The Data-Information-Time (DIT) Model

H. Kopetz
March 2022
Outline

• Introduction
• Data in the Communication among Humans
• Data in Cyber Space
• Data in Biological Systems
• Consequences for System Design
• Conclusions
Purpose of the DIT Model

It is the purpose of the DIT Model

• to establish the semantics of the terms *Data* and *Information*.

• to provide a framework to show how the flow of time can effect the *meaning of a data item* and the *sense of an information item*.

• to investigate the communication among humans, the communication among a human and a machine, and the communication among machines.

• to explain the differences between *archival data* and *real-time data*.

• to show how the understanding of the DIT model can help to improve the design of a computer system.
The *Merriam Webster* dictionary defines communication as follows:

- “A process by which information is exchanged between individuals through a common system of symbols, signs or behavior;”
- *the function of pheromones in insect communication.*”

In human communication, sequences of natural language words, i.e., sentences, are exchanged between involved *speakers* and *listeners*.

The *meaning* of a natural language *word* (a *data item*) is the associated *concept* —*the unit of thought*— in the mind of the speaker. A word has a *denotation* and a *connotation*.

In the DIT model the *sense* of a natural language proposition, i.e., the *idea* communicated by a *propoition composed of meaningful words*, is called an *Itom* (*Information Item*).

*A word has a meaning, a proposition makes sense.*
Concept Formation—Grounding of Words

According to embodied cognition humans develop a basic concept and therewith assigns meaning to—i.e. ground— the sounds of the identifying word when a plethora of bodily sensations (visual, acoustic, haptic, smell, taste) of a thing are perceived during a purposeful process.

Examples:

sensual impression associated with the word “pacifier”

This is a “tree”

After a set of basic concepts has been formed in this way, the grounded words can be used to ground new words and thus form more advanced concepts.
The totality of all concepts, relations among the concepts, and mental models forms the conceptual landscape in the mind of a human that is built up during the lifetime of an individual, partly by human nature and partly by nurture, i.e., the lived and remembered personal experiences.

The conceptual landscape is in continuous flux.

Each fully developed concept has a rich set of conscious and unconscious (tacit) relations to other concepts in the conceptual landscape.

The conceptual landscape is the dynamic subjective knowledge base in the mind of a human.
More Fundamental Terms

Entity: A thing in the real world or a construct of the human mind.

System: A system is a collection of related entities (subsystems) that forms a whole and provides a service to its environment. It is encapsulated by a physical or virtual skin that separates the system from its environment and contains interfaces.

Property: A characteristic of a thing or of a construct. e.g., the weight of a stone

Value: A refinement of a property. e.g., a value of the weight in measurement units.

Category: A construct that denotes a set of things or a set of constructs that share identified properties. e.g., lime-stone.

Time: A construct that can be depicted by a line that proceeds from the past to the future and is marked by the ticks of an UTC synchronized clock (Newtonian).
Model of *Time* in the *DIT Model*

*Happening* is an umbrella term denoting an event or a process.
Symbols in the DIT Model

**Symbol:** A construct that consists of two parts, a *signifier* and a *signified*. The *signifier* denotes the *signified* (the something).

- The *signifier of a symbol* can be a picture, a sound, a gesture, a letter, a digit, a word, or any kind of *physical pattern* that stands for *something*. The *signifier of a symbol* is also called the *name of the symbol* that identifies the *something*.

