YYT-C3001 Management of environmental data and information

Lecture 8: Cloud services and infrastructures

Aalto University
School of Engineering

Jussi Nikander

8.10.2020

Contents of this lecture

Cloud services

- **Technological enablers**
- **Advantages and disadvantages**

Knowledge management

- **The DIKW-pyramid**

Big picture of managing spatial data in organizations

Learning goals for this lecture

Understand why could computing is a useful

Remember the advantages and disadvantages of cloud computing

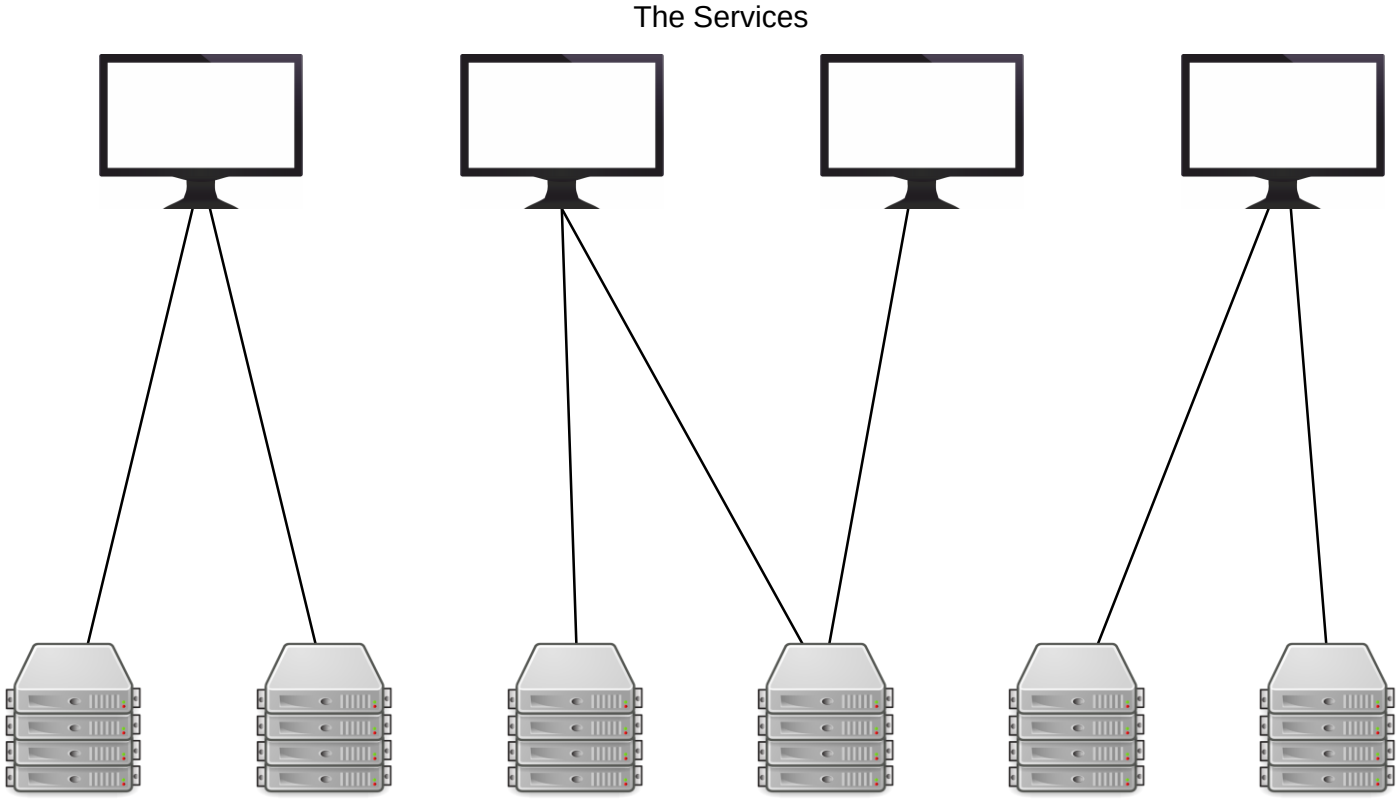
Understand the basics of knowledge management and its relevancy to spatial data management

Cloud services

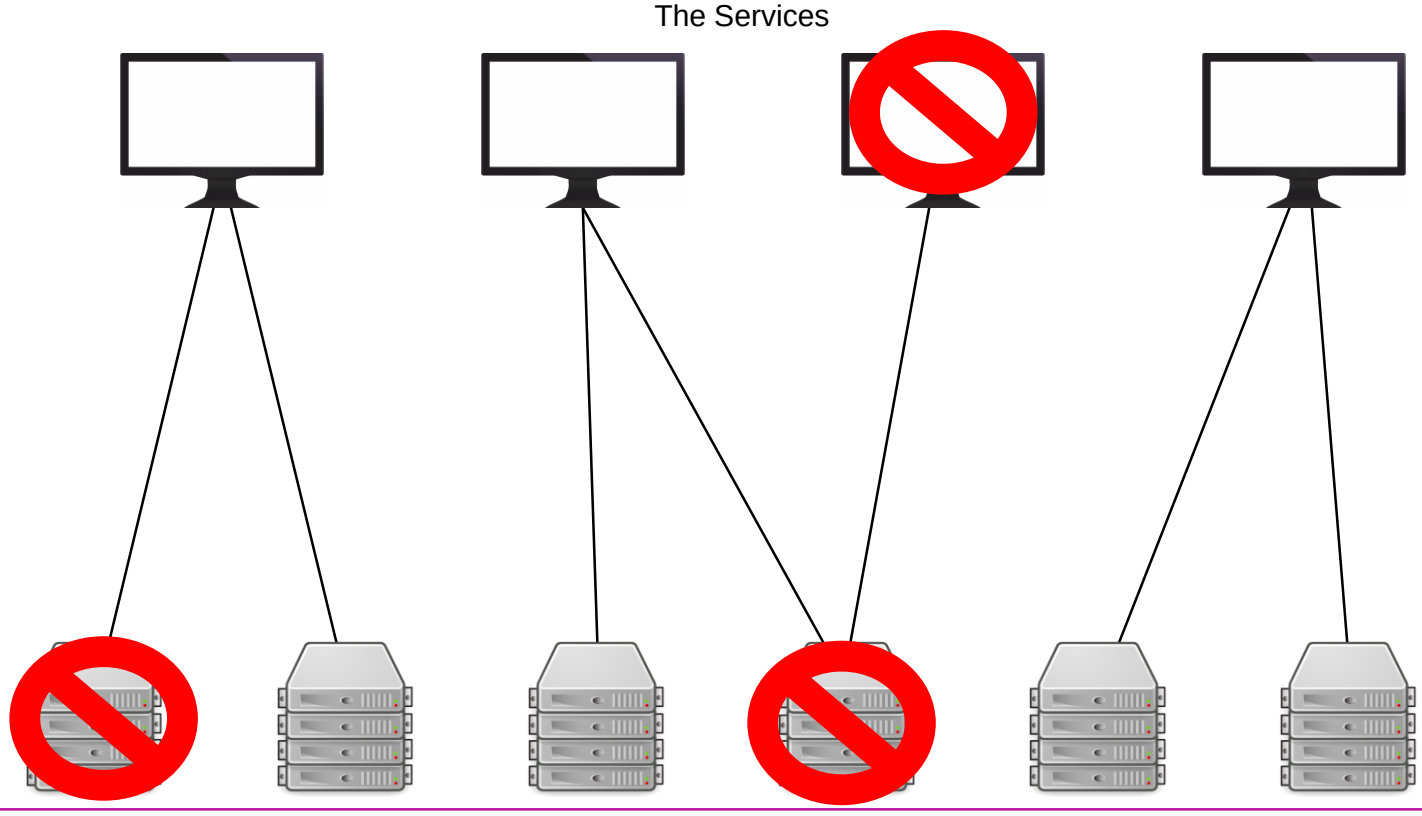
Cloud Computing: a very short introduction

