We are building the next generation of
Big Data and Analytics solutions!

(C) Copyright Cloud Explorers Solutions AS 2016

cloud explorers

Background

26 years Experience IT Industry

- 12 Years Solutions Architect - International
- 10 Years - IT Director - Private Banking OSLO
- 4 Years - CEO - Cloud Explorers - Big Data International

Stephen Karl Ranson
CEO

Profile

Passionate about Technology

Genuine Interest In All Things Digital

Resourceful

Innovative

Out of the Box Thinker

Disruption

I ❤ DATA SCIENCE

(C) Copyright Cloud Explorers Solutions AS 2016

Introduction
“DATA”

(C) Copyright Cloud Explorers Solutions AS 2016

DATA Timeline - 72 Years of Data Growth

1944 - Books - 20m - 40 buildings 2040

The acknowledgement of **big data** was identified by Fremont Rider, Wesleyan Univeristy Librarian, who estimated that libraries in America universities would **expand to over 20,000,000** volumes by 2040 Today, Yale Library alone has approximately 12.5 million volumes across 20 buildings on campus.

1961 - Knowledge - 15yr x2

Derek Price´s research on scientific knowledge concluded that scientific journals had **doubled every 15 years**. This is now better known as "**law of exponential increase**".

1975 - Words - Unit of Measure

The Ministry of Posts and Telecommunications in Japan began conducting the **Information Flow Census**, tracking the volume of information circulating in Japan. The census introduces "**amount of words**" as the unifying unit of measure across all media.

1983

Author Ithiel de Sola Pool looked at the growth trends in 17 major communications media from 1969 to 1977, and concluded that the **flow of information had exponentially grown by 2.9%** throughout that period due to broadcasting and media.

2001 - 3 V´s of Data

Doug Lanely published a research note titled "3D Data Management: Controlling Data **Volume, Velocity, and Variety.**" A decade later, these **"3 V's"** became the defining dimensions of big data.

1997 - Big Data - Phrase Coined

The term "big data" was used for the first time when researchers M.Cox and D. Ellsworth wrote an article identifying that the **rise of data** was becoming an issue for current computer systems. In other, words, the problem of **"big data"**.

2008 - Exaflood - US IP Traffic 2015

Economists Bret Swanson and writer George Gilder published "**Estimating the Exaflood**" which stated U.S. IP traffic could reach one zettabyte by 2015, and that the U.S. Internet of 2015 would be at least **50 times** larger than it was in 2006.

2006

Amazon Launched AWS, To Become Worlds Largest IT Cloud Actor.
Hadoop Created at Yahoo!

2011 - 25% Growth pyr - 94% digital

Martin Hilbert and Priscila Lopez, estimated that the **world´s information storage capacity grew** at a compound annual growth rate of **25% per year** between 1986 and 2007. They also estimated that in **1986, 99.2%** of all storage capacity was **analog**, but by **2007, 94%** of all storage capacity was **digital**.

2009 - 12 hours - 34 GB

Researchers Roger E. Bohn and James E. Short found that modern Americans **consumed information** for an average of almost **12 hours** per day. Consumption totalled 3.6 zettabytes and 10,845 tillion words, corresponding to 100,500 words and **34 gigabytes for an average person on an average day**.

2016

Petabyte pr. day!

1940´s | **1950´s** | **1960´s** | **1970´s** | **1980´s** | **1990´s** | **2000** | **2010**

1876 - Recording Data - Systematic

Statistisk sentralbyrå
Statistics Norway

1949 - Reaserch Storage Capacity - Analog

Claude Shannon, known as the "**Father of Information**", carried out research on big storage capacity on items such as punch cards and photographic data. One of the largest items on Shannon's list was the Library of Congress, measuring over 100 trillion bits of data.

1971 - Bits - Measure of Privacy

Arthur R.Miller, author of the book '**Assault on Privacy,**' identified that "too many information handlers seem to **measure a man by the number of bits** of storage capacity his dossier will occur."

1981 - Enterprise Data - Growth in bits

The Hungarian Central **Statistics Office** carried out a research project that is still on-going today. This involved accounting for the country's information for industries via measuring data volumes in bits.

1994
Founded

1995 - Org Nr - Norway

Org. Nr. Introduced
Registered as AS

Statistisk sentralbyrå
Statistics Norway

1996 - Cost Effective - Digital

Digital storage became more **cost-effective** for storing data than paper.

1998
Founded

2000 - 1.5 exabytes - New Data

Peter Lyman and Hal R. Varian published the first study that quantified, in **computer storage** terms, the total amount of new and original information created in the world annually. The study concluded in 1999, a year in which the world had produced approximately **1.5 exabytes of unique information**.

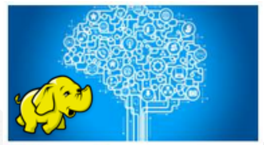
2004
Founded

2005
Doug Cutting
"Nutch"
Google
Whitepapers

2005 - The New Web - Infoware

"**What is Web 2.0**" was published by writer Tim O'Reilly, in which he started that "data is the next intel inside and SQL is the new HTML. Database management is a core competency of Web 2.0 companies, so much so that they have sometimes referred to these applications as "infoware' rather than merely software.

2007
Steve Jobs
Smartphone
Revolution



2013/Ongoing - Big Data Age Begin

Businesses are beginning to implement **Big Data** to **analyze** and optimize mass quantities of data. Deriving **unprecedented** value and **advantage**.

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is “Big Data”

Byte : one grain of rice

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is “Big Data”

Byte : one grain of rice

Kilobyte : cup of rice

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is “Big Data”

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Megabyte

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is “Big Data”

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi Trucks of rice

Desktop

Gigabyte

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is “Big Data”

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi Trucks of rice

Desktop

Terabyte : 2 Container Ships of rice

Server

Terabyte

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is “Big Data”

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi Trucks of rice

Desktop

Terabyte : 2 Container Ships of rice

Server

Petabyte : Blankets Manhattan

Hadoop

Petabyte

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is "Big Data"

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi Trucks of rice

Desktop

Terabyte : 2 Container Ships of rice

Server

Petabyte : Blankets Manhattan

Hadoop

Exabyte : Pacific Ocean

Multiple
Hadoop Clusters

Exabyte

(C) Copyright Cloud Explorers Solutions AS 2016

Understanding When DATA is "Big Data"

