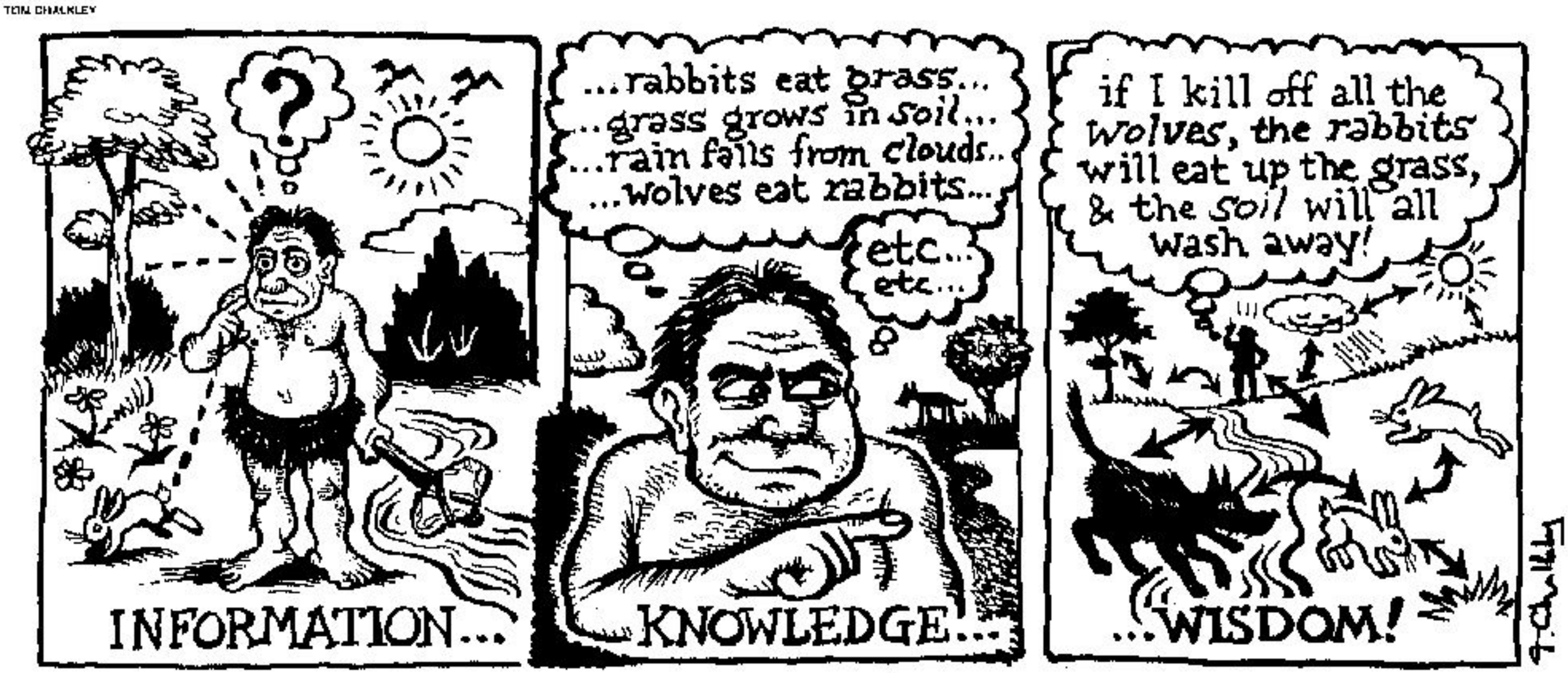
Data – information – knowledge (D-I-K)

prof, dr sc, inż Oleg Zaikin
dr sc, prof. Emma Kusztina
dr sc, prof. Przemyslaw Rozewski

Wydział Informatyki
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
ul. Żołnierska 49, 71-230 Szczecin

Warsaw, 2010

Data – information – knowledge



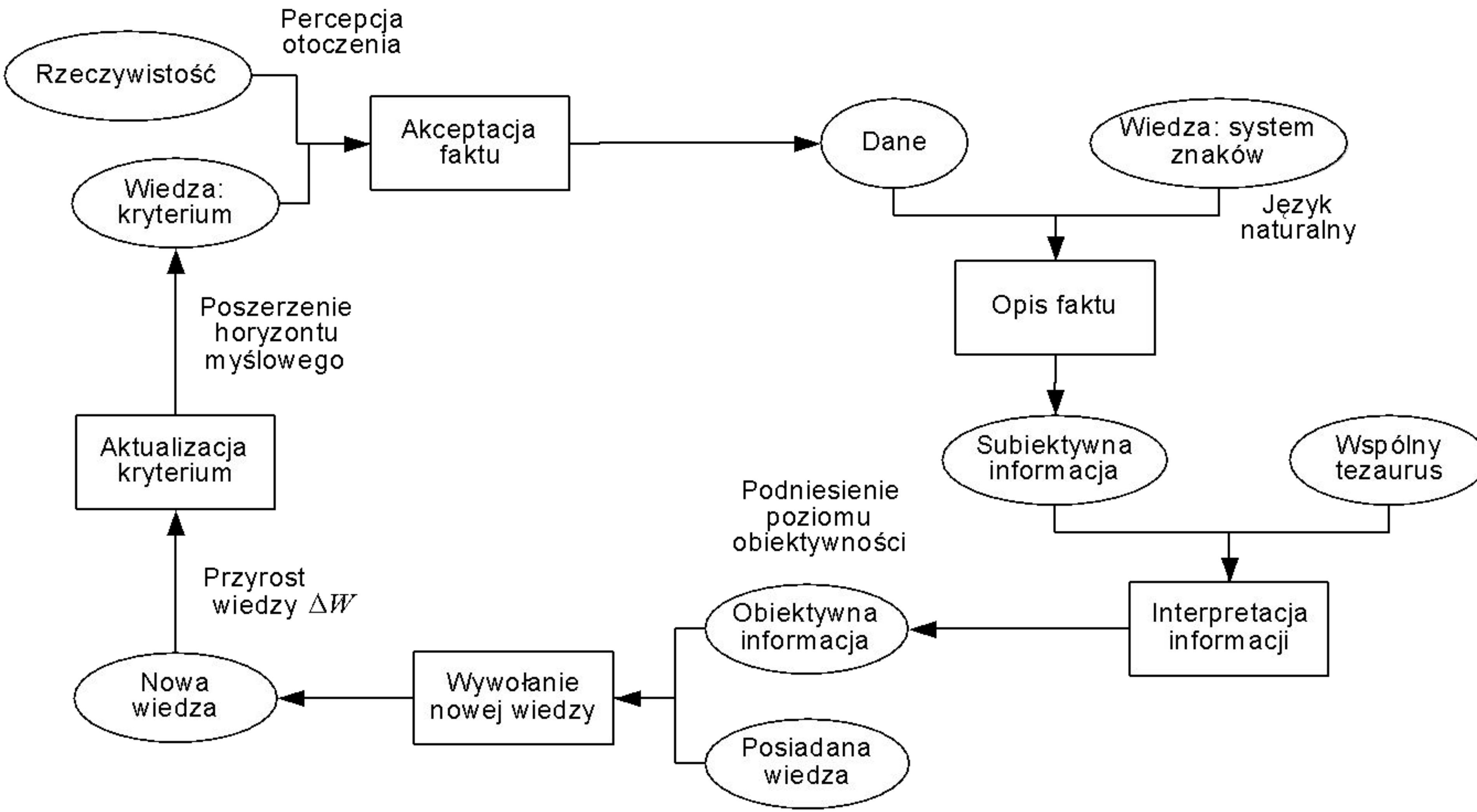
Example

- 1234567.89 is given as data;
- "Your account status changed by 8087% to 1234567.89" is information;
- "No one is so big debtor for me" is knowledge;
- And to finish the discussion we can add that "I better contact the bank before issuing this sum" which is already an example of human wisdom.

Introductory statesments

- ***Informatics*** states the aim the modeling and control of the process represented by a chain: *data - information - knowledge.*
- The following generation of information systems is an attempt at analyzing and decomposing a chain into separate parts - finding indices and criteria that allow them to divide accurately.
- Formalization and subsequent automation of operations by means of cyclic transformation of individual parts of the chain is the development of information systems.

Sequence D-I-K



Źródło: opracowanie własne

Definitions D-I-K

Author(s)	Data	Information	Knowledge
Wiig [41]	-	Facts organised to describe a situation or condition	Truths and beliefs, perspectives and concepts, judgements and expectations, methodologies and know.how
Nonaka and Takeuchi [23]	-	A flow of meaningful messages	Commitments and beliefs created from these messages
Spek and Spijkervet [32]	Not yet interpreted symbols	Data with meaning	The ability to asing meaning
Davenport [15]	Simple observations	Data with relevance and purpose	Valuable information from the human mind
Davenport and Prusak [16]	A set of descrete facts	A message meant to change the receiver's perception	Experiences, values, insights, and contextual information
Quigley and Debons [28]	Text that does not answer questions to a particular problem	Text that answers the questions who, when, what, or where	Text that answers the questions why and how
Choo *et al.* [12]	Facts and messages	Data vested with meaning	Justified, true beliefs

