Revista GCTI
Gestão do Conhecimento e Tecnologia da Informação
RGCTI 5:46-65, 2021
ISSN: 2526-8384

Plataformas de *e-Science*: necessidades e desafios para um programa científico para a Amazônia

Plataformas de e-Science: Necessidades e desafios para um programa científico para a Amazônia

Ronaldo Ferreira da Silva[a], Edilson Ferneda[b], Fernando William Cruz[c], José Laurindo Campos dos Santos[d], Ana Paula Bernardi da Silva[e], Luiza Beth Nunes Alonso[f]

[a] *ronaldofsilva1@gmail.com*

[b] *eferneda@gmail.com*

[c] *fwcruz@gmail.com*

[d] *laurocampos2004@gmail.com*

[e] *anap.bernardi@gmail.com*

[f] *luiza.alonso@yahoo.com*

Resumo: O avanço da tecnologia da informação possibilitou a otimização de processos e maximização de resultados em diversas áreas, possibilitando, sobretudo, a quarta revolução industrial que tem seus pilares nas inovações tecnológicas, surgindo o termo indústria 4.0. Neste contexto, outras áreas também cooptaram e cunharam o termo "4.0" em novas definições onde as atividades finalísticas receberam em seu processo de concepção suportes importantes da tecnologia, dentre elas a produção científica, que faz cada vez mais de forma intensa o uso de tecnologia, sobretudo da computação de alto de desempenho (HPC - High-performance computing), fazendo, desta forma, surgir o termo Ciência 4.0 ou *e-Science*. O presente trabalho tem por finalidade analisar um conjunto de plataformas computacionais, selecionadas com base em critérios como disponibilidade para a comunidade científica brasileira e citações em estudos científicos, e apontar os componentes e características presentes nelas capazes de atender um agrupamento de dimensões conceituais para gestão de dados científicos e do conhecimento científico produzidos no contexto de um programa científico da Amazônia.

Palavras chave: e-Science, Plataformas computacionais, Gestão do conhecimento científico, Gestão de dados científicos.

Abstract: The advancement of information technology has enabled the optimization of processes and maximization of results in various fields, allowing, above all, the fourth industrial revolution which has its pillars in technological innovations, giving rise to the term Industry 4.0. In this context, other areas have also adopted and coined the term "4.0" in new definitions where finalistic activities received important technological support in their conception process, among them scientific production, which increasingly makes intensive use of technology, especially high-performance computing (HPC), thus giving rise to the term Science 4.0 or e-Science. The present work aims to analyze a set of computational platforms, selected based on criteria such as availability to the Brazilian scientific community and citations in scientific studies, and to identify the components and characteristics present in them capable of meeting a grouping of conceptual dimensions for the management of scientific data and knowledge produced in the context of a scientific program in the Amazon.

Keywords: e-Science, Computational platforms, Scientific Knowledge Management, Scientific Data Management.

1. Introdução

O Brasil tem despontado como um dos países com crescimento mais expressivo quanto a publicações científicas indexadas nas principais bases científicas mundiais, como *Institute for Scientific Information* (ISI) e *Scopus*. Esse crescimento justifica-se por fatores que incluem (Ganapati & Reddick, 2018): *(i)* investimentos em infraestruturas tecnológicas para suporte à pesquisa; *(ii)* apoio de agências de fomento; *(iii)* crescimento dos números de

ingressos em programas de pós-graduação; *(iii)* acirramento da cobrança por melhor desempenho dos pesquisadores e consequentemente dos programas de pós-graduação pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Entretanto, os centros de pesquisa brasileiros sempre contaram com pesquisadores de altíssimo nível, porém em quantidade insuficiente (Muccioli et al., 2007).

Por outro lado, o crescimento qualitativo deu-se de forma menos expressiva, distanciando o Brasil da elite da produção científica de grande relevância, como EUA, Alemanha, Canadá e França, que são protagonistas dos principais avanços científicos mundiais (Zago, 2008). Mas não cabe menosprezar os avanços quantitativos obtidos, e sim buscar incessantemente melhor qualificação dessa produção (Guimarães, 2011). Trata-se de uma rota inexorável: o Produto Interno Bruto (PIB), salvo casos pontuais, comanda a posição dos principais países na produção científica.

Outro aspecto importante no avanço da ciência é a interação entre os cientistas. Há colaboração científica quando dois ou mais cientistas trabalham juntos em um projeto de pesquisa e compartilham recursos intelectuais, econômicos ou físicos (Vanz & Stumpf, 2010, p. 42-55). Além do trabalho colaborativo, o surgimento de redes de pesquisa também tem se intensificado. Essas redes estão permitindo que grupos de pesquisadores concentrem esforços na solução de problemas cada vez mais complexos. Basta que dois cientistas sejam coautores em um texto para que exista uma conexão entre eles, uma rede (Newnam, 2001).

Com a vertiginosa evolução das tecnologias da informação e comunicação (TIC), é natural o surgimento de iniciativas para a disponibilização de recursos tecnológicos voltados para a construção, o uso e a disseminação de dados e conhecimento científico. É o caso, por exemplo, de *Data Observation Network for Earth* (DataONE)[1], projeto apoiado pela *US National Science Foundation* (NSF) para oferecer um *framework* distribuído e infraestrutura cibernética que visa atender as necessidades da ciência e da sociedade para acesso aberto, persistente, robusto e seguro a dados bem descritos e de fácil manipulação, relativos a observações de nosso planeta.

Plataformas de *e-Science* enquanto ambientes computacionais que permitem a gestão de dados, de informação e de conhecimento científico para potencialização dos trabalhos de pesquisa, são vistas como um ferramental importante para garantir a gestão da grande quantidade de dados de pesquisa que são gerados. Elas permitem que os dados primários sejam compartilhados entre os grupos de pesquisa para, por exemplo, reuso, modelagem, simulação, análise, e por gestores públicos na elaboração de políticas públicas e, de projetos educacionais.

Na última década, a viabilização de ciberinfraestruturas para *e-Science* no Brasil, sobretudo quanto à gestão de dados, tem como exemplos o Sistema Nacional de Processamento de Alto Desempenho (SINAPAD)[2], a GridUNESP[3] e a Rede Galileu[4]. Iniciativas que possibilitam a gestão e o tratamento de dados em larga escala proporcionando agilidade e precisão nos resultados de pesquisas científicas. No entanto, mesmo no contexto limitado à gestão de dados, não foram encontrados trabalhos que apontem os requisitos de uma plataforma de computação que possibilite uma gestão de dados que integre o contexto e o ambiente da pesquisa, a correlação dos dados com os resultados e o reuso destes dados em outras pesquisas.

Por outro lado, a Gestão do Conhecimento (GC) oferece um conjunto de processos que possibilitam, sobretudo, o compartilhamento do conhecimento organizacional. Na Gestão do Conhecimento Científico (GCC), esses processos dizem respeito à captura, organização, armazenamento e compartilhamento do conhecimento científico.

Para Ferreira (2010), a GCC, prioritariamente deve-se considerar a questão do capital humano e da colaboração, pois são as pessoas e a interação e colaboração entre elas que possibilitam a criação de conhecimento.

Este trabalho tem por finalidade analisar as

[1] *Data Observation Network for Earth* (DataONE) é uma fundação para novas ciências ambientais inovadoras por meio de uma estrutura distribuída e ciberinfraestrutura sustentável que atende às necessidades da ciência e da sociedade para acesso aberto, persistente, robusto e seguro à observação terrestre bem descrita que permite a descoberta de dados facilmente. (https://www.dataone.org)

[2] https://www.lncc.br/sinapad

[3] https://www.ncc.unesp.br/gridunesp/docs/v2

[4] http://redegalileu.lccv.ufal.br

características disponíveis em um grupo de plataformas de *e-Science*, categorizadas, nesta pesquisa, em *e-Infraestrutura*, *middleware* para serviços de nuvens e *frameworks* (arcabouços) baseados em ontologias para realizar a gestão de dados e de conhecimento científico alusivos ao seguinte conjunto de dimensões conceituais: *(i)* Compartilhamento de Dados, *(ii)* Conectividade *(iii)* Segurança; *(iv)* Governança e Gestão de Dados, *(v)* Armazenamento e Replicação de Dados, *(vi)* Curadoria Digital; *(vii)* Relação Semântica, *(viii)* Colaboração Científica, *(ix) workflow* científico e *(x)* Interdisciplinaridade) – inerentes à um programa científico da Amazônia, o LBA (*Large-scale Biosphere-Atmosphere Experiment in Amazonia*), coordenado pelo INPA (Instituto Nacional de Pesquisas da Amazônia)[5]. O LBA, criado em 1996 por meio de acordos internacionais de cooperação científica, é um programa interdisciplinar extenso que tem como pressuposto o estudo integrado entre Natureza e Intervenção Humana. (Avissar et al., 2002; Emilio & Luizão, 2014; Keller et al., 2004).

2. Referencial

Os princípios modernos de pensamento devem contemplar simultaneamente as contribuições das abordagens epistemológicas evolutivas, complexas e eticamente orientadas (Knyazeva, 2017). Do ponto de vista metodológico, o holismo significa construção de pontes entre ciências naturais e humanidades, ciência e engenharia, ciência e cultura, e até convergência dos valores mais elevados da humanidade: verdade, virtude e beleza (Knyazeva, 2017). Essa visão holística leva à busca por soluções complexas, desenvolvidas por grupos de pesquisadores de diferentes áreas do conhecimento e geograficamente dispersos. Nesse contexto, a *e-Science* abarca um conjunto interconectado e distribuído de hardware e software que dá suporte à criação de novos conhecimentos ancorados nos preceitos da complexidade, interdisciplinaridade e colaboração.

Considerando-se que os dados coletados especificamente no âmbito dos programas científicos da Amazônia revelam resultados que têm impacto em diversas regiões do país e em diferentes áreas do conhecimento, avaliar os componentes das plataformas computacionais de *e-Science* necessários para esse contexto é relevante para o aprimoramento desses serviços, essenciais aos avanços da ciência e da sociedade.

2.1 Pesquisa científica e complexidade

O contexto e abordagem da pesquisa científica, assim como a construção humana, vem mudando de acordo com desenvolvimento histórico da ciência, ancorados no surgimento de fenômenos e problemas que exibem uma face complexa, exigindo, desta forma, pensamentos, análises e soluções interdisciplinares. Os princípios que orientavam o tratamento sistemático de um tema ou problema no tempo de René Descartes diferem fundamentalmente dos princípios em construção hoje no cenário de uma ciência complexa e interdisciplinar (Almeida, 2009). Necessita-se, hoje, de um modo diferente de articular as informações para a construção do conhecimento. Este cenário complexo afasta os postulados do saber estritamente compartimentado que norteou o velho paradigma do Ocidente no século passado, dando sustentação à neutralidade científica, com a separação entre objeto e sujeito, e, sobretudo, criando um modelo para compreender a realidade baseado na observação, demonstração, verificação, experimentação e comprovação.

