

A gestão dos dados de pesquisa no contexto dos pesquisadores brasileiros na Ciência da Informação

The management of research data in the context of Information Science Brazilian researchers

Guilherme Ataíde Dias^a

Transcrição da apresentação

Muito obrigado, bom dia para todos, hoje nós vamos discutir aqui um assunto que eu acho muito importante para comunidade da ciência da informação. Então no terceiro e último dia do WIDAT, mostra que todas as palestras têm um eixo em comum, que é a questão da Ciência da Informação (CI) e da questão dos Dados. Nos dois primeiros dias não, nos temos um foco diferente, mas hoje é a Ciência da Informação e a questão dos Dados. Porque eu acho que isso é importante para a área? Historicamente a comunidade da CI foi responsável pela introdução de uso dela na área, visto os portais de periódicos.

Eles foram abordados por essa área e eu acho que no Brasil os periódicos abertos é um verdadeiro sucesso. Talvez seja um dos países que a ferramenta mais deu certo. É o uso do PKP para você disseminar periódicos.

A questão dos repositórios também passa pela Ciência da Informação. Eu acho que em escala menos. Os periódicos foram os mais populares, foi abraçado mais rapidamente. Tanto que eu acho que até hoje a questão dos repositórios ainda não está consolidada, mas isso passou pelos colegas da CI.

E um assunto que eu acho que precisa ser tocado é a questão dos dados de pesquisa, dados científicos. É um assunto que a área precisa investigar, precisa entender para que outras áreas se aproveitem disso. Todos nós temos condições de trabalhar com as questões que envolvem os dados científicos, e de forma mais notória a gestão desses dados.

Outra forma de dizer isso, eu acho que o tempo do profissional da informação, seja arquivista, bibliotecário ou outros, tenham condições de atuar em todas as caixinhas do ciclo de vida dos dados, todas as partes do fluxo, não importa qual ciclo de vida, nós temos como trabalhar.

Para eu entender o que a comunidade acha disso, nós fizemos uma pesquisa com os pesquisadores da área. Fomos entender o que os professores têm a oferecer para a comunidade, a gente não pode dizer que o aluno não sabe trabalhar, porque talvez tenha faltado o treinamento, então nós fomos entender o que os pesquisadores acham dessa área.

Esta palestra, os dados dela, já foram apresentados em outras palestras, mas eu só consegui compilar tudo agora. Nessa compilação de dados só foi finalizada agora, aqui tem 70% dos dados compilados.

^a Universidade Federal da Paraíba (UFPB). E-mail: guilhermeataide@ccsa.ufpb.br. ORCID: <https://orcid.org/0000-0001-6576-0017>. Currículo: <http://lattes.cnpq.br/9553707435669429>

Na nossa pesquisa, qual foi o universo: todos os pesquisadores vinculados a programas de Pós-Graduação em CI listados na plataforma Sucupira da CAPES. Muito provavelmente alguns de vocês aqui responderam a essa investigação.

Quero dizer que o questionário foi baseado no questionário do USGS. Aqui estão todos os programas da área que foram submetidos ao survey. Isso foi em 2017.

Foi um questionário enviado por e-mail para todos os pesquisadores listados nas páginas dos programas brasileiros de CI. O questionário foi construído e disseminado com o uso de uma ferramenta de software livre. O período de aplicação da coleta de dados foi entre 17 de julho a Agosto de 2017. Eu optei por utilizar uma ferramenta de software livre mas o Google Forms também faz isso, mas preferi ficar aderente ao software livre.

Foram enviados 341 questionários, obtivemos em todo de 190 questionários completos. Taxa de retorno 55,4% o que eu acho que foi excelente. Historicamente fica entre 20% a 15%. Então eu acho que a comunidade nos prestigiou com isso, mas também fizemos 3 chamadas para que pudessem enviar o questionário.

A faixa etária dos pesquisadores, a maioria é de 45 a 54 anos, mostrando que o pesquisador não tem uma faixa etária muito baixa não [risos].

Tempo de atuação na pós, a maioria de 10 a 19 anos, e 27,7% entre 5 e 9 anos. Isso é novo né, é gente nova.

A pergunta foi: realiza a gestão de dados científicos?

Antes não foi explicado o que é isso, mas na segunda etapa era perguntado se você queria continuar e alguns já saíam. O total de 188 vai caindo, a partir de algumas questões feitas esse número vai caindo.

Aparentemente tá todo mundo igual, nós temos 62,8% dos entrevistados fazendo gestão de dados científicos.

Realiza a gestão de dados científicos por faixa etária na pós-graduação?

Aqui vocês podem ver na tabela por tempo de pós-graduação quem faz a gestão de dados científicos por faixa etária.

Então, se você faz gestão ou não, eu perguntei por que você não faz? Surgiu a teoria do desconhecimento. A maior parte da resposta é desconhecimento por não fazer a gestão, consideram crescente mas ainda não tem tanto conhecimento para fazer. Falta de prática, já li a respeito, desconheço métodos adequados de armazenamento, foram algumas respostas que os pesquisadores foram dando, mas o motivo de desconhecimento foi o mais evidente.

Mais razões: desconhecimento de procedimentos sobre como gerenciar e armazenadas os dados científicos; ainda tenho pouca familiaridade com este procedimento; preciso conhecer estratégias necessárias para o preparo e disponibilização dos dados de pesquisa em repositórios; porque sou obrigado a fazer e de fato não sei nem começam; a única coisa que faço é guardar meus dados brutos de questionários e transcrições de entrevistas em pastas físicas e digitais.

Criei uma categoria “diversos”, para o motivo de não fazer gestão. O foco foi desconhecimento, depois diversos.

