

Guia de Abertura de Dados

2ª Edição
15/05/2012

Sumário

Introdução.....	3
1. Objetivos.....	3
2. Público Alvo.....	3
3. Como usar o documento.....	3
4. Processo de construção do documento.....	3
5. Governo Aberto – Dados como Infraestrutura.....	4
5.1. Conceitos Básicos.....	4
5.2. Melhores Práticas.....	5
Guia de Abertura de Dados.....	7
Parte I – Aspectos Gerenciais.....	7
1. Alinhamento com a EGTI.....	7
2. Alinhamento estratégico para viabilizar o processo de abertura.....	7
3. Definir os atores da iniciativa e seus papéis.....	8
4. Levantar os grupos de dados para a abertura.....	8
5. Plano de Divulgação e Participação.....	9
6. Plano de continuidade e manutenção.....	10
Parte II – Aspectos Técnicos.....	11
7. Modelar os dados antes da abertura.....	11
8. Determinar as interfaces e formatos dos pacotes de dados.....	12

Introdução

1. Objetivos

Este documento pretende contextualizar e orientar as instituições detentoras de dados públicos no processo de disponibilização desses dados de acordo com os princípios de Dados Abertos. As orientações cobrem aspectos gerenciais do processo de abertura, e devem ser utilizadas como instrumento norteador para definição de um conjunto de ações para implantar uma cultura de abertura de dados sustentável, alinhada com os objetivos estratégicos da organização e com as demandas sociais por governos mais abertos.

O documento também fornece várias orientações técnicas, auxilia na formação de uma equipe com habilidades essenciais para Dados Abertos, e na sua gestão para atingir os objetivos de publicação de dados. Para uma orientação tecnológica mais aprofundada na construção da solução de dados abertos, a SLTI disponibiliza também um documento relatando a experiência de abertura dos dados do Sistema de Cadastramento Unificado de Fornecedores - SICAF. Este documento está em anexo, e segue todos os padrões propostos e deve ser enxergado como arquitetura modelo de referência para soluções de dados abertos.

2. Público Alvo

A publicação deste documento vem para sustentar as ações de abertura de dados em consequência da crescente demanda social por governos mais abertos. Neste sentido, o público alvo são gestores responsáveis pela administração de sistemas de informação e de seus dados nos mais diversos órgãos públicos interessados em fornecer acesso à esses dados em formatos abertos, reutilizáveis, com facilidade de cruzamentos com outras fontes de dados seguindo as melhores práticas atuais.

Apesar da mudança de cultura iniciar no alto escalão das organizações, é fundamental que os responsáveis diretos pela manutenção dos sistemas de informação estejam envolvidos no planejamento e desenvolvimento das ações de abertura de dados. Grande parte dos sistemas de informação dos órgãos públicos, assim como seus respectivos bancos de dados, têm sua manutenção terceirizada. Nessas condições é fundamental o entendimento desses atores quanto aos objetivos pretendidos no processo de abertura de dados.

A estratégia de abertura de dados aqui apresentada defende que, uma vez que a sociedade é o maior beneficiário desses dados, deve considerar a participação social processo fundamental na construção dos canais ideais de disseminação desses dados.

3. Como usar o documento

Este documento deve ser usado como um modelo e/ou roteiro dos procedimentos organizacionais necessários para publicação de dados abertos governamentais. Pode e deve ser usado em conjunto com o Manual de Dados Abertos (http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf), traduzido e adaptado a partir do Open Data Manual (<http://opendatamanual.org>) da Open Knowledge Foundation.

4. Processo de construção do documento

Esta minuta é uma proposta elaborada pelo Time de Dados Abertos da Secretaria de Logística e Tecnologia da Informação – SLTI e pelo grupo de trabalho da Infraestrutura Nacional de Dados Abertos – INDA.

A proposta é baseada em algumas experiências de Publicação de Dados Abertos na Dataprev e no Ministério do Planejamento. São considerados, ainda, casos de sucesso internacionais, tais como os padrões, as licenças, os catálogos e os formatos de dados publicados nas iniciativas de Dados Abertos dos Estados Unidos da América e do Reino Unido, bem como as ferramentas da Open Knowledge Foundation.

