

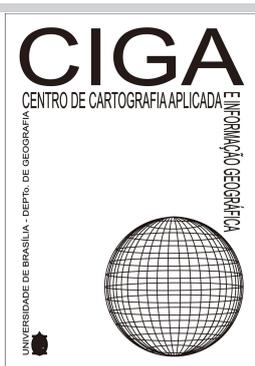
Artigo

A QUEM INTERESSA TER AS PRODUÇÕES ACADÊMICAS DESORGANIZADAS? AS INFRAESTRUTURAS DE DADOS ESPACIAIS E CIENTÍFICOS NAS INSTITUIÇÕES DE ENSINO SUPERIOR DO BRASIL.

José Augusto Sapienza Ramos
Victor Valentim Lassaval Farias

p. 01 - 68

Revista



Revista Eletrônica:
Tempo - Técnica - Território,
V.8, N.2 (2018), 28:47
ISSN: 2177-4366

Como citar este artigo:

Ramos, A.S.R.; Farias, V.V.L.; A QUEM INTERESSA TER AS PRODUÇÕES ACADÊMICAS DESORGANIZADAS? AS INFRAESTRUTURAS DE DADOS ESPACIAIS E CIENTÍFICOS NAS INSTITUIÇÕES DE ENSINO SUPERIOR DO BRASIL. Revista Eletrônica: Tempo - Técnica - Território, v.9, n.2 (2018), p. 28:47
ISSN: 2177-4366.

Disponível em:

<http://periodicos.unb.br/index.php/ciga/issue/view/1905>

Este obra está licenciado com uma Licença Creative Commons Atribuição - Não Comercial 4.0 Internacional.

A QUEM INTERESSA TER AS PRODUÇÕES ACADÊMICAS DESORGANIZADAS? AS INFRAESTRUTURAS DE DADOS ESPACIAIS E CIENTÍFICOS NAS INSTITUIÇÕES DE ENSINO SUPERIOR DO BRASIL.

José Augusto Sapienza Ramos

Coordenador Acadêmico do Núcleo de Geotecnologias (Sistema Labgis) da Universidade do Estado do Rio de Janeiro.

Bacharel Ciência da Computação pela Universidade Federal Fluminense, Brasil.

Mestre em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro, Brasil.

Mestre em Geographic Information Systems & Sciences pela Universidade de Salzburg, Áustria.

Doutorando em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro, Brasil.

sapienza@labgis.uerj.br

Victor Valentim Lassaval Farias

Geógrafo do Núcleo de Geotecnologias (Sistema Labgis) da Universidade do Estado do Rio de Janeiro.

Bacharel em Geografia pela Universidade Federal Fluminense, Brasil.

valentimvictor@labgis.uerj.br

Endereço para correspondência:

Rua São Francisco Xavier, 524, Maracanã, Rio de Janeiro (RJ), Brasil. Sala 2044/F do Pavilhão João Lyra Filho. CEP: 20550-013.

ABSTRACT: The concept about Data Infrastructure handles socio and technical challenges to organize a new scale of production and data sharing in digital era. The majority of higher-level education institutions (if not all) in Brazil maintain data repositories to gather documents such as thesis and papers. Nevertheless, few institutions has similar repositories for other type of academic production: the spatial data. This scenario results in disperse data holdings and archives (with high data loss) that make these databases unfit for further use in academic research, by governments, by businesses or by any other potentially interested sectors. This paper addresses three issues in order to contribute for this debate about academic data infrastructures, namely: (1) present a current debate about Data Infrastructures, Scientific Data Infrastructures (SDI) and Spatial Data Infrastructures (also SDI); (2) describe the current scenario of Academic SDIs in Brazil; and (3) apply a survey to measure the perception of the academic community about its technical and scientific production. The results show a link between the current scenario about Academic SDIs and a negative perception related to spatial data sharing, as well as demonstrate that the community has important values to data infrastructure initiatives such as the information value, its user and

producer rules, and the importance of data sharing. At end, we conclude that both debates about scientific and spatial data infrastructures must be closer.

RESUMO: O conceito sobre Infraestrutura de Dados tange aos desafios sociotécnicos para organizar uma nova escala de produção e compartilhamento de dados na era digital. A maioria das instituições de ensino superior no Brasil, se não todas, mantêm repositórios de dados para reunir documentos tais como teses e artigos. Mesmo assim, poucas instituições possuem repositórios similares para outros tipos de produções acadêmicas como, por exemplo, os dados geoespaciais. Esse cenário resulta em repositórios de dados dispersos – com elevada perda de informação – que tornam essas bases de dados impróprias para futuras utilizações em pesquisas acadêmicas, por governos, por empresas ou por qualquer outro setor em potencial que esteja interessado. Esse trabalho endereça três questões a fim de contribuir com esse debate sobre infraestrutura de dados acadêmicos, a saber: (1) apresentar um debate atual sobre Infraestrutura de Dados, Infraestrutura de Dados Científicos (IDCs) e Infraestrutura de Dados Espaciais (IDEs); (2) descrever o cenário atual das IDEs Acadêmicas no Brasil; e (3) aplicar um questionário para medir a percepção da comunidade acadêmica sobre sua produção técnica e científica. Os resultados apresentam uma relação entre o cenário das IDEs Acadêmicas e a percepção negativa associada à eficiência do compartilhamento de dados geoespaciais, tão bem como demonstra que a comunidade detém importantes valores a iniciativas de infraestrutura de dados tais como o valor da informação, seu papel de usuário e de produtor e a importância do compartilhamento de dados. Por fim, concluímos que ambos os debates sobre IDCs e IDEs precisam ser aproximados.

1. INTRODUÇÃO

Ao longo da história da ciência humana algumas tecnologias deram saltos quantitativos na escala de produção de informações e de circulação dessas informações dentro das sociedades. Podemos citar como exemplos o advento da escrita, da prensa de Gutenberg, do computador eletrônico e da Internet (Ramos, 2016). Junto com o crescimento da escala de produção também cresce o desafio de organizar e disponibilizar esse novo volume de informações de uma forma eficiente, tão como compreender os impactos sociais dessa evolução informacional (Kitchin, 2014). Segundo o mesmo autor, embora a geração e comunicação do conhecimento seja um fenômeno social básico, o conceito de sociedade da informação surge com a ascensão recente das tecnologias da informação e seus impactos globais.

Segundo O'Carroll et al (2013), Demchenko et al (2013) e David (2003), a informação é insumo importante para produção de conhecimento, governança e capital nas sociedades modernas, resultando assim em um grande desprendimento de recursos nas últimas décadas para o desenvolvimento e a implantação de tecnologias da informação. Segundo o levantamento do trabalho de Csorny (2013), apenas nos Estados Unidos da América, a indústria de serviços de informática empregava no ano de 2011 cerca de 1,5 milhão de pessoas, movimentando um mercado bilionário.

Para além da dimensão econômica, entende-se que a informação hoje possui um papel social inexorável – veja uma discussão sobre o conceito da informação em sua dimensão social em Capurro e Hjørland (2015). Não por menos, conceitos como *open data* e *infraestrutura de dados* são debatidos em trabalhos como Kitchin (2014) e O’Carroll et al (2013) a fins de enfrentar o desafio sociotécnico para se realizar o potencial da informação, tornando-a melhor qualificada e disponível para uma parcela maior da sociedade (Ramos e Ferreira, 2015). Como exemplo, temos as iniciativas brasileiras como a Infraestrutura Nacional de Dados Abertos (INDA) e Infraestrutura Nacional de Dados Espaciais (INDE).

O impacto da informação na ciência também é importante. Segundo o trabalho de Tolle et al (2011), após os paradigmas de pesquisa anteriores, a saber: experimental, teórico e computacional, presenciamos um quarto paradigma: *data-intensive scientific discovery* (tradução livre: descobertas científicas com uso intensivo de dados). Nesse quarto paradigma toda a literatura científica está online, todos os dados científicos estão online e eles são interoperáveis uns com os outros. Certamente esse não é o cenário atual da ciência brasileira.

