

A Ciência da Informação e os dados de pesquisa

Information Science and research data

Luana Farias Sales Marques^a

Transcrição da apresentação

Obrigado Guilherme pela apresentação, obrigado pelo convite, estou muito feliz de estar aqui, adoro essa terra e falo para todo mundo isso. Então agradeço o convite pois fiquei super empolgada desde o primeiro momento em que o Guilherme me falou, e agradeço ainda mais por esse tema que o Guilherme sugeriu, porque me deu a oportunidade de rever, assim como a professora Plácida falou que voltou lá nos arquivos antigos, eu também acabei fazendo um mix das coisas que eu venho falando mas de uma forma de repensar e reconfigurar aquele conhecimento que a gente vem desenvolvendo, eu, o professor Sayão e o nosso grupo de pesquisa. Eu coloquei aqui o nome de todos deles, depois se alguém quiser participar e interagir com a gente tem e-mail aqui.

Eu preparei uma agenda, porque assim, a Ciência da Informação apesar de ser uma área pequenininha a gente tem um mundo de coisas para estudar. Então eu fiquei pensando no que eu vou falar sobre Ciência da Informação nas pesquisas, é muita coisa, como é que eu vou fazer esse recorte, como eu vou organizar isso.

Pensei bem e decidi organizar isso pelos GT's da ANCIB porque é a nossa organização das pesquisas em Ciência da Informação no Brasil, e aqui gente nós temos os nossos GT's. E eu agradeço a oportunidade ainda porque ontem quando saí daqui o pessoal achou que fomos passear mas não ontem eu estava tendo ideias, vocês vão falando aqui e a gente vai tendo ideias o tempo todo.

Eu queria trazer para vocês, como a gente está num ambiente de pós-graduação, perspectivas de estudo, trazer um monte de questões, questões que podem passar por todos os GT's da ANCIB, porque eu sei que aqui não deve ter pessoas só do GT8, eu mesma originalmente sou do GT2, de Organização do Conhecimento, agora estou interagindo com o pessoal do GT8 de tecnologia e com o de Comunicação Científica, e no meio de tanta coisa tentando organizar essas ideias e eu espero que vocês gostem e que a gente consiga aprender um pouquinho junto, porque eu estou no aprendizado também.

Então eu preparei uma agenda, um sumário da minha apresentação, eu vou começar falando primeiro do nosso lugar, mas é porque eu achei importante entender o que eu estava compreendendo por Ciência da Informação, porque eu sei que vocês estão estudando Ciência da Informação aqui na Federal da Paraíba.

Escolhi um autor para a gente partir dele, escolhi um clássico para definir. E depois o que são dados de pesquisa? Dados de pesquisa podem ser diversas coisas, então nós optamos por trazer uma definição também para falar um pouquinho nessa apresentação sobre os dados de pesquisa. Depois eu faço uma relação entre os dados de pesquisa e publicação científica, para a

^a Universidade Estadual Paulista (UNESP). E-mail: placida@marilia.unesp.br. Currículo: <http://lattes.cnpq.br/7408791408049766>

gente compreender um pouco essa relação. Depois a gente apresenta um pouco da problemática dos dados de pesquisa. E aí a gente fala da Ciência da Informação no Brasil, às perspectivas de estudos, e os dados de pesquisa nesse passeio pelos GTs da ANCIB.

Eu escolhi uma frase do Borko que é uma definição que gente trabalha muito na Ciência da Informação do IBICT, e eu destaquei aqui uma parte importante da fala dele onde ele define a Ciência da Informação como aquela ciência que está preocupada em como este conhecimento relacionado a origem, correção, organização, armazenamento, recuperação, interpretação, transmissão e utilização da informação, e diz ainda que envolve pesquisas em representação da informação, uso de códigos para transmissão eficiente da mensagem, bem como o processamento das técnicas aplicadas aos computadores e sistemas de informação.

É bem o que a gente está estudando agora só que estamos falando ed dados de pesquisa, e a gente vai ver um pouquinho também desse caminho, dos dados até o conhecimento, o dado que se transforma em informação e a partir da informação a gente tem o conhecimento.

A definição da informação a gente trabalhou com outro clássico de Belkin e Robertson que dizem que a informação é tudo aquilo capaz de transformar estruturas, e o Buckland que apresenta três perspectivas da informação. Buckland fala da informação como processo que é o ato de transformar alguém, a informação como conhecimento que é aquilo que a gente percebe quando se sente informado, e a informação como coisa, que é a informação registrada e representada de alguma forma.

