



CHAMADA ABERTA

01 de julho de 2021

A Rede Nacional de Ensino e Pesquisa (**RNP**), o Instituto Brasileiro de Informação em Ciência e Tecnologia (**IBICT**) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (**CNPq**) convidam Instituições de Ciência, Tecnologia e Inovação (ICTs¹) a colaborarem com projeto de Pesquisa e Desenvolvimento (P&D) que contempla ações de **transferência de conhecimento e Incubação de Repositórios de Dados de Pesquisa**.

Conceitos preliminares

Repositório de dados de pesquisa²: Estrutura tecnológica e gerencial que fornece a melhor opção para armazenar e publicar dados de pesquisa a longo prazo. Repositórios específicos podem ser recomendados por financiadores ou editores, enquanto alguns financiadores operam centros de dados para as pesquisas que financiam.

Dataset³: Qualquer coleção organizada de dados. “dataset” é um termo flexível e pode se referir a um banco de dados inteiro, uma planilha ou outro arquivo de dados, ou uma coleção relacionada de recursos de dados.

Identificador persistente⁴: Um identificador persistente (PID) é uma referência duradoura a um recurso. Esse recurso pode ser uma publicação, datasets ou pessoas.

¹ De acordo com o Marco Legal da Ciência, Tecnologia e Inovação (Lei nº 13.243/2016), as ICTs são órgãos ou entidades de administração pública ou privada, sem fins lucrativos, cuja principal missão institucional é desenvolver e executar atividades de pesquisa (básica ou aplicada) de caráter científico ou tecnológico. Dentre as quais, as universidades públicas.

² <https://www.sheffield.ac.uk/library/rdm/repositories>

³ <https://opendatahandbook.org/glossary/en/terms/dataset/>

⁴ <https://www.openaire.eu/what-is-a-persistent-identifier>

1. Introdução

Atualmente, a “Ciência Aberta” (um termo guarda-chuva que abrange as práticas de abertura de dados seguindo os princípios da transparência e colaboração na Ciência) é um tema estratégico para diversos países, que já a adotam como política pública para organizações de pesquisa, que gradativamente passam a incorporá-la como princípio orientador de suas atividades⁵.

Adicionalmente, comunidades de prática, instituições nacionais e regionais também passaram a identificar o valor estratégico da abertura de dados, fomentando o desenvolvimento de infraestruturas e tecnologias que estimulam sua disponibilização e facilitam o intercâmbio e a interoperabilidade entre os diferentes sistemas.

Um dos temas tratados na Ciência Aberta é o compartilhamento de dados de pesquisa, o qual já é estimulado por agências financiadoras, tais como o CNPq, que está fomentando a criação de um Consórcio Nacional para Ciência Aberta (CoNCienciA). Este pretende congrega instituições que desejam ter *Digital Object Identifiers* (DOIs) associados a seus conjuntos de dados (também chamados *datasets*), além de desenvolver seu próprio repositório de dados (LattesData) para que seus beneficiários depositem ali os dados de pesquisa associados a projetos financiados pelo CNPq.

Além de trazer diversos benefícios ao processo de comunicação científica, como maior celeridade, confiabilidade e redução de custos, os dados depositados em repositórios são citáveis, configurando-se então como produções científicas legítimas e reconhecidas. Desta forma, eles contribuem para aumento da visibilidade da produção científica nacional. Repositórios de dados de pesquisa são, como os demais repositórios, uma base de dados digital voltada para suprir a necessidade de armazenar, organizar e disponibilizar objetos digitais, neste caso, eles abrigam os dados coletados durante uma pesquisa científica.

A coordenação executiva desse edital contará com a participação de integrantes da RNP, IBICT e CNPq além do suporte consultivo do Grupo de Trabalho Rede de Dados de Pesquisa Brasileira (GT-RDP Brasil). Este grupo foi selecionado pela RNP, em parceria com IBICT, no atendimento a carta-convite RNP/IBICT de 01/2017. Sua atuação está direcionada à identificação de práticas de Acesso Aberto a Dados de Pesquisa (AADP) no Brasil e no desenvolvimento de experimentos tecnológicos, visando um futuro serviço nacional de AADP.

⁵ Exemplos: France’s National Plan for Open Science (<https://libereurope.eu/article/frenchopenscienceplan/>) e OECD Principles and Guidelines for Access to Research Data from Public Funding

1.1. A quem se destina este edital

Este edital destina-se à ICTs que já compreenderam os benefícios gerados pela abertura dos dados de pesquisa e que já tenham iniciado algum planejamento interno para a implantação de seu repositório institucional de dados de pesquisa.