No contexto mais específico deste trabalho, a comunidade acadêmica e de outras instituições de ensino superior composta de alunos, docentes, técnicos, entre outros, produz uma valiosa e extensa base de dados científicos em monografias, teses, artigos, relatórios técnicos, trabalhos de campo e outras atividades – o trabalho de Gray et al (2005) afirma que o volume produzido aproximadamente dobra a cada ano. Como aponta Wilkinson et al (2016) e Ramos e Ferreira (2015), as empresas, o governo, os cidadãos e, é claro, a própria academia são usuários em potencial dessa base de dados científicos produzida dentro do ambiente acadêmico. Entretanto, como argumentam também esses mesmos trabalhos, boa parte dos dados científicos é produzida para atender projetos específicos e se encontram em computadores pessoais, mídias como pen drivers e DVDs, bancos de dados não compartilhados, ou em outras formas de armazenamento não apropriados para boa prospecção e, logo, para melhor reutilização – veja um debate em Machado (2016), Wilkinson et al (2016) e Tenopir (2011). Em outras palavras, uma parte da produção intelectual das instituições de ensino superior se torna indisponível ou mesmo é perdida. Como aponta Ramos e Ferreira (2015), grande parte dessas instituições de ensino mantém infraestruturas formais para trabalhos de final de curso como monografias, teses e dissertações, porém essa organização não é replicada para outros tipos de produções intelectuais.

Esse desperdício de informação resulta em perda de diferentes oportunidades como, por exemplo, insumos para pesquisas científicas, interação entre a academia e outros setores da sociedade por meio de compartilhamento de informações, economicidade de recursos públicos e privados

desprendidos na produção da informação, exercício da responsabilidade social para a democratização de informação e para a promoção da ciência, entre outros.

Alguns trabalhos versam sobre Infraestrutura de Dados Espaciais (IDE) em ambientes de instituições de ensino superior como Machado (2016), Muratori e Samakovlija (2015), Ramos e Ferreira (2015), Smyth (2015), Kitchin e Lauriault (2014), Lecarpentier et al (2013), Oliveira e Ramos (2013), Burton et al (2012), Dessers (2012), Ellul et al (2012), Hill e Trimble (2012) e Grill et al (2010). Entretanto a maioria das questões endereçadas nesses trabalhos são técnicas ou operacionais, ou seja, focadas em caminhos sobre *como* implementar e pouco sobre *porquê* implementar uma IDE Acadêmica. Por outro lado, trabalhos que endereçam Infraestruturas de Dados Científicos (IDC) como Wilkinson et al (2016), Kitchin et al (2015), Demchenko et al (2013), Tenopir et al (2011), entre outros que serão citados adiante, percebemos uma discussão mais frequente desse *porquê*.

É evidente as semelhanças entre uma IDE Acadêmica e uma IDC, a ponto de podermos questionar se em geral é razoável coexistir essas duas iniciativas de infraestruturas em uma mesma instituição de ensino superior. A discussão de infraestruturas de dados está recente e intensa, vide que 78,8% das referências deste trabalho possuem menos de 10 anos e que algumas delas figuram entre periódicos de grande prestígio como Science e Nature. Ormeling (2011) afirma assertivamente: os avanços tecnológicos têm deixado a teoria para trás. A comunidade acadêmica precisa perseguir essa distância.

A contribuição desse trabalho reside em: (a) fazer uma discussão atual dos conceitos de Infraestrutura de Dados, Infraestrutura de Dados Científicos e Infraestrutura de Dados Espaciais; (b) estabelecer um cenário sobre IDEs nas instituições de ensino superior no Brasil; e (c) debater a percepção pela comunidade acadêmica sobre sua produção por meio de um questionário.

Para este fim o presente trabalho segue dividido da seguinte maneira: uma revisão bibliográfica e conceitual com exemplos iniciativas de infraestruturas; uma cenarização das IDEs Acadêmicas no Brasil; apresentação do questionário aplicado e seus resultados; uma discussão dos resultados deste trabalho; e as conclusões alcançadas.

2. CONCEITOS E INICIATIVAS ATUAIS SOBRE INFRAESTRUTURA DE DADOS, INFRAESTRUTURA DE DADOS CIENTÍFICOS E INFRAESTRUTURA DE DADOS ESPACIAIS

A nova escala de produção de dados traz impactos aos negócios, à indústria, à ciência, aos governos e às sociedades como um todo. Especificamente no campo das ciências, antes tínhamos métodos científicos que focam essencialmente em extrair informações e formular conclusões a

Revista Eletrônica: Tempo - Técnica - Território, V.9, N.2 (2018), 28:47 ISSN: 2177-4366

partir de um conjunto limitado de observações, porém hoje há novas disciplinas que lidam com grandes volumes de dados (Miller, 2010). Por outro lado, novos campos de conhecimento como a Ciência da Informação trazem questões sociais para o centro da discussão: como a sociedade e seu conhecimento influenciam na produção e consumo da informação tão bem como a informação pode influenciar as sociedades e seu conhecimento. Entretanto para além da produção do conhecimento, hoje o conceito sobre o termo informação está tão também associado ao poder, à vantagem estratégica, à cidadania, à gestão pública mais eficiente, à tomada de decisão qualificada, entre outros.

Um dos problemas endereçados frequentemente por trabalhos sobre gestão da informação é como uma base de dados pode ser eficientemente planejada, identificada, gerida e compartilhada observando tanto aspectos técnicos como sociais. É nesse contexto que o conceito de *infraestrutura de dados* é formulado. Segundo Kitchin (2014), o termo infraestrutura de dados pode ser entendido como “*meios institucionais, físicos e digitais para armazenamento, compartilhamento e consumo de dados por meio de tecnologias ligadas em rede.*” (p. 32, tradução nossa), possuindo diferentes níveis de organização em relação às tecnologias adotadas e, principalmente, à governança sobre os dados.

Kitchin (2014) e O’Carroll et al (2013) propõem a seguinte divisão para infraestruturas de dados: (a) *data holdings*, que são coleções informais de dados; (b) *data archives*, que são coleções formais, isto é, dados estruturados, com curadoria, documentados e com metadados; (c) *catalogues, directories e portals*, que são recursos que associam e centralizam diferentes data holdings e data archives, principalmente como agregadores de metadados; (d) *data repositories*, que objetiva que cada data holding e cada data archive compilado atenda requisitos determinados a fim de garantir determinadas integridades sobre os dados, porém esses dados são produzidos em diferentes padrões e sob diferentes políticas; e (e) *cyberinfraestruturas*: um conjunto de hardware e tecnologias em rede dedicado e integrado, incluindo serviços *middleware*, software interoperáveis, ferramentas analíticas, serviços compartilhados, visualização de dados e políticas compartilhadas (Cyberinfrastructure Council, 2007).

Uma vez que é importante considerar os aspectos de produção, armazenamento e consumo desses dados para diferentes fins, então podemos falar de infraestruturas de dados especializadas em determinados tipos de informação ou determinados campos de aplicação. Como interesse desse trabalho fita sobre a produção e compartilhamento de informações geoespaciais no ambiente universitário, há destaque aqui às Infraestruturas de Dados Espaciais (IDE ou SDI em inglês) e Infraestrutura de Dados Científicos (IDC ou coincidentemente também SDI em inglês). Ressalta-se que ao contrário do termo Infraestrutura de Dados Espaciais, o termo Infraestrutura de Dados

Revista Eletrônica: Tempo - Técnica - Território, V.9, N.2 (2018), 28:47 ISSN: 2177-4366

Científicos não é utilizado em todos os trabalhos aqui referenciados; esse termo é usado em Demchenko et al (2013), Katal et al (2013), Lane e Bertuzzi (2011) e Arzberger et al (2004). Nos demais trabalhos outras composições de palavras são aplicadas como arquivo ou repositório de dados científicos ou acadêmicos, porém sempre alinhados com o conceito de infraestrutura de dados aqui apresentado.

