

A construção do Plano de Dados Abertos de uma organização pública de Pesquisa e Desenvolvimento e o desafio de uma Ciência Agropecuária Aberta

Building up an Open Data Plan for a public Research and Development organization and the challenge of an Open Agricultural Science

La construcción del Plan de Datos Abiertos de una organización pública de Investigación y Desarrollo y el desafío de una Ciencia Agropecuaria Abierta

Patrícia Rocha Bello Bertin | patricia.bertin@embrapa.br

Embrapa, Secretaria de Gestão e Desenvolvimento Institucional. Brasília, DF, Brasil.

Claudia Delaia Machado | claudia.delaia@embrapa.br

Embrapa Solos. Rio de Janeiro, Brasil.

Marcos Cezar Visoli | marcos.visoli@embrapa.br

Embrapa Informática Agropecuária. Campinas, Brasil.

Debora Pignatari Drucker | debora.drucker@embrapa.br

Embrapa Informática Agropecuária. Campinas, Brasil.

Daniela Maciel Pinto | daniela.maciell@embrapa.br

Embrapa Monitoramento por Satélite. Campinas, Brasil.

Resumo

Em atendimento à Política de Dados Abertos do Poder Executivo Federal e à Lei de Acesso à Informação, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) iniciou discussões para a construção do seu Plano de Dados Abertos – um instrumento de planejamento para a implementação e a racionalização dos processos de publicação de dados abertos em organizações públicas. Esse trabalho visa relatar a estratégia geral que foi estabelecida para essa finalidade e os desafios que têm sido encontrados nos estágios preliminares de sua implantação. Como a abertura de dados por instituições de pesquisa no Brasil é ainda incipiente, espera-se que a experiência da Embrapa possa contribuir com iniciativas de organizações congêneres.

Palavras-chave: dados abertos; ciência agropecuária; ciência aberta; e-science; dados de pesquisa; política de dados abertos; Lei de Acesso à Informação.

Abstract

In keeping with the Open Data Policy of the Federal Executive Branch and the Law on Access to Information, the Brazilian Agricultural Research Corporation (Embrapa) has initiated discussions on the construction of its Open Data Plan – a planning tool for implementation and rationalization of open data publishing processes in public organizations. This paper aims to report the general strategy that was established for this purpose and the challenges that have been encountered in the preliminary stages of its implementation. As the opening of data by research institutions in Brazil is still in its infancy, Embrapa's experience is expected to contribute to similar initiatives from other organizations.

Keywords: open data; agricultural science; open science; e-science; research data; open data policy; Law on Access to Information.

Resumen

En cumplimiento de la Política de Datos Abiertos del Poder Ejecutivo Federal y la Ley de Acceso a la Información, la Empresa Brasileña de Investigación Agropecuaria (Embrapa) inició discusiones para la construcción de su Plan de Datos Abiertos – un instrumento de planificación para la implementación y la racionalización de los procesos de publicación de datos abiertos en organizaciones públicas. Este trabajo pretende relatar la estrategia general que se estableció para este propósito y los desafíos que se han encontrado en las etapas preliminares de su implantación. Como la apertura de datos por instituciones de investigación en Brasil es todavía incipiente, se espera que la experiencia de Embrapa pueda contribuir con iniciativas de organizaciones congéneres.

Palabras clave: datos abiertos; ciencia agropecuaria; ciencia abierta; e-science; datos de investigación; política de datos abiertos; Ley de Acceso a la Información.

Introdução

Há um movimento global para que instituições de ciência e tecnologia incentivem os dados abertos como um importante mecanismo para maximizar a visibilidade da pesquisa, reduzir os custos da produção e avanço do conhecimento científico, facilitar a transferência e propiciar mecanismos de verificação deste conhecimento, garantindo que permaneça como um bem público global. O compartilhamento de dados científicos (ou 'de pesquisa') consistiria em elemento fundamental do novo paradigma da e-Science¹⁻³.

