

# Spatial-Temporal Data Watch to Digital Media

Proposta de Tese Doutoral em Engenharia Informática

## *Spatial-Temporal Data Watch to Digital Media*

*PhD Thesis Proposal in Informatics Engineering*

Leonel Dias, Alexandre Carvalho, António Coelho

[pro11006@fe.up.pt](mailto:pro11006@fe.up.pt), [alexandre.valle@fe.up.pt](mailto:alexandre.valle@fe.up.pt), [acoelho@fe.up.pt](mailto:acoelho@fe.up.pt)

DEI – Departamento de Engenharia Informática,  
FEUP – Faculdade de Engenharia da Universidade do Porto,  
R. Dr. Roberto Frias, 4200-465 Porto, Portugal

**Resumo** — Este artigo apresenta um trabalho de doutoramento em Engenharia Informática, com previsão para conclusão em Julho de 2018. Esta tese de doutoramento insere-se na área dos Sistemas de Informação Espaço-Temporais, com aplicabilidade nas ferramentas tecnológicas de comunicação e representação visual do conhecimento, para os Media Digitais (jornais, rádios e televisões). Pretende-se potenciar a eficiência e a eficácia do valor da informação heterogénea, multivariável e multidimensional característica deste contexto, produzida e partilhada por diversas fontes e em diferentes formatos. O projeto de investigação encontra-se na fase de estudo do estado da arte e da elaboração da proposta de tese. Espera-se com a participação neste Simpósio Doutoral enriquecer e atualizar o trabalho em curso.

**Palavras Chave** – *Big Data; Sistemas de Informação Espaço-Temporais; Representação Visual; Recuperação de Conhecimento; Media Digitais; Jornalismo Online;*

**Abstract** — This paper presents a PhD thesis proposal in Informatics Engineering, scheduled for completion in July 2018. This PhD thesis is part of Spatio-Temporal Information Systems, with applicability in technological communication tools and visual representation of knowledge, for Digital Media (newspapers, radio and television). It is intended to maximize the efficiency and effectiveness of the value of heterogeneous, multivariate, multidimensional information characteristic of this context, produced and shared by different sources, in different formats. It is hoped that participation in this Doctoral Symposium will enrich and update the work in progress and help the preparation of the PhD thesis proposal.

**Keywords** – *Big Data; Spatio-Temporal Information Systems; Visual Representation; Knowledge Recovery; Digital Media;*

### I. OBSERVAÇÕES DO MUNDO E DO ESTADO DA ARTE

Os primeiros Sistemas de Informação Geográfica (SIG) [1] lidaram com representações geográficas exatas e estanques no tempo e suportaram análises rigorosas ao nível dos objetos representados [2]. Desde então, os SIG e os Sistemas de Informação espaço-temporal, evoluíram e adquiriram maior importância no auxílio e suporte à decisão e em diversos problemas do quotidiano. Ao mesmo tempo, como Heckscher

[3] afirmou, dia após dia, a informação valoriza-se cada vez mais independentemente da tecnologia e do contexto, e tornou-se um bem económico muito expressivo e sem precedentes, capaz de se sobrepor aos serviços e bens tangíveis, cuja existência é vital para qualquer processo de tomada de decisão de estados, empresas, instituições, entre outras organizações. Já Sun Tzu, que viveu na China entre os anos 400 e 320 a.C. e cujos seus pensamentos ainda influenciam as culturas e políticas dos mundos Asiático e Ocidental, escreveu sobre a importância e valor da informação [4]. Não obstante, como Bernard [5] salienta, com a crescente popularidade, massificação e diversidade das plataformas Web, combinando com a diversidade de serviços e de milhões de utilizadores diários, estes novos ecossistemas digitais estão a conduzir a uma enorme partilha e produção de dados, caracterizados por atingirem rapidamente grande volume, velocidade de produção, variedade, veracidade e valor [6]. A este conjunto de informação, a comunidade científica atribui o nome de Big Data [7].

Segundo [8] vivemos num mundo que produz uma quantidade cada vez mais impressionante de dados e a uma velocidade cada vez maior. Por isso, reunir e avaliar a informação produzida para encontrar factos relevantes é um desafio que exige soluções e metodologias capazes de suportar o volume e acompanhar a rapidez dessa evolução. Com as mudanças aceleradas no universo dos Media Digitais, fomentadas sobretudo pelo uso intensivo das novas tecnologias de informação e também com o aparecimento da informação como produto de alto valor na sociedade do conhecimento, na última década, o Big Data foi mudando e transformando a maneira como vivemos, trabalhamos, e pensamos.

