

# A COCÔZAP

um guia sobre  
geração cidadã de dados



data\_  
labe



**Por que vocês usam o WhatsApp para coletar as queixas?**

**O que precisa ser perguntado ao morador que quer fazer uma queixa?**

**Como vocês organizam as queixas recebidas?**

Essas são algumas perguntas que sempre ouvimos ao longo destes cinco anos de Cocôzap de quem queria replicar o projeto ou experimentar a geração cidadã de dados em outros temas.

Por isso, **neste documento vamos explicar para você como desenvolvemos os dados que fizeram o Cocôzap ser um projeto de geração cidadã de dados.** Aqui você não vai encontrar a história do projeto nem os resultados dele, mas sim o desenrolo, o como, a **METODOLOGIA** de coleta e armazenamento de dados.

**Pra gente o compartilhamento é fundamental.** Esse processo nos ajuda a contar nossa história e construir memória e ainda cria essa troca contigo, que pode sugerir, criticar ou se inspirar um pouquinho nesta experiência.



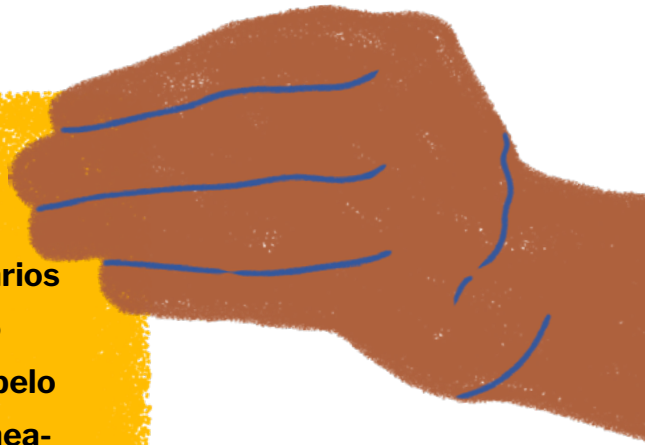
## RESUMÃO:

O Cocôzap é um projeto de geração cidadã de dados e que, portanto, se caracterizou como um levantamento de dados primários (1). No nosso caso, os dados eram exclusivamente sobre irregularidades de saneamento básico no conjunto de favelas da Maré, e

os agentes de coleta eram os próprios moradores. A coleta de dados foi feita por meio de conversa semiestruturada no aplicativo de mensagens *Whatsapp* realizada entre a pessoa que enviou a denúncia e uma pessoa colaboradora do Cocôzap.

## O QUE SÃO DADOS PRIMÁRIOS?

**Pesquisadores diferenciam dados primários e secundários para entender melhor de onde vêm os dados que serão trabalhados. Os dados primários são aqueles gerados pelo próprio projeto. No caso do Cocôzap, as queixas de saneamento básico são geradas pelo próprio projeto e pessoas envolvidas. Já os dados que a gente usa do Instituto Brasileiro de Geografia e Estatística (IBGE) e de outros centros de pesquisa são dados secundários. Embora tratem sobre o mesmo tema, os dados desses órgãos não foram coletados pelo próprio Cocôzap, sacou? Segue o baile, então!**



## POR QUE O WHATSAPP?

A galera que escreveu o projeto pensou muito em qual seria a melhor forma para o morador enviar uma queixa sobre saneamento básico. Muitas ideias foram levantadas, como por exemplo um aplicativo que tivesse os campos para a pessoa anexar a foto e informar a localização. Porém, a maioria das pessoas mal tem internet durante o mês todo para ainda ter que gastar dados baixando app e mandando foto de esgoto a céu aberto.

**Como o *WhatsApp* é um aplicativo que praticamente todo mundo tem e gratuito em muitos planos de internet no celular, a gente achou que seria mais fácil usar uma tecnologia que já estivesse no dia a dia das pessoas.** Outra vantagem que o aplicativo fornecia era a variedade de formatos, visto que as queixas poderiam ser enviadas por texto, áudio e vídeo, o que permitiria alcançar, inclusive, pessoas que não são alfabetizadas.

Esta decisão pode até soar controversa quando comparamos com outros projetos do data\_labe, pois aqui sempre tentamos usar tecnologias livres e abertas. Contudo, entendemos que era importante usar uma ferramenta presente no cotidiano das pessoas e que não demandasse muitas ações para sua participação. Assim, compramos um chip e saímos espalhando o número do celular pela Maré.