A ciência agropecuária se preocupa em compreender sistemas naturais complexos e balancear a produção de alimentos com a preservação ambiental. Sendo um domínio científico multidisciplinar, complexo e heterogêneo, é possível que a ciência agropecuária não realizará todo o seu potencial no contexto da e-Science exclusivamente por meio da abertura e compartilhamento de dados de pesquisa. Para o benefício e avanço do setor agrícola, outros tipos de dados que se relacionam com o contexto de aplicação do conhecimento agropecuário precisam também ser disponibilizados abertamente.

Organizações internacionais como a Global Open Data for Agriculture and Nutrition (Godan) têm conduzido estudos que evidenciam como dados abertos podem contribuir para a solução de problemas nos setores agrícolas e de alimentos⁴, buscando com isso estimular a discussão sobre os potenciais usos, as necessidades e os desafios que pessoas, organizações e governos enfrentam ao iniciarem ações para a abertura de dados no contexto agrícola.

No Brasil, com a publicação da Lei de Acesso à Informação (Lei nº 12.527, de 18 de novembro de 2011) e, mais recentemente, do Decreto no. 8.777 (em 11 de maio de 2016) – que estabeleceu a Política de Dados Abertos do Poder Executivo Federal –, o debate sobre a abertura de dados ultrapassou o contexto da administração direta para alcançar empresas públicas de Pesquisa e Desenvolvimento, como é o caso da Empresa Brasileira de Pesquisa Agropecuária – Embrapa. Criada em 1973 e vinculada ao Ministério da Agricultura, Pecuária e Abastecimento, a Embrapa tem por missão ‘viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira’, tendo contribuído para o desenvolvimento de um modelo de agricultura e pecuária tropical que superou as barreiras que limitavam a produção de alimentos, fibras e energia no país. A Embrapa está estruturada em 17 unidades administrativas, localizadas em Brasília, e 46 unidades descentralizadas distribuídas no território nacional – entre as quais estão cinco unidades especiais de serviços e 41 unidades de Pesquisa e Desenvolvimento.

A Política de Dados Abertos do Poder Executivo Federal tem por objetivos: (i) promover a publicação de bases de dados de órgãos e entidades da administração pública federal direta, autárquica e fundacional sob a forma de dados abertos; (ii) aprimorar a cultura de transparência pública; (iii) franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelo Poder Executivo federal, sobre os quais não recaia vedação expressa de acesso; (iv) facilitar o intercâmbio de dados entre órgãos e entidades da administração pública federal e as diferentes esferas da federação; (v) fomentar o controle social e o desenvolvimento de novas tecnologias destinadas à construção de ambiente de gestão pública participativa e democrática e à melhor oferta de serviços públicos para o cidadão; (vi) fomentar a pesquisa científica de base empírica sobre a gestão pública; (vii) promover o desenvolvimento tecnológico e a inovação nos setores público e privado e fomentar novos negócios; (viii) promover o compartilhamento de recursos de tecnologia da informação, de maneira a evitar a duplicidade de ações e o desperdício de recursos na disseminação de dados e informações; e (ix) promover a oferta de serviços públicos digitais de forma integrada. A implementação da Política de Dados Abertos, conforme estabelece o Decreto 8.777, se dará por meio da elaboração e execução de ‘Plano de Dados Abertos’ no âmbito de cada órgão ou entidade da administração pública federal. ‘Dados abertos’, segundo o Decreto, são aqueles “acessíveis ao público, representados em meio digital, estruturados em formato aberto, processáveis por máquina, referenciados na internet e disponibilizados sob licença aberta que permita sua livre utilização, consumo ou cruzamento, limitando-se a creditar a autoria ou a fonte”.

O Plano de Dados Abertos (PDA), por sua vez, é um documento orientador das “ações de implementação e promoção de abertura de dados, inclusive os geoespecializados, que deverão obedecer aos padrões mínimos de qualidade, de forma a facilitar o entendimento e a reutilização das informações”, conforme define o Ministério de Planejamento, Desenvolvimento e Gestão⁵. No processo de elaboração do PDA, as diversas áreas de negócio têm um papel fundamental, por melhor conhecerem os dados e informações que geram e que obtêm no exercício de suas funções.

