

ARQUIVAMENTO E DISTRIBUIÇÃO DE CONJUNTOS DE DADOS GEOESPACIAIS

RAIAN VARGAS MARETTO¹

*¹Divisão de Processamento de Imagens, DPI
Instituto Nacional de Pesquisas Espaciais, INPE
Av. dos Astronautas, 1758 – 12227-010 – São José dos Campos – SP – Brasil
raian@dpi.inpe.br*

Resumo: Com o avanço nas diversas áreas de estudo de fenômenos espaciais, há uma produção cada vez maior de dados geoespaciais digitais, tais como provenientes de Sensoriamento Remoto e de processamento em Sistemas de Informação Geográfica, dentre outros tipos. Com o passar dos anos, estes dados tornam-se cada vez mais valiosos, sendo de grande importância no estudo de fenômenos dinâmicos. Isto torna necessário o arquivamento dos mesmos, assim como sua distribuição, de forma a evitar a perda de dados e torná-los mais acessíveis para as diversas áreas de pesquisa.

1. Introdução

Dados geoespaciais digitais consistem de informações digitais que identificam a localização espacial e as características de fronteiras naturais ou artificiais na Terra. Incluem conjuntos de dados provenientes de Sistemas de Informação Geográfica (SIGs) e de Sensoriamento Remoto, tais como fotografias aéreas ou imagens de satélites, mapas digitalizados, dados tabulares de localidades específicas, e vários outros relacionados. Como muitos destes existem em formas complexas em que não são toleradas impreviões, a preservação em longo prazo envolve também a migração dos dados para formatos mais facilmente sustentáveis (MORRIS e TUTTLE, 2007).

Os avanços nas diversas áreas geradoras de dados geoespaciais fazem com que haja uma produção cada vez maior dos mesmos. Estes dados, com o passar do tempo tornam-se muito valiosos, devido à necessidade do acompanhamento em longo prazo de alguns tipos de fenômenos para que seja possível entender ecossistemas dinâmicos. Além disto, alguns destes dados identificam também fenômenos raros ocorrentes na natureza, de forma que sua perda pode ocasionar a impossibilidade do estudo destes fenômenos.

Uma importante questão no arquivamento e distribuição de dados geoespaciais é “como prover o acesso permanente” (DOORN e TJALSMA, 2007). Assim como os vários processos de arquivamento digital, o arquivamento de dados geoespaciais é fortemente

influenciado pelas dinâmicas da tecnologia da informação atual e suas rápidas mudanças, tornando-o totalmente vulnerável às obsolescências da tecnologia.

Este trabalho apresenta a importância do arquivamento de dados geoespaciais, os principais desafios enfrentados nesta tarefa e descreve dois exemplos de sistemas utilizados para este fim, citando assim suas principais características e objetivos.

2. A importância de se arquivar dados geoespaciais

A maioria das pesquisas científicas é baseada na aquisição e análise de dados provenientes de medições. Os conjuntos de dados, frequentemente combinados com modelos e parâmetros, tem-se tornado mais importantes e podem, às vezes, serem vistos como saída primária intelectual da pesquisa. Dessa forma, a publicação e a preservação dos mesmos devem ser cuidadosamente consideradas, especialmente nos casos em que estes dados não podem ser reproduzidos (p. ex., quando são originados de eventos únicos) e podem ser úteis no futuro (DEKKER, 2006).

Com o passar dos anos, grupos de cientistas vêm estudando, cada vez mais processos hidrológicos e ecológicos através do sensoriamento remoto óptico e de microondas, estudos estes que resultam em valiosos conjuntos de dados geoespaciais digitais. O grande valor dos mesmos é atribuído ao fato de oferecerem escalas espaciais variando desde medidas pontuais até coberturas regionais e escalas temporais variando desde imagens instantâneas a dados coletados em médio prazo e medidas contínuas coletadas em longo prazo. Estes dados possuem uma potencial utilidade no entendimento de ecossistemas dinâmicos, assim como em prognósticos sobre os mesmos (MORAN et al., 2008).

Os dados geoespaciais podem ser atuais ou históricos, e a grande quantidade deles que podem ser preservados é de grande valor na análise de mudanças de longo prazo nos ecossistemas, tais como a reconstituição de uma floresta (BLEAKLY, 2002).

