A Theoretical Base for the Simulation of Information Systems

Isomorphically Acted Organization Scenarios

Joel Palmius

A Theoretical Base for the Simulation of Information Systems
- Isomorphically Acted Organization Scenarios

Joel Palmius

Thesis for the degree of Licentiate of Philosophy

ISSN: 1652-8948
Mid Sweden University Licentiate Thesis 8
ISBN: 91-85317-13-6

Department of Information Technology And Media
Mid Sweden University
SE-83125 Östersund
Sweden

Phone: +46(0)63165946
E-mail: joel.palmius@miun.se

Printed by Kopieringen Mittuniversitetet, Östersund, Sweden, 2005

Preface

Preface

ABSTRACT: Information is an important commodity in the modern organization. To support the internal flow of information, the organization's information system is important. Thus, designing the information system becomes an important task when planning an organization. However, most systems development method today are prescriptive and do not include predictive measurements. Without a predictive measurement, it is difficult to know in advance if a planned information system is likely to be efficient. One way to get a prescriptive measurement would be through simulating the intended information system. This thesis suggests a theoretical framework for supporting the construction of such a simulation. The important entities are analyzed and operationalized in order to be possible to represent in a simulation. The topic of simulation itself is addressed. A theoretical foundation for managing the complexity of a simulation such as this is presented, based in notion of isomorphic equivalence. A framework with operationalizations of the important entities of an information system is presented as the main result of the thesis.

1. Included material

This thesis builds on four articles previously published in other circumstances. These articles are attached at the end of the thesis. Other sources have been related works. These are available on the thesis homepage (see below).

1.1. Articles

- (Article 1) Egonsdotter G and Palmius J (2002): *A Base for Simulating Information Distribution*. From Ragsdell R (ed) (2002): *Systems Theory and Practice in the Knowledge Age*. Kluwer Academic / Plenum Publishers. USA.

- (Article 2) Palmius J, Egonsdotter G and Asproth V (2003) *Simulations of Highly Complex Social Systems as a Tool for Designing Information Systems*. From Dubois D (ed) (2004): *International Journal of Computing Anticipatory Systems*. Vol 14.
- (Article 3) Palmius J (2005): *A Theoretical Base for High-level Simulation of Organisational Information Systems*. From *Systems Research and Behavioral Science*. No 22.
- (Article 4) Palmius J (2005): *Defining the "information" part of "information system" - A Base for Simulation*. From *Proceedings of IRIS 2005*

1.2. Internally related material

- (Work 1) Palmius J and Silvergran J (2003): *An Evaluation of Ideal Route Models of Indoor Navigation*. Bachelor thesis in Psychology. [http://www.palmius.com/joel/lic/psykc.pdf]
- (Work 2) Egonsdotter G and Palmius J (2001): *AMSIDO-1 end report*. Unpublished report to SwNDC. [http://www.palmius.com/joel/lic/a1report.pdf]
- (Work 3) Egonsdotter G, Palmius J and Öberg LM (2002): *AMSIDO-2 end report*. Unpublished report to SwNDC. [http://www.palmius.com/joel/lic/a2mainreport.pdf]

1.3. Externally related material

Finally, there are also two master thesis written by students who I guided. I have been influenced by these works and will refer to them in the thesis.

- (Thesis 1) Kristoferson M and Bergkvist J (2002): *Information node characteristics for model construction*. Master thesis in Informatics. [http://www.palmius.com/joel/lic/infonode.pdf]
- (Thesis 2) Holmström J and Koli T (2002): Thesis 2: *Making the concept of information operational*. Master thesis in Informatics. [http://www.palmius.com/joel/lic/infoop.pdf]

1.4. Homepage

Articles in original format, and PDFs of related material, are available on a homepage made for this thesis. There you will also find a digital version of this document. The address of this homepage is http://www.palmius.com/joel/lic/.

2. Acknowledgements

As with most research, much of the foundations for this thesis has been made through teamwork. During the first few years of the process, Gunilla Egonsdotter was a very important component of this teamwork. I can safely say, that without her hard work, the work with the Swedish National Defence College would not have had come half as far as it did. Often Gunilla has also played an important part in helping me retain at least a somewhat realistic perspective on the work and my studies, something which decidedly has helped a lot.

