

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO MULTIDISCIPLINAR
PROGRAMA DE PÓS-GRADUAÇÃO INTERDISCIPLINAR EM
HUMANIDADES DIGITAIS**

DISSERTAÇÃO

**PRINCÍPIOS FAIR E PROVENIÊNCIA EM REPOSITÓRIOS DE DADOS
VOLTADOS PARA AGRICULTURA DIGITAL: UM ESTUDO EXPLORATÓRIO À
LUZ DAS HUMANIDADES DIGITAIS**

ALESSANDRA CASTRO FIORINI BESSA



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO MULTIDISCIPLINAR
PROGRAMA DE PÓS-GRADUAÇÃO INTERDISCIPLINAR EM
HUMANIDADES DIGITAIS**

**PRINCÍPIOS FAIR E PROVENIÊNCIA EM REPOSITÓRIOS DE
DADOS VOLTADOS PARA AGRICULTURA DIGITAL: UM ESTUDO
EXPLORATÓRIO À LUZ DAS HUMANIDADES DIGITAIS**

ALESSANDRA CASTRO FIORINI BESSA

Sob a orientação do Professor
Sérgio Manuel Serra da Cruz

E coorientação do Professor
Alexandre Fortes

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre em Humanidades Digitais** no Programa de Pós-graduação Interdisciplinar em Humanidades Digitais, Área de Concentração em Análise Qualitativa e Quantitativa de Dinâmicas Sociais.

Nova Iguaçu
Novembro de 2021

Universidade Federal Rural do Rio de Janeiro
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada
com os dados fornecidos pelo(a) autor(a)

B557p

Bessa, Alessandra Castro Fiorini, 1973-
Princípios FAIR e Proveniência em repositórios de
dados voltados para a Agricultura Digital: um estudo
exploratório à luz das Humanidades Digitais. /
Alessandra Castro Fiorini Bessa. - Rio de Janeiro,
2021.

88 f.: il.

Orientador: Sérgio Manuel Serra da Cruz.
Coorientador: Alexandre Fortes.
Dissertação (Mestrado). -- Universidade Federal
Rural do Rio de Janeiro, Programa de Pós-Graduação
Interdisciplinar em Humanidades Digitais, 2021.

1. Humanidades Digitais. 2. Agricultura Digital.
3. Dados FAIR. 4. Proveniência de dados. 5.
Repositórios digitais. I. Cruz, Sérgio Manuel Serra
da, 1965-, orient. II. Fortes, Alexandre, 1966-,
coorient. III Universidade Federal Rural do Rio de
Janeiro. Programa de Pós-Graduação Interdisciplinar em
Humanidades Digitais. IV. Título.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
DEPARTAMENTO DE COMPUTAÇÃO



TERMO Nº 138 / 2022 - DCOMP (11.39.97)

Nº do Protocolo: 23083.010206/2022-94

Seropédica-RJ, 16 de fevereiro de 2022.

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
INSTITUTO MULTIDISCIPLINAR
PROGRAMA DE PÓS-GRADUAÇÃO INTERDISCIPLINAR EM
HUMANIDADES DIGITAIS
ALESSANDRA CASTRO FIORINI BESSA

Dissertação submetida como requisito parcial para obtenção do grau de **Mestre em Humanidades Digitais**, no Programa de Pós-Graduação Interdisciplinar em Humanidades Digitais, Área de Concentração em Análise Qualitativa e Quantitativa de Dinâmicas Sociais.

DISSERTAÇÃO APROVADA EM: 30/11/2021

Conforme deliberação número 001/2020 da PROPPG, de 30/06/2020, tendo em vista a implementação de trabalho remoto e durante a vigência do período de suspensão das atividades acadêmicas presenciais, em virtude das medidas adotadas para reduzir a propagação da pandemia de Covid-19, nas versões finais das teses e dissertações as assinaturas originais dos membros da banca examinadora poderão ser substituídas por documento(s) com assinaturas eletrônicas. Estas devem ser feitas na própria folha de assinaturas, através do SIPAC.

Sérgio Manuel Serra da Cruz, Dr. UFRRJ - Orientador/Presidente da Banca

Marcelo Panaro de Moraes Zamith, Dr. UFRRJ - Avaliador Interno

Renato Cerceau, Dr. INC - Avaliador Externo

(Assinado digitalmente em 17/02/2022 12:53)
MARCELO PANARO DE MORAES ZAMITH
PROFESSOR DO MAGISTERIO SUPERIOR
DeptCC/IM (12.28.01.00.00.83)
Matrícula: 1581041

(Assinado digitalmente em 16/02/2022 16:54)
SERGIO MANUEL SERRA DA CRUZ
PROFESSOR DO MAGISTERIO SUPERIOR
DCOMP (11.39.97)
Matrícula: 362436

(Assinado digitalmente em 16/02/2022 20:05)
RENATO CERCEAU
ASSINANTE EXTERNO
CPF: 033.048.597-05

Para verificar a autenticidade deste documento entre em
<https://sipac.ufrj.br/public/documentos/index.jsp> informando seu número: **138**, ano:
2022, tipo: **TERMO**, data de emissão: **16/02/2022** e o código de verificação: **a43514d792**

DEDICATÓRIA

Dedico este trabalho a todos que um dia possam precisar dele. Que seja uma boa semente a germinar, crescer, dar bons frutos e gerar novas sementes, formando ciclos.

AGRADECIMENTO

Agradeço ao Pai, ao Filho e ao Espírito Santo de Deus, sem esse mistério, nada seria possível, nunca.

Agradeço ao meu esposo e a minha filha, por todos os momentos, por todo suporte... sem o apoio e compreensão de vocês, o quanto vocês ouviram, o quanto fizeram por mim e comigo, sem vocês não daria certo. E que período atípico! Só Deus sabe!

Agradeço ao meu mestre e professor orientador, Sérgio Serra, que aceitou esta aventura: “ao infinito e além!”

Agradeço ao meu professor e coorientador, Alexandre Fortes, que contribuiu com suas aulas relacionadas ao tema.

Agradeço ao Programa do Mestrado, PPGIHD/UFRRJ, pela oportunidade que nos levou a esta turma pioneira, onde tenho a satisfação e chance de apresentar a primeira (neste curso) pesquisa deste tema.

Agradeço aos professores da Banca Examinadora, que desde o primeiro momento, foram simpáticos ao convite e ao meu tema, trazendo contribuições muito valiosas para o sucesso da pesquisa.

Agradeço a minha irmã e primas, que se cansaram de me ouvir perguntar: “Vê se tá bom?”

Agradeço a minha colega de trabalho e profissão, por ter me avisado sobre a abertura do processo seletivo, eu nem sabia.

Agradeço a cada pessoa da minha convivência, que em algum momento me transmitiu uma palavra de incentivo ou me cedeu um ouvido amigo.

E por fim, não menos importante, agradeço a todos que de uma forma direta ou indireta, participaram deste meu processo, inclusive, você, leitor deste momento, por ter interesse neste trabalho que foi produzido com muito esforço, dedicação e amor.

Meu muito obrigada!

RESUMO

BESSA, Alessandra Castro Fiorini. **Princípios FAIR e Proveniência em repositórios de dados voltados para a Agricultura Digital: um estudo exploratório à luz das Humanidades Digitais**. 2021. 88 p. Dissertação (Mestrado em Humanidades Digitais). Instituto Multidisciplinar, Universidade Federal Rural do Rio de Janeiro, Nova Iguaçu, RJ. 2021.

A humanidade, desde sempre, teve necessidade de se comunicar. Registros de pinturas rupestres, surgimento da escrita cuneiforme, uso de suportes como papiro, aparecimento do alfabeto, meios de comunicação como rádio, telefone e televisão, além de jornais e revistas, computadores e, posteriormente, a *internet*, são saltos evolutivos muito significativos neste processo de comunicação. No mundo contemporâneo, a gigantesca quantidade de informação digital existente demanda organização para que estes dados não se percam e ao mesmo tempo, possam ser usados e validados como seguros e passíveis de credibilidade, já que muitos resultados de pesquisas são totalmente dependentes do que hoje podemos encontrar na *web*. O dilúvio de dados que vamos vivenciando a cada dia faz com que a sociedade atual esteja assoberbada com dados que vêm de todos os lugares, relacionados a instrumentos, experimentos, pessoas e rede sociais. A fim de organizar a produção científica, surgem os Repositórios Digitais, institucionais ou temáticos. No contexto das Humanidades Digitais e seus impactos sociais, esta pesquisa destina-se ao estudo do Repositório Digital da plataforma *OpenSoils*, e o armazenamento de grandes volumes de dados na área de segurança de solos, visando ao desenvolvimento sustentável e as grandes demandas da Agricultura Digital. As Humanidades Digitais podem se conectar com os setores mais digitalizados da grande cadeia do agronegócio, por conta de suas práticas de gestão e análises de dados, reverberando em impactos sociais que a agricultura promove. Considerando o aspecto interdisciplinar desta pesquisa, o impacto social promovido pela agricultura e a emergência das Humanidades Digitais no Brasil, investigar correlações é uma das questões mais importantes. Estas devem ser analisadas visando compreender como um funcionamento encadeado destas etapas da cadeia poderão se beneficiar desta investigação, considerando principalmente os grandes esforços em relação à gestão de dados ligados à sustentabilidade e aos solos – uma das propostas da plataforma *OpenSoils*. Desta forma, o problema de informação dispersa, difícil de ser encontrada, acessada ou reusada e ainda sem origem confiável, dentro dos repositórios de dados na área de solos, é uma problemática que investigamos. A plataforma *OpenSoils*, com seus aportes para dados mais FAIR e com Proveniência de Dados pode contribuir para esta solução, dentro da perspectiva das Humanidades Digitais.

Palavras-chave: Humanidades Digitais. Agricultura Digital. Dados FAIR. Proveniência de Dados. Repositórios Digitais.

ABSTRACT

BESSA, Alessandra Castro Fiorini. **FAIR Data Principles and Data Provenance in Digital Agriculture repositories: an exploratory study in light of Digital Humanities**. 2021. 88 p. Dissertation (Master Science in Digital Humanities). Instituto Multidisciplinar, Universidade Federal Rural do Rio de Janeiro, Nova Iguaçu, RJ. 2021.

Societies always needed to communicate. Records of cave paintings, the emergence of cuneiform writing, use of supports such as papyrus, the appearance of the alphabet, means of communication such as radio, telephone, and television, as well as newspapers and magazines, computers, and, later, the internet, are significant evolutionary leaps in this communication process. In the contemporary world, the huge amount of existing digital information demands organization so that this data is not lost and, at the same time, can be used and validated as safe and credible since many search results are totally dependent on what we can find on the web today. The deluge of data that we are experiencing every day means that today's society is overwhelmed with data that come from everywhere, related to instruments, experiments, people, and social networks. To organize scientific production, Digital Repositories emerge, institutional or thematic ones. In the context of Digital Humanities and their social impacts, this research intends to study the Digital Repository of the *OpenSoils* platform and the storage of large volumes of data in soil security, aiming at sustainable development and the great demands of Digital Agriculture. Digital Humanities can connect with the most digitized sectors of the large agribusiness chain, due to their management practices and data analysis, reverberating in social impacts that agriculture promotes. Considering the interdisciplinary aspect of this research, the social impact promoted by agriculture, and the emergence of Digital Humanities in Brazil, investigating correlations is one of the most important issues. This shall be analyzed in order to understand how the interconnected functioning of these steps in the chain can benefit from this investigation, mainly considering the great efforts in relation to data management related to sustainability and soils - one of the proposals of *OpenSoils* platform. Thus, the problem of dispersed information, difficult to be found, accessed, or reused and still without a reliable source, within data repositories in the soil area, is a problem that we investigated. The *OpenSoils* platform, with its contributions to more FAIR data and with Data Provenance, can contribute to this solution, from the perspective of Digital Humanities.

Keywords: Digital Humanities. Digital Agriculture. FAIR Data. Data Provenance. Digital Repositories.

ABREVIACOES E SIGLAS

ANDS	Australian National Data Service
ASBRAAP	Associao Brasileira de Agricultura de Preciso
AD	Agricultura Digital
BD	Banco de Dados
C4AI	Centro de Inteligncia Artificial da USP
CABI	Centre for Agriculture and Biosciences International
CARE	Collective benefit, Authority to control, Responsibility, Ethics
CODATA	Committee on Data International Science Council
CPDOC	Centro de Pesquisa e Documentao de Histria Contempornea do Brasil
CGIAR	Consultative Group on International Agricultural Research
DCMI	Dublin Core Metadata Initiative
DOI	Digital Object Identifier
DTL	Digital Life Sciences
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuria
ENAP	Escola Nacional de Administrao Pblica
FAIR	Findable, Accessible, Interoperable, Reusable
FAO	Food and Agriculture Organization
FAPESP	Fundao de Amparo  Pesquisa do Estado de So Paulo
FDP	Fair Data Points
FDT	Fair Data Train
FGV	Fundao Getlio Vargas
FIOCRUZ	Fundao Oswaldo Cruz
GIDA	Global Indigenous Data Alliance
GODAN	Global Open Data for Agriculture & Nutrition
HD	Humanidades Digitais
HTTP	Hiper Text Transfer Protocol
IBICT	Instituto Brasileiro de Informao em Cincia e Tecnologia
LARHUD	Laboratrio em Rede de Humanidades Digitais
LAVIHD	Laboratrio Virtual de Humanidades Digitais
LGPD	Lei Geral de Proteo de Dados Pessoais
LOD	Linked Open Data
MAPA	Ministrio da Agricultura, Pecuria e Abastecimento
MCTIC	Ministrio da Cincia, Tecnologia, Inovaes e Comunicaes
NISO	National Information Standards Organization
ODI	Open Data Institute
ODS	Objetivos do Desenvolvimento Sustentvel
OKF	Open Knowledge Foundation
ONU	Organizao das Naes Unidas
ORCID	Open Research and Contributor ID
OWL	Web Ontology Language

PGD	Plano de Gestão de Dados
PID	Persistent Identifiers
RD	Repositórios Digitais
RDA	Research Data Alliance
RDF	Resource Description Framework
ROR	Research Organization Registry
SBIAGRO	Simpósio Brasileiro de Informática na Agricultura
SIBCS	Sistema Brasileiro de Classificação de Solos
SPARQL	SPARQL Protocol and RDF Query Language
TADIRAH	Taxonomy of Digital Research Activities in the Humanities
TIC	Tecnologia de Informação e Comunicação
UFRRJ	Universidade Federal Rural do Rio de Janeiro
URI	Uniform Resource Identifier
USP	Universidade de São Paulo
W3C	World Wide Web Consortium
WDS	World Data System

DIAGRAMAS E FIGURAS

- Diagrama 1 Proveniência - modelo simplificado em W3C PROV-O
- Figura 1 Contextualização da multidisciplinaridade na pesquisa
- Figura 2 Mapa conceitual - Captura - TADiRAH
- Figura 3 Evolução das fases da agricultura
- Figura 4 17 ODS da Agenda 2030
- Figura 5 Ciclo de vida dos dados (DataONE)
- Figura 6 Ciclo de vida dos dados ligados
- Figura 7 Processo de FAIRificação
- Figura 8 O que são Dados FAIR
- Figura 9 Taxonomia da Ciência Aberta
- Figura 10 Proveniência como característica do princípio de Reutilização
- Figura 11 Repositórios multidisciplinares sob Princípios FAIR
- Figura 12 Visão conceitual da arquitetura *OpenSoils*
- Figura 13 Representação conceitual dos três ambientes de coleta de dados (*in situ*, *in vitro* e *in silico*) e fluxo de dados entre os experimentos de segurança de solos e exemplos da abordagem computacional adotada pelo *OpenSoils*
- Figura 14 Representação conceitual (estrela) com as etapas de um projeto de AD e correlação com os princípios FAIR e proveniência de dados e diagrama que representa as etapas ao eixo do tempo

QUADROS

Quadro 1	Iniciativas em Agricultura e Alimentação
Quadro 2	Consórcios e iniciativas em compartilhamentos de dados
Quadro 3	Expectativa de comportamento para os dados ligados
Quadro 4	Tipos de metadados
Quadro 5	Guia de Princípios FAIR e ANEXO 1, comparando o <i>OpenSoils</i>
Quadro 6	<i>Checklist</i> como base para Processo de FAIRificação do <i>OpenSoils</i>

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Motivação	17
1.2	Justificativa	21
1.3	Problema	25
1.4	Objetivo geral	25
1.5	Objetivos específicos	25
1.6	Organização da dissertação	25
2	REFERENCIAIS TEÓRICOS	26
2.1	Humanidades Digitais	26
2.2	Agricultura Digital	30
2.3	Ciclo de gestão do dado científico	35
2.4	Metadados e sua identificação	40
2.5	Princípios FAIR	44
2.6	Proveniência de dados	51
2.7	Repositórios de dados	54
2.8	Base de dados da plataforma <i>OpenSoils</i>	58
3	METODOLOGIA DE PESQUISA	60
4	TRABALHOS RELACIONADOS	62
4.1	Bancos de dados de solos	62
4.2	Outras iniciativas em AD	64
4.3	Farm Data Train	65
5	CONTRIBUIÇÕES	66
5.1	Correlação da HD com AD	66
5.2	Aspectos de FAIRificação de plataformas da AD	68
5.3	Os caminhos da FAIRificação da Plataforma <i>OpenSoils</i>	71
5.4	Artigos produzidos	77
6	CONCLUSÕES	78
	ANEXO 1	80
	REFERÊNCIAS	81

1 INTRODUÇÃO

Ao longo de milhares de anos, a humanidade tem atravessado diversos momentos que necessitaram refinar o processo de comunicação, fato que ocorre e se intensifica até os dias de hoje. A comunicação e a troca de informações são processos dinâmicos. Podemos citar as pinturas rupestres, as quais eram feitas em paredes de cavernas como uma das mais antigas maneiras de se registrar informação em suporte material como forma primitiva de linguagem. Esses elementos possibilitaram, a partir de sítios arqueológicos, permitiram descobertas e o desenvolvimento de pesquisas importantes ao longo da história.

O surgimento da escrita utilizando-se o alfabeto e o uso de suportes distintos, indo do barro passando pelo papiro e alcançando meios digitais, também foram um grande salto evolutivo e primordial para que registros pudessem ser feitos e transmitidos a outras gerações de modo cada vez mais veloz, promovendo assim, a disseminação da informação em escala global.

“A Escrita, como um sistema de signos [...] para exprimir graficamente a linguagem, constitui uma das grandes conquistas da Humanidade. Desde os primeiros signos conhecidos até chegar aos sistemas alfabéticos atualmente em uso, a escrita passou por inúmeras mudanças e transformações. ” (MUSEU VIRTUAL DA IMPRENSA, PORTUGAL)

Segundo Gomes (2011), a escrita causou uma revolução tão significativa nas comunicações, que os historiadores estabeleceram o encerramento da Pré-História e o nascimento da História no período em que o homem começou a escrever e a gerar dados persistidos em meios físicos.

A história da humanidade, afirma Charles Higounet (2003), se divide em duas imensas eras: antes e a partir da escrita. Todas as nossas sociedades baseiam-se sobre os escritos. A lei escrita substituiu a lei oral, o contrato escrito substituiu a convenção verbal, a religião escrita se seguiu à tradição lendária. E, sobretudo não existe história que não se funde sobre textos.

Por toda a história, os processos da transmissão de informação entre indivíduos vêm passando por diversas mudanças, utilizando-se de meios como telégrafo, telefone, rádio, televisão, satélites, computadores e todas as suas derivações até a contemporaneidade. Pierre Lévy (2020) informa que a nossa sociedade está em transformação rumo a uma “sociedade datacêntrica”.

Nos dias atuais, a gigantesca quantidade de informação digital, demanda equipes, organização, difusão e gestão de dados para que não se percam os dados e, ao mesmo tempo, possam ser usados e validados para serem passíveis de credibilidade, já que muitos resultados de pesquisas são totalmente dependentes do que hoje podemos encontrar na *web* ou em bases de dados com pouco controle.

Sayão e Sales (2016), Costa e Cunha (2014) e Medeiros (2011), usam o termo “dilúvio de dados”, para descrever essa a gigantesca quantidade de informação digital cotidiana. Os autores se alinham com Lévy quando afirmam que vivenciamos isto a cada dia, fazendo com que a sociedade atual esteja "assoberbada com dados que vêm de todos os lugares, relacionados a instrumentos, experimentos, pessoas e rede sociais".

Medeiros (2011), complementa com muita propriedade, afirmando o que tem sido visto tão claramente, a respeito do volume informacional que sequer conseguimos selecionar direito. Isto nos leva a confiar em motores de buscas que vão nos retornar resultados que talvez não sejam os melhores. Isso porque eles vão retornar aquilo que, estatisticamente, os outros buscam mais.

Hoje, existem várias formas, normas e técnicas de organização de dados digitais no Brasil e no mundo. Uma delas se dá a partir da reunião de informação de todo tipo e origem nos chamados Repositórios Digitais – RDs (IBICT, 2018), por exemplo, sendo organizadas portema ou área temática, com arquivos de diversos formatos.

Os Repositórios Digitais foram originalmente criados para facilitar o acesso à produção científica. São sistemas e bases de dados especialmente desenvolvidos para reunir, organizar e tornar mais acessível a produção científica dos pesquisadores. Podem ser institucionais ou temáticos. Os repositórios institucionais lidam com a produção científica de uma determinada instituição e os repositórios temáticos, com a produção científica de uma determinada área, sem limites institucionais (IBICT, 2018).

Uma iniciativa nacional muito expressiva nesta direção consolidou-se em dezembro de 2019 através da Agência FAPESP, com o estabelecimento da Rede de Repositórios de Dados Científicos do Estado de São Paulo, envolvendo no projeto seis Universidades públicas do Estado. A plataforma permite acesso aos dados gerados em pesquisas científicas, não somente aos pesquisadores, mas também ao contribuinte paulista, que paga por pesquisas feitas no

Estado. Essa forma de organização tem sido estruturada de modo alinhado ao Movimento de Acesso Aberto à informação científica (FIOCRUZ, 2020).

O movimento propõe mudanças estruturais na forma como o conhecimento científico é produzido, organizado, compartilhado, encontrado e reutilizado. É um novo modo de fazer ciência, mais colaborativo, mais transparente e sustentável. Segundo Moura e Araújo (2020), esse movimento poderá trazer impactos positivos para a área de Humanidades Digitais.

As Humanidades Digitais estão situadas na área de pesquisa contemporânea e multidisciplinar que apesar de ainda não ser um campo de estudo totalmente consolidado, tem interseção com diversas áreas do conhecimento, sendo notória por trazer benefícios amplos para a sociedade (CAVALCANTI; SALES; PIMENTA, 2018). Segundo Pimenta (2020), as HD têm crescido no Brasil desde o início dos anos 2000, elas atuam em questões e objetos ligados às diversas disciplinas das ciências humanas, sociais e sociais aplicadas e encontram apoio nos recursos e ferramentas oriundos da área das TIC, permitindo a possibilidade de desdobramentos para a produção de novos conhecimentos.

Segundo Cavalcanti, Sales e Pimenta, (2018) e Bessa e Cruz (2020), apesar da retórica da multidisciplinaridade estar presente na maioria das iniciativas de pesquisa em HD, a prática nos mostra que a interdisciplinaridade reduz sua abrangência real quando o assunto trata de acesso à informação e reuso de dados em RD em áreas emergentes, como por exemplo na Agricultura Digital (MASSRUHÁ *et al*, 2020).

Diante do contexto supracitado e aliado a necessidade de contribuir com o amadurecimento da pesquisa sobre RD nas áreas de Humanidades Digitais e compreender a emergência da Agricultura Digital no Brasil, esta pesquisa destina-se a estudar os fundamentos de RD e correlações com a perspectiva de criar repositórios para a área de AD.

A Agricultura Digital é uma área dinâmica e multidisciplinar que também se baseia em Tecnologia de Informação e Comunicação (TIC) e grandes volumes de dados, requer gestão de dados e se apoia fortemente nos repositórios digitais (MASSRUHÁ *et al.*, 2020). HD e AD são áreas muito amplas, ativas e recentes. Logo, escolhemos como elo, o conceito “Repositório Digital”.

Em especial, avaliaremos o Repositório Digital da plataforma *OpenSoils* (CRUZ *et al.*, 2018) à luz de conceitos das HD. Atualmente, ele é um banco de dados relacional capaz de

armazenar grandes volumes de dados curados na área de solos, utilizando dados pedológicos como sua base.

