Relatório final de atividades

Projeto: Savits - Sistema de Avaliação e Validação de Impactos em Tecnologias Sociais.

Coordenador do Projeto:	**Washington Luís Ribeiro de Carvalho Segundo**		
Nome do(a) pesquisador(a):	**Rafael Antunes dos Santos**		
Vigência da Bolsa de Pesquisa:	**01/06/2026**	**Até:**	**31/08/2026**
Encerramento em:	**31/08/2026**		
Ref. Fundação:	**1870454**	**Ref. Ibict:**	**32725**

Descrição das Atividades da Bolsa de Pesquisa:

Investigar e analisar de forma conceitual a consistência metodológica de descritores, vocabulários controlados e tesauros para o Caso Piloto, mapeando teoricamente a representatividade temática, a interoperabilidade semântica e a padronização informacional do projeto; Realizar o levantamento bibliográfico e a estruturação de guias conceituais, documentação FAIR (incluindo dicionários de dados) e manuais teóricos de biblioteconomia de dados, definindo os requisitos metodológicos de curadoria e preservação digital aplicáveis aos ativos informacionais e registros científicos; Mapear requisitos de compatibilidade, interoperabilidade sistêmica (BrCris, Oasisbr, BDTD, Bibliodata, dARK, ORCID, ROR) e critérios de identificação persistente, organizando o arcabouço conceitual de dados para subsidiar de forma científica as ações iniciais das Metas 1, 2 e 3.2. Elaborar relatório técnico final de bolsa, consolidando resultados, atividades desenvolvidas e contribuições metodológicas da pesquisa.

1 INTRODUÇÃO

A avaliação de políticas públicas no Brasil ocorre, na maior parte dos casos, depois do encerramento dos ciclos de implementação, quando os efeitos das ações já se consolidaram e as possibilidades de correção de rumo se estreitaram. Esse caráter reativo se agrava pela fragmentação das informações em silos institucionais, o que dificulta a leitura integrada dos fenômenos sociais sobre os quais as políticas incidem. O Sistema de Avaliação e Validação de Impactos em Tecnologias Sociais (SAVITS), iniciativa do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict/MCTI) conduzida no âmbito da Coordenação-Geral de Informação Científica e Técnica (CGIC), propõe um deslocamento desse modelo em direção à análise preditiva em tempo real. Para tanto, articula três camadas tecnológicas complementares: o monitoramento contínuo de indicadores sociais por meio de *business intelligence*, a predição de riscos com modelos de aprendizado de máquina e a análise semântica de dados qualitativos em larga escala com grandes modelos de linguagem (IBICT, 2026a). A concepção do projeto teve origem em emenda parlamentar e resultou de processo colaborativo entre atores institucionais e acadêmicos, com execução organizada em núcleos técnicos especializados e caso piloto previsto para o Distrito Federal.

O objeto do SAVITS, a tecnologia social, constitui um campo conceitual em construção. A literatura latino-americana distingue o artefato do processo que o produz, de modo que um mesmo objeto técnico, como uma cisterna de concreto, só adquire o estatuto de tecnologia social quando sua construção e sua gestão são democráticas, conduzidas pela comunidade e ajustadas ao saber situado do território (IBICT, 2026b). Dessa distinção decorre a noção de replicabilidade adaptativa, segundo a qual não há solução universalizável a ser copiada, e sim experiências que cada território tem o direito de reinventar conforme suas condições. Dagnino (2019) inscreve essa discussão no quadro da tecnociência solidária e da adequação sociotécnica, referencial que orienta a concepção do SAVITS ao deslocar a avaliação de impacto do plano exclusivamente métrico para o da democratização do conhecimento e da aprendizagem coletiva. No plano institucional, o Ibict já mantém a plataforma *Cívis*, dedicada ao registro e à articulação de iniciativas de ciência cidadã no país, o que aproxima o instituto do universo de práticas que o SAVITS pretende avaliar.

Um sistema com essa ambição depende de fundações informacionais rigorosas. A plataforma adota como diretrizes de governança os princípios FAIR, TRUST, CARE e DEIA, em conformidade com a Lei Geral de Proteção de Dados (LGPD), e prevê arquitetura baseada em interoperabilidade e em identificadores persistentes, com destaque para o dARK (IBICT, 2026a). É nesse ponto que se situa a contribuição da biblioteconomia de dados ao projeto. O presente relatório consolida as atividades desenvolvidas pelo bolsista no Núcleo de Dados e Validação, cujas atribuições contratuais compreendem a análise conceitual de padrões de metadados e vocabulários

controlados, a produção de guias de curadoria e documentação orientados aos princípios FAIR, o mapeamento de interoperabilidade com infraestruturas científicas brasileiras e a elaboração deste relatório técnico final.

O trabalho dá continuidade ao primeiro relatório de entrega de atividades, que inventariou dezesseis serviços de informação do Ibict e os organizou em dois eixos, o da padronização e descrição bibliográfica, ancorado na Coordenação de Serviços Bibliográficos (COBIB), e o da agregação e do tratamento analítico e semântico, ancorado na Coordenação de Tratamento, Análise e Disseminação da Informação Científica (CODIC). Aquele exercício, incorporado ao presente documento como Anexo A, formulou a pergunta que esta pesquisa retoma e responde de forma ampliada. Como os serviços estudados, em especial os vinculados à COBIB, à CGIC e à CODIC, incluindo infraestruturas de inteligência artificial e repositórios como *Aleia*, *Pinakes* e RIDI, podem se articular de forma mais integrada com a futura plataforma SAVITS? Para responder a essa pergunta, definiram-se os seguintes objetivos específicos, derivados das atribuições contratuais da bolsa:

a) investigar e analisar de forma conceitual a consistência metodológica de descritores, vocabulários controlados e tesauros para o Caso Piloto;
b) mapear teoricamente a representatividade temática, a interoperabilidade semântica e a padronização informacional do projeto;
c) realizar o levantamento bibliográfico e a estruturação de guias conceituais, documentação FAIR, incluindo dicionários de dados, e manuais teóricos de biblioteconomia de dados;
d) definir os requisitos metodológicos de curadoria e preservação digital aplicáveis aos ativos informacionais e registros científicos;
e) mapear requisitos de compatibilidade e interoperabilidade sistêmica com as infraestruturas *BrCris*, *Oasisbr*, BDTD, *Bibliodata*, dARK, ORCID e ROR, além de critérios de identificação persistente;
f) organizar o arcabouço conceitual de dados para subsidiar de forma científica as ações iniciais das Metas 1, 2 e 3.2;
g) elaborar o relatório técnico final de bolsa, consolidando resultados, atividades desenvolvidas e contribuições metodológicas da pesquisa.

Dois marcos empíricos sustentam a análise. O primeiro é a matriz de requisitos informacionais construída a partir do documento de planejamento do projeto, com vinte e quatro requisitos rastreáveis a metas, etapas e atividades (Anexo B), complementada por uma nota técnica de confronto entre essa matriz e a estrutura de indicadores do projeto europeu MICS, dedicado à avaliação de impacto da ciência cidadã (Anexo C). O segundo é o *Workshop* de Integração do Projeto SAVITS, realizado de forma remota entre 21 e 24 de julho de 2026, que reuniu a equipe de execução em quatro oficinas temáticas e produziu, entre outros registros, um levantamento colaborativo de nove políticas públicas candidatas ao caso piloto no Distrito Federal. A convergência entre o

que a matriz antecipou por dedução analítica e o que o *workshop* confirmou pela via da construção coletiva constitui, ela própria, um dos resultados desta pesquisa.

O relatório está organizado conforme segue. A seção 2 descreve o percurso metodológico, de natureza documental e exploratória, com observação participante no *workshop* e apoio de ferramentas de levantamento bibliográfico assistido. A seção 3 apresenta os resultados em sete subseções, correspondentes aos objetivos específicos, e culmina na resposta consolidada à pergunta de pesquisa. A seção 4 reúne as considerações finais e os desdobramentos propostos para as etapas subsequentes do projeto. Completam o documento as referências e os anexos, entre os quais o primeiro relatório de entrega e a matriz de requisitos adaptada.

2 MÉTODO

A pesquisa que fundamenta este relatório é de natureza aplicada e abordagem qualitativa, com caráter exploratório e descritivo. Trata-se de uma investigação situada em contexto institucional e de projeto, que combina pesquisa documental, observação participante e análise de artefatos produzidos em oficina. A articulação de diferentes fontes e procedimentos qualitativos é metodologicamente pertinente para compreender fenômenos sociotécnicos complexos sob a ótica da complementaridade e do aprofundamento do contexto, permitindo ampliar a visão sobre o fenômeno investigado em vez de apenas buscar a confirmação de hipóteses únicas (Farquhar; Michels; Robson, 2020).

2.1 *Corpus* documental

O *corpus* reuniu três conjuntos de fontes. O primeiro compreende os documentos de planejamento e execução do próprio projeto, com destaque para o documento de planejamento denominado 5D, que detalha metas, etapas e atividades, e para o relatório consolidado de execução em sua versão 2.1 (IBICT, 2026a), que apresenta a fundamentação estratégica, o modelo de execução híbrido, a estrutura organizacional em núcleos técnicos e o processo de formação da equipe. O segundo conjunto abrange a documentação técnica das infraestruturas de informação do Ibict com potencial de articulação com o SAVITS, incluindo os padrões de metadados e o guia do usuário da plataforma *Aleia*, o guia do serviço *Deposita Dados*, a produção científica

sobre o grafo de conhecimento *BrCris* e o modelo *Pinakes*, além do organograma institucional. O terceiro conjunto é composto pelos registros do *workshop*, entre eles a síntese do primeiro dia (IBICT, 2026b), a apresentação sobre plataformas e catálogos de ciência cidadã e tecnologia social e o quadro colaborativo descrito na subseção 2.4. Os documentos foram tratados como fontes de contextualização e confronto analítico com os registros da observação e da oficina, reconhecendo-se que foram produzidos para finalidades institucionais próprias e exigem interpretação situada em relação ao seu contexto original de produção (Farquhar; Michels; Robson, 2020; Runeson; Höst, 2009)

2.2 Instrumentos analíticos

Dois instrumentos estruturaram a análise. O primeiro foi a matriz de requisitos informacionais (Anexo B), construída por extração sistemática a partir do documento 5D. Cada requisito identificado recebeu código próprio, descrição, vínculo rastreável às metas, etapas e atividades do projeto, indicação dos princípios de governança aplicáveis (FAIR, TRUST, CARE e DEIA) e classificação de prioridade. O resultado consolidou vinte e quatro requisitos, organizados em planilha eletrônica com recursos de filtragem e contagem que permitem leituras transversais por atribuição contratual e por princípio. A matriz cumpre função estruturante no conjunto das entregas, pois oferece a referência comum contra a qual os demais produtos foram confrontados.

O segundo instrumento foi a nota técnica de confronto entre a matriz e a estrutura de indicadores do MICS (*Measuring the Impact of Citizen Science*), projeto europeu de medição de impacto de iniciativas de ciência cidadã financiado pelo programa *Horizon 2020* sob o acordo de subvenção nº 824711 e executado entre 2019 e 2022. A íntegra do confronto encontra-se no Anexo C deste relatório. Cabe registrar que o acrônimo coincide com o de um programa de pesquisas domiciliares da UNICEF (*Multiple Indicator Cluster Surveys*), sem relação com o projeto europeu aqui referido. O confronto estabeleceu correspondências entre seis requisitos da matriz e os indicadores do MICS, identificou três lacunas e funcionou como confronto analítico externo e exploratório do instrumento, na medida em que examinou sua aderência e complementaridade em relação a um quadro internacional consolidado de avaliação.

2.3 Observação participante no *workshop*

O *Workshop* de Integração do Projeto SAVITS ocorreu de forma remota entre 21 e 24 de julho de 2026, organizado em quatro oficinas temáticas dedicadas, respectivamente, à apresentação e ao alinhamento institucional, à caracterização dos

núcleos e papéis da equipe, às ferramentas de gestão e boas práticas de comunicação e ao levantamento de políticas públicas para o caso piloto. O bolsista participou das quatro sessões na condição de integrante do Núcleo de Dados e Validação, o que caracteriza a observação como participante. Os dados decorrentes dessa participação consistem em anotações de campo sobre as exposições da coordenação e dos convidados, na apresentação do próprio bolsista durante a sessão de núcleos e na percepção das discussões coletivas. Em pesquisas qualitativas situadas em ambientes de trabalho e projetos organizacionais, a posição de pesquisador-integrante (*insider-researcher*) favorece o acesso direto às práticas, linguagens e racionalidades locais, mas exige postura reflexiva sobre o envolvimento do pesquisador com o campo e sobre as dinâmicas relacionais com os demais atores (Costley; Elliott; Gibbs, 2010). Não houve gravação nem ata formal das sessões dos dias dois a quatro, de modo que o relato dessas etapas se apoia nos registros pessoais do pesquisador, cotejados, sempre que possível, com os materiais documentais disponíveis. Essa condição é assumida como limite metodológico: os registros de campo permitem uma interpretação contextualizada dos processos, mas não uma reconstituição exaustiva das interações (Runeson; Höst, 2009).

2.4 Análise do quadro colaborativo de políticas públicas

Durante a quarta oficina, a coordenação conduziu uma dinâmica de levantamento estruturado em quadro digital colaborativo na plataforma *EasyRetro,* de acesso público. O quadro organizou nove políticas públicas candidatas ao caso piloto no Distrito Federal e, para cada uma, seis categorias de atributos: pontos fortes, desafios, oportunidades, dados e indicadores, parcerias e articulações, dúvidas e sugestões. Os participantes contribuíram com cartões anônimos e votos de priorização. Para fins de análise, o conteúdo do quadro foi exportado, e os cartões foram associados às respectivas políticas pela correspondência de numeração e codificação cromática, com conferência visual do documento original. O material foi então sintetizado em leitura transversal, com atenção à distribuição dos votos, à densidade informacional de cada categoria e à recorrência de temas entre políticas distintas. O quadro foi abordado como

artefato colaborativo e fonte empírica complementar. Na análise de processos de trabalho interdisciplinares, artefatos dessa natureza desempenham papel ativo na mediação da colaboração e na estabilização provisória de entendimentos coletivos, devendo sua interpretação considerar a situação prática e as finalidades específicas para as quais foram gerados (Nicolini; Mengis; Swan, 2012). Por se tratar de registro anônimo e coletivo, os cartões são tratados como manifestações agregadas da equipe, sem atribuição individual.

2.5 Levantamento bibliográfico e gestão de referências

O levantamento bibliográfico combinou a curadoria manual das fontes institucionais com buscas assistidas pela ferramenta *Undermind.ai*, orientadas por descrições de busca formuladas em inglês para cada subseção de resultados e para o próprio método. Os resultados recuperados passaram por triagem de pertinência conduzida pelo pesquisador. O uso de ferramentas de inteligência artificial restringiu-se ao apoio à recuperação e à organização da informação, permanecendo a seleção, a interpretação e a redação analítica sob responsabilidade do pesquisador.

3 RESULTADOS

Os resultados estão organizados em sete subseções, correspondentes aos objetivos específicos definidos na Introdução. As seis primeiras apresentam, em sequência, a análise da consistência metodológica de descritores, vocabulários controlados e tesauros para o Caso Piloto, o mapeamento da representatividade temática e da interoperabilidade semântica, a estruturação de guias conceituais e documentação FAIR, os requisitos de curadoria e preservação digital, os requisitos de interoperabilidade sistêmica e de identificação persistente e o arcabouço conceitual que subsidia as Metas 1, 2 e 3.2. A sétima subseção retoma a pergunta de pesquisa e a responde de forma consolidada, articulando os achados anteriores aos subsídios do *Workshop* de Integração e ao levantamento colaborativo de políticas públicas para o piloto no Distrito Federal. Ao longo do capítulo, os resultados da pesquisa bibliográfica são confrontados com os documentos do projeto e com as evidências produzidas durante o *workshop*, de modo que cada subseção combine fundamentação teórica e aplicação direta ao contexto do SAVITS.

3.1 Consistência metodológica de descritores, vocabulários controlados e tesauros para o Caso Piloto

A construção de vocabulários controlados para os domínios da tecnologia social, da inovação social, da inovação de base e da ciência cidadã enfrenta um problema anterior à escolha dos termos. Esses campos ainda disputam seus próprios limites conceituais. Na literatura brasileira sobre tecnologia social, convivem uma visão instrumental, que trata a tecnologia como produto ou ferramenta aplicada a problemas sociais, e uma visão sociotécnica, que a compreende como processo coletivo de transformação, articulado ao conhecimento local, à autogestão e à economia solidária (Duque; Valadão, 2017). Essa polissemia ultrapassa a dimensão terminológica, pois expressa posições políticas distintas sobre quem define o problema, quem produz o conhecimento e quem controla a solução (Bagattolli; Pezarico, 2025). Um tesauro que convertesse automaticamente essas formulações em sinônimos apagaria justamente a disputa que constitui o campo.

A própria forma gramatical pode carregar uma decisão teórica. Vasconcellos (2025) observa que "Tecnologia Social" no singular preserva a ideia de um projeto político-metodológico comum, enquanto "tecnologias sociais" no plural, tende a sugerir um conjunto de ferramentas. Essa distinção foi tematizada de modo independente durante o *Workshop* de Integração do Projeto SAVITS, na explanação conceitual do primeiro dia, que diferenciou as duas formas e as vinculou à oposição entre o artefato e o processo que o produz. O exemplo discutido foi o da cisterna de concreto, objeto técnico que só adquire o estatuto de tecnologia social quando sua construção e sua gestão são conduzidas de forma democrática pela comunidade e ajustadas ao saber do território (IBICT, 2026b). A convergência entre a revisão de literatura e a exposição conceitual do *workshop* indica que a distinção deve ser registrada no vocabulário como escolha de uso, com notas de escopo, e não resolvida por normalização mecânica de plural para singular.

O mesmo tratamento relacional vale para os conceitos vizinhos. Tecnologia social e inovação social compartilham a orientação para problemas coletivos, mas Medeiros *et al*. (2017) distinguem a primeira, mais local e tangível, da segunda, de alcance micro, meso ou macro e associada também a conhecimentos intangíveis. A inovação social digital acrescenta outra camada, pois pode ser modelada como rede organizacional apoiada por tecnologias de informação e comunicação cujos meios e fins são sociais, o que a afasta de uma sinonímia simples com tecnologia social (Rodrigo; Palacios; Ortiz-Marcos, 2019). No vocabulário do Caso Piloto, esses conceitos devem ser conectados por relações associativas e notas de escopo que explicitem as diferenças de alcance e de natureza, reservando as relações de equivalência para variantes efetivas de um mesmo conceito.

A garantia literária do domínio pode se apoiar em fontes já consolidadas no país. Catálogos institucionais de tecnologias sociais, como os elaborados pela Universidade Federal Fluminense (Patrício *et al*., 2024), selecionaram iniciativas por meio de chamadas públicas, formulários, análise documental, entrevistas e visitas técnicas. A partir disso, registraram-se produtos, processos, serviços e metodologias que atendessem aos critérios de reaplicabilidade, interação com comunidades e orientação à transformação social. Esse procedimento combina a garantia literária extraída de bibliografia, políticas públicas e documentos institucionais com a evidência empírica das iniciativas e com a garantia do usuário. Ele também demonstra por que a oposição entre artefato e processo precisa estar explícita na estrutura do vocabulário, uma vez que uma mesma entrada pode descrever o produto, a metodologia de sua produção, a prática comunitária que o sustenta e os resultados esperados. No contexto do SAVITS, essa base de garantia se amplia com o *Transforma!*, do Banco do Brasil, com o *Observatório de Tecnologias Sociais* do Museu Paraense Emílio Goeldi, com a atuação da ABEPETS e, em posição privilegiada, com a plataforma *Cívis*, mantida pelo próprio Ibict, cujo diretório de iniciativas de ciência cidadã constitui fonte direta de termos em uso pelas comunidades de prática.

A ciência cidadã torna ainda mais visível a dimensão política da terminologia. Eitzel *et al*. (2017) demonstram que termos como cidadão, voluntário, amador e parte interessada incluem certos participantes, excluem outros e sugerem relações hierárquicas distintas entre ciência institucional e público. Hacking, Lewis e Evans (2024) mostram que ciência cidadã e ciência comunitária carregam origens e projetos epistemológicos parcialmente diferentes, com a primeira enfatizando a contribuição pública para agendas científicas institucionais e a segunda, a produção de conhecimento pela própria comunidade e para ela. Um vocabulário metodologicamente consistente conserva ambos como conceitos relacionados e distintos, acompanhados de notas históricas, políticas e de aplicação. Essa sensibilidade terminológica conversa diretamente com os princípios CARE e DEIA adotados pelo projeto, pois a escolha dos rótulos define, na prática, quem se reconhece nos registros da plataforma. O levantamento colaborativo de políticas públicas realizado na quarta oficina do *workshop* oferece uma amostra concreta dessa exigência, já que a política mais aderente ao perfil institucional do Ibict, dedicada às práticas de ciência aberta e ciência cidadã, precisará dialogar com públicos tão distintos quanto comunidades escolares, usuários da assistência social e populações migrantes do Sul Global.

Quanto à operacionalização, a taxonomia em ciência cidadã proposta por Silva *et al*. (2024) exemplifica a estratégia adequada, pois combinou na busca bibliográfica termos como *citizen science*, *community science, participatory science e co-created science*, organizando os resultados em facetas e relações hierárquicas que preservam as distinções. A norma ISO 25964-1 fornece o quadro de referência para representar conceitos, termos preferidos, variantes, relações hierárquicas, associativas e notas de escopo (ISO, 2011). A manutenção, por sua vez, deve ser tratada como ciclo contínuo de

identificação de novos termos, avaliação de equivalências, revisão de relações, validação especializada, registro histórico e atualização (Ferreira; Maculan, 2020). Para a publicação na web semântica, a conversão a SKOS representa os conceitos com URIs e propriedades de rotulagem e relacionamento, ampliando a interoperabilidade, desde que as decisões de mapeamento e as extensões adotadas sejam documentadas (Mochón Bezares; Méndez Rodríguez; Bueno de la Fuente, 2020). A experiência de transformação de listas planas de termos em vocabulários SKOS mostra ainda que a validação das relações se beneficia de entrevistas e grupos focais com os usuários do domínio (Firmino; Baptista, 2013), procedimento que, no SAVITS, pode aproveitar as instâncias participativas já previstas no modelo de execução. No plano dos dados, Simonis (2018) acrescenta que a consistência semântica de registros de ciência cidadã requer identificadores persistentes e URIs resolvíveis para propriedades observadas, procedimentos e metadados, exigência que se alinha à adoção do dARK prevista na arquitetura do projeto e aos requisitos de identificação persistente da matriz informacional.