Definitions

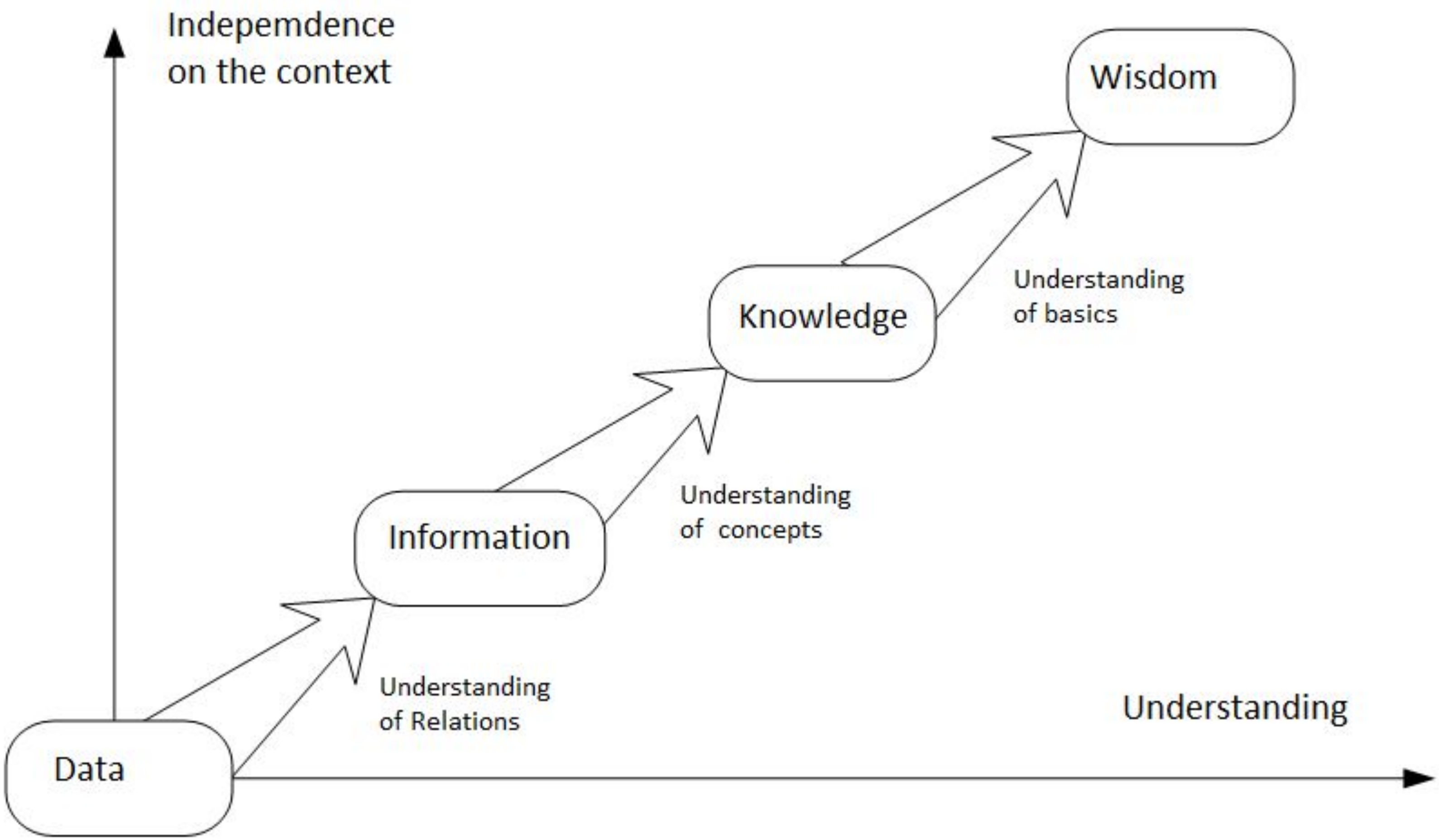
Data - this is an object or event that has no context or relationships to other elements or events

Information - is represented by the relationship between data and possible other information

Knowledge - is represented by a pattern between data, information and possible other knowledge. A given pattern is not knowledge before it is understood.

Wisdom - This is the realization that knowledge patterns come from fundamental principles and understanding what these principles are.

The process of developing wisdom depends on the dimension of understanding and the context



Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX., .

Data

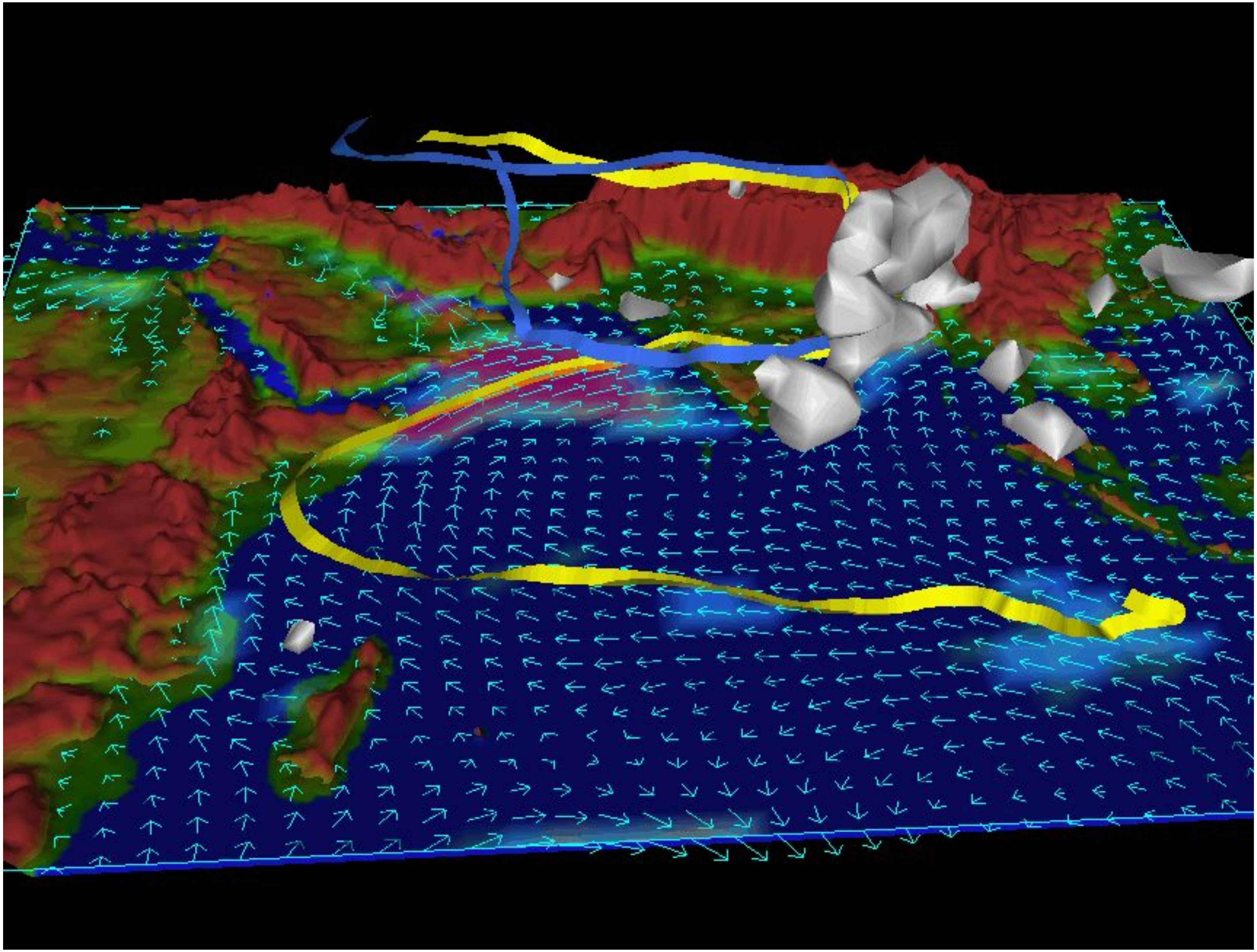
- Data is defined as: "raw" material from which information is extracted (use of extraction operations) .
- Chiew (2002) defines data as "raw" pieces of abstract elements and things.
- The information is defined as: data that has been assigned attributes along with limited relationships between data.
- Bryant (2003) concludes that the only rational definition of data is something that is stored as an object.

Data Acquisition (observer)

