

## e-Science, GIS e Curadoria Digital de dados arqueológicos e históricos: o passado conectado

Marcio Teixeira-Bastos\*  
Ivan Esperança Rocha\*\*

TEIXEIRA-BASTOS, M.; ROCHA, I. E. e-Science, GIS e Curadoria Digital de dados arqueológicos e históricos: o passado conectado. R. Museu Arq. Etn., 32: 131-142, 2019.

**Resumo:** Neste artigo, vamos apresentar uma visão condensada do estado atual da e-Science no país, o uso de Sistemas de Informação Geográficos (SIG) em Arqueologia e História, assim como a tentativa de esboçar tópicos de curadoria digital de dados, a fim de salientar seu potencial nas análises que envolvem o passado. Nesse sentido, o passado precisa ser entendido como algo nunca definitivamente acabado, que é coexistente com o presente, a todo momento sendo evocado pela memória (seja individual ou coletiva). O corrente papel do SIG e da análise espacial para estudos arqueológicos e históricos demonstra enorme potencial para mudar perspectivas teóricas e hermenêuticas de pesquisa. E, finalmente, vamos tentar definir uma agenda provisória para o futuro da pesquisa.

**Palavras Chave:** e-Science; Arqueologia e História; Dados e Arqueologia Digital.

### e-Science e dados em grande escala

A e-Science estuda, emprega e aprimora o processo contínuo de inovação em métodos de pesquisa intensivos em computação ou intensivos em dados. Normalmente, isso é feito de forma colaborativa usando uma infraestrutura distribuída. O termo e-Science (e-Ciência, em Português)<sup>1</sup> engloba todos os campos de pesquisa e aborda todas as etapas do

ciclo de vida da pesquisa, desde a formulação das questões de pesquisa, passando por simulações e análises de dados em grande escala, descoberta científica, compartilhamento em longo prazo, reutilização e replicação dos resultados, bem como as ferramentas, processos e conhecimentos relevantes. A e-Science promove, portanto, a inovação em pesquisa colaborativa, computacional ou intensiva em dados em todas as disciplinas, durante todo o ciclo de vida da pesquisa.<sup>2</sup>

A aplicação da informática à realização de investigação científica moderna, inclui a

(\*) UNESP-Assis / Stanford University. Pesquisador Associado do Laboratório de Arqueologia Romana Provincial, LARP-MAE/USP <marcio\_quisleu@yahoo.com.br>

(\*\*) UNESP-Assis <ivan.rocha@unesp.br>

(1) O termo foi criado em 1999 por John Taylor, Diretor Geral do Escritório de Ciência e Tecnologia do Reino Unido.

(2) A definição oferecida é fruto do IEEE eScience Conference Series (2014). Veja <https://escience-conference.org>

preparação, experimentação, coleta de dados, disseminação de resultados, armazenamento a longo prazo e acessibilidade de todos os materiais gerados por meio do processo científico, que podem incluir conjuntos de dados brutos e/ou ajustados, digitalização de manuscritos e artefatos, modelagem e análise de dados, modelagem 3D de estruturas, artefatos e paisagens, produção de versões preliminares, pré-impressões e publicações impressas ou eletrônicas. É dessa forma que a ciência eletrônica é amplamente interpretada e aplicada. No Brasil, diversas infraestruturas para e-Science (e-Infraestrutura ou Cyberinfraestrutura, são termos similares) estão sendo desenvolvidas. Este ambiente de eScience envolve o compartilhamento de instrumentos científicos, dados distribuídos, visualização remota e interpretação colaborativa de dados e resultados, se adequando perfeitamente às características de uma infraestrutura de computação em *grid*, grade em português (Buyya *et al.* 2000).

Na computação em grade o poder de processamento de muitos *clusters* é reunido para se obter um desempenho muito superior ao que seria possível caso os aglomerados fossem mantidos isolados. A possibilidade de resolver problemas mais complexos com o tratamento de conjuntos de imagens coletadas, biologia computacional, mineração de dados, simulação de modelos científicos e de engenharia, são maximizados pela computação em *grid*. Atualmente, existem diversas tecnologias para suportar a computação em grade, dentre as quais destacamos o Globus, Gradebus, Condor, OurGrade e Anthill (Goes *et al.* 2005).

Entre as iniciativas da computação em grade em âmbito nacional está o SINAPAD (Sistema Nacional de Processamento de Alto Desempenho), que é uma rede de centros de computação de alto desempenho, geograficamente distribuídos, instituída pelo Ministério da Ciência e Tecnologia (MCT), contemplando nove unidades, denominadas «Centros Nacionais de Processamento de Alto Desempenho» (CENAPADs), operadas pela UFRGS, UFMG, UFC, UNICAMP, UFRJ, UFPE, INPE, INPA e LNCC. Este último coordena o sistema por delegação

do MCT.<sup>3</sup> Outra iniciativa de grande porte é o CCES,<sup>4</sup> um centro de pesquisa multidisciplinar de classe mundial dedicado à modelagem computacional avançada em ciências moleculares físicas e químicas, engenharia mecânica e materiais computacionais, biologia computacional e bioinformática, geofísica computacional e ciência da computação.

