



















Visando melhor fixação e compreensão das atividades da dinâmica persona, sugerimos que assistam ao vídeo a seguir:

Clique no ícone abaixo para assistir ao vídeo (é preciso estar conectado).



## 2.2 Dinâmica 2 – Seleção e priorização dos conjuntos de dados

Esta dinâmica é uma forma prática de classificar os conjuntos de dados por critérios, que levam em consideração vários aspectos, como a relevância para o cidadão, por exemplo.

## Dinâmica 2 - Seleção e Priorização dos Conjuntos de Dados

### Objetivos:

A atividade tem por objetivo classificar os dados utilizando critérios de priorização coerentes, para, assim, obter uma lista priorizada de conjuntos de dados a serem abertos.

### Quantidade de integrantes:

Grupos de até 6 pessoas.

### Tempo exigido

30 minutos, aproximadamente.

### Materiais utilizados

Para cada grupo: 10 pedaços de cartolina de 10cmx10cm para cada uma das 6 cores diferentes e 10 pedaços de cartolina de 20cmx10cm de uma cor.

**Enap**

Enap  
Enap

### Desenvolvimento da Dinâmica

1. Primeiramente, pensaremos em alguns conjuntos de dados de um órgão que podem ser úteis e reaproveitados pelo cidadão. Uma das melhores práticas adotadas pelos órgãos é verificar os conjuntos de dados mais solicitados na ouvidoria ou via SIC, ou até mesmo em uma consulta pública, em que a sociedade pode explicitar quais dados do órgão ela julga mais interessante para serem disponibilizados em formato aberto. A ideia é se colocar no lugar do cidadão e a partir daí pensar quais dados seriam úteis para a sociedade em geral.



## DICA

**Lembrando que:** é aconselhável que os órgãos, antes da elaboração do PDA, realizem uma consulta pública, uma vez que, dessa forma, saberão quais são os seus conjuntos de dados preferidos e úteis para a sociedade em geral. No entanto, visando à celeridade do processo de elaboração da execução do PDA, o órgão pode elaborar o PDA com os dados que ele julga úteis e relevantes, para depois realizar uma consulta pública, na qual se poderão identificar novos dados com potenciais usos pela sociedade e, durante a revisão periódica do PDA, incluir esses novos dados no documento. É o que fez, por exemplo, o Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC).

[Clique aqui](#) para visualizar o PDA do MCTIC.

2. Assim, de posse desses conjuntos de dados selecionados, faremos a priorização dos mesmos, obedecendo aos seguintes critérios:
  - ✓ Grau de relevância para o cidadão;
  - ✓ Há obrigatoriedade legal ou compromisso assumido pelo órgão na disponibilização daquele dado/informação/serviço;
  - ✓ O conjunto de dados se refere a projetos estratégicos do órgão ou de suas unidades;



CONJUNTO DE DADOS	CRITÉRIO 1	CRITÉRIO 2	CRITÉRIO 3	CRITÉRIO 4	CRITÉRIO 5	SOMATÓRIO
Fundo de garantia por tempo de serviço.	2	2	2	1	1	8
Rendimento médio real efetivo dos empregados do setor público	2	2	1	1	1	7
Índice de desenvolvimento humano municipal	2	2	2	2	2	10



Enap  
Enap

**Enap**

Enap  
Enap

Visando melhor fixação e compreensão das atividades da dinâmica vista acima, sugerimos que assistam ao vídeo a seguir:

Clique no ícone abaixo para assistir ao vídeo (é preciso estar conectado).



Ao finalizar a dinâmica, você deve obter uma lista priorizada de conjuntos de dados a serem abertos. Em nosso exemplo, o dado prioritário é o Índice de Desenvolvimento Humano municipal, pois o somatório dos critérios atingiu a maior nota, 10 pontos.

## 2.3 Dinâmica 3 – Rascunho do plano de ação

Enap  
Enap

No módulo anterior nós estudamos que, dentre as partes do PDA, existe o Plano de Ação, você se lembra? O propósito dessa dinâmica é que você aprenda como iniciar a elaboração do Plano de Ação e consequentemente a do PDA.