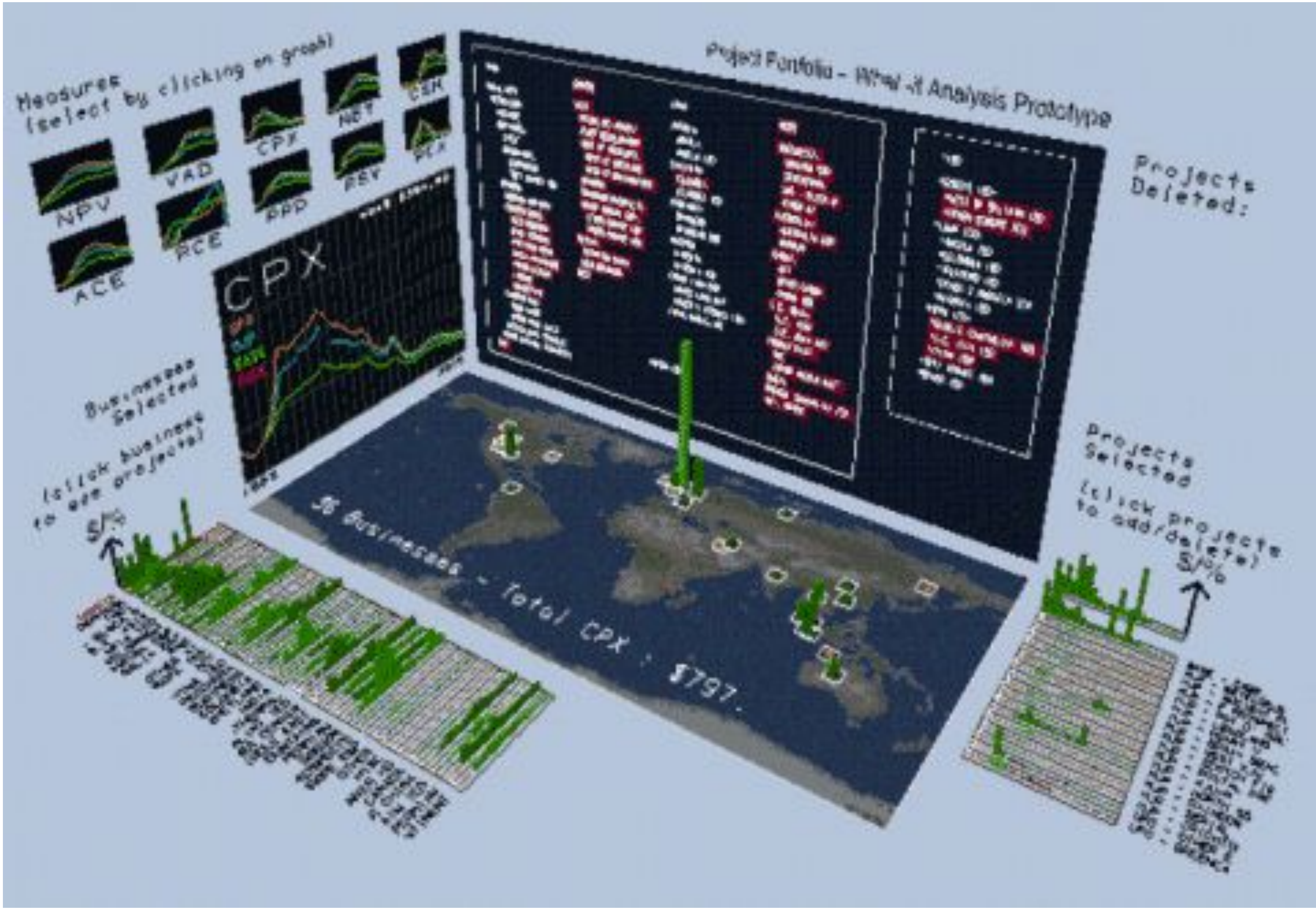
- The most important data-related operation is a data acquisition operation - understood as determining the boundaries of an object based on the prepared procedure.
- We base in this case on **the task of observer** described in philosophy and physics, which analyzes **the objectivity of observations** made by the observer in relation to the system.
- An observer is a model of a subject learning to collect data from a test system that **uses measurement or observation as the primary method of data acquisi**tion.

Data properties

- Data can be obtained either as a result of routine or ad hoc procedures in an automatic or "manual" manner.
- Data that is subjective or objective depending on the measurement method.
- **High quality data** allows for a high degree of comparability, meaning "referring to data of the same meaning, that is, the same definitions".
- Equally important is the provision of a **high degree of representativeness** which allows for the generalization of the expression of specific data to a population larger than the population studied.
- Data Visualization



Sourse: Indian monsoon, water vapor tracers; Source: NASA Data Assimilation Office



Źródło: EFFECTIVE INFORMATION VISUALIZATION Guidelines and Metrics for 3D Interactive Representations of Business Data
http://www3.sympatico.ca/blevis/thesis49observations.html

D

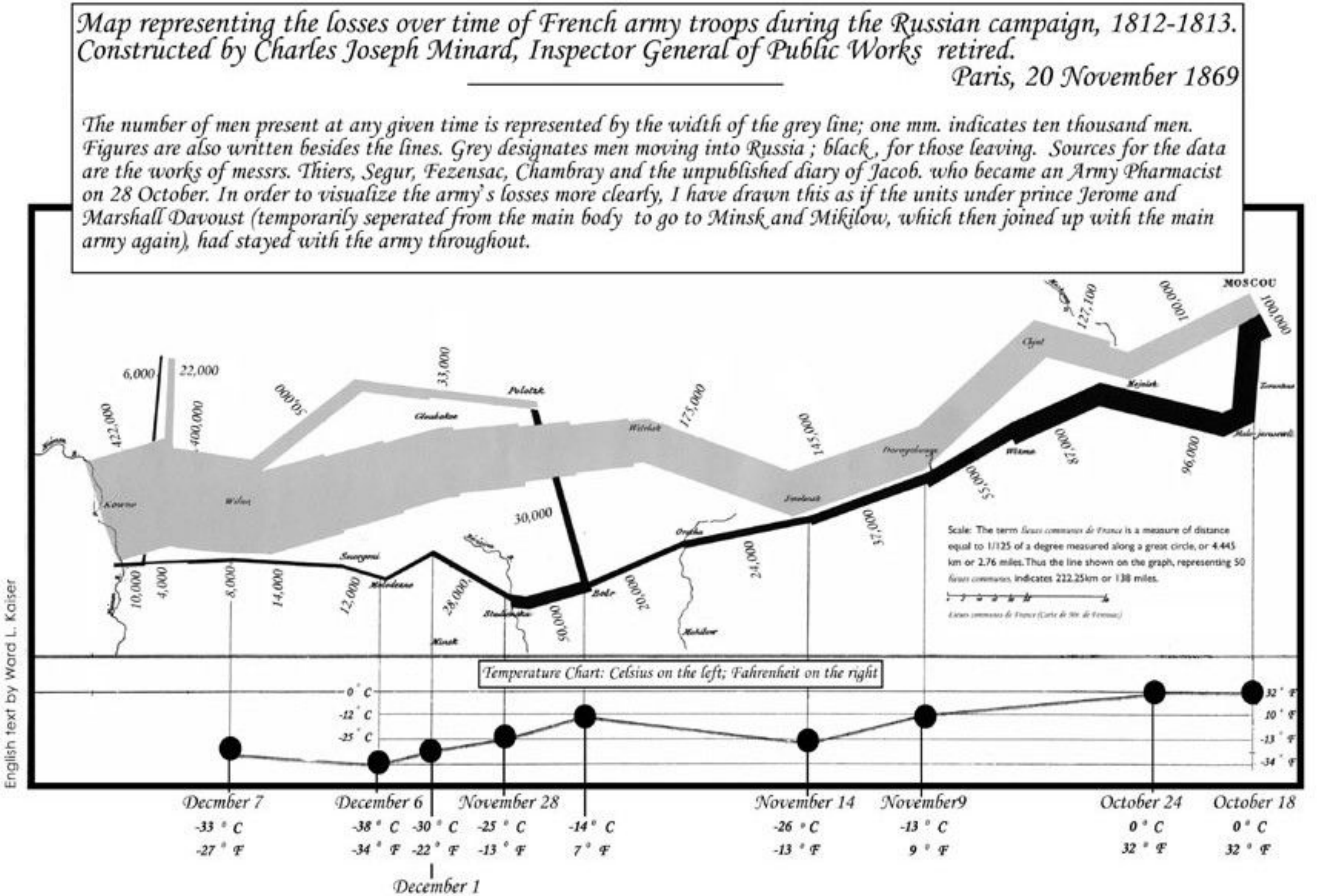


Figure 58. ***Minard s map of Napoleon s Russian campaign.***
This graphic has been translated from French to English and modified to most effectively display the temperature data.

Data as a research object

The definition of data that is treated as a research object consists of the following elements:

- Subject of data
- Used units
- The method used to extract data and its characteristics, time, place, etc.

Semistructural data

- Semi-structured data consciously ignore the serialization process (understood as processing data into the bit stream).

- In semi-structured data, otherwise known as self-describing data, the value is stored with the corresponding description.
- {name: "Jan", age: "33", phone: "4223424"}

- **Advantages**: value association with the right description, data independence from the format of representation

- **Disadvantages:** the most important thing is to increase the demand for the required space memory (data compression can be used which greatly reduces this inconvenience).

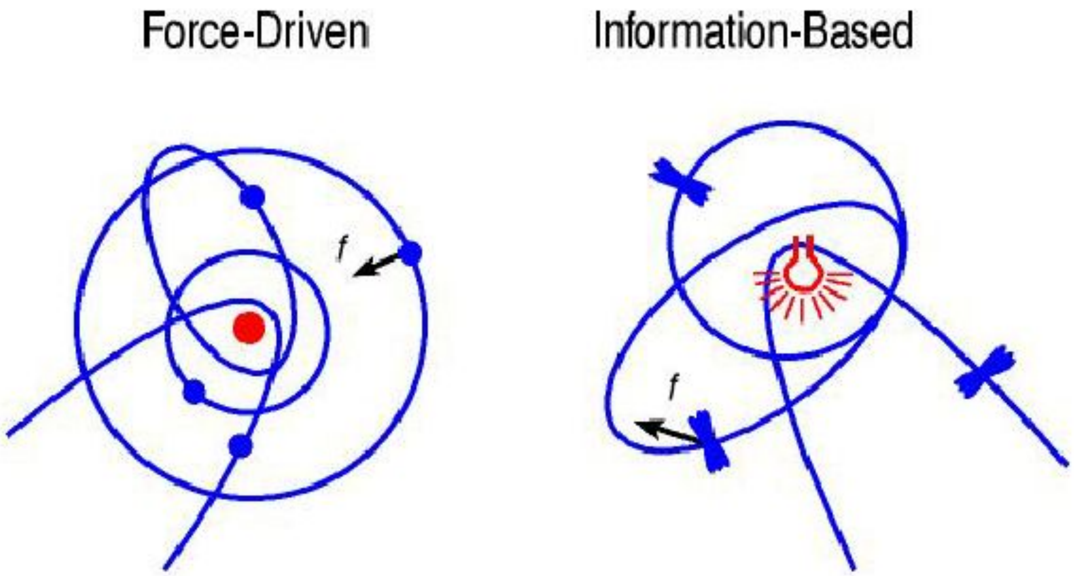
Information

- The word „**Information**" comes from the Latin **informare**, which means "to form". Etymologically information is the creation of a certain structure in a certain indeterminate chaos.
- Information has all the physical qualities, ie:
- (i) the information can be measured: there is a method that allows us to calculate the volume that we call the amount of information,
- (ii) the information is objective: the result of measuring the amount of information does not depend on other factors.