Os autores [9] referem que o avanço e a evolução das tecnologias Web de informação e comunicação aliadas às necessidades da sociedade atual, na procura de ferramentas cada vez mais visuais e interativas, fazem surgir a necessidade de melhorar e implementar sistemas de informação espaço-temporal com formas inovadoras de captura, análise, armazenamento e representação desses grandes volumes de

informação produzidos por terceiros, para aquisição de conhecimento, auxílio na resolução de problemas e tomada de decisão em diversos domínios da atividade humana.

## II. DEFINIÇÃO DO PROBLEMA

Recentemente, a sociedade começou a perceber o valor que existe na identificação e análise de padrões sobre grandes quantidades de informação produzidas e começou a estudar novas formas de encontrar e obter mais conhecimento a partir desses dados, para que possam ser utilizados de forma mais eficiente e proveitosa. Um exemplo concreto são as empresas de comunicação social, que todos os dias competem com os correntes para melhor informar o público em geral e assim obter mais clientes. Mais ainda, a abundância de informação produzida atualmente no mundo digital, está a mudar a forma de trabalhar dos jornalistas e a maneira de como medem as tendências atuais e os principais temas de interesse na sociedade. Contudo, no quotidiano de um jornalista, este novo método de trabalho, resulta em tarefas recorrentes e demoradas que visam analisar milhões de dados que estão disponíveis através de pesquisas na *World Wide Web* e transformar os resultados em informação útil de qualidade que seja do interesse da sociedade atual. Sociedade esta, cuja curiosidade insaciável por mais conhecimento na internet, levou Ahamed [11] a propor uma nova adaptação da divisão hierárquica das diferentes necessidades humanas proposta por Abraham Maslow em [10].

Neste domínio, para a obtenção desse conhecimento, recorre-se a ferramentas que permitem “minerar” no meio dos grandes volumes de dados e extrair a informação mais completa para os temas que realmente podem ter interesse jornalístico. Além disso, obter a informação não é suficiente, sendo também necessário comunicar essa informação, de forma eficaz e permitir uma imediata e rápida leitura através da visualização dos dados. Atualmente a consulta de grandes volumes de dados não estruturados é realizada manualmente e, para concretizar estes objetivos, geralmente os jornalistas recorrem às pesquisas *online*. Através dos principais motores de busca (como o Google [12]) ou em páginas especializadas (como é o caso do *Google News* [13]) para obter o máximo de informação possível sobre um determinado tema em diferentes fontes (website, blogues, redes sociais, etc.), extraem-se eventuais conclusões através da análise visual dos resultados obtidos, utilizando para isso ferramentas tecnológicas, que permitem a representação da informação, como é o caso do *Tableau Public* [14]. Contudo, esta metodologia muito comum exige um enorme esforço de pré-processamento da informação para ser devidamente representada e que frequentemente, pressupõe conhecimentos técnicos que muitas vezes não fazem parte das competências dos jornalistas. Além disso, e tendo em conta a velocidade de geração de informação potencialmente relevante no mundo atual, hoje em dia, não é viável este método de obtenção de conhecimento sobre um determinado acontecimento, num determinado espaço e tempo.

A isto acresce o problema da proliferação de Informação Espacial e Temporal [15], decorrente da utilização massiva dos Sistemas de Informação Baseados na Localização [16]. Este tipo de

*Com base em grandes volumes de informação heterogénea, multivariável, multidimensional proveniente múltiplas fontes e através de representações visuais dos fenómenos, da informação geográfica e da respetiva evolução espacial conjugada com propriedades da dimensão temporal, é possível, com o auxílio de novas metodologias e um Sistema de Informação Espaço-Temporal, gerar conhecimento e informação sobre um determinado tópico de interesse”*

informação, multidimensional por natureza, baseada numa dimensão espacial, numa dimensão temporal e numa dimensão de conteúdo, tem um elevado valor na sociedade atual e por conseguinte, para os principais Media Digitais.