Byte : one grain of rice

Kilobyte : cup of rice

Megabyte : 8 bags of rice

Gigabyte : 3 Semi Trucks of rice

Desktop

Terabyte : 2 Container Ships of rice

Server

Petabyte : Blankets Manhattan

Hadoop

Exabyte : Pacific Ocean

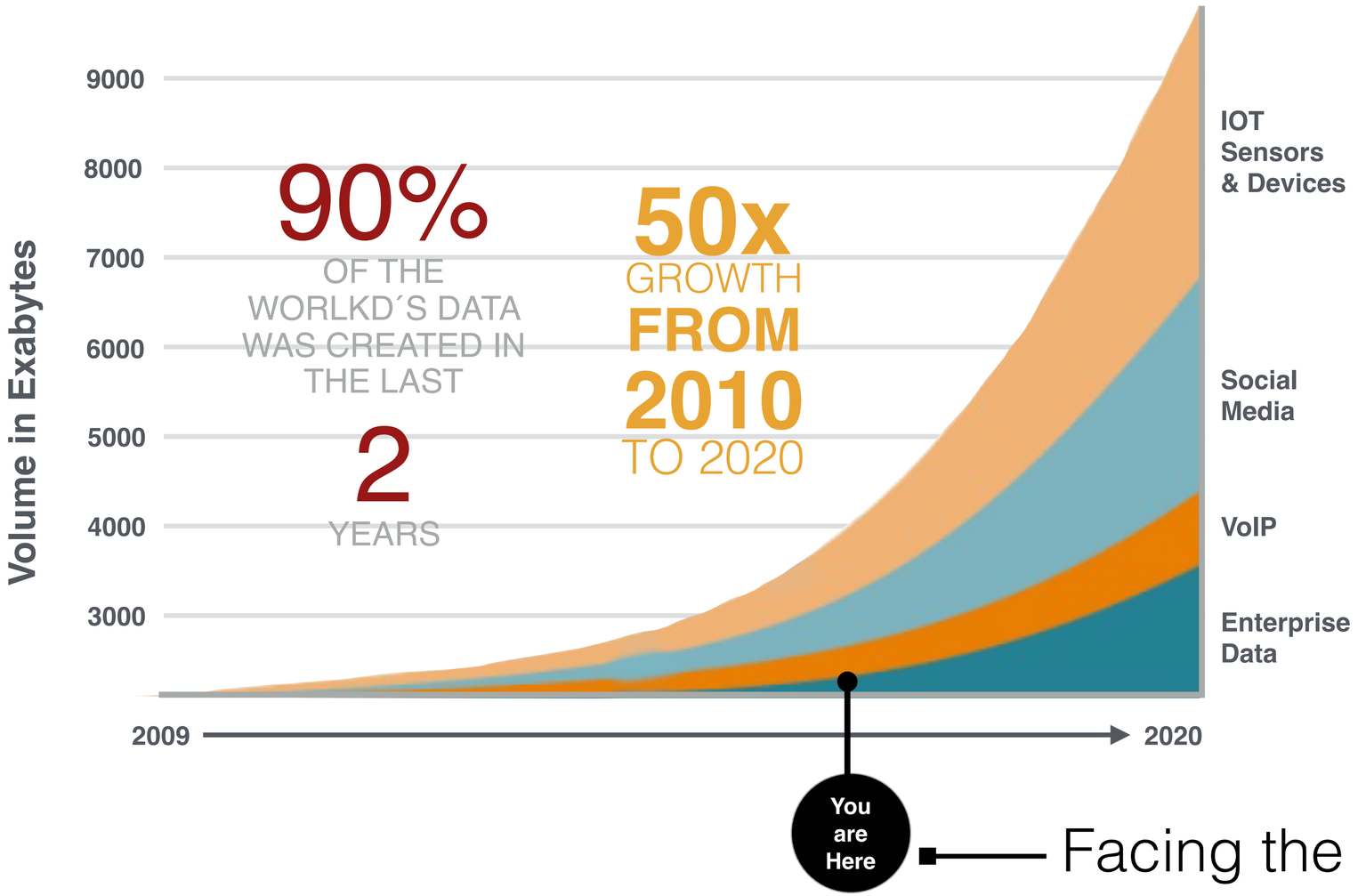
Yottabyte : Earth Size Ball of rice

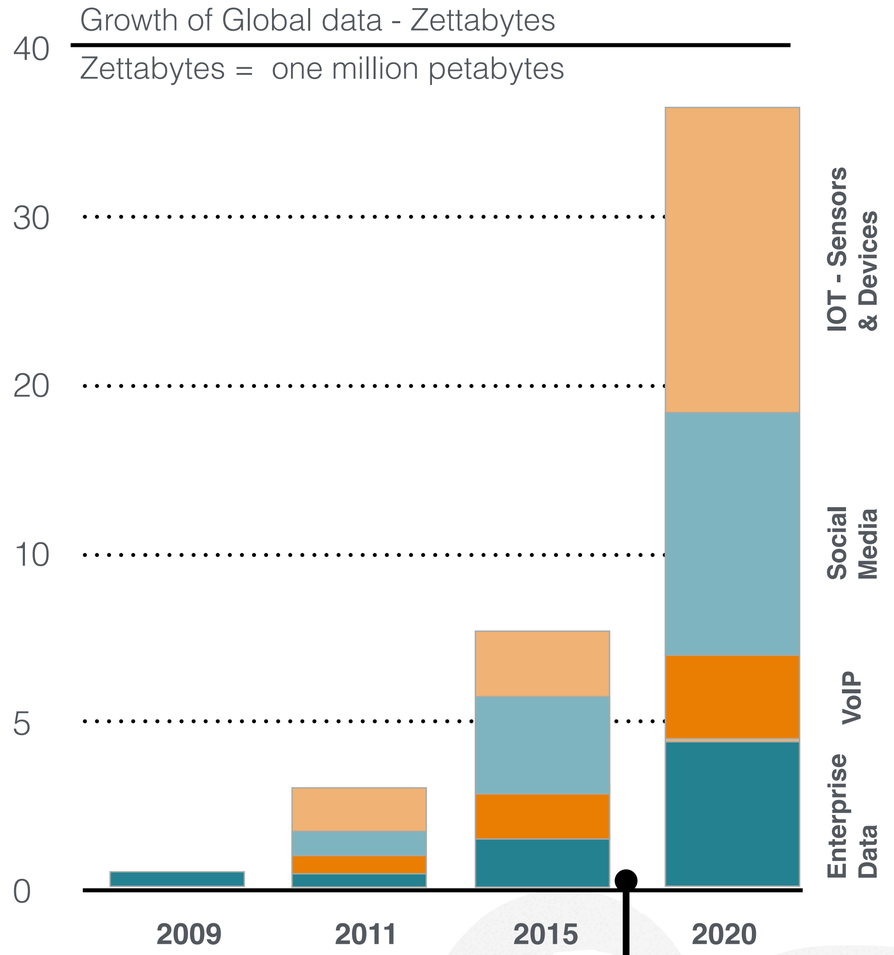
Multiple
Hadoop Clusters

Yottabyte

(C) Copyright Cloud Explorers Solutions AS 2016

DATA is Growing Fast!





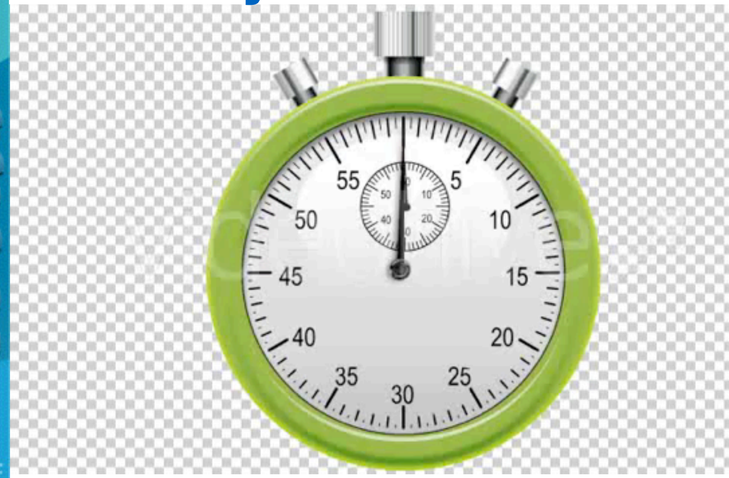
Facing the “Data Tsunami”

(C) Copyright Cloud Explorers Solutions AS 2016

cloud explorers

DATA Growth in 60s

Every 60 seconds



- **98,000+** tweets
- **370,000+** Skype Voice Minutes
- **13,000+** Apps Downloaded
- **100+** new LinkedIN Accounts
- **600+** new YouTube Vids
- **25hrs** New Video
- **695,000+** status updates
- **168 million+** emails sent
- **79,365** wall posts
- **694,445** Google searches
- **1,500+** Blog Posts
- **6,600+** New Pictures Flickr
- **320+** New Twitter Accounts
- **13,000+** Hours Music Pandora

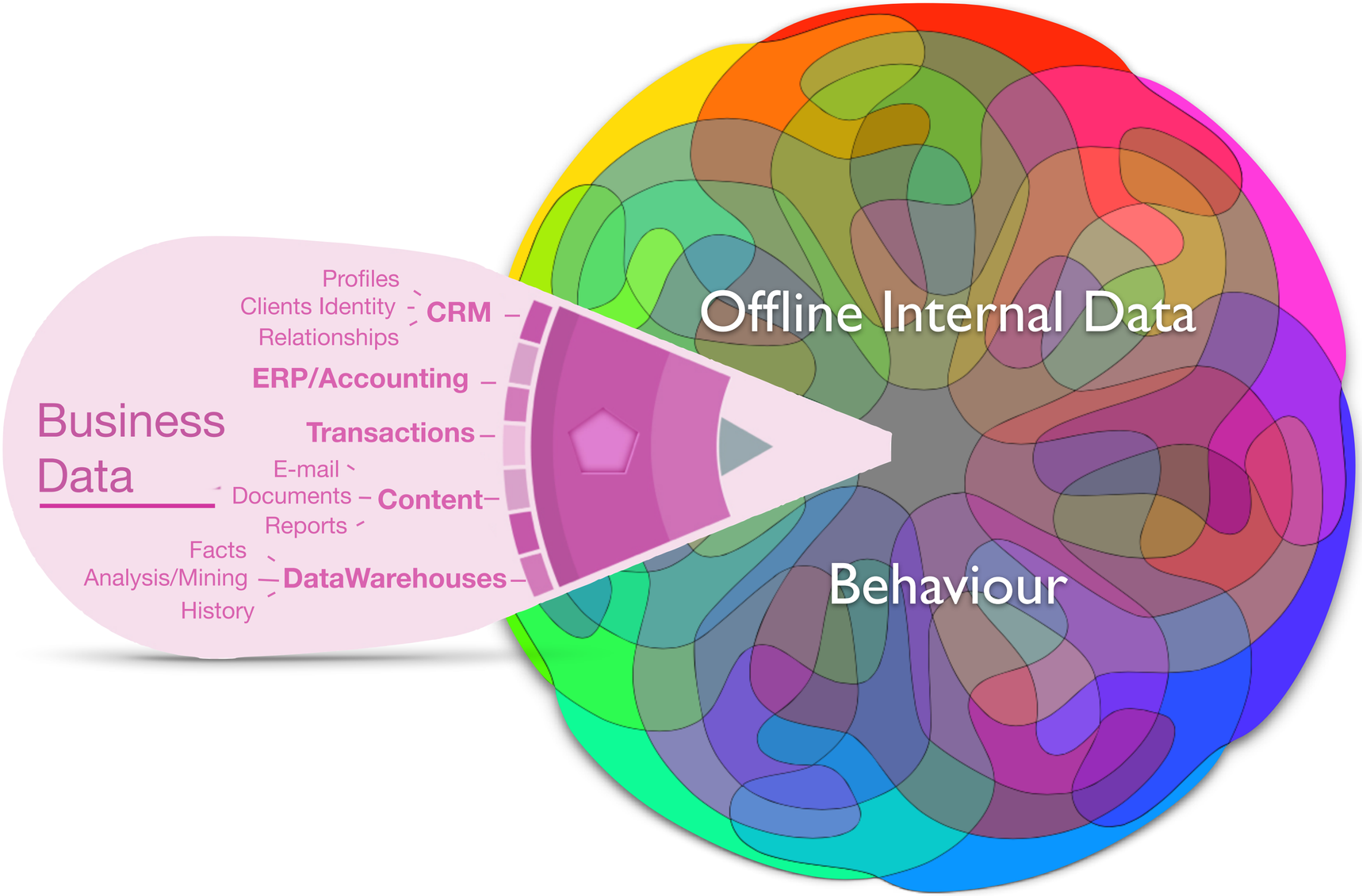
1,820 TB of data created

(C) Copyright Cloud Explorers Solutions AS 2016

Gathering the Data
“The DATALAKE”