Other persons have been involved in the process now and then. Students have been doing wonderful master theses, colleagues have participated in practical work and in discussions, and various contact persons have kept up the connection with external organizations. Of particular note here are Tuomas Koli, Jens Holmström, Johanna Bergkvist and Maria Kristoferson for writing theses within the frame of the AMSIDO project, Lena-Maria Öberg for her participation within AMSIDO, and Anders Christensson and the AQUA team for constituting the National Defence College's part of AMSIDO.

Outside the frame of the university I would like to specifically thank "A" for a lot of interesting discussions. While she might not believe so, they have influenced both me and this thesis in a significant manner.

Finally and most importantly I would like to thank my adviser Viveca Asproth. Believe me, I have never met a person with such super-human patience anywhere else. Combined with her keen eyes for what makes sense, this has supported me and my work in more ways than I can remember.

Cover Paper

List of Figures

Table of Contents

Cover Paper

1. Introduction

During the last decades, information has become a critical commodity for most organizations. It is commonly accepted that we live in a culture which can be characterized as "the information society". Organizations need information in order to be able to react and manage their environments. The way information is managed is one important base for the viability of the modern organization. In consequence, a good deal of the requisite variety of modern organizations is handled by their internal information systems. The full definition of "information system" is presented in a later section, but can for now be understood as all that contribute to distributing information within an organization.

Accepting that the management of information is essential, the question of how to construct information systems becomes interesting. It is common knowledge that there are a lot of methods for constructing *software systems*, such as Agile (Martin R, 2002) and RUP (Kruchten P, 2003), and slightly fewer for modeling information systems in the organizational sense, such as SSM (Checkland P & Scholes J, 1997) and VSM (Espejo R & Harnden R, 1989). What most of these have in common is that they are prescriptive, and have very few points of pre-hoc evaluation.

Such a method can certainly be used in a good manner to construct an information system which is reasonably likely to be sufficient for the purpose for which it was constructed. However, the problem is that we simply do not know in advance that it will be good. Further, if we have several similar or functionally equal suggestions for how to organize an information system, these methods do not provide a pre-hoc predictive evaluation of the coming information system: there is no easy way to quantitatively compare different possible models with each other or with the currently existing information system.

The consequences of this problem is that information systems are constructed with insufficient advance knowledge of whether they are likely to be efficient, ie whether they will distribute the right information to the right person with a minimum of fuzz. During the development process, and on paper, the system may look good, but in the end it might end up being suboptimal or right-out disastrously bad. This is a known problem within the area; many development projects fail for various reasons (Ewusi-mensah K, 1997). The lack of predictive measurements does not explain all these failures, but could certainly address some of them.

Now, if we have come so far that we are about to start a development process for a new information system, or have even started it and have a suggestion, how could we gain a predictive measurement of the model to complement the prescriptive design guidelines set out by the development method we have chosen? In this thesis, I will suggest that simulation is one way of doing such a prediction. With a simulation, we could evaluate and compare system sketches in order to gain additional clues on whether they are likely to be efficient.

The question of simulations is not all that easy though. While a wide variety of simulations have been built and successfully used, there are many questions left to address before they can be viable tools for the form of prediction mentioned here. As an axiom we can say that a simulation is never better than the model it builds on. This model must by nature build on a set of concepts and definitions which mold the reality which is the focus for the modeling. Before beginning to build a simulation we must thus make these concepts and definitions distinct, operational and explicit.

1.1. Purpose and Intended Audience

The purpose of this thesis is to provide a theoretical base for a future implementation of simulation of information systems. This consists of two parts: A thorough theoretical overview of the area, and a suggestion on an operational model.

The intended audience is readers who either have knowledge of the target area, or of the approach. In other words, the intended reader is an informatician with interest for the organizational aspects of information system, an organization scientist who is literate in the problems around information systems, or simulationists who have a basic understanding of organizations and information systems.

My intention is that the cover paper should be possible to read by people who do not have specialist knowledge of simulation approaches.

1.2. Method and Point of View

The following is a specification of my approach for finding the findings in the following text. This method is not to be confused with the method for the separate articles, for which one should look in the articles. This said, the approach is largely the same for all of the work, both the cover paper and the articles.

My basic approach is construction, an iterative testing of mental models. In practise this has meant that I formulate a definition and then test it for internal consistency and apparent (face) validity.

In essence this approach co-incide with evolutionary prototyping from the design area, although I have used it solely on mental models. A prototype for a model is formulated from what is known. Then it is evaluated and tested for deficiencies, and new conclusions are drawn from the prototype. Finally it is discarded in favor of the construction of another model which incorporates the solutions found for the previous deficiencies.

