

Cartilha Técnica para Publicação de Dados Abertos da Cidade de Fortaleza

Versão 1.0

Elaborado por:

Vlândia Pinheiro

Revisado por:

xxxxx

Aprovado por:

<Patrocinador do
projeto>

Data modificação:

06/05/yyyy



HISTÓRICO DE REVISÕES

Data	Versão	Descrição	Autor
15/04/2014	1.0	Versão Inicial	Vlândia Pinheiro



1. Apresentação

A Prefeitura Municipal de Fortaleza incluiu em seu Plano Plurianual para 2014 o projeto FORTALEZA DADOS ABERTOS como resposta a um clamor da sociedade por matéria-prima tão importante para inovação, pesquisa científica e desenvolvimento de aplicações para toda a sociedade.

Esta iniciativa está em consonância com o Governo Federal que, desde 2009, vem realizando ações para o desenvolvimento de uma política de disseminação de dados e informações governamentais para o livre uso pela sociedade. Por sua vez, estas ações estão alinhadas com um crescente movimento global para democratização do acesso à dados e informações no paradigma de dados abertos.

O movimento por dados abertos é mundial e defende a criação de um ecossistema fundamentado no acesso à dados que possibilite a participação da sociedade no desenvolvimento de um estado mais eficiente, com oferta de melhores serviços, e fazendo amplo uso de tecnologias.

A Fundação de Ciência, Tecnologia e Inovação de Fortaleza - CITINOVA, apresenta esta Cartilha Técnica para Publicação de Dados Abertos da Cidade de Fortaleza, com o objetivo de orientar os órgãos, entidades e empresas do governo municipal de Fortaleza quanto às boas práticas de publicação de dados na Internet e o devido cumprimento dos critérios técnicos relacionados ao paradigma de dados abertos.

Este documento é inspirado na Cartilha Técnica da Infraestrutura Nacional de Dados Abertos (INDA) do Governo Federal Brasileiro. Depositamos aqui os créditos e méritos desta exemplar contribuição à comunidade de dados abertos do Brasil.

2. Introdução

Dados Abertos é a publicação e disseminação de dados e informações públicas na Web, seguindo alguns critérios que possibilitam sua reutilização e o desenvolvimento de aplicativos por toda a sociedade. A maior parte dos dados e informações geradas ou mantidas pelo governo, são públicas. Disponibilizar dados na Web não é uma prática recente no governo, porém com uma política de dados abertos, o governo sinaliza que pretende padronizar e alavancar a disseminação de dados públicos por todos os órgãos. O paradigma de dados abertos está fundamentado na constatação de que o dado quando compartilhado abertamente tem seu valor e seu uso potencializados. Com isso o governo pretende desenvolver um ecossistema de dados e informações que beneficia a sociedade e possibilita o envolvimento de todos seus setores, inclusive a iniciativa privada, o setor acadêmico e o próprio governo.

O movimento de dados abertos no mundo foi alavancado há pouco mais de 4 anos quando os governos da Inglaterra e dos Estados Unidos iniciaram a construção de seus portais de dados abertos. Desde então políticas de dados abertos vêm sendo implementadas em diversos países. Em setembro de 2011 oito nações se uniram para firmar uma parceria com o objetivo de promover governos mais transparentes e eficientes, através de uma maior participação e colaboração social. A Parceria para Governo Aberto (OGP do inglês) é uma parceria co-liderada por Brasil e EUA. Em novembro de 2011 foi sancionada a Lei de Acesso à Informação (12.527/2011) que tem o propósito de regulamentar o direito constitucional de acesso dos cidadãos às informações públicas.

3. IMDAFOR

A Fundação de Ciência, Tecnologia e Inovação de Fortaleza (CITINOVA) vem desenvolvendo a Infraestrutura Municipal de Dados Abertos de Fortaleza - IMDAFOR. A IMDAFOR é um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos. O principal projeto da IMDAFOR é o Portal Fortaleza Dados Abertos, que tem o objetivo de ser o ponto central para a publicação, a busca e o acesso de dados públicos no Brasil.

