

Perfil e competências do profissional da Informação para a gestão de dados em massa (Big data)

Luísa Alvim

CIDEHUS - Universidade de Évora, Portugal, mluisa.alvim@gmail.com

Resumo: Neste artigo discute-se o perfil e as competências para o profissional da informação face ao fenómeno dos dados em massa (*Big data*). O conceito de gestão de dados em massa é utilizado no âmbito da informação e documentação e sistemas de informação. Considera-se que os dados em massa são uma tendência no futuro que deve ser percebida pelas instituições privadas e públicas. Apresenta-se uma revisão bibliográfica de introdução ao tema de investigação para sistematização do conhecimento sobre esta matéria e realiza-se uma revisão crítica da literatura sobre as competências específicas para a gestão de dados em massa, revendo o projeto europeu *EDISON Data Science Competences Framework*. Discute-se as transformações indispensáveis ao perfil e às competências para o profissional da informação e por fim apontam-se caminhos para uma investigação futura.

Palavras-chave: Big data, Dados em massa, Perfil do gestor de dados em massa, Perfil do profissional da informação, Formação profissional, Projeto EDISON

1. Introdução e método

Neste artigo discute-se o perfil e as competências para o profissional da informação face ao fenómeno dos dados em massa (*Big data*). O conceito de gestão de dados em massa é utilizado em diferentes âmbitos, mas é aqui circunscrito às áreas da informação e documentação e sistemas de informação.

Dados em massa é um termo utilizado para referir-se à explosão de uma quantidade e diversidade de dados digitais que são um elemento essencial para o crescimento da produtividade, inovação e para a possibilidade de gerar valor excedente para a sustentabilidade da sociedade (Erway & Rinehart, 2016). Considera-se que é uma tendência do futuro que deve ser percebida pelas instituições privadas e públicas, empresas e universidades.

Este campo de trabalho à volta dos dados em massa deverá ser considerado para a reflexão acerca da formação científica e académica adequada aos futuros profissionais da informação e para a formação profissional contínua dos profissionais

já em exercício. Considera-se que o profissional da informação é um especialista que poderá ser de máxima importância para a gestão e para outras tarefas associadas à gestão dos grandes volumes de dados.

Neste artigo faz-se uma revisão bibliográfica para introdução do tema de investigação junto da comunidade científica da Ciência da Informação, realiza-se uma sistematização do conhecimento sobre esta matéria e uma revisão crítica da literatura sobre as competências específicas para o gestor de dados em massa, discutindo-se o projeto EDISON e as competências adequadas ao perfil do profissional da informação. No final apontam-se caminhos para uma investigação futura (Pickard, 2007).

2. Os Dados em massa e novas competências

2.1 Definição de dados em massa

Qualquer atividade pública ou privada proporciona grandes conjuntos de dados provenientes, por exemplo, da informatização de serviços, dos dispositivos móveis, das transações na bolsa, das pesquisas e da navegação na Web, da Internet das coisas, dos eventos da rede de comunicações, das marcações de identificação de radiofrequência, das redes sociais, etc. A Internet de coisas gera uma enorme quantidade de dados a partir de qualquer tipo de dispositivo ou sensor. Em 2012, Torres i Vinals afirmou que para produzir dados, por exemplo, o *Twitter* obtinha mais de 90 milhões de tuítes por dia, representando um total de 8 terabytes de dados. Os dados transacionais de aplicativos Web também crescem a uma velocidade muito grande. No mundo científico, o acelerador de partículas do CERN gerava 40 terabytes de dados por segundo durante as experiências efetuadas nele. Tudo isso implica um crescimento exponencial da informação disponível que produz terabytes de dados. É necessário processá-los e transformá-los em informações com valor e este é um grande desafio e uma oportunidade para quem trabalha com dados em massa (Torres i Vinals, 2012).

Não existe uma definição única e consensual para este novo fenómeno dos dados em massa. Poderá ter vários significados, não estando limitada ao mundo da tecnologia. É atualmente uma prioridade comercial dada a sua capacidade de influenciar profundamente o mercado numa economia integrada à escala global (Alonso Arévalo & Vásquez Vásquez, 2016).