**Enap**

Agora iremos ensinar como fazer a última dinâmica do módulo. Ao final dela, você será capaz de elaborar um rascunho do plano de ação de seu órgão. Basicamente, de posse dos conjuntos de dados selecionados e devidamente priorizados, iremos construir esse rascunho que poderá orientar a execução do PDA. Essa parte do PDA é importante, porque é por meio dela que são atribuídos responsabilidades e prazos de cumprimento das atividades contidas no mesmo. Ele também será consultado no monitoramento da execução do respectivo PDA.

Enap  
Enap

### Dinâmica 3 - Rascunho do Plano de Ação

#### Objetivos:

A atividade tem por objetivo elaborar um rascunho do plano de ação do Plano de Dados Abertos de seu órgão.

#### Quantidade de integrantes:

Grupos de até 6 pessoas.

#### Tempo exigido

30 minutos, aproximadamente.

#### Materiais utilizados

Para cada grupo: 10 pedaços de cartolina de 10cmx10cm para cada uma das 6 cores diferentes e 10 pedaços de cartolina de 20cmx10cm de uma cor.









Além dos requisitos técnicos para a abertura de dados, é fundamental que, para os dados serem considerados abertos, alguns princípios sejam seguidos.

Abordamos esses princípios no primeiro módulo, você ainda se lembra? Tente recordá-los e em seguida clique em revelar resposta, para visualizar os princípios.



## REVELAR RESPOSTA®



1. Os dados devem estar em seu formato mais bruto possível, isto é, sem qualquer tratamento ou agregação. O órgão pode publicar alguma versão do conjunto de dados que esteja tratada ou agregada, mas é indicado também que ele publique a versão bruta do mesmo;
2. Os dados devem estar em formato aberto, não proprietário, estável e de amplo uso; Não deve existir nenhum instrumento jurídico que impeça a reutilização e a redistribuição do conjunto de dados;
3. Para os dados que estão estruturados, deve-se preservar ao máximo a estrutura original. Os arquivos em planilhas, por exemplo, devem estar nos formatos CSV ou ODS, e não no formato PDF;
4. É recomendável a disponibilização dos dados em diversos formatos. Os formatos mais corriqueiros são CSV, XML, JSON, ODF e RDF, que serão detalhados mais adiante;
5. Cada conjunto de dados deve possuir um identificador único e persistente, seguindo uma padronização de URL. Esse requisito é fundamental para que o conjunto de dados seja referenciável e eventualmente consumido por uma aplicação de forma automática, ou seja, se a URL ficar mudando, o dado não conseguirá ser acessado de maneira automática;
6. É recomendável a utilização de considerações dinâmicas na definição de URLs, de forma que seja possível deduzir o conteúdo de um conjunto de dados apenas lendo o seu identificador;

7. É extremamente desaconselhável a utilização de mecanismos antirrobôs, como captchas, para acesso ao conjunto de dados. Isso prejudica, por exemplo, o acesso automático e continuado;
8. É recomendável que os nomes dos arquivos sigam as boas práticas de formação de um slug (parte da URL que aparece depois do nome do domínio principal, por exemplo: no sítio dados.gov.br/ministério-da-fazenda, ministério-da-fazenda é o slug), pois isso facilita a indexação dos dados por mecanismos de busca. 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 ("\_");
  - d. cada conjunto de dados deve ter informações sobre seus dados e metadados. Deve ser possível recuperar o significado dos dados;
  - e. 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;
  - f. 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.



Você se recordava dos princípios? São todos muito importantes!