It should here be noted that I use the terms "construction" and "design" as equivalent. Design theory should largely apply to the motivation. The reason I use 'construction', is because 'design' is not normally applied to the formulation of theories, and since I use a constructivist approach. Readers who are more comfortable with 'design' can in practise read 'design' where it says 'construction' in the following.

The construction process is not linear, and not reproducible in the positivist sense. It does not claim to be. What it says is that through concept formulation, analysis and interpretation a gradually better model is evolved from an initial sketch.

This is the approach for finding the end results. For the presentation, I use logical reasoning to motivate these results. What is visible in this text are the end results, and the motivation for them, while the whole process is not documented.

I do herein work with both ontology and epistemology. In order to work with the world, the world must be conceptually defined through the formulation of an ontology. This makes the world as we know it somewhat different. Further to make the work possible to formalize, we need a language to describe the world. This is the epistemology. The epistemology is the set of terms which we translate into entities.

Having stated the above, we still need to treat the concept of "validity". The term "validity" is largely dependent on a positivist ontology and epistemology. Validation implies that there is an objectively knowable reality against which we can compare measurements and models.

When our findings consist of interpretations, points of view and constructions, the question about validity becomes more complicated. A design cannot be objectively "validated" in the positivist sense. It is one way of doing it, and it may work better or less good than any other arbitrary way of doing it. We cannot claim that our design is the objectively best or only way to do things.

The question of validation in relation to design largely becomes a question about "does it work?". For the theoretical phase of a design, this translates into face validity:

If the construction is acceptable to the person or persons it is directed to, then it is valid. If the construction is not acceptable, then it is not valid. This makes validity rather subjective.

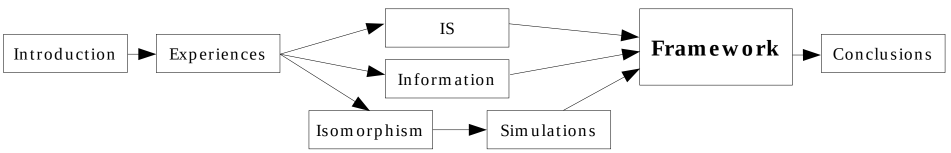
For the later, practical, phases of design (in other words the implementation), the question of validity becomes more tangible. Here the construction is valid if it works in a satisfactory way. This is still subjective: What is satisfactory to one person may be unacceptable to another.

As a necessary consequence of the approach and world view, results are mainly motivated and supported by reasoning, and sometimes by references to other works. The end result is a designed prototype of a model, and it is not claimed in this work that this is the only model or the best model. It is *a* model which it is reasonable to believe is viable to use in order to address the stated problem.

1.3. Format and Disposition

The following is organized into one "cover paper" (the text you are reading now) and four articles. The cover paper consists of 8 major sections.

Figure 1. Disposition



The main product of the work is the framework, while the other sections are (mainly) requirements for the presentation of this framework. The *introduction* is what you are reading now, and it presents the necessary background for understanding the purpose, base and method for the coming sections. The *experiences* section presents an overview of the experiences and decisions which has underlied the end result. The purpose of this section is to provide an understanding of why the current approach was chosen in the end, and to give an insight in the experiences that led to the conclusion that other approaches were not viable.

The four following sections do not depend on each other in a strict linear fashion. The IS (information system) section defines the concept of "information system", which is the whole context of the simulation. The "information" section treats two major issues: The definition of information and an operational model of information. The "isomorphism" section presents the basic philosophy on which the simulation approaches are built. This is a necessary prerequisite for the "simulation" section to be readable. The "simulation" section presents approaches for how the isomorphism basics can be implemented in practise.

These four sections do together constitute the theoretical base, which is one of the products of the work. They also are necessary pre-requisites for the "framework" section. The "framework" section presents suggestions on how the theoretical base can be concretely operationalized and provides a base for a future implementation.

With the existence of the "framework", the "conclusions" section sums up the work and presents concluding remarks and a meta-discussion.

1.4. Contribution

To further motivate the reasons for why the work as a whole is important, I will in the following specify what I see as the work's major contributions.

1.4.1. Problematization and Background

One of the problems encountered at the start of this project was the lack of understanding of the complexity involved with simulating an information system. This was partly due to different cultures and different previous experiences. Looking at current articles in, for example, Journal of Artificial Societies and Social Simulation, one can see that social simulations usually concern one *aspect* of a system, rather than trying to simulate the system as such.

