

# From defensive to offensive data-driven engineering

– data strategy, examples of defensive and offensive data management activities, method for identify AI solutions, and AI in healthcare

Erik Perjons

Department of Computer and Systems Sciences

Stockholm University

# Questions

- **What is a data strategy?**
- **What should a data strategy include?**
- **Why do organisations need a data strategy?**

# **Data strategy**

# What is a data strategy?

- A **data strategy** - is a plan to organize, manage and govern the data assets in an organization

# What is the core of the data strategy?

- The **data strategy** needs to:
  - 1) clarify the **goal of the data strategy** for organizations
  - 2) given the goal, **provide data management activities**

# What is the core of the data strategy?

- DalleMule & Davenport (2017) claim that **an organization's data strategy** should have a **proper balance** between **offensive** and **defensive activities**

# Defensive part of the data strategy

## Goals for the defensive part of the data strategy:

- Ensure data security, privacy, integrity, quality, regulatory compliance, and governance

## Data management defensive activities:

- Ensuring that data is in compliance with regulations
- Introduce data access control
- Detect and limit fraud and theft
- Ensure data integrity of data flows
- Provide a single source of truth

# Offensive part of the data strategy

## Goals for the offensive part of the data strategy:

- Improve innovation, the competitive position and increase profitability, revenue, and customer satisfaction

## Data management offensive activities:

- Generate customer insights by using data analysis, advanced data modelling and data science (including AI) work
- Integrate customer and market data for supporting decision making
- Include real time analysis



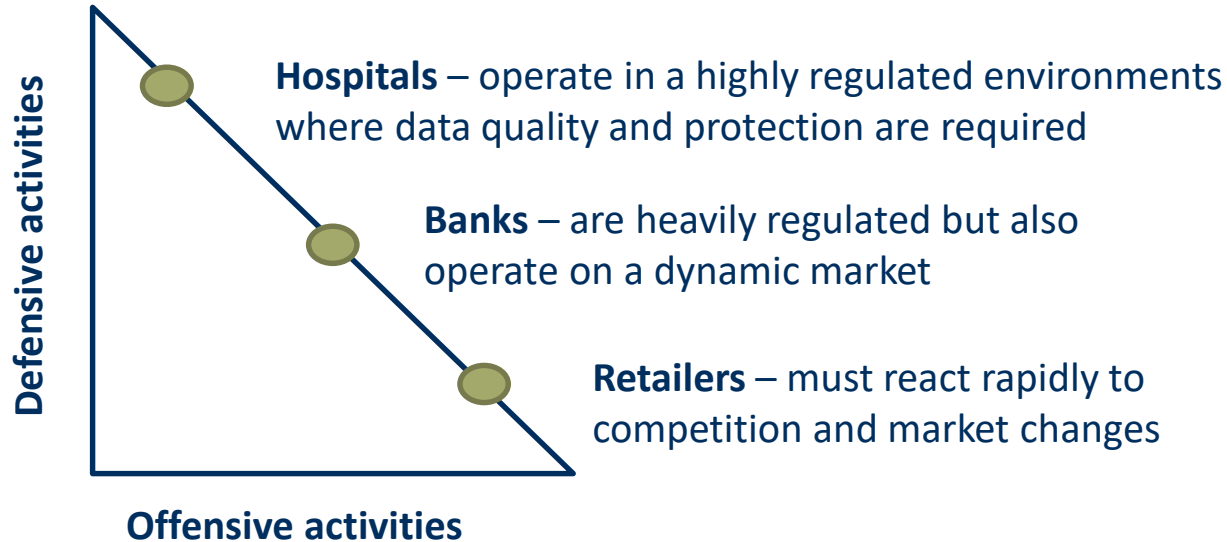
# The proper balance depends on a number of factors

- Market competition and dynamic
- Regulatory environment



External factors

# External factors



# The proper balance depends on a number of factors

- Market competition and dynamic
  - Regulatory environment
  - The overall strategy of the organisation
  - Maturity of data management
  - Centralized or decentralized data management
  - Size of data budget
- External factors
- Internal factors

# Focusing on just defensive activities can inhibit flexibility

- There is a risk that organisation focus too much on defensive activities – and data is not transformed into info that can be used by organizations strategically

