DOI: 10.5433/1981-8920.2016v21n2p40

PRODUÇÃO INTERNACIONAL SOBRE CIÊNCIA ORIENTADA A DADOS: ANÁLISE DOS TERMOS DATA SCIENCE E E-SCIENCE NA SCOPUS E NA WEB OF SCIENCE

INTERNACIONAL SOBRE LA CIENCIA ORIENTADA A LOS DATOS: ANÁLISIS DE LOS TÉRMINOS DATA SCIENCE E E-SCIENCE EN SCOPUS Y LA WEB OF SCIENCE

Leilah Santiago Bufrem*
Fábio Mascarenhas e Silva**
Natanael Vitor Sobral***
Anna Elizabeth Galvão Coutinho Correia****

RESUMO
Introdução: A atual configuração da dinâmica relativa à produção e à comunicação científicas revela o protagonismo da Ciência Orientada a Dados, em concepção abrangente, representada principalmente por termos como "*e-Science*" e "*Data Science*".
Objetivos: Apresentar a produção científica mundial relativa à Ciência Orientada a Dados a partir dos termos "*e-Science*" e "*Data Science*" na *Scopus* e na *Web of Science*, entre 2006 e 2016.
Metodologia: A pesquisa está estruturada em cinco etapas: a) busca de informações nas bases *Scopus* e *Web of Science*; b) obtenção dos registros

*Doutora em Ciências da Comunicação pela Universidade de São Paulo. Professora no Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP-Marília) e no Programa de Pós-Graduação em Educação da Universidade Federal do Paraná. E-mail: santiagobufrem@gmail.com
**Doutor em Ciência da Informação pela Universidade de São Paulo (USP). Professor do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Pernambuco (UFPE). E-mail:natanvsobral@gmail.com
***Doutorando em Ciência da Informação pela Universidade Federal da Bahia. E-mail: fabiomascarenhas@yahoo.com.br
****Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais (UFMG). Professora do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco (UFPE). E-mail: aegcc3@gmail.com

bibliométricos; c) complementação das palavras-chave; d) correção e cruzamento dos dados; e) representação analítica dos dados.
Resultados: Os termos de maior destaque na produção científica analisada foram *Distributed computer systems* (2006), *Grid computing* (2007 a 2013) e *Big data* (2014 a 2016). Na área de Biblioteconomia e Ciência de Informação, a ênfase é dada aos temas: *Digital library* e *Open access*, evidenciando a centralidade do campo nas discussões sobre dispositivos para dar acesso à informação científica em meio digital.
Conclusões: Sob um olhar diacrônico, constata-se uma visível mudança de foco das temáticas voltadas às operações de compartilhamento de dados para a perspectiva analítica de busca de padrões em grandes volumes de dados.

Palavras-chave: Data Science. E-Science. Ciência orientada a dados. Produção científica.

1 INTRODUÇÃO

A forma de fazer ciência tem sido transformada ao longo dos anos. É perceptível que num círculo de inter-relações, existe um constante e dinâmico processo de influência mútua entre a ciência e as tecnologias utilizadas nas diversas rotinas do fazer científico. Estas mudanças foram potencializadas a partir dos avanços das tecnologias de informação e comunicação, em particular das redes computacionais. Dentre os avanços, ressalta-se a implementação/evolução do compartilhamento de dados por parte dos cientistas, que há tempos já aspiravam a possibilidade de poderem compartilhar, utilizar e reutilizar dados com os seus pares, conforme se percebe em livro editorado por Fienberg, Martin e Straf (1985).

Os avanços se revelaram mais concretos a partir de algumas iniciativas, tais como a introdução da Ciência Orientada a Dados (COD), em 2001, por William S. Cleveland, que a lança como uma disciplina independente: a "*Data Science*". Este campo relaciona-se ao da Estatística e incorpora nessa relação, temas por ela abrangidos, como pesquisas multidisciplinares, modelos e métodos para dados, computação de dados, pedagogia, ferramenta de avaliação e teoria. Expressos em termos, esses e outros conceitos relacionados permitem definições que lhes dão respaldo em domínios

específicos, mas podem conjugar-se a modo de interdomínio (CLEVELAND, 2001).

É o caso, por exemplo, dos termos selecionados na constelação deste número da Revista Informação & Informação, que sob uma concepção abrangente dada ao título "*Data Science*: ciência orientada a dados" compõe um fascículo especial, com artigos de pesquisa relacionados aos temas: Curadoria de Dados de Pesquisa (*Research data curation*); Mineração de Dados no Âmbito Científico (*scientific research*); Compartilhamento e Reuso de Dados de Pesquisa (*Research data sharing* e *Research data reuse*); Ciberinfraestrutura de Apoio à *Data Science* (*cyberinfrastructure*); Dados Interligados (*Linked Data*); Publicações Ampliadas (*Enhanced Publications*) e Ferramentas para transformação e visualização de dados.

O apoio na literatura para a discussão dos principais aspectos dessa relação é justificado pelos especialistas no tema, que argumentam sobre a necessidade de maiores esforços para a compreensão dos fundamentos das atuais práticas de curadoria dos dados de pesquisa, por meio de uma compreensão de longo prazo. E, igualmente importante quanto compreender a natureza e atributos desta moderna configuração da dinâmica científica, aqui denominada orientada a dados, é apresentar um panorama daquilo que se foi publicado sobre a égide da referida temática. Destarte, o objetivo deste artigo é apresentar a produção científica mundial relativa à COD - aqui representados pelos termos "*e-Science*" e "*Data Science*" – a partir de uma perspectiva de análise centrada nas temáticas de maior destaque.

Este estudo utiliza métodos bibliométricos para analisar a produção científica sobre o tema. Realiza uma busca nas bases de dados *Web of Science* (*WoS*) e *Scopus*, limitada ao período mais profícuo para o tema que corresponde ao horizonte temporal entre 2006 e 2016 para identificar os estudos, autores, periódicos e temáticas mais proeminentes, categorizando-os e relacionando-os, com vistas à identificação de sua constelação temática. Justifica-se uma incursão mais ampla para esclarecimentos dos conceitos vinculados ao domínio, pois esse esforço conceitual favorece o reconhecimento das características e dos desafios encontrados pelos pesquisadores, no

esforço para compreender o potencial desses estudos para a produção e aperfeiçoamento dos métodos de pesquisas.

A contribuição deste estudo também está fundamentada na apresentação de um recente passado da COD, revelando um retrospecto das compreensões, proposições, experimentos e relações, almejando-se melhor entender o pretérito na intenção de vislumbrar possíveis horizontes de futuro.

2 QUADRO TEÓRICO

A leitura da produção científica voltada às questões teóricas sobre o tema inclui textos selecionados da literatura brasileira e internacional. Aqui são destacados apenas aspectos coincidentes com o propósito deste trabalho, que se volta ao entendimento e à clarificação de conceitos e termos associados ao domínio relativo à COD (*Data Science*), em sua evolução no período de pouco mais de dez anos.

Essa tentativa de enriquecer a discussão conceitual para a melhor compreensão das atividades relacionadas à *e-Science* é uma das dimensões da pesquisa de Karasti, Baker e Halkola (2006), em estudo etnográfico sobre os esforços para a constituição de um trabalho consistente por meio da curadoria de dados de pesquisa. Em cada contexto analisado no estudo, é identificado um conjunto de características salientes de pesquisa e dados ecológicos que se moldam pela rede de Pesquisa Ecológica em Longo Prazo (*Long Term Ecological Research - LTER*).

Considerando a sinergia entre o *Computer Supported Cooperative Work* (*CSCW*) e a *e-Science*, as autoras discutem o *compartilhamento* devido aos interesses de pesquisa similares relativos à colaboração científica mediada pela tecnologia e suas áreas de especialização complementares, concluindo que elas carecem de uma compreensão da perspectiva em longo prazo e da multiplicidade de escalas temporais.

