

Brasília e a memória em registros digitais: traços da paisagem e a preservação de dados

Maria de Fátima Duarte Tavares

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, DF, Brasil
fatimatavares@ibict.br

Miguel Márdero Arellano

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, DF, Brasil
miguel@ibict.br

Bruno Nakagomi

Instituto Brasileiro de Informação em Ciência e Tecnologia, Brasília, DF, Brasil
brunonakagomi@ibict.br

DOI: <https://doi.org/10.26512/rici.v11.n1.2018.8474>

Recebido/Recibido/Received: 2017-11-22

Aceitado/Aceptado/Accepted: 2017-12-04

Resumo: A gestão de dados de pesquisa, em paralelo à disponibilidade pública e informatizada de coleções científicas, contempla a emergência do problema da organização do conhecimento e do reuso de dados digitais. Neste estudo de caso, está em foco o tratamento de dados reutilizados, considerando a problematização de seu acesso futuro e estratégias de preservação digital em longo prazo. O processo de investigação, em análise, está relacionado à urbanização do território do Distrito Federal, após a implantação de Brasília, em contraposição à formação de uma memória científica local. O quadro espaço-temporal, entre 1960 a 2016, visa identificar campos de saber em construção, seus agentes e suas inter-relações a partir de séries de registros visuais, dados populacionais, cartográficos e coleções botânicas. Essas fontes estão sujeitas a políticas de acesso e de disseminação distintas, que colocam restrições ao segundo ciclo de vida dessa documentação. O reuso de dados de coleções científicas, vinculadas ao Cerrado no Distrito Federal, foi orientado por metodologia histórico-geográfica e foi objeto de requalificação analítica, com suporte de um sistema de informações geográficas. Os resultados de mapeamento em cartas temáticas podem ser visualizados na plataforma digital Dataverse, tendo em vista garantir acesso, reuso e preservação digital.

Palavras-chave: Acesso livre; Coleções científicas; Dataverse; Distrito Federal; Gestão de dados de pesquisa; Preservação digital; Reuso de dados.

Brasília and memory in digital records: landscape features and data preservation

Abstract: The management of research data, in parallel to the public and computerized availability of scientific collections, contemplates the emergence of the problem of the organization of knowledge and the reuse of digital data. In this case study, the treatment of reused data is in focus, considering the problematization of its future access and long-term digital preservation strategies. The research process, in analysis, is related to the urbanization of the territory of the Federal District (Brazil), after the implantation of Brasília, in counter position to the formation of a local scientific memory. The space-

temporal frame, between 1960 and 2016, aims to identify fields of knowledge in construction, their agents and their interrelationships from series of visual records, population data, cartographic and botanical collections. These sources are subject to distinct access and dissemination policies, which pose restrictions on the second life cycle of this documentation. The reuse of data from scientific collections, linked to the Cerrado in the Federal District (Brazil), was guided by historical-geographic methodology and was object of analytical requalification, with support of a Geographic Information System. The results of mapping in thematic charts can be viewed on the digital platform (Dataverse), in order to ensure access, reuse and digital preservation.

Keywords: Data reuse; Dataverse; Digital preservation; Distrito Federal; Open access; Research Data Management, Scientific collections.

Brasília y la memoria en registros digitales: rasgos del paisaje y la preservación de datos

Resumen: La gestión de datos de investigación, en paralelo a la disponibilidad pública e informatizada de colecciones científicas, contempla la emergencia del problema de la organización del conocimiento y del reuso de datos digitales. En este estudio de caso, está en foco el tratamiento de datos reutilizados, considerando la problematización de su acceso futuro y de las estrategias de preservación digital a largo plazo. El proceso de investigación, en análisis, está relacionado a la urbanización del territorio del Distrito Federal, después de la implantación de Brasília, en contraposición a la formación de una memoria científica local. El cuadro espacio-temporal, entre 1960 a 2016, busca identificar campos de saber en construcción, sus agentes y sus interrelaciones a partir de series de registros visuales, datos poblacionales, cartográficos y colecciones botánicas. Estas fuentes están sujetas a políticas de acceso y diseminación distintas, que ponen restricciones al segundo ciclo de vida de esa documentación. El reuso de datos de colecciones científicas, vinculadas al Cerrado en el DF, fue orientado por metodología histórica - geográfica y fue objeto de recalificación analítica, con soporte de un sistema de informaciones geográficas - SIG. Los resultados de mapeo en cartas temáticas se pueden ver en la plataforma digital (Dataverse), con el fin de garantizar el acceso, la reutilización y la preservación digital.

Palabras clave: Acceso libre; Colecciones científicas; Dataverse; Distrito Federal; Gestión de los datos de investigación; Preservación digital; Reutilización de datos.

1. Introdução

A informatização de coleções científicas e de coleções representativas do patrimônio cultural do Brasil, com disponibilidade pública, criou novas possibilidades de reuso de material documental e promoveu novas frentes de investigação, para além dos domínios de saber originários. Políticas nacionais de acesso digital, além de estratégias de agregação de acervos em redes distribuídas, tornam acessíveis materiais anteriormente de uso restrito e contribuem não só para agilizar processos de investigação, no domínio da história e áreas afins, quanto para introduzir novas questões às práticas consolidadas de organização do conhecimento em suas vertentes disciplinares, nomeadamente identificadas com as Humanidades Digitais.

A disponibilidade de dados digitais para o reuso em novas pesquisas científicas depende, porém, da confiabilidade da sua proveniência e de curadoria instituída que assegure o caráter de autenticidade de acervos ao longo do tempo, como parte do primeiro ciclo de vida de qualquer documentação. Note-se que, neste estudo de caso, com foco em informações

botânicas, registros visuais e cartográficos, não são objeto de análise os resultados da investigação e sim a relevância de ordenar correlações e os processos de requalificação de informações para a proposição de um modelo experimental de gestão de dados de pesquisa, com utilização de arquivamento em plataforma digital.