- **Cloud computing is a style where massively scalable computation capabilities are delivered as a service over internet**
- **Instead of buying computational capability as new hardware, computational capability is bought as a service from a service provider as needed**
  - *Typically the majority of hardware capability is unused as capability must be estimated according to peak demand*
  - *Buying capability as service removes the need to maintain hardware and utility software*
  - *Allows cost reductions (no need to buy hardware) and organizational agility (capability can easily be scaled according to need)*

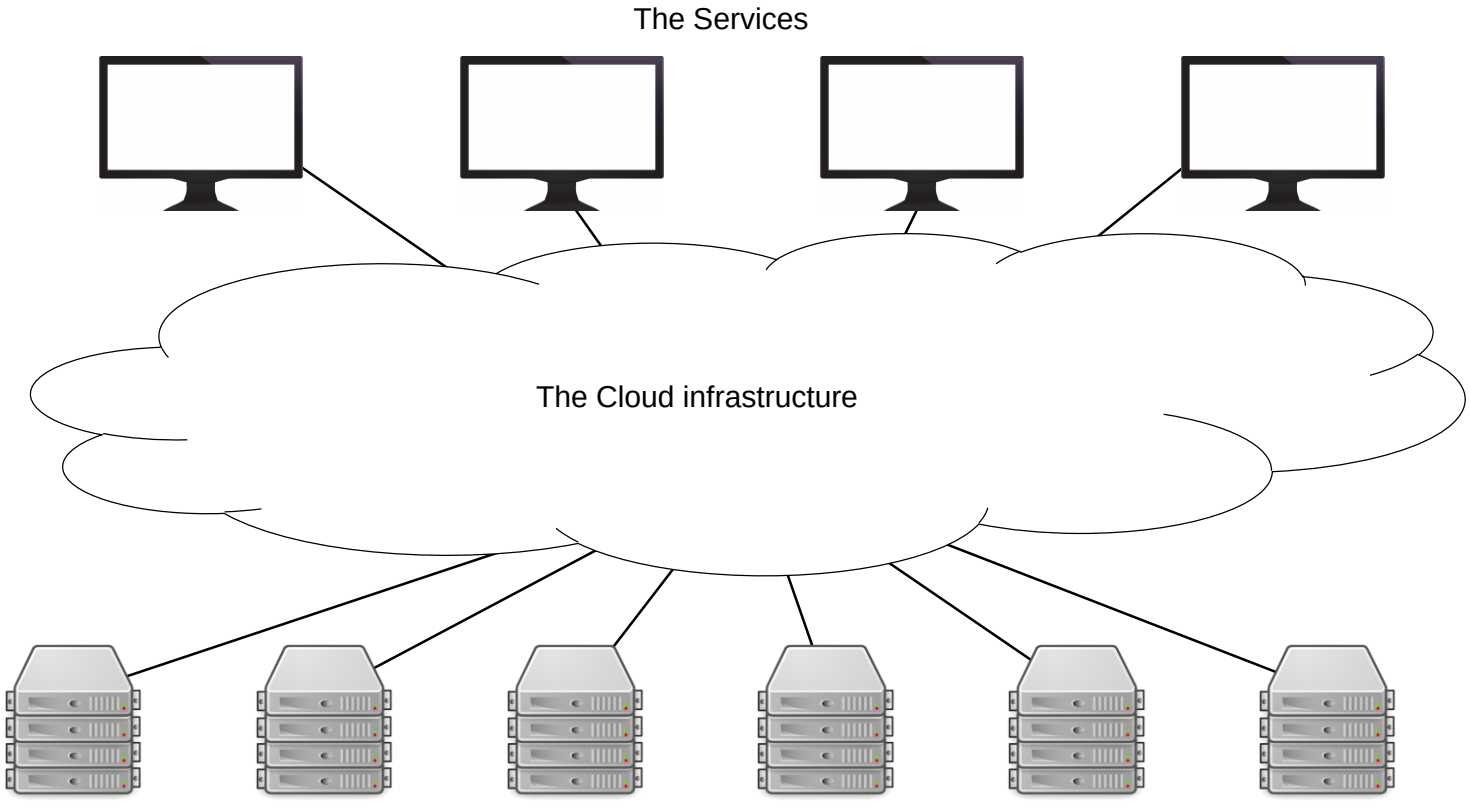
A (very) abstract model of traditional web services



