

REFLEXÕES SOBRE A REPRESENTAÇÃO NO CICLO DE VIDA DOS DADOS

Reflections on the representation in the life cycle of the data.

Ricardo César Gonçalves Sant'Ana

*Doutor em Ciência da Informação
Livre-Docente em Sistemas de Informações Gerenciais
Universidade Estadual Paulista - UNESP
Contato: ricardosantana@marilia.unesp.br*

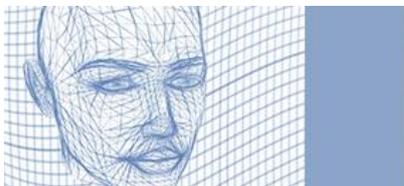
Resumo

Com o advento dos suportes digitais, várias questões ganharam destaque no âmbito da Ciência da Informação e a criação de pontes entre os usuários e a informação passa pela necessidade de representação de conteúdos. A possibilidade de coleta e armazenamento de grandes quantidades de informações pelos recursos digitais ampliou os limites e possibilitou o tratamento de dados em grandes volumes, com ampla variedade de formatos e fontes e, ainda, com novos patamares de velocidade de acesso, contexto que vem recebendo a denominação de Big Data por ser caracterizado pelo descompasso entre a capacidade de coleta e armazenamento em relação à capacidade de recuperação destes dados. Deste novo cenário, emergem questões como: quando, quem e como realizar a representação destes dados, questões alvo de reflexão neste texto. Destaca-se que a representação, diretamente na fonte ou naqueles que atuam como os detentores dos recursos de armazenamento, geraria custos que inviabilizariam o processo. A intermediação de agentes que concentrariam esta tarefa, entre os detentores e usuários, como aconteceu com os demais conteúdos informacionais, enfrenta a barreira da ausência de semântica intrínseca suficiente nestes grandes volumes de dados. Já a possibilidade de distribuição destas funcionalidades junto aos próprios usuários precisa vencer as limitações de capacidade de comunicação, armazenamento e processamento, fatores que vêm sofrendo grandes transformações. No entanto, este modelo aponta para um caminho inverso ao que vêm sendo adotado, em que as funcionalidades tendem a serem acessadas remotamente, deixando na camada de usuário o menor peso tecnológico possível.

Palavras-chave: Representação de Dados; Ciclo de Vida dos Dados; Ciência da Informação; Big Data.

Abstract

With the advent of digital media, several issues have gained prominence in the field of Information Science and the creation of bridges between users and information requires the representation of content. The possibility of collecting and storing large amounts of information through digital resources has broadened the limits and enabled the treatment of data in large volumes, with a wide variety of formats and sources, and also with new levels of access speed, a context that has been receiving the denomination of Big Data because it is characterized by the mismatch between the capacity of collection and storage in relation to the capacity of recovery of this data. From this new scenario emerge questions such as when, who and how to perform the representation of these data, issues for reflection in this text. It is emphasized that the representation directly in the source or in those that act as the holders of the storage resources would generate costs that would make the process unfeasible. The intermediation of agents that would concentrate this task between the holders and users, as



happened with the other information contents, face the barrier of the absence of sufficient intrinsic semantics in these large volumes of data. Already the possibility of distribution of these functionalities with the users themselves would have to overcome the limitations of communication, storage, and processing capacity, factors that have been going through major transformations, however, this model points to a reverse path to what has been adopted in which the functionalities tend to be accessed remotely, leaving the user layer the least technological weight possible.

Keywords: Data Representation; Data Life Cycle; Information Science; Big Data

Resumen

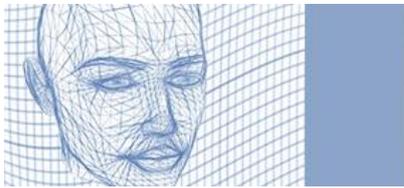
Con el advenimiento de los medios digitales, varias cuestiones han ganado prominencia en el campo de la Ciencia de la Información y la creación de puentes entre los usuarios y la información requiere la representación del contenido. La posibilidad de recopilar y almacenar grandes cantidades de información a través de recursos digitales ha ampliado los límites y ha permitido el tratamiento de datos en grandes volúmenes, con una amplia variedad de formatos y fuentes, y también con nuevos niveles de velocidad de acceso, un contexto que ha sido Recibiendo la denominación de Big Data porque se caracteriza por el desajuste entre la capacidad de recolección y almacenamiento en relación con la capacidad de recuperación de estos datos. De este nuevo escenario emergen preguntas como cuando, quién y cómo realizar la representación de estos datos, cuestiones para reflexión en este texto. Se destaca que la representación directa en la fuente o en aquellos que actúan como poseedores de los recursos de almacenamiento generaría costos que harían inviable el proceso. La intermediación de agentes que concentraría esta tarea entre los titulares y usuarios, como ocurrió con Los demás contenidos informativos, enfrentan la barrera de la ausencia de suficiente semántica intrínseca en estos grandes volúmenes de datos. Ya la posibilidad de distribución de estas funcionalidades con los propios usuarios tendría que superar las limitaciones de la capacidad de comunicación, almacenamiento y procesamiento, factores que han estado pasando por grandes transformaciones, sin embargo, este modelo apunta a un camino inverso a lo que se ha adoptado en el que las funcionalidades tienden a ser accesibles remotamente, dejando a la capa de usuario el menor peso tecnológico posible.