1.2. Benefícios ao proponente

Este edital não prevê repasse de recursos, bolsas de pesquisa ou qualquer ressarcimento de custeio e/ou de capital às ICTs selecionadas. A RNP, IBICT e CNPq fornecerão apoio técnico, conhecimento e expertise adquiridos nos últimos anos no intuito de acelerar a implantação do repositório de dados de pesquisa da ICT.

Será fornecida infraestrutura de TI para hospedagem, nos 9 primeiros meses, para fins de incubação, observando as restrições descritas na Seção 10.

Também serão fornecidos identificadores DOIs, via consórcio estabelecido com a organização DataCite, limitados à 10.000 registros anuais, às ICTs que obtiverem êxito durante as atividades de incubação previstas neste edital.

2. Objetivos e resultados esperados

Os principais resultados esperados com a publicação deste edital são:

- 1) Selecionar até 3 ICTs interessadas em implantar repositórios de dados de pesquisa em âmbito institucional.
- 2) Apoiar as ICTs selecionadas na implantação de seus repositórios institucionais de dados de pesquisa por meio de ações de capacitação e transferência de conhecimento, com o compromisso das ICTs selecionadas de manter e operar os repositórios de dados gerados após um período de incubação⁶ de 9 meses.
- 3) Ampliação das instituições participantes do Consórcio Nacional para Ciência Aberta (CoNCienciA) com a inclusão das ICTs selecionadas, que poderão ter identificadores persistentes DOI associados a seus conjuntos de dados, com o custo inicial da emissão de DOIs patrocinado pelo CNPq.
- 4) Estruturação de uma Rede Federada Nacional de Repositórios de Dados de Pesquisa, com a participação dos repositórios de dados do IBICT, CNPq, RNP e de um conjunto de instituições de pesquisa convidadas, entre elas as ICTs selecionadas por meio deste edital.

Para o atingimento do objetivo deste edital, será necessário capacitar as equipes das ICTs selecionadas em diversos aspectos relacionados à gestão de repositórios de dados de pesquisa. Desta forma, espera-se a participação das ICTs selecionadas em atividades de

⁶ No contexto deste edital, definimos “incubação” como: “período durante o qual um acontecimento ou uma criação se elabora, à surdina, antes de assumir existência efetiva; preparação, elaboração.”

capacitação e transferência de conhecimento relacionadas à instalação, manutenção, operação, utilização e governança de repositórios de dados de pesquisa, a ser realizada em um período de 9 meses.

2.1. Objetivos secundários

- Apoiar instituições de ensino e pesquisa brasileiras na abertura de dados de pesquisa.
- Fomentar o debate acerca da importância do desenvolvimento de repositórios de dados de pesquisa em instituições brasileiras.
- Apoiar atividades de P&D e experimentação em repositório de dados de pesquisa a partir de estudos de caso com *datasets* de diferentes áreas do conhecimento.
- Apoiar ações da RNP no processo de descoberta e estruturação de serviços para a implantação de repositórios de dados de pesquisa.
- Apoiar ações do MCTI voltadas para a elaboração de uma política nacional para ciência aberta.

3. Datas importantes

- Publicação deste edital: **01/07/2021**
- Webconferência pública para tirar dúvidas sobre este edital: **13/07/2021**, quarta-feira, às 14:00 (horário de Brasília) no endereço: <https://conferenciaweb.rnp.br/webconf/dacca>
- Prazo para submissão das propostas: **01/08/2021**
- Divulgação dos resultados: até **13/08/2021**
- Reunião inaugural (kick-off) do projeto com as equipes das ICTs selecionadas: **24/08/2021**.

4. Período de execução

As atividades previstas terão duração aproximada de 12 meses (1 ano) a partir de agosto de 2021, sendo os 9 primeiros meses dedicados ao período de incubação (capacitação, transferência do conhecimento e testes com repositório de dados em ambiente de experimentação) e os 3 meses finais dedicados à migração do repositório para infraestrutura de TI própria da ICT selecionada.

Desta forma, somam-se às datas importantes:

- Previsão de fim do período de incubação e avaliação dos resultados: 31/05/2022.
- Período de suporte para migração do repositório para a administração da ICT (início da fase de pré-produção): de junho a agosto de 2022.