Com o grande aumento no tamanho dos bancos de dados, para estes conjuntos de dados, mídias de armazenamento físico são impraticáveis, pois os requisitos de armazenamento excedem suas atuais capacidades. Outro ponto importante é a capacidade de se acessar

os dados através de múltiplas plataformas computacionais, com diferentes sistemas operacionais (ATKINSON e ISBELL, 2003). Assim, devido à grande capacidade de armazenamento de servidores online, e da facilidade de acesso através de qualquer plataforma computacional, com qualquer sistema operacional, o armazenamento dos conjuntos de dados em sistemas online torna-se uma alternativa muito mais viável. Outros pontos fortes dos sistemas online são: a possibilidade de acesso a partir de qualquer localidade; acesso rápido aos dados, provendo meios eficientes de distribuição dos mesmos.

3. Desafios enfrentados

Fica clara a importância de se arquivar e distribuir conjuntos de dados geoespaciais, para promover o acesso livre e aberto a eles, assim como evitar a perda de valiosos conjuntos de dados coletados em longo prazo. Nesta seção, são descritos os desafios enfrentados nesta tarefa.

Segundo Moran et al. (2009) os principais desafios enfrentados no desenvolvimento de sistemas para arquivamento e distribuição de dados geoespaciais são:

- Primeiramente, o arquivamento em um formato digital em um local seguro e com suficiente padronização dos metadados para assegurar que haja aplicações apropriadas no presente e no futuro. Isto requer um investimento substancial das unidades de pesquisa, havendo, inclusive uma grande emergência para se evitar a perda de dados.
- Promover o acesso completo e aberto com um sistema de distribuição online. Isto é também uma tarefa difícil, devido ao fato de que os conjuntos de dados geoespaciais estão comumente armazenados em diferentes mídias e hospedados em diferentes domínios, com uma variedade de restrições de compartilhamento, protocolos e metadados. É necessário que se institua um processo de planejamento formal para determinar quais dados devem ser distribuídos e para se planejar a distribuição dos devidos bancos de dados e a interface online.
- Integrar os diversos, porém próximos conjuntos de dados pela interconexão dos sistemas de distribuição online. Isto requer um grande esforço, que ultrapassa

as fronteiras federais, das universidades, centros de pesquisa e outros fornecedores de dados, assim como requer também o investimento de recursos por parte destes para um objetivo em comum, prover o máximo acesso para o maior grupo de usuários possível.

- Planejar a sustentabilidade dos sistemas de arquivamento e distribuição de dados, levando em conta o futuro incerto da tecnologia, da aplicação dos dados e dos recursos para manutenção dos sistemas. Esta sustentabilidade assegura a continuidade de dados geoespaciais coletados em longo prazo, agregando assim, grande valor a eles.

4. Sistemas de Arquivamento e distribuição de conjuntos de dados geoespaciais

O movimento de criação e manutenção dos sistemas de arquivamento e distribuição de dados geoespaciais tem sua motivação no anseio de tornar estes dados disponíveis para pesquisa e desenvolvimento e também no receio de que alguns deles, insubstituíveis, sejam perdidos ao longo do tempo (MORAN et al., 2009). Definida a importância de se arquivar conjuntos de dados geoespaciais e os desafios enfrentados nesta tarefa, esta seção descreve dois destes sistemas com suas principais características.

ARIA: Arquivo Regional de Imagens do Arizona

O sistema ARIA (do inglês *Arizona Regional Image Archiving*) é um sistema online para o arquivamento e a distribuição de imagens digitais e mapas da região do deserto de Sonoran, incluindo o sudoeste dos Estados Unidos e o Norte do México (<http://aria.arizona.edu>). Criado em 1997 na Universidade do Arizona, foi originalmente desenvolvido com o intuito de facilitar o compartilhamento de dados sobre a região e de evitar a aquisição duplicada de imagens de satélite, fotografias aéreas, modelos de elevação de terreno e mapas. Vem tendo um crescimento considerável, ultrapassando o volume de 2 TB (2 Terabytes) de dados (MORAN et al., 2009).

