

2022

Relatório Anual

Liderança

Crislane Alves (Institucional / Dados)
Diego Oliveira (Infraestrutura)
Fernanda Scovino (Diretora Executiva / Comunicação)
João Carabetta (Infraestrutura)
Matheus Ávila (Captação, Parcerias e Produtos / Financeiro)
Ricardo Dahis (Presidente / Dados)

BD+

Comunicação

Giovane Caruso
Luiza Vilas Boas
Nayara Moraes

Dados

Arthur Gusmão
Gabriel Pisa
Gabrielle Carvalho
Laura Amaral
Lucas Cavalcanti Rodrigues
Lucas Moreira Nascimento
Luiza Vilela Campos
Mauricio Fagundes
Patrick Teixeira

Devops

Guilherme Alves

Plataforma

Lucas Lessa
Nayara Moraes

Captação, Projetos & Parcerias

Captação

Mariana Bonfanti

Novos Produtos

Gustavo Alcântara

Ex-membros

Fábio Souza Vasconcellos (Captação, Projetos & Parcerias)
Fred Israel (Liderança)
Gustavo Aires Tiago (Dados)
Ingrid de Sales Rabelo (Dados)
Isadora Montenegro Bugarin (Infra)
Kaline Gomes Santos (Dados)

Carta da Liderança	5
Nossos números de 2022	7
Produtos	7
Comunidade e Conteúdo	7
Serviços e Parcerias	7
Gestão e Financeiro	8
Equipe	8
Administrativo e Financeiro	8
Resultados financeiros	9
BD+	9
Comunicação	9
Mídias e Redes	10
Nova BDletter	10
Produção de conhecimento	10
Campanhas	10
Eleições 2022	10
Dados de educação	10
Copa do Mundo no Catar	11
Comunidade	11
Eventos	11
Outras atividades	11
Dados	12
Atividades	12
Copa do Mundo FIFA	13
Dados eleitorais	13
Time e atuação	13
Infraestrutura	14
Pipelines	14
Novos pacotes	15
Estabilização e suporte dos pacotes	15
Nova infraestrutura	15
Plataforma	16
Novas funcionalidades da plataforma e website	16
Página Transparência	16

Página Quem somos	16
Página de Estudos de caso	16
Outros	16
CPP	17
Captação	17
Projetos	18
Editais e financiamento internacional	19
Doações	19
Projetos executados	20
Fundação Lemann	20
Fiance	20
Souk Analytics	20
JOTA	21
Parlascan	21
Instituto República	21
Gesto	21
Novos produtos	22

Carta da Liderança

Para a comunidade *databaser*,

Quando a Base dos Dados surgiu tudo era empolgação. A primeira versão do site. O primeiro conjunto de dados no *datalake*. O primeiro doador. O primeiro contrato de serviços. Mas essa é a história de 2020 e 2021. Em 2022 nós transformamos empolgação em consistência e resultados. A história de 2022 merece uma volta de comemoração.

O que aconteceu em 2022?

Formamos um time talentoso, engajado e inspirador. 10 novos colaboradores se somaram a um time de peso que esteve focado em entregar um produto de qualidade para você, usuário da BD. Foram 120 tabelas em nosso *datalake*. E sim, nós sabemos, algumas delas vieram de pessoas como você que trabalharam para presentear outros usuários com dados tratados. Queremos deixar um agradecimento especial para os nossos colaboradores da comunidade.

A BD existe para a sua comunidade, e por isso buscamos nos aproximar de vocês. Foram 6 eventos para interagir, ouvir e compartilhar com vocês um pouco mais da BD – sejam vocês antigos colegas ou parte das mais de 12 mil pessoas que passaram a nos acompanhar em nossas redes e plataformas sociais em 2022.

Nesse ano também tivemos um impacto *quase* imensurável. É impossível redigir nessa carta a diferença que cada uma das mais de 1,8 milhão de consultas realizadas e mais de 250 mil acessos ao nosso site fizeram no Brasil e no mundo, em projetos de pesquisa, matérias jornalísticas, avaliações de impacto e formulação de políticas públicas. O que temos certeza – e muito orgulho – é dizer que, além do uso espontâneo da BD, todas as parcerias que firmamos em 2022 com entidades do terceiro setor trouxeram, e ainda trarão, ainda mais informação, transparência e impacto para a nossa sociedade. 2022 foi, realmente, um ano de muitas realizações.