Desses elementos resulta uma diretriz de consistência metodológica para o Caso Piloto. Consistência, aqui, não equivale à imposição de uma terminologia única a um campo em disputa. Ela consiste em separar conceito e rótulo, distinguir equivalência de associação, registrar proveniência e contexto de cada decisão terminológica, combinar as garantias literária, institucional e do usuário e submeter o sistema a revisão conforme o domínio evolui. O estudo de Moraes (2012) sobre as representações sociais da tecnologia social no Brasil reforça essa orientação ao evidenciar que os significados atribuídos ao campo variam entre a comunidade científica, o Estado e as organizações sociais. Um vocabulário adequado ao Sul Global preserva essas assimetrias culturais e políticas em vez de tratar as traduções entre português, espanhol e inglês como equivalências perfeitas. A interoperabilidade desejável para o SAVITS é a que torna as diferenças visíveis e documentadas. As decisões estruturais aqui propostas, notas de escopo para as formas singular e plural, relações associativas entre os conceitos vizinhos, facetas para artefato e processo e ancoragem em identificadores persistentes, preparam o terreno para o mapeamento de representatividade temática e interoperabilidade semântica desenvolvido na subseção seguinte.

3.2 Representatividade temática, interoperabilidade semântica e padronização informacional

A literatura sobre sistemas de informação de pesquisa converge para uma conclusão que orienta toda esta subseção. A interoperabilidade semântica resulta da combinação entre um modelo contextual comum, perfis de aplicação, vocabulários controlados, mapeamentos explícitos e mecanismos de validação, e nenhum desses componentes a produz isoladamente, tampouco a adoção de um suposto esquema

universal. Nesse arranjo, o CERIF (*Common European Research Information Format*) ocupa posição de referência para representar pessoas, organizações, projetos, resultados e infraestruturas, enquanto as diretrizes *OpenAIRE* operacionalizam princípios semelhantes em escala europeia, por meio de orientações para repositórios e sistemas CRIS e de um grafo que integra resultados, fontes, financiadores e organizações (Jörg, 2010; Houssos *et al*., 2014; Jeffery *et al*., 2014; Manghi *et al*., 2019). Antes de percorrer esses componentes, convém fixar a distinção terminológica que este relatório adota. Vocabulário controlado é o instrumento que normaliza os termos de indexação, conforme discutido na subseção 3.1. Perfil de aplicação é o conjunto de regras que seleciona, restringe e documenta os elementos de metadados de um ou mais padrões para um contexto de uso determinado. *Crosswalk* é o mapeamento explícito entre elementos de esquemas distintos. A precisão no uso dessas três categorias evita ambiguidades recorrentes na documentação técnica e estrutura os requisitos correspondentes da matriz informacional do projeto.

A principal contribuição do CERIF está na separação entre entidades-base e entidades de ligação. Estas últimas registram relações qualificadas por papel, classificação e intervalo temporal. Isso evita reduzir informações como autoria, participação institucional ou financiamento a atributos simples, enquanto uma camada semântica mantém os significados desses papéis e permite associá-los a esquemas de classificação externos (Jörg, 2010; Jeffery *et al*., 2014). As comparações com a plataforma *VIVO* mostram, contudo, que a interoperabilidade exige mais do que equivalências nominais entre modelos. O *VIVO* organiza suas especializações em hierarquias OWL, ao passo que o CERIF recorre a uma camada semântica relacional e relativamente plana, de modo que a conversão entre ambos é viável para as entidades centrais e sofre perdas em atributos, suporte multilíngue, herança e relações complexas (Lezcano; Jörg; Sicilia, 2012; Lezcano *et al*., 2013). A experiência brasileira oferece uma resposta prática a essa tensão. O *BrCris* adota o CERIF como metamodelo lógico e, para a camada semântica, um subconjunto da ontologia *VIVO*-ISF fundamentada na *Basic Formal Ontology*, acrescido de extensão local para o contexto nacional, com publicação dos dados como *Linked Open Data* e vinculação de pessoas a seus registros no *Lattes* e no ORCID e de organizações ao ROR (Carvalho Segundo *et al*., 2024). A arquitetura demonstra que as duas tradições podem operar em camadas complementares de um mesmo grafo, e é essa solução já validada em ambiente de produção que o primeiro relatório de entrega identificou como referência conceitual mais próxima do SAVITS.

Os estudos sobre CRIS nacionais e regionais mostram que os perfis de aplicação traduzem necessidades políticas e administrativas em regras concretas de metadados. No espaço flamengo, um perfil baseado no *DataCite* foi adaptado para incluir disciplina de pesquisa, palavras-chave obrigatórias, identificadores ORCID, vínculos com projetos e publicações e indicadores de abertura e de aderência aos princípios FAIR, com transmissão ao sistema nacional em CERIF-XML e regras de obrigatoriedade e validação acordadas entre as instituições participantes (Neyens; Vancauwenbergh, 2021; Vancauwenbergh, 2021). A experiência austríaca acrescenta a dimensão operacional, ao procurar ligar sistemas financeiros, agências de fomento e universidades sob o princípio da coleta única, esbarrando na customização local, nas diferenças contábeis e na necessidade de limpeza dos dados (Harbich *et al*., 2023). A padronização depende, portanto, de governança contínua, e a especificação técnica é apenas seu ponto de partida. O registro das definições, alternativas e argumentos mobilizados pelos diferentes atores, como propõe a linguagem RIDAL, preserva a memória dessas negociações (Riechert, 2017). Para o SAVITS, a lição se traduz em um caminho concreto. O perfil de aplicação da plataforma deve nascer da matriz de requisitos informacionais, que já vincula cada exigência às metas e atividades do projeto, e ser pactuado nas instâncias de validação dos núcleos, com registro das decisões e de suas justificativas, prática que o confronto com o projeto MICS iniciou em caráter exploratório.

Os *crosswalks* ampliam esse quadro ao permitir que esquemas distintos participem de ecossistemas comuns, e a literatura delimita com clareza seus alcances. Elementos básicos como título, identificador, descrição e criador tendem a ser mapeáveis entre DCAT, *Dublin Core*, *DataCite*, DDI e *Schema.org*, enquanto relações estruturais, direitos, proveniência e vocabulários disciplinares se mostram bem menos estáveis (Wu *et al*., 2023; Ohmann *et al*., 2023). O mapeamento entre as infraestruturas ECRIN e CESSDA evidencia que termos semelhantes podem esconder modelos diferentes, pois um sistema se orienta ao estudo e o outro, ao recurso, e a preservação do contexto, da granularidade e da relação entre o recurso e o estudo que o gerou é decisiva para evitar uma interoperabilidade apenas superficial (Ohmann *et al*., 2023). Essa lição interessa diretamente ao SAVITS, que precisará distinguir a iniciativa de tecnologia social, entendida como processo avaliado, dos conjuntos de dados e demais recursos que a documentam, mantendo explícita a relação entre ambos, em coerência com a faceta de artefato e processo estabelecida na subseção 3.1. No domínio do *software*, o vocabulário *CodeMeta*, baseado em *Schema.org*, demonstra a viabilidade de transportar metadados de pacotes de código para repositórios como *DataCite* e *Zenodo*, com cobertura dependente da qualidade dos metadados de origem (Boettiger, 2017), referência pertinente aos componentes computacionais que a plataforma produzirá e que a matriz já contempla. Por fim, iniciativas de registro FAIR de esquemas e mapeamentos atribuem identificadores persistentes, versões e interfaces de reutilização aos próprios *crosswalks* (Suominen; Kesäniemi; Koivula, 2023), diretriz que o esqueleto de *crosswalk* previsto entre as entregas subsequentes desta bolsa deverá incorporar desde o início, com versionamento e identificação persistente.

Na camada dos vocabulários e ontologias de suporte, o DDI-RDF mostra como a combinação de RDF, OWL, SKOS, *Dublin Core* e vocabulários estatísticos permite consultar variáveis distribuídas, verificar a consistência e alinhar conceitos a tesauros externos (Bosch; Mathiak, 2015). A escolha de classes e propriedades em perfis apoiados em *Dublin Core*, FOAF e SKOS deve considerar adoção, manutenção, escopo, identificadores persistentes e compatibilidade vocabular (Kneebone, 2020). O DCAT 2 ilustra a estratégia de manter um núcleo comum reutilizável e remeter as restrições específicas aos perfis de aplicação (Albertoni *et al*., 2024). O ecossistema do Ibict já pratica essa coexistência de padrões. A plataforma *Aleia* descreve seus conjuntos de dados em *Dublin Core* e DCMI *Terms*, com identificação por DOI e ORCID dos criadores, e articula a preservação com *PREMIS* e METS por meio do *Archivematica*, enquanto o *Deposita Dados*, apoiado na infraestrutura *Dataverse*, herda a capacidade de exportação em DDI, característica desse ambiente. Para o SAVITS, o reaproveitamento desses perfis institucionais reduz o custo de especificação e assegura, desde o projeto, a compatibilidade com os serviços que a subseção 3.5 examina do ponto de vista sistêmico.

Resta a camada dos grafos de conhecimento, na qual a representatividade temática se decide. Arquiteturas baseadas em RDF e *Linked Open Data* convertem fontes heterogêneas, desambiguam autores, ligam entidades e oferecem consulta estruturada (Leiva-Mederos *et al*., 2017; Fermoso García *et al*., 2022), e grafos científicos fragmentados podem ser federados por interfaces como *GraphQL*, ligando metadados, conteúdos e serviços sem centralizar todos os dados (Aryani *et al*., 2020; Haris *et al*., 2021). A integração de planos de gestão de dados ao grafo *OpenAIRE* mostra o valor de entidades tipadas e relações explícitas entre projetos, conjuntos de dados, custos e pesquisadores (Papadopoulou *et al*., 2023). Um alerta metodológico, porém, encerra essa camada. A igualdade de rótulos entre grafos é insuficiente para inferir sobreposição real, pois categorias homônimas podem abranger conjuntos de registros muito diferentes, o que exige análise baseada em instâncias e validação especializada (Ciuciu-Kiss; Garijo, 2024). O levantamento colaborativo do *workshop* torna esse alerta tangível para o Caso Piloto, na medida em que as políticas públicas candidatas mobilizam fontes tão heterogêneas quanto o Censo Escolar, o Censo SUAS, o *CadÚnico* e os microdados da pesquisa TIC Educação, cada qual com seus modelos, granularidades e vocabulários próprios. A representatividade temática do SAVITS dependerá, assim, de quatro movimentos articulados, sendo a definição de um núcleo comum por perfil de aplicação ancorado na matriz de requisitos, a adoção dos vocabulários controlados delineados na subseção anterior, a documentação versionada dos *crosswalks* entre as fontes e a validação empírica, em nível de instância, da correspondência entre categorias. A subseção seguinte trata do instrumento que torna essas decisões transparentes e reutilizáveis, a documentação orientada aos princípios FAIR.

3.3 Guias FAIR, dicionários de dados e biblioteconomia de dados

A documentação orientada aos princípios FAIR integra a curadoria desde a origem do dado, e tratá-la como anexo tardio ao depósito compromete a reutilização que esses princípios prometem. Para que os dados sejam localizáveis e reutilizáveis, os metadados gerais precisam vir acompanhados de contexto, estrutura legível por máquina, proveniência, condições de acesso e limitações declaradas. A leitura dos princípios FAIR como adequação ao uso implica que a documentação responda às necessidades de comunidades e tipos de dados concretos, de modo que indicadores quantitativos de aderência informem o processo sem substituir o julgamento sobre a capacidade real de um usuário compreender e empregar os dados (Bishop; Hank, 2018). Essa orientação encontra respaldo na produção institucional recente do próprio Ibict, que sistematiza a tipologia dos metadados em descritivos, estruturais, administrativos, de proveniência e de contexto integrado. Esta última categoria está dedicada aos aspectos científicos do fenômeno documentado e aos métodos de criação, coleta, tratamento e análise (Lima; Sunye, 2025). É justamente essa camada de contexto integrado que os esquemas genéricos deixam descoberta e que os instrumentos discutidos nesta subseção procuram suprir.

No plano das práticas profissionais, a principal tensão dos repositórios institucionais está entre plataformas generalistas e dados que requerem descrição especializada. As equipes acumulam papéis de curador, especialista em metadados, gestor, desenvolvedor e interlocutor disciplinar. A comunicação com os pesquisadores, somada ao conhecimento da prática de pesquisa, pesa tanto quanto as competências técnicas (Lee; Stvilia, 2017). Na rotina, os bibliotecários coletam contexto junto a pesquisadores, publicações e propostas de projeto, recorrendo a arquivos *README*, dicionários de dados e pacotes documentais para preencher as lacunas dos esquemas genéricos. Essa limitação tem caráter estrutural. Em uma amostra internacional de repositórios institucionais, predominou o *Dublin Core*, o padrão DDI esteve ausente e as orientações de depósito mencionavam dados em pouco mais da metade dos casos (Nicholson; Bennett, 2021), quadro que reforça a necessidade de requisitos explícitos para métodos, variáveis e conceitos. A garantia de qualidade, por sua vez, opera como um ciclo contínuo de avaliação, intervenção e comunicação entre provedores, curadores e usuários (Stvilia; Lee, 2024). No SAVITS, esse conjunto de achados delimita com precisão o papel do bibliotecário de dados no Núcleo de Dados e Validação, ao qual cabe desenhar os instrumentos documentais, mediar a relação entre depositantes e a plataforma e sustentar o ciclo de qualidade.

Os serviços de dados sociais oferecem o modelo de maior granularidade documental, no qual o padrão DDI, os questionários, os códigos de variáveis, os rótulos de valores e os guias do usuário integram o próprio objeto curado. A experiência do *UK Data Service* com o repositório *ReShare* mostra que a transferência de parte do trabalho de depósito ao pesquisador preservou a revisão profissional de metadados, formatos,

completude e risco de divulgação (Eynden; Corti, 2017). No arquivo ICPSR, maior intensidade curatorial, o maior número de termos de assunto e a presença do texto das perguntas estiveram associados ao maior uso dos conjuntos de dados, ainda que sem demonstração de causalidade (Hemphill *et al*., 2022). Já a análise de registros de trabalho torna as ações curatoriais comparáveis em escala (Lafia *et al*., 2021) enquanto as ontologias de processos ajudam a coordená-las entre domínios (Minamiyama *et al*., 2024). Para variáveis de pesquisas sociais, vocabulários controlados dedicados, como o desenvolvido pelo GESIS, distinguem alterações de nome, rótulo, pergunta e resposta entre ondas de um estudo, apoiando a harmonização, a navegação e a construção de grafos de conhecimento (Estevao; Klas, 2024). O Caso Piloto do SAVITS se beneficia diretamente desse repertório, pois as políticas públicas levantadas no *workshop* mobilizam justamente séries estatísticas longitudinais, como o Censo SUAS, coletado desde 2007, e o Censo Escolar, cujas variáveis mudam de nome, escala e enquadramento ao longo dos anos. Documentar essas variáveis no nível do estudo e da variável, com a rastreabilidade de suas mudanças, é condição para que a camada preditiva da plataforma opere sobre séries harmonizadas em vez de agregados incomensuráveis.

Os planos de gestão de dados antecipam essas decisões documentais para o início do ciclo de vida. A evidência sobre seus efeitos ainda é limitada, e a evolução dos planos narrativos para artefatos acionáveis por máquina, os maDMPs, concentra a agenda atual (Vitale; Moulaison-Sandy, 2019). Avaliações de planos europeus identificaram cobertura apenas parcial dos princípios FAIR, com deficiências concentradas em metadados e interoperabilidade (Henning *et al*., 2021). A ontologia DCSO estrutura os conceitos de maDMP e seus vínculos com vocabulários controlados (Cardoso *et al*., 2022), enquanto a análise automática de planos ganha eficiência ao custo de granularidade em comparação com avaliações manuais (Pham *et al*., 2023). No contexto do projeto, os protocolos e formulários de uso de dados anunciados na abertura do *workshop* convergem com essa agenda, e o guia do serviço *Deposita Dados*, que já incorpora as definições da LGPD ao fluxo de depósito, oferece o ponto de partida institucional para um modelo de plano de gestão de dados do SAVITS, um entregável que pode nascer estruturado, com campos controlados e validação, em vez de puramente narrativo.

A verificação da aderência aos princípios FAIR dispõe hoje de modelos de maturidade que a tornam comparável e priorizável (Bahim *et al*., 2020), de implementações que combinam autoavaliação e testes automatizados (Devaraju *et al*., 2021; Wilkinson *et al*., 2019), de listas de verificação manuais com níveis de implementação (Phegley; Kellam, 2025) e de modelos de metadados comunitários convertidos em formulários e validadores acionáveis por máquina, com o cuidado de que as exigências não sobrecarreguem os depositantes (Musen *et al*., 2022). A automação, contudo, só verifica o que é tecnicamente visível, e estudos com dados eleitorais mostram pontuações automáticas superiores para a localizabilidade e inferiores para a interoperabilidade e a reusabilidade quando comparadas à revisão

manual (Zenk-Möltgen, 2025). O Ibict já acumulou experiência própria nesse terreno, pois os dezesseis conjuntos de dados do repositório *Aleia* foram avaliados com a ferramenta automática *F-UJI* entre janeiro e maio de 2024, em uma escala de aderência que vai de incompleto a avançado, e o próprio relatório dessa avaliação está depositado como conjunto de dados aberto na plataforma. O precedente recomenda ao SAVITS um protocolo combinado, com avaliação automática periódica para acompanhamento e revisão manual orientada por lista de verificação para as dimensões que a máquina avalia de forma limitada.

Sistemas de organização do conhecimento de campos de pesquisa variam consideravelmente em escopo, estrutura, curadoria e ligações externas (Salatino *et al*., 2024), avaliações da hierarquia de conceitos do *OpenAlex* identificam problemas semânticos e hierárquicos em ramos específicos (Stankovic; Sun; Wemp, 2025) e estudos de classificações geradas por máquina apontam os limites dessas classificações para a descoberta de conhecimento (Zafar, 2025). A arquitetura recomendável combina essas classificações amplas, na condição de camada auxiliar, com os vocabulários disciplinares curados discutidos na subseção 3.1 e com a documentação local. O bolsista experimentou essa articulação em artefato próprio, o *Delinéia*, no qual a arquitetura de tópicos hierárquicos do *OpenAlex* operou como dicionário de dados de referência para a busca de literatura assistida por inteligência artificial e para o retorno de sugestões focadas no delineamento da pesquisa. A experiência ilustra, em escala reduzida, o desenho documental que esta subseção propõe para o SAVITS, com a classificação ampla orientando a recuperação, o vocabulário curado garantindo a precisão e a documentação local preservando o contexto. O Quadro 1 sintetiza a leitura transversal da literatura examinada, organizada em seis dimensões que estruturam as recomendações desta subseção.

Quadro 1 - Leitura transversal da literatura sobre documentação FAIR e curadoria

Dimensão	**Estudos centrais**	**Síntese**
Repositórios institucionais	Lee e Stvilia (2017); Nicholson e Bennett (2021); Stvilia e Lee (2024)	Serviços generalistas precisam combinar metadados mínimos com documentação contextual, mediação e garantia de qualidade.
Serviços de dados sociais	Eynden e Corti (2017); Hemphill *et al.* (2022); Lafia *et al*. (2021); Estevao e Klas (2024); Zenk-Möltgen (2025)	A curadoria pode operar em nível de estudo e de variável, articulando documentação, controle de qualidade e reuso.
Planos de gestão de dados	Vitale e Moulaison-Sandy (2019); Henning *et al*. (2021); Cardoso *et al*.	O avanço dos planos narrativos para artefatos acionáveis por máquina

	(2022); Pham *et al*. (2023)	depende de estrutura semântica e
Aderência FAIR operacional	Wilkinson *et al*. (2019); Bahim *et al*. (2020); Devaraju *et al*. (2021); Musen *et al*. (2022); Phegley e Kellam (2025)	Indicadores e modelos tornam os princípios FAIR verificáveis, com interpretação humana e padrões definidos pelas comunidades.
Vocabulários e classificações	Estevao e Klas (2024); Salatino *et al*. (2024); Stankovic; Sun; Wemp (2025); Zafar (2025); Strecker (2026)	Estruturas controladas ampliam a interoperabilidade quando granularidade e adequação local são validadas.
Curadoria como trabalho sociotécnico	Bishop e Hank (2018); Lee e Stvilia (2017); Lafia *et al*. (2021); Stvilia e Lee (2024); Minamiyama *et al*. (2024)	A qualidade emerge da coordenação entre pessoas, processos, infraestrutura e normas, e a ferramenta é apenas um dos componentes.

Fonte: Dados da pesquisa assistida por *Undermind.ai*.

Dessa leitura decorre a estrutura dos guias conceituais propostos como entregáveis da segunda linha de trabalho da bolsa. O conjunto compreende um guia de curadoria FAIR para o SAVITS, um modelo de dicionário de dados com documentação no nível da variável e rastreabilidade das mudanças entre séries, um padrão de *README* e *codebook* para os conjuntos de dados do Caso Piloto e um modelo de plano de gestão de dados estruturado, todos ancorados nos requisitos correspondentes da matriz informacional e alinhados aos perfis institucionais da *Aleia* e do *Deposita Dados*. A observação transversal de que a granularidade se perde nos mapeamentos entre esquemas disciplinares e genéricos (Strecker, 2026) conecta esta subseção à seguinte, pois a preservação digital, tema da subseção 3.4, depende de que o contexto documentado aqui sobreviva aos fluxos entre sistemas ao longo do tempo.