Uma análise sobre o significado de "*compartilhar"* e de como o processo é realizado é empreendida por Cragin et al. (2010), que defendem a necessidade dos serviços de *curadoria de dados* para acomodar uma ampla

gama de características de dados subdisciplinares e práticas de compartilhamento. Como parte de um conjunto maior de estratégias emergentes entre as instituições acadêmicas, os repositórios institucionais (RI), segundo os autores, contribuem para a gestão e mobilização de dados de investigação científica, para a pesquisa e a aprendizagem.

Já o termo *Publicações reforçadas* (*enhanced publication*) remete a uma compreensão mais completa do processo pela qual os dados e as informações são utilizados e aplicados na geração de conhecimento. A definição de uma publicação reforçada é emprestada por Farace et al. (2012) ao projeto DRIVER-II, para significar uma publicação que é fortalecida com três categorias de informações: dados da pesquisa, materiais extras, e dados pós-publicação. As publicações, assim aprimoradas, contribuem inerentemente ao processo de revisão da literatura cinzenta, bem como à replicação de pesquisa, melhorando a visibilidade dos resultados da investigação na cadeia de comunicação científica conforme argumentam Farace et al. (2012).

Tendo como um dos destaques do seu trabalho o conceito de *publicação reforçada* e seu valor científico, Van Den Heuvel et al. (2010) destacam o reuso dos dados como altamente desejável para as pesquisas no que tange à verificação e à usabilidade de dados previamente coletados.

A reflexão sobre essas possibilidades parte de uma das seis experiências dos autores com dados de entrevistas por meio de *Enhanced Publications* (EP), concretizadas no projeto *Veteran Tapes* VP, sobre uma variedade de questões de pesquisa. As publicações eletrônicas dele resultantes permitem que as citações adicionais a um texto sejam disponibilizadas por meio de links que remetem às entrevistas com o texto transcrito.

Observa-se que os autores relacionam o conceito de publicação reforçada tanto aos processos de arquivamento e curadoria de dados de pesquisa, quanto às ferramentas de processamento de linguagem para facilitar a criação de EP e ainda às questões de direitos de propriedade intelectual relacionadas com a reutilização dos dados das entrevistas.

O conceito de publicação reforçada relaciona-se tematicamente com a ideia de que os resultados da investigação científica possam ser publicados e compartilhados em formatos estruturados. Conforme concepção de Marcondes e Costa (2016), o rastreamento desses resultados é realizado por agentes de software, graças à integração da web semântica à tecnologia dos dados interligados (*linked data*), conceito proposto por Bernes-Lee, em 2006, para se referir a um estilo de publicar e interligar dados estruturados de diferentes fontes na Web (MARCONDES; COSTA, 2016). Desse modo, argumentam os autores, os resultados da pesquisa científica podem ser publicados e compartilhados em formatos estruturados, ensejando a mineração por meio de *softwares*, recuperação semântica, reuso do conhecimento, validação dos conhecimentos científicos e identificação de traços de descobertas científicas.

Numa concepção abrangente, proposta por esta edição especial, os temas acima analisados representam parte importante das temáticas discutidas no universo da COD. De todo modo, ressalta-se que ainda permanecem incipiências conceituais que impedem o consenso acerca de algumas definições dos assuntos que circundam o tema. Costa e Cunha (2014), ao discutirem tais questões, afirmaram que é comum a aparição dos termos *e-Science*, ciência orientada a dados, computação fortemente orientada a dados e ciberinfraestrutura como sinônimos. Para este trabalho, adota-se a definição de *e-Science* proposta por Hey e Trefethen (2005) quando afirmam que a *e-Science* não pode ser considerada como uma nova disciplina científica; em vez disso, sua concepção deve estar ligada à infraestrutura capaz de permitir aos cientistas que seus trabalhos sejam desenvolvidos de modo mais preciso, mais rápido e/ou diferente, tendo forte relação com o incremento da produtividade e da inteligência científica através da partilha, replicação e gestão de dados científicos.

Quanto ao *Big data*, por mais que seja um tema recente na produção científica, sua essência remete a meados da década de 1940 com as discussões geradas pelo período subsequente à segunda guerra mundial. O fenômeno da “explosão da informação em Ciência & Tecnologia” é um exemplo disto, em que o foco dos esforços intelectuais foi canalizado para o

desenvolvimento de meios de organização da informação, destacando-se as iniciativas de Bush (1945). Atualmente, por mais que existam grupos consolidados centrados na organização da informação científica, a ênfase principal tem sido direcionada à capacidade de extração de conhecimento e descoberta de padrões em grandes volumes de dados científicos, aparentemente desconexos, armazenados em fontes, suportes e lugares diferentes, e aparentemente incomunicáveis. É deste desafio que parte o *Big data*.

Howe et al. (2008) ao discutirem o crescimento exponencial na quantidade de dados científicos em Biologia, apontam que são necessárias medidas revolucionárias para a gestão, análise e acessibilidade destes dados. Os autores propõem três ações para a solução desta problemática: 1) autores, revistas e curadores devem começar imediatamente a trabalhar em conjunto para facilitar o intercâmbio de dados entre as publicações de periódicos e bases de dados; 2) nos próximos cinco anos, os curadores, pesquisadores e administrações universitárias devem desenvolver uma estrutura de reconhecimento para facilitar os esforços de curadoria de base comunitária; 3) curadores, pesquisadores, instituições acadêmicas e agências de financiamento devem nos próximos dez anos, aumentar a visibilidade e apoio de curadoria científica como uma carreira profissional.

Por mais que um dos espaços temporais determinado pelos autores tenha expirado e o outro esteja por expirar, percebe-se que a problemática posta continua ativa e desafiadora às diversas áreas do conhecimento. Lidar com grandes volumes de dados científicos continua sendo uma árdua tarefa, haja vista, questões de padrões físicos e de armazenamento, indexação e descrição do conhecimento, cultura de acesso e uso, critérios de preservação e disseminação, políticas de seleção, adição de semântica aos dados, infraestrutura de processamento e várias outras questões não menos importantes no universo da COD.

O recorte teórico aqui analisado é representativo de uma parcela restrita da vasta literatura sobre o tema. Entretanto, foi estruturado com a intenção de

evidenciar as múltiplas relações apenas sugeridas introdutoriamente para que se compreendam como empiricamente elas se concretizam no corpus.

3 TRAJETÓRIA METODOLÓGICA

Distinguem-se cinco etapas relativas à trajetória metodológica desenvolvida para este artigo, a seguir enumeradas.

1) Busca de informações: a *priori*, buscaram-se os termos "Data Science" OR "e-Science", limitando-se a busca ao período de 2006 a 2016 e considerando-se apenas os artigos de periódicos. A restrição aos artigos de periódicos justifica-se pelo protagonismo das revistas científicas como fontes privilegiadas de produção de conhecimento, apresentando virtudes que as consagram como modelo exemplar de veículo científico, frutos de pesquisa e reflexão em estágios avançados e consolidados, avaliadas por comissões de distinto saber científico. Ademais, como se trata de um estudo temático, considera-se que os artigos publicados em periódicos apresentam uma maior diversidade de palavras-chave nas selecionadas bases, tornando os estudos desta natureza mais exequíveis.

As bases selecionadas para a busca foram a *Scopus* e a *WoS*. Para Chadegani et al. (2013) a *WoS* leva vantagem na literatura histórica, enquanto a *Scopus* privilegia a literatura recente, sendo superior em número de revistas indexadas, porém com impacto menor do que a sua concorrente. Ambas as bases de dados permitem pesquisar e classificar os resultados por parâmetros específicos, tais como, primeiro autor, citação, instituição, entre alternativas. Além disso, as duas bases proveem a possibilidade de analisar as informações a partir de *rankings* e métricas construídos com base no alto grau de sistematização dos dados contidos nas bases (CHADEGANI et al., 2013). Na *Scopus*, obtiveram-se 825 registros a partir da busca realizada, enquanto que, na *WoS*, foi possível localizar 639 registros.

2) Obtenção dos dados: em seguida, realizaram-se os *downloads* dos registros bibliométricos apresentados pelas bases de dados. O formato selecionado foi o texto sem formação, compatível com o *Microsoft Excel*®.