A área dos dados em massa além de fornecer soluções para os desafios no âmbito dos negócios, inspira novas formas de transformar processos, empresas, setores inteiros e até mesmo a própria sociedade. As empresas usam dados em massa para obter resultados centrados no cliente, aproveitam os seus dados internos e criam um melhor ecossistema de informações (Schroek, Shockley, Smart, Romero Morales, & Tufano, 2012).

Os dados em massa referem-se à acumulação e ao armazenamento de grande volume de dados, ao seu processamento, à análise e interpretação. É um conjunto de dados que excede a capacidade de processamento das bases de dados convencionais num tempo determinado.

O volume dos dados não é a única característica que os explica, acrescentando-se outras características e distinções, à primeira definição dos três Vs: volume, velocidade e variedade, atribuída ao analista Laney (2001) (Torres i Vinals, 2012; Alvarez, Viana, & Silva, 2016; Cavique, 2014; Morais, Cunha, & Mendonça, 2017). Os dados em massa são um termo em formação e evolução (The CASRAI Dictionary, 2015), assim como os instrumentos, as linhas de investigação, a formação de um corpo teórico, etc., obtendo-se uma proliferação de aceções possíveis (García Alsina, 2017).

Prosseguindo o esclarecimento sobre o sentido desta definição e partindo da letra V que Doug Laney (2001) encetou, a bibliografia apresenta outras características sobre os dados em massa, que tem vindo a atualizar-se (Chen, Mao, & Liu, 2014; García Alsina, 2017; Torres i Vinals, 2012): Volume - os dados crescem exponencialmente e aumentam o volume dos mesmos; Velocidade - os dados são geradas e processadas muito rapidamente; Variedade - os dados procedem de fontes diversas e podem ter níveis diferentes de estruturação ou não terem nenhuma estrutura predefinida. A variedade dos dados orienta cada vez mais os investimentos sobrepondo-se ao volume e velocidade, e será esta a tendência dos próximos anos (Tableau Software, 2017); Veracidade - os dados têm que ser fiéis à realidade, confiáveis, estarem protegidos e atualizados; Valor - os dados ao serem analisados são orientados à criação de valor; Visualização - os dados têm que ser visualizados de forma prática e dinâmica; Verificação - os dados têm que ser verificados por questões de segurança e para garantirem a sua integridade; Variabilidade - os dados têm que ser mantidos atualizados tendo em conta a velocidade com que são produzidos para que a geração de novos transformem outros dados em obsoletos; Viabilidade - os dados e a sua gestão relacionam-se com a viabilidade económica da manutenção das infraestruturas, armazenamento e ferramentas necessárias; Vulnerabilidade - alguns dos dados podem ser considerados dados pessoais e são uma preocupação para os gestores na manutenção da privacidade.

A gestão de dados em massa é constituída por uma série de processos de transformação dos dados até à extração de valor. As diferentes fases deste processo são compostas pela recolha e extração da informação, pelo armazenamento dos dados e pela sua análise. Todas estas fases implicam tarefas particulares e exigem ferramentas e softwares muito específicos (García Alsina, 2017).

A extração dos dados, que podem ter origens diferentes, duplicados ou conterem erros, inicia-se por uma recolha em bruto dos mesmos. É necessário tratá-los para se obter fiabilidade, uniformidade, qualidade e dar-lhes consistência para os integrar em bases de dados. Estas tarefas são realizadas com ferramentas de processamento

(Extract Transform Load, federação de dados, etc.) e técnicas de limpeza especializadas (Chen et al., 2014).

A fase seguinte do processo é o armazenamento dos dados, o que implica a obtenção de uma infraestrutura tecnológica num determinado formato para a rápida recuperação, posterior análise e extração de valor. Nestes empreendimentos estão envolvidas tarefas associadas à Informação e Documentação, como a utilização de linguagens controladas para a boa recuperação dos dados, análise de metadados, usabilidade e a visualização dos dados (García Alsina, 2017).

A fase final do processo designa-se por extração de valor dos dados em massa. O objetivo desta fase de gestão de dados - *data mining* - é criar valor, um conhecimento para a ação com vista à tomada de decisões e à inovação. Esta análise é muito especializada, invoca um conjunto de procedimentos informáticos e modelos estatísticos. A análise de dados é objeto de investigação para se obter melhores resultados nos desenhos de tecnologias e softwares, nos métodos de análise e visualização da informação (Schroeck et al., 2012).