Contudo, com o estudo do estado da arte verificou-se que as técnicas e métodos atuais não são os mais adequados para lidar com este tipo de informação e como tal é necessário, propor novas metodologias e novos métodos que permitam um processo de recuperação e contextualização de informação espaço-temporal mais eficaz e eficiente para benefício dos Media Digitais.

Consequente, neste trabalho doutoral procura-se dar resposta ao seguinte problema:

*No âmbito dos Media Digitais, como contextualizar de forma expedita, no tempo e no espaço, os grandes volumes de informação heterogénea existente sobre um determinado tópico de interesse?*

O adjetivo “expedito” refere-se à necessidade de existir uma clara facilidade de utilização por parte de pessoal não técnico, e mesmo alguma automatização do processo, isto é, que permita que os utilizadores consigam, facilmente e sem recorrer a terceiros (geralmente especialistas em TI), criar representações visuais de informação e a partir daí extrair a informação relevante para um determinado domínio. Nos dias de hoje esta agilidade pode representar uma valorização da capacidade de comunicação, nomeadamente a capacidade de “minerar” em grandes volumes de dados, conseguir transformá-los em informação útil e apresentá-la de forma gráfica, para que seja compreensível a todos. Um de muitos exemplos de instanciação deste tipo de problema poderá ser: “*Quero ver toda a informação disponível sobre “assaltos” nos últimos seis meses numa determinada região, verificando a sequência espaço-temporal de acontecimentos e os métodos de assalto utilizados?*”

## III. HIPÓTESE

A representação visual de informação contribui de forma significativa para a correta interpretação e partilha de conhecimentos. Sinteticamente, a extração e representação

visual permite gerar “gráficos” visuais, fáceis de interpretar, a partir de grandes volumes de informação que, embora com grande complexidade, possuem informação muito relevante. A adoção da visualização de dados permite a redução de tempo na procura e análise de dados [17].

Assim sendo, e tendo em conta o problema de investigação, do qual se pretende partir de um grande volume de dados heterógeno para contextualizar de uma forma expedita, clara e precisa, o contexto espacial e temporal de um determinado assunto de interesse, através metáforas visuais adequadas para a geração de conhecimento e tomada de decisão, sugere-se a seguinte hipótese cuja validade se irá demonstrar ao longo deste trabalho:

#### IV. QUESTÕES DE INVESTIGAÇÃO

A definição das questões a investigar é um dos passos mais importantes na investigação científica e define o modo como o problema é abordado. Assim, para procurar a resposta ao problema identificado, foram delineadas as seguintes questões que guiarão esta investigação:

1. De que forma os Sistemas de Informação Espaço-Temporal são capazes de lidar com os grandes volumes de informação, heterogénea, multivariável e multidimensional, produzidos e mantidos por múltiplas fontes?
  - 1.1. Como garantem a interoperabilidade de informação heterogénea?
  - 1.2. Como relacionam a informação proveniente de fontes distintas?
  - 1.3. Como representam a evolução espacial conjugada a dimensão temporal?
  - 1.4. Como lidam com a atualização de informação em tempo-real?
  - 1.5. Como definem os vários graus de relevância e valor da informação mantida?
2. Quais são as principais normas (*standards*) internacionais desenvolvidas pelas comunidades científicas e pela indústria de produtos *geoespaciais*, que garantem desde logo uma base interoperável para inúmeras e distintas fontes de informação?
3. Como é que um Sistema de Informação Espaço-Temporal se pode tornar numa eficaz ferramenta de contextualização espaço-temporal de arquivos textuais em histórico?
4. De que forma a contextualização espaço-temporal dos arquivos históricos assume um papel relevante na geração de conhecimento, facilita uma eficaz aquisição de conhecimento e aumenta o grau de certeza em tomadas de decisão?

5. Como representar de forma expedita, no tempo e no espaço, a informação sobre um determinado tópico de interesse?
6. Poderão as propostas elaboradas ser igualmente vantajosas na compreensão e resolução de problemas semelhantes fora do âmbito dos Media Digitais?