Eu trouxe também essa definição aqui de Belkin e Robertson onde eles falam que um texto é um conjunto de signos organizados por um emissor com o objetivo de mudar a estrutura da imagem do receptor. A gente compreende que um texto é um conjunto de signos organizados, e a gente vê que o signo é o significante que assume um significado, e isso é informação, a informação é o dado organizado, que é o dado mais o metadado.

A gente pensa que um conjunto de signos organizados se transforma em informação quando ele modifica a estrutura cognitiva do receptor, isso é o que a gente está compreendendo como dados. Então os dados é um conjunto de signos e para ele se transformar em informação ele precisa ser representado, organizado, ou seja, precisa de metadados, de ontologias, ele precisa de estrutura.

Afinal o que é dado de pesquisa? Segundo Christine que é uma pessoa muito citada quando a gente fala de dados de pesquisa, e ela fala assim, informação é um conceito complexo com dezenas de definições, já o dado por sua vez é um conceito simples, com poucas definições mas sujeito a muitas e diferentes interpretações.

Atribuir uma definição processual a dados de pesquisa é um fato idiossincrático que pode dizer muitas coisas diferentes para pessoas e coisas diferentes. Isso acontece porque o dado de pesquisa é independente de interpretação.

Tudo pode ser dado de pesquisa dependendo do conceito e da pessoa, e ele pode se configurar em formas muito diferentes. Aqui a gente traz um exemplo, uma sequência de bits provenientes de um sensor sísmico é dado de pesquisa para os sismólogos, às amostras de rochas podem ser dados de pesquisa para as geomorfológicas, às conversas gravadas são dados de

pesquisa dos sociólogos, então o dado de pesquisa é independente de interpretação e pode dizer muitas coisas diferentes para pessoas e circunstâncias diferentes.

Aquí a gente tem outro exemplo, um unicórnio por exemplo, pode ser dados de pesquisa para um ambientalista e também pode ser dado para um arqueólogo, e aí a gente começa a ver essa habilidade de reuso.

Os dados sísmicos também podem ser dados para os biólogos, e os dados são sempre tratados tomando como base algum interesse ou perspectiva e a prática que determina seu significado e sua prática em distintos contextos.

A gente trata por meio de metadado, ela é a base da representação, a forma de tratar o dado vai ser completamente diferente da forma como a gente representa a informação. Quando a gente pensa em informação a gente pensa naquela coisa já registrada, artigos de periódicos, os livros, que já tem um contexto e uma estrutura. É muito simples a gente tratar um livro ou uma obra bibliográfica que você tem lá todo tipo de data, agora como a gente faz o tratamento desse dado de pesquisa?

Dos dados de pesquisa a informação científica, entendendo um pouco essa relação. A informação científica e tecnológica eu trouxe aqui uma definição compreendendo isso, é aquela informação que serve como insumo para atividade de pesquisa ou tecnológica, e ela resulta resultados de atividades científicas. Eu gosto de falar sobre isso porque como a gente está falando de dados, eles aparecem em diversos contextos por exemplo Big Data, mas nós estamos falando sobre um dado específico que é o dado produzido sob resultados de pesquisa, que é diferente de você tratar dados comerciais e a gente tem que entender um pouco isso.

A informação científica ele serve para divulgar conhecimento novo obtido a partir de uma pesquisa científica, ela serve como insumo para novos projetos de pesquisa, serve para explicitar a metodologia empregada nos processos de pesquisa. Então os dados vão ser uma forma de você documentar e transformar aquele conhecimento produzido dentro dos laboratórios em uma informação científica.

É importante a gente lembrar de uma coisa, os dados de pesquisa eles são tanto insumo quanto produtos da pesquisa científica, a gente tem tratado muito dado como insumo mas a gente esquece que ele também é produto, e ele passa por diversas transformações. Eu costumo dar exemplo no trabalho com pesquisa na área nuclear, a gente lá um domínio da realidade virtual e eu tinha dados que eram diferentes, o dado que o cara coletava filmando às atividades do pesquisador dentro do reator, duvido que ele tenha um resultado que é o vídeo em 3D. Assim você tem os estágios de tratamento do dado que também era diferente.

