

Estado da publicação: O preprint não foi publicado em outro meio.

# Infraestrutura Semântica Aberta e Soberana: Convergência entre Ciência da Informação, Ciência de Dados e Inteligência Artificial

Ana Carolina Simionato Arakaki, Washington R. de Carvalho Segundo, Priscila Machado Borges Sena

<https://doi.org/10.1590/SciELOPreprints.15856>

Submetido em: 2026-04-17

Postado em: 2026-04-23 (versão 1)  
(AAAA-MM-DD)

A moderação deste preprint recebeu o(s) endosso(s) de:

- Sigmar de Mello Rode (ORCID: <https://orcid.org/0000-0002-4261-4217>)
- André Luiz Appel (ORCID: <https://orcid.org/0000-0002-9608-803X>)
- Walter Couto (ORCID: <https://orcid.org/0000-0002-6356-9304>)

**PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

**Infraestrutura Semântica Aberta e Soberana:  
Convergência entre Ciência da Informação, Ciência de  
Dados e Inteligência Artificial**

**Open and Sovereign Semantic Infrastructure:  
Convergence between Information Science, Data Science  
and Artificial Intelligence**

**Infraestructura semántica abierta y soberana:  
convergencia entre la ciencia de la información, la ciencia  
de datos y la inteligencia artificial**

Ana Carolina Simionato Arakaki  
Doutora em Ciência da Informação  
Universidade de Brasília - UnB  
ORCID: <https://orcid.org/0000-0002-0140-9110>  
Lattes: <http://lattes.cnpq.br/9896600626524397>  
[ana.arakaki@unb.br](mailto:ana.arakaki@unb.br)

Washington Luís Ribeiro de Carvalho Segundo  
Doutor em Informática  
Instituto Brasileiro de Informação em Ciência e Tecnologia - Ibict  
ORCID: <https://orcid.org/0000-0003-3635-9384>  
Lattes: <http://lattes.cnpq.br/9453481318889500>  
Email: [washingtonsegundo@ibict.br](mailto:washingtonsegundo@ibict.br)

Priscila Machado Borges Sena  
Doutora em Ciência da Informação  
Instituto Brasileiro de Informação em Ciência e Tecnologia - Ibict  
ORCID: <https://orcid.org/0000-0002-5612-4315>  
Lattes: <https://lattes.cnpq.br/0155235005204514>  
Email: [priscilasena@ibict.br](mailto:priscilasena@ibict.br)

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

### **Resumo**

A ascensão da Inteligência Artificial (IA) no século XXI tem sido amplamente interpretada como uma revolução tecnológica centrada em algoritmos e capacidade computacional. No entanto, essa abordagem tende a fragilizar um elemento estrutural decisivo: a qualidade, organização e semântica dos dados que sustentam esses sistemas. Este artigo analisa a convergência entre Ciência da Informação (CI), Ciência de Dados (CD) e Inteligência Artificial, defendendo que a CI constitui a condição de possibilidade para a eficácia, confiabilidade e ética dos sistemas inteligentes. A pesquisa adota abordagem qualitativa, fundamentada em análise documental de materiais institucionais e revisão teórica interdisciplinar. Os resultados evidenciam que a relação entre essas áreas configura uma cadeia de valor estruturada, na qual a CI organiza, a CD analisa e a IA automatiza, tendo o dado como elemento central de articulação. Destaca-se o papel estratégico do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) na construção de uma infraestrutura nacional baseada em grafos de conhecimento, ontologias e lagos de dados, essenciais para a soberania informacional. Conclui-se que a integração dessas áreas é indispensável para o desenvolvimento de políticas públicas de IA alinhadas à Ciência Aberta, à transparência e ao desenvolvimento científico e tecnológico.

### **Palavras-chave:**

Ciência da Informação; Inteligência Artificial; Ciência de Dados; Soberania Informacional.

### **Abstract**

The rise of Artificial Intelligence (AI) in the 21st century has largely been interpreted as a technological revolution centred on algorithms and computational power. However, this approach tends to overlook a crucial structural element: the quality, organisation and semantics of the data underpinning these systems. This article analyses the convergence between Information Science (IS), Data Science (DS) and Artificial Intelligence, arguing that IS constitutes the enabling condition for the effectiveness, reliability and ethics of intelligent systems. The research adopts a qualitative approach, based on a documentary analysis of institutional materials and an interdisciplinary theoretical review. The results show that the relationship between these areas forms a structured value chain, in which IS organises, DS analyses and AI automates, with data as the central element of coordination. The strategic role of the Brazilian Institute of Information in Science and Technology (Ibict) in building a national infrastructure based on knowledge graphs, ontologies and data lakes, which are essential for informational sovereignty, is highlighted. It is concluded that the integration of these areas is indispensable for the development of public AI policies aligned with Open Science, transparency, and scientific and technological development.

