

Plano de Gestão de Dados (PGD)

Projeto: Avaliação de ambientes de inovação: modelos, indicadores e práticas.

1. Quais dados digitais serão gerados pelo projeto?

Este projeto irá coletar dados de inteligência competitiva dos ambientes de inovação cadastrados no Sistema Paulista de Ambientes de Inovação (SPA) e/ou na rede de Ambientes Paulistas de Inovação (A.P.I). A Plataforma será um sistema de tecnologia de comunicação e informação para a coleta e o tratamento de dados, quantitativos e qualitativos, para a sistematização do processo de monitoramento e desenvolvimento destes ambientes de inovação, bem como de suas empresas incubadas, graduadas e aceleradas.

O projeto coleta, armazena e utiliza informações, por meio dos seguintes mecanismos:

- Informações coletadas através de manifestação espontânea de respondentes, explicitamente através de formulários eletrônicos disponíveis na plataforma
- Informações coletadas através de mecanismos de aquisição manual ou automatizada de fontes de dados secundárias de domínio público incluindo serviços governamentais, websites eletrônicos ou fontes impressas.
- Informações coletadas por meio de pesquisas qualitativas ou quantitativas não anonimizadas, com consentimento dos participantes.

As seguintes categorias de dados são coletadas e mantidas pelo projeto:

ID	Categoria de Dados	Exemplo
A	Atributos Organizacionais	Razão Social, data de fundação, natureza legal, endereço de Website, CNPJ e outros
B	Dados Cientométricos	Informações referentes a volumes e tipologia de produção acadêmica, intelectual, patentes e quaisquer outras modalidades de conhecimento tecnológico tácito ou explícito derivados de pesquisa científica.
C	Dados Econométricos	Informações de natureza não confidencial, incluindo setores de atuação econômica, portes empresariais, situação fiscal, níveis de investimento, ou outras informações derivadas de atividades econômicas.
D	Dados georreferenciados	Número IP, coordenadas GPS, CEP, localização (cidade, estado, país) referentes a atores, ativos e atributos do Sistema Paulista de Inovação

E	Características protegidas por leis federais e estaduais	Gênero, Raça, Idade, Data de Nascimento, Dados de contato como telefone, email ou endereço de redes sociais, CPF ou outras formas de identificação pessoal (ORCID, etc)
F	Manifestações Espontâneas	Informações complementares, autodeclaradas associadas a qualquer uma das categorias anteriores ou não, qualitativas ou quantitativas e associáveis a perfis do usuário ou de sua organização.

O projeto armazena dados brutos, incluindo dados anonimizados e não anonimizados com consentimento explícito dos participantes. Estima-se que o volume de dados obtidos ao longo de todo o ciclo de execução do projeto seja entre 1 e 2 (TB) terabytes de informações. O compartilhamento de qualquer informação gerada pelo projeto somente ocorre de forma anonimizada e agrupada (macro-dados), sem a possibilidade de identificação de qualquer dos participantes. Não há compartilhamento de dados brutos.

Entre outros objetos digitais resultantes, o projeto materializa uma plataforma de inteligência competitiva para geração de relatórios analíticos sobre os dados obtidos, os quais estarão acessíveis através de endereços Web caracterizados por Localizadores Universais de Recursos (URL). Autoridades, funcionários e colaboradores diretos das instituições participantes deste serão cadastrados na plataforma

2. Como serão gerenciados internamente (cobrindo as etapas de geração, limpeza e armazenamento seguro e confiável)?

Os dados e informações sobre os ambientes de inovação, bem como das empresas e empreendimentos vinculados estarão disponíveis na plataforma para o acesso restrito da Secretaria de Ciência, Tecnologia e Inovação, da Rede Paulista de Ambientes de Inovação e outros parceiros participantes do projeto, em tempo real. Ao mesmo tempo que as informações de acesso livre, estarão no formato de uma vitrine tecnológica para acesso de empresários e empreendedores, investidores, gestores, pesquisadores, agências de fomento, governos e demais interessados nesses ambientes.

A principal função da Plataforma será a de uma vitrine tecnológica, ou seja, demonstrar o potencial de inovação do Estado de São Paulo, contribuindo, cada vez mais, para a atração de novas empresas e geração de novos negócios para os ambientes de inovação e para o desenvolvimento socioeconômico do país e suas regiões.

O protocolo utilizado para preservação digital, incluindo caracterização dos metadados é o padrão internacional PREMIS, baseado no modelo de referência Open Archival Information System (OAIS) (ISO 14721). O modelo de informação OAIS fornece uma base conceitual na forma de uma taxonomia de objetos e pacotes de informações para objetos arquivados e a estrutura de seus metadados associados para manter a viabilidade, capacidade de reprodução, comprehensibilidade, autenticidade e identidade em um contexto de preservação digital.

Todos os dados gerados estarão disponíveis para reuso através de sua organização em formato público CSV (comma separated value), de fácil interoperabilidade e reutilização entre diferentes padrões de bancos de dados estruturados (SQL) e portabilidade para exportação em planilhas eletrônicas e outros mecanismos de tratamento e visualização de dados.