A pedologia é a “ciência da gênese, morfologia e classificação dos solos. Busca compreender a interação entre os fatores e processos de formação do solo e sua influência nos atributos morfológicos, físicos, químicos e mineralógicos do solo” (PEREIRA *et al.*, 2019). Segundo Coelho *et al.* (2014), os dados pedológicos são produzidos a partir de estudos sobre a distribuição, a gênese e evolução, a morfologia e a classificação dos solos bem como aspectos relacionados à aptidão agrícola, zoneamento e outras interpretações das terras para diversos fins.

Nesta pesquisa, estima-se compreender e correlacionar facetas ainda não explicitamente estudadas em Agricultura Digital sob a ótica das Humanidades Digitais, para aprimorar este repositório com vistas a que seja passível de sofrer processos como a FAIRificação, que de acordo com Sales (2021), significa tornar os dados compatíveis com os Princípios FAIR (WILKINSON *et al.*, 2016) e investigar os possíveis níveis de *FAIRness*, segundo Henning *et al.* (2019), ou seja, o quanto se está mais próximo dos Princípios norteadores (FAIR) no Banco de Dados escolhido.

1.1 Motivação

Segundo Alves (2016), as HD são uma comunidade de práticas e não apenas um conceito fechado em si e que necessita de uma problemática para existir e atuar. A partir desta premissa, entendemos que seria valiosa uma pesquisa sobre como contribuir no tema que também atravessou as fronteiras da transformação do espaço analógico para o terreno digital: a agricultura.

A transformação digital é um paradigma do nosso tempo. Trata-se um processo em que as mudanças alteraram fundamentalmente serviços, expectativas e comportamentos, pressionando as organizações e instituições a interromperem em vários mercados, revendo seus métodos e procedimentos; é um paradigma dinâmico e que, *a priori*, não pode ser previsto no horizonte do tempo, de acordo com Venkatraman (2017).

Percebemos que o processo de transformação digital afeta tanto as HD quanto a AD (ALLEA, 2020), e está deixando de ser uma opção, tornando-se um caminho imprescindível entrecortado por investigações multidisciplinares e que poderá beneficiar diversos segmentos

da sociedade brasileira, além de possibilitar que se descortinem novas linhas de investigações na academia.

A transformação digital foi alavancada pela era da Indústria 4.0 ou Quarta Revolução, afirmado por Nascimento (2019), e vem demandando sistemas computacionais que integram o digital e o físico para satisfazer as novas necessidades das pessoas com novas tecnologias, fundindo o físico, digital e biológico, impactando todas as disciplinas, economias e indústrias, e até mesmo desafiando ideias sobre o que significa ser humano.” (TRANSFORMAÇÃO DIGITAL, 2020)

No cenário internacional, começam a surgir as primeiras iniciativas que buscam compreender como a transformação digital se materializa nas HD e na agricultura. Essencialmente, se busca compreender qual a relação entre diferentes setores ligados ao agronegócio e os grandes *datasets* produzidos por práticas da AD (KLERSKX et al, 2019).

Abaixo, vejamos alguns exemplos, concentrados na área de Agricultura e Alimentação, com objetivo de disponibilizar dados agrícolas relevantes, com programas de pesquisas agroalimentares, trabalhando em defesa de dados abertos, promovendo mudanças significativas em âmbito mundial:

Instituição	Exemplos de Iniciativas/Projetos – Casos
GODAN	Iniciativa de Dados Abertos na Agricultura – A qualidade do solo influencia muitas decisões em nível de fazenda, incluindo a seleção de safras e a necessidade de entrada e gestão da terra. Muitos governos coletam informações do solo para obter um melhor entendimento das condições ambientais em diferentes áreas do país. No entanto, a confiabilidade dos dados do solo depende da precisão e do nível de detalhe das informações. Geralmente as informações são grosseiras (resolução de 1 km ou 250 m), tornando-as relevantes apenas para aconselhamento agrícola generalizado ou planejamento estratégico. Um mapa de solo detalhado fornece uma melhor compreensão das características do solo, resultando em uma melhor seleção de culturas, conselhos direcionados e práticas de gerenciamento de fazendas. Os dados do solo são úteis para o planejamento de safras e safras e para informar as avaliações de risco. (GODAN...2021)
CGIAR	Plataforma Global de Conhecimento de Árvores – A <i>Global Tree Knowledge Platform</i> é sobre as árvores em paisagens 'arborizadas'. Seu objetivo é apoiar o melhor uso das espécies de árvores - promover a árvore certa no lugar certo para o propósito certo - para trazer maiores benefícios para os humanos e o meio ambiente. Dispõem de bancos de dados de árvores que incluem informações sobre o uso das árvores, ecologias e outros recursos. (CGIAR...2021)

ODI	Habilitando o acesso a dados para apoiar a inovação na agricultura – este projeto de dois anos se baseia em trabalho anterior na criação de ecossistemas de dados agrícolas FAIR (localizáveis, acessíveis, interoperáveis e reutilizáveis) e abertos. Tem como objetivo criar ferramentas, diretrizes e programas de aprendizagem que darão suporte ao gerenciamento, compartilhamento e governança de dados em ecossistemas de dados agrícolas. (ODI...2021)
-----	---

Quadro 1 – Iniciativas em Agricultura e Alimentação.

Verifica-se que essas iniciativas, por sua vez, estão se aproximando das temáticas estudadas pelos consórcios de gestão de dados, conforme quadro abaixo, e que ressaltam a importância da gestão adequada de grandes volumes de dados científicos tendo como elemento central a importância do ciclo de vida dos dados, a curadoria de dados, os Princípios FAIR¹ (WILKINSON *et al.*, 2016) e Proveniência de Dados² (BUNEMAN *et al.*, 2001), que serão abordados em seção específica desta pesquisa:

RDA	Criada em 2013 pela Comissão Europeia, a National Science Foundation e o National Institute of Standards and Technology do Governo dos Estados Unidos, e o Departamento de Inovação do Governo Australiano com o objetivo de construir a infraestrutura social e técnica para permitir o compartilhamento aberto e a reutilização de dados. (RDA...2016)
CODATA	Promove o avanço da Ciência Aberta e melhoria da disponibilidade e usabilidade dos dados para todas as áreas de pesquisa. Dados devem ser tão abertos quanto possível e tão fechados quanto necessário. O CODATA também trabalha para promover a interoperabilidade e a usabilidade de tais dados: os dados de pesquisa devem ser FAIR (<i>Findable, Accessible, Interoperable and Reusable</i>). (CODATA...2021)
GIDA	Criada com objetivo de gerir dados indígenas, criando o acrônimo para CARE, baseados nos princípios de Benefício coletivo, Autoridade para controle, Responsabilidade, Ética. Promovendo a soberania e governança de dados indígenas, afirmando os direitos e interesses dos povos indígenas nos dados, advogando por dados para o bem-estar autodeterminado dos Povos Indígenas e reforçando os direitos de se envolver na tomada de decisões de acordo com os valores indígenas e interesses coletivos. (GIDA...2021)

¹ Acrônimo para *Findable, Accessible, Interoperable, Reusable*. Disponível em: <https://www.go-fair.org/fair-principles/>

² Caminho pelo qual se possibilita que dados e metadados usados, permitam transmitir detalhes sobre o histórico de dados aos usuários. Disponível em: https://www.w3.org/Translations/DWBP-pt-BR/#data_provenance

GO-FAIR	Iniciativa com características <i>bottom-up</i> (funcionando a partir de um grande grupo, trabalhando em conjunto e articulando para tomada de decisão), autogovernada e orientada para as partes ou grupo de interesse e baseada nos Princípios FAIR. (GOFAIR...2021)
---------	--

Quadro 2 – Consórcios e iniciativas em compartilhamento de dados.

Manouselis (2016) questiona como as Humanidades se relacionam com as Ciências Agrárias. O autor informa que há um entrelaçamento essencial e que ainda são pouco investigados, ele alega que é possível explorar os relacionamentos, conflitos e contradições da agricultura contemporânea e os aspectos sociais, técnicos e humanos.

Em estudo recente, Bessa et al (2020) verificaram que já existem os primeiros sinais ligando as áreas de HD e AD, mesmo que fracos e ainda pouco explorados na academia e no próprio agronegócio, já se percebe a existência de entrelaçamentos pouco compreendidos entre as HD e AD. Estes, por vezes, podem ser explicáveis através da perspectiva da gestão de dados, dos princípios FAIR e da Proveniência de Dados. Segundo a autora, as HD podem se conectar com os setores mais digitalizados da cadeia do agronegócio por conta de necessidade de adotar práticas de gestão de grandes volumes de dados, adoção de RD e de análises de dados. A associação dessas práticas digitais pode reverberar em impactos sócio-econômicos-ambientais no meio físico.

De forma inicial, pode-se verificar que o agronegócio brasileiro envolve cada vez mais cadeias inovadoras e ricas em dados (MASSRUHÁ, 2019). Com o avanço da transformação digital no campo, esta cadeia demandará uma infraestrutura que poderá beneficiar-se de novos artefatos computacionais trazendo mais benefícios para os diversos tipos de usuários presentes nesses processos. Por exemplo, a aquisição de insumos com uso da tecnologia digitais, produção intensiva de *commodities* de maior qualidade, beneficiamentos pós-produção gerando novos agroprodutos, a distribuição com seu *timing* adequado, até finalmente, a fase em que o produto chega ao consumidor final, são partes essenciais da cadeia do agronegócio centrado no dado (SNA, 2020; EMBRAPA, 2018).

Considerando o aspecto multidisciplinar desta dissertação, os possíveis desdobramentos futuros em aspectos socioeconômicos da agricultura e a emergência da AD e das HD no Brasil, a iniciativa de investigar novas correlações é uma das questões mais importantes.

Estas correlações podem ser analisadas considerando vários prismas. No caso dessa dissertação, optamos pelo prisma da gestão de dados ligados, dos solos do Brasil.

Este trabalho é de natureza eminentemente exploratório e multidisciplinar, visa a investigar a existência das interconexões entre as HD (ALVES, 2016; CAVALCANTI *et al.*, 2018) e a AD (MASSRUHÁ *et al.*, 2020; EMBRAPA, 2020) e ferramentas da área de Tecnologia da Informação e Comunicação (TIC), tendo como ponto de partida o banco de dados da plataforma *OpenSoils* (www.opensoils.org) (CRUZ *et al.*, 2018).

Temos como meta ampliar o compartilhamento e reuso de dados de solos do Brasil. Por esse motivo iremos necessitar de apoio de temas tradicionais da TIC, por exemplo: Proveniência de Dados e Princípios FAIR (BUNEMAN *et al.*, 2001; WILKINSON *et al.*, 2016).

De modo muito resumido, os dados de solos são um dos motivadores dessa pesquisa, como mencionado anteriormente. Os solos têm uma ligação direta com o tripé: sustentabilidade, sociedade civil e desempenho agronegócio no Brasil (COELHO *et al.*, 2014). Logo, são um elemento central que demandam estudos técnicos mais aprofundados e ações de conscientização sobre seu uso responsável e sua conservação por partes da sociedade.

1.2 Justificativa

A justificativa deste trabalho surgiu em função de um cenário complexo que envolve por um lado as demandas da agência de fomento CNPq e por outro lado as necessidades de transformação digital nas HD demandadas pela sociedade moderna (em tempo, estas foram demasiadamente aceleradas pelo advento da COVID-19) (OECD, 2020).

Do ponto de vista do CNPq, a dissertação se alinha com o projeto de pesquisas denominado “OPENSOILS: USANDO E-CIÊNCIA E CIÊNCIA DE DADOS EM PROJETOS DE SEGURANÇA DE SOLOS DO BRASIL”, processo número: 315399/2018-0.

Por outro lado, a transformação digital e o avanço de novas tecnologias de gestão de dados e *big data* – conjunto de propriedades que começam com 'v': volume, velocidade e variedade [big], veracidade, visualização e valor [data] (MEDEIROS, 2019), sobretudo na HD e na AD, justifica novos estudos. Atualmente, há a necessidade de elevada produção agropecuária baseada em princípios que não sejam dissociados da sustentabilidade ambiental para atender às crescentes demandas de alimentos e fibras pela população mundial (FAO, 2018).

A pesquisa também se justifica se consideramos a Agenda 2030, com seus 17 ODS³, onde as formas de produção integrada (com impactos reduzidos sobre o meio ambiente e recursos naturais utilizados na produção) tem papel importante. Destacamos dois ODS, segundo nossa percepção, que são centrados em gestão e análise de grandes volumes de dados, conforme abaixo:

Objetivo		Breve descrição
2	Fome zero e agricultura sustentável	Ligado à necessidade de acabar com a fome, alcançar a segurança alimentar viado à melhoria da nutrição, promovendo a agricultura sustentável, apoiando-se em características de sustentabilidade e preocupando-se em garantir a qualidade e segurança do solo.
12	Consumo e produção responsáveis	Ligado à necessidade de padrões de consumo e produção sustentáveis, visando à gestão sustentável e uso de recursos naturais de forma eficiente. A partir do manejo ambientalmente saudável de produtos químicos e resíduos, redução significativa da liberação destes para o ar, água e solo, para minimizar seus impactos negativos sobre a saúde humana e o meio ambiente.

Objetivos 2 e 12 (resumidos) – (<http://www.agenda2030.org.br>)

O grande volume de dados, em crescimento exponencial, e a necessidade de sua manipulação (extração de conhecimento) ratificam a necessidade de uso de novas investigações e criação de ferramentas computacionais. O objetivo é compreender propriedades tais como interoperabilidade, reprodutibilidade e reuso por humanos e máquinas, de modo confiável nas práticas das HD e AD.

Este trabalho também se justifica pelas necessidades de avaliar as aproximações entre as Humanidades Digitais e Ciências Agrárias, áreas importantíssimas no âmbito da UFRRJ. Consideramos as possibilidades de existirem novas práticas e ferramentas computacionais e tecnologias (ainda não exploradas) que permitirão atender às demandas da AD tendo o ser humano como elemento central.

³ Plano de ação para as pessoas, o planeta e a prosperidade, cujos 17 objetivos visam à erradicação da pobreza e promoção da vida digna para todos. Disponível em: <http://www.agenda2030.com.br/>

No contexto da multidisciplinaridade, as Humanidades Digitais, a Computação e as Ciências Agrárias compartilham características e peculiaridades onde a plataforma *OpenSoils* se posiciona numa interseção, conforme representado na Figura 1.

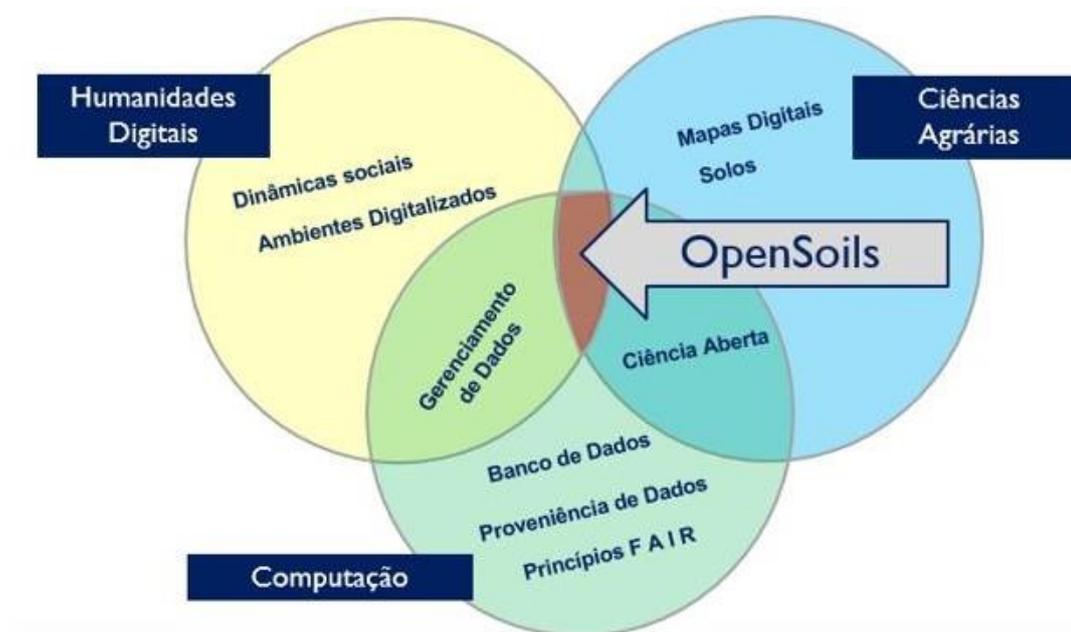


Figura 1 – Contextualização da multidisciplinaridade na pesquisa (BESSA; CRUZ, 2020)

Na AD, busca-se cada vez mais a capacidade de tomada de decisão, principalmente sobre enfrentar a crescente demanda mundial de alimentos e o uso racional do solo, com o menor impacto ambiental possível:

A expectativa de a população mundial atingir 9 bilhões de habitantes em 2050, demandando uma quantidade crescente de alimentos, impõe um grande desafio para a agricultura, como a necessidade de aumentar a produtividade na mesma área plantada com redução de custos e respeitando a conservação dos recursos naturais. Ao mesmo tempo, eventos climáticos extremos afetam a produção agrícola, consumidores exigentes demandam alimentos mais nutritivos, funcionais e produzidos de forma sustentável, a população concentra-se cada vez mais nos centros urbanos e torna-se mais longeva, enquanto cresce o deslocamento dos indivíduos do campo para as cidades e o envelhecimento da população rural. Além disso, surge a ameaça do corona vírus impactando todos os setores, desde a saúde, a educação e o agronegócio até a economia, entre outros. Para superar todos esses desafios são necessários a geração e o uso de novas tecnologias, agregando mais valor em todos os elos das cadeias produtivas. (MASSRUHÁ *et al.*, 2020)

De acordo com a *Food and Agriculture Organization*⁴ (FAO), agência especialista das Nações Unidas, a população mundial tem expectativa de crescimento em até 9,6 bilhões, até 2050. Isso significa que a demanda por alimentos também crescerá.

A Agricultura de Conservação é um sistema de cultivo que promove perturbação mínima do solo (ou seja, plantio direto), manutenção de uma cobertura permanente do solo e diversificação das espécies de plantas. Aumenta a biodiversidade e os processos biológicos naturais acima e abaixo da superfície do solo, o que contribui para o aumento da eficiência do uso de água e nutrientes e para uma produção agrícola melhorada e sustentada. (FAO)

A *Food and Agriculture Organization* (FAO) afirma que com este crescimento populacional, cresce em 70% a demanda por alimentos. Já o Brasil seria responsável por 40% desse incremento e teria que dobrar tudo o que produz atualmente para atender à nova demanda, segundo o Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2019).

Este trabalho considera a importância da adoção de Princípios FAIR (WILKINSON *et al.*, 2016) e Proveniência de Dados (BUNEMAN *et al.*, 2001), para o enriquecimento dos Repositórios de Dados em plataforma voltada para a AD, neste caso, teremos como estudo de caso a plataforma *OpenSoils* (CRUZ *et al.*, 2018) que tem como missão contribuir para o aprimoramento do gerenciamento de dados na área de solos do Brasil.

A abertura de dados junto com o compartilhamento e reuso de *datasets* (arquivos de conjuntos de dados) por humanos e máquinas são temas centrais quando se estuda a temática da gestão de dados em inúmeras áreas do conhecimento.

Como um exemplo de prática, a iniciativa GODAN, em seu *website*, considera que “uma melhor compreensão das características do solo, resultando em uma melhor seleção de culturas, conselhos direcionados e práticas de gerenciamento de fazendas” pode colaborar em informação acerca de safras e planejamento de risco.

Quanto mais fácil de se localizar um *dataset* e acessá-los, maior será a possibilidade de ser reusado por humanos e por máquinas, ou seja, mais facilmente serão reutilizados tanto pelo autor ou colaboradores quanto por terceiros. Essa faceta ampliará a transparência e poderá assegurar mais ampla divulgação dos resultados de uma pesquisa, conferindo à mesma mais confiabilidade e sua reprodutibilidade.

⁴ Disponível em: <http://www.fao.org/about/what-we-do/en/>

1.3 Problema

A quantidade de conteúdo informacional desconhecido e sem metadados de proveniência ou com baixa confiabilidade, tem crescido cada vez mais na AD. Esta pesquisa irá trabalhar com o desafio da gestão do grande volume de dados considerando o cenário da transformação digital para afiançar que possam ser gerenciados adequadamente, promovendo seu compartilhamento e reuso por humanos e máquinas.

1.4 Objetivo geral

O objetivo geral dessa dissertação é elucidar as correlações entre as áreas de Humanidades Digitais e Agricultura Digital, levando em consideração a gestão de dados científicos segundo os princípios FAIR e Proveniência de Dados, aplicando esses saberes a um objeto de estudo, neste caso, o repositório de dados da plataforma *OpenSoils*.

1.5 Objetivos específicos

Os objetivos específicos dessa pesquisa são:

Realizar levantamento na literatura sobre os temas Humanidades Digitais e Agricultura Digital;

Compreender o repositório de dados da plataforma *OpenSoils*;

Compreender como aplicar os Princípios FAIR e Proveniência de Dados no repositório da plataforma;

Considerar o repositório digital da plataforma *OpenSoils* como base de experimentos para estudar conjuntos de métricas que possam ser avaliadas para indicar o nível de *FAIRness* dos objetos digitais a serem FAIRificados nas próximas versões da plataforma;

Investigar possíveis caminhos de preparar o repositório para integrá-lo ao repositório digital institucional e posteriormente, a outras instituições de pesquisa.

1.6 Organização da dissertação

A dissertação está organizada em capítulos. No primeiro capítulo, Introdução, são

apresentados os seguintes tópicos: Motivação, Justificativa, O Problema, Objetivos, Relevância da Pesquisa e a Organização da dissertação; no segundo capítulo é apresentado os Referenciais Teóricos da pesquisa; já o terceiro capítulo apresenta a metodologia da pesquisa; o quarto capítulo aborda trabalhos relacionados à temática da pesquisa; o quinto traz a contribuição e comparação com ferramentas e discussão; o sexto e último capítulo conclui; e, por fim, são disponibilizadas as referências bibliográficas utilizadas.

2 REFERENCIAIS TEÓRICOS

Neste capítulo apresentaremos assuntos correlacionados com essa pesquisa, os quais trazem conceitos e temas necessários à compreensão do problema que esta dissertação investiga. A multidisciplinaridade que permeia as práticas das HD, permite-nos costurar áreas e domínios periféricos para entendermos através de uma abordagem *top-down*, isto é, a partir de uma análise geral que permeia várias áreas e domínios, até chegarmos em uma análise específica, incluindo onde e como a pesquisa pode situar-se.

2.1 Humanidades Digitais

O termo Humanidades Digitais surgiu há pouco mais de uma década, representando uma forte ligação entre as investigações em Humanidades e a utilização de métodos e ferramentas das tecnologias digitais, segundo Alves (2016).

Esta denominação foi bem aceita e bem-sucedida pelo fato de reunir um conjunto de outros conceitos que descreviam esta junção, anteriormente, como Computação para as Humanidades, Informática Aplicada à História, Linguística Computacional, Patrimônio e Computação, Arte Digital, entre outras, como aponta Terras (2014).

Como em qualquer outra área, as definições nem sempre são consensuais, havendo opções que se encaixam em muitas condições. Outra excelente definição foi a cunhada pela professora da área das Ciências da Informação, na *University College London*, Susan Hockey (2007). Esta refere-se às Humanidades Digitais como uma área acadêmica interdisciplinar que fornece metodologias específicas da área das tecnologias digitais para serem incorporadas na investigação nas Humanidades como um todo.

Ainda segundo Hockey (2007), a computação em ciências humanas passou por um processo inicial importante na questão de trazer visibilidade à área, que foi liderado pelo padre

jesuíta Roberto Busa, em 1949, ao desejar preparar um *index verborum* de todas as palavras nas obras de Santo Tomás de Aquino e autores relacionados, totalizando cerca de 11 milhões de palavras da época medieval, em Latim.

Roberto Busa foi o primeiro a receber um prêmio em reconhecimento à esta aplicação de tecnologia da informação à pesquisa humanística, falando mais tarde (1998) sobre o potencial da *World Wide Web* para oferecer material acadêmico multimídia acompanhado por sofisticadas ferramentas de análise (BUSA, 1999).