O momento atual é de submissão para apreciação dos participantes, tomadores e formadores de opinião da e-PING, de governo, da academia e da sociedade organizada, visando formar uma comunidade para o desenvolvimento colaborativo do guia e de um movimento coordenado de Abertura de Dados do Brasil.

5. Governo Aberto – Dados como Infraestrutura

As tecnologias disponibilizadas nos anos recentes têm proporcionado o desenvolvimento de diversas ferramentas que aumentam a capacidade da sociedade em assumir seus direitos e obrigações cívicas. A inclusão digital, a informatização dos procedimentos governamentais e a integração entre os diversos repositórios de dados públicos gera crescentes demandas da população por mais transparência e participação através de meios tecnológicos.

Nessa direção governos de diversos países estão definindo políticas e desenvolvendo plataformas tecnológicas no intuito de promover a disseminação das informações públicas de maneira a possibilitar a reutilização, principalmente pela sociedade, de forma automatizada através da Internet. Principais referências são os governos americano e britânico.

De maneira geral a sociedade (grupos organizados, iniciativa privada, grupos acadêmicos, etc.) têm a capacidade de cruzar informações de diversas fontes e desenvolver aplicações e serviços que resolvam demandas de grupos específicos da sociedade. Essa estratégia beneficia não só a sociedade mas também o governo uma vez que transfere para a sociedade o provimento de serviços que muitas vezes não estão dentre suas prioridades. É muito importante perceber a gama de aplicações que a sociedade é capaz de desenvolver.

Por outro lado, na medida em que mais dados governamentais estejam disponíveis de forma aberta, espera-se que o próprio governo passe a gradualmente utilizar esses dados abertos como plataforma ágil de integração entre sistemas de informação, assim promovendo a interoperabilidade dentro da esfera pública. Além disso, essas integrações são benéficas, ainda, no sentido de tornar as iniciativas de dados abertos algo permanente e sustentável, na medida em que o próprio governo passe a depender delas. Portanto, faz sentido como estratégia de Abertura de Dados estimular o consumo pelo governo de dados abertos governamentais.

5.1. Conceitos Básicos

O movimento internacional de abertura de dados governamentais está embasada em 3 leis propostas por David Eaves:

Se o dado não pode ser encontrado e indexado na web, ele não existe.

Se não estiver aberto e em formato compreensível por máquina, ele não pode ser reaproveitado.

Se algum dispositivo legal não permitir sua reaplicação, ele não é útil.

Para maior esclarecimento o Open Government Working Group definiu 8 princípios:

Completos: Todos os dados públicos estão disponíveis. Dado público é o dado que não está sujeito a limitações válidas de privacidade, segurança ou controle de acesso.

Primários: Os dados são apresentados tais como os coletados na fonte, com o maior nível de granularidade e sem agregação ou modificação.

Atuais: Os dados são disponibilizados tão rapidamente quanto necessário à preservação do seu valor.

Acessíveis: Os dados são disponibilizados para o maior alcance possível de usuários e para o maior conjunto possível de finalidades.

Compreensíveis por máquinas: Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado.

Não discriminatórios: Os dados são disponíveis para todos, sem exigência de requerimento ou cadastro.

Não proprietários: Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo.

Livres de licenças: Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso devem ser permitidas.

5.2. Melhores Práticas

Abaixo, uma seleção de boas práticas na publicação de dados abertos, levantada em uma análise nos portais de disseminação de dados, em conformidade com as orientações do W3C:

1. Identificador Persistente

A não persistência dos identificadores dificulta a busca e o encontro dessas informações, segundo as três leis de David Eaves, “se o dado não pode ser encontrado e indexado na web, ele não existe.” Além disso, os identificadores devem ser tais que o usuário possa inferir alguma informação útil sobre aquilo que identificam.

2. Metadados

Segundo o W3C, “Sem documentação, os dados não são muito úteis. Quando possível, deve-se usar padrões da indústria, tais como aqueles baseados em XML/RDF, que normalmente são auto Documentáveis.” A atribuição de Metadados facilita o processamento por máquinas e o correlacionamento de dados pelas pessoas.