5. Incubação e Experimentação

As ICTs proponentes deverão indicar uma equipe encarregada de executar as atividades previstas no projeto. A equipe deverá ser composta por pelo menos duas pessoas, de acordo com os seguintes perfis:

- i. Pessoa que será responsável pela administração do repositório de dados e por liderar a sua oferta como um serviço para os pesquisadores da ICT após o período de incubação. Tipicamente, pessoa com formação em ciência da informação ou biblioteconomia. Essa pessoa será alvo de ações de transferência de conhecimento relacionadas ao processo de utilização e administração do repositório de dados, incluindo a representação de informações acerca dos dados (metadados) conforme suas comunidades produtoras, categorização, compartilhamento, curadoria e publicação de conjuntos de dados. Essa pessoa também deverá participar de atividades relacionadas à elaboração da política de uso do repositório.
- ii. Pessoa com perfil técnico, idealmente um membro da equipe de TI institucional da ICT. Essa pessoa será alvo de ações de transferência de conhecimento relacionadas à instalação, configuração, operação e manutenção evolutiva da ferramenta de software Dataverse [<https://dataverse.org/>].

A escolha do *software* Dataverse baseia-se, principalmente: (i) no estudo comparativo de soluções tecnológica para compartilhamento de dados realizado pelo GT-RDP e disponibilizado em: [<http://hdl.handle.net/10183/185126>]; (ii) na aceitação e direcionamento do IBICT pelo uso do *software* Dataverse; e, (iii) pela experiência adquirida pelas equipes da RNP e IBICT na instalação e operação do *software* Dataverse, que será usada para o suporte das atividades previstas do projeto.

Durante o período de incubação, a ICT selecionada também deverá fornecer conjuntos de dados para testes e estudos de caso, idealmente de domínios de conhecimento diferentes. Para isso, será exigido que cada proponente indique pelo menos 3 pesquisadores da ICT, que deverão colaborar com o projeto por meio do fornecimento de *datasets* para testes, e que participem do processo de validação da organização dos *datasets* no repositório e dos termos e condições assumidas pelo pesquisador e pela infraestrutura de testes fornecida.

5.1. Infraestrutura de apoio

Para execução das atividades de experimentação será fornecida infraestrutura computacional com o *software* Dataverse pré-instalado pela RNP.

Caso a ICT selecionada queira instalar o *software* Dataverse em infraestrutura própria desde o início do projeto, todo o suporte de consultoria será dado para execução das atividades do proponente. Espera-se que os conjuntos de dados de todos os estudos de caso de cada proponente (*datasets* de teste) não ultrapassem um total de 750 GB de armazenamento.

Na fase de migração, após os 9 meses iniciais, serão fornecidos identificadores persistentes para os *datasets* do tipo DOI, via consórcio com a DataCite [<https://datacite.org/doi.html>], de forma a permitir a publicação de dados pelo Dataverse. Os identificadores persistentes atribuídos manterão seus registros armazenados pelo período de **dez anos após a vigência deste edital**. A emissão de DOIs fornecidos pelo CNPq será limitada em 10.000 (dez mil) registros por ano. Esse limite de registros deverá ser respeitado sob pena de suspensão imediata do fornecimento de novos DOIs.

6. Contrapartida esperada das ICTs selecionadas

As ICTs selecionadas deverão:

- Alocar equipe (pelo menos 2 pessoas, de acordo com os perfis descritos na seção 5) para:
 - Participar das reuniões mensais de acompanhamento do trabalho de cada ICT (carga horária estimada em 2h por mês);
 - Participar das atividades de capacitação e transferência de conhecimento organizada pela RNP e IBICT (carga horária média estimada em 2h por semana);

Após o fim do período de incubação (9 meses após o início do projeto), será realizada uma avaliação dos resultados e será facultado à ICT escolher se deseja iniciar a migração do repositório para a fase de pré-produção. Nesse caso, a ICT deverá:

- Fornecer infraestrutura de TI própria (“on premises” ou virtualizada em nuvem pública) após 9 meses de execução do projeto, para a migração do repositório de dados de pesquisa.
- Comprometer-se em manter os conjuntos de dados identificados por meio de DOIs de produção fornecidos pelo CNPq por meio do consórcio CoNCiência pelo período mínimo de 10 anos.
 - Na eventualidade de descontinuidade e desligamento do repositório de dados da ICT, esta deverá comunicar essa decisão ao CNPq, com antecedência mínima de 90 dias. Além disso, a ICT deverá migrar seus conjuntos de dados para o repositório LattesData.