Pode-se dizer que é “evidente a inclinação de que a noção geral sobre as Humanidades Digitais possui vertentes implícitas na direção dos estudos informacionais tão característicos da Ciência da Informação” (CASTRO; PIMENTA, 2018), e segundo Russell (2011), um dos grandes objetivos das HD é criar bases de dados com recursos digitais relevantes para as Humanidades. Isto inclui a captura, estruturação, documentação, preservação e disseminação dos dados; desenvolver metodologias que permitam gerar novos elementos derivados desses dados; gerar pesquisa e conhecimento para aumentar nossa compreensão das Humanidades.

Sendo assim, e considerando a necessidade tecnológica para contribuir com as questões em dinâmicas sociais, as Humanidades Digitais oferecem novas formas para articular as pesquisas nesse sentido, possibilitando criação, análise e disseminação do conhecimento.

Centros de pesquisa desenvolvida por múltiplas comunidades caracterizam lugares de aplicação específica desta abordagem. A construção de ciber-infraestruturas evolutivas que respondam a demandas reais são uma necessidade. Estas devem ser construídas de maneira interativa, apoiando-se sobre o reconhecimento de métodos e abordagens já comprovados pelas comunidades de investigação, segundo o Manifesto das Humanidades Digitais (2012).

De acordo com abordagens que se complementam, podemos ainda afirmar que:

As humanidades digitais compreendem o estudo do que acontece na intersecção das ferramentas da computação com artefatos de ferramentas culturais de todos os tipos e como essas ferramentas comuns podem ser usadas para fazer novos conhecimentos do nosso conhecimento de herança cultural e do mundo contemporâneo. (COLEMAN, 2007, tradução nossa)

Historicamente, pode-se afirmar que cada campo do saber vem passando pela influência da digitalização da informação e “após quinze anos da aceitação irrestrita dos sistemas computacionais conectados às redes avançadas de troca de dados, as estruturas de saber foram sendo lentamente tocadas e desafiadas.” (SILVA *et al.*, 2016)

No Brasil, já é possível identificar importantes iniciativas nesta área, como o Grupo de Pesquisas em Humanidades Digitais⁵, desde 2009, cuja origem deu-se na Universidade de São Paulo (USP), liderando projetos de pesquisa nas áreas da filologia e linguística computacional, ciência da informação, história da ciência e tradução, todos com o objetivo comum da extroversão do acervo de obras raras da Biblioteca para o ambiente digital e outros projetos atuais, focando na organização digital da informação.

Neste contexto de atuação, o LaViHD – Laboratório Virtual de Humanidades Digitais – desenvolve ambientes de apoio e ferramentas para acervos digitais e corpora eletrônicos, com um Corpus que deverá ser publicado em 2022. A frente principal de trabalho, atualmente, acontece junto ao Centro de Inteligência Artificial (C4AI) da USP.

Ainda nesta direção de pesquisas, um fator impulsionador deste campo, é a disponibilidade existente de muitos repositórios digitais, principalmente na área de Ciências Humanas e Sociais. A aplicação de técnicas computacionais de visualização, consulta e análise conduzindo a novos conhecimentos, abriu possibilidades também para que o Laboratório de Humanidades Digitais⁶ fosse criado em 2016, tendo como berço a expertise do Centro de Pesquisa e Documentação de História Contemporânea do Brasil (CPDOC) e da Fundação Getúlio Vargas (FGV), contemplando plataformas, acervos e mídias digitais; análise de texto, redes sociais e dados quantitativos através de tecnologias computacionais; visualização de dados; jogos digitais, difusão e educação patrimonial; ensino de História e de Ciências Sociais e dados abertos. O Laboratório disponibiliza em sua página, atividades realizadas e resumos de projetos de pesquisas para conhecimento da sociedade em geral.

Uma outra iniciativa brasileira muito consistente e muito significativa para a área de Humanidades Digitais é a atividade do Laboratório em Rede de Humanidades Digitais⁷ (LARHUD) do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Iniciativa do Grupo de Pesquisa Informação, Memória e Sociedade (IMeS), mantido na Coordenação de Ensino e Pesquisa (COEPE) que tem como objetivo incentivar a produção científica e tecnológica além de desenvolver ferramentas e metodologias mediadas no vasto campo das Humanidades intermediada pelas plataformas, ferramentas e registros digitais (IBICT).

⁵ Disponível em: <http://lavihd.fflch.usp.br/>

⁶ Disponível em: <https://cpdoc.fgv.br/laboratorios/lhud>

⁷ Disponível em: <http://www.larhud.ibict.br/>

Uma de suas grandes contribuições nesta área é a taxonomia TADiRAH – *Taxonomy of Digital Research Activities in the Humanities* – que emerge como instrumento apropriado à compreensão das *digital humanities* como "uma taxonomia compartilhada da pesquisa em humanidades digitais, na qual se pode identificar ‘objetivos e métodos’ (como captura e análise), ‘objetos’ (como dados, imagens, manuscritos), e ‘técnicas’ (como análise de cluster, codificação, modelagem de tópico) aproximando atores das HDs e suas práticas relacionadas.” Por definição:

a TADiRAH corresponde à classificação mais expressiva e fundamentada sobre as atividades de pesquisa digital nas humanidades, certificando senão todos, a parte mais significativa da estrutura das humanidades digitais, se qualificando como principal expoente taxonômico para área o que justifica sua abordagem enquanto recurso complementar à compreensão deste domínio. (LARHUD, 2021)

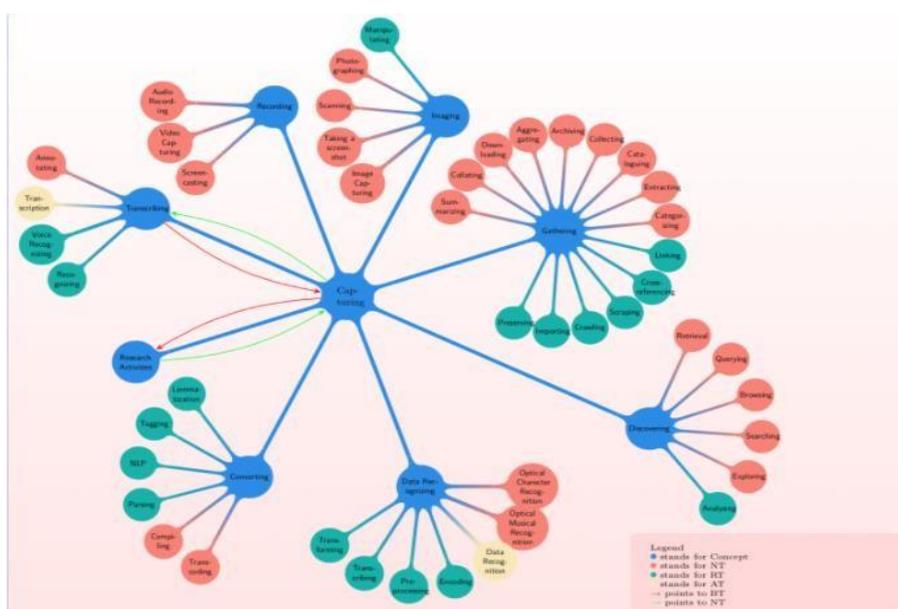


Figura 2 – Mapa conceitual - Captura - TADiRAH (Disponível em: <https://dhd-blog.org/?p=13108>)

Vocabulários controlados desempenham um papel fundamental, pois garantem interoperabilidade entre projetos e seus dados. Segundo Dodebei (2002), relações conceituais do ponto de vista epistemológico, podem ser analisadas segundo as ações mentais, comparando algo novo com conhecimento adquirido, produzindo relações de equivalência, hierárquicas, partitivas, de oposição e funcionais, por exemplo.

As análises sobre como o contexto digital tem influenciado a sociedade fazem-se mais necessárias à medida em que alguns estudos vão se descortinando, timidamente. As necessidades vão surgindo no campo das humanidades, a considerar história, literatura, filosofia, artes, educação, ciências sociais, música etc. Por outro lado, temos as ciências duras, com suas questões, investigação e rigores específicos. Entretanto, a característica interdisciplinar inerente às Humanidades Digitais amplia suas possibilidades de soluções práticas, colaborando nestas áreas.

Estudos em Humanidades Digitais com suas Associações estão presentes na América do Sul, por exemplo, no Brasil, em países da América Latina como México e Colômbia e em diversos países para além destas fronteiras, reunindo de 60 a 70 distintas Associações na Europa, Canadá, Austrália, Taiwan, Japão. Segundo o Gayol e Flórez (2017):

“La expectativa se resume en intentar construir un campo en el cual el diálogo horizontal entre la informática y las ciencias humanas permita construir nuevas interpretaciones que no podrían ser elaboradas de otra manera.”

É com este pensamento que pretendemos seguir nas análises desta pesquisa, contribuindo no campo das Humanidades Digitais, sob perspectiva das análises qualitativas nas dinâmicas sociais, correlacionadas ao uso de dados de solos.

2.2 Agricultura Digital

Quando se trata de agricultura, o Brasil tem um histórico muito importante e interessante – nos últimos 40 anos, nosso país saiu da condição de importador de alimentos para se tornar um grande provedor para o mundo, de acordo com pesquisas feitas pela EMBRAPA (2018).

Contextualizando, a atividade agrícola começou há milhares de anos, quando os homens percorriam lugares distantes em busca de alimentos ofertados pela natureza, além de serem movidos pela necessidade de caça e pesca. Devido às condições climáticas, por exemplo, podia haver fartura ou escassez e a busca continuava. O tempo foi passando e o homem foi percebendo que ao lançar sementes ao solo, elas germinavam, dando origem à novas plantas que produziram frutos e que estes serviram para sua alimentação (ARAÚJO, 2010 *apud* SENAR, 2015).

Podemos considerar que esse é o início da agropecuária, o que permitiu que o homem se fixasse em lugares preestabelecidos, deixando de se deslocar por grandes distâncias em busca de comida, um marco significativo na história, segundo o autor.

Portanto, o fato de o homem precisar de alimento gerou uma atividade que vem passando por diversos momentos de adaptação, a fim de suprir cada vez mais intensamente suas necessidades. O setor agrícola atravessou diversas etapas como, produzir para subsistência, modernizar-se fortalecendo relações com a indústria, especializar-se em determinadas atividades, movimento de êxodo rural, investimentos na tecnologia, investimentos e créditos facilitados para agricultura e até pela crise econômica (SENAR, 2015).

De acordo com Massruhá *et al.* (2020), a história agrícola no mundo deve ser comemorada como uma das grandes conquistas da humanidade. Se no início do século 20 tinha-se a agricultura 1.0, onde a força de trabalho era provida pela mão de obra das famílias, utilizando instrumentos manuais, ajudada pela tração animal, hoje temos a Agricultura 4.0.

Segundo a autora, a 1ª Revolução Industrial e o crescimento da população urbana, a demanda por alimentos aumentou, exigindo que os novos processos da produção agrícola também evoluíssem. Ou seja, desde essa época o método científico e as tecnologias avançadas já vem sendo aplicados à agricultura.

A agricultura brasileira era rudimentar em meados do século passado, por volta de 1950 e 1960. Prevalia o trabalho braçal na produção agropecuária. Porém, ela passou por grandes mudanças. O crescimento da agricultura exigia que extensas áreas naturais fossem convertidas em lavouras e pastagens, mas atualmente, tenta-se mudar esse cenário com o aporte de tecnologias, segundo Massruhá *et al.* (2020).

Ainda nesta linha, a autora explica que entre 1950 e 1960 (época de ineficiência no campo) existiam muitos problemas de abastecimento em todo o país, o qual atravessava um momento de transição para a industrialização e com escassez de alimentos. Esta fase originou a revolução verde que conseqüentemente levou a implantação de técnicas agrícolas mecanizadas, levando ao surgimento da agricultura 2.0, após a Segunda Guerra Mundial, no início do século XX.

Em 1973, a Embrapa foi criada para garantir segurança alimentar como principal atribuição, nessa época, o Brasil era grande importador de alimentos. Houve um progresso tecnológico produzindo intensificação nas atividades agrícolas, com tendências que levaram ao

surgimento da agricultura de precisão, dando início à agricultura 3.0. De acordo com a Associação Brasileira de Agricultura de Precisão, a atividade é definida como:

uma estratégia de gestão que reúne, processa e analisa dados temporais, individuais e espaciais e os combina com outras informações para apoiar as decisões de gerenciamento de acordo com a variabilidade estimada para melhorar a eficiência no uso de recursos, produtividade, qualidade, rentabilidade e sustentabilidade da produção agropecuária. (AsBraAP)

Desta forma, a adoção de tecnologias (agrárias, digitais e mecânicas) na Agricultura de Precisão favoreceu grandemente o avanço e até mesmo o fortalecimento do agronegócio nacional. A partir daí se explicitou a necessidade da coleta e análises de inúmeros tipos de dados, como informações de solos, clima, características das plantas e dos animais, aplicação de insumos, colheita, produção, entre outras. O grande volume de dados levantados por meio da agricultura de precisão constitui uma fonte de informação do campo (BERNARDI et al., 2014 apud MASSRUHÁ et al., 2020). Ainda de acordo com a autora, paralelamente a todas estas tecnologias e ao processo de transformação digital, originou-se o conceito da Agricultura Digital, levando a mais uma revolução tecnológica, a chamada Agricultura 4.0 ou Agricultura Digital.

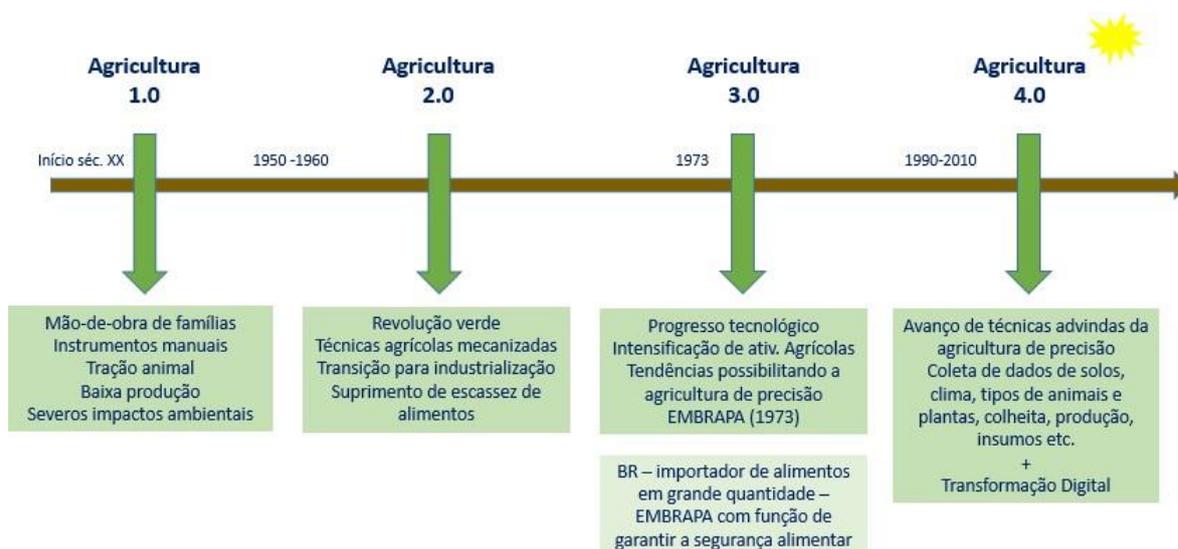


Figura 3 – Evolução das fases da agricultura (BESSA, 2021, apresentação)

As mudanças e adaptações nas ciências vêm ocorrendo muito rapidamente, isso não é exceção nas Ciências Agrárias. São muitas informações geradas ao mesmo tempo, em diversos lugares e com diversos resultados e aplicações, que podem ser usadas no dia a dia do produtor,

se este tiver alguma pré-condição, por exemplo, formação, financiamento, acompanhamento e treinamento, pois as tecnologias digitais muitas vezes são assimétricas e não são neutras por si só (BOYD; CRAWFOD, 2012).

O acesso aos recursos da Agricultura 4.0 são pouco mais complexos (e até mesmo excludentes para determinados segmentos produtivos), nesse caso temos:

Essas novas tecnologias e elementos promovem o uso mais sustentável dos recursos naturais, a mitigação dos impactos da mudança do clima e adaptação a eles, a conservação da diversidade biológica e utilização sustentável de seus componentes, bem como a repartição justa e equitativa dos benefícios derivados da utilização dos recursos genéticos, entre outros. Assim sendo, influenciam os esforços empenhados na intensificação e sustentabilidade da produção agrícola nacional. (EMBRAPA, 2018)

A Agricultura 4.0 revestida pela tecnologia e seus benefícios para determinados setores, pode favorecer a questão da sustentabilidade e dos negócios, onde a produção de alimentos deve ser apoiada em esforços para manter equilíbrio entre o meio-ambiente e tudo o que se utiliza para esta produção, bem como gestão sustentável e uso eficiente dos recursos naturais, conforme enunciado nos “17 Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos sob a coordenação da Organização das Nações Unidas (ONU), visando garantir, até 2030, um planeta mais equitativo e saudável.” (EMBRAPA, 2018)

Pode-se concluir que diante de tantos desafios, com tanta informação disponível e necessidades imperativas ditando os caminhos do desenvolvimento, a agricultura tradicional vem sendo transformada na Agricultura 4.0 ou Agricultura Digital.

Desta forma, assuntos com peso e relevância mundial, em especial, sustentabilidade e segurança de solos, podem ser devidamente alinhados aos temas centrais da Agenda 2030 da ONU, conforme abaixo, com destaques nos itens 2 e 12, por exemplo, como foi mencionado na seção da Justificativa.



Figura 4 – 17 ODS da Agenda 2030 (<http://www.agenda2030.org.br>)

Os solos são assunto central nas cadeias agroalimentares e industriais das sociedades modernas. Estão sendo produzidos grandes volumes de dados coletados por sensores, experimentos de campo e imagens de satélites e mapas, em especial, ajudando consumidores, agricultores, profissionais da área, pesquisadores e formuladores de políticas a tomar decisões mais inteligentes e racionais sobre usos, recomendações e mapeamentos dos solos nas áreas urbanas e rurais. (L'HÉNAFF; SMITH, 2018 *apud* MARINHO; PINHEIRO; BESSA, 2020).

Um outro exemplo sobre o papel central que correlaciona a conectividade do campo ao “*data deluge*” ou dilúvio de dados, na agricultura é destacado pelo *Open Data Institute* (ODI). Segundo o instituto com a melhoria no acesso aos dados, como novos (re)usos e compartilhamento de dados agrícolas será possível trazer maior inovação no setor e, em última instância, contribuir promover a sustentabilidade na agricultura e apoiar as tomadas de decisões importantes nas diversas fases da produção agrícola. Destacamos que o ODI também apoia a criação de um ecossistema de dados agrícolas que sejam abertos e aderentes aos Princípios FAIR, os quais veremos mais adiante.

Tais considerações ratificam o fato de que as novas tecnologias digitais e facilidade de acesso e comunicação via internet, além de aplicativos com interface amigável, colaboram para que o produtor (em qualquer escala) se torne mais receptivo às mudanças e deseje se adequar, aprendendo a atuar neste novo contexto.

Para Massruhá, a conectividade no campo aliada ao empreendedorismo e inovação, tem sido um assunto largamente discutido. Uma das mais novas iniciativas dentro desta seara, foi criada em agosto de 2019 – a Câmara do Agro 4.0⁸ – resultado de um acordo de cooperação técnica entre os Ministérios da Agricultura, Pecuária e Abastecimento (MAPA) e Ciência, Tecnologia, Inovações e Comunicações (MCTIC). “O objetivo da Câmara do Agro 4.0 é implementar ações destinadas à expansão da internet no meio rural, ao aumento da produtividade no campo, e à difusão de novas tecnologias e serviços inovadores nas propriedades rurais.” (MAPA, 2019).

A AD representa uma mudança de referência na área agropecuária. O secretário de Inovação, Desenvolvimento Rural e Irrigação do Mapa, Fernando Camargo destacou, durante a programação da 16ª Semana Nacional de Ciência e Tecnologia, “que a Agricultura Digital

⁸ Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/camara-do-agro-4-0-apresenta-as-principais-demandas-para-ampliar-o-uso-de-novas-tecnologias-no-campo>

representa uma mudança de paradigma na agropecuária e que a Câmara do Agro 4.0 é fundamental para revolucionar a forma de produzir no país”. Afirmou ainda que o Brasil é um dos principais *players* na produção de alimentos, mundialmente falando, basicamente graças à tecnologia e inovação. Neste cenário, é possível perceber a necessidade iminente desta mudança e atualização de técnicas no dia a dia da agricultura. O Plano Nacional de Internet das Coisas⁹ é um dos alvos para receber investimentos em tecnologia de informação, visando ao desenvolvimento e potencialização do uso de internet no campo, já que somente 5% da área de agricultura irrigável do país está conectada à internet, segundo estudo desenvolvido no Programa de Pós-graduação em Fitotecnia, da Escola Superior de Agricultura Luiz de Queiroz (Esalq/USP), afirma Albuquerque (2019).

Dentre vários desafios apresentados pela mudança de paradigma na AD, por limitações de escopo, focaremos apenas nos relacionados aos dados digitais sobre segurança dos solos, cujas características poderão ser acompanhadas pela plataforma a ser analisada nesta dissertação, o *OpenSoils*, com seus repositórios de dados na área de solos.

2.3 Ciclo de gestão do dado científico

A gestão do dado científico não pode ser dissociada do conceito de Ciência Aberta. Ciência Aberta preconiza o compartilhamento do conhecimento e abertura do processo de produção de conhecimento e da própria comunicação do conhecimento. Assim, além do acesso aos dados abertos (resultados de pesquisas), o conceito é amplo e pode contemplar práticas como: o compartilhamento e o acesso aos métodos e dados utilizados nas pesquisas, a abertura e publicação dos dados de pesquisa, o uso de repositórios de dados, e mesmo a participação dos cidadãos na definição de prioridades da agenda de pesquisa e no processo de produção e difusão de conhecimento (RNP, 2021; FIOCRUZ, 2021).

Adicionalmente, para que os dados sejam acessados, preservados e reusados em novas pesquisas, é necessário que se tenha um conjunto de boas práticas, norteados por um Plano de Gestão de Dados¹⁰ que nada mais é do que um “documento formal, dinâmico e vivo que

⁹ Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D9854.htm

¹⁰ Plano de Gestão de Dados - O PGD é um documento formal, dinâmico e atualizável com perguntas orientadoras que estimulam o pesquisador a planejar, de maneira intencional, como se dará a gestão de dados ao longo de todo o seu ciclo de vida - período que é maior que a execução e finalização da pesquisa. Por limitações

descreve detalhadamente como os dados de pesquisa serão tratados durante todo o ciclo de vida da pesquisa e possivelmente após a sua conclusão.” (VEIGA *et al.*, 2019).

Algumas questões importantes para a gestão de dados e persistência de dados, de acordo com Sant’Ana (2016), não podem ser dispensadas, como: Quais são os dados necessários a serem coletados? De quais fontes serão coletados? Em que formato estão os dados? Como identificar sua proveniência? São dados abertos? Estão sendo coletados dados que permitam que estes venham a ser identificáveis e recuperáveis em um momento futuro?