One of the contributions of this thesis is thus the problematization of the area: An analysis of information system simulations with an holistic approach, and ways to go forward with it.

1.4.2. Re-focus of Simulation into a Holistic Approach

Another problem was the lack of understanding of what the target actually was. Gradually during the process of working with the simulation area, I have moved towards a more holistic approach. I have realized that there is a balance to be maintained between necessary complexity and necessary reduction, and that the reduction is not an easy thing to do.

One contribution of this thesis is a simulation model where this conflict is essentially resolved.

1.4.3. A Theoretical Base for Practical Implementation

In order to make the holistic approach work, a theoretical foundation for it needs to be formulated. This includes reformulations of important concepts into an operational epistemology and ontology. One contribution of this thesis is that reformulation. It should be noted that this is different than a summary of literature within the are: The reformulation is more to be likened with a design of a working toolbox of concepts.

More practically, the theoretical base is summarized in one of the later sections: It is formulated into a framework for the base for further work for the implementation of simulations of information systems.

One contribution of this work is an operational framework, possible to use as a base for a practical implementation.

1.4.4. Suggestion for Modified Approach in ISD

Traditionally, expensive mistakes in the information systems design practise have had to be done in the target organization: The specifications have been done according to prescriptive principles and then implemented in the hope that they will work.

This can be extended with pre-hoc evaluation. Another method step can be added so that expensive mistakes can be discovered inside a simulation rather than at or after actual implementation.

One contribution of this work is that I open the door to, or at least hint at the possibility of, pre-hoc evaluation as a method step in ISD.

1.4.5. Long-term Impact

In the long run, the results presented in the following are intended to be implemented in a running simulation. The following studies are a few examples of what such a simulation could be used for.

One study the designers could conduct would be comparing two culturally different approaches in the same intended information system: One where "push" is the main principle, and one where "pull" is the main principle. With the simulation it would be possible to study whether it would be more effective if information is pressed out to the target, or if the target has to be active and go and search for the information.

One study could be whether people should be distributed physically according to a horizontal or vertical organization policy: Is it more efficient to place the leaders of the organization together, and the lower levels together in another part of the building? Or should people sit close together according to which task they have in the organization, so that for example the leader in charge of production sits close to the people who conduct the actual production?

One study could be if the organization needs a computerized information system at all. Maybe it would work better with frequent meetings and a physical bulletin board in the lunch room?

1.5. A Case Scenario

The following is an illustration of how a finished simulation suite could be used during a development project. The case is theoretical, as the simulation suite does not exist yet. It is, however, inspired by a pilot project conducted at SwNDC. The present tense is used for convenience only.

The organization is disaster management headquarter. It is contained in a building in a larger city. The purpose of the organization is to co-ordinate interventions in the event of larger disasters, such as fires, floodings or disease outbreaks. In the organization, representatives of police, firefighters, military and city government work together to direct the work forces. Each section has one leading representative, and several command members. For example, the police staff has one leading representative in the cross-section committee, four phone operators, one materiel's co-ordinator and one press officer.

The building containing the headquarters is a two-level office building, where each section has a few rooms available as private offices. There are conference rooms, a cafeteria, a lunchroom, toilets and a reception desk. There is a central command room where section representatives meet and discuss the situation. This room contains aggregated situation awareness information, such as maps, order logs and decision logs.

Before beginning the development, the system is completely analog. Co-ordination takes place through hand-written documents, information about ground units are submitted to the organization via phone calls. Situation awareness is handled through pen marks and needles on a paper map.

However, it is noted that the information system is sub-optimal. Sometimes confusion occurs because the sections have no efficient way of sharing situation awareness. Further, as most information is written down on pieces of paper and then carried from the phone operators to the co-ordinators, it is difficult to track messages and get a good overview. In general, it is felt that the IS could be optimized with current-day technology and a better way of arranging the information flows.

A change analysis is begun through creativity sessions where section representatives meet and discuss how a desired system should look. Several options are specified through a rich picture approach. However, as a continued process would require significant investments in time and resources, it is decided that sandbox implementations of the three most popular alternatives be modeled in simulations. It is further decided that the current situation should also be modeled as a base for comparison.

The sandbox models are tinkered with until the developers agree that they correctly describe the alternatives. The model of the current situation is carefully modeled and compared with the reality until organization representatives agree that it is a fair view of how things work.