(C) Copyright Cloud Explorers Solutions AS 2016

Big Data - The Data Lake



Enterprise Business Data

(C) Copyright Cloud Explorers Solutions AS 2016

Free & Diverse Data (Offline External Data)

Statistisk sentralbyrå
Statistics Norway

World Health Organization

unicef
70 YEARS FOR EVERY CHILD

datacatalogs.org

Meteorologisk institutt
150 år

FDF Financial Data Finder

Google

GAPMINDER
a fact-based worldview

WIKIPEDIA

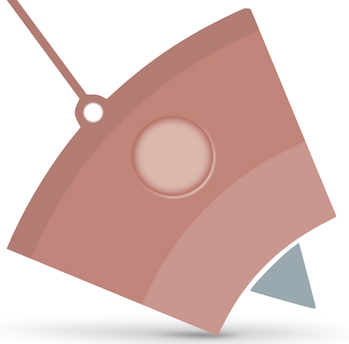
Crowdsourced
Cool
Geodata
Social
Interesting
Government
Environments
Society
Infrastructure
Events
Fun
Finance
Business
News
Health

OPEN Data

NOAA

csc infochimps
A CSC BIG DATA BUSINESS

TOPSY



DATA.GOV.UK
Opening up Government

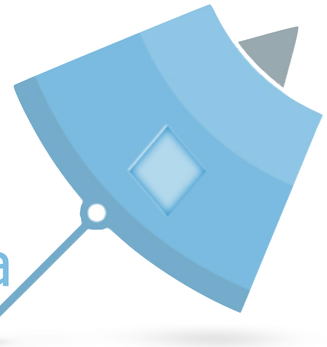
DATA.GOV

Freebase

CENTRAL INTELLIGENCE AGENCY

HealthData.gov

The New York Times



GEOGRAPHIC Data

facebook

amazon web services™

DBpedia

(C) Copyright Cloud Explorers Solutions AS 2016

Quality Data (Offline & Online External Data)

TNS

Bisnode

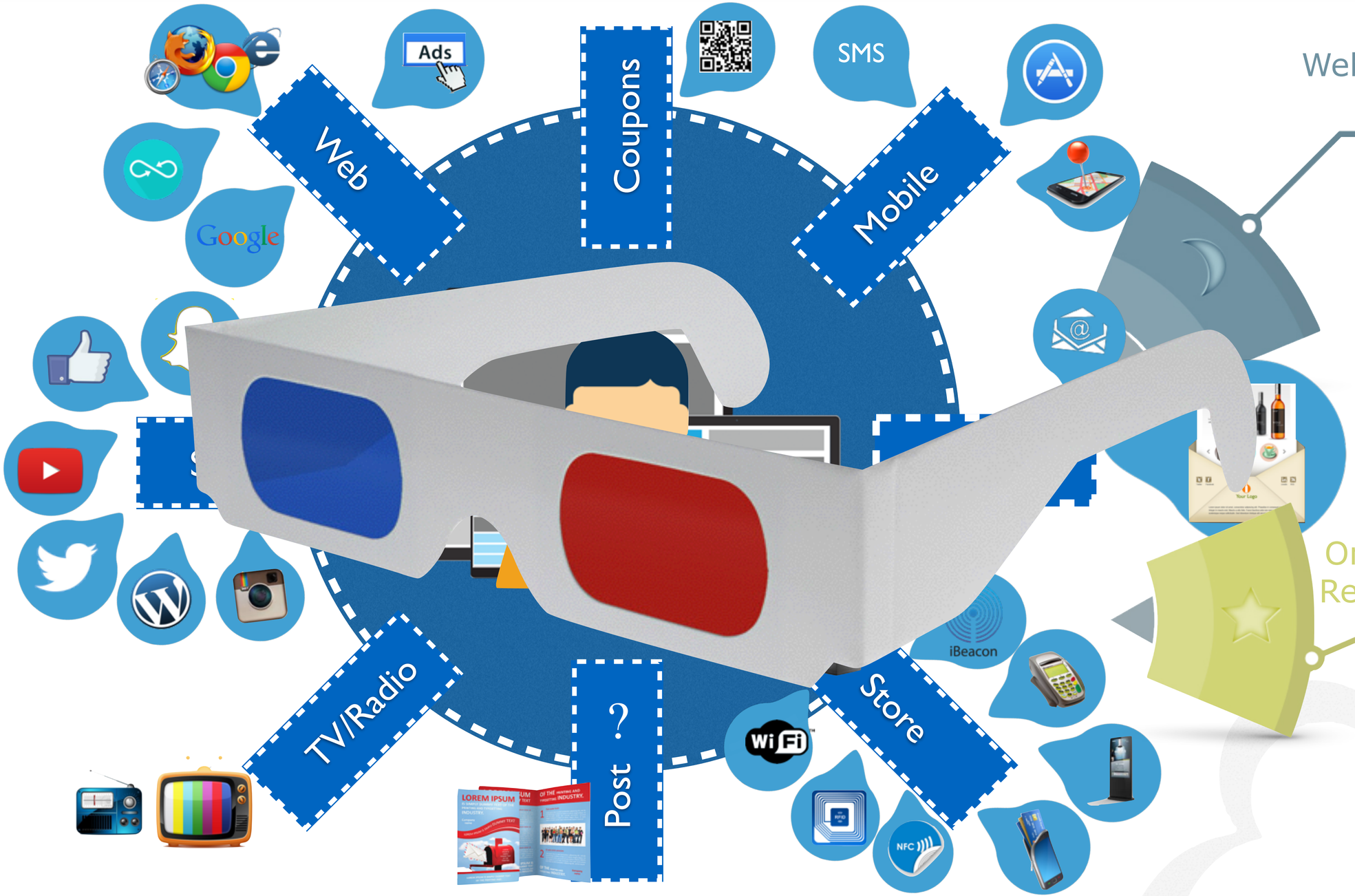
Commercial Data

(C) Copyright Cloud Explorers Solutions AS 2016

Market and Customer Contact Points (Online Behavior)

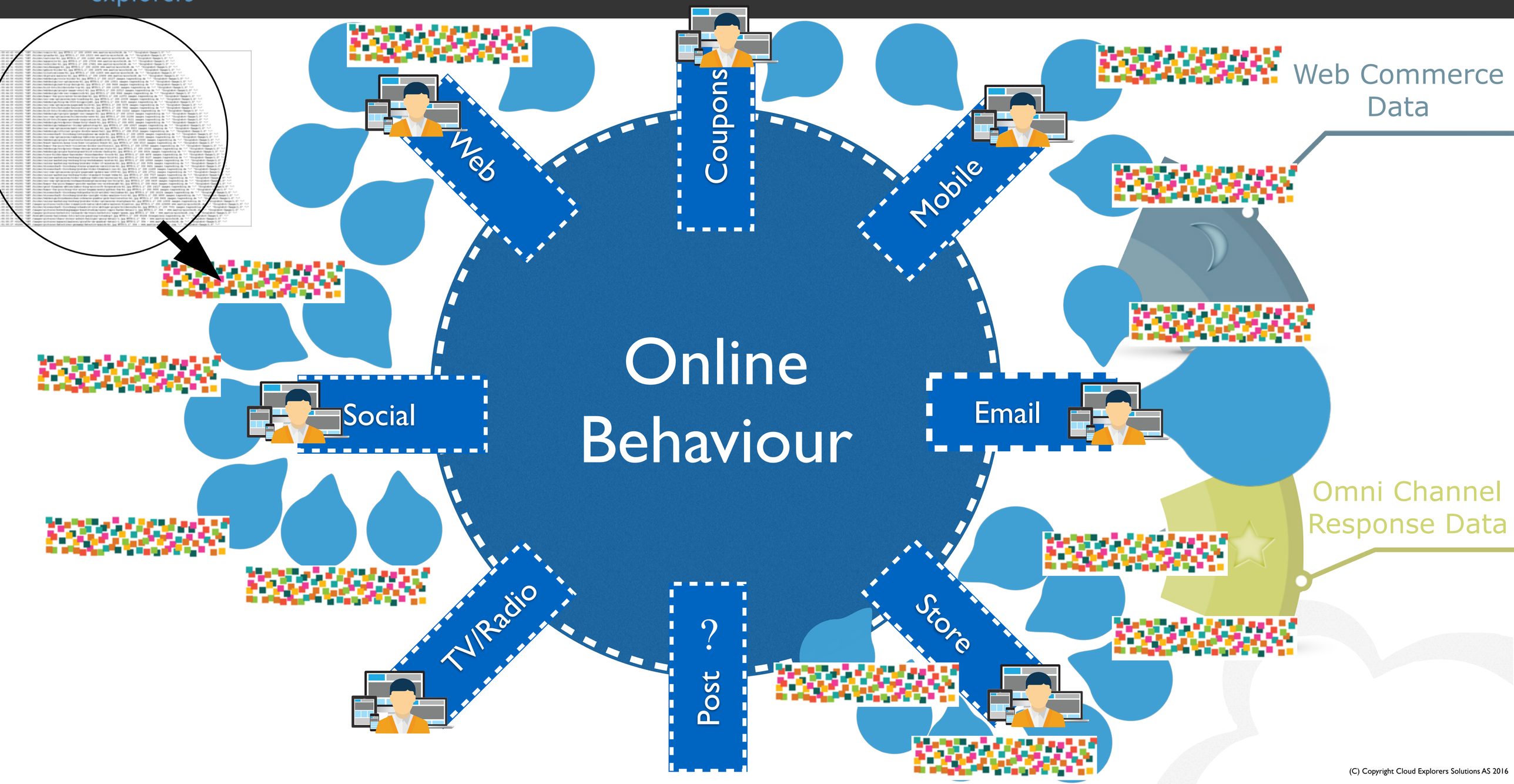
(C) Copyright Cloud Explorers Solutions AS 2016