Como já se afirmou, a origem dos dados é diversificada, as fontes são diversas, podem ser oriundas das organizações, das empresas, de atividades privadas, das redes sociais, etc. Qualquer conjunto de dados capturado e armazenado em sistemas de informação pode ser tratado nos processos de gestão de dados em massa.

Encontra-se atualmente uma fonte muito importante de dados, representada nas plataformas de informação, disponibilizada pelos governos de vários países. Estes dados públicos podem ser relatórios, mapas, estatísticas, estudos, dados criados e gerados pela administração pública em todas as áreas, desde a saúde, a economia, a educação, a cultura, etc., que são de grande interesse social e económico (Torres i Vinals, 2012).

No setor público, a reutilização da informação, nomeadamente a partilha de dados em massa com valor, é exemplo desta filosofia universal de acesso aberto e partilha de dados. Pretende-se que as instituições públicas exponham os dados públicos que estão na sua posse de forma reutilizável para que terceiros possam criar serviços derivados desses dados. Como consequência, os conjuntos de dados expostos são oferecidos sob licenças de propriedade aberta que permitem a sua redistribuição, reutilização e uso para fins comerciais. Existe todo um conjunto de pacotes de código aberto que é uma oportunidade para os empreendedores gerarem novas aplicações com base nestes dados (Torres i Vinals, 2012; García Alsina, 2017).

Em Portugal, a Agência para a Modernização Administrativa, no âmbito das suas responsabilidades de promoção e desenvolvimento de políticas de dados abertos para a Administração Pública, o que inclui a gestão do portal nacional de dados abertos - *dados.gov*, dedica-se a explicitar o que são e o que se pretende com os dados abertos (AMA - Agência Portuguesa para a Modernização Administrativa, 2016b) e (AMA - Agência Portuguesa para a Modernização Administrativa, 2016b) através da

publicação de guias que permitem aprofundar o conhecimento sobre o tema e abordam várias questões teóricas e práticas pertinentes para o desenvolvimento do *Open Data*.

2.2 Novas competências para gerir dados em massa

2.2.1 O perfil do gestor de dados em massa

O ciclo de gestão de dados em massa está vinculado à gestão dos documentos, da informação e do conhecimento. Os dados são a base para gerar informação e são o passo prévio à criação de conhecimento para ação. Toda a gestão de dados em massa faz-se através de uma cadeia de valor orientada ao conhecimento (García Alsina, 2017).

O mercado laboral tem estado a ter um incremento considerável na procura de profissionais com competências para gerir dados em massa. Trabalhar em todas as fases do ciclo da gestão exige conhecer as ferramentas de armazenamento e de computação que podem processar o grande volume de dados. O armazenamento é muito importante para as organizações que procuram migrar os seus dados, armazenar e calcular operações, em softwares do tipo repositório ou em serviços sediados numa plataforma na nuvem. É necessário um conhecimento sobre as novas ferramentas para o processamento rápido e inteligente dos dados em tempo real e de um conhecimento que esteja também focado na visualização e na análise dos dados para os poder comunicar posteriormente à administração ou direção.

Constata-se uma dificuldade elevada na implementação de projetos de gestão de dados em massa que exige perfis profissionais multidisciplinares. Estes projetos são tecnológicos e necessitam de distintos conhecimentos, desde os saberes associados à engenharia informática (arquiteturas, plataformas tecnológicas, sistemas de informação), de tecnologias de aquisição, de armazenamento e transporte de dados, de prospeção e análise de dados, de tecnologias de visualização e de tecnologias orientadas aos serviços oferecidos na nuvem, etc.

Portanto, em síntese, os processos requeridos na gestão de dados em massa são desenvolvidos em quatro âmbitos: desenvolvimento e implementação de soluções de gestão; desenvolvimento de atividades para gerir os dados, armazenar, tratar e manutenção da qualidade; administração e manutenção da arquitetura dos sistemas de informação; extração da informação e criação de conhecimento para obter benefício dos dados mediante a análise (García Alsina, 2017).

