

Big data

"BIG DATA is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it..."

Professor Dan Ariely, Duke University

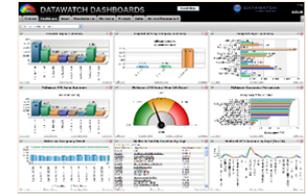
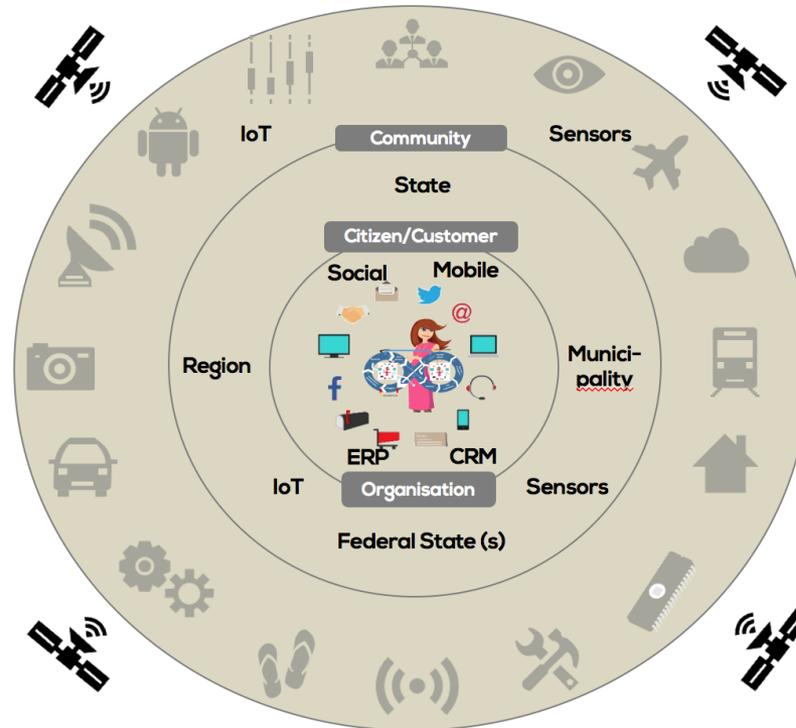
Big Data

In balance between strategy, system and process support

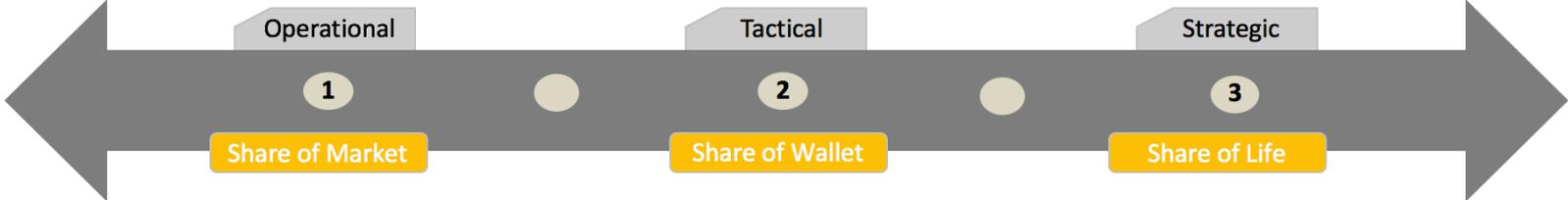
We see a development from employees keys data into systems to users themselves create and upload data via various devices and social medias.

Then add a growing "machine" collected data!

And finally combine this with public data. Then you become knowledge and insights?