Enap  
Enap

**Enap**

Enap  
Enap

### 3.1 Principais formatos

No âmbito do tema de Dados Abertos, não existe uma lista taxativa de formatos permitidos. O que é sugerido é que os formatos sejam amplamente conhecidos. Vale ressaltar que existe um erro recorrente cometido por diversos órgãos quando eles publicam em PDF algumas planilhas de dados. O formato PDF não é estruturado e, ao se publicar planilhas em PDF, os dados perdem a estruturação e ficam de difícil processamento para uma eventual automatização da publicação dos mesmos. Os formatos mais utilizados pelos órgãos da APF, uma vez que eles atendem a todos os princípios de dados abertos vistos no módulo anterior, são os seguintes:

#### JSON (JavaScript Object Notation)

É um padrão aberto de estruturação de dados baseado em texto e legível por humano (autodescritivo), sendo de fácil entendimento para pessoas leigas. A serialização em JSON é muito simples e resulta em uma estrutura pouco verbosa, o que se mostra uma ótima alternativa para o XML. O 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 (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. 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 dos dados. Em substituição, recomenda-se utilizar JSON.

#### CSV (Comma-Separated Values)

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. Arquivos CSV são processáveis diretamente por editores de planilhas, como o OpenOffice e o MS Excel.

#### ODS (Open Document Spreadsheet)

É um formato não proprietário de arquivo baseado em XML. É comumente chamado de planilha, similar ao XLS do MS Office Excel, porém aberto, por isso deve ser utilizado em substituição ao XLS. 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 (Resource Description Framework)

É um modelo de dados estruturado em grafos e possui diversos formatos de serialização. 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, no qual se constituirá a *Web* semântica.



Enap  
Enap

Enap

Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap  
Enap

Na plataforma CKAN, os dados estão organizados utilizando as estruturas de conjuntos de dados e recursos. Os conjuntos de dados são as entidades principais retornadas a partir das buscas. Cada conjunto de dados possui uma descrição, um ou mais recursos, e uma série de outros metadados, como periodicidade de atualização e órgão responsável. Cada recurso compreende uma fonte de dados. Pode ser uma planilha, um método de webservice ou um documento. A princípio cada recurso poderia se catalogado como um conjunto de dados separado, porém é preferível que os recursos sejam agrupados, sempre que forem oriundos da mesma base de dados, ou possuírem metadados em comum, para facilitar a busca e o entendimento de seu conteúdo.

Uma das dúvidas mais recorrentes dos órgãos, ao serem publicados os conjuntos de dados, é exatamente a questão de quais metadados devem ser utilizados sobre esses conjuntos. Cada conjunto de dados, ao ser disponibilizado, necessita de uma descrição, isto é, uma ficha com as informações, características e contexto aplicável. Chamamos isso de metadados - 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, metadados agregam mais valor ao conjunto de dados, além de melhorar sua classificação e a busca sobre ele.

Portanto, metadados são informações que tornam possível que outras pessoas encontrem mais facilmente os dados. Continue o curso, pois no próximo tópico trataremos mais sobre esse conceito.



### 3.3 Metadados

Conforme visto no módulo 1, o governo federal desenvolveu o [Portal Brasileiro de Dados Abertos](#), que é o portal de referência (catálogo central) para o acesso aos catálogos de dados do governo pela sociedade. Cada órgão ou entidade pode catalogar direto no portal central





Por fim, sobre o segmento administrativo, sugere-se a abertura das bases:

- Lista de imóveis/equipamentos (patrimônio);
- Lista de imóveis para aluguel.

Lembrando que essas sugestões não excluem a inclusão de outras bases de dados que a instituição achar conveniente abrir para o uso da sociedade.



## SAIBA MAIS

Para conhecer mais detalhes do referido documento acesse [Sugestão de base de dados](#) para as instituições federais de ensino.

Qualquer outra dúvida ou sugestão em relação não só ao processo de planejamento do PDA, mas também ao processo de execução do PDA (que envolve a publicação e a catalogação dos conjuntos de dados), você pode entrar em contato com a equipe de Dados Abertos do Ministério do Planejamento, Desenvolvimento e Gestão por meio do e-mail [contato@dados.gov.br](mailto:contato@dados.gov.br) ou dos telefones 2020-1012 / 2020-1319.