O Decreto nº XX, de XX de XXXXXX de 9999 que institui a IMDAFOR define em seu art. 3º os órgãos e entidades que a integram obrigatória ou facultativamente. Cada órgão ou entidade integrante da IMDAFOR deverá indicar um servidor responsável pela catalogação dos dados de sua instituição. O órgão ou entidade que não estão previamente obrigados e que desejam aderir à IMDAFOR devem preencher Termo de Adesão, disponível no endereço xxxxxxxxxxxx e encaminhar a CITINOVA por meio de ofício.

4. Objetivo

O objetivo deste documento é o estabelecimento de diretrizes para que os dados públicos da cidade de Fortaleza constituam o Catálogo de Dados Abertos de Fortaleza, contribuindo para a democratização do acesso à informação na sociedade. As diretrizes aqui dispostas compreendem requisitos mínimos para que uma organização seja capaz de preparar e validar os conjuntos de dados, e submeter os mesmos para catalogação no Portal Fortaleza Dados Abertos.

5. Portal Fortaleza Dados Abertos

O governo municipal de Fortaleza desenvolveu o Portal Fortaleza Dados Abertos (<http://dados.fortaleza.ce.gov.br/>) com o principal objetivo de ser o ponto central para a busca e o acesso a dados públicos de governo e da cidade de Fortaleza, os quais compõem o Catálogo de Dados Abertos de Fortaleza.

O Catálogo de Dados Abertos de Fortaleza consiste em um repositório com os conjuntos de dados abertos publicados pelo governo e entidades com adesão à IMDAFOR (conforme definido no inciso II do Art. 3º do Decreto nº XX, de XX de XXXXXX de 9999) e em serviços para que o usuário tenha acesso aos conjuntos de dados catalogados. O catálogo tem o objetivo de simplificar a busca e o acesso aos conjuntos de dados através de ferramentas. O catálogo também pode ser visto como a organização dos metadados dos conjuntos de dados do repositório.

Alguns requisitos para o catálogo:

1. O catálogo deve conter uma lista exaustiva dos conjuntos de dados presentes no repositório (ou seja, todos os dados publicados no repositório devem estar acessíveis através do catálogo).
2. É desejável que o catálogo possibilite a navegação e a busca simplificada dos dados. A busca pode ser através do uso de categorias, busca por texto, busca por formato, etc.
3. O catálogo deve utilizar, no mínimo, o conjunto de metadados obrigatórios disponíveis nesta norma, e poderá extê-lo a seu critério.

O Portal Fortaleza Dados Abertos tem sua estrutura baseada na arquitetura da Web e a identificação e publicação de itens do Catálogo de Dados Abertos de Fortaleza (conjuntos de dados abertos ou *datasets*) seguem as diretrizes abaixo:

1. Cada conjunto de dados possui um identificador único e persistente, seguindo uma

padronização de URL. Esse requisito é imprescindível para que esse conjunto de dados seja referenciável e eventualmente consumido automaticamente por um aplicativo;

2. São utilizadas considerações semânticas na definição URLs, de forma que seja possível deduzir o conteúdo de um conjunto de dados apenas lendo seu identificador;
3. É desejável que o repositório dos dados possibilite a composição de filtros dentro da URL, seguindo algum padrão de API, permitindo que o usuário restrinja o volume dos dados para aqueles que ele deseja;
4. Não são usados mecanismos antirrobôs, como captchas, para acesso ao Catálogo de Dados Abertos de Fortaleza;
5. Considerando que é desejável facilitar a indexação dos dados por motores de busca, sendo esta uma importante forma do cidadão encontrar os dados que procura, é recomendável que os nomes dos arquivos sigam as boas práticas de formação de um slug [WIKIPEDIA http://en.wikipedia.org/wiki/Slug_%28web_publishing%29], tal como é realizado em muitas soluções de software para blog ou gerenciadores de conteúdo. Mais especificamente, o título deve ser convertido para slug da seguinte maneira:
 - a) substituem-se todos os caracteres acentuados pelos seus correspondentes não acentuados;
 - b) transformam-se todos os caracteres maiúsculos em minúsculos;
 - c) substituem-se cada sequência de um ou mais espaços por um único hífen ("-"). Usa-se hífen em lugar de sublinhados ("_"), pois estes fazem com que os mecanismos de busca tratem o texto como um só termo. O mesmo aconteceria caso fossem utilizadas palavras concatenadas, no modo chamado "camelCase". Por outro lado, o hífen permite que as palavras sejam indexadas individualmente [GOOGLE <https://www.youtube.com/watch?v=AQcSFsQyct8>].