Market and Customer Contact Points (Online Behavior)



(C) Copyright Cloud Explorers Solutions AS 2016

Critical Relationship DATA!



(C) Copyright Cloud Explorers Solutions AS 2016

Big Data "DATALAKE"

Professional Data

TNS, **Bisnode**, Quality Data Brokers,
High quality lists, Survey, Panel Data,
Clickstream, Directories, Telephone numbers,
Board and Company roles....

Open Data

Free Public Data
Environment, Infrastructure, Finance,
Health, Education, Reference, Society

Web Commerce

Existing Web Commerce Sites
Logs, Trackers, Purchases,
Abandoned Carts, Interest, Usage....

Bring Your Own Data

- Clients Identity Relationships Profiles – CRM
- Economy
- Communication
- Transactions/ERP/POS
- DataWarehouses

Omni Channel Response Data

Communication Data from all Channels
Web Logs, Campaign, Member, Mail Logs, Telefoni
Creates Red Thread across all channels...

Geographic Data

Mapping, Location, Geo spatial Data,
Property, size of house,
POI nearest shops, cafes.
GEO Data, Nordica…

Social

Social Data, FB, Twitter, Instagram,
Snapchat, LinkedIN, Google+, Blogs,
NLP Sentiment Analysis, Networks,
Likes, Interests, Trends, Discussions,
Product Awareness and Feedback

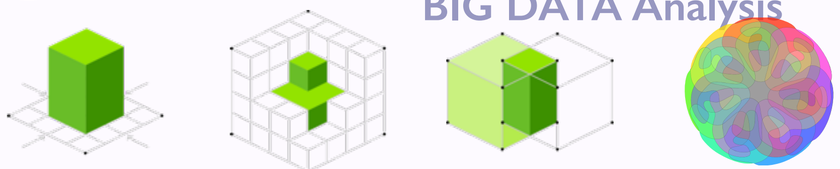
Real Time IOT - "STORM"

IOT, iBeacon, Wifi, Web Sessions, Sensors
POS, RFID, Houses, Cars, Payment Terminals,
Temperature, Weather, Health,
Traffic, News Events, Trends….

EXTRACT

1 **Import**
Wash / Enrich

2 **BIG DATA Analysis**



SINGLE VIEW OF ENTITY | DATA DISCOVERY | PREDICTIVE ANALYTICS | ADVANCED SEGMENTATION

3 **Operationalize**
Publish
Integrate / Orchestrate

ACTIONS

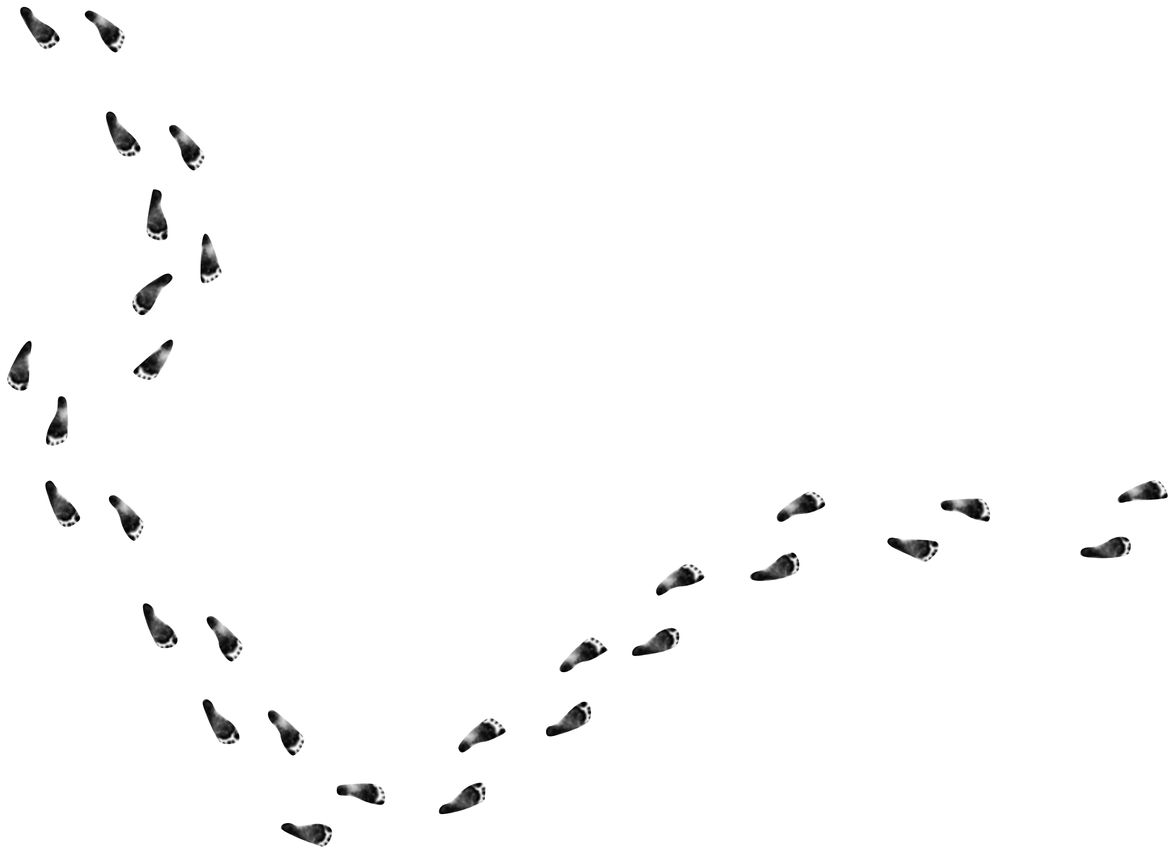
(C) Copyright Cloud Explorers Solutions AS 2016

Turning Data into Value
“DATA Science”
-or-
“The Digital Crime Scene”