A Ciência de Dados é uma área científica emergente e atrai vários especialistas em tecnologias da informação, profissionais da área dos negócios, engenheiros, etc. Há uma tendência para centrar no perfil do cientista de dados as tarefas e trabalhos

associados à cadeia de valor da gestão de dados em massa. Este perfil exige o cruzamento de habilidades analíticas e de comunicação com as habilidades técnicas próprias do engenheiro e do programador de software (O'Reilly Media, 2017). O cientista de dados possui conhecimentos avançados e experiência em pelo menos três disciplinas: tecnologia da informação, negócios e analíticas avançadas e usa este conjunto de habilidades multidisciplinares na resolução de problemas envolvidos no ciclo da gestão de dados (Cavique, 2014).

O perfil profissional do cientista de dados está assente nas estruturas de competências/habilidades reconhecidas no campo das Tecnologias de Informação e Comunicação, nomeadamente *European e-Competence Framework* (CEN European Committee for Standardization, 2014) e *The Skills Framework for the Information Age* (SFIA Foundation, 2003). Uma parte significativa da base de conhecimento e do conjunto de habilidades do cientistas de dados relaciona-se com as competências de tecnologias de informação e comunicação combinadas com as contribuições de diferentes áreas, como a ciência da computação, a estatística e a matemática (Costa & Santos, 2017).

No mundo de hoje, em que os negócios são cada vez mais complexos, em que as empresas lidam diariamente com um elevado número de dados, ser cientista de dados é uma profissão com cada vez mais procura. Nos Estados Unidos há inúmeros cursos de Ciência de Dados nas mais conceituadas universidades. As empresas do setor ajudam a desenvolver os referidos cursos que resultam da interação salutar entre a universidade e o mundo empresarial ficando estes cursos mais ajustados à realidade.

A criação de uma equipa de profissionais multidisciplinares para gerir os trabalhos do ciclo de gestão de dados em massa, por parte das organizações, é uma estratégia interessante, mas requer um consenso no mercado dos negócios e entre os especialistas. É um processo que contribui para aprofundar o conhecimento e as intervenções do saber específico de cada profissão e pode ser uma alternativa ao perfil do cientista e dados.

2.2.2 Projeto EDISON

Desde 2015 que a União Europeia financia o projeto *EDISON Data Science Competences Framework* para definição das competências relacionadas com o profissional da gestão de dados em massa (EDISON Project UE, 2015). O projeto implementa mecanismos para trabalhar a definição de perfis e competências e conhecimentos associados à gestão de dados em massa para o aumento do número de cientistas de dados competentes e qualificados em toda a Europa. O projeto apresenta uma série de documentos e matérias de informação para auxiliar educadores, formadores, empregadores e para o recrutamento de profissionais na área da Ciência de Dados, nomeadamente apresenta bases para a definição do

quadro de competências, do corpo de conhecimento e de um modelo de currículo e certificação do profissional.

O projeto pretende que as competências sejam definidas em conformidade com *European e-Competence Framework* (CEN European Committee for Standardization, 2014) e fornece sugestões para se realizar uma extensão desta com as competências e habilidades relacionadas com a Ciência de Dados. O projeto EDSION fornece uma orientação e uma base para que as universidades definam os seus currículos e cursos nesta área, e por outro lado, para que as empresas definam melhor um conjunto de competências e habilidades necessárias para a criação de perfis de gestores de dados (Demchenko & Belloum, 2017).

Como se afirmou anteriormente, EDISON apresenta as competências e as habilidades habituais atribuídas ao cientista de dados, como a análise dos dados, a engenharia ou a programação e o domínio de um setor do conhecimento referente à área de negócio, mas acrescenta-lhe grupos de competências adicionais: a gestão, a preservação e a curadoria de dados, por um lado e os métodos científicos versus os modelos de negócios. Outras habilidades são igualmente requeridas e afirmadas nas competências, como a inteligência emocional.