### **Keywords:**

Information Science; Artificial Intelligence; Data Science; Information Sovereignty.

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

### **Resumen**

El auge de la Inteligencia Artificial (IA) en el siglo XXI se ha interpretado ampliamente como una revolución tecnológica centrada en los algoritmos y la capacidad computacional. Sin embargo, este enfoque tiende a pasar por alto un elemento estructural decisivo: la calidad, la organización y la semántica de los datos que sustentan estos sistemas. Este artículo analiza la convergencia entre la Ciencia de la Información (CI), la Ciencia de Datos (CD) y la Inteligencia Artificial, defendiendo que la CI constituye la condición de posibilidad para la eficacia, la fiabilidad y la ética de los sistemas inteligentes. La investigación adopta un enfoque cualitativo, basado en el análisis documental de materiales institucionales y en una revisión teórica interdisciplinar. Los resultados evidencian que la relación entre estas áreas configura una cadena de valor estructurada, en la que la CI organiza, la CD analiza y la IA automatiza, teniendo el dato como elemento central de articulación. Cabe destacar el papel estratégico del Instituto Brasileño de Información en Ciencia y Tecnología (Ibict) en la construcción de una infraestructura nacional basada en grafos de conocimiento, ontologías y lagos de datos, esenciales para la soberanía informacional. Se concluye que la integración de estas áreas es indispensable para el desarrollo de políticas públicas de IA alineadas con la ciencia abierta, la transparencia y el desarrollo científico y tecnológico.

### **Palabras clave:**

Ciencias de la Información; Inteligencia Artificial; Ciencia de Datos; Soberanía informativa.

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

### **Introdução**

A transformação digital contemporânea tem sido frequentemente descrita como uma revolução impulsionada pela Inteligência Artificial (IA), cujos avanços parecem decorrer, à primeira vista, do aperfeiçoamento de algoritmos e do aumento da capacidade computacional.

A Ciência da Informação (CI), desde sua origem, dedica-se à compreensão dos fluxos, sistemas e usos da informação; a Ciência de Dados (CD) concentra-se na modelagem e na extração de padrões a partir dos dados; e a IA atua na automatização de processos de representação, inferência e ação (Cox; Mazumdar, 2022; Wang, 2018; Warner, 2001). A aproximação entre esses campos intensifica-se no ambiente digital, onde as fronteiras entre documento, dado, sistema e decisão tornam-se cada vez mais difusas. Ainda assim, cada área mantém seu próprio eixo epistemológico, e é justamente essa diferença que torna sua articulação produtiva: não se trata de campos equivalentes, mas de domínios que se complementam ao enfrentar, sob perspectivas distintas, os desafios contemporâneos de organização, análise, automação e uso social da informação (Marchionini, 2023; Qin, 2020; Shah, 2023).

Nesse contexto de convergência, a narrativa dominante que associa o avanço da IA sobretudo ao desenvolvimento de algoritmos e à ampliação da capacidade computacional revela-se insuficiente, pois tende a obscurecer o papel estruturante dos dados e, principalmente, das formas de sua organização, interpretação e contextualização. A inteligência produzida por sistemas de IA não é intrínseca às máquinas, mas emerge da qualidade e da estrutura semântica dos dados que as alimentam, evidenciando a centralidade da CI nesse processo.

Nesse contexto, torna-se fundamental compreender a convergência entre CI, CD e IA, como uma reconfiguração das infraestruturas abertas de informação científica e tecnológica. Longe de constituírem áreas isoladas, esses campos articulam-se em uma cadeia funcional e conceitual. Essa cadeia não apenas redefine a produção do conhecimento, mas reforça o papel da informação como ativo estratégico central no século XXI.