# SSOT and MVOT

- The data strategy can include both defensive and offensive activities by introducing:
  - a single source of truth (SSOT) and
  - a multiple version of the truth (MVOT)
- Therefore, the framework could be seen as a **SSOT-MVOT model**

# Singe source of truth (SSOT)

- **Singe source of truth (SSOT)** - is a repository that contains one authorative copy of crucial data, such as customers, suppliers and product details (often called the master data)

# More about SSOT

- SSOT requires **data governance activities to ensure that the data is accurate and timely so that data can be relied on for both defensive and offensive activities**
- For example customers, suppliers and product details need to be specified in an agreed-upon way - supported by, for example a **master data management system**

# More about SSOT

- **If a SSOT does not exist** – the company may not understand:
  - what the relationships to customers and suppliers are
  - what details are correct about its customers, suppliers and products
- SSOT is often implemented by introducing 1) a **master data management system** or 2) **decide which systems are the master for different types of data**



# Multiple versions of truth (MVOT)

- **Multiple versions of truth (MVOT)** – provide different data for different business units
- **MVOT is based on a SSOT but adapted to different units' need.**
- That is, **SSOT data have to be transformed, enriched and adapted** to be useful for the different needs – for example, use different attributes for different concepts

# More about MVOT

- For example, the **marketing and financial department** are both interested in **ad spending**
- The **marketing department** is interested in **the effectiveness of advertise product and services**
- The **financial department** is interested **cash flow**, for example, when **the invoices were payed**
- That is, different departments are interested in different numbers, and therefore, their reports differs

# The need for MVOT

- According to DalleMule and Davenport (2017), **the need for SSOT is well understood, but not the need for MVOT is not**

# The need for MVOT

- **Different business units have different needs**
- Therefore, **SSOT data need to be transformed, enriched and adapted for different business unit**
- **MVOT is the result such business-specific transformation**
- However, **MVOT must diverge from SSOT in a carefully controlled way otherwise siloed and uncontrolled MVOT will be created**

# Centralized or a decentralized data management?

- If an organization should develop a **centralized or a decentralized data management** depends on the organizations position on the offence-defence spectrum.
- Organisations with a **defensive strategy** usually prefer a **centralized data management**
- Organization with a **offensive strategy** has a more **decentralised data management**, where Unit Chief Data Officers have responsibility to MVOT and an Enterprise Chief Data Officer owns the SSOT

# The elements of data strategy

	Defensive	Offensive
Key objectives	Ensure data security, privacy, integrity, quality, regulatory compliance, and governance	Improve innovation, competitive position and profitability
Core activities	Activities that optimize data quality, data extraction, standardization, storage, access	Activities that optimize data analytics, modeling, visualization, transformation and enrichment
Data management orientation	Focus on control	Focus on flexibility
Enabling architecture	Single source of truth (SSOT)	Multiple sources of truth (MVOT)

