

Ciência de Dados Espaciais

Aonde Vamos?

Joana Simões¹

¹Eurecat, Centro Tecnológico da Catalunha

September 15, 2015

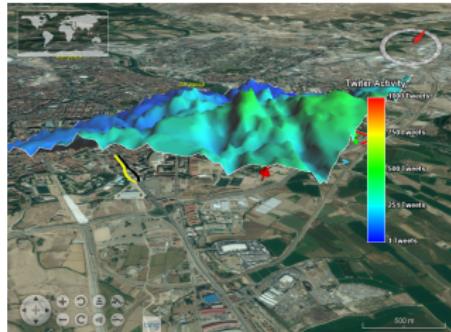


Tabela de Conteúdos

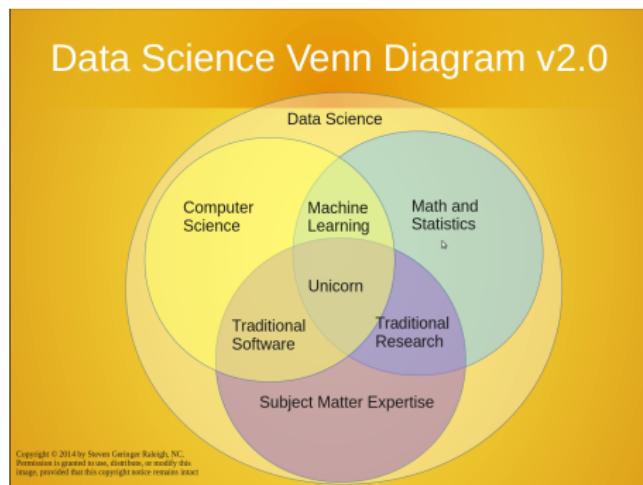
1 Introdução

2 Tendências

3 Considerações Finais

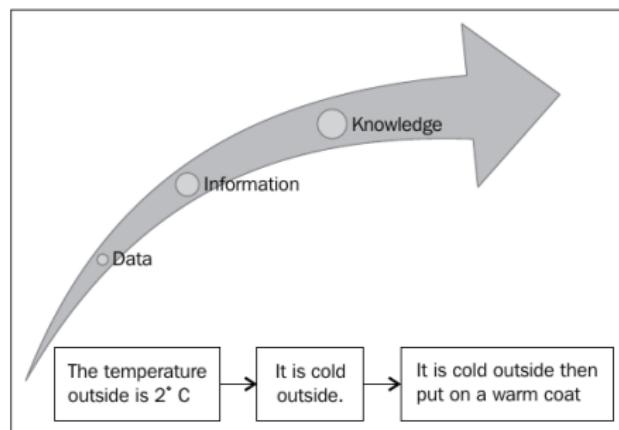
Cientistas & Unicórnios

- “Data Scientist” is a Data Analyst who lives in California.
- A data scientist is someone *who is better at statistics than any software engineer and better at software engineering than any statistician.* (Wills, Cloudera)



Dados, Informação e Conhecimento

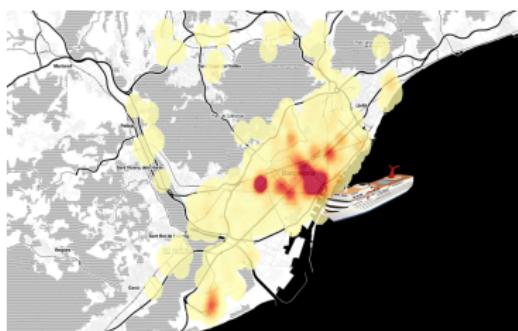
- Dados são os factos que descrevem o mundo.
- A informação surge quando transformamos esses valores em algo relevante.
- Conhecimento implica uma generalização dos dados e da informação, de forma a criar regras.



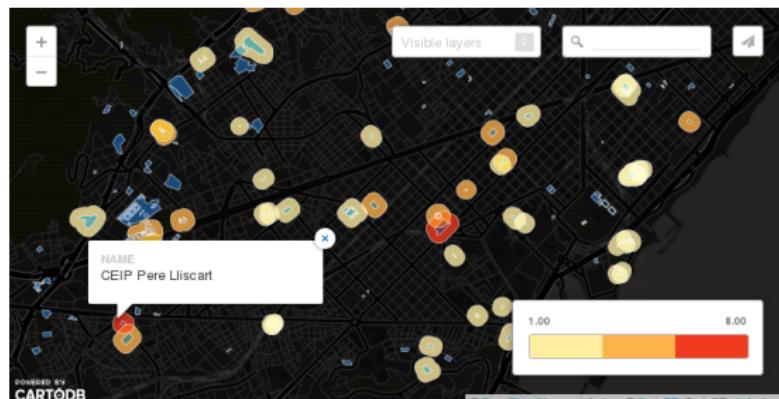
Algo espacial...

