

RBPDRevista Brasileira de
Preservação Digital**RBPD**Brazilian Journal of
Digital Preservation

doi: 10.20396/rbpd.v1i00.14881



Entrevista¹ com o pesquisador Pascal Aventurier² sobre o processo de preservação digital no Institut de Recherche pour le Développement (IRD)

² Institut de Recherche pour le Développement (França)

<https://orcid.org/0000-0003-0211-4549> 

E-mail: pascal.aventurier@ird.fr – Montpellier – França

1

Entrevista em: 16/12/2020

Publicado em: 23/12/2020



¹ Entrevista concedida à Revista Brasileira de Preservação Digital por meio do editor Gilденir Carolino Santos, também bibliotecário e pesquisador do Grupo de Pesquisa e Estudos sobre Preservação Digital – Dríade, via e-mail em dezembro de 2020.

Interview with researcher Pascal Aventurier on the process of digital preservation in Institut de Recherche pour le Développement (IRD)

RBPD [Gildenir]: Poderia nos falar como se processou a preservação digital dos dados de pesquisa do Institut de Recherche pour le Développement (IRD) em que trabalha?

Pascal Aventurier: No Institut de Recherche pour le Développement (IRD), as pesquisas estão focadas nas relações entre o homem e o seu ambiente, na perspectiva de um desenvolvimento sustentável para informação, o IRD trabalha em parceria com instituições brasileiras há mais de 60 anos e tem uma representação em Brasília. É um pequeno instituto² com 800 pesquisadores e 2.400 empregados. Nós não trabalhamos nesse momento sobre preservação de pesquisa de longo prazo. Estamos muito ativos nas etapas anteriores que são muito importantes. Como verificar que os dados estão seguindo os princípios FAIR, é muito importante verificar os formatos de dados em formato aberto e os documentos para usá-los. Os documentos que detalham como reutilizar os dados estão disponíveis no mesmo local que os dados, mas também a descrição dos dados (metadados) para a citação que facilitam a descoberta. Por fazer isso estamos incentivando os pesquisadores a depositar os dados dentro do nosso repositório de dados DataSuds que também assegura a preservação de curto meio prazo (10 anos). A preservação de longo prazo será contratada com um centro nacional: O CINES³.

2

RBPD [Gildenir]: Você e sua equipe foram capacitados para realizar a preservação digital dos dados de pesquisa da sua instituição? Quais foram as competências essenciais para formar a equipe do IRD para a gestão dos dados?

Pascal Aventurier: São 4 pessoas (3 em tempo parcial) que trabalham no acompanhamento e na formação aos pesquisadores. Temos um grupo com gente da Missão Ciência Aberta, dirigido por Jean-Christophe Desconnets, do setor de infraestruturas da computação, da divisão jurídica e da valorização. É essencial ter uma equipe com várias competências. As competências essenciais dos chamados bibliotecários de dados (*data librarian*) correspondem às etapas do ciclo da vida dos

² <https://brasil.ird.fr/>

³ <https://www.cines.fr/>

dados (coletar, preservar disseminar, etc.). Para ajudar ou treinar, tem que dominar a redação de plano de gestão de dados: um documento que descreve a gestão de dados do projeto, os data paper e os formatos de metadados padrão. É importante também ter noções de gestão de projetos. A gestão de dados é diferente em cada comunidade de pesquisa. O IRD é multidisciplinar, por isso ele apresenta muitos casos diferentes.

RBPD [Gildenir]: Quais foram os obstáculos estruturais e tecnológicos que enfrentaram para realizar a preservação digital dos dados de pesquisa do IRD?

Pascal Aventurier: Nós não encontramos obstáculos estruturais por diversas razões: o movimento da ciência aberta é muito forte na França desde 2018. Trabalhamos com outras regiões do sul, isso está oficializado, mas também faz parte da nossa cultura de compartilhar material científico com Horizon Plein Textes para textos, DataSuds para dados e Indigo para imagens. Também, nós recebemos apoio de um administrador do Dataverse, que é um novo cargo no IRD. A direção da computação nos ajudou bastante e nós trabalhamos com a missão da ciência aberta. Nós temos uma política de reestruturação do Dataverse e estamos planejando uma atualização com o CINES para preservação de longo prazo.