3. Padrões de arquivo aberto

Para que o acesso a esses dados ocorra de maneira isonômica, é necessário que eles possam ser interpretados por ferramentas gratuitas e possuam formatos abertos.

4. Clareza nos direitos de Uso

Levando em consideração que os dados estão sendo divulgados publicamente, é necessário que sejam explicitados os direitos de uso dessas informações, e que seja analisado quais são possíveis de se utilizar abertamente.

5. Ferramentas para consulta e indexação

Com o atual volume de dados existente nas bases do governo, nada mais adequado que sua correta organização e indexação, o aspecto mais importante é manter esses dados ordenados de uma forma que o leigo possa encontrar o que quer.

Tão importante quanto a organização dos dados é o provimento de uma ferramenta que facilite seu encontro, uma ferramenta de busca bem estruturada ajuda o interessado a encontrar seus dados de forma mais rápida e garante seu retorno quando precisar de novas informações.

6. Ambiente de feedback

Não menos importante que os demais, o ambiente de feedback é a ferramenta essencial para proporcionar a melhoria contínua da qualidade do portal de disseminação.

Ao passo que existe sua importância, é evidente que existe um grande custo envolvido em ouvir as sugestões e ter o comprometimento de implementar a melhoria.

Esse ambiente, caso seja de fato administrado e utilizado, é um marco para a maturidade de um portal ou sistema de disseminação.

Guia de Abertura de Dados

1 Parte I – Aspectos Gerenciais

Este documento propõe uma atividade transversal e longitudinal às diversas áreas e hierarquias da organização, portanto sentiu-se a necessidade de dividi-lo em duas linhas de visão, gerencial e técnica, os tópicos a seguir orientam a primeira delas.

1. Alinhamento com a EGTI

A Estratégia Geral de TI (<http://www.governoeletronico.gov.br/sisp-conteudo/estrategia-geral-de-ti>) é o instrumento base para que os órgãos do Sistema de Administração dos Recursos de Informação e Informática (SISP) elaborem os seus Planos Diretores de Tecnologia da Informação (PDTI), visando o aprimoramento institucional e a maturidade da governança de Tecnologia da Informação.

Na última edição deste documento, a Estratégia Geral de TI (EGTI) ainda estava em fase de desenvolvimento. Entretanto, como ferramenta de planejamento e alinhamento do SISP, proverá algumas diretrizes que gerarão a necessidade de publicar dados públicos em padrões abertos. Dentre elas, a principal é a da transparência governamental.

Essa necessidade resultou na elaboração deste documento, que pretende atingir parte desse objetivo e ir além das expectativas, trazendo não só os previsíveis benefícios das combinações dos dados governamentais, mas criando uma plataforma de colaboração, firmando um marco na participação do cidadão nas práticas do Estado.

2. Alinhamento estratégico para viabilizar o processo de abertura

Entende-se que esse momento de articulações é muito específico para cada instituição, é aqui que o coordenador da iniciativa garante, através do patrocinador, o apoio das áreas envolvidas para prosseguir com o trabalho.

Para a obtenção desse alinhamento, devem ser apresentados os objetivos da publicação de dados abertos e seus benefícios, a visão de que os dados abertos são um recurso estratégico substitui a antiga idéia de que dado é um subproduto da atividade do governo.

A cultura da inteligência de negócio, atualmente utilizada como um diferencial competitivo na iniciativa privada mundial, é um caso de sucesso que demonstra de que maneira o governo pode e deve ser mais eficiente. O reuso dos indicadores de resultados já é uma cultura nacional, o que falta é um processo coordenado no qual a essa reutilização e análise possa ser feita de diversas maneiras diferentes. Um exemplo prático: ao invés de investir em tratamento e análise de dados, direcionar esse investimento na publicação desses dados e promoção da análise de maneira colaborativa com a sociedade organizada, transformar o governo numa plataforma de colaboração, o governo 2.0.