A diversidade de fontes e o conjunto de distintos dados seriais que advém, originalmente, de processos de sistematização e classificação diversos, serão traduzidos na própria catalogação dos objetos digitais, conforme áreas de saber, mas o processo de investigação e seus níveis de análise, também configuram dimensões que a estrutura de dados na plataforma digital Dataverse deverá reproduzir, constituindo, assim, um segundo ciclo de vida do material processado.

Destaque-se que como dados de pesquisa, nesta abordagem, entende-se um conjunto de informações quantitativas, textuais e de representações cartográficas ou imagéticas, além de dados de coleta botânica, que remetem a um referente ou uma situação dada.

Esses conjuntos de dados resultam de atividades de pesquisa sobre o conflito entre urbanização e áreas protegidas no bioma Cerrado, na perspectiva de história urbana e no âmbito do Projeto Saberes do Cerrado. Uma das linhas temáticas de investigação trata da transformação da paisagem e do reconhecimento territorial no Distrito Federal, nos termos da aplicação práticas de saberes, e está direcionada temporalmente ao período posterior à implantação da nova capital, Brasília, ao fim dos anos 1950, do século XX.

Fundamentalmente, os subconjuntos de dados de pesquisa, considerados nos termos de um processo de gestão que contempla arquivamento, disseminação e reuso, constituem na sua origem coleções de dados observacionais ou derivados (SAYÃO E SALES, 2016). Porém, nas pesquisas vinculadas ao Saberes do Cerrado, de caráter multidisciplinar, a dimensão espaço-temporal é determinante para a compreensão de práticas e saberes instituídos. Assim, realizam-se, como parte da metodologia adotada e com o apoio de um Sistema de Informações Geográficas (SIG), aproximações espaciais e correlações entre diversos registros de observação. Como resultado intermediário, as informações históricas analisadas puderam ser reproduzidas em mapeamentos, que traduzem com maior facilidade relações complexas.

As séries de dados analisadas para reconstituir historicamente práticas e saberes no DF pertencem a uma problemática territorial, que é da relação entre urbanização progressiva e incorporação de extensas áreas à produção intensiva agrícola, com perda da cobertura vegetal e fragmentação do bioma Cerrado no DF, de forma irreversível. Nos anos 1970, período do incremento agrícola sobre os cerrados, ainda persistia uma perspectiva de que fosse dada alguma atenção seletiva para a ocupação regional. AB'Saber, no entanto, já explicitava que o limite tolerável, para garantir a preservação do patrimônio natural, ao fim do ano de 2000, já

tinha sido superado pela “devastação antrópica” que afetou a área nuclear dos domínios paisagísticos dos cerrados (AB’SABER, 2003, p. 43).

O quadro desfavorável para a qualidade de vida urbana no DF é patente. Em 2017, o governo local implantou procedimentos de racionamento do abastecimento de água para a totalidade da população. A interrogação sobre como se instituem as relações entre cidade e Cerrado no período da implantação da nova Capital, e nos períodos subsequentes, conduz as linhas de investigação do Projeto Saberes do Cerrado, do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT).

A seleção das fontes e os resultados dos mapeamentos alcançados demonstram que as metodologias utilizadas, apoiadas por SIG, trouxeram visibilidade a questões que tendem a ficar restritas aos níveis especializados de abordagem ecológica, botânica ou urbanística. A adoção da plataforma digital Dataverse, que visa implementar a gestão de dados de pesquisa, viabiliza o processo de sistematização das fontes e de sua análise, além de tornar mais explícitos os procedimentos para a continuidade ou avanços da própria investigação, com implicações de controle sobre licenciamentos de origem, de citação e de reuso dos dados por terceiros.

2. Plataformas digitais e gestão de dados de pesquisa

2.1 O repositório Dataverse

O Dataverse é uma plataforma tecnológica de código aberto para compartilhar, encontrar e preservar dados de pesquisa. Ela serve para conectar trabalhos publicados e periódicos com conjunto de dados de pesquisas. Suas funcionalidades de coleta e de exportação de metadados permitem a circulação de grandes quantidades de dados, expandindo seu arquivamento e preservação (KING, 2007).

Uma solução tecnológica como o Dataverse, voltada para a gestão de dados científicos deve conter entre suas funções a atribuição de identificadores persistentes e possuir interface multilíngue para sua comunidade de usuários. Pesquisadores, curadores e instituições representados na plataforma precisam de orientações básicas para seu uso. A estrutura hierárquica das coleções deve ser definida por uma política de gestão de dados de pesquisa que descreva as áreas, projetos e o tipo de materiais que o sistema contemplará. O Dataverse é um sistema de informação e de controle de dados de pesquisa.

Por outro lado, o controle do acesso, deve ser aberto procurando o compartilhamento benéfico das informações. A confiabilidade que se deseja alcançar por parte dos usuários não parte do uso que ele vai dar aos dados, mas pela certeza de que a origem e permanência desses dados é responsabilidade de uma instituição comprometida com a pesquisa científica.

Pode acontecer a necessidade de uso opcional de formulários para acesso aos materiais sensíveis, os quais mais tarde permitirão aplicação de ferramentas para análise de uso.

Após criar seus próprios “Dataverses” a gestão dos depósitos focaliza nos formatos e na descrição dos conjuntos de dados e os arquivos, com a finalidade de que eles sejam facilmente recuperáveis e reusáveis. Inclusive os arquivos de dados tabulares, cuja exploração no Dataverse deve ser especificada.

Um dos aspectos mais relevantes do uso dos repositórios de dados de pesquisa é a sua capacidade de incluir metadados de citação para cada conjunto de dados submetido. Os metadados de citação do Dataverse para todos os conjuntos de dados são compatíveis com DataCite, Dublin Core, DDI (Data Documentation Initiative) (ALTMAN; CROSAS, 2013).

As instituições usuárias do Dataverse podem criar suas nuvens de armazenamento, com Harvard que usa a nuvem aberta de Massachusetts da Universidade de Boston. A integração com sistemas de preservação é uma das funcionalidades que o Dataverse permite para garantir a preservação dos conjuntos de dados. A política de gestão de dados do repositório deverá garantir a disponibilidade de uma infraestrutura mínima para o repositório de pesquisa, propondo também que sejam aplicadas medidas de segurança e de proteção da privacidade.