Existem muitos exemplos em que a aplicação de conhecimento "compartimentado", ou disciplinar, não é suficiente (Alonso et al., 2011). Muitos dos problemas atuais exigem soluções que a lógica tradicional linear não pode resolver, em especial no que tange aos problemas que envolvem questões sociais e ambientais, pois afetam variáveis de diversas disciplinas de forma independente.

Há, portanto, a necessidade de mudança de paradigmas que deve se sobrepor à busca por ordem e organização do pensamento cartesiano, insuficiente para produzir um conhecimento científico em face da sua atual complexidade (Morin, 1999; Almeida, 2009). Diante da complexidade de certos produtos científicos e contextos da pesquisa científica, por envolverem diversas disciplinas, a observação de forma monolítica não é apropriada (Alonso et al., 2011). Por exemplo, no contexto científico da Amazônia, cada disciplina oferece um conjunto de sensores que permitem coletar e registrar fatos que serão cientificamente interpretados e produzir previsões de diversos aspectos

[5] http://lba.inpa.gov.br

fenomênicos.

2.2 *e-Science*

O termo *e-Science* surgiu no Reino Unido para designar tecnologias desenvolvidas para apoiar pesquisas colaborativas e multidisciplinares. Mendes et al. (2011) destacam que esse termo é utilizado para externalizar a criação de uma plataforma computacional imbuída de instalações remotas compostas por recursos de computação distribuída, armazenamento em larga escala e o compartilhamento de grandes volumes de dados, resultados e conhecimento, indo, portanto, para além da definição de Medeiros e Caregnato (2012), para quem "[...] *e-Science* pode ser entendida como a infraestrutura para que cientistas e pesquisadores tenham acesso a dados científicos primários distribuídos, utilizando acesso remoto a esses conteúdos".

Com as novas TIC surgiram iniciativas para a disponibilização de recursos tecnológicos voltados para a construção, o uso e a disseminação de dados e conhecimento científico. Mattoso et al. (2008) consideram que fazer pesquisa científica moderna "implica, dentre outros aspectos, ubiquidade e distribuição, visando o desenvolvimento e execução de soluções com alto desempenho, baseadas em reutilização, gerência de dados e experimentos". Para atender às necessidades da ciência e da sociedade, é necessário, por um lado o acesso aberto, persistente, robusto e seguro a dados bem descritos e de fácil manipulação. Por outro lado, é preciso viabilizar a Gestão do Conhecimento Científico (GCC), ou seja, a captura, organização, armazenamento, compartilhamento e disseminação desses dados e do conhecimento científico, sintetizado pelo processamento e análise de dados primários. Dá-se o nome de *e-Science* à infraestrutura que reúne métodos e ferramentas computacionais voltados à produção do conhecimento científico nesse contexto.

Na pesquisa científica, as TIC têm papel essencial na captura, processamento e publicação de dados científicos. De acordo com Andronico et al. (2011), o método científico passa a prever a utilização de plataformas digitais para prover a aquisição de dados e resultados das pesquisas. Epistemologicamente, essa nova forma de concepção cria o que vem sendo considerado o quarto paradigma da ciência: *a ciência de dados*. As infraestruturas de *e-Science* podem ser representadas em três camadas (Andronico et al., 2011): *(i)* uma primeira camada composta pelos instrumentos científicos e experimentos que fornecem uma grande quantidade de dados; *(ii)* uma camada intermediária composta de grade (*grid*), centros de processamento de dados de alto desempenho em rede e de softwares de *middleware; (iii)* a camada superior, que inclui a interação com os pesquisadores, que realizam suas atividades independentemente da localização geográfica, interagindo com outros cientistas, realizando trabalhos em grupo, criando redes de colaboração geograficamente distribuídas. Andronico et al. (2011) destacam ainda que em face dos cada vez mais complexos problemas que norteiam a sociedade contemporânea, são necessários grupos de pesquisa interdisciplinares. Alonso et al. (2011) e Motloch (2016) corroboram e aprofundam a discussão sobre a complexidade na pesquisa, indicando que as investigações científicas precisam subsidiar as soluções para os problemas sociais e ambientais de difícil abordagem segundo os métodos reducionistas da ciência tradicional.

Devido à complexidade dos problemas, à necessidade de trabalhos colaborativos subsidiados por grupos de pesquisa e à necessidade de processar um volume cada vez maior de dados, a ciência vem se tornando dependente de infraestruturas eletrônicas Nesse sentido, Sampaio (2007) constata "o aumento expressivo na quantidade de dados coletados a serem validados e interpretados, fenômenos a serem estudados e resultados a serem analisados". Dessa dependência de recursos tecnológicos nasceu a *e-Science*. Segundo Peach (2004), *e-Science* é a "ciência que pode ser alcançada pelo uso de computadores para conectar diferentes fontes de dados sobre um assunto, usualmente independentemente coletados, para extrair nova informação, a fim de gerar novo conhecimento e entendimento". Assim,

> *A e-Science, como estrutura que visa à colaboração entre cientistas a partir do compartilhamento e gerenciamento de dados científicos primários, parece ganhar corpo, uma vez que é parte essencial de uma descentralização do conhecimento e da aplicação efetiva de recursos públicos em um país com pretensões de avanços significativos em ciência e tecnologia, possibilitando que cientistas de diversos*

> *ramos tenham acesso a conteúdo já mapeado* (Medeiros & Caregnato, 2012).

A *e-Science* também pode ser considerada como a combinação de três desenvolvimentos (Segura 2009): *(i)* recursos de computação em larga escala; (*ii)* acesso a conjuntos de dados maciços, distribuídos e heterogêneos, e *(iii)* uso de plataformas digitais para colaboração e comunicação. A Figura 1 ilustra esta afirmação.

Evidencia-se a integração de diversas fontes de informação para criação de bibliotecas digitais, utilizadas como um ambiente virtual de aprendizagem para interação entre cientistas e estudantes. Neste contexto, elas são essenciais para a compilação do conhecimento oriundo de fontes diversas possibilitadas, em especial para aquilo que Segura (2009) descreve como fontes de "conjunto de dados maciços, heterogêneos e distribuídos".

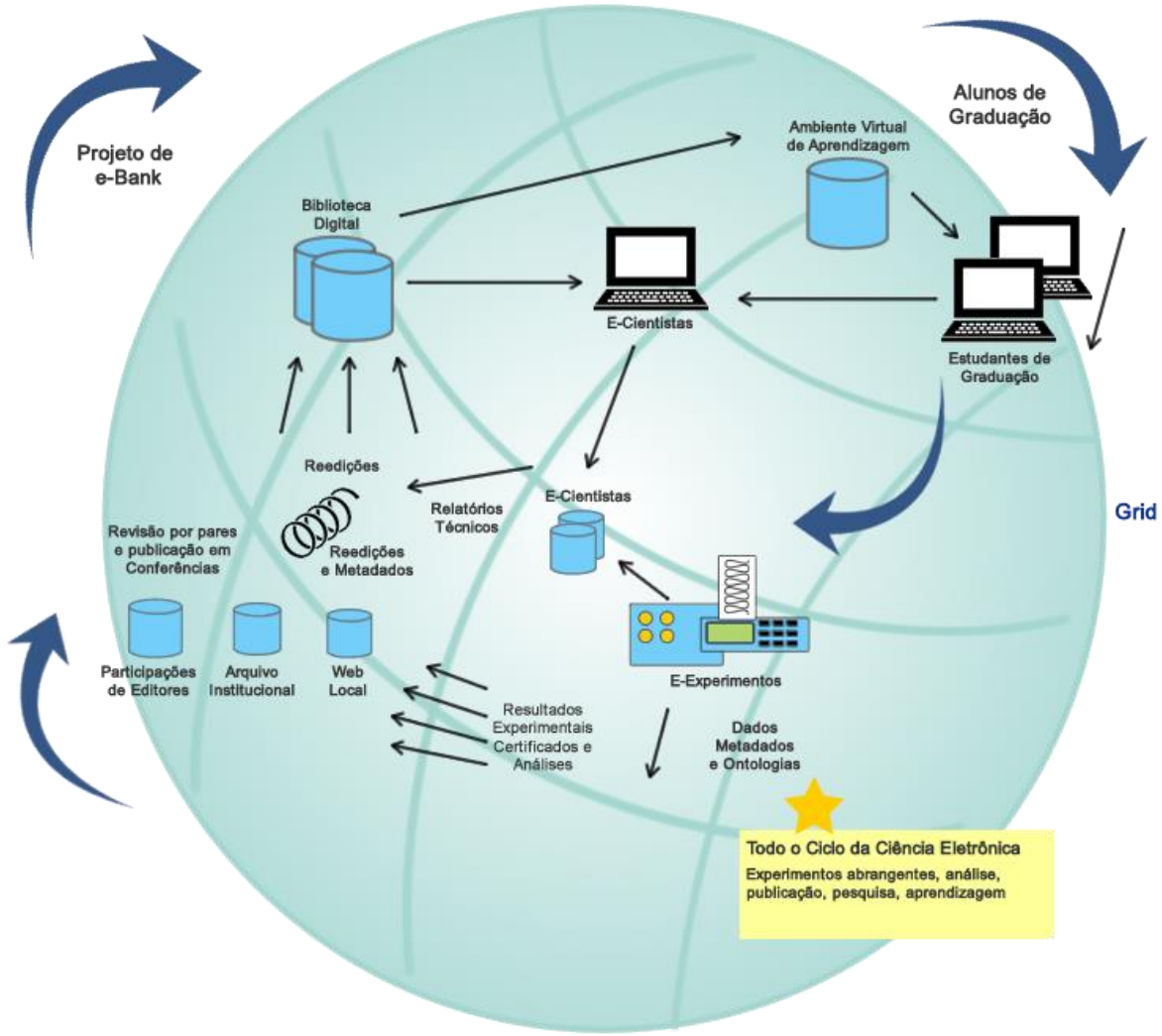
Ribes e Lee (2010) corroboram, ampliam os conceitos e descrevem os aspectos que caracterizam as transformações proporcionadas pela *e-Science*: *(i)* grupos de pesquisa colaborativos e interdisciplinares; *(ii)* coleta, representação e análise de dados dirigidos por computação intensiva; *(iii)* integração *end-to-end*.