O autor levanta uma longa série de questões típicas da computação sobre o processo denominado de persistência dos dados. O autor destaca uma série de preocupações e aspectos que devem ser detalhadamente planejados. Dentre outras questões fundamentais desta fase o autor destaca: Quais são os dados disponíveis? Quais destes dados deverão ser armazenados? Qual a estrutura (física e lógica) poderá ser utilizada para o armazenamento? Como garantir a permanência dos dados complementares sobre a coleta para que se tenha garantido o contexto de sua obtenção? Estes dados podem representar um risco a privacidade dos indivíduos ou instituições neles referenciados de alguma forma? Como as partes de sua estrutura lógica serão interligadas e como serão mantidas as interligações com outros conjuntos de dados? Como garantir que os elementos que sustentam a sua qualidade sejam mantidos? Tem-se o direito de armazenar estes dados? Todos os aspectos que podem contribuir para sua encontrabilidade estão sendo armazenados? Todos os fatores para sua utilização ao longo do tempo estão sendo mantidos? (SANT’ANA, 2016)

Pensando nas novas características da ciência, onde a abertura dos dados é tão importante, bem como seu compartilhamento e reuso, é que se torna essencial o cuidado com este ciclo de vida dos dados. O ato de gerenciá-los de modo que eles se tornem utilizáveis com eficácia e recuperáveis para reutilização não é um processo simples. A Escola Nacional de Administração Pública - ENAP (2019) considera que o gerenciamento de dados “é o desenvolvimento, execução e supervisão de planos, políticas, programas e práticas que entregam, controlam, protegem e aprimoram o valor dos ativos de dados e informações durante todo o seu ciclo de vida.”

de escopo esse tema não será aprofundado nessa dissertação. Disponível em:
<https://mooc.campusvirtual.fiocruz.br/rea/ciencia-aberta/serie3/curso2/aula5.html>

Exemplos internacionais tais como o *DataOne* (2017) estabeleceu seu próprio modelo de ciclo de vida dos dados, para nortear as fases relevantes desta questão, conforme figura abaixo:

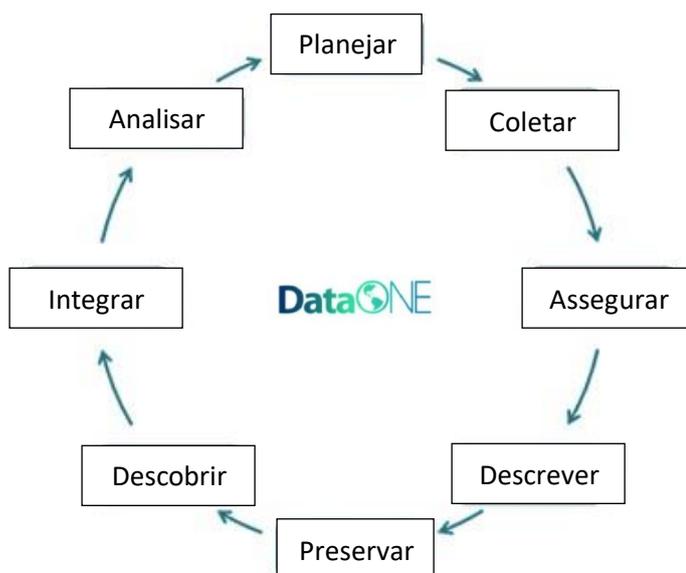


Figura 5 – Ciclo de vida dos dados (DataONE, *website*, adapt. e trad. nossa)

De acordo com a representação da Figura 5, o ciclo de vida de dados é um conjunto de etapas que os permitem serem úteis em sua totalidade e conforme as necessidades preestabelecidas nas políticas e diretrizes de um PGD. Fornecem uma visão geral de alto nível dos estágios envolvidos no gerenciamento e preservação bem-sucedidos de dados para uso e reutilização. Existem várias versões de um ciclo de vida de dados com diferenças atribuíveis à variação nas práticas entre domínios ou comunidades. (DataONE...2021)

Os dados abertos também se relacionam com o tema gestão de dados. Os dados abertos promovem o acesso à informação e tem sido o grande propulsor de desenvolvimento ao longo do século XXI. O movimento *Open Data* abriu precedentes inclusive para que a Lei de Acesso à informação nº 12.527, de 18 de novembro de 2011, fosse instituída no Brasil.

Segundo a *Open Knowledge Foundation*, os dados abertos são blocos de construção do conhecimento aberto, significando dizer que os dados se tornam úteis, utilizáveis e usados, ou seja, “Conhecimento aberto é qualquer conteúdo, informação ou dado que as pessoas são livres

para usar, reutilizar e redistribuir - sem qualquer restrição legal, tecnológica ou social.” (OKF, *website*).

Acerca de dados abertos, de modo a fazer sentido na *web* semântica, atendendo ao inciso da Lei de Acesso à Informação (que indicam a possibilidade de os dados serem processados por máquina, além da ligação entre informações de bases diferentes através de relacionamentos semânticos), percebe-se que os dados tornam-se, não apenas acessíveis e processáveis por máquinas, mas passíveis de processos de organização que podem facilitar a geração de novos dados, apresentação de resultados, relação com outros grupos de dados, aumento do conhecimento para tomadas de decisão, novos modelos de dados gerados a partir do relacionamento e cruzamento de dados de várias esferas governamentais, além da geração de novos modelos mentais de apresentação da informação de forma a facilitar o acesso dos dados pela sociedade civil, afirma Santarém (2015).

Considerando as características da *web* semântica, o “LOD (*Linked Open Data*) é um conjunto de melhores práticas para publicação e conexão de dados estruturados na *Web*, permitindo ligações entre itens de diferentes fontes de dados para formar um único espaço de dados global.” (HEATH; BIZER, 2011 *apud* SEGUNDO SANTARÉM, 2015)

Segundo Berners Lee (2006), trabalhar com *Linked Open Data* significa fazer ligações, possibilitando exploração de uma teia entre os dados, permitindo que estes sejam encontrados por pessoas ou máquinas, pois quando se tem um pouco de dados, é possível encontrar outros que estejam relacionados. Estes links são âncoras de relacionamentos.

O autor ainda descreve quatro regras, às quais chama de expectativa de comportamento e que possibilitam tornar os dados interconectados, dando ao usuário a possibilidade de reaproveitamento inesperado da informação (característica de valor agregado da *web*):

Descrição	Orientação sobre a descrição
Use URIs como nomes para coisas.	Conjunto de símbolos universal na área de <i>web</i> semântica, para identificação de coisas.
Use URIs sob HTTP para que as pessoas possam pesquisar esses nomes.	O não uso está ligado ao fato de não querer se comprometer com o Sistema de Nome de Domínio Estabelecido para delegação de autoridade, mas para construir algo sob controle separado.

Quando alguém procurar um URI , forneça informações úteis, usando os padrões (RDF, SPARQL) ¹¹	Em geral, é possível pesquisar as propriedades e classes encontradas nos dados e obter informações das ontologias RDF, RDFS e OWL, incluindo as relações entre os termos na ontologia.
Inclua <i>links</i> para outros URIs para que possam descobrir mais coisas.	<i>Links</i> em outros lugares, é necessário para conectar os dados que temos em uma web, uma web séria e ilimitada na qual podemos encontrar todo tipo de coisas, assim como na web de hipertexto que conseguimos construir.

Quadro 3 – Expectativa de comportamento para os dados ligados. (BERNERS-LEE, 2006, trad. nossa)

Como é possível perceber, Dados Ligados Abertos são essenciais na conexão da *web* semântica.

Na figura abaixo, identificamos etapas importantes que caracterizam o ciclo de vida dos dados ligados.

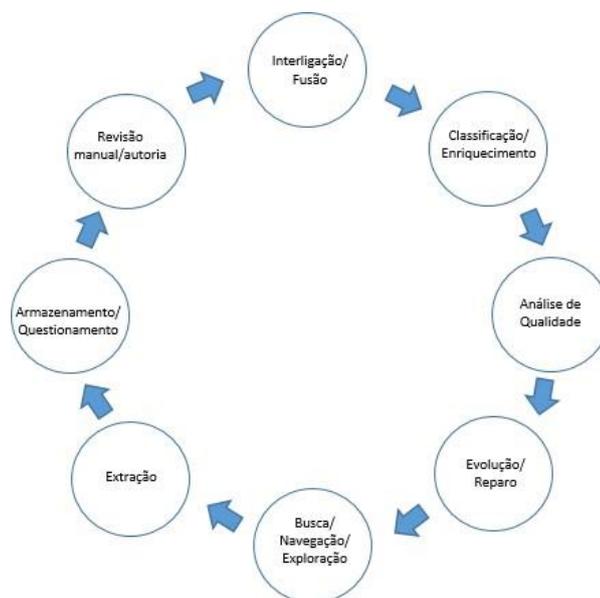


Figura 6 – Ciclo de vida dos dados ligados (AUER *et al.*, 2011, adap. e trad. nossa)

¹¹ RDF é uma linguagem de proposição geral para a Web, unificando dados de diversas fontes. SPARQL, uma linguagem de consulta para RDF, pode juntar dados de diferentes bancos de dados, bem como documentos, mecanismos de inferência ou qualquer outra coisa que possa expressar seu conhecimento como um gráfico rotulado direcionado. SPARQL é uma excelente linguagem para unificar dados em bancos de dados relacionais com outros bancos de dados, bem como essas outras fontes de dados.
Disponível em: <https://www.w3.org/2007/03/VLDB/>

Segundo o autor, o uso de dados ligados abertos oferece benefícios significantes, dentre os quais, uniformidade, coerência, integridade e oportunidade.

Os dados ligados abertos, tecnicamente, referem-se aos dados publicados na *Web*, tornando-se legíveis por máquina. Segundo Bizer et al. (2009), seu significado é definido explicitamente e está vinculado a outros conjuntos de dados externos, e que por sua vez, podem se vincular a outros conjuntos de dados externos também. Dados ligados “permitem que os cientistas combinem metadados públicos expressos como triplas de dados vinculados de maneira *ad-hoc*, sem a necessidade de primeiro coletar entradas de metadados e armazená-los localmente.” (MOREIRA *et al.*, 2019)

Esta compreensão do ciclo de vida dos dados importa para que se torne possível identificar a necessidade de guarda, compartilhamento e reuso de cada projeto e seus conjuntos de dados.

2.4 Metadados e sua identificação

O termo metadados surge num contexto em que são inúmeros os canais de publicação e disseminação de informação digital. Consequentemente, é necessário colocar ordem neste universo, de modo que o crescimento acelerado e caótico da *web* não seja um problema para identificação, recuperação e avaliação de uma infinidade de recursos neste ambiente digital. (MARCONDES, 2005).

De acordo com Segundo e Araújo (2019), os metadados explicitam os diferentes aspectos do recurso que descrevem. Ou seja, sua estrutura, seu conteúdo, sua qualidade, seu contexto, origem, propriedade e condição. Além disso, auxiliam na organização, favorecem a interatividade, validam as identificações e asseguram a preservação e principalmente, otimizam o fluxo informacional melhorando o acesso aos dados. Os metadados são vitais para se entender bem como armazenar recursos de dados, eles descrevem informações semânticas e sintáticas sobre o recurso, os metadados podem ser comparados com um sistema de rotulagem.

Desta forma, os metadados possibilitam uma forma de se descrever dados e repositórios de dados disponíveis na *internet* e mesmo nas redes corporativas. Representam um conjunto de elementos com significância padronizada, possibilitando descrição das informações eletrônicas (dados) ou recursos eletrônicos de maneira bibliográfica.

O *World Wide Web Consortium* (W3C) descreve metadados como informações sobre objetos da *web* compreensíveis por máquinas, com ênfase no processamento automático dos dados. Segundo Dempsey e Heery (1998) “metadados são dados associados a objetos que dispensam seus usuários potenciais de ter um conhecimento prévio completo de sua existência ou características. Um usuário pode ser um programa ou uma pessoa.”

Riley (2017) afirma que “metadados são a chave para a funcionalidade dos sistemas que mantêm o conteúdo, permitindo que os usuários encontrem itens de interesse, registrem informações essenciais sobre eles e compartilhe essa informação com outras pessoas.”

Felisberto (2016) informa que a Norma ISO 26324:2012 define que metadados são dados específicos associados com uma referência, com base em um modelo de dados estruturado persistindo a associação com os objetos de qualquer grau de precisão desejado e granularidade para suportar identificação, descrição e serviços.

O autor resume que a NISO (2004) conceitua metadados como informações estruturadas que descrevem, explicam, localizam, ou de outra forma torna mais fácil recuperar, usar ou gerenciar um recurso de informação. No ambiente de bibliotecas são comumente usados para qualquer esquema formal de descrição de recursos, aplicando-se a qualquer tipo de objeto, digital ou não digital.

Verifica-se na bibliografia, diversas categorias para metadados. Listamos abaixo, os tipos elencados pela autora da publicação NISO Primer Series (2017), baseados na categorização proposta pela NISO (2004):

Metadado descritivo	Para encontrar ou compreender um recurso
Metadado administrativo - Metadados técnicos - Metadados de preservação - Metadados de direitos	- Para decodificar e renderizar arquivos - Gerenciamento de arquivos a longo prazo - Direitos de propriedade intelectual associados ao conteúdo
Metadado estrutural	Relações de partes de recursos entre si

Quadro 4 – Tipos de Metadados (RILEY, 2017, trad. nossa)

Sayão (2010) corrobora com esta categorização de metadados, que é a seguida pela maioria dos autores que se dedicam à pesquisa sobre o assunto.

Metadados descritivos: é a face mais conhecida dos metadados, são eles que descrevem um recurso com o propósito de descoberta e identificação; podem incluir elementos tais como título, autor, resumo, palavras-chave e identificador persistente.

Metadados estruturais: são informações que documentam como os recursos complexos, compostos por vários elementos, devem ser recompostos e ordenados. Por exemplo, como as páginas de um livro, digitalizadas separadamente, são vinculadas entre si e ordenadas para formar um capítulo.

Metadados administrativos: fornecem informações que apoiam os processos de gestão do ciclo de vida dos recursos informacionais. Incluem, por exemplo, informações sobre como e quando o recurso foi criado e a razão da sua criação. Nessa categoria, estão metadados técnicos que explicitam as especificidades e dependências técnicas do recurso; inclui também os metadados voltados para apoio à gestão dos direitos relacionados ao recurso. (SAYÃO, 2010)

Arellano (2004) afirma que as bibliotecas empregam um grande esforço no trato de coleções digitais visando a interoperabilidade de acervos digitais e um dos grandes fatores que contribui no gerenciamento desses arquivos, são arquitetura, metadados e formatos padronizados. O autor cita a importância de padronizar os metadados de preservação, eles são uma forma especializada de administrar metadados que podem ser usados como um meio de estocar a informação técnica que suporta a preservação dos objetos digitais.

Sayão e Marcondes (2008) destacam que o acesso integrado e transparente a recursos informacionais heterogêneos, armazenados em bibliotecas digitais e repositórios digitais distribuídos e gerenciados de forma autônoma por instituições diferentes, é um dos grandes desafios das pesquisas, os autores também destacam que é de suma importância o uso dos metadados nas bibliotecas e repositórios digitais, para uma boa gestão dos dados de uma pesquisa.

Uma das formas de preservar e gerir dados de pesquisa em bibliotecas digitais e repositórios digitais é através da adoção de padrões de metadados. Um dos exemplos de padrão é o *Dublin Core*. O padrão foi criado a partir da iniciativa *Dublin Core Metadata Initiative* (1994), em Chicago, Estados Unidos, durante a segunda Conferência Internacional sobre a *Web* e uma reunião em Ohio (1995), após a qual, originou-se o conjunto de padrões mais utilizado para a descrição de documentos eletrônicos, simples e auto-explicativo, conforme Morato e Moraes (2010).

Além de metadados é necessário identificar os dados propriamente ditos. Muitas vezes se adota o conceito de identificação de dados (ou *datasets*) através de identificador persistente.

No âmbito de objetos digitais, encontramos ainda os identificadores de documentos, por exemplo, os *Digital Object Identifier* (DOI), representando um “padrão para identificação de documentos em redes digitais. Composto por números e letras, atribuído a um objeto digital para que este seja identificado de forma única e persistente no ambiente Web” usados para livros, capítulos de livros, periódicos, artigos etc (FERREIRA *et al.*, 2015).

Essas características são necessárias e importantes, acerca de persistência por parte das instituições que irão trabalhar com esses objetos digitais, entretanto, não se pode deixar de mencionar como parte desse apoio, a tecnologia *Handle System* (<http://handle.net/>):

Sistema distribuído de grande alcance, voltado para identificação persistente de objetos digitais e outros recursos na Internet - independente de localização - 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 URN –Uniform Resource Names. (SAYÃO, 2007)

O *Handle System* é um sistema com grande amplitude e pronto para a interoperabilidade. Segundo Sayão (2007), essa conversão em nomes pode ser transformada em informação necessária para localizar e acessar objetos na *internet*. O sistema foi criado inicialmente no âmbito de fornecer estrutura básica para bibliotecas digitais e documentos eletrônicos, no entanto, sua abrangência tornou-se tão ampla, de forma que pode ser usado como base para várias aplicações neste contexto.

Na categoria de identificadores de autores, por exemplo, encontramos o *Open Researcher and Contributor ID – ORCID*, cuja combinação de números e letras formam um ID único, gratuito, persistente que desambigua o autor e seu trabalho, facilita ao preencherem informações e formulários sobre suas produções e melhora a forma como as informações são compartilhadas, possibilitando conexões transparentes e confiáveis entre pesquisadores, suas contribuições e afiliações.” (ORCID...2021)

Na identificação de instituições, lançado em janeiro do ano de 2019, é possível encontramos o *ROR – Research Organization Registry*, cujos identificadores são fornecidos de forma aberta, sustentáveis, utilizáveis e únicos para cada organização de pesquisa no mundo. Grandemente utilizado pela comunidade de pesquisa global, permitem interconexões em vários sistemas. (ROR...2021)

2.5 Princípios FAIR

Em 2011, na cidade de Dagstuhl (Alemanha), um grupo composto por uma comunidade de acadêmicos, bibliotecários, arquivistas, editores e financiadores de pesquisas organizou-se em função de busca da criação e compartilhamento de conhecimento de maneira mais aprimorada. Uma das ideias desenvolvidas e registradas em Manifesto foi a seguinte:

Para possibilitar essa visão, precisamos criar e usar novas formas de publicação acadêmica que funcionem com artefatos acadêmicos reutilizáveis. Dois aspectos principais podem ser distinguidos. Primeiro, precisamos revisar os artefatos de comunicação. Como ponto de partida, nossa visão envolve a criação de uma forma nova e enriquecida de publicação acadêmica que permite a criação e gerenciamento de relacionamentos entre conhecimento, afirmações e dados. Significa também a criação de uma infraestrutura de conhecimento que permite o compartilhamento de componentes executáveis computacionalmente, como workflows, código de computador e cálculos estatísticos, como componentes de conteúdo cientificamente válidos; e uma infraestrutura que permite que esses componentes sejam disponibilizados, revisados, referenciados e atribuídos. (FORCE11, 2011)

Ou seja, chegou-se à conclusão de que era necessário que conhecimento fosse compartilhado de maneira mais eficaz, gerando assim uma nova era diferente daquela de Gutenberg, imprensa e modelo centrado em papel, para um modelo centrado em rede mais distribuído.

Acerca de produção do conhecimento e compartilhamento eficaz para uso, em 2014, “foram discutidos os principais aspectos de uma infraestrutura global para publicação, descoberta, compartilhamento e reutilização de dados eficazes para experimentação em *eScience*” (DTL, 2014), em um *workshop* com quatro dias de duração, o *Jointly designing a data FAIRPORT*.

Após um consenso, baseado em necessidades reais, identificou-se a urgência de uma infraestrutura global para publicação, descoberta, troca e reutilização de dados profissionais, como sendo fatores essenciais para uma pesquisa baseada em dados, segundo este *workshop* tão expressivo para a continuidade deste avanço tecnológico.

Trilhando este caminho, visando à urgência de uma infraestrutura com aporte para a interoperabilidade e a reutilização de dados e metadados, representantes da academia, indústria, agências de financiamento e editores, uniram-se para projetar um conjunto de regras para descrever os Princípios FAIR, um acrônimo para *Findable, Accessible, Interoperable, Reusable*, que são

Princípios norteadores para ativos digitais, visando à capacidade de ação da máquina (ou seja, a capacidade dos sistemas computacionais de encontrar, acessar, interoperar e reutilizar dados com nenhuma ou mínima intervenção humana) porque os humanos dependem cada vez mais do suporte computacional para lidar com os dados como resultado do aumento no volume, complexidade e velocidade de criação de dados. (GoFAIR...2021)

Em 2016, publica-se o então *FAIR Guiding Principles for scientific data management and stewardship*, no periódico *Scientific Data* com a prerrogativa de se estabelecer diretrizes para melhorar a aplicação dos Princípios FAIR – (WILKINSON *et al.*, 2016). Este artigo traz a abordagem para um bom gerenciamento de dados com a finalidade de beneficiar novas descobertas e inovação do conhecimento, bem como integração e reutilização dos dados.

Neste contexto, os princípios FAIR, um acrônimo para Findable, Accessible, Interoperable e Reusable, apresentam-se como elementos norteadores ao processo de gestão de dados de pesquisa. Estes princípios vêm sendo amplamente difundidos por comunidades internacionais nos últimos anos e, recentemente chegando ao Brasil como mais uma vertente de estudos e aplicação por gestores de dados. As discussões têm sido encaminhadas com o intuito de proporcionar o grau máximo de reuso de dados científicos, a partir da adoção de padrões, metadados, vocabulários controlados, ontologias e identificadores persistentes que proporcionam significado preciso aos dados e aos demais objetos a eles vinculados. (HENNING *et al.*, 2018)

Cada princípio norteador de alto nível, tem suas divisões e subdivisões, com as quinze características de sua aplicabilidade em repositórios, conforme abaixo:

Findable (encontrabilidade)

F1 – (Meta) dados são atribuídos a identificadores globais únicos e persistentes.

F2 – Os dados são descritos com metadados ricos.

F3 – Metadados incluem clara e explicitamente o identificador dos dados que descrevem.

F4 – (Meta) dados são registrados ou indexados em um recurso pesquisável.

Accessible (acessibilidade)

A1 – (Meta) dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado.

A 1.1 – O protocolo é aberto, gratuito e universalmente implementável.

A 1.2 – O protocolo permite um procedimento de autenticação e autorização, quando necessário.

A2 – Os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis.

Interoperable (interoperabilidade)

I1 – Os (meta) dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.

I2 – (Meta) dados usam vocabulários que seguem os princípios FAIR.

I3 – (Meta) dados incluem referências qualificadas a outros (meta) dados.

Reusable (reusabilidade)

R1 – (Meta) dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes.

R 1.1 – (Meta) dados são liberados com uma licença de uso de dados clara e acessível.

R 1.2 – (Meta) dados são associados à proveniência detalhada.

R 1.3 – (Meta) dados atendem aos padrões da comunidade relevantes para o domínio.

As novas formas de fazer ciência, espelhadas em compartilhamento e reuso de dados de pesquisa, vêm colocando em evidência a necessidade de substituir a ideia dos dados apenas como insumos intermediários das atividades científicas. Este novo comportamento científico está postulando novos processos e métodos que elevam os dados para um patamar de destaque nas áreas das ciências exatas, sociais e humanas, na arte ou na literatura, segundo Henning *et al.* (2018).

A iniciativa GO-FAIR busca o desenvolvimento de um ambiente global compartilhado para a pesquisa e inovação baseada em dados. Para atingir este objetivo, é necessário cumprir-se alguns passos essenciais. Desta forma, elaborou-se um processo de FAIRificação, que será apresentado abaixo (Figura 7), o qual tem como foco os dados, mas também indica o trabalho necessário para os metadados. A iniciativa GO-FAIR tem como papel, apoiar as comunidades na implementação deste processo.

O processo de FAIRificação é fundamentado em etapas que irão tornar os objetos digitais FAIR, tanto quanto for possível. Vejamos a seguir:

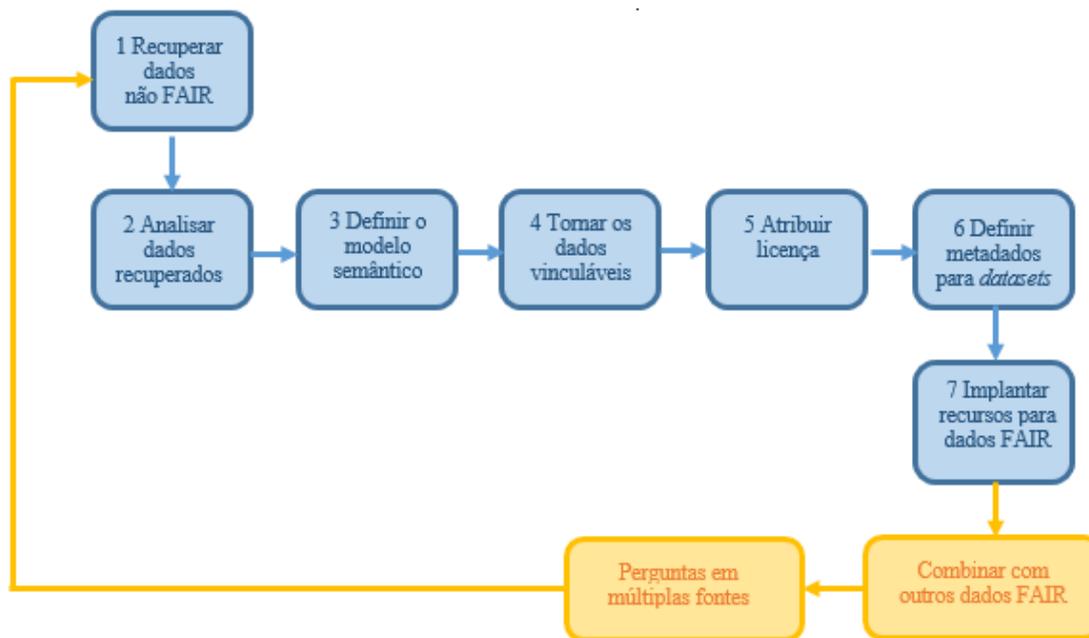


Figura 7 – Processo de FAIRificação (GoFAIR, *website*, trad. nossa)

Cada etapa, tem suas aplicações e devem ser minuciosamente verificadas para que o objetivo seja atendido o máximo possível. Na etapa 1, é importante definir os dados que serão utilizados. Na etapa 2, deverá ser feita uma análise nos conceitos, estruturas e relações dos dados escolhidos. A etapa 3 é muito importante, pois vai tratar de definir em qual modelo semântico os dados devem se apoiar, e trabalhar com modelos existentes é um bom caminho, já que deve haver uma visão consensual em um domínio particular.