Palabras clave: Representación de Datos; Ciclo de Vida de los Datos; Ciencia de la Información; Big Data.

1 Introdução

O advento dos recursos tecnológicos para acesso e tratamento de conteúdos digitais propiciou o surgimento de novos patamares no acesso a dados com potencial transformador sem precedentes nas dimensões social, cultural e econômica.

A multiplicação de formas de obtenção de informações sobre o ambiente, relações sociais, transações comerciais e até mesmo de situações resultantes de simulacros multiplicam o 'volume' de dados que são coletados, que se apresentam nos mais 'variados' formatos e granularidades. Acrescente-se a este cenário recursos mais rápidos, disseminados e altamente conectados, formando um amalgama de diferentes tecnologias que permitem a dissolução das limitações



temporais no acesso a estes dados e, portanto, agregando um terceiro 'v', de velocidade, aos conceitos definidores deste novo contexto tecnológico de coleta de dados, que vem sendo definido como Big Data para se referir a

[...] dados que são grandes demais para um único servidor, muito diversos para se adequar a uma base de dados estruturada em linhas e colunas, ou cujo fluxo seja tão intenso que não permita adequação a um datawarehouse estático (DAVENPORT, 2014).

Todo este conteúdo coletado não teria serventia caso não houvesse meios de persistir estes dados de forma eficiente e viável, e esta funcionalidade, relacionada aos recursos digitais, vem ganhando não só em capacidade, mas também em velocidade e custo, oportunizando o armazenamento de abundantes quantidades de conteúdos.

Com a ampliação massiva dos meios de armazenamento e acesso à informação, o foco pode então ser direcionado para a recuperação dos novos patamares de informação que passaram a estar disponíveis e

[...] convencionou-se afirmar que o primeiro desafio [...] foi o da aquisição, em períodos históricos em que os livros e registros materiais do conhecimento eram escassos e frequentemente perdidos ou intencionalmente destruídos. Ao longo dos séculos, e especialmente após o desenvolvimento da imprensa, tal desafio foi sendo cada vez mais superado, a ponto de se começar a ter coleções e acervos relativamente consistentes. A grande questão passou a ser não mais a aquisição de livros, mas sua organização, de forma a se conseguir promover a sua recuperação e o seu uso. (ARAUJO, 2009)

No caso dos dados, a questão não é diferente. O volume cada vez maior de dados ao alcance de acesso direto passou a apresentar problemas com relação a como encontrar os dados desejados e, mais ainda, como interpretá-los.

Mas, ao se referir a estes grandes volumes de dados, é preciso delinear, mesmo que sem fechar uma definição para este termo notoriamente polissêmico, características que permitam a compreensão de suas especificidades. O que caracterizaria um dado?

Com vistas a atender aos objetivos deste texto, destaca-se duas de suas principais características: a tendência de tratar de conteúdos altamente estruturados e sua inerente baixa carga semântica, conforme ilustrado na figura 1.

Nesta figura pode-se perceber que não há uma linha claramente demarcatória



mas, antes, um degradê relativo a propriedades como a estrutura e a semântica, relacionados a um determinado conteúdo.

Destacam-se, ainda, fatores resultantes desta caracterização, tais como a maior aderência em relação a eficiência de atuação, tanto de agente humano quanto de agentes tecnológicos, com matizes invertidas quando se considera o espectro de estruturação e semântica dos conteúdos.