Lei de Tobler:

Everything is related to everything else, but near things are more related to each other.



Heatmap de Tweets perto de um Cruzeiro.



Buffers de acidentes ao redor de escolas.

Problemas Fundamentais

Longley *ET AL* (2005):

Problemas Fundamentais

Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padrões passados.

Problemas Fundamentais

Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padroes passados.
- A explicacao no tempo apenas necessita de olhar para o passado, mas a explicacao no espaco necessita de olhar em todas as direccoes simultaneamente.

Problemas Fundamentais

Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padroes passados.
- A explicacao no tempo apenas necessita de olhar para o passado, mas a explicacao no espaco necessita de olhar em todas as direccoes simultaneamente.
- Embora alguns fenomenos espaciais variem de forma gradual atraves do espaco, outros podem exibir uma extrema irregularidade.

Problemas Fundamentais

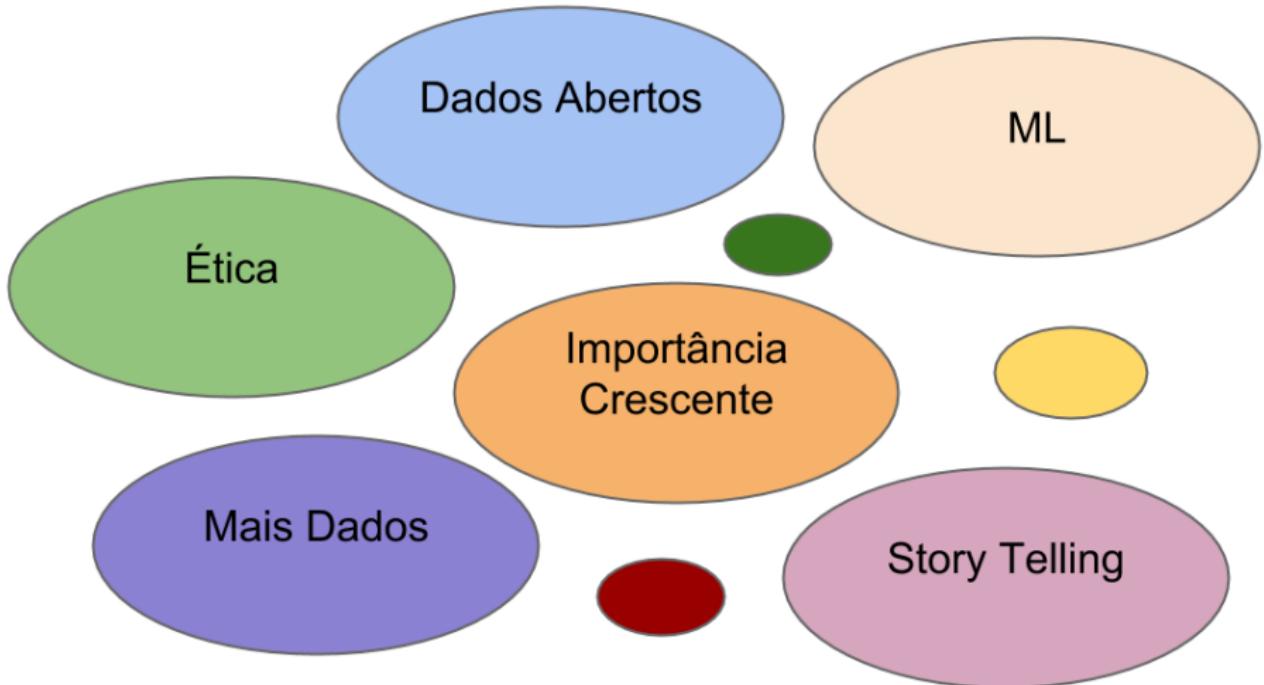
Longley *ET AL* (2005):

- O comportamento espacial actual, muitas vezes reflecte padroes passados.
- A explicacao no tempo apenas necessita de olhar para o passado, mas a explicacao no espaco necessita de olhar em todas as direccoes simultaneamente.
- Embora alguns fenomenos espaciais variem de forma gradual atraves do espaco, outros podem exibir uma extrema irregularidade.
- Embora a autocorrelacao espacial nos ajude a construir representacoes, ela pode frustrar os nossos esforcos de predicao.