Features information (Wang, 2003)

- Information is an abstract artifact: Information is created by observing physical elements, building relationships between physical or abstract objects. Artefacts are created intentionally and their meaning is usually built on the basis of context.
- Information is not subject to the laws of physics: Based on physics, matter and energy can not be destroyed or amplified, only transformation from one state to another (the second law of thermodynamics) is possible. Information may, however, be destroyed, duplicated or merged. Accumulation of information allows for its continuous evolution.
- Infinite Usability: Information without quality loss can be used by many different users an infinite number of times.
- Information has no dimension: information does not have a physically meaningful spatial dimension. No matter how big or small the physical object is, the information counter is dealing with a similar frame, which may differ from another frame by resolution only.
- Information has no weight: the physical weight of information is always zero. An empty or filled floppy disk weighs the same, the information contained therein has no bearing on the physical weight.
- Multiple possible forms of representation: information can be represented in different forms: analog (eg audio), abstract (eg spoken and written language), digitally (eg xml file). The most important is a digital representation that stores information in discrete form. Digital representation enables information to be effectively stored and processed.
- The number of possible forms of transmission: information can be transmitted in the following modes: 1-1 (transmission), 1-n (broadcasting), n-1 (infiltration), and n-m (infraction).
- Generic Information Sources: Every object in an investigated universe can generate information.

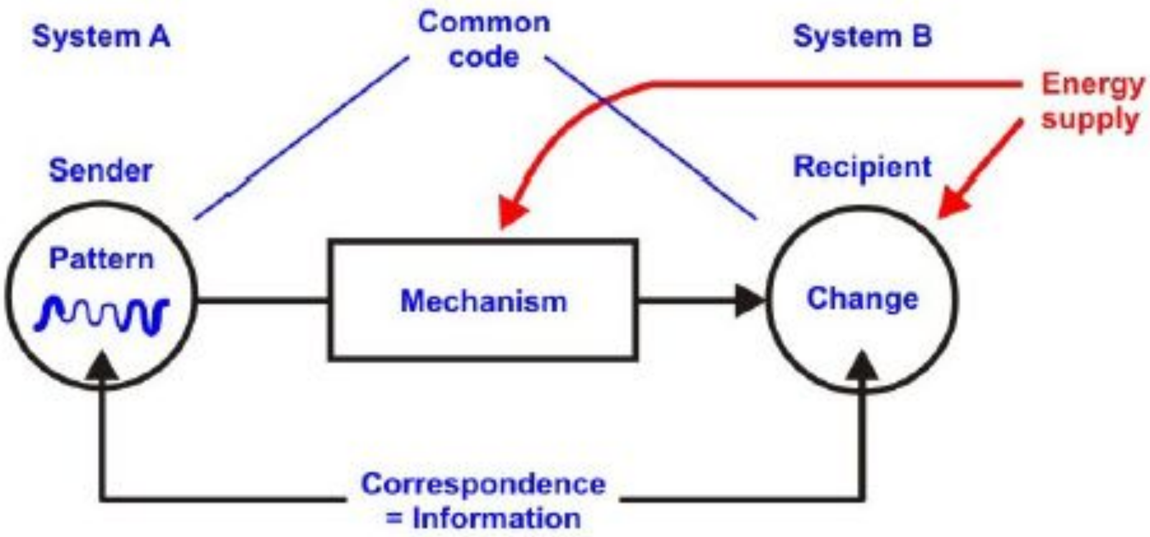
Information-based interactions



Examples of the two main types of interactions: force-field driven (satellites in orbit around a central body, left), and information-based (insects “in orbit” around a light source, right).

Information and information-processing play no role in the former, whereas in the latter we have the chain light emission -> pattern detection -> pattern analysis -> muscle activation, in which neither force nor energy but information is the controlling agent throughout.

Źródło: Juan G. Roederer (2003), On the **Concept of Information and Its Role in Nature**, Entropy 2003, 5[1], 3-33



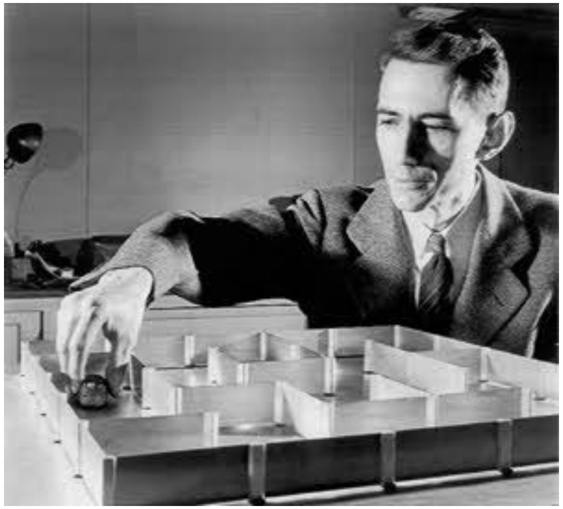
In an information-based interaction a *correspondence* is established between a pattern in the “sender” and a specific change (structural or dynamic change) in the “recipient”.

Information is the agent that represents this correspondence. The pattern could be a given spatial sequence of objects (e.g., chemical radicals in a molecule), a temporal sequence (e.g., the phonemes of speech), or a spatiotemporal distribution of events (e.g., electrical impulses in a given region of the brain).

The mechanism for a natural (not artificially made) information-based interaction must either emerge through evolution or be developed in a learning process, because it requires a common code (a sort of memory device) that could not appear by chance. There is no direct energy transfer between the sender and the recipient, although energy, to be supplied externally, is involved in all intervening processes.

Źródło: Juan G. Roederer (2003), On the **Concept of Information and Its Role in Nature**, Entropy 2003, 5[1], 3-33

Information Theory of Shannon

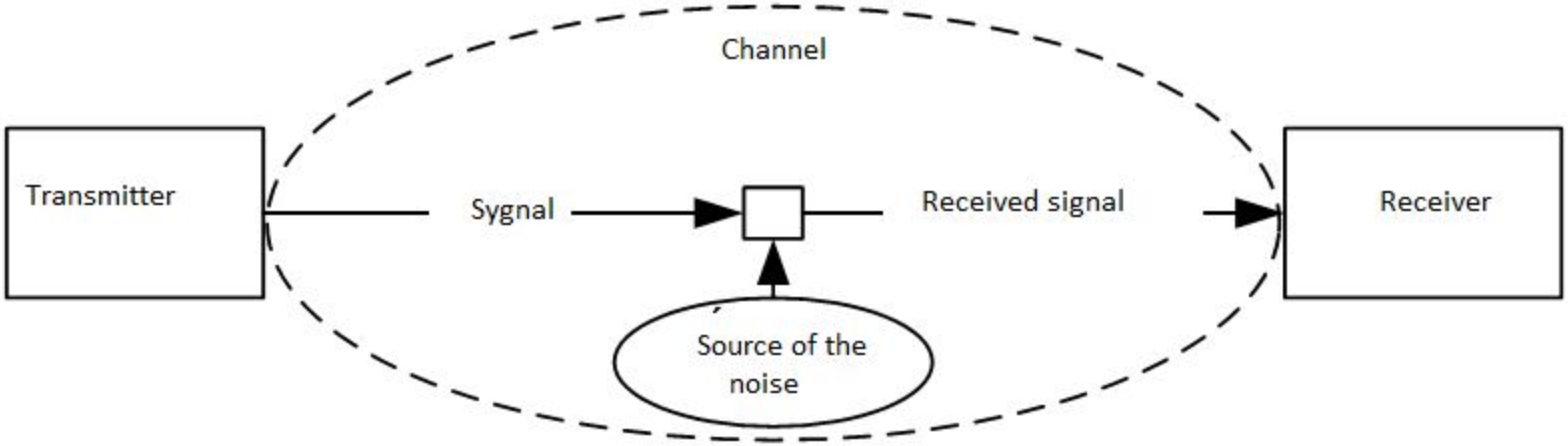


In the 1940s, there was a need for a coherent theory to analyze the information transmitted in the form of electrical signals via telecommunication lines.

The advancement of technology related to the construction of efficient transmitters and receivers has allowed for a new level of quality and transmission speed, which has led to difficulties such as:

- determining the degree of maximum use of the telecommunication channel or
- determining the degree of data compression.

.



- Shannon built up a transmission channel model within which the information is sent.
- The source (S) transmitter transmits information from the transmitter to the receiver as a destination for the information.
- The transmission channel is not an ideal medium for transmitting lossless signals; the noise is infiltrated into the information due to the physical characteristics of the track.
- In the analyzed model the information is treated as a message, which is characterized by its value but does not have such qualitative characteristics as semantic features of information.

Information content of the message

Each message is characterized by a load of information determined by the information content of the message and expressed in bits. According to Shannon, the expected message provides us little information, while the surprise message is characterized by a large amount of information.