Os arquivos secundários já identificados por DOI e que forem obtidos pelo projeto preservarão sua identidade. Para novos objetos gerados na execução deste projeto não serão gerados identificadores DOI, salvo quando ocorrer publicação de artigos científicos resultantes desta pesquisa, os quais serão associados a identificadores DOI no ato da respectiva publicação científica.

3. Como serão preservados e compartilhados (interna e externamente), considerando questões éticas, legais, privacidade e segurança?

Sobre as questões éticas, o protocolo ético adotado para esta pesquisa é a Resolução CNS Nº 510 (CNS, 2016).

No que concerne a aquisição de dados secundários de domínio público, o Art.1 da CNS nº 510 estabelece normas aplicáveis à pesquisa em Humanidades e Ciências Sociais, nas quais os procedimentos metodológicos utilizam dados primários dos participantes ou dados não anonimizados que exporiam os participantes a níveis de risco mais altos do que os existentes em suas vidas cotidianas. A Resolução CNS n.510 isenta explicitamente projetos de pesquisa baseados em informações de domínio público da obrigação de submissão de análise sistemática pela Comissão Nacional de Ética em Pesquisa (CONEP) ou dos Comitês de Ética em Pesquisa locais (CEP).

No tocante à aquisição de dados primários, o projeto está em conformidade com a Política Nacional de Dados Abertos, estabelecida pelo Decreto n. 8.777 de 11 de maio de 2016, e com a Lei Geral de Proteção de Dados Pessoais, Lei n. 13.709, de 14 de agosto de 2018, visando aplicação com os instrumentos regulatórios vigentes para a garantia das boas práticas de governança em processos de aquisição, armazenamento e recombinação de dados, em especial a utilização de manifestações explícitas de consentimento de coleta, armazenamento e compartilhamento de informações dos participantes da pesquisa.

Visando assegurar privacidade e segurança dos dados, o projeto implementa as seguintes medidas e técnicas para prevenir acesso não autorizado:

- Métodos de Verificação CSRF “cross-site request forgery”, contra falsificação de requisições;
- Mecanismos de prevenção de sessões múltiplas e autenticação centralizada “single sign-on” (SSO);
- Armazenamento criptografado de senhas e informações de identificação pessoal sensíveis;
- Implementação de protocolo SSL (https) como parte integrante da Plataforma;
- Monitoramento de tentativas de acesso à plataforma.

- Métodos de Autenticação de usuários;
- Testes de prevenção de intrusão;

Estes mecanismos de segurança, combinados, visam proteger a confidencialidade, segurança e integridade de dados pessoais de usuários, e a ocorrência de eventuais danos em virtude do tratamento desses dados.

4. Caso o Centro decida criar o próprio repositório institucional de dados, como planeja fazê-lo e gerenciá-lo?

Todos os dados produzidos pelo projeto, serão armazenados em nuvem computacional, sob controle direto do Comitê de Gestão de Dados (CGD) deste projeto, formado por representantes formalmente indicados pela da Secretaria de Ciência e Tecnologia do Estado de São Paulo (SCTI/SP) e representantes indicados pelos ambientes de inovação participantes da Rede dos Ambientes Paulistas de Inovação. O compartilhamento interno e externo, será estabelecido a através de política de acesso a ser definida para estas finalidades, considerando as seguintes premissas :

- O compartilhamento interno será efetuado através da definição de perfis especializados de acesso à plataforma analítica a ser desenvolvida no âmbito deste projeto e somente usuários autorizados pelo comitê poderão acessar integralmente os dados gerados pelo projeto, incluindo micro-dados não anonimizados e macro-dados (dados agrupados).
- O compartilhamento externo será público, também efetuado através de perfil “público” de acesso á plataforma analítica a ser desenvolvida no âmbito deste projeto, entretanto em visões limitadas exclusivamente a macro-dados (dados agrupados) e dados anonimizados.

4.1 Gerenciamento dos dados – responsabilidades, preservação e segurança

O Comitê de Gestão de Dados (CGD) nomeará o responsável pelo gerencialmente da plataforma, bem como dados gerados por este projeto. Este Gestor, deverá implementar a Política de Gestão de Dados a ser definida para este projeto, em conformidade com os protocolos e normas regulatórias vigentes para controle de privacidade, ética e segurança de dados públicos, em conformidade especialmente com a resolução CNS n.510 de 2016, a Política Nacional de Dados Abertos, estabelecida pelo Decreto n. 8.777 de 11 de maio de 2016 e a Lei Geral de Proteção de Dados Pessoais n. 13.709, de 14 de agosto de 2018.

Os dados mantidos pelo projeto são cumulativos e não descartados, visando a construção de séries temporais, para construção de painéis estatísticos históricos e análises de tendências.

A monitoramento da execução do Plano de Gestão de Dados (PGD) será atribuição do Comitê de Gestão de Dados (CGD) como parte de suas atividades regulares de acompanhamento da execução e da evolução deste projeto.