**Essa parte deu um trabalho, viu?!** A Maré é simplesmente o 9º maior bairro do Rio de Janeiro, maior que muita cidade. A estratégia foi bem raiz: imprimimos um folder pequeno

que dobrava e cabia no bolso e distribuimos nas casas, associação de moradores, comércios. O bom e velho “boca a boca”, explicando que agora havia um número de *WhatsApp* (2) na favela que coletava queixas de saneamento básico e sobre como é importante gerar dados sobre o território.

O telefone com o chip do Cocôzap ficava na responsabilidade de uma pessoa da equipe, com revezamento entre integrantes. Quando o volume de queixas aumentou, percebemos que era hora de padronizar o recebimento. Então, **criamos um protocolo de uso do WhatsApp do Cocôzap**, que abrangia:










## CUIDADO COM SEU ZAP



Um detalhe bem de bastidor e que quase ninguém percebeu foi que, em 2021, o nosso número de WhatsApp foi cancelado por falta de recarga. Tu acredita? Logo a gente, que fala de tecnologia todos os dias da nossa vida, esqueceu de ficar fazendo recarga no celular, porque sempre usávamos o celular por conexão via wi-fi. Descobrimos num dia em que não dava nem para enviar nem para receber mensagens, daí tivemos que comprar outro chip, imprimir novos materiais de divulgação e sair por aí avisando que o “Cocôzap estava de número novo”. Deu muito trabalho, mas acabou dando tudo certo. Então, se você também tem um projeto de geração cidadã de dados via *WhatsApp*, se liga, hein?! Recarga nesse número!

## A CONSTRUÇÃO DA QUEIXA

Durante a conversa entre o morador realizador da denúncia e a pessoa do próprio Cocôzap, era

necessário **identificar informações básicas sobre o problema:**

- 1  **O tipo de denúncia: se era lixo, água ou esgoto;**
- 2  **A localização do problema;**
- 3  **A regularidade do problema;**

- 4  **Foto da denúncia;**
- 5  **Um relato expandido sobre o problema que era opcional.**

Essa coleta era feita de forma livre, ou seja, a conversa começava de acordo com a vontade do morador, e a gente ia desenrolando e

perguntando todas as informações que precisávamos. Veja um exemplo a seguir:

## EXEMPLO FICTÍCIO DE CONVERSA COM O MORADOR (baseado em fatos reais)

Nessa etapa, o morador manda a queixa pelo número do COCÔZAP e precisamos respondê-lo o quanto antes. Essa conversa precisa ser guiada de modo que se consiga gerar os dados para o mapeamento da queixa. Além disso, o morador precisa ter sua denúncia acolhida entendendo os objetivos do projeto.

Exemplo de Recebimento de Queixa e Padronização de Respostas:



**COCÔZAP:**

/Auxiliar o morador no processo de resolução inicial do problema/  
**(item 2.1)**

Você gostaria de fazer parte da nossa lista de transmissão e receber conteúdos do Cocôzap? **(item 4)**

**MORADOR:**

Sim

**COCÔZAP:**

Agradecemos o seu contato!

O Cocôzap é um projeto que procura contribuir com a melhoria do saneamento básico no Complexo da Maré levantando dados de como se encontra a situação do esgoto, lixo e falta d'água, por exemplo. Com esses dados pretendemos chamar atenção da sociedade e do Estado buscando melhores políticas de saneamento básico para as favelas.

Se quiser saber mais sobre o trabalho do COCÔZAP e ficar por dentro das nossas ações, nos siga nas nossas redes:

**Instagram:** @cocozapmare

**Facebook:** Cocôzap Maré

**Twitter:** Cocozapmare

**Site:** <https://cocozap.datalabe.org/>



as agências públicas responsáveis pelo problema. Também buscamos estabelecer uma relação para além daquela primeira conversa, pois, em geral, uma pessoa que já enviou uma denúncia tem interesse pelo tema e pode ser uma grande aliada no processo de geração de dados e defesa de melhorias.



Como você pôde ver no exemplo acima, muita coisa podia acontecer na conversa. Por isso, fomos inserindo outras possibilidades de desdobramento, como a encaminhamentos para a Associação de Moradores ou para



## ORGANIZANDO AS QUEIXAS

Após a conversa, uma pessoa colaboradora do Cocôzap devia preencher de forma manual uma tabela com as informações sobre a denúncia coletada na conversa. Os dados pessoais, como telefone do morador e nome da pessoa, entravam em outra tabela dos contatos dos moradores, mas esse é um papo pra daqui a pouco.