Para Jankowski (2007), *e-Science* inclui ainda:

> [...] *(i) a colaboração internacional entre pesquisadores; (ii) o aumento do uso de computadores de alta velocidade interconectados, aplicando arquitetura de grids; (iii) a visualização de dados; (iv) desenvolvimento de ferramentas e processos baseados na Internet; (v) construção de estruturas organizacionais virtuais para a realização de pesquisas; (vi) a distribuição eletrônica e publicação de resultados.*

Como se pode observar, os autores conver-

Figura 1. Processo da *e-Science*



Fonte: Segura (2009)

gem quanto à necessidade das plataformas de *e-Science* possibilitarem o trabalho colaborativo de forma interdisciplinar e estender suas funcionalidades até o processo da comunicação científica.

As disciplinas que norteiam o conhecimento científico, integradas ao conjunto de recursos que possibilitam sua criação ou ampliação, assim como ocorre no contexto da *e-Science*, necessitam de fundamentos epistemológicos e filosóficos para caracterizar sua existência e validade dentro dos preceitos para além do senso comum. Saracevic (2009) aborda a confluência da TI com a Ciência da Informação e a importância da criação de ciberinfraestruturas para a ciência contemporânea:

> [...] *a ciência e a prática (recursos que permitem sua aplicação) lidam com a efetiva coleta, armazenamento, recuperação e uso de informações. Preocupa-se com informações e conhecimentos registráveis, e as tecnologias e serviços relacionados que facilitam a sua gestão e utilização.*

Ao afirmar que "a tecnologia revolucionou os métodos científicos e possibilitou o surgimento de novas metodologias que mudaram como os estudos científicos são feitos", Kaplan (2004) reforça a ideia de que os instrumentos científicos suportados por dispositivos tecnológicos são a chave para as práticas científicas contemporâneas. Ressalta também que a tecnologia em si não soluciona todos os problemas da ciência e da humanidade, e explica que a neutralidade tecnológica define a tecnologia em termos de suas propriedades técnicas: "este modelo apresenta tecnologia como uma forma de ciência aplicada. Esta abordagem considera a tecnologia simplesmente como uma ferramenta que pode ser usada para uma variedade de finalidades, sejam elas boas ou ruins".

Na perspectiva filosófica da tecnologia, há preceitos determinísticos sobre seu papel na construção e mudanças sociais, sinalizando que a evolução tecnológica precede desenvolvimentos da sociedade. Kaplan (2004) afirma que a tecnologia impõe percursos específicos na sociedade e é, portanto, responsável por mudanças culturais, sociais e políticas.

Kaplan (2004), no entanto, vai de encontro a essa hipótese afirmando que a tecnologia não pode ser uma força autônoma, pois a mudança social é não uma mera questão técnica, mas um assunto a ser tratado sob uma perspectiva humanística.

O *Research Counsils UK* (2009)[6] justifica a contribuição das plataformas de *e-Science* para o avanço da pesquisa e pontua suas características:

> [...] *considera não somente os sistemas que atendem às definições tradicionais da computação (software), mas também outras formas importantes de infraestrutura. As estruturas organizacionais (grupos formais e informais que oferecem serviços de e-Science), assim como o capital humano (conhecimento e experiência) e os recursos de dados e informações (sistemas que suportam o crescente volume de dados gerados pela pesquisa).*

Configura-se, portanto, a necessidade intrínseca da utilização de recursos tecnológicos para aportar melhores e mais abrangentes resultados na atividade científica. Considerando as particularidades do processo de criação do produto científico, as plataformas tecnológicas precisam considerar, dentre outros fatores: *(i)* o trabalho colaborativo geograficamente distribuído e *(ii)* a interdisciplinaridade do conhecimento gerado, oriundo de múltiplas fontes de dados e heterogêneos. Além disso, as plataformas habilitadas pela *e-Science* precisam considerar características de hardware e software, de forma que apoiem a comunicação, experiências e troca de ideias entre os cientistas.

Para Mattoso *et al*. (2008), "processos experimentais isolados, interligados apenas na concepção do cientista que conduz a análise, não são atualmente suficientes para tratar a complexidade imposta pelos problemas a serem analisados."

No que concerne os sistemas gerenciadores de dados científicos, há alta complexidade na utilização de modelos de dados convencionais, tais como o relacional ou hierárquico, para representação de dados e experimentos científicos. Appel (2014) afirma que "o armazenamento e análise de tais dados requer, por sua vez, a preexistência de uma infraestrutura computacional robusta, expansível e, prefe-

[6] O *UK Research and Innovation* tem o objetivo de otimizar as formas em que os Conselhos de Pesquisa trabalham em conjunto para cumprir seus objetivos: melhorar o desempenho geral e o impacto da pesquisa, treinamento e transferência de conhecimento do Reino Unido e ser reconhecido pela academia, empresas e governo para a excelência na promoção da pesquisa.

rencialmente, que possa ser organizada ou acessada de forma distribuída, para que os cientistas possam contribuir".

2.3 Ciberinfraestruturas para *e-Science*

Os avanços das TIC permitiram elevar a relevância dos resultados das pesquisas pelo tratamento de um volume cada vez maior de dados e, consequentemente, pela forma de fazer pesquisa. Price (1976) já antevia as repercussões da utilização das TIC na ciência em nossa vida cotidiana e no destino das nações. Nesse contexto, o compartilhamento de dados científicos primários possibilita que pesquisadores tenham acesso a conteúdos já mapeados e devidamente anotados (Medeiros & Caregnato, 2012). Este é o grande benefício obtido com o aumento do poder de processamento das máquinas e do surgimento de sistemas informáticos cada vez mais eficientes, possibilitando acesso em tempo real, compartilhamento e interação entre pesquisadores geograficamente distantes.

Nesse sentido, o Brasil apresentou, até meados da década passada, um considerável aumento de investimentos em infraestrutura de TI, como aqueles oriundos do Decreto nº 5.156 de 26 de julho de 2004, que institucionalizou o Sistema SINAPAD, uma rede distribuída geograficamente em oito unidades, os CENAPAD, operadas por diversas universidades e centros de pesquisa brasileiros (UFRGS, UFMG, UFC, UNICAMP, UFRJ, UFPE, INPE), coordenados pelo Laboratório Nacional de Computação Científica (LNCC).

Além do SINAPAD, diversas instituições brasileiras têm implementado centros de computação intensiva ou participam de redes internacionais distribuídas para oferta de processamento de alto desempenho, como: *(i)* Rede Galileu, já citada; *(ii)* EELA - *E-Infrastructure Shared Between Europe and Latin America* (EELA)[7]; *(iii)* CCES-CEPID - *Center for Computational Engineering & Sciences* (Unicamp)[8]. Outras iniciativas oferecem infraestrutura aberta para gestão de dados científicos, como o DataONE, *Open Archive Initiative* (OAI)[9] e DRIVER II[10].

A Rede Galileu, montada com recursos da Petrobrás e coordenada por Centro de Pesquisa (CENPES), é composta por cinco instituições âncoras envolvidas: Universidade de São Paulo (USP), Universidade Federal do Rio de Janeiro (UFRJ), Pontifícia Universidade Católica do Rio de Janeiro (PUC/RJ), Instituto Tecnológico da Aeronáutica (ITA) e Universidade Federal de Alagoas (UFAL). A temática da Rede Galileu engloba quatro áreas de conhecimento: Engenharia Naval, Estruturas, Computação Gráfica e Geomecânica.

O CCES-CEPID, centro de computação intensiva da Universidade de Campinas (Unicamp), mantido com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por meio do programa de pesquisa CEPID (Centros de Pesquisa, Inovação e Difusão), tem como objetivo integrar os diversos grupos de pesquisa, dando suporte em recursos computacionais e humanos para realização de pesquisas nas áreas de ciências moleculares computacionais, engenharia mecânica computacional, bioinformática, geofísica computacional e informática. O escopo da pesquisa no CCES abrange uma variedade de problemas intensivos em informática em diferentes domínios.

Criada no ano de 2006, EELA consiste em uma rede integrada e ambiente de processamento/armazenamento (*e-Infrastructure*) para realização de pesquisa colaborativa global. Por meio do compartilhamento de conhecimentos e recursos computacionais disponíveis na Europa (Itália, Portugal e Espanha), já integrados no âmbito da *Enabling Grids for E-sciencE* (EGEE) e América Latina (Argentina, Brasil, Chile, Cuba, México, Peru e Venezuela), viabilizou-se a criação de uma rede de pesquisa com o desenvolvimento de uma infraestrutura eletrônica de grid subjacente para aplicativos de e-Science.

O projeto EELA está dividido em pacotes de trabalho encarregados da criação de um banco de teste de *grids* interoperáveis comum usando recursos distribuídos existentes na América Latina e Europa, distribuídos em 15 centros de computação. Esta camada de teste depende da infraestrutura de rede fornecida pela Geant na Europa e RedCLARA na América Latina.

No âmbito da Amazônia, existem outras iniciativas, como o Sistema de Informações sobre a Biodiversidade Brasileira (SIBBr) (25), plataforma online concebida com o propósito de reunir a maior quantidade de dados e informa-

[7] https://www.eu-eela.eu

[8] https://cepid.fapesp.br

[9] https://www.openarchives.org

[10] https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/driver-ii

ções existentes sobre a biodiversidade do Brasil. Esse sistema trabalha em conjunto com o Metacat[11], do DataONE. Atualmente, o sistema armazena e compartilha mais de 10 milhões de registros.

A pesquisa científica contemporânea, abarcada sobretudo por um grande volume de dados primários e subsidiada por plataformas computacionais, evidencia a relação intrínseca entre redes de pesquisa, complexidade, interdisciplinaridade interfaceando com a necessidade de métricas e componentes computacionais para gestão de dados científicos e mecanismos para acurar a relação entre esses dados e os resultados produzidos.

3. Abordagem metodológica

O estudo partiu de uma revisão da literatura em que se buscou conectar *e-Science*, com suas infraestruturas propostas ou disponíveis. Em seguida, entrevistas semiestruturadas *in loco* com atores técnicos e científicos envolvidos com o programa de pesquisa LBA foi realizado, de forma a subsidiar uma discussão crítica sobre as aquelas plataformas à luz das necessidades desse Programa

Na primeira fase, foram analisadas publicações dos últimos dez anos, selecionadas por sua relevância, considerando o número de citações. As bases consultadas foram *IEEExplorer, Web Of Science, Springer, Science Direct* e *SciELO*. Os termos utilizados foram: *"complexidade da pesquisa científica + interdisciplinaridade"* e *"arquiteturas de e-Science para Gestão do Conhecimento Científico"* e *"Curadoria Digital"*, no idioma inglês. Esses mesmos termos em português foram considerados na busca por trabalhos acadêmicos na Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).