Diferentes facetas do desafio sociotécnico de alcançar uma IDC eficiente são endereçadas por trabalhos recentes. Por exemplo, Wilkinson et al (2016) aponta a necessidade de se adotar os quatro princípios FAIR para dados científicos: *Findability* (prospectabilidade), *Accessibility* (acessibilidade), *Interoperability* (interoperabilidade) e *Reusability* (reutilizabilidade), enquanto Tenopir et al (2011) e Gray et al (2005) endereçam questões sobre hábitos do usuário e organização de repositórios de dados científicos. Já os trabalhos de Simmhan (2005), Bose e Frew (2005) e Buneman et al (2004) discutem sobre o problema da documentação, versionamento e identificação das fontes primárias dos dados científicos, enquanto Demchenko et al (2013) e Ailamaki et al (2010) abordam cenários de Big Data. Em uma abordagem mais econômica, David (2003) debate sobre a lógica econômica de licenças de uso sobre dados científicos, Lane e Bertuzzi (2011) discursa que IDCs são importantes para se estabelecer medidas de avaliação do retorno de investimento em pesquisas, Kitchin et al (2015) escreve sobre o financiamento de repositórios de *open data* e Kehrer e Hauser (2013) defende que a natureza dos acervos de dados científicos possui características multifacetadas: espaço-temporal, multivariado, multimodal, multiprocessado e multimodelado. Ressalta-se essas discussões podem ser levadas a cenários de IDEs.

Hoje existem expressivas iniciativas¹ globais de IDCs como Genbank (Benson et al, 2012), Data Portal of World Data System (WDS)², Worldwide Protein Data Bank (wwPDB³) (Berman et al, 2003), Harvard Dataverse Project³, EUDat (Lecarpentier, 2013) e Dataverse (Crosas, 2011). Por outro lado, diferentes trabalhos como Wilkinson et al (2016), Muratori e Samakovlija (2015), Kitchin (2014) e Gray et al (2005) apontam que ainda há um grande volume de dados científicos desorganizado e, logo, indisponível a grande parte dos potenciais usuários. Além disso, essa desorganização favorece ao plágio e outras quebras de direitos autorais. Como IDCs brasileiras podemos apontar o Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr)⁴, o CarpeDIEN⁵ com dados científicos sobre energia nuclear, ou mesmo o Acervo Digital da UNESP⁶ e

¹ Vide uma listagem em <https://www.nature.com/sdata/policies/repositories> (acesso: dezembro de 2017)

² <https://www.icsu-wds.org/services/data-portal> (acesso: dezembro de 2017)

³ <https://dataverse.harvard.edu/> (acesso: dezembro de 2017)

⁴ <http://www.sibbr.gov.br> (acesso: dezembro de 2017)

⁵ <http://carpedien.ien.gov.br/> (acesso: dezembro de 2017)

⁶ <https://acervodigital.unesp.br/> (acesso: dezembro de 2017)

o Repositório da Produção Científica e Intelectual da Unicamp⁷ que armazenam produções técnicas e científicas para além do tradicional armazenamento de monografias, teses e dissertações.

No que tange Infraestrutura de Dados Espaciais – vide Dessers (2012), Harvey et al (2012) e Hendriks et al (2012) para uma revisão conceitual e história – diferentes trabalhos como Machado (2016) e Brasil (2010) sustentam que a particularidade de uma IDE reside na natureza da informação envolvida: a informação geoespacial. Ressalta-se que o termo informação geoespacial também é encontrado na literatura como informação geográfica, geoinformação e informação espacial. Veja um debate não consensual sobre as diferenças e semelhanças desses termos em Longley et al. (2015), Brasil (2010), Castiglione (2009) e Raper (2000). Vide próprio nome IDE, que se firmou com um uso menos usual do termo: *informação espacial*.

Segundo Goodchild (2011), em poucas palavras, a informação geoespacial é uma das formas de se representar um conhecimento sobre a superfície da Terra. Assim como ocorreu em outras áreas, os adventos de tecnologias nas últimas décadas como Sistemas de Informações Geográficas, Banco de Dados Geográficos, Sistemas de Posicionamento Global por Satélites e Sensoriamento Remoto permitiram um salto quantitativo e qualitativo na produção e compartilhamento da produção de informações geoespaciais, as tornando disponíveis para um público maior e sendo utilizadas em um espectro mais amplo de aplicações.

Se o debate sobre IDE avançou em muito associado a iniciativas nacionais, ou seja, para compartilhamento de dados principalmente governamentais e oficiais produzidos por um país, o debate mais recente – que alguns trabalhos chamam de terceira geração de IDEs (Machado, 2016) – amplia a discussão mais profusamente para outros níveis de IDE e para formas descentralizadas de produção e compartilhamento desses dados (Gao et al. 2017; Ramos e Ferreira, 2015; Dessers 2012; Budhathoki, 2008). Ressalta-se que a arquitetura da plataforma tecnológica de uma IDE é comumente orientada a serviços (*Service-Oriented Architectures* – SOA) (Machado, 2016; Brasil, 2010; David et al, 2009), sendo um arquétipo ao conceito de *cyberinfrastructures* (Kitchin e Lauriault, 2014). Entretanto o debate sobre infraestrutura de dados como apresentado em Kitchin (2014) e O’Carroll et al (2013) aceitam também outros tipos paradigmas de plataforma tecnológica.

As iniciativas de IDEs no Brasil aumentam a olhos vistos nos últimos anos, principalmente depois da instituição da INDE em 2008. Hoje temos IDEs em nível de estados como Alagoas⁸, Bahia⁹, Distrito Federal¹⁰, Espírito Santo¹¹ e São Paulo¹² e também em nível municipal como Belo

⁷ <http://repositorio.unicamp.br> (acesso: dezembro de 2017)

⁸ <http://dados.al.gov.br/> (acesso: dezembro de 2017)

⁹ <http://geoportal.ide.ba.gov.br/geoportal/> (acesso: dezembro de 2017)

¹⁰ <https://www.geoportal.segeth.df.gov.br/mapa/> (acesso: dezembro de 2017)

Horizonte (MG)¹³ e São Paulo (SP)¹⁴. Em complemento, também há iniciativas importantes que envolvem dados científicos e geoespaciais como a Embrapa¹⁵ e o Serviço Geológico do Brasil (CPRM)¹⁶.

Grande parte dos trabalhos aqui citados direcionam seus debates mais especificamente para IDEs ou para IDCs, apesar de suas evidentes similaridades principalmente quando olhamos uma IDE no ambiente acadêmico. Não por menos trabalhos como Machado (2016) e Kitchin e Lauriault (2014) já endereçam essas duas linhas de debate em conjunto.

3. Infraestrutura de dados nas instituições de ensino superior brasileiras

Especificamente no cenário brasileiro, a Infraestrutura Nacional de Dados Espaciais (INDE), que foi criada pelo Decreto Federal 6.666 de 2008 (Brasil, 2008), figura como a mais importante IDE no Brasil. Apesar do documento intitulado “Plano de Ação da INDE” (Brasil, 2010) definir o papel das instituições de ensino superior apenas como ente capacitador – “*o setor acadêmico é responsável pela educação, capacitação, treinamento e pesquisa em IDE;*” (p. 61, grifo original) – qualquer instituição pode aderir à INDE como produtora ou custodiante de informações geoespaciais. Em adição, algumas instituições de ensino superior no Brasil começam a debater que sua produção de dados geoespaciais deve ser mais bem disponibilizada à sociedade (Machado, 2016).