- Depending on what a signifier stands for we distinguish between four types of signifiers in the DIT model:
  1. a *word*: of a natural language that *denotes* a *concept*
  2. a *proper name*: *denotes* a specific person, thing or construct
  3. a *literal*: the form of the signifiers *denotes* the *signified*
  4. a *token name*: *denotes* a meaningless placeholder, a *token*
**Literals:** The *Gestalt* of the *signifier* denotes the *signified*

<table>
<thead>
<tr>
<th>Signifier</th>
<th>Signified</th>
</tr>
</thead>
<tbody>
<tr>
<td>a pattern (acoustic, visual)</td>
<td>an abstraction</td>
</tr>
</tbody>
</table>

**Example:** sequence of digits

<table>
<thead>
<tr>
<th>(decimal)</th>
<th>roman</th>
<th>(binary)</th>
<th>base-2</th>
<th>base-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>base 10</td>
<td>roman</td>
<td>base-2</td>
<td>base-4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>V</td>
<td>101</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>III</td>
<td>11</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

What determines whether “11” means five, or eleven, or three? *the context!*
Context

The context of a symbol is the general situation that relates to it, and which helps it to be understood.

In the DIT model we distinguish between the

- **Outer Context**: The outer context of a word is the objective reality in the current situation. It is determined by the prevailing physical, cultural and social environment.

- **Inner Context**: The inner context of a word consists of those parts of the conceptual landscape in the brain of a conscious human that are related to the concept denoted by the word and the current perceptions of relevant events in the physical environment (i.e. the outer context) that are delivered to the conceptual landscape by the human senses.
What is Consciousness? The TNGS Theory

- According to the Theory of Neuronal Group Selection (TNGS) consciousness is an emergent property as a consequence of the intense interactions among a large number of neurons in the thalamocortical system of the human brain.
- At any moments, a conscious person is only aware of the small part of the conceptual landscape—called the dynamic core—that relates to the momentary situation.
- By shifting the attention from one topic to another topic, the composition of the dynamic core is changed. The attention shift time is in the order of hundred of milliseconds to seconds.
- The dynamic core interacts with the other areas of the brain, where a plethora of powerful concurrent autonomous neural processes are active, by input ports and output ports.
- The current perceptions of the physical environment are preprocessed and categorized by unconscious autonomous neural processes before they are delivered in the form of a high-level concept to an input port of the dynamic core.
Communication among Humans

Human Communication uses predominately sentences composed of natural language words. A *basic natural language* sentence communicates an *idea*—an *Information Item* (*Itom*)—by the use of *symbols* (words). It consists of a *subject*, a *predicate* and an *object phrase*.

![Diagram of sentence structure](image)

- **<Subject>**
  - Often a proper name of an entity
    - a person
    - a thing
    - a construct

- **<Predicate>**
  - Often a relation, e.g.
    - a property
    - tensed

- **<Object>**
  - Often a refinement of the relation

A sentence only makes sense, if all words of the sentence are grounded.
The DIT Model of Human Communication

Information, i.e., an *Idea* in the Mind of the Sender

- Encoding into Words
- Larynx
- Motor Output

Stream of Sounds

Written Text

Data in the Physical World

Ear

Decoding of the Words

Eye

Information, i.e., an *Idea* in the Mind of the Receiver

- Autonomous Neural Circuitry

Dynamic Core of the Sender

Dynamic Core of the Receiver

Message

Generation of signifiers, i.e., generation of a message

Explanation of signifiers, i.e., explanation of a message
An Example

Tom possesses a valid passport.

This sentence is semantically meaningful! But is it true?

This depends on whether the zero point (i.e., the utterance event) lies within the time interval contained in the passport.

In order to make a sentence denotationally meaningful we have to add to a semantically meaningful sentence the the instant (the zero point) when the sentence has been uttered.
When looking at a written text, we have two deal with two events:

- the *write event* ($W$-event), i.e., the instant when the text is written
- the *read event* ($R$-event), i.e., the instant when the text is read.

The interval between the $W$-event and the $R$-event can have an impact on the meaning of the text and the truth-value of a proposition.

If the $R$-event of the previous sentence, where the tense is the *present*,

\[
\text{Tom possesses a valid passport}
\]

is outside the validity interval contained in the passport, then this sentence is false.
Stigmergic Information Flow

We call the transmission of Itoms between a sender and a receiver via an observable modification of the state of a shared physical environment *stigmergic communication*. **Example:** Communication among Drivers.

