

A emergência dos dados de pesquisa na ciência contemporânea

The emergence of research data in contemporary science

Wagner Junqueira de Araújo^a, Ana Alice Baptista^b, Marcello Peixoto Bax^c, Pedro Luiz Pizzigatti Corrêa^d e Luís Fernando Sayão^e

Transcrição da apresentação

Abertura:

Vamos dar início a mesa-redonda intitulada “A emergência dos dados de pesquisa na ciência contemporânea”, para moderar gostaria de convidar o professor Dr. Wagner Junqueira, do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba. Convidamos a professora Dra. Ana Alice Batista, professor Marcello Bax da UFMG, professor Dr. Pedro Luiz Corrêa e professor Dr. Luís Fernando Sayão.

Wagner:

Boa noite a todos, é um imenso prazer estar aqui, agradecer o professor Guilherme e todos que estão organizando. Surgiu um assunto muito interessante e que a Ciência da Informação vem trabalhando já a alguns anos com isso daí, e o contexto da ciência no Brasil é uma coisa delicada com esse compartilhamento de dados que nós vimos durante o dia. Então eu vou fazer alguns questionamentos aqui na mesa, vou passar para cada um para comentar, porque nós vimos alguns desafios, como dos metadados descritivos, o professor Sayão fez uma exposição sobre os elementos que convergem na área de ciência, e Pedro expôs algumas dificuldades e desafios desse tema.

Mas aqui no Brasil a gente teve alguns processos que foram superados, um exemplo seriam os periódicos eletrônicos, a alguns anos atrás ninguém falava em publicar em periódico eletrônico porque não tinha estruturada. Hoje a quantidade de revistas são incontáveis. Por outro lado nós tivemos uma experiência muito ruim que foram os repositórios digitais das instituições de pesquisa das universidades e no final a gente acabou todo mundo nadando e morrendo na praia.

Aí minha pergunta é na seguinte linha, aqui no Brasil, como é que nós fazemos para gerenciar uma infraestrutura de dados, e não discorrer sobre os ciclos de dados que nós discutimos hoje, mas o que esperar das instituições de pesquisa, dos órgãos de pesquisa, de grupos de interesse comum, algum órgão de governo, a gente tem o arquivo nacional que o Sayão e conselheiros estão trabalhando nas questões de repositórios digitais, quem seriam esses órgãos?

-
- a Universidade Federal da Paraíba (UFPB). E-mail: wagnerjunqueira.araujo@gmail.com. ORCID: <http://orcid.org/0000-0002-2301-4996>. Currículo: <http://lattes.cnpq.br/6762905361803183>
 - b Universidade do Minho (Uminho). E-mail: analice@dsi.uminho.pt. ORCID: <https://orcid.org/0000-0003-3525-0619>. Currículo: <http://www.degois.pt/visualizador/curriculum.jsp?key=4103065722022437>
 - c Universidade Federal de Minas Gerais (UFMG). E-mail: bax@eb.ufmg.br. ORCID: <https://orcid.org/0000-0003-0503-3031>. Currículo: <http://lattes.cnpq.br/1864473087690223>
 - d Universidade de São Paulo (USP). E-mail: pedro.correa@usp.br. ORCID: <https://orcid.org/0000-0002-8743-4244>. Currículo: <http://lattes.cnpq.br/3640608958277159>
 - e Centro de Informações Nucleares (CIN). E-mail: lsayao@cnen.gov.br. Currículo: <http://lattes.cnpq.br/3422623122948389>

Porque esperava parte de nossa política que a gente tenha uma atividade dos profissionais da área que começam a trabalhar e percebem esse problema e desafio, então quem seria esse órgão gestor ou se teria algum órgão para trabalhar isso aí?

Mas a ideia é essa, nós teremos alguma organização para isso, alguma organização de pesquisadores no Brasil que criam estruturas, o Sayão citou o ResearchGate que além de rede social de pesquisadores serve como repositório, plataforma de serviço, como é que seria, eu vou passar para vocês. Passar primeiro para a professora Ana que tem uma visão de fora porque a gente aqui no Brasil fica um pouco angustiado com essa situação.

Ana Alice:

Eu não vou falar especificamente sobre a organização, a coisa que eu queria dizer é que eu gosto dos modelos brasileiros. Eu acho que a gente precisa atuar em vários momentos, agora precisa de alguém para administrar alguém, alguém fazendo as coisas andarem. Essa é minha primeira perspectiva, sem pensar muito vamos dizer assim.