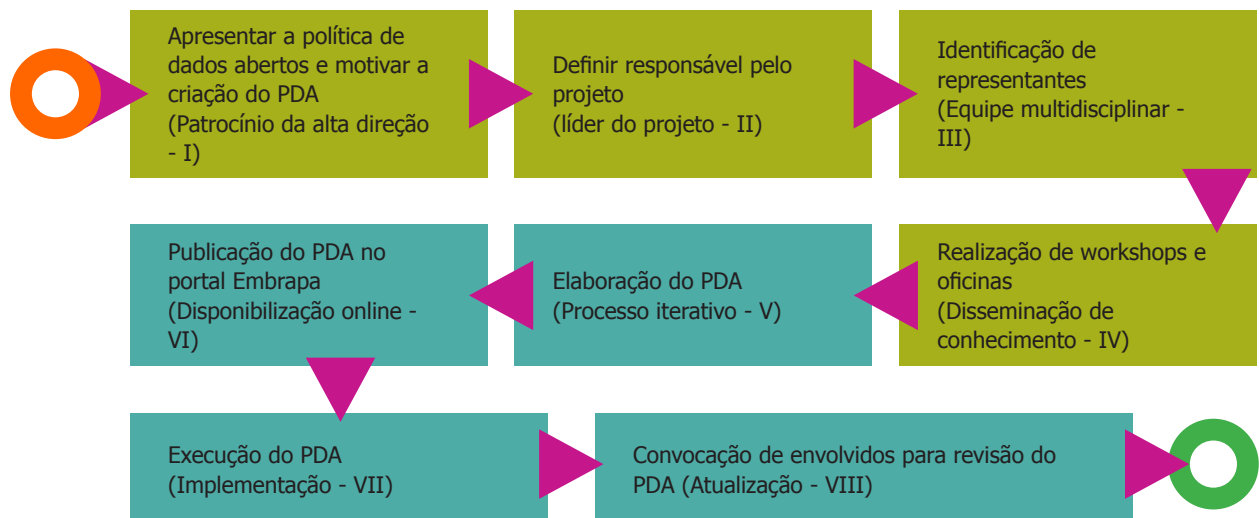
Face à importância e urgência de um debate aprofundado sobre dados abertos, a Embrapa acatou, recentemente, o desafio de construir seu PDA, seguindo procedimentos recomendados pelo Ministério do Planejamento, Desenvolvimento e Gestão e orientações complementares da Infraestrutura Nacional de Dados Abertos (INDA). Na Embrapa, a atividade de construção do PDA integra um projeto de desenvolvimento institucional mais amplo, intitulado ‘Inserção de ativos de soluções tecnológicas da Embrapa no mercado de negócio digital’, coordenado pelo Departamento de Tecnologia da Informação em colaboração com a Secretaria de Gestão e Desenvolvimento Institucional, e que conta com a participação das diversas unidades centrais e descentralizadas da Empresa.

O objetivo do presente trabalho é relatar a estratégia geral que foi estabelecida para a construção do PDA da Embrapa e os desafios encontrados nos estágios preliminares de planejamento, a partir das percepções dos participantes no processo.

Métodos

A construção do PDA da Embrapa segue os procedimentos recomendados no ‘Manual para a Elaboração de Plano de Dados Abertos’ editado e publicado no final de 2013 pelo Ministério do Planejamento, Desenvolvimento e Gestão - MP, para servir de instrumento orientador aos órgãos da Administração Pública Federal no planejamento das ações de publicação de dados abertos. O documento incorpora recomendações do Comitê Gestor da INDA e aprendizados obtidos a partir da construção do Plano de Dados Abertos no âmbito do próprio MP. A Figura 1 destaca as etapas que compõem o processo de construção do PDA, conforme instrui o manual. Em verde, as etapas já cumpridas pela Embrapa.