A UNESP, por sua vez, recentemente implantou uma das maiores iniciativas do Campus Grid na América Latina, com recursos computacionais amplamente dispersos em sete cidades diferentes do Estado de São Paulo. Os computadores do GridUnesp utilizam processadores Intel Xeon de quatro núcleos e foram adquiridos da Sun Microsystems do Brasil. A estrutura permitirá a grupos de pesquisa acesso a elevados níveis de capacidade de processamento e armazenamento de dados nos mais diversos campos. O GridUNESP,<sup>5</sup> como é conhecido o projeto, capacitará grupos de pesquisa universitários de várias áreas de investigação científica, entre elas o sequenciamento genético,<sup>6</sup> previsão do tempo, simulações numéricas em grande escala, e a física de alta energia, por exemplo, permitindo que os pesquisadores acessem sistemas avançados de processamento e armazenamento de dados. O *cluster* central do Núcleo de Computação Científica da Unesp (NCC/Unesp) está instalado no novo campus da UNESP na Barra Funda, em São Paulo, e possui 3104 núcleos de processamento, atingindo um pico teórico de cerca de 77 teraflops (trilhões de cálculos por segundo). Esses recursos estão disponíveis por meio de computação de alto desempenho (*clusters HPC*), armazenamento e processamento em grade (*grid computing*).

(3) <https://www.lncc.br/sinapad/sinapad.php?pg=sinapad#>  
(4) <http://cces.unicamp.br>

(5) <https://ncc.unesp.br/gridunesp/docs/introducao.html> considerada atualmente a maior rede de alto desempenho da América Latina.

(6) Na UNESP-Assis o grupo é liderado pelo Prof. Dr. José Celso Rocha, através do projeto: «Classificação de embriões humanos mediante as técnicas de time-lapse, processamento de imagens digitais e inteligência artificial». Disponível em: <<https://bv.fapesp.br/pt/auxilios/100832/classificacao-de-embrioes-humanos-mediante-as-tecnicas-de-time-lapse-processamento-de-imagens-digit/>>. Acesso em 10 maio 2019.

A e-Science possui, dessa forma, a capacidade de revolucionar ambos os fundamentos do método científico: a pesquisa empírica e a teoria científica. A primeira, especialmente através de big data digital, e a segunda, através da construção de modelos de simulação computacional. O método mais tarde foi dividido em dois ramos principais, abordagens dedutiva e empírica. Hoje, uma revisão teórica do método científico moderno deve incluir um novo ramo, o da abordagem computacional. A e-Science reclama uma mudança de paradigma, pois visa combinar tradições empíricas e teóricas, enquanto as simulações computacionais podem criar dados artificiais, e big data em tempo real pode ser usado para calibrar modelos teóricos de simulação.

As plataformas de e-Science procuram fornecer suporte a um novo paradigma de fazer ciência e novas regras para cumprir os requisitos de preservar e disponibilizar os resultados de dados computacionais (digitalizados) de uma maneira que são reproduzíveis em etapas lógicas rastreáveis, como um requisito intrínseco para a manutenção da integridade científica moderna. Nesse caso, simplesmente modelar a proveniência dos dados não é suficiente. É preciso modelar a proveniência das hipóteses e resultados gerados a partir da análise dos dados, bem como fornecer evidências que suportem novas descobertas. Fluxos de trabalho científicos foram, portanto, propostos e desenvolvidos para auxiliar os cientistas a rastrear a evolução de seus dados, resultados intermediários e resultados finais, como um meio de documentar e acompanhar a evolução das descobertas dentro de uma pesquisa científica. A ciência intensiva em dados ou “e-Science” pode, sim, ser vista como outro “paradigma” na ciência moderna (a empírica, a teórica e a computacional baseada em dados). Tudo sobre ciência está mudando por causa do impacto da tecnologia da informação e do dilúvio de dados (Bell *et al.* 2009).

Técnicas de computador nos permitem analisar dados e construir modelos que vão muito além da capacidade do cérebro humano e, portanto, são essenciais para alargar os limites do nosso conhecimento do passado. Mas se não considerarmos adequadamente as ferramentas como objeto de investigação, somente nos res-

tará a completa dependência dos especialistas, em vez de definir nossa própria agenda de uso do digital na Arqueologia e História. As técnicas de computador são extremamente adequadas para a formulação e teste de hipóteses, cumprindo um papel importante como dispositivos heurísticos, como ferramentas que nos ajudam a pensar mais profundamente sobre nossas suposições.

Assim, como existe uma tendência mundial em dar acesso livre aos resultados de pesquisa por meio da criação de repositórios digitais e da publicação de periódicos livres, existe também uma demanda em torno do acesso livre aos dados gerados pela pesquisa científica. Esse é um assunto complexo e fomenta posicionamentos distintos que não são o foco nesse artigo. A importância desses dados, por outro lado, é cada vez mais patente, visto que são eles que dão sustentação aos resultados que serão discutidos nas publicações acadêmicas tradicionais e podem servir de base para novos projetos. Por isso, o uso intensivo de computadores, de tecnologias de rede, de instrumentação avançada e de simulação na pesquisa científica proporciona o surgimento de uma ciência inteiramente baseada nos fluxos de dados e conjunto de objetos digitais armazenados em repositórios distribuídos globalmente. Essa nova forma de gerar e disseminar conhecimento é chamada de eScience. Os dados digitais gerados a partir dessa forma de praticar ciência precisam ser tratados e gerenciados de modo que a preservação por longo prazo, o acesso, a autenticidade e o reuso para novas pesquisas desses dados possam ser assegurados. Este trabalho encontra no conceito de “Curadoria Digital” um novo instrumento de tratamento e representação da informação que, desenvolvida de forma correta, poderá contribuir significativamente para o ciclo da comunicação científica e o estudo interdisciplinar.