Trabalhos como Smyth (2015), Muratori e Samakovlija (2015), Ellul et al (2012), Hill e Trimble (2012), Burton et al (2012) e Grill et al (2010) debatem questões principalmente focadas em padronizações, tecnologias e metadados de IDEs em ambientes acadêmicos fora do Brasil, a saber: Reino Unido, Espanha, União Europeia, Canadá e República Tcheca. Já os trabalhos de Silva (2017), Machado (2016) e Camboim e Brandalise (2013) discursam sobre questões relacionadas a uma IDE para a Universidade Federal do Paraná¹⁷, enquanto Ramos e Ferreira (2015) e Oliveira e Ramos (2013) falam sobre uma IDE para a Universidade do Estado do Rio de Janeiro¹⁸. Para além dessas duas instituições, Silva (2017) e Machado (2016) identificaram projetos de IDEs na Universidade de São Paulo¹⁹ e na Universidade Federal da Bahia²⁰. Em complemento, foi realizada

¹¹ <https://geobases.es.gov.br/> (acesso: dezembro de 2017)

¹² <http://www.idesp.sp.gov.br/> (acesso: dezembro de 2017)

¹³ <http://geoportail.pbh.gov.br/> (acesso: dezembro de 2017)

¹⁴ <http://geosampa.prefeitura.sp.gov.br/> (acesso: dezembro de 2017)

¹⁵ <http://geoinfo.cnpm.embrapa.br/> (acesso: dezembro de 2017)

¹⁶ <http://geosgb.cprm.gov.br/> (acesso: dezembro de 2017)

¹⁷ <http://www.idea.ufpr.br/> (acesso: dezembro de 2017)

¹⁸ <https://metadados.labgis.uerj.br:8080/> (acesso: dezembro de 2017)

¹⁹ <http://www.cesadweb.fau.usp.br/> (acesso: dezembro de 2017)

²⁰ <http://ideufba3.webnode.com/> (acesso : dezembro de 2017)

uma consulta na lista de instituições participantes da INDE junto ao seu Catálogo de Metadados²¹ no dia 07 de dezembro de 2017, onde apenas três universidades figuravam entre as instituições participantes, a saber: Universidade do Estado do Rio de Janeiro, Universidade Federal do ABC e Universidade de Brasília, totalizando 28 entradas de metadados. É evidente que tanto o número de universidades como o número de entradas no catálogo são muito pequenos frente à dimensão da produção de informações geoespaciais dentro das instituições universitárias. Todavia há um movimento recente e crescente de criação de IDEs Acadêmicas que parece ganhar força no Brasil – vide como exemplo essa edição espacial do presente periódico.

4. Avaliando a percepção da comunidade acadêmica sobre suas produções

Entendendo que o nível de sensibilização da comunidade de usuários é fundamental para o sucesso de qualquer infraestrutura de dados, inclusive acadêmica, foi confeccionado e passado um questionário online entre fevereiro de 2016 e novembro de 2017 para avaliar a percepção da comunidade acadêmica sobre a gerência e compartilhamento de produções técnicas e científicas nas instituições de ensino superior (IES). O convite à participação ocorreu pelos canais de comunicação oficiais deste núcleo de pesquisa (e-mail, site e mídias sociais), onde o participante era apresentado a uma contextualização da pesquisa e um Termo de Consentimento Livre e Esclarecido. A Figura 1 ilustra o fluxo entre os blocos de perguntas.

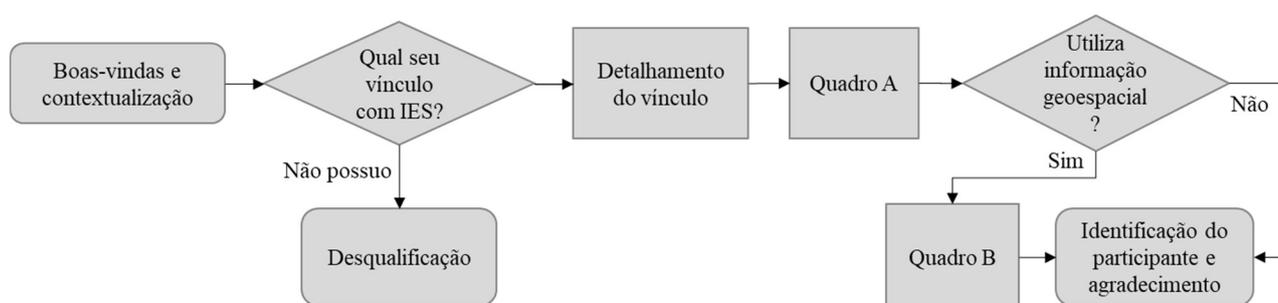


Figura 1 – Fluxograma de preenchimento do questionário para pessoas que integram a comunidade das instituições de ensino superior (IES).

Num total de 640 pessoas responderam, sendo que 117 (18,3%) informaram que não tinham vínculo no momento uma instituição de ensino superior e foram direcionados para a página de desqualificação. Os demais 523 declararam o seguinte vínculo com instituições de ensino superior: 223 alunos de graduação (42,6%), 182 alunos de pós-graduação *lato* ou *stricto sensu* (34,8%), 81 professores (15,5%) e 37 outros como técnicos, bolsistas ou administrativos (7,1%). As instituições com mais participantes em ordem decrescente foram: UERJ (13,9%), UFRJ (7,8%), USP (5,1%), UFPR (4,3%), UNESP (4%) e UFF (3,9%).

²¹ <http://www.metadados.inde.gov.br/geonetwork/srv/por/main.home>

Perguntou-se também para os alunos e professores qual o curso graduação ou pós-graduação que eles estão associados. A resposta foi informada como texto livre, onde as palavras mais frequentes nas respostas foram: Geografia (39,9%), Engenharia (15,5%) e Ciências (9,5%).

Na sequência de perguntas, o entrevistado foi exposto ao Quadro A com o seguinte título: *'Sobre a produção técnica ou científica nas instituições de ensino, expresse a sua opinião sobre cada afirmação abaixo com "Concordo", "Discordo" ou "Não sei opinar"'*. As afirmações apresentadas ao entrevistado foram:

- A1.** A produção técnica ou científica da minha instituição de ensino possui valor apenas dentro de seu âmbito.
- A2.** A minha instituição de ensino disponibiliza a sua produção técnica ou científica à sociedade como um todo.
- A3.** Qualquer cidadão deve ter acesso irrestrito à produção técnica ou científica da minha instituição de ensino.
- A4.** As instituições de ensino devem compartilhar suas produções técnicas ou científicas entre si.
- A5.** A minha instituição de ensino organiza e disponibiliza sua produção técnica ou científica de forma satisfatória.
- A6.** Eu utilizo as bases de dados, bibliotecas e outros repositórios de produção técnica ou científica mantidos pela minha instituição.
- A7.** É parte de minhas atividades realizar produção técnica ou científica no âmbito da minha instituição de ensino.
- A8.** A produção técnica ou científica de minha autoria deve estar totalmente à disposição da minha instituição, se desenvolvida em seu âmbito.

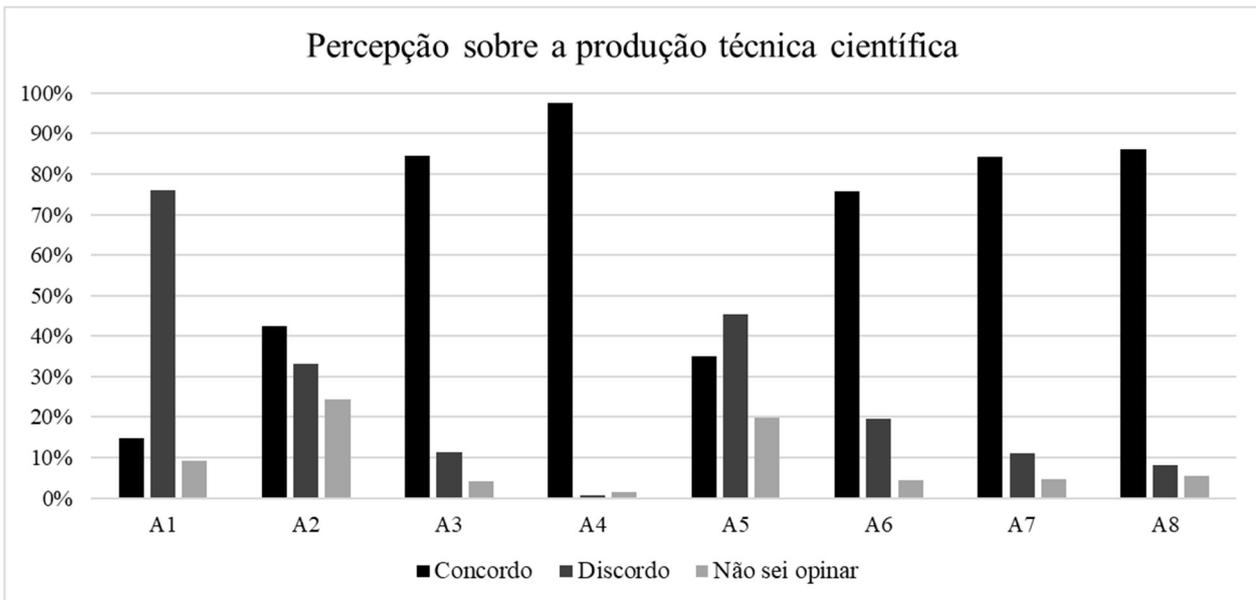


Figura 2 – Gráfico com as respostas às afirmações do Quadro A sobre a percepção da produção técnico e científica na instituição de ensino superior.