Onde Vamos?



(Algumas) Tendências



Importância Crescente

- A ciencia de dados ganha torna-se cada vez mais importante, a medida que entra em novos campos.



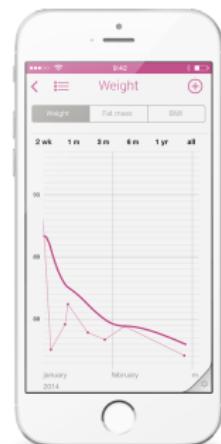
Importância Crescente

- A ciencia de dados ganha torna-se cada vez mais importante, a medida que entra em novos campos.
- A Geografia tambem benefia desse impulso, a medida que o publico em geral toma consciencia de que a maior parte das coisas, acontecem num lugar.



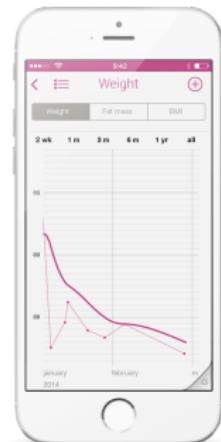
Importância Crescente (cont.)

- *Whitings* é uma balança “inteligente”.



Importância Crescente (cont.)

- *Whitings* e uma balança “inteligente”.
- Recolhe medições corporais precisas: peso, massa gorda e batimentos cardíacos.



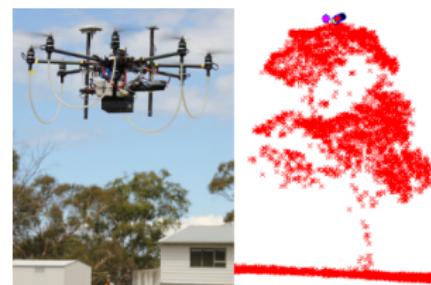
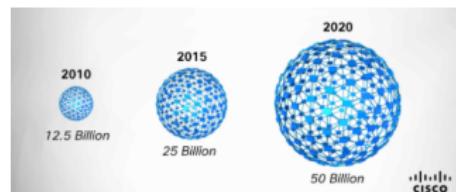
Importância Crescente (cont.)

- *Whitings* é uma balança “inteligente”.
- Recolhe medições corporais precisas: peso, massa gorda e batimentos cardíacos.
- Uma app analisa estes dados, mostra tendências e permite gerar planos e monitorizar metas.



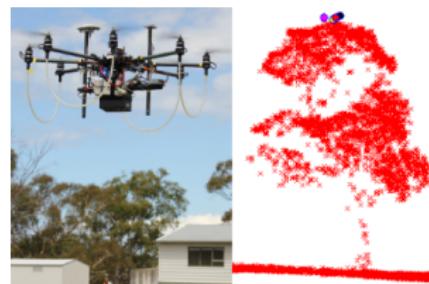
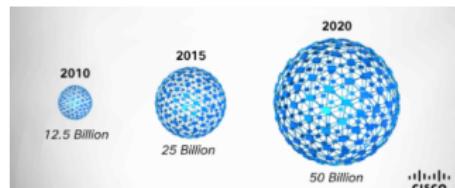
Mais Dados

- Existe uma explosão na quantidade de dados gerados por sensores (IoT, IoE).