3.4 Curadoria e preservação digital

A curadoria digital de dados sociais sensíveis constitui um processo contínuo de planejamento, proteção, documentação, preservação e mediação do acesso, e reduzi-la à anonimização seguida de depósito de um repositório desfigura sua natureza. Para uma plataforma dedicada a tecnologias sociais comunitárias, esse processo precisa conciliar três objetivos que entram em tensão com frequência: a proteção de pessoas e grupos vulneráveis, a preservação do valor contextual dos dados e a viabilização de formas

responsáveis de reutilização científica e social. O SAVITS parte, nesse terreno, de uma vantagem institucional concreta. O Ibict já opera uma pilha de preservação alinhada ao modelo de referência OAIS por meio da plataforma *Aleia*, cujos pacotes são processados pelo *Archivematica* com metadados de preservação em *PREMIS*, empacotamento estrutural em METS e descrição em *Dublin Core*, de modo que os requisitos de preservação registrados na matriz informacional do projeto encontram uma implementação de referência em casa.

A proteção deve ser incorporada desde a criação dos dados. Pesquisadores que atuam em campo com populações vulneráveis recorrem com frequência a soluções improvisadas, com pouco apoio institucional, o que fragiliza a segurança e pode inviabilizar o compartilhamento posterior, razão pela qual o planejamento do consentimento, a escolha de ferramentas seguras, a definição de responsabilidades, a previsão de preservação e a consulta prévia a especialistas em dados, segurança e ética precisam anteceder a coleta (Karcher; Secen; Weber, 2023). O modelo de ciclo de vida proposto para os dados das Ciências Sociais reforça essa orientação ao situar o planejamento no centro de etapas conectadas de coleta, organização, descrição, garantia, preservação, descoberta, reutilização e análise (Rodrigues *et al*., 2023). Essa centralidade do planejamento conecta a presente subseção ao modelo de plano de gestão de dados proposto na subseção 3.3 e aos protocolos e formulários de uso de dados anunciados na abertura do *workshop*, que operacionalizam, no fluxo do projeto, o momento em que essas decisões devem ser tomadas.

Sob a Lei Geral de Proteção de Dados, a gestão abandona a lógica da retenção ilimitada. Privacidade, transparência e visibilidade atravessam todo o ciclo, com coleta fundada em finalidade legítima, armazenamento seguro, acesso autenticado e registrado e descarte ao término da finalidade ou mediante solicitação cabível do titular (Bagatini; Guimarães; Sant'Ana, 2021). Para os repositórios institucionais, a adequação exige mapear fluxos de dados e metadados, identificar bases legais, aplicar a minimização, rever políticas, capacitar equipes e publicar informações claras sobre tratamento, segurança e direitos dos usuários (Nascimento; Silva, 2023). A legislação brasileira ainda guarda lacunas específicas para os arquivos de ciência, sobretudo na definição das responsabilidades de custódia e preservação, e a LGPD oferece parâmetros relevantes sem substituir uma política arquivística e institucional própria (Costa; Roncaglio, 2020). No ecossistema do Ibict, o guia do serviço *Deposita Dados*[1] já traduz as definições da lei para o fluxo de depósito de dados de pesquisa, e esse instrumento pode servir de matriz textual para a política de tratamento de dados pessoais do SAVITS, poupando a redação do zero e assegurando coerência entre os serviços da instituição.

A adequação regulatória tampouco conduz à eliminação ou à abertura indiscriminada dos dados pessoais. Um modelo europeu de classificação demonstra que

1Disponível em: https://depositadados.ibict.br/guide.xhtml;jsessionid=5b181447ab5e8af12ef693f21d77.

a reutilização sob o *General Data Protection Regulation* (GDPR)[2] pode ser organizada por níveis de sensibilidade, consentimento e risco, com dados não pessoais abertos e dados pessoais ou pertencentes a categorias especiais sujeitos a autenticação, permissões graduais, criptografia, separação de chaves e, nos casos críticos, impedimento de descarga (Noguera; Uroz; Juan, 2025). Essa lógica sustenta o princípio do acesso proporcional, no qual os metadados permanecem visíveis para promover a descoberta e a transparência, enquanto os dados primários ficam sob acesso condicionado. O diagnóstico coletivo do *workshop* confirma a centralidade desse princípio para o Caso Piloto. Na coluna de desafios do levantamento colaborativo, os cartões da equipe registram a presença de dados sensíveis de saúde de menores de idade, a necessidade de anonimização em massa, as restrições impostas pela LGPD e a tensão entre os prazos de aprovação pelo comitê de ética e o cronograma do projeto. As três políticas mais votadas dependem justamente de bases administrativas sensíveis, como o *CadÚnico*, cuja cessão de dados identificados segue rito formal próprio. O acesso proporcional permite que a plataforma cumpra sua vocação pública de transparência sem expor os titulares dos registros.

Para os dados qualitativos, narrativos e comunitários, que documentam a dimensão de processo das tecnologias sociais, a desidentificação é salvaguarda necessária e insuficiente quando isolada. Ela deve ser conduzida como análise contextual, atenta a combinações de eventos, lugares, relações, diagnósticos e experiências que permitam o reconhecimento por pessoas do círculo social do participante, para além de nomes e datas (*Campbell et al*., 2023). Estratégias de generalização, supressão seletiva e substituição de trechos por resumos analíticos reduzem riscos e, ao mesmo tempo, alteram a densidade interpretativa do material, de modo que a curadoria deve documentar as transformações realizadas e avaliar a utilidade remanescente para análise secundária (*Campbell et al*., 2023; Dubois *et al*., 2023). Fluxos automatizados de processamento de linguagem natural ampliam a escala desse trabalho e já identificam categorias contextuais como organizações, doenças raras, orientação sexual e eventos únicos, mantendo indispensável a revisão humana final (Gupta *et al*., 2021). A camada de análise semântica por grandes modelos de linguagem prevista na arquitetura do SAVITS pode ser mobilizada para apoiar essa triagem, desde que o desenho do fluxo consagre a revisão humana como etapa obrigatória. A meta realista, nesse domínio, consiste em reduzir o risco a um nível justificável combinando a modificação dos dados com controles de acesso. Já o modelo de referência para dados qualitativos é direto quanto ao critério, pois quando a remoção do contexto destrói o valor analítico do material, é preferível limitar quem acessa os dados, em quais condições e para qual propósito (Kirilova; Karcher, 2017).

2 Em português, a sigla é traduzida e conhecida como o Regulamento Geral sobre a Proteção de Dados, da União Europeia. Trata-se do principal marco regulatório e legal europeu para garantir a proteção, a privacidade e o tratamento seguro de dados pessoais de indivíduos.

Os repositórios precisam, portanto, de uma gradação explícita de acesso. A experiência da Universidade de Bristol distingue dados abertos, restritos, controlados e fechados, articulando avaliação de solicitantes, acordos de acesso, comitê de decisão, infraestrutura segura e transferência protegida (Hiom *et al*., 2017). O enquadramento *Five Safes* amplia essa lógica ao considerar simultaneamente a segurança dos dados, dos projetos, das pessoas, dos ambientes e dos resultados. Sua operacionalização envolve acordos de uso, análise de risco de divulgação, treinamento e revisão dos produtos analíticos. Há, contudo, o alerta de que tais exigências podem reproduzir desigualdades de acesso em prejuízo de pesquisadores sem vínculo institucional forte ou sem financiamento (Jang *et al*., 2023), ressalva que dialoga com o princípio DEIA assumido pelo projeto. Avaliações de impacto sobre a privacidade complementam o sistema ao mapear fluxos de informação, riscos para participantes e usuários e medidas técnicas e organizacionais de mitigação (Mhaidli *et al*., 2020). Deriva daí uma recomendação objetiva para o SAVITS, a formulação de uma política de acesso em níveis, inspirada nas quatro categorias de Bristol e no *Five Safes*, acompanhada de avaliação de impacto sobre a privacidade específica para o Caso Piloto antes do início da coleta.

Em projetos participativos e de tecnologia social, a governança dos dados ultrapassa a dimensão técnica. A curadoria pode ser centralizada sem que o poder decisório o seja, com participantes mantendo controle individual sobre usos futuros ou comunidades deliberando por meio de representantes e instâncias coletivas (Cooper *et al*., 2023). Este é um princípio de especial relevância quando a reutilização pode afetar grupos já expostos ao estigma, à discriminação ou à apropriação externa de conhecimento. Consentimento, acesso e benefício passam a ser tratados como compromissos de longa duração, revisáveis ao longo da vida dos dados, em vez de uma autorização única obtida no início da pesquisa. Essa concepção materializa os princípios CARE na camada de curadoria e converge com a orientação sociotécnica apresentada pela coordenação do projeto no *workshop*, que situa a participação das comunidades no registro da coprodução e a avaliação de impacto no horizonte da democratização e da aprendizagem coletiva.

Fecham o quadro os indicadores de preservação verificável. Para repositórios de dados de ciência cidadã, acessibilidade, autenticidade, integridade, interoperabilidade e sustentabilidade orientam a avaliação contínua da infraestrutura (Silva; Umpierre; Witt, 2023). O ciclo de vida dos dados nesse domínio já reconhece planejamento, documentação, garantia de qualidade, descoberta e preservação como componentes estabelecidos, com privacidade e preservação demandando atenção adicional (Shwe, 2020). E políticas claras de depósito, embargo, licenças, responsabilidades e tratamento de dados sensíveis dão previsibilidade a depositantes e usuários (Silveira; Vanz, 2023). Os requisitos metodológicos que esta subseção consolida para o SAVITS compreendem, assim, a herança da pilha de preservação da *Aleia* com OAIS, *Archivematica*, *PREMIS* e METS, a política de tratamento de dados pessoais derivada do guia do *Deposita Dados*, a política de acesso em níveis com avaliação de impacto à privacidade, a governança

participativa de longa duração alinhada aos princípios CARE e a documentação de proveniência das transformações curatoriais, para a qual o esquema *PROV-O*, já previsto na matriz de requisitos, fornece a representação formal. A persistência desses compromissos no tempo depende, contudo, de que os objetos preservados permaneçam identificáveis e localizáveis através das mudanças de sistemas e endereços, exigência que conduz aos identificadores persistentes e à interoperabilidade sistêmica tratadas na subseção seguinte.

3.5 Interoperabilidade sistêmica e identificação persistente

A interoperabilidade em infraestruturas nacionais de pesquisa depende da capacidade de conectar, atualizar e governar relações entre pessoas, organizações, produtos científicos, projetos e fontes de financiamento. A atribuição de identificadores persistentes é condição necessária para esse arranjo, mas, sem jamais esgotá-lo. DOI, ORCID, ROR e ARK operam como componentes de um grafo de informação de pesquisa, no qual os identificadores tornam entidades e vínculos legíveis por máquinas. A utilidade do grafo se decide nos metadados relacionais, nos vocabulários alinhados, nos serviços de resolução confiáveis e nos fluxos de trabalho que reduzem a duplicação de registros (Dappert *et al*., 2017). Essa formulação descreve com precisão o desafio do SAVITS, que precisará situar cada iniciativa de tecnologia social avaliada em uma rede de relações com pessoas, organizações, territórios, políticas públicas e conjuntos de dados, e a matriz de requisitos informacionais do projeto já reserva aos identificadores persistentes o papel de costura dessa rede.

Os sistemas CRIS podem operar como a camada de integração desse ecossistema. A experiência portuguesa do *PTCRIS* adota o ORCID como ponto de sincronização entre currículos, repositórios, fontes bibliográficas internacionais e sistemas nacionais, sob o princípio de registrar uma vez e reutilizar muitas vezes, o que reduz o retrabalho administrativo e mantém a consistência entre serviços por meio da importação de registros externos e da exportação seletiva de informações validadas (Moreira; Cunha; Macedo, 2015). A sincronização, contudo, convive com a necessidade permanente de curadoria, pois os metadados dos perfis ORCID podem chegar incompletos, duplicados ou conflitantes, o que exige regras sobre fontes prioritárias e validação pelos usuários ou pelas instituições. O princípio da coleta única reapareceu, aliás, na experiência austríaca discutida na subseção 3.2, e sua tradução para o SAVITS é direta, já que os atores das iniciativas avaliadas não devem ser cadastrados novamente a cada política ou ciclo avaliativo, mas resolvidos com base nos identificadores que o ecossistema nacional já mantém.

A experiência croata do *CroRIS* alarga a concepção do que um CRIS pode ser, ao posicioná-lo como elemento de ligação entre a infraestrutura digital nacional e os serviços de ciência aberta. Os vínculos entre pesquisadores, projetos, repositórios, periódicos e recursos computacionais sustentam desde o acesso a serviços de nuvem e

computação de alto desempenho até o depósito de dados e a integração com catálogos nacionais, com exposição e reutilização de metadados por APIs REST, OAI-PMH e semânticas derivadas do CERIF, dentro do país e em agregadores internacionais como o *OpenAIRE* (Udovičić *et al*., 2024). O caso também evidencia que os modelos padronizados precisam evoluir com os requisitos nacionais, pois a adesão semântica ao CERIF conviveu com adaptações técnicas sempre que surgiram objetos, relações ou processos não contemplados pelo modelo original. O paralelo brasileiro é imediato. O *BrCris* já resolve pessoas contra o *Lattes* e o ORCID e organizações contra o ROR, como visto na subseção 3.2, e os agregadores *Oasisbr* e BDTD expõem a produção nacional por OAI-PMH, de modo que o SAVITS não precisa construir essa camada de ligação, e sim conectar-se a ela, herdando identificadores e vínculos que o ecossistema do Ibict mantém em produção.

A padronização decide também os usos analíticos e avaliativos, dimensão que interessa de modo particular a uma plataforma de avaliação de impactos. Sistemas CRIS e agregadores nacionais permitem acompanhar a produção científica e o acesso aberto em escala institucional ou nacional, e os indicadores só se tornam comparáveis quando definições, regras de contagem e a qualidade dos dados são harmonizadas, pois diferenças sobre o que conta como publicação em acesso aberto, variações entre soluções técnicas e registros inseridos manualmente alteram substancialmente as métricas produzidas. Os identificadores ampliam a capacidade de integração e rastreamento e deixam intacta a tarefa de tornar os dados adequados para avaliação. Para o SAVITS, a lição converge com o que o confronto entre a matriz de requisitos e o projeto MICS antecipou, na medida em que os indicadores de impacto de tecnologias sociais exigirão definições pactuadas e regras de contagem explícitas antes de qualquer comparação entre políticas, territórios ou ciclos.

Os limites da automação confirmam essa cautela. Em duas instituições alemãs, a combinação de ROR, ORCID e DOI recuperou parte das informações de pesquisadores em fontes abertas, com cobertura fortemente variável entre instituições, disciplinas e bases, e mesmo onde a adoção interna do ORCID era alta, a recuperação externa por afiliação permaneceu incompleta, de modo que esses dados já enriquecem os CRIS locais e ainda dependem de revisão humana para sustentar relatórios oficiais (Schnieders *et al*., 2022). A identificação persistente precisa, assim, de governança de afiliações, controle de qualidade e mecanismos de proveniência. A advertência se estende à camada preditiva do SAVITS, cujos modelos consumirão registros ligados por identificadores, e a qualidade das inferências acompanhará a qualidade dessas ligações.

No contexto latino-americano, a interoperabilidade segue como agenda de consolidação. Um levantamento regional identificou sistemas CRIS em 31 das 133 instituições respondentes, todos com alguma interação com repositórios institucionais, e apenas uma pequena parcela declarou uso do CERIF ou de CERIF-XML para intercâmbio (Vázquez Tapia, 2022). A expansão dos repositórios, por si só, não produz uma

infraestrutura nacional integrada, e a consolidação requer políticas coordenadas, apoio técnico e padrões compartilhados. As desigualdades de acesso aos identificadores aprofundam o problema, pois a concentração da emissão de DOIs em grandes editoras do Norte Global limita a inserção de parcela relevante da produção dos países em desenvolvimento nos fluxos internacionais de descoberta, bibliometria e altmetria (Turki *et al*., 2023). Essa assimetria foi tematizada no quarto dia do *workshop*, na discussão sobre colonialismo de dados e custos de processamento de artigos, e reapareceu nos cartões do levantamento colaborativo que apontam a dependência de infraestruturas proprietárias como risco para o projeto. É nesse quadro que o dARK ganha sentido estratégico. A proposta implementa o esquema ARK de forma descentralizada, sobre consórcio permissionado, para reduzir custos, evitar pontos únicos de falha e distribuir a governança entre instituições, encontrando-se em estágio de prova de conceito, ainda sem evidência de impacto operacional em larga escala (Carvalho Segundo *et al*., 2023). A adoção do dARK prevista na arquitetura do SAVITS representa, portanto, uma resposta de soberania informacional ao diagnóstico do próprio *workshop*, e o estágio de maturidade da tecnologia recomenda que a política de identificadores do projeto preveja a convivência com identificadores estabelecidos e critérios de migração.

A confiança no ecossistema de identificadores exige, por fim, sustentabilidade financeira, transparência e planos de contingência, pois os identificadores persistentes são infraestruturas sociotécnicas cuja persistência depende tanto de instituições e políticas quanto de protocolos e resolutores (Castro *et al*., 2022). Dessa constatação deriva o entregável que esta subseção fundamenta, a minuta de política de identificadores persistentes do SAVITS, prevista entre as entregas subsequentes da bolsa, com definição dos tipos de identificador por classe de entidade, pessoas por ORCID, organizações por ROR, objetos digitais por DOI ou dARK conforme o caso, das responsabilidades de atribuição e manutenção, das regras de convivência e migração entre esquemas e dos compromissos de sustentação a longo prazo. Com os identificadores garantindo que entidades e vínculos permaneçam resolvíveis, o passo seguinte é reunir esses elementos no arcabouço conceitual que subsidia as metas iniciais do projeto, objeto da subseção 3.6.

3.6 Conceitos e requisitos para as metas 1, 2 e 3.2

O arcabouço conceitual que subsidia as ações iniciais das Metas 1, 2 e 3.2 do projeto repousa sobre dois instrumentos construídos durante a bolsa, sendo a matriz de requisitos informacionais e a nota técnica de confronto com o projeto MICS, aqui confrontados com a literatura internacional sobre avaliação de impacto. A matriz organiza vinte e quatro requisitos em nove categorias, que vão dos metadados e da descrição à capacitação, passando por semântica e vocabulários, identificadores persistentes, interoperabilidade, governança e ética, coleta e fontes, curadoria e preservação, acessibilidade e disseminação e reuso. Cada requisito recebe identificador

próprio, rastreabilidade explícita em relação às metas, etapas e atividades do documento 5D, indicação dos princípios e marcos associados, vínculo com as atribuições contratuais do bibliotecário de dados, prioridade e situação. Essa rastreabilidade bidirecional, do requisito ao planejamento e do planejamento ao requisito, é a propriedade que transforma a planilha em instrumento de governança em vez de inventário.

A literatura de avaliação de impacto confirma a orientação geral do instrumento e a enriquece. O impacto de iniciativas de ciência cidadã e tecnologias sociais escapa à contagem de participantes, dados recolhidos ou publicações, pois constitui uma cadeia causal e contextual na qual recursos e condições viabilizam atividades. As atividades produzem saídas, as saídas podem gerar resultados para participantes, instituições e territórios e apenas parte desses resultados se consolida como mudança de longo prazo. Modelos lógicos explicitam essa cadeia entre insumos, atividades, produtos, resultados e impacto e revelam as dependências técnicas, organizacionais e sociais que condicionam a intervenção (Ammenwerth; Bindel; Liebe, 2023). A modelagem de percursos de impacto em programas de ciência cidadã deve declarar pressupostos e elos causais, de modo que a avaliação identifique onde e por que um percurso foi interrompido, além de verificar se houve mudança (Skarlatidou; Haklay, 2021). O objeto avaliado é multidimensional, com impactos na sociedade, na economia, no ambiente, na ciência e tecnologia e na governança que se cruzam sem seguir trajetórias lineares (Wehn *et al*., 2021). A metodologia modular do projeto ACTION acrescenta o potencial transformador e permite selecionar as dimensões pertinentes a cada projeto (Passani *et al*., 2022) e o quadro aberto de avaliação separa as condições de processo dos resultados e impactos (Kieslinger *et al*., 2017). Dessa separação decorre um princípio que a matriz incorpora, o de que um indicador de atividade ou de produto não comprova, por si, uma alteração social, ambiental ou política.

A teoria de mudança deve, portanto, preceder a escolha de indicadores, ligando objetivos, públicos, mecanismos esperados, riscos e evidências necessárias. Os nove percursos propostos para a ciência cidadã mostram que dados mais numerosos, ricos e abertos, aprendizagem, capital científico, empoderamento e novas parcerias conduzem a benefícios públicos distintos, e a diversidade dos participantes opera como condição causal desses percursos, cuja ausência pode enviesar dados, restringir quem aprende ou se empodera e tornar decisões menos representativas (Pateman; West, 2023). A avaliação precisa registrar quem participa, quem fica de fora, quais barreiras existem e como essas desigualdades afetam benefícios e danos. Essa exigência conecta o princípio DEIA do projeto aos requisitos de instrumentos de coleta (RQ-13) e de mapeamento de fontes complementares (RQ-14), ambos rastreados à Meta 1, dedicada à compreensão do contexto e das necessidades.

Os indicadores, por sua vez, precisam combinar comparabilidade com adequação ao contexto. O inventário de produtos científicos da ciência cidadã oferece medidas concretas de produtos e práticas de dados, incluindo metadados,

compartilhamento, apoio à decisão e ação regulatória (Wiggins *et al*., 2018). A autoavaliação vinculada aos *Objetivos de Desenvolvimento Sustentável* (ODS) distingue a contribuição para o monitoramento formal dos indicadores da contribuição direta para o alcance das metas por ação, aprendizagem ou influência em políticas, com a ressalva de que metas globais exigem tradução para a escala, a linguagem e o horizonte temporal de cada projeto (Parkinson *et al*., 2022). Essa ressalva encontra eco preciso no confronto com o MICS. A plataforma europeia ancora seus indicadores nas metas 8.2, 9.4, 12.2, 12.7, 12.8 e 13.3, enquanto o SAVITS elege os ODS 3, 4, 8 e 10, e a interseção restrita ao ODS 8 indica que o caminho produtivo é a adaptação da arquitetura com substituição do conteúdo, aproveitando a forma madura do indicador e repovoando-a com metas pertinentes às tecnologias sociais. Cada indicador deve ainda ser definido com linha de base, fonte de evidência, momento de coleta, responsável, critério de qualidade e limite interpretativo, com métodos mistos combinando métricas, relatos, entrevistas, observação e evidências documentais (Schaefer *et al*., 2021; Wehn *et al*., 2021).

