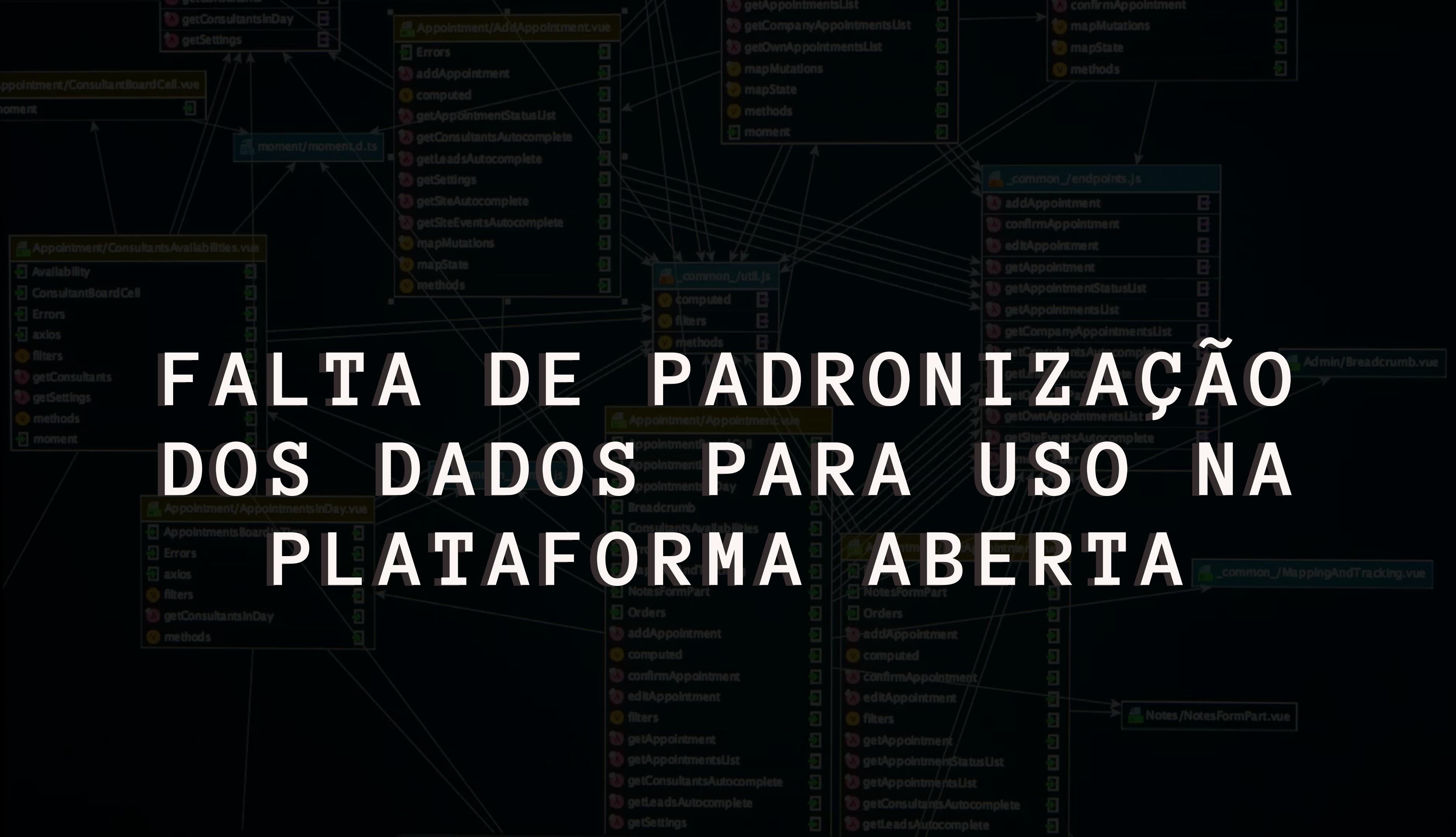


PROJETO DE CONCLUSÃO: MANUAL DE BOAS PRÁTICAS NA CATALOGAÇÃO DE DADOS

Gabriel dos Santos Nascimento
Yago Rusef da Silva

FALTA DE PADRONIZAÇÃO DOS DADOS PARA USO NA PLATAFORMA ABERTA



MANUAL DE BOAS PRÁTICAS NA CATALOGAÇÃO DE DADOS

Deise Lins
Gabriel Nascimento
Yago Rusef

DEZEMBRO de
2024

Nossa proposta de solução consiste em um **manual de boas práticas para a catalogação de dados públicos**. O documento é estruturado em **dez capítulos**, cada um dedicado a uma boa prática específica, acompanhada de suas diretrizes. A organização do conteúdo segue um formato progressivo, dividido em etapas: prévia, inicial, intermediária e avançada. Esse modelo permite uma progressão linear, em que, ao cumprir as recomendações de uma boa prática, o leitor está apto a avançar para a próxima etapa.