A formalização do compromisso da ICT será efetuada por meio de um instrumento jurídico específico, como um Termo de Cooperação ou um Memorando de Entendimento (MoU), a ser discutido entre as partes após o nono mês.

7. Apresentação das propostas

A submissão das propostas deverá ser realizada por meio do preenchimento e envio, por parte da ICT proponente, do formulário eletrônico disponível no endereço: <https://forms.office.com/r/ewWU5fKv6Z>

Serão consideradas as propostas enviadas até às 23:59 do dia 01/08/2021, no fuso horário GMT-5.

Além do preenchimento do formulário eletrônico, cada proponente deverá apresentar:

- Anuência do gestor de TI da instituição em apoiar as atividades de migração do repositório de dados para a infraestrutura de TI gerida pela ICT após o período de incubação. Os requisitos técnicos mínimos necessários são apresentados no Anexo I.
- Manifestação da alta gestão da instituição (reitor, pró-reitor ou chefe de unidade) em apoio à proposta.

Tais manifestações, do gestor de TI e da alta gestão, deverão ser enviadas via e-mail para o endereço editaldadosdepesquisa@rnp.br até o dia **02/08/2021**.

O envio poderá ser feito em arquivo PDF com a assinatura do declarante ou por de meio de uma mensagem escrita e enviada por e-mail pelo próprio declarante. Recomenda-se a utilização do modelo de “carta de suporte” disponibilizado no Anexo II.

8. Critérios de avaliação

O processo de avaliação e seleção das propostas será realizado em etapa única, abrangendo a verificação dos seguintes requisitos:

- Atendimento dos critérios de elegibilidade estabelecidos por este edital (vide seção 9).
- Grau de maturidade da ICT proponente no processo de abertura dos dados científicos da instituição. Essa avaliação será baseada em um conjunto de indicadores propostos para essa aferição, resultado do trabalho do Marco 9, do Compromisso 3 do 4º Plano de Ação Nacional em Governo Aberto⁷;
- Diversidade de áreas de conhecimento dos conjuntos de dados (*datasets*) que serão fornecidos para testes durante o período de incubação do repositório.

As propostas serão avaliadas por um Comitê de Avaliação, composto por um painel de especialistas indicados pela RNP, CNPq e IBICT, que apresentará como resultado uma lista ranqueada das propostas. Caso haja mais de 3 propostas elegíveis e selecionadas de acordo com os critérios de avaliação descritos acima, o Comitê de Avaliação poderá adotar os seguintes critérios de priorização:

1. É desejável que as instituições selecionadas possuam experiência prévia com a operação de repositórios institucionais (de publicações científicas).
2. É desejável que as instituições selecionadas sejam sediadas em diferentes regiões geográficas do Brasil.

⁷ <https://wiki.rnp.br/pages/viewpage.action?pageId=108957292>

9. Critérios de Elegibilidade

Cada proposta deverá:

- Ter como proponente e coordenador da proposta um funcionário/servidor da ICT proponente.
- Indicar equipe executora do projeto, formada por pelo menos duas pessoas, de acordo com os perfis descritos na seção 5. Em resumo:
 - Pessoa responsável pela gestão do repositório de dados de pesquisa.
 - Pessoa com perfil técnico para receber capacitações sobre instalação, manutenção e operação do software Dataverse;
- Indicar pelo menos 3 grupos de pesquisa ou pesquisadores individuais da instituição, que colaborarão com o projeto por meio do fornecimento de *datasets* para testes durante o período de incubação.
- Apresentar anuência do gestor de TI da instituição.
- Apresentar manifestação da alta gestão da instituição em apoio à proposta.
- Declarar a inexistência de repositório de dados de pesquisa implantado em âmbito institucional, mesmo em fase de pré-produção.