**Data governance – focusing on a defensive data strategy – but is a good base for an offensive as well**

# Data governance

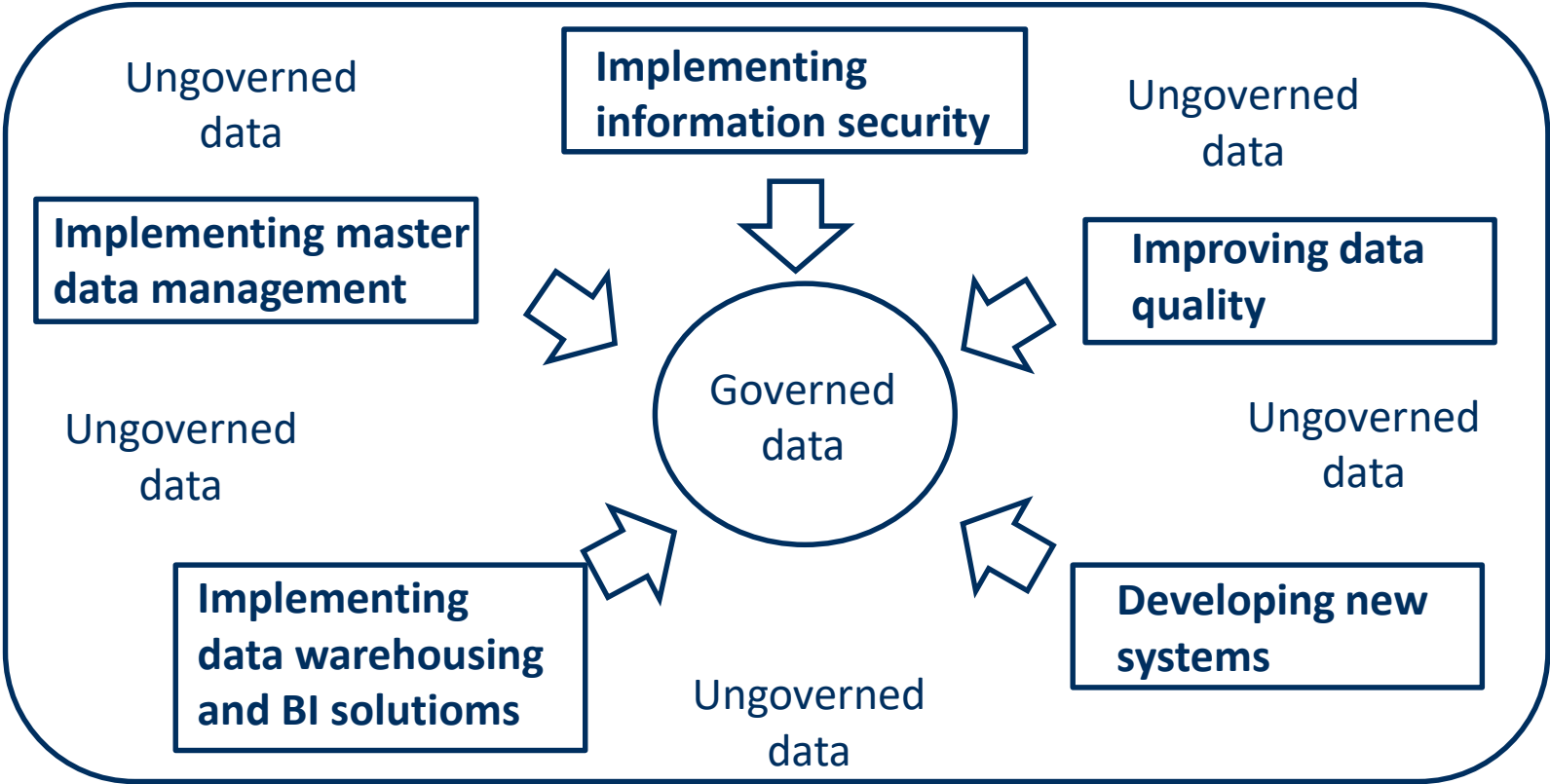
- Data governance aims to move data from an **ungoverned state** to a **governed state**, meaning:
  - data shall be owned
  - data shall be understood, inventoried and quality checked as well as corrected when data-related issues appear
  - data shall be wisely used



# Governed data

- **Governed data require:**
  - standardized business names
  - standardized business definitions
  - specified rules for data creation – specifying what is needed for creating certain data
  - specified rules for usage of the data – specifying for which purpose the certain data can or cannot be used
  - specified rules of data quality (in order to achieve and check such quality)
  - documented physical location of the physical instances of the data
  - specified data governors and data stewards responsible for the data

# Drivers for moving data to a governed state



**Data science – focusing on an offensive data strategy**

# Questions

- **How to identify new data-driven solutions, including AI, in an organization?**

**Method for identifying, architecting and developing data-driven solutions, including AI, in an organization**

# Method for identifying, architecting and developing data-driven solutions, including AI solutions

- The method is presented in Schmarzo (2013) and are developed for big data – not explicitly for AI, but I have adapted it for AI as well
- **Problem addressed by the adapted method:** It is not clear for organizations **how they can identify, architect and develop AI, big data - and other data-driven - solutions**
- Therefore, there is a need of a **solution engineering method supporting the organizations** addressing this problem