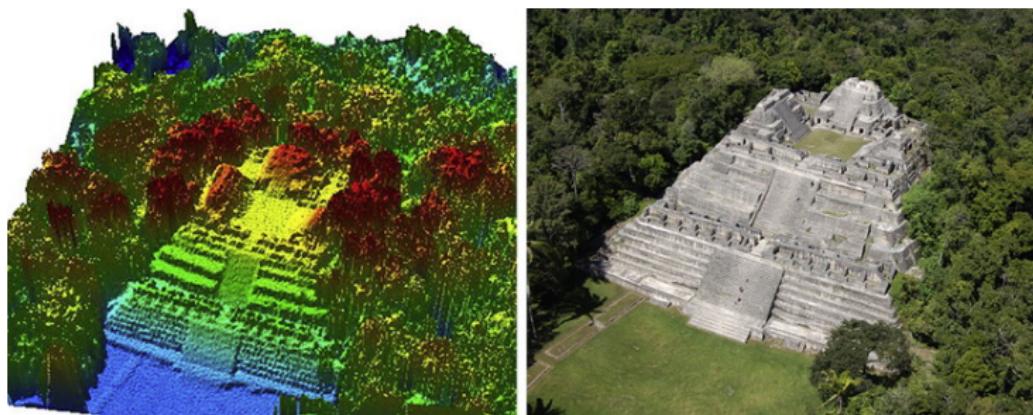
Mais Dados

- Existe uma explosão na quantidade de dados gerados por sensores (IoT, IoE).
- Graças a tecnologias de posicionamento mais baratas e generalizadas (e.g.: receptores de GPS), uma proporção grande destes dados está georeferenciada.



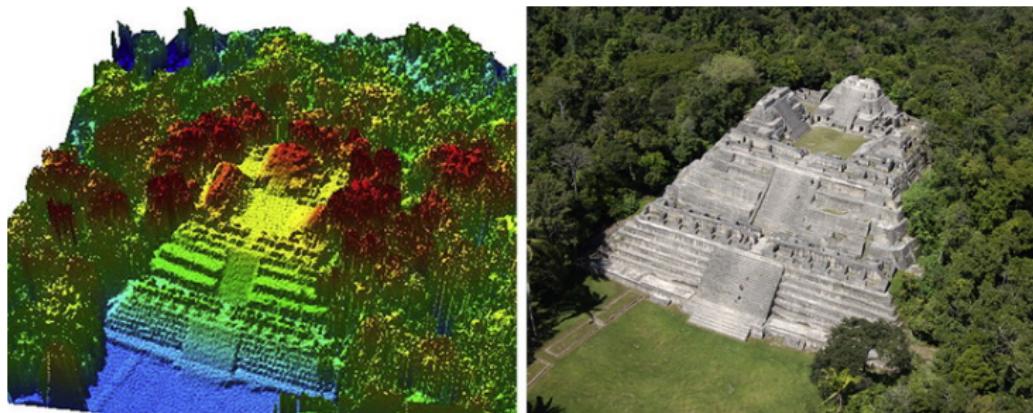
Mais Dados (cont.)

- LIDAR: mede propriedades da luz reflectida de modo a obter a distancia ou outra informacao a respeito um determinado objecto distante.



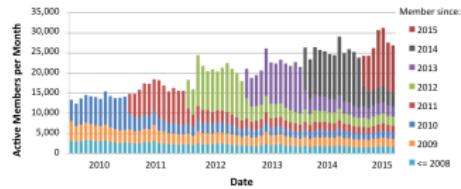
Mais Dados (cont.)

- LIDAR: mede propriedades da luz reflectida de modo a obter a distancia ou outra informacao a respeito um determinado objecto distante.
- O UAV LIDAR aplica esta tecnica a partir de um drone.



E Ainda Mais Dados

- Existe uma explosão na quantidade de UGC.
- voluntário: citizens as sensors e VGI (p.e: OSM).
- Não-voluntário (?): gerados nas redes sociais (p.e: Twitter).

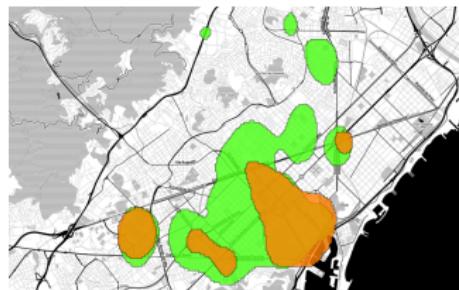


Evolução do número de membros do OSM.

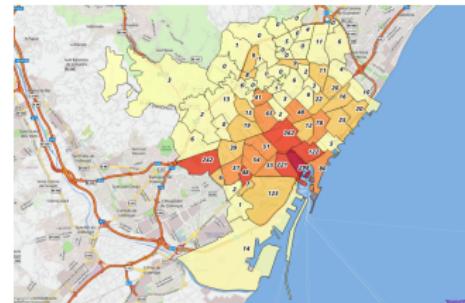


Mapa com a localização do último milhão de Tweets.

E Ainda Mais Dados (cont.)



