NOVA IMS Information Management School

Research Methodologies

Master in Data Science and Advanced Analytics

Design Science Research Methodology (DSRM)

Roberto Henriques

2022/2023

Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Agenda

- What is Design Science Research (DSR)
- DSR Context
- Research outputs
- DSR Methodology

Agenda

- What is Design Science Research (DSR)
- DSR Context
- Research outputs
- DSR Methodology

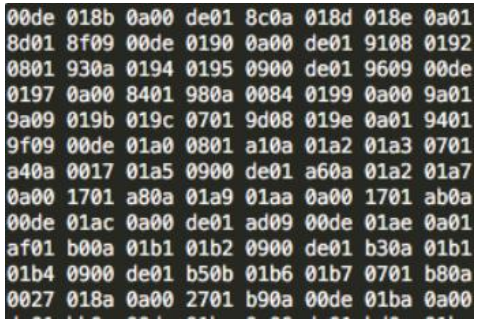
What is Design?

- **Design** is the **instructions based on knowledge** that turns **things into value** that people use. It embodies the instruction for making the things.
- However, **design is not the thing**.

(Henver & Chatterjee, 2010)

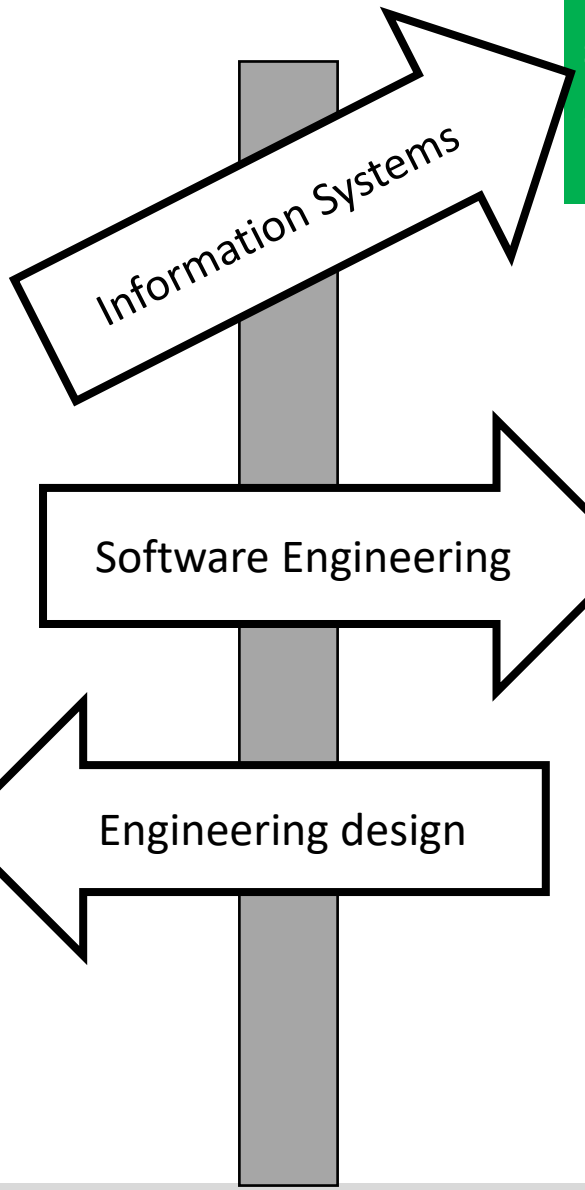
e.g.

source code=design



compiled code= thing itself

Design means diferent things to diferent areas



Design is the **building software artifacts** which solve a human problem.
(Walls et al., 1992)

Design is the "thing" as well as the "process" which keeps human concerns in the center.
(Winograd, 1996)

Design is the systematic intelligent generation and evaluation of specifications for artifacts that meet the stated objectives.
(Dym & Little, 2000)

What is research?

Research can be defined as an **activity** that contributes to a better **understanding** of a **phenomenon**.

Phenomenon= set of behaviours of some entity that is found interesting by researchers.

Understanding= is the knowledge that allows prediction of the behaviour of some aspects of the beahviour.

(Kuhn, 1970; Lakatos, 1978; Henver & Chatterjee, 2010)

What is Design Science Research (DSR)?

*"Design science research (DSR) is a **research paradigm** in which a designer **answers questions** relevant to human problems via the **creation of innovative artifacts**, thereby contributing new knowledge to the body of scientific evidence. The designed artifacts are both useful and fundamental in understanding that problem."*

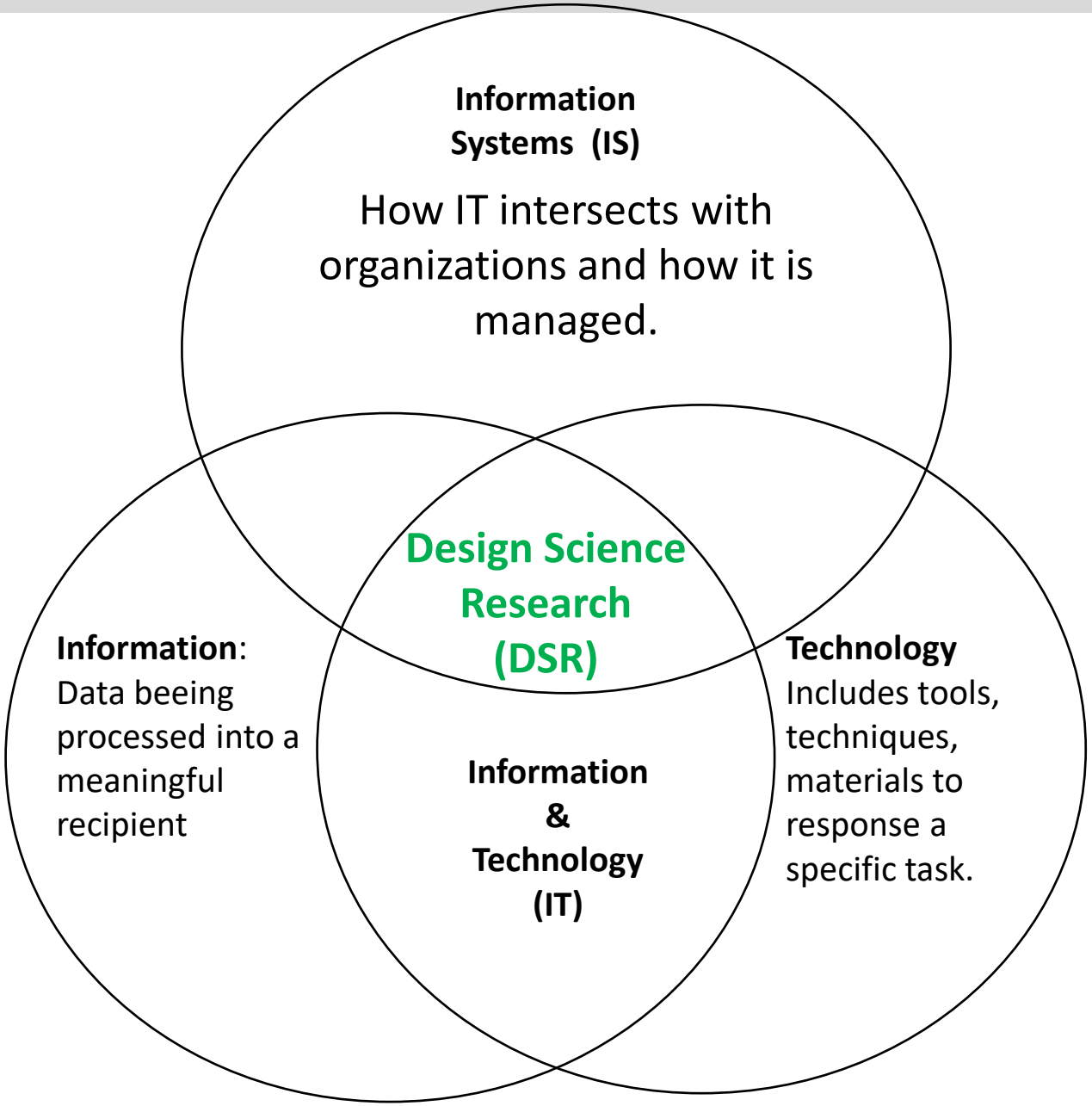
(Henver & Chatterjee, 2010, p.5)