Nesse momento de reflexão sobre o ciclo que encerramos é inevitável ponderar sobre o nosso futuro. Como reforçar nossas principais qualidades? Como ser mais ágil frente às mudanças por vir? Como sustentar financeiramente um projeto completamente gratuito? Como crescer e se manter relevante? Em resumo: para onde vai a BD?

O foco da Base dos Dados hoje é assegurar que temos a capacidade de ser um serviço autossustentado que provê soluções para universalizar o acesso a dados, de forma cada vez mais eficiente e em escala. E, assim como em 2022, contamos com vocês para realizar o sonho de ser a maior e melhor plataforma de dados públicos do Brasil.

O seu apoio será necessário para atravessarmos 2023 com o mesmo sucesso anterior. Esse ano nos reserva novos desafios: novos experimentos para diversificação da nossa receita por novos produtos, desafios frente à evolução da Lei Geral de Proteção de Dados, e um cenário de retomada, ainda que lenta, do investimento e foco em ciência e dados no Brasil. Nosso papel é mover a comunidade na construção de algo ainda maior, sustentável e democrático.

Seguimos sonhando. Seguimos em frente.

Obrigado,

Liderança do Instituto Base dos Dados

Nossos números de 2022

Produtos

1.821.856

consultas realizadas à BD+

129

tabelas tratadas em 50 conjuntos

268.456

acessos ao site

1

pacote adicionado (Stata)

33

colaboradores no GitHub

494.343

linhas de código líquidas adicionadas

Comunidade e Conteúdo

MATÉRIAS USANDO A BD

24

reportagens

ARTIGOS E TUTORIAIS PUBLICADOS

8

artigos

1

tutorial

PALESTRAS MINISTRADAS

3

palestras

PARTICIPAÇÕES EM EVENTOS

5

eventos sobre dados abertos e transparência

WORKSHOPS E LIVES NO YOUTUBE

1

workshop

1

live

REDES E MÍDIAS SOCIAIS

YouTube

14.600

views

945

novos inscritos

Twitter

6.705

novos seguidores

LinkedIn

3.743

novos seguidores

Discord

511

novos membros

Serviços e Parcerias

7

projetos realizados junto a ONGs ou empresas

Gestão e Financeiro

Equipe

Nossa equipe passou por contratações e reformulações ao longo do ano de 2022. Crescemos enquanto instituição em praticamente todas as perspectivas, inclusive em tamanho do time. Em suma, as mudanças da nossa equipe foram:

- 1 pessoa para a equipe de Administrativo;
- 1 pessoa para a equipe de Comunicação;
- 5 pessoas para a equipe de Dados e/ou Analytics (7 ao todo no ano);
- 1 pessoa para a equipe de Devops (3 ao todo no ano);
- 1 pessoa para a equipe de Plataforma;
- 1 pessoa para a equipe de Captação.

No total, finalizamos o ano com 15 colaboradores contratados, além de 5 voluntários ativos.

Administrativo e Financeiro

Alguns fatos relevantes sobre o nosso administrativo são:

- Conseguimos criar aplicações financeiras para o nosso caixa, de forma a obter resultados financeiros do estoque da capital nas contas correntes do Instituto Base dos Dados;
- Conseguimos a isenção do ITBI (Imposto sobre Transação de Bens e Imóveis), o que ajuda a desonerar nossos custos quando recebemos doações;
- Como visto acima, temos agora uma pessoa focada exclusivamente em nossos processos administrativos (contratos, RH, etc);
- Disponibilizamos nossa contabilidade atualizada mensalmente de maneira transparente em nosso website;
- Tivemos nossa primeira auditoria aprovada para as atividades de 2021, e estamos realizando o mesmo processo para 2022.

Resultados financeiros

Receitas	
Receita de Serviços	\$ 585,734.30
Receita de Contribuições e Doações	\$ 223,164.76
Receita Financeira	\$ 20,669.14
Total das Receitas	\$ 829,568.20

Despesas	
Remuneração por Serviços de Terceiros	\$ (514,771.60)
Despesas Administrativas Diversas	\$ (33,786.58)
Despesas com Localização e Funcionamento	\$ (174.09)
Custos com Vendas	\$ (2,749.50)
Total das Despesas	\$ (551,481.77)

Resultado do exercício	
RECEITAS	\$ 829,568.20
DESPESAS + CUSTO	\$ (551,481.77)
RESULTADO LÍQUIDO DO EXERCÍCIO	\$ 278,086.43

Demonstração do Resultado do Exercício de 01/01/2022 até 31/12/2022

BD+

Nos referimos aos nossos produtos e atuação pública como BD+. É uma forma de diferenciar nossas atividades daquelas que envolvem a prestação de serviços ou produtos monetizados que desenvolvemos. É importante frisar que, mesmo em atividades comerciais, a Base dos Dados reforça não apenas seu caráter absolutamente sem fins lucrativos como também o impacto socioeconômico e ambiental positivo de seus projetos.