A partir deste momento, a conversa livre era convertida em categorias pré-estabelecidas pela equipe do data\_labe. Já a foto enviada na conversa precisava ser salva em um drive online. Para proteger os dados pessoais dos moradores, a gente adotou um protocolo de criptografia. A foto era catalogada de acordo com um conjunto de números que representava a data e hora de envio da foto. Dessa forma, conseguiríamos identificar qual foto pertencia a qual denúncia.

Por fim, os dados pessoais dos moradores eram preenchidos em uma tabela que ficava

salva localmente no computador da pessoa responsável pelo preenchimento das denúncias, também de forma criptografada. Essa tabela continha o nome do morador, o telefone usado na conversa e uma chave identificadora gerada por hashing (vamos explicar o que é hashing mais à frente). Esta tabela, além de ser armazenada apenas de forma local, ficava guardada em um container criptografado via [VeraCrypt](#), uma plataforma de código aberto. Pelo VeraCrypt é possível enviar por email ou salvar no drive online, porém somente a pessoa operadora do Cocôzap e a responsável pelos dados do data\_labe tinham acesso à senha.

A [tabela](#) de denúncias era alimentada em um drive online no formato de um arquivo Google Sheets. Escolhemos manter esta tabela online devido à facilidade de compartilhamento entre as demais pessoas envolvidas no projeto e também para fins de automatização de relatórios e visualizações.



## CATEGORIAS E SUBCATEGORIAS

Durante o período do projeto, as denúncias foram mantidas em categorias gerais que se desdobram em subcategorias em três eixos de irregularidades de saneamento básico que se relacionam com “água”, “esgoto” ou “lixo”.

As categorias de denúncia foram construídas a partir do conhecimento técnico e empírico dos moradores da Maré que trabalhavam no data\_labe e eram especialistas no tema, que atinge de forma semelhante outras favelas e áreas periféricas dos grandes centros urbanos brasileiros. **As categorias macro eram:**



**Abastecimento de água**



**Alagamento**



**Drenagem**



**Esgoto**



**Lixo**



**Vazamento de água**

A definição sobre o que tratava cada denúncia era feita com base no relato e na foto enviados pelos moradores, mas a seleção final era feita por uma pessoa colaboradora do projeto, aqui denominado como “operador da denúncia”. Essa escolha foi realizada para deixar a conversa com o morador mais rápida e fluida, partindo do pressuposto que o momento de troca entre o morador e operador precisava ocupar o menor tempo possível, bem como poderia ser feita de forma retroativa. Isto é, o canal de WhatsApp não possuía cobertura 24 horas e sete dias da semana, o que acabava possibilitando que um morador realizasse o envio da denúncia e, apenas após um certo período, o operador entrasse em contato.

A subcategoria da denúncia era preenchida de forma dependente da categoria, ou seja, as subcategorias relacionadas a esgoto eram exclusivas e não poderiam ter uma segunda categoria relacionada a lixo, por exemplo. Eram elas:

Acúmulo de lixo

Falta d'água

Acúmulo de lixo e entulho

Problemas com bueiro

Alagamento

Qualidade da água

Bueiro entupido

Transbordamento de água

Esgoto a céu aberto

Transbordamento de esgoto

## VALIDAÇÃO E EXCLUSÃO DE DADOS

O processo de ter certeza que determinado dado é o mais próximo possível da realidade é chamado de validação e pode ser feito de diversas formas. Por exemplo, o SUS tem uma base de dados que tem somente os nascimentos do país, o Sistema de Informação sobre Nascidos Vivos - SINASC. Uma das formas de validar a informação é não ter nascidos vivos com menos de 500gr ao nascer, porque a Organização Mundial da Saúde (OMS) estabelece que este é um dos critérios necessários. Então, cada sistema vai