# Method for identifying, architecting and developing data-driven solutions, including AI solutions

1. Understand what make the organization successful – now, and in the future



2. Understand key business initiatives or opportunities



3. Brainstorm how AI can support a business initiative or an opportunity in focus



4. Break down the business initiative into use cases where AI is used – and for each use case define requirements



5. Validate the feasibility of the AI enhanced initiative (and the including use cases)

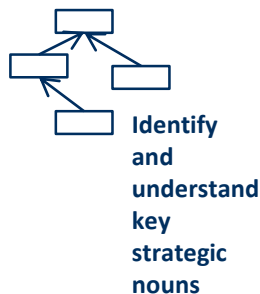


6. Design och implement the solution

# Step 1: Understand the organisation

1. Understand what make the organization successful – now, and in the future

(Schmarzo: Understand how the organisation makes money)



- Identify the most important strategic nouns and understand how they drive success, and envision how they, in the future, can drive further success
- **Examples of important strategic nouns:**
  - the major products and services
  - the revenue and cost drivers
  - the key issues to address
  - the key processes and activities
  - the business stakeholders and their roles
  - the major IT systems and their roles



# Step 2: Understand ongoing business initiatives

2. Understand key business initiatives or opportunities

(Schmarzo: Understand your organisation's key business initiatives)

**Identify and understand ongoing key initiatives or opportunities, based on step 1, but also based on:**

- Reading business reports, such as annual reports
- Reading presentations by executives
- Interviewing key employees

# Step 3: Brainstorm about AI impact

3. Brainstorm how AI can impact a business initiative or an opportunity

(Schmarzo: Brainstorm big data business impact)

**Four ways that AI, big data and advanced analytics can impact a business initiative or an opportunity:**

- "Mining" more detailed transaction data
- Integrate unstructured internal and external data - for more accurate and complete decision
- Improve real time delivery of data - for more timely decision
- Apply different forms of predictive analytics to uncover causality hidden in the data – for more actionable and predictive decision

# Step 4: Design use cases where AI is used

4. Break down the business initiative into use cases where AI is used – and for each use case define requirements

(Schmarzo: Break down the business initiative into use cases)

**Design use cases where AI, big data and analytics could enhance a business initiative in focus, and specify the following for each use case:**



- targeted stakeholders, including their roles and responsibilities
- business questions that the stakeholders try to answer
- business decisions that the stakeholders try to make
- requirements on data and data analysis algorithms/models as well as user experiences
- design key performance indicators (in order to make it possible to measure the success of the use case)

# Step 4: Design use cases where AI is used

4. Break down the business initiative into use cases where AI is used – and for each use case define requirements

(Schmarzo: Break down the business initiative into use cases)

**Design use cases where AI, big data and analytics could enhance a business initiative in focus, and specify the following for each use case:**



- targeted stakeholders and their roles and responsibilities
- business questions that the stakeholders try to answer
- business decisions that the stakeholders try to make
- requirements on data and data analysis algorithms/models as well as user experiences
- design key performance indicators (in order to make it possible to measure the success of the use case)

**Prioritize among the use cases**

# Step 5: Validate the AI enhanced initiative and included use cases

5. Validate the feasibility of the AI enhanced initiative - and the including use cases

(Schmarzo: Prove out the use case)

**Validate the feasibility of the AI enhanced initiative (and the including use cases) by deploy data and technology (like a prototype), and for the initiative:**

- Carry out a ROI/cost-benefit analysis
- Perform a feasibility study:
  - Make a plan to manage data – manage source systems, transformations, cleaning of data, decide master data, etc
  - Make a plan to test and fine tune analytical models
  - Develop mockups and wireframes to help the stakeholders understand the solution and its role in the daily business processes

# Step 6: Design and implement the solution

## 6. Design och implement the solution

(Schmarzo: Design and implement the big data solutions)

### **Design, plan for and implement the solution in form of one or a set of use cases, including, for example:**

- Capture and the store the data needed, including internal and external data, structured and unstructured data.
- Capture and the store additional data about customers, products and operations, for further data analysis. This data is mainly found outside the existing business processes.
- Implement real-time data access when required.
- Implement the AI solution

# **AI in healthcare**

## AI in healthcare - benefits and issues



# Why AI in healthcare?

- AI has the **potential to transform healthcare** since healthcare is producing **a large amount of clinical and administrative data**
- **This large amount of data can be used for analysis**
- Moreover, research studies have shown that **AI can carry out many key healthcare activities better than, or as well as, humans**, such as **diagnosing diseases**, for example by **analyzing radiology images**