ETAPA PREVIA

1.Compreensão dos Problemas e Desafios

A construção de um catálogo de dados públicos requer a compreensão de como os dados são gerados, armazenados e gerenciados. Esse processo pode ser conduzido por meio da designação de representantes de todas as secretarias, que serão responsáveis por realizar um levantamento interno dessas informações em colaboração com os demais integrantes.

DEFINIÇÕES

- **Data Warehouse:**
 - Repositório centralizado, organizado, estruturado e robusto;
- **Data Lake:**
 - Repositório flexível, não estruturado e unificado;
- **Data Mesh:**
 - Repositório descentralizado, organizado, estruturado e robusto;

DEFINIÇÕES

- API:
 - Serviço que permite conectar dois ou mais softwares a partir de instruções.
- ETL:
 - Processo onde os dados são extraídos de uma fonte, tratados e carregados.
- ELT:
 - Processo onde os dados são extraídos, carregados, e então tratados.

ETAPA INICIAL

2.Identificação e Levantamento de Fontes de Dados

MAPEAR

mapear todas as fontes de dados existentes, que ainda são predominantemente físicas

DIGITALIZAR

A digitalização das fontes é crucial para garantir a coleta de dados de forma eficaz

GOVERNANÇA E PADRONIZAR

A gestão dos dados deve ser descentralizada, mas com padrões claros de organização

ETAPA INICIAL

3. Padronização dos Dados

Com os dados digitais mapeados, o próximo desafio é a padronização e normalização dessas informações.

Ferramentas



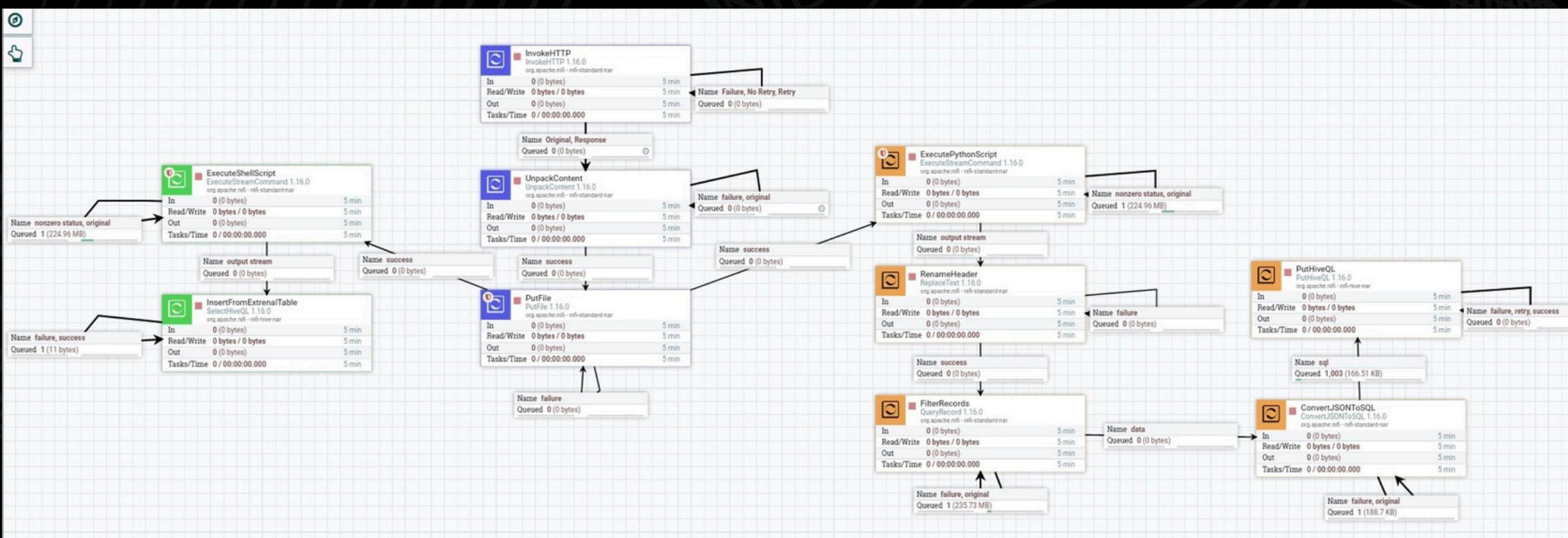
exemplos

Para tanto, no guia sugerimos ferramentas que podem auxiliar nesse processo de padronização e exemplos de aplicação com metadados.