(C) Copyright Cloud Explorers Solutions AS 2016

AN ORIGINAL CRIME STORY……

Hansel & Gretel



(C) Copyright Cloud Explorers Solutions AS 2016

AN ORIGINAL CRIME STORY……

Hansel & Gretel

(C) Copyright Cloud Explorers Solutions AS 2016

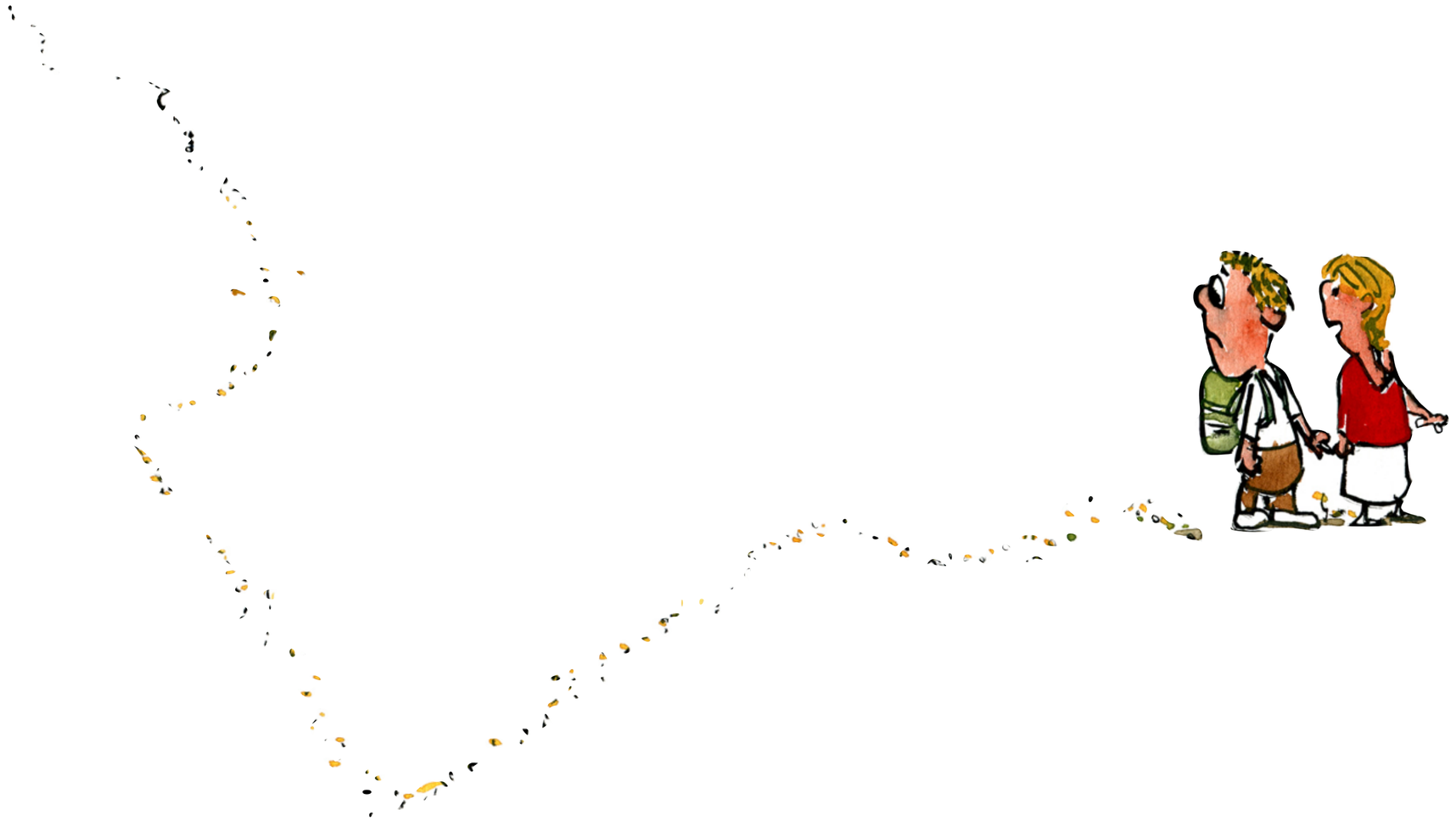
AN ORIGINAL CRIME STORY……

Hansel & Gretel Creators of “Cookie” tracking

(C) Copyright Cloud Explorers Solutions AS 2016

AN ORIGINAL CRIME STORY……

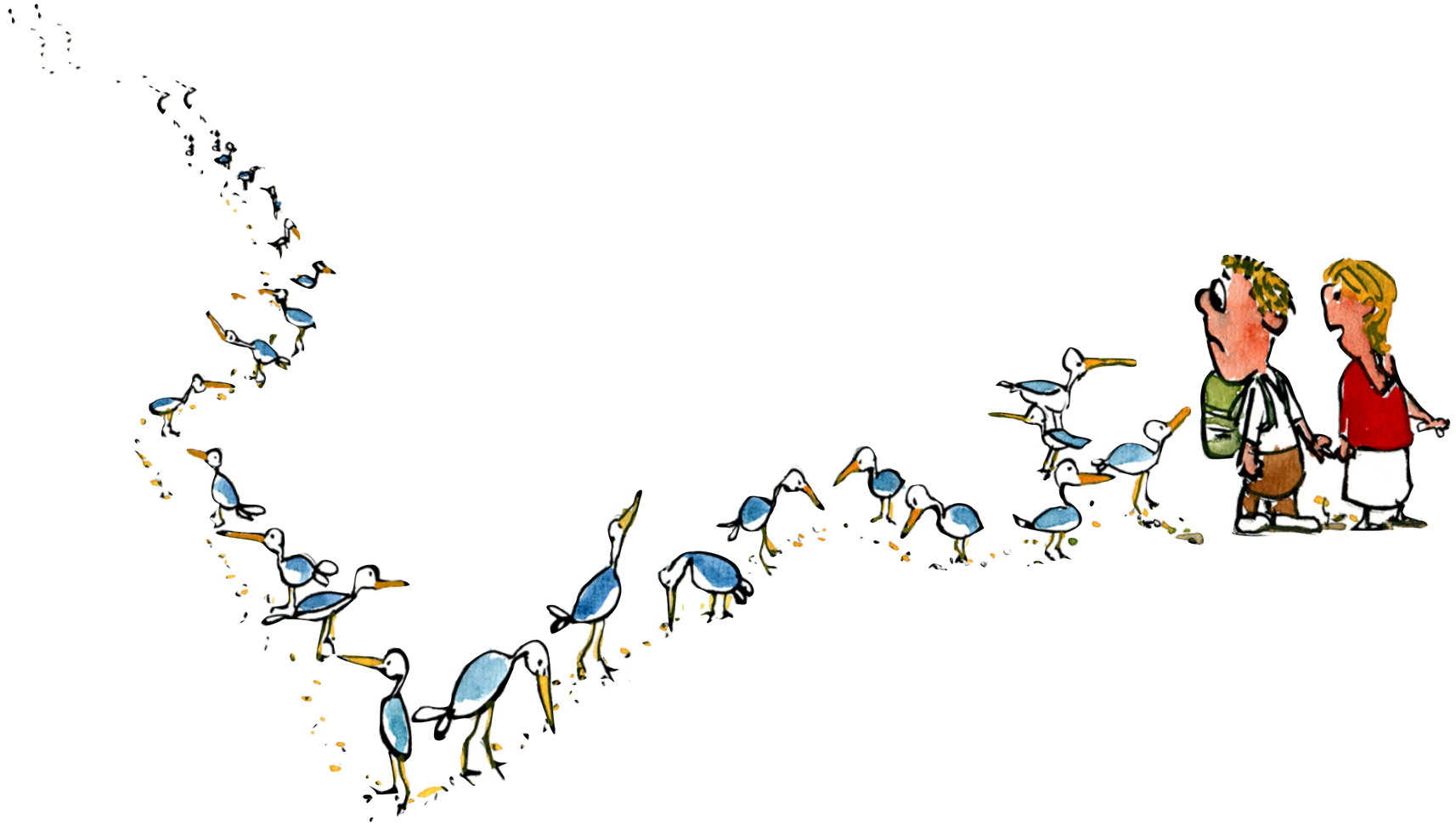
Hansel & Gretel Creators of “Cookie” tracking



(C) Copyright Cloud Explorers Solutions AS 2016

AN ORIGINAL CRIME STORY……

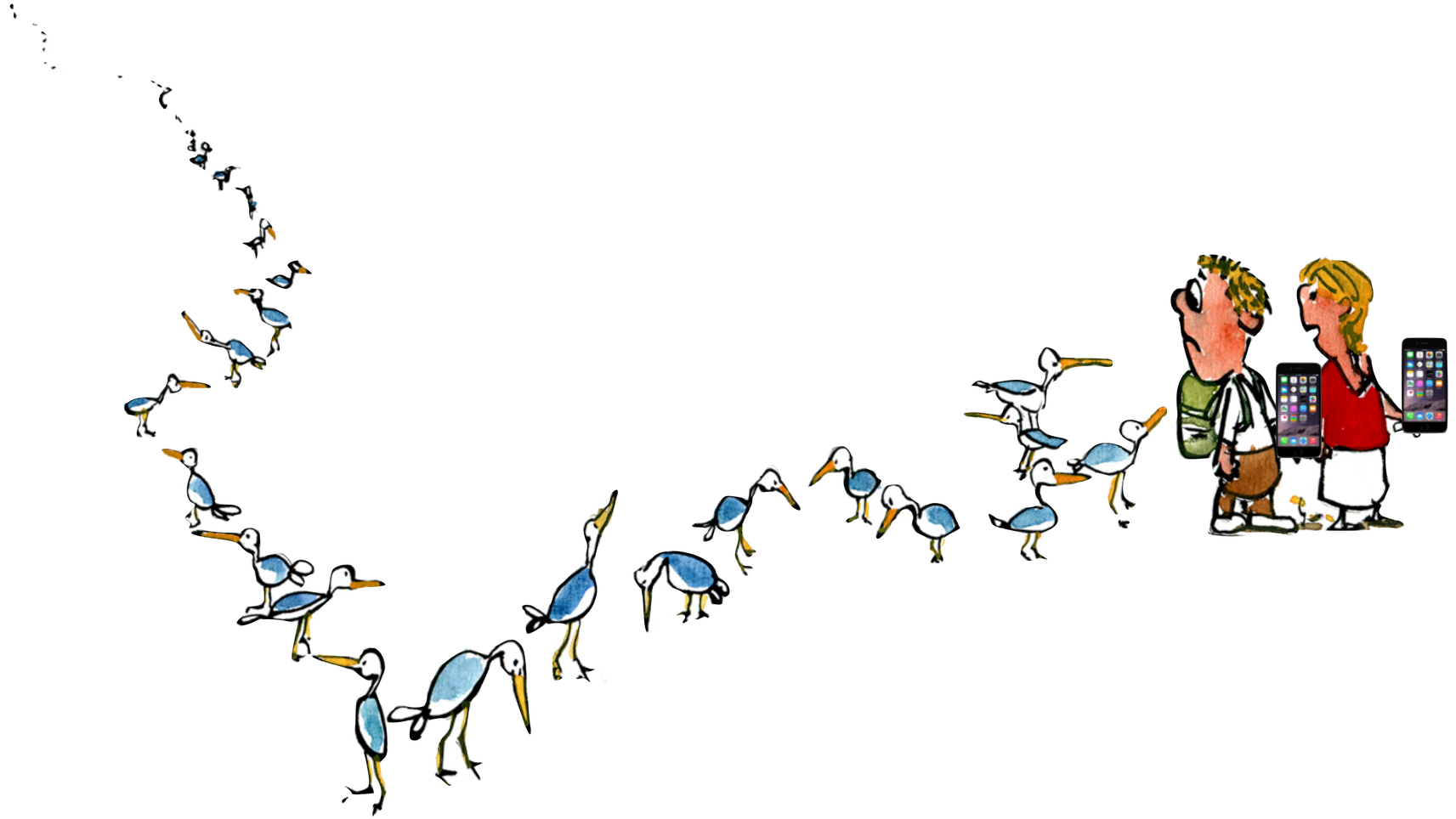
Hansel & Gretel Creators of “Cookie” tracking