A avaliação é também um processo participativo. O modelo *Payback* adaptado à ciência cidadã incorpora o grau de participação externa em cada etapa, da contribuição de dados à cocriação, e vincula essa participação aos benefícios produzidos (Uude; Kurzhals; Wesbuer, 2022), enquanto o projeto *Step Change* distingue a avaliação analítica de cada iniciativa de uma camada transversal voltada à ancoragem social, transepistêmica e institucional da ciência (D'Andrea *et al*., 2022). Uma plataforma de avaliação, nessa concepção, apoia reflexão, adaptação e prestação de contas ao longo do projeto, além de recolher respostas. A orientação converge com a questão que a coordenação do SAVITS formulou no *workshop*, sobre se a participação se dará por consulta ou por coprodução, e o levantamento colaborativo de políticas públicas realizado na quarta oficina funcionou, nesse sentido, como primeiro exercício prático da ancoragem participativa que a plataforma pretende institucionalizar.

No plano do sistema de informação, o requisito central é a rastreabilidade entre documentos de planejamento, teoria de mudança, indicadores, fontes de dados, evidências e alegações de impacto, formando uma base de evidências auditável na qual cada indicador remete ao objetivo e à hipótese causal que o justificam e aos dados que sustentam ou limitam a conclusão. A plataforma MICS operacionaliza parte dessa arquitetura com seu instrumento de cerca de duzentas perguntas ponderadas, combinadas em indicadores por domínio, com relatórios, pontuações e recomendações (Sprinks *et al*., 2022). Para a etapa inicial de elicitação, a mineração de texto sobre documentos de política reduz o universo documental e liga intenções a tópicos, documentos e trechos relevantes, com interpretação e validação humanas indispensáveis (Massey *et al*., 2013), caminho que interessa ao Caso Piloto na medida em que a extração manual praticada sobre o documento 5D precisará escalar para o corpus normativo das políticas públicas candidatas, do Plano Decenal aos marcos da assistência social e da educação digital.

O confronto sistemático entre a matriz e o MICS, registrado em nota técnica de caráter exploratório e sujeita à validação da coordenação, produziu três ordens de resultados. A primeira é a constatação da ortogonalidade entre as camadas, pois os indicadores do MICS definem o que se mede, camada substantiva da avaliação, enquanto a matriz define como o que se mede deve ser descrito, identificado, curado, preservado e interoperado, camada infraestrutural, e é essa complementaridade que torna o contraste produtivo. A segunda ordem reúne as correspondências diretas, sintetizadas no Quadro 2. A terceira é o diagnóstico do que cada instrumento silencia. O MICS deixa fora de escopo os identificadores persistentes, os metadados de proveniência, a preservação de longo prazo, o consentimento como metadado e a acessibilidade, exatamente as camadas em que a matriz agrega valor e que delimitam a contribuição específica da biblioteconomia de dados ao projeto. O silêncio inverso também informa, pois a matriz nada afirma sobre a definição substantiva do impacto social, terreno em que a experiência europeia oferece repertório testado de indicadores e de métodos participativos. Cada instrumento cobre, assim, a lacuna constitutiva do outro, e a leitura conjunta antecipa uma plataforma em que avaliação e infraestrutura amadurecem de forma solidária.

Quadro 2 - Correspondências entre MICS e SAVITS

Elemento do MICS	**Requisito(s) da matriz**	**Natureza da relação**
Taxonomia de cinco domínios e dez indicadores	RQ-04	Fonte candidata de vocabulário controlado de indicadores
Anatomia documental dos indicadores	RQ-13	Modelo publicado de codebook, serializável em SKOS
Indicador de ODS no domínio da governança	RQ-04	Sobreposição parcial de metas, com adaptação de arquitetura
Página de projeto padronizada na plataforma	RQ-01, RQ-02	Captura de metadados acoplada ao instrumento de avaliação
Barreira de política pública por desconfiança nos dados	RQ-07, RQ-08	Validação externa da prioridade da interoperabilidade
Publicação aberta do instrumento sob licença CC BY-SA	RQ-20	Precedente de abertura de algoritmos e instrumentos
Indicador de sustentabilidade financeira	RQ-18	Convergência sobre sustentação de longo prazo
Pontuação por rede neural declaradamente inexplicável	RQ-12	Contraste que fortalece o requisito de auditabilidade

Fonte: Dados da pesquisa.

Em sentido inverso, o confronto revelou três lacunas na matriz, hoje registradas como candidatas a novos requisitos, a avaliação como evento informacional datado e versionado, sem o qual a série longitudinal da Meta 3 perde comparabilidade interna, a recomendação como objeto informacional com metadados de geração, exigência que a interpretação por modelos de linguagem prevista no projeto torna incontornável, e a documentação dos modelos preditivos na linha dos *model cards* e *datasheets*. O contraste de maior valor argumentativo está na camada de aprendizado de máquina, pois o MICS declara inexplicáveis suas pontuações por rede neural e as apresenta como estimativas aproximadas, postura que um sistema voltado à avaliação de políticas públicas com implicações distributivas dificilmente poderia acomodar, e o requisito de rastreabilidade e auditabilidade integral do fluxo informacional (RQ-12) ganha, com isso, justificativa comparativa concreta.

Dessa articulação resulta o arcabouço conceitual consolidado. A matriz opera como espinha dorsal de requisitos rastreáveis, os vocabulários controlados da subseção 3.1 fornecem a camada terminológica, os perfis de aplicação e *crosswalks* da subseção 3.2 asseguram a padronização, os instrumentos documentais da subseção 3.3 preservam o contexto, os protocolos de curadoria da subseção 3.4 protegem pessoas e dados e os identificadores da subseção 3.5 garantem a persistência das ligações, enquanto a camada substantiva dos indicadores de impacto, informada pelo MICS e pela literatura aqui examinada, define o conteúdo que essa infraestrutura sustentará. O respaldo externo que o exercício proporcionou, somado à convergência entre o que a matriz antecipou e o que o *workshop* confirmou em temas como grafos de conhecimento, governança e participação, indica que o arcabouço está maduro para subsidiar as ações iniciais das Metas 1, 2 e 3.2, restando à subseção seguinte demonstrar como ele responde à pergunta que originou esta pesquisa.

3.7 Articulação dos serviços do Ibict com a plataforma SAVITS

Esta subseção retoma a pergunta que originou a pesquisa, sobre como os serviços estudados, em especial os vinculados à COBIB, à CGIC e à CODIC, incluindo as infraestruturas de inteligência artificial e os repositórios *Aleia*, *Pinakes* e RIDI, podem se articular de forma mais integrada com a futura plataforma SAVITS. A resposta amadurecida ao longo das subseções anteriores se organiza em três planos complementares, o infraestrutural, o institucional e o político-epistêmico, e a literatura examinada adverte, de saída, contra a redução do problema ao primeiro deles. A integração de repositórios, serviços informacionais e grafos de conhecimento amplia a capacidade de avaliar políticas quando sustentada por arranjos institucionais duráveis, que combinam regras de acesso, curadoria, competências analíticas e mediação entre produtores de dados, analistas e decisores, e a interoperabilidade técnica, sozinha, não produz esse resultado.

No plano infraestrutural, a resposta consolida o fluxo que o primeiro relatório de entrega identificou e que as subseções anteriores detalharam. O eixo da COBIB padroniza e organiza a descrição bibliográfica, por meio da catalogação cooperativa do Bibliodata e da modelagem do Pinakes, enquanto o eixo da CODIC agrega e trata analítica e semanticamente, com os agregadores Oasisbr e BDTD e as competências em ciência de dados. Ao final da cadeia, o SAVITS acrescenta a camada avaliativa e preditiva, transformando registros padronizados e tratados em subsídio para decisão. Nessa divisão de trabalho, cada serviço assume papel definido. A *Aleia* e o *Deposita Dados* fornecem os perfis de metadados e a pilha de curadoria e preservação que as subseções 3.2 e 3.4 recomendaram herdar. O *BrCris* opera como camada de resolução de entidades, ligando pessoas, organizações e produtos a *Lattes*, ORCID e ROR, na função de elemento de ligação que a experiência croata demonstrou viável para um CRIS nacional. O dARK responde pela identificação persistente sob governança distribuída. O *Cívis* funciona como porta de entrada das iniciativas de ciência cidadã e fonte de garantia literária do vocabulário, papel que o levantamento colaborativo do *workshop* confirmou ao indicá-lo como apoio à seleção de práticas para o piloto. O RIDI, por fim, acolhe e dá visibilidade à produção técnico-científica institucional, embora seu escopo atual não contemple relatórios técnicos de projeto, tipologia em que se restringe aos relatórios de pesquisa de pós-doutorado. A restrição constitui, ela própria, um achado desta análise. A produção documental que o SAVITS gerará, como relatórios, guias de curadoria e notas técnicas, ainda não dispõe de destino institucional definido, e a ampliação das tipologias aceitas pelo RIDI desponta como caminho natural para que a cadeia aqui descrita se complete também no plano da preservação e do acesso. O SAVITS, nessa arquitetura, consome os serviços existentes e lhes devolve valor, sem duplicá-los.

O plano institucional é o que a literatura internacional ilumina com maior força. No sistema integrado de dados para a primeira infância de Iowa, as agências preservaram o controle sobre suas próprias bases, enquanto um conselho de governança, uma instância de custódia e equipes de implementação autorizavam e operacionalizavam os usos compartilhados, com comitês específicos aproximando pesquisa e prática para enfrentar a fragmentação entre educação, saúde e assistência social (Rouse *et al*., 2018). O levantamento de quarenta iniciativas norte-americanas de vinculação intersetorial de dados mostra que estruturas de governança, financiamento, quadros legais e comunidades de prática sustentam a produção de relatórios para formuladores de políticas, com ganho relatado na capacidade avaliativa e na confiança entre agências (Berkowitz *et al*., 2024), e o *Networked Data Lab* britânico articula equipes analíticas inseridas nos próprios sistemas de saúde e assistência, com prioridades locais, patrocínio de dirigentes e participação de usuários (Knight *et al*., 2024).

A tradução para o Caso Piloto é direta e encontra, na coluna de parcerias do levantamento colaborativo, seus atores já nomeados, das secretarias distritais de educação, ciência e tecnologia e desenvolvimento social aos CRAS e ao Ministério da

Saúde. A recomendação que decorre é a constituição de uma instância de governança de dados do piloto, na qual as agências custodiem suas bases e pactuem os usos, com os núcleos do projeto operando como equipes de implementação, desenho que evita tanto a centralização extrativa quanto a fragmentação que o SAVITS nasceu para superar.

A literatura acrescenta uma advertência que interessa à identidade profissional deste relatório. As plataformas informacionais funcionam como infraestruturas de conhecimento sustentadas por trabalho humano contínuo, e o arquivo nacional holandês DANS ilustra a dependência de arquivistas que curam metadados, convertem formatos, protegem dados sensíveis e apoiam o reuso, de modo que a abertura efetiva se realiza como prática de mediação e ultrapassa o depósito de arquivos (Borgman; Scharnhorst; Golshan, 2018). A avaliação da plataforma *GO-SPIN* da UNESCO, dedicada a informação sobre políticas científicas, identificou lacunas simultâneas de comunidade, sistema técnico e sustentabilidade, com pouca compreensão dos usuários, proveniência insuficiente, baixa atualização pelos países e dependência de equipe restrita (Carrillo *et al*., 2023), quadro que serve de alerta direto a uma plataforma brasileira de avaliação de políticas. O grafo parlamentar finlandês *ParliamentSampo* demonstra contribuições analíticas e históricas consolidadas, com a ligação direta à avaliação de políticas permanecendo no horizonte de possibilidades (Hyvönen *et al*., 2022), e os serviços seguros de vinculação de dados interagências nos Estados Unidos reportam resultados ainda preliminares (Mirel *et al*., 2024). Esses precedentes recomendam ao SAVITS ambição calibrada e investimento contínuo na mediação humana, papel no qual a biblioteconomia de dados, conforme delineado na subseção 3.3, tem contribuição permanente e insubstituível.

O plano político-epistêmico foi tematizado pela coordenação no quarto dia do *workshop*, na articulação entre ciência da informação e políticas públicas que discutiu injustiça epistêmica, colonialismo de dados, trabalho não remunerado da avaliação por pares e custos do ecossistema de publicação. A literatura recuperada dá densidade a essa pauta. O caso ugandense mostra que infraestruturas digitais de desenvolvimento podem institucionalizar dependência, vigilância e extração de valor quando agendas, contratos e categorias são definidos externamente (Kalema, 2025), e a experiência de bibliotecas africanas registra que exigências de financiadores por métricas e abertura deslocaram para atores externos o controle sobre validade, visibilidade e risco (Lynch *et al*., 2023). O ecossistema de comunicação científica reproduz dinâmicas análogas, com a concentração das taxas de processamento de artigos em grandes conglomerados do Norte Global, custos que no Brasil ultrapassaram dezenas de milhões de dólares em poucos anos (Pavan; Barbosa, 2018) e um modelo de negócio que se tornou ainda mais lucrativo que o das assinaturas (Zhang *et al*., 2022), enquanto a própria ciência aberta, quando os padrões e a infraestrutura permanecem sob controle de poucos, pode reproduzir as desigualdades que pretende combater (Silva *et al*., 2024). A camada de inteligência artificial adiciona sua parcela extrativa, com trabalho invisível em cada etapa da cadeia de montagem dos modelos (Pasquinelli; Joler, 2021). Os cartões do

levantamento colaborativo demonstram que a equipe do projeto já internalizou esse diagnóstico, ao registrar como risco a dependência de infraestruturas proprietárias de modelos de linguagem e os custos de participação no ecossistema de comunicação científica. A resposta do SAVITS a esse plano se compõe dos elementos que as subseções anteriores fundamentaram, a infraestrutura pública nacional herdada do Ibict, a identificação persistente descentralizada do dARK, a governança participativa de longa duração alinhada aos princípios CARE e a auditabilidade integral do fluxo informacional, e converge com a orientação da literatura, segundo a qual uma plataforma orientada à política pública deve coproduzir problemas e categorias com as instituições e comunidades afetadas, explicitar proveniência e limites interpretativos e repartir o poder sobre acesso, classificação e benefícios. Essa convergência devolve a discussão ao referencial da adequação sociotécnica que fundamenta o projeto, pois repartir o poder de classificar é a tradução informacional da tecnociência solidária (Dagnino, 2019).

O levantamento colaborativo de políticas públicas oferece, por fim, a demonstração empírica de que esses três planos já operam no projeto. As nove políticas candidatas foram analisadas pela equipe em seis categorias de atributos, e a distribuição dos votos revelou um critério implícito de decisão de natureza informacional, uma vez que as três políticas mais votadas, as ações de assistência social do eixo PNAS, SUAS e SCFV, a implementação de práticas de ciência aberta e ciência cidadã e a educação digital, são precisamente as que exibiram maior maturidade de bases de dados, séries históricas e indicadores disponíveis. A preferência coletiva por políticas com lastro informacional confirma, pela via da participação, a centralidade dos requisitos que a matriz estabeleceu por dedução analítica. A segunda política mais votada confirma ainda a antecipação do confronto com o MICS, pois os cartões que a sustentam apontam a existência de estruturas internacionais consolidadas para avaliação de ciência cidadã, exatamente o objeto da nota técnica produzida semanas antes do *workshop*.

E as dúvidas registradas sobre execução ética e legalmente segura reiteram os requisitos de curadoria da subseção 3.4. A convergência entre o instrumento analítico e a construção coletiva constitui evidência de que o arcabouço proposto descreve necessidades reais do projeto, e responde à pergunta de pesquisa em sua formulação mais sintética. Os serviços do Ibict se articulam com o SAVITS quando a plataforma os assume como camadas de uma cadeia institucional única, na qual a COBIB padroniza e descreve, a CODIC agrega, trata e analisa, as infraestruturas transversais identificam, ligam e preservam, e o SAVITS avalia e prediz, sob governança pactuada com as agências custodiantes e com as comunidades participantes, sustentada pela mediação contínua de

profissionais de informação e orientada pelo compromisso de que os dados das políticas públicas brasileiras sejam tratados por infraestrutura pública brasileira.

3 CONSIDERAÇÕES FINAIS

Este relatório consolidou as atividades desenvolvidas durante a vigência da bolsa no Núcleo de Dados e Validação do Projeto SAVITS, orientadas pela pergunta sobre a articulação dos serviços de informação do Ibict com a futura plataforma. Os sete objetivos específicos derivados das atribuições contratuais foram cumpridos ao longo das subseções de resultados. A análise da consistência metodológica de descritores, vocabulários controlados e tesauros estabeleceu que a terminologia do Caso Piloto deve preservar as disputas conceituais do campo das tecnologias sociais em vez de dissolvê-las, com notas de escopo para as formas singular e plural, facetas para artefato e processo e garantia literária ancorada nos catálogos nacionais e na plataforma *Cívis*. O mapeamento da representatividade temática e da interoperabilidade semântica definiu a estratégia de perfil de aplicação derivado da matriz de requisitos, com crosswalks documentados e versionados e validação de correspondências em nível de instância.

A estruturação de guias conceituais e documentação FAIR delimitou quatro instrumentos, o guia de curadoria, o modelo de dicionário de dados com rastreabilidade de variáveis, o padrão de *README* e *codebook* e o modelo de plano de gestão de dados estruturado. Os requisitos de curadoria e preservação digital articularam a herança da pilha da *Aleia*, a política de tratamento de dados pessoais derivada do guia do *Deposita Dados*, a gradação de acesso com avaliação de impacto à privacidade e a governança participativa de longa duração. O mapeamento de interoperabilidade sistêmica atribuiu papéis às infraestruturas nacionais e fundamentou a minuta de política de identificadores persistentes, com ORCID para pessoas, ROR para organizações e convivência entre DOI e dARK para objetos digitais. O arcabouço conceitual consolidou a matriz de vinte e quatro requisitos como espinha dorsal rastreável das Metas 1, 2 e 3.2, respaldada em caráter exploratório pelo confronto com o projeto MICS. E a resposta à pergunta de pesquisa organizou a articulação em três planos, com o fluxo institucional no qual a COBIB padroniza e descreve, a CODIC agrega, trata e analisa e o SAVITS avalia e prediz, sob governança pactuada e mediação profissional contínua.

Três contribuições metodológicas merecem destaque. A primeira é a própria matriz de requisitos informacionais, cuja rastreabilidade bidirecional entre requisitos e planejamento a converte em instrumento de governança reutilizável pelas etapas seguintes do projeto. A segunda é o confronto analítico com o MICS, que demonstrou a complementaridade entre a camada substantiva dos indicadores de impacto e a camada infraestrutural dos requisitos informacionais, produziu sustentação externa às prioridades da matriz e posicionou a auditabilidade como avanço deliberado do SAVITS sobre a limitação que sua principal referência internacional reconhece em si mesma. A

terceira é a convergência verificada entre a antecipação dedutiva dos instrumentos analíticos e a construção coletiva do *Workshop* de Integração, na qual o levantamento colaborativo de políticas públicas confirmou, pela via da participação, a centralidade dos requisitos informacionais como critério de decisão da equipe. Essa convergência sugere que o método empregado, de imersão documental seguida de elicitação sistemática e teste comparativo, produz instrumentos aderentes às necessidades reais de projetos dessa natureza.

Os desdobramentos propostos organizam a agenda das etapas subsequentes. O confronto com o MICS revelou três lacunas registradas como candidatas a novos requisitos, sujeitas à validação da coordenação, a avaliação como evento informacional datado e versionado, a recomendação como objeto informacional com metadados de geração e a documentação dos modelos preditivos na linha dos *model cards* e *datasheets*. A elas se somam os entregáveis fundamentados ao longo do relatório, o vocabulário controlado do Caso Piloto, o perfil de aplicação da plataforma, os quatro instrumentos documentais, a política de acesso em níveis com avaliação de impacto à privacidade, a minuta de política de identificadores persistentes e a proposta de instância de governança de dados do piloto com as agências custodiantes. A incorporação formal da nota técnica ao conjunto de entregas permanece condicionada à validação pela coordenação do projeto, conforme registrado no Método.

Quatro limites condicionam a leitura destes resultados. A matriz de requisitos constitui um exercício analítico ainda pendente de validação formal pela coordenação. O relato das sessões do *workshop* posteriores ao primeiro dia se apoia em observação participante sem registro documental equivalente, condição declarada e tratada com a cautela correspondente. O confronto com o MICS se baseou na documentação pública do projeto europeu, sem acesso aos entregáveis internos do consórcio. E o Caso Piloto se encontrava, ao término da vigência, em fase de seleção de políticas, de modo que as aplicações aqui propostas descrevem preparação fundamentada em vez de uma implementação verificada.

O percurso da bolsa confirma, por fim, a pertinência da biblioteconomia de dados como função permanente de projetos de infraestrutura informacional voltados a políticas públicas. As camadas que a literatura internacional aponta como decisivas para que plataformas dessa natureza produzam evidência utilizável, a curadoria contínua, a mediação entre produtores e usuários de dados, a documentação que preserva o contexto e a governança que reparte poder de classificação, correspondem ao núcleo da atuação profissional aqui relatada. Que o sistema brasileiro de avaliação de tecnologias sociais nasça sobre infraestrutura pública nacional, com dados tratados segundo princípios que protegem as comunidades participantes e lhes devolvem benefícios, é um compromisso que ultrapassa a dimensão técnica deste relatório e que a continuidade do projeto terá a oportunidade de honrar.