### GIS e dados digitais históricos e arqueológicos

A Análise Espacial é vista pela maioria dos arqueólogos atuantes como ferramenta essencial para explorar, analisar e interpretar dados coletados em atividades de campo ou outra base

de dados arqueológico-históricos. Tornou-se, assim, um elemento integrante de muitos projetos de pesquisa, utilizando, principalmente, os sistemas SIG para maior eficiência na realização de trabalhos arqueológicos “tradicionais” de pesquisa. Os Sistemas de Informação Geográficas (SIG-GIS) foram apresentados à Arqueologia em meados da década de oitenta (e.g. Hasenstab 1983), impulsionando a assim denominada “virada espacial” (*spatial turning*), segmento que vem delineando suas prerrogativas no domínio das Ciências Humanas nos últimos trinta anos. Pearson e Collins (1994: 3-5; Hodder 1994) afirmam que “os espaços comandam os corpos, prescrevem gestos, rotas e distâncias a ser cobertas”, de maneira que até mesmo a noção de paisagem e meio-ambiente passa pela compreensão de artefato cultural, ou melhor, das relações entre humanos e coisas (Hodder 2018).

Obviamente, dificuldades são encontradas na configuração de protocolos de análise padronizados e os desafios nos obrigam a ser muito críticos não somente a respeito dos dados originais, mas também dos resultados da modelagem produzida, uma vez que os erros nos dados serão propagados em futuras análises. Outra barreira é o entendimento corrente entre historiadores e arqueólogos de ver métodos e técnicas da “ciência dura” como ferramentas auxiliares que fornecem dados úteis para serem usados na construção de uma narrativa histórica, ao invés de um possível foco de pesquisa. Por esta razão, a Arqueologia Digital ainda se circunscreve no âmbito de um grupo de especialistas. A análise de redes arqueológicas está ganhando importância e se tornará muito melhor integrada com os sistemas SIG em um futuro próximo. Um desenvolvimento semelhante pode ser esperado no que diz respeito à combinação de SIG e modelagem de simulação dinâmica, que é um campo crescente de pesquisa arqueológica e ciência social (Kohler e Van der Leeuw 2007, Romanowska 2105, 2018). Soluções baseadas em técnicas de SIG e análises que levam em conta modelagem baseada em agentes (Wurzer *et al.* 2015), assim como o avanço em relação a softwares de processamento estatísticos, prometem levar a uma maior integração e extensão das técnicas de análise espacial.

A ideia de usar análise espacial formalizada como uma linha principal de pesquisa para arqueólogos e historiadores procura abarcar o melhor entendimento das dinâmicas regionais e diacrônicas dos assentamentos (sítios arqueológicos) e manejos da terra. A História, Arqueologia ou mesmo as Ciências Sociais, de maneira geral, não se qualificam ainda no país como um campo com dados realmente grandes, mas se utilizam de muitos dados “confusos” vindos de uma multiplicidade de fontes com distintas peculiaridades históricas (Cooper e Green 2015). Os dados ainda estão em sua infância, mas inevitavelmente começam a desempenhar um papel mais importante no futuro da pesquisa científica, visando em essência à análise de dados de várias fontes por meio de plataformas institucionais, até mesmo em cooperações internacionais (Wilcke 2015; Chiarcos *et al.* 2016).

O SIG certamente continuará contribuindo para o debate sobre a utilidade do digital e a utilização de métodos quantitativos em Arqueologia e História, mas precisamos estar cientes de que tem suas limitações. No final, é a atitude multidisciplinar referida por Hacgüzeller (2012), combinando conceitos do SIG com outras abordagens teórico-metodológicas, que fomentará novas práticas e aplicações em Arqueologia e História. A análise espacial é tão natural quanto interpretar a estratigrafia, teorizar sobre identidades a partir de um conjunto cerâmico ou mesmo debruçar-se sobre manuscritos para o entendimento de um período histórico determinado, solidificando esses esforços mutuamente. Os usuários de SIG em ciências sociais devem estar cientes disso e manter sua atenção para o desenvolvimento de novas ferramentas de forma a combinar a análise espacial com outras abordagens (Fig.1).

O significado dos sistemas e da ciência da informação geográfica parece indicar que não será negligenciado nos estudos atuais, principalmente quando se pensa em uma abordagem de questões no nível suprarregional através de dados. A ciência social computacional aborda os efeitos que a digitalização tem por meio do trabalho teórico e o uso de simulações computacionais para modelar cenários futuros que não necessariamente existem em dados empíricos, mas em teoria (Fig. 2). Combinando evidências