Comunicação

O time de Comunicação expandiu a visibilidade da Base dos Dados e trabalhou para cultivar a comunidade que se formou ao redor da organização e do debate sobre a acessibilidade de dados públicos no Brasil. A equipe manteve em 2022 uma produção constante de conteúdo com objetivo de divulgar as principais soluções da Base dos Dados e instruir usuários sobre como explorar a potencialidade de nossos dados.

Mídias e Redes

Nova BDletter

A BDletter, newsletter mensal da Base dos Dados, foi criada inicialmente para manter os usuários atualizados com informações gerais sobre novos conjuntos de dados disponibilizados, eventos e novidades da organização. Em 2022, reformulamos a publicação para ser mais que um boletim informativo, passando a trazer análises inéditas, entrevistas com especialistas e novidades relevantes dentro do universo da análise de dados e transparência pública. A nova versão da newsletter contou com [6 edições](#) em 2022.

Produção de conhecimento

Para tornar dados públicos ainda mais acessíveis, criamos regularmente conteúdos que ajudam usuários a utilizar toda a potencialidade de nossas soluções. Em 2022, foram publicados [7 artigos](#), entre tutoriais e análises detalhadas com dados da organização.

Campanhas

Eleições 2022

A Base dos Dados se tornou uma fonte valiosa de informações relevantes para análises eleitorais em 2022. Com dados do TSE organizados e atualizados automaticamente, foram realizadas campanhas para divulgar essa nova fonte de acesso prático a informações como receitas e despesas de candidatos, pesquisas eleitorais e o perfil de candidatos.

Realizamos uma campanha de marketing com o JOTA para divulgar o projeto [Siga o Dinheiro](#), painel interativo criado para trazer mais transparência e acessibilidade para o financiamento eleitoral. Além do lançamento da plataforma, foram lançadas nas redes sociais algumas análises que demonstram a potencialidade do painel para analisar as despesas e receitas de candidatos com recortes geográficos, raciais, de gênero e muito mais. A campanha também envolveu a publicação de artigos e análises em uma coluna específica na plataforma de notícias do JOTA.

Para demonstrar como usuários podem começar a criar análises eleitorais com os dados do TSE, lançamos uma [apresentação ao vivo](#) no YouTube, com a presença de Arthur Fischer, analista político do Centro de Política e Economia do Setor Público da Fundação Getulio Vargas.

Dados de educação

A parceria entre Base dos Dados e Fundação Lemann resultou na disponibilização de importantes conjuntos de dados sobre a educação brasileira em nosso *datalake* público. Entre

os dados disponibilizados estão os do Sistema de Avaliação da Educação Básica (Saeb), da Avaliação Nacional de Alfabetização (ANA) e do Censo da Educação Básica. Para demonstrar como o usuário pode utilizar esses dados em suas análises, foi articulada uma campanha nas redes sociais com exemplos de análises utilizando dados que tratam do tema.

Copa do Mundo no Catar

Ao longo do período da Copa do Mundo de 2022, no Catar, foi articulada uma campanha nas redes sociais para divulgar os dados históricos do campeonato, tratados e padronizados no *datalake* público da Base dos Dados. A campanha envolveu a publicação de diversas análises com os dados do campeonato e de consultas SQL que puderam ser copiadas e adaptadas por usuários em suas próprias análises.

Comunidade

Temos uma comunidade heterogênea e distribuída por todo o Brasil. A equipe de Comunicação promoveu a participação da Base dos Dados em eventos sobre dados públicos e transparência, além de executar estratégias de engajamento da comunidade através das nossas redes sociais.

Eventos

- **Open Data Day (Curitiba):** A Base dos Dados participou da maratona de inovação cívica que promoveu o engajamento dos cidadãos para usar dados e impactar o desenvolvimento de Curitiba.
- **Cerveja com Dados (Rio de Janeiro):** Apresentamos nosso *datalake* público no evento Cerveja com Dados, organizado pela Open Knowledge Brasil.
- **CODA 2022:** Apresentamos duas atividades na Conferência Brasileira de Jornalismo de Dados e Métodos Digitais.
- **CDPD - LatAM:** Nossa co-fundadora, Fernanda Scovino, mediu o painel Open Data x Data Privacy na CDPD Latam.
- **Meet Up - Think Big Think Data (Campinas):** Apresentamos a Base dos Dados no evento MeetUp, organizado pela Think Big Think Data na Escola Korú, em Campinas.
- **Hacktoberfest:** Participamos do evento internacional Hacktoberfest, que busca incentivar a contribuição em projetos *open source* ao longo do mês de outubro.