Marcelo:

Eu acho que esse evento é fundamental para a Ciência da Informação, a gente precisa continuar e precisa ser estudado porque é uma grande oportunidade, acho que todos nós que estamos aqui enxergamos esse tema como uma grande oportunidade que a Ciência da Informação não pode perder.

Obviamente sem se descaracterizar, sem perder identidade como o professor Ricardo colocou, mas abraçando com toda legitimidade que a Ciência da Informação tem para lidar com esse problema dos dados. Diria até sobrevida das fases iniciais do ciclo de gestão dos dados, dados científicos, dados de negócio, principalmente dados científicos, que são aquelas fases de aquisição e preparação dos dados.

As palestras do professor Sayão, do professor Pedro, foram na direção de mostrar que há muita coisa a ser feita antes de que o dado possa ser analisado. A questão de métrica, de registro, proveniência, alinhamento semântica, a importância das ontologias nesse tema para integração e reuso dos dados científicos, para preservação dos dados científicos.

Essa pergunta em questão do Wagner foi interrogada por alguém da plateia, alguém que deve ser monitor de um repositório de alguma instituição de uma universidade, eu fiquei pensando nisso e a ideia aqui é que fazemos um bate-bola e vocês também participem, opinem, que a mesa seja bem aberta aqui para ouvir vocês. Então alguém perguntou isso e a minha opinião é que deve ser bottom up e top down, a gente tem que ter às sociedades científicas de alguma forma se organizando como a gente vê em países mais desenvolvidos, isso acontecendo massivamente. Acabo de retornar de um pós-doc nos EUA onde tem milhões e milhões de dólares investidos no desenvolvimento de repositório, sobretudo na saúde, porque às pesquisas na área de saúde são extremamente caras, então é muito importante utilizar os dados em questão econômica de racionalização de gastos é fundamental para a economia da pesquisa americana hoje. São milhões de dólares no apoio a construção desses repositórios.

Acho que nesse ponto nós temos que ter iniciativas de todos os lugares, eu acho que é legítimo que a área como de repente metodologia em uma universidade se valha de uma ferramenta de software livre e promova dentro da área o incentivo de compartilhamento de dados. Mas é importante também que a Reitoria da universidade estejam pensando no sentido de comentar, formar recursos humanos, eu acho que é uma oportunidade gigantesca de informação, formar bibliotecários de dados com todos os conhecimentos, e isto não está restrito a nenhuma disciplina específica. Lá na computação e administração me parece que é uma ação da Ciência da Informação, acho que mais uma vez legitimar deve surgir de vários locais.

Eventos como esse, eu acho que deve um futuro brilhante para a Ciência da Informação, fico contente em participar e estar aqui falando com vocês sobre um tema que é importante para a Ciência da Informação.

Pedro:

Mais ou menos na linha que foi colocada aqui, seguindo a questão de ser distribuída, orquestrada, a questão de que a gente tem hoje instituições preocupadas com o tema no Brasil, cada um tem sua responsabilidade. Acho que esse evento tem responsabilidade no sentido de criar demandas de pesquisa pro Brasil, começar a mostrar visões, eu acho que esse é o papel desse evento, e articular a comunidade da Ciência da Informação com a Computação e outras áreas, tem esse papel, e os órgãos de fomento eles têm um papel importante no sentido de valorizar a disponibilização e qualidade dos dados científicos, começar a reconhecer a questão da disponibilização dos dados para ofertar bolsas, na questão do reconhecimento da carreira do pesquisador, a gente precisa começar a colocar isso de maneira mais incisiva e cobrar isso dos órgãos de fomento.

As instituições de pesquisa de uma maneira geral, tem conversado muito com as entidades para que os seus pesquisadores passem a gerar dados, disponibilizar esses dados com qualidade, e tenham mecanismos de revisão, esse é um ponto importante, não significa apenas disponibilizar mas significa saber da qualidade daquilo que está sendo disponibilizado.

As universidades também tem que se preocupar com o tema, eu vejo por exemplo, na USP a gente conhece um programa que eu gosto muito porque ele não só cria na universidade baseado em modelos, mas disciplina e envolve a comunidade o tempo todo, a comunidade ativa com relação aquele assunto, define às boas práticas. Isso é fundamental para as instituições de pesquisa e universidades criarem. Repositório novo a gente não vai ter com um intuito só, nós vamos ter vários repositórios, o importante e aí vem a visão do profissional em Ciência da Informação de identificar junto com a demanda do seu usuário qual é a melhor solução.