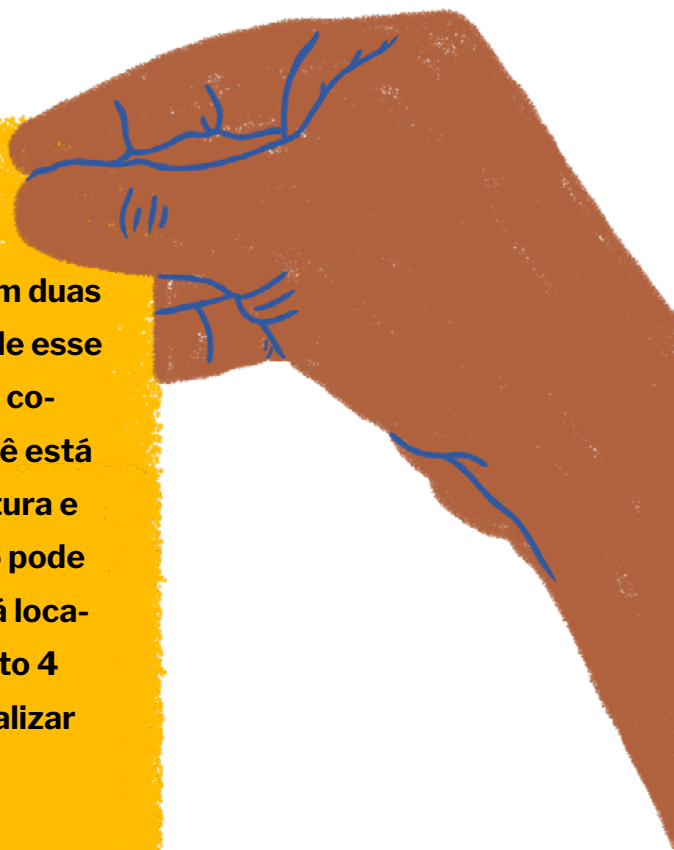
testando e elaborando a melhor forma para que seus dados sejam válidos e verdadeiros.

No caso do Cocôzap, nós precisávamos de dados que contivessem pelo menos a foto, a categoria e a localização da queixa, pois isso nos permitiria validar alguns critérios. Um dos mais fundamentais era que a localização informada pelo morador tivesse geolocalização (4) dentro da Maré em acordo com a localidade descrita pelo morador.

### O QUE É GEOLOCALIZAÇÃO?

Qualquer ponto na superfície do planeta Terra contém duas informações que permitem localizar exatamente onde esse ponto está. Isso é um rolê antigo, que vem desde que começamos a mapear o nosso espaço. Imagina que você está num grande retângulo, que por sua vez possui sua altura e comprimento, cada pontinho dentro desse retângulo pode fazer referência à altura e comprimento de onde está localizado. Você pode estar no ponto 3 de altura e no ponto 4 do comprimento e assim sempre conseguiremos localizar qualquer ponto desse retângulo.

No caso do planeta Terra, quando o achatamos em um mapa, ele se transforma em uma imagem plana que também vai possuir uma latitude e uma longitude, assim como o retângulo tem uma altura e comprimento. Então, a partir do endereço que o morador informava, a gente pesquisava no *Google Maps* quais eram a latitude e a longitude daquele ponto informado. Às vezes o próprio morador já mandava a localização compartilhada pelo *WhatsApp*, o que facilitava o nosso trabalho.



Outra forma de validação era a coerência entre a categoria informada e a foto enviada, que na maioria das vezes estava correta. Se a pessoa denunciava “acúmulo de lixo”, a foto era claramente de um acúmulo de lixo. Em outros casos, surgiam dúvidas. Por exemplo: um esgoto a céu aberto poderia se tratar de um bueiro entupido?

A gente entendeu que essa dúvida poderia ser conversada com o morador, mas caso não fosse possível, a foto valia mais que o relato. Quando a gente ainda ficava com dúvidas, duas pessoas da equipe avaliavam essa queixa até entrarem em consenso e conseguirem reclassificar de forma unânime.

Nossa perda de dados ou incoerência sempre foi muito baixa. Acreditamos que isso se deva ao fato do morador estar convivendo com a queixa que enviava, então tínhamos uma memória e engajamento muito peculiares, o que fazia com que a informação enviada possuísse alto grau de **confiabilidade**.

Era esperado também que as pessoas que enviam a foto e constroem uma conversa sobre aquela queixa estavam de fato interessadas em serem ouvidas. Portanto, a predisposição das pessoas que enviavam suas queixas também foi um fator determinante para a **qualidade** dos dados das queixas.