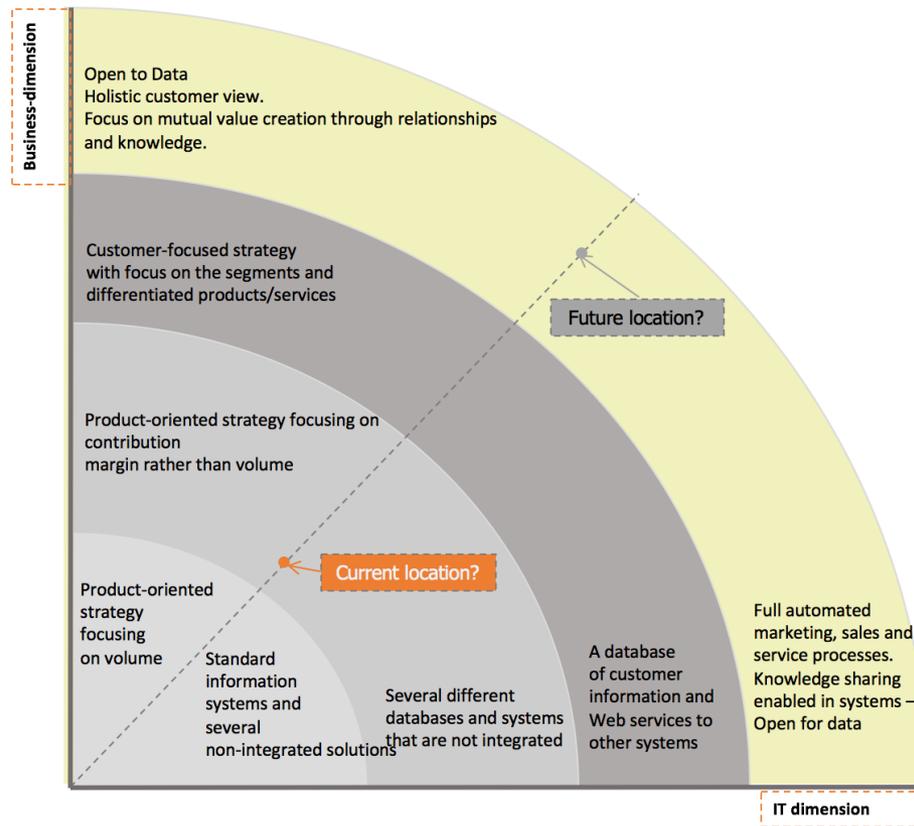


Different levels of Big Data



The company must be Ambidextrous

in balance between strategy, system and process support



Kilde: Efficiens – Per Ø. Jacobsen

Big Data 3V-5V

Gartner Group defines Big Data as (3V model):

“Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.”

3V

Finally Bernard Marr Author, Keynote Speaker and Data Expert, have this definition (5 V model):

“Big Data is a big thing. It will change our world completely and is not a passing fad that will go away.

To understand the phenomenon that is big data, it is often described using five Vs:

Volume, Velocity, Variety, Veracity and Value”

5V

Then IBM invented this model (4V model) -

4V

Big data gives you the ability to achieve superior value from analytics on data at higher volumes, velocities, varieties or veracities.

With higher data volumes, you can take a more holistic view of your subject’s past, present and likely future. At higher data velocities, you can ground your decisions in continuously updated, real-time data.

With broader varieties of data, you can have a more nuanced view of the matter at hand. And as data veracity improves, you can be confident that you’re working with the truest, cleanest, most consistent data. Unlock the value of your big data.

We see Big Data in a different way* –
Therefore we have developed our own 6V model inspired by the 3V, 4V and 5V models.

* Professor Torsten Ringberg, CBS & Per Østergaard Jacobsen

Our model

Big Data is defiantly not just about a huge amount of data from different sources in a structured and unstructured dataset. Big Data is a concept giving the opportunity to find new insight into your existing data as well guidelines to capture and analysis your future data, and the ability to predict customer attitude. This makes business more agile and solid so it can adapt and overcome business challenges to fulfill the big data dogma there is or should be included in the organizations strategy.

Many see data as given free – They might be, but from a customer view, the balance between sharing information and data should be in balance with the value the customer gets back in value and convenience? If not why should customers maintain to give and share information? This in line with the discussion we had about CRM seen as a strategy rather than operational tool. It is the same with Big Data it is strategy.

One main difference in our 6 V model is the “Value” – We see this as the fundamentals for working with Big Data. Having access to Big Data is not valuable unless we can turn it into mutual value, it is useless and just ad on cos if not. So we can safely argue that 'value' is the fundamental basic for Big Data also as a securement for still having access for customer generated data.

Source: Professor Torsten Ringberg, CBS & Per Østergaard Jacobsen

Big data Dogma – 6 V model

Big Data dogma:

- Big data creates mutual value for the citizen/customer and community/business

Value = Volume x Variety x Velocity x Veracity x Virtuety x Visuality



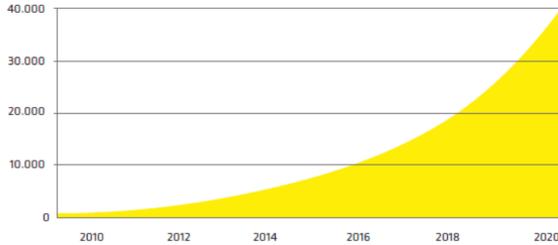
Source: Professor Torsten Ringberg, CBS & Per Østergaard Jacobsen