A seleção dos dados para inclusão no arquivo foi originalmente baseada nos conjuntos de dados disponíveis entre os pesquisadores da Universidade do Arizona e organizações locais, estaduais e federais na área. Após este conjunto inicial, foram adquiridos novos conjuntos de dados com base na disponibilidade dos mesmos e em uma avaliação de

necessidade. Em particular, foram adicionadas fotografias aéreas de alta resolução espacial devido à alta demanda, e mais recentemente, o ARIA foi atualizado com dados do WGEW (do inglês *Walnut Gulch Experimental Watershed*), assim como com dados proprietários, como imagens SPOT (MORAN et al., 2009)

O sistema original de metadados do ARIA foi criado em 1997 para suportar os requisitos de aplicação (pesquisa, pré-visualização das imagens, e downloads), requisitos de usuário (processamento histórico, referência espacial e organização dos dados) baseado no Conteúdo Padrão para Metadados Geoespaciais Digitais (CSDGM – do inglês “*Content Standard for Digital Geospatial Metadata*”) do Comitê Federal de Dados Geoespaciais (FGDC – do inglês “*Federal Geospatial Data Committee’s*”) (MORAN et al., 2009) (Federal Geographic Data Committee, 1998).

A figura 1 mostra a arquitetura do sistema ARIA, à qual este trabalho não entrará em detalhes. É tido como um ponto forte do ARIA, sua interconexão com os sistemas WIGDA (do inglês *WGEW Image and Ground Data Archive*), e SRER (do inglês *Santa Rita Experimental Range*), que possibilita a aquisição, através do ARIA, de dados destes sistemas.

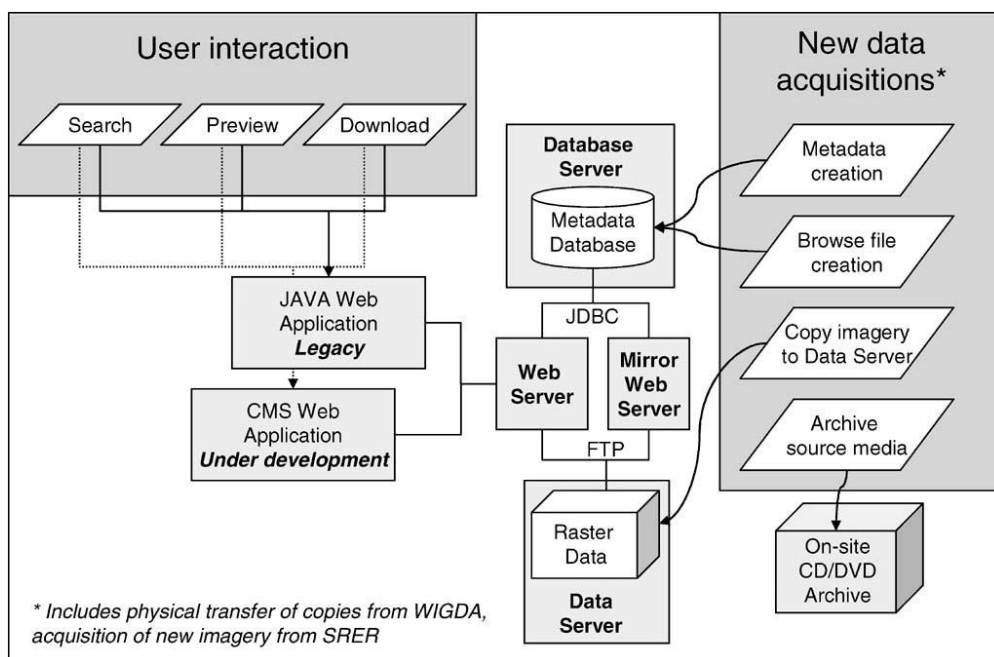


Figura 1 – Arquitetura do sistema ARIA, assim como suas interconexões com WIGDA e aquisição de imagens do SRER (MORAN et al., 2009)

NCGDAP: Projeto de Arquivamento de Dados Geoespaciais da Carolina do Norte

O sistema NCGDAP (do inglês *North Carolina Geospatial Data Archiving Project*), disponível no endereço <http://www.lib.ncsu.edu/ncgdap> é outro exemplo deste tipo de sistema, que opera desde 2004. Segundo Morris et al. (2006), o sistema tem seu foco na identificação, aquisição, preservação e distribuição de dados geoespaciais digitais de agências governamentais estaduais e locais no estado da Carolina do Norte, nos Estados Unidos. Foi criado com os seguintes objetivos:

- Identificação das fontes de dados disponíveis através de inventários locais;
- Aquisição de dados geoespaciais em risco (dados valiosos que correm o risco de serem perdidos), incluindo dados estáticos, tais como ortofotos, e de séries temporais;
- Desenvolvimento da arquitetura de um repositório digital utilizando-se softwares livres e de código aberto;
- Melhoramento de metadados geoespaciais existentes, para a melhor preservação dos mesmos;
- Investigação de identificação mecanizada e captura de fontes de dados utilizando as especificações OGC (do inglês *Open Geospatial Consortium*) para interação do cliente com os dados em servidores remotos;
- Desenvolvimento de um modelo para o arquivamento de dados e séries temporais.