O Portal Fortaleza Dados Abertos fornece serviços para publicação, atualização e consulta ao Catálogo de Dados Abertos de Fortaleza. A arquitetura do portal é centralizada e tanto a infraestrutura física como a responsabilidade pela manutenção do portal são de responsabilidade da CITINOVA.

Para isto, devem ser garantidos os seguintes requisitos técnicos:

1. **Controle de acesso:** Somente as pessoas autorizadas devem ter acesso para publicação e modificação dos dados e metadados de um órgão ou entidade. Estas são denominadas de Administradores de Dados Abertos. É mandatória a utilização de mecanismos de autenticação (usuário e senha).
2. **Integridade:** Deve assegurar que os dados não sejam adulterados durante a transferência. Esse requisito não é mandatório, porém para dados mais críticos é recomendável o uso de SSL.
3. **Autenticidade:** Deve assegurar que os dados provêm de uma fonte legítima da instituição. Para isso, todo conjunto de dados e seus respectivos metadados deve ser incluído ou alterado somente pelo administrador de dados abertos, autorizado de cada órgão ou entidade.

6. Publicação e Manutenção de Dados Abertos

• Preparação e Validação

Cada órgão ou entidade é responsável pela preparação, validação e publicação dos conjuntos de dados públicos que estão em sua posse. Apesar do termo Dados Abertos estar em evidência, a publicação de dados públicos por instituições governamentais não é uma prática nova. Diversos órgãos têm a cultura de publicar dados na Internet, dentre eles podemos citar: IBGE, DataSus,

IPEA, INEP, entre outros. Apesar de muitos órgãos e entidades não possuírem dentre suas atribuições finalísticas a divulgação de dados ao público, é improvável que este não possua dados públicos. Conforme dispõe a Lei 12.527 / 2011, em seu Art. 8º, §1º, no mínimo, precisam ser publicadas as seguintes informações que são comuns a todos os órgãos e entidades:

- dados referentes a registros das competências e estrutura organizacional, endereços, telefones e horários de funcionamento;
- registro de quaisquer repasses, transferências de recursos financeiros e despesas; informações sobre licitações, inclusive editais publicados e contratos firmados; e
- dados sobre acompanhamento de programas, ações, projetos e obras.

De maneira simplificada, sem entrar no mérito operacional de cada órgão, para que um conjunto de dados esteja apto a constituir o Catálogo de Dados Abertos de Fortaleza, o Administrador de Dados Abertos daquele órgão deve garantir que este conjunto de dados cumpra as seguintes condições gerais:

1. Os dados devem estar em seu formato mais bruto possível, ou seja, antes de qualquer cruzamento ou agregação. Mesmo que o órgão ou entidade ache importante e já tenha publicado alguma visão de agregação desses dados, existe grande valor no dado desagregado. Dessa forma o órgão ou entidade pode publicar esses dados nas duas formas.
2. Os dados devem estar em formato aberto, não proprietário, estável e de amplo uso (vide subseção Formatos).
3. Não deve existir nenhum instrumento jurídico que impeça sua reutilização e redistribuição por qualquer parte da sociedade.
4. Para os dados que são estruturados ou estão em planilhas na sua fonte, deve-se preservar ao máximo a estrutura original. Por exemplo, não deve-se publicar planilhas em arquivo PDF, neste caso utilize CSV ou ODS.
5. É recomendável a disponibilização dos dados em diversos formatos.
6. Cada conjunto de dados deve ter informações sobre seus dados e metadados. Deve ser possível recuperar o significado dos dados.
7. Para conjunto de dados muito grandes, recomenda-se a divisão em conjuntos menores, permitindo uma fácil manipulação. Recomenda-se fazer a divisão pela dimensão temporal (ano ou mês), pela dimensão geográfica (estado ou município), ou por outra dimensão.