(C) Copyright Cloud Explorers Solutions AS 2016

UPDATING THE STORY……

Hansel & Gretel 2.0 and NO “Cookie” tracking



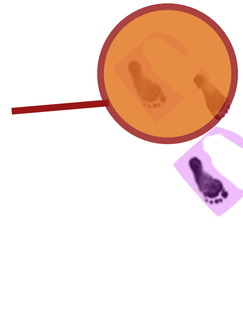
(C) Copyright Cloud Explorers Solutions AS 2016

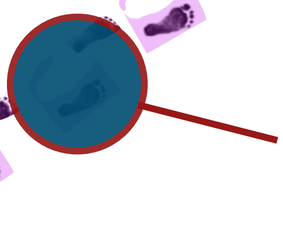
UPDATING THE STORY……

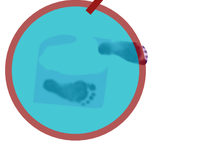
Hansel & Gretel 2.0 and NO “Cookie” tracking

GEO: 59.927262
10.961176

Hansel Checked In
Scary Forest

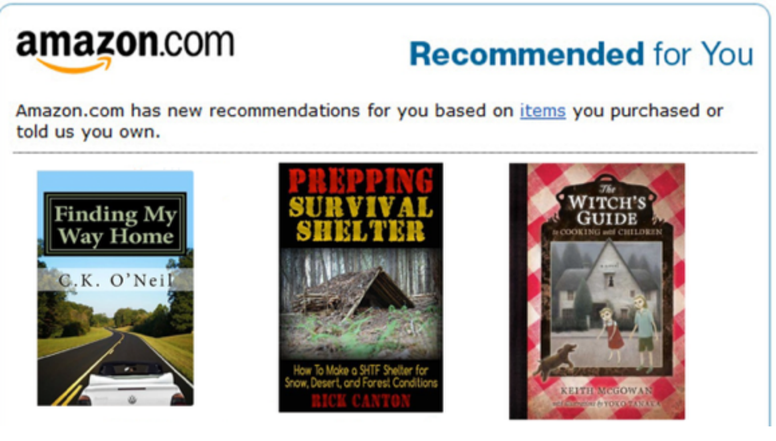






Google
Norway

Are Gingerbread Houses Good To Eat



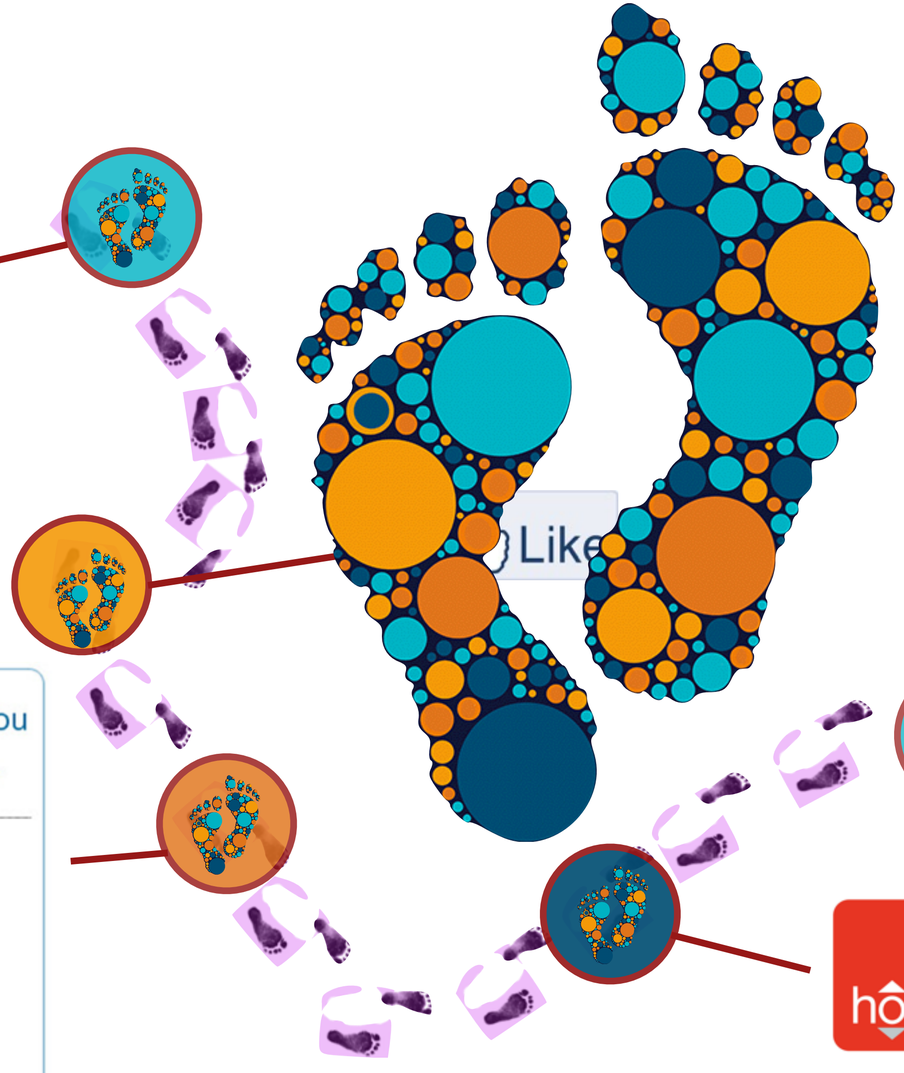
(C) Copyright Cloud Explorers Solutions AS 2016

UPDATING THE STORY……

Hansel & Gretel 2.0 and NO “Cookie” tracking

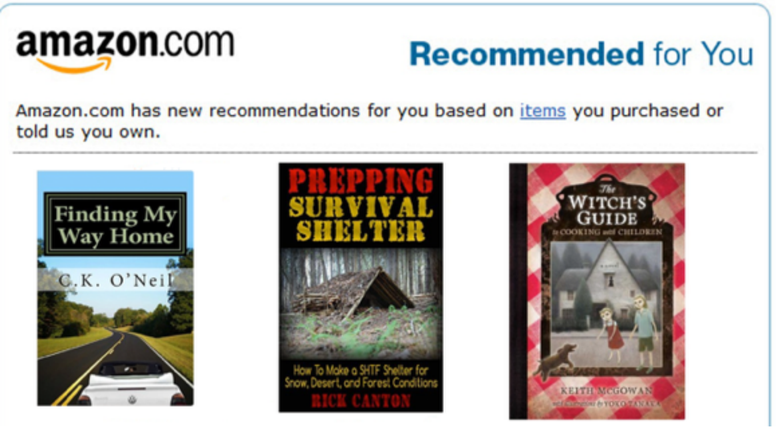
GEO: 59.927262
10.961176

Hansel Checked In
Scary Forest



Google
Norway

Are Gingerbread Houses Good To Eat



hotels.com

(C) Copyright Cloud Explorers Solutions AS 2016

UPDATING THE STORY……

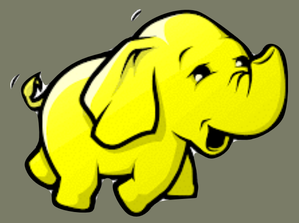
Hansel & Gretel 2.0 and NO “Cookie” tracking

“The Digital Crime Scene”

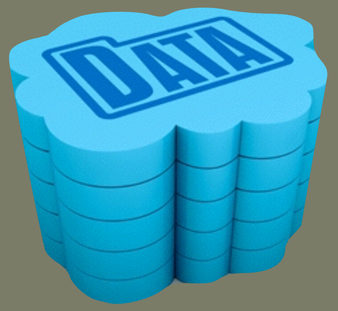
(C) Copyright Cloud Explorers Solutions AS 2016