Em visita à sede do INPA buscou-se, por um lado, observar os recursos físicos disponíveis do Programa LBA/INPA e, por outro lado, a percepção dos pesquisadores sobre esses recursos e sobre a gestão de dados científicos implantadas.

As entrevistas semiestruturadas tiveram como objetivo aferir a percepção, por meio de perguntas subjetivas, das necessidades da comunidade científica do LBA/INPA com relação a infraestrutura de hardware e software disponível, além de políticas institucionais para gestão e governança de dados científicos, e identificar os desafios dos dirigentes e da equipe técnica com relação a implantação e gestão dos recursos informáticos para infraestrutura de *e-Science*.

A escolha das plataformas a serem analisadas baseou-se nos seguintes critérios: *(i)* apareceram nas pesquisas bibliográficas com maior recorrência; e *(ii)* fazem parte do arcabouço tecnológico disponibilizado para a comunidade científica brasileira. Foi o caso, por exemplo, do SINAPAD, da Rede Galileu, do CCES-CEPID e EELA, que embora seja uma plataforma da comunidade europeia, possui centros de processamento no Brasil, com a finalidade de promover sua convergência com as plataformas nacionais.

A partir do referencial teórico considerado, foram identificados quatro eixos temáticos que se configuraram como fundamentais para as práticas de *e-Science*, a saber: *(i)* trabalhos colaborativos em grupos de pesquisa; *(ii)* governança, gestão e política de dados; *(iii)* infraestrutura de TIC; *(iv)* desenvolvimento da pesquisa e publicação dos resultados. Esses eixos nortearam a análise documental das características proeminentes no conjunto de plataformas de *e-Science* selecionadas.

O objetivo da entrevista foi identificar as reais demandas por infraestrutura tecnológica para prover a gestão de dados e de conhecimento científicos para o Programa LBA, com o intuito de detectar as plataformas de *e-Science* que podem apoiar o referido Programa e apontar cenários possíveis, com ênfase na eficiência, economicidade e perenidade, para auxiliar no desenvolvimento da ciência no contexto da região Amazônica.

4. Resultados

Os resultados deste trabalho são apresentados a seguir na seguinte sequência: *(i)* identificação das plataformas de *e-Science* consideradas para este trabalho; *(ii)* apresentação das dimensões pertinentes às necessidades do Programa LBA/INPA; e *(iii)* análise das plataformas de *e-Science* de acordo com essas dimensões.

[11] Metacat é um catálogo flexível de metadados de código aberto e um repositório de dados que atende a dados científicos, particularmente de ecologia e ciência ambiental.

4.1 Identificação das plataformas de *e-Science*

As plataformas de *e-Science* consistem em um aparato de hardware e software para computação intensiva e distribuída, destinada ao processamento paralelo de dados em grandes escalas. A sua operacionalização remete a conceitos e plataformas computacionais já consolidadas e largamente utilizadas, como a computação distribuída e uso de protocolos de comunicação já presentes em diversas plataformas de tecnologia, a exemplo da Internet e de outras plataformas específicas. São, portanto, aparatos tecnológicos escaláveis e de alta disponibilidade que permitem o processamento de grandes volumes de dados em tempo aceitável.

No contexto de plataformas para *e-Science*, Andronico et al. (2011) destacam os *grids*, ou seja, um grande número de dispositivos de computação e armazenamento ligados entre si por redes de alta velocidade, dotado de um software mediador, denominado *middleware*. Este sistema permite que os recursos sejam compartilhados e utilizados como um único e grande computador "distribuído" que se "dissolve" na Internet e pode ser acessado de forma ubíqua por meio de serviços virtuais e interfaces de usuário. O *grid* e as redes subjacentes são denominados e-Infraestrutura.

Diante do aspecto heterogêneo destes sistemas, é necessário a adoção de regras e mecanismos que garantam a comunicação e operação entre eles, a base para essa interoperabilidade é a cooperação entre os provedores de tecnologia de *grids* computacionais (p.ex., gLite, Globus, UNICORE e ARC) e a implantação de infraestruturas distintas, mas de propósitos comuns (p.ex., EGEE, TeraGrid, DEISA, NorduGrid) representam um importante requisito social (Crichton et al., 2011). Para isso, o *Open Grid Forum* (OGF) propôs a *Open Grid Services Architecture* (OGSA), um conjunto de padrões que estendem *Web services* e arquitetura orientada a serviços[12] para o ambiente de computação em *grid*. Esses padrões estão descritos em um documento (Roberts et al. 2019; Roberts 2019) que detalha os elementos essenciais para implementação de ambientes de HPC, garantindo, desta forma, a interoperabilidade entre os sistemas. Os recursos descritos são gerenciamento de execução, gerenciamento de dados, gestão de recursos, segurança, autogestão de informações. A OGSA aborda ainda questões e desafios enfrentados atualmente na ciência ou em ambientes corporativos, como autenticação, autorização, negociação e execução de políticas, administração de acordos de nível de serviço, gerenciamento de organizações virtuais e integração de dados.

Nesse aspecto, o UNICORE, conforme descreve Kang et al. (2011), é uma tecnologia de *grid* que fornece software de grade que combina recursos de centros de supercomputadores e disponibiliza-os por meio da Internet. O sistema está em conformidade com o OGSA e vários padrões abertos, exigência da OGF, e é um meio importante de promover a interoperabilidade com o outro *middleware grid*, ampliando assim o horizonte dos usuários e desenvolvedores de aplicativos.

Corroborando a sua utilização em ambientes científicos, Kang et al. (2011) apresentam um estudo sobre um ambiente integrado de solução de problemas de cunho científico (PSE - *Problem Solving Environment*), baseado em UNICORE, disponibilizando uma interface para recursos de *grid*, oferecendo *plug-in* otimizado para aplicativos, viabilizando a criação soluções interoperáveis para computação intensiva de alto desempenho.

UNICORE é uma suíte de componentes modularizados para *middleware grid* contendo um cliente em ambiente gráfico e linha de comando, além de um portal *Web* de administração, componentes de servidor (*firewall, Web services*, serviços de *grid*), segurança, *workflow* e um conjunto de possíveis extensões.

O DataONE, desenvolvido pela NSF, foi o primeiro projeto de compartilhamento de dados de larga escala de ciências biológicas, da terra e ambientais do mundo. Compreende uma rede com dez *datacenters*, ou nós membros, incluindo centros de pesquisa e universidades norte-americanas, sendo nove com sede no território norte-americano e apenas um em outro país, sediado na África do Sul. Iniciativas como o DataONE evidenciam os benefícios promovidos pela *e-Science* no fomento das redes de pesquisa, com consequências positivas na produção de novos conhecimentos a partir de análises detalhadas e interdisciplina-

[12] Arquitetura Orientada a Serviços (SOA – *Service-Oriented Architecture*) é um modelo de arquitetura de software cujo princípio afirma que as funcionalidades de uma aplicação computacional devem ser disponibilizadas por meio de serviços.

res, compartilhando ciberinfraestrutura[13], possibilitando o aumento exponencial da capacidade de processamento de dados, uma vez que os dados compartilhados nestes ambientes podem ser utilizados por pesquisadores de diversas áreas e localidades.

A plataforma nacional para HPC, SINAPAD, utiliza o *middleware* CSGrid, uma instanciação do *framework* CSBase. Conforme destacam Gomes et al. (2015), "A estrutura CSBase é baseada no servidor central (também é possível uma implementação de *farm* de servidores) responsável por gerenciar todos os recursos computacionais subjacentes disponíveis para seus usuários." A plataforma é flexível e possibilita a criação de *gateways* científicos, além de fácil configuração por meio de arquivos *eXtensible Markup Language* (XML).

Considerando a complexidade na implementação de um ambiente de *grid* apontada por Meyer (2006), diversas soluções surgiram em ambientes de pesquisa com a finalidade de facilitar a implementação destes ambientes por meio da abstração de recursos de hardware utilizando softwares específicos. Dentre estas soluções, além dos *middleware* já apontados, está a *Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems* (Eucalyptus), uma infraestrutura ofertada como serviço (IaaS). Chaganti (2010) descreve-a como "uma infraestrutura de software livre para implementar computação em nuvem[14], de utilidade e elástica com o uso de *clusters* em nuvem ou de *farms* de estações de trabalho".

A estrutura do *middleware* é composta por um controlador de *nodes* (NC) que tem a finalidade de controlar a execução, inspeção, encerramento de instâncias onde a Máquina Virtual (MV) é executada. O controlador de cluster (CC) reúne informações e programa a execução da MV em controladores de nós específicos e gerencia a rede de instâncias virtuais. O controlador de armazenamento é o serviço que implementa a interface S3 (*Simple Storage Service*) da Amazon *Web Services*, sendo responsável pelo mecanismo de armazenamento e fornecimento de imagens e dados do usuário. O controlador de nuvem é o ponto de entrada na nuvem para usuários e administradores.

Diversos serviços computacionais cresceram ao passo que as TIC evoluíram. Ancorada nesse crescimento, a *Web* expandiu-se a partir de conceitos incorporados em sua segunda geração, passando a oferecer um conjunto de serviços que permitiram a comunicação e interação entre sistemas. Essas características são consequências dos chamados *Web services*, que utilizam protocolos e regras para padronização e interoperabilidade entre agentes de software, além de beneficiar-se das infraestruturas complexas de hardware e as abstrações proporcionadas pelos serviços de nuvem.

As aplicações de ecossistemas da *Web* evoluem constantemente, proporcionando um crescimento vertiginoso no volume de informações. Desta forma, fica visível a necessidade de ação de sistemas interoperáveis e a subjacente demanda de automação dos aspectos relacionados aos *Web services*, para que a ligação e composição dos dados sejam feitos de forma cada vez mais autônoma. Neste sentido, Mendes et al. (2011) defendem que "o uso de tecnologias da *Web* semântica, como *Web Services* semânticos, agentes e ontologias, podem auxiliar na construção de ferramentas que permitem essa automação". Em resumo, a *Web* semântica consiste na aplicação de uma camada adicional na *Web* para promover o trabalho colaborativo entre humanos e máquinas por meio da ligação e interpretação automática de informações, por meio de agentes de softwares, devidamente anotadas para essa finalidade.