In addition, we may give you a chance to guess the probability of a particular message. For the expected message, the probability P will be high, but for an unexpected message the probability P will be low. The relationship between I and P is as follows:

$$I = \log_2 \frac{1}{P}$$

Information content of the message

If we assume that the source S (represented by the transmitter) has a set of possible **states** whose **probability of occurrence** is then the information content generated by the source by the occurrence of the state is:

$$I(s_i) = \log_2 \frac{1}{p(s_i)}$$

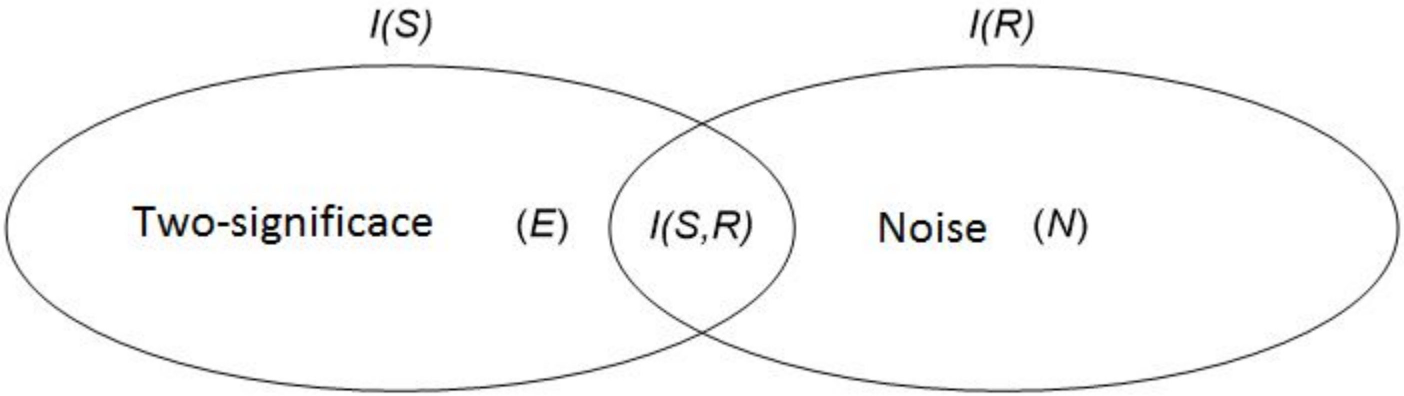
Information content of the message

The communication process is usually treated as a whole, so all messages generated by the source are treated. In this case we can calculate the average content of information generated by source I (S) according to the formula:

$$I(S) = \sum p(s_i) I(s_i) = \sum p(s_i) \log_2 \frac{1}{p(s_i)}$$

Similarly for the receiver

$$I(r_i) = \log_2 \frac{1}{p(r_i)} \qquad I(R) = \sum p(r_i) I(r_i) = \sum p(r_i) \log_2 \frac{1}{p(r_i)}$$



In the expression the variable ***E*** is the **ambiguity of the information**, which is interpreted as the **average value of the information generated by the source *S* and not received by the receiver *R***.

Likewise, the variable ***N*** means noise and is interpreted as the **average value of information received by the receiver *R* but not generated by the source *S***.

$$I(S,R) = I(S) - E = I(R) - N$$

Calculating the values ***N*** and ***E*** requires consideration of the characteristics of the telecommunication channel. A channel, understood as a message transfer medium, is the cause of transmission errors.

The unfavorable properties of the channel s_i are expressed in the form of a matrix $\lfloor p(r_j \mid s_i) \rfloor$ where $p(r_j \mid s_i)$ is a conditional probability of the event r_j provided the event s_i

.

$$E = \sum p(r_j) \sum p(s_i \mid r_j) \log_2 \frac{1}{p(s_i \mid r_j)}$$

$$N = \sum p(s_i) \sum p(r_j \mid s_i) \log_2 \frac{1}{p(r_j \mid s_i)}$$

Entropy

A system with a high entropy value is more likely to have a low degree of ordering, whereas a highly ordered system has a low entropy value.

Entropy is the average amount of information per symbol representing the occurrence of an event from a certain set. Events in this set are assigned the probability of occurrence.

Pattern for entropy:

$$H(x) = \sum_{i=1}^{n} p(i) \log_r \frac{1}{p(i)} = -\sum_{i=1}^{n} p(i) \log_r p(i)$$

Systems with high entropy values have a low degree of ordering. The greater the freedom of choice, the better the quality of information. So there is a greater probabilitty that there is some kind of information in the series of random symbols than when the series has some unexpected structure. Surprising us information is carrying a lot of information, the expected message provides us with little information, Simmonds (1999).

Shannon's theory- summary

- The basic task of communication is to accurately or approximatly reproduce a message in a certain place, which has been selected elsewhere to be transmitted. Often messages have content ie refer to a system that has a physical or mental meaning. These semantic aspects of the message do not refer the technical side of the issue.
- It is important only that the message being sent is the message selected from a certain set of messages. The communication system should be designed so that it can be used to transmit any possible message, not just the one that will actually be selected, as the result of this choice is not known at the time of design. ... "Shannon (1948).

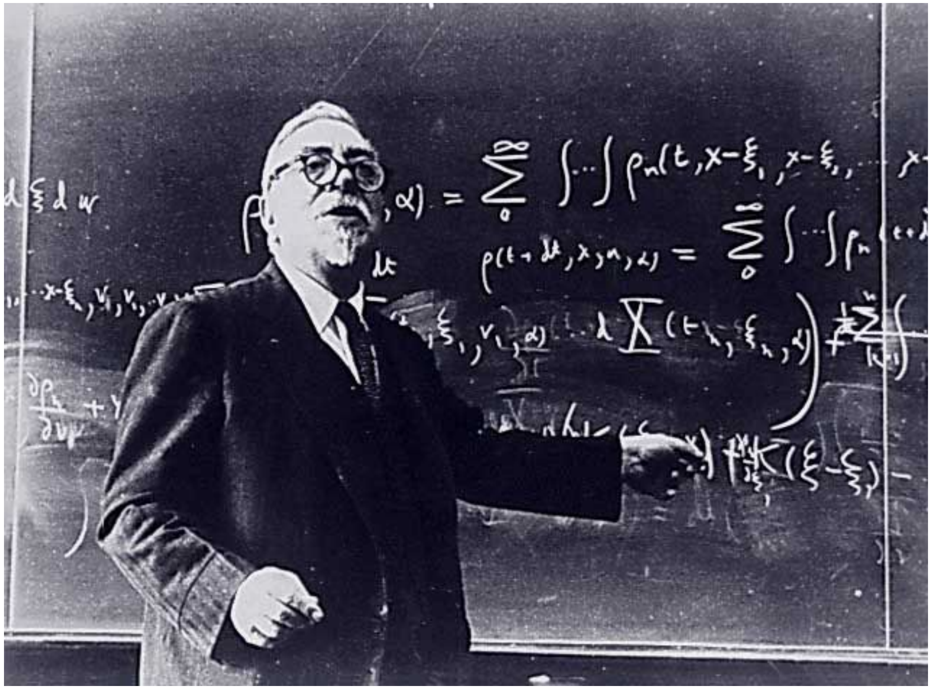
Cybernetics

***Cybernetics** (from grees word ‚kybernetes') - learning about control systems and related processing and communication*

Cybernetics is a science which

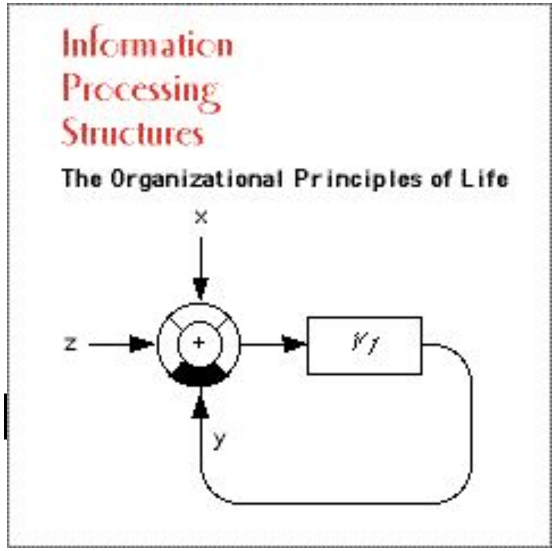
- analyzes analogues (homologies) between the principles of living organisms, social systems (societies) and machines (holism)
- discovers general laws common to various sciences and enables the transfer of these rights from one domain to another;

It is therefore an interdisciplinary science, which has many practical applications.



Norbert Wiener
Cybernetics or Control and Communication in the Animal and the Machine (1948)

Cybernetics - information



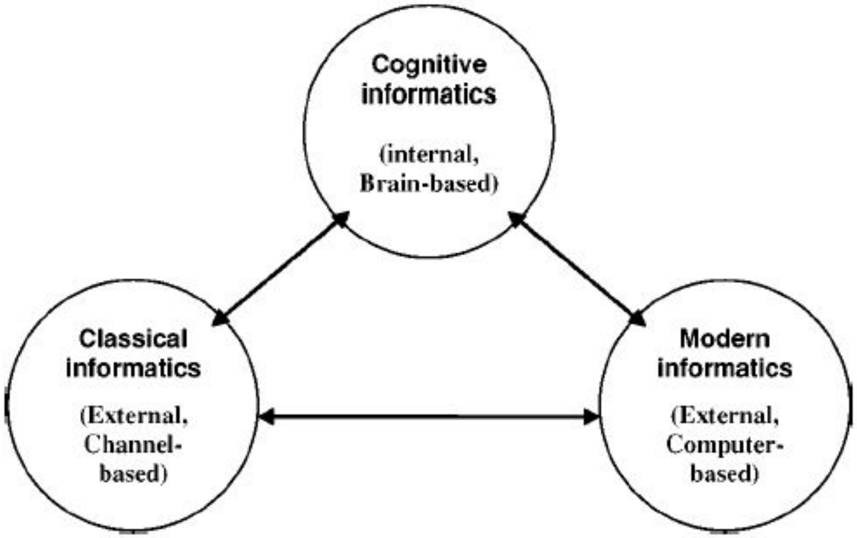
According to Wiener, the information is "content taken from the outside world in our process adapting to him.

Another theorist of cybernetics Couffignal (1963) defines the information in cybernetics as any action accompanied by physical action

- Information is a set of media and semantics, where semantics is understood as the psychic effect of information, while the media is treated as a physical phenomenon associated with semantics to create information.