Então você tem o movimento em que o dado é capturado, esse dado já passa por uma curadoria, e você tem os dados que depois são transformados a partir dessa curadoria, dessa análise. Sabia como dados, informações e conhecimento são definidos é saber quais são as formas de transformar dados em conhecimento, eu queria definir isso com vocês.

Dados representa fatos desorganizados ou processos. Já informação é uma agregação dos dados processados. O conhecimento é a compreensão humana sobre o assunto adquirido sobre o estudo ou experiência. Os dados também geralmente são de natureza estática, já a informação possui algum significado e propósito, como falei a informação é o fato mais o metadado, a

representação. Então a gente fala de representação de um dado que já tem uma representação e consequentemente um significado, e é representado para algum propósito específico.

Já o conhecimento é baseado na aprendizagem, e compreensão do problema, é a interpretação que a pessoa faz com essa informação. O dado pode representar um conjunto de fatos sobre eventos, já a informação precisa de um contexto, para o dado se transformar em informação eu preciso ter isso contextualizado. O dado é pré-requisito para que haja informação, já a informação é pré-requisito para que haja conhecimento.

Um exemplo, 45%, se eu falar apenas isso eu tenho um dado, mas 45% está relacionado a que? Ah, 45% da população consome doces e frituras, aqui eu já tenho a informação que é o dado contextualizado.

E aí o conhecimento é o que? 45% da população não é saudável, é a inferência que eu fiz aqui com a informação que eu recebi a partir do dado que me foi dado. A gente tem outro exemplo, os dados como caracteres ou imagens como atributos de um fenômeno, eles vão se transformar em informação quando combinados de maneira em potencial de regular padrões sobre um fenômeno. E o conhecimento é um sistema de inferências não triviais e verdadeiras acerca de um fenômeno. Aqui eu tenho os números medindo a altura de uma montanha, esses números são os dados, e aí eu tenho a fórmula que vai dizer que a altura do pico pode ser deduzida a partir dos dados se transformando em informação. Eu combino essas informações e aí eu tenho o conhecimento de que às rochas da montanha.

Aqui para falar sobre a problemática dos dados. Os dados com essa zona, o estudo a pesquisa sobre dados de pesquisa está uma zona, a gente ainda não sabe como vai organizar isso, eu fiz uma proposta de levar para um DT da ANCIB, porque a pesquisa em Ciência da Informação se dá através de programas de pós-graduação e o país possui uma instância de representação científica e política importante para o debate que é a ANCIB, e a ANCIB atua em duas frentes, nos programas de pós-graduação e no encontro de pesquisa através dos GTs.

A temática dos dados de pesquisa pode ser estudada sob vários aspectos, escolhemos a abordagem sob a perspectiva dos GTs. Aqui como falei para vocês são 11, e começaremos então essa organização por GTs.

O GT1 é o GT que trata dos estudos históricos e epistemológicos da Ciência da Informação e essa é a ementa. Eu levante algumas perspectivas de estudo que me veio na cabeça, mas talvez vocês analisando a ementa encontre outras questões. Eu percebi olhando aqui que nós estudar a epistemologia dos dados de pesquisa, quais dados de pesquisa são conceituados a medida com que são gerados, a gente precisa estudar a história dos dados e das bases de dados na Ciência da Informação, o conceito de dados de pesquisa, a cadeia dado-informação-conhecimento como eu mostrei, a necessidade de tratamento interdisciplinar de dados de pesquisa e a possibilidade de reúso.

É importante a gente compreender a partir da perspectiva da epistemologia da Ciência da Informação como os dados se originam para poder compreender a história dos dados, em uma determinada apresentação a gente criou essa definição para dizer que os dados podem ser observacionais, isso é só uma abordagem de classificação, tem diversas.

A gente classificou aqui os dados quanto a sua origem e percebeu que o dado pode ser observacionais como aqueles gerados através da erupção de um vulcão, eles são críticos porque não tem como reproduzir essa pesquisa e o levantamento desses dados. Então eles precisam ser armazenados e a gente tem um problema crítico aqui, os dados podem ser também experimentais, provenientes de situações controladas em laboratórios, e os dados podem ser computacionais.

Aí vamos ter algumas barreiras disciplinares para tratamento e organização desses dados, porque cada dado vai se manifestar de forma diferente de acordo com sua área e com o contexto de pesquisa, de acordo com a metodologia utilizada.