REFERÊNCIAS

ALBERTONI, R.; BROWNING, D.; COX, S.; GONZALEZ-BELTRAN, A. N.; PEREGO, A.; WINSTANLEY, P. The W3C Data Catalog Vocabulary, version 2: rationale, design principles, and uptake. *Data Intelligence*, [*s. l.*], v. 6, n. 2, p. 457-487, 2024. DOI: https://doi.org/10.1162/dint_a_00241.

AMMENWERTH, E.; BINDEL, M.; LIEBE, J.-D. Logic models for evaluation of complex health information systems. *In*: PFEIFER, B.; SCHREIER, G.; BAUMGARTNER, M.; HAYN, D. (ed.). *dHealth 2023*. [*S. l.*]: IOS Press, 2023. p. 131-132. DOI: https://doi.org/10.3233/SHTI230026.

ARYANI, A.; FENNER, M.; MANGHI, P.; MANNOCCI, A.; STOCKER, M. Open science graphs must interoperate! *In*: BELLATRECHE, L. *et al*. (ed.). *ADBIS, TPDL and EDA 2020 common workshops and doctoral consortium*. Cham: Springer, 2020. p. 195-206. (Communications in Computer and Information Science, v. 1260). DOI: https://doi.org/10.1007/978-3-030-55814-7_16.

BAGATINI, J. A.; GUIMARÃES, J. A. C.; SANT'ANA, R. C. G. Gerenciamento dos dados pessoais em arquivos: uma perspectiva centrada no indivíduo com base na LGPD. *Acervo*, Rio de Janeiro, v. 34, n. 3, p. 1-20, set./dez. 2021. DOI: https://doi.org/10.64729/an.acervo.v34i3.1749.

BAGATTOLLI, C.; PEZARICO, G. Tecnologia social: apropriações conceituais e implicações políticas. *InterAção*, Santa Maria, v. 16, n. 1, e90666, p. 1-22, jan./mar. 2025. DOI: https://doi.org/10.5902/2357797590666.

BAHIM, C.; CASORRÁN-AMILBURU, C.; DEKKERS, M.; HERCZOG, E.; LOOZEN, N.; REPANAS, K.; RUSSELL, K.; STALL, S. The FAIR data maturity model: an approach to harmonise FAIR assessments. *Data Science Journal*, [*s. l.*], v. 19, n. 41, p. 1-7, 2020. DOI: https://doi.org/10.5334/dsj-2020-041.

BERKOWITZ, E.; JENKINS, D.; NELSON, A. H.; ZANTI, S.; KEMP, D.; BURNETT, T. C.; BENITEZ, J. R. The AISP Network of cross-sector data linkage efforts: results & lessons from a national survey of 40 state and local systems in the US. *International Journal of Population Data Science*, [*s. l.*], v. 9, n. 5, 2024. DOI: https://doi.org/10.23889/ijpds.v9i5.2829.

BISHOP, B. W.; HANK, C. Measuring FAIR principles to inform fitness for use. *International Journal of Digital Curation*, [*s. l.*], v. 13, n. 1, p. 35-46, 2018. DOI: https://doi.org/10.2218/ijdc.v13i1.630.

BOETTIGER, C. Generating CodeMeta metadata for R packages. *Journal of Open Source Software*, [*s. l.*], v. 2, n. 19, p. 454, 2017. DOI: https://doi.org/10.21105/joss.00454.

BORGMAN, C. L.; SCHARNHORST, A.; GOLSHAN, M. S. Digital data archives as knowledge infrastructures: mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, [*s. l.*], 2018. DOI: https://doi.org/10.1002/asi.24172.

BOSCH, T.; MATHIAK, B. Use cases related to an ontology of the Data Documentation Initiative. *IASSIST Quarterly*, [*s. l.*], 2015. DOI: https://doi.org/10.29173/IQ898.

CAMPBELL, R.; JAVORKA, M.; ENGLETON, J.; FISHWICK, K.; GREGORY, K.; GOODMAN-WILLIAMS, R. Open-science guidance for qualitative research: an empirically validated approach for de-identifying sensitive narrative data. *Advances in Methods and Practices in Psychological Science*, [*s. l.*], v. 6, n. 4, p. 1-17, 2023. DOI: https://doi.org/10.1177/25152459231205832.

CARDOSO, J.; CASTRO, L. J.; EKAPUTRA, F. J.; JACQUEMOT, M. C.; SUCHÁNEK, M.; MIKSA, T.; BORBINHA, J. DCSO: towards an ontology for machine-actionable data management plans. *Journal of Biomedical Semantics*, [*s. l.*], v. 13, n. 21, p. 1-12, 2022. DOI: https://doi.org/10.1186/s13326-022-00274-4.

CARRILLO, E.; FRIGERIO, C.; VALENZUELA, M. J.; AQUARO, A.; MAUDUIT, J.-C.; STEENMANS, I.; SANDOVAL, M. P. The performance gap of policy information systems: a knowledge infrastructure assessment framework. *Journal of Science Policy & Governance*, [*s. l.*], v. 22, n. 1, 2023. DOI: https://doi.org/10.38126/jspg220105.

CARVALHO SEGUNDO, W. L. R.; DIAS, T. M. R.; SOUZA, M. G.; CAMPOS, F. F.; ANDRADE, D. A. F. Inovação e conectividade: uma perspectiva sobre o Projeto BrCris e suas tecnologias para tratamento de dados científicos. *In*: WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA (WIDaT), 7., 2024, Porto Velho. *Anais* [...]. Porto Velho: UNIR, 2024. p. 1-11. DOI: https://doi.org/10.22477/vii.widat.182.

CARVALHO SEGUNDO, W. L. R.; MATAS, L.; NÓBREGA, T.; SILVA FILHO, J. E.; MENA-CHALCO, J. P. dARK: uma implementação descentralizada de identificadores persistentes ARK baseada em blockchain. *BiblioCanto*, Natal, v. 9, n. 2, p. 147-158, 2023. DOI: https://doi.org/10.21680/2447-7842.2023v9n2id33759.

CASTRO, P.; HERB, U.; ROTHFRITZ, L.; SCHÖPFEL, J. Some reflections on the current PID landscape – with an emphasis on risks and trust issues. *Procedia Computer Science*, [*s. l.*], v. 211, p. 28-35, 2022. DOI: https://doi.org/10.1016/j.procs.2022.10.173.

CIUCIU-KISS, J. T.; GARIJO, D. Assessing the overlap of science knowledge graphs: a quantitative analysis. *In*: REHM, G. *et al.* (ed.). NSLP 2024. [*s. l.*]: Springer, 2024. p. 171-185. DOI: https://doi.org/10.1007/978-3-031-65794-8_11.

COOPER, C.; MARTIN, V.; WILSON, O. R.; RASMUSSEN, L. Equitable data governance models for the participatory sciences. *Community Science*, [*s. l.*], v. 2, e2022CSJ000025, p. 1-10, 2023. DOI: https://doi.org/10.1029/2022csj000025.

COSTA, T. A.; RONCAGLIO, C. Uma análise da legislação brasileira referente à produção, custódia, preservação e acesso aos arquivos de ciência. *Ágora: Arquivologia em Debate*, Florianópolis, v. 30, n. 61, p. 499-511, jul./dez. 2020. Disponível em: https://agora.emnuvens.com.br/ra/article/view/920. Acesso em: 17 jul. 2026.

COSTLEY, C.; ELLIOTT, G.; GIBBS, P. *Doing work based research*: approaches to enquiry for insider-researchers. London: SAGE Publications, 2010. DOI: https://doi.org/10.4135/9781446287880.

DAGNINO, R. *Tecnociência solidária*: um manual estratégico. Marília: Lutas Anticapital, 2019.

D'ANDREA, L.; SCHMIDT, E. K.; BUŽAN, E.; MERINO, M. V.; DALL, E.; COLONNELLO, C.; GRAVERSEN, E. K.; CERRI, J.; IACOLINA, L.; FEUDO, F. Evaluating citizen science initiatives through a citizen science-based approach. *fteval Journal for Research and Technology Policy Evaluation*, [*s. l.*], n. 54, p. 43-50, set. 2022. DOI: https://doi.org/10.22163/fteval.2022.570.

DAPPERT, A.; FARQUHAR, A.; KOTARSKI, R.; HEWLETT, K. Connecting the persistent identifier ecosystem: building the technical and human infrastructure for open research. *Data Science Journal*, [*s. l.*], v. 16, n. 28, p. 1-16, 2017. DOI: https://doi.org/10.5334/dsj-2017-028.

DEVARAJU, A.; MOKRANE, M.; CEPINSKAS, L.; HUBER, R.; HERTERICH, P.; VRIES, J. de; AKERMAN, V.; L'HOURS, H.; DAVIDSON, J.; DIEPENBROEK, M. From conceptualization to implementation: FAIR assessment of research data objects. *Data Science Journal*, [*s. l.*], v. 20, n. 4, p. 1-14, 2021. DOI: https://doi.org/10.5334/dsj-2021-004.

DUBOIS, J. M.; MOZERSKY, J.; PARSONS, M. V.; WALSH, H. A.; FRIEDRICH, A. B.; PIENTA, A. Exchanging words: engaging the challenges of sharing qualitative research data. *Proceedings of the National Academy of Sciences of the United States of America*, [*s. l.*], v. 120, n. 42, p. 1-12, 2023. DOI: https://doi.org/10.1073/pnas.2206981120.

DUQUE, T. O.; VALADÃO, J. A. D. Abordagens teóricas de tecnologia social no Brasil. *Revista Pensamento Contemporâneo em Administração*, [*s. l.*], v. 11, n. 5, p. 1-22, 2017. DOI: https://doi.org/10.12712/rpca.v11i5.962.

EITZEL, M. V.; CAPPADONNA, J. L.; SANTOS-LANG, C.; DUERR, R. E.; VIRAPONGSE, A.; WEST, S. E.; KYBA, C. C. M.; BOWSER, A.; COOPER, C. B.; SFORZI, A.; METCALFE, A. N.; HARRIS, E. S.; THIEL, M.; HAKLAY, M.; PONCIANO, L.; ROCHE, J.; CECCARONI, L.; SHILLING, F. M.; DÖRLER, D.; HEIGL, F.; KIESSLING, T.; DAVIS, B. Y.; JIANG, Q. Citizen

science terminology matters: exploring key terms. *Citizen Science: Theory and Practice*, [*s. l.*], v. 2, n. 1, p. 1-20, 2017. DOI: https://doi.org/10.5334/cstp.96.

ESTEVAO, J. S. B.; KLAS, C.-P. Enhancing FAIR compliance: a controlled vocabulary for mapping Social Sciences survey variables. *IASSIST Quarterly*, [*s. l.*], v. 48, n. 2, p. 1-15, 2024. DOI: https://doi.org/10.29173/iq1118.

EYNDEN, V. van der; CORTI, L. Advancing research data publishing practices for the social sciences: from archive activity to empowering researchers. *International Journal on Digital Libraries*, [*s. l.*], v. 18, p. 113-121, 2017. DOI: https://doi.org/10.1007/s00799-016-0177-3.

FARQUHAR, J.; MICHELS, N.; ROBSON, J. Triangulation in qualitative case study research: widening the scope. *Industrial Marketing Management*, [*s. l.*], v. 86, p. 160-170, 2020. DOI: https://doi.org/10.1016/j.indmarman.2020.02.001.

FERMOSO GARCÍA, A. M.; MANZANO GARCÍA, M. I.; BERJÓN GALLINAS, R.; MATEOS SÁNCHEZ, M.; BEATO GUTIÉRREZ, M. E. Integration and open access system based on semantic technologies: a use case applied to university research facet. *International Journal on Semantic Web and Information Systems*, [*s. l.*], v. 18, n. 1, p. 1-25, 2022. DOI: https://doi.org/10.4018/IJSWIS.309422.

FERREIRA, A. C.; MACULAN, B. C. M. S. Metodologias para revisão e atualização de tesauros: mapeamento da literatura. *Informação & Informação*, Londrina, v. 25, n. 1, p. 229-253, jan./mar. 2020. DOI: https://doi.org/10.5433/1981-8920.2020v25n1p229.

FIRMINO, H. N.; BAPTISTA, A. A. Making sense of a flat list of terms into Linked Open Data SKOS vocabularies. *Information Services & Use*, [*s. l.*], v. 33, p. 73-75, 2013. DOI: https://doi.org/10.3233/ISU-130699.

GUPTA, A.; LAI, A.; MOZERSKY, J.; MA, X.; WALSH, H.; DUBOIS, J. M. Enabling qualitative research data sharing using a natural language processing pipeline for deidentification: moving beyond HIPAA Safe Harbor identifiers. *JAMIA Open*, [*s. l.*], v. 4, n. 3, p. 1-10, 2021. DOI: https://doi.org/10.1093/jamiaopen/ooab069.

HACKING, N.; LEWIS, J.; EVANS, R. Mapping approaches to 'citizen science' and 'community science' and everything in-between: the evolution of new epistemic territory? *Minerva*, [*s. l.*], v. 62, p. 549-572, 2024. DOI: https://doi.org/10.1007/s11024-024-09529-z.

HARBICH, M.; BACHER, A.; HICKER, U.; ROHR, L.; SPRINGER-BRIEM, S. Nationale Standards und Schnittstellen zur Übertragung von Forschungsinformationen. *Zeitschrift für Hochschulentwicklung*, [*s. l.*], v. 18, p. 109-121, 2023. DOI: https://doi.org/10.21240/zfhe/sh-f/07.

HARIS, M.; FARFAR, K. E.; STOCKER, M.; AUER, S. Federating scholarly infrastructures with GraphQL. *ArXiv*, [*s. l.*], 2021. DOI: https://doi.org/10.1007/978-3-030-91669-5_24.

HEMPHILL, L.; PIENTA, A.; LAFIA, S.; AKMON, D.; BLECKLEY, D. A. How do properties of data, their curation, and their funding relate to reuse? *Journal of the Association for Information Science and Technology*, [*s. l.*], v. 73, n. 10, p. 1432-1444, 2022. DOI: https://doi.org/10.1002/asi.24646.

HENNING, P.; SILVA, L. O. B.; PIRES, L. F.; SINDEREN, M. van; MOREIRA, J. L. R. The FAIRness of data management plans: an assessment of some European DMPs. *Revista Eletrônica de Comunicação, Informação e Inovação em Saúde*, Rio de Janeiro, v. 15, n. 3, p. 1-20, 2021. DOI: https://doi.org/10.29397/reciis.v15i3.2270.

HIOM, D.; GRAY, S.; STEER, D.; MERRETT, K.; SNOW, K.; BECKLES, Z. Introducing safe access for sensitive data at the University of Bristol. *International Journal of Digital Curation*, [*s. l.*], v. 12, n. 2, p. 26-36, 2017. DOI: https://doi.org/10.2218/ijdc.v12i2.506.

HOUSSOS, N.; JÖRG, B.; DVOŘÁK, J.; PRÍNCIPE, P.; RODRIGUES, E.; MANGHI, P.; ELBÆK, M. K. OpenAIRE guidelines for CRIS managers: supporting interoperability of open research information through established standards. *Procedia Computer Science*, [*s. l.*], v. 33, p. 33-38, 2014. DOI: https://doi.org/10.1016/j.procs.2014.06.006.

HYVÖNEN, E.; LESKINEN, P.; SINIKALLIO, L.; MELA, M. L.; TUOMINEN, J.; ELO, K.; DROBAC, S.; KOHO, M.; IKKALA, E.; TAMPER, M.; LEAL, R.; KESÄNIEMI, J. Finnish Parliament on the Semantic Web: using ParliamentSampo data service and semantic portal for studying political culture and language. *DHNB Publications*, [*s. l.*], 2022. DOI: https://doi.org/10.5617/dhnbpub.11261.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. *Projeto SAVITS*: Sistema de Avaliação e Validação de Impactos em Tecnologias Sociais: relatório consolidado de execução, versão 2.1. Brasília, DF: Ibict, 2026a. 40 p. Documento interno, não publicado.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. *Síntese do Workshop de Integração do Projeto SAVITS*: dia 1. Brasília, DF: Ibict, 21 jul. 2026b. 12 p. Evento realizado de forma remota entre 21 e 24 jul. 2026. Documento interno, não publicado.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964-1: thesauri and interoperability with other vocabularies - part 1: thesauri for information retrieval. Geneva: ISO, 2011.

JANG, J. B.; PIENTA, A. M.; LEVENSTEIN, M. C.; SAUL, J. Restricted data management: the current practice and the future. *Journal of Privacy and Confidentiality*, [*s. l.*], v. 13, n. 2, p. 1-20, 2023. DOI: https://doi.org/10.29012/jpc.844.

JEFFERY, K. G.; HOUSSOS, N.; JÖRG, B.; ASSERSON, A. Research information management: the CERIF approach. *International Journal of Metadata, Semantics and Ontologies*, [*s. l.*], v. 9, n. 1, p. 1-13, 2014. DOI: https://doi.org/10.1504/IJMSO.2014.059142.

JÖRG, B. CERIF: the Common European Research Information Format model. *Data Science Journal*, [*s. l.*], v. 9, p. CRIS4-CRIS12, 2010. DOI: https://doi.org/10.2481/dsj.CRIS4.

KALEMA, N. L. The 'digital transformation for development' anti-politics machine: a case study on global digital development governance and public-sector digital transformation in Uganda. *Policy & Internet*, [*s. l.*], 2025. DOI: https://doi.org/10.1002/poi3.436.

KARCHER, S.; SECEN, S.; WEBER, N. Protecting sensitive data early in the research data life cycle. *Journal of Privacy and Confidentiality*, [*s. l.*], v. 13, n. 2, p. 1-20, 2023. DOI: https://doi.org/10.29012/jpc.846.

KIESLINGER, B.; SCHÄFER, T.; HEIGL, F.; DÖRLER, D.; RICHTER, A.; BONN, A. The challenge of evaluation: an open framework for evaluating citizen science activities. [*S. l.: s. n.*], 2017. Preprint. DOI: https://doi.org/10.31235/osf.io/enzc9.

KIRILOVA, D.; KARCHER, S. Rethinking data sharing and human participant protection in social science research: applications from the qualitative realm. *Data Science Journal*, [*s. l.*], v. 16, n. 43, p. 1-7, 2017. DOI: https://doi.org/10.5334/dsj-2017-043.

KNEEBONE, L. A class focused approach to research outputs and policy literature metadata. *International Journal of Digital Curation*, [*s. l.*], v. 14, n. 1, p. 250-261, 2020. DOI: https://doi.org/10.2218/ijdc.v14i1.640.

KNIGHT, H.; FRASER, C.; PRENDERGAST, T.; KIFETEW, C. The Networked Data Lab: lessons from four years of analytical collaboration on common problems in health on social care. *International Journal of Population Data Science*, [*s. l.*], v. 9, n. 5, 2024. DOI: https://doi.org/10.23889/ijpds.v9i5.2788.

LAFIA, S.; THOMER, A.; BLECKLEY, D. A.; AKMON, D.; HEMPHILL, L. Leveraging machine learning to detect data curation activities. *IEEE International Conference on eScience*, [*s. l.*], v. 17, p. 1-10, 2021. DOI: https://doi.org/10.1109/eScience51609.2021.00025.

LEE, D. J.; STVILIA, B. Practices of research data curation in institutional repositories: a qualitative view from repository staff. *PLoS ONE*, [*s. l.*], v. 12, n. 3, e0173987, p. 1-32, 2017. DOI: https://doi.org/10.1371/journal.pone.0173987.

LEIVA-MEDEROS, A.; SENSO, J. A.; HIDALGO-DELGADO, Y.; HIPOLA, P. Working framework of semantic interoperability for CRIS with heterogeneous data sources. *Journal of Documentation*, [*s. l.*], v. 73, n. 3, p. 481-499, 2017. DOI: https://doi.org/10.1108/JD-07-2016-0091.

LEZCANO, L.; JÖRG, B.; LOWE, B.; CORSON-RIKERT, J. Promoting international interoperability of research information systems: VIVO and CERIF. *Journal of Universal Computer Science*, [*s. l.*], v. 19, n. 12, p. 1854-1867, 2013. DOI: https://doi.org/10.3217/jucs-019-12-1854.

LEZCANO, L.; JÖRG, B.; SICILIA, M.-A. Modeling the context of scientific information: mapping VIVO and CERIF. *In*: BAJEC, M.; EDER, J. (ed.). CAiSE 2012 Workshops. Heidelberg: Springer, 2012. p. 123-129. (LNBIP, v. 112). DOI: https://doi.org/10.1007/978-3-642-31069-0_11.

LIMA, K. C. R.; SUNYE, M. S. Proposta de automação da curadoria de metadados com machine learning. *In*: ARAÚJO, P. C.; LIMA, K. C. R. (org.). *Práticas de ciência aberta*. Brasília, DF: Ibict, 2025. p. 96-117. DOI: https://doi.org/10.22477/9788570131966.

LYNCH, R.; YOUNG, J. C.; JOWAISAS, C.; SAM, J.; BOAKYE-ACHAMPONG, S.; GARRIDO, M.; ROTHSCHILD, C. 'The tears don't give you funding': data neocolonialism in development in the Global South. *Third World Quarterly*, [*s. l.*], v. 44, n. 5, p. 911-929, 2023. DOI: https://doi.org/10.1080/01436597.2023.2166482.

MANGHI, P.; BARDI, A.; ATZORI, C.; BAGLIONI, M.; MANOLA, N.; SCHIRRWAGEN, J.; PRÍNCIPE, P. The OpenAIRE Research Graph Data Model. [*S. l.: s. n.*], 2019. DOI: https://doi.org/10.5281/zenodo.2643199.

MASSEY, A. K.; EISENSTEIN, J.; ANTÓN, A. I.; SWIRE, P. P. Automated text mining for requirements analysis of policy documents. *In*: IEEE INTERNATIONAL REQUIREMENTS ENGINEERING CONFERENCE, 21., 2013. [*S. l.*]: IEEE, 2013. p. 1-10. DOI: https://doi.org/10.1109/RE.2013.6636700.