A etapa 4 é aonde dados NÃO FAIR podem ser transformados em dados interoperáveis e reutilizáveis, através de tecnologias específicas como, por exemplo, *Linked Data* ou *Semantic Web*. Deve haver a análise de viabilidade por parte do usuário, pois pode funcionar bem para dados estruturados mas pode não ser relevante para dados como áudios em imagem, dados de áudio ou vídeo, por exemplo. A etapa 5 reforça a ideia de atribuir aos dados uma licença de uso clara. A etapa 6 reafirma que metadados adequados e ricos suportam todos os aspectos do FAIR. Finalmente, a etapa 7, promovendo dados FAIRificados juntamente com os metadados relevantes e uma licença, possibilitando indexação dos metadados por mecanismos de pesquisa e para que os dados possam ser acessados, ainda que com autenticação e autorização.

Esta iniciativa sustenta-se em três pilares, os quais vêm servindo de base para três categorias de atividades: cultura, treinamento e tecnologia, respectivamente:

GO CHANGE: com foco em prioridades e incentivos para a implementação do FAIR;

Uma mudança sociocultural envolvendo as partes interessadas relevantes em todos os níveis relevantes para o florescimento da Ciência Aberta.

GO TRAIN: coordenando o treinamento de conscientização e desenvolvimento de habilidades da FAIR;

Treinar os administradores de dados necessários, capazes de projetar e implementar planos de gerenciamento de dados adequados, incluindo dados e serviços FAIR.

GO BUILD: coordenação da tecnologia FAIR;

Projetar e construir os padrões técnicos, melhores práticas e componentes de infraestrutura necessários para implementar os princípios de dados FAIR. (GoFAIR)

Atualmente existem alguns escritórios regionais do GO-FAIR. O Escritório de Apoio de Coordenação GO-FAIR Brasil, que iniciou suas atividades em 2018, é um dos escritórios regionais de apoio e coordenação do GO-FAIR. Tem a responsabilidade de difundir, apoiar e coordenar as atividades relacionadas às suas redes temáticas em apoio à adoção da estratégia de implementação dos princípios FAIR em todo o território brasileiro. O GO-FAIR Brasil está organizado em redes temáticas que surgem em domínios disciplinares específicos, voltadas para práticas, ferramentas e padrões apropriados aos domínios de suas redes. Maiores informações sobre o GO-FAIR Brasil estão presentes no manifesto denominado Declaração GO-FAIR Brasil (BR).

As redes temáticas do GO-FAIR possuem gestão própria, podendo interagir com redes de implementação de outros domínios para troca de experiência e colaboração. Vale destacar duas redes, uma que se aproxima das HD e outra da AD.

No caso da rede GO-FAIR Humanidades, esta tem a responsabilidade de elaborar estratégias e mobilizar parcerias para a implantação dos princípios FAIR no universo dos dados da pesquisa das Humanidades (Ciências Humanas, Sociais, Sociais Aplicadas, Linguística, Letras e Artes). Essa rede é coordenada pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), através do Laboratório em Rede de Humanidades Digitais (LARHUD - larhud.ibict.br) localizado no Rio de Janeiro e que tem seu lançamento previsto para o segundo semestre de 2021.

A outra rede de interesse deste trabalho é a GO-FAIR Agro, lançada em 2021 durante o Simpósio Brasileiro de Informática na Agricultura (SBIAGRO) (Drucker et al, 2021). Os objetivos específicos propostos são: i) Promover pesquisas e inovações sobre repositórios de dados de pesquisa nacionais e internacionais que poderão armazenar e compartilhar, de forma

confiável, os dados de pesquisa em Ciências Agrárias; ii) Desenvolver metodologias voltadas para as práticas dos produtos e serviços FAIR, que atendam às Ciências Agrárias; iii) Promover difusão do conhecimento através de encontros, cursos, *workshops* e seminários visando impulsionar e disseminar os princípios FAIR entre os membros da Rede GO-FAIR Brasil Agro e sociedade em geral e iv) Trabalhar de forma articulada com a Rede GO-FAIR Brasil e com a Rede de Implementação *Internacional Food Systems*.

Para uma compreensão resumida, com prática de aplicação sobre os Princípios FAIR, segue a figura abaixo:



Figura 8 – O que são Dados FAIR? (MARTÍNEZ-LAVANCHY *et al.*, 2019, trad. nossa)

Assim, é possível concluir, segundo os autores, que os princípios FAIR permitem que o pesquisador contribua para a Ciência Aberta com dados de pesquisa de alta qualidade, tornando seus dados reutilizáveis para outras pessoas e mais suscetíveis a serem descobertos através de seus metadados. Estes devem facilitar o encontro de informação, com formato de dados legíveis por máquinas, dados passíveis de acesso, isto é, com tipo de acesso permitido definido.

Considerando a Ciência Aberta, convém destacar: dados FAIR não são necessariamente Dados Abertos, conforme descrição da *Open Knowledge Foundation*. Ou ainda “abertos tanto quanto possível e fechados tanto quanto necessário” (ALLEA, 2020).

Os Dados Abertos referem-se a qualquer conteúdo, informação ou dado que possa ser usado, reusado e distribuído por qualquer pessoa sem nenhuma restrição legal, tecnológica ou social.

A Ciência Aberta melhora a qualidade, eficiência e capacidade de resposta de uma investigação. Conhecimento e dados compartilhados desde o começo de uma pesquisa contribuem com a divulgação dos resultados mais recentes, segundo a política da Comissão Europeia (*website*).

A Ciência Aberta é um movimento que incentiva a transparência da pesquisa científica desde a concepção da investigação até o uso de softwares abertos. Também promove esclarecimento na elaboração de metodologias e gestão de dados científicos, para que estes possam ser distribuídos, reutilizados e estar acessíveis a todos os níveis da sociedade, sem custos. Propõe, ainda, a colaboração de não cientistas na pesquisa, ampliando a participação social por meio de um conjunto de elementos que dispõem de novos recursos para a formalização da comunicação científica. (SILVA; SILVEIRA, 2019)

A Taxonomia da Ciência Aberta (FOSTER...2021), torna visível a necessidade de dados FAIR, cada vez maior, considerando três vertentes – acesso aberto, dados abertos e pesquisas com dados reprodutíveis, pois estes conceitos tem grandes relações entre si. Analisemos a figura abaixo, e o destaque para as divisões:



Figura 9 – Taxonomia da Ciência Aberta (FOSTER, *website*, trad. nossa)

Dados FAIR, tem como condição em seu ‘A’ serem ‘Acessíveis sob condições bem definidas’. Segundo MONS *et al.* (2017), mesmo que os Princípios FAIR tenham sido

inspirados na Ciência Aberta, explicitamente e deliberadamente não entram no mérito moral e ético, com os quais se relacionam a abertura de dados.

Sobre dados FAIR, aborda-se a necessidade de links duradouros para os dados e licenças que permitam que os dados sejam utilizados por outros. Observados estes quesitos e os princípios de reutilização dos dados, é importante observar a manutenção de registros de Proveniência para dados e metadados, pois será de suma importância para a reutilização destes, o que é contemplado no escopo desta pesquisa.

2.6 Proveniência de dados

A Proveniência está presente em tudo o que nos cerca. É um conceito que determina origem e histórico de uma informação, em qualquer meio ou suporte. O conceito origina-se no mundo de Belas Artes, e segundo Tan (2007) é especialmente importante para determinar diretamente o valor de uma obra arte.

A Proveniência de Dados, do ponto de vista da computação, é definida como “informações sobre entidades, atividades e pessoas envolvidas na produção de um dado ou coisa, que podem ser usadas para formar avaliações sobre sua qualidade, confiabilidade” (Consórcio W3C) e segundo Sayão (2010), é a “história do objeto [digital] armazenado; pode incluir informações sobre sua fonte ou origem, sua cadeia de custódia; registra também as ações de preservação sofridas pelo objeto e seus efeitos, por exemplo: as migrações efetuadas”.

O *Australian National Data Service* classifica a proveniência de dados como documentação de onde um dado vem e os processos e metodologia pelos quais ele foi produzido e afirma ainda que registrar a proveniência dos dados, a qual é um tipo de metadado, é importante para confirmar a autenticidade dos dados e permitir que eles sejam reutilizados.

O *W3C Provenance Incubator Group* define a proveniência de dados da seguinte forma:

um registro que descreve entidades e processos envolvidos na produção e entrega ou de outra forma influenciando esse recurso. A procedência fornece uma base crítica para avaliar a autenticidade, permitindo a confiança e permitindo a reprodutibilidade. As afirmações de proveniência são uma forma de metadados contextuais e podem se tornar registros importantes com sua própria proveniência. (W3C, 2010, trad. nossa)

Verificamos, abaixo, estruturas fundamentais e suas inter-relações, atribuídas à proveniência de dados, no diagrama.

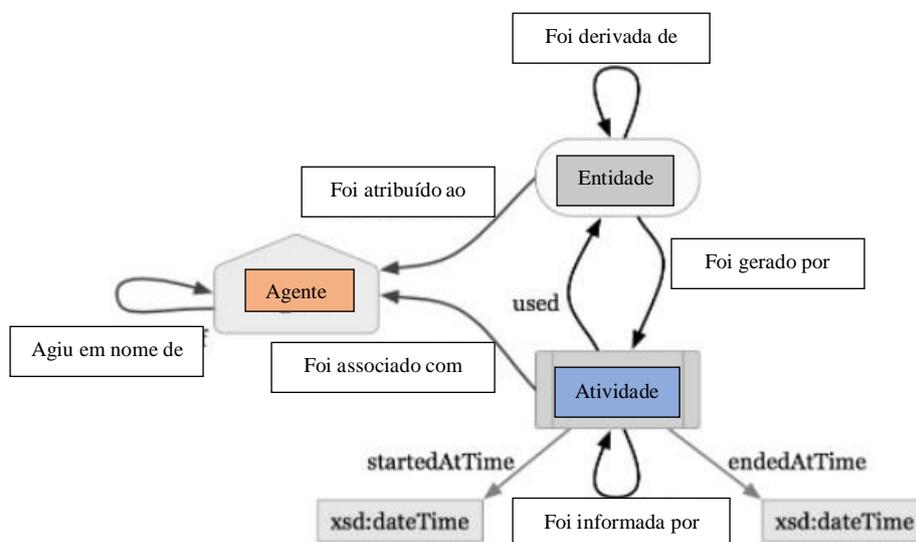


Diagrama 1 – Proveniência - modelo simplificado em W3C PROV-O (ANDS, website, trad. nossa)

Pomerantz (2015) *apud* Segundo e Araújo (2019), afirma que os padrões de metadados de proveniência “categorizam as relações entre recursos e entidades e que as três estruturas fundamentais nesses modelos de metadados são entidade, agente e atividade. Ou seja, não basta conhecer o agente originário, o intermediário ou o final, mas também quais as atividades que foram exercidas por eles durante o fluxo de interação com os dados.

De acordo com os autores, pode-se afirmar que **entidade** é o recurso, o objeto, o dado. O **agente** é o ator, o indivíduo, a instituição que influenciou o ciclo de vida dos dados e a **atividade** é a descrição dessa influência sofrida.

Simplificando, a proveniência responde às perguntas por que, como, onde, quando e por quem os dados foram produzidos. Neste contexto, segundo Davidson e Freire (2008), existem ainda a proveniência prospectiva que “captura a especificação de uma tarefa computacional (ou seja, um fluxo de trabalho) - e corresponde às etapas que precisam ser seguidas (ou uma receita) para gerar um produto de dados ou classe de produtos de dados” e a proveniência retrospectiva que “captura as etapas que foram executadas, bem como informações sobre o ambiente de execução usado para derivar um produto de dados específico - um registro detalhado da execução de uma tarefa computacional.”

Em consonância com a linha de pensamento acima, segundo Buneman, Chapman e Cheney (2014), existem questões particularmente importantes como, por exemplo, como e de onde, as informações foram obtidas quando se trata de registros científicos:

essas questões são particularmente importantes porque a tecnologia de banco de dados é empregada não apenas para fornecer acesso aos dados de origem, mas também para o conhecimento derivado de cientistas que interpretaram os dados. Muitos cientistas acreditam que a proveniência, ou metadados que descrevem criação, gravação, propriedade, processamento ou histórico de versão, é essencial para avaliar o valor de tais dados.” (BUNEMAN; CHAPMAN; CHENEY, 2014, tradução nossa)

Ainda conforme os autores, a tecnologia de banco de dados atual oferece pouca ajuda para gerenciar proveniência. Neste trabalho, utilizaremos como referencial teórico de proveniência, o modelo da PROV, utilizado pelo W3C¹², o qual define um modelo, serializações correspondentes e outras definições de suporte para permitir o intercâmbio interoperável de informações de proveniência em ambientes heterogêneos como a web. (CONSÓRCIO W3C...2021).

Considerando a Proveniência de Dados, pode-se dizer que o processo de FAIRificação, visto anteriormente, pode ser grandemente beneficiado a partir desta análise, proporcionada pelos caminhos investigados pelos métodos PROV.

Somando-se a este fato, segundo Wilkinson (2016), o princípio “*Reusable*”, uma condição para Reutilização de Dados Digitais é a adoção de metadados de Proveniência, conforme figura abaixo:



Figura 10 – Proveniência como característica do princípio de Reutilização (AUSTRALIAN NATIONAL DATA SERVICE, *website*)

Vale ressaltar que nesta pesquisa, o princípio **R 1.2** é de extrema relevância, pois está diretamente ligado à interpretação da proveniência detalhada dos dados. Retomaremos em detalhes, abordando a descrição do princípio na Metodologia desta pesquisa.

¹² W3C - *Wide Web Consortium* (W3C) é uma comunidade internacional que desenvolve padrões abertos para garantir o crescimento da web a longo prazo. Disponível em: <https://www.w3.org/Consortium/>

Segundo Tan (2007), “a palavra proveniência é usada como sinônimo da palavra linhagem na comunidade de banco de dados”. Então, durante as etapas de busca, recuperação e análise dos dados não-FAIR, podemos afirmar que a proveniência será muito importante para se gerar bancos de dados FAIRificados visando a uma confiabilidade muito maior, dando à pesquisa produzida, uma credibilidade inerente a esta característica.

Nesta pesquisa, a proveniência é um fator determinante para conferir credibilidade aos dados disponibilizados para reuso, pois na *e-Science*, é uma das questões predominantes, considerando a origem de um dado e qual o nível de confiança atribuído à informação gerada em uma pesquisa e seu resultado. Tratando-se de dados pedológicos, a proveniência tem o seu papel crucial no sentido de que cada vez mais os dados abertos podem ser acessados e examinados, levando à melhor qualidade, progresso e economia de tempo e recursos.

2.7 Repositórios de dados

Os repositórios de dados (RD) surgem como ambientes informacionais digitais para gerenciamento e controle da produção acadêmica e científica de instituições e/ou comunidades, oferecendo vantagens como acesso e preservação da informação a longo prazo (CAMARGO; VIDOTI, 2008)

O armazenamento, preservação e acesso irrestrito aos dados de pesquisa são elementos fundamentais para o ciclo de vida da própria pesquisa e da valoração do investimento em pesquisas. Logo, é essencial pensar desde o início de um projeto como e onde armazenar e preservar os dados da pesquisa. Portanto, compreender o tema RD é um elemento primordial nesta dissertação.

Segundo Moreira, Bonino e Pires (2019), os repositórios de dados são geralmente projetados para armazenar e gerenciar as saídas de pesquisa após a produção dos dados, desempenhando um papel importante para permitir a reutilização desses dados, em que a interoperabilidade é um “cidadão” de primeira classe. Repositórios de dados garantem que a autoria dos dados seja mantida, enquanto a responsabilidade de manter os dados seguros e intactos ao longo dos anos é transferido para o repositório e equipes técnicas que o mantêm.

Os RD oferecem visibilidade para instituição, interoperabilidade de dados e permitem que terceiros repliquem análises ou resultados de estudos prévios. Além disso, estimulam pesquisas adicionais com conjuntos de dados pré-existentes, conforme afirmam Camargo e

Vidoti (2008). Por fim, são um facilitador do compartilhamento de dados de pesquisa, pois pode contribuir para ampliar o conhecimento da sociedade evitando que se canalizem recursos para duplicar os esforços de coleta de dados pré-existentes, minimizando de custos de publicação, entre outras vantagens.

De acordo com Agência USP de Gestão da Informação Acadêmica (2021) e Camargo e Vidoti (2008), existem diversos tipos classificações de RD, variando desde os públicos até os privados ou de escopo distinto.

Vários desses RD, são gerais (multidisciplinares) em termos de escopo, enquanto que outros são mais especializados e estruturados em áreas de conhecimento e atendem principalmente às ciências sociais, humanas, biológicas ou mesmo exatas. O escopo pode ajudar a definir qual o tipo de repositório os pesquisadores podem selecionar, dependendo do público-alvo, para reutilizar seus dados.

Um outro tipo de classificação dos repositórios envolve os custos, se exige associação institucional ou individual ou ainda ter taxas associadas ao depósito de dados de pesquisa.

Café *et al.* (2003) relatam que os repositórios institucionais são a reunião de repositórios temáticos hospedados em uma instituição. No caso de uma universidade, cada departamento trata de uma área do conhecimento e, portanto, seu repositório temático será específico no assunto deste departamento. A união de todos os repositórios das diversas unidades de pesquisa comporá o repositório institucional, caracterizando-o como multidisciplinar. No contexto dessa dissertação os RD institucionais são os que despertam maiores interesses. As principais ferramentas para o desenvolvimento de repositórios institucionais públicos e gratuitos estão organizados em torno do *DSpace*, *Dataverse*,¹³ *DataHub*¹⁴, *DANS*¹⁵.

Por outro lado, existem diversos RD multidisciplinares disponíveis na internet. São exemplos de RD multidisciplinares as ferramentas *Figshare*¹⁶, *Dryad*¹⁷ e *Zenodo*¹⁸, destacamos que estes últimos seguem os Princípios FAIR. A figura 10 apresenta um quadro comparativo entre esses tipos de repositórios. As principais características desses repositórios estão ilustradas no quadro a seguir.

¹³ Disponível em: <https://dataverse.org/>

¹⁴ Disponível em: <https://datahub.io/>

¹⁵ Disponível em: <https://dans.knaw.nl/nl/>

¹⁶ Disponível em: <https://figshare.com/>

¹⁷ Disponível em: <https://datadryad.org/stash>

¹⁸ Disponível em: <https://zenodo.org/>

Repositório	Figshare	Data Dryad	Zenodo
URL (Localizador Uniforme de Recursos)	https://figshare.com	http://datadryad.org	http://www.zenodo.org
Publicação de Dados	Sim	Sim	Sim
Identificador persistente	Sim (DOI)	Sim (DOI)	Sim (DOI)
ORCID (Id de Pesquisador Aberto e Contribuinte)	Sim	Sim	Sim
Custos para depósito	Sob consulta	Sob consulta	Gratuito
Metadado	Tipo de	Dublin Core	Esquema Dublin Core OpenAir
Metadado de domínio	Não	Não	Em campos personalizados
Colaboração do dado	Sim, em projetos	Não	Sim, em coleções de dados
Tipo de Licenças	Creative Commons CC0 Licenças personalizadas	Creative Commons CC0	Creative Commons CC0 Licenças personalizadas
Recursos de pesquisa	Tipo Google	Tipo Google	Tipo Google
Confidencialidade/ Embargo	Sim	Sim	Sim
Funcionalidade na queda	Não	Não	Integração com Dropbox
Provedor de armazenamento	Amazon	Organização sem fins lucrativos	OpenAir e CERN
Comentário		Dado associado à publicação do artigo	

Figura 11 – Repositórios multidisciplinares sob Princípios FAIR (<https://www.aguia.usp.br/apoio-pesquisador/dados-pesquisa/lista-repositorios-dados-pesquisa/>)

O conceito central por trás dos RD públicos estruturados é o refinado suporte para a curadoria de dados, os serviços de curadoria de dados são amplos e envolvem muitos processos que validam os dados garantindo que haja alinhamento com o acervo de dados de projetos de pesquisa existentes no repositório. Os dados também podem ser disponibilizados em vários padrões (por exemplo, FAIR) ou formatos de arquivo. Os serviços de curadoria de dados também podem servir como uma verificação adicional antes de os dados serem disponibilizados para terceiros.

Para um conhecimento mais amplo acerca da função dos RD estruturados e organizados em torno de áreas de conhecimento apresentamos dois exemplos – o *Dataverse* e o *DSpace* – de soluções tecnológicas em softwares, comumente adotadas por repositórios digitais confiáveis (Rocha et al. 2021).

O *Dataverse*, está sendo desenvolvido pelo *Institute for Quantitative Social Science* (IQSS) de Harvard, junto com muitos colaboradores de todo o mundo. O *Dataverse* não é uma plataforma que partiu do zero, foi construído com base na experiência prévia no projeto anterior *Virtual Data Center* (VDC) (ROCHA *et al.*, 2021). O principal objetivo do *Dataverse* é atender os problemas de compartilhamento de dados por meio da criação de tecnologias que permitam às instituições reduzir a carga de trabalho de pesquisadores e editores de dados e incentive-os a compartilhar seus dados pois possui recursos para configuração de vários tipos de ambientes de repositório, incluindo hierarquias organizacionais e políticas de gestão distintas para unidades ou grupos, com esquemas de metadados e licenças.

O *DSpace* foi lançado em 2007 como um esforço conjunto entre os desenvolvedores do *MIT Libraries* e do *HP Labs*. Com o crescimento da comunidade de usuários, HP e MIT. O *software* é mantido por uma organização sem fins lucrativos. O *DSpace* (versão 6.5 ou anteriores) também permite diversas configurações, porém com adaptações por ter sido desenvolvido para repositórios institucionais e estar estruturado a partir do conceito de coleção de itens (ROCHA *et al.*, 2018). O *DSpace* pode ser considerado como um sistema de repositório digital destinado a receber, preservar e redistribuir a produção intelectual de investigação de uma universidade em formatos digitais (CAMARGO; VIDOTI, 2008). Em novembro de 2021, existem aproximadamente 136 instalações de *DSpace* no Brasil registradas na comunidade *Duraspace* (DURASPACE, 2021¹⁹). O *DSpace* está presente principalmente em instituições acadêmicas e governamentais. Entretanto, a grande maioria destas instalações não são dedicadas a repositórios de dados, mas a documentos.

Segundo o autor Rocha *et al.* (2018), os repositórios *DSpace* e *Dataverse* podem adotar diversas políticas de submissão e formas peculiares de organização dos materiais submetidos, bem como estratégias pontuais de funcionamento, visando aos seus objetivos. Outro ponto importante que deve ser considerado na política de um repositório, considerando seus objetivos, é o ciclo de vida dos dados. Alguns poderão ser armazenados para o compartilhamento dos dados que acompanham uma pesquisa ao longo de seu andamento ou somente aqueles dados de pesquisa que devem ser preservados a longo prazo.