Mendes et al. (2011) afirmam que *Web* Semântica "é apropriada para o desenvolvimento de aplicações complexas e distribuídas em ambientes que incorporem inúmeros componentes com diferentes conhecimentos e interesses". O autor indica uma inserção dos conceitos que definem as plataformas de *e-Science*, pois estas possibilitam trabalho em conjunto e interdisciplinar, reunindo dados de diversos locais. Os recursos da *Web* Semântica podem auxiliar na formação de um conjunto informacional interligado automaticamente, criando uma relação semântica entre fontes distintas, promovendo a interdisciplinaridade dos trabalhos científicos. Com vistas nessa

[13] Alguns trabalhos utilizam também o termo e-Infraestrutura para designar infraestruturas de *e-Science*. Este trabalho adota como padrão o termo ciberinfraestrutura.

[14] Computação em nuvem é definida por Chaganti (2010) como "o uso de recursos de computação escaláveis fornecidos como um serviço de fora do seu ambiente no esquema de pagamento de acordo com o uso".

possibilidade, Mendes et al. (2011) propuseram uma plataforma denominada SASAgent, cujo principal objetivo é criar uma arquitetura baseada em agentes inteligentes para pesquisar, recuperar e compor modelos de simulação, gerados no contexto de projetos de pesquisa relacionados ao domínio científico.

A arquitetura SASAgent é descrita como uma camada múltipla, composta por três módulos principais, onde a ontologia CelO atende aos requisitos colocados por projetos de *e-Science*, representados principalmente pela base de conhecimento semântico (Mendes et al. 2011). A arquitetura propõe ainda ferramentas para: *(i)* coletar informações de modelos científicos e, com base nesses modelos, *(ii)* fornecer, consumir e compor novos modelos, conforme a disponibilização da informação semântica, e *(iii)* simulação científica, para compartilhar e reutilizar o conhecimento que são implícitos ou explícitos nesses artefatos.

A arquitetura SASAgent é baseada em três componentes básicos: *(i)* agentes com seus respectivos papéis; *(ii)* a base de conhecimento que, por meio do uso de ontologias, possibilita inferências lógicas e pode descobrir informações implícitas; e *(iii)* artefatos científicos (principalmente modelos CellML e *Web services* semânticos).

Outro projeto relevante que utiliza conceitos da *Web* Semântica consiste em um *framework* baseado em nuvem para suporte à gestão do conhecimento científico entre comunidades biomédicas: COWB (*COllaborative Workspaces in Biomedicine*). Segundo Dessi et al. (2015), "O COWB é baseado em um modelo centralizado de ontologia multicamada que aproveita tanto o conhecimento semântico capturado a partir dessas ontologias, quanto o conhecimento funcional sobre recursos capazes de ampliar o domínio, com base em relações semânticas". O projeto utiliza ontologias para modelagem e gestão do conhecimento de forma colaborativa dentro das comunidades biomédicas. Essa abordagem descobre, importa e publica o conhecimento semântico extraído.

Diante das plataformas apresentadas, observa-se um aparato tecnológico composto por soluções de hardware (*grids* e *clusters*) e software para subsidiar as necessidades inerentes a ciência contemporânea ancorada em grande volumetria de dados. Neste aspecto aponta-se também elementos essenciais para criar inteligência nos dados de pesquisa por meio de conceitos da inteligência artificial, como a ligação semântica de dados para criação e inferência de novos conhecimentos.

As plataformas selecionadas foram classificadas em três grupos, conforme suas características e finalidade (Quadro 1, em apêndice): *(i)* e-Infraestrutura, *(ii) middleware* para serviços de computação em nuvem para gestão de dados científicos e *(iii) frameworks* para GCC baseado em ontologias. A categorização das plataformas teve como objetivo analisar conceitualmente características distintas, para uma posterior análise holística e a sua pertinência para auxiliar na solução de problemas científicos enfrentados pelo Programa LBA/INPA.

4.2 Conjunto de dimensões pertinentes às necessidades do Programa LBA/INPA

Considerando as associações e ligações identificadas por meio do conjunto de análises aplicadas ao *corpus* e inferências corroboradas por meio da visita *in loco*, foram identificas nove dimensões conceituais necessárias para uma efetiva GCC e GDC do Programa LBA/INPA: *(i)* Compartilhamento de Dados, *(ii)* Conectividade e *(iii)* Segurança, essas dimensões foram identificadas na análise de similitude com conexão entre os termos Compartilhamento, Conectividade e Estabilidade; *(iv)* Governança e Gestão de Dados, *(v)* Armazenamento e Replicação de Dados, *(vi)* Curadoria Digital, estas dimensões foram identificadas com a conexão entre os termos Governança, Armazenamento e Preservação; *(vii)* Relação Semântica, identificada com a conexão entre os termos Relação, Semântica e Ontologia; *(viii)* Colaboração Científica, *(ix)* workflow científico e *(x)* Interdisciplinaridade, identificadas com a conexão entre Pesquisa e Colaboração. No Quadro 2 (em apêndice) são apresentadas as descrições conceituais para as dimensões propostas. Cada dimensão conceitual está relacionada a um ou mais eixos temáticos, conforme apontou as análises.

4.3 Análise das plataformas de *e-Science*

A análise das plataformas de *e-Science* levou, inicialmente, em consideração seu compromisso declarado com os padrões da OGF para ambientes de computação de alto de-

sempenho. Sob esse enfoque, em primeira análise, o INPA faz parte apenas da rede SINAPAD. Assim, as demais teriam contribuições limitadas quanto à dimensão conectividade. É sabido que mesmo com os investimentos feitos para criação de uma rede de alta velocidade por meio do Programa Amazônia Conectada[15], à exceção da cidade de Manaus, o restante da área de atuação do INPA não é coberto por rede de alta velocidade para tráfego de dados em grande escala. Diante disso, mesmo essa rede possuindo *nodes* no Brasil, não é viável para uso dos projetos de pesquisa no contexto do Programa LBA/INPA. Constataram-se investimentos por parte da Rede Nacional de Pesquisas (RNP) no intuito de solucionar os problemas relacionados à conexão com a Internet nessa região, o que pode viabilizar a utilização outras plataformas no futuro. O Instituto também faz parte de um projeto piloto de instalação de nuvens científicas, no qual já está instalada uma solução com capacidade de armazenamento de quatro *petabytes*, ainda pequena diante da expectativa de volume de dados do Programa.

Ao apontar as limitações referentes a banda de Internet, ressalta-se que, conforme relato da equipe de TIC do Programa LBA/INPA, o projeto recebeu investimentos relevantes para estruturação do parque tecnológico: *(i)* aquisição de servidores de alto desempenho, *(ii)* instalação de uma solução de nuvem científica privada, *(iii)* aquisição, com transferência de tecnologia, de um sistema de governança e gestão de dados e *(iv)* implementação de sistema para gerenciamento de repositório de dados científicos na *Web*. Por outro lado, a equipe ressalta que mesmo havendo uma considerável infraestrutura no Instituto, os investimentos não foram acompanhados de uma ampla discussão com a comunidade científica sobre uma política de governança e gestão dos dados produzidos pelos diversos projetos de pesquisa. Desta forma, para que os projetos tenham maior êxito, são necessários investimentos na evolução da cultura organizacional no sentido da popularização da utilização de recursos de *e-Science* como suporte às pesquisas do INPA. A equipe mostra ainda preocupação com a falta de planejamento nas políticas públicas do país como um todo, pois, embora já haja discussões sobre essas políticas (Costa 2017), elas ainda são insipientes.

De acordo com relatos, a adoção de uma solução de nuvem pública seria economicamente mais viável do que a atual solução de nuvem privada[16], mas isso não é atualmente possível em função do problema de conectividade da região. Foi apontada como uma solução eficiente o modelo híbrido, com armazenamento de dados na nuvem privada e posteriormente "replicado" em outros sistemas e infraestrutura, para garantir a perenidade e disponibilidade *full time* dos dados.

A governança, gestão e políticas de dados envolvem uma relação direta entre equipe técnica e pesquisadores, sendo necessária ampla discussão entre todos os envolvidos no Programa. Dentre os pontos de debate estão a publicação de dados de pesquisa em formato aberto, utilizando padrões como *Resource Description Framework* (RDF)[17]. Os relatos apontam que para alcançar esse resultado, além das discussões envolvendo toda a comunidade científica concernente e equipe técnica, é necessário o investimento em recursos humanos habilitados para trabalhar com *Big Data*[18] em *e-Science*.

[15] Projeto que prevê a interligação com fibra ótica do Estado por meio da criação de redes de alta velocidade instaladas nos leitos dos principais rios da região: Rio Negro, Solimões, Madeira, Purus e Juruá.

[16] O serviço de computação em nuvem, conforme definem (Katzan Jr, 2010; Ryan & Loeffler, 2010), pode ainda ser classificado quanto aos níveis de acesso, de acordo com as categorias: *(i)* Nuvem Pública: a infraestrutura da nuvem é disponibilizada para o público em geral; *(ii)* Nuvem Privada: o gerenciamento e operação da nuvem são realizados por uma organização e o acesso às informações pode ser restrito por políticas de segurança; *(iii)* Nuvem Comunitária: a infraestrutura da nuvem é administrada por um conjunto de organizações e cujo gerenciamento pode estar sujeito a regras estabelecidas pela comunidade proprietária; *(iv)* Nuvem Híbrida: corresponde a um grupo de nuvens, embora estas nuvens mantenham sua identidade diferenciada entre o grupo, podem ser do tipo privada, pública ou comunitária. As nuvens pertencentes a esta categoria podem estar associadas entre si por protocolos ou padrões técnicos.

[17] RDF é um modelo padrão para intercâmbio de dados na Web (W3C, 2014).

[18] *Big data* reúne uma grande coleção de dados estruturados, semiestruturados e não estruturados, a variedade de dados é uma das características que o conceituam, juntamente com o volume e velocidade. Sha e Carotti-Sha (2016) definem o conceito

O armazenamento dos dados produzidos pelo Programa constitui uma tarefa básica, porém, diante da quantidade de áreas de pesquisa e consequente diversidade de estruturas e formatos de dados empregados, essa atividade torna-se não trivial. A possibilidade de utilização de relações semânticas entre dados, ventilada anteriormente e corroborada por trabalhos como o de Futrelle et al. (2011), abre perspectivas interessantes nesse sentido, mas configura-se como um desafio considerável.