3) Complementação das palavras-chave: sabe-se que os dados provenientes destas bases já vêm indexados com palavras-chave atribuídas pela base e pelos autores dos artigos, porém, é comum que muitos registros não apresentem palavras-chave devido às normas de algumas revistas que não exigem tais descritores. Assim, de um total de 1464 registros, 420 (28,6%) não traziam palavras-chave, razão pela qual foram indexados por um profissional Mestre em Ciência da Informação com experiência nas referidas bases e no tema pesquisado. A política de indexação pautou-se em três palavras no mínimo e cinco no máximo. O assunto foi identificado pela leitura do título e do *abstract*. Na *WoS*, 78% dos artigos apresentavam palavras-chave, enquanto, na *Scopus*, 66% dos artigos estavam indexados. Após este trabalho, todos os registros passaram a contar com palavras-chave.

4) Correção e cruzamento dos dados: por meio da ferramenta *Vantage Point* (VP)[1], foi possível corrigir os registros e realizar eventuais agrupamentos de dados, tais como junção de palavras-chave com significados similares e autores que estavam representados com grafias diferentes, mudanças de nomes, nomes incompletos e abreviações, ocorrências prejudiciais aos resultados encontrados. O *VP* é uma ferramenta de mineração de textos para a organização do conhecimento resultante de busca em bases de dados. Sua principal funcionalidade é o estabelecimento de relações entre os dados, cruzando os campos, e propiciando a criação de matrizes legíveis por ferramentas de análise de redes sociais. Nesta etapa, os termos "e-Science" e "Data Science" foram removidos da lista de termos por se tratarem de palavras comuns a toda a produção.

5) Representação analítica dos dados: no intuito de apresentar os resultados, foram utilizadas as ferramentas *UCINET* e *NetDraw* (BORGATTI; EVERETT; FREEMAN, 2002). Por meio da técnica de análise de redes sociais as relações tornaram-se visíveis, o que permitiu a análise dos dados frente aos comportamentos verificados e aos referenciais teóricos adequados.

[1] https://www.thevantagepoint.com/: neste artigo foi utilizada a versão 9.0 do *software* no Laboratório Otlet CI da Universidade Federal de Pernambuco (UFPE).

Para a identificação dos termos vinculados às publicações da área de Ciência da Informação foram utilizados os filtros: *Information Science & Library Science* na *WoS* e *Social Sciences* na *Scopus*. Neste último caso, optou-se por filtrar apenas as revistas vinculadas à Ciência da Informação, reconhecendo os termos "*information*" e "*library*" no título e analisando o escopo dos periódicos visitando os sites. Com isto, restaram 52 artigos na *Scopus* e 29 na *WoS*. Para a criação da nuvem de *tags* utilizou-se o *software VP*.

Ademais, ressalta-se que na *Scopus* deu-se preferência às palavras-chave definidas pela base (*index keywords*), enquanto na *WoS* optou-se por utilizar as palavras-chave definidas pelos autores e validadas pelas revistas (*keywords*). O motivo de tal distinção é que as palavras-chave produzidas pela *WoS*, denominadas *Keyword Plus* são extremamente genéricas e não representam o conteúdo de maneira adequada, pois, como afirmam Hoppen e Vanz (2014), estas são geradas pela *WoS* a partir de algoritmo que extrai palavras ou expressões significativas de todos os títulos citados nos *papers* da pesquisa (GARFIELD, 1990; THOMSON REUTERS, 2010). Buscando-se minimizar as discrepâncias entre os universos analisados, todos os descritores foram revisados pelo profissional indexador, que atribuiu palavras-chave aos registros que não dispunham delas. Com isto, assegurou-se a qualidade dos termos obtidos para esta pesquisa.

4 ANÁLISE E DISCUSSÃO

Esta seção está dividida em quatro etapas, a primeira exibe a evolução cronológica dos temas em destaque nos tópicos "e-Science" e "Data Science" na *Scopus* e na *WoS,* no decorrer dos anos analisados (2006 a 2016). A segunda diz respeito à análise dos dados da *Scopus*, enfatizando as relações entre os temas, autores e temas, periódicos e temas. A terceira destaca estas mesmas relações, porém na base *WoS*. Por fim, na última etapa, apresentam-se as análises dos temas mais representativos, somando-se as palavras-chave das duas bases estudadas, todavia, considerando apenas a área de Biblioteconomia e Ciência da Informação (BCI).

Tabela 1 - Evolução cronológica dos temas em destaque nos tópicos de *e-Science* e *Data Science* na *Scopus* e na *WoS* ao longo dos anos analisados (2006 a 2016)*

Ano	Temas (Scopus)	Temas (WoS)
2006 (41)	Distributed computer systems (8) Grid computing (7) World Wide Web (7)	Distributed computer systems (9) Semantic web (7) Ontology (4)
2007 (57)	Grid computing (17) Semantics (7) Research (5)	Grid computing (8) Web service (3) Workflow (3)
2008 (64)	Grid computing (15) Distributed computer systems (9) Software (6)	Grid computing (12) Workflow (6) Semantic web (6)
2009 (97)	Grid computing (23) Research (10) Management (9)	Grid computing (16) Workflow (6) Web (6)
2010 (72)	Grid computing (18) Management (11) Software (11)	Grid computing (13) Workflow (9) Cyberinfrastructures (6)
2011 (72)	Grid computing (14) Research (11) Software (10)	Grid computing (13) Data provenance (7) Workflow (4)
2012 (56)	Information Management (7) Grid computing (6) Software (6)	Grid computing (6) Workflow (3) Data provenance (2)
2013	Information Management (9)	Workflow (5)

(73)	Research (9) Grid computing (7)	Research (4) Grid computing (3)
2014 (102)	Big data (11) Human (9) Information Management (7)	Big data (9) Workflow (5) Data provenance (5)
2015 (124)	Big data (27) Human (14) Data mining (9)	Big data (9) Machine learning (8) Workflow (7)
2016 (65)*	Big data (8) Human (6) Distributed computer systems (6)	Big data (7) Workflow (3) Cloud computing (3)

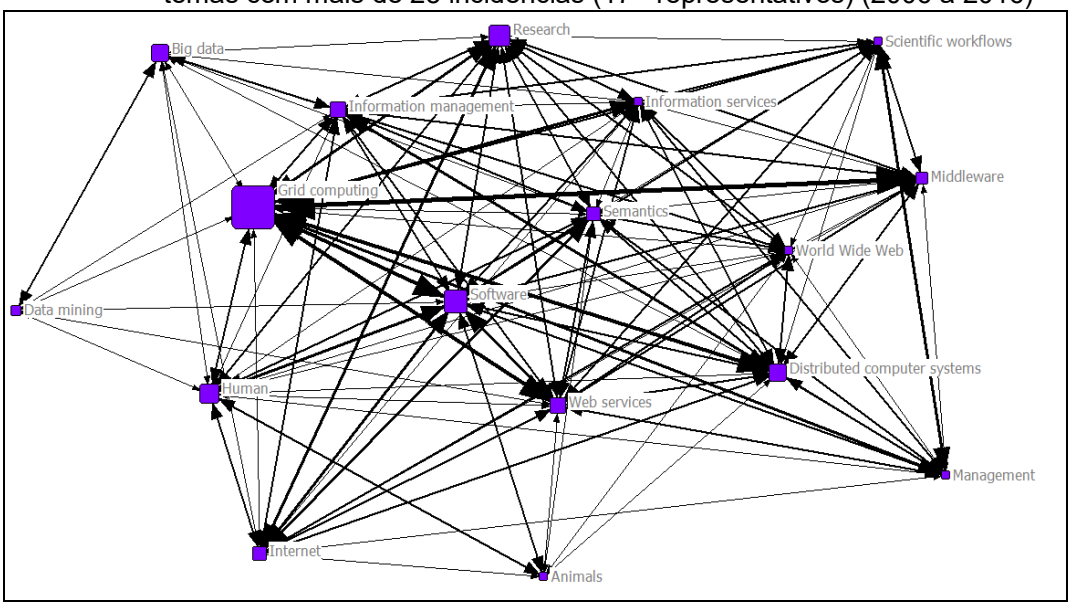
Fonte: dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.