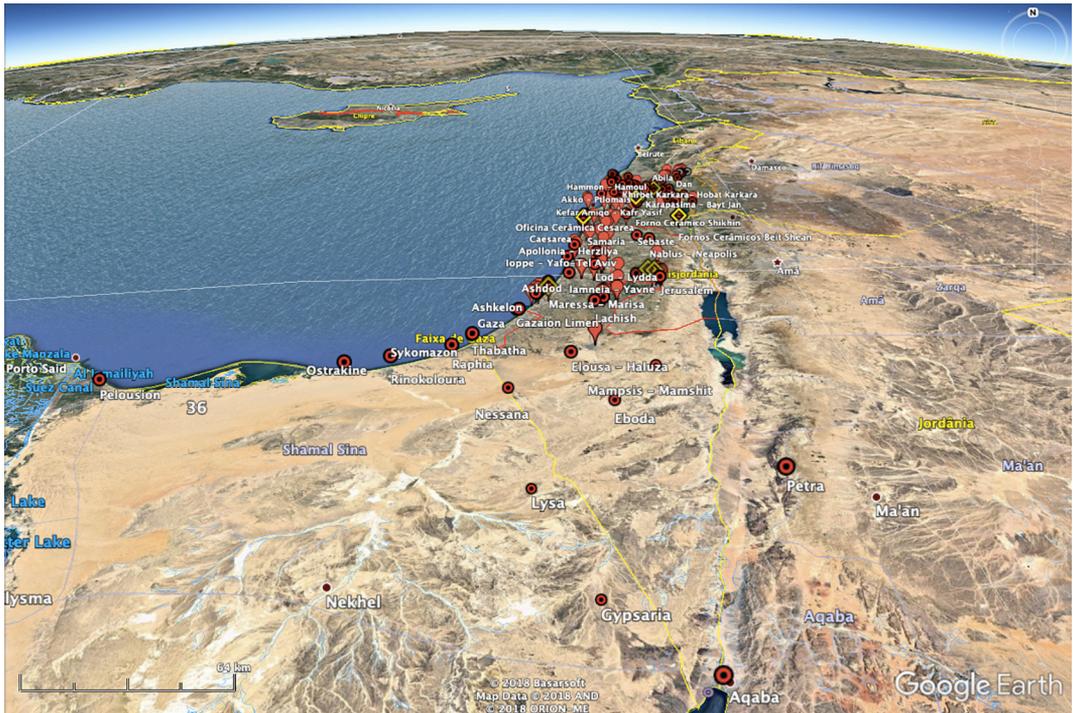


Fig. 1. Dados SIG Período Romano em Israel em Google Earth.

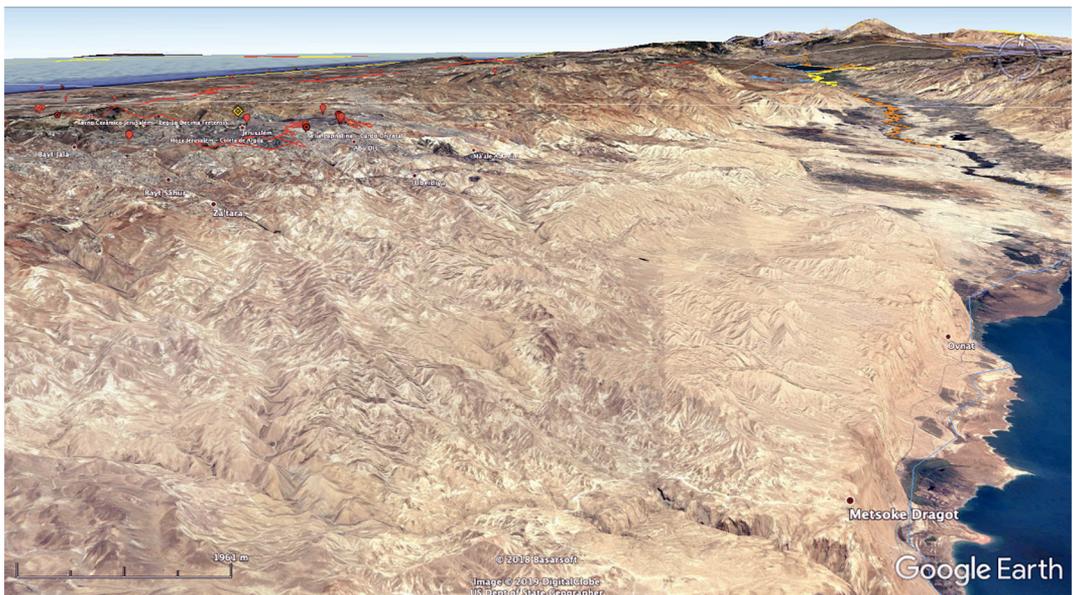


Fig. 2. Dados SIG Período Romano com destaque para a área montanhosa de Jerusalém.

empíricas com modelagem teórica, criamos a exploração digital da realidade (e.g. *Google Earth* ou *Google Maps*). Essa abordagem, que não é exclusiva das ciências sociais, entusiasma muitas

disciplinas científicas sobre as possibilidades que a digitalização traz à mesa.

Mesmo com o êxito alcançado pela Google e Microsoft Research em termos tecnológicos, a

ideia de modelar a vida na Terra (em termos de espécies) para obter uma visão ampla do ecossistema e saber o que acontece com as diferentes espécies ainda enfrenta muitos problemas. O maior obstáculo é obter os dados para parametrizar e validar o modelo. Os métodos sofisticados utilizados (câmeras automatizadas, tecnologias de reconhecimento de imagem, câmeras ativadas por movimento na selva, gravadores etc.) apenas conseguiram captar alguns dados para poder parametrizar e validar o modelo a fim de afinar suas suposições sobre como a vida na Terra realmente funciona (Purves *et al.* 2013). E esse é o maior problema: obter os dados.

Contudo, em relação às Humanidades, lidamos atualmente com o avanço digital em uma escala técnica tão evoluída, que a cada momento nos diz exatamente onde estamos, nos geo-localizando (o mapa que localiza o indivíduo, não o indivíduo que se localiza no mapa, como na cartografia clássica). A telefonia móvel mudou radicalmente o uso da cartografia moderna e também das análises espaciais. Hoje, além da geolocalização, é possível coletar dados a respeito das estruturas, materiais, designers e dietas alimentares. Em relação a este último caso, a maioria de suas despesas é registrada através de dinheiro de plástico para que saibamos o que consomem. Sendo assim, as ciências sociais tradicionalmente nunca tiveram uma larga escala de dados, porém, atualmente podemos observar a realidade em tempo real, podemos observar as interações sociais em tempo real à medida que acontecem e a proliferação dos mais diversos dados. Então, hoje em dia, nós certamente temos muito mais dados do que podemos analisar nas ciências sociais.