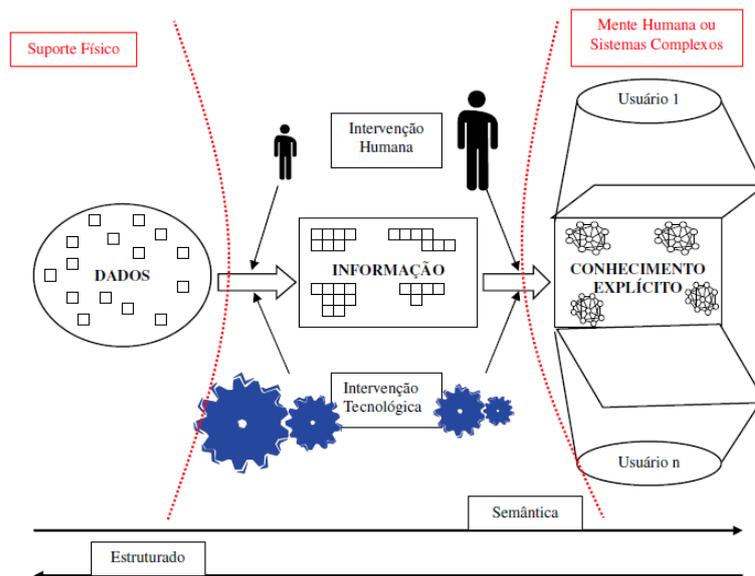
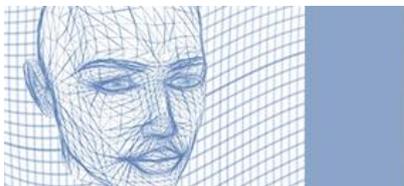


Figura 1. Semântica e Estruturação dos Dados (SANTOS & SANT'ANA, 2002).

Por sua baixa carga semântica (SANTOS e SANT'ANA, 2013), os dados requerem esforços adicionais ainda maiores para sua representação, já que não bastam elementos que o descrevam como um todo e que propiciem sua recuperação. São necessários, ainda, elementos que permitam a sua interpretação por quem os acesse, com informações que detalhem sua estrutura e possibilitem a interpretação de cada atributo que os compõe.

Ao considerarmos a recuperação dos dados na análise do processo de sua representação emergem, entre outras, as seguintes questões: Como representar os dados? Em que momento os dados devem receber o tratamento da representação? Quem deve ser o responsável pela representação dos dados?

Neste texto, não se tem a pretensão de apresentar respostas a estas questões, mas, antes, estabelecer parâmetros iniciais para que se possa refletir sobre em que momento os dados poderiam ser representados. Estas reflexões implicarão, em



momentos futuros, na possibilidade de delinear os demais elementos envolvidos neste processo, tais como a identificação de quem seria o responsável por esta representação e, ainda, como ela deverá ser operacionalizada.

2 Desenvolvimento

Entender e refletir sobre as questões propostas neste texto gera impactos diretos em todas as ações envolvidas na contextualização dos dados. São estes fatores (quem, como e quando) que nortearão o processo de representação na gestão destes dados. Conforme destacado, busca-se, inicialmente, apresentar parâmetros sobre qual o momento do processo de acesso a dados poderia ancorar a representação destes conteúdos.

Para amparar esta análise, o Ciclo de Vida dos Dados - CVD (SANT'ANA, 2013; 2016) oferece uma estrutura baseada em fases que permite entender o “quando” e, por conseguinte, estabelecer parâmetros para o “quem”, definindo, ainda, um cenário para o “como” em função da contextualização da realização do processo.

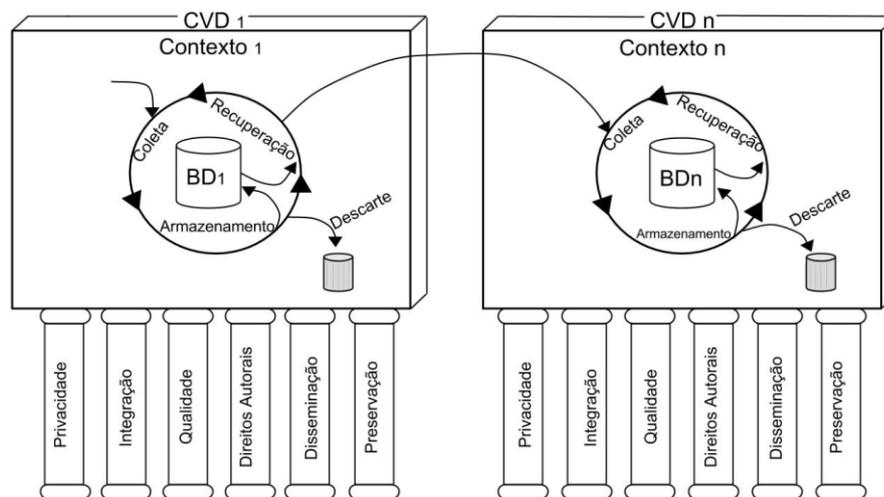
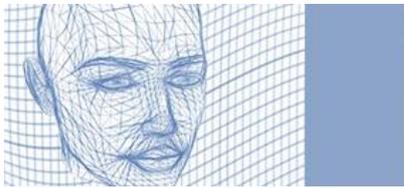


Figura 2. Ciclo de Vida dos Dados. (SANT'ANA, 2016).