Um exemplo de produto dessa fase é um documento formalizando o início do processo de

abertura, com apoio e aprovação dos tomadores de decisão da organização.

3. Definir os atores da iniciativa e seus papéis

A definição dos responsáveis é importante para formalizar a vinculação do ônus aos representantes das áreas envolvidas, a seguir a representação dos que são considerados os principais responsáveis no processo de abertura:

3.1. Órgão central de suporte e controle às ações

Promove encontros periódicos para criação e evolução dos modelos de plano de ação e manutenção dos dados, com fim de formalizar um processo integrado de governo: SLTI.

3.2. Patrocinador

Responsável pela iniciativa, principal interessado na abertura dos dados, Executivo de nível superior do órgão.

3.3. Coordenador/Articulador

Responde pela execução do processo de abertura, articula entre os responsáveis técnicos e gestores de negócio.

3.4. Gestor Negocial

Tem a capacidade de identificar os dados a serem abertos, os dados de risco, dados confidenciais. Tem conhecimento dos processo de negócio.

3.5. Gestor Técnico

Deve dominar, na organização, os assuntos de ambiente e segurança e conhecer os sistemas de suporte à área de negócio, responde também pelas ações técnicas do processo de abertura(extração e apresentação dos dados).

3.6. Usuário do dado

Cliente final do dado aberto, pessoa física ou jurídica que tem interesse de acesso àquele dado: Sociedade civil, em especial os grupos de cidadãos organizados com capacidade técnica e *hackers* cívicos; iniciativa privada e os próprios gestores públicos.

4. Levantar os grupos de dados para a abertura

Após o mapeamento dos envolvidos no processo de abertura, são iniciadas as atividades conjuntas de levantamento dos dados passíveis de abertura. Antes mesmo que os dados sejam preparados e efetivamente publicados, é preciso encontrá-los, simplificá-los e adquirir permissão para disponibilizá-los.

Em teoria o movimento de abertura de dados prega a publicação total dos dados o mais rápido possível. Porém, os recursos disponibilizados aos órgãos possuidores dos dados não prevêm ações de publicação dos dados nos moldes atualmente discutidos. As ações sugeridas a seguir consideram a colaboração como peça importante na priorização dos dados a serem publicados, otimizando a utilização dos recursos e reduzindo as incertezas quanto ao impacto das publicações.

4.1. Levantar os grupos de dados possuídos pela organização:

O primeiro passo é conhecer os ativos do órgão. Para que todos os grupos de dados sejam

elicitados é essencial o conhecimento dos negócios operacionais geridos pelo órgão. O levantamento deve cobrir quais são os dados, gerados ou mantidos, cujo o órgão é responsável. O Gestor Negocial deve analisar os processos de negócio do órgão e identificar todos os grupos de dados produzidos no processo.

Participantes: Coordenador e Gestor Negocial.

4.2. Obter autorização para publicação dos grupos de dados do órgão:

Para que um grupo de dados seja disponibilizado publicamente é imprescindível um levantamento jurídico e permissão formal da instância responsável, isso porque podem existir instrumentos jurídicos, portarias, instruções normativas ou afins que proíbam a publicação de certas informações. Deve-se tomar como premissa que todos os dados podem ser liberados, e remover aqueles que possuem alguma restrição. É importante enfatizar que mesmo parcialmente publicado, um grupo de dados agrega algum valor. É aceitável que motivos relacionados a segurança nacional e estratégia de governo restrinjam os dados, mas justificativas fortalecem a atuação transparente do órgão.

Participantes: Coordenador e Gestor Negocial.

4.3. Definir metadados preliminares:

O movimento de abertura de dados está embasado em algumas leis e princípios que defendem o acesso irrestrito e automatizável aos dados. Porém, para o sucesso do movimento é importante facilitar a compreensão dos dados. Se um pacote de dados publicado não está gerando conclusões ou algum serviço, então ele não atingiu a expectativa inicial. É papel do provedor dos dados fornecer artefatos suficientes permitindo a máxima utilização do valor agregado nos dados.

Sugerimos a utilização de uma abordagem incremental. No primeiro momento é importante a construção de metadados suficientes para que um usuário externo conheça a definição e escopo dos dados.