A ciência das redes e a modelagem baseada em agentes são duas novas abordagens na caixa de ferramentas do arqueólogo e historiadores, que nos permitem estudar fenômenos relacionais e simular nossas teorias sobre o comportamento humano passado. A ciência das redes pode ser usada para representar dados arqueológicos como nós e ligações, para explorar a estrutura do nosso conjunto de dados a partir de uma perspectiva relacional (Brughmans e Brandes 2017; Brughmans e Poblome 2016; Brughmans, Collar e Coward 2016: 3-19). Com a modelagem baseada em agentes, pode-se

representar formalmente o comportamento de indivíduos passados que supostamente levaram aos padrões observados em conjuntos de dados arqueológicos e históricos; por meio de experimentos de simulação, é possível testar se o comportamento teorizado oferece uma boa explicação no registro arqueológico e/ou histórico. Ambos os métodos estão ancorados em uma base compartilhada de técnicas formais da ciência da complexidade, e o aprendizado dessas abordagens em conjunto proporciona uma compreensão mais profunda dos pontos fortes e das diferenças entre ambos. No entanto, a curva de aprendizado para essas abordagens é íngreme (Davies e Romanowska 2018; Romanowska *et al.* 2019).

É possível encontrar humanos adaptados a praticamente qualquer bioma da Terra e os humanos parecem estar em todo lugar. Embora alguns desses locais possam ser mais agradáveis do que outros, nossa espécie parece ter uma distribuição geográfica tão ampla, virtualmente sem precedentes no reino animal. No curso da evolução humana, os humanos foram capazes de encontrar respostas a alguns dos desafios enfrentados, como a rápida mudança climática, as migrações em massa e a resiliência social diante de desastres naturais. Modelar o passado humano fornece dados ao presente e possibilita simulações de futuro que nos auxiliam a compreender melhor a formação do tecido social humano. No entanto, as questões mais importantes dizem respeito a como tratar diferentes escalas de análise, registro de dados desiguais e o desafio de traduzir modelos e dados entre fronteiras disciplinares.

### Curadoria Digital

O Centro de Curadoria Digital (Digital Curation Cen ter, DCC)<sup>7</sup> do Reino Unido oferece um modelo para o ciclo de vida da curadoria (Fig.3), amplamente aceito atualmente entre os pesquisadores, que reflete uma visão de alto nível dos estágios necessários para

(7) <http://www.dcc.ac.uk>

o sucesso do processo de curadoria e de preservação de dados de pesquisa. As bases de dados são definidas como coleções estruturadas de registros ou de dados armazenados em sistemas de computadores. O modelo proposto está orientado para o planejamento das atividades de curadoria nas organizações ou em consórcios ajudando a garantir que todos os passos do ciclo serão cumpridos e os dados gerenciados. No centro do ciclo de vida da curadoria está o dado digital, que é qualquer informação codificada em formato binário. A ideia de dado inclui os objetos compostos por um único arquivo (objetos digitais simples), além do identificador e metadados (objetos digitais complexos) que, por sua vez, podem se encontrar em uma página *web* integrada pela combinação de outros objetos digitais (formando uma unidade discreta).

Os elementos fundamentais do modelo são os dados, os objetos digitais e as bases de dados.

A operacionalização dos estágios depende das áreas de necessidade de cada organização e pesquisa. A descrição e a representação da informação, efetivada pela atribuição de metadados, devem acontecer sob um plano de preservação, cujo espectro englobe todo o ciclo de vida da curadoria digital (Fig.4). A participação na escolha e desenvolvimento de ferramentas e de *software* adequados ao problema alerta para o contínuo fluxo desse tipo de pesquisa e ações de preservação digital. Dessa forma, se pode verificar a disseminação da curadoria digital de dados e avaliar a nova alteração do ciclo da comunicação científica. Ainda que se trate de um conceito em evolução, “*envolve a gestão atuante e a preservação de recursos digitais durante*