• Formatos

A filosofia de dados abertos não define exaustivamente uma lista de formatos permitidos. Para ser considerado um dado aberto, o conjunto de dados deve estar disponível em um formato de especificação aberta, não proprietário, e estruturado, ou seja, que possibilite seu uso irrestrito e automatizado através da Web. Além disso, é imprescindível que seja utilizado um formato amplamente conhecido.

É importante enfatizar a importância de dados estruturados. Um erro recorrente cometido por diversas instituições é a publicação em formato PDF de planilhas de dados. O PDF é um formato não estruturado, e ao fazer isso - desestruturação dos dados - o publicador está inviabilizando, ou dificultando, a reutilização daqueles dados.

A seguir uma lista de diversos formatos não proprietários para dados abertos:

- **JSON** é um acrônimo para JavaScript Object Notation. É um padrão aberto de estruturação de dados baseado em texto e legível por humano. A especificação é a RFC 4627. JSON ganhou maior utilização com o advento do Ajax. A serialização em JSON é muito simples e resulta em uma estrutura pouco verbosa o que se mostra uma ótima alternativa para o XML.

JSON possibilita serialização de estrutura de objetos complexos, como listas e subpropriedades. JSON está se tornando o padrão mais utilizado para integração de dados entre repositórios e frameworks, também está se tornando o padrão nativo de armazenamento em alguns bancos de dados modernos.

- **XML** significa Extensible Markup Language. É um conjunto de regras para codificar documentos com estrutura hierárquica e em um formato legível por máquina. É baseado em texto e tem como principais objetivos simplicidade, extensibilidade e usabilidade. XML é largamente utilizado como formato de troca de dados nos clássicos Web Services SOAP. Possui uma ampla gama de ferramentas associadas, tais como o padrão XSLT que permite transformar para outra estrutura XML ou outro formato. Apesar de sua ampla utilização, tem sido menos encorajada a utilização desse formato para integração de aplicações na Web, por utilizar mais recursos para transmissão e para o processamento do dados. Em substituição, recomenda-se utilizar JSON.
- **CSV** significa Comma-Separated Values, ou valores separados por vírgula, e é um formato para armazenamento de dados tabulares em texto. A codificação é muito simples: cada linha do arquivo representa uma linha na tabela, e as colunas são separadas por vírgula. Campos que podem conter vírgula devem ser delimitados por aspas. CSV é recomendado para representação de estrutura de dados mais simples, de natureza tabular, onde não existem subpropriedades ou listas, gerando um arquivo menor e mais leve para processamento. Arquivos CSV são processáveis diretamente por editores de planilhas, como o OpenOffice e o MS Excel.
- **ODS** significa Open Document Spreadsheet, é um formato não proprietário de arquivo baseado em XML, padronizado pela ABNT sob a norma NBR ISO/IEC 26300:2006. É comumente chamado de planilha, similar ao XLS do MS Office Excel, porém aberto, por isso deve ser utilizado em substituição ao XLS. Planilhas são largamente utilizadas, são de fácil utilização e manipuláveis por diversos aplicativos. Apesar de ser um formato estruturado, é muito flexível, possibilitando manipulação e mistura de diversos tipos de dados, como imagens e textos formatados. Para a publicação de dados abertos tabulares, é recomendável a utilização de CSV, pela sua simplicidade e padronização.
- **RDF** significa Resource Description Framework, é um modelo de dados estruturado em grafos e possui diversos formatos de serialização, tais como RDF/XML, Notation 3 e Turtle. Os formatos baseados em RDF têm seus dados descritos em vocabulários disponíveis na Web. Apesar da grande qualidade dos dados disponibilizados em RDF, a construção de vocabulários para seu uso não é trivial. Numa escala de níveis de qualidade/complexidade de dados abertos, o RDF está no último nível, onde se constituirá a Web semântica.

Além desses existem outros formatos, como: SVG (Scalable Vector Graphics) utilizado para dados vetoriais e geográficos, GML (Geography Markup Language) útil para exprimir características e exportação de dados geográficos, HTML/RDFa é a incorporação de marcações semânticas com uso de RDF sobre as tags HTML de uma página.

Para saber mais sobre formatos abertos consulte a e-PING (Padrões de Interoperabilidade de Governo Eletrônico) em <http://eping.governoeletronico.gov.br/>.