2.2. A vida dos dados de pesquisa

A quantidade de dados de pesquisa que foi perdida até hoje e a que está em risco de desaparecer é incalculável. Os dados de centenas de estudos guardados em disquetes, fitas VHS, fitas cassetes, CD e DVD estarão inutilizados se não houver uma ação de resgate por parte das instituições responsáveis pela memória da ciência no país. O fato de alguns desses dados já estarem na Web não garante sua continuidade e integridade, é inegável que as URL deixam de existir por vários motivos (VINES *et al.* 2014)

Parte do problema são também os formatos de arquivos proprietários (SPSS, Excel, SAS, Word, Stata, ASCII, Access, Matlab, VHS, File Marker Pro, JPG, etc.). Assim como as mídias de armazenamento usadas comumente pelos pesquisadores, sendo os hard drives pessoais os mais comuns, e os mais fáceis de perder. O problema não é só perder os arquivos pessoais, mas perder aqueles dos quais somos institucionalmente responsáveis.

Nos Estados Unidos a National Digital Information Infrastructure and Preservation Program (NDIIPP) mostrou que naquele país a maior parte dos dados coletados em pesquisas financiadas pelo governo não têm sido arquivadas de forma responsável (LEE, 2012). Muitos dados foram perdidos sem nenhuma chance de recuperação. Quando os dados não são

arquivados de forma rápida tornam-se um desafio para sua preservação e acesso, a menos que seja feito um planejamento inicial o risco para o avanço da ciência é grande.

Fazer um plano de gestão de dados no final do ciclo, quando o projeto terminou leva ao fracasso sua preservação. Uma das mais importantes mudanças é o estabelecimento de uma política que obrigue ao compartilhamento dos dados (STRASSER *et al.* 2012). Mas o maior desafio é fazer a comunidade científica pensar no alto valor da preservação.

As instituições de pesquisa têm a responsabilidade de comunicar à comunidade científica sobre as ações que devem ser tomadas desde o início do ciclo de vida dos dados. A primeira tarefa é a avaliação de que os materiais são publicáveis, úteis e se as expensas na sua preservação a longo prazo são justificáveis.

O registro científico dos dados brutos da pesquisa precisa estar voltado para sua criação e reprodução. Os dados empíricos, estatísticos e computacionais são centrais para o método científico e a busca pela ausência de erro. A simulação de dados em grande escala movimenta a ciência computacional atual, reunindo um conjunto de instruções e de dados que determinarão parte dos resultados. A integridade na pesquisa é garantida pela quantidade de informação necessária para o conhecimento pessoal. Um dos componentes dessa garantia de integridade são os metadados de preservação, os quais proveem informação para dar suporte ao processo de permanência dos registros.

Um sistema de gestão de dados de pesquisa garante que esses dados, sua codificação e fluxo poderão estar disponíveis e serem citados. Sua documentação é um dos princípios que destaca a reprodutividade dos dados (STODDEN, 2010). Mas os pesquisadores são responsáveis por decidir quando e em quais termos os dados de pesquisa podem ser acessados. Eles devem produzir seus planos de gestão de dados para cada projeto de pesquisa que gerará um conjunto de dados. Cada vez mais os projetos científicos se tornam intensivamente computacionais e por isso mais transparentes. As opções de customização das interfaces devem atender às demandas de cada projeto, descrevendo as especificações de cada *template*, conjunto de dados e metadados (STRASSER, 2015).

Outro aspecto que envolve a preservação dos dados de pesquisa é incluir na política de gestão institucional que existem mecanismos para garantir que os dados com valor de uso a longo prazo estejam depositados num repositório de dados para seu arquivamento, tendo em conta que o papel do pesquisador é transitório em comparação com o papel atribuído às instituições.

2.3 O serviço de preservação de dados de pesquisa do IBICT

Em países como os Estados Unidos, existem instituições que armazenam arquivos de dados de pesquisa a várias décadas. É o caso do Inter-university Consortium for Political and Social Research (ICPSR) da University of Michigan em funcionamento desde 1960 e que é considerado o maior arquivo de dados das ciências sociais do mundo. Na University of North Carolina o Odum Institute mantém o arquivo mais antigo de dados das ciências sociais fundado em 1920. E também The National Archives and Records Administration lidera as iniciativas de preservação de acervos de dados especializados governamentais. Na Universidade de Harvard existe a primeira aliança nacional de preservação de dados de pesquisa, Data-PASS, especializada no desenvolvimento de soluções e políticas de gestão de dados científicos (ALTMAN *et al.* 2009).

Na América Latina existem instituições emergentes que requerem dos seus pesquisadores que seus dados sejam publicados em repositórios abertos, caso esses projetos recebam algum tipo de financiamento. No Brasil a agência de financiamento do estado de São Paulo, a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pede aos pesquisadores que solicitam auxílio financeiro pelo fundo do programa E-Science que apresentem seus planos de gestão de dados de pesquisa (FAPESP, 2017). Mas ainda falta em iniciativas como essa a integração do processo de preservação em todo o ciclo de vida dos dados. Qualquer custo de armazenamento e gestão deve ser explicitado nos pedidos de financiamento, incluindo o local onde os dados serão disponibilizados e aplicadas as técnicas de curadoria, por anos após do término do projeto.

O Ibict é uma instituição de pesquisa da América Latina que realiza esforços para contribuir com o compartilhamento de dados e seu reuso. Essas atividades se iniciaram no Instituto com a implementação da Biblioteca Digital de Teses e Dissertações, a segunda maior biblioteca digital de seu tipo, e com a promoção da plataforma Open Journal Systems (OJS), sendo atualmente o Brasil o país com maior número de instalações dessa plataforma.

Em 2015 iniciaram-se os trabalhos para a criação de um serviço de preservação de dados na rede Cariniana. Os estudos realizados confirmaram que o Dataverse prioriza e reforça as propriedades de preservação que devem ser verificadas no funcionamento de um sistema de informação que inclui ações de preservação digital, são provas de autenticidade e de integridade dos dados nos processos de comunicação. A partir dessa comprovação as atividades de suporte à preservação de dados de pesquisa nas universidades parceiras da Rede começaram a ser estruturadas.