A ICT proponente deverá fazer parte do Sistema RNP⁸

10. Restrições e Escopo

Ao submeter proposta em atendimento a este edital, a ICT proponente, caso selecionada, deverá estar de acordo e aceitar as seguintes restrições:

- A infraestrutura de TI fornecida para a incubação dos repositórios de dados de pesquisa (incluindo atividades de teste e capacitação), é uma infraestrutura dedicada às atividades de pesquisa, desenvolvimento e experimentação, sujeita a falhas de hardware e interrupção de serviços sem aviso prévio. Suporte técnico poderá ser fornecido pela equipe técnica da RNP em horário comercial e em dias úteis. Portanto, em hipótese alguma a ICT proponente deverá considerar o repositório de dados a ser criado no âmbito do projeto como uma infraestrutura de pesquisa em produção.
- Em função de limitações de tempo e escopo, apenas a ferramenta de software Dataverse será suportada, mesmo que a ICT selecionada possua larga experiência na instalação e manutenção de outras ferramentas, como por exemplo o software DSpace.
- A URL dos repositórios de dados a serem incubados deverá ser definida pela instituição proponente com a anuência do gestor de TI da instituição, que por sua

⁸ <https://www.rnp.br/sistema-rnp>

vez deverá proceder à configuração do registro de domínio no serviço DNS apontando para os endereços IPs fornecidos pela RNP (ex.: *repositorio.ICT.edu.br*).

- Após o nono (9º) mês de execução do projeto, será realizada uma avaliação parcial dos resultados alcançados. Se bem avaliada, a ICT poderá optar por iniciar o processo de migração e pré-produção do repositório de dados em infraestrutura de TI própria.

11. Coordenação do projeto e Informações Públicas

As atividades de coordenação do projeto e acompanhamento do trabalho das equipes de cada ICT serão conduzidas pela RNP. Informações classificadas como públicas poderão ser utilizadas em ações de disseminação da RNP, CNPq e/ou IBICT, podendo também ser utilizadas pelas assessorias de imprensa das ICTs selecionadas. Entre as informações públicas, temos:

- Produtos resultantes do projeto, tais como: manuais, relatórios, *templates*, vídeo-aulas ou componentes de software (APIs, plug-ins, scripts etc).
- Apresentações, artigos e demais bibliografias que sejam geradas a partir dos resultados do projeto, durante a vigência do projeto, e
- Atividades realizadas em workshops abertos ou em eventos organizados pela RNP, IBICT, CNPq ou MCTI.

Entre as informações restritas, temos:

- Mensagens trocadas por meio das listas de discussão do projeto;
- Produtos e documentos ainda em versões preliminares e não finalizados;
- Reuniões de gestão e acompanhamento das atividades, a serem realizadas por meio de webconferência, e que eventualmente poderão ser gravadas e mantidas pelo período máximo de 3 meses.

12. Viagens

Este edital não prevê recursos para o custeio de viagens de membros da equipe do projeto. Para a gestão e acompanhamento das atividades serão utilizadas ferramentas colaborativas, como o serviço Conferência Web da RNP [<http://conferenciaweb.rnp.br>]

Entretanto, ao longo do desenvolvimento do projeto, e a depender do arrefecimento das restrições de mobilidade impostas pela COVID-19, a RNP poderá convidar, a seu critério e arcando com todos os custos, a participação de membros da equipe em reuniões específicas ou em eventos organizados pela RNP.

13. *Dúvidas*

Dúvidas poderão ser sanadas na Webconferência pública para tirar dúvidas sobre este edital (vide seção 3) ou poderão ser enviadas para o e-mail: editaldadosdepesquisa@rnp.br

=====

ANEXO I – Requisitos técnicos mínimos para a instalação do software Dataverse e expectativa de capacidade de armazenamento

É desejável como requisitos técnicos mínimos para a instalação do Dataverse: um servidor (ou máquina virtual) com 2 processadores de 2.8 GHz , 8 GB de RAM e 50 GB de armazenamento em disco.

Mais detalhes sobre os pré-requisitos para instalação do Dataverse podem ser consultados em: <https://guides.dataverse.org/en/latest/installation/prep.html>.

No escopo das atividades de incubação previstas nesse edital, o espaço em disco dedicado para o armazenamento dos conjuntos de dados (*datasets*) será limitado a 750GB. Desta forma, para a fase de migração, recomenda-se que o espaço a ser disponibilizado para o armazenamento de *datasets* seja superior a 750GB.

ANEXO II – Modelo de texto para a carta de suporte

Estou de acordo com a proposta submetida por <nome do proponente>, <cargo do proponente e vínculo com a instituição>, referente à carta convite que objetiva a promoção de ações de transferência de conhecimento e Incubação de Repositórios de Dados de Pesquisa.

O/A <nome da ICT> compromete-se em apoiar as atividades do projeto a fim de acelerar o processo de implantação de seu repositório institucional de dados de pesquisa.

<local e data>

<nome do declarante>

<cargo do declarante: | Reitor, Pró-reitor ou chefe de unidade| >