A relevância dessa discussão se amplia quando inserida no contexto do chamado “Quarto Paradigma” da ciência, caracterizado pela centralidade dos dados na geração de conhecimento (Gray, 2009). No quarto paradigma, o desafio não reside apenas na capacidade de processar grandes volumes de dados, mas na habilidade de atribuir significado a esses dados, evitando que a abundância informacional se converta em ruído. É nesse ponto que a tradição da CI, historicamente dedicada à organização, mediação e recuperação da informação, assume papel fundamental (Saracevic, 1996; Marchionini, 2023).

Assim, este artigo parte da hipótese de que a CI não é uma disciplina auxiliar ou periférica, mas a base estruturante que possibilita a existência de sistemas de IA confiáveis, éticos e socialmente relevantes. No contexto brasileiro, essa discussão ganha contornos ainda mais estratégicos ao ser articulada com políticas públicas recentes voltadas à transformação digital e à IA. O Plano Brasileiro de Inteligência Artificial (PBIA) explicita diretrizes orientadas ao bem-estar social, à soberania tecnológica e de dados, à governança participativa e à redução de desigualdades,

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

indicando que o desenvolvimento de sistemas inteligentes deve estar alinhado a princípios éticos, inclusivos e socialmente responsáveis. Ao enfatizar a necessidade de infraestrutura computacional, formação de capacidades nacionais e desenvolvimento de modelos baseados em dados locais, o Plano reforça a centralidade das infraestruturas informacionais na construção de uma IA confiável e alinhada às demandas do país (Brasil, 2025).

Nesse sentido, essas diretrizes convergem com a perspectiva defendida neste artigo, ao evidenciar que a qualidade, organização e governança dos dados constituem elementos estruturantes para a efetividade e legitimidade dos sistemas de IA. A partir dessa premissa, busca-se analisar a convergência entre CI, CD e IA, enfatizando o papel das infraestruturas informacionais e da soberania de dados no contexto brasileiro, com destaque para a atuação do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) no campo.

### **Percurso metodológico**

A investigação desenvolvida nesta comunicação adota uma abordagem qualitativa, de caráter exploratório e interpretativo, fundamentada na análise integrada de documentos institucionais e literatura científica especializada. Foram examinados materiais estratégicos produzidos pelo Ibict, incluindo relatórios técnicos, projetos de pesquisa, bem como estudos acadêmicos que discutem a evolução da CI, CD e IA.

A análise foi orientada por uma estrutura conceitual derivada de modelos de convergência entre os três campos, considerando três dimensões principais: a base epistemológica das disciplinas, a centralidade do dado como elemento articulador e a infraestrutura tecnológica e semântica necessária para a operacionalização dessa convergência. Essa abordagem permitiu compreender o fenômeno não apenas como intersecção disciplinar, mas como um sistema sociotécnico complexo, no qual tecnologia, informação e sociedade se entrelaçam. A discussão realizada foi pautada pela análise das intersecções das três áreas, observando subáreas derivadas nessas intersecções, tais como: modelagem de dados, modelos preditivos e processamento de linguagem natural (ver Figura 1).

## PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)

Fonte: Aatoria (2026).

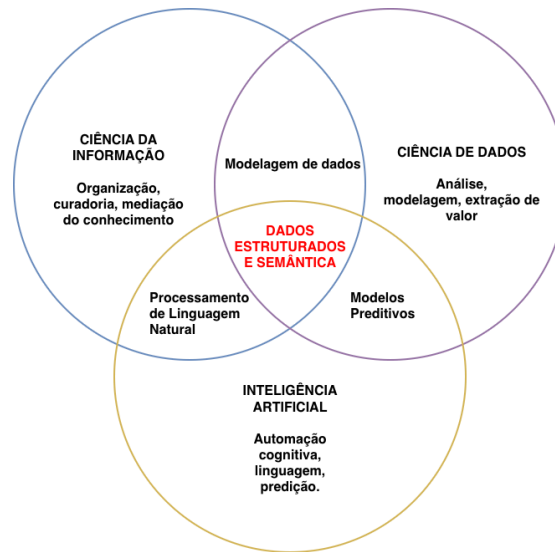


Figura 1. Diagrama de Venn com exemplos de subáreas emergentes nas intersecções entre os três campos CI, CD e IA.

### Resultados e discussão

A análise evidencia que a convergência entre CI, CD e IA configura-se como uma cadeia de valor informacional na qual cada campo desempenha funções complementares e interdependentes: a CI organiza, qualifica e contextualiza os dados; a CD os analisa e extrai padrões; e a IA automatiza processos cognitivos, viabilizando previsões e decisões em escala.