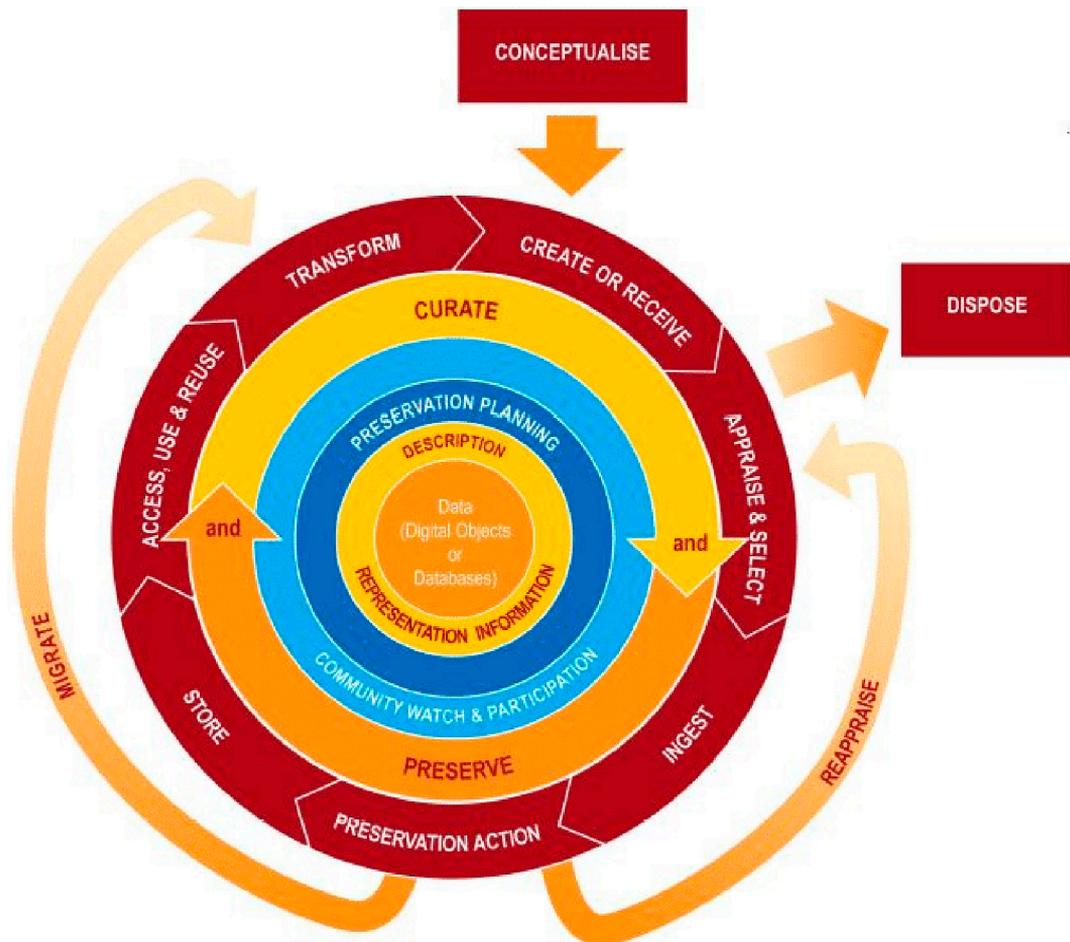


Fig. 3. Modelo proposto de Curadoria Digital (DCC, Digital Curation Center).



Fig. 4. Dados arqueológicos administrados digitalmente por sistemas GIS.

todo o ciclo de vida de interesse do mundo acadêmico nessas dados, tendo como perspectiva o desafio temporal de atender a gerações atuais e futuras de usuários” (Sales e Sayão 2012:125; Sayão 2009; 2012).

A sequência de ações do modelo de ciclo de vida da curadoria digital proposto pelo DCC tem os seguintes estágios: • **Conceituar** – conceber e planejar a criação do dado, incluindo os métodos de captura e as opções de armazenamento; • **Criar e receber** – criar o dado incluindo o elenco de metadados necessários à sua gestão e compreensão; • **Avaliar e selecionar** – avaliar o dado e selecionar o que será objeto dos processos de curadoria e de preservação em longo prazo; manter-se aderente às práticas, às políticas pertinentes e às exigências legais; • **Capturar** – transferir o dado para um arquivo, repositório, centro de dados ou outro custodiantes apropriado; • **Ação de preservação** – promover ações para assegurar a preservação de longo prazo e a retenção do dado de natureza oficial. Essas ações de preservação incluem a limpeza do dado e a sua validação, a adição de metadados de preservação.; **Armazenar** – armazenar o dado de forma segura, mantendo a aderência

aos padrões relevantes; • **Acessar, usar e reusar** – assegurar que o dado possa ser cotidianamente acessado tanto pela sua comunidade alvo quanto pelos demais usuários interessados no reuso do dado; • **Transformar** – criar novos dados a partir do original, por exemplo, pelo processo de migração para diferentes formatos ou pela criação de subconjuntos, derivando novos resultados que podem ser publicados.

Por fim, ocasionalmente o DCC aconselha a • **Eliminar** – eliminar o dado que não foi selecionado para curadoria e preservação de longo prazo de acordo com políticas documentadas, diretrizes e exigências legais; • **Reavaliar** – retornar ao dado cujos procedimentos de avaliação foram falhos para nova avaliação e possível seleção. • **Migrar** – migrar os dados para um formato diferente; isso pode ser feito no sentido de compatibilizá-lo com o ambiente de armazenamento ou para assegurar a imunidade do dado em relação à obsolescência de hardware e de software.

No sentido cooperativo e de preservação de dados arqueológicos e históricos, a UNESP-Assis e o MAE-USP desenvolvem o fomento

dessas práticas digitais conjuntas. Existe a necessidade urgente de melhorar a infraestrutura que suporta a reutilização de dados acadêmicos entre as Universidades brasileiras. Um grupo diversificado de partes interessadas (representando a academia, a indústria, as agências de financiamento e os editores acadêmicos europeus) se uniram para projetar e endossar um conjunto de princípios. O consórcio de cientistas e organizações denominou-os de Princípios de Dados FAIR<sup>8</sup> (Fig.5). A intenção é que estes princípios possam funcionar como uma diretriz para aqueles que desejam aumentar a capacidade de reutilização de seus dados. Diferentemente das iniciativas de pares que concentram o acadêmico no humano, os Princípios FAIR enfatizam especificamente o aprimoramento da capacidade das máquinas de encontrar e usar os dados automaticamente, além de apoiar sua reutilização por indivíduos.

Os Dados FAIR são dados que atendem a padrões de encontrabilidade, acessibilidade, interoperabilidade e reusabilidade. Em 2016 os líderes do G20 emitiram uma declaração aprovando os princípios FAIR para a pesquisa, além de outras instituições como a Association of European Research Libraries. Ultimamente a valorização do conceito FAIR aumentou entre vários pesquisadores e institutos, sendo que os guias sobre a implementação de práticas de dados FAIR recomendam que o custo de um plano de gestão de dados em conformidade com as práticas de dados do FAIR deve ser de 5% do orçamento total da pesquisa.