MEDEIROS, C. B.; GALVÃO, C. E. S.; CORREIA, S. É. N.; GÓMEZ, C. R. P.; CASTILLO, L. Inovação social além da tecnologia social: constructos em discussão. *RACE: Revista de Administração, Contabilidade e Economia*, Joaçaba, v. 16, n. 3, p. 957-982, set./dez. 2017. DOI: https://doi.org/10.18593/race.v16i3.13606.

MHAIDLI, A.; JORDAN, C.; HEMPHILL, L.; SCHAUB, F.; THOMER, A. Privacy impact assessments for digital repositories. *International Journal of Digital Curation*, [*s. l.*], v. 15, n. 1, p. 1-5, 2020. DOI: https://doi.org/10.2218/ijdc.v15i1.692.

MINAMIYAMA, Y.; TAKEDA, H.; HAYASHI, M.; ASAOKA, M.; YAMAJI, K. A study on formalizing the knowledge of data curation activities across different fields. *PLoS ONE*, [*s. l.*], v. 19, n. 4, e0301772, p. 1-32, 2024. DOI: https://doi.org/10.1371/journal.pone.0301772.

MIREL, L.; GOLDEN, C.; ZYBRICK, R.; WANG, R.; TADLER, C.; COX, C. Linking data in a shared service environment. *International Journal of Population Data Science*, [*s. l.*], v. 9, n. 5, 2024. DOI: https://doi.org/10.23889/ijpds.v9i5.2548.

MOCHÓN BEZARES, G.; MÉNDEZ RODRÍGUEZ, E. M.; BUENO DE LA FUENTE, G. Conversión normalizada (SKOS) de sistemas de organización del conocimiento interoperables en la web. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, México, v. 34, n. 82, p. 53-86, 2020. DOI: https://doi.org/10.22201/iibi.24488321xe.2020.82.58071.

MORAES, C. A. Representações sociais da comunidade científica brasileira sobre tecnologia social. 2012. Tese (Doutorado em Psicologia) – Pontifícia Universidade Católica de Goiás, Goiânia, 2012. Disponível em: https://tede2.pucgoias.edu.br/handle/tede/1756. Acesso em: 9 jun. 2026.

MOREIRA, J. M.; CUNHA, A.; MACEDO, N. An ORCID based synchronization framework for a national CRIS ecosystem. *F1000Research*, [*s. l.*], v. 4, art. 181, p. 1-19, 2015. DOI: https://doi.org/10.12688/f1000research.6499.1.

MUSEN, M. A.; O'CONNOR, M. J.; SCHULTES, E.; MARTÍNEZ-ROMERO, M.; HARDI, J.; GRAYBEAL, J. Modeling community standards for metadata as templates makes data FAIR. *Scientific Data*, [*s. l.*], v. 9, art. 696, p. 1-13, 2022. DOI: https://doi.org/10.1038/s41597-022-01815-3.

NASCIMENTO, B. L. C.; SILVA, E. M. Lei Geral de Proteção de Dados (LGPD) e repositórios institucionais: reflexões e adequações. *Em Questão*, Porto Alegre, v. 29, e-127314, p. 1-27, 2023. DOI: https://doi.org/10.1590/1808-5245.29.127314.

NEYENS, E.; VANCAUWENBERGH, S. Towards a semantic interoperable flemish research information space: development and implementation of a flemish application profile for research datasets. *International Journal of Digital Curation*, [*s. l.*], v. 16, n. 1, p. 1-16, 2021. DOI: https://doi.org/10.2218/ijdc.v16i1.762.

NICHOLSON, S. W.; BENNETT, T. B. Do institutional repository deposit guidelines deter data discovery? *Evidence Based Library and Information Practice*, [*s. l.*], v. 16, n. 3, p. 2-17, 2021. DOI: https://doi.org/10.18438/eblip29913.

NICOLINI, D.; MENGIS, J.; SWAN, J. Understanding the role of objects in cross-disciplinary collaboration. *Organization Science*, [*s. l.*], v. 23, n. 3, p. 612-629, 2012. DOI: https://doi.org/10.1287/orsc.1110.0664.

NOGUERA, E. M.; UROZ, R. O.; JUAN, I. L.. Enabling the reuse of personal data in research: a classification model for legal compliance. *ArXiv*, [*s. l.*], p. 1-10, 2025. DOI: https://doi.org/10.5334/dsj-2026-018.

OHMANN, C.; MOILANEN, K.; KLEEMOLA, M.; CANHAM, S.; PANAGIOTOPOULOU, M. ECRIN – CESSDA strategies for cross metadata mappings in selected areas between life sciences and social sciences and humanities. *Open Research Europe*, [*s. l.*], v. 3, p. 1-26, 2023. DOI: https://doi.org/10.12688/openreseurope.16284.2.

PAPADOPOULOU, E.; BARDI, A.; KAKALETRIS, G.; TZIOTZIOS, D.; MANGHI, P.; MANOLA, N. Data management plans as linked open data: exploiting ARGOS FAIR and machine actionable outputs in the OpenAIRE research graph. *Journal of Biomedical Semantics*, [*s. l.*], v. 14, art. 17, p. 1-12, 2023. DOI: https://doi.org/10.1186/s13326-023-00297-5.

PARKINSON, S.; WOODS, S. M.; SPRINKS, J.; CECCARONI, L. A Practical approach to assessing the impact of citizen science towards the sustainable development goals. *Sustainability*, [*s. l.*], v. 14, n. 8, art. 4676, p. 1-26, 2022. DOI: https://doi.org/10.3390/su14084676.

PASQUINELLI, M., JOLER, V. The nooscope manifested: AI as instrument of knowledge extractivism. *AI & Society: knowledge, culture and communication*, [*s. l.*], v. 36, p. 1263-1280, 2021. DOI: https://doi.org/10.1007/s00146-020-01097-6.

PASSANI, A.; JANSSEN, A.; HÖLSCHER, K.; LISIO, G. Di. A participatory, multidimensional and modular impact assessment methodology for citizen science projects. *fteval Journal for Research and Technology Policy Evaluation*, [*s. l.*], n. 54, p. 33-42, 2022. DOI: https://doi.org/10.22163/fteval.2022.569.

PATEMAN, R. M.; WEST, S. E. Citizen science: pathways to impact and why participant diversity matters. *Citizen Science: Theory and Practice*, [*s. l.*], v. 8, n. 1, art. 50, p. 1-15, 2023. DOI: https://doi.org/10.5334/cstp.569.

PATRÍCIO, L.; SOUSA, L. S.; LOPES, E.; SOUZA, F.; OLIVEIRA, H. V. A elaboração de um catálogo de tecnologias sociais: a prática para além da teoria. *Sitientibus*, [*s. l.*], v. 1, n. 64, e5748, 2024. DOI: https://doi.org/10.13102/sitientibus.v1i64.5748.

PAVAN, C.; BARBOSA, M. Article processing charge (APC) for publishing open access articles: the Brazilian scenario. *Scientometrics*, Budapest, v. 117, p. 805-823, 2018. DOI: https://doi.org/10.1007/s11192-018-2896-2.

PHAM, N.-M.; MOULAISON-SANDY, H.; BISHOP, B. W.; GUNDERMAN, H. Data management plans: implications for automated analyses. *Data Science Journal*, [*s. l.*], v. 22, art. 2, p. 1-14, 2023. DOI: https://doi.org/10.5334/dsj-2023-002.

PHEGLEY, L.; KELLAM, L. How are we FAIR-ing?: creating a FAIR self-assessment checklist for data repositories. *IASSIST Quarterly*, [*s. l.*], v. 49, n. 3, p. 1-9, 2025. DOI: https://doi.org/10.29173/iq1152.

RIECHERT, M. T. RIDAL – a language for research information definition argumentation. *Data Science Journal*, [*s. l.*], v. 16, art. 5, p. 1-17, 2017. DOI: https://doi.org/10.5334/dsj-2017-005.

RODRIGO, L.; PALACIOS, M.; ORTIZ-MARCOS, I. Digital social innovation: analysis of the conceptualization process and definition proposal. *Dirección y Organización*, [*s. l.*], n. 67, p. 59-66, 2019. DOI: https://doi.org/10.37610/dyo.v0i67.545.

RODRIGUES, D. G. A.; ANJOS, R. L.; DIAS, G. A.; SIEBRA, S. A. Modelo referencial de ciclo de vida dos dados científicos: uma proposta metodológica. Informação & Informação, Londrina, v. 28, n. 3, p. 55-80, jul./set. 2023. DOI: https://doi.org/10.5433/1981-8920.2023v28n3p55.

ROUSE, H. L.; DORIUS, C.; ANDERSON, J. B.; RICHEY, E. Integrated administrative data for Early Childhood Iowa: a governance model to inform policy and program collaboration. *International Journal of Population Data Science*, [*s. l.*], v. 3, n. 5, art. 27, p. 1-2, 2018. DOI: https://doi.org/10.23889/ijpds.v3i5.1067.

RUNESON, P.; HÖST, M. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, [*s. l.*], v. 14, n. 2, p. 131-164, 2009. DOI: https://doi.org/10.1007/s10664-008-9102-8.

SALATINO, A.; AGGARWAL, T.; MANNOCCI, A.; OSBORNE, F.; MOTTA, E. A survey on knowledge organization systems of research fields: resources and challenges. *ArXiv*, [*s. l.*], 2024. DOI: https://doi.org/10.1162/qss_a_00363.

SCHAEFER, T.; KIESLINGER, B.; BRANDT, M.; BOGAERT, V. van den. Evaluation in citizen science: the art of tracing a moving target. *In*: VOHLAND, K. *et al*. (ed.). *The science of citizen science*. Cham: Springer, 2021. p. 495-514. DOI: https://doi.org/10.1007/978-3-030-58278-4_25.

SCHNIEDERS, K.; MIERZ, S.; BOCCALINI, S.; MEYER ZU WESTERHAUSEN, W.; HAGEMANN-WILHOLT, S.; SCHULZE, S. ORCID coverage in research institutions: readiness for partially automated research reporting. *Frontiers in Research Metrics and Analytics*, [*s. l.*], v. 7, art. 1010504, p. 1-11, 2022. DOI: https://doi.org/10.3389/frma.2022.1010504.

SHWE, K. M. Study on the data management of citizen science: from the data life cycle perspective. *Data and Information Management*, [*s. l.*], v. 4, n. 4, p. 279-296, 2020. DOI: https://doi.org/10.2478/dim-2020-0019.

SILVA, F. C. C.; UMPIERRE, L. W.; WITT, A. S. Indicadores de preservação digital em projetos de ciência cidadã. *Revista Brasileira de Preservação Digital*, Campinas, v. 4, e023013, p. 1-14, 2023. DOI: https://doi.org/10.20396/rebpred.v4i00.18010.

SILVA, F. C. C.; WITT, A. S.; WENDT, L. G.; GABRIEL JUNIOR, R. F.; LEAL, L. R. C. Taxonomia em ciência cidadã: uma análise de terminologias fundamentais. *Informação & Informação*, Londrina, v. 29, n. 1, p. 172-201, jan./mar. 2024. DOI: https://doi.org/10.5433/1981-8920.2034v29n1p172.

SILVEIRA, J. I.; VANZ, S. A. S. Diretrizes para políticas de depósito, acesso e uso de dados de pesquisa: proposta a partir da análise de repositórios de dados universitários internacionais. *AtoZ: novas práticas em informação e conhecimento*, Curitiba, v. 12, e87331, p. 1-17, 2023. DOI: https://doi.org/10.5380/atoz.v12i0.87331.

SIMONIS, I. Standardized information models to optimize exchange, reusability and comparability of citizen science data: a specialization approach. *International Journal of Spatial Data Infrastructures Research*, [*s. l.*], v. 13, p. 38-47, 2018. DOI: https://doi.org/10.2902/1725-0463.2018.13.art5.

SKARLATIDOU, A.; HAKLAY, M. Citizen science impact pathways for a positive contribution to public participation in science. *Journal of Science Communication*, [*s. l.*], v. 20, n. 6, A02, p. 1-19, 2021. DOI: https://doi.org/10.22323/2.20060202.

SPRINKS, J.; CECCARONI, L.; WOODS, S.; PARKINSON, S.; GUMIERO, B. Design of a platform to measure the impact of citizen science. *Proceedings of Science*, [*s. l.*], v. 418, art. 006, p. 1-8, 2022. DOI: https://doi.org/10.22323/1.418.0006.

STANKOVIC, S.; SUN, I.; WEMP, K. An evaluation of OpenAlex Level 0: Art. *Proceedings of the Annual Conference of CAIS / Actes du congrès annuel de l'ACSI*, [*s. l.*], 2025. DOI: https://doi.org/10.29173/cais1979.

STRECKER, D. Metadata conflicts and their impact on DataCite metadata completeness in disciplinary research data repositories. *ArXiv*, [*s. l.*], p. 1-28, 2026. DOI: https://doi.org/10.48550/arXiv.2603.25468.

STVILIA, B.; LEE, D. J. Data quality assurance in research data repositories: a theory-guided exploration and model. *Journal of Documentation*, [*s. l.*], v. 80, n. 4, p. 793-812, 2024. DOI: https://doi.org/10.1108/JD-09-2023-0177.

SUOMINEN, T.; KESÄNIEMI, J.; KOIVULA, H. Making schemas and mappings available and FAIR: a metadata and schema crosswalk registry from the FAIRCORE4EOSC project. *Biodiversity Information Science and Standards*, [*s. l.*], v. 7, e112223, p. 1-3, 2023. DOI: https://doi.org/10.3897/biss.7.112223.

TURKI, H.; FRAUMANN, G.; HADJ TAIEB, M. A.; BEN AOUICHA, M. Global visibility of publications through digital object identifiers. *Frontiers in Research Metrics and Analytics*, [*s. l.*], v. 8, art. 1207980, p. 1-6, 2023. DOI: https://doi.org/10.3389/frma.2023.1207980.

UDOVIČIĆ, P.; MATIJAŠEVIĆ, M.; VUSIĆ, L.; HANIĆ, M.; BALENOVIĆ, N.; OREL, O. The role of CRIS as a glue for national e-infrastructure and digital services for open science. *Procedia Computer Science*, [*s. l.*], v. 249, p. 167-172, 2024. DOI: https://doi.org/10.1016/j.procs.2024.11.060.

UUDE, K.; KURZHALS, K.; WESBUER, A. The 'payback' of citizen science: a participatory evaluation and impact-assessment model for social innovation projects. *fteval Journal for Research and Technology Policy Evaluation*, [*s. l.*], n. 54, p. 61-67, set. 2022. DOI: https://doi.org/10.22163/fteval.2022.572.

VANCAUWENBERGH, S. FAIR and open research metadata as leverage for digital libraries: the Flemish case. *IntechOpen*, [*s. l.*], p. 1-18, 2021. DOI: https://doi.org/10.5772/intechopen.97192.

VASCONCELLOS, A. A. Tecnologia social: uma revisão de literatura. *InterAção*, Santa Maria, v. 16, n. 1, e89357, p. 1-21, jan./mar. 2025. DOI: https://doi.org/10.5902/2357797589357.

VÁZQUEZ TAPIA, R. Development and characterisation of CRIS systems in Latin America: preliminary results of diagnostic survey. *Procedia Computer Science*, [*s. l.*], v. 211, p. 267-276, 2022. DOI: https://doi.org/10.1016/j.procs.2022.10.201.

VITALE, C. H.; MOULAISON-SANDY, H. Data management plans: a review. *DESIDOC Journal of Library & Information Technology*, [*s. l.*], v. 39, n. 6, p. 322-328, 2019. DOI: https://doi.org/10.14429/djlit.39.06.15086.

WEHN, U.; GHARESIFARD, M.; CECCARONI, L.; JOYCE, H.; AJATES, R.; WOODS, S.; BILBAO, A.; PARKINSON, S.; GOLD, M.; WHEATLAND, J. Impact assessment of citizen science: state of the art and guiding principles for a consolidated approach. *Sustainability Science*, [*s. l.*], v. 16, n. 5, p. 1683-1699, 2021. DOI: https://doi.org/10.1007/s11625-021-00959-2.

WIGGINS, A.; BONNEY, R.; LEBUHN, G.; PARRISH, J. K.; WELTZIN, J. F. A science products inventory for citizen-science planning and evaluation. *BioScience*, [*s. l.*], v. 68, n. 6, p. 436-444, 2018. DOI: https://doi.org/10.1093/biosci/biy028.

WILKINSON, M. D.; DUMONTIER, M.; SANSONE, S.-A.; SANTOS, L. O. B. da S.; PRIETO, M.; BATISTA, D.; MCQUILTON, P.; KUHN, T.; ROCCA-SERRA, P.; CROSAS, M.; SCHULTES, E. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data*, [*s. l.*], v. 6, art. 174, p. 1-12, 2019. DOI: https://doi.org/10.1038/s41597-019-0184-5.

WU, M.; RICHARD, S. M.; VERHEY, C.; CASTRO, L. J.; CECCONI, B.; JUTY, N. An analysis of crosswalks from research data schemas to Schema.org. *Data Intelligence*, [*s. l.*], v. 5, n. 1, p. 100-121, 2023. DOI: https://doi.org/10.1162/dint_a_00186.

ZAFAR, H. Exploring a machine-generated concept hierarchy through the lens of 'naive' classification. *Proceedings of the Annual Conference of CAIS / Actes du congrès annuel de l'ACSI*, [*s. l.*], p. 1-5, 2025. DOI: https://doi.org/10.29173/cais1978.

ZENK-MÖLTGEN, W. The role of FAIR principles in high-quality research data documentation: looking at national election studies. *IASSIST Quarterly*, [*s. l.*], v. 49, n. 2, p. 1-22, 2025. DOI: https://doi.org/10.29173/iq1119.

ZHANG, L.; WEI, Y.; HUANG, Y.; SIVERTSEN, G.. Should open access lead to closed research?: the trends toward paying to perform research. *Scientometrics*, Budapest, v. 127, p. 7653-7679, 2022. DOI: https://doi.org/10.1007/s11192-022-04407-5.

Anexo A – Relatório de entrega de atividades

Período do estudo: 15 / 06 / 2026 a 29 / 06 / 2026

Projeto: SAVITS

Data de Entrega: 30 / 06 / 2026

1 ECOSSISTEMA IBICT

A presente seção apresenta um levantamento sistemático dos principais serviços, sistemas e produtos de informação que compõem o ecossistema do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Para cada item, foram identificados sua natureza e seu objetivo institucional, com base na documentação oficial e nos materiais consultados durante o período de imersão. Este exercício tem por finalidade alinhar a terminologia adotada pela equipe executora do projeto SAVITS - Sistema de Avaliação e Validação de Impactos em Tecnologias Sociais, estabelecendo um vocabulário comum que sustente as discussões técnicas das próximas reuniões e assegure precisão conceitual nas etapas subsequentes da pesquisa.

Quadro 1 - Produtos e serviços do Ibict

Serviço	O que é?	Principal objetivo?
Portal Ibict (gov.br/ibict)	Portal institucional do Ibict, unidade de pesquisa vinculada ao MCTI dedicada à produção, organização, preservação, disseminação e uso da informação científica e tecnológica no Brasil.	Centralizar o acesso institucional aos serviços, notícias, projetos e canais de transparência do Instituto, apoiando a ciência aberta, a gestão da informação e o acesso ao conhecimento.
Biblioteca do Ibict (Lydia de Queiroz Sambaquy)	Biblioteca especializada em Ciência da Informação, Biblioteconomia e áreas afins, com acervo de mais de 40 mil volumes, incluindo raridades nacionais e internacionais.	Oferecer ao público serviços como consulta ao acervo, levantamento bibliográfico, catalogação na fonte, empréstimo, treinamentos e acesso a bases de dados.
Canal do Ibict no *YouTube*	Canal oficial de divulgação científica e institucional do Ibict em vídeo.	Disseminar conhecimento científico e ações institucionais em formato audiovisual.
Bibliodata	Rede de cooperação de catalogação bibliográfica que compartilha registros e recursos entre bibliotecas brasileiras, sendo uma das mais antigas iniciativas de catalogação	Integrar os acervos de bibliotecas em um repositório unificado, padronizado e de acesso público. A iniciativa otimiza a catalogação cooperativa e a pesquisa

Serviço	O que é?	Principal objetivo?
	cooperativa do Brasil.	bibliográfica.
Catálogo Coletivo Nacional (CCN)	Catálogo de acesso público que reúne informações sobre coleções de publicações seriadas (periódicos) nacionais e estrangeiras disponíveis em bibliotecas brasileiras.	Difundir, identificar e localizar publicações seriadas no Brasil e promover o intercâmbio entre bibliotecas, articulado pelo *Comut*.
Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	Portal que reúne e dá acesso aberto ao texto completo de teses e dissertações defendidas em instituições brasileiras de ensino e pesquisa, assim como as defendidas no exterior por brasileiros.	Reunir, num só ponto de busca, a produção de teses e dissertações no país, ampliando sua visibilidade e o acesso aberto. O Ibict atua como agregador, coletando os metadados, enquanto o documento original permanece na instituição de defesa.
Repositório Institucional Digital do Ibict (RIDI)	Repositório que disponibiliza um acervo crescente de publicações produzidas no próprio Ibict, proporcionando maior visibilidade.	Reunir, preservar e difundir em acesso aberto a produção científica e técnica do Ibict.
Programa de Comutação Bibliográfica (*Comut*)	Programa que permite a obtenção de cópias de documentos técnico-científicos disponíveis em bibliotecas brasileiras e internacionais.	Facilitar o acesso à documentação científica por meio da solicitação de cópias entre bibliotecas. Os usuários podem localizar as obras através do CCN.
Memória Técnica do Ibict	Acervo que reúne os relatórios e a produção técnica e documental histórica do Ibict, assim como monografias, anais, seriados e obras de referência. Inclusive, é parte integrante do acervo especializado da Biblioteca Lydia de Queiroz Sambaquy, do Instituto.	Preservar e dar acesso à memória institucional e à produção técnica do Ibict. Atualmente, apoia-se com o reforço do Laboratório *LaDAACRI*, vinculado à COBIB, com ações nas áreas de digitalização e preservação.
Rede de Bibliotecas de Pesquisa do MCTI	Rede que articula as bibliotecas das unidades de pesquisa vinculadas ao Ministério de Ciência, Tecnologia e Inovação.	Integrar e fortalecer os acervos e serviços das bibliotecas das unidades de pesquisa do MCTI, de forma cooperativa e fortalecer o intercâmbio de informações.
Centro Brasileiro do ISSN	Centro nacional responsável pela atribuição e gestão do ISSN (identificador internacional de publicações seriadas) no Brasil, integrado à rede internacional do ISSN.	Atribuir e controlar os ISSN das publicações seriadas brasileiras, garantindo sua identificação única e padronizada internacionalmente.
Pinakes	Ferramenta/projeto que utiliza inteligência artificial e web semântica para uma nova experiência de interação, pesquisa, navegação e	Reestruturar e modernizar os serviços bibliográficos tradicionais (CCN, Bibliodata e Comut, garantindo interoperabilidade e

Serviço	O que é?	Principal objetivo?
	recuperação de dados bibliográficos.	recuperação semântica.
Repositório *Aleia*	Repositório oficial do Ibict para custodiar, publicar, preservar e disseminar conjuntos de dados das pesquisas geradas pela comunidade científica do Ibict. Certificado pelo *CoreTrustSeal*.	Promover acesso aberto e a gestão confiável (FAIR) dos dados de pesquisa produzidos pela comunidade científica do Ibict.
Repositório *Deposita Dados*	Repositório nacional de dados de pesquisa desenvolvido pelo Ibict para reunir e dar visibilidade à produção de pesquisadores vinculados a instituições que ainda não possuem repositórios próprios.	Democratizar o acesso aos dados e incentivar a criação de novos repositórios institucionais em universidades e institutos de pesquisa.
Ecossistema de Informação da Pesquisa Científica Brasileira (*BrCris*)	Plataforma online que reúne, integra, transforma e disponibiliza vários sistemas de informação de suporte à atividade científica, com análise e visualização de dados técnicos-científicos do Brasil (modelo CRIS - *Current Research Information System*).	Estabelecer um modelo único de organização da informação científica de todo o ecossistema de pesquisa nacional, conectando *Lattes*, *Sucupira*, *OasisBr* e bases internacionais. Atua a partir de uma nova versão, desde maio de 2026.
Plataforma *Cívis*	Plataforma de ciência cidadã desenvolvida pelo Ibict em código aberto a partir do EU-Citizen.Science.	Oferecer infraestrutura e conteúdos para ampliar o entendimento sobre ciência cidadã e apoiar iniciativas e metodologias do campo, com foco na América Latina e Caribe.