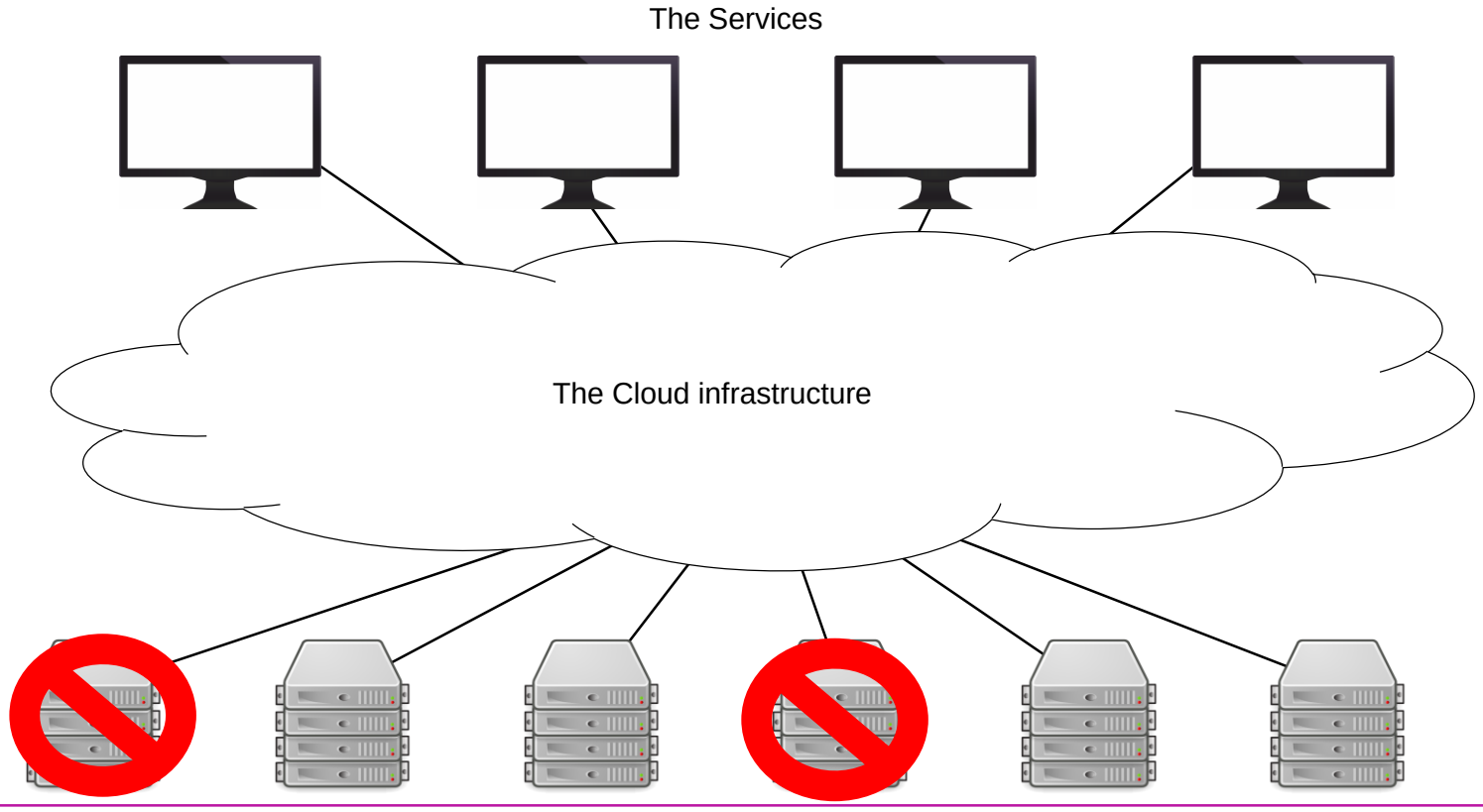
A (very) abstract model of traditional web services



A (very) abstract model of cloud computing



A (very) abstract model of cloud computing



Technological enablers of Cloud computing

- **Broadband networks**
  - *High bandwidth, low latency, reliable, and fault-tolerant data delivery*
- **Data center technology**
  - *Modular hardware grouped together and maintained together with high level of automation and reliability*
  - *Provides enough computation power for a large number of services*
- **Virtualization**
  - *Decouples physical hardware from network services provided*
- **Web technologies**
  - *Modern web technologies allows services to be easily deployed, shared, and accessed on the internet*

Different types of cloud services

- **Cloud services can be characterized in many different ways, this is merely one**
- **Software as a Service (SaaS): cloud is used to run an application – the application can be accessed through some client program**
  - *Webmail, virtual desktop,* ***engineering software***
- **Platform as a Service (PaaS): cloud is used to provide a platform to run specific customer software**
  - *Database, web server*
- **Infrastructure as a Service (IaaS): cloud is used to provide the customer with fundamental computing resources**
  - *Cloud storage, virtual machine*
- **etc.**

Advantages of the Cloud

- **(Practically) unlimited computation and storage capacities**
  - *Most useful in calculation that can be parallelized*
- **Ability to quickly solve significantly larger problems than locally**
- **Capacity can be scaled according to need**
- **Redundancy, security, and reliability can be better**
- **Ability to access the functionality from any computer**
- **Many maintenance issues are automatized**

