Infraestrutura brasileira para dados de pesquisa: reflexões

Brazilian infrastructure for research data: reflections

Pedro Luiz Pizzigatti Corrêa[a]

Transcrição da apresentação

Inicialmente eu gostaria de agradecer ao organizador do evento, professora Bernardina e professor Guilherme pelo convite, é um prazer enorme estar aqui com vocês e compartilhar um pouquinho do que a gente aprendeu e aprender também com vocês.

Eu sou mais da área de computação e da engenharia, o que envolve muito aplicação, mas a gente tem aprendido muito nesses últimos anos com a comunidade da Ciência da Informação. Então gostaria de agradecer essa oportunidade porque para mim já é uma consolidação desse aprendizado e dessa evolução que tivemos do ponto de vista do conhecimento.

Sou da escola Politécnica da USP, do departamento de Engenharia da Computação, que trabalha temas de desenvolvimento de aplicação da computação nas engenharias, a gente vem trabalhando com a comunidade científica no sentido de dar apoio, armazenamento, acesso, nos últimos 10 anos. Desenvolvimento de ferramentas e técnicas para aprendizagem, e com o tempo a gente foi entendendo que o problema era maior, não era simplesmente disponibilizar repositórios, ferramentas de análise, a gente viu que o problema era maior e que existe uma demanda enorme por recursos que permitam agregar esses dados, disponibilizar essas ferramentas para a comunidade. Existe um problema sério nessa área e que essa interação com a Ciência da Informação é a chave para resolver esses problemas.

O nosso grupo de pesquisa Big Data tem uma atuação nos últimos quatro anos trabalhando com projetos na área de e-Science, em conjunto com o Ministério do Meio Ambiente, ANEL e recentemente a gente disponibilizou um centro de dados com parceria do Itaú.

A ideia desse centro é disponibilizar bolsas de mestrado e doutorado para incentivar o ecossistema de Ciência de Dados em especial para pesquisa. Do ponto de vista de Pós-Graduação, a gente tem atuado no programa de Pós-Graduação em Engenharia da Computação, Engenharia Elétrica da Escola Politécnica da USP. Dentro desse programa a gente criou uma linha de ciência dos dados com o objetivo de tratar a questão de gestão de dados e análise dos dados. Também temos trabalhado com o pessoal da Universidade do Tennessee que também tinha acabado de criar um programa de Pós-Graduação com Mestrado e Doutorado em Ciência dos Dados, então a gente tem procurado estabelecer cooperação com as instituições no sentido de que a gente precisa aprender com os outros, é uma área que precisa de colaboração, então a ideia é que a gente possa estar estabelecendo colaborações para auxiliar na formação de recursos humanos.

Essa é a ideia de fazer essa colaboração com a Universidade do Tennessee e com um Laboratório de Energia do Governo Americano, é um laboratório que participou do Projeto

a 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

Manhattan na década de 40, eles desenvolveram toda a parte das pesquisas na área de energia nuclear e atualmente conta com 5.000 pesquisadores, e uma parte desse laboratório cuida da questão de pesquisas climáticas para entender o papel do clima para a energia.

Temos alguns cursos de extensão em Big Data voltado para o mercado, cursos de curta duração e cursos gratuitos voltado para a comunidade. A gente tem montado alguns cursos na área de gestão de dados e análise de dados.

Vou falar bastante da experiência que a gente teve com o RGs, Universidade do Tennessee e que já vem desenvolvendo a algum tempo, pelo menos nos últimos 20 anos, pesquisas na área de análise e gestão de dados científicos.

Temos feito Workshops com eles, a gente realizou três Workshops internacionais, na USP, na escola Politécnica, o último foi em 2017 com a professora Suzie Allard, que é da Ciência da Informação.

Na minha apresentação eu vou fazer uma rápida introdução sobre como a gente vê esse conceito de Ciência dos Dados, eu vou falar sobre gestão dos dados científicos, eu vou repetir um pouquinho pro Sayão algumas coisas que ele já colocou, inclusive é bom porque a gente aprende quando a gente repete. E eu vou falar um pouquinho na prática como esses conceitos estão presentes nesses dois grandes projetos, que são a cabeça como o DataONE.

