



4º Plano de Ação do Brasil para Governo Aberto
Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

10ª REUNIÃO BIMESTRAL DO COMPROMISSO PELA CIÊNCIA ABERTA

Data: 24.04.2020

Horário: 9h às 11h

Local: Reunião Remota

Pauta: Relato das atividades realizadas pelo coordenador do compromisso e relato das atividades executadas pelos Marcos no bimestre (15/02/2020 a 24/04/2020).

Participantes:

Nome	Instituição/Setor	E-mail
Embrapa		
Patrícia Bertin	Supervisão de Governança da Informação e Transparência	patricia.bertin@embrapa.br
Juliana Meireles Fortaleza	Supervisão de Governança da Informação e Transparência	juliana.fortaleza@embrapa.br
Débora Pignatari Drucker	Embrapa Informática Agropecuária	debora.drucker@embrapa.br
Alessandra Silva	Secretaria-Geral	alessandra.silva@embrapa.br
Massayuki Okawachi	Supervisão de Governança da Informação e Transparência	massayuki.okawachi@embrapa.br
Lígia Sardinha Fortes	Supervisão de Governança da Informação e Transparência	ligia.sardinha@embrapa.br
Marcia de Oliveira Cardoso	Supervisão de Governança da Informação e Transparência	marcia.cardoso@embrapa.br
Isaque Vacari	Embrapa Informática Agropecuária	isaque.vacari@embrapa.br
Capes		
Tatiane Pacanaro Trinca	Coordenação de Programas, Cursos e Formação em Ensino a Distância (CPCF)/ Diretoria de Educação à Distância (DED)	tatiane.pacanaro@capex.gov.br
Katyusha Madureira Loures de Souza	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior	katyusha.souza@capex.gov.br
FIOCRUZ		
Josué Laguardia	Vice-Presidência de Educação, Informação e Comunicação (VPEIC)	josue.laguardia@fiocruz.br
Vanessa de Arruda Jorge	Vice-presidência de Educação, Informação e Comunicação (VPEIC)	vanessa.jorge@fiocruz.br
CNPq		
Adriana Maria Tonini	Diretoria de Engenharias, Ciências Exatas, Humanas e Sociais do CNPq	adriana.tonini@cnpq.br
Paulo Henrique de Assis Santana	Coordenação de Informação e Estudos Internacionais - COINF/CGCIN	paulo.assis@cnpq.br
MCTIC		



4º Plano de Ação do Brasil para Governo Aberto

Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

Luis Fernando Fauth	Assessoria da Secretaria-Executiva (SEEXEC)	luz.fauth@gmail.com
Ibict		
Bianca Amaro	Coordenação-Geral de Pesquisa de Manutenção de Produtos Consolidados (CGPC)	bianca@ibict.br
Tainá Batista de Assis	Coordenação do Laboratório de Metodologias de Trat. e Disseminação da Informação (COLAB)	taina@ibict.br
Vanderlino C. Barreto Neto	Coordenação do Laboratório de Metodologias de Trat. e Disseminação da Informação (COLAB)	vanderlinoneto@ibict.br
Lucas Paganine	Coordenação do Laboratório de Metodologias de Trat. e Disseminação da Informação (COLAB)	lucaspaganine@ibict.br
ABEC		
Sigmar de Mello Rode	Presidente	sigmarrode@uol.com.br
Eloísa Príncipe	Conselheira	eloisaprincipe@gmail.com
SciELO		
Solange Maria dos Santos	Coordenadoria de Produção e Publicação	solange.santos@scielo.org
Arquivo Nacional		
Luana Faria Sales	Coordenadoria-Geral de Acesso e Difusão de Acervos	luanafsales@gmail.com
RNP		
Gustavo Neves Dias	Gerência de Comunidades e Aplicações Avançadas (GCAA)	gustavo.dias@rnp.br
Leandro Neumann Ciuffo	Diretoria de Pesquisa e Desenvolvimento (DPD)	leandro.ciuffo@rnp.br
Carolina Howard Felicíssimo	Gerência de Comunidades e Aplicações Avançadas (GCAA)	carolina.felicissimo@rnp.br
Instituto de Pesquisa do Jardim Botânico do Rio de Janeiro		
Eduardo Dalcin	Coordenadoria do Núcleo de Computação Científica e Geoprocessamento	edalcin@jbrj.org
Outros		
Débora Rejane Fior Chadi*	USP	drfchadi@usp.br
Samile Vanz	UFRGS	samilevanz@terra.com.br
Sonia Caregnato	UFRGS	sonia.caregnato@ufrgs.br
Rene Faustino Gabriel Junior	UFRGS	rene.gabriel@ufrgs.br
Fernanda Scovino	GT da Sociedade Civil	fscovinom@gmail.com