# AI is sparsely implemented

- Today, AI solutions are **sparsely implemented in practical healthcare**
- Existing AI solutions are **mainly supporting the individual functions in healthcare, like radiology and pathology image analysis**

# Why is AI sparsely implemented? 1(2)

According to Davenport and Kalakota (2019), **two major reasons for AI being sparsely implemented in practical healthcare** are:

- **AI solutions are focusing on limited tasks** and are **rarely integrated into the clinical processes**
- Moreover, **AI is not implemented in electronic record systems (EHR)**. Therefore, AI is not part of the system that most healthcare personnel use for their day-to-day work

# Why is AI sparsely implemented? 2(2)

Panch et al. (2014) **add additional important reasons for AI being sparsely implemented in practical healthcare are:**

- **Healthcare systems are complex and fragmented, and will not easily change as a result of new technology**
- **Healthcare organisations lack the capacity to collect the necessary training data of sufficient quality - while also respecting ethical principles and legal constraints**

# AI technologies in healthcare

# AI technologies in healthcare

Note, according to Devenport and Kalakota (2019), AI is not one technology, but rather a collection of them.

Examples pf AI technologies:

- Machine learning
- Natural language processing
- Rule based expert system
- Physical robots
- Robotic process automation

# Machine learning

- **Traditional machine learning** is the most common application in healthcare. This application is mostly **using supervised learning**, which **requires a training datasets** to be used to be able to do the work
- Supervised learning systems are **supporting the making of diagnosis**, and **predicting what treatment protocols are likely to be successful** for a patient, **based on various patient attributes and the treatment context**

# Neural network and deep learning

- A more complex form of supervised machine learning is the **neural network**. **Neural network** make use of a **network of variables** that **associate inputs with outputs** and create **weights on these associations, in order to predict outcome**
- A **neural network** can also have **variables on many different so called hidden layers**, called **deep neural network** or **deep learning**
- Deep learning has been very successful for identifying **clinically relevant features in imaging data** - beyond what can be perceived by the human eye
- **Deep learning** is also increasingly used for **speech recognition in NLP**, see next slide



# Natural language processing

- **Natural language processing (NLP) aims to make sense of human language. NLP includes application such as speech recognition, text analysis, translation.**
- In healthcare, NLP can, for example, be used for **analyzing unstructured clinical notes** and supporting the **transformation from speech to text**

# Rule based expert system 1(2)

- **Rule based expert system** - require human experts and knowledge engineers to **construct a series of rules in a particular knowledge domain, which will be the base for the expert system**
- **Rule based expert systems in healthcare - are the base for many clinical decision support system**
- **Rule based expert systems - are also be part of many medical record systems (i.e. EHR systems), for example, they provide **functionality to warn for drug-to-drug interactions**, and **support the physician of making diagnoses****

# Rule based expert system 2(2)

## The limitation of rule based expert systems:

- Rule based expert systems - **work well if the rules are not so many.**
- However, **if number of rules is over several thousand, it is hard to maintain the rules**, for example, **the rules soon start to conflict with each other.**
- Moreover, **if the knowledge domain changes, rules need to change, which may be time-consuming**, especially if the rules are many, and related on each other
- Therefore, **due to this limitations, rule based expert systems are being replaced by systems based on ML algorithms**

# Physical robots

- **Physical robots** - perform pre-defined tasks in factories and warehouses, like lifting and assembling objects
- Applied in healthcare are **surgical robots** – which can **improve the surgeons ability to see and make tasks more precise**
- **Moreover, physical robots are also becoming more intelligent,** as other **AI capabilities are being embedded in their operating systems.**

# Robotic process automation (RPA)

- **Robotic process automation (RPA) – record the keyboard and mouse actions of a human being, and repeat these actions automatically**
- **RPA does not involve physical robots – instead RPA is a form of software**
- **RPA act like a semi-intelligent user of the systems, following a script or a set of rules based on actions done by human beings**
- **RPA can be used in healthcare for updating patient records, billing or other administrative tasks**
- **Moreover, RPA can be used in combination with other technologies, for example combining image recognition and RPA, where RPA can be used for extract data from the recognitions of images and update EHR system with this data**

# AI technologies can be combined

- **AI technologies are being more and more combined and integrated**, for example:
  - physical robots are getting AI-based features
  - image recognition is being integrated with RPA.