With the available simulations, the developers present the alternatives to the city council to get a decision on how to proceed. The council members can study the simulations in which they can see the personnel moving around, carrying messages, answering phones, taking coffee breaks and meeting in corridors. They can see how information pile up and cause congestion in the "current" model, while things seem to flow smoothly in the alternative simulations.

In the end it is decided that one alternative is too costly, and that another where section personnel is mixed, is not culturally feasible. The remaining alternative is selected for more detailed specification.

A more detailed specification is developed using contemporary systems development methods. In parallel with this, the simulation model of the selected alternative is changed and evaluated. While the specification could have been done without the simulation, an additional layer of evaluation is added to the design process, and the simulation forces the developers to be very explicit in their specifications.

Continuously, representatives of the future intended users are invited to participate in the design process. With the simulation they can see in real-time who is supposed to do what and in which order. This gives the users the opportunity to get a down-to-earth insight in the design specification, and also allows them to give detailed suggestions.

At the point where the specification is finished enough for realization and implementation, the finished simulation model is used to complement the end-user training. In the simulation the end-user can watch different scenarios and see how the technology is supposed to be used to handle those scenarios.

1.6. Delimitation

While this work could in theory be used for areas I have not so far studied or considered, the work has been done with a certain focus. The focus has all along been on a *geographically delimited* information system. In practise this has meant an organization contained within one building. All theories and models have been formulated with this as focus. This necessarily narrows the number of application areas. For example, for now only those organizations which do in fact reside within one building are treated. Many organizations do not, and the models presented in this work may or might not need to be modified to take those into account. I have not studied this and make no claims either way. It has been suggested to me that I could as well apply my models to "virtual rooms" and not only "physical rooms", and that might very well be true. However, if it is, then it is the sake for a later study to address.

Further, it is a base assumption that it is a good idea that the right person gets the right information at the right time, and that the main purpose of an information system is to see to it that this happens. The definition of "right" is of course case-specific. I do not take into account opposite views in this question, and do not consider other tentative purposes with the information system.

The view of an information system presented here does not take into account that there might be additional purposes of an information system. There may very well be, and thus what I look at might be considered *a part of* an information system. However, for reasons of convenience I equal the information distribution with the whole of the information system.

Finally, I have accepted as an axiom that an efficient information system is good for organization viability. It has been pointed out to me that there are many viable organizations with bad information systems, and conversely many dying organizations with excellent information systems. I accept this but does not take it into account in this thesis. The thesis is about aiding the design of a good information system, something I pre-suppose is a good thing and do not question.

2. Experiences and Decisions

The following is a description of the work process which has led to the resulting material. An understanding of what happened and what consequences this had is necessary in order to understand why the work has ended up as it did, the relations between the attached articles, and between the articles and the cover paper.

2.1. Background

The classical disposition where background, problem, purpose, method, results and conclusions are stated in order, may certainly be correct and useful in a logical sense. However, it is not necessarily correct in a temporal sense: Things did not necessarily happen in that order. In the beginning of a study, the view of the background and the problem might be not entirely clear. Thus mistakes are made, and the researcher has to backtrack, change his mind, and reformulate large parts of the curriculum of the work. This is not a problem; it is my belief that most social science research happens this way. Part of the whole idea with it is its exploratory nature.

A problem will, however, arise when a text gives the impression that none of this happens, that the way was clear-cut from the start, and no reformulations had to take place. For the interpretation of a text, this leaves out important parts of the underlying process that lead to the stated conclusions.

To paraphrase Cullberg (Cullberg J, 1975), research can be said to be driven by chaos and crises. It is when you get stuck that you learn something.

The following sections describe the major phases in the project so far. They usually coincide with places I got stuck and had to learn something. A part of this discussion is to point out in which context the attached articles were written, and why they sometimes contradict each other.

Before reading on, two things should be noted. First, some terms are here used without being defined. Terms like "information" and "simulation" are defined in later sect1s. I do not feel that full definitions of these terms are necessary for the understanding of this sect1. Secondly, I have consciously given this chapter a rather personal touch, with the explicit view that you cannot fully separate the researcher and the research. Thus, what affected me personally in the research process, also affected the research results.

2.2. Cellular Automata and Military Applications

A frame project called CASID (see below) was initiated in the summer 2001. Apart from my own doctoral studies (which were financed by the university), SwNDC financed travels, materiel and salary for another participant in the project. The closer details of the case scenario and the project settings are not described here, as they can be found in verbatim both in the first article and in the CASID and AMSIDO reports.