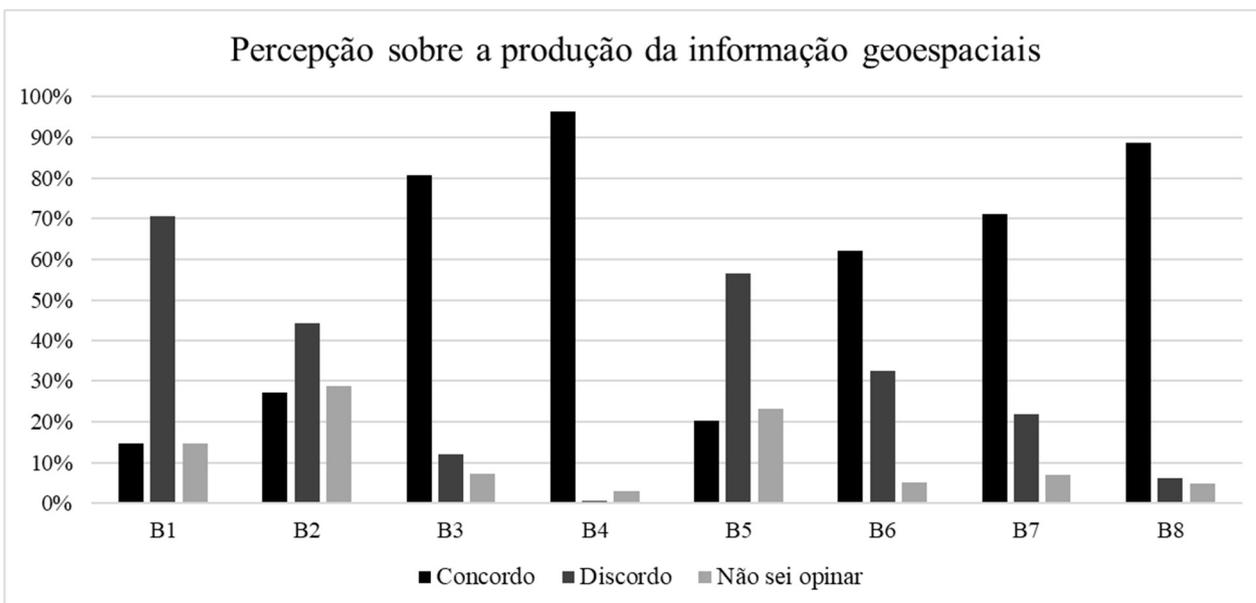


Figura 3 – Gráfico com as respostas às afirmações do Quadro B sobre a percepção da produção de informações geoespaciais na instituição de ensino superior.

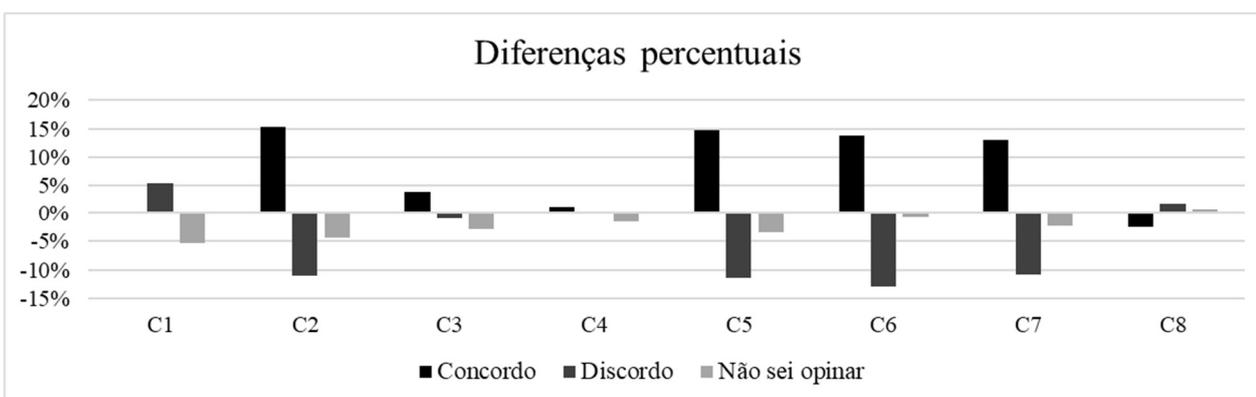


Figura 4 – Gráfico com a diferença percentual das respostas às afirmações dos quadros de percepção, onde Ck é calculado a partir dos valores percentuais das respostas para a afirmação Ak menos os valores percentuais das mesmas respostas para Bk .

As afirmativas foram construídas evitando o uso de termos técnicos, pois o desconhecimento destes poderia comprometer as respostas. Como exemplo tem-se a afirmação A3, onde foi apresentada uma frase sem termos técnicos em vez de apresentá-la usando termos como *open data*. Também se buscou a criação de afirmações simples que pudessem ser mais facilmente transportadas para o cotidiano do entrevistado em sua instituição de ensino.

Na sequência do questionário online, se perguntou *‘Você utiliza ou produz informação geoespacial dentro ou fora da sua instituição de ensino?’*. O objetivo aqui era avaliar se o entrevistado se entendia como consumidor ou gerador de informação geoespacial, onde 387 (83,6%) responderam afirmativamente, 76 (16,4%) negativamente e 60 participantes abandonaram o questionário após responder o primeiro quadro. O questionário direcionou para o fim quem respondeu negativamente essa questão – vide Figura 1. Mas para os demais 387 participantes o Quadro B sobre a percepção da informação geoespacial foi apresentado com o título: *‘Sobre a produção de informações geoespaciais nas instituições de ensino, expresse a sua opinião sobre cada afirmação abaixo com "Concordo", "Discordo" ou "Não sei opinar"’*. As afirmações foram:

- B1.** As informações geoespaciais produzidas pela minha instituição de ensino superior possuem valor apenas no seu próprio âmbito.
- B2.** A minha instituição de ensino disponibiliza suas informações geoespaciais para a sociedade.
- B3.** Qualquer cidadão deve ter acesso irrestrito a produção de informações geoespaciais produzidas pela minha instituição de ensino superior.
- B4.** A informação geoespacial das instituições de ensino deve ser compartilhada entre elas.
- B5.** A minha instituição de ensino organiza e disponibiliza sua produção de informações geoespaciais de forma satisfatória.
- B6.** Já utilizei informações geoespaciais produzidas por minha instituição de ensino em minhas atividades técnicas ou científicas.
- B7.** É parte de minhas atividades junto à minha instituição de ensino produzir informações geoespaciais.
- B8.** A produção de informações geoespaciais de minha autoria deve estar totalmente à disposição da minha instituição, se desenvolvida em seu âmbito.

Observe que há proposital correspondência entre as afirmações do primeiro e do segundo quadro. As Figuras 2, 3 e 4 mostram as respostas ambos os quadros, tão como as diferenças percentuais entre as respostas de cada um.

Duas perguntas adicionais também foram apresentadas junto ao Quadro B. A primeira foi *'Você conhece o conceito sobre Infraestrutura de Dados Espaciais (IDE)?'*, onde 49,5% responderam afirmativamente e 50,5% negativamente. A segunda pergunta foi *'Considerando a importância da produção de informação geoespacial e das produções bibliográficas como artigos acadêmicos, monografias ou teses, qual possui mais relevância para a produção técnica ou científica da instituição de ensino?'*. As opções para responder e seus percentuais de marcação foram: (1) *'São mais importantes as produções bibliográficas como artigos acadêmicos, monografias e teses.'* com 7,2%; (2) *'É mais importante a produção de informação geoespacial.'* com 2,2%; (3) *'Ambas possuem importância equivalente.'* com 27,6%; (4) *'Cada uma possui importância em diferentes contextos.'* com 61,3%; e (5) *'Não sei opinar.'* com 1,7%.