O CVD propõe que o acesso a dados é composto por quatro fases: coleta, armazenamento, recuperação e descarte. Assim, pode-se entender que, em um primeiro momento, existem as ações relacionadas ao planejamento, projeto e ação de obtenção de um determinado conteúdo, que pode ser um dado ou um conjunto de



dados que tenha características tais que possam atender a uma necessidade ou conjunto de necessidades correlatas. Esta primeira fase, identificada no CVD como coleta, vai até o ponto em que o conteúdo é obtido e mantido provisoriamente em meios fugazes, como a memória de curta duração, e que, portanto, não estarão disponíveis em momentos futuros.

Estes conteúdos podem, por diferentes razões, mostrarem-se relevantes para acesso posterior ao processo de coleta, o que leva a uma nova fase que exige ações relacionadas ao planejamento, projeto e ação, relacionadas ao processo de persistência destes dados. No CVD, esta fase é denominada como fase de armazenamento.

Uma vez coletados e armazenados, estes dados podem vir a ser requisitados para acesso, o que leva à necessidade de ações relacionadas ao planejamento, projeto e ação relacionadas ao processo de potencialização das possibilidades de acesso a estes dados, inclusive por elementos humanos ou tecnológicos, exógenos ao próprio CVD, em seus próprios processos de coleta de seus respectivos CVDs. Esta fase é denominada como fase de recuperação.

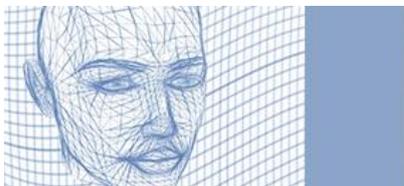
Uma quarta fase ainda mantém espaço para um momento em que se realizam ações relacionadas à eliminação de conteúdos e que é denominada fase de descarte.

Cabe destacar que estas fases não são autoexcludentes, ou seja, eles podem ocorrer concomitantemente e o foco das atenções pode passar de uma fase posterior a uma anterior em função de necessidades surgidas durante o processo. O próprio planejamento e projeto relacionado a cada uma delas pode ocorrer desde o início ou desde o momento em que o CVD é percebido como necessário.

Assim, a função desta proposta de divisão em fases se justifica como forma de identificar momentos específicos do dado em relação à necessidade que fundeu a existência do ciclo, delimitando elementos envolvidos e características específicas de cada momento dentro do processo.

Fatores envolvidos

Um dos fatores-chave envolvido nos processos que ocorrem ao longo do CVD está relacionado ao Custo da Transação, ou, de forma mais aderente a este contex-



to, ao custo envolvido para cada determinada ação a ser realizada no acesso a dados. O fator Custo da Transação foi estudado pelo Nobel de Economia Ronald Harry Coase (1988) e pode-se considerá-lo no centro do processo de representação, já que para se obter um ganho em função do aumento do potencial de recuperação se faz necessário investimento na representação.

Este custo será composto principalmente pelo tempo de agente humano responsável pela realização da representação, tempo este que pode ser convertido na estruturação de sequências de instruções para que a ação seja automatizada por processamento computacional, ou seja, por meio de algoritmos. No entanto, esta alternativa também apresenta características limitantes.

Ao considerar o custo da transação, os recursos digitais permitem, ainda, a potencialização de aspectos como a colaboração, ao disponibilizar funcionalidades de comunicação para integrar fases distintas de diferentes CVDs em tempo real, o que favorece o compartilhamento de recursos e a economia das redes de informação parece aumentar as capacidades de cooperação e coordenação de grandes grupos das partes interessadas (MANSELL;NOLAN;WEHN, 1998; BERRIO-ZAPATA & SANT'ANA, 2015).