First Principle of DSR

*"The **fundamental** principle of design science research is that **knowledge** and **understanding** of a design problem and its solution are acquired in the **building and application of an artifact**."*

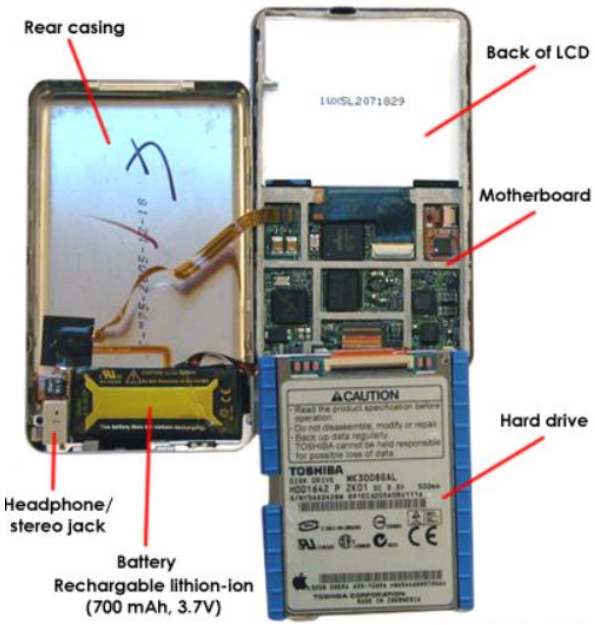
(Henver & Chatterjee, 2010, p.5)

DSR Context


Information Systems (IS)
How IT intersects with organizations and how it is managed.
Design Science Research (DSR)
Information: Data beeing processed into a meaningful recipient
Information & Technology (IT)
Technology Includes tools, techniques, materials to response a specific task.

Routine Design vs. DSR

Components



Artifact

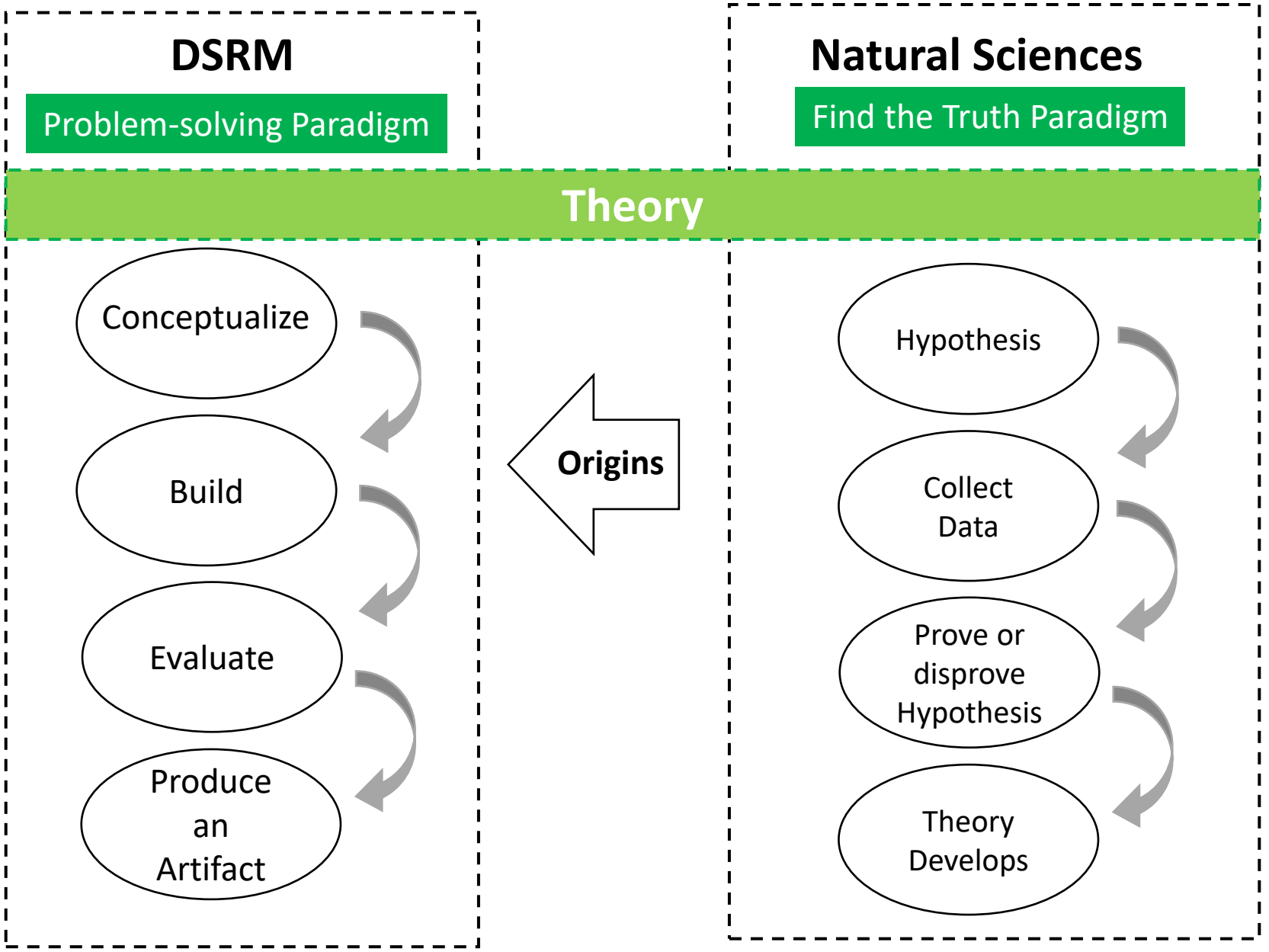
Design Research

Document the artifact, with rigorous evaluation method comparing with other artifacts-> knowledge created