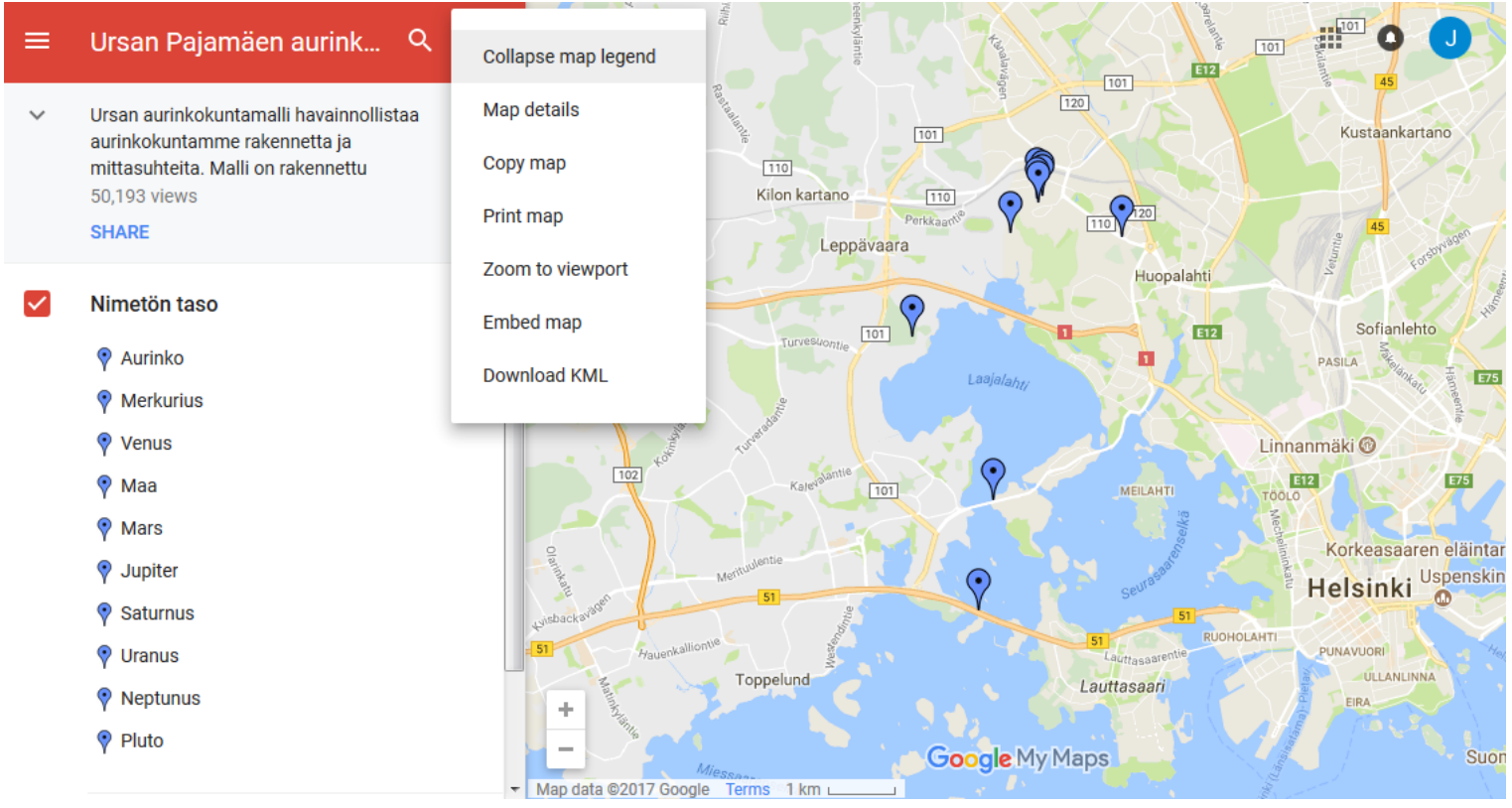
Disadvantages of the Cloud

- **Most useful when there is a need for very large, highly parallel computation**
- **Virtualization induces overheads**
- **Very dependent on high-speed internet connection, especially when large amounts of data are transferred between local and remote machines**
  - *Latency can be an annoyance*
- **Data location may not be known**
- **Some of the control over the service is lost, since it is (partially) outsourced**
- **Typically involves several organizations, which can cause organizational friction**

Some Spatial Cloud Services

- GIS Cloud
- ArcGIS Online
- QGIS Cloud
- CARTO
- Mapbox
- Etc.
- Etc.
- Etc.
- Etc.

Spatial Cloud Service Example: Google Maps



Spatial Cloud Service Example: ArcGIS Online

Break and small exercise

Let’s have a 10 minute break. Stand up a bit.

Also, add to Jamboard at least one spatial cloud service (try to add a new service to the list)

https://jamboard.google.com/d/1kbrQwDV94T3rrsU2iXaXoTenNBb9jUuc7CTd_x3pAhc/edit?usp=sharing

Data, information, knowledge, wisdom, and their management

The context: organizatorial point of view

- **Knowledge management** aims to identify, use, store, and exploit the knowledge an organization has
- **Knowledge (wisdom)** is information that has been applied in a specific context
- **Knowledge (information)** is a justified, true belief
- **Knowledge (wisdom and information)** are used in order to improve the organization
- **Knowledge** of an individual can be either **tacit** or **explicit**
  - *Tacit knowledge is experience-based (hiljainen tieto)*
  - *Explicit knowledge is articulated (eksplisiittinen tieto)*