O conjunto de termos apontados na Tabela 1 traz uma meta representação do conjunto de artigos indexados nas bases *WoS* e *Scopus* a partir do universo definido neste artigo. Sabe-se que quaisquer representações, por definição, acarretam considerável perda da essência motriz do objeto representado. Dito isto, o conjunto das relações apresenta uma linha de tempo dos termos mais recorrentes em ambas as bases. Como estratégia de análise, buscou-se perceber inicialmente termos coincidentes por ano e por base, procurando-se ainda, contextualizá-los em cada período.

Considerando-se o recorte cronológico do *corpus* analisado, os primeiros estudos focaram em *Distributed computer systems*, seguidos pelos trabalhos de *Grid computing*. O segundo tornou-se proeminente ao longo do período de 2007 a 2013. Ambos os temas caracterizam-se como modelos computacionais orientados à distribuição de tarefas com o intuito de alcançarem uma alta capacidade de processamento de dados em estruturas globais. Neste período - de 2006 a 2013 – percebe-se então a coexistência destes conceitos com processos de gestão e fluxos de trabalho, destacando-se os desígnios da web

semântica. A partir de 2014, os estudos relacionados a *Big Data* despontam de forma expressiva, sabe-se que o conceito referente ao mencionado termo é convergente à tecnologia de *Grid computing*, sendo o *Big Data* um estágio que envolve processamento/descobertas a partir de dados advindos de uma infraestrutura possibilitada pela segunda – a *Grid computing*.

Figura 1 - Relações temáticas nos tópicos de *e-Science* e *Data Science* na *Scopus*: temas com mais de 25 incidências (17+ representativos) (2006 a 2016)*



Fonte: Dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.
Nota: O tamanho das nós é proporcional à representatividade quantitativa dos termos e a intensidade das linhas é proporcional às relações.

Na figura 1 nota-se que são mais intensas as correlações entre os seguintes pares de termos: *Middleware* e *Grid computing* (19), *Animals* e *Nonhuman* (18), *Grid computing* e *Web services* (16), *Distributed computer systems* e *Grid computing* (13), *Grid computing* e *Software* (12), *Internet* e *Software* (11), *E-research* e *Research* (11), *Information services* e *Grid computing* (11), *Semantic web e Semantics* (11), *Management* (11) e *Scientific Workflows*.

Ponderando-se a maior incidência dos termos, o quadro 1 evidencia que as temáticas com maior destaque foram *Grid computing, Software*, *Research*,

Human, *Big data* e *Distributed computer systems*, como se pode observar no Quadro 1.

Quadro 1 - Temas com maior destaque nos tópicos de *e-Science* e *Data Science* na *Scopus*

>100	*Grid computing* (116)
63 a 50	*Software* (63), *Research* (59), *Human* (53), *Big data* (50), *Distributed computer systems* (50)
43 a 30	*Information management* (43), *Web services* (42), *Internet* (39), *Semantics* (37), *Middleware* (35), *Data mining* (30)
26 a 23	*Animals* (25), *Information services* (26), *Scientific workflows* (26), *World Wide Web* (26), *Management* (25), *Computer Models* (23), *Computer simulation* (23), *Metadata* (23), *Ontology* (23)

Fonte: Dados da Pesquisa (2016).

Outros temas se destacaram, porém, visando proporcionar uma melhor representação visual da figura 1, foram suprimidos. São eles: *Bioinformatics* (22), *Cloud computing* (22), *Cyber infrastructures* (22), *E-research* (22).

Chama a atenção a presença dos termos *Animals* (25) e *Non human* (22) nas relações de maior destaque. Todavia, ao se analisar o contexto de aparição dos termos, ficou evidente que estes se referem, especialmente, à utilização de dispositivos implantáveis sem fio para monitoramento e coleta de dados fisiológicos. A empresa líder neste segmento chama-se *Data Science International* (*DSI*)[2] e desenvolve pesquisa biomédica focada em sistemas de fisiologia e farmacologia. Seu trabalho é oferecer equipamentos de telemetria, instrumentação, *software*s e serviços que contribuam para o avanço da ciência. Conforme destacam Pagés et al. (2009), esta técnica é especialmente útil no campo da farmacologia, toxicologia e fisiologia, havendo modelos de sensores que permitem a implantação em animais de vários tamanhos, como ratos, cães e porcos. Desse modo, assim podem ser explicadas as relações enfatizadas entre *Animals* e *Non human* neste estudo.

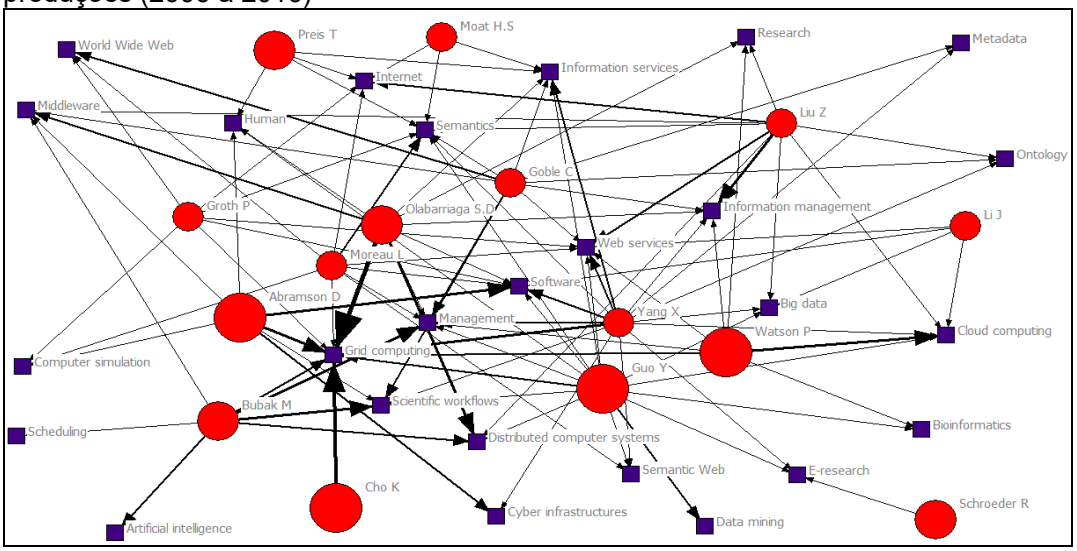
[2] www.datasci.com

No mais, com base nos resultados encontrados na tabela 1, torna-se evidente que os estudos pressupunham ser profícua a *e-Science* e/ou *Data Science* a construção de um arcabouço teórico e aplicado que unisse estudos, majoritariamente com vieses tecnológicos, que buscasse em recursos de *hardware* e *software*, o estabelecimento de uma estrutura de compartilhamento e processamento de dados orientados à pesquisa.

O horizonte temporal contemplado na coleta de artigos coincide com um momento de grandes transformações no *modus operandi* cientifico. As dinâmicas dos ambientes e comunidades científicas mantiveram um processo de incorporação das tecnologias de informação iniciado com o advento dos computadores e das de redes de comunicação, movimento este desencadeado mais intensamente a partir dos anos de 1990. Os números mais proeminentes evidenciam que o referido processo coaduna com a perspectiva de projetos coletivos e universais de pesquisa pautados na distribuição de esforços cognitivos e de processamento/distribuição de dados.

A seguir, são exibidas as principais relações entre temas e autores na temática analisada.

Figura 2 - Relações entre temas e autores nos tópicos de *e-Science* e *Data Science* na *Scopus*: grupo dos 30 temas mais representativos e autores com mais de 5 produções (2006 a 2016)*



Fonte: Dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.
Nota: O tamanho dos nós é proporcional à representatividade quantitativa dos termos e a intensidade das linhas é proporcional às relações.