Rocha *et al.* (2018) afirma ser crescente a importância de termos cada vez mais conhecimentos sobre conjunto de dados, suas atualizações, transformações ou ainda, o fato de serem resultados de transformações. Nós complementamos essa percepção do autor ao verificar

¹⁹ Disponível em: <https://duraspace.org/dspace/download>

que incorporar Proveniência de Dados e Princípios FAIR, podem ser diferenciais de confiabilidade dos RD.

2.8 Base de dados da plataforma *OpenSoils*

OpenSoils é uma plataforma computacional gratuita, aberta, elástica, distribuída, multiusuário e multicamada cuja arquitetura computacional é ilustrada pela Figura 11. É orientada para armazenar dados curados e harmonizados (primários e secundários) de solos do Brasil e seus metadados de proveniência (CRUZ *et al.*, 2018; 2019).

Atualmente o *OpenSoils* não se beneficia de nenhum tipo de repositórios de dados do tipo *Dspace* ou *Dataverse*. Ele mantém uma base de dados relacional sobre dados de solos, com aporte para Proveniência de Dados e algum grau de aplicação dos Princípios FAIR, sendo capaz de contribuir parcialmente com alguns aspectos das Humanidades Digitais, no que diz respeito a algumas questões sociais na área de Ciências Agrárias.

O objetivo do *OpenSoils* é hospedar, conectar e compartilhar grandes quantidades de dados e conhecimento de solo com curadoria em nível brasileiro e sul-americano. A e-infraestrutura consiste em várias camadas de serviços, um banco de dados de perfis de solo, uma estrutura computacional baseada em nuvem para computar e compartilhar dados de solo integrados com ferramentas de visualização de mapas. O *OpenSoils* é uma infraestrutura eletrônica aberta, elástica, orientada para proveniência e leve que coleta, armazena, descreve, organiza, harmoniza e direciona para vários tipos de recursos de solo: grandes conjuntos de dados de perfis de solos, serviços / aplicações, documentos, projetos e *links* externos. *OpenSoils* é o primeiro *framework* computacional de segurança de solos baseado em Ciência Aberta na literatura. (CRUZ *et al.*, 2018)

Na prática, o *OpenSoils* necessita se expandir e se incorporar uma estrutura de RD institucional para continuar a fornecer soluções essenciais e adequadas em consonância com sua proposta de rastreamento e análise de solos, oferecer maior capacidade de ingestão e transformação de dados legados, promover acesso seguro aos dados, agregar reprodutibilidade e transparência de ensaios e, idealmente, ser integrado com práticas educacionais da área de solos da Universidade Federal Rural do Rio de Janeiro, permitindo o uso colaborativo e ainda o acesso personalizado à agricultores, tomadores de decisão e cientista de solos.

Conta ainda com o *OpenSoils Field*, um aplicativo para dispositivos móveis que permite a coleta e análise de dados de solos diretamente do campo e o *OpenSoils Edu*, desenvolvido para ser usado em atividades educacionais, possibilitando por exemplo, a consulta dos dados (abertos originalmente e legados) no banco de dados. (CRUZ *et al.*, 2019)

Na figura 12, apresentamos uma visualização da arquitetura conceitual, visando à capacidade do banco de dados capaz em fazer armazenamento de longo prazo com proveniência de seus metadados. O destaque desta pesquisa está na identificação do banco de dados e aplicação nesta estrutura. No entanto, os padrões de metadados e aporte para proveniência, no *OpenSoils*, permitem um estudo sistêmico futuro, de modo a ampliar sua utilização atendendo ao padrão *Dublin Core*, por exemplo.

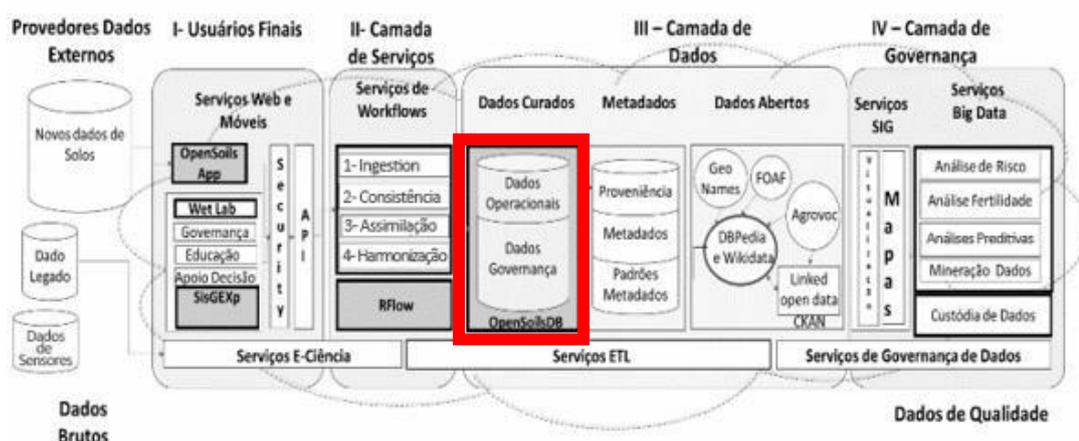


Figura 12 – Visão conceitual da arquitetura *OpenSoils* (CRUZ, 2018)

O banco de dados da plataforma *OpenSoils* armazena grandes volumes de dados curados e harmonizados (primários e secundários) de solos do Brasil e seus metadados de proveniência (CRUZ *et al.*, 2018; 2019). Seu BD é do tipo relacional usando o MySQL no armazenamento de dados pedológicos, seguindo as normas da taxonomia da quinta edição do SiBCS - Sistema Brasileiro de Classificação de Solos²⁰.

O *schema* completo do BD possui um total de 46 tabelas, contendo dados pedológicos utilizados em projetos da categoria, baseadas em relações como: “integração de dados coletados em campo (as tabelas *projeto*, *observação*, *relevo*, *descrição geral*, *horizontes*, *morfologia* entre

²⁰ SiBCS. Disponível em: <https://www.embrapa.br/solos/sibcs>

outros) com dados obtidos em laboratório através de experimentos físicos (curva de retenção de água) e químicos (ataque sulfúrico, pasta saturada).” (OLIVEIRA *et al.*, 2021)

3 METODOLOGIA DE PESQUISA

Neste capítulo descrevemos a metodologia e os procedimentos utilizados para avaliar o repositório digital da plataforma *OpenSoils* e como podemos torná-lo mais próximo das características dos princípios norteadores de dados FAIR. É importante salientar que esforços foram empregados de forma que esta metodologia possa servir de base para outros projetos, onde o processo de FAIRificação possa ser aplicado. A metodologia tem como foco exclusivo, o banco de dados, como objeto digital. Não há uma aplicação sistêmica, portanto, nesta pesquisa.

Iniciamos o trabalho com o levantamento de artigos na literatura (KITCHENHAM; CHARTERS, 2007) que envolveu os domínios da Humanidades Digitais e Agricultura Digital, realizado a partir de literatura da área e documentação disponível em repositórios e acervos públicos. Neste trabalho, incorporamos conceitos sobre a percepção dos Repositórios Digitais, os quais nos ajudarão a compreender com mais clareza o funcionamento do repositório desta plataforma.

A aplicação dos Princípios FAIR e análise de Proveniência de Dados é um elemento importante na metodologia, permitirá avaliar quais as características necessárias de um repositório digital na área de AD. Nas etapas para avaliação da FAIRificação do banco de dados da plataforma *OpenSoils*, adotamos o *checklist* de questões norteadoras de Sarah e Marjan (2017). O *checklist* é intitulado “Quão FAIR são seus dados”, o *checklist* aborda as etapas de verificação, o qual apresenta-se no ANEXO 1.

Segundo Henning *et al.* (2019), os métodos de avaliação das métricas FAIR podem definir, com alguma precisão, um conjunto mensurável de propriedades e comportamentos que irão avaliar o nível de *FAIRness* dos objetos digitais, ou seja, o quanto um recurso digital adere aos Princípios FAIR.

Este conjunto de propriedades e comportamentos, segundo Wilkinson *et al.* (2017), é determinado por métricas que deverão considerar o grau de encontrabilidade, acessibilidade, interoperabilidade e usabilidade de um recurso digital por máquinas, sem nenhuma ou com pouca intervenção humana.

Deste modo, a metodologia apoia-se nesta base teórica, ao avaliar os objetos digitais dispostos na plataforma *OpenSoils*, considerando os condicionantes abaixo (WILKINSON *et al.*, 2017) para a formulação de uma métrica FAIR, que deve seguir quatro condições:

- Clara : qualquer pessoa pode entender o propósito da métrica;
- Realística : não deveria ser indevidamente complicado para um recurso estar em conformidade com a métrica;
- Mensurável : a avaliação pode ser feita de forma objetiva, quantitativa, interpretável por máquina, escalonável e reproduzível, garantindo a transparência do que está sendo medido e como se mede;
- Universal : a métrica deve ser aplicada a todos os recursos digitais.

Toda esta base teórica foi utilizada para nortear um diagnóstico inicial do Banco de Dados da plataforma *OpenSoils* e que poderá adaptar-se a outros Bancos de Dados, de modo a torna-los o mais FAIR possível.

Em resumo, reafirmamos a metodologia em tópicos, para melhor visualização:

<ul style="list-style-type: none">• Levantamento de artigos nos domínios da pesquisa: Humanidades Digitais e Agricultura Digital
<ul style="list-style-type: none">• Percepção de conceitos dos Repositórios Digitais, para compreender o papel do <i>OpenSoils</i>
<ul style="list-style-type: none">• Adoção de um checklist, intitulado “Quão FAIR são seus dados?” (ANEXO 1), usado como base para as etapas de FAIRificação do <i>OpenSoils</i>
<ul style="list-style-type: none">• Avaliação do grau de encontrabilidade, acessibilidade, interoperabilidade e usabilidade dos recursos digitais e sua aderência aos Princípios FAIR.

Todo este diagnóstico tem o objetivo de aproximar o Repositório Digital do *OpenSoils*, das métricas FAIR (WILKINSON *et al.*, 2017), que devem ser, tanto quanto for possível:

Claras : qualquer pessoa pode entender o propósito da métrica;

Realísticas : não deveria ser indevidamente complicado para um recurso estar em conformidade com a métrica;

Mensuráveis : a avaliação pode ser feita de forma objetiva, quantitativa, interpretável por máquina, escalonável e reproduzível, garantindo a transparência do que está sendo medido e como se mede;

Universais : a métrica deve ser aplicada a todos os recursos digitais.

4 TRABALHOS RELACIONADOS

4.1 Bancos de dados de solos

Esta seção dedica-se a apresentar alguns trabalhos que vêm sendo desenvolvidos no intuito de oferecerem soluções práticas, relacionadas à área de solos, que segundo Cruz et al. (2018), “são provavelmente o recurso natural e biossistema mais importantes que apoiam a vida humana e terrestre. É um recurso natural finito primário que deriva outros recursos, bens e serviços.”

Ainda segundo o autor, “o solo é considerado um sistema aberto que interage com outros componentes do ciclo geológico. As características de um solo são função do material original, clima, relevo, organismos e tempo.” (PANSU; GAUTHEYROU, 2006 apud CRUZ et al, 2018).

Atualmente, o Brasil dispõe apenas de levantamentos de solo de caráter geral, com mapas de pequena escala, segundo o PronaSolos²¹, o que não permite uma análise em grande escala de detalhes, conforme em grande parte dos países desenvolvidos, baseado em escalas maiores. “Estes levantamentos mostram as propriedades dos solos que afetam o uso da terra, bem como a localização exata de cada tipo de solo de uma determinada área (fazenda, município, microbacia, estado) em forma de mapas.”

O uso sustentável dos solos beneficia toda a sociedade em diversos aspectos. Na AD, são produzidas enormes quantidades de dados e quanto mais houver qualidade nesses dados, mais os agricultores e demais profissionais de diversas áreas poderão ser beneficiados neste contexto, em suas rotinas diárias com tomadas de decisão assertivas envolvendo diversas questões, com resultados mais positivos.

Entretanto, para que este grande volume de dados seja bem utilizado e bem estruturado em um banco de dados, não se pode ignorar o fato de que, segundo Cruz et al. (2018), as

²¹ Programa Nacional de Solos do Brasil. Disponível em: <https://www.embrapa.br/pronasolos/sobre-o-programa>

propriedades dos solos são muitas e difíceis de serem coletadas, mapeadas, analisadas, armazenadas e compartilhadas como dados de solos em bancos de dados, sendo imprescindíveis ações de campo e laboratórios científicos úmidos.

O OpenSoilsDB usa a recomendação W3C PROV-DM para armazenar a proveniência prospectiva e retrospectiva de fluxos de trabalho e scripts (MOREAU & MISSIER, 2013). Além disso, o OpenSoilsDB oferece suporte às diretrizes FAIR (localizáveis, acessíveis, interoperáveis e reutilizáveis) para gerenciamento e compartilhamento de dados científicos. (WILKINSON et al., 2016 *apud* CRUZ et al., 2018)

O autor afirma ainda que vários bancos de dados de solos apresentam-se inacessíveis à consultas estruturadas ou apresentam-se como “planilhas simples ou arquivos de texto, sendo dificilmente compartilhados e reutilizados por fazendeiros e formuladores de políticas.” (ARROUAYS et al., 2017 *apud* CRUZ et al., 2018).

Podemos encontrar exemplos como o BDSolos, pouco eficiente para fazer armazenamento e recuperação de dados, além da interface disponível que não permite a entrada de dados novos por parte dos pesquisadores e o Fe.BR, com centenas de planilhas armazenadas, cujos arquivos não compartilham uma mesma estrutura, armazenados em um disco virtual. (CRUZ *et al.*, 2018)

Atesta também que a estrutura do *OpenSoils* pretende transpor estas dificuldades, pois utiliza a recomendação W3C (PROV-DM) para armazenar proveniência retrospectiva e prospectiva de fluxos de trabalho e *scripts* (MOREAU; MISSIER, 2013 *apud* CRUZ et al., 2018), além de oferecer suporte aos Princípios FAIR para gerenciamento e compartilhamento dos dados científicos (WILKINSON et al., 2016 *apud* CRUZ et al., 2016).

Na figura abaixo, é possível verificar a estrutura conceitual e aportes do *OpenSoils*. Esta figura ilustra de forma resumida, o fluxo de dados na infraestrutura, que possibilita experimentos científicos mais reprodutíveis, uso de grande quantidade de dados, com distribuição baseada em serviços *web* e dispositivos móveis e possibilidade de uso em fluxos de trabalho científicos que funcionem em ambiente de nuvens de computadores, provendo recursos sob demanda. (CRUZ *et al.*, 2018)



Figura 13 – Representação conceitual dos três ambientes de coleta de dados (*in situ*, *in vitro* e *in silico*) e fluxos de dados entre os experimentos de segurança de solos e exemplos da abordagem computacional adotada pelo OpenSoils (CRUZ *et al.*, 2018)

Embora existam esforços em apresentar bancos de dados que contribuam nesta área de solos, através de aplicativos por exemplo, é necessário focar em uma boa gestão dos dados integrados. Seus repositórios devem visar à algumas características importantes como armazenamento, versionamento de dados, certificação institucional, preservação, curadoria, visibilidade, fontes para compartilhamento, fonte para reuso de dados, citação e métricas. (SAYÃO; SALES, 2012; ROCHA *et al.*, 2021)

4.2 Outras iniciativas em AD

Segundo Cruz *et al.* (2015), podem ser encontrados um crescente número de aplicativos móveis, na área de agricultura familiar, que oferecem alguns serviços relacionados a dados, mas ainda desconectados dos princípios FAIR e Data Provenance.

O *Farmers Text Center* (nas Filipinas) que permite ao agricultor a transmissão de imagens (via SMS) para a avaliação de arrozais por especialistas; os serviços e-Sagu e m-Krishi (na Índia) que fornecem informações técnicas (via SMS) aos agricultores sobre clima, cotações, ocorrência de doenças e formas de manejos de diversas lavouras.

No cenário brasileiro existem uma série de aplicativos. Por exemplo: O *AccuWeather*, com previsão do tempo precisa, relatórios ao vivo, alertas de chuvas e outras funcionalidades; O *AgroMercado* que permite ao produtor rural acompanhar o tempo e cotações agropecuárias; O *Adama Alvo*, que possui um grande banco de imagens para auxiliar na identificação de pragas no campo ou ajuda técnica por parte da equipe, caso não haja imagem disponível, e informações sobre as principais ameaças às lavouras brasileiras, entre outros.

No mesmo contexto, há disponível o aplicativo ligado a plataforma *OpenSoils*, no caso, o *OpenSoils field* que facilita a coleta e a transmissão de dados diretamente do campo para a nuvem, além de consulta a dados curados e etiquetados com proveniência, extração e compartilhamento em formato aberto, totalmente compatível com os princípios da Ciência Aberta. (opensoils.org)

4.3 Farm Data Train

Uma iniciativa na área agrícola tem sido desenvolvida entre três instituições – DTL (*Dutch Techcentre for Life Sciences*)²², GODAN (*Global Open Data for Agriculture and Nutrition*)²³, CABI (*Centre for Agriculture and Biosciences International*)²⁴ com participação da *Wageningen University & Research* – o *Farm Data Train*. A iniciativa criou um projeto que concebeu um documentário animado (<https://vimeo.com/215975839>) que visa explicar a possibilidade de se conectar dados agrícolas para torná-los mais reutilizáveis sob a forma de um imenso ecossistema de dados globais para agricultura e alimentos (BROUWER et al., 2020).

O documentário animado do *Farm Data Train* serve como uma representação visual dos usos potenciais dos Princípios de Dados FAIR no domínio da Agricultura e Nutrição, ele incentiva as partes interessadas nesses domínios a aplicá-los em contextos reais para conectar dados agrícolas e torná-los mais reutilizáveis.

Segundo Brouwer et al., (2020), o conjunto de dados do *Farm Data Train* é combinado a partir de diversas fontes de dados, de uma forma amigável possibilitando localização e reuso simplificados. Conexões entre os dados são estabelecidos entre os usuários, fontes de dados e pesquisadores, as conexões corresponderiam a uma “estação” (carga e descarga de dados). Os

²² Disponível em: <https://www.dtls.nl/fair-data/farm-data-train/>

²³ Disponível em: <https://www.godan.info/>

²⁴ Disponível em: <https://cabiagbio.biomedcentral.com/>

metadados são fornecidos por uma “estação centralizadora”, mas os dados permanecem armazenados em suas estações de origem evitando transferências na Web e permitindo o transporte de identificadores. Além disso, o *Farm Data Train* usaria protocolos e Princípios FAIR no suporte à padronização e adequada dos dados, tornando a iniciativa cada vez mais próxima a estes princípios norteadores.

5 CONTRIBUIÇÕES

Neste capítulo apresentamos as principais contribuições desta dissertação.

5.1 Correlação da HD com AD

Cientistas da informação e profissionais das áreas agrárias atuam em ambientes cada vez mais digitalizados, datacêtricos e enfrentam os crescentes desafios de sustentabilidade e do gerenciamento de enorme volume de dados em projetos de ciência aberta (ALBAGLI *et al.*, 2014). Isso demanda novos olhares que considerem abordagens que ainda não são amplamente internalizadas nas universidades, empreendimentos agrícolas ou laboratórios de pesquisas.

Bessa e Cruz (2021) e Marinho *et al.*, (2020) pontuam que apesar do crescente interesse sobre os diferentes recursos de digitalização da agricultura (*big data, internet of things, realidade aumentada, robótica, sensores, impressão 3D, integração de sistemas, inteligência artificial, digital twins, blockchain* entre outros) e seus impactos na cadeia produtiva do agronegócio ainda existem poucos estudos correlacionando-os com as Humanidades Digitais.

Uma das contribuições desta pesquisa foi o questionamento das possíveis interseções, em ambientes digitais, entre a área de Humanidades Digitais e a área de segurança de solos (via plataforma *OpenSoils*). Abordou-se a questão sobre como tais áreas poderão se beneficiar dos princípios FAIR e metadados de proveniência, seguindo na direção multidisciplinar das Humanidades Digitais.

Diante do grande desafio na área de segurança de solos, acreditamos que as Humanidades Digitais podem auxiliar a explorar demandas dentro do contexto de grandes mudanças por conta do paradigma da transformação digital. Muitas das práticas são comuns às HD e AD quando visam a análises de dados, em prol de ações que envolvem sustentabilidade, e do impacto social.

Defendemos que as Humanidades Digitais se relacionam mesmo que indiretamente, com a Agricultura Digital através de aspectos sociotécnicos, artefatos computacionais e tecnologia digital. Essa percepção foi adquirida ao analisar alguns projetos internacionais comparando-os com o projeto *OpenSoils*.

No contexto internacional, o projeto *Europeana Cloud*²⁵ é um portal da Web criado pela União Europeia que contém coleções digitalizadas de patrimônio cultural de mais de 3.000 instituições em toda a Europa. Inclui registros de mais de 50 milhões de artefatos culturais e científicos, reunidos em uma única plataforma e apresentados de várias maneiras (interfaces e repositórios digitais) para os usuários. O protótipo da *Europeana* foi a *European Digital Library Network* (EDLnet), lançada em 2008.

Ao avaliarmos o portal *Europeana Cloud*²⁶, foi investigada a conexão entre o repositório digital do patrimônio cultural da Europa com foco na Política Agrícola Comum (PAC) e as cooperativas agrícolas (MANOUSELIS, 2016). O PAC é voltado para a área agrícola, facilitando aos cientistas das áreas de agricultura e alimentos acesso aos recursos disponíveis. Verificamos que o portal não mantém coleções de dados de solos ou de segurança de solos.

Nossa investigação sobre o PAC se apoiou nos questionamentos de Manouselis (2016). O autor indagou como as Humanidades Digitais se relacionam com as Ciências Agrárias e Alimentares na *Europeana*. O autor defende que elas têm um relacionamento essencial. A investigação socioeconômica e a pesquisa política são parte integrante da pesquisa agrícola, da inovação e da tomada de decisões. Eles ajudam a trazer a dimensão e questões humanas no processo de pesquisa, contribuindo de forma crítica para condução de pesquisas em temas transversais como pobreza, segurança alimentar, desequilíbrios nutricionais e gestão de recursos naturais. Complementamos o autor ao destacar que todos esses aspectos têm como elemento comum a geração de grandes volumes de dados e a necessidade de repositórios digitais.

A criação dos repositórios digitais desta plataforma se deve ao setor de patrimônio cultural e tornou-se possível a partir da colaboração de três organizações – *Fundação Europeana*, *Europeana Network Association* e *Fórum de Agregadores*²⁷, e segue condições como “ativos digitais de boa qualidade em formatos padronizados, sendo possível compartilhar,

²⁵ Disponível em: <https://pro.europeana.eu/project/europeana-cloud>

²⁶ Disponível em: <https://pro.europeana.eu/project/europeana-cloud>

²⁷ Disponível em: <https://pro.europeana.eu/about-us/mission#who-we-are>

explorar, interrogar e usar suas coleções de maneiras que cumpram suas próprias missões” (EUROPEANA PRO, *online*).

Verificamos que existe muita ressonância entre os Princípios do FAIR e os repositórios de dados do portal *Europeana*. A *Europeana* agrega registros de metadados sobre objetos digitais. Para tal, ela e sua rede de parceiros desenvolveram um “Quadro de Licenciamento” que usa padrões de metadados abertos e rotulagem obrigatória de conteúdo com declarações de direitos de uso padronizadas e interoperáveis, o que torna clara as possibilidades de reutilização de cada item de dado ou *dataset* (ISAAC; FREIRE, 2020). Estes registros fornecem ligações para conteúdos digitalizados acessíveis nos repositórios de dados dos parceiros da *Europeana* e alimentam o serviço de pesquisa que facilita a descoberta de materiais para uma grande variedade de públicos distintos.

Neste caso da *Europeana*, é possível ver o impacto das práticas das Humanidades Digitais em áreas com temas transversais, diante da necessidade de gestão de grandes volumes de dados que variam desde atividade culturais até os envolvidos na produção de alimentos, no uso sustentável dos solos e segurança alimentar.

5.2 Aspectos de FAIRificação de plataformas da AD

Baseado na literatura, reiteramos que as Humanidades Digitais podem propiciar novas possibilidades na investigação cruzada com a Agricultura Digital devido a sua conexão com práticas arquivísticas, bibliométricas, humanísticas e computacionais no que diz respeito ao trato dos ativos informacionais. Esta seção traz uma contribuição ligada aos aspectos conceituais, que pode ser aplicada em projetos de AD, tais como o *OpenSoils*.