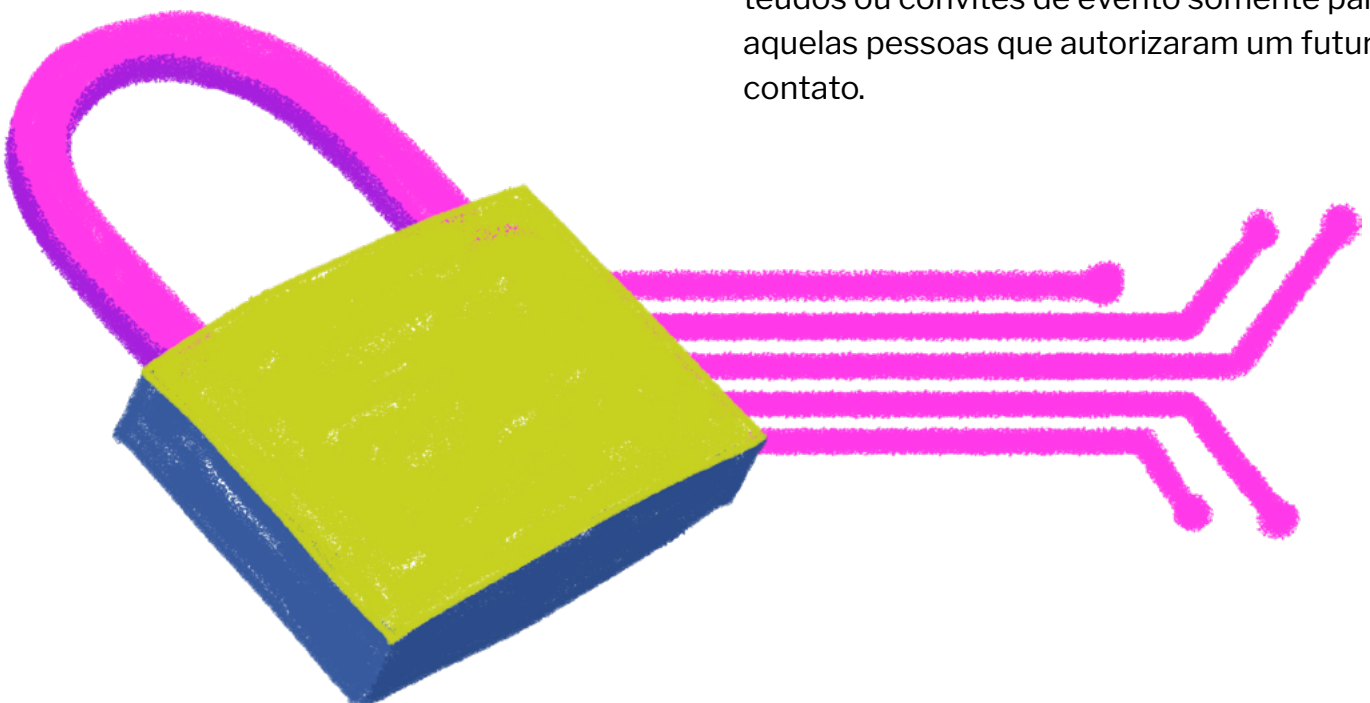
## PROTEÇÃO E PRIVACIDADE DOS DADOS PESSOAIS

Pelo fato da coleta ser feita por meio do *WhatsApp*, de cara a gente já tinha acesso ao número da pessoa. Esta era uma grande preocupação pra gente, porque não tem coisa pior do que quando seu número acaba indo pra algum telemarketing que fica te oferecendo coisas que você nem tem interesse. Então, a gente sabia que precisava guardar essa informação com muito cuidado. Além disso, a gente sempre perguntava o nome da pessoa pra poder saber como se dirigir a quem estava falando com a gente. Essas pessoas eram muito aliadas ao nosso trabalho e na sua maioria, eram também quem frequentava nossos encontros, eventos e precisávamos manter essa lista de contato muito bem guardadas.

Todo cuidado é pouco quando estamos lidando com dados de pessoas de favelas e comunidades periféricas. Por isso, desenvolvemos uma rotina offline para guardar e proteger esses dados. Primeiro, a pessoa do Cocôzap deveria colocar o nome informado e o telefone do morador numa planilha, que ficaria disponível no arquivo local do computador.

A partir daí, nossa equipe de dados usaria um script - que ela mesma elaborou - para pegar essas duas informações e gerar uma chave que mistura letras, números e símbolos. É como se a gente gerasse um número único e exclusivo para cada pessoa com seu telefone. Esta chave é conhecida como **CHAVE HASHING**, porque utiliza algoritmos para pegar as informações e gerar uma nova chave, tornando impossível voltar ao valor ou à palavra original. A chave seria inserida ao lado do nome e telefone desse morador e, então, a planilha estaria criptografada. Este arquivo criptografado não era compartilhado com ninguém. Caso fosse necessário identificar de quem era uma determinada denúncia, poderíamos acionar a chave hashing. Saiba como criar uma chave *hashing* [aqui](#).