Outras Atividades

Além das atividades relacionadas a marketing e comunidade, em 2022 a equipe de Comunicação

auxiliou outras frentes da organização com produção de conteúdo de texto e visual, como elaboração de apresentações focadas na captação de recursos, criação de conteúdo e design para nossa plataforma, revisão de textos e edição de arte em projetos como BD Educação.

Dados

O time de Dados se fortaleceu como *core* da Base dos Dados em 2022. Além de serem os principais responsáveis por subirem novos conjuntos de dados para o nosso repositório, atuaram também no suporte à comunidade, em projetos de *analytics*, e na formulação do modelo experimental da BD Educação.

Atividades

Assim como em 2021, um dos grandes focos do nosso trabalho em 2022 foi a ampliação e sofisticação do nosso *datalake* público, onde disponibilizamos bases públicas já limpas e integradas de forma gratuita.

Durante o ano de 2022, tivemos 1,82 milhão de consultas, sendo:

- 1,71 milhão em SQL, através da interface online do *datalake* no Google BigQuery;
- 67 mil pelo pacote em Python e Stata*;
- 42 mil pelo pacote em R.

A nossa evolução de acessos passou de 13 mil consultas em janeiro para mais de 300 mil consultas apenas em dezembro. Esse crescimento vertiginoso de consultas ao nosso *datalake* pode ser explicado por diversos fatores como crescimento do repositório em si, marketing e manutenção da qualidade e atualização das bases existentes.

Considerando a manutenção da qualidade dos conjuntos de dados disponíveis, tivemos ações como o mutirão de correção de problemas em diversos conjuntos, que ajudou a elevar a régua na diligência e qualidade do nosso trabalho de dados.

Ao longo de 2022, nossa equipe de Dados, com ajuda de usuários e colaboradores, limpou, tratou e compatibilizou mais de 50 conjuntos de dados, adicionando ao todo 129 novas tabelas no *datalake*.

*Os dois pacotes utilizam o mesmo processamento, então, são compilados em conjunto.

Entre as bases adicionadas e atualizadas tivemos algumas que valem destaque:

Copa do Mundo FIFA

Em 2022 tivemos a Copa do Mundo FIFA sediada no Catar. Por esse motivo, adicionamos dois conjuntos de dados sobre o tema: estatísticas das copas do mundo masculinas e femininas. O tema gerou interesse nas mídias sociais a partir das visualizações construídas pela equipe. Esperamos que a inclusão dos dados da categoria feminina seja de interesse para 2023, ano que ocorre a Copa do Mundo feminina.

É importante mencionar que os conjuntos da Copa do Mundo FIFA fazem parte de uma nova padronização de nomenclatura e metadados para dados internacionais – utilizada também em outros conjuntos como o PISA –, crucial para a expansão e alcance da Base dos Dados a nível internacional.

Dados eleitorais

Os dados eleitorais já existiam em nosso *datalake* desde 2021, mas, em ano eleitoral, construímos *pipelines* para automação dos dados tanto do TSE sobre filiação e financiamento de campanha, quanto das pesquisas eleitorais do Poder 360.

Não surpreendentemente, o conjunto de dados do TSE foi o mais acessado do ano, com mais de 341 mil consultas. Esse número é também resultado do painel construído em parceria com o JOTA. O conjunto das pesquisas eleitorais, ficou em terceiro lugar com aproximadamente 186 mil consultas – sendo os microdados do conjunto a tabela mais acessadas do *datalake* durante o ano. Os dois conjuntos representam 29% das consultas de 2022.

Para um relatório completo e detalhado da utilização da nossa plataforma em 2022 ou em períodos específicos, temos um painel público [aqui](#).

Time e atuação

Como dito anteriormente o time passou um crescimento considerável, terminando o ano com 5 pessoas a mais do que começou, significando que mais do que dobrou de tamanho em 12 meses. Houve uma mudança na liderança do time, mas que ocorreu sem maiores problemas ou disrupções graças à qualidade do time de ambas as lideranças.

O time de Dados também se conectou com outras instituições de pesquisa e desenvolvimento de

dados como IEPS (Instituto de Estudos para Políticas de Saúde) e MADE-USP (Centro de Pesquisa em Macroeconomia das Desigualdades), criando conexões institucionais com os mesmos.