FERRAMENTA RECOMENDADA

Apache NiFi

Uma ferramenta de ETL/ELT aberta, com interfaces gráficas, que melhoram a acessibilidade, sobretudo para pessoas sem conhecimento em programação



ETAPA INICIAL

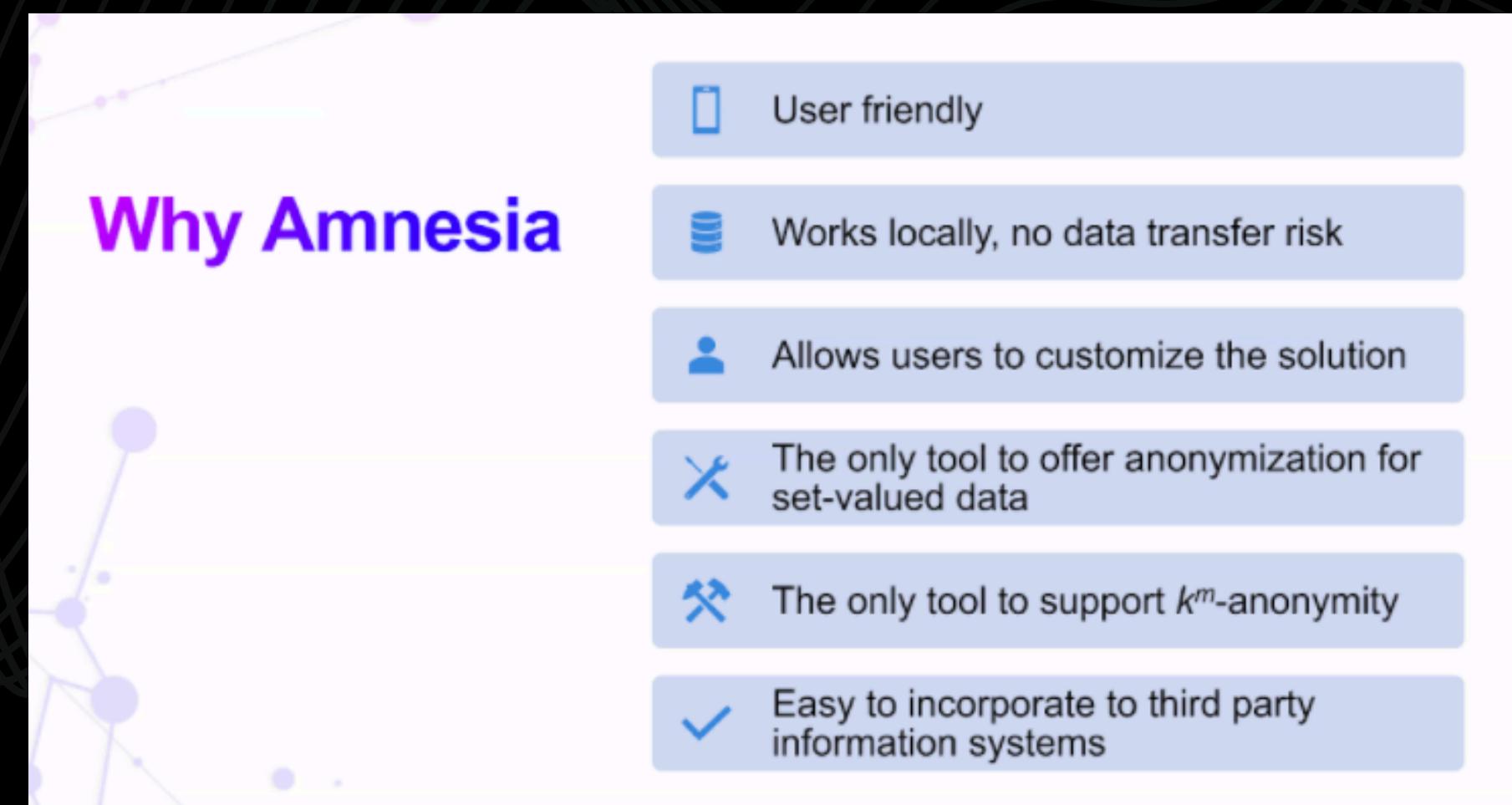
4. Privacidade e Proteção de Dados Pessoais

A Lei Geral de Proteção de Dados (LGPD) exige que a prefeitura tome medidas rigorosas para garantir que dados pessoais sejam protegidos e não utilizados de forma indevida.