O importante é você publicar esses dados e deixar disponível seus dados e seus metadados. Eu vejo assim, protagonismo da Ciência da Informação nessa área é fundamental na área de Ciência dos Dados que tem uma demanda enorme, a gente está vendo um levantamento de Recursos Humanos, só na Europa nos próximos anos a demanda de 100 mil profissionais na área. A média de salário dessa área gira em torno de 10 mil dólares por mês, então é um profissional valorizado.

O profissional da informação tem um papel fundamental na gestão dos dados científicos com a habilidade na área de negócio, essa gestão de dados acabam gerando conhecimento para outra, as áreas se conversam. Só para vocês terem uma ideia, nos projetos que a gente participa no pós-doc as pessoas da Ciência da Informação no nosso projeto de pesquisa acabaram sendo contratados no meio do projeto para fazer a gestão, e a gente tava perdendo gente do time para as empresas, porque é normal isso.

A valorização tanto do pesquisador quanto do profissional, o mercado vai colar rapidamente, eu vejo a questão do profissional da Ciência da Informação com o papel de protagonista nessa área, e a gente não pode perder essa oportunidade no sentido de gerar condições para que tudo isso aconteça, do ponto de vista da legislação, do ponto de vista da estrutura, do ponto de vista do fomento, do ponto de vista da formação de recursos humanos, a gente tem que olhar nossas responsabilidades para não perder essa onda.

Porque essa onda uma hora passa, nós já vimos várias ondas mais importantes passando na vida da gente, então essa onda vai passar também. Se a gente não for protagonista e assumir nós vamos perder essa oportunidade. A gente está com a faca e o queijo na mão como relação a esse tema.

Sayão:

Eu queria falar um pouco de algo que eu acho importante nos dados, especialmente esses dados da cauda longa, é em relação às bibliotecas de pesquisa ou especializadas, a investigação das literaturas têm mostrado que às bibliotecas de pesquisa estão em uma crise terrível. A gente observou em um artigo alguns anos atrás, será que o futuro da biblioteca de pesquisa não é o e-Science, e a gente fez um levantamento dos problemas e desafios. Um dos problemas é que acabou o relacionamento amoroso entre os usuários e sistemas de bibliotecas, e a biblioteca não é um lugar que o pesquisador procura para descobrir informação. Ele sempre procura na Internet e na Web e a biblioteca acaba sucumbindo e não acompanha esse desenvolvimento.

A crise da biblioteca de pesquisa não foi uma coisa local, é universal, às instituições que tem pós-graduação de alguma forma tem mas os pesquisadores não vão à biblioteca. E assim acaba que essa questão dos dados traz de volta o protagonismo dos objetos científicos, especialmente por não ser retrabalho nessa questão da cauda longa.

O que eu observo é que no mundo todo as bibliotecas de pesquisa começam a retomar esse papel importante em relação ao status. Isso se dá diante da seguinte situação, meu lugar, a gestão de dados, ela é pró-comunicação, sempre divulgou, tratou e são publicadas, só que agora a biblioteca tem que preocupar com a pré-comunicação, com o planejamento dos dados, com um plano de gestão de dados, tem que ajudar o pesquisador a fazer o plano de gestão de dados quando ele cria o projeto dele para o CNPq ou FAPESP.

Você tem que ter essa preocupação e a biblioteca passa a atuar junto, em uma fase anterior à publicação, e nessa situação ela vai ter que atuar muito próxima ao pesquisador, aí aquele clima amoroso em biblioteca, bibliotecário e usuário retoma. A gente teve essa experiência e impulsiona, a biblioteca é parte do processo dentro dos centros de pesquisa, a biblioteca oferece

cursos de gestão de dados, cursos para o pesquisador do bibliotecário de dados, e vai ser também uma referência.

O autor da referência muda nessa situação, então a gente tem uma aproximação do bibliotecário da biblioteca de pesquisa com o pesquisador, a gestão de dados dentro da biblioteca vai funcionar dessa forma, senão a gente volta a situação anterior desse distanciamento. Essa aproximação também se dá com o profissional da informática, então a gente tem uma nova situação em que essa aproximação é fundamental. Isso que tem que levar para dentro da biblioteca, a biblioteca como um apoio a gestão, como apoio a treinamento do pesquisador, como uma referência.