Cognitive Informatics

- Intellectual Foundations of Computer Science
- Internal information processing mechanisms
- Models of brain memory
- Cognitive models of the mind
- Descriptive Mathematics
- Semantic Networks and Intellectual Roots of Computer Science
- Cognitive basis of software engineering
- Law of Software Informatics
- Representation of knowledge
- Expansion of human memory
- New approach to computer science
- IT applications
- Applications in cognitive science



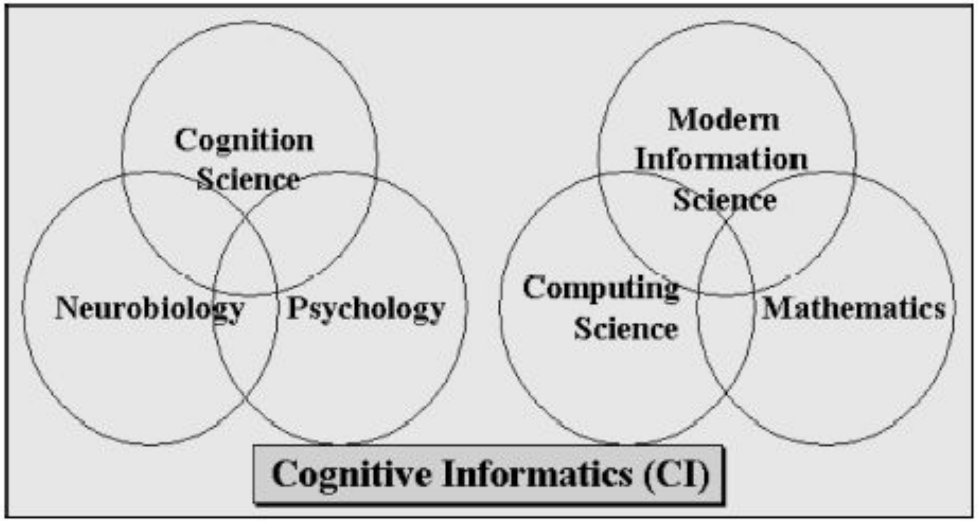
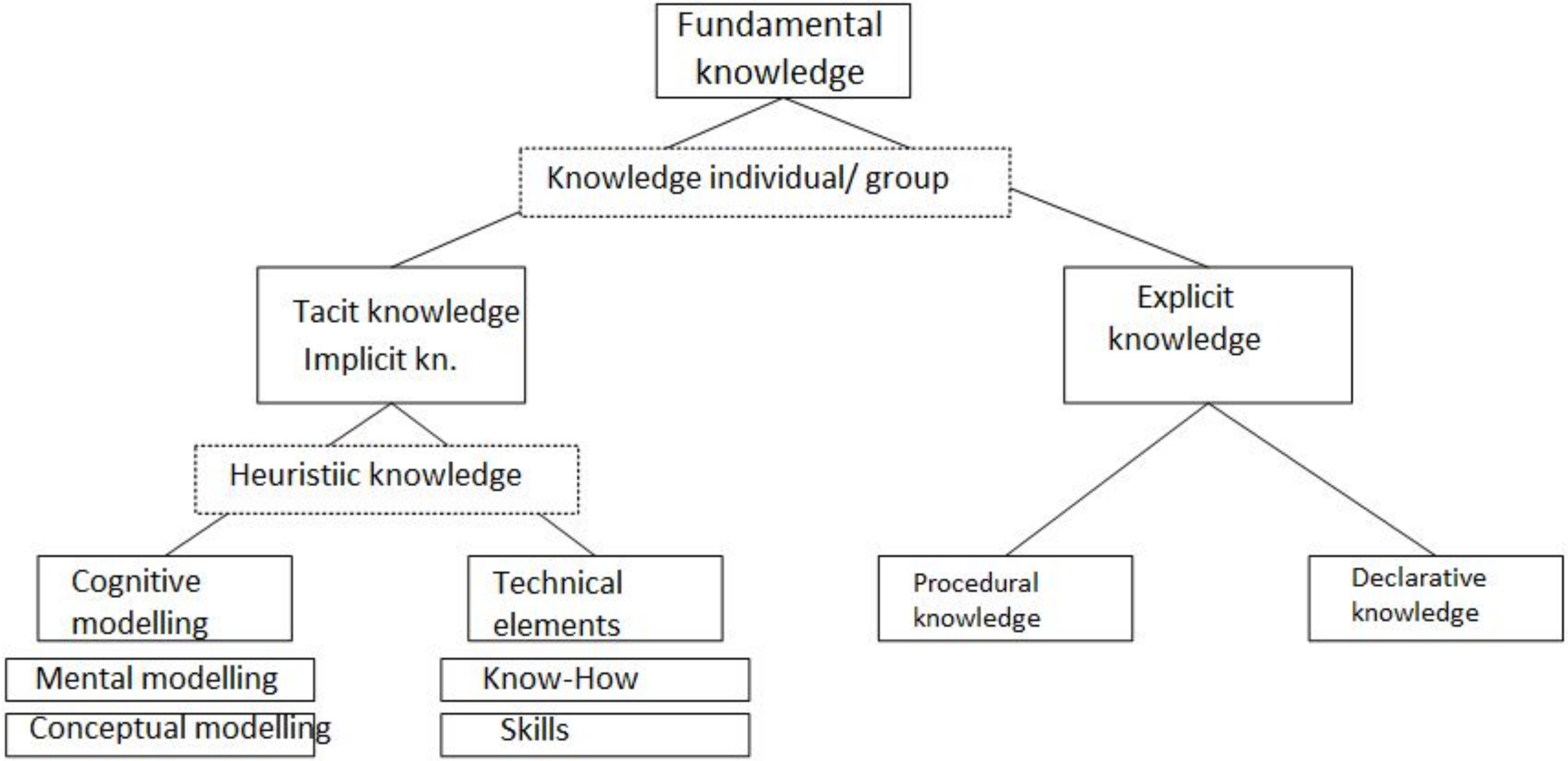
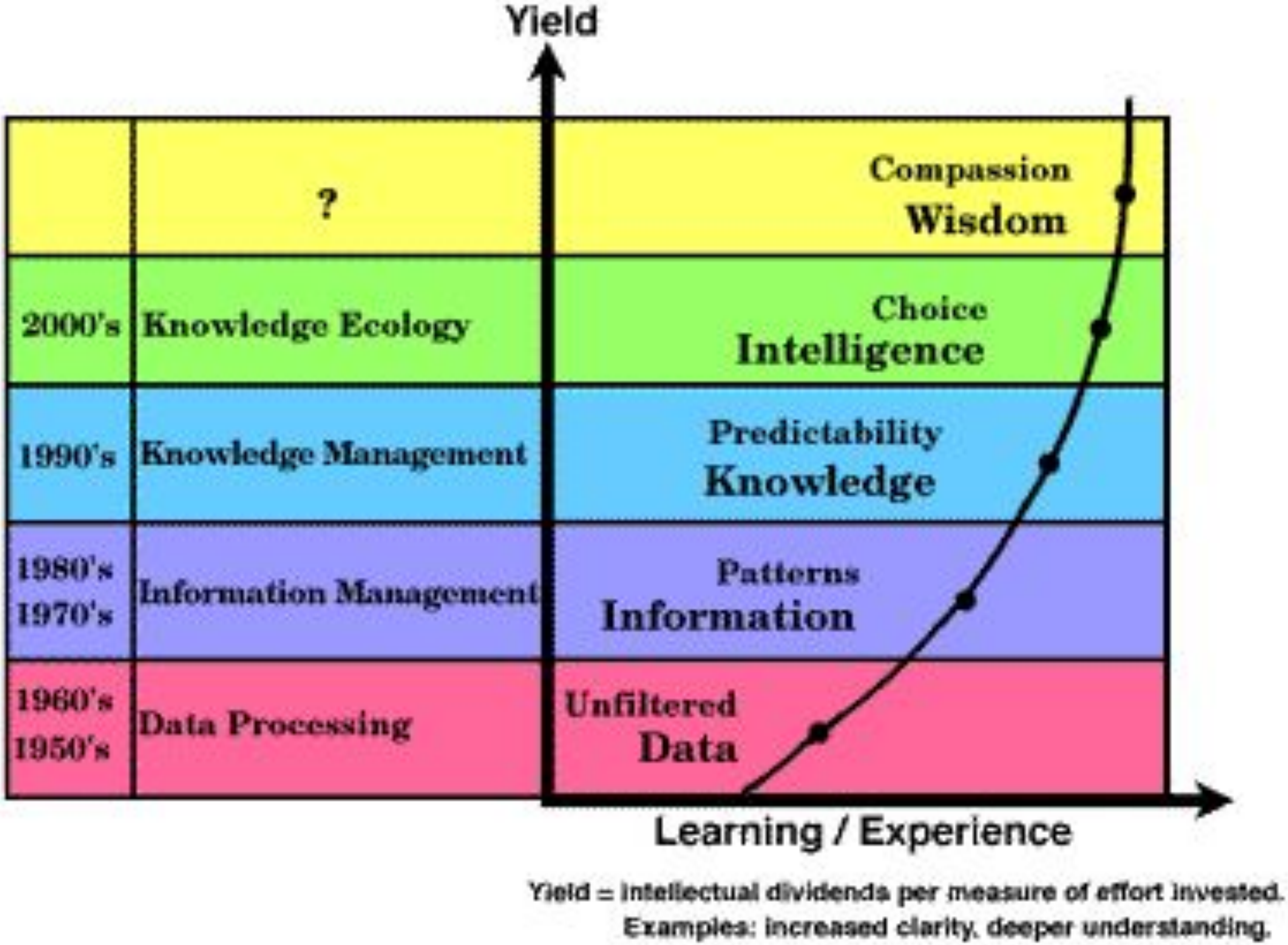