Vale mencionar o papel da equipe enquanto executor dos projetos descritos mais à frente atuando da concepção à entrega final do projeto. O time de *Analytics*, braço apartado do time de Dados, é crucial para o impacto do time de Dados em outras instituições além da Base dos Dados.

Finalmente, tivemos uma participação surpreendente no Coda.br (Conferência Brasileira de Jornalismo de Dados e Métodos Digitais). Entre as dezenas de workshops organizados pela Escola de Dados e diversos parceiros, tivemos presença importante. Ministramos dois minicursos, outros 7 minicursos utilizaram a plataforma da Base dos Dados como ferramenta necessária para sua execução. Tudo isso em um ambiente presencial em que fomos reconhecidos por dezenas de pessoas presentes.

Infraestrutura

As funções básicas desse time durante 2022 foram:

- Organizar e operar a nossa infraestrutura de cloud / GCP (Google Cloud Platform);
- Manter os pacotes de dados (Python, R e Stata) para acesso e subida de dados em nosso *datalake*;
- Suportar o time de Plataforma (antigo Website), organizando e operando o back-end do site, e mantendo-o atualizado e estável;
- Integrar os serviços do GCP que utilizamos em nossa plataforma;
- Construir *pipelines* em cooperação com o time de Dados.

Os principais entregáveis do time durante de 2022 foram:

Pipelines

O time trabalhou em substituir os processos semi manuais de subida de dados para o nosso repositório por *pipelines* automatizadas. O objetivo geral era ter maior robustez e automação no processo de atualização dos conjuntos de dados existentes na Base dos Dados.

A automação foi focada nos conjuntos utilizados em projetos pelo time de Analytics, de forma a garantir um nível de excelência no serviço prestado. A meta era implementar tal substituição em

60% dos conjuntos prioritários. O time atingiu 65% da meta, ou seja, 39% dos conjuntos estratégicos foram migrados.

Novos pacotes

Em 2021 construímos nossos pacotes em Python e R, uma forma alternativa de consulta ao repositório da Base dos Dados via SQL no BigQuery. Em 2022 o time buscou ir ainda mais além com novos pacotes para deixar tudo ainda mais acessível.

Lançamos então o pacote Stata. A linguagem possui bastante penetração no meio acadêmico – nicho com forte presença entre os usuários da Base dos Dados – e representa então mais um passo para a popularização do uso da plataforma.

Estabilização e suporte dos pacotes

Além de lançar novos pacotes, uma parte importante de toda operação é a manutenção dos produtos e ferramentas existentes. Os nossos pacotes em Python e R receberam então ajustes e melhorias pontuais para sua estabilização.

Além disso, trabalhamos para manter o suporte ao pacote Python ágil e prestativo. Nossa meta é ter uma taxa de resposta em até 7 dias. Em 2022 ainda tivemos problemas para atingir a meta estabelecida, estando, segundo nossas estimativas, com uma média de 14 dias para retorno do suporte sobre o pacote. Já sobre o tempo de resposta para suporte aos metadados, conseguimos atingir essa meta, com aproximadamente 7 dias para resposta.

Nova infraestrutura

O crescimento da Base dos Dados implica em preocupações e complexidades crescentes na hospedagem e manutenção dos dados. Por isso, uma das principais atividades do time de Infraestrutura foi a implementação da infraestrutura em Kubernetes, transformando nossa infraestrutura em uma versão mais escalável e elástica. Com essa implementação, conseguimos manter a taxa de utilização dos recursos entre 50% e 75%, nível considerado seguro e eficiente em termos de custos e *performance*.

Além disso, realizamos implementação parcial do acompanhamento histórico da utilização dos recursos e das aplicações presentes no nosso *cluster* e no nosso painel de custos. Foi implementado também um sistema de alerta pelo Discord para eventuais problemas ou mudanças inesperadas.

Plataforma

O time de Plataforma (anteriormente referenciado como Website) é responsável por desenvolver e manter o nosso site, que é a ferramenta pela qual os usuários utilizam o nosso mecanismos de busca e são levados à consulta do nosso *datalake* propriamente dito pelos pacotes ou BigQuery. Nosso website teve 268.456 acessos em 2022, por 188.811 usuários únicos.

Em 2022 o foco do time foi elevar a barra nos padrões de qualidade visual além de utilizar o nosso site como plataforma de visibilidade de diferentes informações para além dos nossos dados.