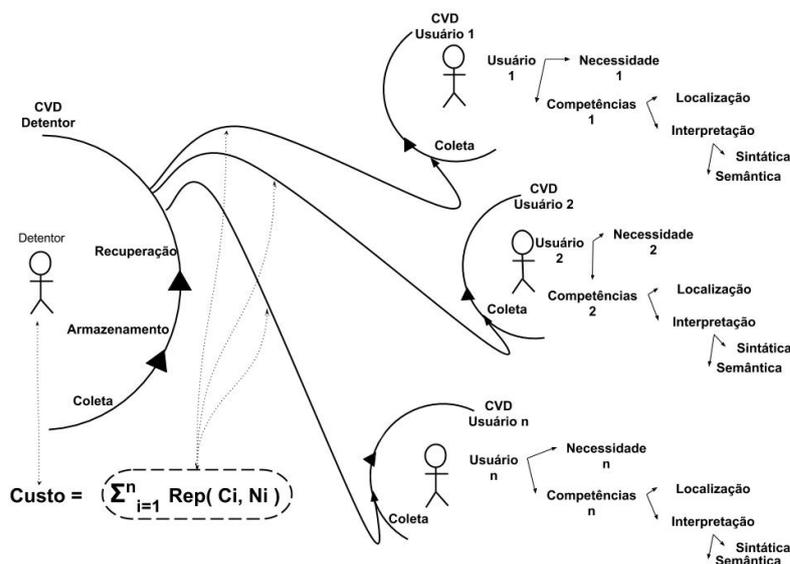
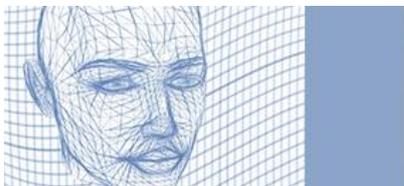


Figura 3. Custo da Representação de Dados na Interação entre CVDs

Para efeito de interpretação deste cenário, considera-se detentor aquele que, achando-se em relação de dependência para com outro, conserva a posse em nome

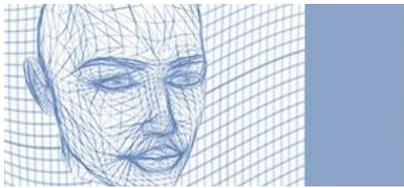


deste e em cumprimento de ordens ou instruções suas. Portanto, aquele que começou a comportar-se de modo específico, em relação ao bem e à outra pessoa, presume-se detentor, até que prove o contrário, ou seja, entidade (personificada por um indivíduo ou instituição) que é responsável pela manutenção de todas as fases de um determinado CVD.

Na questão da interpretação inerente às competências de cada usuário, cabe salientar que esta interpretação, em um primeiro momento, terá de atender às necessidades iniciais relacionadas à identificação de elementos mínimos da estrutura necessária para que os dados possam ser utilizados, ou seja, ao menos a tríade (entidade, atributo, valor $\langle e,a,v \rangle$) (SANTOS & SANT'ANA, 2013) e, conforme destacado na figura 3, ao se manter o processo de representação sob a responsabilidade do detentor, ou mesmo por outrem que tenha por finalidade realizá-la, deve-se considerar que esta representação deverá conter elementos que permitam sua adequação às especificidades de cada um dos potenciais usuários, que por meio de seus CVDs, mais especificamente na sua fase de coleta, precisam ter não só suas necessidades atendidas como, ainda, terem consideradas suas especificidades relacionadas as suas competências para localizar e interpretar estes dados coletados.

Mais uma vez, as características relacionadas à baixa carga semântica e à alta estruturação inerentes aos dados, tornam o processo de recuperação de dados diferente dos processos de recuperação da informação via mecanismos de buscas (JANOWICZ et. al. 2012) e, em especial, em algumas propriedades, como na correspondência da pesquisa, na inferência, no modelo matemático, na classificação dos resultados, na linguagem utilizada para a elaboração de uma pesquisa, na recuperação dos dados e nas respostas aos possíveis erros no processo (VAN RIJSBERGEN 1999).

Para tornar mais clara estas diferenças, pode-se recorrer às diferenças mais marcantes, que podem ser entendidas ao destacar-se pontos como: (a) a recuperação da informação lida com conteúdos não estruturados ou semiestruturados, enquanto a recuperação de dados com dados estruturados; (b) os resultados de processos de recuperação da informação podem ser apresentados por meio de listas organizadas com características como aderência ao resultado esperado, ou seja, resultados parcialmente aderentes ao que se buscava são permitidos, enquanto que



na recuperação de dados os resultados devem ser exatos (precisos) ou não retornar conteúdos caso não encontre correspondências exatas.

Estas propriedades, que diferenciam o processo de recuperação dos dados, tornam o vínculo com especificidades dos usuários muito mais forte, o que leva à necessidade de se acrescentar custos extras a este processo de representação para o atendimento de particularidades relacionadas à interpretação sintática e semântica dos resultados (RODRIGUES; SANT'ANA; FERNEDA 2015).

Assim, para atender aos usuários em potencial, deve-se considerar que o processo de representação não poderá seguir um procedimento padrão ou modelos previamente estabelecidos como na recuperação da informação, o que leva à identificação de uma composição dos elementos envolvidos que pode ser sintetizada como:

$$\text{Custo} = \sum_{i=1}^n \text{Rep}(C_i, N_i) \quad (1)$$

Agregue-se, portanto, ao custo da transação, a necessidade de replicação de funcionalidades e de investimento de horas em cada um dos detentores, conforme explicitado na figura 4, o que multiplica os esforços necessários para a representação destes conteúdos, inviabilizando este processo, e que pode ter a composição dos elementos envolvidos sintetizado como:

$$\text{Custo Total} = \sum_{i=1}^n \text{Custo}(\text{Detentor } i) \quad (2)$$

Esta síntese pode ser melhor visualizada por meio da figura 4.