2

Síntese dos assuntos discutidos:

Abertura da Reunião e relato das atividades do coordenador do Compromisso 3

A abertura da reunião foi feita pela coordenadora do Compromisso, a pesquisadora da Embrapa Patrícia Bertin. Ela apresentou a execução do compromisso, conforme informações do último Relatório de Execução. Atualmente, o Compromisso 3 está 78,27% de suas ações concluídas. De acordo com o andamento dos marcos, a



4º Plano de Ação do Brasil para Governo Aberto

Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

coordenadora acredita que o Compromisso 3 irá atingir 100% das suas dentro do prazo. Os Marcos 6 e 9, são os que estão com a porcentagem de conclusão mais baixa. O Marco 6 devido às mudanças político-institucionais, mas que o novo responsável pelo marco tem feito um esforço grande para compensar os atrasos e alcançar a proposta do marco, que é a promover a articulação dos editores científicos. O Marco 9, que começou em setembro de 2020 e que já adiantou bastante suas ações no último bimestre, conforme será apresentado nesta reunião. A proposta de data para próxima reunião do compromisso, é dia 30 de junho. A intenção é que a reunião ocorra presencialmente na Embrapa. Se não for possível, será feita virtualmente.

Assunto 1: Relato do Marco 4

Relatora: Vanessa Jorge, Fiocruz

Vanessa Jorge informou que Josué Laguardia irá representar a Fiocruz na RDA Brasil e, também, irá acompanhar e participar das ações no Compromisso, referente ao Marco 4, juntamente com a Vanessa. Josué Laguardia é o novo coordenador de Informação e Comunicação na vice-presidência da Fiocruz.

Avanços do último bimestre:

O marco não tem atrasos. Está tudo dentro do cronograma.

- i) Lançamento em 19 de fevereiro de 2020 do curso 2- Dados Abertos, da série 3 do Curso EAD de Ciência Aberta. Está com inscrições abertas.
- ii) 2º Encontro Nacional de Governo Aberto e Ciência Aberta para entregar até julho de 2020. Com a situação da pandemia, não poderá ser mais presencial, e será uma webinar com a temática compartilhamento e abertura de dados (governamentais e científicos) na pandemia do coronavírus. Essa webinar acontecerá em 18 de junho de 2020, às 14h, via webconferência RNP.
- iii) A tradução das cartilhas de Ciência Aberta para o inglês e o espanhol estão em andamento e serão entregues até julho de 2020.
- iv) O lançamento da cartilha sobre gestão de dados com foco no bibliotecário também está em andamento e será entregue até julho de 2020.
- v) O vídeo sobre Ciência Aberta está em produção. Encontra-se na fase de gravação dos vídeos do colaboradores.
- vi) Glossário da Ciência Aberta está em andamento. A ontologia foi finalizada e, no momento, as definições dos termos estão sendo redigidas, as quais serão enviadas aos especialistas para validação em maio de 2020.

3

Dificuldades/Causas/Possíveis soluções

Dificuldade: A realização do 2º Encontro Nacional de Governo Aberto e Ciência Aberta

Causas: Cancelamento de eventos durante a pandemia do novo coronavírus.



4º Plano de Ação do Brasil para Governo Aberto

Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

Solução: organizar evento on-line com nova configuração.