Essa iniciativa de gestão de dados ela renova a Ciência da Informação sem esquecer os pressupostos, todos os pressupostos que parte da Biblioteconomia, Arquivologia e Ciência da Informação, são tomados e são valorizados pelas áreas de informática e de pesquisa. Você tem uma renovação apoiada nos pressupostos básicos desta metodologia. Então nós temos duas coisas positivas que podem ser atendidas, é isso que eu tinha em relação a isso.

Wagner:

Obrigado Sayão, recentemente a gente fez um trabalho de gestão de dados dentro das bibliotecas e a gente entrou em contato o pessoal do STI, com o pessoal da reitoria e da biblioteca central, então na visão do bibliotecário eles não percebem que isso está mudando, na cabeça deles parece que vão trabalhar com livro pro resto da vida. Mas isso está mudando e o bibliotecário tem que se renovar.

Um outro questionamento para vocês que são questões que me angustiam aqui nessa universidade, como é que a gente busca financiamento para tirar essas ideias do papel? Através de parcerias ou financiamentos coletivos? Talvez esse seja o caminho que a gente consiga seguir, ou vocês que pensam em fazer um pós-doc fora, como é que a gente trabalha essa questão. Aqui no Brasil a gente vê que sem ter um órgão fomentador as coisas acontecem muito esparsa. Então qual seria o caminho para conseguir esses financiamentos e esses recursos, até mesmo como treinamento para o pessoal, não para comprar equipamento mas para preparar o pessoal, treinar essa nova geração de bibliotecários, arquivistas e cientistas da informação, cientistas de dados que estão sendo preparados para os próximos anos. Depois eu abro aqui para os participantes fazerem suas questões para mesa.

Marcelo:

Eu acho que uma questão bem pertinente essa questão de como é que a gente ensina esses agentes. Uma coisa é verificar os agentes envolvidos nesses processos, então olhar tudo. A gente tem o olhar dos financiadores, o olhar das agências de pesquisa, as escritoras, as empresas, existe um ecossistema aí.

Nesse ecossistema cada um tem muito interesse, primeiro uma coisa que a gente identificou é que o setor privado tem interesse nessa área. Por exemplo, as empresas de consultoria montam modelos e times de ciência dos dados hoje. O banco Itaú foi montar um setor de ciência dos dados, por exemplo.

Agora qual é o perfil do profissional que eu preciso? As empresas teriam colocado um profissional da Ciência da Informação porque é ele que vai articular e saber o fluxo de dados na empresa, como ocorre o acesso. Então a gente precisa começar a trabalhar mais com o mercado, acho que isso é um ponto importante. O mercado vai financiar isso porque ele tem interesse nisso, ele precisa desse skill, ele precisa ter essa habilidade.

O problema é os entraves que existem para as instituições e universidades, por exemplo, na USP a gente demorou para montar o projeto quase um ano, porque a coisa vem e volta, e a passa, em um dois anos as pessoas passam. A gente precisa melhorar um pouco os nossos mecanismos de trabalhar os recursos privados, mas existe e a gente precisa aproveitar essas oportunidades, a demanda é grande.

Em termos de instituições do governo eu não vejo muito a possibilidade de aumentar recursos que a gente tem, os recursos são escassos, o que o governo tem são as facilidades e esse é um ponto. Criar mecanismos que viabilizem esse tipo de coisa, da mesma forma que existem mecanismos de isenção de impostos, você criar ações para ter outras estratégias de financiamento.

A gente falou que existe esse interesse do setor privado, e um outro ponto que a gente identificou, é que os órgãos de financiamento de uma maneira geral, eles têm criados esses programas especiais. A gente na universidade fica muito aquém disso, por exemplo, o pessoal do ORCID, que desenvolve os verificadores de objeto hoje eles estão trabalhando e querem valorização, é importante no Brasil e qual é a colaboração do Brasil nisso, como a gente está colaborando de fato. Nós estamos participando das discussões?

E tem recursos internacionais envolvidos hoje, a gente precisa estar atento nisso. Existe disponibilidade de recursos por isso é importante a gente ter essa articulação internacional. Eu sei que a gente não está no mesmo nível que eles, a gente pode participar de projetos com ele para entender os problemas e começar a ajudar e contribuir nas soluções.