A tradução do *software* para o português foi uma iniciativa da coordenação da Rede Cariniana no Ibict para facilitar o acompanhamento das ações de curadoria e do cumprimento da política de gestão da instituição. O fluxo da curadoria de dados inicia-se na observância dos

padrões de depósito estabelecidos pela equipe responsável. Segundo o modelo de ciclo de vida da curadoria Digital Curation Centre (DCC), o rigor na submissão permite o controle das informações sobre os conjuntos de dados, seria uma auto curadoria prévia ao uso dos repositórios de dados. A Rede Cariniana ainda atua na produção de guias de usuário, desenvolvimento de tutoriais com sugestões sobre o preparo do material a ser submetido, fluxo da informação e melhores práticas (MÁRDERO ARELLANO; OLIVEIRA, 2016)

O projeto do serviço de preservação digital de dados de pesquisa da Cariniana está direcionado à replicação de cópias dos arquivos dos dados armazenados nos repositórios das instituições parceiras. Como uma atividade de preservação distribuída, a guarda segura de pelo menos quatro cópias dos documentos digitais em instituições geograficamente distantes faz o sistema de arquivamento como o LOCKSS (Lots of Copies Keep Stuff Safe) funcionar em rede (MANIATIS *et al.* 2005).

A rede Cariniana adotou o sistema LOCKSS como um processo ativo de preservação onde cópias são validadas automaticamente em diferentes locais. Um primeiro desafio a ser enfrentado é a assimetria do tamanho dos arquivos, sendo que algumas instituições guardaram acervos de dados maiores dos que elas podem chegar a produzir. O segundo desafio é o fato de lidar com material confidencial ou sigiloso, que pode ser crucial para a pesquisa, mas não podem ser compartilhados na Internet. Esse tipo de replicação deve ser regulado para proteger a confidencialidade dos dados. A preservação digital distribuída de dados de pesquisa conta com instrumentos de segurança no armazenamento compartilhado que permita a replicação segura dos arquivos.

No serviço de preservação de dados de pesquisa a adoção de um sistema consorciado e distribuído de arquivamento parte da ideia de que nenhuma instituição pode manter e preservar isoladamente seus materiais por longos períodos de tempo. As parcerias da Rede são de colaboração técnica e científica, que seguem normas e padrões internacionais de seleção, aquisição, arquivamento e acesso compartilhados.

3. Repositório de dados e processos de investigação

3.1. O caso do projeto Saberes do Cerrado

Se a cultura digital propicia a disponibilidade de informações de forma massiva, já a oferta de dados qualificados e representativos no Brasil tem sido objeto seletivo de políticas públicas. A determinação de maior amplitude de organização e acesso a dados de coleções científicas tem tido destaque para algumas áreas do conhecimento, atendendo a compromissos internacionais vinculados à Convenção sobre Diversidade Biológica (CDB), de 1992. Uma infraestrutura informatizada complexa foi criada para as áreas relacionadas à

biodiversidade, com aproximadamente seis milhões de registros (MCTIC, 2017). A valorização de coleções científicas como documentos preservados da flora brasileira e sua disponibilização internacionalizada pelos serviços informatizados em rede é parte constitutiva dessas ações (PEIXOTO; MORIM, 2003). Por outro lado, outros projetos, no campo do patrimônio cultural, têm disponibilizado de forma integrada acervos fotográficos de instituições diferenciadas, públicas e privadas, porém, recriando coleções de forma seletiva para visualização conjunta, com 4409 itens (BRASILIANA, 2017). Outro relevante exemplo é o da Biblioteca Nacional Digital, com 1.504.339 registros, que avança na disponibilização do extenso universo iconográfico, situação bem distinta do acesso limitado aos materiais originais no contexto anterior ao domínio digital (BIBLIOTECA NACIONAL. BNDigital, 2017). O que estes projetos têm em comum? O caráter institucional e a avaliação certificada por procedimentos internos de curadoria digital de documentação com acesso livre.

Estes rápidos exemplos de porte nacional evidenciam que disponibilizar conteúdos digitais relacionados à memória científica e cultural exige um conjunto de pré-requisitos e criação de políticas próprias de gestão. Enquanto o universo documental digital de acervos públicos, ou de relevância pública, está inscrito em normativas do CONARQ (CONARQ, 2005), quando o centro da questão envolve os dados de pesquisa, financiados com recursos públicos, o quadro é distinto. Inexistem políticas instituídas para a estruturação de ações, que necessitam ser determinativas, apesar da relevância da gestão de dados para a produção científica nas suas diferentes escalas (SAYÃO, SALES, 2016).

Na atualidade, apesar do movimento internacional de dados abertos, a dimensão de arquivamento ou a disponibilização de dados de pesquisa estão sujeitas às prerrogativas criadas por instituições financiadoras e poucas indicações de obrigatoriedade externas estão colocadas, a exemplo de proposição europeia no Programa Horizonte 2020 e, no Brasil, o caso recente da Fapesp que adotou a exigência de Plano de Gestão de dados, já referido (VIDOTTI *et al*, 2017).

Dessa forma, além das condições técnicas e operacionais oferecidas por serviços de arquivamento digital e das configurações de padronização de metadados para adequada recuperação de conteúdos, a gestão pública de dados de pesquisa depende também das condições definidas previamente quanto às licenças de utilização atribuídas pelos originários detentores da propriedade intelectual de acervos.

Na atenção à elaboração de políticas para gestão de dados de pesquisa cabe a distinção entre a proposição de arquivamento, em que bastaria o interesse declarado pela memória institucional da entidade ou unidade de pesquisa, e a segunda opção, conjugada ou não, ao reuso dos dados científicos, em que se preconiza o acesso aberto.