Para a publicação de dados cartográficos e espaciais, consulte padronização na Infraestrutura Nacional de Dados Espaciais - INDE em <http://inde.gov.br/>.

A escolha do padrão a ser utilizado pelo órgão na publicação dos dados abertos deve levar em consideração a capacidade tecnológica do órgão e o público alvo que utilizará aqueles dados.

É desaconselhável o empacotamento de diversos arquivos assim como a compressão de arquivos.

Em casos especiais, na existência de muitos arquivos, ou manipulação de arquivos com grande capacidade de compressão, recomenda-se a utilização de formatos abertos de compactação como o 7Z, TAR/GZIP ou ZIP.

• Metadados

Metadados são dados sobre os dados, ou seja, são informações que possibilitam organizar, classificar, relacionar e inferir novos dados sobre o conjunto de dados. A quantidade e a qualidade dos metadados de um conjunto de dados podem determinar a utilidade daquele conjunto de dados. Em outras palavras, mais e melhores metadados agregam mais valor ao conjunto de dados, além de melhorar sua classificação e a busca sobre ele.

A seguir são definidos os metadados obrigatórios e os metadados desejáveis.

Metadados obrigatórios

1. **Título:** Nome do conjunto de dados.
2. **Descrição:** Uma breve explicação sobre os dados.
3. **Órgão responsável:** Nome e sigla do órgão ou entidade responsável pela publicação do conjunto de dados.
4. **Categorias no VCGE:** O Vocabulário Controlado de Governo Eletrônico é uma lista hierarquizada de assuntos do governo que utiliza termos comuns e é voltada para a sociedade. Para navegar e escolher as categorias acesse o VCGE em <http://vocab.e.gov.br/2011/03/vcge>.
5. **Recursos:** Um conjunto de dados pode ser composto por mais de um arquivo de dados. O critério básico para separar um conjunto de dados em mais de um recursos ou arquivo de dados é a constatação de que eles divergem em vários metadados.
 - a) **Título:** Nome do recurso.
 - b) **Formato:** Formato do recurso. Ex.: XML, JSON, CSV, etc.
 - c) **Descrição:** Breve detalhamento sobre o conteúdo do recurso.

Metadados desejáveis:

1. **Etiquetas:** Lista de palavras chaves relacionadas ao conjunto de dados, e que são úteis na classificação e busca dele.
2. **Autoria:** Instituição ou pessoa responsável pela produção do conjunto de dados.
3. **Mantenedor:** Instituição ou pessoa responsável pela manutenção do conjunto de dados
4. **Documentação:** URL de documento que expõe detalhes sobre o conjunto de dados.
5. **Cobertura geográfica:** Localização ou região geográfica a que se referem os dados. Ex.: Recife.
6. **Cobertura temporal:** Data ou período à que referem os dados. Ex.: 03/2012.
7. **Metodologia:** Processo de criação dos dados.
8. **Vocabulário/ontologia:** Documentos estruturados com metadados específicos do conjunto de dados.

• Cadastro do Administrador de Dados Abertos

O órgão ou entidade integrante da IMDAFOR deve solicitar o cadastro de seus Administradores de Dados Abertos, os quais são autorizados a publicar e manter seus conjuntos de dados no Catálogo de Dados Abertos de Fortaleza. Estes, após validação e finalização de seus cadastros, receberão as informações de login e senha de acesso à área restrita do Catálogo de Dados Abertos de Fortaleza. Recomenda-se o cadastro de 02 (dois) administradores.

A solicitação de cadastro dos Administradores de Dados Abertos do órgão ou entidade é feita através de ofício à CITINOVA encaminhando o Formulário de Cadastro de Administrador de Dados Abertos, disponível no endereço xxxxxxxxx, preenchido e devidamente assinado pelo representante legal do órgão ou entidade, com os seguintes dados:

1. Dados pessoais do Administrador de Dados Abertos
2. Dados funcionais do Administrador de Dados Abertos
3. Dados pessoais e funcionais do solicitante do cadastro

Figura 1 apresenta o fluxo de solicitação de cadastro do Administrador de Dados Abertos de um órgão ou entidade.