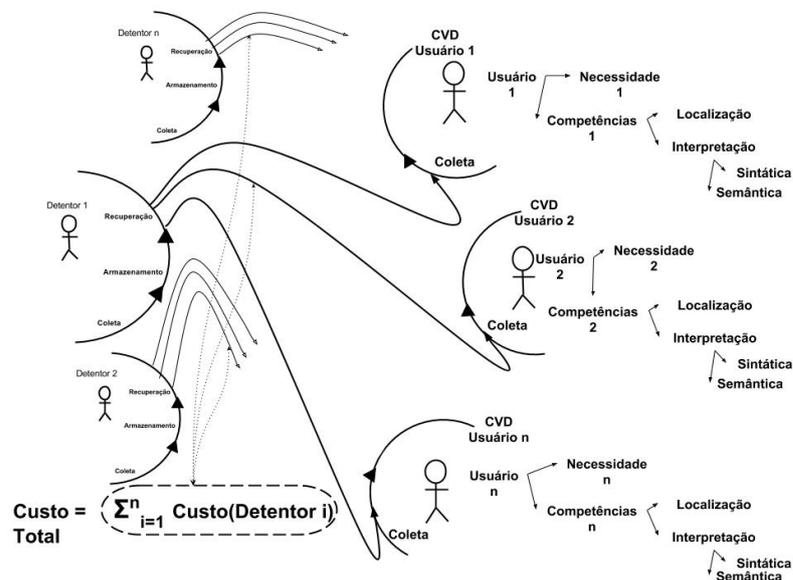
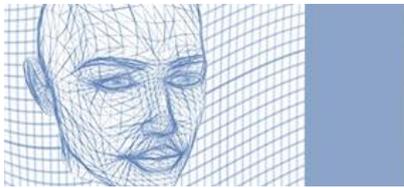


Figura 4. Custo total de Representação de Dados em múltiplos detentores.



Esta inviabilização se concretiza em função, em parte, pela desvinculação dos objetivos relacionados à fase de coleta e armazenamento dos dados e à fase de recuperação destes dados e, ainda, se justifica pelo que o linguista George Zipf chamou de “Princípio do Menor Esforço” (ZIPF, 1949), em que o comportamento informacional humano, como qualquer outro processo biológico, tende a salvar energia mesmo ao custo de redução da qualidade da informação envolvida no processo.

Uma alternativa a este modelo poderia ser a adoção de um elemento intermediário e exógeno aos dois extremos do processo, como acontece na recuperação da informação, em que mecanismos de busca atendem às demandas dos dois lados, evitando custos por parte dos detentores, conforme explicitado na figura 5.

No entanto, mais uma vez a especificidade inerente à recuperação de dados impede que a participação de um elemento agregador destes conteúdos possa realizar sua tarefa de representar os conteúdos, parte pela falta de semântica do lado dos dados, parte pela característica de exatidão que se espera dos resultados destes processos de recuperação.

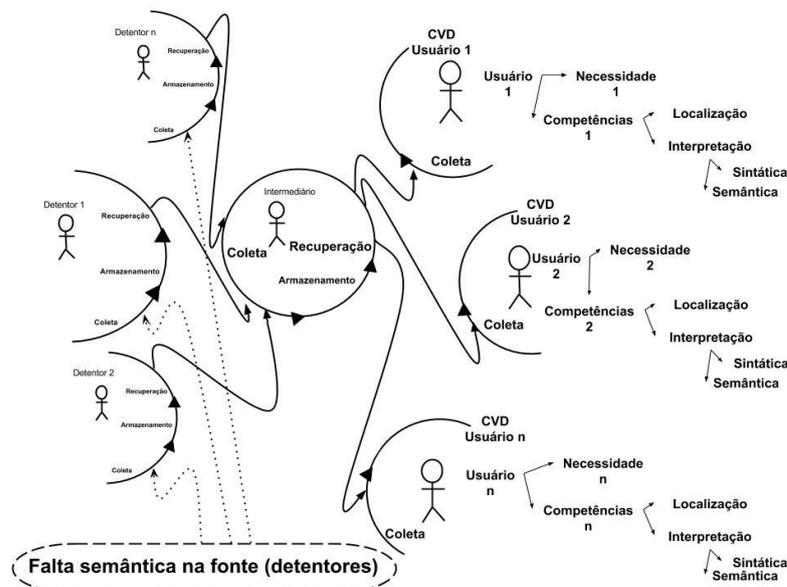


Figura 5. Intermediário entre detentores e usuários.