Estas questões estão colocadas na presente proposta exploratória de estruturação de uma plataforma digital, que tem por objetivo a preservação de dados pesquisados e a possibilidade de reutilização por terceiros dos materiais processados em SIG do Projeto Saberes do Cerrado. Tem-se em conta o problema da obsolescência tecnológica e dos custos inerentes aos recursos humanos necessários para a manutenção e disseminação na INTERNET dos resultados criados no referido sistema. Mas a perspectiva da preservação digital em longo prazo ganha preponderância não só pela memória institucional, mas pelo conjunto de fontes agregadas e resultados analíticos alcançados que remetem à situação ambiental de vulnerabilidade dos remanescentes do Cerrado e de áreas protegidas no DF.

As séries de fontes pesquisadas que, no quadro da investigação de caráter histórico e urbanístico, associado ao reconhecimento territorial, tornam-se fontes primárias, são compostas, fundamentalmente, por registros textuais, registros iconográficos, registros cartográficos, dados populacionais, dados de coleta botânica e dados de focos de calor. Essa diversidade de dados, como vestígios de uma determinada temporalidade, torna-se fundamental para a compreensão do objeto da investigação e não é evidentemente arbitrária. Ela corresponde às questões previamente colocadas sobre dinâmicas espaço – temporais, que se originam da situação crítica ambiental vivenciada no presente.

Como há longo tempo ensinou Marc Bloch (1974), nossas questões sobre o passado invariavelmente passam pelas condições do presente e por nossa capacidade de neste reconhecer problemas, e, assim, inversamente olhar a transformação histórica.

Outro aspecto a considerar, desse ponto de vista, é que a relação que as sociedades estabelecem com natureza se altera no tempo, assim como também a percepção que dela se faz não é unitária, transita entre as condições que se criaram para a vida em sociedade (THOMAS, 2010). O caso de Brasília, como cidade-parque, instaurada, segundo a narrativa dominante sobre um território vazio, merece que se dê mais atenção às questões que envolvem a preservação da biodiversidade do bioma Cerrado.

A relação que se pretende instituir, entre o projeto Saberes do Cerrado, em suas diferentes linhas de atuação, e o repositório de dados, é de espelhamento dos processos de investigação. As fontes de pesquisa que serão organizadas, em conjuntos temáticos, advêm, em geral, de grandes bases de dados nacionais anteriormente referidas. Explorar essas bases digitais, em paralelo a acervos locais, para dar conta do universo urbano e suas representações espaciais no DF, se deve ao objetivo de reconstituir relações que dizem respeito às práticas e saberes aplicados no passado relativamente recente. De outra forma, contudo, esse percurso da investigação teria sido inviável.

O campo de atividades do projeto, que envolve uso de serviços informatizados para apoio analítico à investigação, depende fundamentalmente de dados digitais, ou objetos digitais, em suportes tecnológicos diferenciados. Uns, arquivados eletronicamente em computadores do Arquivo Público do Distrito Federal, derivam da digitalização de imagens fotográficas, que demandam consulta local e reprodução autorizada. Outros, disponíveis em serviços na internet, representam de forma eletrônica dados de coleções científicas físicas e seus catálogos sistemáticos, pertencentes a inúmeras instituições; ou, parte de acervos disponíveis em *sites* governamentais, que podem, ou não, estar disponíveis no futuro, como os relevantes mosaicos de séries temporais de imagens aéreas do Sistema Cartográfico do DF (SICAD), resultado de monitoramento sistemático, durante décadas, sobre a área de implantação de Brasília (DISTRITO FEDERAL, 2017).

Esse quadro introduz um questionamento sobre a natureza das fontes de informação. Sendo em sua maioria dados de observação, disponíveis em plataformas de acesso aberto, e considerando que haja, em cada instituição detentora, uma responsabilidade presumida de manter indefinidamente esses materiais, qual a razão de se promover a experimentação de preservação digital associada a este projeto? A resposta é a possibilidade de garantia de acesso no longo prazo aos dados digitais referidos diante da potencial ausência de alguns dos elementos que traçam a complexidade das relações estudadas. Fatores como a obsolescência tecnológica, mudanças nos aparatos institucionais de compartilhamento de acervos ou perda acidental de materiais digitais podem introduzir quadros de incerteza sobre a reprodução futura dessas séries. Sendo a indicação precisa das fontes um dos requisitos da prática historiográfica, a sua disponibilidade integrada às análises efetuadas e aos resultados alcançados tornam mais nítidos os processos de investigação. Ou, nos termos de Sales e Sayão (2012), cria-se um meio digital de aferição ou reuso, pela natureza da publicação ampliada.

3.2 O reuso, sistematização e reclassificação de objetos digitais

A estrutura proposta para o Dataverse do projeto Saberes do Cerrado envolve procedimentos inerentes aos fluxos da investigação e à necessidade de agregar o material coletado, sem perda das situações de sistematização e classificação originárias.

Ordenar correlações entre registros e observar os processos de requalificação de informações são etapas relevantes para a criação de procedimentos de análise. Neste caso, mais do que incorporar procedimentos *a posteriori*, pretende-se criar um instrumental experimental de gestão de dados de pesquisa, tendo em vista o arquivamento dos objetos digitais, independente da sua unidade ou complexidade interna, visando o potencial reuso futuro de dados.

O acesso progressivo a bases digitais dá margem a que a análise sobre conjuntos seriais se dê sobre um volume maior de registros. Essa amplitude, no caso dos dados de coleções científicas botânicas ou de registros visuais, pertencentes à documentação urbana de Brasília, garante maior controle sobre processos comparativos, temporais ou espaciais, mas gera dificuldades para sua sistematização analítica.

Tendo em conta as diversas vertentes de saberes aplicados, o botânico e o urbanístico, que hipoteticamente se encontram em contraposição, estão em debate dois níveis de abordagem: o primeiro, da construção social de uma memória urbana referente ao tempo histórico da análise, e o da segunda instância representada na retomada interpretativa desses materiais e sua nova disseminação e circulação como objetos digitais. Portanto, do ponto de vista da estrutura do sistema de arquivamento no Dataverse, importa deixar bem evidente essa diferenciação de temporalidades e de reusos dos registros.