Table I. Subject Areas of Cognitive Informatics and their relationships with Informatics and Cognitive Sciences

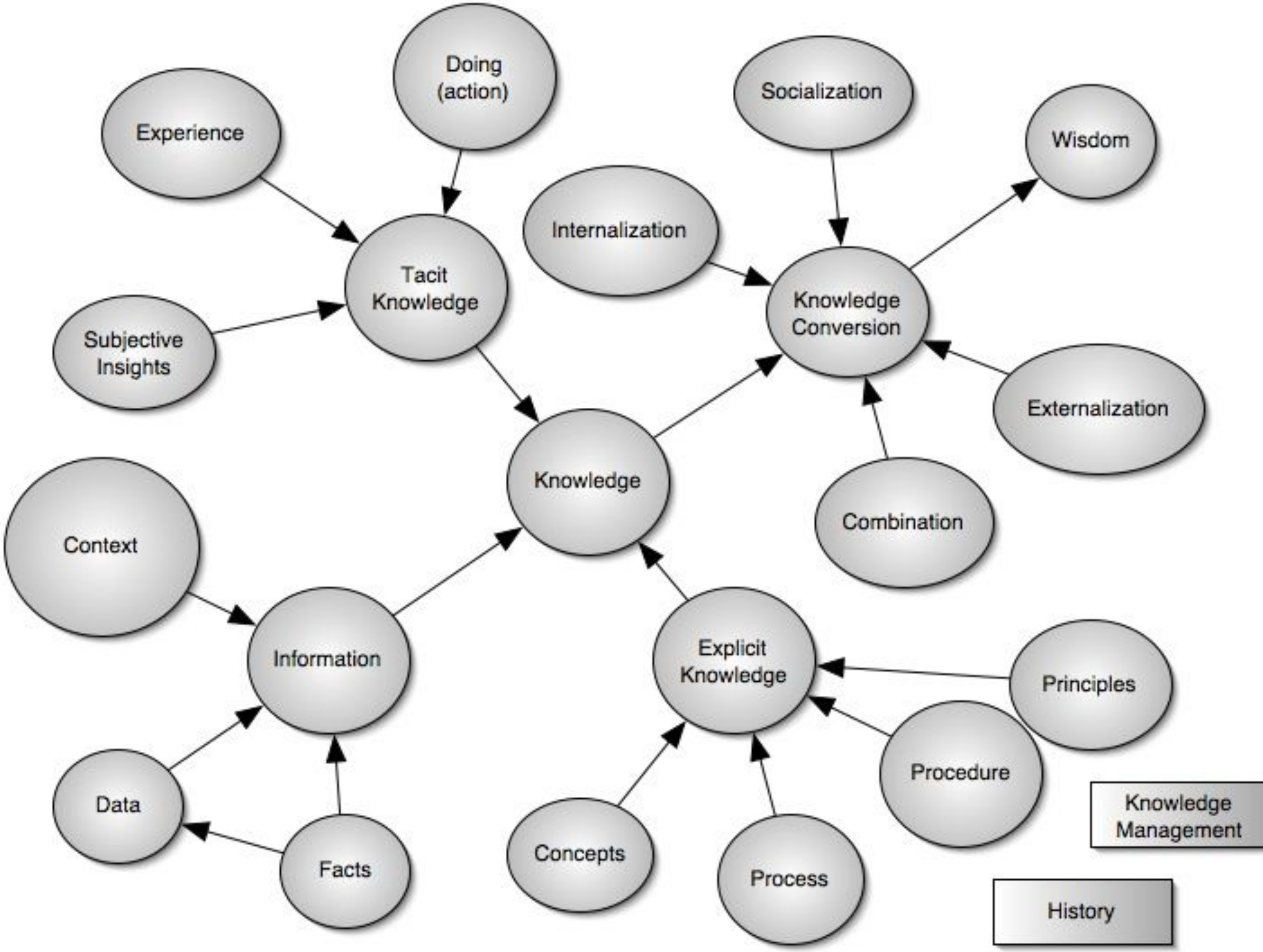
Informatics	Cognitive Science	Cognitive Informatics (CI)
Nature of software	Natural intelligence	Intellectual foundations of informatics
Knowledge engineering	Philosophy of mind	Internal information processing
Bioinformatics	Brain organization	Information models of the brain
Quantum information processing	Cognitive mechanism and process	Informatics foundations of software engineering
Artificial intelligence	Memory	Expressive mathematics
Fuzzy logic	Learning	Intellectual roots of computing
Machine learning	Thinking	Comparative programming languages
Neural networks	Reasoning	Extension of human memories
Pattern recognition	Problem-solving	Informatics laws of software
Agent technologies	Cognitive linguistics	Knowledge representation
Web-based information systems	Neuropsychology	Neural computation

Knowledge classification



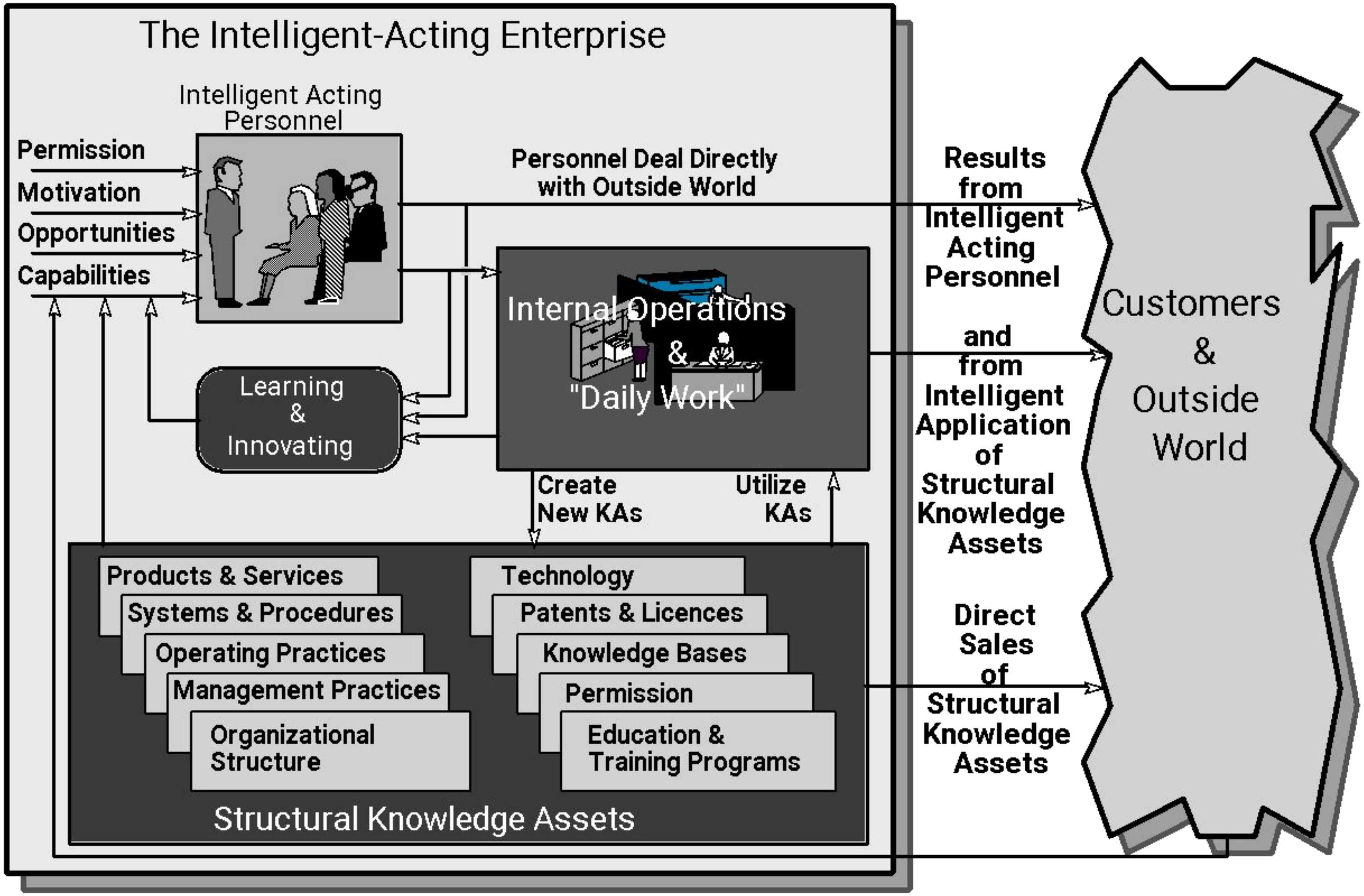


Źródło: Oregon Technology in Education Council,
http://otec.uoregon.edu/data-wisdom.htm



