



# WIKIDATA LAB III

COMO PROPOR E UTILIZAR IDENTIFICADORES

# Introdução

- Lembrar que o Wikidata se propõe a organizar e armazenar dados estruturados sob licença livre, mais especificamente a CC-0. Textos de documentação seguem a tradicional licença CC-BY-SA.

- Os dados estruturados seguem o modelo:

— Item ( <i>item</i> )	→ Propriedade ( <i>property</i> )	→ Valor ( <i>value</i> )
— Brasil	→ Capital	→ Brasília
— Q155	→ P36	→ Q2844

<https://www.wikidata.org/wiki/Q155> / <https://www.wikidata.org/wiki/Property:P36>

- Os dados estruturados acima, contudo, dizem respeito a um grupo de informações chamado “**Declarações**” (*Statements*). Hoje veremos um segundo grupo de informações: os **Identificadores** (*Identifiers*)

# Identificadores

- No Wikidata, os **Identificadores** servem tanto para organização interna do banco de dados (bd) quanto para atrelar o bd do Wikidata a outros bancos.
- Todos os identificadores ficam no domínio *Property* e, portanto, são compostos pela letra “P” seguida por algarismos.
- No dia de hoje, veremos de modo mais específico identificadores externos, ou seja, que atribuem às propriedades em questão links externos. Isso incluiria, por exemplo, até mesmo propriedades como contas no *Facebook*, *Twitter*, *Instagram* ou *Tumblr*.
- Mas passemos a alguns identificadores externos ainda mais específicos: os **controles de autoridade**.

**Leia mais:** `[[Wikidata:Identifiers]]`

# São controles de autoridade reconhecidos



Todos os logos nesta página possuem  
marca registrada.

**Leia mais:** [\[\[Help:Authority control\]\]](#)



# Por que utilizá-los?

- Mais do que atrelar outros bancos de dados a um item no Wikidata, a atribuição de um link externo no Wikidata permitirá num segundo momento cotejar informações e até mesmo trazer informações novas àquele item, principalmente via **raspagem de dados** (*data scrapping*).
- Mas peraí! Que o que é *data scrapping*?
- *Data scrapping* é uma técnica computacional na qual um programa extrai dados de saída legível somente para humanos, proveniente de um serviço ou aplicativo. Os dados extraídos geralmente são minerados e estruturados em um formato padrão como CSV, XML ou JSON.



## *Hands on* e a tarefa desta tarde

- Antes do nosso primeiro *hands on* desta manhã, embora ainda algo mais expositivo, a primeira tarefa é pensar **quais possíveis controles de autoridade você imagina que possa ser útil** aos itens no Wikidata.
- Faremos antes do almoço uma **proposta de controle de autoridade** para conhecermos o “caminho das pedras”.
- Durante a tarde, cada participante do evento será responsável por um controle de autoridade recém aprovado e deverá criar uma tabela de correspondência entre um item no Wikidata e seu link equivalente no banco de dados de um determinado controle de autoridade.
- Mas por que não fazer isso de maneira automática? Porque na maior parte dos casos essa é uma etapa que não é passível de automatização.

# Mas como fazer isso?

- Simples! Basta abrir ir à página de testes atrelada à conta na Wikipédia e criar uma tabela de apenas três colunas, com os campos “nome”, “item no Wikidata” e “controle de autoridade”; cujo preenchimento ficará algo assim:

— Marilena Chaui	-> Q3891586	-> 922
— Mayana Zatz	-> Q2920198	-> 98
— Theatro Municipal	-> Q2297627	-> 79
— Ladeira da Misericórdia	-> Q10315482	-> 38398

- Outra forma de fazer isso é abrir um bloco de notas qualquer, seja no Windows ou no Mac. Para diferenciar as colunas, no entanto, ao invés de dar um espaço entre elas, basta aperta o *Tab*. Ah! Alt shift G exhibe o item Wikidata.