## Conclusão

O uso de conceitos relacionais e pensamento relacional no estudo do passado não é incomum. Arqueólogos e historiadores frequentemente discutem redes passadas, as interações de entidades, a conectividade e o papel que as

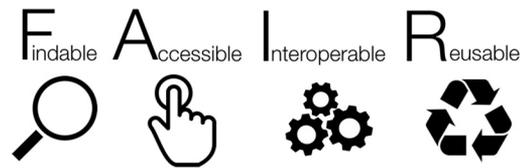


Fig. 5. Representação dos Princípios de Dados FAIR.

entidades desempenham como partes de uma rede. Gordon Childe (1925), por exemplo, procurou entender o papel desempenhado pela ilha de Creta e seus habitantes nas regiões do Mar Mediterrâneo e no Mar Negro, considerando que a estruturação geográfica das entidades permitiu e determinou a disseminação de inovações, criando as dependências de poder e fluxo de material e recursos. Renfrew (1972), por sua vez, esteve interessado nas hostilidades inter-aldeias e interregionais e como isso é expresso na arquitetura e no planejamento de assentamentos. Ele supõe que mudanças na natureza das hostilidades deram origem a fortificações que, por sua vez, mudaram a natureza dessas hostilidades.

Ian Hodder (1990), por sua vez, está mais interessado no papel dos modelos de casa e forno, e nos paralelos simbólicos entre eles, e supõe que a analogia conceitual é expressa por semelhanças morfológicas, permitindo-lhe fazer interpretações sobre a centralidade do forno dentro da casa. Todos os três exemplos exploram fenômenos passados definidos por múltiplas entidades interagentes (por exemplo, arranjos geográficos, rotas de comércio e difusões culturais entre Creta, Egito e Suméria; arranjos espaciais de regiões, aldeias e edifícios individuais; e paralelos simbólicos e analogias entre casas, fornos e pessoas) através de um quadro mais abstrato de conceitos.

Os conceitos de rede podem ser intuitivamente atraentes e particularmente adequados para descrever muitos dos fenômenos passados que pretendemos compreender. As entidades podem ou não ser conceituadas como coisas físicas claramente delimitadas e separáveis, mas são consideradas analiticamente distintas para os propósitos dos argumentos específicos de cada caso (Brughmans, Collar e Coward 2016). Entretanto, usar conceitos de “rede” vagamente definidos raramente levam a *insights* críticos, uma vez que não está claro quais vantagens eles oferecem frente a outras abordagens. Esses

(8) Certamente levando em consideração o significado de fair (justo) em inglês. Veja Wilkinson, M.D. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3. doi:10.1038/sdata.2016.18

critérios precisam ser entendidos e apreendidos para que as perspectivas formais da rede possam contribuir com seu potencial de novas perspectivas sobre o passado: um passado conectado.

Seguindo a visão de Hacigüzeller (2012), arqueólogos e historiadores em um centro de pesquisa ganham muito adotando os Sistemas de Informações Geográficas (SIG/GIS) para gestão e representação dos dados coletados quando estudam o passado. Nesta visão, o passado não é algo que possa ser entendido de forma estática e definitiva, mas sim algo que continuamente muda e é repetidamente reconstruído no presente. Portanto, é um apelo para o entendimento que se trata de uma prática de pesquisa em constante mudança, mais do que uma solução instantânea orientada para a tecnologia na hora de criar mapas. Isso reforça a ideia de que a análise espacial baseada em SIG (GIS) e modelagem digital nunca pode ser uma abordagem independentemente, mas deve ser uma parte integrante do que se pode chamar de pesquisa arqueológica colaborativa e de apelo à interdisciplinaridade na ciência moderna.

A importante característica de possuir bases técnicas com computadores apropriados

e a capacidade de lidar com grandes conjuntos de dados, o chamado “big data”, diferenciam centros de pesquisa. Os softwares SIG podem lidar com big data e acrescentam a dimensão espacial na forma de discernir padrões e simular teorias do comportamento humano em grandes áreas geográficas. Com o aumento do fluxo de informações ocasionado pelo uso das novas tecnologias da informação, os arqueólogos e historiadores podem ir além dos limites do local, das análises intra-sítio ou micro-regionais. A informação digital é vulnerável, visto que as tecnologias da informação evoluem a todo o momento, sendo assim, a preservação dos dados e informações coletadas sobre o período romano e bizantino em Israel torna-se um desafio no ambiente web. Isso significa dizer que a curadoria disponibiliza dados tratados, acompanhados por metadados semânticos e estruturais – que assegurariam o seu significado e a reconstrução de sua apresentação correta – e metadados de preservação – que mantêm sua integridade, precisão e autenticidade desses dados. É, portanto, com toda a probabilidade uma das próximas fronteiras da tecnologia espacial.

TEIXEIRA-BASTOS, M.; ROCHA, I. E. e-Science, GIS and Digital Curation of Archaeological and Historical Data: the connected past. *R. Museu Arq. Etn.*, 32: 131-142, 2019.

**Abstract:** In this article, we will present a condensed view of the current state of e-Science in the country, the use of Geographical Information Systems (GIS) in Archeology and History, as well as an attempt to approach topics related to digital data curation, in order to highlight their potential to the study of the past. In this sense, the past must be understood as something never finished, which is coexistent with the present, constantly being evoked by memory (either individual or collective). The application of GIS and Spatial Analysis for archaeological and historians held a significant role to increase the scholars’ methods and the hermeneutics of research. Finally, we hope to approach a possible agenda for the future of this type of research in Brazil.