Figura. 1. *Etapas do processo de construção do Plano de Dados Abertos da Embrapa*



Uma vez cumpridas as etapas I a III, que dizem respeito ao comprometimento da alta direção e definição da governança do processo, iniciaram-se os eventos presenciais e virtuais de sensibilização e alinhamento para a construção do PDA da Embrapa. Como parte da etapa IV, foi realizado em Brasília, em março de 2017, o ‘Workshop de Alinhamento Conceitual e Estratégia para Dados Abertos’, cujo objetivo foi alinhar conceitos e apresentar o panorama nacional e internacional de Dados Abertos; apresentar a proposta metodológica para elaboração do PDA da Embrapa; e discutir perspectivas e iniciativas voltadas para dados abertos na Empresa. Ao workshop seguiram-se reuniões presenciais e conferências-web com representantes de cada uma das unidades centrais e descentralizadas da Empresa – os perfis, a formação e a área de atuação desses profissionais eram dos mais diversos, propositadamente na intenção de melhor representar o conjunto de empregados da Embrapa.

As reuniões culminavam com o seguinte questionamento: “Do universo de dados produzidos pela sua unidade, quais conjuntos possuem maior potencial para abertura?”. Os representantes eram então instruídos quanto aos critérios que deveriam considerar para priorização dos dados para abertura, a saber: (i) o grau de relevância para o cidadão ou segmentos de público da Embrapa; (ii) a eventual obrigatoriedade legal ou compromisso assumido pela unidade para a disponibilização dos dados; (iii) a associação às diretrizes estratégicas da unidade (especificidade); (iv) dados já disponibilizados ou que podem ser encontrados em bancos ou sistemas sob a gestão da unidade; e (v) o nível de maturidade na organização dos dados. Os conjuntos de dados selecionados eram então caracterizados em relação à localização e situação do dado, ao setor responsável, à frequência de atualização do dado (se alguma) e ao cronograma sugerido para publicação.

Embora o processo ainda esteja em andamento, a análise temática das discussões ocorridas até o presente permitem identificar as principais dificuldades que a organização enfrentará tanto no processo de

elaboração quanto na execução do seu PDA. Os temas que dominaram o debate e que, portanto, representam as principais preocupações dos participantes no que concerne à abertura de dados pela Empresa são explorados na seção a seguir.

Resultados e Discussão

A abertura de dados na administração pública visa, sobretudo, garantir os princípios da publicidade e da transparência. Como organização atuante no segmento de Pesquisa e Desenvolvimento, no entanto, a Embrapa precisa tratar os dados, a informação e o conhecimento que produz com acrisolado zelo e responsabilidade, implementando ações que garantam tanto a proteção e a segurança do conhecimento sensível quanto a ampla divulgação de conhecimentos e informações de interesse social. Essa abordagem é condizente com a Lei de Acesso à Informação, que estabeleceu a publicidade como preceito geral e o sigilo como exceção na administração pública – este último, aplicável apenas nos casos previstos em lei (como nos sigilos contábil e bancário, segredo industrial, de justiça, etc.).

Com a elaboração e execução do seu PDA, a Embrapa terá intensificado os esforços de transparência ativa, ou seja, a disponibilização de dados e informações de interesse da sociedade por iniciativa da própria Empresa. Não é difícil imaginar que muitos desafios e conflitos são encontrados no caminho para abertura de dados em uma empresa de Pesquisa e Desenvolvimento, os quais podem ser sumarizados na forma das seguintes constatações:

- Dificuldades em distinguir os dados com vocação para a abertura daqueles que precisam ser protegidos e classificados como sigilosos, por representarem um diferencial competitivo para a Empresa ou cuja divulgação possa causar risco a projetos de P&D.
- Objeções à abertura de dados oriundos ou relacionados à atividade de P&D.
 - a. À menção do termo ‘abertura de dados’ em uma organização de P&D, o primeiro impulso é apontar para dados de natureza administrativa, como atos de gestão, contratos, etc., os quais compõem o rol mínimo de informações cuja disponibilização é exigida pela Lei de Acesso à Informação. Ou seja, há dificuldades em conceber e uma resistência imediata para a abertura de dados de natureza técnico-científica.
 - b. Para além dos dados de natureza administrativa, dados georreferenciados e dados meteorológicos ocupam o segundo e terceiro lugar entre aqueles que são mais facilmente associados à noção de abertura.
 - c. Dentre os dados oriundos ou relacionados à atividade de P&D, conjuntos de dados associados a herbários e outras coleções científicas são aqueles cujo potencial para abertura é mais facilmente percebido.
- Observa-se uma tendência em oferecer, como dado aberto, aqueles dados que estão excessivamente processados, agregados e tratados, em vez de dados primários.
- Há desafios no que concerne à maturidade na organização dos dados, visto que dados com vocação para abertura não estão necessariamente estruturados, podendo inclusive estar de posse do pesquisador e não da organização.
- Percebe-se a necessidade de aporte de recursos humanos e orçamentários, para a consecução das atividades que conduzirão à abertura de dados. O desenvolvimento de competências ou a contratação de perfis profissionais específicos como os do ‘bibliotecário de dados’ e ‘engenheiro de dados’ mostra-se essencial.
- Resistências culturais demandam um plano de comunicação arrojado, com foco em públicos específicos como gestores, pesquisadores, analistas de TI, entre outros, e com o desejável envolvimento alta-direção, para maior legitimidade das ações.

- Há uma carência pela definição de métricas que tanto sirvam de incentivo quanto facilitem o monitoramento das ações de abertura de dados.
- A ausência de uma política institucional ou diretriz específica nos documentos de planejamento estratégico da organização prejudica a internalização e o aculturamento da noção de dados abertos. Nesse sentido, também a omissão da perspectiva de gestão e abertura de dados nas políticas de agências nacionais de fomento à pesquisa prejudica a emergência de uma cultura mais aberta de produção do conhecimento agrícola.

Além das dificuldades e conflitos apontados acima, foi possível alcançar, nos estágios preliminares de construção do PDA da Embrapa, os seguintes aprendizados:

- **Importância de inserção da Empresa nos fóruns relevantes, nacionais e internacionais**, para discussão e avanço na temática dos dados abertos, juntamente com instituições congêneres e outros interessados. A transição para um modo mais aberto de produção do conhecimento científico e tecnológico se beneficiará grandemente da aproximação e do intercâmbio com outras organizações e iniciativas exitosas no cenário nacional e internacional, como a Research Data Alliance (<http://rd-alliance.org>) e a Godan (<http://www.godan.info/>), que oferecem espaços neutros para discussão de dados abertos no contexto científico e tecnológico agrícola. Além disso, ressalta-se a necessidade de integrar as ações que a Empresa venha a desenvolver com as iniciativas governamentais, como a INDA e a Infraestrutura Nacional de Dados Espaciais (INDE), para reforço mútuo e maior visibilidade das ações.
- **O bibliotecário ou o profissional da informação possui papel central nas ações que conduzem à abertura de dados.** Dados bem catalogados, documentados e organizados são pré-requisitos fundamentais para a abertura dos dados, sua recuperação e reuso.
- Para que a estratégia de abertura de dados da Embrapa logre êxito, faz-se necessário o **envolvimento direto das diversas áreas da Empresa** no processo de construção do PDA, em particular, as áreas de Pesquisa, Gestão Estratégica, Negócios, Transferência de Tecnologia, Comunicação, Tecnologia da informação, Inteligência e Análise de Dados.
- Urge implementar **ações de sensibilização e esclarecimento do público interno quanto à aparente dicotomia entre a abertura de dados e os processos que visam garantir a segurança da informação, o sigilo e a proteção intelectual.**
- **A abertura de dados deve ser feita de modo planejado e coordenado**, a partir de critérios definidos por uma estrutura de governança de dados abertos e uma política corporativa para a gestão de dados científicos que esteja articulada à política de segurança da informação e à norma de classificação da informação. Questões como ‘a quem cabe decidir’, na Empresa, se o dado tem ou não vocação para abertura devem, também, ser tratadas na política corporativa de gestão de dados e eventuais dúvidas dirimidas pelos gestores das unidades centrais e descentralizadas, em parceria com aquele que concretamente produziu ou obteve os dados.
- **Os mecanismos de organização, gestão e estruturação dos dados da Empresa devem ser aprimorados**, particularmente aqueles produzidos pela atividade de P&D, para que estejam aptos ao compartilhamento interno e à abertura mais ampla, quando assim convier. Como proposta concreta, propõe-se que o Sistema Embrapa de Gestão (SEG) passe a requerer, ao final dos projetos financiados, o depósito dos dados produzidos ou obtidos em um repositório institucional. Na mesma linha, é recomendável que, no momento da submissão de projetos ao SEG, o proponente seja requisitado a apontar, em linhas gerais, a forma como a gestão dos dados produzidos pelo projeto será feita – o que seria um primeiro passo para a solicitação, em momento posterior (e quando a cultura organizacional estiver melhor preparada), de um Plano de Gestão de Dados de Pesquisa para cada projeto.
- **Mecanismos de incentivo à gestão e abertura de dados precisam ser implementados**, com valorização e fortalecimento das experiências já existentes na Empresa