O NCGDAP trás consigo um *feedback* para a comunidade de geradores de dados geoespaciais sobre a importância da qualidade do conteúdo e da qualidade dos metadados, com o objetivo também de estimular o processo de desenvolvimento de sistemas de arquivamento e distribuição de dados geoespaciais (MORRIS et al., 2006).

5. Conclusão

Com o grande crescimento da produção de dados geoespaciais, cresce também a demanda pelo desenvolvimento de sistemas para o arquivamento e distribuição dos mesmos. A distribuição é um fator importante nestes sistemas, pois visa facilitar o

acesso aos dados para pesquisadores ou outros interessados. Outro fator importante neste tipo de arquivamento é a preservação dos dados, evitando perdas importantes.

Apesar de necessário, o desenvolvimento de sistemas de arquivamento e distribuição de dados geoespaciais enfrenta alguns desafios, como a obsolescência da tecnologia, a segurança do local de armazenamento, a promoção do acesso livre e completo aos dados, a integração entre os sistemas e o planejamento de sua sustentabilidade. Através disto, é possível concluir que este é um campo que necessita de um investimento considerável por parte das unidades interessadas no mesmo, no desenvolvimento e também na manutenção destes sistemas.

6. Referências Bibliográficas

ATKINSON, C., H.; ISBELL, C., E., *Preservation of digital data from high volume data repositories. Proceedings of American Geophysical Union Conference*. San Francisco, CA, USA, 2003. Disponível em: <http://trs-new.jpl.nasa.gov/dspace/bitstream/2014/8024/1/03-3319.pdf>. Acesso em maio de 2009.

BLEAKLY, D., R., *Long-Term spatial data preservation and archiving: What are the issues? Sandia National Laboratories*, Albuquerque, NM, USA, SAND Report 2002-0107, 2002.

DEKKER, R. *The importance of having data-sets. Proc. Conf. Int. Assoc. Technol. Univ. Libr.*, 2006, p. 1-4. Disponível em http://www.iatul.org/doclibrary/public/Conf_Proceedings/2006/Dekkerpaper.pdf. Acesso em maio de 2009.

DOORN, P.; TJALSMA, H. *Introduction: archiving research data. Arch Sci*, vol. 7, n.1, p. 1-20, Mar 2007.

Federal Geographic Data Committee, *Content Standard for Digital Geospatial Metadata – FGDC Geographic Metadata Committee*, 1998. Disponível em: <http://www.fgdc.gov/metadata/csdgml/>. Acesso em Maio de 2009.

MORAN, S., M.; HUTCHINSON, B., S.; MARSH, S., E.; MCCLARAN, P., M.; OLSSON, A., D. *Archiving and Distributing Three Long-Term Interconnected Geospatial Data Sets. IEEE Transactions on Geosciences and Remote Sensing*, vol. 47, n. 1, p. 59-71, Jan 2009.

MORAN, S., M.; PETERS, D., C., P.; MCCLARAN, M., H.; NICHOLS, M., H.; ADDAMS, M., B., *Long-Term Data Collection at USDA Experimental Sites for Studies of Ecohydrology*. **J. Ecohydrology**, vol. 1, p. 377-393, 2008.

MORRIS, S., P., *Geospatial Web Services and Geoarchiving: New Opportunities and Challenges in Geographic Information Services*. **Library Trends**, vol. 55, n. 2, p. 285-303, 2006.

MORRIS, S., P.; TUTTLE, J.; FARREL, R., *Preservation of State and Local Government Digital Geospacial Data: The North Carolina Geospatial Data Archiving Project*. **Proceedings of Archiving**, Ottawa, Canada, 2006. Disponível em <http://www.lib.ncsu.edu/ncgdap/documents/archiving2006.pdf>. Acesso em Maio de 2009.

MORRIS, S., P.; TUTTLE, J. *Curation and Preservation of complex data: The North Carolina geospatial data archiving project*. **Proceedings of Digital Content Curation Symposium**, 2007. Disponível em: <http://repository.lib.ncsu.edu/publications/handle/1840.2/105>. Acesso em Maio de 2009.