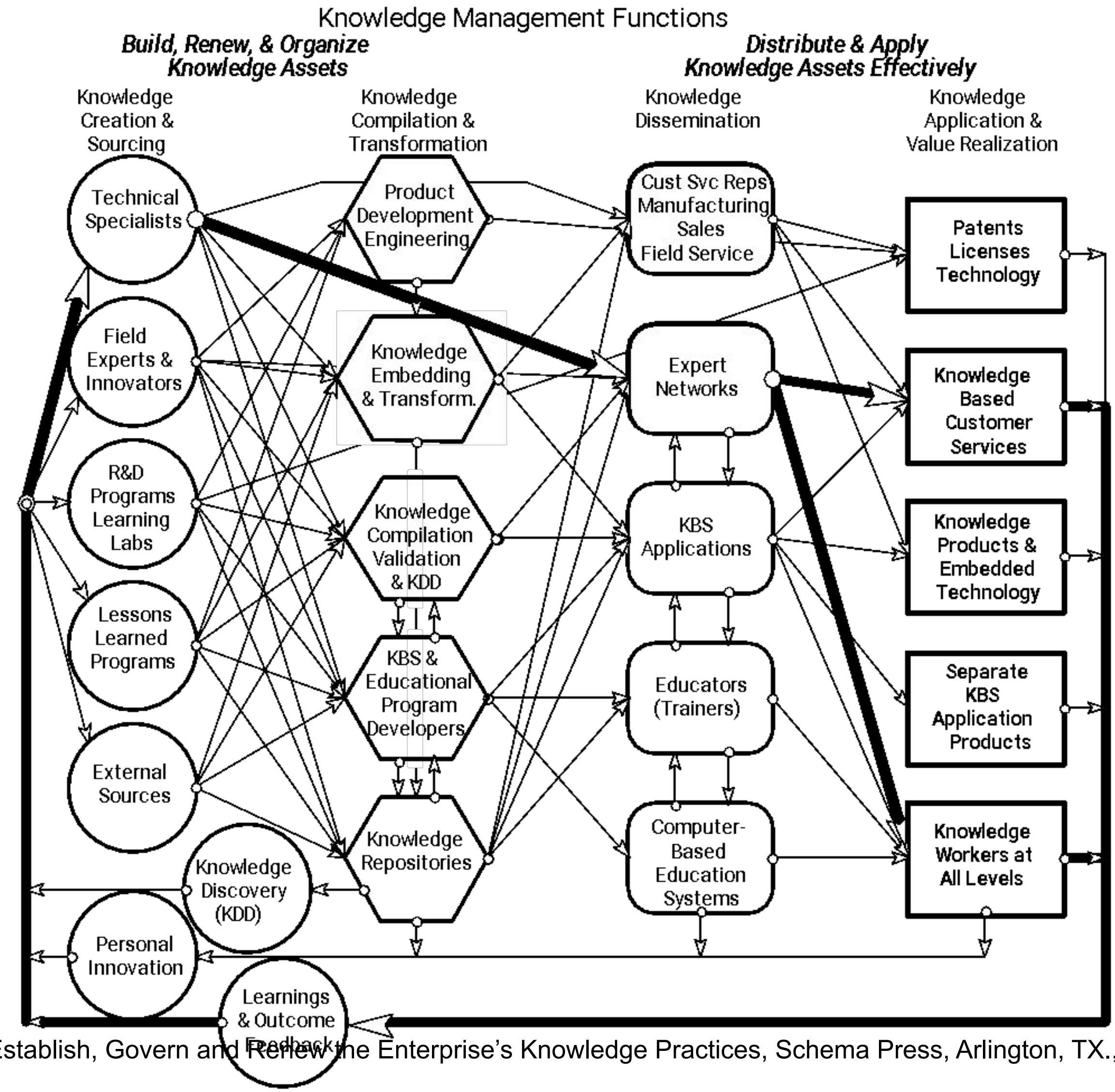
Źródło: Performance, Learning, Leadership, & Knowledge,
http://www.nwlink.com/~donclark/knowledge/knowledge_typology.html

Role of Individuals, Knowledge Assets, Learning and Innovation, and Internal Operations for Enterprise-Wide Intelligent-Acting Behavior



Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX.,
.

Knowledge Functions and Pathways in an Integrated Transfer Program

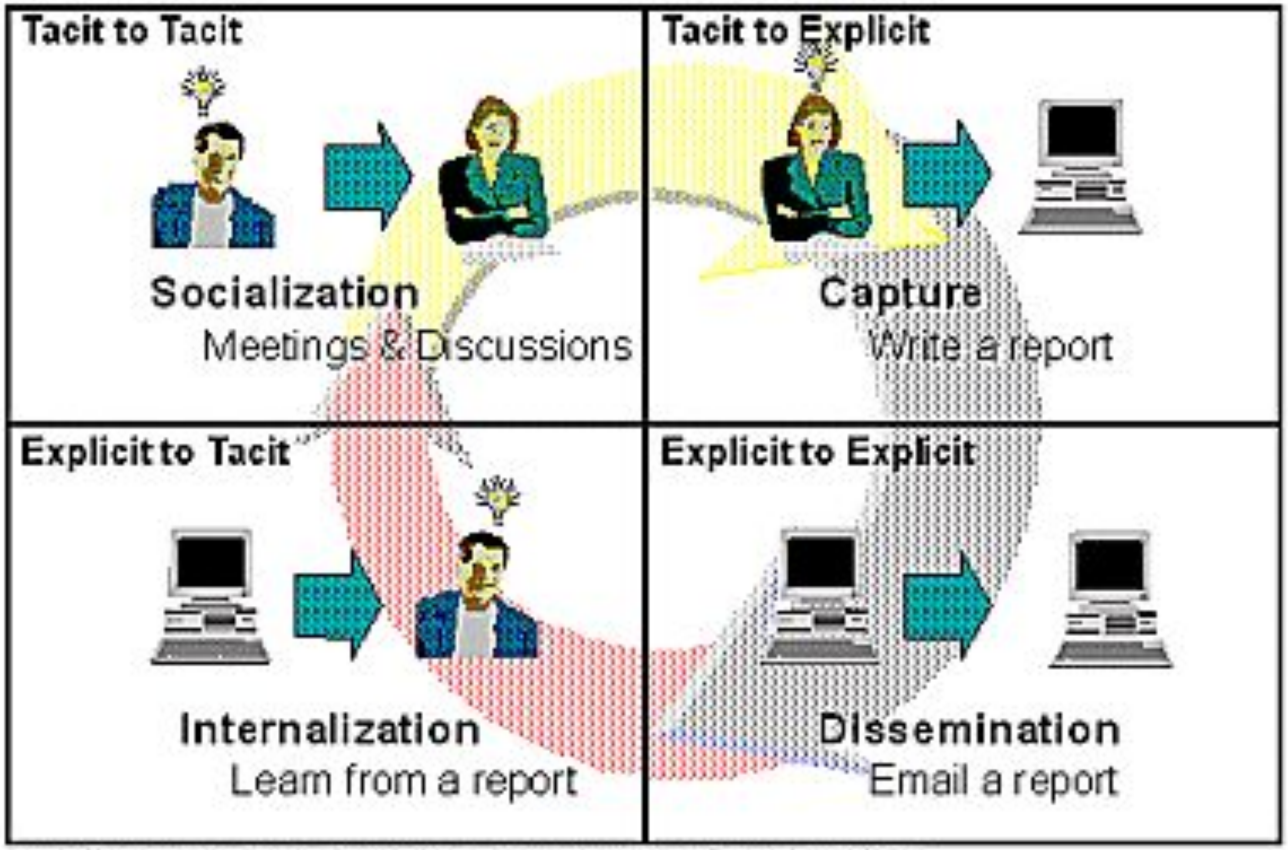


Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX., .

Wiedza ukryta (ang. Tacit knowledge)

"We know more than we can tell."

Michael Polanyi



Source: "The Knowledge Creating Company" Ikujiro Nonaka and Hirotaka Takeuchi (OUP, 1995)

Selected examples
Activities related with KM

1. Gathering knowledge
2. Automate knowledge transfer
3. Building computer education systems
4. Construction of a corporate university
5. Building knowledge base
6. Building a portfolio of knowledge-based activities
7. Collaborate to combine the right knowledge
8. Compilation of knowledge in knowledge bases
9. Comprehensive multi-path knowledge transfer programs
10. Conducting research and development
11. Creating and organizing knowledge repositories
12. Expert networking - design, targeting, budgeting, access mechanisms
13. Creating and developing the KBS educational program
14. Creating knowledge strategies

Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX.,
.

Selected Examples of KM-Related Activities (2/4)

16. Create Lessons Learned programs
17. Build staffs of technical specialists
18. Determine knowledge requirements for specific tasks
19. Determine knowledge-related benefits
20. Develop and deploy KBS applications
21. Develop educators (trainers)
22. Develop information technology (IT) infrastructure
23. Develop products with valuable knowledge contents
24. Discover & innovate - constantly
25. Create programs for effective knowledge capture
26. Embed knowledge in services
27. Embed knowledge in systems and procedures
28. Embed knowledge in technology
29. Build a program for enterprise-wide formal education and training
30. Establish KM professional consulting team

Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX.,
.

Selected Examples of KM-Related Activities (3/4)

31. Make available the expertise of field experts & innovators
32. Implement incentives to motivate knowledge creation, sharing, & use
33. Establish knowledge acquisition program
34. Discover knowledge in data bases (KDD)
35. Build knowledge inventories
36. Provide incentives to motivate employees to share knowledge
37. Maintain knowledge bases
38. Make knowledge available to customer service representatives
39. Make knowledge available to field service
40. Manage intellectual assets
41. Place high expertise in conceptual sales situations
42. Promote personal innovation
43. Provide best knowledge to workers at all levels
44. Provide companion KBS application products
45. Provide knowledge-based customer services

Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX.,
.

Selected Examples of KM-Related Activities (4/4)

46. Provide learnings & outcome feedback
47. Pursue knowledge-focused strategy
48. Restructure operations & organization
49. Sell knowledge embedded in technology
50. Sell knowledge products
51. Sell or license patents and technology
52. Sell products with high knowledge content
53. Sell separate KBS application products
54. Set knowledge activity priorities
55. Share knowledge throughout enterprise
56. Survey & map the knowledge landscape
57. Transform knowledge
58. Use external sources for valuable knowledge
59. Utilize technical specialists
60. Validate & verify knowledge

Źródło: Wiig, K. (1999), Establish, Govern and Renew the Enterprise's Knowledge Practices, Schema Press, Arlington, TX.,
.