CASID, Cellular Automata Simulation of Information Distribution, was conducted in two phases during the later half of 2001. The first phase consisted of a pre-study of the problem and a base for a possible implementation. The second phase consisted of the implementation of a prototype simulation.

The first article attached to this thesis is the result of the pre-study phase. It was submitted to a conference with a very long lead time, and was thus not published until almost a year later.

At the time of the pre-study phase several things were given, something which later on was realized to be unfortunate. One of these things is implied by the name of the project: Even before starting, the simulation method was decided upon. The reason for choosing cellular automata as an approach was that this was at the time an established method, and several of the other project members at SwNDC were using CA or derivates thereof.

That the CA world view was a drastic limitation soon became evident. The main problem with it was its implied severe reductionist approach: a cell (for example representing an individual) could only have very few states, usually only on or off. This implied a drastic reduction of everything related to an information system: no geographical aspects could be taken into account, no planning, very limited variety in the description of the information entity and a very basic description of information behavior.

As these things were run into, we gradually started to steer away from an orthodox CA view. As is evident by the first article, after the pre-study phase we had already adopted a view more close to mobile automata, and were in fact looking longingly towards a view approaching intelligent agents or multi-agent systems.

However, the second phase of CASID started, and an attempt at making a prototype implementation was done. In the beginning of 2002 this prototype was demonstrated at SwNDC. The findings and results were summarized in a report, now called AMSIDO with the sub-title CASID. AMSIDO was an evolution of the original view: Agent-based Micro-world Simulation of Information Distribution in Organizations. At the time of the presentation of the prototype we had thus come to realize that what we were after were really agent based simulations. Still, the prototype built on the former mobile automata view.

While the prototype worked technically, it had lead to the realization that a drastic reductionist approach was not a viable method for making a simulation of a social system. It was, however, felt that the co-operation with SwNDC could be furthered into a second take on the AMSIDO project.

2.3. Social Simulations, Validity and Agents

The project group was expanded and conducted in several parallel tracks, which are visible in the related material: One track focused on previous research in social simulations (Egonsdotter G, 2002), one on validation/evaluation of the constructed simulation (Egonsdotter G & Öberg LM, 2002), one on operationalizations of models (Palmius J, 2002) and one on operationalizations of major concepts (Holmström J & Koli T, 2002) (Bergkvist J & Kristoferson M, 2002). This work was conducted during the spring 2002, and the results were presented at SwNDC in a report collection called AMSIDO-2 (Egonsdotter G & Palmius J & Öberg LM, 2002). To make a somewhat sweeping generalization, perhaps we came up with more problems than solutions in these reports. The conclusions of the main part of the report were that before trying to make a new implementation we would need to make a thorough study on philosophy-of-science aspects of simulation and complexity. Several major problematic areas were identified, such as the problem with how to study a complex system such as an organization, and such as the problem with ensuring the validity of a reduction.

All in all, the various parts of the AMSIDO-2 project pointed towards the realization that the central part of the simulation was the agent, or in other words the human in the organization. This was also the source of the complexity of the simulation.

The survey of social simulations literature had not shown us a viable way forward, since, hardly surprising, none of it had been conducted with a systems-theoretical view. We felt that while the various techniques available were useful, we would need to add something more atop of them to make it all work together.

In fall 2002, the co-operation with SwNDC virtually ended. This happened for various reasons, but in retrospect we probably had different views of the purpose of the project. Unfortunately, the organizational and communicational problems stalled most work for almost half a year, before it was finally accepted that the AMSIDO project would not continue in its current form.

During this period the work that was conducted was focused on the continuation of agent models. As had been suggested, sub-systems of agents should be formalized, and thus I made a work concerning measurement and quantitative modelling of navigation. This was later presented as a bachelor thesis in psychology (Palmius J & Silvergran J, 2003).

2.4. Complexity and Cost

We have now arrived at spring 2003. At this time I was seriously beginning to doubt the sanity of the research area. The status at the time was that we had a list of problems which seemed very difficult to solve, the main funding of the project had been withdrawn and I was essentially left to my own devices to think up a viable continuation.