Este caminho nos proporcionava uma ótima qualidade da informação e a identificação da autoria da denúncia. Para garantir a proteção e a privacidade dos dados dos moradores, nós consideramos mais dois critérios ao protocolo: jamais compartilhar esses dados com terceiros e nem mesmo com outras áreas do *data\_labe*; e enviar outros conteúdos ou convites de evento somente para aquelas pessoas que autorizaram um futuro contato.





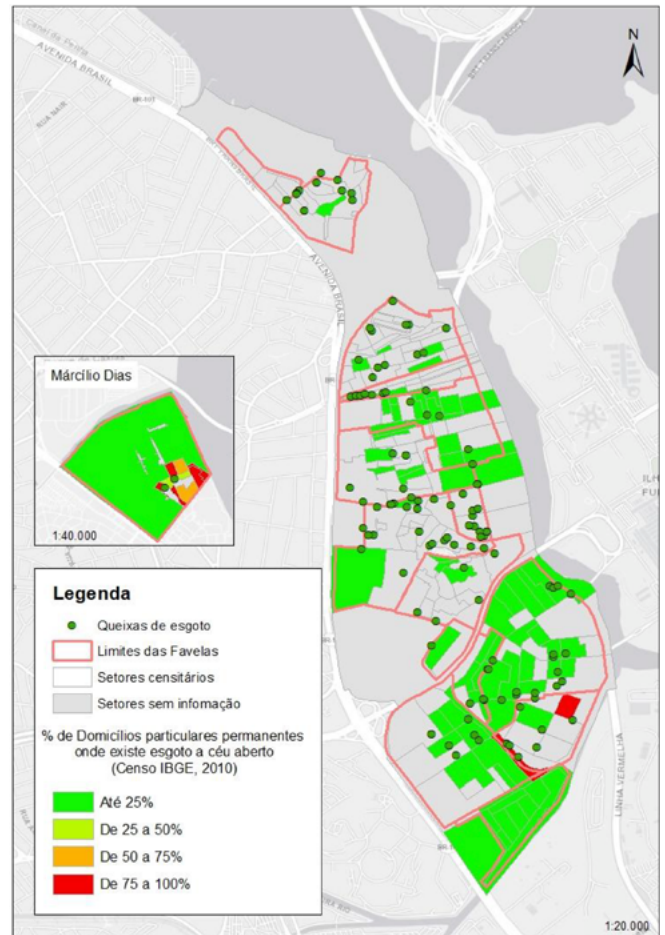
## ANÁLISES E PAINÉIS AUTOMATIZADOS

A tabela das queixas foi crescendo com o passar dos anos, às vezes mais, às vezes menos, porém constante, como diz o sambista “quem anda devagar anda sempre na frente”. Por já usar uma tabela estruturada, a gente conseguiu desenvolver scripts automatizados que faziam o download desse arquivo, aplicavam comandos para gerar tabelas e gráficos e produziam relatórios. Este processo nos deu um maior controle de como os dados e queixas estavam chegando e se desenvolvendo.

Mais recentemente nós conseguimos elaborar um painel que, além de fazer esses relatórios, produz um documento que pode ser hospedado na internet e de forma interativa, pois permite que o usuário acesse e extraia informações por ano, por categoria ou por favela.

Para chegar a este ponto de construir tecnologias, foi preciso tempo e expandir a equipe, até que, em 2021, conseguimos adicionar ao nosso mapeamento de queixas, os dados secundários produzidos pelo IBGE e pelo Instituto Pereira Passos. Esta adição de “camadas” possibilitou uma comparação mais apurada entre os dados públicos disponibilizados e os nossos, gerados de forma cidadã. A comparação torna mais evidente como processos de geração cidadã de dados possibilitam diagnósticos mais próximos à realidade cotidiana de territórios de favelas e comunidades, como a Maré. Você pode ter um exemplo disso na próxima foto aonde os pontos verdes são denúncias de esgoto a céu aberto que recebemos pelo Cocôzap e a informação das quadras (setores censitários) por trás dos pontos em ver-

de é a informação oficial do IBGE baseada no Censo 2010 também para esgoto a céu aberto.



Vejam que existem muitos setores censitários sem informação alguma e que temos denúncias recebidas pelo Cocôzap. Em diálogo com o IBGE, eles informaram que esses setores podem se tratar do que era chamado “aglomerado subnormal” e por isso essa informação não é prevista de ser coletada, ou ainda que na época do Censo esse setor não teve a informação coletada. Independente do que se tratava, é possível vermos a disparidade entre as informações, principalmente na região central do território, onde temos mais denúncias recebidas e nenhuma informação disponível pelo Censo 2010.