Conforme relato dos pesquisadores entrevistados, não há padrões institucionais para a coleta e armazenamento de dados, sendo definidos pelo próprio pesquisador ou grupo responsável pelo experimento. Por exemplo, o grupo de pesquisa de Micrometeorologia, além de participar do LBA/INPA, tem projetos em outros programas de pesquisa. Em seus projetos, a coleta de dados é feita em vários pontos da floresta amazônica, tanto de forma manual como por telemetria via rádio ou satélite. Alguns desses dados são enviados diretamente a instituições parceiras, que disponibiliza os dados aos demais pesquisadores do Programa LBA/INPA via *File Transfer Protocol* (FTP). Ou seja, não existe uma política de dados unificada para o Programa. Nestes casos não há utilização direta da infraestrutura tecnológica do Programa.

Sobre preservação e política de acesso a dados de pesquisa, Sayão e Sales (2015) destacam que:

> *Os dados e as coleções de dados de pesquisa possuem um tempo de vida maior que os projetos de pesquisa que os criaram. Isso significa que pesquisadores, professores, estudantes e outros profissionais podem continuar a trabalhar sobre esses dados após os projetos e financiamentos tenham sido cessados.*

Jirotka e Olson (2013) afirmam que trabalhos interdisciplinares e multidisciplinares não foram inicialmente previstos em ambientes de *e-Science*, mas que, atualmente, é necessária uma apreciação dos contextos nos quais os dados científicos são produzidos e reutilizados. Além disso, os dados devem ser compartilhados entre grupos de pesquisa e instituições de acordo com requisitos que incluem preceitos éticos e legais. Atualmente, o trabalho colaborativo no contexto do LBA/INPA, se dá de forma segmentada em cada um dos grupos envolvendo pesquisadores do INPA e suas instituições parceiras. A interação é feita por *e-mail*, não sendo utilizadas ferramentas específicas, o que dificulta o desenvolvimento de trabalhos em coautoria entre grupos de instituições distintas. Quando ocorre o compartilhamento de dados, o nome dos colaboradores é referenciado na forma de agradecimentos na publicação dos resultados.

Sobre isso, Jirotka e Olson (2013) destacam a importância da incorporação de ferramentas de Trabalho Colaborativo Apoiado por Computadores (*Computer Supported Cooperative Work* – CSCW) em plataformas de *e-Science*, viabilizando a colaboração em grande escala, geograficamente distribuída e copresente. O trabalho realizado de forma colaborativa com suporte da CSCW ainda pode auxiliar em projetos com abordagens interdisciplinares, sendo esse uma ação de grande relevância no contexto do Programa.

A realização de trabalho colaborativo e integrado entre diversos grupos de pesquisa requer padronização de procedimentos via *workflow* científico. Para Mattoso et al. (2008):

> *Processos experimentais isolados, interligados apenas na concepção do cientista que conduz a análise, não são atualmente suficientes para tratar a complexidade imposta pelos problemas a serem analisados. O problema se agrava quando o experimento científico ocorre de modo distribuído e em larga escala. Faz-se necessário um sistema que gerencie a composição de processos e dados num fluxo coerente, descrito através de um workflow científico, e que registre as etapas realizadas com escolhas de parâmetros de execuções bem sucedidas do experimento, independente do local de execução.*

O grupo de pesquisa do INPA relacionado a Hidrologia relata que sua coleta de dados se dá de forma ininterrupta a cada 30 minutos por meio de equipamentos que armazenam dados primários nos próprios locais de coleta. Esses dados são transferidos manualmente para os servidores do Programa LBA/INPA para compartilhamento com os pesquisadores e alunos de pós-graduação do INPA e de outras instituições. Neste caso, as pesquisas são realizadas de forma colaborativa (Vanz & Stumpf, 2010, p. 42-55), e seus resultados são publicados em coautoria. Mas, conforme relato dos entrevistados pertencentes ao quadro técnico,

como sendo o campo que estuda gerenciamento e processamento de qualquer conjunto de dados muito grande para interpretação direta e individual.

há uma dificuldade operacional, de recursos materiais e, principalmente, de recursos humanos para que os dados coletados possam ser disponibilizados para a comunidade científica de forma ágil e com melhor qualidade.

Para o grupo de Informática para Biodiversidade, a adoção de um modelo para curadoria digital baseada em ontologias e interoperabilidade semântica, proporcionaria um grande avanço quanto a política de dados. Para o grupo, a possibilidade de fazer uma associação dos resultados da pesquisa aos dados primários e ao seu contexto de coleta seria essencial para que os projetos do Instituto tivessem maior alcance. Atualmente, o foco do grupo está no desenvolvimento de aplicações baseadas em ontologias que proporcionem a integração entre base de dados distintas. Este trabalho está avançando, porém, é dificultado pelo grande volume e heterogeneidade de dados envolvidos, sendo a relação semântica entre essas bases de dados essencial para seu avanço.

Quanto à dimensão Curadoria Digital, todos os pesquisadores informaram que existem iniciativas de preservação dos dados, até mesmo devido à natureza dos projetos, porém não há política de dados específica quanto a perenidade, mídias empregadas ou local de armazenamento.

A Relação Semântica destacada neste trabalho faz parte de um estudo em andamento que inclui a criação de relações semânticas entre os trabalhos científicos (como artigos, teses, dissertações) desenvolvidos no Programa por meio de uma biblioteca digital semântica. Iniciativas em andamento, como a construção de uma ontologia de biodiversidade (OntoBIO) no INPA, vão ao encontro de projetos que possibilitam uma correlação semântica entre conhecimentos científicos e a integração entre dados e resultados de pesquisa.

Nesse aspecto, Elsayed e Brezany (2013) propõem mecanismos associados para gerenciar relacionamentos semânticos por meio de recursos RDF entre fontes de dados científicas (dados primários) e seus achados correspondentes (dados derivados), que resultam de um conjunto de atividades, definindo métodos concretos de pré-processamento e análise que foram aplicados em um conjunto de dados. Os autores também apontam mecanismos para acompanhar experiências científicas e formas de vinculá-las com informações do usuário.

No Quadro 3 (em apêndice) é apresentada a correlação entre as plataformas de *e-Science* e as dimensões identificadas, classificadas em 4 eixos temáticos: *(i)* trabalhos colaborativos em grupos de pesquisa; *(ii)* governança, gestão e política de dados; *(iii)* infraestrutura de TIC; *(iv)* desenvolvimento da pesquisa e publicação dos resultados. Esses eixos são bastante abrangentes e englobam os preceitos da ciência moderna baseada no quarto paradigma, a Ciência de Dados, envolvendo também a inferência e lógica formal com a finalidade de criar relações e ligações semânticas entre dados e resultados e entre pesquisas de áreas distintas.

A relação entre as plataformas estudadas as dimensões elencadas como essenciais para o Programa LBA/INPA, aponta que as soluções que podem ser utilizadas na gestão de seus dados. No concernente à e-Infraestrutura, considerando principalmente as características do Programa, o modelo recomendado é uma solução híbrida, projetada localmente e seguindo o modelo proposto pela OGSA, de forma que garanta a interoperabilidade, com implementações específicas sob os preceitos da *Web* Semântica, e replicações em projetos como o DataONE, para garantir a preservação digital e especificidades do contexto regional. Parte das necessidades apontadas leva em consideração a utilização de soluções distribuídas sob a licença *General Public License* (GPL), que permitem a modificação de softwares e implementação de extensões, como por exemplo, conectividade para comunicação e transferência de dados com sítios de coletas, por meio de dispositivos e protocolos de *Internet of Things* (IoT) (15)[19].

5. Conclusão

O objetivo deste trabalho foi analisar características em plataformas de *e-Science* para atender um conjunto de dimensões conceituais alusivas à gestão de dados e conhecimento científicos em um contexto específico, neste caso para a região Amazônica, especificamen-

[19] A *International Telecommunication Union* (ITU), define IoT como "uma infraestrutura global para a Sociedade da Informação, permitindo serviços avançados (físicos e virtuais) interconectados por meio de tecnologias de informação e comunicação interoperáveis, tanto existentes, como em desenvolvimento" (ITU, 2012).

te para a parte do Programa LBA coordenado pelo INPA. Considerou-se que, em um ambiente computacional, onde esse conjunto de características esteja habilitado, tenha-se um cenário ideal para a gestão de dados e do conhecimento científicos, desde a sua concepção, *data acquisition*, passando pela qualidade do dado coletado, *data quality*, até a publicação dos resultados da pesquisa, suportado por recursos computacionais.

Devido ao estudo ter sido realizado no contexto específico do Programa LBA/INPA, a pesquisa *in loco* permitiu uma análise e avaliação detalhada sobre a atividade científica dos diversos grupos de pesquisa e da infraestrutura tecnológica que oferece suporte às estas atividades e a percepção dos pesquisadores sobre o aparato tecnológico para subsidiar o trabalho científico. Outros pontos também foram observados, como o trabalho colaborativo e a relação entre os pesquisadores e outros profissionais do Instituto. Por meio das entrevistas semiestruturadas, foi possível fazer a análise e mapeamento sobre as necessidades de infraestrutura de *e-Science* do programa, sendo que algumas delas já foram implementadas e outras em fase de implantação.

Quanto às plataformas para concepção de *e-Science* destacadas para este trabalho, foram classificadas em três categorias: *(i)* e-Infraestrutura, *(ii) middleware* para serviços de nuvem para gestão de dados científicos e *(iii) framework* para gestão do conhecimento científico baseado em ontologias, cujos principais componentes foram elencados com base em uma análise documental e ainda considerando que alguns destes componentes são padronizados em soluções de *grid*, para possibilitar a interoperabilidade entre ambientes de HPC, com base no que propõe a OGSA.

É importante ressaltar que esta pesquisa possui uma abordagem conceitual, podendo ser continuada em trabalhos futuros com a criação de um modelo de gestão de dados científicos com a propositura de um *framework* que englobe os elementos indicados neste cenário.

Referências

Almeida, M. C. (2009). Método complexo e desafios da pesquisa. In: Almeida, M. C. & Carvalho E. A. (Eds). *Cultura e pensamento complexo*, Natal: EDUFRN, pp. 97–111.

Alonso, L., Sallantin, J., Ferneda, E., & Luzeaux, D. (2011). Scientific Knowledge Management Anchored on Socioenvironmental Systems". *TripleC*, 9(2), 610–623.

Andronico, G. et al. (2011). e-Infrastructures for e-Science: A Global View. *Journal of Grid Computing*, 9(2), 155–184.