Percentual de conclusão: 78,4%

Discussões:

Adriana Tonini (CNPq) perguntou se os cursos EAD da Fiocruz estão disponíveis. A Vanessa esclareceu que sim. O objetivo inicial era encerrar as inscrições dos cursos da Série 1, que foram lançados em dezembro de 2018, para fosse feita uma avaliação. Mas, com a questão da pandemia, a Fiocruz decidiu deixar todos os cursos disponíveis durante o período de isolamento social. A instituição percebeu que nesse período, a procura pelos cursos está grande. A Adriana propôs fazer uma divulgação nos comitês assessores do CNPq sobre o curso EAD da Fiocruz, para que os pesquisadores divulguem em suas instituições de ensino superior e promova a participação de alunos de pós-graduação. Vanessa irá providenciar um folder de divulgação do curso para enviar para o CNPq.

Em relação aos vídeos, Juliana (Embrapa) esclareceu que, a princípio, as gravações seriam feitas no estúdio e pela equipe de comunicação do CNPq. Porém, por causa da necessidade de isolamento social, cada colaborador teve que gravar os seus vídeos. O prazo para envio dos vídeo era dia 25 de abril e foi prorrogado por mais sete dias. Alguns colaboradores já enviaram seus vídeos. Os vídeos estarão disponíveis para qualquer instituição que deseja usar em seus eventos técnico-científicos.

Bianca sugeriu que todas as instituições divulguem e disponibilizem esses vídeos em suas páginas. Gustavo (RNP) sugeriu que os vídeos fossem colocados na plataforma de vídeos da RNP de acesso aberto, direto e livre. Débora Drucker sugeriu divulgar os vídeos na RDA Brasil também.

4

Relato do Marco 6. Articulação com editores científicos para a implantação de ações em apoio à Ciência Aberta

Relator: Sigmar de Mello Rode, Abec

Avanços do último bimestre:

- i) Os parceiros têm implementado várias ações. O repositório de pre-prints piloto da SciELO em funcionamento. O primeiro manuscrito a ser depositado no servidor foi o relatório de vários pesquisadores sobre o Covid-19, em que um dos autores era o então Ministro Luiz Mandetta. Outros manuscritos foram depositados sobre Covid, os quais têm sido divulgados na mídia. Alguns já estão em processo de avaliação, mas que ainda não estão publicados. A SciELO tem feita grande divulgação desse repositório.
- ii) Outra ação que foi feita em parceria com o Ibict, foi o desenvolvimento da plataforma para pré-prints EmeRI (Emerging Research Information), que será lançada no início de maio e alimentada pelos editores dos periódicos científicos, para preprints do Covid. Bianca complementou que a ideia é



4º Plano de Ação do Brasil para Governo Aberto Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

ampliar o seu escopo de atuação. Ela destaca que a interação como parceiros dentro marco tem sido direta e efetivas, o marco 6 está deslançando para o bom caminho.

- iii) Os vídeos *homemade* sobre a Ciência Aberta estão sendo produzidos.
- iv) Algumas webinars sobre pre-prints serão realizadas em maio e junho, para o editores científicos: Webinar sobre preprints (em maio); Webinar sobre os requisitos mínimos para os periódicos científicos visando a ciência aberta (em junho); Webinars em colaboração com publishers (Clarivate, Elsevier, Springer, entre outros) a partir de maio.

Discussões:

Patrícia (Embrapa) destacou sobre a aproximação entre os parceiros dos Marcos que aconteceu nos últimos bimestres. Juliana (Embrapa) comentou sobre a visibilidade dos pre-prints na pandemia. Solange (SciELO) explicou que há um processo de controle de qualidade dos pre-prints para que o repositório não caia em descredibilidade. Qualquer pesquisador pode depositar no repositório, não apenas aqueles que irão publicar nos periódicos da SciELO, a diferença é que haverá um link entre o pre-print e o periódico.