Organizatorial knowledge, an example

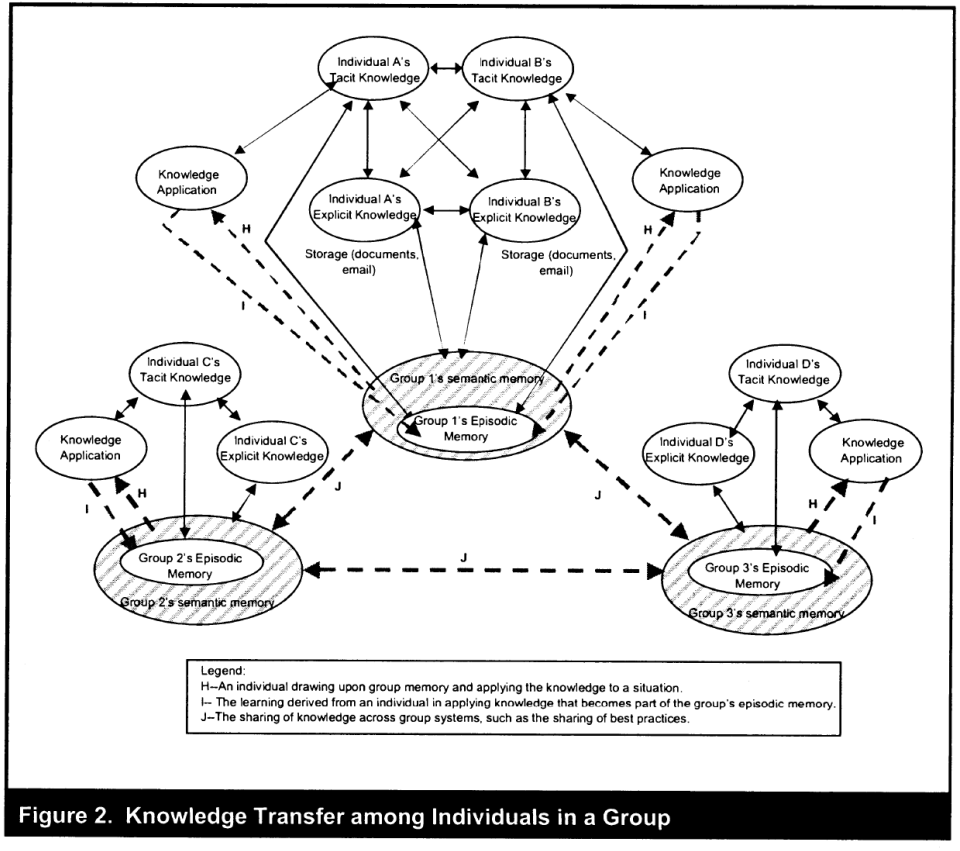


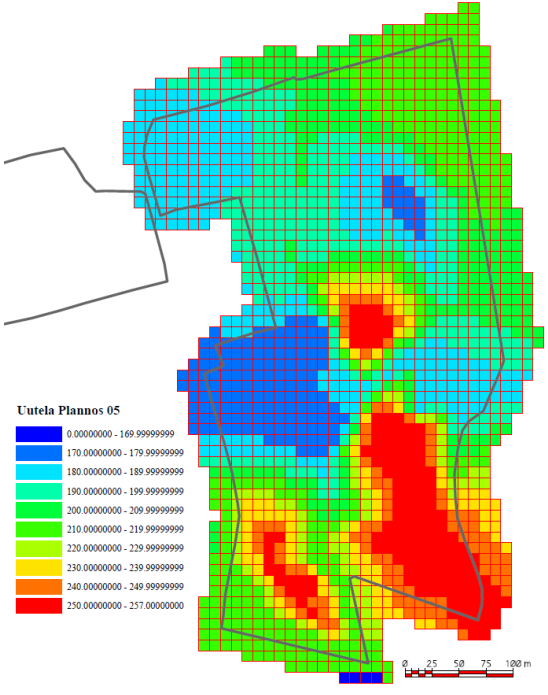
Figure 2. Knowledge Transfer among Individuals in a Group

Source: Alavi, M. and Leidner, D.E., 2001. Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS quarterly*, pp.107-136.

Knowledge application: a simple example

- **Base knowledge:** fertilization tends to help plants to grow
- **Context:** agriculture
- **Applied knowledge:** fertilization improves yield
- **Organizatorial improvement:** Fertilization causes k € expenses and is expected to increase the value of the harvest by (k+m) €. Therefore fertilizer is applied to the fields

Better knowledge that is applied well can improve processes even more



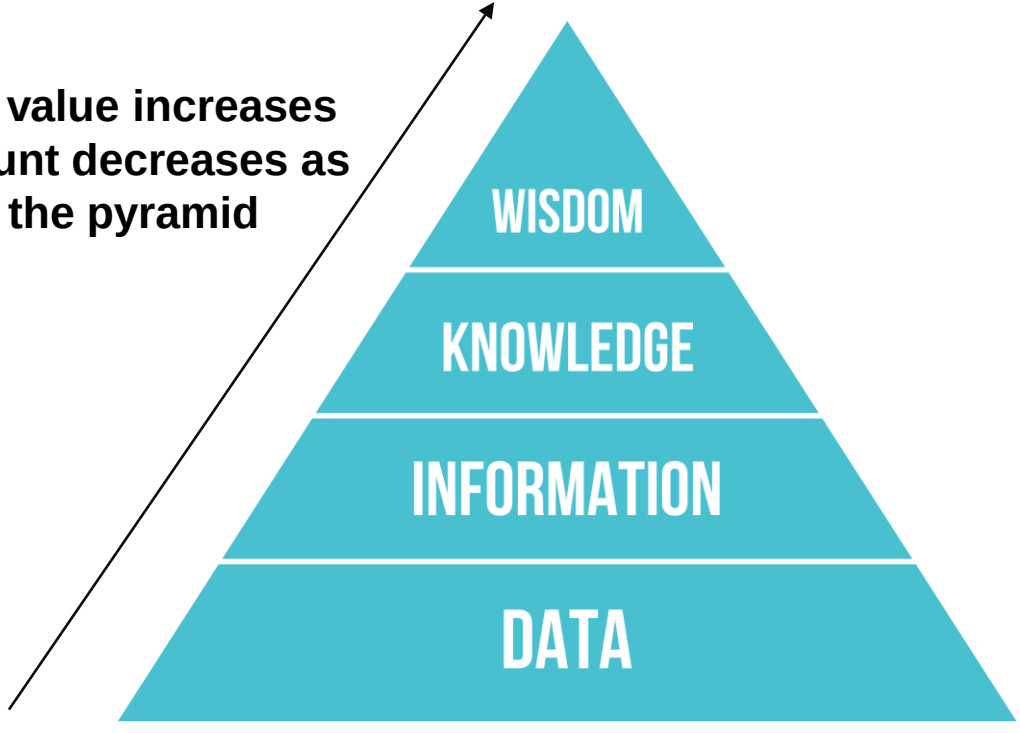
The yield potential in a field is not constant; different amount of fertilizer should be applied in different parts of the field

The context: information science point of view

- **Data** is symbols or signals; values
  - *For example 20201008*
- **Information** is data that has been given a meaning
  - *For example 20201008 represents the date October 8th, 2020*
- **Knowledge** is information put into a context
  - *For example, 20201008 is the date of this lecture*
- **Wisdom** is the ability to use the knowledge
  - *Today, October 8th 2020, is a day when I must remember to attend a lecture*
- The definitions for knowledge and wisdom vary a lot, depending on who is talking and in what context

The DIKW pyramid

Typically value increases and amount decreases as we climb the pyramid



Data, information, and databases

- Digital information is managed using **databases**
- A database is an ordered and indexed group of data elements that have a specific meaning (information), often used in specific context (knowledge)
- Database makes it easy to use the information
  - *It is possible to **retrieve**, **insert,** and **delete** data in a database*
- Databases are not a new invention
  - *We used to call them archieves*
  - *The user interface was a person called archivist*