A professora Plácida mostrou aqui, os dados podem ser números ou diversas naturezas, o que é diferente de tratar um documento bibliográfico. E esses dados podem ser facilmente reutilizados em contextos diferentes, os dados podem ser reusados e transportados através mix setoriais, podem ultrapassar limites disciplinares.

Algumas perguntas: O que é um dado de pesquisa? Qual o valor dos dados na pesquisa contemporânea? Isso é importante a gente investigar, eu pus algumas dicas de pesquisa para vocês. Como os dados podem alimentar o interdisciplinar? E aí então nós entramos no GT2, que é meu GT de origem.

O GT2 estudo teorias, metodologias e práticas relacionadas à organização e preservação dos documentos e da informação enquanto conhecimento registrado em ambientes informacionais tais como arquivos, museus, bibliotecas e congêneres. Compreende também os estudos relacionados aos processos e produtos de representação do conhecimento e às relações inter e transdisciplinares verificadas. Então a perspectiva desse estudo eu levantei que é a curadoria de dados, que nada mais é que tratamento e representação do conhecimento, a gente às vezes confunde o conjunto de curadoria com o conceito de gestão mas é diferente, a questão da organização dos dados e da representação temática desses dados de pesquisa.

Aí então às pesquisas de metadados, taxonomias, vocabulários, e uma questão nobre que surge aqui é às anotações de dados, que são às anotações dos pesquisadores enquanto estão fazendo experimentos da pesquisa.

Aquí para mostrar um pouquinho do ciclo da curadoria dos dados nós temos perspectivas de estudos sobre todas essas fases do ciclo da curadoria, os metadados ele tem um forte capacidade para dados digitais de transmitirem conhecimento e possibilitam que estes sejam interpretados e usados.

Aquí tem alguns exemplos de padrão de metadados, cada área vai ter seu padrão de metadado, aqui é uma área que não tem seu padrão de metadados mas vai ter que constituir, e o profissional da informação vai ser importante para isso, lembrando que são os profissionais da informação que estivemos na base de criação do Dublin Core e hoje nós temos por exemplo o Darwin Core que é um padrão de metadados utilizados na área de Biodiversidade e que está apoiado no Dublin Core e às outras áreas também vão ter que começar.

Ainda sobre o GT2 eu trouxe aqui um monte de questão. Que padrões de metadados e ontologias são necessários e adequados para tratar dados de pesquisa? Como integrar semanticamente dados e publicações de outros objetos digitais, e esse foi meu tema de Tese de

Doutorado quando estudei as publicações ampliadas. Você tem que dar semântica, dar contexto e definição para as relações e links. Quais os padrões de metadados utilizados para tratar os dados nos domínios disciplinares? Como visualizar e modelar os sistemas para recebimento desses dados? Como integrar semanticamente dados representados sob a perspectiva de domínios distintos, e isso é muito importante porque a gente está pensando no reuso dos dados para promover pesquisas interdisciplinares que é um tema tão importante para a Ciência da Informação. Que padrões utilizar para anotar os dados e permitir o processamento inteligente, porque o cara está anotando no laboratório mas está fazendo tudo sem padrão, e quando a gente criar as bases de conhecimento utilizando ontologia e a gente precisar processar esse conhecimento, essa linguagem precisa ser padronizada, precisa de um formato para essa anotação.

E a gente entra então no GT3, Mediação, Circulação e Apropriação da Informação, que estuda os processos e relações da mediação, circulação e apropriação da informação em diferentes contextos e tempos históricos, considerando sua complexidade e abrangência, bem como relacionado com a construção do campo científico da Ciência da Informação.

Perspectivas de estudos, a relação da Big Science e da Small Science e a geração de dados, são formas distintas de gerar dados científicos e que precisam se relacionar, o Sayão falou um pouquinho no primeiro dia dos dados gerados lá na cauda longa que são dados heterogêneos que precisam se combinar com dados lá da cabeça da ciência que é a Big Science para que a inovação seja gerada no meio da ciência.

Falei um pouco disso da Big Science e Small Science, essas duas ciencias precisam estar integradas para a geração de conhecimento, e reuso desses dados em contextos diferentes para gerar inovação. Isso é um ecossistema de dados, e nós temos às perguntas: Quais às diferentes na gestão de dados gerados pela grande e pela pequena ciência? Como vou integrar dados diferentes e gerar inovação? Quais dados devem ser preservados para sempre? Como avaliar o potencial de reuso de dados e estabelecer prioridade para coleções de dados. Na prática a gente gera tantos dados e às vezes não tem tempo, por isso precisa priorizar como estabelecer as prioridades.