Por fim, analisaram-se as respostas para os Quadros A e B separando por alunos de graduação, alunos de pós-graduação e professores. Aplicou-se então o teste *t de Student* para selecionar quais são as diferenças significantes estatisticamente para um limiar de *p-valor* igual a 0,05. Em outras palavras, se aplicou esse teste de hipótese para se avaliar quais as respostas são significativamente diferentes entre as três categorias selecionadas de entrevistados, a saber:

- A resposta *'Não sei opinar'* à afirmação A2, onde 30,1% dos alunos de graduação responderam contra 17,3% dos professores.
- A resposta *'Concordo'* e *'Não sei opinar'* à afirmação A3, onde respectivamente os alunos de graduação responderam 79,3% e 7,2% e os alunos de pós-graduação responderam 88% e 1,8%.
- A resposta *'Concordo'* à afirmação A6 entre alunos de pós-graduação (81,4%) e professores (68%).
- A opção *'Concordo'* à afirmação A7 para alunos de graduação (77,7%) em relação a alunos pós-graduação (89,8%) e professores (89,3%).
- A opção *'Não sei opinar'* à afirmação A7 para alunos de graduação (8,8%) em relação a alunos pós-graduação (2,4%) e professores (1,3%).
- A opção *'Concordo'* à afirmação A8 entre alunos de graduação (81,9%) e professores (93,3%); e a opção *'Não sei opinar'* entre alunos de graduação (9,3%) e alunos de pós-graduação (3,6%).

- A opção ‘*Discordo*’ à afirmação B2 entre alunos de graduação (37%) em relação a alunos de pós-graduação (50%) e professores (50,9%).
- A opção ‘*Não sei opinar*’ à afirmação B2 entre alunos de graduação (34,8%) e professores (19,3%).
- A opção ‘*Discordo*’ à afirmação B5 entre professores (70,2%) em relação a alunos de graduação (51,8%) e pós-graduação (54,9%).
- A opção ‘*Não sei opinar*’ à afirmação B5 entre professores (10,5%) em relação a alunos de graduação (25,2%) e pós-graduação (27,4%).
- A opção ‘*Discordo*’ à afirmação B6 entre alunos de graduação (26,7%) em relação a alunos de pós-graduação (38,7%).
- A opção ‘*Não sei opinar*’ à afirmação B5 entre alunos (11,8%) em relação a alunos de pós-graduação (2,11%) e professores (0%).

Também houve diferença significativa para a pergunta ‘*Você conhece o conceito sobre Infraestrutura de Dados Espaciais (IDE)?*’, onde os alunos de graduação, de pós-graduação e professores marcaram 61%, 44% e 38,7% respectivamente à resposta negativa.

5. DISCUSSÕES

Os resultados do questionário nos permite apontar que há um ambiente favorável para implantação de uma IDE, apesar de que as intenções nem sempre se apresentam melhores do que a ação quando, por exemplo, um professor precisa preencher uma ficha de metadados – vide uma discussão sobre as dificuldades dos pesquisadores em organizar e documentar seus dados em Tenopir et al (2011). Pelo menos 75% dos entrevistados no Quadro A e 70% no Quadro B entendem que essa produção técnica e científica possui valor não apenas para dentro das instituições (A1 e B1), que ela deve ser compartilhada com outras instituições de ensino (A4 e B4) e também com o cidadão (A3 e B3), que o entrevistado se considera um consumidor dessas produções (A6²²) e também um autor (A7 e B7), sendo que as produções de sua autoria devem estar à disposição da instituição de ensino (A8 e B8). Além disso, mais de 60% dos entrevistados afirmam que a informação geoespacial detém seu valor em um contexto diferente do que monografias, teses ou dissertações. Em suma, as respostas são condizentes a valores importantes às iniciativas de infraestruturas de dados no ambiente acadêmico.

Por outro lado, as perguntas que tocam a questão do compartilhamento dos apontam para uma percepção menos positiva e ao mesmo tempo condizente com o cenário descrito sobre as IDEs

²² B6 ficou com 62,2% para a resposta ‘Concordo’, desta forma abaixo do piso de 70% delimitado anteriormente no texto.

Acadêmicas no Brasil. Se 42,5% concorda que a instituição disponibiliza sua produção técnica e científica a sociedade (afirmação A2), esse percentual cai para 27,1% quando a afirmação se refere a informações geoespaciais (B2). Quando se coloca a afirmação sobre organizar e disponibilizar de forma satisfatória, ou seja, com algum nível de qualificação e com uma dimensão mais prática, os percentuais para a resposta '*Concordo*' caem para 34,9% (A5) e 20,2% (B5) respectivamente. Por fim, as respostas '*Discordo*' se apresentaram com diferenças estatisticamente significativas em B2 e B5 entre alunos de graduação e professores, este último grupo com valores maiores do que o primeiro em cerca de 15%. Talvez essa maior percepção negativa possa ser explicada, pelo menos em parte, pela interação diferente que o professor possui com a instituição de ensino e também por seu maior conhecimento técnico – vide que os percentuais de conhecimento sobre o conceito de IDE é 1,6 vezes maior entre os professores (61%) em comparação aos alunos de graduação (38,7%).

Mais especificamente sobre o gráfico da Figura 4, as diferenças entre o resultado do Quadro A e do Quadro B residem principalmente nas respostas '*Concordo*' e '*Discordo*' novamente sobre a segunda e quinta afirmação, mas também na sexta e na sétima afirmação. Em todos esses quatro casos o percentual de '*Concordo*' está entre 13% e 15% menor no Quadro B em relação ao Quadro A, podendo indicar duas circunstâncias. A primeira é a ainda pior percepção da organização e do compartilhamento de informações geoespaciais frente a outras produções técnicas ou científicas; a segunda é que os entrevistados se entendem menos usuários e autores das informações geográficas do que usuários e autores de outras produções técnicas e científicas em suas respectivas instituições de ensino.

Os resultados nos parecem bem condizentes com o cenário descrito sobre as IDEs Acadêmicas, de todo o modo, chama a atenção que 20,2% dos entrevistados tenham concordado com a afirmação B5. Ao se analisar as diferenças das entrevistas entre esse 1/5 das respostas para os 4/5 demais, não foi percebido nenhuma diferença expressiva como tipo de vínculo, conhecimento do conceito IDE ou demais respostas aos Quadros A e B – exceto a presença de variações percentuais parecidas às perguntas A2, A5 e B2, que são evidentemente relacionadas ao contexto de B5. Ou seja, há consistência nessa variação. Talvez esses entrevistados acreditem que infraestrutura de dados do tipo *data holding* ou esparsos *data archives* sejam suficientes para organizar e disponibilizar os dados científicos e geoespaciais da sua instituição de ensino. Outra possibilidade é que esses entrevistados pertençam às universidades identificadas anteriormente já com projetos de IDEs Acadêmicas, porém descartou-se essa hipótese uma vez que apenas 8,2% desses 20,2% - ou seja, 0,17% do total dos entrevistados os são.

6. CONCLUSÕES

A quem interessa ter as produções acadêmicas desorganizadas? Talvez seja mais simples responder ao quê não interessa: a uma ciência de ponta; a uma transparência na produção científica; a uma economicidade dos recursos envolvidos, que em importante parte são públicos; a uma graduação e pós-graduação mais qualificada; a um ensino com técnicas de ensino-aprendizagem versáteis; a um respeito aos direitos autorais; a um mercado com mais insumos para decidir; a um ambiente acadêmico que dialogue mais com a sociedade; a uma democratização do acesso à informação. Em suma, essa desorganização não interessa a uma sociedade mais informada e empoderada. Acreditamos que essa deva ser a maior motivação para a criação de infraestruturas de dados acadêmicas e fazer conceitos como *Open Science* (David, 2003) se afastar cada vez mais do imaginário das utopias e ganhar contornos de realidade dentro da academia brasileira.