No new knowledge was created, applied best practices

Routine Design Practice

Research Paradigms



Research question

"In **natural and social sciences**, the main question has been: **What or which kind is the world?**

Concerning an **artifact** and its construction process we ask: **Why and how we build an artifact ?**"

(Jarvinen, 2004)

"To define a problem as any situation where a gap exists between the actual and the desired ideal states."

(Sekaran, 2003)

Spectrum of DSR

Artifact is used to describe something that is **artificial**, or **constructed by humans**, as opposed to something that occurs naturally. Such **artifacts**, must **improve upon existing solutions** to a problem or perhaps provide a first solution to an important problem.

(Simon 1996)

IT artifacts:

- Constructs (vocabulary and symbols)
- Models (abstractions and representations)
- Methods (algorithms and practices)
- Instantiations (implemented and prototype systems)

Research Output

Constructs (vocabulary, concepts and symbols)

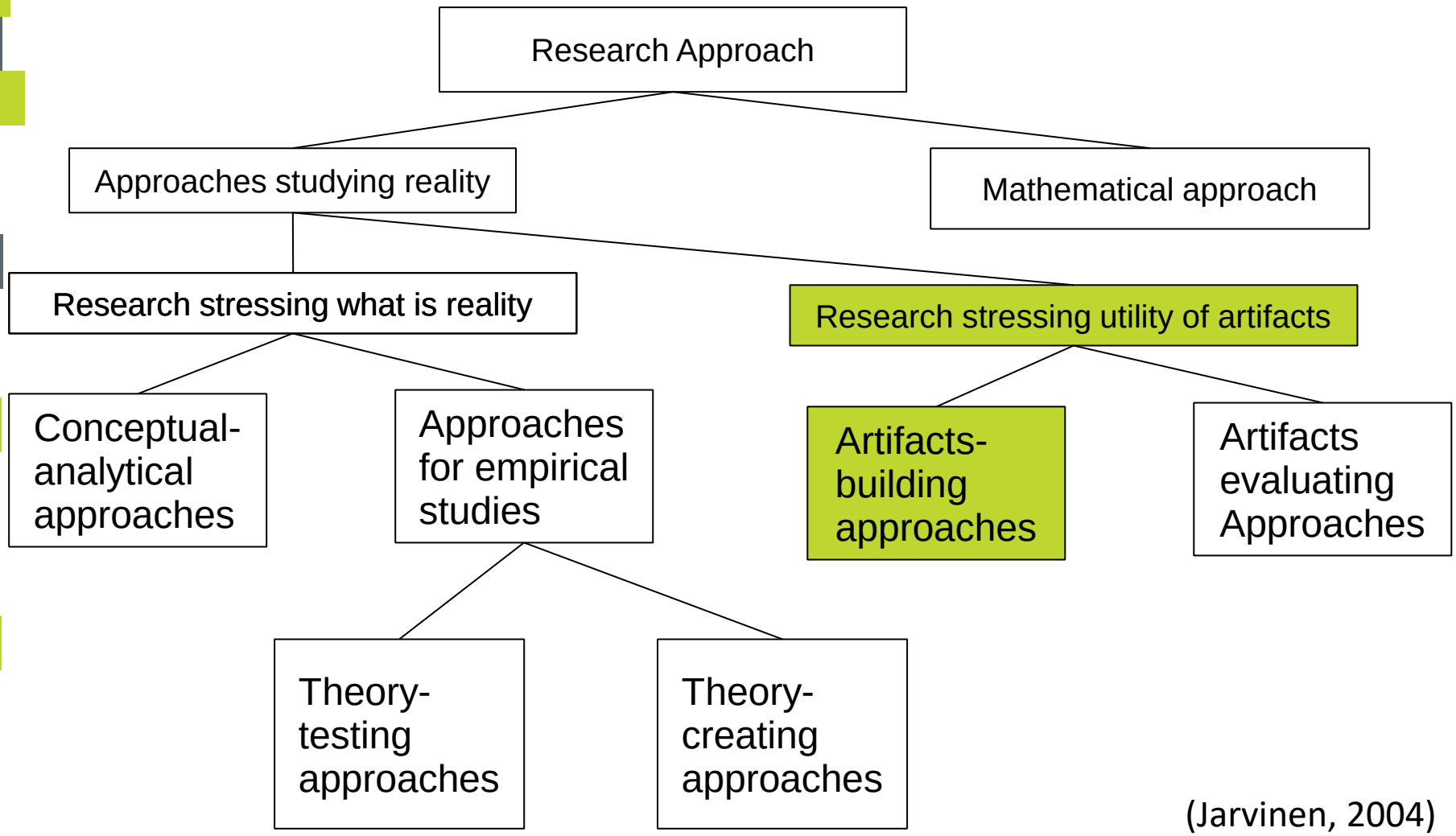
Models is a set of propositions or statements expressing relationships among constructs (abstractions and representations)

Methods is a set of steps (an algorithms, guidelines and practices) used to perform a task

Instantiations is the realization of an artifact in its environment (implemented and prototype system), verifying its feasibility & effectiveness

(March & Smith, 1995; Simon, 1996)

Research Output



(Jarvinen, 2004)