Faltam recursos que permitam acesso a dados, tempo é um dos principais dele. Concordo, porque tempo é complicado para o pesquisador.

Público o resultado da pesquisa, mas a gente não quer só os resultados a gente quer os dados também.

“Infelizmente ainda não faz parte das atividades do fazer científico”, pensamento de um colega nosso.

“Ainda não é tradição na área”, “guardei por vários anos em fita magnética os dados de pesquisa, entretanto nunca recebi uma solicitação para compartilhar dados”. Assim como o tempo esses dados o suporte de preservação ficaram defasados. Desinteresse, na categoria “diversos”, do por que não fazer.

Se você não faz gestão eu optei pela decisão de não continuar participando, as questões do questionário eram pertinentes a gestão de dados, então se você não faz gestão de dados por que continuar?

Você elabora um PGD?

Agora o PGD é um instrumento que está mais popular, principalmente lá em São Paulo agora porque a FAPESP se for um projeto temático ela exige. Geralmente o que começa na FAPESP dissemina pelas outras fundações de apoio científico no Brasil. Acho que o CNPQ também em breve deve estar pedindo. Então, 45% não elabora o PGD, 55% sim.

Vou adiantar aqui uma coisa que achei interessante. Às vezes você faz uma pergunta, o pesquisador diz que sim e aí o discurso não condiz com a resposta, mas todo assunto novo é alguma coisa.

Razões para não elaborar o Plano de Gestão de Dados?

“Não incorporei ainda esta prática, mesmo que considere importante”. Mais uma vez a teoria do desconhecimento foi a que prevaleceu.

“Falta de capacitação técnica para tal”, “precisa saber mais sobre o tema”, “falta de conhecimento a respeito do assunto”, desinteresse.

Você faz o uso de alguma ferramenta computacional para utilizar na elaboração do PGD? Sim com 55% e com 45% não.

“Por falta de software específico da faculdade”; “prazos e outras habilidades deixam o PGD em última prioridade sendo ao final não realizado”; “já possuo prática de procurar e curar os dados de pesquisa”; “já faço milhões de outras coisas absolutamente sozinho”, etc.

“A função do pesquisador não é gestão de dados”; “demanda uma política institucional com pessoas qualificados”, eu acho isso muito plausível. Todos os processos de gestão eu entendo que o pesquisador não tenha como fazer isso pela dificuldade.

Quem é essa pessoa qualificada para resolver a gestão? É aí que vocês profissionais da informação tem que entrar, não pode perder esse trem, de perder essa possibilidade de conhecimento, se vocês não forem atrás alguns irão, certo?

Você faz uso de metadados para descrever seus dados?

A maioria não faz, meio a meio quase. Vamos ver dos nossos colegas como está o uso de padrão de metadados. “Uso etiquetas com palavra-chave para catalogar a documentação física”; “também uso recurso para organização do dado no computador”; etc.

Lembrando que a pergunta era: Padrão de metadados utilizado na gestão de dados? Tiveram várias respostas mas ninguém respondeu qual padrão de metadados, apenas dois pesquisadores colocaram Dublin Core.

Estou disposto a guardar todos meus dados em um repositório central sem restrição de acesso? Então discordo fortemente e discordo é a maior parte, como vocês vêm aqui. A maioria dos colegas acha que todos, não a maioria, todos os dados não devem ser compartilhados. Eu mudei a pergunta para alguns dados, aí nesse caso ficou mais positivo.

Usaria conjunto de dados de outros pesquisadores se esses conjuntos de dados estivessem facilmente acessíveis? A maioria sim, entende que eu não compartilho mas eu usaria. Talvez a questão de compartilhar ganhe visibilidade quando os órgãos de fomento estimule isso. Mas em geral a visão da área é que todos os dados não devem ser compartilhados.

Os pesquisadores devem definir um período de embargo?

O período de embargo é o tempo em que você depositou em um repositório mas esses dados não estão disponíveis para todos, estão preservados e você pode controlar com o software do repositório, você faz preservação, a sua comunidade de pesquisadores pode acessar, como meus orientandos, mas para o mundo está fechado. Mas então por que é importante o embargo?

Você publica seu dataset aí alguém pode pegar esse dataset e utilizar para fazer descobertas antes que o pesquisador que coletou e processou, então iria se beneficiar desse dataset antes que quem criou faça isso, por isso a importância do embargo de dados.

Foi rápida apresentação mas basicamente é isso, não só a visibilidade mas quanto mais dados, eu acho que a nossa comunidade não está preparada para repassar isso para os discentes. Como que você começa com 188 respostas e cai para 66, e a partir do momento que você faz aquelas perguntas mais indiscretas como “Qual o padrão de metadados você usa”, as respostas não condizem com o que o pesquisador afirmou.

Vocês que são pesquisadores e trabalham com isso, você tem uma missão, a missão de levar isso aí, é orientar as pessoas acerca dessa temática. Então muito obrigado, espero ter sido claro, e obrigado pela presença.

Vídeo da apresentação


Título: A Gestão dos dados de pesquisa no contexto dos pesquisadores brasileiros na CI.



Disponível em: http://dadosabertos.info/enhanced_publications/idt/video.php?id=34

Slides da apresentação

Título: A Gestão dos dados de pesquisa no contexto dos pesquisadores brasileiros na CI.



WIDaT 2018
II WORKSHOP DE INFORMAÇÃO,
DADOS E TECNOLOGIA

**A Gestão dos dados de pesquisa
no contexto dos pesquisadores
brasileiros na área da Ciência da
Informação**

Guilherme Ataíde Dias
PPGCI-MPGOA/UFPB
PPGGOC/UFGM

Disponível em: http://dadosabertos.info/enhanced_publications/idt/presentation.php?id=34