para disponibilização de dados abertos. Algumas iniciativas já identificadas são os repositórios de informação científica e tecnológica ‘Alice’ e ‘Infoteca-e’, respectivamente, e a Infraestrutura de Dados Espaciais da Embrapa, que objetiva publicar dados espaciais abertos na INDE.

- Dentro do ‘guarda-chuva’ da Ciência Agropecuária, a Embrapa atua em uma vasta gama de especialismos científicos, o que se reflete na produção de dados dos tipos mais variados. Por isso, é relevante **realizar estudos que conduzam a um entendimento mais aprofundado dos outputs produzidos pelas culturas científicas representadas na Embrapa**, a fim de definir parâmetros e descritores que garantam maior qualidade na organização e recuperação dos dados.

Considerações finais

A construção do PDA da Embrapa acontece no momento em que se intensificam os debates, em nível global, sobre a emergência de um novo paradigma científico, comumente designado Ciência Aberta ou e-Science, entre outros. O novo fazer científico, mais transparente, intenso e dependente do compartilhamento e análise de grandes volumes de dados, está aos poucos alcançando espaço, desafiando os modelos tradicionais de produção do conhecimento.

Embora as tendências globais apontem para a emergência de um modo científico mais aberto, com a intensificação da pesquisa em rede, o fortalecimento das colaborações em escala global, e o aumento da produção, análise e visualização de dados – ou seja, o fortalecimento de uma ciência ‘dirigida por’ e ‘intensiva em’ dados –, a realidade das organizações de P&D aponta diversos desafios e conflitos que precisam ser melhor analisados e abordados. O processo de construção do PDA de uma organização de P&D auxilia na identificação desses desafios e no entendimento de como estes influenciam a percepção e o comportamento dos pesquisadores, particularmente no tocante à abertura de dados de pesquisa.

Referências

1. Research Councils UK (2010). Delivering the UK’s e-Infrastructure for Research and Innovation. Available at: <http://www.rcuk.ac.uk/documents/research/esci/e-infrastructurereviewreport-pdf/>.
2. Gray, J. (2009). Jim Gray on eScience: A transformed scientific method. In *The fourth paradigm: Data-intensive scientific discovery*, T. Hey, S. Tansley, and K. Tolle, eds. (Redmond: Microsoft Research).
3. Wouters, P. (2004). c. In *4S & EASST Conference: Public Proofs, Science, Technology and Democracy* (Ecole des Mines, Paris).
4. Carolan, L., Smith, F., Protonotarios, V., Schaap, B., Broad, E., Hardinges, J., and Gerry, W. (2015). How can we improve agriculture, food and nutrition with open data? (London: Open Data Institute) Available at: <http://www.godan.info/sites/default/files/old/2015/04/ODI-GODAN-paper-27-05-20152.pdf>.
5. Ministério do Planejamento, Desenvolvimento e Gestão. (2013). Manual para a elaboração de Plano e Dados Abertos (Brasília: Ministério do Planejamento, Desenvolvimento e Gestão).