Appel, A. L. (2014). *A e-Science e as atuais práticas de pesquisa científica*. Dissertação (Mestrado em Ciência da Informação), Rio de Janeiro: IBICT-UFRJ.

Chaganti, P. (2010). Serviços em nuvem para sua infraestrutura virtual, Parte 1: Infrastructure-as-a-Service (IaaS) e Eucalyptus. *IBM DeveloperWorks*. https://www.ibm.com/developerworks/br/library/os-cloud-virtual1/index.html.

Costa, M. M. (2017). *Diretrizes para uma política de gestão de dados científicos no Brasil*. Tese (Doutorado em Ciência da Informação), Universidade de Brasília.

Crichton, D. J., Mattmann, C. A., Hughes, J. S., Kelly, S. C., & Hart, A. F. (2011). A Multidisciplinary, Model-Driven, Distributed Science Data System Architecture". In: Yang, X., Wang, L., & Jie, W. (Eds). *Guide to e-Science: Next Generation Scientific Research and Discovery*, Springer, pp. 117–143.

Dessi, N., Milia, G., Pascarielo, E., & Pes, B. (2016). COWB: A cloud-based framework supporting collaborative knowledge management within biomedical communities. *Future Generation Computer Systems*, 54, 399–408.

Ferreira, M. A. (2010). *Estudo sobre a utilização de ferramentas de colaboração em redes de pesquisa científica*. Dissertação (Mestrado em Gestão do Conhecimento e da Tecnologia da Informação). Universidade Católica de Brasília.

Futrelle, J., Gaynor, J., Plutchak, J., & Bajcsy, P. (2008). Knowledge Spaces and Scientific Data. *4th IEEE International Conference on eScience*.

Gomes, A. T. A., Bastos, B. F., Medeiros, V., & Moreira, V. M. (2015). Experiences of the Brazilian national high-performance computing network on the rapid prototyping of science gateways. *Concurrency and Computation: Practice and Experience*, 27(2), 271-289.

ITU-T Study Group (2012). *New ITU standards define the Internet of Things and provide the blueprints for its development*. International Telecommunication Union.

Jankowski, N. W. (2007). Exploring e-science: an introduction. *Journal of Computer-Mediated Communication*, 12(2), 549–562.

Jirotka, M. & Olson, G. (2013). Supporting Scientific Collaboration: Methods, Tools and Concepts. *Computer Supported Cooperative Work*. 22, 667–715.

Kang, H., Yoon, K., Kim, S., Kim, Y., & Kim, C. (2011). An e-Science problem solving environment for scientific numerical

study. *13th International Conference on Advanced Communication Technology* (ICACT).

Kaplan, D. M. (2004). *Readings in the Philosophy of Technology*. Lanham: Rowman & Littlefield Publishers, Inc.

Katzan Jr., H. (2010). On an ontological view of cloud computing. *Journal of Service Science*, 3(1),1-6.

Knyazeva, H. (2017). Complexity Studies: Interdisciplinarity in Action. In: Pietsch, W., Wernecke, J., & M. Ott (Eds). *Berechenbarkeit der Welt?* Springer.

Mattoso, M., Werner, C., Travassos, G. H., Braganholo, V., & Murta, L. (2008). Gerenciando Experimentos Científicos em Larga Escala. *XVIII Congresso da SBC - Seminário Integrado de Software e Hardware*.

Medeiros, J. S. & Caregnato, S. E. (2012). Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. *LIINC em Revista*, 8(2), 311–322.

Mendes, L. F., Silva, L., Matos, E., Braga, R., & Campos, F. (2011). SASAgent: An agent based architecture for search, retrieval and composition of scientific models. *Computers in Biology and Medicine*, 41(7), 449–462.

Meyer, L. A. V. C. (2006). *Uma visão geral dos sistemas distribuídos de cluster e grid e suas ferramentas para o processamento paralelo de dados*. IBGE. https://ww2.ibge.gov.br/confest_e_confege/pesquisa_trabalhos/CD/palestras/368-1.pdf

Morin, E. (1999). *O Método 3: O conhecimento do conhecimento*. Porto Alegre: Sulina.

Motloch, J. L. (2016). Disciplining interdisciplinarity: integration and implementation sciences for researching complex real-world problems (and the deeper science challenge to co-evolve with complexity). *Journal of Environmental Studies and Sciences*, 6(2), 445–447.

Muccioli, C., Campos, M., Goldchmit, M., Dantas, P. E. C., Bechara, S. J. & Costa, V. P. (2007). A Produção Científica no Brasil. *Arquivos Brasileiros de Oftalmologia*, 70(4). https://doi.org/10.1590/S0004-27492007000400001

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. Santa Fé: The Santa Fé Institute. https://doi.org/10.48550/arXiv.cond-mat/0104209

Peach, K. J. (2004). *The Impact of eScience*. https://cds.cern.ch/record/865540/files/p30.pdf.

Price, D. S. (1976). *A ciência desde a Babilônia*. São Paulo: EDUSP.

Ribes, D. & Lee, C. P. (2010). Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. *Computer Supported Cooperative Work*, 19(3-4), 231–244.

Roberts, G. (2019). *NSI Connection Service v2.0 to v2.1 Delta*. Open Grid Forum, GWD-I.238. https://www.ogf.org/documents/GFD.238.pdf.

Roberts, G., MacAuley, J., Kudoh, T., & Guok, C. (2019). *NSI Connection Service v2.1*, Open Grid Forum, GWD-R-P.237, https://www.ogf.org/documents/GFD.237.pdf.

Ryan, W. M., & Loeffler, C. M. (2010). Insights into cloud computing. Intellectual. *Property & Technology Law Journal*, 22(11), 22-28.

Sampaio, J. O. (2007). *METHEXIS: uma abordagem de apoio à Gestão do Conhecimento para ambientes de e-Science*. Tese (Doutorado em Ciências). UFRJ, Rio de Janeiro.

Saracevic, T. (2009). Information science. In: Bates, M. J. & Maack M. N. (Eds) *Encyclopedia of Library and Information Sciences*. 3rd ed., Abingdon: Taylor & Francis, pp. 2570–2585.

Sayão, L. F. & Sales, L. F. (2012). Curadoria digital: um novo patamar para preservação de dados digitais de pesquisa. *Informação & Sociedade: Estudos*, 22(3).

Segura, J. V. (2009). Computational Epistemology and e-Science: a new way of thinking. *Minds and Machines*, 19, 557–567.

Sha, X. W. & Carotti-Sha, G. (2016). Big Data. *AI & Society*. https://doi.org/10.1007/s00146-016-0662-7

Vanz, S. A. S. & Stumpf, I. R. C. (2010). Colaboração científica: revisão teórico-conceitual. *Perspectivas em Ciência da Informação*, 15(2), 42–55.

W3C (2014). Resource Description Framework. https://www.w3.org/RDF/

Zago, A. M. (2008). *Perfil da Produção Científica no Brasil*. http://www.fapesp.br/eventos/2011/06/Marco_Antonio.pdf.

Apêndice

Quadro 1. Plataformas de *e-Science*

Categoria	Plataforma	Descrição	Financiamento
e-Infraestrutura	EELA - *E-Infrastructure Shared Between Europe And Latin America*	Rede de pesquisa de alta velocidade com infraestrutura *Grid* subjacente que fornece energia e armazenamento de computação distribuída em domínios geográficos e administrativos. Esta rede integrada e ambiente de processamento / armazenamento (*e-infraestrutura*) fornece uma plataforma poderosa para novos métodos de pesquisa colaborativa global (*e-Science*).	Comunidade Europeia e instituições parceiras

Categoria	Plataforma	Descrição	Financiamento
	CCES-CEPID - *Center For Computational Engineering & Sciences*	O Centro de Computação e Ciências Computacionais (CCES-CEPID) é um centro multidisciplinar da Universidade de Campinas (Unicamp), cuja missão é realizar pesquisa científica de alcance mundial em simulações computacionais de alto desempenho, modelagem computacional e computação intensiva em dados para resolver problemas de fronteira em ciências moleculares e engenharia.	FAPESP e Unicamp
	Rede Galileu	O projeto Galileu surgiu de uma iniciativa conjunta do CENPES / PETROBRAS (Centro de Pesquisas e Desenvolvimento Leopoldo Américo Miguez de Mello) e de centros de excelência em pesquisa: COPPE/UFRJ, USP, PUC/RJ, UFAL e ITA - com o objetivo de desenvolver soluções em Visualização Científica, Modelagem Computacional e Computação de Alto Desempenho.	Petrobrás e instituições participantes
	SINAPAD – Sistema Nacional de Processamento de Alto Desempenho	Uma rede de centros de computação de alto desempenho, geograficamente distribuídos, instituída pelo Ministério da Ciência e Tecnologia e Inovação e Comunicações (MCTIC). São nove unidades, denominadas "Centros Nacionais de Processamento de Alto Desempenho" (CENAPAD), operadas respectivamente pela UFRGS, UFMG, UFC, UNICAMP, UFRJ, UFPE, INPE, INPA e LNCC. Este último coordena o sistema por delegação do MCTIC.	Governo Federal Brasileiro e instituições participantes
	DataONE	É a fundamentação de novas ciências ambientais inovadoras por de uma estrutura distribuída, sustentável que atenda às necessidades da ciência e da sociedade para acesso aberto, persistente, robusto e seguro à observação da Terra bem descrita e dados facilmente acessíveis.	National Science Foundation (NFS)
Middleware para Serviços de Nuvem	Eucalyptus - *Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems*	É uma estrutura de software *open source* para computação em nuvem que implementa o que é comumente referido como Infraestrutura como um Serviço (IaaS); sistemas que oferecem aos usuários a capacidade de executar e controlar instâncias de máquinas virtuais inteiras implantadas em uma variedade de recursos físicos.	Universidade da Califórnia, (Santa Barbara) - distribuído sob a licença GPL.
	Nimbus/ OpenStack/ Chamaleon	O Nimbus é um conjunto de ferramentas de código aberto focado no fornecimento de capacidades de Infraestrutura como serviço (IaaS) para a comunidade científica. Para alcançar isso, estabeleceu três objetivos: *(i)* permitir que os provedores de recursos criem nuvens IaaS privadas ou comunitárias; *(ii)* permitir que usuários usem nuvens IaaS; *(iii)* permitir que os desenvolvedores estendam, experimentem e personalizem IaaS. A partir de 2010, a infraestrutura Ninbus passou a integrar os projetos OpenStack e Chamaleon, ambos com foco em nuvens científicas.	Distribuído sob a licença GPL.
	UNICORE - *Uniform Interface to Computing Reources*	Plataforma de middleware que oferece um sistema pronto para executar incluindo software cliente e servidor. A UNICORE disponibiliza recursos de computação e dados distribuídos de forma transparente e segura nas intranets e na internet.	Distribuído sob a licença GPL.