Participantes: Gestor Negocial, Gestor Técnico e Usuário.

4.4. Determinar quais grupos de dados são mais relevantes para abertura:

O movimento de abertura de dados preconiza que todos os dados devem ser disponibilizados o mais rápido possível. Entretanto, frente à possível escassez de recursos disponíveis diante do montante de dados que podem ser disponibilizados, é necessário um planejamento priorizando a abertura dos dados mais importantes. A definição do que é importante está relacionada ao valor percebido pelo consumidor dos dados.

Recomenda-se a utilização de uma abordagem colaborativa a definição dos dados a serem abertos, de modo a construir um cronograma participativo de prioridades na abertura dos dados. Com o registro das atividades, quantidade de recursos utilizados, dificuldades encontradas e sucessos obtidos, sendo itens transparentes para os usuários interessados, é possível criar uma agenda de abertura e reduzir as incertezas quanto ao impacto e alcance da abertura posterior, buscando o melhor balanço entre a demanda e a possibilidade de oferta. A definição dos metadados preliminares, descrita no tópico 4.3, é essencial para que os usuários vislumbrem as possibilidades.

Participantes: Coordenador, Gestor Negocial, Gestor Técnico e Usuário.

5. Plano de Divulgação e Participação

O objetivo dos esforços de divulgação é maximizar o alcance e o impacto das ações de abertura de dados. Por isso, considerando que o paradigma de dados abertos é orientado à promoção da transparência, colaboração e participação, torna-se essencial a existência de um plano de divulgação para alcance dessa meta.

O usuário tem um papel essencial no processo de abertura de dados, portanto o plano deve conter ações para seu envolvimento ainda na etapa de priorização e escolha dos domínios de dados a serem publicados. Esse processo colaborativo aumenta a transparência da ação e reduz as incertezas quanto aos resultados da abertura dos dados.

O plano de divulgação pode ser especializado para cada público de usuário específico dos dados. A definição de papéis propõe o melhor aproveitamento dos colaboradores de acordo com o propósito de uso dos dados.

5.1. Sociedade civil

O cidadão é o principal usuário dos dados disponibilizados e suas aplicações. O plano deve conter ações informativas do movimento como um todo, esclarecendo motivação, benefícios e meios de acesso. Também deve promover a utilização das aplicações que fazem uso dos dados disponibilizados. Essas aplicações fornecem as informações contextualizadas aumentando a percepção dos impactos das ações de governo, ou seja, maior transparência.

5.2. Grupos de cidadãos organizados, com capacidade técnica e Hackers cívicos

O plano deve prever a colaboração com esses grupos em várias etapas do processo de abertura. Têm a capacidade de especificar os requisitos técnicos de formato e meio de acesso aos dados. São considerados pontes facilitadoras para acesso à informação contida nos dados.

5.3. Acadêmicos

As universidades e centros de pesquisa constituem grupos com capacidade técnica e visão de tendência. São parceiros importantes para geração e aprimoramento de práticas e ferramentas. O processo também tem muito a ganhar através análises multidisciplinares que podem ser feitas por especialistas nas mais variadas áreas do conhecimento.

5.4. Gestores públicos

Apesar de o movimento primar pela disponibilização dos dados aos cidadãos, os gestores públicos também se beneficiarão. A abertura organizada e orquestrada dos diversos órgãos do governo, constitui uma oportunidade única de interoperabilidade. Além disso, os serviços e aplicações desenvolvidas sobre os dados abertos também podem ser utilizados como ferramenta de gestão. Essas aplicações devem ser vistas como investimento privado em prol do Governo.

5.5. Imprensa

A imprensa tem um papel importante no ecossistema de dados abertos, ela potencializa o alcance dos dados públicos, por já ser um meio de comunicação conhecido e buscado pelos cidadãos. Dados de interesse da imprensa têm chances de se tornarem muito visualizados e utilizados pela sociedade. Quem for publicar dados abertos não deve temer o impacto causado por alguma má interpretação dos dados se mantiver a fonte sempre disponível para conferência dos interessados.