## AI application areas in healthcare

# AI application areas

## Example of AI application areas in healthcare:

- Diagnosis and treatment
- Patient engagement and adherence
- Administrative activities



# Diagnosis and treatment 1(3)

- **IBM's Watson has received a lot of attention** for its **application in diagnosis and treatment area**, particularly **cancer diagnosis and treatment**
- Watson consisted **of a set of 'cognitive services'**, employing a combination of **machine learning and NLP technologies**
- However, **IBM's Watson's application in healthcare has not been a success:**
  - Watson has **not been able to handle different types of cancer**
  - Watson has also **been hard to integrate into care processes and systems**

# Diagnosis and treatment 2(3)

## Other examples of the use of AI for diagnosis and treatment:

- Several organizations **work on ML based solutions** to better **understand the how different genetic variants of humans will response to different treatments, such as drugs and protocols.**
- Organizations are also **working on ML based solution to predict populations at risk of particular diseases, high-risk conditions or to predict hospital readmission**

# Diagnosis and treatment 3(3)

## Drawbacks of using AI in the application of diagnosis and treatment:

- **To embed AI-based diagnosis and treatment recommendations into clinical workflows and EHR systems has not been successful**
- According to Davenport and Kalakota (2019), **“such integration issues have probably been a greater barrier to broad implementation of AI than any inability to provide accurate and effective recommendations”**



# Patient engagement and adherence 1(2) Stockholms universitet

- **Patients engagement in their own well-being and care are important for receiving better outcome** in healthcare
- **The major problem is that the patient may not make necessary behavioral adjustment, that is, does not follow a course of treatment or take the prescribed drugs**

# Patient engagement and adherence 2(2)

- Therefore, ML and business rules engines can be used to support patient engagement and adherence, by:
  - sending message alert to patients,
  - providing targeted content given the patients' status and characteristics,
  - tailoring recommendations by comparing patient data to other effective treatment pathways for similar cohorts
  - nudging patient behavior in a more anticipatory way

# Administrative activities

Different AI technologies can be used for administrative tasks:

- RPA can be used for a variety of applications in healthcare, like managing medical records
- NLP can be applied in chatbots for patient interaction
- ML could be used to verify whether millions of insurance claims are correct, for example, by applying probabilistic matching of data across different databases

## Healthcare workers

# Implication for healthcare workforce 1(2)

- According to Davenport and Kalakota (2019) estimate that **it will take 20 years before will see any substantial change in healthcare employment due to AI**
- **Instead, there is also the possibility that new jobs for working with AI technologies are created**



# Implication for healthcare workforce 2(2)

- The area where **most healthcare jobs will be automated** are those dealing with **digital information, radiology and pathology**
- **However, for example, not even radiologist jobs will not disappear in the near future, and maybe not in the long term either – see next slides**

# Implication for radiology 1(2)

- Today, **radiology AI systems can only perform single tasks.**
- **Radiology AI systems cannot fully identify all potential findings in medical images.**  
**Radiologist are still needed for that**
- **Radiologists also do a lot of other thing than just read and interpret images:**
  - radiologists **relate findings from images to other medical records and test results**
  - radiologists **consult with other physicians** regarding diagnosis and treatment
  - radiologists **discuss procedures and results with patients**
  - radiologists **define the technical parameters of imaging examinations.** The parameters need to be tailored to the patient's condition

# Implication for radiology 2(2)

- **Moreover, for employing full scale AI-based image work:**
  - **clinical processes need to be changed**, which will take time
  - **an aggregated repository of radiology images is required for training the AI system**, but such **an aggregated repository is lacking today**
  - **changes in medical regulation and health insurance contracts for automated image analysis are needed**

## A brief summary

# To summarize 1(2)

- The **greatest challenge to AI** is to **ensure its adoption in daily clinical practice.**
- **There are a number of challenges to overcome to achieve this.**
- Therefore, Davenport and Kalakota (2019) **estimate that we will see a limited use of AI in clinical practice within 5 years and more extensive use within 10**

## To summarize 2(2)

- Moreover, **“AI systems will not replace human clinicians on a large scale, but rather will augment their efforts to care for patients”**.
- According to Davenport and Kalakota (2019) **it might take 20 years before will see any substantial change in healthcare employment**