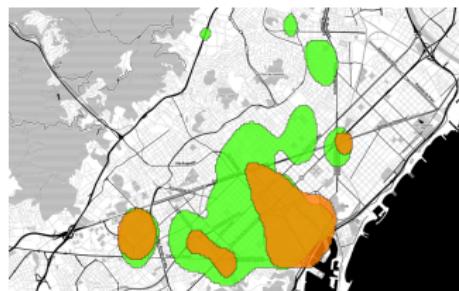
Clusters de Tweets de locais e estrangeiros.



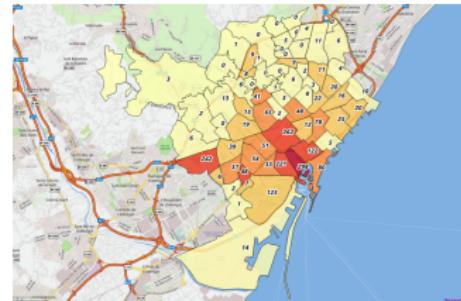
Distribuição de densidades de Tweets de estrangeiros.

E Ainda Mais Dados (cont.)

- Um metodo de deteccao de origem, permite diferenciar os Tweets gerados por locais dos gerados por estrangeiros.



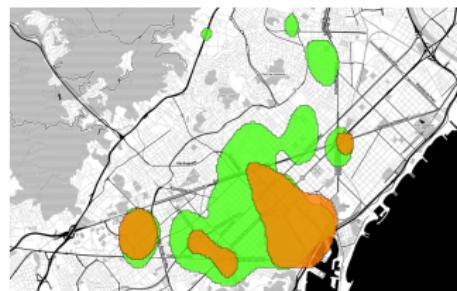
Clusters de Tweets de locais e estrangeiros.



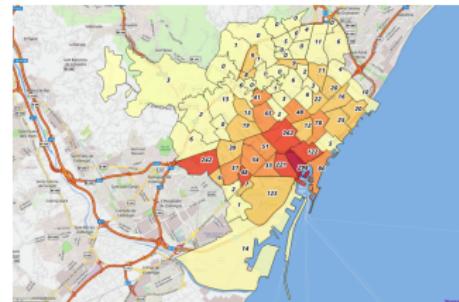
Distribuicao de densidades de Tweets de estrangeiros.

E Ainda Mais Dados (cont.)

- Um metodo de deteccao de origem, permite diferenciar os Tweets gerados por locais dos gerados por estrangeiros.
- Podemos observar que ha um acrescimo no numero de Tweets gerados por estrangeiros, apos a chegada dos cruzeiros ao Porto de Barcelona.



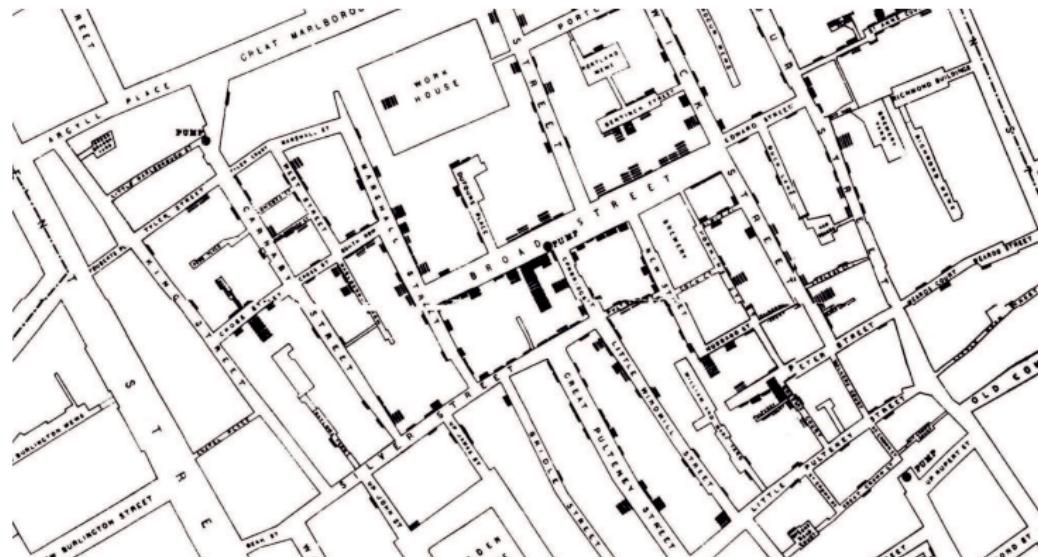
Clusters de Tweets de locais e estrangeiros.