Essa estrutura reforça que a IA não constitui um ponto de partida, mas o estágio final de um processo informacional mais amplo, cuja eficácia depende diretamente da qualidade da base de dados que a sustenta. Como argumenta Marchionini (2023), a transformação de dados em conhecimento significativo exige não apenas volume, mas sobretudo contexto, estrutura e interpretação, elementos tradicionalmente associados à CI. Nesse sentido, a ausência de organização semântica compromete não apenas a precisão técnica dos sistemas, mas também sua legitimidade social, ao amplificar vieses e reduzir a confiabilidade dos resultados.

O dado, nesse cenário, emerge como elemento central de articulação entre os três campos, assumindo naturezas distintas conforme a perspectiva disciplinar: na CI, é concebido como entidade contextual e sociotécnica; na CD, como insumo analítico; e na IA, como representação formal manipulável por sistemas computacionais. Essa multiplicidade revela que o principal desafio contemporâneo não é apenas técnico, mas epistemológico, exigindo abordagens integradas que articulem dimensões quantitativas e qualitativas da informação (Hjørland, 2021).

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

Parte da literatura em CI sustenta que dados não são elementos brutos sem contexto, mas entidades produzidas, selecionadas, descritas e usadas em práticas sociais e técnicas. Isso aproxima a CI da CD, porque a utilidade analítica dos dados depende de descrição, proveniência, qualidade, contexto de uso e interpretação (Marchionini, 2023; Moutinho et al., 2024; Shah, 2023; Wang, 2018). Nessa perspectiva, CD entra não como campo externo, mas como extensão metodológica de problemas clássicos do domínio informacional (Marchionini, 2016; Mayernik, 2023; Virkus & Garoufallou, 2020). É importante mencionar que os dados passam a ser tratados como objetos que exigem descrição, preservação, interoperabilidade, documentação, políticas de acesso e suporte ao reuso, portanto, esse movimento recoloca a CI no centro da infraestrutura de pesquisa e aproxima o campo de competências antes associadas sobretudo à computação e à estatística (Cox et al., 2017; Cox & Pinfield, 2014; Federer, 2018; Semeler et al., 2019).

Do mesmo modo, a literatura recente rejeita a ideia de que a IA tornaria dispensáveis as funções da CI; o argumento dominante é o oposto: quanto mais automatizados os sistemas, maior a necessidade de profissionais capazes de avaliar dados, critérios de classificação, opacidade algorítmica, vieses e efeitos institucionais. Nesse quadro, a mediação profissional muda de forma, mas não perde centralidade (Cox & Mazumdar, 2022; Cox, 2022; Ridley; Pawlick-Potts, 2021).

Outro ponto destacado pela literatura foi que os desafios recorrentes evidenciam tensões estruturais no processo de convergência entre a CI e as tecnologias orientadas por dados. Destaca-se, em primeiro lugar, o risco de subordinação da identidade epistemológica da área a agendas predominantemente tecnicistas, que tendem a reduzir sua complexidade teórica e social (Marchionini, 2023; Wang, 2018). Soma-se a isso a persistência de lacunas formativas em competências como programação, estatística e IA, tanto nos currículos quanto na qualificação profissional, o que limita a atuação em contextos *data-driven* (Bridges et al., 2025; Ortiz-Repiso et al., 2018; Tait & Pierson, 2022). Observa-se ainda a dificuldade de articular, de forma integrada, fundamentos teóricos, práticas aplicadas e infraestruturas institucionais em programas formativos já densos e historicamente consolidados (Wu et al., 2022; Zhang et al., 2022). Paralelamente, ampliam-se as exigências éticas relacionadas a vieses algorítmicos, opacidade dos sistemas e desigualdades no acesso e uso da Inteligência Artificial, tensionando os princípios tradicionais da área (Coleman, 2020; Cox, 2022, 2024). Por fim, impõe-se a necessidade de renovação das práticas de organização do conhecimento diante de ambientes marcados por dados massivos, embeddings e processos automatizados de classificação, que desafiam modelos conceituais clássicos (Hjørland, 2021; Lima, 2025; Qin, 2020). Longe de indicarem incompatibilidade, essas tensões revelam a materialidade da convergência em curso, sinalizando uma reconfiguração efetiva das dimensões de ensino, pesquisa e prática profissional.