Em termos de recursos eu vejo essas duas oportunidades que estão disponíveis, é questão da gente olhar para essas oportunidades e começar a trabalhar de maneira mais colaborativa com o exterior e com o setor privado. E um ponto que eu queria falar mais é sobre como nossa comunidade colabora, eu mesmo acho que sou culpado disso, mas é a primeira vez que estou em um evento oficial da área de Ciência da Informação e eu estou sentindo falta de ter esses eventos e participar desses eventos. Eu preciso ter uma colaboração maior com nossa comunidade, uma maior articulação. Isso não é financiamento, mas é importante convencer os órgãos oficiais de ter programas nessa área.

Infelizmente a gente não pode ficar dependendo de um programa da FAPESP, quais são as opções do CNPq que existem nessa área para financiar recursos de projetos colaborativos, a gente precisa começar a olhar para isso e a gente como comunidade aqui pode estabelecer essas demandas, então eu vejo essas três oportunidades aqui, obrigado.

Sayão:

Eu queria colocar aqui alguns números importantes nessa questão de dados e financiamento. Nós temos uma situação muito complicada aqui no Brasil que é financiar os dados com verbas de projetos que incluam no máximo 3 anos. Uma coisa que é perene, por

exemplo na área pública, você tem o dinheiro aplicado a 3 anos. Na verdade você está dando dinheiro do projeto para poder criar estruturas de gestão de dados, isso é um complicador.

Mas tem uma coisa que é importante que um relatório que saiu o ano passado fala o seguinte, uma plataforma de dados ou um repositório de dados, custa quatro vezes mais caro que um repositório de publicações, então o custo para tratar os dados é mais caro. Outro ponto é que os custos da gestão de dados é 14% do custo do projeto, da verba para pesquisa. Outro número importante, o pesquisador para poder organizar, limpar, processar, retrabalhar os dados antes de publicar, segundo o pessoal do FAIR ele gasta metade do tempo fazendo isso.

Existe aí uma coisa super complicada, é que o pesquisador gasta muito tempo e muito recurso para poder organizar os dados e publicar. Só que ele não tem uma contrapartida em termos de recompensa, então ele prefere fazer o artigo e deixar os dados desorganizados para lá.

A gente precisa de uma política que considere essa organização dos dados com valor de uma forma que você no artigo ou revista de qualis A é importante, senão a gente não vai adiante nisso. Precisamos ter uma política ampla no nível elementar até as políticas comunitárias para poder ir adiante. Publicar é complicado, gerenciar dados é muito mais complicado, então é essa questão.

Marcello:

Só complementando o que o Sayão acabou de dizer, há uns dois anos atrás eu participei durante um ano e meio da COEP na UFMG, você sabe que toda instituição que realiza pesquisa tem que ter um comitê de ética, isso é uma resolução nacional e eu me lembro bem que nessas discussões do aspecto ético da pesquisa, acho que todos que coletam dados sobretudo dados relativos ao ser humano, não necessariamente da área de saúde mas desde que haja alguma relação ética entre o pesquisador, deve passar pelo comitê de ética. Eu quero dizer o seguinte, a comissão de ética uma das coisas que ela exige é o que o pesquisador pretende fazer com os dados, e isso a gente analisava lá durante um ano e meio participando, e todo pesquisador que submetia pra COEP mas todo pesquisador deve dizer o que ele vai fazer com os dados depois no final da pesquisa.

Sabendo o que ele vai fazer com os dados é importantíssimo para obter sucesso lá na coleta, então eu acho que eu vejo aí um potencial enorme de auxílio nas bibliotecas, assim como o pesquisador envia para COEP ele deveria chamar um bibliotecário e perguntar o que fazer com os dados, que ferramentas de software livre existe. A Ciência da Informação precisa formar esse sujeito para auxiliar os pesquisadores.

Só terminando eu acho que a gente está vivendo uma revolução na área de comunicação e informação, a gente está vendo o Washington Post foi salvo pela Amazon, que comprou o Washington Post e mudou o modelo de negócio e salvou o Washington Post.

A gente tem a Saraiva uma das maiores empresas de cultura de informação quebrando, a Saraiva pediu judicial por 350 milhões de reais, e a Saraiva tem mais de 100 anos. Será que a Ciência da Informação não está vendo isso, eu acho que não. Acho que eventos como esse que vão abrir os olhos.