Bessa e Cruz (2020) desenvolveram um modelo de representação conceitual descrevendo as etapas de um projeto AD e a produção de dados de pesquisa com correlação com Princípios FAIR e Proveniência de Dados. Graficamente, as principais correlações estão ilustradas a seguir na Figura 14.

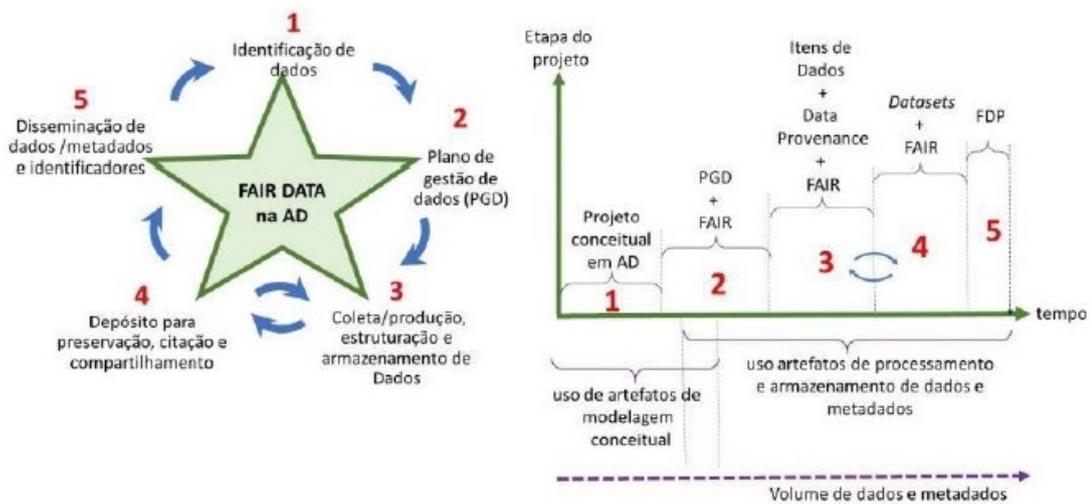


Figura 14 – Representação conceitual (estrela) com as etapas de um projeto AD e correlação com princípios FAIR e proveniência de dados e diagrama que representa as etapas ao eixo do tempo. (BESSA; CRUZ, 2020)

Segundo os autores, as correlações iniciam-se na fase de elaboração e identificação dos projetos de AD, definição de suas regras, atores e seus papéis ao longo do ciclo de vida do projeto de geração dos dados é essencial. No entanto, observamos que, em geral não há grandes esforços de armazenamento e ou processamento desse tipo de informações. Defendemos que maiores esforços sejam depreendidos levando em conta a *definição e a identificação* dos atributos de dados (que sejam semanticamente bem definidos), seus tipos, formatos e definição das estruturas e *datasets* a serem produzidos.

Esta proposta conceitual foi concebida tendo a Agricultura Digital e o *Opensoils* como pontos de partida, por serem inerentes a esta pesquisa. Entretanto, esta prática poderá ser aplicada em projetos de diferentes domínios, adaptada às diversas áreas de conhecimento.

A segunda etapa do ciclo se alia a necessidade de definição de Planos de Gestão de Dados (PGD) para os projetos de Agricultura Digital. Destacamos que ainda existem poucos trabalhos na literatura que explicitamente correlacionam PGD com AD e papéis e atores. Os PGD em projetos de AD podem ser elaborados por ferramentas digitais e devem incluir as metodologias da criação/coleta de dados, modelos de dados, documentação e metadados; aspectos legais e éticos, políticas de armazenamento durante a pesquisa, além de indicações de como os dados serão preservados e compartilhados e descrição das responsabilidades e proprietários dos dados. A partir deste ponto do ciclo o processamento das tarefas tende a ser

massivo, a adoção dos princípios FAIR é mais evidente e passa a desempenhar um papel central para o sucesso do projeto.

A terceira etapa do ciclo se aproxima da operacionalização da coleta de itens de dados e execução de tarefas computacionais, além da estruturação e armazenamento dos *datasets*. Essa etapa é central, há um ciclo interno de produção de itens de dados com os *datasets*. Ela assegura a confiabilidade e a acessibilidade dos dados ao longo do tempo. A partir deste ponto a proveniência de dados, do tipo retrospectiva, passa a desempenhar um papel crucial, visto que sua coleta e armazenamento junto aos metadados e *datasets* são mecanismos complementares que visam assegurar os aspectos de qualidade dos dados, proveniência, reprodutibilidade dos processos geradores de dados em AD.

A quarta etapa está intimamente relacionada com a preservação e compartilhamento dos *datasets* ao longo do tempo. Inevitavelmente, a etapa suscita novas situações que estão além das questões tecnológicas anteriores, levantam muitos pontos relacionados que se alinham com o sucesso do projeto que vão desde a propriedade intelectual, aspectos legais de licenças de uso/compartilhamento até pontos ligados à privacidade e autoria (citação) dos *datasets*.

Por fim, a quinta etapa trata dos requisitos necessários a assegurar a curadoria e eficiência da disponibilidade dos *datasets* ao longo do tempo, destacamos que aos *datasets* devem ser atribuídos os *persistent identifiers* (PID) além de etiquetas de metadados para assegurar a acessibilidade automática por máquinas através de RD do tipo *Fair Data Points* (FDP), ampliando a visibilidade desejada dos dados e do projeto propriamente dito, além de aprimorar a preservação dos ativos digitais presentes nos *datasets*.

Os *Fair Data Points* (BONINO, 2020) consistem em uma aplicação no servidor que expõe metadados na internet seguindo os princípios FAIR. Permitem exposição de metadados de objetos digitais (por seus donos, criadores ou editores); possibilitam que usuários ou consumidores descubram informações sobre os objetos digitais que os interessam; e apoiam interações com agentes humanos e *softwares*.

Aproximar o *OpenSoils* o máximo possível, destas características, é um dos caminhos para que esta plataforma possa contribuir cada vez mais na área de segurança de solos, seguindo moldes das iniciativas da área.

5.3 Os caminhos da FAIRificação da Plataforma *OpenSoils*

Nesta seção, apresentamos primeiros estudos de FAIRificação da base de dados *OpenSoils*, como contribuição inicial para a área de Humanidades Digitais no sentido de demonstrar como o uso de artefatos computacionais na Agricultura Digital podem alinhar-se com o foco da discussão.

Os métodos escolhidos para avaliar o grau de *FAIRness* e a condução de um futuro processo de FAIRificação desta plataforma e sua base de dados são essenciais. Segundo a literatura levantada, nossa escolha está baseada no *checklist* do ANEXO 1. Na ausência de fundamentação conceitual mais consolidada, adotar o *checklist* viabilizará esforços para tornarem os dados acessíveis, reutilizáveis e interoperáveis. Ou seja, alcançar o estado mais FAIR tanto quanto for possível, visando o reuso de dados científicos em pesquisas relacionadas.

Nesta discussão, apresentamos abaixo, quadro comparativo com uma proposta inicial que visa a identificar na plataforma *OpenSoils*, através deste diagnóstico, sua **aderência** a cada Princípio FAIR norteador.

Princípio	Descrição da Aplicação	OpenSoils
<i>Findable</i>	Para serem encontrados	
<i>F1</i> – (Meta) dados são atribuídos a identificadores globais, únicos e persistentes	Identificadores únicos para conjuntos de (meta)dados. Ex.: DOI; ARK; RRID; PID	Cumpriu parcialmente
<i>F2</i> – Os dados são descritos com metadados ricos.	Os metadados devem ser generosos e extensivos para que a partir da indexação, os dados possam ser encontrados, mesmo sem um identificador.	Cumpriu
<i>F3</i> – Metadados incluem clara e explicitamente o identificador dos dados que descrevem.	Os metadados e o conjunto de dados que eles descrevem são geralmente arquivos separados. Esta associação deve ter um identificador persistente e globalmente único.	Cumpriu

<i>F4</i> – (Meta) dados são registrados ou indexados em um recurso pesquisável.	Por si só não garantem facilidade de localização, portanto dependem de indexação de baixa granularidade, indexados por mecanismos de busca.	Cumpriu parcialmente
<i>Accessible</i>	Para serem acessíveis	OpenSoils
<i>A1</i> – (Meta)dados são recuperáveis por seu identificador usando um protocolo de comunicação padronizado.	Dados e metadados devem ser recuperados independente de seus identificadores, impedindo barreiras de acesso, como protocolos ou intervenção humana.	Cumpriu
<i>A 1.1</i> – O protocolo é aberto, gratuito e universalmente implementável.	Para maximizar a reutilização de (meta)dados o protocolo deve ser gratuito e aberto, favorecendo sua implementação global.	Cumpriu
<i>A 1.2</i> – O protocolo permite um procedimento de autenticação e autorização, quando necessário.	O ‘A’ em FAIR não necessariamente significa aberto ou livre. Mas sim fornecer condições (protocolo de comunicação) exatas para acesso, mesmo para dados altamente protegidos.	Cumpriu
<i>A2</i> – Os metadados são acessíveis, mesmo quando os dados não estão mais disponíveis.	Links inválidos gastam o tempo do usuário, que procuram dados que podem não estar mais lá. A afirmação significa que os metadados devem persistir mesmo quando os dados não são mais sustentados.	Cumpriu
<i>Interoperable</i>	Para serem Interoperáveis	OpenSoils
<i>II</i> – Os (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.	A troca e a interpretação de dados pelos computadores necessita minimamente de vocabulários controlados comumente usados, ontologias, tesouros e um bom	Cumpriu

	modelo de dados (boa estrutura para dados e metadados).	
I2 – (Meta) dados usam vocabulários que seguem os princípios FAIR.	O vocabulário controlado usado para descrever o conjunto de dados deve ser facilmente encontrado pelos usuários, usando identificadores globais únicos e persistentes.	Cumpriu
I3 – (Meta) dados incluem referências qualificadas a outros (meta) dados.	Estas referências cruzadas entre conjuntos de dados visa a enriquecer o conhecimento contextual entre os dados, vinculando-os entre si.	Cumpriu
Reusable	Para serem Reutilizáveis	OpenSoils
RI – (Meta) dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes.	Este princípio determina a capacidade de um usuário (humano ou máquina) decidir se um determinado contexto é útil. Quanto mais generoso for o autor do metadado, mais condições de gerar metadados úteis.	Cumpriu parcialmente
R 1.1 – (Meta) dados são liberados com uma licença de uso de dados clara e acessível.	Trata-se de interoperabilidade legal. As condições de direito de uso e licenciamento devem ser claras.	Cumpriu parcialmente
R 1.2 – (Meta) dados são associados à proveniência detalhada.	Para que outros utilizem os dados, eles deverão saber de onde vieram os dados, sua origem e histórico, quem deverá ser citado e a quem deverão ser atribuídos os créditos. Deverão ser incluídas informações sobre como os dados foram gerados ou coletados; como foram processados; se já foram publicados	Cumpriu

	anteriormente; se contém dados de outros autores que possam ter sido modificados ou completados. Notar que todo este fluxo deve ser legível por máquina.	
RI.3 – (Meta) dados atendem aos padrões da comunidade relevantes para o domínio.	É mais fácil utilizar conjuntos de dados semelhantes. Padrões da comunidade ou melhores práticas para arquivamento ou compartilhamento também devem ser seguidos.	Cumpriu

Quadro 5 – Guia de Princípios FAIR e ANEXO 1, comparando o *OpenSoils* (WILKINSON *et al.*, 2016; adaptado pela autora)

De acordo com o quadro comparativo, percebe-se a necessidade de uma continuidade na análise, visto que características dos objetos digitais, da plataforma *OpenSoils*, estão em contínuo processo de aproximação aos princípios norteadores – FAIR. O uso de um sistema como *Dublin Core* é uma das possibilidades para tornar o RD do *OpenSoils* mais alinhado aos objetivos desta pesquisa. Este, é um administrador confiável de vocabulários de metadados, esquemas de conceito e outros artefatos de metadados (DCMI, *homepage*)

O processo de FAIRificação de uma base de dados é um trabalho dinâmico, longo e complexo, pois é necessário considerar todos os relacionamentos que estes objetos devem ter entre si e com as políticas às quais se submetem.

A plataforma *OpenSoils* foi desenvolvida com aportes para captura de proveniência e princípios norteadores (FAIR), conferindo mais credibilidade e segurança às pesquisas e dando origem a resultados satisfatórios e confiáveis.

Após análise da aderência aos Princípios FAIR, temos as seguintes conclusões, tendo o ANEXO 1 como métrica para cada princípio norteador:

Princípio Localizável	<i>OpenSoils</i>
Critério 1 (:F1)	Parcialmente. Os dados ainda não estão totalmente linkados. Somente as tabelas

Existem identificadores persistentes atribuídos aos dados?	do BD tem suas colunas relacionadas a identificadores.
Critério 2 (:F2) Existem metadados ricos que descrevem os dados?	Sim. Os metadados são utilizados com o máximo da especificidade inerente ao domínio (solos), padronizados e sob um dicionário de dados.
Critério 3 (:F3) Os metadados estão online em um recurso pesquisável, por exemplo um catálogo ou repositório de dados?	Sim. Os metadados e conjuntos de dados associam-se pelas tabelas, que se relacionam, cumprindo o requisito.
Critério 4 (:F4) O registro de metadados especifica o identificador persistente?	Parcialmente. Os objetos digitais são pesquisáveis em mecanismos de busca embora ainda estejam em um BD relacional e não plenamente linkados na web de dados. Ainda assim, são pesquisáveis.
Princípio Acessível	<i>OpenSoils</i>
Critério 5 (:A1) Seguir o ID persistente levará aos dados ou metadados associados?	Sim. Os (meta)dados encontram-se disponíveis.
Critério 6 (:A1.1) O protocolo pelo qual os dados podem ser recuperados segue padrões reconhecidos?	Sim. O acesso a plataforma é aberto e gratuito, favorecendo o acesso.
Critério 7 (:A1.2) O procedimento de acesso inclui etapas de autenticação e autorização?	Sim. Através de um login e senha, o usuário tem seu acesso permitido, conforme a configuração de segurança do acesso.
Critério 8 (:A2) Os metadados são acessíveis, sempre que possível, mesmo que os dados não o sejam?	Sim. Os metadados se mantêm disponíveis, para acesso.
Princípio Interoperável	<i>OpenSoils</i>
Critério 9 (:I1) Os metadados fornecidos seguem padrões relevantes?	Sim. O uso de ontologias e tesouros, por exemplo, vem se tornando extremamente importante de modo a refinar cada vez mais o uso de termos precisos.

Critério 10 (:I2) Vocabulários controlados, palavras-chave, tesouros ou ontologias são usados sempre que possível?	Sim. A partir de harmonização de dados via workflows ETHL, há uma harmonização e maior qualidade dos dados, que os tornam prontos para a representação do conhecimento.
Critério 11 (:I3) Referências qualificadas e links são fornecidos para outros dados relacionados?	Sim. Os dados referenciam-se entre si, na plataforma. Integram relacionamentos persistentes entre si.
Princípio Reutilizável	<i>OpenSoils</i>
Critério 12 (:R1) Os dados são curados e bem descritos com atributos relevantes?	Parcialmente. A partir de uma análise (ex.: mapas conceituais), o valor semântico dos objetos digitais vem sendo cada vez mais refinado, cumprindo a proposta.
Critério 13 (:R1.1) Os dados têm uma licença de uso de dados clara e acessível?	Parcialmente. Será incorporada ao projeto a anonimização dos dados permitindo sua reutilização, garantindo cumprimento da LGPD. ²⁸
Critério 14 (:R1.2) Está claro como, por que e por quem os dados foram criados ou processados?	Sim. A fim de garantir qualidade, precisão, contexto de aplicação, além de veracidade dos objetos digitais, é muito importante que se tenha cada vez mais o caminho histórico de um dado digital. A plataforma possui métodos de registro e coleta de proveniência retrospectiva bem definidos apoiada na especificação PROV.
Critério 15 (:R1.3) Os dados e metadados atendem aos padrões de domínio relevantes?	Mesmo com carga de novos datasets ou atualização dos existentes, os valores são devidamente guardados e atualizados, de modo a continuarem sendo reconhecidos.

²⁸ Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

Quadro 6 – Checklist (SARAH; MARJAN, 2017) como base para Processo de FAIRificação do *OpenSoils* (BESSA, 2021)

Como foi visto na análise, serão necessários ajustes para que o Banco de Dados esteja cada vez mais alinhado às necessidades do contexto atual, tanto da Ciência Aberta como às demandas de órgãos internacionais na área agrícola (CRUZ *et al.*, 2018)

A metodologia usada nesta pesquisa pode servir como base para o processo de FAIRificação de outros bancos de dados. Entretanto, vale ressaltar que certas características ou conclusões são estudos iniciais e reflexões primárias, tendo o banco de dados do *OpenSoils* como base. Além do fato de que este estudo é um dos primeiros a serem desenvolvidos na área de Humanidades Digitais.

5.4 Artigos Produzidos

Outras contribuições relevantes dessa dissertação são os artigos produzidos ao longo do estudo. No ano de 2020/21, foram produzidos três artigos, sendo que o primeiro da lista abaixo tem foco específico no tema desta pesquisa de mestrado sendo apresentado no II Congresso Internacional em Humanidades Digitais (HDRIO).

BESSA, A. C. F.; CRUZ, S. M. S. da. Investigando a adoção de Princípios FAIR e Proveniência de Dados na Agricultura Digital sob perspectiva das Humanidades Digitais: um estudo de caso na plataforma *OpenSoils*. In: HDRIO2021, 2021, Rio de Janeiro. II Congresso Internacional em Humanidades Digitais, 2021.

É importante ressaltar que o texto supramencionado foi escolhido para fazer parte de edição especial de uma revista mantida pela Associação Brasileira de Humanidades Digitais (<https://ahdig.hypotheses.org/>). Até o momento da defesa dessa dissertação o novo texto revisado e ampliado ainda não havia sido publicado pelos editores.

Os outros dois trabalhos têm foco em aspectos complementares a esta dissertação, a saber:

MARINHO, E. C.; PINHEIRO, A. G.; BESSA, A. C. F.; CRUZ, S. M. S. da; SCHMITZ, E. A. Proteção de Dados: Proposta de gerenciamento de dados de solos usando os princípios FAIR e a tecnologia blockchain. In: 4º Encuentro Latinoamericano de e-Ciencia - TICAL, 2020, Cuenca. Anais da Décima Conferencia de Directores de Tecnología de Información y Comunicación en Instituciones de Educación Superior, TICAL2020, 2020.

MARINHO, E. C.; PINHEIRO, A. G.; BESSA, A. C. F.; SCHMITZ, E. A.; CRUZ, S. M. S da. Agricultura 4.0: Uma abordagem a partir dos Princípios FAIR e tecnologias Blockchain para a gestão de dados de levantamentos pedológicos. In: 12º Congresso Argentino de AgroInformática - CAI 2020, 2020, Buenos Aires. Anais da 49 Jornadas Argentinas de Informática- 12 Congreso Argentino de AgroInformática, 2020.

6 CONCLUSÕES

Desenvolver pesquisas de caráter multidisciplinar e dinâmico é uma tarefa complexa. Nossa linha de pesquisa ainda está começando, mas já se vislumbra abertura de uma nova oportunidade de estudo mais aprofundado na interseção das áreas de Humanidades Digitais e Agricultura Digital.

Nosso estudo visou a buscar compreensão e soluções informacionais aplicáveis a artefatos computacionais e tecnologias digitais para a área da Agricultura Digital, através do estudo de indicadores para os futuros aprimoramentos da plataforma *OpenSoils*.

Os Princípios FAIR e a Proveniência de Dados quando aplicados na área de AD podem representar ganhos importantes. Ao serem aplicados em sua maior parte ou totalidade na plataforma *OpenSoils*, poderão melhorar o uso, o acesso, o compartilhamento, a localização e possivelmente o reuso de dados na área da segurança de solos, trazendo uma possível ampliação da quantidade de usuários a se beneficiarem da própria plataforma e dos seus dados, possivelmente agregando alguns novos aspectos do cotidiano da área, naquilo que diz respeito as dinâmicas sociais, nas quais se inserem as HD.

Neste processo de alinhamento, no contexto de Ciência Aberta e dados FAIR, não se pode deixar de mencionar que os dados devem ser abertos tanto quanto possível e fechados tanto quanto necessário, além de anonimizados, garantindo que a Lei Geral de Proteção de Dados Pessoais (LGPD - Lei nº 13.709, de 14 de agosto de 2018) seja devidamente observada, bem como preservados os direitos de quem fornece os dados. Essas condições devem ser respeitadas na base de dados do repositório da Plataforma *OpenSoils*.

Nosso estudo foi um dos primeiros do Brasil envolvendo aspectos pouco explorados pelas HD e AD. Ele aponta para a possibilidade de ampliar a integração da plataforma com outros repositórios de dados, idealmente um RD institucional na UFRRJ ou possivelmente, a outros externos à instituição.

A metodologia de verificação de FAIRificação e *FAIRness* para Base de Dados, seja na área de Humanidades Digitais seja na área de Agricultura Digital, também representa uma contribuição importante, visto que Repositórios Digitais institucionais necessitam cada vez mais estarem alinhados aos Princípios FAIR, de modo a se adequarem às crescentes demandas da Ciência Aberta, onde “o gerenciamento de dados de pesquisa é uma prática que vem sendo adotada como uma exigência das agências de fomento às pesquisas internacionais.” (HENNING *et al.*, 2018)

Concluimos assim, esta dissertação, a qual acreditamos ser somente o primeiro passo para o desenvolvimento de um estudo mais detalhado e ampliado com vistas a uma futura integração deste Repositório Digital com o Repositório de Dados institucional da UFRRJ. Percebemos que há uma clara correlação entre as Humanidades Digitais e Agricultura Digital, o que possibilitou e motivou esta pesquisa e, ao final, demonstramos os aspectos da FAIRificação da plataforma utilizada, o *OpenSoils*, com aporte para a Ciência Aberta.

Como produtos durante a pesquisa, foram apresentados e publicados, dentro da temática da pesquisa, um artigo de âmbito internacional na área de Humanidades Digitais e dois artigos em Congressos Internacionais nas áreas de Informática e Agroinformática.

Os próximos passos deverão ser dados na direção da aplicação técnica das métricas, para que a Plataforma *OpenSoils* esteja o mais alinhada possível aos Princípios FAIR e com Proveniência de Dados, tendo a Agricultura Digital como contexto real, e dando continuidade ao projeto do CNPq e ao estudo nas práticas das Humanidades Digitais.

ANEXO 1

Questionário norteador - Quão FAIR são seus dados?

Princípio - Localizável

Critério 1 - Existem identificadores persistentes atribuídos aos dados?

Critério 2 - Existem metadados ricos que descrevem os dados?

Critério 3 - Os metadados estão online em um recurso pesquisável, por exemplo um catálogo ou repositório de dados?

Critério 4 - O registro de metadados especifica o identificador persistente?

Princípio - Acessível

Critério 5 - Seguir o ID persistente levará aos dados ou metadados associados?

Critério 6 - O protocolo pelo qual os dados podem ser recuperados segue padrões reconhecidos?

Critério 7 - O procedimento de acesso inclui etapas de autenticação e autorização?

Critério 8 - Os metadados são acessíveis, sempre que possível, mesmo que os dados não o sejam?

Princípio - Interoperável

Critério 9 - Os metadados fornecidos seguem padrões relevantes?

Critério 10 - Vocabulários controlados, palavras-chave, tesauros ou ontologias são usados sempre que possível?

Critério 11 - Referências qualificadas e links são fornecidos para outros dados relacionados?

Princípio - Reutilizável

Critério 12 - Os dados são curados e bem descritos com atributos relevantes?

Critério 13 - Os dados têm uma licença de uso de dados clara e acessível?

Critério 14 - Está claro como, por que e por quem os dados foram criados ou processados?

Critério 15 - Os dados e metadados atendem aos padrões de domínio relevantes?

REFERÊNCIAS

ALLEA. Sustainable and FAIR Data Sharing in the Humanities (2020). Disponível em: <https://repository.dri.ie/catalog/tq582c863>

AHDig. Disponível em: <<https://ahdig.hypotheses.org/historico>>.

ALBAGLI, S., APPEL, A. L., MACIEL, M. L. E-science, ciência aberta e o regime de informação em ciência e tecnologia. Disponível em: <https://ridi.ibict.br/bitstream/123456789/854/1/124-540-1-PB.pdf>

ALVES, D. As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português, *Ler História*, 69 | 2016, 91-103.