MARKETING WANTS TO SOLVE THE MYSTERY

MODERN DATA SHERLOCK



BIG DATA



MATH & STATISTICS

- Machine learning
- Statistical modelling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Database SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any visualization tools e.g. Flare, D3.js, Tableau

(C) Copyright Cloud Explorers Solutions AS 2016

DATA SCIENCE - The Different Types Of Analysis

fuzzy boundaries

rule-based decision making

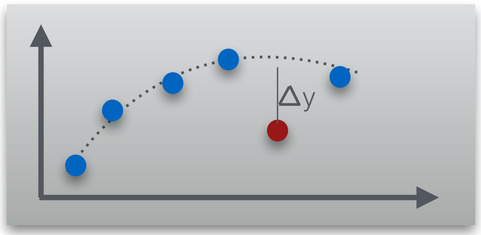
if condition fulfilled ***then***
activity 1
else
activity 2

boolean data
(yes or no)

Examples:
- phone notifiction
- time- or threshold-based alarms
- simple pattern
- matching

every programmer

statistical reasoning

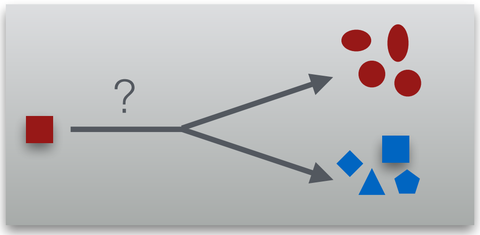


simple regression

numerical data
allowing for
curve fitting

Examples:
- extra- and interpolation
- outlier detection
- predictive maintenance

machine learning



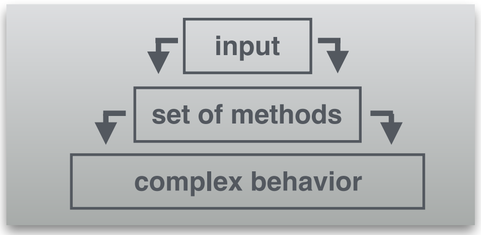
classification tasks

arbitrary data
that needs to be
abstracted into
numbers

Examples:
- identification of relevant features from large input datasets
- quality control using various metrics

data science types

artificial intelligence



dynamic adaption
to novelty

autonomous selection
of best methodology
when presented with
arbitrary data

Examples:
- autonomous vehicles
- human-like conversational skills
- intelligent digital assistant

complex systems specialists

(C) Copyright Cloud Explorers Solutions AS 2016

Data Science - The Future Of Marketing

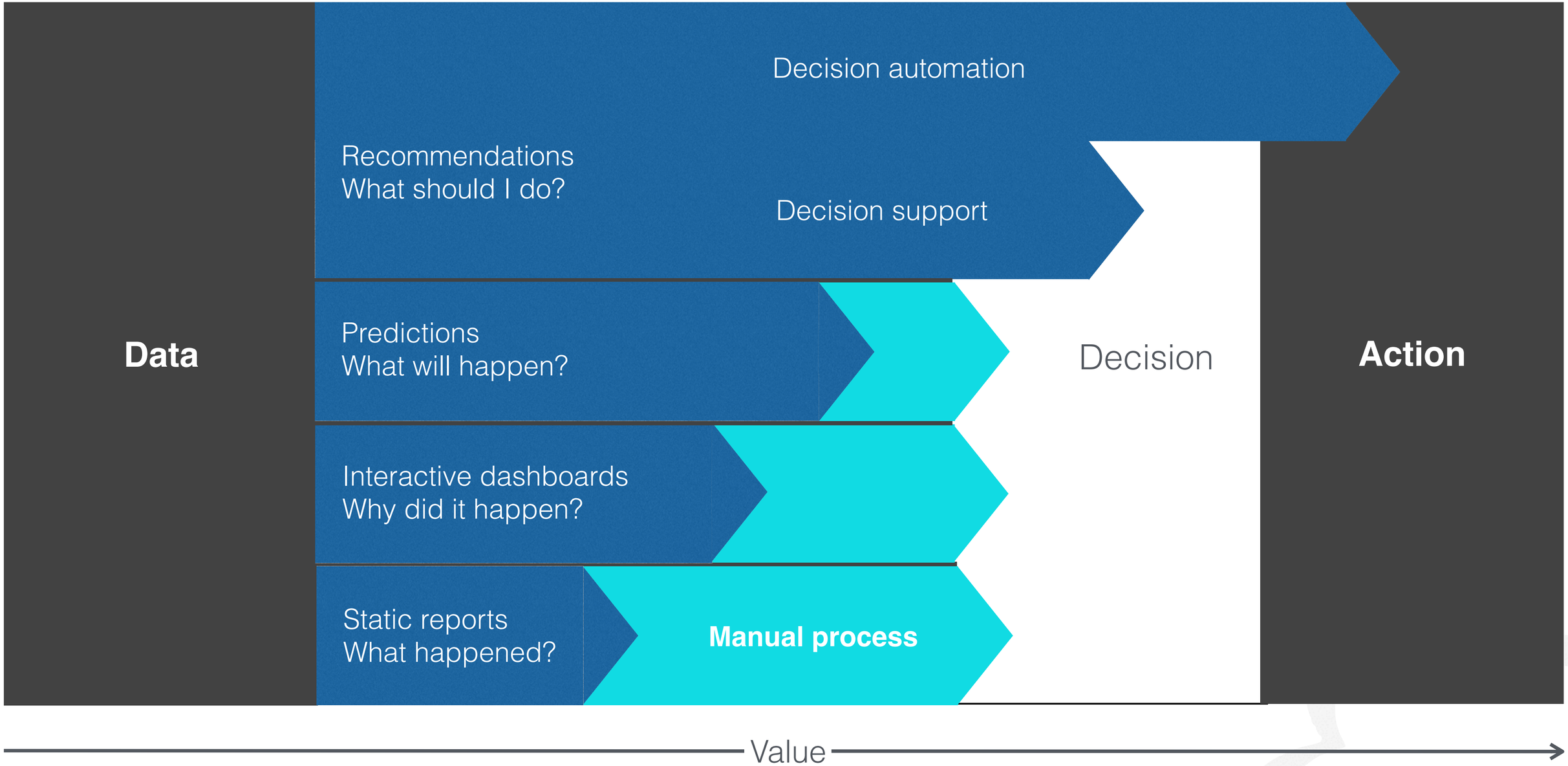
Marketing Teams Will Soon Include “Engagement Scientists”

Move over, data scientists, Columnist Paul Ford explains why engagement science is set to redefine the way marketers do business.

marketing expert + data scientist = engagement scientist

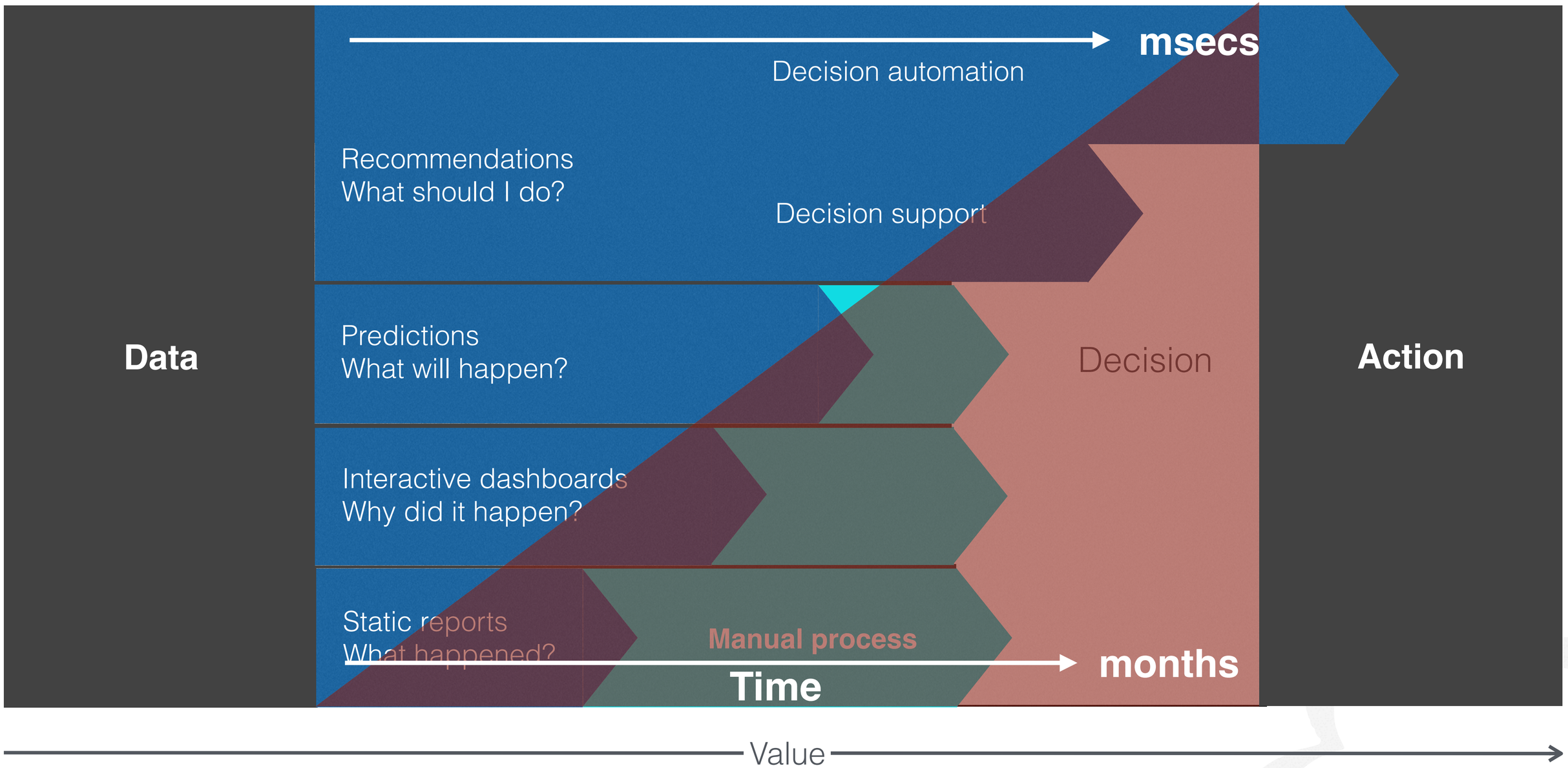
(C) Copyright Cloud Explorers Solutions AS 2016