FERRAMENTA RECOMENDADA

Amnesia:

Ferramenta aberta usada para anonimização de dados. Com ela podemos transformar os dados sensíveis, que, pela LGPD, não poderiam ser divulgados, em chaves genéricas.



ETAPA INICIAL

5. Estruturação do Catálogo

Nesse capítulo, devemos começar a estruturar uma governança de dados, dividindo as áreas onde cada servidor deve atuar, como:

- Segurança;
- Educação;
- Saúde;
- Infraestrutura;
- Etc...

ETAPA INICIAL

6. Acessibilidade e Disponibilidade

Nesse momento, devemos dar inicio à disponibilização dos dados, tornando-os públicos em uma plataforma acessível.

Para manter a atualização constante e automatizada, podemos utilizar uma API, que conecte os dados da prefeitura com a plataforma.

ETAPA INICIAL

7. Transparência e Abertura:

Nessa etapa, devemos garantir que os dados da etapa anterior estão com uma boa documentação. Nela deveremos informar alguns metadados como:

- Frequência de atualização;
- O que cada atributo representa;
- Qual a área responsável pelos dados.

Além disso, deve-se garantir que as informações divulgadas estejam respeitando os pontos mencionados na LGPD.

ETAPA INTERMEDIARIA

8. Atualização Contínua

Este capítulo aborda a importância da perenidade dos dados, destacando a necessidade de estabelecer regras claras para a divulgação de dados públicos. Discute-se a periodicidade ideal para a atualização das informações e apresentam-se ferramentas que podem apoiar a implementação eficaz dessas diretrizes.

ETAPA INTERMEDIARIA

9. Reutilização de Dados

Este capítulo destaca a importância não apenas de disponibilizar informações, mas também de considerar as melhores formas para que os cidadãos possam acessá-las e utilizá-las de maneira eficiente.

Como exemplo de uma prática bem-sucedida, apresentamos o caso da API de dados do Instituto Brasileiro de Geografia e Estatística (IBGE), que se destaca por sua abordagem inovadora e acessível.



ETAPA INTERMEDIARIA

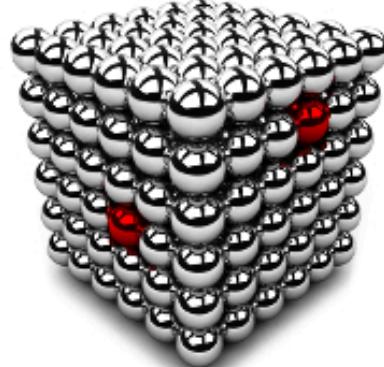
9. Reutilização de Dados

IBGE

API de serviço de dados







Agregados
Análise multidimensional ao seu alcance.

3.0



Banco de Dados Geodésicos
API para acesso às informações das estações geodésicas

1.0



BNGB
API para acesso aos dados do Banco de Nomes Geográficos do Brasil

1.0

ETAPA INTERMEDIARIA

9. DEMONSTRAÇÃO

```
"codigo":"AA",
"nome":"Pesquisa Nacional de Saúde do Escolar",
"nome_ingles":null,
"situacao":"Ativa",
"categoria":"Especial",
"periodicidade_coleta":"Eventual",
"periodicidade_divulgacao":"Eventual",
"classificacoes_tematicas": [...]
},
{
"codigo":"AM",
"nome":"Pesquisa de Assistência Médico-Sanitária",
"nome_ingles":null,
"situacao":"Ativa",
"categoria":"Especial",
"periodicidade_coleta":"Eventual",
"periodicidade_divulgacao":"Eventual",
"classificacoes_tematicas": [...]
},
```

A	B
1 Regiao	Quantidade de Estados
2 Norte	7
3 Sul	3
4 Nordeste	9
5 Sudeste	4
6 Centro-Oeste	4
7	
8	
9	