*A stigmergic information flow* can form the final link in the formation of a causal loop.
The Impact of Time

(i) The predicate of a proposition (the *Zeitwort*) is tensed, causing a displacement of the Itom on the time line.

(ii) Over time, the contents of the conceptual landscape are changed, causing a change in the *denotation* and *connotation* of words.

(iii) During the interval between the *W-event* and the *R-event* the context is altered, causing a change in the truth value of the proposition.

In oral communication where the *zero-point of an utterance* is also the *instant of interpretation*, point (iii) is of no concern.
Handling of *Time* in the *DIT model*

In the DIT model, the *predicate of a proposition* must be attributed with temporal parameters:

- If the communicated happening refers to an event, then the timestamp of the occurrence of the event must be attached to the predicate.

- If the communicated happening refers to a state or a process, then two timestamps, the timestamp of the start event and the timestamp of the end event of the happening must be attached to the predicate.

It is assumed, that all involved parties have access to the global UTC time, distributed by GPS, such that the precise interpretation of the timestamps is possible.
A *token name* denotes a meaningless placeholder, a *token*.

Consider the Statement

<table>
<thead>
<tr>
<th>token (variable) names:</th>
<th>a</th>
<th>x</th>
<th>b</th>
<th>=</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>values of the tokens (<em>ungrounded</em>):</td>
<td>3</td>
<td>x</td>
<td>5</td>
<td>=</td>
<td>15</td>
</tr>
</tbody>
</table>

If we consistently change the names of all tokens in a complete program, e.g., d,e,f, instead of a,b,c, then the result of the computation is not changed.

The *token names* and the *values* are *not grounded*!
A meaningless **token name** *denotes* the placeholder of a *value*. Consider the Program

```

token (variable) names:       a    x    b    =    c
Values of the Variables:    3    x    5    =    15
```

