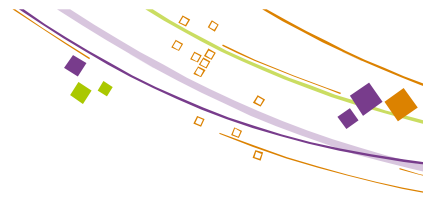


La science des données, une discipline émergente

Faire parler les mesures, de la capture aux premiers mots

Séminaire doctoral, 2015/03/30



How big is a Yottabyte?

TERABYTE	Will fit 200,000 photos or mp3 songs on a single 1 terabyte hard drive.
PETABYTE	Will fit on 16 Backblaze storage pods racked in two datacenter cabinets.
EXABYTE	Will fit in 2,000 cabinets and fill a 4 story datacenter that takes up a city block.
ZETTABYTE	Will fill 1,000 datacenters or about 20% of Manhattan, New York.
YOTTABYTE	Will fill the states of Delaware and Rhode Island with a million datacenters.

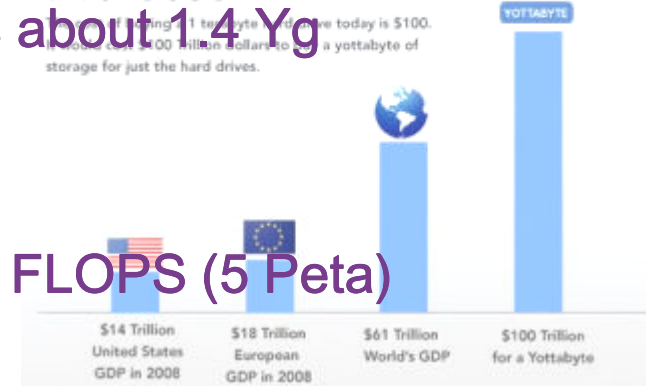


On data avalanche (infobesity)

- Un trait danger, deux traits sécurité (J. Cherhal)
 - The world is continuous (seemingly)
 - We listen to it in a discrete way
- Cheap sensors, tons of data
 - Mass of the Earth: 5972.6 Yg; Oceans is about 1.4 Yg
 - tera-, peta-, exa-, zetta-, yotta-, bronto-?
- Models become increasing complex
 - US Weather Forecasting Finally Gets Its FLOPS (5 Peta)

The Cost

Storing 1 terabyte of data today is \$100. It would cost \$100 trillion dollars to store a yottabyte of storage for just the hard drives.



- Here comes the data scientist, or datamancer
 - From mere sounds to gibberish though words toward insights



Toward datamancy?

- **Machines Teach Astronomers About Stars**
 - Sort thousands of stars in our galaxy; learn their sizes, compositions and other basic traits (JPL, Jan. 2015)





Toward datamancy?

- Early Stage Influenza Detection from Twitter
 - arxiv.org/abs/1309.7340





Toward datamancy?

- How Shazam uses big data to predict music's next big artists (The Guardian, Dec. 2013)