Depois então a gente vai falar sobre dados, dados são fatos brutos, as informações são aqueles dados que são significados, e o conhecimento são as informações relacionadas e estruturadas.

Como a gente está falando sobre dados, ele acaba perdendo essas três áreas, dados normalmente são as observações, esses dados tratados e úteis também são passíveis de gerenciamento, e o conhecimento gerado também é passível de gestão.

A nossa definição de dados do ponto de vista científico é o resultado de uma pesquisa que não está publicado no texto dos artigos, teses, monografias ou dissertações. A gestão é uma iniciativa global que reconhece a importância dos dados como produtos que podem ser gerenciados.

Do ponto de vista de dados de pesquisa, a questão dos dados de pesquisa envolve coleções de registros que são utilizados pelos pesquisadores para realizar um registro de uma determinada evidência de sua pesquisa. De maneira geral tem alguns atributos, são digitais, seguem algum formato, do ponto de vista de conteúdo tem um contexto que é válido, tem um domínio de aplicação, e tem um valor.

Do ponto de vista de ciência, a gente tem uma variedade de dados de vários domínios de aplicação, desde a área e energia por exemplo, que temos dados que são observações e modelos também, na área biológica e ambiental, temos dados numéricos e tabulares, dados gerados através de sistemas de georreferenciamento de imagens, dados de gravações, então a gente tem uma variedade grande em cada um dos domínios de aplicações.

Quando a gente discute a questão da análise e gerenciamento dos dados, por que a gente chegou nisso? Na palestra anterior o Sayão comentou sobre um pesquisador na área da computação e banco de dados que publicou um livro chamado o quarto paradigma. O que seria o quarto paradigma da ciência?

Primeiro é preciso conhecer os paradigmas anteriores. Começou com a ciência empírica voltada para descrever fenômenos naturais, céus, estrelas, etc. A partir do século XV a ciência passou a ter uma abordagem teórica, representava fenômenos através de modelos. Nas últimas décadas, a partir de 40 e 50, a ciência foi auxiliada pela Computação, que passou criar modelos de simulação para descrever fenômenos, e isso foi um avanço do ponto de vista de novos conhecimentos. O quarto paradigma coloca o uso intensivo de dados e a simulação para gerar novos conhecimentos na ciência de maneira geral.

Ciência dos dados veio dessa ciência, hoje tem uso intensivo dos dados na ciência, que nasceu a partir dessa visão de que o conhecimento vão ser gerados a partir de simulações e experimentação baseado em dados.

Existe junto com essa visão a ideia de que a ciência seja convergente, a ciência é multidisciplinar, as novas ideias vem de campos diferentes a ideia é que você gere inovação e descoberta utilizando a colaboração multidisciplinar de várias áreas da ciência. É essa a abordagem que a ciência dos dados traz pra gente descobrir e gerar novos conhecimentos.

Em termos de ciência dos dados, tem algumas definições para entender como situa esse conceito. A ciência dos dados pode ter algumas definições, por exemplo, a ciência que estuda os dados científicos, dados de negócio, que une áreas do conhecimento como Estatística, Ciência da Informação e Tecnologia da Informação para resolver problemas por meio de extração de conhecimento a partir dos dados. Essa é uma definição que a gente entende objetos mas não entende processos.

Uma outra definição que nos dá a ideia dos processos é que envolve o processo de manipulação, tratamento, processamento, análise que visa a descoberta de novos conhecimentos.

Nessa visão de processo a gente pode entender a ciência dos dados como a possibilidade de estabelecer questionamentos gerais, adquire dados em diferentes fontes, esses dados de certa forma não estão organizados nem estruturados, a gente passa para uma etapa de organização e tratamento dos dados. Esses dados então estão prontos para alimentar um modelo, esse modelo gera um entendimento, esses dados entendidos podem ser visualizados, e com a visualização desses dados é possível analisar resultados e refinar o problema as hipóteses.

Esse é um ciclo de experimentação, que de maneira geral todo cientista passa, agora a gente tem que imaginar que em cada uma dessas etapas a gente está manipulando dados. Coletamos dados, colocamos nos programas, e a ideia é que a gente possa estar fazendo a gestão desse processo. Quando falamos de dados científicos estamos pensando em todas essas etapas aqui.