Uma alternativa seria a disseminação das funcionalidades de representação dos conteúdos em cada um dos usuários. No entanto, apesar da possibilidade de



redução da lacuna semântica em função do conhecimento prévio do usuário sobre o contexto ao qual os dados tendem a atender em função da proximidade com sua pesquisa, volta a cena a questão da multiplicação do custo de transação conforme ilustrado na figura 6.

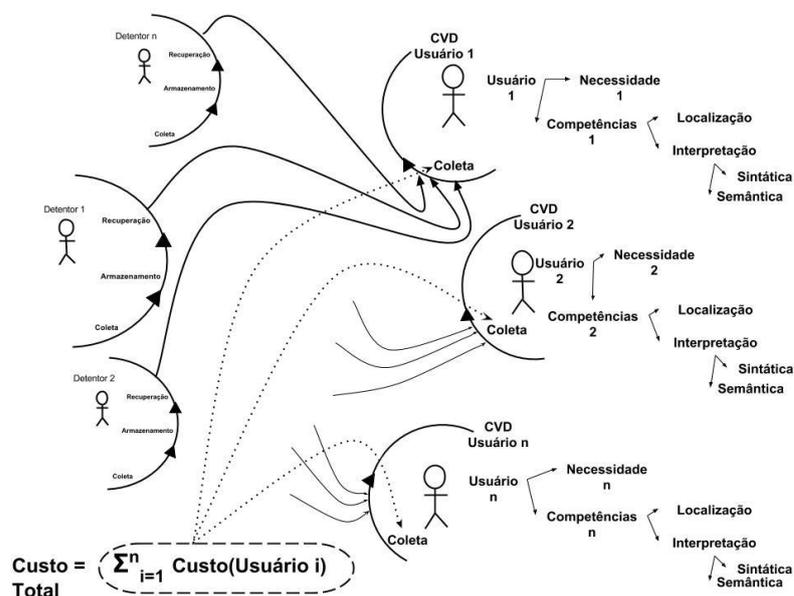


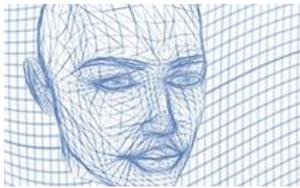
Figura 6. Distribuição das funcionalidades de representação em cada usuário.

Assim, percebe-se que o processo de coleta dos usuários ficaria complexo e altamente dependente de grande capacidade computacional para executar o processo de localização, tratamento e interpretação de conteúdos que viriam de muitas fontes distintas, que apesar de sua semântica mínima, representada pela tríade <e,a,v>, tornaria o processo altamente custoso, conforme pode ser observado em uma síntese da composição dos elementos envolvidos:

$$\text{Custo Total} = \sum_{i=1}^n \text{Custo}(\text{Usuário } i) \quad (3)$$

Esta alternativa ainda levaria à requisição de grande potencial de comunicação e armazenamento em toda a rede e, em especial, nas camadas mais próximas do usuário, o que leva a um caminho inverso ao que vem sendo adotado pela tecnologia digital que tem buscado tornar os dispositivos do usuário os mais leves possíveis e ainda manter todo o armazenamento e processamento por acesso remoto.

O mesmo conceito de custo de transação foi aplicado aos sistemas de recuperação computacionais pela Lei de Mooers, que propõe que, independentemente da excelência de qualquer sistema de informação, os usuários irão descartá-lo se a



recuperação for mais custosa (ou incômoda) do que não ter a informação de qualquer forma (LIU & YANG, 2004; MOORE, 1996; BERRIO-ZAPATA & SANT'ANA, 2015).

3 Conclusão

A partir de (1), (2) e (3), percebe-se que não se apresentam soluções neste texto, mas somente a proposta de elementos que devem ser considerados no estabelecimento de contextos para reflexão sobre o processo de representação de dados.

São descritos cenários com possíveis alternativas para representação sendo feita no detentor, em um intermediário ou no usuário, todas elas mostrando-se de difícil viabilização com base nos modelos e tecnologias atuais.

Vale destacar que, por suas especificidades em relação ao processo de recuperação da informação, o processo de recuperação de dados provavelmente não seguirá o mesmo caminho trilhado pela recuperação da informação, fato confirmado pela falta de alternativas que tenham sido amplamente adotadas até o momento, mesmo considerando-se o grande potencial econômico que tal conquista traria.