Data Science - From Data to Decisions to Actions



(C) Copyright Cloud Explorers Solutions AS 2016

Data Science - From Data to Decisions to Actions

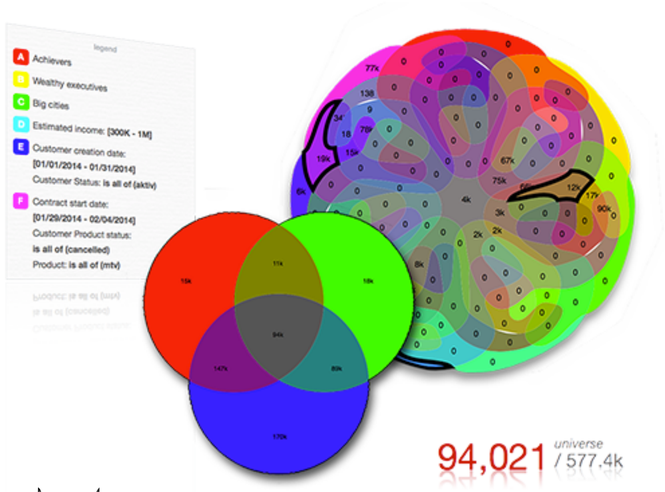


(C) Copyright Cloud Explorers Solutions AS 2016

Seeing is believing!

(C) Copyright Cloud Explorers Solutions AS 2016

Seeing is believing! PRISM Analytics & Viking DATALAKE



Drive Big Data exploration with an easy to use analytic interface

Anyone can start rapid exploration of big data repositories. With an intuitive use interface, you can create segments based on simple criteria, or business-driven filters. Exploration is as simple as looking for a holiday trip on hotel.com! Ajust slider controls, pick values, enter your criteria, and you see your target group segmented live in a graphical way, including in an incredible venn diagram.

Open the doors to A/B Testing. Use categorised filters to create additional effects, such as global filters, quarantine filters.

Create advanced analysis, detect trends, find patterns, and use your findings as a source of information to refine segmentation down the road. It's as simple as using the result of a previous segment as a filter criteria !

PRISM

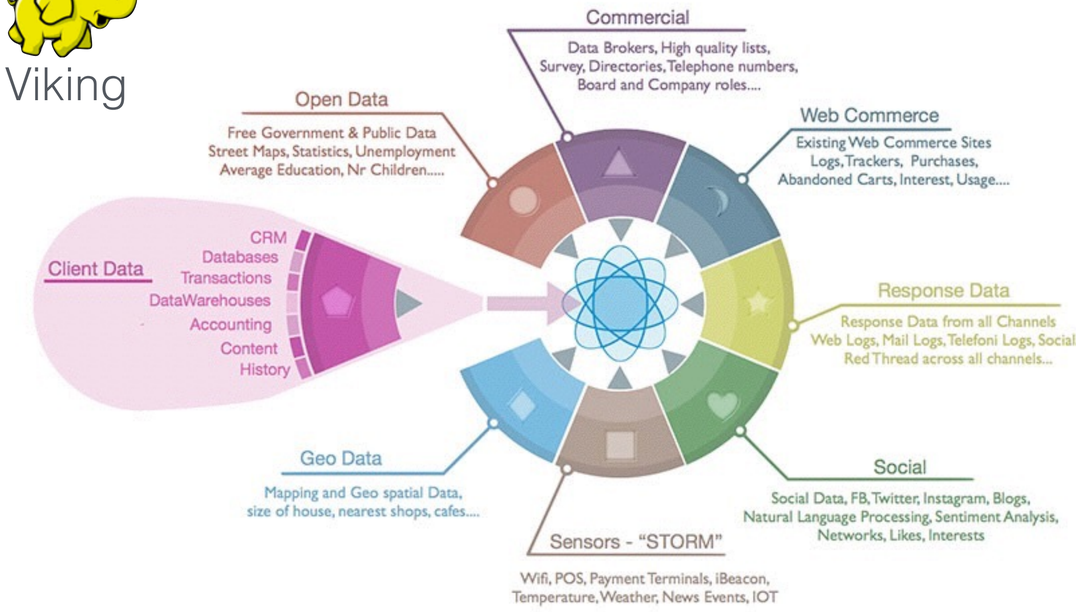
All your activities, simple to find and work with

No need for complex classification or structure. It's never been easier to organise all your work.

The user interface has been tailored to support facetted searches. Filtering on authors, tags, dates is always just one click away! Or you can do free text search too!

today	2	customers	24	prospecting	16
this week	7	sales	5	performances	7
this month	24	campaign	8	sales	4
last month	29	christmas	10	customer satisfaction	6
this year	356	gold	19	loyalty	10
last year	2410	sms	9	complains	5
				analysis	19
				fraud	2

Viking



BIG DATA LAKE

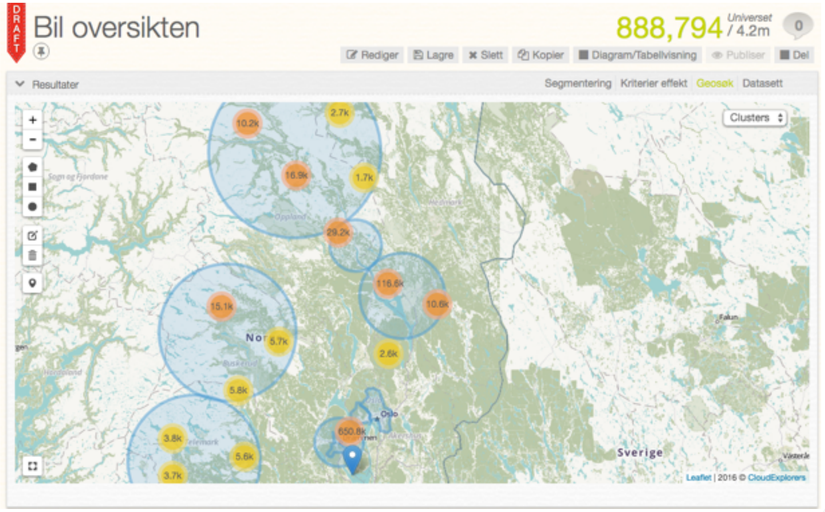
Any data, any volume, any velocity

No need for predefined data structure

There is no need for a large data warehouse projects that spread across months. There is no need to know your data in advance. Prism analytics can crunch any unstructured dataset for any type of industries. Only simple configuration is needed to get started.

And since it's cloud based, you can get started now, no need for software installation.

GEO Search



(C) Copyright Cloud Explorers Solutions AS 2016

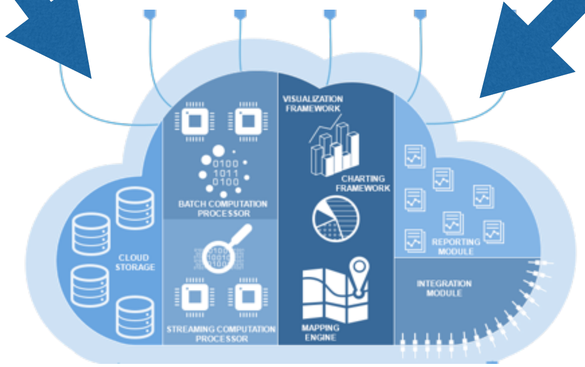
Suggestions on Getting Started

Typical Approach To Big Data

STRATEGY

DATA SOURCES

POC



PLATFORM

PRODUCTION

(C) Copyright Cloud Explorers Solutions AS 2016

THANK YOU :-)

cloud explorers

Stephen Karl Ranson

CEO
Founder and Partner

stephen@cloudexplorers.com
mobile : +47 992 88 221

(C) Copyright Cloud Explorers Solutions AS 2016