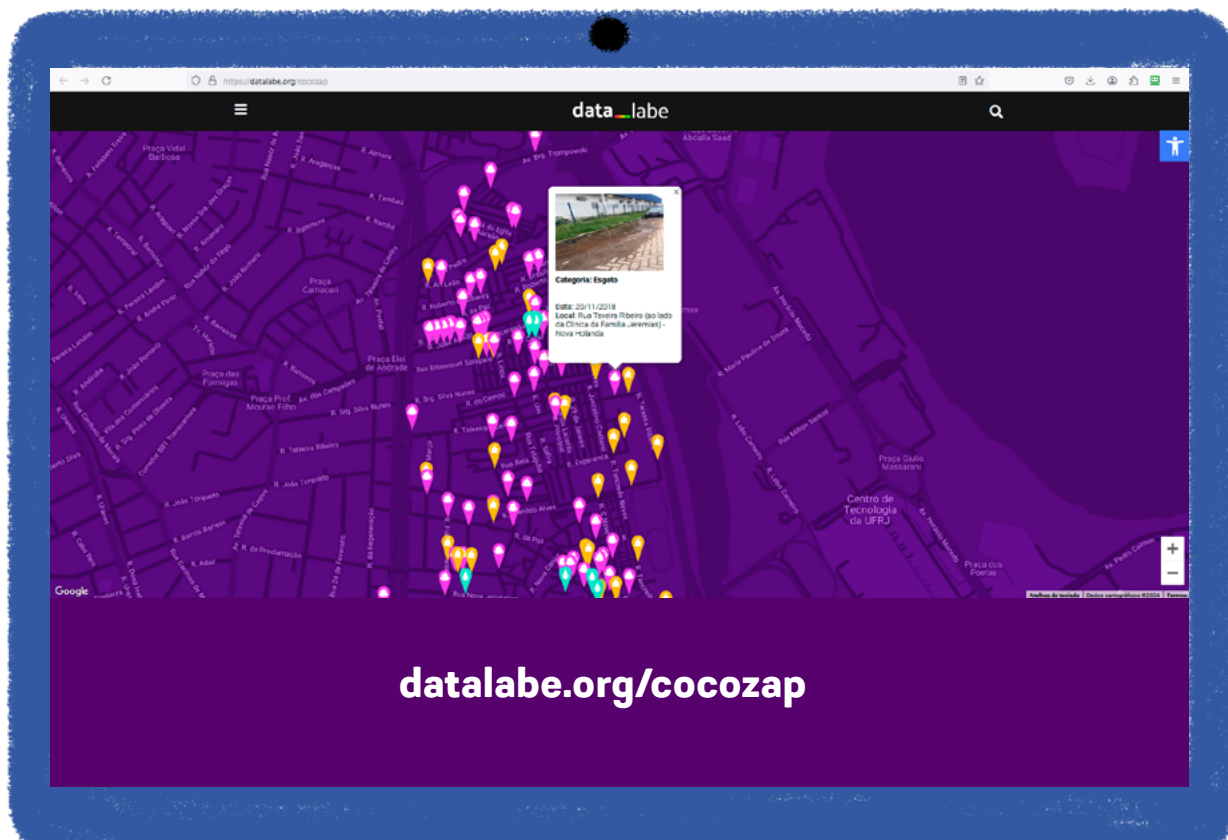
## OS DADOS NA INTERNET

Um dos nossos compromissos desde o início do projeto foi com a **transparência e abertura dos dados**. Para nós, não tem sentido um processo gerado pelos cidadãos não estar disponível para eles. Para isso, a gente desenvolveu um site que, além de explicar os nossos processos, contém todas as nossas publicações em diversos formatos e possibilita a visualização de todos os nossos resultados. Nele é possível visualizar o mapa com as denúncias recebidas e as respectivas fotos.

Nosso site foi desenvolvido em Wordpress, com plugins que facilitam o envio dessas informações. Então, por questões de segurança e acesso, uma pessoa da equipe ficava responsável por olhar a planilha de queixas, enviar os dados e fotos para o site e, assim, manter o mapa atualizado.

Para desenvolver o mapa utilizamos o *plugin WP Go Maps*. Ele é responsável pelas funcionalidades de mapeamento do site, facilita a exibição de locais no mapa e a inclusão de informações detalhadas sobre pontos de interesse. Também criamos marcadores personalizados para cada tipo de queixa, a fim de dar uma visão geral e objetiva dos diferentes tipos de problemas relatados pelos moradores e ajudar na identificação de áreas problemáticas no mapa.