Artifacts-building approaches

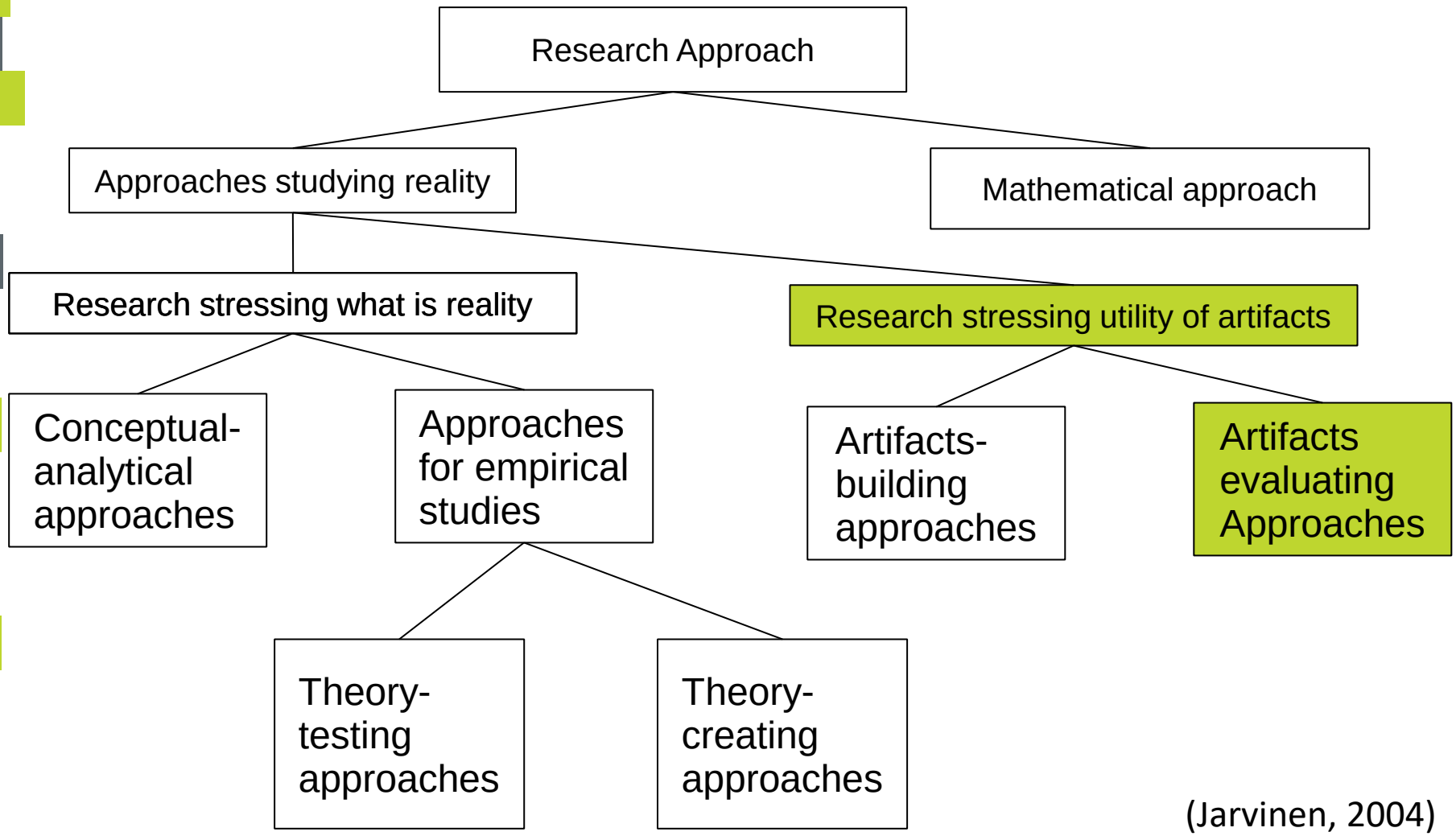
Research question could be:

- Is it possible building a certain artifact?

General output:

- A certain abstract or concrete artifact (predictive model, normative method or measurement instrument) is built.

Research Output



(Jarvinen, 2004)

Artifacts-evaluating approaches

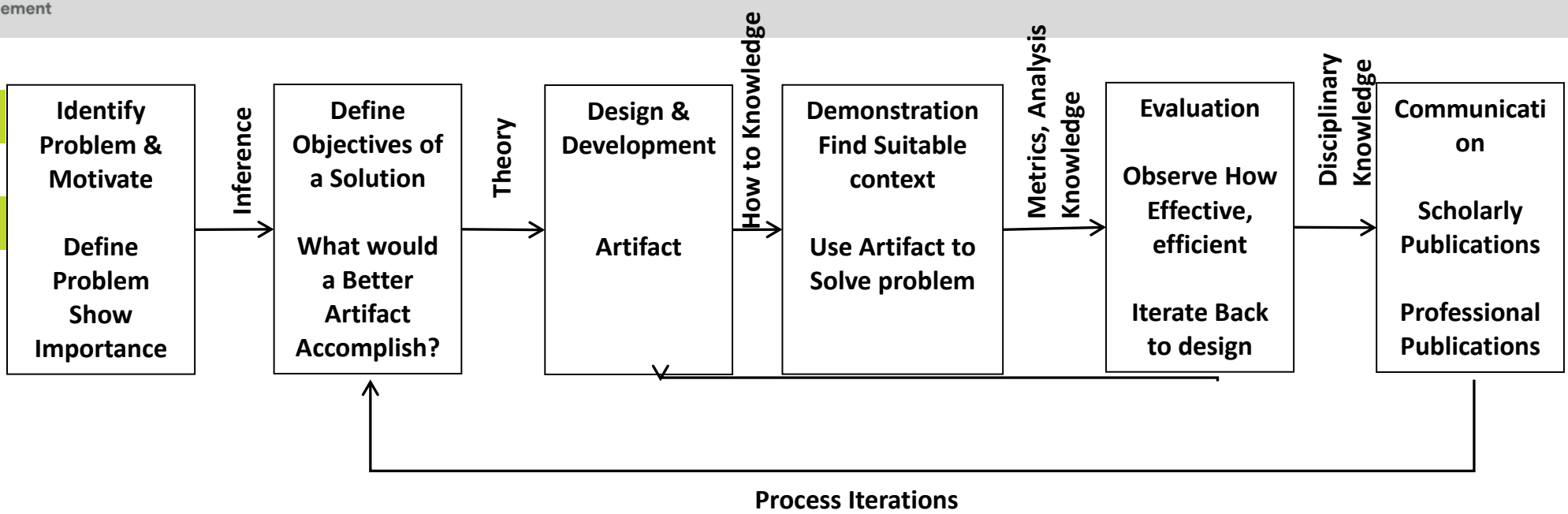
Research question could be:

- How effective is the artifact?