Novas funcionalidades da plataforma e website

Página Transparência

Um dos pontos mais importantes do ano foi a criação da nossa página de transparência. Nela está contido um resumo da nossa contabilidade, além de referências para o relatório de atividades existentes. Foi um marco importante para exemplificar nossa missão e visão sobre transparência sobre todos os aspectos que nos rodeiam.

Página Quem somos

Presente na maioria das organizações, a página institucional apresenta mais informações sobre nossa missão, visão, história da Base dos Dados e seus membros, além de números de acesso e uso. É importante ressaltar essa página como forma de dar visibilidade externa à nossa organização, transparecendo mais confiança e credibilidade, e também internamente trazendo todas as pessoas da equipe para a sua devida visibilidade como parte da organização.

Página de Estudos de caso

Em um trabalho colaborativo com os times de Comunicação e de *Analytics*, a página de estudos de caso é hoje nossa maior vitrine de serviços. Contendo o *case* da Fundação Lemann, ele conta em mais detalhes o impacto que nosso time desenvolve através da Base dos Dados nessas instituições. A presença de marcas parceiras fortes e importantes em nosso site ajuda na construção da credibilidade e qualidade dos nossos serviços.

Outros

Além das adições e melhorias já citadas, houve um trabalho constante de aperfeiçoamento da plataforma focando em sempre otimizar a experiência dos nossos usuários. Alguns exemplos são:

- Reformulação da *landing page* com carrossel que funciona como vitrine dos temas disponíveis em nossos dados, e também como filtros rápidos em nosso mecanismo de busca;
- Além do filtro rápido em carrossel, foram adicionados novos filtros na busca por conjuntos de dados, facilitando uma busca mais específica e agilizada. Entre os filtros adicionados, temos: Cobertura espacial (com segmentação hierárquica), Nível da observação, Frequência de atualização e Qualidade da fonte original;
- Trabalhamos em melhorias na organização e visualização dos metadados e colunas na página de conjunto de dados. Isso ajuda os usuários a entenderem melhor os dados antes de despender maiores esforços de exploração dos dados em nossa plataforma para extrair as mesmas informações – além de diminuir o risco de entendimentos errados;
- Durante o ano também houve uma reavaliação da arquitetura de informação e reforço da marca. Entre as melhorias dessa categoria, tivemos reestruturação do menu de navegação, correções visuais de espaçamento, cor, entre outros, criação de ícones e componentes visuais autorais, sistematização de fontes (são mais de 30 utilizadas na forma e contexto correto);
- Por último, no fim do ano foi iniciado o processo de migração da nossa plataforma do sistema CKAN para o Django com o objetivo de criar mais flexibilidade e possibilidades. Essa migração é foco dos trabalhos do início de 2023.

CPP

O time de CPP (Captação, Parcerias e Projetos) é a vertical da Base dos Dados responsável por (1) angariar financiamento para a operação da instituição, e (2) estender o impacto da instituição por vias além dos usuários finais. A área está dividida entre uma frente de Captação, focada na prospecção de diferentes oportunidades, e outra de Projetos e Produtos, que concebem e executam projetos e novos serviços a fim de diversificar a fonte de renda da Base dos Dados enquanto causamos impacto.

Captação

Iniciamos o trabalho de forma coordenada em Captação e Parcerias em agosto de 2022, quando efetuamos a contratação da nossa coordenadora de captação. Realizamos um diagnóstico do que havia sido feito até aquele momento e de quais seriam as frentes mais promissoras com o objetivo de captar recursos para manter a Base dos Dados a longo prazo.

Nos quatro meses seguintes experimentamos diferentes frentes de trabalho, que serão listadas a seguir. Todas elas têm como gatilho inicial o **fortalecimento da Base dos Dados como instituição de**

referência quando o assunto for engenharia de dados públicos, independente do volume de recursos no curto prazo, pois esse é um pressuposto que favorece a captação em todas as frentes.

É importante pensar na captação de recursos e no estabelecimento de parcerias como ações perenes, que dependem da construção de relações de confiança, seriedade e boa reputação. As melhores oportunidades aparecem se estamos ativos com outros parceiros, com boa capacidade de comunicação e enraizados como uma referência no tema. Em termos quantitativos, criamos um total de 58 oportunidades, concretizando 7 projetos e submetendo 3 editais.

Foi possível estabelecer uma hierarquia inicial e algumas prioridades levando em conta retorno e viabilidade. Optamos por utilizar o Coda, que se mostrou útil para o *workflow* em CPP. Produzimos novo *deck* de captação, versões português e inglês, que também têm sido usado na maioria das prospecções. Dessa forma, iniciamos 2023 com a área mais estruturada e com mais clareza em relação a estratégia e plano de ação.