Do ponto de vista do gerenciamento, de como a gente trata esses dados, temos várias questões que podem ser feitas em cada uma das etapas, desde o planejamento, o que é um PGD, como eu garanto a qualidade, quais serão os metadados para descrever, como eu armazeno, como eu coloco os dados armazenados em outra plataforma, e, por fim, todos os mecanismos e estratégias de análise. Em tudo isso estaremos gerando dados e a ideia é que possamos fazer o gerenciamento de todas essas coisas.

Eu gosto de colocar esses slides aqui porque motiva, por que a gente aplica as técnicas de gestão de dados? por que hoje a gente está fazendo ciência dos dados? Tem a questão de proteção dos dados. Aqui tem um exemplo do Governo Americano que passou os dados de

segurança para o Governo Britânico, a gente teve aquele caso recente no governo da Dilma do pré-sal, que foram coletados e disponibilizados de acordo com o interesse econômico das empresas que iam para o leilão do pré-sal.

Existem questões de segurança e isso é gestão dos dados. A questão de reproduzir as análises é uma necessidade de manter os dados. Outra questão é a substituição se tiver problemas técnicos, a gente teve aqui recentemente o museu nacional do Rio, a gente perdeu várias coleções que provavelmente não vai recuperar, o próprio Butantã da USP perdeu os dados em 5 ou 6 anos atrás uma coleção enorme. Esse é o problema, não ter digitalizado e não fazer gestão desses dados.

Do ponto de vista não institucional, o pesquisador guarda em pendrive ou discos externos vários projetos como projetos de grande porte por exemplo LDA que está fazendo coleta de dados atmosférico da região amazônica, boa parte está na Alemanha, nos EUA, outra parte no pendrive dos pesquisadores, então não estão integrados, e vocês sabem dados climáticos sobre a região da amazônia são fundamentais para que a gente possa criar modelos e prever o futuro.

A questão da gestão dos dados acaba sendo uma disciplina fundamental, ele coloca questões de como eu defino, planejo, executo estratégias, procedimentos e práticas para gerenciar de forma efetiva recursos de dados e de informações dentro das organizações.

Uma questão que já comentaram aqui é a importância de que muitos dados acabam sendo perdidos, são dados em mídias que ficam defasadas, tem a questão de dados que são de pesquisadores, a gente teve contato com pesquisadores que disseram estar aqui toda a minha vida, trabalhei por 30 anos nessa organização guardado em uma caixa com todos os pendrives dos dados coletados, porque senão as pessoas se desligam das instituições e você vai perder esse conhecimento.

Conforme o Sayão comentou, a gente teria esse problema de que muita parte desses dados um dia vão ser órfãos ou serão perdidos, boa parte dos dados que são da ciência, da Big Science, são coletados aqui, por exemplo, os dados de energia nuclear boa parte estão aqui, são dados que você consegue capturar dos sensores e são simples de fazer gerenciamento.

Agora um outro desafio é o volume de dados, quais desses dados são úteis? Quais armazenar? A gente tem coletas que são locais, outras mais extensivas que envolve imagens de satélites, então isso tudo gera um volume de dados grandes. Quais são os dados significativos e que podem ser importantes para se manter a longo prazo?

Para a gente entender a importância de não perder esses dados, temos que imaginar que o conhecimento dos dados a partir do momento que vai passando o tempo a gente vai perdendo de maneira natural. A ideia de fazer gestão é que a gente impeça que haja perda de informação com o passar do tempo.

Aqui um exemplo, em 1889 teve um projeto que fez gestão e coleta de dados, que foi sistematizado e organizado. Hoje esse mesmo processo desde 1889 até hoje encontra-se no sistema deles. Agora a questão é o seguinte, a ciência de maneira geral foca na interpretação e nas conclusões que são sintetizadas, mas esses dados todos e esse legado estão em caixas desse tipo, essa é a ciência normal que é a que vem acontecendo nos últimos tempos.

Tenho alguns dados interessantes, por exemplo, a probabilidade de você encontrar a origem de dados ela decai 17% a cada ano em uma determinada publicação. Então uma publicação que disponibilizou os dados depois de um ano o site parou de funcionar ou o pesquisador não atualizou. Todo ano você vai perdendo 17% dos dados e a gente precisa mudar essa situação.