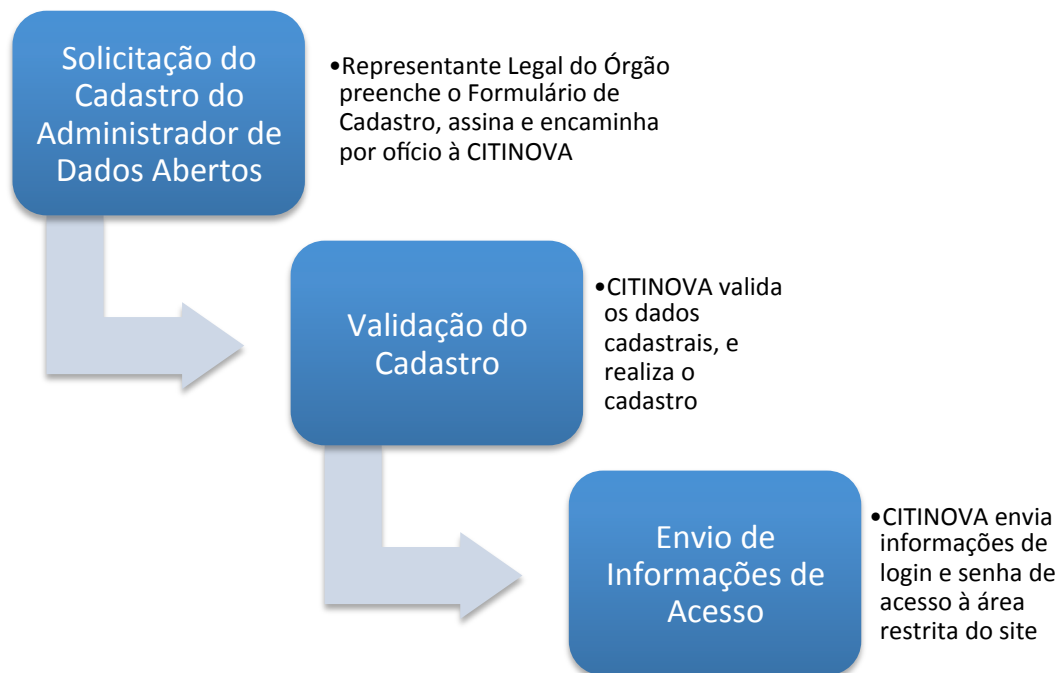


Figura 1. Fluxo de Solicitação de Cadastro de Administrador de Dados Abertos

• **Publicação de Dados e Metadados**

A publicação é o processo de tornar disponível de forma permanente um conjunto de dados no Catálogo de Dados Abertos de Fortaleza, para uso irrestrito da sociedade. Este processo compreende não apenas a publicação dos dados em si, mas também dos seus metadados.

A IMDAFOR se reserva o direito de comissionar representantes para analisar e validar os formatos, tamanhos e metadados do conjunto de dados publicados. Caso alguma não-conformidade seja identificada, os ajustes são comunicados por e-mail ao órgão ou entidade de origem dos dados para providências.

Figura 2 apresenta o fluxo de publicação de conjuntos de dados no Catálogo de Dados Abertos de Fortaleza.