Categoria	Plataforma	Descrição	Financiamento
Framework baseado em ontologias	COWB - *COllaborative Workspaces in Biomedicine*	É um *framework* que apoia a gestão colaborativa do conhecimento no contexto das comunidades biomédicas. O COWB é fundamentado em um modelo centrado em ontologia de várias camadas. Ele aproveita tanto o conhecimento semântico capturado pelas ontologias quanto o conhecimento funcional sobre recursos que ampliam o conhecimento do domínio e apoiam seu gerenciamento. Espaços de trabalho públicos e privados fornecem uma representação acessível do conhecimento coletivo que é criado incrementalmente e permite que o conhecimento atravesse os limites de informações locais fechadas	–
	SASAgent	A arquitetura SASAgent é descrita como uma camada múltipla, composta por três módulos principais, onde a ontologia CelO atende aos requisitos colocados por projetos de *e-Science*, representados principalmente pela base de conhecimento semântico.	–

Quadro 2. Descrição dos conceitos relativos às dimensões propostas

Dimensão	Descrição	Fonte bibliográfica
Compartilhamento de Dados	Liberação de dados primários de pesquisas para o uso de outros cientistas.	Borgman (2012)
Armazenamento e Replicação de Dados	A replicação é útil na melhoria da disponibilidade de dados. O caso mais relevante é a replicação do banco de dados inteiro em cada *site* no sistema distribuído, criando assim um banco de dados distribuído totalmente replicado. Isso garante também maior segurança.	Elmasri e Navathe (2011) Appel (2014)
Governança e Gestão de Dados	Governança de dados é uma estrutura que orienta e estabelece estratégias, políticas e objetivos com a finalidade de gerenciar os dados, como se fossem qualquer outro recurso de uma organização. Gestão de dados científicos é um termo geral capaz de cobrir a organização, a estrutura, o armazenamento e o cuidado da informação gerada durante o processo de pesquisa.	Loftis (2014) Universidade de Oxford (2013)
Relação Semântica	As relações semânticas são um importante componente para organização do conhecimento, sendo a unidade básica entre dois conceitos. A relação semântica, neste aspecto, tem por finalidade criar uma nova instância de um conhecimento (ou dado), possibilitando a expansão informacional sobre determinado *corpus*, ou ainda sua relação com dados primários. No contexto da ciência da informação, Khoo e Na (2006) consideram relações semânticas como relações significativas entre dois ou mais conceitos, entidades ou conjunto de entidades, podendo ainda fazer referência a relações entre conceitos mentais, entre elementos lexicais e entre parágrafos.	Hjørland (2003) Khoo e Na (2006)
Curadoria Digital	Conjunto das todas as atividades existentes no gerenciamento de dados, desde o planejamento da sua criação, passando pela digitalização (transformação digital) ou criação, procurando assegurar a disponibilidade e adequação para a recuperação e reuso futuro destes dados.	Abbot (2008)
Conectividade	O fator conectividade, neste contexto, refere-se a interconexão de diversos equipamentos tanto de coleta em sítios como de armazenamento para uso e reuso de dados primários como dados derivados. Essa conectividade significa uma maior quantidade de dados, recolhidos a partir de mais lugares, com muitas maneiras	Fabricio et al. (2016)

Dimensão	Descrição	Fonte bibliográfica
	de aumentar a eficiência e melhorar a proteção e segurança.	
Workflow Científico	Redes de processos tipicamente utilizadas como dutos de análises de dados ou ainda para comparar dados observados ou previstos, e que podem incluir uma vasta gama de componentes, como para consultar bancos de dados, para transformar ou minerar dados, para executar simulações em computadores de alto desempenho, etc.	Ludäscher et al. (2006)
Segurança (SLA)	*Service Level Agreement* ou Contrato de Nível de Serviço. Contrato assinado pelo fornecedor e cliente de serviços de nuvem pública, com a finalidade de garantir, dentre outros requisitos, confidencialidade, disponibilidade, Qualidade do Serviço (QoS).	Dastjerdi et al. (2011).
Colaboração Científica	Há colaboração científica quando dois ou mais cientistas trabalham juntos em um projeto de pesquisa e compartilham recursos intelectuais, econômicos ou físicos.	Vanz e Stumpf (2010)
Interdisciplinaridade	A interdisciplinaridade está imbricada com as demais dimensões, uma vez que se deve considerar, além da comunidade científica, diversos outros atores da sociedade.	Minayo (2007, 2010), Philippi Jr e Silva Neto (2010), Alonso et al. (2011)

Quadro 3. Correlação entre as plataformas e e-Science e as dimensões propostas

Eixo temático	Dimensão	Plataforma de *e-Science*		
		e-Infraestrutura [EELA, Rede Galileu, CCES-CEPID, SINAPAD, DataONE]	**Middleware** [Eucalyptus, Nimbus/OpenStack, UNICORE]	***Framework* para GCC** [COWB, SASAgent]
Infraestrutura de TIC	Armazenamento e replicação de dados	×	×	
	Curadoria Digital		×	
	Conectividade		×	
Governança, Gestão de Dados	Governança e gestão de dados	×	×	
	Compartilhamento de dados	×	×	
	Segurança (SLA)	×	×	
Desenvolvimento da Pesquisa	Relação Semântica			×
	Workflow Científico		×	
Trabalhos Colaborativos	Colaboração científica	×	×	
	Interdisciplinaridade			×

Sobre os autores

Ronaldo Ferreira da Silva
Graduado em Tecnologia em Processamento de Dados pela Universidade Estadual de Goiás (2003), com Especialização em Marketing pela Universidade Cândido Mendes (2006) e Mestrado em Gestão do Conhecimento e da Tecnologia da Informação pela Universidade Católica de Brasília (2018). É Professor da Universidade Estadual de Goiás (Câmpus Posse).

Edilson Ferneda
Graduado em Tecnologia de Computação pelo Instituto Tecnológico de Aeronáutica - ITA (1979), Mestre em Sistemas e Computação pela Universidade Federal da Paraíba - UFPB (1988) e Doutor em Ciência da Computação pelo Laboratoire dInformatique, Robotique et de Microélectronique de Montpellier - LIRMM, França (1992). Entre 1986 e 2004, foi professor do Depar-

tamento de Sistemas e Computação da Universidade Federal de Campina Grande - UFCG (antiga UFPB), tendo atuado nos cursos de Tecnologia em Processamento de Dados, Bacharelado em Ciência da Computação, Mestrado em Informática e Doutorado em Engenharia Elétrica. Desde 2001 é professor titular da Universidade Católica de Brasília, onde atua no Curso de Bacharelado em Ciência da Computação e no Mestrado em Governança, Tecnologia e Inovação (antigo Mestrado em Gestão do Conhecimento e Tecnologia da Informação).

Fernando William Cruz

Possui graduação em Processamento de Dados pela UnB (1988), Mestrado em Informática pela UFPB (1992) e Doutorado em Ciências da Informação pela UnB (2008). É professor adjunto da Universidade de Brasília com as seguintes atuações: *(i)* na Graduação em Engenharia de Software - FGA em disciplinas básicas de computação (sistemas operacionais, sistemas distribuídos, estruturas de dados, arquitetura e redes de computadores), *(ii)* na Pós-Graduação em Ciência da Informação - FCI realiza pesquisas voltadas para Organização e Representação da Informação, Estudos de Usuários e Bibliotecas Digitais. Participações recentes (desde 2012) em projetos ligados a sistemas colaborativos e web semântica.

José Laurindo Campos dos Santos

Graduado em Engenharia Modalidade Construção Civil pelo Instituto de Tecnologia do Amazonas (1984), Mestrado em Ciência da Computação pela Universidade Federal da Paraíba (1988) e Doutorado em Ciência da Computação - Universidade de Twente e Instituto Internacional de Pesquisa Aeroespacial e Observação da Terra - ITC, (2003) - Holanda. É Analista C & T Senior do Ministério de da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), e Coordenador substituto de Ações Estratégicas (COAE) do Instituto Nacional de Pesquisas da Amazônia (INPA), Membro do Conselho Técnico Consultivo do Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr) e Membro da Sociedade Brasileira para o Progresso da Ciência (SBPC). Coordena as atividades de Gestão de Dados Científicos do Programa de Grande Escala da Biosfera-Atmosfera na Amazônia (LBA), desenvolvendo pesquisas como colaborador nos Grupos - Zoologia na Amazônia: Diversidade, Biogeografia e Coleções (INPA), Biogeofísica da Região Amazônica e Modelagem Ambiental - BRAMA (Universidade Federal do Oeste do Pará (UFOPA) e Laboratório Mídias Eletrônicas (UFOPA), Gestão de Conhecimento Científico na Universidade Católica de Brasília (UCB), Universidade de Brasília (UnB) e em Web Semântica Aplicada a Ciência da Vida no Instituto de Ciências Matemáticas e de Computação (USP).

Ana Paula Bernardi da Silva

Graduada em Licenciatura Plena em Matemática pela Universidade Federal do Rio Grande (1995), Mestrado em Matemática pela Universidade Federal do Rio Grande do Sul (1998) e Doutorado em Engenharia Elétrica pela Univewrsidade de Brasília (2012). É professora do Programa de Pós Graduação Stricto Sensu em Governança, Tecnologia e Inovação. Atua na área de governança e gestão pública em tempos de transformação digital. Suas temáticas de interesse são: *(i)* governança de TIC pública e corporativa (direcionamento, avaliação e monitoramento de estratégias organizacionais) para otimizar recursos e gerar valor; *(ii)* desenvolvimento de capacidades dinâmicas e capacidades digitais; *(iii)* governança de cidades inteligentes e sustentáveis.

Luiza Beth Nunes Alonso

Possui graduação em Ciências Sociais pela Universidade de São Paulo (1975), com Mestrado e Doutorado em Educação pela Harvard University (1985). Atualmente é professora e pesquisadora da Universidade Católica de Brasília. Tem experiência na área de Sociologia do Conhecimento, com ênfase na interface entre domínio conceitual e campos de aplicabilidade. Realiza pesquisas sobre os temas transformação Digital e Gestão Social de Conhecimentos.