Portanto, a consolidação dessa convergência depende da existência de infraestruturas informacionais mais adequadas às necessidades computacionais e dos próprios usuários, capazes de garantir organização, interoperabilidade e preservação dos dados. No Brasil, o

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) destaca-se como ator estratégico na construção dessa infraestrutura, atuando como arquiteto de um ecossistema nacional de informação científica e tecnológica. Iniciativas como o Ecossistema de Informação da Pesquisa Científica Brasileira (BrCris)<sup>1</sup>, que estrutura a produção científica em forma de grafo de conhecimento, o Projeto Laguna (Carvalho-Segundo, Pinto, do Canto & Neubert, 2023), que implementa um lago de dados científicos alinhado aos princípios FAIR (Wilkinson et al, 2016), e a ontologia Pinakes (Arakaki, A. C. S. et al, 2025), que fornece base semântica para representação do conhecimento, estabelecem as condições necessárias para o desenvolvimento de sistemas de Inteligência Artificial confiáveis e transparentes.

Os projetos ServiçosPI e Patente.IA, executados pelo Ibict em parceria com o Instituto Nacional da Propriedade Industrial (INPI), evidenciam o potencial de aplicação ampliada de ferramentas de organização da informação, como lagos de dados, e de interfaces avançadas de recuperação da informação, associadas ao uso intensivo de técnicas de Processamento de Linguagem Natural (PLN) e de Grandes Modelos de Linguagem (LLMs).

Por sua vez, os projetos RDAPP e SAVITS resultam, respectivamente, da colaboração entre o Ibict e o Ministério do Planejamento e Orçamento (MPO), e da execução de emenda parlamentar da bancada do Distrito Federal, voltados à construção de plataformas que integram IA, Inteligência de Indicadores (BI) e Aprendizado de Máquina (ML).

As aplicações desenvolvidas no âmbito desses projetos são utilizadas na gestão e curadoria da informação, bem como na avaliação e meta-avaliação de políticas públicas. O Quadro 1 apresenta uma síntese das principais ferramentas e projetos conduzidos pelo Ibict nos campos da CI, CD e IA.

---

<sup>1</sup> <https://brcris.ibict.br/>

**PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

Quadro 1. Projetos do IbiCT que atuam na intersecção entre CI, CD e IA.

Projeto	Foco Institucional	Tecnologia Base	Impacto Central
Ontologia Pinakes	Padronização e representação do conhecimento científico	Modelagem ontológica, web semântica	Garante interoperabilidade e consistência semântica dos dados
BrCris	Integração e organização da produção científica nacional	Grafos de conhecimento, <i>linked data</i> , ontologias	Estruturação semântica da ciência brasileira e suporte a aplicações de IA e geração de indicadores de CT&I
Projeto Laguna	Gestão e integração de grandes volumes de dados científicos	Lago de Dados, arquitetura distribuída, princípios FAIR	Viabiliza análise em larga escala e treinamento de modelos de IA
Projeto ServiçosPI	INPI	PLN & Busca Avançada	Precisão em buscas de Propriedade Industrial.
Projeto Patente.IA	INPI	LLMs aplicados a texto e imagens	Identificação multimodal de anterioridade em invenções.
Projeto RDAPP	Governo	IA Analítica & Semântica	Transparência em políticas públicas e decisão baseada em evidência.
Projeto SAVITS	Avaliação Pública	BI + Aprendizado de Máquina (ML) + LLM	Predição ética e explicável do impacto de políticas.

Fonte: Aatoria (2026).

Essas infraestruturas, para além de organizar dados, transformam o conhecimento em ativo estratégico, contribuindo para a soberania informacional nacional e reduzindo a dependência de soluções tecnológicas externas. Nesse sentido, a dimensão infraestrutural da informação assume caráter geopolítico, uma vez que o controle sobre dados e sistemas passa a ser determinante para a autonomia científica e tecnológica das nações.

A aplicação prática dessa base semântica torna-se particularmente evidente em arquiteturas de IA baseadas em *Retrieval-Augmented Generation (RAG)*, que integram busca semântica, filtragem ontológica e geração por modelos de linguagem (ver Figura 2). Ao ancorar as respostas em fontes verificáveis, esse modelo reduz significativamente o problema das “alucinações” e amplia a rastreabilidade e a transparência dos sistemas (Cox, 2022). Mais do que uma solução técnica, trata-se de uma solução que recoloca a semântica no centro da IA e reforça o papel da CI como fundamento da inteligência computacional.

## PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)

Fonte: Aatoria (2026).

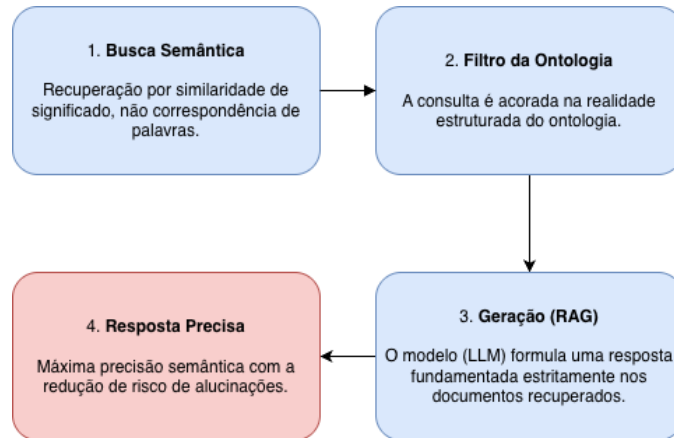


Figura 2. Exemplo de uso de Ontologia com RAG.

Nesse ecossistema, as bibliotecas de ensino superior assumem um papel estratégico que ultrapassa sua função tradicional de mediação e acesso à informação, posicionando-se como infraestruturas críticas na produção de dados que alimentarão sistemas de IA. Atuando no ponto de origem do ciclo informacional, essas instituições são responsáveis pela curadoria, organização e preservação de dados científicos, garantindo qualidade, padronização e interoperabilidade. Ao implementar planos de gestão de dados, aplicar metadados e alinhar-se aos princípios de publicação de dados, como o *Findable, Accessible, Interoperable, and Reusable* (FAIR), *Collective benefit, Authority to control, Responsibility and Ethics* (CARE) (Carroll, 2020), *Transparency, Responsibility, User focus, Sustainability, and Technology* (TRUST) (Lin, 2020) e *Diversity, Equity, Inclusion, and Accessibility* (DEIA) (Rosenberg et al., 2023) as bibliotecas produzem bases de dados estruturadas e semanticamente enriquecidas (ver Tabela 2), fundamentais para análises avançadas e para o treinamento de modelos de IA.

Além disso, as bibliotecas universitárias operam como espaços institucionais de implementação da Ciência Aberta, promovendo o acesso livre ao conhecimento e incentivando práticas de compartilhamento de dados. Essa atuação fortalece a transparência, a reprodutibilidade científica e a colaboração internacional, ampliando o potencial de inovação. Ao mesmo tempo, desempenham papel de mediação sociotécnica, conectando pesquisadores, sistemas e infraestruturas digitais, garantindo que os dados sejam não apenas armazenados, mas compreendidos e reutilizados em diferentes contextos.

Nesse cenário, o profissional da informação passa a atuar como agente de governança de dados, responsável por assegurar qualidade, integridade e ética nas bases informacionais. Essa transformação implica a incorporação de competências em ciência de dados, modelagem semântica e análise crítica da tecnologia, em consonância com as diretrizes das *iSchools*, que defendem uma formação interdisciplinar centrada no humano e na responsabilidade social (Wu et al., 2022).

## PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)

Por fim, a Ciência Aberta emerge como paradigma estruturante dessa convergência, promovendo a democratização do acesso ao conhecimento e fortalecendo a governança de dados. A adoção de princípios de publicação de dados amplia a dimensão ética e política da gestão informacional, incorporando preocupações com equidade, responsabilidade e transparência. Nesse contexto, a soberania informacional torna-se elemento central das políticas científicas, especialmente em países em desenvolvimento, onde a capacidade de controlar e utilizar dados de forma estratégica constitui um diferencial competitivo e um instrumento de autonomia tecnológica. Ver Quadro 2.

Quadro 2. Critérios aplicados a dados.

<b>Princípio</b>	<b>Foco</b>	<b>Significado</b>	<b>Objetivo</b>
FAIR (Wilkinson et al, 2016)	Sistemas e Máquinas	Findable, Accessible, Interoperable, Reusable	Padronização, metadados ricos e automação do reuso.
CARE (Carroll, 2020)	Comunidades e Povos	Collective benefit, Authority to control, Responsibility, Ethics	Proteger direitos coletivos, especialmente de povos indígenas.
TRUST (Lin, 2020)	Repositórios Digitais	Transparency, Responsibility, User focus, Sustainability, Technology	Garantir a robustez, certificação e financiamento da infraestrutura.
DEIA (Rosenberg et al., 2023)	Sociedade e Inclusão	Diversity, Equity, Inclusion, Accessibility	Democratizar a produção e o acesso à ciência.