The big picture: data, technology, services, organizations

The Big Picture: Reference Model of Open Distributed Processing

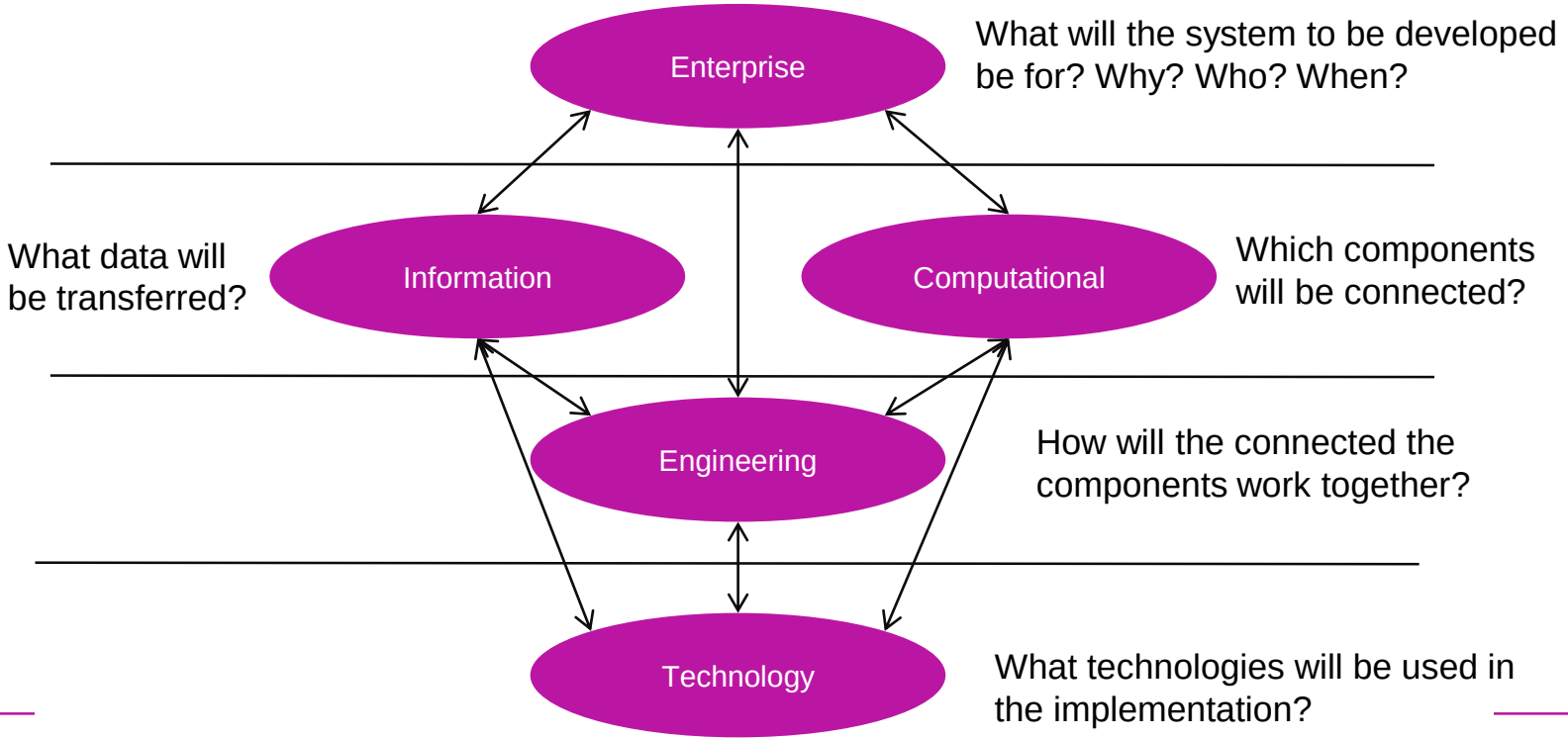
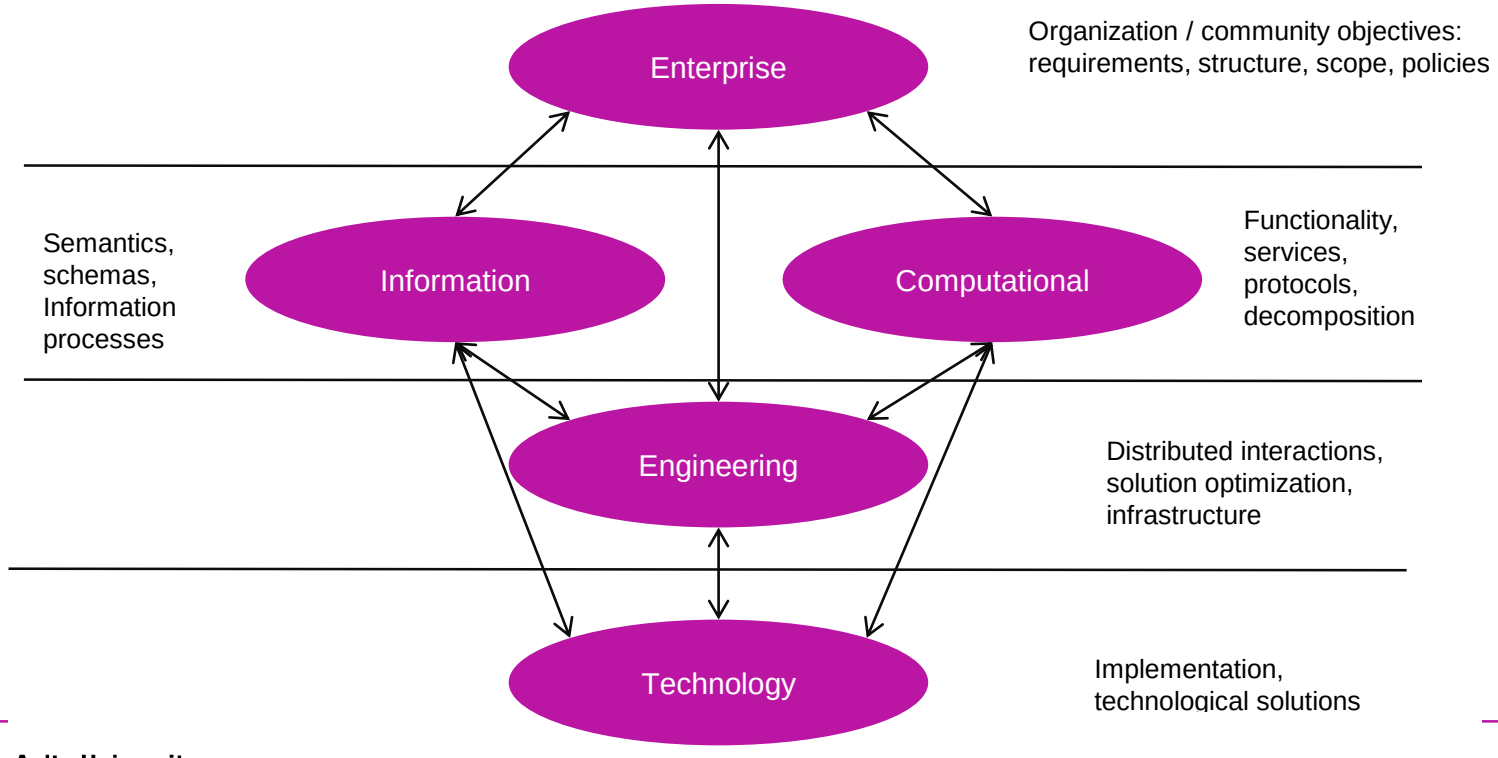