Volume

- Refers to the amounts of data generated. Just look at emails, social media (Facebook, twitter. Instagram etc.), photos, video clips, sensor data (Internal and external), transaction data, weather etc. that we produce and share every second.
- This is huge amount of data. We are not talking Terabytes, even this for may would seem as lot of date. It is Zettabytes or Brontobytes. On Facebook alone we send more than 10 billion messages per day (at Roskilde Festival we had 61 million observation on Facebook alone, click the "like' button 4.5 billion times and upload 350 million new pictures each and every day.
- If we take all the data generated in the world between the beginning of time and until 2008, the same amount of data will soon be generated every minute! This increasingly makes data sets too large to store and analyses using traditional database technology.
- By using Big Data technology we are able to store and use these data sets with the help of distributed systems, where parts of the data is stored in different locations and brought together by software. So without a clear strategy for this your organization might drown in data!

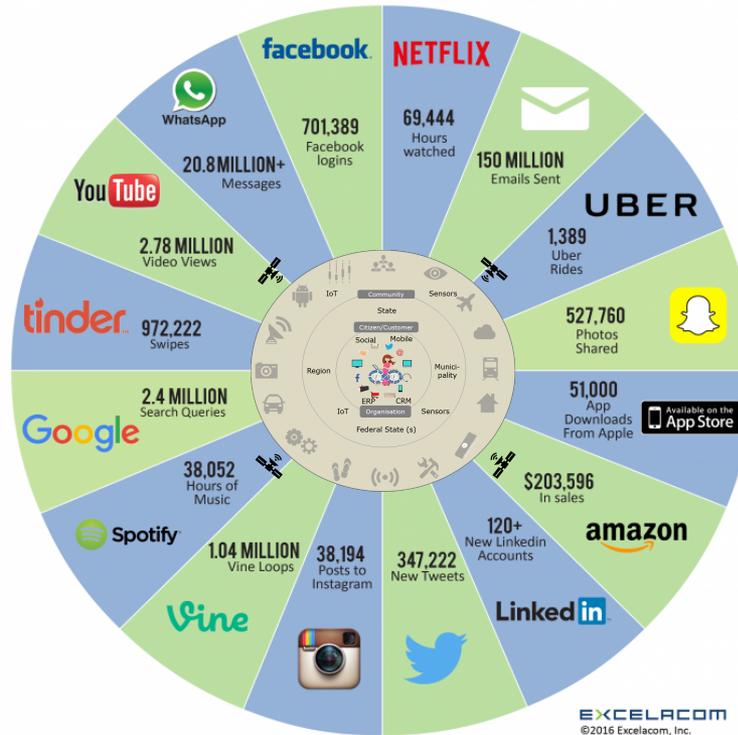
Volume

Angivet i exabyte (mla. gigabyte)



Byte	1	
Kilobyte (KB)	1.000	10(3)
Megabyte (MB)	1.000.000	10(6)
Gigabyte (GB)	1.000.000.000	10(9)
Terabyte (TB)	1.000.000.000.000	10(12)
Petabyte (PB)	1.000.000.000.000.000	10(15)
Exabyte (EB)	1.000.000.000.000.000.000	10(18)
Zettabyte (ZB)	1.000.000.000.000.000.000.000	10(21)
Yottabyte (YB)	1.000.000.000.000.000.000.000.000	10(24)

class	size	manage with	how it fits	examples
small	< 10 GB	Excel, R	fits in one machine's memory	thousands of sales figures
medium	10GB-1TB	indexed files, monolithic DB	fits on one machine's disk	millions of web pages
Big	> 1TB	Hadoop, distributed DBs	stored across many machines	billions of web clicks



EXCELACOM
© 2016 Excelacom, Inc.