Quais são os princípios que direcionam as ações do ponto de vista da gestão de dados. Primeiro, os dados são ativos, incrivelmente valiosos, os dados coletados têm valor porque não tem como você fazer a leitura novamente desse mesmo dado.

A ideia é que a gente possa administrar e proteger os dados coletando de modo que eles estejam acessíveis e compreensíveis, reproduzíveis e utilizáveis. Esses são os princípios que regem a gestão dos dados.

O nosso objetivo é mudar o futuro dos cientistas, para que não seja só sintetizar, gerar conclusões e compartilhar as conclusões, mas você também compartilha os dados que são gerados no processo.

Um ponto importante para a gestão dos dados é a definição de uma política que estabelece todo o modelo e estratégia que as instituições podem utilizar, por exemplo, essas políticas podem ser explícitas, você escrever em um documento, ou implícitas para a comunidade que compartilha os dados, mas é importante ter essa política.

Se a gente pensar em uma instituição de pesquisa, qual é a estratégia dela fazer curadoria dos seus dados? Normalmente essa política estabelece os fundamentos para a gestão dos dados, define ferramentas, softwares, padrões para descrever dados, estabelece o processo para revisão, aprovação e publicação dos dados, e define quais os requisitos para preservar os dados com o tempo.

Por isso eu comentei com você sobre aquelas parcerias, para a gente entender um pouco como eles estão com 100 anos de dados científicos, a gente tem que ter essa experiência trabalhando com quem tem para entender quais são os problemas para que a gente possa avançar nesses problemas. A gente não pode querer refazer a roda, por isso é importante utilizar um pouco da experiência dessas outras instituições que já passaram por esse problema e já avançar mais que a gente.

Um atrativo pro partido de uma política, você faz o plano de acesso, de como será implementada essa política, a RGs definiu um plano de como ela implementou essas políticas, de quais são as estratégias que ela utilizou.

Imagina assim a RGs é uma instituição de pesquisa como outras, a maioria dos cientistas estão preocupadas com essas questões, não estão preocupados com metadados, estão preocupados em como fazer publicações, deadlines, concursos, laboratórios, equipes, congelamento de contratações, resistência de colaborações e prestações de conta, esse é o dia a dia do cientista.

Agora você chega para o cientista e fala que tem que fazer metadados e gestão dos dados, ele fala que isso não interessa e tem resistência, como a gente resolve isso? Claro que tem todo o trabalho de conscientização mas precisamos de ferramentas para ser fácil de usar senão eles não vão fazer.

No caso do RGs, dentro da política de dados existiam um ciclo de vida dos dados com enfoque desde o planejamento até a publicação, e com ferramentas que difundiam melhores práticas, estabelecem comunidades de usuários e ferramentas para apoiar cada uma das etapas, por exemplo, editor de metadados, ferramentas para citação, ferramentas para armazenamento, ferramenta para catálogo de dados, foram criadas para viabilizar a questão da gestão dos dados simples. Tem alguns detalhes interessantes, essas ferramentas são compartilhadas, então o vocabulário de dados aqui é utilizado para indexar os dados dali, é importante a integração dessas coisas por isso precisam ser baseadas em padrões para que você não precise reproduzir dados de uma etapa em outra etapa.

Uma outra questão para que a gente possa fazer proveniência e todo o processo em ciência dos dados, a gente precisa pensar na questão do software, ter informação sobre o software, porque se eu quero reproduzir os dados preciso saber como ele parametrizou aquele software, esse é um desafio do ponto de vista de pensar na questão de tornar os experimentos reprodutíveis.

Um software pode ser disponibilizada de uma maneira informação, você pode deixar o link ou de maneira forma, você coloca em ferramentas que você documenta todos objetos de software para representar e tornar bem documentado aquela ferramenta que foi utilizada. Também tem a questão de licenças, que pode ter domínio público e você pode deixar disponível de alguma maneira, inclusive disponibilizar informações de uso. Tem algumas sugestões do ponto de vista de obtenção de DOI para software para ser referenciado também, mas isso está sendo estruturado, as instituições estão se organizando.