Fonte: Aatoria (2026).

### Considerações finais

A convergência entre CI, CD e IA representa uma transformação estrutural na forma como o conhecimento é produzido, organizado e aplicado na sociedade contemporânea. Esta comunicação demonstrou que a Ciência da Informação desempenha papel central nesse processo, ao fornecer a base semântica e epistemológica necessária para que os dados se transformem em conhecimento confiável e socialmente relevante.

A análise evidenciou que a eficácia da IA não depende exclusivamente de avanços algorítmicos, mas da qualidade das infraestruturas informacionais que a sustentam. Nesse contexto, o papel do Ibict é estratégico, ao liderar a construção de uma infraestrutura nacional que integra dados, semântica e tecnologia, promovendo soberania informacional e alinhamento com as práticas da Ciência Aberta.

Conclui-se que o futuro da IA no Brasil dependerá da capacidade de integrar essas áreas de forma sistêmica, garantindo que o desenvolvimento tecnológico esteja fundamentado em bases

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

informativas sólidas, éticas e transparentes. A CI, longe de ser periférica, assume, portanto, posição central na construção de uma inteligência artificial orientada ao bem público e ao desenvolvimento científico e tecnológico de forma sustentável.

### **Disponibilidade de dados**

Os dados são apresentados no corpo do artigo.

### **Conflito de Interesses**

Exemplo: Os autores declaram não haver conflitos de interesses.

### **CRedit – Contribuições dos Autores**

Ana Carolina Simionato Arakaki | Conceitualização, Escrita – redação original, Supervisão, Recolha de dados

Washington Luís Ribeiro de Carvalho Segundo | Conceitualização, Escrita – redação original, Supervisão, Coleta de dados

Priscila Machado Borges Sena | Escrita – redação original, Supervisão.

### **Referências**

Arakaki, A. C. S., et al. (2025). Ontologia Pinakes: Formalização Semântica para o Catálogo Coletivo Nacional de Publicações Seriadas. *ISKO Brasil*, (8).

Brasil. Ministério da Ciência, Tecnologia e Inovação (MCTI), & Centro de Gestão e Estudos Estratégicos (CGEE). (2025). *IA para o bem de todos: Plano Brasileiro de Inteligência Artificial*. Brasília, DF: MCTI; CGEE.

[https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/plano-brasileiro-de-inteligencia-artificial-pbia-\\_vf.pdf](https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/plano-brasileiro-de-inteligencia-artificial-pbia-_vf.pdf)

Bridges, L. M., Lopezosa, C., Velilla, M. C., & Ferran, N. (2025). Moving towards critical AI literacy. *The Electronic Library*, 43.

Carroll, S. et al. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, v. 19, p. 1-12.

Carvalho-Segundo, W. L. R., Pinto, A. L., do Canto, F. L., & Neubert, P. (2023). Projeto Laguna: infraestrutura de um lago de dados científicos em acesso aberto. *BiblioCanto*, 9(2), 133-138.

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

Cox, A. M. (2022). Artificial intelligence and libraries. *Journal of the Association for Information Science and Technology (JASIST)*, 74.

Cox, A. M. (2022). How artificial intelligence might change academic library work: Applying the competencies literature and the theory of the professions. *Journal of the Association for Information Science and Technology*, 74, 367–380. <https://doi.org/10.1002/asi.24635>

Cox, A. (2024). Algorithmic Literacy, AI Literacy and Responsible Generative AI Literacy. *Journal of Web Librarianship*, 18, 93–110. <https://doi.org/10.1080/19322909.2024.2395341>

Cox, A. M., Kennan, M., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 68. <https://doi.org/10.1002/asi.23781>

Cox, A., & Mazumdar, S. (2022). Defining artificial intelligence for librarians. *Journal of Librarianship and Information Science*, 56, 330–340. <https://doi.org/10.1177/09610006221142029>

Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46, 299–316. <https://doi.org/10.1177/0961000613492542>

Gray, J. (2009). The fourth paradigm: Data-intensive science. *Microsoft Research*.