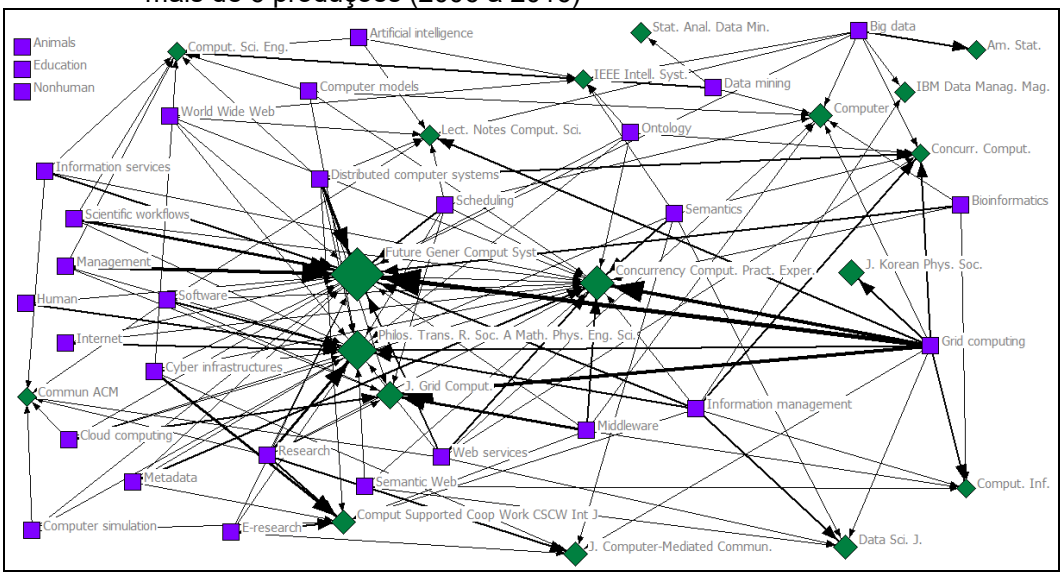
As relações mais intensas entre autores e temas são: Olabarriaga S.D *e Grid computing (5);* Cho K *e Grid computing (4);* Liu Z *e Information management (3);* Olabarriaga S.D *e Distributed computer systems (3);* Abramson D *e Software (3);* Abramson D *e Grid computing (3);* Bubak M *e Management;* Bubak M *e Scientific workflows (3);* Watson P *e Cloud computing (3).*

Os autores mais representativos são: *Abramson D, Cho K, Guo Y e Watson P com 7 artigos; Bubak M, Olabarriaga S.D, Preis T e Schroeder R com 6 produções; e por fim, Goble C, Groth P, Li J, Liu Z, Moat H.S, Moreau L e Yang X* com 5 artigos.

Ao verificar o perfil de pesquisa dos 4 autores mais representativos no Google Scholar (*GS*), percebeu-se que: David Abramson (*University of Queensland*) trabalha com alta performance de computação distribuída, seu trabalho mais citado é *Nimrod/G: An architecture for a resource management and scheduling system in a global computational grid*, assim como a temática de *Grid computing* e suas citações estão decrescendo desde 2010, provavelmente pela inversão da tendência de pesquisa já apontada nas relações temáticas, em que o *Grid computing* tem sido superado pelo *Big data* na produção de conhecimento; Kihyeon Cho (*Korea Institute of Science Technology*) não possui perfil no *GS* e sua produção versa sobre os temas Física de altas energias, ciberinfraestrutura, processamento de dados e *e-Science*; Yike Guo (*Imperial College London*) tem como temas principais a mineração de dados, o aprendizado de máquina e a inteligência artificial e suas citações vêm evoluindo desde 2012; Paul Watson (*Newcastle University*) tem como temas principais bases de dados, *web services* e computação em grade. Suas citações desde 2009 vêm caindo sistematicamente no *GS*.

A figura 2 reitera as observações anteriores que apontaram a maciça presença da temática G*rid computing* e respectivas interfaces com outros temas. Tal apontamento parte do princípio de que os autores que focaram suas publicações na mencionada temática elaboraram estudos contemplando também outras discussões, como por exemplo: *Scientific workflows*, *Management* e *Semantics*.

Figura 3: Relações entre temas e periódicos nos tópicos de *e-Science* e *Data Science* na *Scopus*: grupo dos 30 temas mais representativos e dos periódicos com mais de 6 produções (2006 a 2016)*



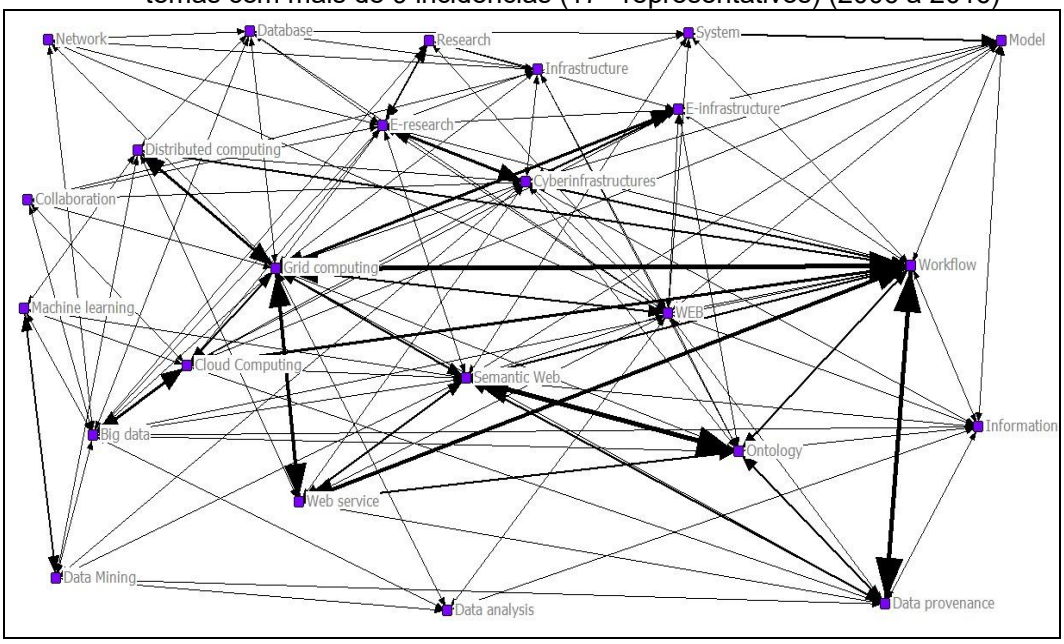
Fonte: Dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.
Nota: O tamanho das nós é proporcional à representatividade quantitativa dos termos e a intensidade das linhas é proporcional às relações.

As relações mais expressivas entre temas e periódicos são: *Grid computing* e *Future Gener Comput Syst* (12), *Management* e *Future Gener Comput Syst* (9), *Grid computing* e *Concurrency Comput. Pract. Exper*. (9), *Grid computing* e *J. Grid Comput*. (8), *Distributed computer systems* e *Future Gener Comput Syst* (7), *Research* e *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci* (7), *Scientific workflows* e *Future Gener Comput Syst* (6), *Cyber infrastructures* e *Comput Supported Coop Work CSCW Int J* (6) e *Middleware* com *J. Grid Comput.* (6).

Os periódicos mais representativos são: *Future Gener Comput Syst* (34), *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* (23), *Concurrency Comput. Pract. Exper.* (18), *J. Grid Comput.* (13), *J. Korean Phys. Soc.* (12), *Comput Supported Coop Work CSCW Int J* (10), *Computer* (10), *Data Sci. J.* (10), *J. Computer-Mediated Commun.* (10), *IBM Data Manag. Mag.* (7), *Lect. Notes Comput. Sci.*(7), *Stat. Anal. Data Min.* (7), *Am. Stat.* (6), *Commun ACM* (6), *Comput. Inf.* (6), *Comput. Sci. Eng.* (6), *Concurr. Comput.* (6), *IEEE Intell. Syst.* (6).