Fonte: Dados da pesquisa.

2 REFLEXÕES SOBRE OS SERVIÇOS

A compreensão da articulação entre os serviços de informação do Ibict e a futura plataforma SAVITS exige, preliminarmente, o reconhecimento da estrutura organizacional que os sustenta. No âmbito da Coordenação-Geral de Informação Científica e Técnica (CGIC), à qual o projeto SAVITS se vincula, organizam-se duas coordenações complementares, sendo elas, a Coordenação de Tratamento, Análise e Disseminação da Informação Científica (CODIC) e a Coordenação de Serviços Bibliográficos (COBIB). Em conjunto, a CGIC é responsável pela criação, manutenção e aprimoramento dos sistemas e serviços de informação científica e bibliográfica, além da

condução de iniciativas estruturantes relacionadas à Ciência Aberta no Instituto. À COBIB cabe a gestão dos serviços bibliográficos tradicionais, entre eles o CCN, a Rede *Bibliodata* e o *Comut*, hoje em processo de modernização pelo Projeto *Pinakes*, enquanto a CODIC responde pelo tratamento, análise e disseminação da informação científica, abrangendo repositórios de dados de pesquisa como o *Aleia* e o *Deposita Dados*, bem como plataformas analíticas como o *BrCris*. É nesse arranjo institucional, orientado à interoperabilidade e às práticas abertas, que o SAVITS se insere e a partir do qual estabelece suas possibilidades de articulação. Este relatório técnico propõe-se a responder a seguinte pergunta: Como os serviços estudados podem se articular de forma integrada com a futura plataforma SAVITS apresentada no projeto?

2.1 COBIB: Serviços bibliográficos e a base de padronização informacional

A COBIB reúne os serviços de tradição mais consolidada do ecossistema Ibict e constitui, sob a perspectiva da biblioteconomia de dados, a camada de padronização informacional sobre a qual o SAVITS poderá apoiar a consistência semântica do Caso Piloto. Integram este eixo a *Rede Bibliodata*, o Catálogo Coletivo Nacional (CCN), o Programa de Comutação Bibliográfica (*Comut*), o Centro Brasileiro do ISSN e o Projeto *Pinakes*, que atualmente reestrutura e moderniza esse conjunto.

A Rede Bibliodata, como iniciativa cooperativa de catalogação, materializa a aplicação sistemática de padrões de representação descritiva e de vocabulários controlados em escala nacional. Essa experiência acumulada é diretamente pertinente à investigação e à análise da consistência metodológica de descritores, vocabulários controlados e tesauros para o Caso Piloto. Os padrões e as autoridades já consolidados na Bibliodata oferecem referência empírica para mapear a representatividade temática e a padronização informacional que o projeto demanda. De modo análogo, o Centro Brasileiro do ISSN, ao gerir um identificador internacional unívoco para publicações seriadas, exemplifica em escala operacional o princípio da identificação persistente, fundamento que se estende, no escopo do projeto, ao mapeamento de compatibilidade com identificadores como dARK, ORCID e ROR.

O Projeto *Pinakes* ocupa posição central neste eixo. Ao empregar inteligência artificial e tecnologias da web semântica para modernizar CCN, *Bibliodata* e *Comut*, ele persegue exatamente o tipo de interoperabilidade semântica que constitui requisito metodológico do SAVITS. O intercâmbio com a COBIB, nesse ponto, é de mão dupla, pois a experiência do *Pinakes* em recuperação semântica e reestruturação de bases bibliográficas informa a definição dos requisitos de curadoria e dos guias conceituais a serem elaborados, ao passo que o arcabouço conceitual de dados estruturado para o Caso Piloto pode dialogar com os vocabulários e modelos adotados na modernização desses serviços.

Por fim, o Comut e o CCN, voltados à localização e ao intercâmbio de documentos e publicações seriadas, evidenciam a importância de metadados consistentes e interoperáveis como condição para a circulação da informação. Esta premissa sustenta a documentação FAIR e os dicionários de dados a serem estruturados no âmbito da pesquisa. Assim, o eixo COBIB fornece ao SAVITS um repertório metodológico consolidado de padronização, sobre o qual se assenta o trabalho de curadoria e representação dos ativos informacionais do Caso Piloto.

2.2 CODIC: Tratamento analítico, dados de pesquisa e interoperabilidade sistêmica

A CODIC abriga os serviços que mais se aproximam, em natureza e finalidade, daquilo que o SAVITS pretende realizar. Enquanto a COBIB sustenta a base de padronização, a CODIC opera na camada de tratamento analítico e de gestão de dados de pesquisa, território no qual a atuação do bibliotecário de dados encontra sua expressão mais plena. Integram esse eixo a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), o Repositório *Aleia*, o Repositório *Deposita Dados* e o *BrCris*, além das iniciativas estruturantes de Ciência Aberta conduzidas pela coordenação.

O *BrCris* constitui a referência mais imediata para o SAVITS dentro do ecossistema. Sendo a implementação brasileira de um *Current Research Information System*, ele reúne, integra e disponibiliza dados científicos provenientes de fontes

heterogêneas, modelando suas entidades pela Ontologia *VIVO* e empregando identificadores únicos para vincular e deduplicar registros oriundos do *Lattes*, da CAPES, de bases de patentes e de fontes internacionais como *Wikidata*, *OpenAIRE* e DOAJ. Essa arquitetura dialoga de forma direta com a atribuição de mapear requisitos de compatibilidade e interoperabilidade sistêmica, uma vez que o *BrCris* já resolve, em ambiente de produção, parte dos desafios de integração que o Caso Piloto enfrentará. A experiência acumulada na conversão de metadados entre padrões distintos e na adoção de identificadores persistentes oferece um modelo concreto a partir do qual o arcabouço conceitual de dados do projeto pode ser organizado, subsidiando cientificamente as ações iniciais das Metas 1, 2 e 3.2.

Os repositórios *Aleia* e *Deposita Dados* acrescentam a esse eixo a dimensão da curadoria e da preservação digital de dados de pesquisa. O *Aleia*, certificado internacionalmente pelo *CoreTrustSeal*, evidencia a aplicação prática dos princípios FAIR e dos requisitos de confiabilidade que devem orientar a documentação e a curadoria dos ativos informacionais do projeto. O *Deposita Dados*, por sua vez, ao acolher a produção de pesquisadores ainda sem repositório institucional próprio, ilustra estratégias de democratização do acesso e de fomento à criação de novas infraestruturas, preocupações que se alinham ao caráter replicável previsto para o SAVITS. Ambos oferecem parâmetros valiosos para a definição dos requisitos metodológicos de curadoria e preservação digital, bem como para a estruturação dos dicionários de dados e da documentação FAIR sob responsabilidade do bibliotecário de dados.

A BDTD complementa o conjunto ao demonstrar, em escala nacional, a viabilidade de um modelo agregador no qual o Ibict coleta e harmoniza metadados enquanto os documentos permanecem custodiados em suas instituições de origem. Esse arranjo de coleta automatizada e padronização de metadados, sustentado por protocolos de interoperabilidade, fornece referência metodológica para pensar como o SAVITS poderá incorporar dados distribuídos sem comprometer a consistência semântica. Articulado aos demais serviços da CODIC, o eixo evidencia que o SAVITS encontra no

Ibict uma infraestrutura madura de tratamento e disseminação de dados científicos, sobre a qual o trabalho de interoperabilidade, identificação persistente e curadoria conceitual poderá se desenvolver com solidez metodológica.

Considerados em conjunto, os dois eixos revelam que a articulação do SAVITS com o ecossistema Ibict se sustenta em uma cadeia informacional contínua. A padronização descritiva, os vocabulários controlados e a identificação persistente consolidados no âmbito da COBIB, hoje renovados pelo Projeto *Pinakes*, constituem a camada inicial que assegura consistência aos dados. Essa consistência, por sua vez, viabiliza o tratamento analítico, a integração de fontes heterogêneas e a curadoria conduzidos pela CODIC, da qual o *BrCris* é a expressão mais avançada. Os dados assim padronizados e tratados tornam-se aptos a alimentar o SAVITS com a qualidade semântica e a interoperabilidade que sua camada preditiva exige, de modo que cada eixo prepara as condições para o seguinte e o conjunto opera como um fluxo único. Essa convergência situa o trabalho do bibliotecário de dados em um ponto estratégico do projeto, no qual a consistência semântica herdada da tradição bibliográfica encontra as exigências de interoperabilidade sistêmica e documentação FAIR próprias da gestão contemporânea de dados científicos. É a partir desse diagnóstico que se delineiam as oportunidades de aprimoramento e as sugestões apresentadas a seguir.

3 OPORTUNIDADES E SUGESTÕES

As proposições reunidas nesta seção resultam fundamentalmente da imersão realizada nas duas últimas semanas junto à documentação e aos produtos de informação do Ibict, razão pela qual se apresentam com caráter primário. Esse caráter merece explicitação. Os serviços e sistemas analisados encontram-se, em sua maioria, consolidados e maduros, ainda que parte de suas funcionalidades permaneça em estado de elaboração, como demonstra a própria modernização em curso pelo Projeto *Pinakes*. Da mesma forma, a equipe executora ainda constrói a clareza necessária quanto aos

objetivos práticos para a execução das demandas do SAVITS. Por essas razões, as sugestões a seguir devem ser compreendidas mais como um produto reflexivo do período de ambientação do que como diretrizes de direcionamento do projeto, cuja consolidação dependerá do amadurecimento coletivo das metas e da definição progressiva do escopo do Caso Piloto.

Feita essa ressalva, a principal oportunidade identificada consiste em ancorar o *BrCris* como fundamento conceitual e arquitetural para o desenvolvimento do SAVITS. Durante a imersão, a exploração direta da plataforma permitiu realizar buscas por entidades de diferentes naturezas e avaliar a forma como os resultados são dispostos e relacionados. Foram consultadas, por exemplo, entidades do tipo pessoa, incluindo meus orientadores de mestrado e de doutorado, Sônia Elisa Caregnato e Eliseo Berni Reategui, respectivamente, além do meu próprio registro. A busca alcançou também grupos de pesquisa, com a localização do GTech.Edu, e programas de pós-graduação, dimensão na qual constatei a ausência de dados do Programa de Pós-Graduação em Informática na Educação (PPGIE), observação que, em si, já aponta uma oportunidade de aprimoramento da cobertura de fontes da plataforma. A disposição dos resultados mostrou-se clara e bem estruturada, evidenciando a maturidade do *BrCris* no tratamento e na visualização de dados científicos, qualidade que decorre de sua arquitetura de integração. Uma vez que, independentemente do formato de origem, os dados tratados são reunidos em um único banco e, por meio de identificadores únicos, os conjuntos são vinculados e deduplicados, viabilizando a interoperabilidade independentemente da fonte e do formato, o *BrCris* oferece ao SAVITS um modelo já validado em ambiente de produção para os desafios de integração que o Caso Piloto enfrentará. Considerando que o SAVITS também terá de reunir fontes heterogêneas e assegurar consistência semântica, a adoção do *BrCris* como referência reduz incertezas metodológicas e acelera as ações iniciais das Metas 1, 2 e 3.2.

A essa oportunidade central soma-se o Projeto *Pinakes*, cuja contribuição se mostrou igualmente instigante. A modernização que ele conduz parte da construção de

uma ontologia específica, ancorada em modelos consolidados internacionalmente, na medida em que a estrutura conceitual foi desenvolvida com base na definição das entidades e atributos provenientes do Library Reference Model (LRM), desenvolvido pela IFLA. A forma de visualização das entidades que compõem essa ontologia e o vocabulário controlado chamou particularmente a atenção. Durante a exploração da plataforma, surpreendeu-me constatar o uso da ferramenta de acesso aberto *Annif* como propulsora das indexações automáticas realizadas nesse âmbito, recurso que evidencia o quanto a indexação assistida por aprendizado de máquina já se incorpora às práticas de representação temática do Instituto. A convergência entre a experiência do *Pinakes* em representação semântica de domínios bibliográficos e a experiência do *BrCris* em integração de dados de pesquisa configura, a meu ver, um repertório metodológico valioso para o exercício do bibliotecário de dados no projeto. A combinação dos modelos dessas duas iniciativas pode fundamentar tanto a definição de vocabulários controlados e descritores consistentes quanto a organização interoperável dos ativos informacionais do Caso Piloto.

Os repositórios de dados de pesquisa do Ibict acrescentam uma terceira frente de oportunidades, voltada à curadoria e à preservação. O *Aleia* e o *Deposita Dados*, ambos estruturados sobre o software de código aberto *Dataverse*, oferecem fluxos de trabalho já testados que podem inspirar a gestão dos dados do SAVITS. No *Deposita Dados*, por exemplo, o povoamento ocorre por meio de autoarquivamento por parte dos autores, sendo a publicação realizada pela equipe gestora após uma etapa de curadoria, momento no qual o conjunto de dados recebe o identificador persistente DOI. Esse encadeamento, que articula autoria, revisão de metadados e atribuição de identificador persistente, fornece um parâmetro concreto para os requisitos de curadoria e preservação digital previstos nas atribuições da pesquisa. O *Aleia*, por sua vez, ao adotar esquemas de metadados de preservação como *PREMIS*, METS e *Dublin Core* e ao submeter seus conjuntos a avaliações de aderência aos princípios FAIR, demonstra na

prática o padrão de qualidade que se espera replicar no tratamento dos dados do projeto.

Os serviços bibliográficos tradicionais, embora menos diretamente ligados à camada analítica, completam o quadro de oportunidades pela sólida base de padronização que oferecem. O CCN, o Comut e o Centro Brasileiro do ISSN representam décadas de experiência acumulada em representação descritiva, vocabulários controlados e identificação unívoca de recursos, justamente os fundamentos sobre os quais a consistência semântica do Caso Piloto deverá se apoiar. A análise desses serviços, conduzida durante a imersão, reforçou a percepção de que a tradição bibliográfica do Ibict constitui um patrimônio metodológico aproveitável pelo SAVITS, especialmente no que se refere à definição de autoridades e ao mapeamento de compatibilidade com identificadores persistentes como dARK, ORCID e ROR.

Por fim, todas essas oportunidades convergem para o modelo de gestão de dados que o próprio projeto estabelece como obrigatório, os princípios FAIR. Tais princípios, formulados pela FORCE11 e disseminados globalmente pela iniciativa GO FAIR, cujo escritório regional no Brasil é sediado no próprio Ibict, organizam-se em quatro requisitos complementares. Os dados devem ser encontráveis, de modo a serem facilmente localizados por humanos e máquinas por meio de identificadores únicos persistentes e metadados ricos. Devem ser acessíveis, recuperáveis mediante protocolos padronizados, abertos e gratuitos, garantindo inclusive que os metadados permaneçam disponíveis mesmo quando os dados não estiverem. Devem ser interoperáveis, estruturados com linguagens e formatos comuns que permitam sua combinação com outros conjuntos. E devem ser reutilizáveis, documentados com proveniência clara e permissões de uso bem definidas. No âmbito do SAVITS, esses princípios não constituem mera recomendação, uma vez que sua aplicação é compulsória para todo o ciclo de vida dos dados do projeto, ao lado dos princípios TRUST e CARE, e se materializa na exigência de Linked Open Data, de grafos de conhecimento e de identificadores persistentes de nível internacional. É precisamente nesse ponto que as oportunidades aqui levantadas

encontram seu sentido comum, pois tanto o *BrCris* quanto o *Pinakes*, o *Aleia* e o *Deposita Dados* já incorporam, em graus distintos, os fundamentos FAIR que o SAVITS precisará observar de maneira integral.

Cabe registrar, ao final, a motivação e o entusiasmo com que recebo a expectativa de contribuir para esse esforço. A riqueza do ecossistema informacional do Ibict, percebida ao longo da imersão, revela um terreno fértil para o trabalho do bibliotecário de dados, e é com disposição para aprender com as iniciativas já consolidadas, e ao mesmo tempo somar a elas, que encaro as tarefas vindouras do projeto.

4 FONTES CONSULTADAS

ARAKAKI, A. C. S.; COSTA, B. C. C.; MACHADO, D. O. F.; SANTOS, G. K.; MENDONÇA, A. C. N. Integração do modelo CCN/Pinakes ao IFLA Library Reference Model. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 24., 2024, Vitória. *Anais* [...]. Vitória: Ancib, 2025. Disponível em: https://enancib.ancib.org/enancib/pt_BR/article/view/605. Acesso em: 29 jun. 2026.

ARAÚJO, P. C.; LIMA, K. C. R. (org.). *Práticas de ciência aberta*. Brasília, DF: Ibict, 2025. DOI: https://doi.org/10.22477/9788570131966.

DIAS, T. M. R.; SOUZA, M. G.; CARVALHO SEGUNDO, W. L. R.; SILVA, J. B. M.. Como os dados exportados pela plataforma BrCris impulsionam análises bibliométricas e cientométricas. *In*: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 9., 2024, Brasília. *Anais* [...]. Brasília, DF: UnB; Ibict, 2024. DOI: https://doi.org/10.22477/ix.ebbc.373.

DIAS, T. M. R.; SOUZA, M. G.; NEUBERT, P, S.; CARVALHO SEGUNDO, W. L. R.; SILVA, J. B. M. Análises métricas potencializadas: a contribuição da plataforma BrCris na integração de repositórios. *In*: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 9., 2024, Brasília. *Anais* [...]. Brasília, DF: UnB; Ibict, 2024. DOI: https://doi.org/10.22477/ix.ebbc.377.

DIAS, T. M. R.; GABRIEL JUNIOR, R. F.; SILVA, V. S.; SOUZA, M. G.; OLIVEIRA, A. F.; CARVALHO SEGUNDO, W, L. R. Development and implementation of the BrCris and BrCris-T platform: integration of scientific and technological data in Brazil. *Procedia Computer Science*, v. 249, p. 68-77, 2024. DOI: https://doi.org/10.1016/j.procs.2024.11.105.

GABRIEL JUNIOR, R. F.; MENA-CHALCO, J. P.; DIAS, T. M. R. BrCris: evolução e visibilidade do ecossistema de pesquisa brasileiro no contexto da Ciência Aberta. In:

SILVA, F. C. C.; STUEBER, K.; CARVALHO SEGUNDO, W, L. R. (org.). *Ciência aberta no Brasil*: conquistas e desafios. Porto Alegre: Letra1; São Paulo: SciELO, 2025. p. 201-222. DOI: https://doi.org/10.5281/zenodo.15151063.

MOREIRA, J. L. R.; BONINO, L. O.; PIRES, L. F.; VAN SINDEREN, M.; HENNING, P. C. Towards Findable, Accessible, Interoperable and Reusable (FAIR) Data Repositories: improving a data repository to behave as a FAIR Data Point. *Liinc em Revista*, Rio de Janeiro, v. 15, n. 2, p. 244-258, nov. 2019. DOI: https://doi.org/10.18617/liinc.v15i2.4817.