6B Mobile phones worldwide



118B E-mails sent daily from a total of **3.4B** e-mail accounts; volume growing to **168B** in 2015



450B Business transactions / day over the Internet by 2020



1M Wireless sensors over 10 km² for Shell oil exploration project



2.3B Internet users worldwide, growing to **3.5B** in 2017



2.4B Social networking accounts, projected to grow to **3.9B** in 2015



\$155B Worldwide sensor market in 2011, forecasted to grow to **\$240B** in 2016



420M Wearable (wireless) health monitors by 2014 from 12M in 2012

Variety

- Refers to the different types of data to be used in the organization. In the past, they focused on structured data that neatly fits into tables or relational databases. Financial data (e.g. sales by product, region or/and customer) and transactional data (production, logistics and/or service) were typically used.
- These years is more than 80% of the world's data now unstructured and the share is still growing. This challenges the thinking of putting these into tables (think of photos, video sequences or social media updates), because it simply not possible.
- Using Big Data technology we can harness differed types of data (structured and unstructured) including messages, social media conversations, photos, sensor data, video or voice recordings and bring them together with more traditional, structured data. Again, it also demand a strategy for doing this and some clear goals.



Velocity

- Refers on one side to the speed at which new data is generated and on the other side to the speed at which data moves around.
- In addition, this might be fast and complex, as illustrated in the Infinity model (see page xx). Just think of social media messages going viral in seconds, the speed at which credit card transactions are checked for fraud, or the milliseconds it takes trading systems to analyze social media networks to pick up signals that trigger decisions to buy or sell shares.
- Big Data technology gives us the ability to analyze the data while it is generated, and this is without ever putting the data into databases. Again, without any clear goals for doing this out might ending up in nothing

Veracity

- Refers to the reliability or trustworthiness of the data.
- Big Data in its many forms of, quality and accuracy are less controllable (just think of Twitter posts with hash tags, abbreviations, typos and colloquial speech as well as the reliability and accuracy of content),
- The good news is that Big Data and the analytics technology allows us to work with these type of data.
- The volumes often make up for the lack of quality or accuracy and as the amount of data is growing overtime, the data can be considered as more accurate and reliable.

Virtuety

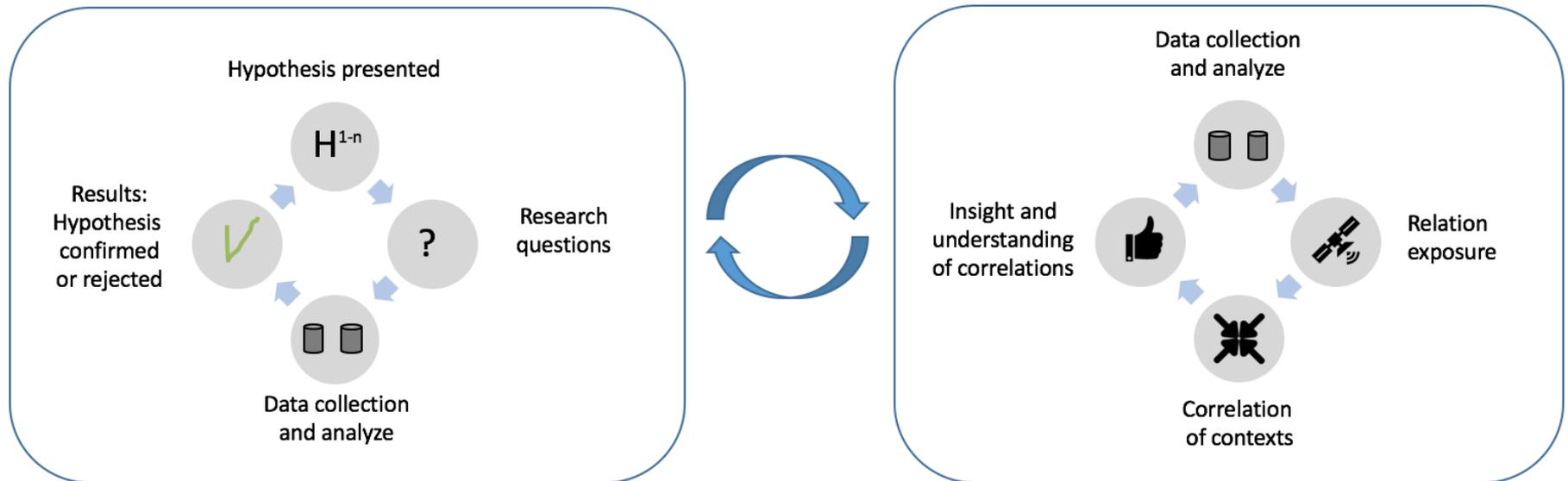
- No this is not a wrong spelling. . If we should continue the V model concept – we simply needed to create a new word – Virtuety.
- This refers to the discussion about the relevance of data in relation to our strategy about creating mutual value. This also address the need for data about customer's behavior in the Infinity model at any stage.
- What is needed to create the mutual value creation between customer/ citizen and business/community? This might really be a learning by doing exercise there never ends depending the customers behaviors and processes.