Dos periódicos representados na Figura 3, o *Future Gener Comput Syst* abarca boa parte das temáticas consideradas no extrato da análise feita neste artigo. Porém, não menos importantes também foram o *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci* e o *Concurrency Comput. Pract. Exper.* Desta forma, do ponto de vista de cobertura, os três mencionados títulos distinguem-se num núcleo de periódicos mais devotados ao tema, conforme sinaliza a Lei de Bradfort, também conhecida como lei da dispersão.

Figura 4 - Relações temáticas nos tópicos de *e-Science* e *Data Science* na *WoS*: temas com mais de 9 incidências (17+ representativos) (2006 a 2016)*



Fonte: dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.
Nota: O tamanho das nós é proporcional à representatividade quantitativa dos termos e a intensidade das linhas é proporcional às relações.

Ao se analisar a *WoS,* observou-se que existem intensas correlações entre as seguinte duplas de termos: *Semantic web* e *Ontology* (11), *Workflow* e *Data provenance* (10), *Grid computing* e *Workflow* (9), *Grid computing* e *Web service* (7), *Workflow* e *Web service* (7), *Distributed computing* e *Grid computing* (6), *E-research* e *Cyberinfrastructures* (5), *E-research* e *Grid computing* (5), *Cloud computing* e *Workflow* (5), *Big data* e *Cloud computing* (5).

Desta feita, os temas com maior destaque foram *Grid computing*, *Workflow* e *Big data.*

Quadro 2 - Temas com maior destaque nos tópicos de *e-Science* e *Data Science* na *WoS*

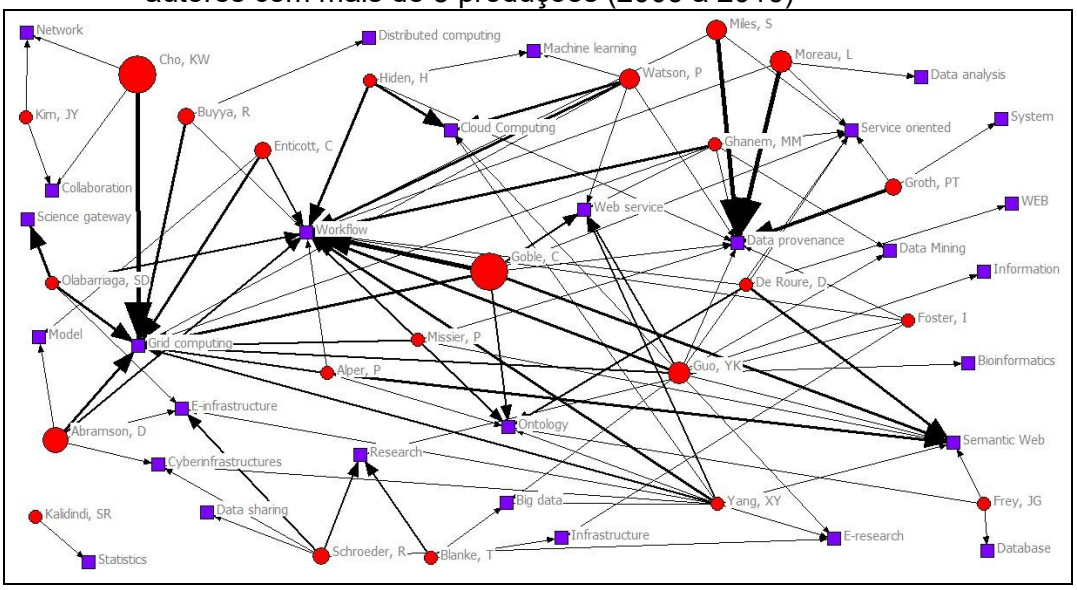
>80	*Grid computing* (87)
52 a 40	*Workflow* (52) e *Big data* (40)
31 a 20	*Semantic web* (31), *Data provenance* (29), *Ontology* (26), *Research* (23), *Web service* (20),
19 a 13	*Cloud Computing* (19), *Cyberinfrastructures* (19), *Model* (17), *Web* (15), *E-research* (14), *Database* (13), *Machine learning* (13).

Fonte: Dados da Pesquisa, 2016.

Assim como na base *Scopus*, foi evidenciada a forte presença dos termos *Grid computing* e *Big data*, confirmando as tendências nas transformações no *modus operandi* científico. Chamou a atenção, a alta posição do termo *Workflow* no *ranking.* Para Deelman (2009), um *workflow* é uma especificação de alto nível de um conjunto de tarefas dependentes a fim de satisfazer um objetivo específico. Pode ser, por exemplo, um protocolo de análise de dados que consiste em uma sequência de pré-processamento, análise, simulação e etapas de pós-processamento, sendo muito utilizado em aplicações ligadas ao *e-Science*. Ainda segundo os autores, do ponto de vista da execução, um *workflow* pode ser operacionalizado como um programa de computador (DEELMAN, 2009).

Assim, entende-se que os *workflows* estão estritamente ligados às temáticas de *e-Science* e *Data Science,* principalmente no que diz respeito ao desenvolvimento de modelos computacionais capazes de executar rotinas de obtenção e distribuição de dados científicos.

Figura 5 - Relações entre temas e autores nos tópicos de *e-Science* e *Data Science* na *WoS*: grupo dos 30 temas mais representativos e autores com mais de 5 produções (2006 a 2016)*



Fonte: Dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.
Nota: O tamanho das nós é proporcional à representatividade quantitativa dos termos e a intensidade das linhas é proporcional às relações.

As interações mais densas entre temas e autores são: Cho, KW e *Grid computing* (6), Goble, C e *Workflow* (6), Miles, S e *Data provenance* (6), Moreau, L e *Data provenance* (6), Groth, PT e *Data provenance* (5), De Roure, D e *Semantic Web* (4), Goble, C e *Semantic Web* (4), Guo, YK e *Workflow* (4), Buyya, R e *Grid computing* (4), Olabarriaga, SD e *Science gateway* (4), Olabarriaga, SD e *Grid computing* (4), Abramson, D e *Grid computing* (4).

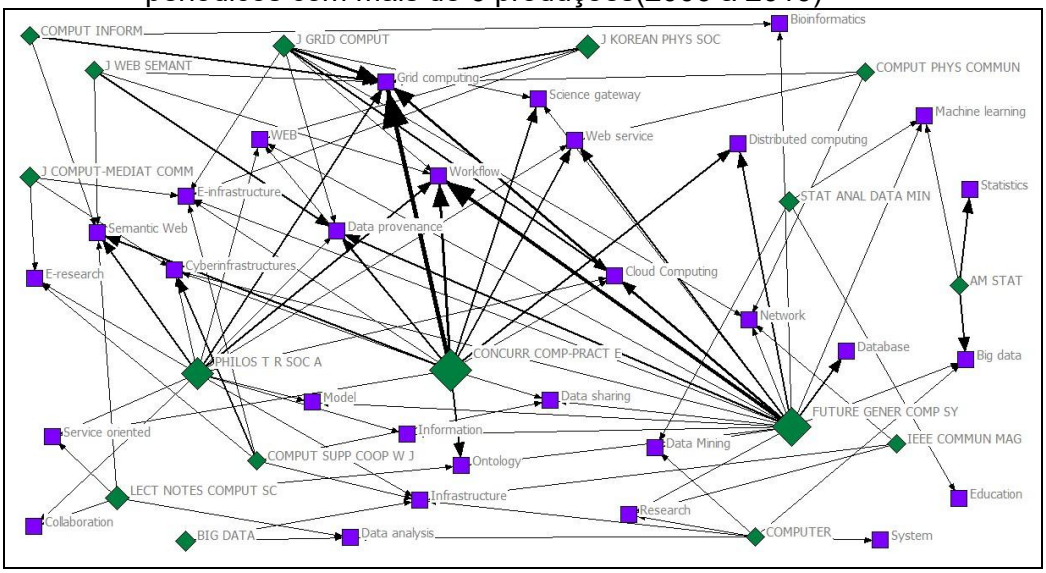
Os autores mais produtivos são: Cho, KW. e Goble, C. apresentaram 13 artigos, Abramson, D. com nove produções, Guo, YK. e Moreau, L. com oito artigos, Miles, S. e Watson, P. com sete, Buyya, R., Enticott, C., Groth, PT. e Schroeder, R. com seis, e Alper, P.,Blanke, T., De Roure, D., Foster, I., Frey, JG., Ghanem, MM., Hiden, H., Kalidindi, SR., Kim, JY., Missier, P., Olabarriaga, SD. e Yang, XY com cinco.