Para ilustrar o cenário corrente, tomemos como exemplo a base de dados produzida a partir do questionário apresentado nesse trabalho. Desconhecemos uma infraestrutura em nossa instituição de ensino onde possamos disponibilizar tal base para quaisquer interessados que desejem conferir nossas análises ou mesmo realizar outras análises não aqui apresentadas. Como acreditamos que seria uma contradição falar sobre o compartilhamento de dados gerados na academia e não disponibilizar a base de dados gerada para a criação desse artigo, a disponibilizamos abertamente na plataforma Mendeley sob a licença Creative Commons Attribution 4.0 International

Todavia uma pergunta que cabe é se essas infraestruturas de dados acadêmicas seriam IDEs ou IDCs. Não que o nome dado à infraestrutura seja mais importante do que sua finalidade, porém parece existir uma discussão hoje demasiadamente distante entre esses dois conceitos que se complementam. Na contramão dessas distantes discussões, estão as plataformas tecnológicas que cada vez mais integram dados geoespaciais dos demais tipos de dados – segundo ideias de Harvey (2013) e Norman (1998), esse é o movimento que fará os Sistemas de Informações Geográficas ficarem invisíveis aos seus usuários e virarem infraestruturas²³. Vide como exemplo dessa integração de diferentes tipos de dados o Data.Rio²⁴ da Prefeitura do Rio de Janeiro (RJ) e o Sistema de Informações Municipais do Estado de Alagoas²⁵, que integram planilhas, mapas, aplicações, documentos, serviços e outros formatos de dados. Também temos como exemplo o trabalho de Silva (2016), que discute a incorporação de algoritmos – hoje um formato não usual de informação – em IDEs. Em suma, uma infraestrutura de dados exitosa provavelmente necessita trabalhar com dados em diferentes formatos.

²³ <https://data.mendeley.com/datasets/ry8j9t9z5m/1> (acesso: setembro de 2018). DOI: 10.17632/ry8j9t9z5m.1

²⁴ Há uma coincidência no uso da palavra *infraestrutura* entre Norman (1998) e os conceitos de infraestrutura de dados.

²⁵ <http://www.data.rio/> (acesso: dezembro de 2017)

²⁵ <http://dados.al.gov.br/> (acesso: dezembro de 2017)

Podemos dizer que um movimento de IDEs Acadêmicas já se iniciou no Brasil, porém ainda há muito ao que se implementar e muitas questões a serem endereçadas. Para além da mudança do paradigma tecnológico, há uma mudança da cultura técnica e científica da comunidade acadêmica, que envolve uma melhor gestão da informação produzida e consumida – um exemplo usual é a necessidade do preenchimento dos metadados. Talvez essa mudança de cultura seja um dos maiores desafios (Ramos e Ferreira, 2015; Tenopir, 2011), onde não por menos as instituições de fomento a pesquisa como, por exemplo, as brasileiras CNPq e CAPES, demandam hoje a divulgação dos resultados gerados pelos projetos financiados (Machado, 2016; Lane e Bertuzzi, 2011). Como ponto positivo, os resultados coletados a partir do questionário apontam que, pelo menos, uma parte da comunidade acadêmica está de alguma forma sensibilizada a valores importantes ao tema de infraestrutura de dados.

Por fim, o trabalho de Kitchin et al (2015) alerta: “*muitos dados gerados ao longo da história foram perdidos ou destruídos porque eles foram armazenados de forma informal, em vez de um arquivo institucional.*” (p. 1, tradução nossa). O mesmo ocorre hoje em muitos nichos da academia: uma quantidade inestimada da produção intelectual é mal aproveitada ou mesmo perdida.

Como expressou Teixeira de Freitas há quase um século em um pensamento que continua atual e pode ser transportado ao ambiente acadêmico brasileiro:

O Brasil, lamentavelmente, ainda é um país que ignora quase tudo de si mesmo. Os dados com que se possa caracterizar qualquer aspecto da vida nacional, ou não existem, ou são difíceis de encontrar, ou se acham elaborados fragmentária e lacunosamente, ou são antiquados, ou se contradizem, quando não se distanciam da verdade. (IBGE, 1939)

Referências Bibliográficas

AILAMAKI, Anastasia; KANTERE, Verena; DASH, Debabrata. Managing scientific data. **Communications of the ACM**, v. 53, n. 6, p. 68-78, 2010.

ARZBERGER, Peter et al. Promoting access to public research data for scientific, economic, and social development. **Data Science Journal**, v. 3, p. 135-152, 2004.

BENSON, Dennis A. et al. GenBank. **Nucleic acids research**, v. 41, n. D1, p. D36-D42, 2012.

BERMAN, Helen; HENRICK, Kim; NAKAMURA, Haruki. Announcing the worldwide protein data bank. **Nature Structural & Molecular Biology**, v. 10, n. 12, p. 980-980, 2003.

BOSE, Rajendra; FREW, James. Lineage retrieval for scientific data processing: a survey. **ACM Computing Surveys (CSUR)**, v. 37, n. 1, p. 1-28, 2005.

BRASIL. **Decreto Nº 6.666 de 27 de Novembro de 2008**. Institui, no âmbito do Poder Executivo federal, a Infra-Estrutura Nacional de Dados Espaciais - INDE, e dá outras providências, 2008. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2007-2010/2008/decreto/d6666.htm (Acesso: Dezembro de 2018)

BRASIL – Ministério do Planejamento, Orçamento e Gestão, **Plano de Ação para implantação da INDE: Infraestrutura nacional de dados espaciais**, Comissão Nacional de Cartografia (CONCAR), Rio de Janeiro, 2010. Disponível em: <http://www.concar.gov.br/pdf/PlanoDeAcaoINDE.pdf> (Acesso: Dezembro de 2017)

BUDHATHOKI, Nama Raj et al. Reconceptualizing the role of the user of spatial data infrastructure. **GeoJournal**, v. 72, n. 3-4, p. 149-160, 2008.

BUNEMAN, Peter et al. Archiving scientific data. **ACM Transactions on Database Systems (TODS)**, v. 29, n. 1, p. 2-42, 2004.

BURTON, A.; GROENEWEGEN, D.; LOVE, C.; TRELOAR, A.; WILKINSON, R.. Making research data available in Australia. **Intelligent Systems, IEEE**, v. 27, n. 3, p. 40–43, 2012.

CAMBOIM, S.P.; BRANDALISE, M.C.B. The Role of Universities in the Brazilian NSDI Capacity Building Plan – The Case of the Open Geospatial Laboratory at UFPR. In: **Proceeding of the 26th ICC – International Cartographic Conference**, Dresden, Germany, 2013.

CAPURRO, Rafael., HJØRLAND, Birger. The concept of information. **Annual Review of Information Science and Technology**. Draft version, vol. 37, p. 343-411, 2003. Disponível em: <http://www.capurro.de/infoconcept.html> (Acesso: Dezembro de 2017).

CASTIGLIONE, Luiz Henrique Guimarães. **Epistemologia da geoinformação: uma análise histórico-crítica** (Tese de Doutorado), Universidade Federal Fluminense, Rio de Janeiro, 2009.

CROSAS, Mercè. The dataverse network®: an open-source application for sharing, discovering and preserving data. **D-lib Magazine**, v. 17, n. 1/2, 2011.

CSORNY, Lauren. Careers in the growing field of information technology services. **Beyond the Numbers** (Vol. 2, No. 9). Washington, DC: Bureau of Labor Statistics, 2013.

CYBERINFRASTRUCTURE COUNCIL 2007, **Cyberinfrastructure vision for 21st century discovery**, National Science Foundation. Disponível em: <http://www.nsf.gov/pubs/2007/nsf0728> (Acesso: Dezembro de 2017).

DAVIS JR, Clodoveu A.; FONSECA, Frederico T.; CÂMARA, Gilberto. Infraestruturas de dados espaciais na integração entre ciência e comunidades para promover a sustentabilidade ambiental. In: **Workshop De Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais**. 2009. p. 1279-1288.

DAVID, Paul A. The economic logic of “open science” and the balance between private property rights and the public domain in scientific data and information: a primer. **The role of the public domain in scientific and technical data and information**, p. 19-34, 2003.