A partir da bibliografia, apresenta-se resumidamente as habilidades identificadas para o perfil do cientista de dados: analisar os dados (usar métodos estatísticos aplicados aos dados para extrair conhecimento; usar análise preditiva, saber analisar o negócio, etc.); Desenvolver software e hardware (usar princípios de engenharia para investigar, desenhar ou desenvolver estruturas e tecnologias, desenvolver ferramentas de análise de dados, desenhar e gerir base de dados, desenvolver soluções para acesso seguro e fiável aos dados, etc.); Possuir competências e conhecimentos sobre temas científicos (compreender o âmbito de negócio e saber fornecer informação, traduzir problemas de negócio desestruturado para um quadro abstrato matemático, usar dados para melhorar serviços ou desenvolver novos, etc.); Gerir dados e preservar (desenvolver e implementar estratégia de dados, desenvolver modelos de dados e metadados, integrar diferentes fontes de dados para análises futuras, desenvolver e manter o repositório de análise de dados, recolher e gerir diferentes fontes de dados, visualizar dados variáveis e complexos, etc.); Conhecer métodos de investigação (criar conhecimento novo e ter capacidades para empregar técnicas do método científico; rever criticamente; levar a cabo um trabalho criativo, criar novas aplicações, etc.) (Demchenko, 2016; García Alsina, 2017).

3. Resultados e Discussão

O projeto EDISON propõe um perfil de gestor de dados em massa com a designação de cientista de dados. Neste artigo recupera-se outro perfil profissional que trabalha a questão da gestão de dados, para além dos já mencionados, como o engenheiro informático, engenheiro de telecomunicações. Apresenta-se o profissional da

informação que trabalha para a qualidade de vida dos dados e pelo ciclo de vida da informação e do conhecimento, analisa os dados, e facilita às organizações a geração do conhecimento para ação mediante a criação de produtos de informação.

Por um lado, no mercado de trabalho exige-se profissionais competentes para o desenvolvimento de projetos neste domínio, chamando atenção para o perfil e competências que requerem especialização em diferentes âmbitos, como o cientista de dados, desde o desenvolvimento e implementação de soluções de gestão de dados em massa, desenvolvimento de atividades para os gerir no tratamento, armazenamento e manutenção, administração e manutenção da arquitetura dos sistemas informáticos até à extração da informação e criação de conhecimento (García Alsina, 2017). Por outro lado, constata-se que esta situação de necessidade de profissionais para gerir dados em massa é uma oportunidade para um progresso no desenvolvimento do perfil do profissional da informação no âmbito da formação de competências e habilidades nesta nova área científica.

Para a gestão de dados em massa, como referido anteriormente, são necessários conhecimentos sobre gestão e administração de empresas, engenharia, informática, estatística, matemática e gestão de documentos e da informação, etc. (Demchenko & Belloum, 2017; SAS, 2013). Outros saberes são também requeridos, como o conhecimento sobre legislação que afeta a proteção de dados, os aspetos éticos do uso da informação, os conhecimentos sobre preservação e recuperação dos dados em sistemas orientados ao utilizador, a facilidade no acesso e a usabilidade. Estes temas são estudados, em geral, nos cursos académicos para obtenção de grau de licenciatura e outros, na área da Informação e Documentação, assim como em cursos de formação contínua.

Esta situação expressa que algumas das habilidades e das competências para o gestor de dados em massa são também inerentes ao profissional da informação. Nalgumas áreas de trabalho, este profissional lida com um grande volume de dados digitais que necessitam de ser geridos e requerem uma readaptação profissional, adoção de novas tecnologias e enriquecimento das suas competências com currículos de outras áreas científicas (Alonso Arévalo & Vásquez Vásquez, 2016).

O conhecimento e as aptidões emergentes necessárias para o perfil dos profissionais da informação são interdisciplinares, transversais, especializados, complexos e exigem uma reflexão sobre o modelo formativo ajustado ao mercado de trabalho. Isto aplica-se ao domínio da gestão de dados em massa e a outros.

De referir a abordagem já efetuada em Portugal em 2016, no *Encontro de Curadoria Digital: Estratégias e Experiências*, na FCSH-UNL, relativamente aos perfis e competências profissionais no âmbito da curadoria digital (Freitas, 2017; Ochôa, 2017) e noutros eventos, como no painel "Gestão da informação: perfis e competências" no 6.º *Encontro de Arquivos Empresariais*.

A área da gestão de dados em massa é emergente e requer uma reflexão profunda, a partir dos referenciais de competências existentes, desde *Euroreferencial ID* (Conselho Europeu das Associações de Informação e Documentação, 2005), *A Imagem das Competências dos Profissionais da Informação-Documentação: Relatório* (Ochôa & Pinto, 2006), *Observatório da Profissão de Informação e Documentação* da Universidade do Porto (Pinto, Cerqueira, Baptista, & Pinto, 2015), etc.