Hjørland, B. (2021). Information retrieval and knowledge organization. *Information*, 12.

Lima, J. A. de O. (2025). A new “sense of position”: from the logic of subject indexing to the geometry of embeddings. *Journal of Documentation*. <https://doi.org/10.1108/jd-07-2025-0204>

Lin, D. et al. (2020). The TRUST principles for digital repositories. *Scientific Data*, v. 7, n. 1, p. 1-5, 2020.

Marchionini, G. (2016). Information Science Roles in the Emerging Field of Data Science. *Journal of Data and Information Science*, 1, 1–6. <https://doi.org/10.20309/jdis.201609>

Marchionini, G. (2023). Information and data sciences. *Data & Information Management*, 7.

Mayernik, M. (2023). Data Science as an Interdiscipline: Historical Parallels from Information Science. *Data Sci. J.*, 22. <https://doi.org/10.5334/dsj-2023-016>

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

Moutinho, S. O. M., Martins, P., Alencar, D., & Coneglian, C. S. (2024). Ciência da Informação e Ciência de Dados: convergências interdisciplinares. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência Da Informação*. <https://doi.org/10.5007/1518-2924.2024.e99127>

Ortiz-Repiso, V., Greenberg, J., & Calzada-Prado, J. (2018). A cross-institutional analysis of data-related curricula in information science programmes: A focused look at the iSchools. *Journal of Information Science*, 44, 768–784. <https://doi.org/10.1177/0165551517748149>

Qin, J. (2020). Knowledge Organization and Representation under the AI Lens. *Journal of Data and Information Science*, 5, 17–13. <https://doi.org/10.2478/jdis-2020-0002>

Ridley, M., & Pawlick-Potts, D. (2021). Algorithmic Literacy and the Role for Libraries. In *Information Technology and Libraries*. <https://doi.org/10.6017/ital.v40i2.12963>

Rosenberg, A. et al. (2023). Plain language summaries: enabling increased diversity, equity, inclusion and accessibility in scholarly publishing. *Learned Publishing*, v. 36, n. 1, p. 38-45.

Saracevic, T. (1996). Relevance in information science. *JASIS*, 26.

Semeler, A. R., Pinto, A., & Rozados, H. B. F. (2019). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, 51, 771–780. <https://doi.org/10.1177/0961000617742465>

Shah, C. (2023). The past, the present, and the future of information and data sciences: A pragmatic view. *Data Inf. Manag.*, 7, 100028. <https://doi.org/10.1016/j.dim.2023.100028>

Tait, E., & Pierson, C. M. (2022). Artificial Intelligence and Robots in Libraries: Opportunities in LIS Curriculum for Preparing the Librarians of Tomorrow. *Journal of the Australian Library and Information Association*, 71, 256–274. <https://doi.org/10.1080/24750158.2022.2081111>

Virkus, S., & Garoufallou, E. (2020). Data science and its relationship to library and information science: a content analysis. *Data Technol. Appl.*, 54, 643–663. <https://doi.org/10.1108/dta-07-2020-0167>

Warner, J. (2001). W(H)ITHER Information Science?/! In *The Library Quarterly* (Vol. 71, pp. 243–255). <https://doi.org/10.1086/603263>

Wang, L. (2018). Twinning data science with information science in schools of library and information science. *J. Documentation*, 74, 1243–1257. <https://doi.org/10.1108/JD-02-2018-0036>

Wu, D., Xu, H., Sun, Y., & Lv, S. (2022). What should we teach? A human-centered data science graduate curriculum model design for iField schools. *Journal of the Association for Information Science and Technology*, 74, 623–640. <https://doi.org/10.1002/asi.24644>

## **PREPRINT SUBMETIDO À REVISTA CIÊNCIA ABERTA LUSÓFONA (ReCAL)**

Xia, J., & Wang, M. (2014). Competencies and Responsibilities of Social Science Data Librarians: An Analysis of Job Descriptions. *Coll. Res. Libr.*, 75, 362–388. <https://doi.org/10.5860/CRL13-435>

Zhang, Y., Wu, D., Hagen, L., Song, I., Mostafa, J., Oh, S., Anderson, T., Shah, C., Bishop, B., Hopfgartner, F., Eckert, K., Federer, L., & Saltz, J. (2022). Data science curriculum in the iField. *Journal of the Association for Information Science and Technology*, 74, 641–662. <https://doi.org/10.1002/asi.24701>

## Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.