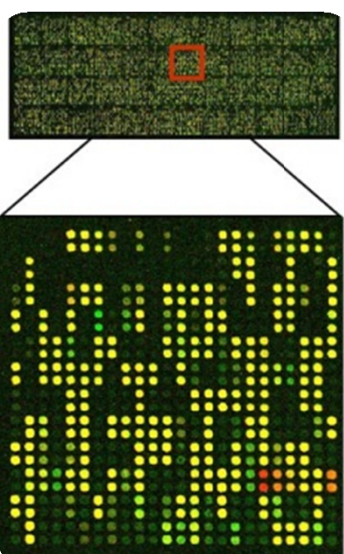
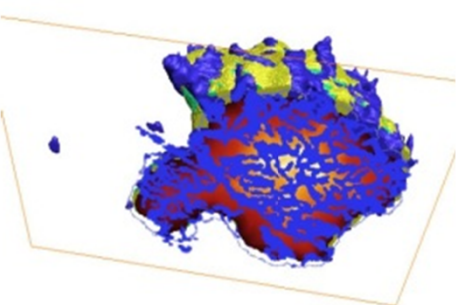
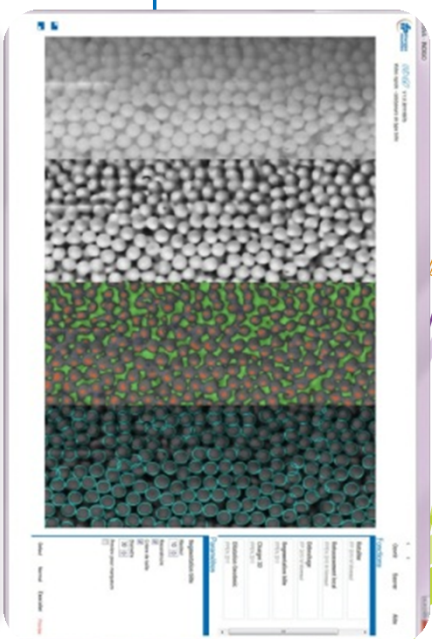
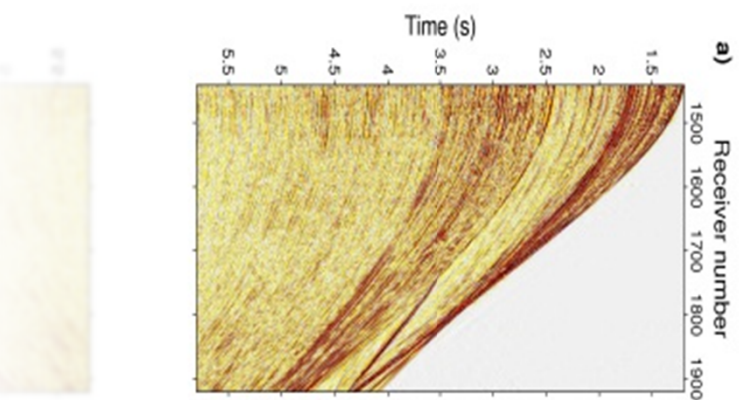
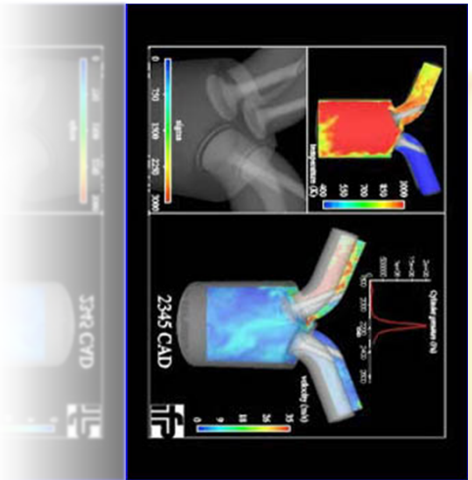
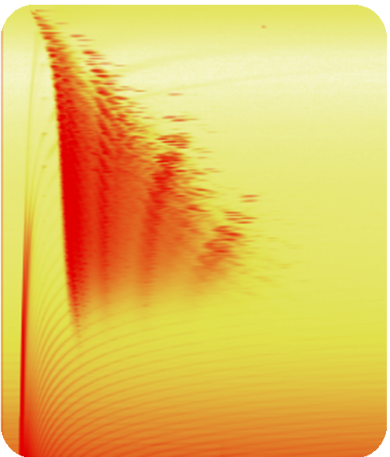


Toward datamancy?

- Building a natural description of images
 - Automatically captioned: “Two pizzas sitting on top of a stove top oven” (Google Research Blog, Nov. 2014)