Mas, por outro lado, como considerar para a história da ocupação do Distrito Federal a massa de dados automaticamente produzida por sensoriamento remoto dos dados de calor da plataforma de queimadas do INPE? Ou como tratar o mosaico de imagens aéreas disponíveis no sistema SICAD? Os dados originários desses serviços disponíveis publicamente correspondem a atuações de monitoramento por imagem, realizados estrategicamente pelo Estado, e envolvem, portanto, outras sistemáticas e técnicas de classificação relacionadas à racionalidade instrumental e suas aplicações (INPE, 2017; DISTRITO FEDERAL, 2017).

A vertente de entendimento de uma exomemória, como todos os registros disponíveis externos à capacidade individual de apropriação, tal como justificada por Gutiérrez (2003), é conceitualmente útil para considerar esse universo de dados à disposição, assim como o próprio serviço de arquivamento objeto deste artigo.

Nesse conjunto de vestígios, que foram enquadrados por suas práticas aplicadas ao reconhecimento territorial, serão identificados os elementos da paisagem e as possíveis indicações valorativas atribuídas à ordem urbana e ao meio ambiente do Cerrado. Considerando as correlações mais aprofundadas entre esses registros, cabe enfatizar a associação entre práticas botânicas e as informações de natureza representativa constituída por ilustrações científicas. A análise combinada dessas séries, com processo seletivo de dados, permitiu superar a incompletude da dimensão espacial (localização) relacionada às coleções botânicas formadoras de acervos institucionais do Distrito Federal.

O problema da identificação de áreas de coleta por uso do mapeamento temático, em SIG, diz respeito às atividades desenvolvidas por Ezechias Paulo Heringer. Suas interações com os taxonomistas Guido Pabst e Pedro Carauta, além do contato com a ilustradora botânica Maria Werneck de Castro, na década de 1960, constituem atividades de caráter

interdependente e associadas às práticas instituídas e sistemáticas do fazer botânico. Essas informações foram estrategicamente utilizadas para extrair conjuntos de dados do volume de mais de 15.000 registros de exsicatas coletadas por Ezechias P. Heringer no DF (INCT, 2017; CASTRO, 2004). O processo de requalificação de informações vinculadas à biodiversidade, restrito até agora ao SIG, pode ser explicitado e continuado na proposição do modelo experimental de gestão de dados de pesquisa, incluindo novas ferramentas de análise, como a mineração de dados.

Um segundo ciclo de vida do material processado está dessa forma configurado em níveis diferenciados de análise, tendo em conta as classificações originárias. Um conjunto mínimo de metadados foi definido para constituir a descrição básica dos itens, em que é essencial a dimensão espacial para qualificar os resultados elaborados em cartas temáticas, além da proveniência e licenciamento de uso de respectivas fontes.

No limite deste trabalho, destaca-se que, exceto os materiais vinculados às coleções científicas e imagens fotográficas, os demais registros referem-se a um território de estudo previamente definido, que privilegiou a envolvimento do Jardim Botânico de Brasília e sua Estação Ecológica. Os exemplos apresentados a seguir de cartas temáticas, elaboradas para o projeto, mostram o resultado intermediário de análise realizado com dados populacionais na Figura 1 e com dados do coletor botânico Ezechias P. Heringer na Figura 2. Essas cartas temáticas são relativamente autoexplicativas por conterem indicações de suas formas de representação, além de categorias utilizadas. Mas ganham outra dimensão se a lógica de sua construção estiver plenamente explicitada, algo possível em uma base de arquivamento de dados de pesquisa como o Dataverse.

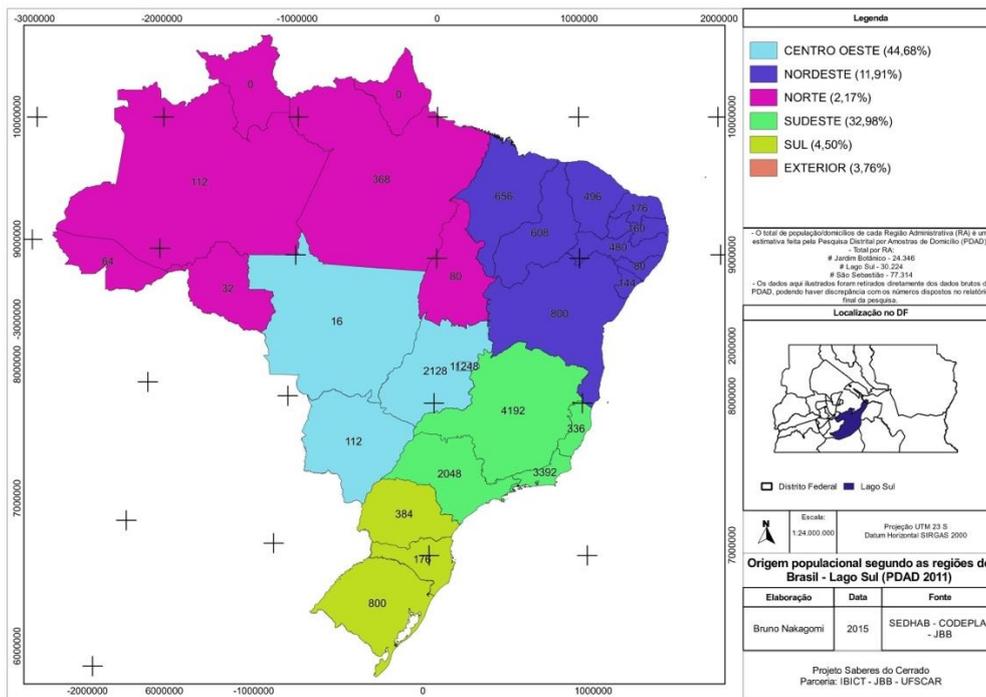


Figura 1. Origem populacional segundo as regiões do Brasil - Lago Sul (PDAD 2011).
 Fonte: IBICT. Projeto Saberes do Cerrado. Elaboração cartográfica Bruno Nakagomi, 2015.