Observe-se o caso dos dados de investigação que é um dos grandes desafios que se coloca aos profissionais das bibliotecas académicas e de investigação (Erway & Rinehart, 2016), aos arquivos e a outros organismos que trabalhem com a informação. De salientar a realização de *Fóruns de Gestão de Dados de Informação*, em 2016 e 2017, em Portugal, organizados pela Secretaria de Estado da Ciência, Tecnologia e Ensino Superior, FCT-FCCN – Fundação para a Ciência e Tecnologia – Fundação para a Computação Científica Nacional e a Universidade do Minho, promovido no âmbito da Política Nacional de Ciência Aberta e das atividades do RCAAP – Repositório Científico de Acesso Aberto de Portugal, que procuraram dinamizar e criar uma comunidade de profissionais que desenvolvem atividades de suporte à gestão de dados nas instituições de investigação. Estes fóruns têm sido um espaço de debate e partilha de iniciativas emergentes, de projetos e de boas práticas relacionadas com a gestão de dados de investigação e para discussão do desenvolvimento de competências próprias ao gestor de dados (Príncipe & Furtado, 2017).

Um outro desafio para o profissional da informação está relacionado com a literacia sobre os dados em massa nas bibliotecas universitárias e de investigação, nos arquivos, etc. Surgem também iniciativas governamentais cujas organizações disponibilizam dados abertos. Atualmente conhecer e realizar ações de literacia sobre os dados em massa requer adaptação do profissional da informação a este novo paradigma de desenvolvimento dos dados (Carlson & Johnston, 2015; O'Reilly Media, 2017) sendo necessário ainda mais bibliografia, relatos de experiências e práticas de literacia de dados para divulgação entre os profissionais (Alonso Arévalo & Vázquez Vázquez, 2016). De realçar o trabalho efetuado por *Research Data Alliance* (2016) que apresentou uma visão geral de recursos úteis e ferramentas livres *online* que podem servir para integrar a gestão de dados científicos no trabalho diário dos profissionais da informação. Por exemplo, a questão da Literacia dos dados com o projeto *Data Information Literacy* que propõe um currículo para apoiar profissionais da informação e docentes a integrar a *Gestão de Dados* nos programas de capacitação de Literacia da informação (Carvalho, Furtado, & Príncipe, 2017).

Conclusões

Constata-se uma tendência para que nos próximos anos haja um enorme desenvolvimento dos conjuntos de dados em massa, com mais organizações, incluindo a administração pública, empresas e outros, a armazenar, a processar e a extrair valor dos dados que poderão ter tamanhos diferentes e formatos estruturados e não estruturados. A convergência da Internet das coisas, da nuvem e dos dados em massa gerará novas oportunidades para as competências e habilidades dos profissionais (Tableau Software, 2017). No mercado empresarial e nos empregos públicos surgirão mais lugares para profissionais especializados neste domínio.

Devido à multiplicidade de tarefas a executar e ao desenvolvimento célere desta disciplina de gestão de dados em massa, pode-se considerar que o perfil do gestor encontra-se em construção, salientando-se o projeto *EDISON building the data science profession*, cuja equipa tem estado a discutir e a produzir documentos com o objetivo de orientar formadores, docentes e empregadores, oferecendo propostas de perfis, conhecimentos e competências associados a estes profissionais (Demchenko & Belloum, 2017). Da leitura de alguns documentos deteta-se uma complexidade de habilidades e competências requeridas para o perfil de gestor (Demchenko & Belloum, 2016).

As novas iniciativas do governo aberto na Europa (European Commission, 2017) que visam promover a igualdade, a coesão social e a cidadania ativa, simultaneamente com as iniciativas de abertura dos dados públicos, também em Portugal, ao escrutínio e à contribuição dos cidadãos, da sociedade civil e das organizações, para que haja um maior envolvimento por parte de todos e criação de valor público (AMA - Agência Portuguesa para a Modernização Administrativa, 2016a), é uma excelente oportunidade para os profissionais da informação.