Anexo B – Matriz de requisitos informacionais

A matriz de requisitos completa encontra-se disponível no link: https://docs.google.com/spreadsheets/d/1GYq6pK07tObrwiFxx2rVYwRh_KBYNXG9OxS0_peny_c/edit?usp=sharing

Quadro B.1 – Matriz de requisitos informacionais do SAVITS: identificação e priorização

ID	Categoria	Requisito informacional	Prioridade	Status
RQ-01	Metadados e descrição	Modelar os fluxos informacionais e os workflows de submissão, validação e publicação de dados e relatórios da plataforma	Alta	Não iniciado
RQ-02	Metadados e descrição	Projetar arquitetura de dados e pipelines de ingestão que tornem os dados localizáveis, acessíveis, interoperáveis e reutilizáveis	Alta	Não iniciado
RQ-03	Semântica e vocabulários	Adotar padrões de Linked Open Data e estruturar grafos de conhecimento para integração semântica ao ecossistema nacional	Alta	Não iniciado
RQ-04	Semântica e vocabulários	Definir vocabulários controlados para indicadores de impacto social alinhados aos ODS 3, 4, 8 e 10 e às políticas públicas setoriais	Alta	Não iniciado
RQ-05	Identificadores persistentes	Implementar identificadores persistentes via dARK para modelos de ML e conjuntos de dados	Alta	Não iniciado
RQ-06	Identificadores persistentes	Integrar ORCID para pesquisadores e ROR para instituições em toda a cadeia de descrição	Alta	Não iniciado
RQ-07	Interoperabilidade	Desenvolver módulos de exportação de metadados compatíveis com Oasisbr, BDTD e BrCris	Alta	Não iniciado
RQ-08	Interoperabilidade	Disponibilizar APIs abertas para integração com o ecossistema nacional de informação científica	Média	Não iniciado
RQ-09	Interoperabilidade	Garantir compatibilidade com a Política de Ciência Aberta do Ibict	Média	Não iniciado
RQ-10	Governança e ética	Estabelecer protocolos padronizados de custódia, anonimização, segurança e governança dos dados desde o ponto de coleta	Alta	Não iniciado
RQ-11	Governança e ética	Definir e validar procedimentos padronizados de consentimento livre e esclarecido, contínuo e adaptativo	Alta	Não iniciado
RQ-12	Governança e ética	Garantir rastreabilidade e auditabilidade de todo o fluxo informacional	Alta	Não iniciado
RQ-13	Coleta e fontes	Desenvolver, documentar e validar instrumentos padronizados de coleta que gerem dados estruturados, semiestruturados e qualitativos	Alta	Não iniciado
RQ-14	Coleta e fontes	Mapear fontes de dados complementares (bases públicas e indicadores oficiais) para enriquecer as análises	Média	Não iniciado
RQ-15	Curadoria e preservação	Preparar base de dados pública anonimizada do programa piloto para pesquisa	Média	Não iniciado
RQ-16	Curadoria e preservação	Definir política de preservação de longo prazo da base longitudinal consolidada	Média	Não iniciado
RQ-17	Curadoria e	Versionar código e dados para garantir	Média	Não

ID	Categoria	Requisito informacional	Prioridade	Status
	preservação	reprodutibilidade científica		iniciado
RQ-18	Curadoria e preservação	Elaborar plano de sustentabilidade da plataforma, custódia e gestão de longo prazo dos dados	Média	Não iniciado
RQ-19	Acessibilidade	Aplicar diretrizes de acessibilidade do Ibict e WCAG em todos os produtos, dos instrumentos de coleta às visualizações	Média	Não iniciado
RQ-20	Disseminação e reuso	Publicar algoritmos e documentação técnica em repositório aberto	Baixa	Não iniciado
RQ-21	Disseminação e reuso	Devolver resultados às comunidades participantes em formatos acessíveis, assegurando benefício coletivo	Baixa	Não iniciado
RQ-22	Disseminação e reuso	Produzir relatórios técnico-científicos, protocolos de replicação e manuais metodológicos	Baixa	Não iniciado
RQ-23	Capacitação	Incluir módulos sobre curadoria de dados, viés em dados, privacidade e diversidade nos treinamentos	Baixa	Não iniciado
RQ-24	Governança e ética	Seguir as diretrizes de Biblioteconomia de dados, assegurando integridade e perenidade dos dados e registros bibliográficos e científicos	Alta	Não iniciado

Fonte: elaboração própria (2026).

Quadro B.2 – Matriz de requisitos informacionais do SAVITS: rastreabilidade e produtos associados

ID	Rastreabilidade no projeto 5D	Atribuição	Produto esperado do bibliotecário de dados
RQ-01	Meta 2, Etapa 2.1, Atividade 2.1.1	A1	Diagrama de fluxo informacional com pontos de captura de metadados em cada transição de estado
RQ-02	Meta 2, Etapa 2.1, Atividade 2.1.2 e Fase 2.1 do método	A1, A2	Perfil de aplicação de metadados do SAVITS com elementos obrigatórios, recomendados e opcionais
RQ-03	Diretriz mandatória de Arquitetura e gestão de dados	A1, A3	Recomendação de serializações (RDF, JSON-LD) e de ontologias de referência, avaliando a ontologia Pinakes como camada institucional
RQ-04	Meta 1, eixo Compreender o contexto e as necessidades	A1	Lista comentada de vocabulários candidatos com recomendação de adoção, adaptação ou criação, preferencialmente em SKOS
RQ-05	Meta 2, Etapa 2.3, Atividade 2.3.2 e Resultado R6	A3	Política de identificação persistente com tipologia de ativos e momento de atribuição no fluxo
RQ-06	Diretriz mandatória de	A3	Mapeamento dos campos de autoria e afiliação nos

ID	Rastreabilidade no projeto 5D	Atribuição	Produto esperado do bibliotecário de dados
	Identificadores persistentes		esquemas do SAVITS com validação de sintaxe ORCID e ROR
RQ-07	Meta 2, Etapa 2.3, Atividade 2.3.1 e Resultado R5	A3	Tabela de crosswalk entre o perfil do SAVITS e os esquemas alvo (Dublin Core via OAI-PMH e modelo do BrCris)
RQ-08	Diretriz de Interoperabilidade do método de trabalho	A3	Especificação dos requisitos informacionais das APIs (formatos de resposta, esquemas, protocolos como OAI-PMH)
RQ-09	Meta 2, Etapa 2.3, Atividade 2.3.3	A2, A3	Checklist de conformidade da plataforma com a política institucional
RQ-10	Meta 1, Etapa 1.2, Atividade 1.2.3	A2	Seção de governança no guia de curadoria, com procedimentos de anonimização e pseudonimização conforme LGPD
RQ-11	Meta 1, Etapa 1.2, Atividade 1.2.2	A2	Requisitos de registro do consentimento como metadado vinculado ao dado coletado (versão do termo, data, escopo)
RQ-12	Meta 1, eixo Implementar governança e ética desde a origem	A1, A2	Especificação de metadados de proveniência (sugerido PROV-O) e de trilha de auditoria
RQ-13	Meta 1, Etapa 1.2, Atividade 1.2.1	A1, A2	Modelo de documentação dos instrumentos (dicionário de variáveis, codebook) apto a alimentar ML e LLM
RQ-14	Meta 1, eixo Conhecer os dados e suas características	A1, A3	Inventário de fontes externas com esquema, licença, forma de acesso e chave de ligação semântica
RQ-15	Meta 3, Etapa 3.2, Atividade 3.2.2 e Resultado R8	A2	Protocolo de curadoria para publicação (anonimização, licenciamento, depósito com PID, sugerido no Deposita Dados ou Aleia)
RQ-16	Meta 3, Etapa 3.2, Atividade 3.2.3	A2	Recomendações de preservação (formatos, checksums, metadados PREMIS) com base nas práticas da Aleia
RQ-17	Meta 3, Etapa 3.1, Atividade 3.1.3	A2	Convenções de versionamento e de citação de dados e modelos (versão como metadado obrigatório)
RQ-18	Meta 3, Etapa 3.2, Atividade 3.2.1	A2, A4	Contribuição com o capítulo informacional do plano (papéis de curadoria, ciclos de revisão de metadados)
RQ-19	Meta 2, Atividade 2.2.3 e diretriz de Acessibilidade integral	A2	Requisitos de metadados de acessibilidade (texto alternativo, formatos acessíveis) nos guias de curadoria
RQ-20	Meta 4, Atividade 4.2.3	A2, A3	Definição do repositório alvo, esquema de descrição de software e licenciamento

ID	Rastreabilidade no projeto 5D	Atribuição	Produto esperado do bibliotecário de dados
RQ-21	Meta 4, Atividade 4.2.2	A2	Diretrizes de linguagem e formato para produtos de devolutiva no guia de curadoria
RQ-22	Meta 4, Atividade 4.2.1 e Resultado R10	A4	Relatório técnico final consolidando padrões adotados, crosswalks, políticas e lições aprendidas
RQ-23	Meta 4, Atividade 4.1.2 e OE6, OE7	A2, A4	Material didático derivado dos guias de curadoria para capacitação de gestores e equipe
RQ-24	Diretriz transversal do método de trabalho	A1, A2, A3, A4	Princípio orientador de todas as entregas. Registrar decisões de curadoria em memória documental do projeto

Fonte: elaboração própria (2026).

Quadro B.3 – Matriz de requisitos informacionais do SAVITS: princípios e marcos associados

ID	FAIR (facetas)	TRUST	CARE	LGPD	DEIA	WCAG
RQ-01	F, A					
RQ-02	F, A, I, R					
RQ-03	I					
RQ-04	I, R				X	
RQ-05	F	X				
RQ-06	F, I					
RQ-07	I					
RQ-08	A, I	X				
RQ-09	F, A, I, R	X				
RQ-10		X	X	X		
RQ-11			X	X	X	
RQ-12	R	X				
RQ-13	R				X	
RQ-14	I, R					
RQ-15	F, A, I, R		X	X		
RQ-16	A, R	X				
RQ-17	R	X				
RQ-18		X				
RQ-19					X	X
RQ-20	R	X				
RQ-21			X		X	
RQ-22		X				

ID	FAIR (facetas)	TRUST	CARE	LGPD	DEIA	WCAG
RQ-23		X	X		X	
RQ-24	F, A, I, R	X				

Fonte: elaboração própria (2026).
Nota: a coluna de observações analíticas de cada requisito é preservada na planilha eletrônica que origina estes quadros, mantida no acervo do projeto.

Anexo C – Nota técnica - comparação MICS - SAVITS

Confronto analítico entre a estrutura de indicadores do projeto MICS e a matriz de requisitos informacionais do SAVITS[3]

1 Finalidade e enquadramento

Esta nota confronta a estrutura de indicadores de impacto do projeto europeu MICS (*Measuring the Impact of Citizen Science*) com a matriz de 24 requisitos informacionais elaborada a partir do documento de projeto 5D do SAVITS.

Cabe uma desambiguação preliminar. A sigla MICS designa também o programa *Multiple Indicator Cluster Surveys*, iniciativa de inquéritos domiciliares criada pela Unicef nos anos 1990 para produzir estimativas sobre a situação de crianças, mulheres e famílias nas áreas de saúde, educação e proteção infantil. Os dois programas não guardam relação entre si. Toda menção a MICS nesta nota e nos demais documentos do SAVITS refere-se exclusivamente ao projeto europeu de avaliação de impacto em ciência cidadã, financiado pelo programa *Horizon 2020* e documentado em *about.mics.tools*. Como o programa da Unicef é anterior, de maior escala e tende a predominar nos resultados de buscas pela sigla isolada, recomenda-se que os documentos do projeto grafem sempre a sigla acompanhada da expansão completa na primeira ocorrência de cada texto. O exercício se justifica pela menção nominal ao MICS no próprio documento 5D, que o apresenta como demonstração da viabilidade de estruturar sistemas de avaliação com amplo conjunto de indicadores e como referência para a construção das

3 Nota técnica exploratória produzida durante o período de imersão. Sujeita a validação pela coordenação do projeto antes de incorporação a entregas formais.

dimensões analíticas e categorias de impacto do SAVITS. O confronto tem dupla função. Serve como teste de consistência externa da matriz, verificando se os requisitos formulados resistem ao contraste com um sistema efetivamente implementado, e produz subsídios antecipados para o relatório técnico final previsto na atribuição contratual A4.

2 A estrutura de indicadores do MICS

O MICS foi um projeto financiado pelo programa *Horizon 2020* (acordo nº 824711, 2019 a 2022) que resultou em uma plataforma de avaliação de impacto de projetos de ciência cidadã. A avaliação se organiza em cinco domínios, a saber, sociedade, governança, economia, ciência e tecnologia, e meio ambiente. Cada domínio recebeu inicialmente dois indicadores baseados em regras, totalizando dez. No domínio da sociedade estão atividade cognitiva (*activeness*) e envolvimento (*involvement*); no de governança, política pública (*policy*) e Objetivos de Desenvolvimento Sustentável; no econômico, produtividade econômica e sustentabilidade financeira; no de ciência e tecnologia, produtividade científica e interdisciplinaridade; no ambiental, pegada ambiental e consciência ambiental.

A mecânica de cálculo é transparente e publicada integralmente. O instrumento reúne cerca de 200 perguntas com respostas pré-definidas. Conjuntos de métricas sobre um mesmo tema são combinados em um indicador, cuja pontuação, de 0 a 42, resulta da soma dos pesos atribuídos às opções de resposta selecionadas. Faixas de pontuação acionam recomendações textuais específicas, transformando o diagnóstico em orientação para melhoria. Sobre essa camada determinística, a plataforma acrescenta uma pontuação experimental por domínio calculada por rede neural treinada com dados de 30 projetos, cujos resultados os próprios autores declaram inexplicáveis, apresentando-os como estimativa aproximada do impacto.

Cada indicador é documentado com uma anatomia constante que compreende definição formal, perguntas associadas, justificativa ancorada em literatura revisada por pares ou em documentos normativos (como as características de ciência cidadã da ECSA e as metas dos ODS), tabela de pesos e faixas de recomendação. Todo o conteúdo é licenciado em *Creative Commons Attribution-ShareAlike 4.0*.

3 Natureza dos dois artefatos

O confronto exige reconhecer primeiro que os dois artefatos operam em camadas distintas. Os indicadores do MICS constituem a camada substantiva da avaliação, ou seja, definem o que se mede quando se fala de impacto. A matriz de requisitos do SAVITS constitui a camada infraestrutural, ou seja, define como aquilo que se mede deve ser descrito, identificado, curado, preservado e interoperado. As camadas são complementares em vez de concorrentes, e é precisamente essa ortogonalidade que torna o exercício produtivo. O MICS oferece um exemplo maduro do conteúdo semântico que a infraestrutura descrita na matriz precisará suportar, enquanto a matriz descreve a infraestrutura que o MICS, por seu recorte, deixou de especificar.

4 Correspondências entre indicadores do MICS e requisitos da matriz

O quadro a seguir sintetiza as correspondências identificadas, comentadas na sequência.

Quadro 1 - Relações MICS - SAVITS

Elemento do MICS	Requisito(s) da matriz	Natureza da relação
Taxonomia de cinco domínios e dez indicadores	RQ-04	Fonte candidata de vocabulário controlado
Anatomia documental dos indicadores (definição, perguntas, pesos, faixas)	RQ-13	Modelo de codebook publicado

Indicador de ODS (governança)	RQ-04	Sobreposição parcial de metas (ODS 8 comum; SAVITS acrescenta 3, 4 e 10)
Página de projeto padronizada na plataforma	RQ-01, RQ-02	Captura de metadados descritivos acoplada ao instrumento de avaliação
Barreira de política pública por desconfiança nos dados	RQ-07, RQ-08	Validação externa da centralidade da interoperabilidade
Publicação aberta do instrumento e licença CC BY-SA	RQ-20	Precedente de abertura de algoritmos e instrumentos
Indicador de sustentabilidade financeira e plano de exploração	RQ-18	Convergência sobre sustentabilidade de longo prazo
Pontuação por rede neural declaradamente inexplicável	RQ-12	Contraste que fortalece o requisito de auditabilidade

Fonte: Dados da pesquisa.

A correspondência de maior rendimento imediato envolve o RQ-04, que prevê a definição de vocabulários controlados para indicadores de impacto social. Os dez indicadores do MICS, com suas definições formais e ancoragem bibliográfica, constituem exatamente o tipo de fonte que a lista comentada prevista naquele requisito deve avaliar sob as alternativas de adoção, adaptação ou criação. A sobreposição com os ODS é apenas parcial. O MICS ancora indicadores nas metas 8.2, 9.4, 12.2, 12.7, 12.8 e 13.3, enquanto o SAVITS elege os ODS 3, 4, 8 e 10. A interseção restrita ao ODS 8 sugere que o caminho será a adaptação da arquitetura com substituição do conteúdo, aproveitando a forma do indicador MICS e repovoando-a com metas pertinentes a tecnologias sociais.

A segunda correspondência relevante é estrutural. A anatomia constante com que o MICS documenta cada indicador equivale, em termos de biblioteconomia de dados, a um codebook publicado, com a particularidade de incluir a camada prescritiva das recomendações. O RQ-13 prevê modelo de documentação de instrumentos apto a alimentar aprendizado de máquina e modelos de linguagem, e a anatomia do MICS oferece um gabarito testado do que esse modelo precisa conter. Registre-se ainda que essa anatomia é diretamente serializável em SKOS, com o domínio como esquema de

conceitos, o indicador como conceito e as perguntas como propriedades associadas, o que conecta o achado ao RQ-04 também no plano técnico.

A terceira correspondência tem valor argumentativo para o relatório final. O material de orientação do MICS registra que a barreira mais comum à influência de projetos sobre políticas públicas reside nas preocupações dos decisores com a qualidade dos dados, e recomenda o alinhamento com os padrões de dados dos formuladores de políticas. Trata-se de respaldo externo, produzido por experiência empírica europeia, à prioridade alta atribuída aos requisitos de interoperabilidade RQ-07 e RQ-08. Para um sistema como o SAVITS, cuja razão de ser é subsidiar a avaliação de políticas públicas, a lição é direta. Sem conformidade com os padrões do ecossistema nacional, os resultados da plataforma enfrentarão a mesma desconfiança que o MICS diagnosticou.

5 O que a matriz cobre e o MICS silencia

O recorte do MICS deixou fora de escopo camadas inteiras que a matriz especifica. Não há no MICS tratamento de identificadores persistentes (RQ-05 e RQ-06), de metadados de proveniência (RQ-12), de preservação de longo prazo (RQ-16), de registro de consentimento como metadado (RQ-11) nem de requisitos de acessibilidade (RQ-19). Essa ausência confirma que a matriz agrega valor além do que a referência internacional oferece, e delimita com precisão a contribuição específica da função de bibliotecário de dados no SAVITS.

O contraste mais instrutivo, porém, está na camada de aprendizado de máquina. O MICS declara que suas pontuações por rede neural não são explicáveis e as apresenta como estimativa aproximada. O SAVITS, por operar sobre avaliação de políticas públicas com implicações distributivas, dificilmente poderia acomodar essa mesma postura. O RQ-12, que exige rastreabilidade e auditabilidade de todo o fluxo informacional, ganha com esse contraste uma justificativa comparativa concreta, pois posiciona o SAVITS

como avanço deliberado sobre a limitação que sua principal referência internacional reconhece em si mesma.

6 O que o MICS sugere e a matriz ainda não formaliza

O confronto revelou três lacunas na matriz, apresentadas aqui como candidatas a novos requisitos ou a observações em requisitos existentes, a critério da coordenação.

A primeira diz respeito à avaliação como evento informacional datado e versionado. A plataforma MICS permite avaliar um projeto do planejamento à conclusão e acompanhar a evolução do impacto no tempo, o que pressupõe que cada aplicação do instrumento seja registrada com data, versão do questionário e estado do projeto. A matriz trata do versionamento de código e dados (RQ-17) e do consentimento versionado (RQ-11), mas não formaliza a avaliação em si como objeto descrito por metadados próprios. Sem esse registro, a série longitudinal prevista na Meta 3 perde comparabilidade interna.

A segunda envolve a recomendação como objeto informacional. No MICS, faixas de pontuação disparam recomendações textuais fixas. No SAVITS, com a camada de interpretação por modelos de linguagem prevista no OE4, as recomendações serão geradas dinamicamente, o que torna ainda mais necessário descrevê-las com metadados de geração (modelo utilizado, versão, data, insumos considerados). A matriz não contém requisito para essa classe de objeto.

A terceira trata da documentação dos modelos preditivos. O treinamento da rede neural do MICS com apenas 30 projetos ilustra o risco de bases de treinamento pequenas e pouco documentadas. O RQ-05 prevê identificadores persistentes para modelos e o RQ-20 menciona descrição de software, mas nenhum requisito especifica a documentação do modelo em si, na linha de padrões emergentes como model cards e datasheets for datasets. Dado o peso do OE3 no projeto, a lacuna merece registro.

7 Implicações para o relatório final

Três usos deste confronto no relatório previsto na atribuição A4 parecem imediatos. O primeiro é fundamentar a seção de referências internacionais com análise própria em vez de resenha, demonstrando apropriação crítica da fonte que o documento 5D indicou. O segundo é justificar decisões de priorização da matriz, em especial a prioridade alta da interoperabilidade, com evidência externa. O terceiro é registrar as três lacunas identificadas como evolução documentada da matriz, o que materializa a diretriz do RQ-24 de preservar a memória decisória do projeto.

8 Limites do exercício

Quatro limites condicionam a leitura desta nota. A matriz confrontada é hipotética e não foi validada pela coordenação. O MICS avalia projetos de ciência cidadã enquanto o SAVITS avaliará impactos de tecnologias sociais e políticas públicas, de modo que as transposições exigem adaptação conceitual e nunca importação direta. A análise se baseou na documentação pública do MICS disponível em julho de 2026, sem acesso aos entregáveis internos do consórcio, que podem detalhar decisões não refletidas no sítio. Por fim, a camada de 200 perguntas foi examinada por meio de sua síntese em indicadores, restando como desdobramento natural o exame pergunta a pergunta para eventual reuso no desenho dos instrumentos de coleta do RQ-13.

Referências

EU-CITIZEN.SCIENCE. MICS: Measuring the impact of citizen science. 2022. Disponível em: https://eu-citizen.science/project/373. Acesso em: 12 jul. 2026.

MICS. Impact indicators. [S. l.], [2022]. Disponível em: https://about.mics.tools/indicators. Acesso em: 12 jul. 2026.

MICS. The MICS project. [S. l.], [2022]. Disponível em: https://about.mics.tools/project. Acesso em: 12 jul. 2026.

Documento assinado digitalmente
RAFAEL ANTUNES DOS SANTOS
Data: 24/09/2026 12:44:26-0300
Verifique em https://validar.iti.gov.br

Bolsista

De acordo:

Documento assinado digitalmente
WASHINGTON LUIS RIBEIRO DE CARVALHO SEG
Data: 28/09/2026 17:47:56-0300
Verifique em https://validar.iti.gov.br

Coordenador do projeto