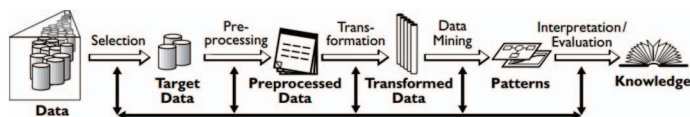
#### V. OBJECTIVOS CIENTIFICOS

Com este trabalho, pretende-se aprofundar o contexto da recuperação e visualização de grandes volumes de informação espacial e temporal, heterogénea, multivariável e multidimensional, através do desenvolvimento de métodos e técnicas de representação visual das dimensões espaço e tempo. Bem como, estudar processos de armazenamento, interoperabilidade, análise e validação de grandes volumes de informação espaço-temporal provenientes de distintas fontes e em diferentes formatos. Desta forma, pretende-se propor como solução, uma nova infraestrutura de apoio, que permita responder aos problemas e limitações existentes nos atuais sistemas de informação espaço-temporal que se encontram ao serviço dos Media Digitais.

Para atingir a finalidade deste trabalho, definem-se como objetivos gerais:

1. Analisar o papel, a cobertura, as vantagens e as limitações decorrentes do uso de Sistemas Espaço-temporal ao serviço dos Media Digitais;
2. Estudar o estado da arte dos Sistemas de Informação Espaço-Temporais, nomeadamente:
  - 2.1. Identificar as normas e padrões atuais no domínio deste tipo de soluções;
  - 2.2. Rever o contexto da recuperação e utilização de grandes volumes de informação heterogénea, multivariável e multidimensional;
  - 2.3. Conhecer as diferentes tecnologias de código aberto que podem suportar a conceptualização de uma infraestrutura, que permita o acesso interoperável e universal, através da Internet, à informação e a sua respetiva representação visual;
3. Estudar e desenvolver abordagens e ferramentas eficazes e eficientes para que pessoal não-técnico ligado à área dos Media Digitais consiga, a partir de Sistemas de Informação Espaço-temporal extrair conhecimento e tomar decisões com base representações visuais de grandes volumes de informação heterogénea, multivariável e multidimensional proveniente de diversas fontes. A concretização deste objetivo divide-se em duas grandes áreas:
  - 3.1. A componente de extração não trivial de conhecimento previamente desconhecido e potencialmente útil, na qual é necessário conhecer, identificar e propor quais as adaptações que são

necessárias em cada etapa proposta pelo KDD - *Knowledge Discovery in Databases* de [17]:



3.2. A componente de representação visual de grandes volumes de informação espaço-temporal, na qual é necessário conhecer, identificar e propor quais as técnicas ou combinações de técnicas que permitem a representação de conhecimento útil para auxílio na tomada de decisão, partindo do estudo feito por [19] em “*A Tour through the Visualization Zoo*”.

4. Submeter as propostas e as soluções estudadas a um conjunto de problemas reais, aleatórios e diversificados, que permita avaliar e interpretar os resultados obtidos;

## VI. METODOLOGIA PROPOSTA

A investigação foi iniciada no segundo semestre de 2013, e desde o início está sendo acompanhada com uma revisão sistemática das referências bibliográficas, com o objetivo de identificar projetos, conhecer os principais conceitos e processos de negócio inerentes ao problema identificado e atualizar o estado da arte no domínio dos Sistemas de Informação Espaço-Temporais.

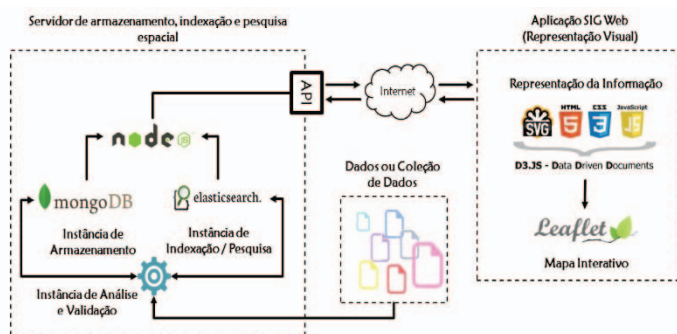


Figura 1 - Perspectiva Geral da Arquitetura de Apoio

Além disso, a estratégia geral de investigação passa por um estudo exploratório em que se recorre à utilização combinada de técnicas de recuperação e representação visual de grandes volumes de informação, tentando procurar, indagar, compreender e apresentar uma solução para o problema de fundo. No âmbito deste estudo, pretende-se propor uma solução científico-tecnológica, única e inovadora, utilizando para o efeito uma arquitetura tecnológica representada na Figura 1 e que tem como objetivo ajudar a dar resposta às questões de investigação identificadas, implementar novas metodologias para resolver as limitações encontradas, nomeadamente na extração e representação visual de conhecimento a partir informação heterogénea, multivariável e multidimensional características dos Media Digitais, produzida e partilhada por diversas fontes e em diferentes formatos.