Visuality

- Refers to the relevance and importance for access to the data. How and to whom is the knowledge shared? How often and in what format.
- In organization with a clear customer oriented strategy it is mandatory that data is available in the interaction with customers and this should be based on customer behavior and processes.
- Dashboards provides a quick and relevant information that provides the basic knowledge for decision making available for all relevant person's in the organization. This again address some strategic issues like rights and authorization in organization.

New normal – New standards

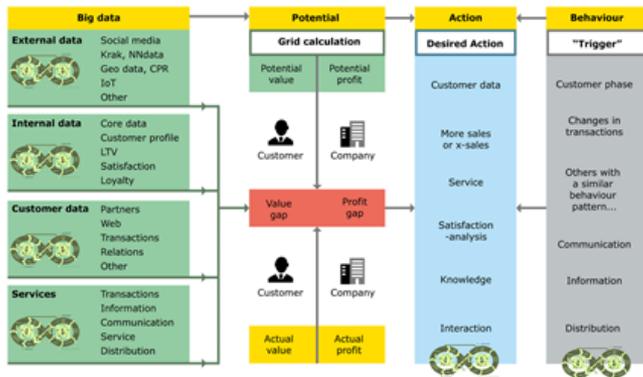
Focus is moved from the search of contexts to proactive to exposé contexts



Adapted from 'The Rise of Big Data' by Kenneth Cukier and Viktor Mayer-Schoenberger, an essay in 'Foreign Affairs' May/June 2013

Source: Professor Torsten Ringberg, CBS & Per Østergaard Jacobsen

DATA is key



Volume:

There is a lot of data, and currently most of it is not utilized (properly). We live in the age of terabytes (TBs), but the future will be measured in Exabyte's (EBs). Today's challenge or opportunity is that analytically use of data is poor, less than 5 pct. of available data is used. Making use of that volume is key.

Variety:

Data comes in many forms and shapes, and from a wide array of sources. Available as structured, unstructured, text, multimedia etc. 80 pct. of the data is not structured. In addition, 98% of the data is digital. To gain a grasp of the variety offers a great opportunity.

Velocity:

Data is in motion! Streaming data from mobile phones, others devices and IoT gives challenges in collecting data from different sources, where you may have milliseconds/ seconds to give response. Globally more than 30 billion RFID sensors provides information. This challenges API integration and Open Data concepts to capture the value of these data.

Veracity:

There is a high uncertainty about the data. Lack of consistency of data formats, ambiguities, latency, from many different sources, validity and security. The result? One in three managers²⁴ doubt the data of which the decisions are based.

"Virtuety":

How relevant is the data. What data do we need to create the mutual value creation between customer & business?