The grounding of statement replaces the *token name* *by a meaningful word phrase* and *grounds the value*, thus forming a (possibly degenerate) *Itom*.

```

\[
\begin{array}{c}
\text{Number of dishes on the table} \\
\text{Itom 1} \\
\text{a}
\end{array} \\
\begin{array}{c}
\text{Number of apples on a dish} \\
\text{Itom 2} \\
\text{x}
\end{array} \\
\begin{array}{c}
\text{Number of apples on the table} \\
\text{Itom 3} \\
\text{b}
\end{array} \\
\begin{array}{c}
\text{=}
\end{array} \\
\begin{array}{c}
\text{=}
\end{array} \\
\begin{array}{c}
\text{c}
\end{array}
```

```
Correctness of an Algorithm

• An algorithm, implemented by a program, establishes formal relations among a set of tokens.

• If we ground the tokens by assigning meaningful words to the tokens, then these *formal relations* model *actual relations* about the assigned concepts that are supposed to exist in reality.

• If these *actual relations* are *true* in reality, then we say that the program is *accurate*. 
If the name of a variable in a program is a meaningful word that denotes (as its signified) an established concept in the mind of a human, then the value of this variable is grounded and can be considered a (possibly degenerate) information item (Itom) that carries sense.
Summary: Characteristics of a *Data Item*

In the DIT model a *data item* is a symbol that consists of a *signifier* and a *signified*. The *signifier* — the name of the symbol — is the physical pattern that represents the *data item* in the physical world.

- In human communication the signifier is a *word* and the *signified* of this symbol is the *meaning of the word*, determined by the assigned concept in the mind of an conscious human receiver, forming a *meaningful data item*.

- In a computer system the *signifier* is a *meaningless token (variable) name* and the *signified* of this symbol is the *value* housed by the token that is operationally explained by the use of the variable name in the computer program, forming a *meaningless data item, if the variable name is not grounded*.

In both cases a *data item* is a *whole* that consists of two parts, the *signifier* and the *signified*. If we take the whole apart and look at each part in isolation, then the notion of a *data item* breaks.
In the DIT model an Itom is a *proposition* that makes *sense*—informs about an *idea*—and has the following characteristics:

- **Temporal Aspects:** The Itom must inform when the predicate of the proposition holds.

- **Truthfulness:** Our conception of an Itom does not make any assumptions whether the information carried by an Itom is *true*.

- **Relativity:** Any representation of an Itom must consider the context, which is given by the current physical and cultural environment.

- **Newness:** The aspect of *newness of information* to the receiver and associated metrics about the subjective value or the subjective utility of the Itom to a receiver are not part of the DIT model.
Archival Information

An archive is a repository that holds documents of permanent historical information (a vast plurality of Itoms) that are dated and stored on different kinds of media in order that this information can be accessed and analyzed at some future date.

An Itom, the sense of a proposition, consists of four data items:

- **Subject**: an entity
- **Predicate**: a relation (property)
- **Time**: timestamps
- **Object**: a value (e.g., a literal)

An Itom consists of a value and the explanation of the value.

In a database, the explanation of the values, which makes the Itoms, is contained in the schema of the data base.
**Explanation of the Value: A simple File**

A simple file is a tabular two-dimensional data structure, containing a

- A file name—should explain the context of the file and the *grounding of the values* (and possibly the time).
- A row name—explains the entities
- A column name — explains the relation (property)
- A value field — contains the value, often a *literal*, at the intersection of *row* and *column*.

Every value field together with its explanation, forms a (degenerate) *Itom*.

*What is often missing is the temporal information.*
Example of a Table

This information can be represented by the following six items:

- On 2021 11 11 Paul has a size of 160 cm.
- On 2021 11 11 Ann has a size of 155 cm.
- On 2021 11 11 Paul weighs 61 kg.
- On 2021 11 11 Ann weighs 53 kg
- On 2021 11 11 Paul belongs to Class 3A
- On 2021 11 11 Ann belongs to Class 3A

<table>
<thead>
<tr>
<th>Name/Property</th>
<th>Size in cm</th>
<th>Weight in kg</th>
<th>---</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul</td>
<td>160</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Ann</td>
<td>155</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
An *Itom* is a *primitive* of an DIT knowledge graph.
Example: Two different Tables for the same *Itoms*

**Size and Weight of the Students in Class 3A, measured on 2021 11 11**

<table>
<thead>
<tr>
<th>Name/Property</th>
<th>Size in cm</th>
<th>Weight in kg</th>
<th>YT -----------</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul</td>
<td>160</td>
<td>61</td>
<td>...</td>
</tr>
<tr>
<td>Ann</td>
<td>155</td>
<td>53</td>
<td>...</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