O incentivo mais percebido para utilização dos dados, diferente dos meios usuais, nos quais

cada ator desenvolve aplicações para atender a seus interesses, é a motivação do desenvolvimento de aplicações sobre dados de utilidade pública através de torneios de programação. O governo, provedor dos dados, promove torneios premiando àquelas aplicações inovadoras socialmente reconhecidas. Alguns exemplos:

- Apps for climate action(A4CA)
- Apps for californians
- Open.Up
- Apps for democracy

6. Plano de continuidade e manutenção

Como esclarecido anteriormente, as ações descritas neste documento constituem esforços iniciais para publicação de dados abertos. Todavia, é imprescindível a utilização de planos de continuidade do processo de abertura dos dados, o maior desafio dos governos que praticam a publicação de dados abertos é tornar suas iniciativas projetos sustentáveis. Nesse sentido, o plano deve cobrir dois aspectos:

6.1. Manutenção do processo de abertura

O modelo de processo sugerido neste documento prevê a priorização dos pacotes de dados a serem abertos tendo em vista a disponibilidade de os poucos recursos. Porém, a abertura de dados precisa ser encarada como assunto estratégico e incorporado aos processos de criação de dados do órgão. O plano de continuidade deve prever ações para melhoria do processo de abertura através do feedback do usuário e troca de experiências com outros órgãos, contemplando o cronograma para que a sociedade possa acompanhar a execução do plano.

6.2. Manutenção dos dados já disponibilizados

O plano deve conter ações de manutenção dos pacotes de dados já publicados. A partir do momento que o órgão disponibiliza um pacote de dados através de um endereço Web, assume o compromisso de mantê-lo. Grande parte dos dados são séries temporais que, provavelmente, são gerados continuamente pelo órgão. O processo de abertura de um pacote de dados deve prever sua atualização indeterminada. A atualização com versões mais recentes dos dados deve acontecer sempre que possível preservando seu valor agregado. A frequência de atualização, e seus custos operacionais, devem ser considerados na etapa de definição da cobertura temporal daquele pacote de dados.

2 Parte II – Aspectos Técnicos

7. Modelar os dados antes da abertura

7.1. Definir quais são os detalhes fundamentais a serem abertos, gerando um modelo inicial.

Dos grupos selecionados e autorizados, é importante determinar quais são os dados centrais, o eixo inicial que traz sentido na abertura daquele conjunto. Tomemos como exemplo o grupo de dados de pagamento de benefícios da Previdência Social: o maior interesse nesse grupo é em cima do benefício em si, seu valor e qual a agência pagadora, no entanto esse dado tem seu valor anulado sem a data do pagamento do benefício, e reduzido sem a localização da agência.

Em torno dos dados fundamentais será discutido o acréscimo dos demais para agregação de valor, essa discussão será muito mais proveitosa se contar com a participação dos usuários que ajudaram na definição dos grupos de dados mais importantes.

Participantes: Coordenador, Gestor Negocial, Gestor Técnico e Usuário.

7.2. Analisar aspectos de Privacidade, Segurança e Legalidade.

É necessário avaliar se a publicação desses dados pode comprometer algum desses itens em seu estado natural:

- Segurança nacional;
- Segredos estratégicos de governo;
- Direito de privacidade do cidadão;
- ou Legislação

Caso positivo será necessária sua adequação: o procedimento comum é o de anonimização, no qual são retirados do grupo os dados responsáveis por esse comprometimento.

Cabe ressaltar que a anonimização não se trata de simplesmente abstrair os dados que identificam prontamente o objeto (normalmente pessoa), o processo usa como subsídio análises das formas de se “desanonimizar” os dados através de cruzamentos com outros conjuntos de dados. Outra ressalva é que quanto mais anônimo o dado, menor sua utilidade, portanto deve-se ser cauteloso ao retringir partes importantes dos dados.

Participantes: Coordenador, Gestor Negocial, Gestor Técnico e Especialista em anonimização.

7.3. Definir conjunto de metadados a ser demonstrado para cada pacote de dados, um informativo que descreve o que será apresentado.

