managers v4 ou Diretrizes para gestores de repositórios de literatura versão 4 devem ser seguidas e descrevem um perfil de aplicabilidade para repositórios, visando a compatibilidade ao OpenAIRE através de metadados e o protocolo OAI-PMH (que é um protocolo de comunicação). Essas diretrizes também são aplicadas à repositórios de dados. Essas diretrizes são descritas incluindo um perfil de aplicação em <https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/latest/index.html>.

- Visibilidade



O repositório deverá estar cadastrado nos diretórios nacionais e internacionais, para tanto deverá ser feita uma pesquisa sobre a existência dos mesmos.

Referências

AFFONSO, E. P.; OLIVEIRA, S. C. de; SANT'ANA, R. C. G. Análise do equilíbrio entre privacidade e utilidade no acesso a dados. *Informação & Sociedade*, v. 27, n. 1, 2017. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/29422/17414>. Acesso em: 02 out. de 2019.

ANGUES, Ivan. O ciclo de vida dos dados de pesquisa. 2018. Disponível em: <https://dadosdepesquisa.rnp.br/o-ciclo-de-vida-dos-dados-de-pesquisa/>. Acesso em: 02 out. de 2019.

AVENTURIER, Pascal. Dados de pesquisa : descrição do ciclo de vida de gestão de dados. 2016. Disponível em: <https://publicient.hypotheses.org/1359>. Acesso em: 02 out. de 2019.

COSTA, Maíra Murrieta. Diretrizes para uma política de gestão de dados científicos no Brasil. 2017. 288 f., il. Tese (Doutorado em Ciência da Informação)—Universidade de Brasília, Brasília, 2017.

BFE, Belmont Forum e-Infrastructures. A Place to Stand:e-Infrastructures and Data Management for Global Change Research. 2015. Disponível em: http://www.bfe-inf.org/sites/default/files/doc-repository/A_Place_to_Stand-Belmont_Forum_E-Infrastructures_Data_Management_CSIP.pdf. Acesso em: 02 out. de 2019.

BRASIL. Lei Geral de Proteção de Dados (LGPD). Lei n. 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei n. 12.965 de 23 de abril de 2014 (Marco Civil da Internet). Diário Oficial da União, Brasília, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8777.htm. Acesso em: 02 out. de 2019.

BRASIL. Política de Dados Abertos do Governo Federal . Decreto n. 8.777/2016, 23 de abril de 2014. Institui a Política de Dados Abertos do Poder Executivo federal. Diário Oficial da União, Brasília, 15 ago. 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 02 out. de 2019.

DCC, Data Curation Centre. Curation Lifecycle Model. Disponível em: <http://www.dcc.ac.uk/drupal/resources/curation-lifecycle-model>. Acesso em: 02 out. de 2019.

DE CAPITANI DI VIMERCATI, S. et al. Data privacy: Definitions and techniques. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, v. 20, n. 06, p. 793-817, 2012. Disponível em: <http://spdp.di.unimi.it/papers/ijufks2012.pdf>. Acesso em: 02 out. de 2019.

EMERSON, Claudia; et al. World Data System (WDS) Data Sharing Principles. 2015. Disponível em: https://www.icsu-wds.org/files/WDS_Data_Sharing_Principles_2015.pdf. Acesso em: 02 out. de 2019.

EMBRAPA, Empresa Brasileira de Pesquisa Agropecuária. Política de Governança de Dados, Informação e Conhecimento da Embrapa. Disponível em: <https://www.embrapa.br/documents/10180/1600893/Pol%C3%ADtica+de+Governan%C3%A7a+de+Dados%2C+Informa%C3%A7%C3%A3o+e+Conhecimento+da+Embrapa/96873780-8e04-b67d-be7c-a0b6a4318556>. Acesso em: 14 ago. 2019.

FAPESP. Fundação de Amparo à Pesquisa do Estado de São Paulo. Plano de Gestão de Dados. Disponível em: <http://www.fapesp.br/gestaodedados/>. Acesso em: 14 ago. 2019.

GUANAES, Paulo Cezar Vieira (Org.). Marcos legais nacionais em face da abertura de dados para pesquisa em saúde: dados pessoais, sensíveis ou sigilosos e propriedade intelectual. Rio de Janeiro: Fiocruz, 2018. 123 p.

JRC, Joint Research Centre. **Open Data Principles**. Disponível em: <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC115832/kjna27163enn.pdf>. Acesso em: 02 out. de 2019.

MINISTÉRIO PÚBLICO, Secretaria de Logística e Tecnologia da Informação. Cartilha Técnica para Publicação de Dados Abertos no Brasil. Disponível em: <http://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>. Acesso em: 14 ago. 2019.

MONTEIRO, Elizabete Cristina de Souza de Aguiar, et. al. Repositório de dados científicos: aspectos sobre privacidade de dados. In: I Workshop de Informação, Dados e Tecnologia, UFSC, 2017. Disponível em: <https://repositorio.ufsc.br/handle/123456789/180294>. Acesso em: 14 ago. 2019.