**Keywords:** e-Science; Archaeology and History; Data and Digital Archaeology.

## Referências bibliográficas

- Bell, G.; Hey, T.; Szalay, A. 2009. Computer Science: Beyond the Data Deluge. *Science*, 323 (5919): 1297-1298. doi:10.1126/science.1170411
- Buyya, R., Abramson, D., Giddy, J. 2000. Nimrod/g: An architecture of a resource management and scheduling system in a global computational grid. In: *The 4th International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia 2000)*, Beijing, China.
- Brughmans, T., Brandes, U. 2017. Visibility network patterns and methods for studying visual relational phenomena in archaeology. *Frontiers in Digital Humanities: Digital Archaeology*, 4(17). doi:10.3389/fdigh.2017.00017
- Brughmans, T., Poblome, J. 2016. MERCURY: an agent-based model of tableware trade in the Roman East. *Journal of Artificial Societies and Social Simulation*, 19(1): 3. Disponível em: <<http://jasss.soc.surrey.ac.uk/19/1/3.html>>. Acesso em 14 de junho 2019.
- Brughmans, T., Collar, A., Coward, F. (Eds.) 2016. *The Connected Past: challenges to network studies in archaeology and history*. Oxford University Press, Oxford.
- Chiarcos, C., Lang, M., Verhagen, P. 2016. IT-assisted exploration of excavation reports using natural language processing in the archaeological research process. In: Campana, S.; Scopigno, R.; Carpentiero, G.; Cirillo, M. (Eds.) CAA 2015. *Keep the revolution going. Proceedings of the 43rd annual conference on computer applications and quantitative methods in Archaeology, Siena*. Archaeopress, Oxford: 87-94.
- Childe, G. 1925. *The Dawn of European Civilization* (1<sup>st</sup> edition). Kegan Paul, London.
- Cooper A.; Green, C. 2015. Embracing the complexities of 'big data' in archaeology: the case of the English Landscape and Identities project. *J. Archaeol Method Theory*, 23(1): 271-304 doi:10.1007/s10816-015-9240-4
- Davies, B. and I. Romanowska. 2018. An Emergent Community? Agent-based Modelers in Archaeology. *The SAA Archaeological Record*, 18(2): 27-32.
- Goes, L. F. W., Guedes, D., Ferreira, R., Cirne, W. 2005. Computação em Grade: Conceitos, Tecnologias, Aplicações e Tendências. In: SBC. (Org.) *Escola Regional de Informática de Minas Gerais (ERI-MG)*: 1-39.
- Hacigüzeller P. 2012. GIS, critique, representation and beyond. *J. Soc Archaeol*, 12: 245-263.
- Hasenstab R.J. 1983. *A preliminary cultural resource sensitivity analysis for flood control facilities construction in the Passaic River basin of New Jersey*. US Army Corps of Engineers, Marietta, GA.
- Hodder, I. 1990. *The Domestication of Europe: Structure and Contingency in Neolithic Societies*. Blackwell, Oxford.
- Hodder, I. 1994. Architecture and meaning: the example of Neolithic houses and tombs. In: Pearson, M., Collins, R. (Ed.) *Architecture and order: approaches to social space*. Routledge, London: 73-86.
- Hodder, I. 2018. *Where are we heading? The Evolution of Humans and Things*. Yale University Press, New Haven.
- Pearson, M.; Collins, R. (Ed.) 1994. *Architecture and order: approaches to social space*. Routledge, London.
- Purves, D.; Scharlemann, J.P.W.; Harfoot, M.; Newbold, T.; Tittensor, D.P.; Hutton, J.; Emmott, S. 2013. Time to model all life on Earth. *Nature*, 493: 295-297. Disponível em: <<https://www.nature.com/articles/493295a>>. Acesso em 14 jun. 2019.
- Kohler T. A.; van der Leeuw, S. 2007. *Model-based archaeology of socio-natural systems*. School of Advanced Research, Santa Fe.
- Renfrew, C. 1972. *The Emergence of Civilisation: The Cyclades and the Aegean in the Third Millennium BC*. Methuen, London.
- Romanowska, I. 2015. So you think you can model. A guide to building and evaluating archaeological simulation models of dispersals. *Human Biology*, 87 (3): 169-192. Disponível em: <[http://digitalcommons.wayne.edu/humbiol\\_preprints/79/](http://digitalcommons.wayne.edu/humbiol_preprints/79/)>. Acesso em 20 maio 2019.
- Romanowska, I. 2018. Using agent-based modelling to infer economic processes in the past. In: Remesal Rodríguez, J.; Revilla Calvo, V.; Bermúdez Lorenzo, J.M. (Eds.) *Cuantificar las economías antiguas. Problemas y métodos*. Universitat de Barcelona: 107-117.
- Romanowska, I.; Crabtree, S.; Davies, B.; Harris, K. 2019. Agent-based Modelling for Archaeologists. Part 1 of 3. *Advances in Archaeological Practice*. v.7, issue 2: 178-184. doi:10.1017/aap.2019.6

- Sales, L. F.; Sayão, L.F. 2012. O Impacto da Curadoria Digital dos Dados de Pesquisa na Comunicação Científica. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 17, n. esp. 2 - III SBCC: 118-135.
- Sayão, L. F. 2009. *Implantação e gestão de repositórios institucional: política, memória, memórias, livre acesso e preservação*. EdUFBA, Salvador.
- Sayão, L. F. 2012. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. *Inf. & Soc.: Est.*, João Pessoa, v.22, n.3: 179-191.
- Wilcke, W. X. 2015. *ARIADNE D16.1: first report on data mining*. VU University, Amsterdam.
- Wurzer, G.; Kowarik, K.; Reschreiter, H. 2015. *Agent-based modeling and simulation in archaeology*. Springer, Cham.