Distribuicao de densidades de Tweets de estrangeiros.

“Arqueologia” de Dados

- Existe uma necessidade crescente de “recuperar” fontes de dados de legacy (por exemplo em formato analógico)



Mapa de casos de colera em Londres, produzido por John Snow (1854).

“Arqueologia” de Dados

- Existe uma necessidade crescente de “recuperar” fontes de dados de legacy (por exemplo em formato analogico)
 - Embora isto implique um desafio tecnologico, muitas vezes estes conjuntos de dados sao extremamente valiosos.



Mapa de casos de colera em Londres, produzido por John Snow (1864).

“Arqueologia” de Dados (cont.)

- Dados de informacao de pescas (biomassas), foram guardados sem metadados.

```

Path line file is "/home/joana/projects/rs/test1/erpl". /home/joana/projects/no/test2/erpl"
RestDir is "/tmp/arg1"
File duplicates:
00777/007771196-4f7-47b7-902d0284 "/home/joana/projects/rs/test1/erpl/arg2/aaest100.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest100.arg" [arg2/aaest100.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest200.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest200.arg" [arg2/aaest200.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest202.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest202.arg" [arg2/aaest202.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest300.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest300.arg" [arg2/aaest300.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest302.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest302.arg" [arg2/aaest302.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest400.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest400.arg" [arg2/aaest400.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest402.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest402.arg" [arg2/aaest402.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest500.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest500.arg" [arg2/aaest500.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest502.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest502.arg" [arg2/aaest502.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest600.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest600.arg" [arg2/aaest600.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest602.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest602.arg" [arg2/aaest602.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest700.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest700.arg" [arg2/aaest700.arg]
2551e/2551e/0225c/0225c/0225c/0225c "/home/joana/projects/rs/test1/erpl/arg2/aaest702.arg" "/home/joana/projects/rs/test1/erpl/arg2/aaest702.arg" [arg2/aaest702.arg]
File duplicates:

```

“Arqueologia” de Dados (cont.)

- Dados de informacao de pescas (biomassas), foram guardados sem metadados.
- Para capturar estes dados, foi necessario efectuar cruzeiros com um custo elevadissimo.

```
Path list file is = ['./home/joana/projects/no/test1/erg2', './home/joana/projects/no/test2/erg1']  
Report list is = ['./erg1.html']  
Output file is = '/tmp/test.html'  
Report duplicates  
00775e7cde5a0642fa9b20f7340f12341 [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxts02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxts02.arg']  
23887537866160279703240f5d63581 [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxtp02.arg', './home/joana/projects/no/test1/erg2/2013402_angla_Pvl_1/sanxtp02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxtp02.arg']  
25115e7cde5a02203b947238bb8bf [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxp02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxp02.arg']  
34115e7cde5a02203b947238bb8bf [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxp02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxp02.arg']  
584bebc11c4d59cc65d47c18051385 [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxt02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxt02.arg']  
688bebc11c4d59cc65d47c18051385 [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxtp02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxtp02.arg']  
688bebc11c4d59cc65d47c18051385 [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxp02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxp02.arg']  
688bebc11c4d59cc65d47c18051385 [./home/joana/projects/no/test1/erg2/2000408_erg2/sanxt02.arg', './home/joana/projects/no/test2/erg1/2000408_erg2/sanxt02.arg']  
1f62017df931385f03f202836518ea [./home/joana/projects/no/test2/erg1/2000_Pelagic_Gobio_Namibia/sanxts02.arg', './home/joana/projects/no/test2/erg1/20006_Pelagic_Gobio_Namibia/Copy of sanxts02.arg']  
2f2000ac02d49a45ba0da3a7f015131 [./home/joana/projects/no/test1/erg2/2000408_erg2/AM05702.MHD', './home/joana/projects/no/test2/erg1/2000408_erg2/AM05702.MHD']  
Path list file is = ['./home/joana/projects/no/test1/erg2', './home/joana/projects/no/test2/erg1']  
Report list is = ['./erg1.html']  
Output file is = '/tmp/test.html'
```

“Arqueología” de Dados (cont.)