General Output:

- The utility (efficiency, effectiveness) of a certain artifact (or prescriptive model or normative method) is evaluated by using some criteria

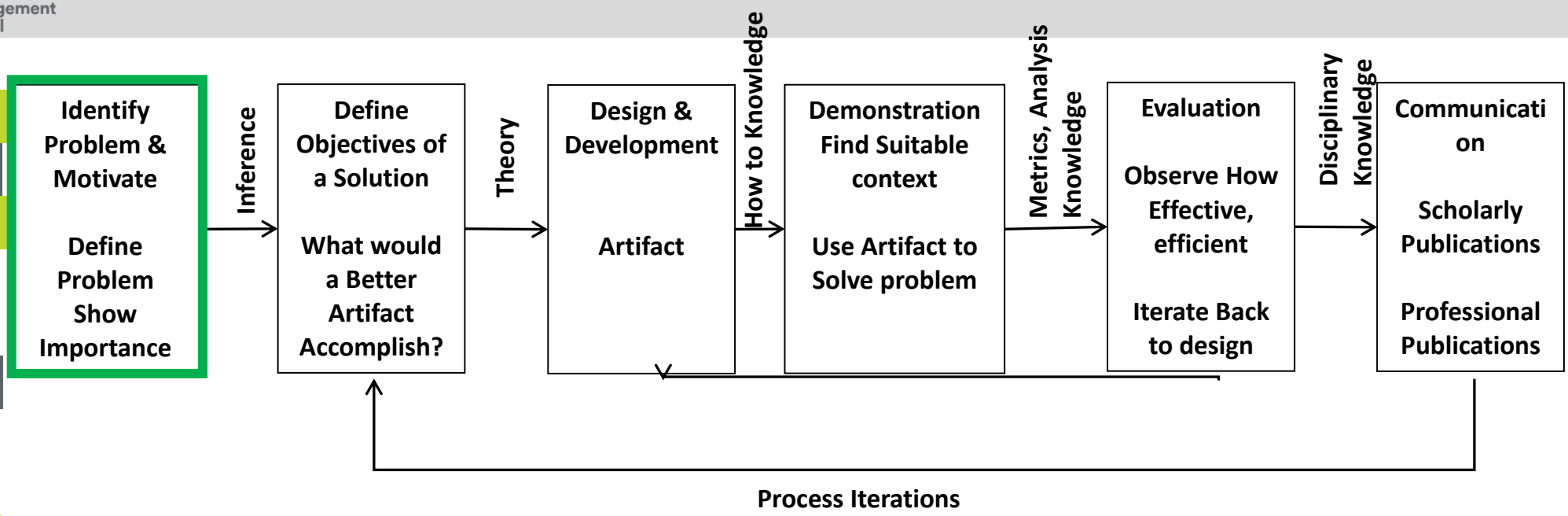
Design Science Research Process



Activity 1 | Activity 2 | Activity 3 | Activity 4 | Activity 5 | Activity 6

(Vaishnavi & Kuechler, 2004)

Design Science Research Process



Activity 1 | Activity 2 | Activity 3 | Activity 4 | Activity 5 | Activity 6

Activity 1. Problem identification and motivation

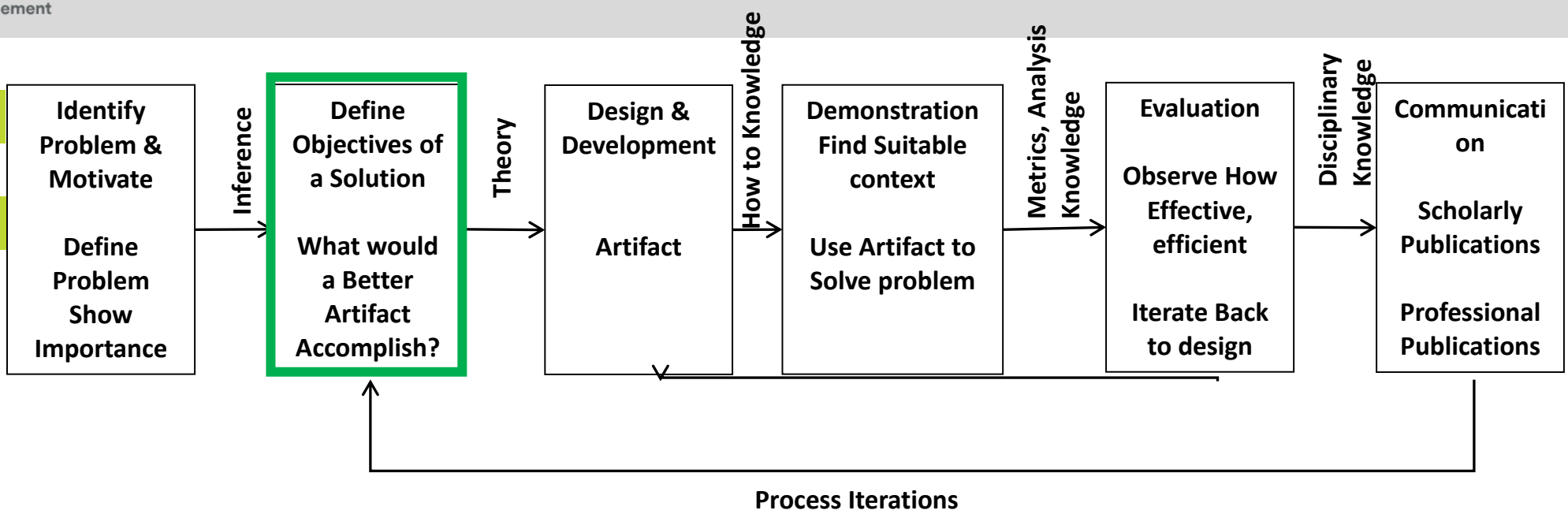
Define the specific research problem and justify the value of a solution

Well-defined problem can be used to develop an effective artifact

Justifying the value motivates the researcher to pursue the solution and accept the result

Resources required: Knowledge of the state of the problem and the importance of its solution