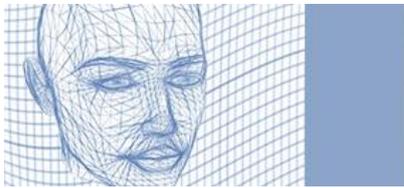
Referências

ARAUJO, C.A.A. **Correntes teóricas da ciência da informação**. Ciência da Informação, v.38, n.3, set./dez. 2009. p.192-204. Disponível em: <<http://revista.ibict.br/ciinf/article/view/1240>> Acesso em: 10 mar. 2015.

BERRIO-ZAPATA, C.; SANT'ANA, R.C.G. **Transparency and open data in the classroom: a pedagogical exercise to construct civic awareness about access to public digital data in Brazil**. Int.J.Electronic Governance, v.07, n. 4. 2015. p. 313-332. Disponível em: <<http://www.inderscienceonline.com/doi/pdf/10.1504/IJEG.2015.074330>> Acesso em: 10 jan. 2017.

COASE, R. **The Firm, the Market, and the Law**. Universidade de Chicago: Chicago, 1988.

DAVENPORT, T. H. **Big data at work: dispelling the myths, uncovering the opportunities**. Harvard: Harvard Business School Publishing, 2014.



JANOWICZ, K. et al. **Geospatial semantics and linked spatiotemporal data: past, present, and future.** Semantic Web, v. 3, n. 4, p. 321-332, 2012.

LIU, Z.; YANG, Z.Y.L. **Factors influencing distance-education graduate students use of information sources: a user study.** The Journal of Academic Librarianship, v.30, n.1, p.24-35.

MANSELL, R.; NOLAN, J.; WEHN, U. **Knowledge Societies: Information Technology for Sustainable Development.** United Nations Publications: New York, 1998.

MOORE, C.N. **Mooers' Law or why some retrieval systems are used and others are not.** Bulletin of the American Society for Information Science and Technology, v.23, n.1, p.22-23. Disponível em:
<<http://onlinelibrary.wiley.com/doi/10.1002/bult.37/full>> Acesso em: 10 jan. 2017.

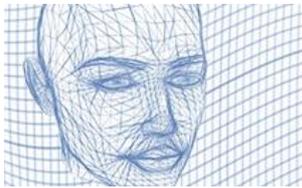
RODRIGUES, F.A.; SANT'ANA, R.C.G; FERNEDA, E. **Análise do processo de recuperação de conjuntos de dados em repositórios governamentais.** InCID: R. Ci. Inf. e Doc., Ribeirão Preto, v. 6, n. 1, p. 38-56, mar./ago. 2015. DOI: 10.11606/issn.2178-2075. v.6. n.1. p.38-56. Disponível em:
<<http://www.revistas.usp.br/incid/article/view/73496/96247>> Acesso em: 10 jan.2016.

SANT'ANA, R. C. G. **Ciclo de vida dos dados e o papel da ciência da informação.** In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. Anais do ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO. Florianópolis, 2013. Disponível em:
<<http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/284/319>>. Acesso em: 5 maio 2014.

SANT'ANA, R. C. G. **Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação.** Informação & Informação, [S.l.], v. 21, n. 2, p. 116–142, dez. 2016. ISSN 1981-8920. Disponível em:
<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940/20124>>. Acesso em: 10 dez 2016.

SANTOS, P. L. A. C.; SANT'ANA, R. C. G. **Transferência da informação: análise para valoração de unidades de conhecimento.** DataGramaZero, v.3, n.2, 2002. Disponível em:
https://www.researchgate.net/publication/316859684_Transferencia_da_Informacao_analise_para_valoracao_de_unidades_de_conhecimento_Transference_of_Information_analysis_for_valuing_units_of_knowledge. Acesso em: 07 mar. 2015.

SANTOS, P. L. V. A. C.; SANT'ANA, R. C. G. **Dado e Granularidade na perspectiva da Informação e Tecnologia: uma interpretação pela Ciência da Informação.** Ciência da Informação, Brasília, v. 42, p. 199-209, 2013. Disponível em:<
<http://revista.ibict.br/ciinf/article/view/1382>> Acesso em: 20 mar. 2015.



I ENCONTRO DE REPRESENTAÇÃO DOCUMENTAL - EnReDo

Grupo de Pesquisa
GPTAI
Tecnologias em Ambientes Informacionais

VAN RIJSBERGEN, C. J. **Information retrieval**. 2. ed. Londres: Butterworths, 1999.
Disponível em: <<http://www.dcs.gla.ac.uk/Keith/Preface.html>>. Acesso em: 15 jan.
2016.

ZIPF, G.K. **Human Behavior and the Principle of Least Effort: an introduction to human ecology** (reimpressão de 2002 de edição de 1949), Addison-Wesley Pres, Oxford, Inglaterra, 1949.