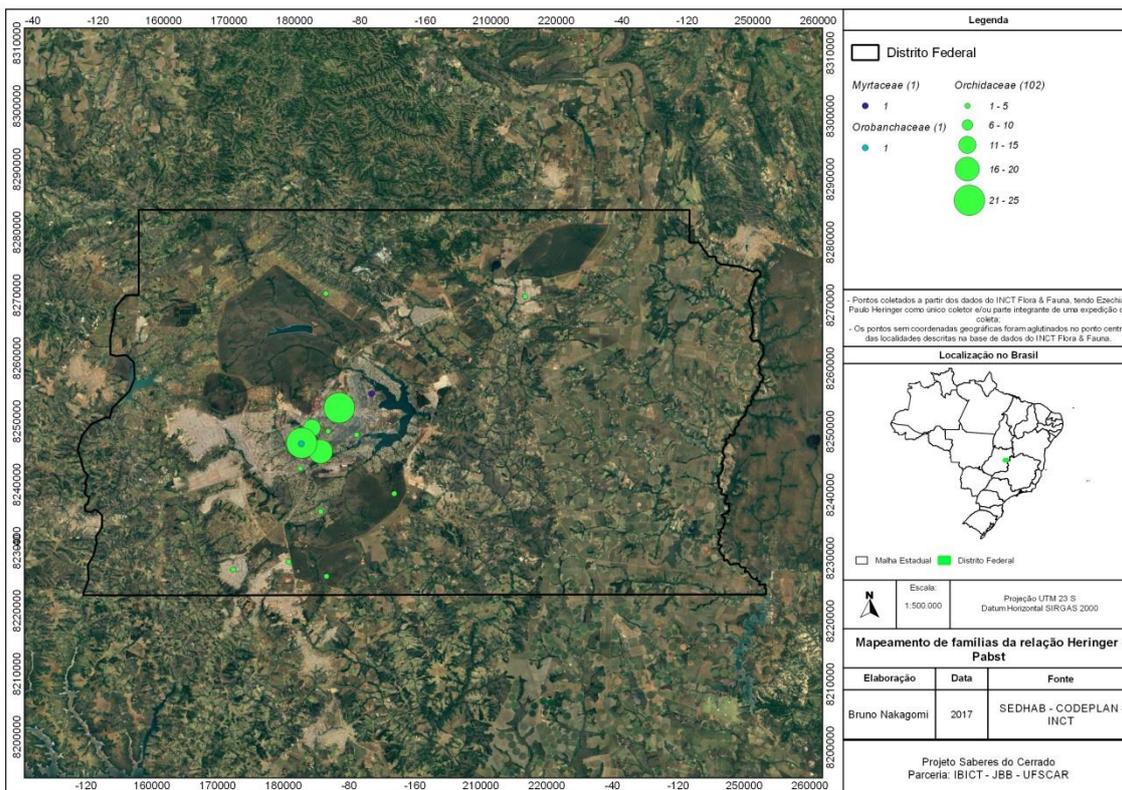


Fig. 2. Mapeamento de famílias da relação E. P. Heringer e Guido Pabst
 Fonte: IBICT. Projeto Saberes do Cerrado. Elaboração cartográfica Bruno Nakagomi, 2017. A partir de dados do INCT- Flora e Fungos e de imagem do Google Maps.

4. Considerações finais

Pode-se concluir que a proposição de um modelo experimental de gestão de dados de pesquisa para atender campos de atuação multidisciplinares necessita estabelecer pré-condições: a possibilidade de representação adequada de descrição de itens seriais, considerando a classificação originária de objetos e seus agrupamentos tipológicos em acervos originais; traduzir a classificação do conjunto documental com suficiente atenção à diversidade de áreas de conhecimento; identificação dos objetos digitais de forma a estabelecer múltiplas interações e correspondências entre itens; e atenção às soluções adequadas para a questão do licenciamento de direitos de propriedade intelectual ou de possibilidades de reuso com acesso aberto.

A proposição de um repositório de dados de pesquisa Dataverse, no âmbito das linhas de investigação do Projeto Saberes do Cerrado, contempla as possibilidades de arquivamento, disseminação e reuso de dados e de materiais abrigados até o presente em um Sistema de Informações Geográficas. Dessa forma, é incorporada, ao mesmo tempo, a dimensão da memória institucional e a outra face, que é a de instaurar mais uma experiência de acesso aberto do IBICT, na perspectiva de contribuir para o reuso de informações sistematizadas que dizem respeito à história da ocupação territorial do DF e sua relação com o bioma Cerrado.

Referências

AB'SABER, Aziz. **Os domínios de natureza no Brasil: potencialidades paisagísticas**. São Paulo: Ateliê Editorial, 2003.

ALTMAN, M.: CROSAS, M. The Evolution of Data Citation: From Principles to Implementation. **IASSIST Quarterly**, 2013. Disponível em http://www.iassistdata.org/sites/default/files/iqvol371_4_altman.pdf Acesso em: 15 nov. 2017.

ALTMAN, Micah *et al.* Digital Preservation through Archival Collaboration: The Data Preservation Alliance for the Social Sciences. **American Archivist**, v. 72, n. 1, p. 170-184, Spring/Summer, 2009. Disponível em: <https://doi.org/10.17723/aarc.72.1.eu7252lhnrp7h188> Acesso em: 13 nov. 2017.

BLOCH, Marc. **Introdução à história**. 2ª. ed. Lisboa: Publicações Europa-América, 1974.

BIBLIOTECA NACIONAL (Brasil). BNDigital. **Acervo digital**. Disponível em: <http://bndigital.bn.gov.br/acervodigital/> Acesso em: 19 nov. 2017.

BIBLIOTECA NACIONAL (Brasil). **Brasiana Fotográfica Digital**. Disponível em: <http://brasianafotografica.bn.br/brasiana/> Acesso em: 20 nov. 2017.

BRASIL. MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES. Sistema de Informação sobre a Biodiversidade Brasileira. **Explorar dados**. Disponível em: <http://www.sibbr.gov.br/> Acesso em: 19 nov. 2017.