</tbody>
</table>

**Size and Weight of Paul**

<table>
<thead>
<tr>
<th>Property/Date</th>
<th>2020 11 10</th>
<th>2021 05 08</th>
<th>2021 11 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size in cm</td>
<td>153</td>
<td>157</td>
<td>160</td>
</tr>
<tr>
<td>Weight in kg</td>
<td>58</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td>Member of Class</td>
<td>2B</td>
<td>2B</td>
<td>3A</td>
</tr>
</tbody>
</table>
Itoms in Real-Time Systems

- In a real-time system, the validity (and utility) of an Itom (which must be part of the Itom) is limited by the dynamics of the controlled object.
- A short response time of the computer system improves the quality of control.
- The communication time can be reduced by a recoding the representation of an Itom for the transport. Separate the specification of an Itom (using natural language) from representation (the encoding) of the Itom.
- The processing time can be reduced by the deployment of anytime algorithms. A satisfying result produced in time is more important than a precise results that arrives too late.
Generation of Data

An Itom consists of a data item and the explanation of the data item:

• **Identification of an entity:** We have to identify an entity or categorize the entities in order to restrict the number of entities in the CSTD.

• **Identification of a property:** If the property specification changes in the IoD, then the value will probably change as well.

• **Timestamp of the observation of the phenomenon:** In a real-time control system, a second timestamp for the end of the validity time of the observation must be part of the temporal data.

• **Value of the property:** this is often a literal (e.g., a string of digits) representing a numeric value. The measuring units of the numeric value must be included in the property specification.
Explanation of Data

The explanation of the data is somehow the inverse of the generation of the data.

• Care must be taken that the context of explanation of data is the same as the context of the generation of data. This is of particular importance when the change of context between the instant of data generation and instant of data explanation has an effect on the explanation of the data.

• At the human-machine interface (HMI) the representation of the values of interest must be framed by an explanation that is grounded in the mind of the human reader. Also, the measurement units of values must be familiar to the intended clientele of the Itom.
Data in Biological Systems

Phases in the Life of a Plant

Seed contains the genome of an organism

One Generation of the Phenotype
- dormant phase
- germinating phase
- growth phase
- mature phase

Next Generation of the Phenotype
- dormant phase
- germinating phase
- growth phase
- mature phase

The genome, stored in the DNA, can be considered as the control data base that controls the development of an organism.
Main Subsystems of a Plant

The *Codons* of the DNA control the structure of the protein production in the Ribosomes.

(a) Main Subsystems of a Seed

- Master Controller (cell nucleus contains the DNA)
- Protein Factory (Ribosome)
- Energy Subsystem (Mitochondria)

(b) Main Subsystems after Germination

- Master Controller (cell nucleus contains the DNA)
- Protein Factory (Ribosome)
- Energy Subsystem (Mitochondria)

Power Plant for the Conversion of Sunlight to stored energy
## Comparison of the Notion of *Data*

<table>
<thead>
<tr>
<th></th>
<th>Computer</th>
<th>Plant</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signifier</strong></td>
<td>variable</td>
<td>codon</td>
<td>word</td>
</tr>
<tr>
<td><strong>Signified</strong></td>
<td>a specific value</td>
<td>a specific amino acid</td>
<td>concept in the human mind</td>
</tr>
<tr>
<td><strong>Meaning of the</strong></td>
<td>its use in its use in</td>
<td>concept</td>
<td></td>
</tr>
<tr>
<td><strong>signified derived from</strong></td>
<td>a program</td>
<td>protein synthesis</td>
<td>formation</td>
</tr>
<tr>
<td><strong>Physical Embodiment generated by</strong></td>
<td>Human</td>
<td>Nature</td>
<td>Nature</td>
</tr>
<tr>
<td><strong>Modification of the Physical Embodiment by its use</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Reproduction capability</strong></td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
An important Conclusion

Humans deal with *I*to*ms* while Machines deal with *Tokens*
Consequences for System Design

- Specification Dilemma
- Human Machine Interface Design
- Benefits of a Global Time—Data Reduction
- Partitioning of a Safety-Critical Control System
The fact that the precise meaning of a natural language word is subjective and unique to a single person and may thus differ among different persons is the cause of the specification dilemma.

The use of natural language is unavoidable in

• the specification of the purpose of a system,
• the grounding of the symbols or
• the interfaces between cyberspace and the real-world

However, the data interfaces within and between the computers in cyberspace can, and should be formally specified.

Consequence: Project Dictionary of all technical terms!
Human Machine Interface Design