Também implementamos janelas com informações incorporadas aos marcadores, o que possibilita a exibição de uma janela com informações detalhadas relacionadas a cada queixa toda vez que um usuário clica em um marcador.



O plugin também inclui o recurso de “*Mapas de Cluster*”. Esta funcionalidade é utilizada quando há muitos marcadores agrupados em uma única região. Ela organiza automaticamente esses marcadores em grupos, tornando a visualização mais clara e facilitando a navegação.

Por fim, implementamos o sistema de geolocalização, que permite que os usuários identifiquem sua própria localização no mapa. Isso ajuda a entender como os problemas se apresentam onde a pessoa está.

Parece muita coisa e realmente é. Pessoas para conversar com os moradores pelo *WhatsApp*, para preencher a planilha, realizar as análises e, agora, mais uma para alimentar o site. Mas nem sempre a realidade é

a ideal. Por muito tempo, havia apenas uma pessoa para dar conta de todo o processo. Conforme as queixas foram aumentando, fomos avançando, buscando mais recursos e redistribuindo as tarefas.

**Assim como a transparência e o envolvimento com as pessoas são princípios importantes da geração cidadã de dados, a inclusão digital também é fundamental.**

Por isso, quanto mais pessoas se envolvem e constroem todo esse processo necessário para gerar o dado, mais conhecimentos serão agregados e também disseminados. Cada pessoa que participa desse processo precisa se apropriar dele e entender o que são cada linha daquela planilha, cada conversa no *WhatsApp* e cada foto no mapa.



## PRÓXIMOS PASSOS DO COCÔZAP

Em 2024 nós paramos de coletar os dados de saneamento na Maré. Entendemos que, ao longo dos anos, conseguimos chegar a uma base de dados válida, consistente e robusta. Esse resultado foi possível graças à participação de muitas pessoas engajadas e que protagonizaram cada etapa de acordo com sua experiência e conhecimento.

A partir dos dados, pudemos comprovar que a precariedade do saneamento básico na Maré obedece a uma temporalidade de acordo com as estações do ano, ora piora a falta d'água, ora aumenta a quantidade de esgoto a céu aberto. O acúmulo de lixo é uma realidade equivocadamente naturalizada no Complexo da Maré, não pelos moradores, pois nenhuma pessoa está de acordo em ter montes de lixo perto de sua casa, mas muito mais por um descaso e falta de adaptação da coleta de resíduos sólidos a território de favelas. Nosso trabalho de ativismo e geração cidadã de dados atualmente possibilitou que as bases de dados do Cocôzap fossem consideradas para as decisões do plano diretor da cidade do Rio de Janeiro, para que sejam consideradas em obras de melhoria e infraestrutura da cidade. Além disso, passamos a integrar o Conselho Consultivo do IBGE para favelas e comunidades urbanas e por fim, em 2022 fomos finalistas do prêmio de tecnologia social replicável do Banco do Brasil, possibilitando que todas as pessoas pudessem consultar e replicar os nossos métodos.

Nesta jornada encontramos parceiros que se inspiraram no Cocôzap para desenvolver seus próprios projetos de geração cidadã de dados. Incentivamos que mais e mais adaptações sejam feitas, por isso sempre que

quiser falar com gente, trocar uma ideia, ensinar, aprender, tomar um café, ou uma cerveja, basta falar com a gente pelas nossas redes sociais ou pelo nosso email institucional [contato@datalabe.org](mailto:contato@datalabe.org).

Como falamos no começo desse documento, aqui você não encontrará os resultados do projeto. Eles estão disponíveis no site, no qual você encontra documentos, relatórios, visualizações de dados interativas e conteúdos.

**A suspensão da coleta não significa que o Cocôzap acabou. A gente segue produzindo dados de forma livre, popular e favelada para democratizar o conhecimento e melhorar as realidades da favela. Acreditamos que projetos como esse nunca têm fim, pois se desdobram, se reinventam e culminam em novos desafios. Um deles tem sido a articulação da Rede GCD, uma rede criada por organizações da sociedade civil no ano passado para conectar outras iniciativas que acreditam na geração cidadã de dados como uma tecnologia social para garantia da nossa democracia. Então, fica o convite para que você conheça a Rede GCD e, quem sabe, seja mais uma pessoa no bonde que trabalha com dados sobre quem importa: você!**

**Autores:** Andreia Luiza  
Kezia Antero  
Polinho Mota

**Rvisão:** Elena Wesley

**Arte:** Nicolás Noel