Design Science Research Process



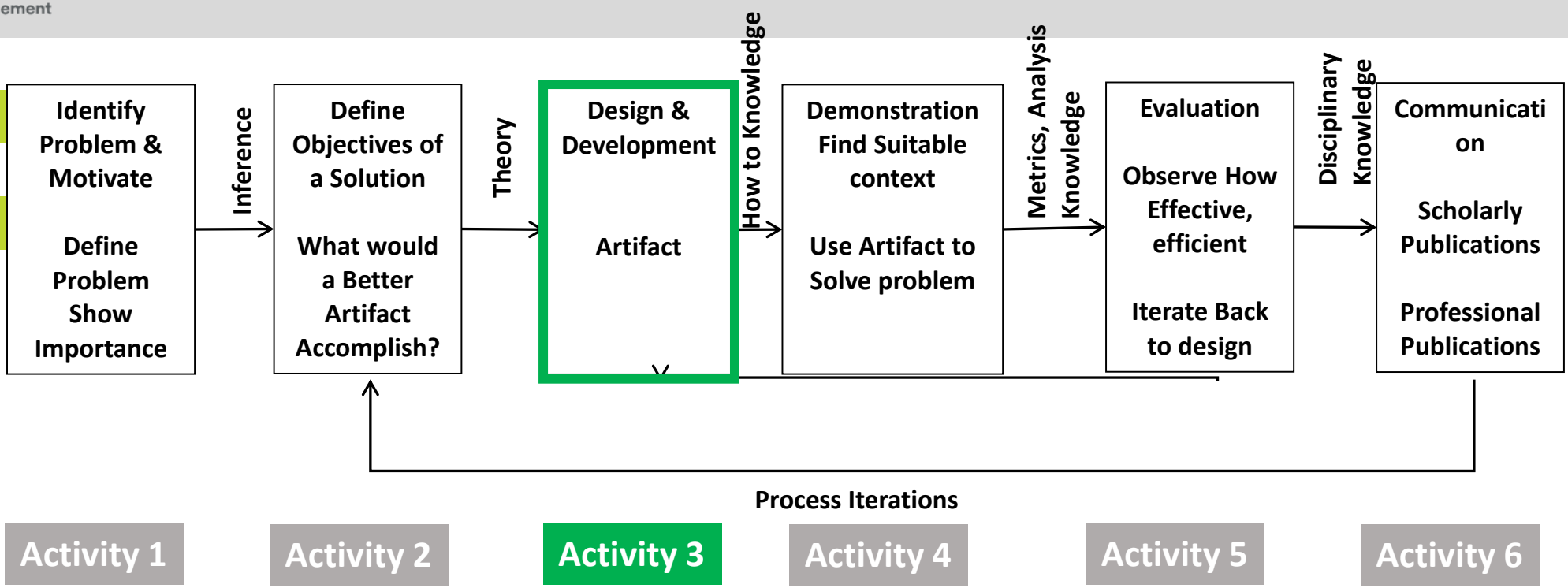
Activity 1 | **Activity 2** | Activity 3 | Activity 4 | Activity 5 | Activity 6

Activity 2. Define the objectives for a solution

Infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible (Metric objectives/Nonmetric objectives)

Resources required: Knowledge of the state of the problem and current solutions and their efficacy

Design Science Research Process

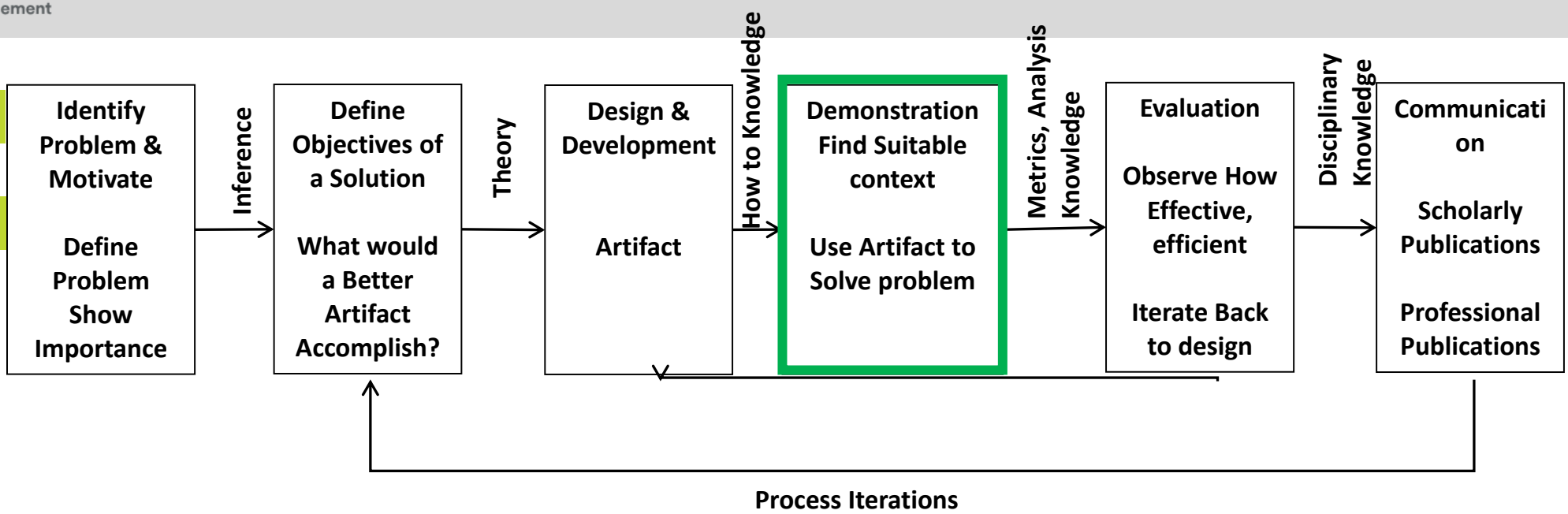


Activity 3. Design and development

Create the artifact, such like constructs, models, methods, or instantiations or "new properties of technical, social, and/or informational resources"

Resources required: Knowledge of theory that can be brought to bear in a solution

Design Science Research Process



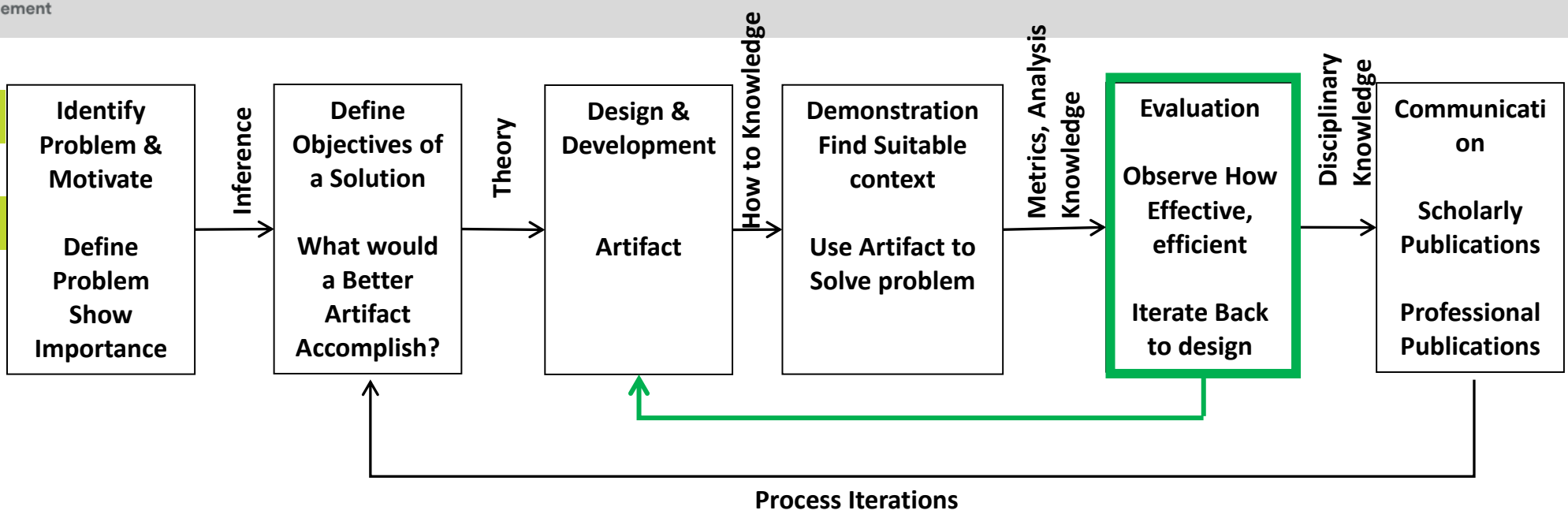
Activity 1 | Activity 2 | Activity 3 | **Activity 4** | Activity 5 | Activity 6