The experiences so far, and our conclusions of these were summarized in an article which was sent to the CASYS conference. This article was later published in IJCAS (Palmius J & Egonsdotter G & Asproth V, 2003) and is included here as article 2. While this article pointed out some of the major problems we had ran into, it did not fully list the scope of the obstacles to a functional simulation. This was partly because we did not entirely agree about these problems within the remaining project group. This was the last production of what could be called the AMSIDO group.

In the middle of spring 2003 I formulated my major criticism of the idea of a simulation of an information system. At the time I was quite convinced that there was no way to make the things work at all.

This criticism focused on three major areas: complexity, cost and validity. Complexity because both the object system (the real information system) and the simulation would contain so many variables that they would be essentially unmanageable. Cost because the effort of trying to study and pin down the object system in order to transfer it to the simulation would be magnitudes larger than the benefit of performing the simulation. Validity since it would be impossible to make a valid reduction of such a complex system such as a sum-total information system.

In the end of the spring I summarized these things in a letter in which I also suggested that the project be terminated and that I discontinue studies of the area altogether in order to move on to something completely different. In the end, however, it was found that changing subject would not be organizationally viable, and I was recommended to try to make what I could of the current subject anyway.

2.5. Re-defining reality

Since one of the major obstacles was complexity, I put much effort in trying to think up approaches through which this complexity could be managed in such a way that both validity and economic viability could be maintained for the simulations. Essentially the consequence of this was that the nature of the focus of the study, the information system, had to be thoroughly examined and reformulated in order to be made operational.

Through the introduction of isomorphism (explained later) and system hierarchies, I was beginning to feel that a way forward could be found. The findings and conclusions were summarized in an article which was sent to Systems Research and Behavioral Science. In this article the major problems where made explicit, and their possible solutions were discussed.

This article was accepted for publication in spring 2005, and is here included as article 3.

All in all, the problems of the simulation did not only consist of the reality as such, but also of the mental models used to study this reality. If the mental view of the reality was not possible to realize in a simulation, then perhaps the problem was not the simulation, but the mental models. Thus, work continued with the effort of continuing to re-define the mental models to better fit the intended use of those models.

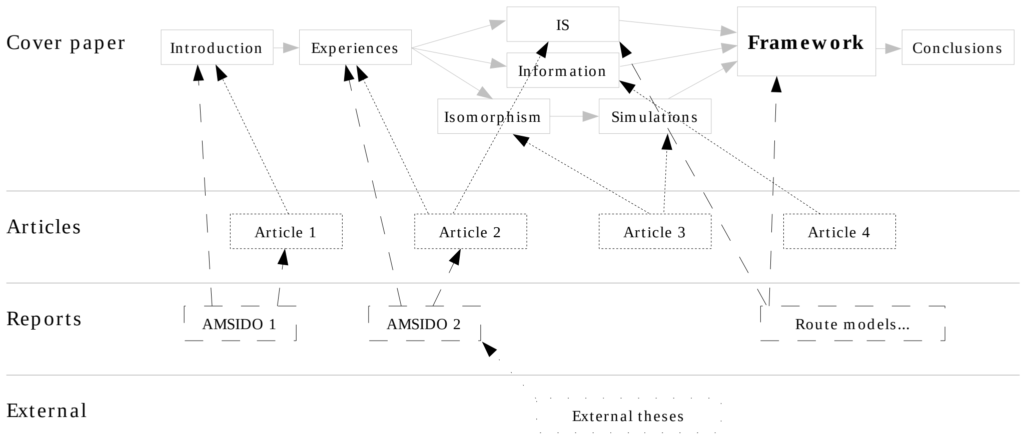
During spring 2005, these operationalizations were made explicit in text. One of the major entities, the "information" was thoroughly treated in an article sent to IRIS. This article is included here as article 4.

At the time of writing this, functional mental models exist for most of the relevant entities of the information system. The information entity is, as mentioned, formalized in one of the articles (and also discussed in this text), while the other entities has not yet been discussed in articles. Discussions about these entities are included further down.

2.6. The Articles and Related Works

To summarize the formal basis for this thesis is the collection of attached articles. Alongside with these, there are a number of unpublished reports, which nevertheless are available on the thesis homepage.

Figure 2. Articles and related works



Article 1 described a technical and somewhat naive view, where the problem with the simulation was viewed as a practical problem of implementation. The basic ideas with the purpose of the simulation and the inspiration for the case scenario are fetched from this article, and can be found in the introduction sections in the cover paper. Article 2 described a more elaborated problematization of the area, a survey of relevant literature, and the realization that more had to be done before the simulation could be make to work. The more elaborate definition of information system, and the base for the simulatory approaches are fetched from this article. Article 3 suggested possible theoretical approaches to managing the obstacles which had been ran into. This is where the base for the isomorphical models were fetched, and also the elaborations of the simulatory approaches. Article 4 presented one of the operationalizations which was one consequence of the suggestions presented in article 3. This is in essence the complete base for the information section.