Relato Marco 7. Implantação de infraestrutura federada piloto de repositórios de dados de pesquisa

5

Relator: Carolina Felicíssimo, RNP

Avanços do último bimestre:

- I. Início da PoC de preservação digital para repositórios de dados. A questão agora é quanto tempo os repositórios vão garantir o acesso aos identificadores persistentes de dados disponibilizados? Esse prazo deve ser definido e explicitado nos “Termos de Uso” do repositório considerando critérios como: disponibilidade de infraestrutura tecnológica (aplicação de armazenamento elástico, por exemplo) e recursos humanos (profissionais capacitados em Archivematica que conseguiriam recuperar os dados preservados em um futuro, daqui a 10 anos, por exemplo). O próximo passo é avaliar soluções tecnológicas específicas para preservação digital de dados avaliando-se critérios de repositórios confiáveis, como estudar microsistemas que garantam a preservação digital, inclusive no Dataverse, e extensões do Dataverse para outras soluções tecnológicas, como Archivematica. O que se deseja fornecer com essa avaliação é propor um conjunto de critérios norteadores para a decisão do investimento necessário para garantir a preservação digital de cada tipo de dado, num determinado período.
- II. O repositório piloto da Fiocruz finalizou sua fase de instalação e experimentação de funcionalidades básicas conforme planejada. A segunda fase consistirá na expansão dos testes de funcionalidades para compartilhamento e publicação de



4º Plano de Ação do Brasil para Governo Aberto

Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

dados de alguns pesquisadores pré-selecionado e convidados para participar do piloto. A RNP continuará acompanhando o processo, sempre que solicitada. Foi realizada uma webconferência, onde uma consultoria externa especializada em segurança, contratada pela Fiocruz, expôs vulnerabilidades do repositório piloto da época. A equipe do GIdLab da RNP, especializada em gestão de identidades, participou também da webconferência e concordou com os pontos levantados. Foi sugerido que tais testes de segurança sejam repetidos quando o repositório piloto da FIOCRUZ tiver a implementação do Shibboleth, evitando-se assim uma série de vulnerabilidades detectadas. Vale ressaltar que os pontos de segurança detectados nos testes do repositório piloto da FIOCRUZ valem para os demais repositórios que estão sendo criados e, por isso, sugere-se sua avaliação com equipe especializada.

- III. Avanços na autenticação Shibboleth do Dataverse. A equipe do GIdLab da RNP, especializada em gestão de identidades, tem feito uma série de atendimentos técnicos para suporte na implementação do Shibboleth no Dataverse. É um assunto que está bem dominado e encontra-se numa fase de categorização de atributos, requisito do Dataverse, que consiste, basicamente, no IdP (provedor de identidades) institucional liberar dois atributos a mais para garantir a autenticação via credenciais institucionais e, não, via criação de contas locais no Dataverse. O assunto é delicado por conta da entrada em vigor da LGPD. Nesse bimestre, a Embrapa solicitou um atendimento técnico especializado para avançar na implementação do Shibboleth em seu repositório piloto. Nessa reunião, percebeu-se que, apesar de ser a primeira vez que a equipe da Embrapa entrava em contato com a tecnologia, que a reunião fluiu bem, o que mostra que a RNP alcançou uma maturidade no processo, o que ajuda a repassar essa solução para quem precisa quando demandada.

Próximos passos:

Sobre o repositório da RNP, ele está sendo repassado para a equipe de infraestrutura da RNP, para garantir que o serviço seja mantido no ar.

Discussões:

Vanessa (Fiocruz) informou que já existem pesquisas avançadas de grupos de preservação digital sobre softwares de repositórios confiáveis. Para chegar na arquivemática já existe um conhecimento acumulado de avaliação de outros softwares para ajudar nesses estudos. Em relação ao relatório citado pela Carolina referente à primeira fase, é um relatório gerencial para entrega de resultado, num formato de lição aprendida para contribuir aos que irão começar nesse processo. Ela ressaltou que irá entrar na fase 2 do projeto, com o DOI, por meio do consórcio resultado do Marco 5.