PANTON PRINCIPLES, Principles for open data in science. Murray-Rust, Peter; Neylon, Cameron; Pollock, Rufus; Wilbanks, John. 2010. Disponível em: <https://pantonprinciples.org/index.html>. Acesso em: 02 out. de 2019.

PORTELA, Durval. Proteção de dados pessoais - Lei nº 13.709/2018. 2018. Disponível em: <https://www.pwc.com.br/pt/sinopse-legislativa/outros-assuntos/protecao-dados-pessoais-lei-13709-2018.html>. Acesso em: 02 out. de 2019.

SANTOS, Paula Xavier (Coord.). Livro Verde - Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Rio de Janeiro: Fiocruz, 2017. 141 p.

UC SANTA CRUZ, University Library. Research Data Management. 2014. Disponível em: <https://guides.library.ucsc.edu/datamanagement>. Acesso em: 02 out. de 2019.

UKRI. Guidance on best practice in the management of research data. 2015. Disponível em: <https://www.ukri.org/files/legacy/documents/rcukcommonprinciplesondatapolicy-pdf/>. Acesso em: 02 out. de 2019.

Anexo I – Orientações para os Princípios FAIR

- Encontrabilidade (qualidade e abertura de metadados)
 - Designar identificador persistente único
 - Descrito com metadados ricos
 - Ser registrado ou indexado em ferramentas de busca
 - Especificar nos metadados o identificador dos dados
- Acessibilidade (Disponível na web)
 - Serem recuperáveis pelo identificados usando protocolos de comunicação padrão
 - Protocolos abertos, gratuitos e universalmente implementáveis
 - O protocolo permite procedimentos de autenticação e autorização
 - Metadados acessíveis mesmo sem o registro
- Interoperável (Vocabulários controlados, referenciação)
 - Utilizar linguagem formal, acessível para representação do conteúdo
 - Usar vocabulários que seguem padrões FAIR
 - Incluir referências qualificadas para outros metadados e dados
- Reutilizável (Proveniência, licenças apropriadas e finalidade clara)
 - Pluralidade de atributos precisos e relevantes
 - Disponibilizar com licenças de uso claras e acessíveis
 - Associar com sua proveniência
 - Estar de acordo com os padrões da área de origem

Anexo II – Orientações PGD

O DMPonline é uma ferramenta online que ao preenchida gera um plano de gestão de dados em diferentes padrões automaticamente e gratuitamente. Ele foi desenvolvido pois além de ajudar a organizar e facilitar a representação dos dados cada vez mais e mais instituições financiadoras estão pedindo planos de gestão de dados dos pesquisadores como requisito para conceder fundos. Ela está disponível em <https://dmponline.dcc.ac.uk/>.

Caso a instituição não possua modelo de PGD disponível, pode ser utilizado o modelo da HORIZON 2020 disponível em: https://dmponline.dcc.ac.uk/template_export/1612436782.pdf, que também pode ser acessado pela ferramenta mencionada. Nele é requisitado que se descreva: Um sumário dos dados; Como os dados irão se adequar à cada princípio FAIR; Alocação de recursos; Segurança de dados; Aspectos éticos; e Outros.



Anexo III – Orientações vocabulários controlados

Os Vocabulários controlados a se utilizar são os *da Confederation of Open Access Repositories* (COAR). A COAR ou Confederação de repositórios de acesso aberto é uma associação internacional que age como voz para a comunidade de repositórios agindo em questões como políticas e práticas. Ela desenvolveu três vocabulários: *Resource Type* ou Tipo de recurso, *Access Rights* ou Direitos de acesso e *Version Type* ou Tipos de versão. Cada vocabulário está disponível gratuitamente juntamente com um guia de implementação em <https://www.coar-repositories.org/news-updates/what-we-do/controlled-vocabularies/#35d2e41d9c1ac3c22>.

Anexo IV – Orientações para metadados

O Datacite desenvolveu um padrão para dados de pesquisa generalizados. O Datacite é uma organização internacional sem fins lucrativos que visa estabelecer um acesso mais fácil aos dados de pesquisa na Internet e aumentar a aceitação dos dados da pesquisa como contribuições legítimas e citáveis. Esse padrão, bem como exemplos de aplicação estão disponíveis em: <<https://support.datacite.org/docs/schema-40>>.

Caso seja necessária a descrição de datasets de campos específicos o *Digital Curation Centre* (DCC) desenvolveu uma lista com padrões de metadados por área do conhecimento que pode ser acessada em: <<http://www.dcc.ac.uk/resources/subject-areas/general-research-data>>.



Anexo V – Orientações para visibilidade

Os diretórios e registros disponíveis, são eles:

- OpenDOAR <<https://v2.sherpa.ac.uk/opensoar/>>;
- ROAR <<http://roar.eprints.org/>>;

E o específico para repositórios de dados:

- RE3DATA <<https://www.re3data.org/>>.