Activity 4. Demonstration

Demonstrate the use of the artifact to solve one or more instances of the problem

Resource required: Knowledge of how to use the artifact to solve the problem

Design Science Research Process



Activity 1 | Activity 2 | Activity 3 | Activity 4 | **Activity 5** | Activity 6

Activity 5. Evaluation

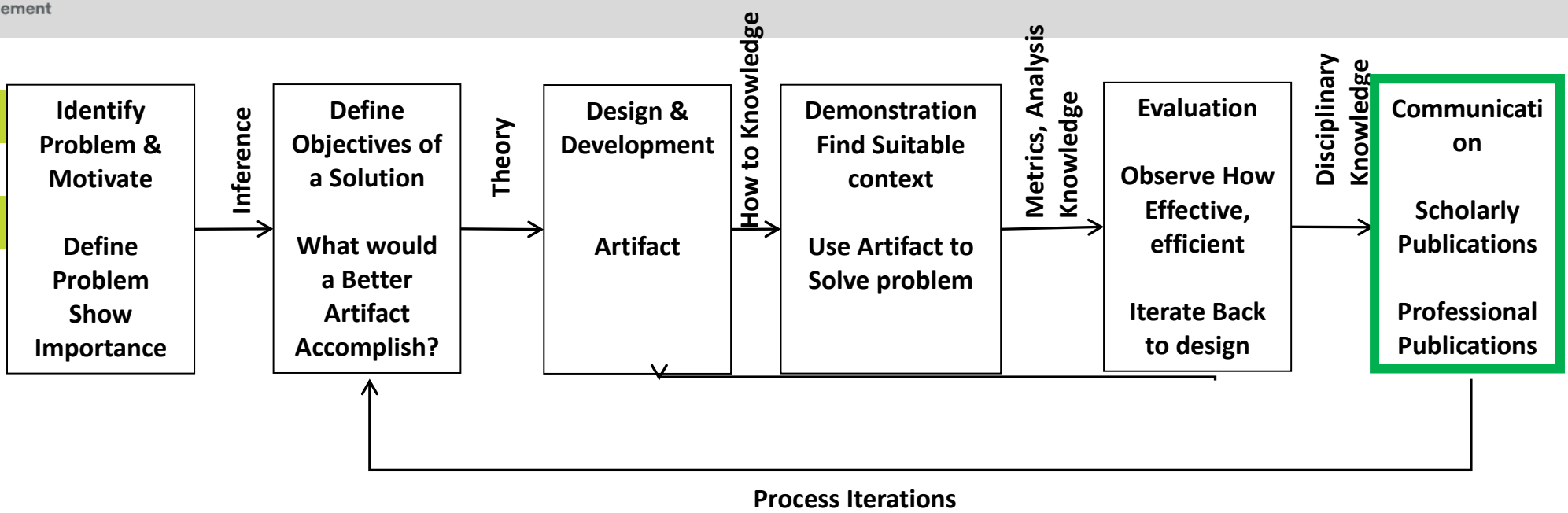
Observe and measure how well the artifact supports a solution to the problem

Comparing the objectives of a solution to actual observed results from use of the artifact in the demonstration

Resources required: Metrics and analysis techniques

Can be back to activity 3 to try to improve artifact

Design Science Research Process



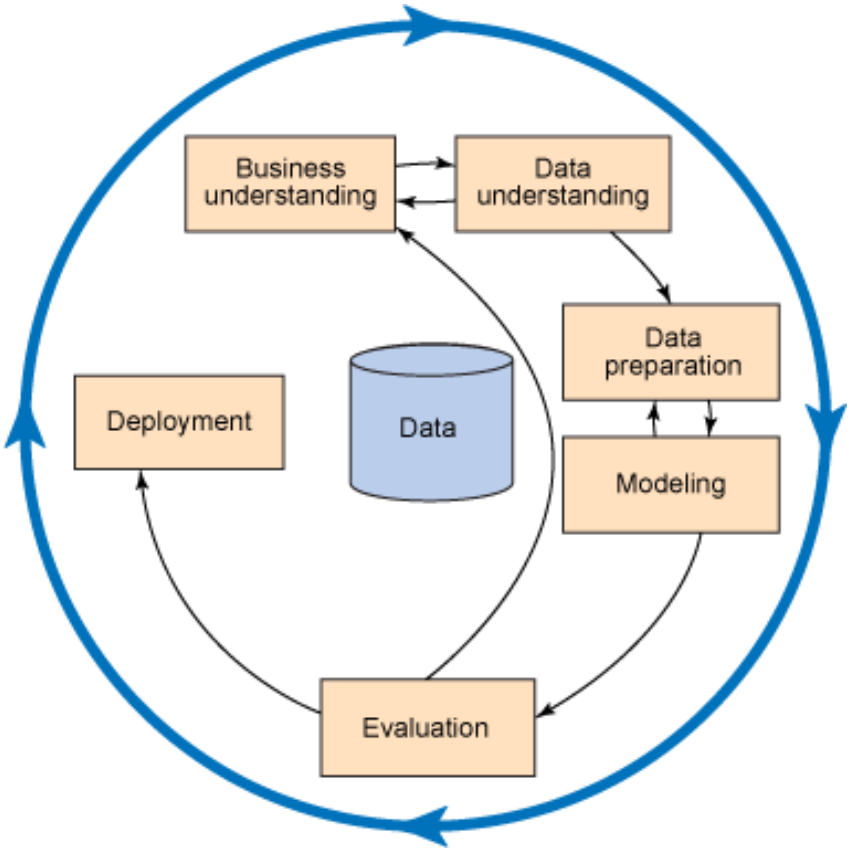
Activity 1 | Activity 2 | Activity 3 | Activity 4 | Activity 5 | Activity 6