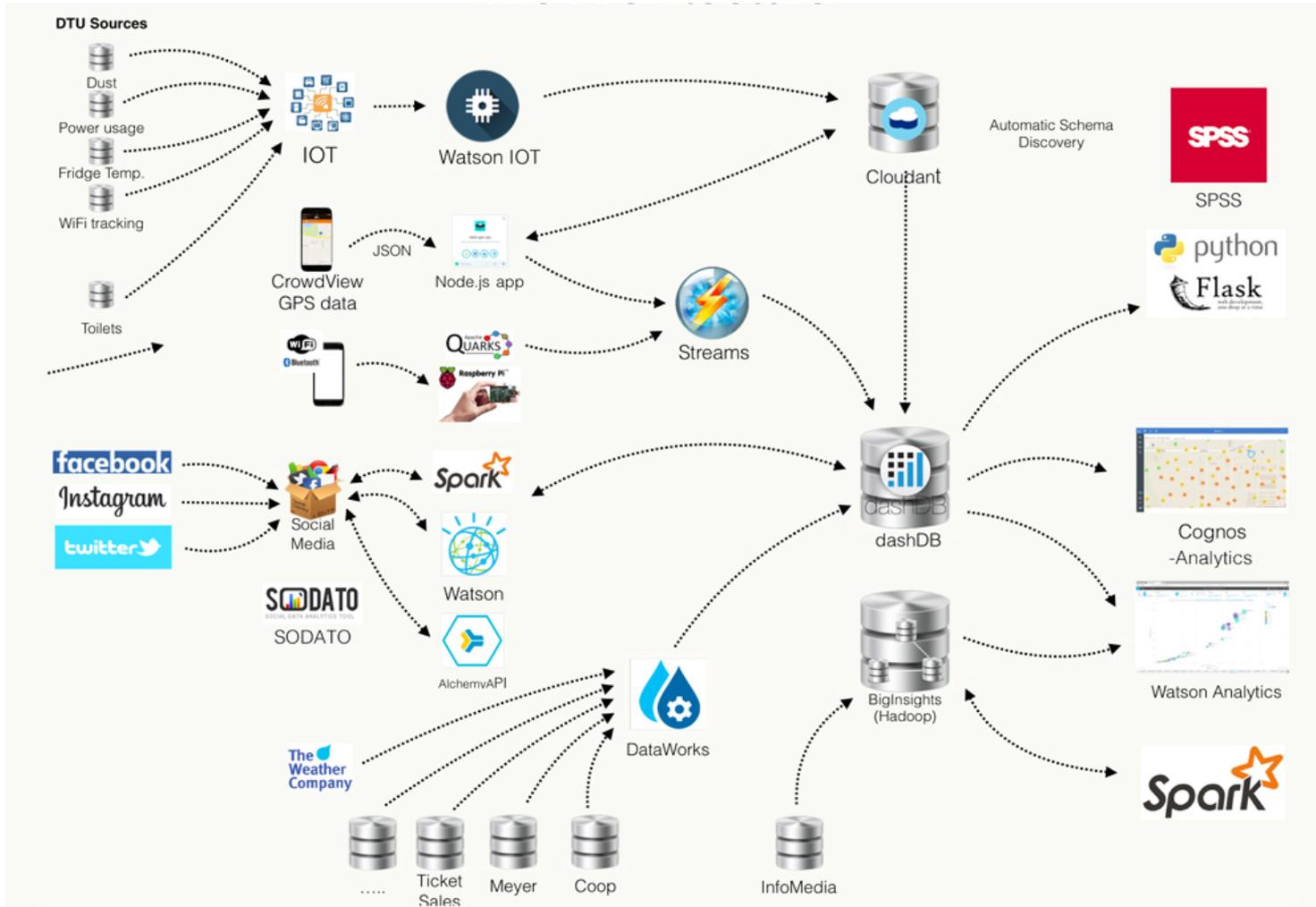
"Visuality":

How and to whom is the knowledge shared? How often and in what format? Dashboards for all relevant touchpoints provides a quick and relevant information that provides the basic knowledge for decision making.

Potential or nuisance?

Only 0.5% of the 2.5 billion GB of data generated every day will be used for **analytical** purposes?

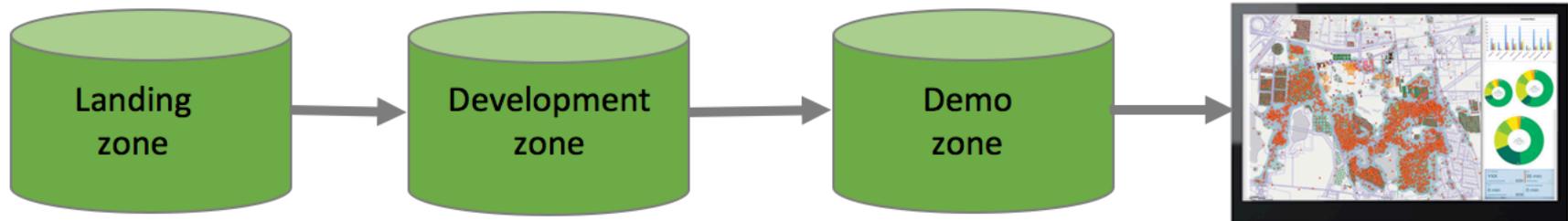
Example – BIG DATA Architecture



FRA RIO
TIL ROSKILDE
TUR RETUR

From Research Project
"Rio til Roskilde" - CBS 2016

Example – Governance structure



FRA RIO
TIL ROSKILDE
TUR RETUR

From Research Project
"Rio til Roskilde" - CBS 2016

Working with Big Data – Do not forget:

- Morale**
- Ethics**
- The General Data Protection Regulation (GDPR) 2018**



That's why a strategy is the right starting point - This is not an IT issues at the beginning

Be open to data before you are open for data!

OO + NT = EOO

Hvordan kan vi hjælpe dig med analyser?

KONTAKT OS FOR EN UFORPLIGTENDE DRØFTELSE AF MULIGHEDERNE

PJA@EFFICIENS.NU
+45 2562 4898