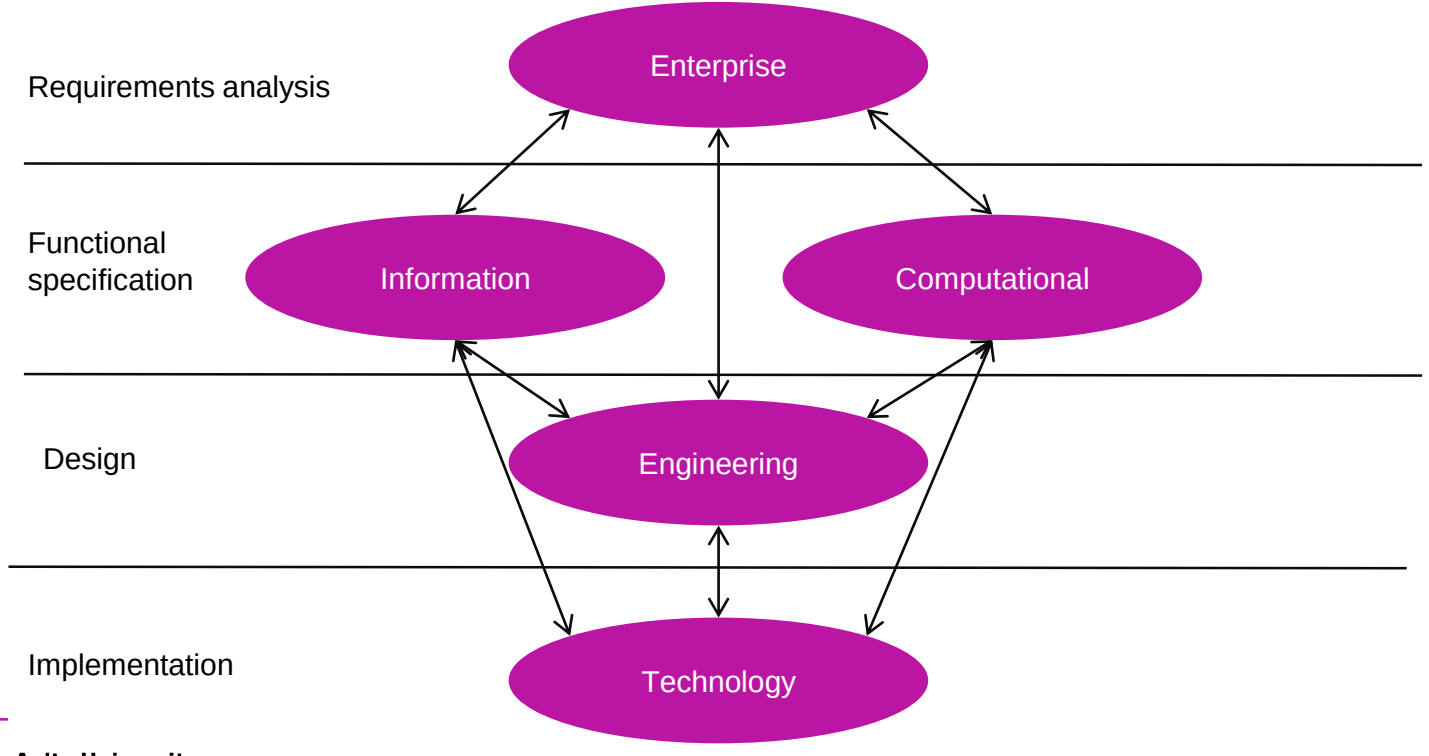
Figure adapted from: Raymond, K., 1995. Reference model of open distributed processing (RM-ODP): Introduction. In *Open distributed processing* (pp. 3-14). Springer, Boston, MA.