Dentre os quatro autores mais produtivos na *WoS*, três coincidem com os quatro primeiros encontrados na *Scopus*, o que aponta para dois comportamentos: primeiro, existe uma grande intersecção de artigos entre a *WoS* e a *Scopus*, segundo, os autores que estão inseridos nas bases de dados

que representam a elite do conhecimento científico nas temáticas aqui analisadas tendem a reafirmar seu prestígio em diferentes *rankings* através do forte grau de inserção de suas produções em veículos indexados. Para estar inserido nas bases internacionais de conhecimento científico é preciso ter mais do que talento, faz-se necessário também entender os esquemas de publicação das bases e conhecer os temas e formatos que são privilegiados pelas revistas que nelas estão indexadas.

A autora com maior destaque na *WoS* que não figura entre os quatro primeiros da *Scopus* é Carole Goble (*University of Manchester*). Seus principais temas de atuação são: *semantic web*, *bioinformatics*, *e-Science*, *social computing*, *workflows*. Seu artigo com maior número de citações no *GS* é o texto *Taverna: a tool for building and running workflows of services*. Suas citações no GS vêm apresentando uma leve queda desde 2012.

Figura 6 - Relações entre temas e periódicos nos tópicos de E-science e Data Science na WoS: grupo dos 30 temas mais representativos e dos periódicos com mais de 6 produções(2006 a 2016)*



Fonte: Dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.
Nota: O tamanho das nós é proporcional à representatividade quantitativa dos termos e a intensidade das linhas é proporcional às relações.

As interações entre periódicos e temas mais latentes são: *Concurr Comp-Pract e Grid computing* (16)*, Future Gener Comp SY* (11), *Concurr*

Comp-Pract e Workflow (9), *J Grid Comput e Grid computing* (8), *Future Gener Comp SY e Grid computing* (7), *Future GenerComp SY e Data provenance* (6), *J Korean PhysSoc e Grid computing* (6).

Os periódicos mais representativos são: *Concurr Comp-Pract* (45), *Future Gener Comp SY* (40), *Philos T R Soc A* (27), *Lect Notes Comput SC* (14), *J Korean Phys Soc* (13), *J Grid Comput* (12), *J Comput-MediatComm* (10), *Big Data* (9), *Stat Anal Data Min* (9), *Comput Inform* (7), *J Web Semant* (7), *AM Stat* (6), *Comput Phys Commun* (6), *ComputSupp Coop W J* (6), *Computer* (6), *IEEE Commun Mag* (6).

Dos periódicos representados na Figura 6, o *Concurr Comp-Pract* e o *Future Gener Comp SY* são os mais representativos dos temas considerados neste artigo, obtendo destaque nos resultados, por oferecerem maior cobertura dos temas encontrados.

Visando contextualizar os resultados obtidos com a área de BCI, apresentam-se abaixo os temas de maior destaque nesta área do conhecimento segundo os achados apontados pelas bases analisadas.

Figura 7 - Nuvem de *tags* gerada a partir dos tópicos *e-Science* e *Data Science* na *WoS* e *Scopus* apenas na área de Biblioteconomia e Ciência da Informação: termos com mais de 3 incidências (2006 a 2016)*

Fonte: Dados da Pesquisa, 2016.
* Os dados de 2016 estão limitados ao mês de junho de 2016.

Associando a temática discutida neste artigo à área de BCI, a figura 7 indica a alta incidência dos termos *Digital library, Open access, Research e Library*. Para Jeng (2005), as bibliotecas digitais são coleções digitais de informações geridas e organizadas, acessíveis através de uma rede e podem

incluir serviços diversos. Deste modo, para a área de BCI, as bibliotecas digitais são os espaços em que a e-Science e o Data Science se apresentam de maneira mais profícua, devido à natureza digital de suas operações. O cerne desta discussão é o acesso a conteúdos científicos digitais disponíveis em acervos que rompem com os métodos analógicos e tradicionais de compartilhamento de informações, permitindo aos pesquisadores a superação dos limites do ambiente físico.

O segundo tema em termos de frequência é o *open access*, que consiste na criação de estratégias para o compartilhamento de informações, partindo do pressuposto de que o acesso aos conteúdos científicos precisa ser livre, em favor do progresso da Ciência. Ao analisar detalhadamente os registros, percebe-se que os principais temas associados ao acesso aberto são: publicações científicas, políticas científicas, *software* livre, repositórios, bases de dados, gestão de dados, preservação e pesquisa.

Alguns temas não menos importantes, que se relacionam com a discussão deste artigo são: *Library, Research, Science, Research data, Grid Computing, Open data, Big data, E-research e Collaborate*.

5 CONSIDERAÇÕES FINAIS

Este artigo teve como propósito apresentar a produção científica mundial relativa à COD - aqui representada pelos termos “e-Science” e “Data Science”, considerando-se um corpus constituído por artigos de periódicos indexados na *Scopus* e na *WoS* (2006 a 2016).

A evolução cronológica dos temas em destaque nos tópicos “*e-Science*” e “*Data Science*” na *Scopus* e na *WoS*, entre 2006 e 2016 permitiu perceber que até 2013 havia uma grande preocupação dos pesquisadores com as questões ligadas à computação em grade (*Grid computing*), enquanto que, a partir de 2014 o foco dos trabalhos esteve mais ligado ao *Big data*. Conforme indicaram Cao et al. (2003), a computação em grade é uma tecnologia para o compartilhamento de recursos em larga escala distribuída. Quanto ao *Big data*,

este consiste na tarefa de encontrar padrões em grandes volumes de dados (SHIFFRIN, 2016).

Foram também analisadas as relações dos temas entre si, entre autores e temas, periódicos e temas, tanto na base *Scopus* quanto na base *WoS* e analisados os temas mais representativos, das duas bases estudadas, considerando apenas as áreas de Biblioteconomia e Ciência da Informação (BCI).

Vale destacar que, entre os quatro autores mais produtivos encontrados na *WoS*, três estão entre os quatro primeiros encontrados na *Scopus*, o que indica sua representatividade e prestígio, graças a forte inserção de suas produções em veículos indexados. Tal fato também é um indicativo do alto grau de intersecção da produção indexada nas duas bases.

Sobre os periódicos mais representativos, ficou evidente a grande contribuição da área de Tecnologia da Informação (TI), sobretudo, com os periódicos *Future Gener Comput Syst* (*destaque na Scopus*) e *Concurr Comp-Pract* (destaque na *WoS*).

Pode-se considerar como ponto restritivo deste estudo a limitação da análise à tipologia documental de artigos. Tal opção impossibilitou que os trabalhos publicados em anais de congresso (veículo científico mais utilizado na área de TI) fossem contemplados. E, conforme se percebeu, a área em pauta tem se destacado nos estudos referentes a *e-Science* e *Data Science*. Por outro lado, o recorte tipológico permitiu o aprofundamento das análises temáticas e, se considerado o trâmite padrão das comunicações científicas, o formato do artigo é antecedido pelos *preprints* e pelas apresentações em eventos, fato gerador de conteúdos repetidos quando se analisam todos os formatos comunicacionais científicos.

No que tange à Ciência da Informação, verificou-se que a área de BCI demonstrou uma identidade própria na ênfase dada aos estudos das temáticas analisadas. Tendo sua origem histórica vinculada ao campo da Biblioteconomia, a BCI apresentou como temas mais representativos as questões ligadas às bibliotecas digitais e acesso aberto, temas que são tendências para os campos em questão.