Vou falar também sobre alguns projetos na prática como essas coisas acontecem. Aqui eu to falando mais sobre a cabeça e não da cauda dos dados, sobre quais são as estratégias que algumas dessas instituições e projetos têm adotado. Esse programa aqui foi onde fiz meu pós doc, é um programa que coleta dados atmosférico do mundo todo, para que você possa prever alguma mudança você capta dados do mundo todo para que criar modelos de como vai ser a evolução e fazer predição no futuro. Existem coletas permanentes nos EUA, e campanhas que são realizadas no mundo todo, no Brasil teve uma campanha em 2014-2015 na região amazônica para coletar dados dessa região e agora terão novas campanhas que serão feitas na Argentina e na região do Ártico, envolve sensores que são levados para esses lugares, têm aeronaves coletando na atmosfera, e satélites fazendo mapeamento da região. Então você acaba tendo uma boa coleta de dados atmosféricos de determinada região.

Esse projeto o foco é para a questão energia e fazer gestão dos recursos que são importantes para a humanidade. Ele definiu um ciclo de vida de dados, estratégias de qualidade, estratégias de monitoramento, e medidas para saber como está sendo sua evolução, quantos dados você disponibiliza? Quantas referências são feitas sobre seus dados? A gestão de dados passa por toda essas questões aqui, ciclo de vida, qualidade, métricas e monitoramento.

Esse projeto coleta atualmente 1.4 petabytes pode ano, a base de dados dele gira em torno de 9 petabytes, e isso é a tendência de ir crescente. Se a gente pensar onde tem essa experiência aqui no Brasil a gente tem pouca experiência para fazer gestão desse volume de dados aqui no Brasil, por isso a gente precisa ter essas parcerias para formar gente que entenda e saiba fazer a gestão

desses dados aqui, por isso a gente precisa de colaborações, essa é a ideia que eu comentei com vocês.

Como funciona esse ambiente? Ele guarda dados e tem que analisar, por exemplo, você não vai fazer download de 9 petabytes no notebook, então funciona assim. Você como a comunidade de usuário requisita parte dos dados, parte desses dados vão para um cluster conectados a essa estrutura, nesse cluster você roda suas análises, ou se precisar trabalhar com dataset menor você pode fazer download. Então veja, estas conectando dados de satélite, você está conectando dados do site, dados enviados por projetos que não necessariamente fazem parte dessas campanhas mas que tem interesse em disponibilizar dados aqui nessa estrutura, e a partir daí é disponibilizado tanto para análises locais em clusters quando fazer download para analisar seus dados.

Essa aqui é para vocês terem ideia como que é a interface de acesso a esses dados. Então aqui você tem uma unidade de um sensor, aqui informações de qualidade dos dados, você passa o mouse aqui e ele mostra se houve falha do sensor em coletar esses dados por exemplo, você seleciona quais dados você quer, colocar os dados em uma cestinha e depois faz escolhe o formato e faz download desses dados.

Em termos de plataforma e infraestrutura de software envolvida. Você tem infraestrutura de software desde o sistema operacional, por exemplo Linux, copiadores padrões, não tem nenhuma ferramenta proprietária, todas são de acesso livre, algumas ferramentas de desenvolvimento como Python, Matlab para controle de qualidade dos dados, ferramentas para distribuir dados no cluster como Cassandra e SPARQL, e ferramentas para você fazer a mescla, porque você não recebe diretamente esses dados, são ferramentas de comunicação via mensageria, que usam para poder executar a extração de dados.

Um outro projeto que eu queria comentar com vocês que passa por todas essas questões de gestão de dados e envolve uma comunidade científica maior, é o projeto DataONE. Uma coisa que é interessante é que a gente pode interagir em função das pessoas, eu passei um tempo nos EUA durante meu pós doc trabalhei com Mike Frame a RGs, boa parte da documentação que eu passei para vocês é da RGs. Aqui está a Suzie Allard uma das PIs desse projeto, as pessoas estão atuando o tempo todo em diversos projetos, essa é uma das chaves do trabalho aqui diferente do Brasil.