GT4, que é o GT de Gestão da Informação e do Conhecimento, acaba sendo meu GT também porque eu estou muito na área de gestão do conhecimento nuclear, abre um parêntese, é a Ciência da Informação que trabalha com outro nome, porque para reunir um grupo de pessoas que tratassem dessas informações, mas para mim é tudo Ciência da Informação, e que lá eles chamam de gestão do conhecimento por conta das modalidades de financiamento internacionais.

Esse GT4 trabalha com ambientes, sistemas, ambientes, produtos de informação, estudos de fluxos e processos em ambientes de informação como instrumento de informação, gestão de conhecimento e aprendizagem informacional na Ciência da Informação, Marketing da informação, estudo de redes para gestão.

Perspectivas de estudo: gestão de dados de pesquisa como um viés para gestão do conhecimento, aqui a gente tem trabalhando lá na área nuclear na gestão de dados mas é um caminho para a gestão do conhecimento e a gente tem que entender isso, porque a cadeia dado-informação-conhecimento, sem o dado não chega até o conhecimento. Estudo e mapeamento dos fluxos de pesquisa nos laboratórios, documentar para criar memória da instituição,

compartilhamento de dados, curadoria digital como serviço de disseminação seletiva, é um tema que não gosto muito mas que está sendo usado muito na área da computação e agora na área de gestão que é o evangelismo de dados, que é você divulgar a importância da gestão dos dados e da curadoria. Curadoria de dados é isso e às pessoas ainda não conseguiram se conscientizar da importância disso. Apesar do nome esquisito o Guilherme e o Sayão dizem que vai se criando uma igreja universal dos dados, [risos].

Para fazer a gestão do conhecimento precisa-se fazer a gestão dos dados e da informação, o conhecimento é derivada da informação da mesma maneira que a informação é derivada do dado, e a gestão dos dados é uma parte essencial dessa gestão do conhecimento.

E quais são as questões que nós temos aqui. Quais as diferenças entre gestão de dados, gestão da informação e gestão do conhecimento, e como integrar esses processos? Como mapear os fluxos de dados e informação nos laboratórios? Quando a gente faz esse mapeamento das atividades de pesquisa dos pesquisadores é importante porque aí a gente vê onde os dados são gerados, onde esses dados estão sendo transformados, como eu apresentei no slide anterior os dados passam por estágios de geração. Eu tenho um dado gerado por um experimento científico que é o primeiro estágio da geração, aí esse dado passa por um processamento e se transforma em outra coisa, são os dados de segunda geração. Aí o pesquisador vai, analisa, e tira às conclusões, eu tenho dados de terceira geração, isso é muito legal para a gente saber como coletar os dados.

Os dados de pesquisa podem servir como um mecanismo de preservação do conhecimento das instituições? Gerenciar o fluxo de dados de pesquisa pode servir como um mecanismo de transformação do conhecimento tácito em conhecimento explícito?

GT5. É o GT de Política e Economia da Informação. Esse GT estuda às políticas emergentes de informação, propriedade intelectual, acesso a informação, poder e cidadania, ética da informação, informação e poder, agora olhando aqui já lembrei de outras questões. Quais as perspectivas de estudo: ambientes de geração de dados, humanidades digitais, e-Science, ciência da cauda longa, que são ambientes que geram dados de pesquisa, assim como a grande ciência e pequena ciência, que são outras formas de classificar esses ambientes de geração de dados. Políticas de gestão de dados, políticas nacionais, locais, políticas de repositórios, porque a gente fica sempre pensando na política nacional e quer que a situação venha de cima para baixa, mas podemos começar criando às políticas para repositórios. A questão da sustentabilidade na gestão dos dados, a criação de laboratórios e redes de colaboração de dados, ética no compartilhamento e no reúso, reúso de dados como mecanismo de economia de recursos.

Aqui como falei os dados de pesquisa existem para serem usados, então a gente tem que transformar esses dados em coisas que podem ser recusadas, se os dados de pesquisa não estão sendo reutilizados perdem o sentido por isso precisam ser estudados também.