REFERÊNCIAS

BORGATTI, Stephen P.; EVERETT, Martin G.; FREEMAN, Lin C. **UCINET for windows**: software for social network analysis. Harvard: Analytic Technologies, 2002.

BUSH, Vannevar. As we may think. **Atlantic Monthly**, Washington, v. 176, n. 1, p. 101-108, 1945.

CAO, Junwei et al. GridFlow: workflow management for grid computing. In: CLUSTER COMPUTING AND THE GRID, 3., 2003, Washington. **Proceedings...** IEEE/ACM, 2003. p. 198-205. Disponível em: <http://dx.doi.org/10.1109/ CCGRID.2003.1199369>. Acesso em: 1 jul. 2016.

CHADEGANI, Arezoo Aghaei et al. A comparison between two main academic literature collections: Web of Science and Scopus databases. **Asian Social Science**, Toronto, v. 9, n. 5, p. 18-26, 2013. Disponível em: <https://arxiv.org/ftp/arxiv/ papers/1305/1305.0377.pdf>. Acesso em: 23 jun. 2016.

CLEVELAND, William S. Data science: an action plan for expanding the technical areas of the field of Statistics. **International Statistical Review**, Malden, v. 69, n. 1, p. 21-26, Apr. 2001.

COSTA, Maira Murriera; CUNHA, Murilo Bastos. O bibliotecário no tratamento de dados oriundos da e-Science: considerações iniciais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 19, n. 3, p. 189-206, set. 2014. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362014 000300010&lng=en&nrm=iso>. Acesso em: 14 jul. 2016.

CRAGIN, Melissa H. et al. Data sharing, small science and institutional repositories. **Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences**, v. 368, n. 1926, p. 4023-4038, 2010. Disponível em: <http://rsta.royalsocietypublishing.org/content/368/1926/4023.short> Acesso em: 13 jul. 2016.

DEELMAN, Ewa et al. Workflows and e-Science: An overview of workflow system features and capabilities. **Future Generation Computer Systems**, Amsterdam, v. 25, n. 5, p. 528-540, 2009.

FARACE, Dominic et al. Linking full-text Grey Literature to underlying research and post-publication data: An Enhanced Publications Project 2011-2012. **Grey Journal (TGJ)**, v. 8, n. 3, 2012.

FIENBERG, Stephen E.; MARTIN, Margaret E.; STRAF, Miron L. (Ed.). **Sharing research data.** Washington: National Academy Press, 1985.

GARFIELD, Eugene. Keywords plus: ISI's breakthrough retrieval method. Part I. Expanding your searching power on current contents on diskette. **Current Contents**, Woodbury, n. 32, p. 5-9, 6 ago. 1990.

JENG, Judy. What is usability in the context of the digital library and how can it be measured?. **Information technology and libraries**, v. 24, n. 2, 2005.

HEY, Tony; TREFETHEN, Anne Elizabeth. Cyberinfrastructure for e-Science. **Science**, New York, v. 308, n. 5723, p. 817-821, 2005.

HOPPEN, Natascha Helena Franz; VANZ, Samile Andréa de Souza. Tendências da pesquisa brasileira em neurociências. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 4., 2014, Recife. **Anais...** Recife: UFPE, 2014. p. 1-4. Disponível em: <http://basessibi.c3sl.ufpr.br/brapci/_repositorio/2014/05/pdf_c714184cab_0014376.pdf>. Acesso em: 27 jun. 2016.

HOWE, Doug et al. Big data: the future of biocuration. **Nature**, London, v. 455, n. 7209, p. 47-50, 2008.

KARASTI, Helena; BAKER, Karen S.; HALKOLA, Eija. Enriching the notion of data curation in e-Science: data managing and information infrastructuring in the long term ecological research (LTER) network. **Computer Supported Cooperative Work (CSCW),** Netherlands, v. 15, n. 4, p. 321-358, ago. 2006.

MARCONDES, Carlos H.; COSTA, Leonardo C. A model to represent and process scientific knowledge in biomedical articles with semantic web technologies. **Knowledge Organization**, Wurzburg, v. 43, n. 2, p. 86-101, 2016.

PAGÉS, Teresa et al. **Monitorización por telemetría**: implantación subcutánea de dispositivos telemétricos en rata: preparación terminal antena negativa. 2009. Disponível em: <http://www.mdx.cat/handle/10503/8894?locale-attribute=es>. Acesso em: 24 ago. 2016.

SHIFFRIN, Richard M. Drawing causal inference from big data. **Proceedings of the National Academy of Sciences**, Washington, v. 113, n. 27, p. 7308-7309, 2016.

THOMSON REUTERS. **Web of science**: quick reference card. Philadelphia: Thomson Reuters, 2010.

VAN DEN HEUVEL, Henk et al. The veterantapes: research corpus, fragment processing tool, and enhanced publications for the e-humanities. In: INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, 7., 2010, Malta. **Proceedings...** Malta, 2010. p. 2687-2692. Disponível em: <http://www.lrec-conf.org/proceedings/lrec2010/pdf/328_Paper.pdf> Acesso em: 27 jun. 2016.

Title

International production on science oriented towards data: analysis of the terms data science and e-science in scopus and the web of science

Abstract
Introduction: current configuration in the dynamics of production and scientific communication reveals the role of Science Oriented Towards Data, a comprehensive conception represented, mainly, by terms such as "e-Science" and "Data Science".
Objective: To present the global scientific production on Science Oriented Towards Data by using the terms "e-Science" and "Data Science" in Scopus and the Web of Science during 2006-2016.
Methodology: The study is divided into five phases: a) search for information in Scopus and the Web of Science data bases; b) obtaining bibliometric records; c) complementing keywords; d) data correction and crossing; e) analytical data representation.
Results: The most important terms within the analyzed scientific production were Distributed computer systems (2006), Grid computing (2007-2013) and Big data (2014-2016). In the area of Library and Information Science, the emphasis was on Digital Library and Open Access issues, highlighting the importance of the field for the discussions on the devices providing access to scientific information in digital media.
Conclusions: Under a diachronic look, it was found a visible shift of focus, from issues approaching data exchange operations to an analytical perspective for finding patterns in large data volumes.
Keywords: Data Science. E-Science. Science oriented towards data. Scientific production.

Título

Producción internacional sobre la ciencia orientada a los datos: análisis de los términos Data Science e e-Science en Scopus y la Web of Science

Resumen
Introducción: La configuración actual en la dinámica de la producción y la comunicación científica revela el protagonismo de la Ciencia Orientada a los Datos, una concepción integral representada, principalmente, por términos como "e-Science" y "Data Science".
Objetivo: Presentar la producción científica mundial relativa a la Ciencia Orientada a los Datos a partir de los términos "e-Science" y "Data Science" en Scopus y la Web of Science en el período 2006 - 2016.
Metodología: El estudio se divide en cinco etapas: a) búsqueda de información en las bases de datos Scopus y la Web of Science; b) obtención de los registros bibliométricos; c) complementación de las palabras clave; d) corrección y cruce de los datos; e) representación analítica de los datos.
Resultados: Los términos más importantes en la producción científica analizada fueron Distributed computer systems (2006), Grid computing (2007-2013) e Big data (2014-2016). En el área de Biblioteconomía y Ciencia de la Información, se hace énfasis en los temas Digital Library y Open Access, evidenciando la importancia del

campo en las discusiones sobre los dispositivos para proporcionar acceso a la información científica en los medios digitales.
Conclusiones: Bajo una mirada diacrónica, se constata un cambio visible de la atención que se prestaba a las temáticas enfocadas en las operaciones de intercambio de datos, para una perspectiva analítica de búsqueda de patrones en grandes volúmenes de datos.

Palabras clave: Data Science. E-Science. Ciencia orientada a los datos. Producción científica.

Enviado em: 17.07.2016
Aceito em: 20.11.2016.