The Big Picture: Reference Model of Open Distributed Processing

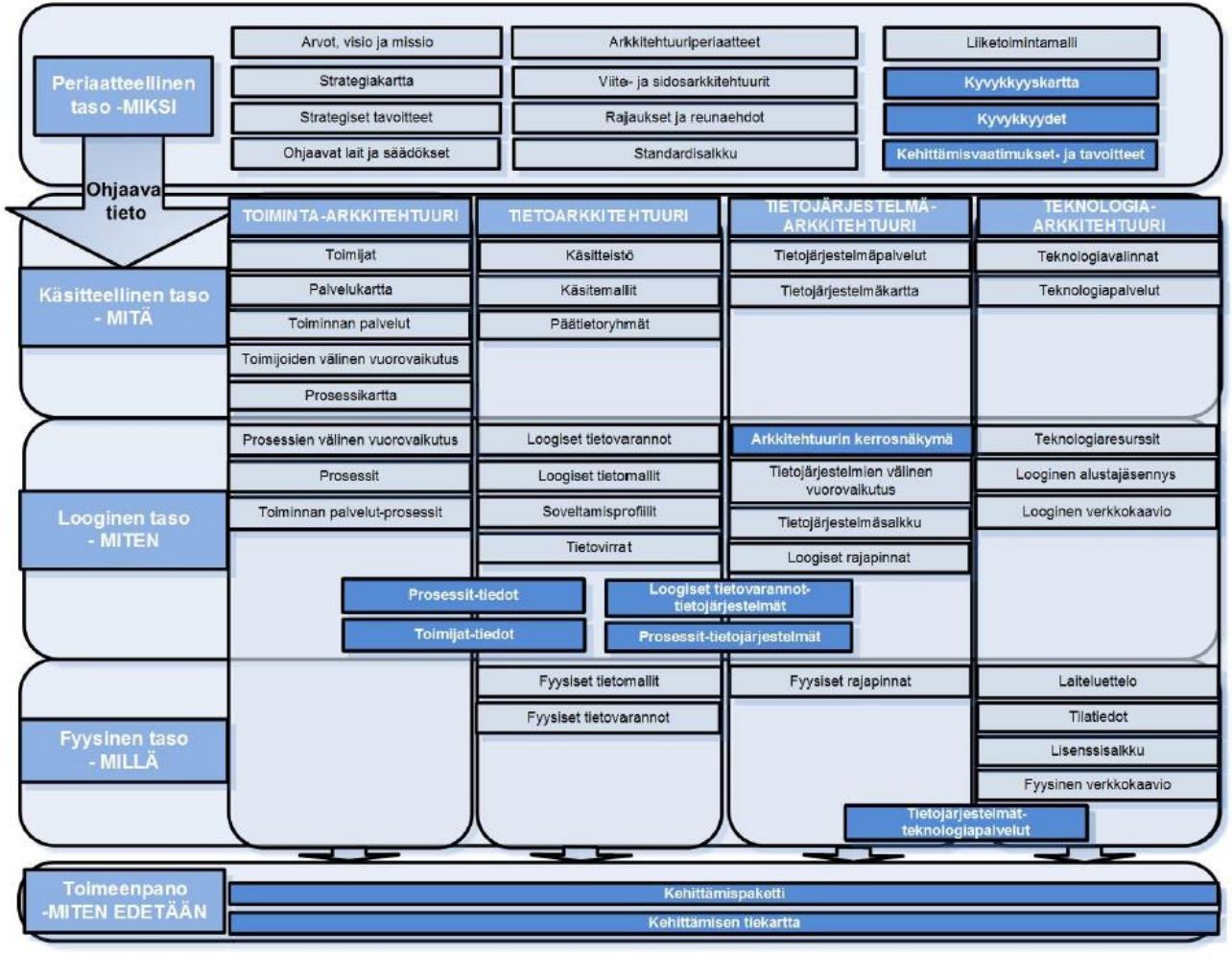


The Big Picture: Reference Model of Open Distributed Processing

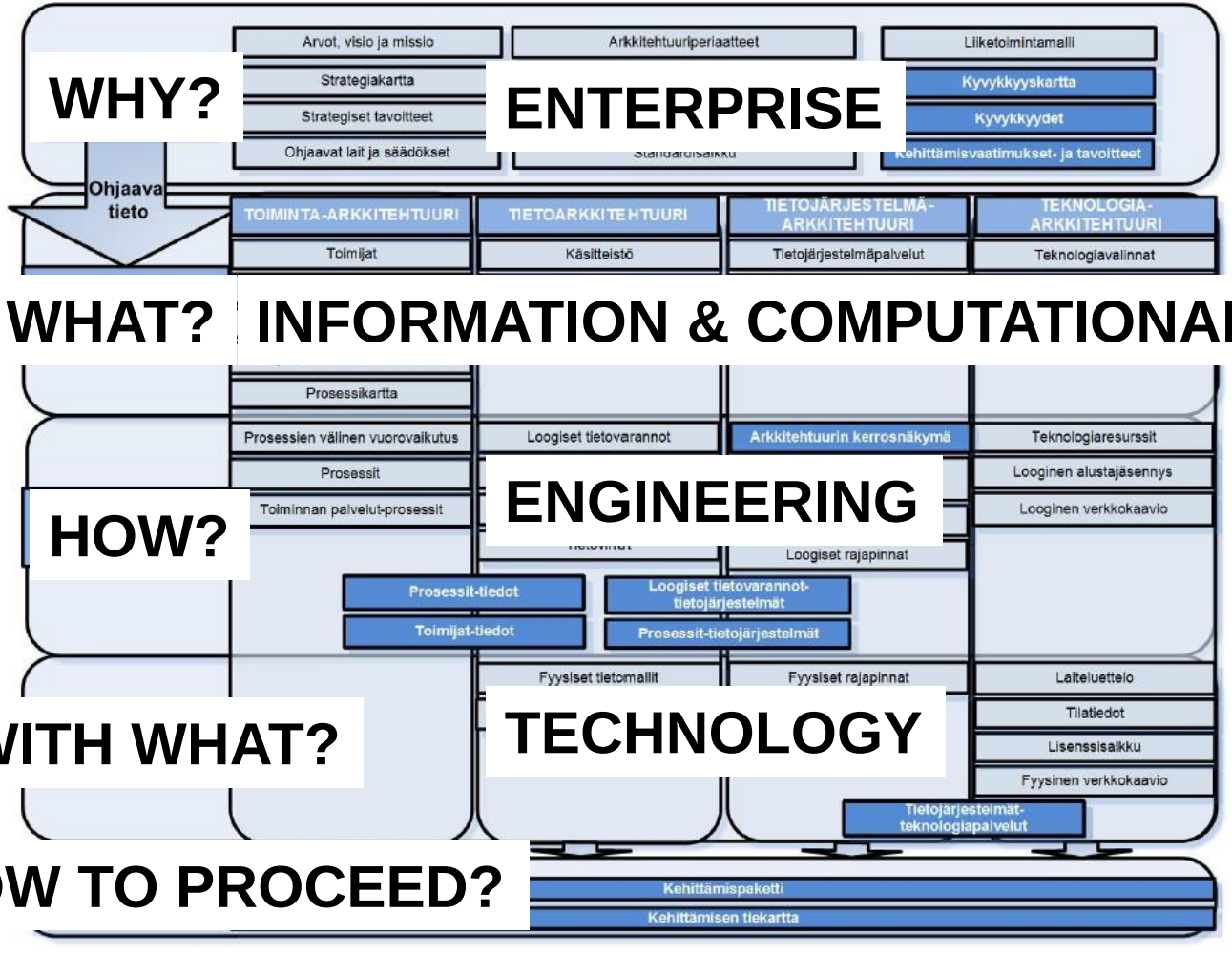
- Comparison to (classical) software engineering



Reference models in action: JHS 179



Reference models in action: JHS 179



For the next time…

Submit the fifth (and final) exercise

Continue writing the learning diary