Eu acho que as bibliotecas, o próprio jornalismo precisa usar as irmãs Ciência da Informação e a Comunicação, também há uma crise no jornalismo a gente viu recentemente, que são movimentos que estão por trás de uma revolução muito maior, que a Ciência da Informação precisa enxergar e as bibliotecas têm protagonismo e potencial gigantesco, mas é preciso formar as pessoas

Eu fico muito feliz que tem 21 anos que eu estou na Ciência da Informação, e eu vejo hoje depois de 21 anos meus alunos se submetendo a fazer BPM e ciência dos dados dentro das empresas mas com uma legitimidade interessantíssima, porque a legitimidade é marcada porque alguém da computação que veio para a Ciência da Informação diz que essas pessoas estão explorando os processos, a própria gestão de documentos eletrônicos, esses domínios, a gente vê os alunos se despontando.

O mercado está demandando mas o mercado não sabe muito bem o que demandar, a verdade é essa, então a reação natural do mercado é ir para a computação, mas não é na computação na verdade eles tem desenho por essa questão, não pela ciência dos dados mas por todas essas questões de organização, da ciência, é secundário. A Administração também não está se preocupando com isso, então é uma área que a Ciência da Informação precisa retornar, e eu acho que eventos como essa mais uma vez fazem isso, portanto tem-se a importância aí.

É explorar novas formas de fomento junto ao mercado, sem dúvida alguma. A CAPES está com um programa agora que acabou com esse negócio de bolsa de pós-doutorado, se quiser fazer parceria internacional vai seguir um edital. Na UFMG tem um movimento enorme lá da pró-reitoria de pós-graduação juntando os PPGs para juntos fazerem parcerias com as instituições internacionais. Precisa ter dinheiro para enviar pesquisadores, mas não é mais aquela coisa de balcão de pedir o pós-doc, porque o grande problema da CAPES é o seguinte, eu dou uma bolsa de pós-doc e o sujeito vai para o exterior, volta e nunca mais, não dá continuidade, por isso eles querem acabar com isso.

Portanto agora quem quer ir para o exterior vai ter que se articular com outros programas para ter parcerias externas e assim nesses contextos institucionais que são contextos que predisõem uma manutenção da evolução da parceria acontecer.

Eu acho que do ponto de vista do fomento, as coisas também estão mudando, e vão mudar drasticamente daqui para frente, e vamos ter que pensar em formas para conseguir isso.

Ana Alice:

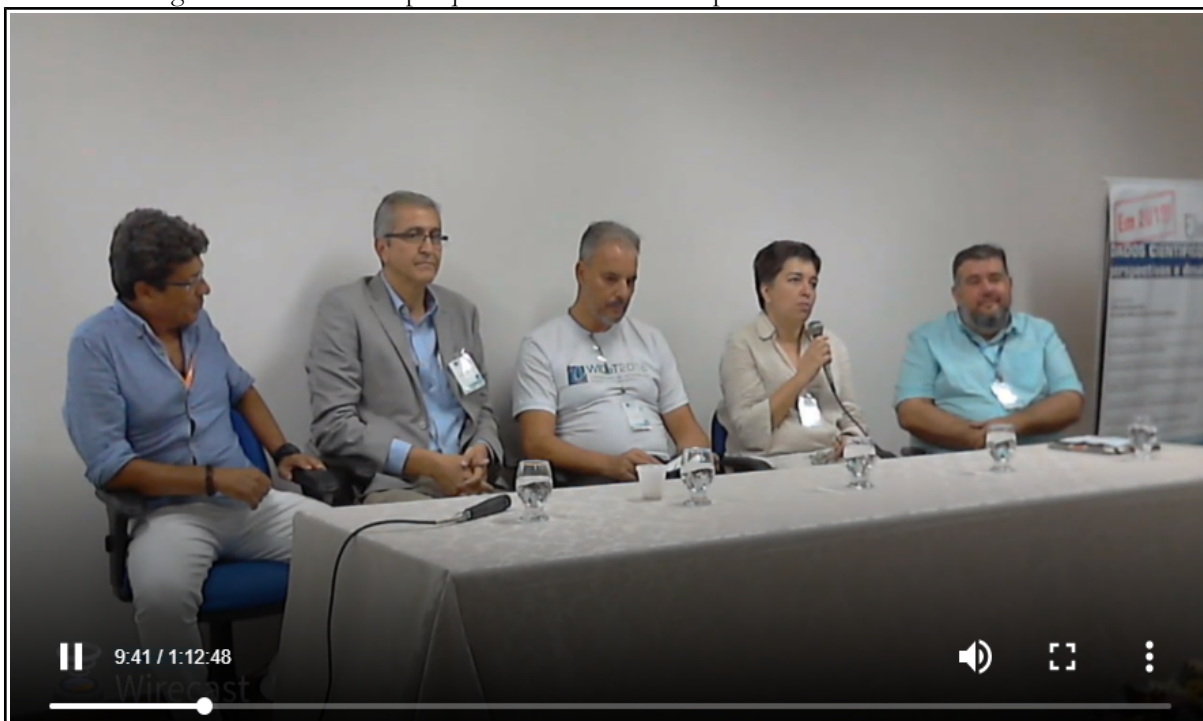
Queria dizer que eu concordo com tudo que foi dito, todas as formas de financiamento são interessantes, eu penso também que o Estado deve financiar pelo menos parte dos dados que são disponibilizados ou do processo de investigação, etc. Porque senão tem coisas que não vão ser disponibilizadas.