Leandro Ciuffo destaca que o escopo alcançado pelo Marco 7 é muito maior do que o que planejado inicialmente. À medida que as ações do marco avançam, percebe-se que há sempre mais coisa a fazer. Na visão dele, esse marco já está 100% concluído, quando se



4º Plano de Ação do Brasil para Governo Aberto Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

compara com o que foi almejado no início. Na verdade, o resultado do marco é um alvo móvel, sempre necessário dar continuidade.

Débora Drucker (Embrapa) agradeceu o apoio da RNP na questão da gestão da identidade. Foi muito importante esse apoio, pois poupou muito esforço que seria empreendido pela equipe da Embrapa nessa questão. Ela finaliza reforçando a ideia do quanto essa parceria é importante.

Paulo Henrique (CNPq) informou que o consórcio é uma realidade e será assinado nos próximos dias.

Relato Marco 8. Proposição de padrões de interoperabilidade para repositório de dados de pesquisa

Relator: Washington Segundo, Ibict

Avanços do último bimestre:

Apesar de o marco já ter alcançado os requisitos gerais, chegou-se à conclusão sobre a necessidade de extensão desse conjunto básico para área específicas do conhecimento, como biologia, saúde, agricultura, energia nuclear e ciências naturais. Na área de saúde não foi possível avançar esse objetivo, devido a uma série de padrões diferentes o que torna difícil delimitar o que é saúde. Na Agricultura, já está bastante avançado, pois já existem padrões definidos para essa área. Na energia nuclear, também é complicado, pois envolve outras áreas como Física e Química e, por isso, decidiu-se trabalhar as Ciências Naturais, devido a existência de repositórios temáticos nessa área.

Foi feita uma comparação entre esses conjuntos básicos existentes na OpenAIRE, que também usa a descrição do DataCite, com o Fair Data Point (FDP), que é um conjunto que já tem olhar para interoperabilidade semântica e com o Data Documentation Initiative (DDI), que é um padrão usado mais nas áreas de humanas.

Nesse primeiro momento é preciso decidir quais os domínios atacar. Diante disso, está sendo trabalho com a Embrapa, pois ela possui um grupo de pesquisa que está trabalhando com análise de corpus para identificação de vocabulários e outras questões relacionadas à websemântica. Então, foi necessário coletar corpous para identificar quais as áreas mais tratadas na ciência em relação a dados de pesquisa. Primeiro foram coletados metadados de repositórios temáticos para construção de conjunto de corpus em determinadas áreas do conhecimento e extrair dados de projetos de pesquisa da Plataforma Carlos Chagas também para construção de corpus em diversas áreas do conhecimento.

Além disso, o marco tem fomentado a criação de uma comunidade brasileira de usuários do Dataverse, pois entende-se que quanto mais usuários e repositórios, mais discussões e mais chance de melhorar a questões da interoperabilidade para implementação de repositório e obter um feedback maior da comunidade brasileira. Foi criado um Fórum (<https://groups.google.com/d/forum/dataverse-brasil>) para usuários Dataverse, que já



4º Plano de Ação do Brasil para Governo Aberto Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

conta com mais de 200 participantes e foi construído o Manual colaborativo do Dataverse (<https://cienciaabertabrasil.github.io/dataversemanual/>) com instruções para instalação e configurações específicas do Dataverse. Pretende-se trabalhar mais nesse manual, reduzindo aquele manual PDF para essa versão web e depois convidar pessoas que já trabalham com o Dataverse para contribuir com as conversar e ter uma comunidade brasileira para discutir problemas e soluções sobre o Dataverse.

O acordo de cooperação técnica entre CNPq e Ibict para criação do repositório do CNPq, o Lattes Data, vem sendo trabalho, o que resultou na criação de um consórcio DataCite para atribuição de DOIs e para governança do DataCite. Haverá uma reunião sobre governança do DataCite em que o CNPq irá participar com direito a voto. Já há um entendimento interno para que outras instituições entrem nesse consórcio para ter um poder de direção de como atribuir o DOI nos repositórios que forem sendo criados.

O repositório do CNEN já foi instalado e está em validação da equipe do prof. Sayão e o do Ibict vai sair em sequência com o repositório Lattes Data.