Data @ IFPEN





Data★Information★Knowledge★Wisdom

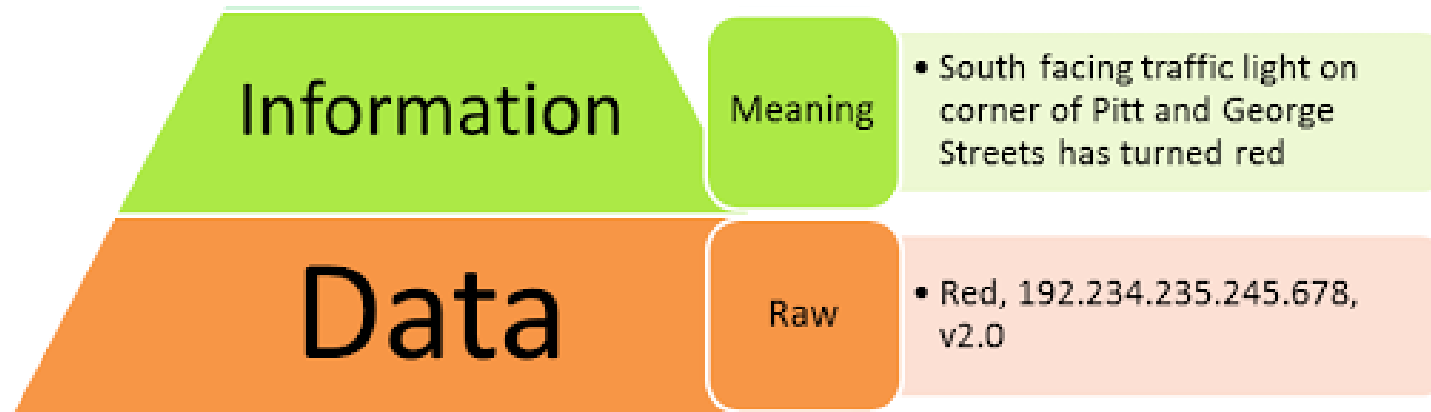
Data

Raw

- Red, 192.234.235.245.678, v2.0

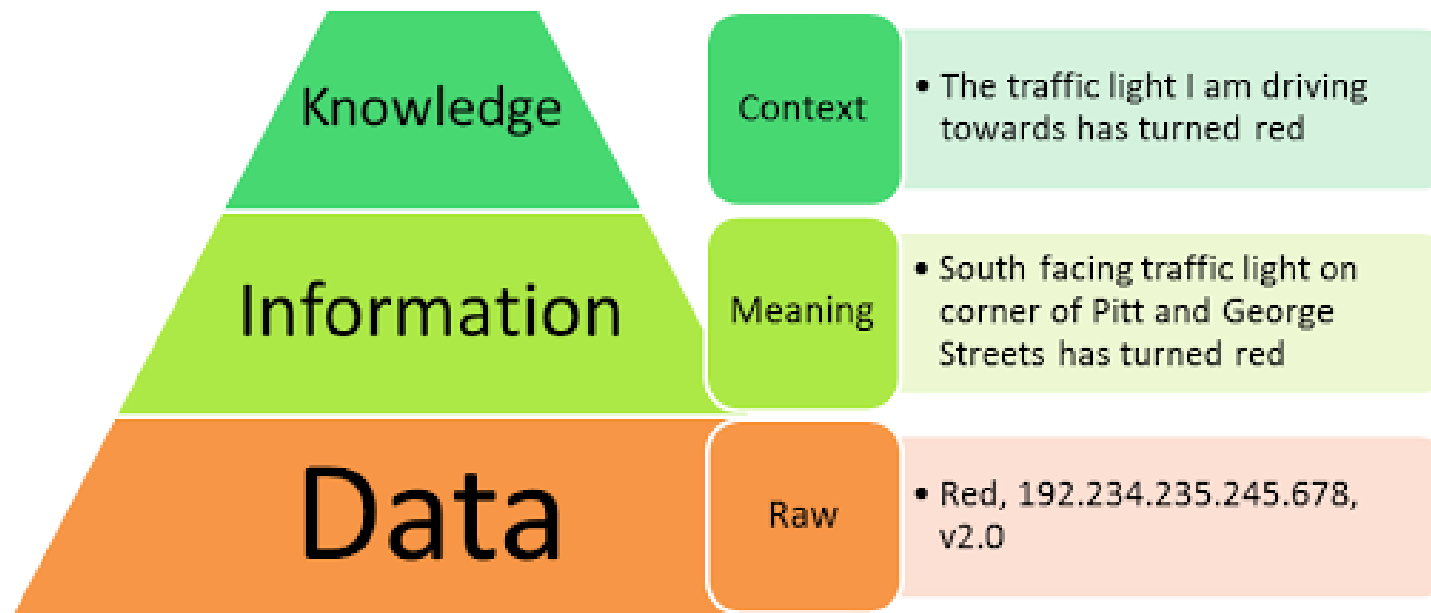


Data★Information★Knowledge★Wisdom



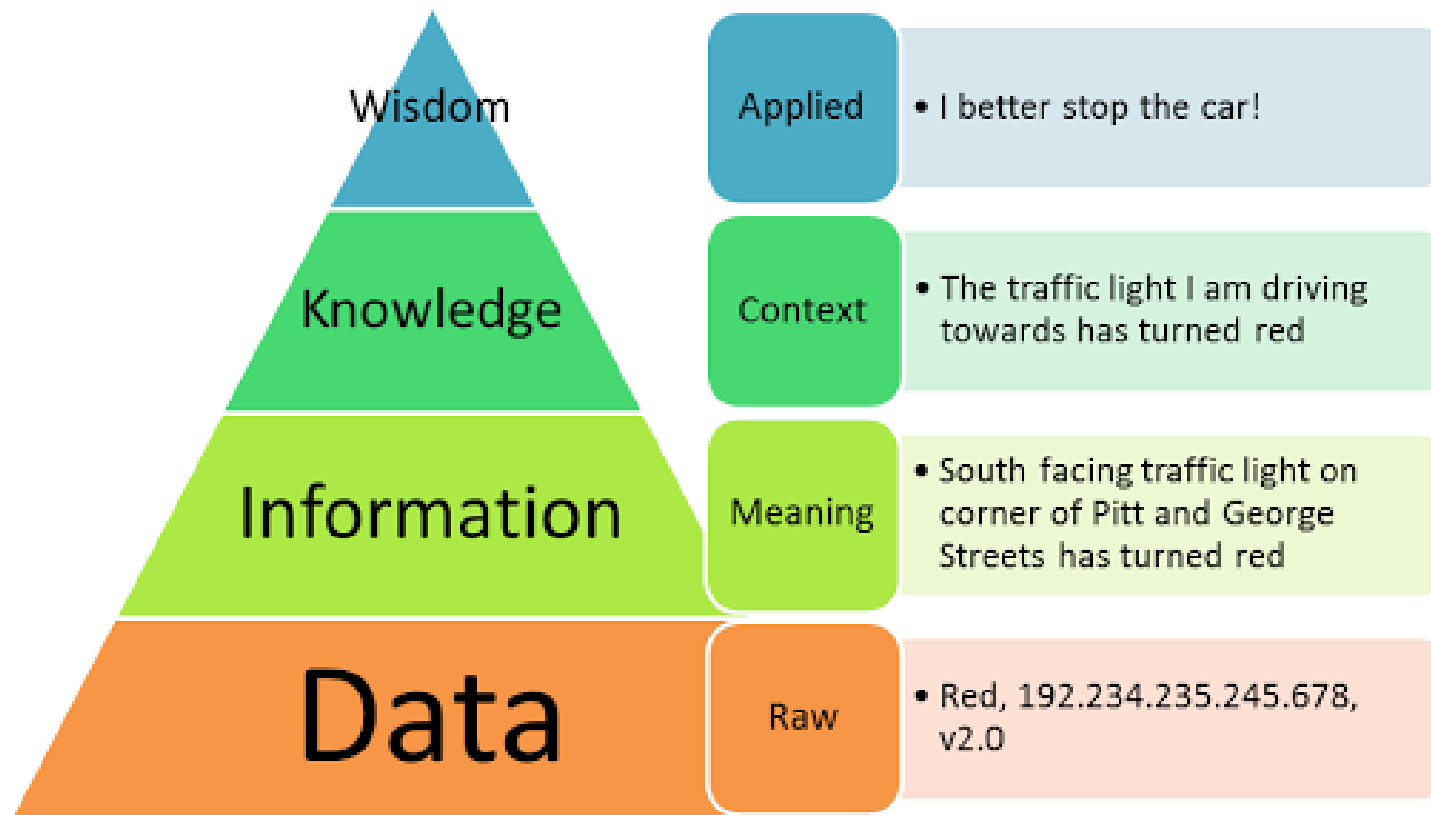


Data★Information★Knowledge★Wisdom





Data★Information★Knowledge★Wisdom





Toward a data-centric research?

- Knowing what we sense
- Using the limits of our systems
- Looking for the needle in the haystack
- Reducing dimensions
- Learning models
- Making sense

The sexiest job in the 21st century? (HBR, Oct. 2012)
Multiple skills required, curiosity first



Program

- 1400-1415: Interlude + introduction
- 1415-1500: Igor Carron (questions)
 - "Ca va être compliqué": Islands of knowledge, Mathematician-Pirates and the Great Convergence
- 1500-1545: Laurent Daudet (questions)
 - Compressed Sensing Imaging through multiply scattering materials
- 1545-1550: Interlude
- 1550-1650: Stéphane Mallat (questions)
 - Learning Signals, Images and Physics with Deep Neural Networks
- 1650-1700: Panel talk



Program

■ Igor Carron

- "Ca va être compliqué": Islands of knowledge, Mathematician-Pirates and the Great Convergence

An unlimited amount of irrelevant data is still irrelevant.
@StatFact



Program

- **Laurent Daudet**

- Compressed Sensing Imaging through multiply scattering materials

There is not a pony at the bottom of every pile of horse manure. @StatFact



Program

- Stéphane Mallat

- Learning Signals, Images and Physics with Deep Neural Networks

The purpose of computing is insight, not numbers.
(Richard W. Hamming, 1962)



Program

- Stéphane Mallat

- Learning Signals, Images and Physics with Deep Neural Networks

The purpose of computing is insight, not numbers.
(Richard W. Hamming, 1962)



Program

■ Panel

Science is the belief in the ignorance of experts.
(Richard Feynman)



Innovating for energy

www.ifpenergiesnouvelles.com