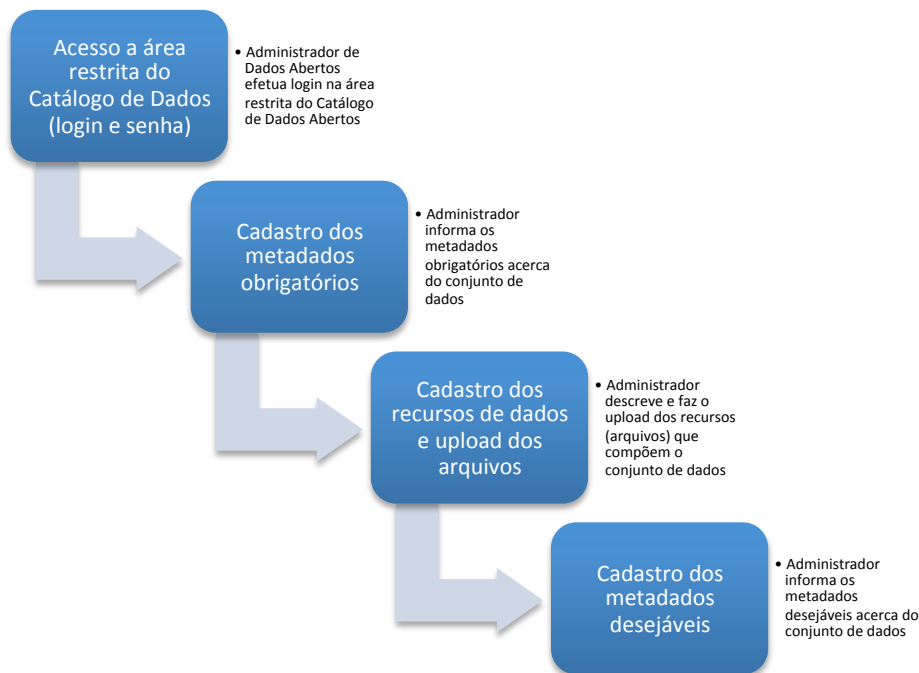


Figura 2. Fluxo de Publicação de Conjuntos de Dados Abertos

• Manutenção de Dados e Metadados

Ao longo do tempo e após sua publicação, um conjunto de dados ou seus metadados podem sofrer modificações ou atualizações e estas devem ser refletidas no Catálogo de Dados Abertos de Fortaleza. O órgão ou entidade pode decidir realizar esta manutenção através da inclusão de um novo recurso do mesmo conjunto de dados ou pela publicação de um novo conjunto de dados. **A seguir algumas diretrizes para auxiliar nesta decisão:**

Figura 3 apresenta o fluxo de manutenção de conjuntos de dados no Catálogo de Dados Abertos de Fortaleza.

7. Coordenação e Manutenção do Portal

A coordenação do Portal Fortaleza Dados Abertos é da IMDAFOR e a manutenção do mesmo é de responsabilidade da Fundação CITINOVA da Prefeitura Municipal de Fortaleza.

Atualmente, o Portal Fortaleza Dados Abertos está em fase beta para avaliação e evolução. Nesta fase, a CITINOVA auxiliará os órgão e entidade na catalogação e publicação de seus dados.

Qualquer pessoa ou entidade pode participar do desenvolvimento do portal através da wiki do projeto em <http://wiki.imdafor.fortaleza.ce.gov.br/>

• **Macrocronograma**

Marco	Prazo
Lançamento do Portal dados.fortaleza.ce.gov.br (Versão beta)	Maio/2014
Avaliação e evolução do portal	Maio/2014 a Julho/2014
Mobilização dos órgãos e crescimento do catálogo	Abril/2014 a Julho/2014

• **Situações Especiais**

A seguir algumas situações especiais do processo de abertura e publicação de dados públicos:

1. O órgão ou entidade possui um conjunto de dados públicos de amplo interesse, tem a intenção de publicá-lo, porém não possui recursos (tecnológicos e humanos) para realizar a preparação e a publicação dos dados. A Fundação CITINOVA conta com um time de consultores de Dados Abertos à disposição do governo. Para saber mais entre em contato através do formulário Fale Conosco do Portal Fortaleza Dados Abertos (<http://www.dados.fortaleza.ce.gov.br>).
2. O órgão ou entidade já tem a cultura de publicar dados, porém os requisitos contidos neste documento não são totalmente contemplados. Por exemplo, o órgão publica dados de forma estruturada porém em formato proprietário. O órgão deve publicar esses dados no Catálogo de Dados Abertos de Fortaleza (vide subseção 6. Publicação, neste documento), e iniciar ações para adaptação e alinhamento com o disposto neste documento.

• **Manutenção do Catálogo**

O Catálogo de Dados Abertos de Fortaleza faz referência à dados publicados por órgãos do governo municipal e entidades integrantes da IMDAFOR. Cada órgão é responsável pela preparação, submissão e atualização dos conjuntos dos dados e metadados que dissemina no Catálogo. Esta atividade deve ser desempenhada por servidor do órgão indicado no momento da adesão deste órgão à IMDAFOR.