Eu trago aqui que nem algumas questões de ética que precisam ser estudados, coletar, usar e compartilhar os dados em ambientes de pesquisa que envolva pessoas obrigam que condições éticas e legais sejam respeitadas, isso deve permear toda a pesquisa incluindo a fase de compartilhamento de dados. Nesse contexto também é importante estudar as diferenças dos

dados pessoais, quais são os dados confidenciais e como disseminar isso, dados sensíveis são outros tipos de dados que precisam ser estudados.

Aquí tem algumas questões: Como criar coleções de dados que atenda a necessidade de públicos específicos? Como se configura o regime de informação nesses ambientes de geração de dados, na e-Science, Ciência Aberta, na Ciência da Cauda Longa? Como garantir a sustentabilidade desses dados? Como se organizam às redes de colaboração e compartilhamento de dados? Como a curadoria de dados pode otimizar os investimentos de pesquisa?

E aí entramos no GT6, que entra um pouco no que a professora Plácida falou aqui a questão do perfil do profissional. A ementa do GT6 eles estudam competências, habilidades requeridas, sociedade do conhecimento, tecnologia e trabalho, saúde no mercado de trabalho e ética nas profissões de informação, formação do profissional em Ciência da Informação, paradigmas predominantes. Então coloquei como perspectivas de estudo esses cenários de geração de dados, e-Science, Ciência Aberta, Ciência da cauda longa, e os atores, quem são os atores que trabalham com esses dados? Qual o perfil do profissional que vai consumir os dados, que pode ter mais de um perfil de acordo com a área? Quem é esse bibliotecário, qual o perfil desse bibliotecário que vai trabalhar com dados, arquivista, profissional de TI, e aqui estão algumas questões tão importantes sobre competência e a gestão de dados.

Qual o perfil do profissional que vai trabalhar com esses dados? Qual o currículo mínimo para formação desse profissional? Nós temos diversos papéis nos dados de pesquisa, temos o papel do pesquisador, do bibliotecário de dados, são novas roupagens para as profissões tradicionais.

Nós temos também o cientista de dados como um papel novo que é o profissional da área de computação, a gente tem o gerente de dados que é o cara da TI responsável pela manutenção e criação de bases de dados. E tem o curador de dados, que pode ser o próprio pesquisador, o cientista da informação com conhecimento interdisciplinar que vai adicionar valor aos dados através da curadoria, representação, anotação, compartilhamento e reúso desses dados. E a competência com dados de pesquisa a gente vem trabalhando isso por conta da pesquisa de uma aluna que é a capacidade de ler, entender, criar e comunicar dados como informação. E como sendo às competências envolvidas no trabalho com os dados de pesquisa desde a coleta até o compartilhamento passando pela rotina de curadoria e tratamento de dados, nós pegamos a literatura e criamos uma definição própria porque partimos de algum lugar para começar nossos estudos.

E a questão da competência de dados, a gente tem duas questões super importantes, que é que habilidades de curadoria e gestão são necessárias para o pesquisador que desenvolve pesquisas no âmbito da e-Science? E como os bibliotecários podem aplicar seus conhecimentos de recuperação, disseminação e preservação da informação para ensinar essas competências para os pesquisadores. Então a gente tem que saber qual o conhecimento o pesquisador precisa, qual o conhecimento o bibliotecário precisa, e treinar esse pessoal.

Qual o papel do GT6. Aqui algumas perguntas: Qual o valor da competência informacional nos dados de pesquisa? Qual o papel do profissional da informação na gestão de dados e quais outros atores importantes? Qual o perfil do profissional para trabalhar com dados de pesquisa?

E aí a gente entra no GT7 que também é um pouco meu GT, agora que eu venho trabalhando muito pesado nas questões de comunicação científica, que estuda os estudos teóricos, metodológicos e aplicados sobre a produção e o uso da informação sobre ciência e tecnologia da informação. Inclui pesquisas relacionadas aos processos de comunicação e análise de indicadores para planejamento e avaliação de C&T.

Tem como perspectivas de estudo novos padrões de comunicação científica, a gente tem alteração no ciclo da comunicação científica, nova configuração do fazer científico que foi reformulada, antes tinha às hipóteses e depois íamos coletar os dados, agora nós temos os dados e a partir dos dados a gente observa e cria novas hipóteses. Nós temos novos modelos de publicação, publicação composta, publicação ampliada, e outros modelos como data papers, em que eu compartilho os dados e os metadados e indico nesses artigos o artigo que apresenta a pesquisa.