Compreende-se que há necessidade de desenvolver competências mais alargadas, diversas das tradicionais, intrínsecas ao profissional da informação. O novo perfil e os novos modos de atuação deste profissional, no século XXI, requerem uma formação renovada e orientada para as problemáticas do campo científico em que se inserem as múltiplas e multifacetadas tarefas do gestor de dados em massa.

O ensino especializado - Informação/Documentação - terá que estar alinhado com os desenvolvimentos do perfil relacionado com os dados em massa porque será uma linha de desenvolvimento profissional muito importante. A reflexão sobre a formação dos profissionais da informação, neste domínio, é sumamente marcante fazer-se atualmente.

A disciplina *dados em massa*, frequentemente chamada de *Big Data*, é um dos campos mais importantes de trabalho para os profissionais associados às tecnologias da informação e comunicação e um repto para os profissionais da informação. Há uma ênfase no termo *dados* que mostra o valor deste como matéria-prima nos setores empresariais e do estado.

Relativamente ao trabalho futuro, pretende-se alargar a discussão sobre o tema junto dos responsáveis pelos cursos de licenciatura, mestrado e doutoramento das universidades portuguesas na área da Ciência da Informação, nomeadamente através de metodologias qualitativas, como entrevistas de tipo - grupo focal - e entrevistas individuais aos professores dos cursos que possuam responsabilidades de direção. Simultaneamente deseja-se realizar um trabalho exaustivo de levantamento de currículos disciplinares de cursos associados à área da gestão de dados em massa em universidades portuguesas e internacionais, assim como proceder-se a um estudo comparado de modo a criar-se um modelo que afira as competências e os conhecimentos indispensáveis aos profissionais da informação para o exercício das funções articuladas com a gestão de dados em massa. No final, pretende-se estruturar uma matriz de competências com identificação de áreas a considerar para a conceção de planos de estudo sobre gestão de dados aplicados ao perfil do profissional da informação.

Referências

Alonso Arévalo, J., & Vásquez Vásquez, M. (2016). Big Data: la próxima «gran cosa» en la gestión de la información. *BiD: Textos Universitaris de Biblioteconomia I Documentación*, (36). Disponível em: <http://bid.ub.edu/es/36/alonso.htm>>

Alvarez, I., Viana, J., & Silva, N. (2016). Big Data: Suporte consistente das decisões de gestão. *Lusíada. Economia & Empresa*, (21).

AMA - Agência Portuguesa para a Modernização Administrativa. (2016a). *Dados.gov.pt*. Disponível em: <http://www.dados.gov.pt/pt/inicio/inicio.aspx#sthash.1xWtAJ09.dpbs>

AMA - Agência Portuguesa para a Modernização Administrativa. (2016b). *Guia Dados Abertos*. Disponível em: https://www.ama.gov.pt/documents/24077/24804/guia_dados_abertos_ama.pdf/a97d8e8-c5fe-47ab-9500-734948c02b19

Carlson, J., & Johnston, L. (ed). (2015). *Data Information Literacy: Librarians, Data and the Education of a New Generation of Researchers*. West Lafayette, Indiana: Purdue Information Press.

Carvalho, J., Furtado, F., & Príncipe, P. (2017). *Kit sobre dados de investigação*. RCAAP. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/46351>

Cavique, L. (2014). Big Data e Data Science. *Boletim APDIO*, (51). Disponível em: <http://hdl.handle.net/10400.2/3918>

CEN European Committee for Standardization. (2014). *European e-Competence Framework 3.0*. Disponível em: http://www.ecompetences.eu/wp-content/uploads/2014/02/European-e-Competence-Framework-3.0_CEN_CWA_16234-1_2014.pdf

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171–209. Disponível em: <http://doi.org/10.1007/s11036-013-0489-0>

Conselho Europeu das Associações de Informação e Documentação. (2005). *Euro-Referencial I-D*. Lisboa: INCITE.

Costa, C., & Santos, M. (2017). The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age. *International Journal of Information Management*.