Projetos

Prospecção de potenciais clientes para a execução de projetos:

1. Para instituições do terceiro setor que precisam de ajuda com infraestrutura ou análise de dados

Prospectamos e identificamos agendas com atuação ampla que requerem ou seriam beneficiadas com inteligência do uso de dados públicos. Esse parece ser o caminho mais promissor e factível para obtenção de recursos a curto e médio prazo. Fizemos algumas aproximações, com boas perspectivas para 2023, quando as instituições entram em fase de planejamento. Algumas agendas com exemplos de instituições:

- Educação ([Projeto Lemann](#), [Fundação Itaú](#), Fundação Educar, Parceiros da Educação)
- Saúde ([IEPS - Instituto de Estudo sobre Políticas de Saúde](#), [Desiderata](#))
- Meio ambiente ([Climate Ventures](#))
- Justiça ([IDDD - Instituto de Defesa pelo Direito de Defesa](#), Fundo Brasil, Rede de Justiça Criminal)
- Cultura ([Fundação Itaú](#), Petrobrás)
- Consultorias ([BMJ](#), [Ponte a Ponte](#))
- Cidadania ([COLAB](#))
- Serviço Público ([Republica BI](#))
- Informação e Pesquisa ([Internet LAB](#))
- Política (Parlamentaria, [Dados do STF](#))

2. Para empresas que precisam de ajuda com infraestrutura de dados e análise de mercado ou em parceria com prestadores de serviços afins

No caso de empresas privadas, notamos que pode haver interesse específico na análise e cruzamento de dados públicos, como é o caso da [Yducs](#) e da [Neotrust](#). Assim com as instituições do Terceiro Setor essas empresas dependem de informações sobre as condições brasileiras em vários setores para fortalecer ou pensar estratégias. Já no caso de prestadores de serviços, podemos nos aproximar dos institutos de pesquisas como o [Instituto Locomotiva](#), que via de regra já possuem um departamento ou setor de engenharia de dados mas não com o alcance e aderência da Base dos Dados. Nesse caso, podemos suprir demandas e facilitar o trabalho desses Institutos. No entanto, de maneira geral, há um interesse menor pelos serviços que podemos prestar.

Editais e financiamento internacional

Essa pode ser uma ótima alternativa, considerando que não utilizamos ainda os mecanismos de obtenção de recurso incentivados. Para o acompanhamento dos editais e grants, existem os portais de divulgação, sendo o maior e mais abrangente o [Portal Prosas](#). Já estamos cadastrados e recebemos alertas quando da divulgação de oportunidades nas áreas de interesse.

Aplicamos para três oportunidades nos últimos dois meses do ano, já recebemos a negativa de duas, mas ainda acreditamos que essa frente possui alto potencial:

- [Artemisia + Umane - 2a Edição | Plataforma de Inovação Aberta em Atenção Primária à Saúde](#)
- [National Endowment for Democracy \(NED\)](#)
- [Ford Foundation](#)

Doações

- PJ: patrocínio de empresas com recursos próprios que tenham interesse no propósito da BD

Prospectar com instituições e empresas com cultura de filantropia e propósitos relacionados à transparência, democracia, acesso à informação. Alguns escritórios de advocacia, por exemplo, se encaixam nesse perfil, como é o caso da [Pinheiro Neto Advogados](#), [Machado Meyer](#), [Demarest, Levy & Salomão Advogados](#). Outras instituições, como o [Instituto Betty & Jacob Lafer](#), que financiam projetos de organizações da sociedade civil em aprimoramento, inovação de gestão e políticas públicas ou ainda instituições financeiras como [Credit Suisse](#) e [Bahia Asset](#).

- PF: doações de pessoas com interesses e visões alinhadas com a BD

Fizemos uma aproximação e orçamento com a [TrackMob](#) para gestão de doação de pessoa física, mas julgamos que não era a prioridade no momento de 2022.