Próximos passos:

No término do compromisso, será entregue um relatório que irá constar como é possível interoperar com o Oasisbr e alguns padrões específicos para algumas áreas de conhecimento consideradas prioridades.

Deseja-se que os repositórios institucionais surjam com mais facilidade. Esse trabalho com a Fiocruz é todo um trabalho de desbravamento para saber onde e como chegar em modo consistente.

O objetivo, portanto, é criar uma rede que interopere com o Oasisbr para que possa exportar o metadados que são coletados desses repositórios com bastante qualidade e, também, conseguir ter um portal onde seja possível recuperar esses metadados de forma simples dentro do Brasil e depois em nível regional, LA Referencia, e depois internacional, via OpenAIRE.

Discussões:

Juliana (Embrapa) lembrou que na reunião anterior, esse marco já tinha alcançado o resultado planejado, porém eles teriam um resultado a mais para entregar (um “plus”) e pediram prorrogação do prazo para junho de 2020.

Conclusão: 95%

Relato Marco 9. Proposição de conjunto de indicadores para aferição da maturidade em Ciência Aberta

Relator: Juliana Fortaleza, Embrapa

Avanços do último bimestre:



4º Plano de Ação do Brasil para Governo Aberto Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

- i) Construção da proposta do conjunto de indicadores.
- ii) Levantamento de fontes para construção desses indicadores. Essa etapa é constante.

Foi apresentada a metodologia para construção dessa proposta. Essa proposta é para aferição do grau de maturidade de abertura de dados científicos nas instituições de pesquisa, que são a geradoras dos dados. Todas as referências estão disponíveis na página do marco 9 da wiki da Ciência Aberta. A proposta é que a aferição do grau de maturidade seja fácil de ser obtida, de fácil resposta.

Os indicadores foram distribuídos em quatro eixos (governança, aspecto cultural, infraestrutura tecnológica e gestão de dados). Esse eixos estão alinhados aos eixos discutidos na primeira oficina de cocriação do compromisso, quando foi construído o cenário desejado. Cada indicador tem uma pontuação de 0, 1 e 2.

Foi apresentada a proposta que está disponível na página do marco 9 na wiki da Ciência Aberta.

Próximos passos:

A proposta está disponível, no formato aberto, para avaliação e sugestões de melhoria pelos parceiros do compromisso, que deverão ser enviadas, por e-mail, para juliana.fortaleza@embrapa.br, até o dia 31 de maio de 2020.

Discussões:

Vanessa (Fiocruz) relata que esse trabalho será muito importante para a sua instituição, pois irá ajudar na tomada de decisão do que é necessário alcançar.

Patrícia lembra que quando esse marco foi proposto, a expectativa era produzir algum tipo de instrumento que fosse útil para as organizações aplicarem e se autoavaliarem e assim conhecerem qual a aptidão e o preparo para implementar as ações de abertura de dados científicos. Dá um pouco mais de concretude, pois o que se vê hoje, especialmente, nos altos níveis hierárquicos declarações que não conferem com a realidade. Fornece um medida ao gestor.

Patrícia sugeriu que, posteriormente, esse instrumento fosse disponibilizado em uma ferramenta web.

Juliana (Embrapa) lembrou que a Tainá propôs a construção de indicadores, pós-OGP, para cada um dos pilares da Ciência Aberta. No compromisso, foi necessário fazer o recorte para o pilar de dados abertos, já que o compromisso era para mecanismos de governança de dados científicos.

Conclusão: 70%

Encaminhamentos:



Open
Government
Partnership

4º Plano de Ação do Brasil para Governo Aberto Compromisso 3: “Estabelecer mecanismos de governança de dados científicos para o avanço da Ciência Aberta no Brasil”

- Previsão da próxima reunião para o dia 30 de junho de 2020, na Embrapa, caso o isolamento social tenha relaxado.
- Parceiros avaliarem a proposta do conjunto de indicadores até o dia 31 de maio de 2020.

ra pesquisar