Wagner:

Obrigado a todos, eu passo então para plateia interagir.

Vídeo da apresentação

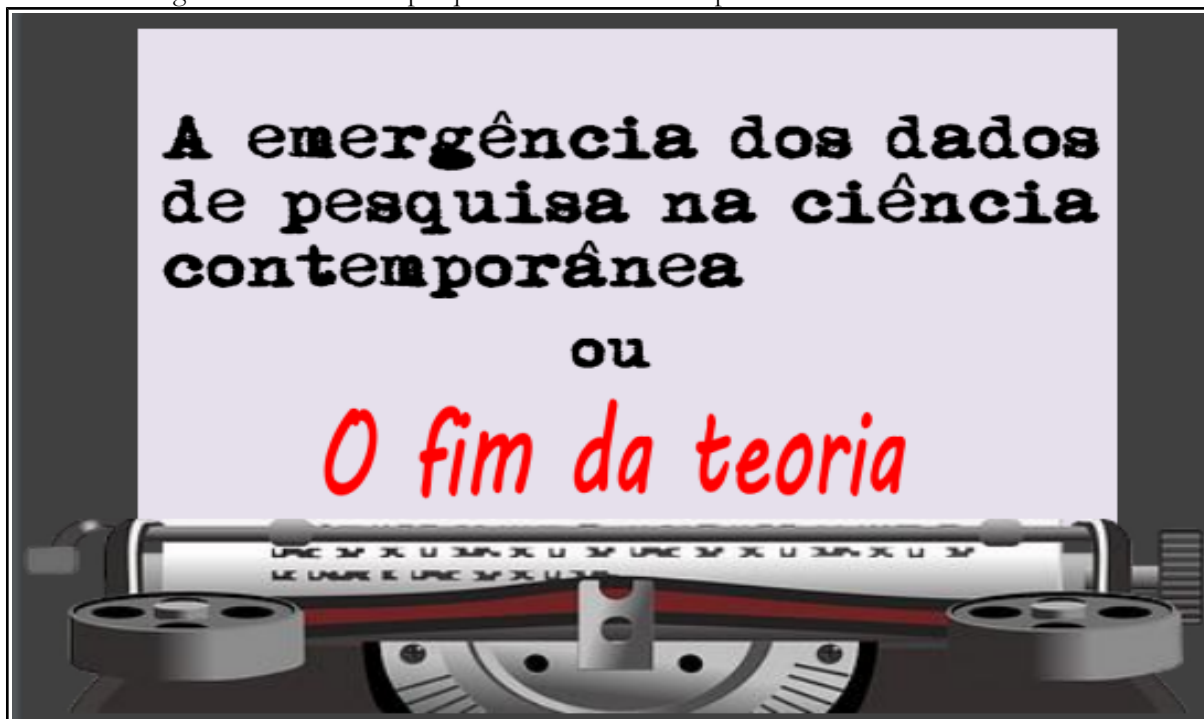
Título: A emergência dos dados de pesquisa na ciência contemporânea.



Disponível em: http://dadosabertos.info/enhanced_publications/idt/video.php?id=38

Slides da apresentação

Título: A emergência dos dados de pesquisa na ciência contemporânea.



Disponível em: http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=38