Projetos executados

A área de Projetos é responsável por gerenciar os projetos de parceria assinados entre a Base dos Dados e agentes externos, articulando a execução dos contratos de consultoria e prestação de serviços. Nesse sentido, sua principal função é fazer os contratos acontecerem, coordenando as equipes responsáveis, promovendo reuniões de alinhamento com clientes e cientistas de dados, identificando demandas, prazos e as respectivas entregas definidas nos contratos. Em 2022, a área de Projetos coordenou 7 contratos de curta e média duração:

Fundação Lemann

Com mais de 61 redes de ensino apoiadas e mais de 28 mil escolas envolvidas, implementar um fluxo de análise de dados e indicadores foi fundamental para a [Fundação Lemann](#) compreender melhor o cenário da educação no Brasil, avaliar os programas apoiados e estabelecer metas estratégicas para potencializar o avanço da educação e o impulsionamento de lideranças no país. A Base dos Dados trouxe soluções práticas para tornar isso possível, com a reestruturação da arquitetura dos dados utilizados pela Fundação, tratamento e disponibilização de novas bases de dados em um repositório online, análise das informações coletadas e construção de painéis interativos atualizados automaticamente. O projeto, que conta com um pequeno [descritivo das atividades](#) no nosso site, começou em abril e seguiu ao longo de todo o ano, tendo ainda continuidade em 2023.

Finance

O trabalho para a [Finance](#) consistiu no mapeamento e levantamento de dados sociais e econômicos do Estado de Alagoas. Os dados ajudaram a compor o relatório final produzido pela Finance, cuja função foi mensurar os impactos sociais do acidente geológico ocorrido em 2018, nas proximidades da Lagoa Mundaú. O projeto, iniciado em junho, foi concluído 30 dias depois com entregas de tabelas, gráficos e dados solicitados pela equipe da Finance.

Souk Analytics

A plataforma de cursos online [Souk Analytics](#) passou a disponibilizar cursos de análise de dados utilizando conteúdos produzidos em vídeo pela nossa equipe. A parceria, que começou em julho entre a Souk e a Base dos Dados, permitiu que os alunos da plataforma pudessem utilizar

gratuitamente os [vídeos](#) para aprenderem a usar a linguagem SQL para análise de dados.

JOTA

O projeto com o site jornalístico [JOTA](#) consistiu na criação de uma plataforma de monitoramento dos recursos de campanha eleitoral distribuído pelos partidos aos seus candidatos. Para isso, foram estruturadas bases de dados segundo os recortes definidos pelas equipes do JOTA e da Base dos Dados. Na sequência, foi produzido o painel [Siga o Dinheiro](#), com acesso livre de todos os leitores interessados em acompanhar a distribuição dos recursos de campanha. O projeto, que durou entre agosto e novembro de 2022, contou ainda com a produção de [artigos](#) de análises para o JOTA sobre financiamento de campanhas, além de uma [newsletter](#) com informações e análises de dados.

Parlascan

O projeto consiste em uma parceria da Base dos Dados com a [Dado Capital](#) para a produção de um novo e robusto painel com informações sobre o Congresso Brasileiro e a atuação dos parlamentares. O Parlascan tem previsão de lançamento da sua primeira versão para o início de 2023. Para isso, as equipes de ambas instituições se dividiram em duas frentes. A Base dos Dados trabalhou estruturando dados sociais e regionais das bases eleitorais dos deputados e senadores eleitos, além das receitas e despesas de campanha dos políticos eleitos em 2022. A Dado Capital, por sua vez, trabalhou no front-end do painel do Parla Scan, além de atualizar bases já existentes sobre o perfil dos deputados e senadores. Outras fases do projeto a serem desenvolvidos em 2023 vão contemplar índices e indicadores de desempenho e atuação política dos parlamentares.

Instituto República

O projeto com a [República](#) começou em outubro e foi finalizado em janeiro de 2023. O projeto consistiu em quatro frentes; tais como padronização de variáveis de questionários de fontes primárias, adição de dados dos questionários no *datalake* da Base dos Dados, criação de agregações para analisar variáveis de interesse, além de criação de painéis interativos para gestão de pessoas nos estados brasileiros.

Gesto

O projeto consistiu na produção de três projeções para o Instituto [Gesto](#), com foco em dados de educação dos territórios sob a coordenação da organização. As projeções consistiram nos percentuais de alunos em aprendizagem adequada/suficiente para ensino fundamental anos iniciais e alfabetização. O projeto começou em dezembro de 2022 e foi finalizado em janeiro de 2023.

Novos produtos

Em uma tentativa de (1) diversificar as nossas fontes de renda, sobretudo de maneira escalável; (2) atingir ainda mais pessoas e alavancar o impacto e missão da Base dos Dados, criamos um produto novo e experimental de formação em análise de dados públicos.

No segundo semestre de 2022, fechamos a parceria com a Vetor Brasil após um experimento de um minicurso no primeiro semestre. O produto ainda está em construção e será completado e compartilhado com os alunos (40, na primeira fase) ainda no primeiro semestre de 2023.