ARELLANO, M. A. Preservação de documentos digitais. *Ciência da Informação*, Brasília, v. 33, n. 2, p.15-27, ago. 2004. Disponível em: <https://bit.ly/2RHf6is>

ANDS. What is data provenance? Disponível em: <https://www.ands.org.au/working-with-data/publishing-and-reusing-data/data-provenance>. Acesso em: 29 maio, 2021.

AUER, S., LEHMANN, J., NGOMO, Axel-Cyrille C. N., ZAVERI, A. Introduction to Linked Data and its Lifecycle on the Web. Disponível em: https://jens-lehmann.org/files/2013/reasoning_web_linked_data.pdf

BERNARDI, A. C. C. *et al.* Potencial de uso das tecnologias de agricultura e pecuária de precisão e automação. São Carlos: Embrapa, 2017. Disponível em: < <https://bit.ly/33WNpFl>>

BESSA, A. C. F.; CRUZ, S. M. S. da. Investigando a adoção de Princípios FAIR e Proveniência de Dados na Agricultura Digital sob perspectiva das Humanidades Digitais: Um estudo de caso na plataforma OpenSoils. In: II Congresso Internacional em Humanidades Digitais - HDRio20/21.

BERNERS-LEE; T. Linked data. (2006). Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>

BIZER, C.; HEATH, T.; BERNERS-LEE; T. Linked data: the story so far. Disponível em: <https://eprints.soton.ac.uk/271285/1/bizer-heath-berners-lee-ijswis-linked-data.pdf>

BONINO, L. Introduction to the FAIR Data Point. (2020). Disponível em: https://www.youtube.com/watch?v=PtS_ek7BXSA

BOYD, D.; CRAWFORD, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, T&F, v. 15, n. 5, p. 662–679, 2012.

BROWER, M.; PAPOUTSOGLU, E.; HENDRICKX, P. *et al.* The Farm data train infrastructure. [version 1; not peer reviewed]. *F1000Research* 2020, 9(ELIXIR):557 (poster) (<https://doi.org/10.7490/f1000research.1117994.1>)

BUNEMAN, P.; CHAPMAN, A.P.; CHENEY, J. Provenance management in curated databases. *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*. 1-12. 10.1109/WISE.2003.1254462.

BUNEMAN, P. Why and where: a characterization of data provenance. (2001). Disponível em: Departmental Papers (CIS) | Department of Computer & Information Science | University of Pennsylvania (upenn.edu)

BUNEMAN, P.; TAN, W.-C. Provenance in Databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD), pages 1171–1173, 2007. (Tutorial Track).

CAFÉ, L.; MELO, B. A.; BARBOZA, E. M. F. *et al.* Repositórios institucionais: nova estratégia para publicação científica na Rede. In: XXVI CBCC. Disponível em: <http://repositorio.febab.org.br/items/show/4158>

CAIO, A. Agricultura 4.0 deve otimizar a agricultura irrigada. Esalq, 2019. Disponível em: <<https://www.esalq.usp.br/banco-de-noticias/agricultura-40-deve-otimizar-agricultura-irrigada>>.

CAMARGO, L. S. A.; VIDOTTI, S. B. G. Uma estratégia de avaliação em repositórios digitais. *Repositório - FEBAB*, acesso em 8 de novembro de 2021, <http://repositorio.febab.org.br/items/show/4158>.

CASTRO, R. M. de; PIMENTA, R. M. Novas práticas informacionais frente às humanidades digitais: a construção de acervos digitais. *Inf. Inf., Londrina*, v. 23, n. 3, p. 523 – 543, set./dez. 2018.

CAVALCANTI, M. T.; SALES, L. F.; PIMENTA, R. M. Compartilhamento e acesso a dados de pesquisa em humanidades digitais. *Cadernos BAD (Portugal)*, n. 1, p. 30-40, 2018. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/110022>>. Acesso em: 17 maio 2021.

CGIAR. A global tree knowledge platform. Disponível em: <https://www.cgiar.org/innovations/a-global-tree-knowledge-platform/>

CHENEY, J., CHITICARIU, L., TAN, W-C. Provenance in databases: why, how, and where. *Foundations and Trends in Databases*. 1 379-474. 10. 1561. 1900000006. Disponível em: <<https://bit.ly/3c21hD8>>

CODATA. CODATA's mission. Disponível em: <https://codata.org/about-codata/our-mission/>

COELHO, M. R. *et al.* O solo e a sustentabilidade agrícola no Brasil: um enfoque pedológico. *Boletim informativo da SBCS*. Set - dez, 2014. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/item/137904/1/Coelho-et-al-2014.pdf>>

COLEMAN, R. A maturing partnership: eHumanities and the digital library. The University of Sydney: Australasia, 2007. Disponível em: <<https://bit.ly/3ohN4F1>>

COMISSÃO EUROPEIA. A política de ciência aberta da União Europeia

COSTA, M. M.; CUNHA, M. B. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. *Perspectivas em Ciência da Informação*; v. 19, n. 3 (2014). Disponível em: <<https://brapci.inf.br/index.php/res/v/37440>>

CRUZ, S. M. *et al.* Proveniência de dados na agricultura (2018). Disponível em: <<https://bit.ly/3o2ixuV>>

CRUZ, S. M. S. da; CEDDIA, M. B.; SCHMITZ, E. A. *et al.* Towards an e-infrastructure for Open Science in Soils Security. In: BRAZILIAN E-SCIENCE WORKSHOP (BRESKI), 12,

2018, Natal. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2018. DOI: <<https://doi.org/10.5753/bresci.2018.3273>>.

CRUZ, S. M. S. da; CRUZ, P.; CEDDIA, M.; MIRANDA, R. C. T. OpenSoils: e-science em segurança de solos. (2018) Disponível em: <https://bit.ly/3lNYG36>

CRUZ, S.; KLINGER, F.; CRUZ, P. et al. Desenvolvendo sistemas agrícolas de próxima geração: um estudo em Ciência de Solos. In: Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais (WCAMA), 10, 2019, Belém. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2019. p. 135-144. ISSN 2595-6124. DOI: <https://doi.org/10.5753/wcama.2019.6428>.

DATAONE. Data life cycle. Disponível em: <https://old.dataone.org/data-life-cycle>

DAVIDSON, B; FREIRE, J. Provenance and scientific workflows: challenges and opportunities. Disponível em: https://www.researchgate.net/publication/221212739_Provenance_and_scientific_workflows_Challenges_and_opportunities

DEMPSEY, L.; HEERY, R. Metadata: a current view of practice and issues. *Journal of Documentation*, 54 (2), March 1998. pp. 145-172. Disponível em: <<http://www.ukoln.ac.uk/metadata/publications/jdmetadata/>>

DODEBEI, V. L. D. Tesouro: linguagem de representação da memória documentária. Rio de Janeiro: Interciência, 2002.

DRUCKER, D. P.; CRUZ, S. M. S. et al. Implantação da rede temática GO-FAIR Agro Brasil: primeiros passos. In: XIII SBIAgro, 2021. Disponível em: <https://eventos.unipampa.edu.br/sbiagro/programacao/trabalhos-aceitos-sessoes-tecnicas/>

DTL Digital Life Sciences. Jointly designing a Data FAIRPORT - Dutch Techcentre for Life Sciences, 2014. Disponível em: <https://www.dtls.nl/2014/01/20/jointly-designing-data-fairport/>

DTL Digital Life Sciences. Farm data train. Disponível em: <https://www.dtls.nl/fair-data/farm-data-train/>

DUBLIN CORE METADATA INITIATIVE. Disponível em: <https://dublincore.org/>

EMBRAPA. Visão 2030: o futuro da agricultura brasileira. (2018) Disponível em: <<https://www.embrapa.br/visao/o-futuro-da-agricultura-brasileira>>

EMBRAPA. Pesquisa mostra o retrato da agricultura digital brasileira. 2020. Disponível em: <<https://bit.ly/3tMeJ4j>>

ENAP. Módulo 2: Princípios, importância e desafios do gerenciamento de dados. ENAP: Brasília, 2019. Disponível em: <<https://bit.ly/3nzoYH2>>

FAO, The future of food and agriculture Alternative pathways to 2050: FAO, 2018. Disponível em : <<http://www.fao.org/3/CA1553EN/ca1553en.pdf>>

FELISBERTO, P. M. Gestão dos metadados contidos nos identificadores de objetos de biblioteconomia. Araranguá: UFSC, 2016. Disponível em: <<https://bit.ly/3tEnxbs>>

FERREIRA, E., FÜHR, F., LIMA, K. C. R., ARAÚJO, P. C. & PEREIRA, S. Z. (2015). Digital Object Identifier (DOI): o que é, para que serve, como se usa?. AtoZ: novas

práticas em informação e conhecimento, 4(1), 5 – 9. Recuperado em: <http://dx.doi.org/10.5380/atoz/v4i1.42369>

FIESP/CIESP. Como estará o mundo em 2030? Quais serão as macrotendências? Disponível em: <https://bit.ly/3pjdIF7>

FIOCRUZ. Política de gestão, compartilhamento e abertura de dados para pesquisa: princípios e diretrizes. (2020). Disponível em: <https://www.arca.fiocruz.br/handle/icict/46408>

FIOCRUZ. Função modular sobre Ciência Aberta. (2021). Disponível em: https://campusvirtual.fiocruz.br/gestordecursos/mod_hotsite/ciencia-aberta

FOSTER. Open Science Taxonomy. Disponível em: <https://www.fosteropenscience.eu/foster-taxonomy/open-science?page=37>

GAYOL, V.; FLÓREZ, J. A. M. Presente y perspectivas de las humanidades digitales en América Latina, *Mélanges de la Casa de Velázquez*, [Online], 47-2 | 2017. Disponível em: <http://journals.openedition.org/mcv/7907>. DOI: <https://doi.org/10.4000/mcv.7907>

GIDA. Disponível em: <https://www.gida-global.org/>

GO FAIR BRASIL. Manifesto Go FAIR Brasil. Disponível em: [7c2b72_3d8e7c8431cb4896912de38b13d243e4.pdf](https://filesusr.com/7c2b72_3d8e7c8431cb4896912de38b13d243e4.pdf) (filesusr.com)

GODAN. Iniciativa de Dados Abertos na Agricultura. Disponível em: <https://www.godan.info/pages/agriculture>

GOMES, E. de C. A escrita na história da humanidade. Disponível em: https://issuu.com/eduardo/docs/eduardo_aspectos_da_escrita_na_historia_da_humanid

HIGOUNET, C. História concisa da escrita. São Paulo: Parábola, 2003. Disponível em: <https://bit.ly/311d7QC>

HENNING, P. C.; RIBEIRO, C. J. S.; BONINO, L. O.; SANTOS, P. X. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/84753>

HENNING, P. C.; RIBEIRO, C. J. S.; SALES, L. F.; MOREIRA, L. R.; BONINO, L. O. Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. *PBCIB*; v. 14, n. 3 (2019). Disponível em: <https://periodicos.ufpb.br/index.php/pbcib/article/view/46969>

HOCKEY, S. The history of humanities computing. In: (Hockey, 2007) Susan Schreibman, Ray Siemens e John Unsworth, eds., *Companion to Digital Humanities*. Oxford, Blackwell, 2004, pp. 15-16.

ISAAC, A.; FREIRE, N. Europeana and the FAIR principles for research data. (2020). Disponível em: <https://pro.europeana.eu/post/europeana-and-the-fair-principles-for-research-data>

JONES, S.; GROOTVELD, M. (2017, November). How FAIR are your data?. Zenodo. <http://doi.org/10.5281/zenodo.1065991>

KITCHENHAM, B.; CHARTERS, S. Guidelines for performing Systematic Literature Reviews in Software Engineering. EBSE Technical Report EBSE-2007-01. Disponível em: <https://bit.ly/3o3EGsR>

KLERKX, L.; JAKKU, E.; LABARTHE, P. A review of social science on digital agriculture, smart farming and agriculture 4.0: New contributions and a future research agenda. Disponível em: <https://doi.org/10.1016/j.njas.2019.100315>

LAKATOS, E. M.; MARCONI, M. A. Metodologia Científica. 2a . ed. São Paulo: Editora Atlas. 242 p. 1991.

LEVY, P. O médium algorítmico. Disponível em: <https://ateliadedhumanidades.com/2021/01/07/fios-do-tempo-o-medium-algoritmico-por-pierre-levy/#sdfootnote2sym>

LIBER. Implementing FAIR data principles: the role of libraries. Disponível em: <https://libereurope.eu/wp-content/uploads/2020/09/LIBER-FAIR-Data.pdf>

MANOUSELIS, N. Digital Humanities and Agricultural Cooperatives: is there a link? EuropeanaPro (2016). Disponível em: <https://pro.europeana.eu/post/digital-humanities-agricultural-cooperatives-is-there-a-link>

MARINHO, E. C.; PINHEIRO, A. G.; BESSA, A. C. F. et al., 2020. Proteção de Dados: Investigações sobre gerenciamento de dados pedológicos usando os princípios FAIR e tecnologias blockchain. Disponível em: < <https://bit.ly/3q9IdYD>>.

MARTÍNEZ-LAVANCHY, P.M., HÜSER, F.J., BUSS, M.C.H., ANDERSEN, J.J., BEGTRUP, J.W. (2019). FAIR Principles. In: Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), Research Data Management (eLearning course). doi: 10.11581/dtu:00000049

MEDEIROS, C. B. e-Ciência pode construir arca digital para enfrentar dilúvio de dados. (2011). Disponível em: <https://www.inovacaotecnologica.com.br/noticias/noticia.php?artigo=e-ciencia-arca-digital-diluvio-dados&id=010150111014#.YafDPdDMLIU>

MEDEIROS, C. B. Big data pode gerar benefícios políticos, sociais e econômicos. Disponível em: <https://agencia.fapesp.br/big-data-pode-gerar-beneficios-politicos-sociais-e-economicos/30333/>

MASSRUHÁ, S. M. F. S.; LEITE, M. A. de A.; OLIVEIRA, S. R. de M. *et al.* (Ed. Téc.). Agricultura Digital: pesquisa, desenvolvimento e inovação nas cadeias produtivas. Brasília, DF: EMBRAPA, 2020. Disponível em: < <https://bit.ly/3uYefIP>>.

MASSRUHÁ, S. M. F. S. Big Data como Ferramenta Estratégica no Desenvolvimento de Aplicações para a Agropecuária. A Mina de Ouro da Embrapa. (2019). Disponível em: <https://www.agrotic.com.br/wp-content/uploads/2019/08/SILVIA-MASSRUHA.pdf>

MORATO, A. de C.; MORAES, M. A. de. Metadados Dublin Core: uma breve introdução. E-LiS, 2010. Disponível em: < <http://eprints.rclis.org/14424/>>.

MOREIRA, J. L. R.; BONINO, L. O.; PIRES, L. F. et al. Towards Findable, Accessible, Interoperable and Reusable (FAIR) Data Repositories: Improving a Data Repository to Behave as a FAIR Data Point. 2019. Disponível em: < <https://binged.it/3ai09cj>>.

MONS, B. FAIR Principles. Disponível em: <https://www.youtube.com/watch?v=K40utIzUzOk>

MOUNIER, P. Manifeste des Digital Humanities. Journal des anthropologues [Online], 122-123 | 2010, Online since 01 December 2012, connection on 21 January 2021. URL:

<http://journals.openedition.org/jda/3652>; DOI: <https://doi.org/10.4000/jda.3652> MANIFESTE 2012

MOURA, M. A.; ARAÚJO, I. A. Humanidades Digitais e Ciência Aberta: perspectiva e desafios para a formação humana no campo da Ciência da Informação. In: IX Encontro Ibérico EDICIC (Barcelona, entre 9 e 11 de julho de 2019). Disponível em: <https://fima.ub.edu/edicic2019/sites/edicic2019/files/2019-05/246.pdf>

NASCIMENTO, S. B. R. TRANSFORMAÇÃO DIGITAL: Uma análise dos fatores envolvidos na evolução empresarial no setor de instrumentos musicais e acessórios no Brasil. Disponível em: <https://bit.ly/3xOJoko>

ODI. Open Data Institute. Enabling data access to support innovation in agriculture. Disponível em: <https://theodi.org/project/enabling-data-access-to-support-innovation-in-agriculture/>.

OECD, Digital Transformation in the Age of COVID-19. 2020 Disponível em: <https://www.oecd.org/digital/digital-economy-outlook-covid.pdf>

OLIVEIRA, S. S. C. de; DUARTE, E. B.; MARINHO, E. C.; CRUZ, S. M. S. da. Integração de *Data Lakes* Pedológicos através de Workflows ETLH. In: ERSI, ago. 2021.

ORCID. About ORCID. Disponível em: <https://info.orcid.org/what-is-orcid/>

PEREIRA, M. G.; ANJOS, L. H. C. dos; PINHEIRO JUNIOR, C. R.; PINTO, L. A. da S. R.; SILVA NETO, E. C. da; FONTANA, A. Formação e caracterização de solos. (2019). Disponível em: <https://bit.ly/3Gd628Z>

PIMENTA, R. M. Por que humanidades digitais na ciência da informação? perspectivas progressas e futuras de uma prática transdisciplinar comum. *Informação & Sociedade: Estudos*, v. 30, n. 2, 2020. DOI: 10.22478/ufpb.1809-4783.2020v30n2.52122 Acesso em: 08 nov. 2021.

POMERANTZ, J. Metadata. Cambridge: MIT Press, 2015.

PRONASOLOS. Programa Nacional de Solos do Brasil. O PronaSolos e seus objetivos. Disponível em: <https://www.embrapa.br/pronasolos/sobre-o-programa>

RDA. The Research Data Alliance (RDA) builds the social and technical bridges to enable the open sharing and re-use of data. (2016). Disponível em: <https://www.rd-alliance.org/about-rda>

RILEY, J. Understanding metadata: what is metadata, and what is it for? NISO Primer Series. Baltimore: NISO, 2017c. Disponível em: <https://bit.ly/3bgYrsR>

RNP. Rede Nacional de Ensino e Pesquisa. O que é Ciência Aberta e como ela pode facilitar a vida de cientistas. (2021). Disponível em: <https://www.rnp.br/noticias/o-que-e-ciencia-aberta-e-como-ela-pode-facilitar-vida-de-cientistas>

ROCHA, R. P. da *et al.* Acesso aberto a dados de pesquisa no Brasil: soluções tecnológicas: relatório 2018. Porto Alegre: UFRGS, 2018. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/185126/001082284.pdf?sequence=1&isAllowed=y>

ROCHA, R. P. da *et al.* Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto, . São Paulo:, v. 17, p. 1-25, 2021. Disponível em: Análise dos sistemas DSpace e Dataverse para repositórios de dados de pesquisa com acesso aberto |

Rocha | Revista Brasileira de Biblioteconomia e Documentação (febab.org.br) RUSSELL, I. G. ¿Qué son las Humanidades Digitales? Revista Digital Universitaria. 12(7), jul. Disponível em: <http://www.revista.unam.mx/vol.12/num7/art68/art68.pdf>

ROR. Research Organization Registry. Disponível em: <https://ror.org/>

SALES, L. F. Princípios FAIR aplicados à repositórios. (2021). Disponível em: https://www.arca.fiocruz.br/bitstream/icict/50571/2/RIAA_Curso_Princ%C3%ADpios_FAIR_20_10_2021.pdf

SALES, L. F.; SOUZA, R. F. de; SAYÃO, L. F. Publicação ampliada: um novo modelo de publicação científica voltada para os desafios de uma ciência orientada por dados. In: Encontro Nacional de Pesquisa em Ciência da Informação, 15., 2014. Anais... Belo Horizonte: ANCIB, 2014. GT7, p.3471-3492. Disponível em: <https://ridi.ibict.br/handle/123456789/824>

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. Inf. Inf., Londrina, v. 21, n. 2, p. 116 – 142, maio/ago., 2016. Disponível em: <https://bit.ly/3dX5oRA>

SANTOS, P. X.; ALMEIDA, B. A.; ELIAS, F. T. S. Livro Verde - Ciência aberta e dados abertos: mapeamento e análise de políticas, infraestruturas e estratégias em perspectiva nacional e internacional. Disponível em: <https://www.arca.fiocruz.br/handle/icict/24117>

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. Encontros Bibli: Rev. Eletrônica Biblioteconomia e Ciência da Informação, pp. 1–31, 2010. Disponível em: <https://bit.ly/2RFXe7E>

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. Transinformação, v. 20, n. 2, p. 133-148, 2008. DOI: 10.1590/S0103-37862008000200002 Acesso em: 08 nov. 2021.

SAYÃO, L. F.; SALES, L. F. Algumas considerações sobre os repositórios digitais de dados de pesquisa. Disponível em: https://brapci.inf.br/_repositorio/2016/12/pdf_ecf54d34cf_0000022200.pdf. DOI: 10.5433/1981-8920.2016v21n2p90

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. Disponível em: AtoZ: novas práticas em informação e conhecimento, 5(2), 67 – 71. Disponível em: <https://bit.ly/3tXuAfH>

SAYÃO, L. F.; SALES, L. F. Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. (2012). Disponível em: <https://bit.ly/3C3SpXg>

SAYÃO, L. F. Interoperabilidade das bibliotecas digitais: o papel dos sistemas de identificadores persistentes - URN, PURL, DOI, Handle System, CrossRef e OpenURL. Transinformação. 2007, v. 19, n. 1, pp. 65-82. Disponível em: <https://www.scielo.br/j/tinf/a/NTr5XbPG7LG5pWH876MmWVN/?format=pdf&lang=pt>. Epub 12 Set 2014. ISSN 2318-0889.

SEGUNDO, S. J. S.; ARAÚJO, J. W. de. Curadoria e ciclo de vida dos dados. (2019). Disponível em: https://www.researchgate.net/publication/340092407_Curadoria_e_Ciclo_de_Vida_dos_Dados

SEGUNDO SANTARÉM, J. E. Web semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente às iniciativas internacionais. In: XVI ENANCIB, 2015. Disponível em: <https://www.brapci.inf.br/index.php/article/download/43838>

SENAR. Introdução ao agronegócio. Senar: Brasília, 2015.

SILVA, C. da; ALMEIDA, J.; HOOPER, S. (2016). As humanidades digitais e as novas formas de disseminação do conhecimento. 10.13140/RG.2.2.17324.92800.

SILVA, F.C.C.; SILVEIRA, L. O ecossistema da Ciência Aberta. Transinformação, v.31, e190001, 2019. <http://dx.doi.org/10.1590/2318-0889201931e190001>

SITE INOVAÇÃO TECNOLÓGICA. e-Ciência pode construir arca digital para enfrentar dilúvio de dados. 14/10/2011. Online. Disponível em: <https://bit.ly/3wdYpKx>

SNA. A força da economia brasileira. (2020) Disponível em: <https://www.sna.agr.br/agronegocio/>

SULAMITA, B. R. N. Transformação digital: uma análise dos fatores envolvidos na evolução empresarial no setor de instrumentos musicais e acessórios no Brasil. (2019). Disponível em: https://www.fpl.edu.br/2018/media/pdfs/mestrado/dissertacoes_2019/dissertacao_sulamita_barbara_ribeiro_nascimento_2019.pdf

TAN, Wang-Chiew. Provenance in databases: past, current and future. Disponível em: C:\DEBull\2007\December\WANG-CHIEW\pdb-ieee.dvi (computer.org)

TERRAS, M. Inaugural Lecture: A Decade in Digital Humanities. (2014). Disponível em: <http://melissaterras.blogspot.com/2014/05/inaugural-lecture-decade-in-digital.html>

TOTVS. O que são cadeias produtivas do agronegócio? (Mar. 2020). Disponível em: <https://bit.ly/3rcYCe1>

TRANSFORMAÇÃO DIGITAL. O que é transformação digital? Disponível em: <https://transformacaodigital.com/o-que-e-transformacao-digital/>

VENKATRAMAN, V. The Digital Matrix: New Rules for Business Transformation Through Technology. Vancouver: Greystone Book, 2017.

WILKINSON, M.; DUMONTIER, M.; AALBERSBERG, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

WILKINSON, M.; SANSONE, S. A.; SCHULTES, E. *et al.* A design framework and exemplar metrics for FAIRness. (2017). Disponível em: <https://doi.org/10.1101/225490>

W3C. The PROV Data Model. Disponível em: <https://www.w3.org/TR/prov-dm/>