É importante informar do que se trata um grupo de dados, isso pode ser feito com uma interface prévia à de apresentação dos dados. Não estamos falando de web-semântica, esse ainda é um passo adiante, aqui os metadados são contextualizadores: faixa de tempo dos dados, data de publicação, direitos de uso, etc.

Participantes: Coordenador, Gestor Negocial e Gestor Técnico.

8. Determinar as interfaces e formatos dos pacotes de dados

A interface de acesso a um grupo de dados é determinante para o sucesso da iniciativa. Sua escolha deve equilibrar dois pontos: as restrições trazidas por uma infraestrutura tecnológica limitada e a necessidade de se publicar dados em quantidade e com a maior qualidade possível.

Alguns grupos de dados podem ser pouco volumosos, o que justifica o uso de tecnologias mais modernas e completas, no entanto, deve-se avaliar a taxa de atualização desses dados e seu consequente crescimento natural antes de adotar qualquer método de publicação.

8.1. Levantamento das limitações de infraestrutura.

Aqui devemos definir com relativa precisão quais são os recursos disponíveis para essa abertura, essas informações, farão parte das variáveis que restringirão as características elegíveis para os pacotes de dados e seu acesso.

Exemplos de limitações:

- Espaço em servidores de dados;

- Disponibilidade de servidores de aplicação;
- Banda de rede;

8.2. Formatos.

A escolha do formato errado, dependendo do volume de dados possuídos, pode gerar um arquivo de tamanho muito grande. A manutenção e o elevado número de *downloads* simultâneos desse tipo de arquivo podem causar sobrecarga nos servidores. Um exemplo comum é a utilização de um arquivo XML ou RDF no lugar de um CSV: essa escolha vai, na melhor das hipóteses, dobrar o tamanho do arquivo. Por outro lado, o valor agregado resultante da riqueza de metadados introduzidos por esses formatos é extremamente vantajoso para o processamento automático.

A decisão do formato deve ser baseada no equilíbrio entre as limitações de infra e a quantidade de valor agregado desejado para os dados.

Exemplos de arquivos:

- XML;
- CSV;
- RDF;
- HTML.

8.3. Protocolo de acesso

Assim como na escolha do formato, o protocolo deve ser avaliado antes da adoção, alguns meios de acesso podem exigir mais ou menos esforço do servidor ou até mesmo menos banda de download.

Quando não se pode evitar a produção de arquivos de tamanhos grandes, uma boa saída para escapar do tráfego exorbitante é o uso de Torrent, o funcionamento do protocolo é baseado no compartilhamento distribuído dos arquivos, dividindo o acesso ao servidor com outras pessoas que já fizeram download desses arquivos.

Exemplos de protocolos:

- HTTP;
- FTP;
- Endpoint SPARQL;
- Bit Torrent.

8.4. Cobertura dos pacotes

Para o consumidor dos dados é custoso obter um arquivo enorme quando seu objetivo é acessar um conjunto de dados restrito. Uma aplicação que precisa consultar dados de Pernambuco, por exemplo, reduz processamento e economiza tempo, se puder acessar um pacote exclusivo do estado ou um da região Nordeste.

Por outro lado, é trabalhoso para o órgão disponibilizar 27 arquivos agrupando dados por estado ao invés de 5 arquivos por região ou 1 para o país.

Segundo esses motivos, a cobertura de cada pacote é uma decisão estratégica para a publicação: ela influi diretamente nos seus tamanhos e organização.

Nos portais data.gov e data.gov.uk, as divisões por coberturas geográficas e temporais são frequentes, explícitas nos metadados de contextualização do usuário. Essas divisões foram feitas após a percepção da vantagem tanto para o usuário quanto para a organização.

Exemplos de dimensões de cobertura:

- Geográfica
 - Estadual
 - Regional
 - Municipal

- Temporal
 - Semanal
 - Mensal
 - Trimestral
 - Anual

É importante enfatizar que cada órgão deve fazer uma análise individualizada e encontrar outras dimensões relacionadas à natureza do dado gerado ou manipulado.