ETAPA AVANÇADA

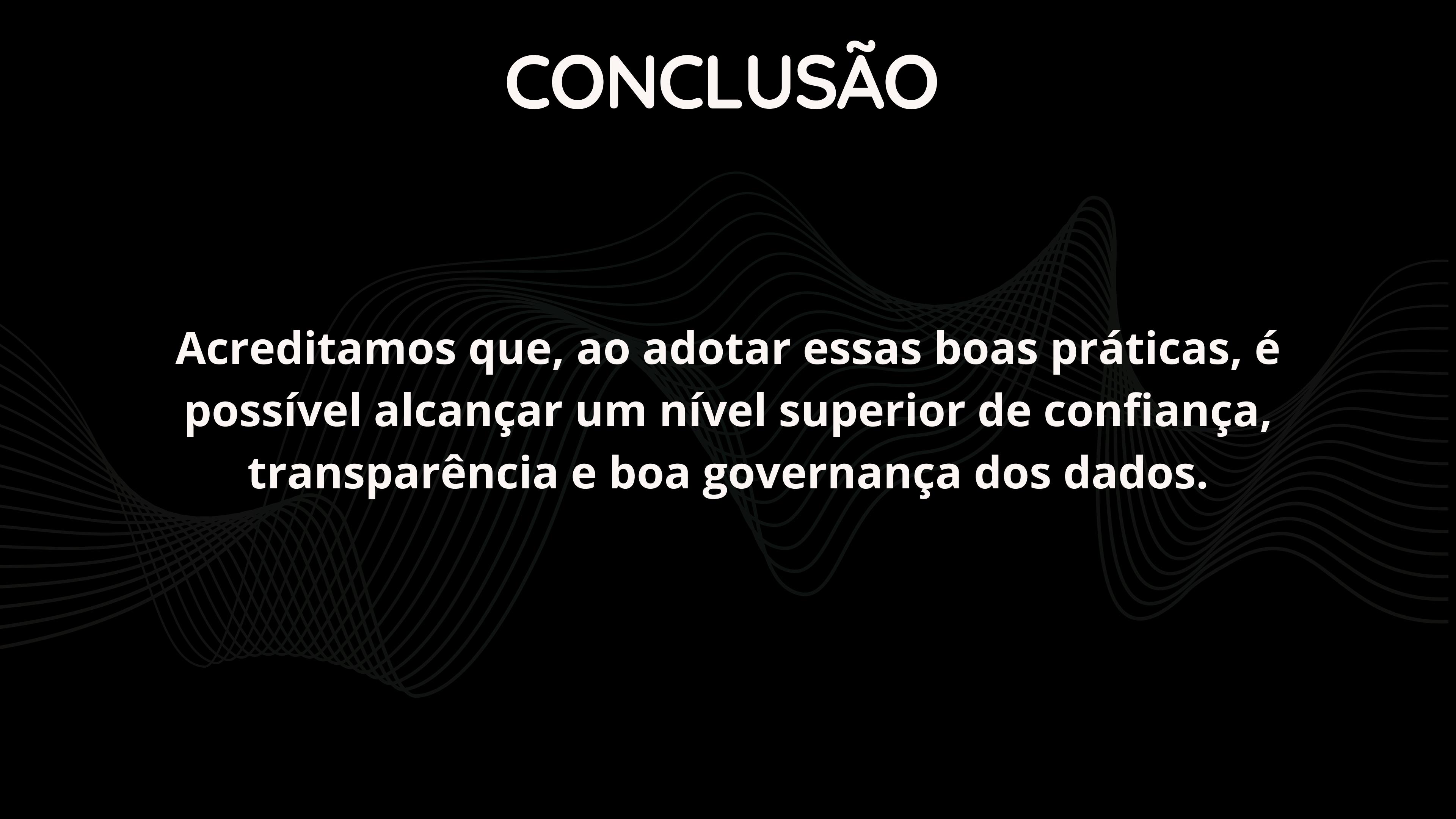
10. Ferramentas de Visualização e Interatividade

Este capítulo apresenta diretrizes para orientar as interações dos cidadãos com os dados disponibilizados, destacando a importância de ferramentas que promovam visualização clara e interatividade.

ETAPA AVANÇADA



CONCLUSÃO



Acreditamos que, ao adotar essas boas práticas, é possível alcançar um nível superior de confiança, transparência e boa governança dos dados.