Temos os cadernos de laboratório, periódicos de resultados negativos, porque a gente precisa dar visibilidade a esses dados e temos estudado a questão da ciência invisível, que é a ciência que está sendo produzida nos laboratórios e ninguém tem conhecimento, tem uma pesquisa que diz que 50% dos achados da pesquisa não são publicados, quais os motivos para o pesquisador não compartilhar?

E aqui algumas questões, como promover visibilidade para esses dados de pesquisa, esses que estão lá na ciência invisível dentro dos pendrives ou no HD dos pesquisadores? Como se configuram os novos modelos de publicação e seu impacto na comunicação científica? A gente vai ter novos periódicos com avaliação por pares abertos, modelo aberto, e tudo isso vai reconfigura a comunicação científica porque não terá mais que esperar porque no futuro esse processo será importado. Hoje a gente começa a produzir uma pesquisa e a gente manda para a revista, e mesmo com toda a tecnologia e automatização ainda é lenta, porque precisa de uma revisão por pares. Como a gente também está do outro lado, quando você recebe um pedido de avaliação e às vezes a gente se enrola com outras atividades e acho que esse é um dos motivos. Como citar esses dados e garantir que o pesquisador receba créditos pelo compartilhamento? Às vezes o cara não quer compartilhar porque ele não vai ser citado, e existem normas e padrões para citação de dados, os bibliotecários da plateia precisa saber como citar esses dados para auxiliar os pesquisadores.

Aí a gente entra no GT8, nosso GT também que é o GT de Informação e Tecnologia, e estuda pesquisas teóricas para desenvolvimento de tecnologias de informação e comunicação, processos de recuperação, armazenamento, disseminação, uso, gestão, segurança e preservação e ambientes digitais. Nós temos aqui a questão dos dados de pesquisa e os metadados, os protocolos de pesquisa de campo, os processamentos de ontologia, processamento dos dados, repositórios e bases de dados, a gente volta com esse conceito de base de dados que ficou esquecido, a gente vem falando muito dos repositórios institucionais como se eles fossem resolver o problema dos dados e estamos começando a ver que não, ele é muito bom para por lá a produção científica e o Google fazer lá o processo de indexação. Quando a gente fala de bases de dados a gente precisa de ambientes específicos para que esses dados sejam recuperados, processados e transformados em conhecimento.

A questão das redes, da representação digital, da gestão dos dados de pesquisa, o armazenamento em nuvens, em quantidade como tera ou petabytes. Às aplicações e geração de novos serviços, muitas vezes os repositórios não estão sendo utilizados porque não tem serviços, então essas são questões a serem discutidas no GT8, e aqui eu já falei dos repositórios e bancos de dados científicos, a Sun Foundation já definiu em 2003 os repositórios como bases de dados científicos digitais para pesquisa e agora no futuro tendo como perspectiva primordial o acesso aberto que é considerado hoje patrimônio da humanidade. A primeira exigência para curadoria é o dado está disponível em um repositório ou base de dados.

Aqui alguns exemplos, como serviços de visibilidade aos dados, serviços de compartilhamento, reúso da informação, serviços de crédito ao autor, serviço de transparência a memória científica, segurança dos dados, outros tipos de serviços inovadores, interoperabilidade, redes de repositório, indicadores de qualidade e produtividade, para citar alguns.

Aqui nós temos algumas questões subjacentes a algumas plataformas de gestão de dados de pesquisa, Quais as tecnologias para promover a interoperabilidade entre as plataformas de gestão de dados? Como a tecnologia pode apoiar a criação de serviços para compartilhamento de banco de dados? Como garantir preservação a longo prazo aos dados em formatos digitais? Quais as tecnologias para armazenamento seguro de grandes quantidades de dados? Aposto que o Guilherme pensou em várias outras questões e vocês também, mas essas foram questões que vieram do meu insight.

E aí a gente vem para o GT9 – Museu e Patrimônio da Informação, que tem como ementa análise das relações no museu, patrimônio, bem simbólico, e informação como processo em múltiplas perspectivas teóricas e práticas de análise. Estuda museus e patrimônios de informação, criação e representações. A gente colocou aqui como perspectiva de estudo: curadoria digital de dados de pesquisa e os dados da pesquisa como patrimônio de ciência e tecnologia, como patrimônio da humanidade.