CASTRO, Maria Werneck. **Natureza viva: memórias, carreira e obra de uma pioneira do desenho científico no Brasil**. Rio de Janeiro: Edições Biblioteca Nacional, 2004.

CONSELHO NACIONAL DE ARQUIVOS (CONARQ). **Carta para a preservação do patrimônio arquivístico digital: preservar para garantir o acesso**. 2005. Disponível em: http://conarq.arquivonacional.gov.br/images/publicacoes_textos/Carta_preservacao.pdf
Acesso em: 20 nov. 2017.

DISTRITO FEDERAL. SECRETARIA DE ESTADO DE GESTÃO DO TERRITÓRIO E HABITAÇÃO. **Sistema cartográfico do DF (SICAD)**. Disponível em: <http://www.sedhab.df.gov.br/sicad.html>
Acesso em: 20 nov. 2017.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO (FAPESP). **Plano de Gestão de Dados FAPESP**. Agencia Notícias, 2017. Disponível em: <http://www.fapesp.br/gestaodedados/> Acesso em: 17 nov. 2017.

GUTIÉRREZ, Antonio García Redes digitales y exomemoria. **Revista Científica de Información y Comunicación IC**, n. 1, 2003. Disponível em: <http://icjournal-ojs.org/index.php/IC-Journal/article/view/140/38> Acesso em: 9 nov. 2017.

INCT Herbário Virtual da Flora e Fungos. Disponível em: <http://inct.splink.org.br/> Acesso em: 8 mar.2017

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). Programa de queimadas: monitoramento por satélites. Disponível em: http://www.inpe.br/queimadas/estatistica_estados Acesso em: 2 mai.2017.

KING, G. An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. **Sociological Methods and Research**, v. 36, n. 2, p. 173-199, November 2007. Disponível em: <http://gking.harvard.edu/files/gking/files/dvn.pdf> Acesso em: 14 nov. 2017.

LEE, Christopher A. **States of Sustainability: A Review of State Projects funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP)**. Chapel Hill, NC: University of North Carolina, 2012. Disponível em: http://www.digitalpreservation.gov/multimedia/documents/ndiipp-states-report032612_final.pdf Acesso em 10 nov. 2017.

MANIATIS, P. *et al.* The LOCKSS Peer-to-Peer Digital Preservation System. **ACM Transactions on Computer Systems**, v. 23, n. 1, p. 2–50, February 2005. Disponível em: DOI 10.1145/1047915.1047917 Acesso em 10 nov. 2017.

MÁRDERO ARELLANO, M. A.; OLIVEIRA, A. Faria de. Scientific Data Management on a Dataverse Network at IBICT. In: **LEARN Toolkit of Best Practice for Research Data Management**. p. 82-86, 2016. Disponível em: <http://discovery.ucl.ac.uk/1546580/> Acesso em 30 out. 2017.

PEIXOTO, Ariane Luna; MORIM, Marli Pires. Coleções botânicas: documentação da biodiversidade brasileira. **Ciência e Cultura**, São Paulo, v. 55, n. 3, Julho-setembro 2003. Disponível em: <http://cienciaecultura.bvs.br/pdf/cic/v55n3/a16v55n3.pdf> Acesso em: 20 nov. 2017.

SANTOS, Milton. **A urbanização brasileira**. São Paulo: Edusp, 2005.

SALES, Luana Faria; SAYÃO, Luís Fernando. O impacto da curadoria digital dos dados de pesquisa na comunicação científica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 17, n. esp. 2 – III SBCC, p. 118-135, 2012. Disponível em: <http://dx.doi.org/10.5007/1518-2924.2012v17nesp2p118> Acesso em: 14 nov.2017.

SAYÃO, Luís Fernando. Uma outra face dos metadados: informações para a gestão da preservação digital. **Encontros Bibli: Revista eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 15, n. 30, p.1-31, 2010 Disponível em: <http://dx.doi.org/10.5007/1518-2924.2010v15n30p1> Acesso em: 9 nov.2017.

SAYÃO, Luís Fernando; SALES, Luana Faria. Algumas considerações sobre os repositórios digitais de dados de pesquisa. **Informação & Informação**, Londrina, v. 21, n. 2, p. 90 – 115, maio/ago., 2016. DOI:10.5433/1981-8920.2016v21n2p90 Disponível em: <http://www.uel.br/revistas/informacao/> Acesso em: 8 nov.2017.

STODDEN, Victoria. **The Scientific Method in Practice: Reproducibility in the Computational Sciences**. Boston, MA: MIT Sloan Research Paper No. 4773-10, 2010. Disponível em: <http://dx.doi.org/10.2139/ssrn.1550193> Acesso em: 14 nov. 2017.

STRASSER, Carly *et al.* **Primer on data management: what you always wanted to know**. University of California, Office of the President, California Digital Library, 2012. (CDL Staff Publications). Disponível em: <http://escholarship.org/uc/item/7tf5q7n3> Acesso em 18 nov. 2017.

STRASSER, Carly. **Research data management: a primer publication of the national information standards organization**. Baltimore: NISO, 2015. (NISO Premier Series). Disponível em <http://wiki.lib.sun.ac.za/images/2/24/PrimerRDM-2015-0727.pdf> Acesso em: 14 nov. 2017.

THOMAS, Keith. **O homem e mundo natural: mudanças de atitude em relação às plantas e aos animais (1500-1800)**. São Paulo: Companhia das Letras, 2010.

VINES, Timothy *et al.* The availability of research data declines rapidly with article age. **Current Biology**, v. 24, p. 94-97, 2014. Disponível em: doi: 10.1016/j.cub.2013.11.014 Acesso em: 14 nov. 2017.

VIDOTTI, Silvana Aparecida Borseti Gregório *et al.* Repositório de dados de pesquisa para grupo de pesquisa: um projeto piloto. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18, 2017. **Anais...** Marília, SP: UNESP, 2017.

Recebido/Recibido/Received: 2017-11-22
Aceitado/Aceptado/Accepted: 2017-12-04