3

RBPD [Gildenir]: Como foi elaborado o plano de gestão de dados do IRD? Você considerou fácil essa elaboração?

Pascal Aventurier: Nos usamos uma ferramenta DMP OPIDOR⁴ que a uma implementação da ferramenta DMP Online. Essa ferramenta oferece formatos padrão das agências de fomento ou de outras instituições. Usamos também as recomendações de outras instituições. Temos um grupo multidisciplinar para atender aos pesquisadores. Trabalhamos juntos também com colegas das instituições parceiras. Os serviços são lista de recomendações, em casos especiais, e às vezes revisão de plano de gestão de dados. Estamos finalizando as recomendações do IRD para implementá-las dentro de um modelo padrão. O importante é reunir todas as competências num grupo de trabalho, ensinar a esse grupo o que é um PGD e manter a dinâmica.

⁴ <https://dmp.opidor.fr/>

RBPD [Gildenir]: Quais instituições validam esse plano de gestão de dados da IRD?

Pascal Aventurier: A validação é feita para agências de fomento. Por exemplo a União Europeia exige o PGD desde a submissão do projeto. O objetivo do PGD é que dentro de um projeto exista uma reflexão sobre gestão de dados e uma avaliação dos custos e das tarefas a serem realizadas.

RBPD [Gildenir]: Quais foram os conjuntos de dados (tipologia) catalogados inicialmente no repositório de dados de pesquisa do IRD, e qual a quantidade existente no repositório hoje?

Pascal Aventurier: O Datasuds recebe conjuntos de dados que pertence na cauda longa dos dados (Long Tail), não são big data ou dados que são produzidos dentro de uma infraestrutura de pesquisa. São dados sobre a região sul. Principalmente. Hoje temos 85 conjuntos de dados, mas tem muito em preparação para divulgar.

RBPD [Gildenir]: Vocês praticam todos os princípios FAIR no repositório de dados de pesquisa do IRD? Poderia falar um pouco como acontece esse processo?

Pascal Aventurier: Tem um dispositivo de formação importante para os pesquisadores na gestão de dados e também um site⁵. Nós temos uma rede de referentes em cada laboratório. Antes da submissão dos dados o administrador ou o referente aconselham sobre descrição de dados, formato de dados, licença, e sobre restrições (dados sensíveis, Lei Geral de Proteção de Dados Pessoais (LGPD) etc.). Temos ferramentas para avaliar se as funcionalidades de DataSuds estão cumprindo os princípios FAIR, é um caso para melhoria dos critérios. O mais difícil é a interoperabilidade. Estamos trabalhando nesse assunto com vocabulário em SKOS e guia de boas práticas.

RBPD [Gildenir]: A equipe do repositório de dados de pesquisa do IRD adota algum vocabulário controlado para padronizar os termos utilizados no âmbito do repositório?

Pascal Aventurier: Sim, estamos trabalhando num vocabulário temático em SKOS (web semântica). Esse vocabulário seria usado para as ferramentas do IRD e disponível numa ferramenta para disponibilização de linked data (SKOSMOS). Também, estamos trabalhando no alinhamento com outros vocabulários.

⁵ <https://data.ird.fr/>

RBPD [Gildenir]: O repositório de dados de pesquisa do IRD é certificado por alguma agência internacional de certificação de repositórios?

Pascal Aventurier: Vamos começar a processo para receber a certificação CoreTrustSeal que também tem relação com a preservação de dados.

RBPD [Gildenir]: Além da preservação digital que fazem para o repositório de dados de pesquisa da IRD, você tem um plano de gestão de risco para o repositório?

Pascal Aventurier: Nós não escrevemos plano de gestão de risco, mas pensamos sobre ele. Tem risco ligados a tecnologia, mas acho que o mais importante é de não fazer um repositório cemitério onde ninguém vem depositar ou reusar os dados. Outro elemento é a integração de outras instituições com o repositório nacional que está em previsão. Também, estamos trabalhando sobre um projeto BRIDGE Science Ouverte⁶ com duas outras instituições para definir regras comuns no caso de projeto em comum (onde depositar os dados, quais regras? etc.).

⁶ <https://ist.blogs.inrae.fr/bridge-science-ouverte/>