Eu trouxe isso daqui porque em um determinado momento a gente fez um estudo do termo curadoria e ele nasce na área de arquivologia e a gente usa por empréstimo. Acredita-se que os museólogos também tem a contribuir com a pesquisa de curadoria, é você criar coleções de dados e integrar o público específico. Antes das práticas acadêmicas se deslocarem para o meio digital, para o paradigma do Big Data, os museus de história nacional já tinham criado conceito de curadoria antecipando a demanda por gestão e aprimoramento dos dados digitais. Eu lembro das bases de dados de um museu que a um tempo atrás era em papel sobre primatas, eu busquei muito essa base durante o doutorado e não tive acesso a ela.

Eu trouxe como perguntas, quais as técnicas de curadoria utilizada em museus podem ser aproveitado na curadoria de dados? Como os dados de pesquisa podem configurar patrimônio para ciência e tecnologia.

Aí a gente entra no GT10 de Informação e Memória, que estuda a relação entre os campos do conhecimento e a Ciência da Informação na memória social, às pesquisas transdisciplinares que envolvem o conceito e prática da informação e memória, memória coletiva, coleções, discurso e memória, preservação da memória social. Quando eu falo que os dados e que a curadoria digital estão sendo estudados nesses GTs que eu recebi já no último GT para avaliar

sobre dados de pesquisa desses GTs, então tem material sendo produzido dentro desses grupos de trabalho.

Dados de pesquisa seria uma questão de estudo como memória da pesquisa científica, como memória institucional, então eu trouxe os dados de pesquisa como possibilidade da gente entender a memória da pesquisa científica.

A questão da preservação dos dados de pesquisa, mas como disse Sayão os dados de pesquisa são incompreensíveis e portanto sem uso a menos que haja uma descrição detalhada e clara de como e quando foram obtidos, e de como foram produzidos. Os dados de pesquisa não falam por si só, e a gente tem uma interação entre os GTs, por exemplo, o GT de organização do conhecimento auxiliando na descrição desses dados para que eles realmente venham ser memória da pesquisa científica.

Perspectivas de estudo no GT de Memória, a questão da preservação, porque a informação digital não sobrevive naturalmente, a gente precisa trabalhar essa questão da preservação. A mesma tecnologia que muda a pesquisa científica coloca os dados gerados em risco como desafio estratégico, gerencial e político de criar, arquivar e tornar disponíveis esses dados. Mas aí tem outra questão, o que tem quer preservar, tem que preservar tudo? Aqui a gente tem uma complexidade crescente da preservação.

Como preservar esses dados de pesquisa considerando toda essa complexidade? Quando a gente fala de preservação vai para além dos metadados descritivos, a gente tem outros metadados que precisam ser considerados. Como incorporar os dados de pesquisa às memórias das instituições de pesquisa, por ser uma outra questão trabalhada.

E aí passamos para o último GT, o GT de Informação e Saúde que estuda as teorias e os métodos informacionais no contexto da saúde considerando a complexidade, estudo da tecnologia, informação e saúde, informação na representação da saúde, informação saúde e sociedade, políticas de informação em saúde, tudo que a gente tem aqui são da área da saúde.

Qual perspectiva de estudo, por exemplo, os dados de pesquisa com genética, e às peculiaridades próprias dos dados da área da saúde. E às perguntas, quais são as peculiaridades dos dados na área da saúde, os dados são gerados de forma diferente, são usados de forma diferente, então precisa ter um estudo específico sobre esses dados. Dados de pesquisa em saúde podem ser baseados em evidência? É uma questão que eu deixo.

Para chegar ao final, a organização da ANCIB em GTs auxiliou a ter uma visão global das pesquisas desenvolvidas no contexto do Brasil, a gente fez apenas uma análise superficial das temáticas abordadas, e a gente concluiu que as pesquisas sobre dados de pesquisa perpassa todas às áreas da Ciência da Informação. Isso revela o potencial das contribuições da Ciência da Informação para os estudos em dados de pesquisa e a solução de diversos problemas dentro dessa temática.

Eu lembrei aqui também quando falei do GT de saúde, que eu acho assim, é uma sub interdisciplinaridade, porque a Ciência da Informação já é interdisciplinar mas é preciso ter trabalho em conjunto desses GTs em que a ANCIB é organizada.

Então é isso, não sei se falei muito rápida, se foi bom para vocês, estou aberta a perguntas e discussão.

Vídeo da apresentação

Título: A Ciência da Informação e os dados de pesquisa.



Disponível em: http://dadosabertos.info/enhanced_publications/idt/video.php?id=36