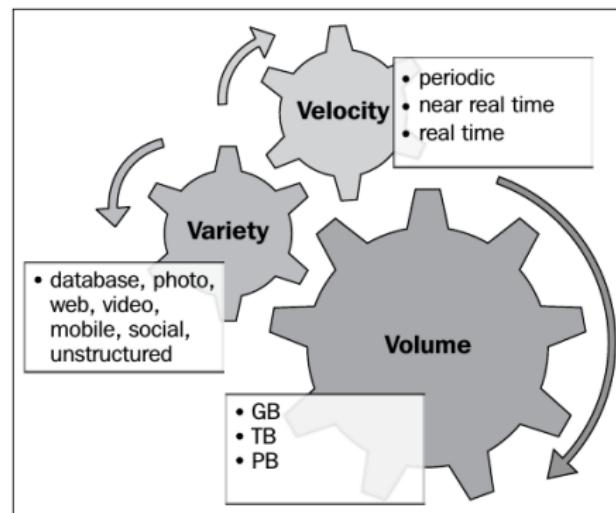
- Dados de informacao de pescas (biomassas), foram guardados sem metadados.
 - Para capturar estes dados, foi necessario efectuar cruzeiros com um custo elevadissimo.
 - A informacao capturada e irrepetivel.

“Arqueología” de Dados (cont.)

- Dados de informacao de pescas (biomassas), foram guardados sem metadados.
 - Para capturar estes dados, foi necessario efectuar cruzeiros com um custo elevadissimo.
 - A informacao capturada e irrepetivel.
 - Os metadados estao encodificados na path do ficheiro (i.e.: timestamp, nome do cruzeiro).

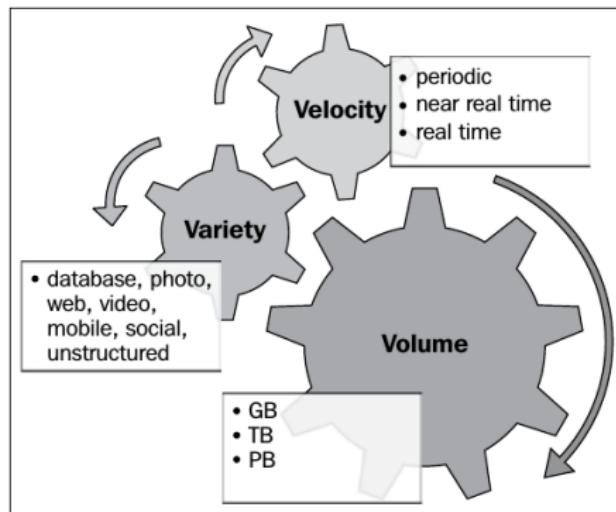
Tecnologias de *Big Data*

- 3 Vs.



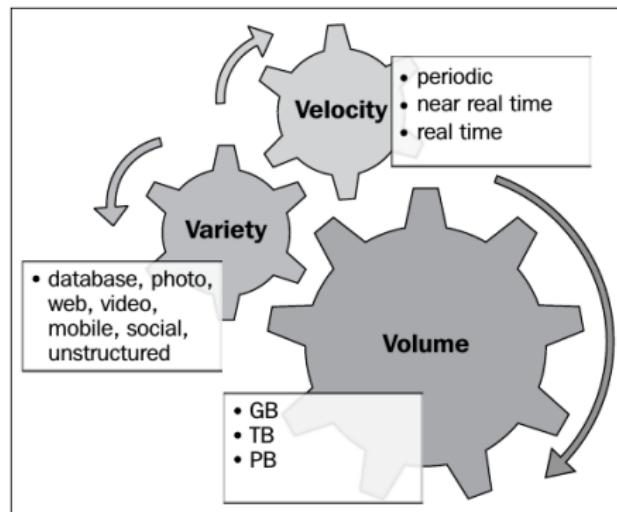
Tecnologias de *Big Data*

- 3 Vs.
- Computacao em ambiente Cloud, NoSQL, Processamento em tempo-real ,Linked Data.



Tecnologias de *Big Data*

- 3 Vs.
- Computacao em ambiente Cloud, NoSQL, Processamento em tempo-real ,Linked Data.



Tecnologias de *Big Data*

- 3 Vs.
- Computacao em ambiente Cloud, NoSQL, Processamento em tempo-real ,Linked Data.
- Big Spatial Data.