DEMCHENKO, Yuri et al. Addressing big data issues in scientific data infrastructure. In: **Collaboration Technologies and Systems (CTS), 2013 International Conference on**. IEEE, 2013. p. 48-55.

DESSERS, Ezra. **Spatial Data Infrastructures at work. A comparative case study on the spatial enablement of public sector processes** (Tese de Doutorado), Katholieke Universiteit

- Leuven, 2012. Disponível em: <https://lirias.kuleuven.be/handle/123456789/346510> (Acesso: Dezembro de 2017)
- ELLUL, Claire et al. Bridging the gap between traditional metadata and the requirements of an academic SDI for interdisciplinary research. **Spatially Enabling Government, Industry and Citizens**, 2012.
- GAO, Song et al. Constructing gazetteers from volunteered big geo-data based on Hadoop. **Computers, Environment and Urban Systems**, v. 61, p. 172-186, 2017.
- GOODCHILD, Michael F. Challenges in geographical information science. In: **Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**. The Royal Society, 2011. p. 2431-2443.
- GRAY, Jim et al. Scientific data management in the coming decade. **ACM SIGMOD Record**, v. 34, n. 4, p. 34-41, 2005.
- GRILL, Stanislav et al. Archive and catalogue system for receiving satellite data as a part of academic SDI. In: **Imagin [e, G] Europe: Proceedings of the 29th Symposium of the European Association of Remote Sensing Laboratories**, Chania, Greece. IOS Press, 2010. p. 150.
- HARVEY, Francis et al. SDI past, present and future: a review and status assessment. In: In: RAJABIFARD, A.; COLEMAN, D. **Spatial Enabling Government, Industry and Citizens**, p. 23-38, 2012.
- HARVEY, F. A New Age of Discovery: The Post-Gis Era. In: **Gi_Forum 2013. Creating the Gisociety**, JEKEL, T., CAR, A., STROBL, J., GRIESEBNER, G. (eds.), p. 272-281. Berlin: Wichmann-Verlag.
- HENDRIKS, Paul HJ; DESSERS, Ezra; VAN HOOTEGEM, Geert. Reconsidering the definition of a spatial data infrastructure. **International journal of geographical information science**, v. 26, n. 8, p. 1479-1494, 2012.
- HILL, Elizabeth; TRIMBLE, Leanne. Scholars GeoPortal: A New Platform for Geospatial Data Discovery, Exploration and Access in Ontario Universities. **IASSIST Quarterly**, v. 36, n. 1, 2012.
- IBGE. Resoluções. Assembleia Geral do Conselho Nacional de Estatística. Rio de Janeiro: IBGE/SERGRAF, t. 4, 379 p, 1939.
- KEHRER, Johannes; HAUSER, Helwig. Visualization and visual analysis of multifaceted scientific data: A survey. **IEEE transactions on visualization and computer graphics**, v. 19, n. 3, p. 495-513, 2013.
- KITCHIN, Rob. **The data revolution: Big data, open data, data infrastructures and their consequences**. Sage, 2014.
- KITCHIN, Rob; COLLINS, Sandra; FROST, Dermot. Funding models for open access digital data repositories. **Online Information Review**, v. 39, n. 5, p. 664-681, 2015.
- KITCHIN, Rob; LAURIAULT, Tracey P. **Small data, data infrastructures and big data**. The Programmable City Working Paper 1, National University of Ireland Maynooth, County Kildare, Ireland, 2014.

LANE, Julia; BERTUZZI, Stefano. Measuring the results of science investments. **Science**, v. 331, n. 6018, p. 678-680, 2011.

LECARPENTIER, Damien et al. EUDAT: a new cross-disciplinary data infrastructure for science. **International Journal of Digital Curation**, v. 8, n. 1, p. 279-287, 2013.

LONGLEY, Paul A.; Goodchild, M. F.; Maguire, D. J.; Rhind, D. W. **Geographic information science and systems**. John Wiley & Sons, 2015.

MACHADO, Adriana Alexandria. IDE acadêmica: proposta para a Universidade Federal do Paraná (UFPR) (Tese de Doutorado). Universidade Federal do Paraná, Brasil, 2016.

MILLER, Harvey J. The data avalanche is here. Shouldn't we be digging?. **Journal of Regional Science**, v. 50, n. 1, p. 181-201, 2010.

MURATORI, Ilaria; SAMAKOVLJA, Marcella. **Sharing geodata through university libraries: the case of Politecnico di Milano**. Geomatics Workbooks n° 12 – FOSS4G Europe Como 2015, 2015. Disponível em: <http://eprints.rclis.org/27995/> (Acesso: Dezembro de 2017).

NORMAN, Donald A. **The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution**. MIT press, 1998.

O'CARROLL, A.; COLLINS, S., GALLAGHER, D., TANG, J.; WEBB, S. Caring for Digital Content, Mapping International Approaches. **Maynooth: NUI Maynooth**, 2013.

OLIVEIRA, Pedro Feijó; RAMOS, Jose Augusto Sapienza. Arquitetura e Desafios de uma Infraestrutura de Dados Espaciais para Setores da Universidade do Estado do Rio de Janeiro. In: **Anais XVI Simpósio Brasileiro de Sensoriamento Remoto – SBSR**. Foz do Iguaçu-PR, Abril 2013, INPE.

ORMELING, Ferjan. Cartography and Geoinformation in the 20 th and 21 st Centuries. **Journal for Theoretical Cartography** ISSN, v. 1868, p. 1387, 2011.

RAMOS, José Augusto Sapienza. Banco de dados geográficos para a gestão ambiental municipal. **Gestão e Governança Local para a Amazônia Sustentável**, v. 2, p. 166-182, 2016. Disponível em: <http://www.amazonia-ibam.org.br/biblioteca/publicacao/detalhe/gestao-e-governanca-local-para-a-amazonia-sustentavel---notas-tecnicas-%7C-volume-02/144> (Acesso: Dezembro de 2017)

RAMOS, Jose Augusto Sapienza; FERREIRA, Carlos Eduardo Goncalves. Discussing the role of university in spatial data infrastructure construction: issues and challenges for UERJ-V-SDI. In: **27th International Cartographic Pre-conference workshop on Spatial data infrastructures, standards, open source and open data for geospatial (SDI-Open)**, 2015. p. 24-28. 2015. Disponível em: http://sdistandards.icaci.org/wp-content/uploads/2015/09/SDI-Open2015_Proceedings.pdf (Acesso: Dezembro de 2017)

RAPER, Jonathan. **Multidimensional geographic information science**. CRC Press, 2000.

SILVA, Eduardo Silverio da. **Métodos de Compartilhamento de Processos Geoespaciais em Infraestruturas de Dados Espaciais** (Dissertação de Mestrado). Universidade Federal do Paraná, Brasil, 2017.

SIMMHAN, Yogesh L.; PLALE, Beth; GANNON, Dennis. A survey of data provenance in e-science. **ACM Sigmod Record**, v. 34, n. 3, p. 31-36, 2005.

SMYTH, C. G. SDI – National to Global: perspectives from the UK academic sector. **Proceedings of a pre-conference workshop of the 27th International Cartographic Conference: Spatial data infrastructures, standards, open source and open data for geospatial (SDI-Open 2015)**, 2015. Disponível em: http://sdistandards.icaci.org/wp-content/uploads/2015/09/SDI-Open2015_Proceedings.pdf (Acesso: Dezembro de 2017)

TENOPIR, Carol et al. Data sharing by scientists: practices and perceptions. **PloS one**, v. 6, n. 6, p. e21101, 2011.

TOLLE, Kristin M.; TANSLEY, D. Stewart W.; HEY, Anthony JG. The fourth paradigm: Data-intensive scientific discovery [point of view]. **Proceedings of the IEEE**, v. 99, n. 8, p. 1334-1337, 2011.

WILKINSON, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. **Nature**, v. 3, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. (Acesso: Dezembro de 2017)

KATAL, Avita; WAZID, Mohammad; GOUDAR, R. H. Big data: issues, challenges, tools and good practices. In: **Contemporary Computing (IC3), 2013 Sixth International Conference on. IEEE**, 2013. p. 404-409.