Activity 6. Communication

Communicate the problem and its importance, the artifact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences when appropriate

Resource required: Knowledge of the disciplinary culture

CRISP-DM



- Data Science methodology
- Process Model
- For anyone
- Provides a complete blueprint
- Life cycle: 6 phases

Phases

- Business Understanding
  - Project objectives and requirements understanding, Data mining problem definition
- Data Understanding
  - Initial data collection and familiarization, Data quality problems identification
- Data Preparation
  - Table, record and attribute selection, Data transformation and cleaning
- Modeling
  - Modeling techniques selection and application, Parameters calibration
- Evaluation
  - Business objectives & issues achievement evaluation
- Deployment
  - Result model deployment, Repeatable data mining process implementation

Phases and tasks

Business Understanding	Data Understanding	Data Preparation	Modelling	Evaluation	Deployment
Determine Business Objectives	Collect Initial Data	Select data	Select modelling technique	Evaluate results	Plan deployment
Assess Situation	Describe Data	Clean data	Generate test design	Review process	Plan monitoring and maintenance
Determine Data Mining Goals	Explore data	Construct data	Build model	Determine next steps	Produce final report
Produce project plan	Verify data quality	Integrate data	Assess model		Review project
		Format data			

Business Understanding

- Determine business objectives
  - thoroughly understand, from a business perspective, what the client really wants to accomplish
  - uncover important factors, at the beginning, that can influence the outcome of the project
  - neglecting this step is to expend a great deal of effort producing the right answers to the wrong questions

- Assess situation
  - more detailed fact-finding about all of the resources, constraints, assumptions and other factors that should be considered
  - flesh out the details

Business Understanding

- Determine data mining goals
  - a business goal states objectives in business terminology
  - a data mining goal states project objectives in technical terms
    - “Increase catalog sales to existing customers.” vs “Predict how many widgets a customer will buy, given their purchases over time”
- Produce project plan
  - describe the intended plan for achieving the data mining goals and the business goals
  - the plan should specify the anticipated set of steps to be performed during the rest of the project including an initial selection of tools and techniques

Data Understanding

- Collect initial data
  - acquire within the project the data listed in the project resources
  - includes data loading if necessary for data understanding
  - possibly leads to initial data preparation steps
  - if acquiring multiple data sources, integration is an additional issue, either here or in the later data preparation phase

- Describe data
  - examine the “gross” or “surface” properties of the acquired data
  - report on the results

Data Understanding

- Explore data
  - tackles the data mining questions, which can be addressed using querying, visualization and reporting including:
    - distribution of key attributes, results of simple aggregations
    - relations between pairs or small numbers of attributes
    - properties of significant sub-populations, simple statistical analyses
  - may address directly the data mining goals
  - may contribute to or refine the data description and quality reports
  - may feed into the transformation and other data preparation needed
- Verify data quality
  - examine the quality of the data, addressing questions such as:
    - “Is the data complete?”, Are there missing values in the data?”

Data Preparation

- Select data
  - decide on the data to be used for analysis
  - criteria include relevance to the data mining goals, quality and technical constraints such as limits on data volume or data types
  - covers selection of attributes as well as selection of records in a table

- Clean data
  - raise the data quality to the level required by the selected analysis techniques
  - may involve selection of clean subsets of the data, the insertion of suitable defaults or more ambitious techniques such as the estimation of missing data by modeling

Data Preparation

- Construct data
  - constructive data preparation operations such as the production of derived attributes, entire new records or transformed values for existing attributes
- Integrate data
  - methods whereby information is combined from multiple tables or records to create new records or values
- Format data
  - formatting transformations refer to primarily syntactic modifications made to the data that do not change its meaning, but might be required by the modeling tool

Modelling

- Select modeling technique
  - select the actual modeling technique that is to be used
  - ex) decision tree, neural network
  - if multiple techniques are applied, perform this task for each techniques separately

- Generate test design
  - before actually building a model, generate a procedure or mechanism to test the model's quality and validity
  - ex) In classification, it is common to use error rates as quality measures for data mining models. Therefore, typically separate the dataset into train and test set, build the model on the train set and estimate its quality on the separate test set

Modelling

- Build model
  - run the modeling tool on the prepared dataset to create one or more models
- Assess model
  - interprets the models according to his domain knowledge, the data mining success criteria and the desired test design
  - judges the success of the application of modeling and discovery techniques more technically
  - contacts business analysts and domain experts later in order to discuss the data mining results in the business context
  - only consider models whereas the evaluation phase also takes into account all other results that were produced in the course of the project

Evaluation

- Evaluate results
  - assesses the degree to which the model meets the business objectives
  - seeks to determine if there is some business reason why this model is deficient
  - test the model(s) on test applications in the real application if time and budget constraints permit
  - also assesses other data mining results generated
  - unveil additional challenges, information or hints for future directions

Evaluation

- Review process
  - do a more thorough review of the data mining engagement in order to determine if there is any important factor or task that has somehow been overlooked
  - review the quality assurance issues
    - “Did we correctly build the model?”

- Determine next steps
  - decides how to proceed at this stage
  - decides whether to finish the project and move on to deployment if appropriate or whether to initiate further iterations or set up new data mining projects
  - include analyses of remaining resources and budget that influences the decisions

Deployment

- Plan deployment
  - in order to deploy the data mining result(s) into the business, takes the evaluation results and concludes a strategy for deployment
  - document the procedure for later deployment

- Plan monitoring and maintenance
  - important if the data mining results become part of the day-to-day business and it environment
  - helps to avoid unnecessarily long periods of incorrect usage of data mining results
  - needs a detailed on monitoring process
  - takes into account the specific type of deployment

Deployment

- Produce final report
  - the project leader and his team write up a final report
  - may be only a summary of the project and its experiences
  - may be a final and comprehensive presentation of the data mining result(s)

- Review project
  - assess what went right and what went wrong, what was done well and what needs to be improved

References

- ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/en/CRISP_DM.pdf
- http://www.kdnuggets.com/
- https://mineracaodedados.files.wordpress.com/2012/04/the-crisp-dm-model-the-new-blueprint-for-data-mining-shearer-colin.pdf


My thesis is written in
ow!
Blood
Sweat
what am i doing with my life??
Tears
JORGE CHAM © 2013
and COFFEE.
WWW.PHDCOMICS.COM
!
?
#@$%.
JORGE CHAM © 2012
WWW.PHDCOMICS.COM

Main Sources