Demchenko, I. (2016). *EDISON Data science Framework*. Disponível em: http://edison-project.eu/sites/edison-project.eu/files/attached_files/node-448/edison2016-07-champs-conf-v02.pdf

Demchenko, I., & Belloum, A. (2016). *Data Science competence framework: Approach and first working*. *EDISON Discussion Document*. Disponível em: http://edison-project.eu/sites/edison-project.eu/files/attached_files/node-29/edison-cf-ds-draft-cc-v06.pdf

Demchenko, I., & Belloum, A. (2017). *EDISON: Discussion Document: Part 1. Data Science Competence Framework (CF-DS) release 2*. Disponível em: <http://edison->



project.eu/sites/edison-project.eu/files/filefield_paths/edison_cf-ds-release2-v08_0.pdf

EDISON Project UE. (2015). *EDISON: Building the data science profession*. Disponível em: <http://edison-project.eu/>

Erway, R., & Rinehart, A. (2016). *If You Build It, Will They Fund? Making research data management sustainable*. OCLC.

European Commission. (2017). *European eGovernment Action Plan 2016-2020*. Disponível em: <https://ec.europa.eu/digital-single-market/en/european-egovernment-action-plan-2016-2020>

Freitas, M. C. (2017). O Futuro é hoje: perfis e competências dos profissionais da informação para a curadoria digital. In *Encontro de Curadoria Digital: estratégias e experiências*. Disponível em: <http://hdl.handle.net/10760/31832>

García Alsina, M. (2017). *Big Data: Gestión y explotación de grandes volúmenes de datos*. Barcelona: Editorial UOC.

Laney, D. (2001). *3-D Data management: Controlling Data volume, velocity and variety*. Stamford: META Group Research Note.

Morais, E., Cunha, C., & Mendonça, V. (2017). Gestão de Big Data: Novos Paradigmas. In *17.a Conferência da Associação Portuguesa de Sistemas de Informação (CAPSI'2017)*. Disponível em: <http://dx.doi.org/10.18803/capsi.v17.00-00>

O'Reilly Media. (2017). *Big Data Now*. United States of America: O'Reilly Media.

Ochôa, P. (2017). Painel Perfis e competências profissionais. In *Encontro Curadoria Digital – Estratégias e experiências: atas*.



Ochôa, P., & Pinto, L. G. (org). (2006). *A Imagem das competências dos profissionais de Informação-Documentação: Relatório*. Lisboa: Observatório da Profissão de Informação-Documentação (OP I-D). Disponível em: <http://apdis.pt/download/RELID2006.pdf#page=176>

Pickard, A. (2007). *Research methods in information*. London: Facet Publisher.

Pinto, A., Cerqueira, A., Baptista, I., & Pinto, M. (2015). Observatório de Ciência da Informação da Universidade do Porto: Um projeto colaborativo de sucesso. *Cadernos BAD*, (1). Disponível em: <http://hdl.handle.net/10216/84774>

Príncipe, P., & Furtado, F. (2017). *Relatório do 2o Fórum de gestão de Dados de Informação*. Disponível em: <http://hdl.handle.net/1822/46338>

Research Data Alliance. (2016). *23 coisas: Bibliotecas e Dados Científicos*.

SAS. (2013). *Big Data Analytics: An assessment of demand for labour and skills, 2012-2017*. London. Disponível em: [file:///C:/Users/luisa/Downloads/BigDataAnalyticsAnassessmentofdemandforlabourandskills2012-2017 \(1\).pdf](file:///C:/Users/luisa/Downloads/BigDataAnalyticsAnassessmentofdemandforlabourandskills2012-2017(1).pdf)

Schroeck, M., Shockley, R., Smart, J., Romero Morales, D., & Tufano, P. (2012). Analytics: el uso de big data en el mundo real. *IBM. Informe ejecutivo*. Disponível em: http://www-05.ibm.com/services/es/bcs/pdf/Big_Data_ES.PDF

SFIA Foundation. (2003). *The Skills Framework for the Information Age - SFIA*. Disponível em: <https://www.sfia-online.org/en>

Tableau Software. (2017). *10 Principais tendências de Big Data para 2017*. Disponível em: <https://www.tableau.com/learn/whitepapers>



The CASRAI Dictionary. (2015). Disponível em:
http://dictionary.casrai.org/Main_Page

Torres i Vinals, J. (2012). *Del Cloud Computing al Big Data: Visión introductoria para jóvenes emprendedores*. Barcelona: Universitat Oberta de Catalunya.