At an interface to a human user, the display must include the value and the explanation of the value in an appropriate display format, such that the human user can grasp the meaning of the value with a minimal mental load:

• Consider the mental models in the mind of the user
• Study the cultural background and the work profile of the user
• Distinguish between occasional and experienced users
• Take advantage of established color codes (green, orange, red)
• Implement the explanation of the data and the provision of the results in separate subsystems.
The DIT model requires a global time in order to arrive at a *non-tensed design*. A global time is required to

- to interpret time-stamps
- to design unidirectional communication protocols. A unidirectional control flow avoids a failure propagation of a faulty receiver back to a correct sender.
- Time-triggered protocol avoid queuing delays in communication systems.

Example: Smart grid
Architecture Design

Given a clearly specified purpose, it is the most difficult—and most important—challenge in the development of a large system to find a structure a that will minimize the cognitive complexity.

At the level of Architecture Design—reason in Itoms, including the temporal parameters. In an highly iterative design process:

• Partition the System into nearly independent subsystems.
• Minimize the interface complexity.
• Define the Itoms that must flow between the subsystems.
• Ensure that the system integration will not produce emergent effects.
• Ensure that the project organization maps the evolving system structure.
Two Techniques for *Complexity Reduction*

**Abstraction:** Find a hierarchy of reduced representations—and of *design artefacts*—primarily from the point of view of the intended system service.

- Formal Hierarchy
- Holonic Hierarchy
- Dependable Control Hierarchy

**Divide and Conquer:** Partition the system into nearly independent subsystems.

- Minimal Interfaces (where ever possible, unidirectional)
- Avoidance of unintended *Emergent Effects*.
Abstraction

Formal Hierarchy

Holonic Hierarchy

Dependable Control Hierarchy

Input/output
Minimize the Interface Complexity

In a real-time system many *Itoms* flow periodically and unidirectionally from the sensors to the actuators.

- Place the interfaces at positions, where a *minimal number of Itoms* must cross the interface and where future changes are unlikely to change the interface.
- Maintain the unidirectionality of the communication at all levels of the communication protocol in order that *by design* a failure of a receiver has no impact on the sender—use *time-triggered protocols*!

Define the *Itoms* in *natural language* and in a *formal notation* in carefully monitored *Interface Control Documents (ICDs)* that meet the *mutual expectations*.
Avoid unintended Emergence

A phenomenon at the system level is called *emergent* if it is of a new kind with respect to the non-relational phenomena of any of its proper subsystems.

- In many cases of emergent behavior, we find a causal loop exists between interacting subsystems.
- Try to avoid unnecessary bidirectional communication links.
Starting point: The purpose of the proposed system, giving rise to set of functional requirements and safety requirements.

Partition the overall system into Fault-Containment Units (FCUs) and specify the dependability properties of the FCUs

• Assign the top level functional and safety requirements to the FCUs
• Specify the $ltoms$ that cross the interfaces of the FCUs.
• Analyze if the required safety will be attained and can be argued by the proposed structure.

Take into consideration widely accepted impossibility results.
**Impossibility Results for Autonomous Driving**

There is strong experimental evidence that it is impossible in ultra-dependable systems to overcome the constraints that are summed up in the following *five impossibility results*:

- It is impossible to find all design faults in a large and complex *monolithic* Software System.
- It is impossible to avoid single event upsets in non-redundant hardware during the life-time of an ultra-dependable system.
- It is impossible to establish the ultra-high dependability of a large monolithic system by testing and simulation.
- It is impossible to precisely specify all edge cases that can be encountered in driving situations.
- It is impossible to shift human attention without a significant *attention shift time*. 
Example of an FT-Architecture for an SAE Level 4 System

FTDSS
Fault-Tolerant
Decision Subsystem

MSS
Monitoring
Subsystem

CCDSS
Computer Controlled
Driving Subsystem

CEHSS
Critical Event Handling
Subsystem

Fault-Tolerant
Actuator

SP: Sensor
Preprocessing

SP

Prospective
Trajectory

SP

SP

SP

SP

SP

SP

SP

SP

Sensors
of the MSS

Sensors
of the CCDSS

Sensors
of the CEHSS

Example of an FT-Architecture for an SAE Level 4 System
Conclusion

• The DIT Model makes a clear distinction between the terms *data item* and *information item*.

• A *data item* is a symbol that is *meaningful* if the signified is grounded, while an information item—an *Itom*—is a timed proposition that is composed of *grounded data items* and carries an *idea*—makes *sense*.

Humans deal with *Itoms* while Machines deal with *Tokens*.

*Thank you.*
Publication Date: April 25, 2022