Aqui também está o Mike Frame que colabora com esse projeto que tenta agregar dados da natureza de maneira geral. O DataONE agrega dados de uma maneira geral e extensiva de diversas fontes, esses são pontos de observação de coleta de dados em diferentes posições geográficas, dados de observação de satélite, esse é o nível de dados que o DataONE agrega.

Ele trabalha hoje financiado pela CNs que completa 10 anos e tem três nós coordenadores, na Universidade do México e na Universidade da Califórnia. Esses catálogos têm por objetivo organizar os metadados, indexar buscas, serviços de rede para compartilhar, assegurar disponibilidade de conteúdo, e a ideia é que você possa preservar em três instituições diferentes.

Além dos nós coordenadores, têm os nós membros, no Brasil tem um que é o PPBio que coleta informações ecológicas, e uma série de instituições e universidades que são nós membros. Os nós membros gerenciam seus dados, fazem capacitações e treinamentos.

A ideia é que ele disponibiliza uma série de ferramentas para cada uma das etapas, no planejamento dos dados, geração de conteúdos, geração dos metadados, ferramentas que auxiliam na integração e ferramentas de análise também.

Do ponto de vista do profissional da informação, em cada uma dessas etapas ele tem alguns desafios. Por exemplo, nessas etapas iniciais quais são os desafios? É definir e auxiliar no plano de gestão de dados. Na etapa de gerar os metadados, qual é o desafio? É verificar qual é o padrão mais adequada para aquela comunidade. Na etapa de agregar os metadados você tem que conhecer algumas ferramentas e repositórios para disponibilizar os dados. E na etapa de análise é conhecer as ferramentas e estratégias de análise para auxiliar o pesquisador. O profissional da informação está envolvido desde o início, em todos os processos.

Aqui eu estou consolidando algumas lições aprendidas por essa equipe da RGs que é a equipe do DataONE o que eles consideram e a gente vem aprendendo com eles. Um ponto é considerar o ciclo de vida dos dados. O ciclo de vida dos dados é qual é o ciclo de vida mais adequado para minha instituição ou para o meu domínio.

Identificar e disponibilizar ferramentas fáceis de usar, de treinar e apoiar cada uma das etapas. A formação de recursos humanos de maneira interdisciplinar, para que possa formar as pessoas com seus novos gestores, analistas e cientistas dos dados. Envolver gente da Ciência da Informação, da Computação, todos trabalhando juntos.

Um outro ponto a questão de recursos específicos é importante porque preservar dados é caro, então tem que ter modelos de apoio específico, pessoas engajadas, que não necessariamente sejam técnicos, e especialistas, pesquisadores, que possam contribuir na geração desse modelo de gestão.

As políticas de dados que apoiam dentro da instituição e que também tem um sistema de valorização que reconheça o fato de que a comunicação daquele dado possa ser reconhecida de alguma maneira, por exemplo o DOI.

Campanhas de engajamento da comunidade é importante, hoje a USP está o tempo todo informando as boas práticas para fazer reciclagem, é a mesma coisa com dados de pesquisa.

Métricas, você está avaliando os resultados e impactos o tempo todo. Do ponto de vista de quantos downloads, quantos acessos, quem está usando, o perfil do uso. Tem muita coisa a ser feita ainda, do ponto de vista de promover os principios FAIR de Findable, Accessible, Integridible e Reutilizable. Deve-se garantir que todo objeto de dados possa ser encontrado com mecanismos de comunicação e isso não está resolvido atualmente. A questão dos dados e metadados acessíveis, as vezes só os metadados, se existem mecanismos para tornar acessíveis esses dados. A questão da interoperabilidade e semântica dos dados, e o fato dele serem usáveis, manter para que sejam úteis.

Estes são os desafios e existem muitas pesquisas em cima disso, por exemplo, para análise e visualização, formação de pessoas sob abordagem interdisciplinar, a questão do uso e reúso na formação, a gente sempre estar praticando essa questão de uso e reúso dos dados, precisa pôr a mão na massa.

E como estratégia existem algumas áreas que são emergentes, Física, Biologia, Química, mas Química um pouco mais atrasada, mas Biologia e Física são áreas estratégicas.

A questão dos avanços culturais e das mudanças, é fundamental mudar a cultura. Investimentos a longo prazo, investimento nessa área para sustentabilidade. E sempre estar promovendo e divulgando.