A estratégia de avaliação dos resultados deste trabalho de investigação, passa por integrar em protótipos funcionais os contributos científico-tecnológicos, os quais serão testados e validados em ambientes reais de utilização junto de um *early adopter* dos Media Digitais, no qual os jornalistas terão uma papel ativo na avaliação dos métodos e técnicas propostas, por comparação com as abordagens atualmente utilizadas.

## VII. RESULTADOS ESPERADOS

Com este trabalho, pretende-se aprofundar o contexto da recuperação e visualização de grandes volumes de informação espacial e temporal, heterogénea, multivariável e multidimensional, através do desenvolvimento de métodos e técnicas de representação visual das dimensões espaço e tempo. Pretende-se ainda estudar processos de armazenamento, interoperabilidade, análise e validação de grandes volumes de informação espaço-temporal provenientes de distintas fontes e em diferentes formatos. Finalmente, pretende-se propor soluções que permitam responder aos problemas e às limitações identificadas, existentes nos atuais sistemas de informação espaço-temporal que se encontram ao serviço dos Media Digitais.

## VIII. AGRADECIMENTOS

Os autores agradecem ao Programa Doutoral em Engenharia Informática da FEUP pelo suporte financeiro e ao Diretor de Programa Doutoral pela sua disponibilidade e colaboração.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [1] R. Tomlinson, An Introduction to the Geo-Information System of the Canada Land Inventory, 1967.
- [2] G. F. Bonham-Carter, Geographic information systems for geoscientists: modelling with GIS, volume 13, Elsevier, 2014.
- [3] C. Heckscher, The postbureaucratic organization: New perspectives on organizational change, Sage Publications, 1994.
- [4] S. Tzu, A Arte da Guerra, Lisboa: Texto Editores, 2011.
- [5] M. Bernard, Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance, John Wiley & Sons, 2015.
- [6] A. McAfee, E. Brynjolfsson e T. Davenport, Big data. The management revolution, Harvard Bus Rev, 2012.
- [7] W. Fan e A. Bifet, “Mining Big Data: Current Status, and Forecast to the Future,” *SIGKDD Explor. Newsl*, pp. 1-5, 2013.
- [8] S. Lohr, “The age of big data,” *New York Times*, 2012.
- [9] X. Jin, B. Wah, X. Cheng e Y. Wang, “Significance and challenges of big data research,” *Big Data Research*, 2015.
- [10] A. Maslow, “A Theory of Human Motivation,” *Psychological Review*, pp. 370-396, 1943.
- [11] S. Ahamed, Next Generation Knowledge Machines: Design and Architecture, Boston: Elsevier, 2014.
- [12] “Google,” [Online]. Available: [www.google.pt](http://www.google.pt). [Acedido em 20 Janeiro 2016].
- [13] “Google Noticias,” [Online]. Available: <https://news.google.pt/>. [Acedido em 20 Janeiro 2016].



- [14] T. Software, “Tableau Public: Free Data Visualization Software,” [Online]. Available: <https://public.tableau.com/>. [Acedido em 20 Janeiro 2016].
- [15] S. Shekhar e H. Xiong, “Modeling and Multiple Perceptions: Spatial, Temporal, Spatiotemporal Objects and Relationships,” em *Encyclopedia of GIS (Springer Reference)*, 2007, p. 685.
- [16] S. Forstall, G. Christie, R. Borchers e K. Tiene, “Location-based services”. Estados Unidos da América Patente US8290513 B2, 28 Junho 2007.
- [17] D. Keim, G. Andrienko, C. Görg, J. Kohlhammer e G. Melançon, “Visual Analytics: Definition, Process, and Challenges,” *Information Visualization*, vol. 4950, pp. 154-175, 2008.
- [18] U. Fayyad, G. Piatetsky-Shapiro e P. Smyth, “Knowledge Discovery and Data Mining: Towards a Unifying Framework,” em *KDD-96 Proceedings*, 1996.
- [19] J. Heer, M. Bostock e V. Ogievetsky, “A Tour through the Visualization Zoo,” *A survey of powerful visualization techniques, from the obvious to the obscure*, vol. 8, 2010.