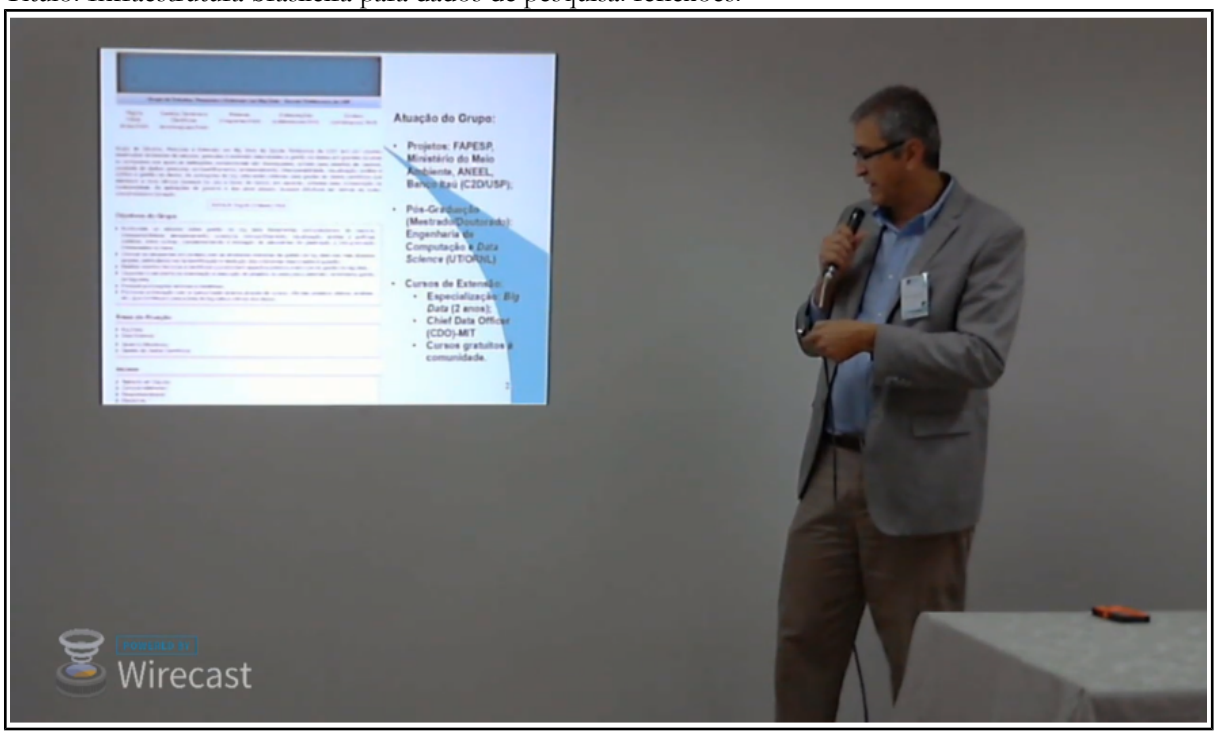
Para concluir, temos um gráfico em duas dimensões, que de um lado fala sobre o armazenamento e do outro do entendimento. Podemos ter armazenamento a curto prazo para o pesquisador fazer uma publicação, beleza resolveu o problema dele aqui. Agora se a gente quer de fato do ponto de vista do armazenamento a gente tem a curadoria ativa, para que esses dados possam servir para gerar novas intervenções na sociedade, ter um nível de maturidade de uso bem alto.

A nossa meta é chegar aqui, dados que tenham curadoria ativa, que de fato estejam armazenados e possam ser utilizados para apoiar tomada de decisão, iniciativa de políticas públicas, etc.

Eu vou deixar esses slides disponíveis para a organização porque tem uma série de links interessantes com relação à política de dados, melhores práticas de gestão de dados, é isso, obrigado.

Vídeo da apresentação

Título: Infraestrutura brasileira para dados de pesquisa: reflexões.



Fonte: http://dadosabertos.info/enhanced_publications/idt/video.php?id=31

Slides da apresentação

Título: Infraestrutura brasileira para dados de pesquisa: reflexões.

Disponível em: http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=31