As mentioned there are other works, for example the AMSIDO reports, which also contain base material for this thesis. These have not been formally published and are not included here, but are available on the homepage for this thesis [http://www.palmius.com/joel/lic/]. This also includes two master theses which I guided (Holmström J & Koli T, 2002) (Bergkvist J & Kristoferson M, 2002), copies of which are available on the homepage, and my bachelor thesis in psychology (Palmius J & Silvergran J, 2003), also available from the homepage.

The AMSIDO reports mainly provided the basis for the first two articles, although ideas presented therein have been implicit for the following articles too. The external theses provided some of the base of the AMSIDO-2 report, mainly the basic operationalizations of what makes up an information system. The bachelor thesis provided parts of the ideas of the geographical aspects of an information system, and some if the ideas for the subsystems of the actor in the framework.

3. Information System

The concept of the "information system" is the focus of the simulation and thus one of the core concepts of this thesis. The following section will define this concept and related entities.

3.1. Background

In the words of this article, an information system is all that within an organization that contribute to the distribution of information. The information system consists of nodes capable of holding information, channels able to distribute information and actors acting and re-acting upon the information. Thus the actors - the humans within the organization - are *parts of* the information system rather than *users of* the information system. With this perspective, there is no such thing as an information system without humans.

Since the actors are parts of the system, and since "the system" is our arbitrary grouping of the activities and components we study, the information system can be characterized as a human activity system in the sense of Checkland and Scholes (Checkland P & Scholes J, 1997). The system is an aggregate function of the organization, its people and its technological artifacts. The higher-order object, the human activity system, the "information system" does not exist as a separate tangible object. It is rather an abstraction, a holon, to which it is convenient to attach properties and behaviors. When further mentioned, the term "information system" shall denote this abstraction as described above.

The short-term purpose of the information system is to distribute the right piece of information to the right actor in the right time. The long-term purpose of the information system is to support the viability of the organization through providing a sound base for decision, information merging and organizational development.

3.2. The Definition of Information System

To summarize, an information system is all that, abstract or not, within an organization that contributes to and shapes the distribution of information: Humans, routines, policies, information nodes, spatial design, the location of the coffee machine and the size of the tables in the lunch room.

There is no such thing as an information system without humans. However, it is very possible to have an information system without a single computer and even without electricity.

3.3. The Parts of an Information System

Having defined the information system as such, we can also make the description of the constituting parts more explicit.

3.3.1. Information technology

The most salient part of an information system is the information technology. This is often even confused with the information system as such.

Information technology is a collective name for the artifacts within the organization, that supports the distribution of information. This is any artifact with this function, there is no requirement that the artifact is computerized or electronic.

To abstract and generalize information technology, I will use the term "node". The node is a point of interaction between actors and artifacts, or between artifact and artifact. In most cases the node and the artifact is the same thing.

All nodes shares some common attributes, for example storage capacity (which can be zero), but must also be viewed as a class hierarchy where some properties and behavior only exist in a sub-class of nodes. The actual properties and the actual class hierarchy are implementation dependent, but a few examples of possible properties follows.

To exemplify the node, I will here try to operationalize the concept of a bulletin board hanging in a corridor. I am well aware that the term "information" has not yet been defined (it will be in the next chapter), but for the sake of the discussion, information can here be understood as an explicitly constructed message in any format.

A bulletin board can store information. In practise this is synonymous with notes pinned to the board. Thus the bulletin board has a storage capacity, which can be reached (when there is no place for more notes).

Several people can interact with the bulletin board at the same time (in practise, several people can stand around it reading notes). However, this capacity is not unlimited. At a certain point there will be too many interactions for new actors to be able to interact with the board. Thus the bulletin board has an output capacity of one to many, where many is restricted by its spatial properties. Consequently it has an input capacity of many to one where many in this case is fewer than in the output case (since not as many actors can pin notes at exactly the same time).

Each operation, input and output, has a normal speed. It takes a certain time to pin a note to the board, and a certain time to read a note. Thus the board has an information transmission capacity in both directions.

The range of interactions have definite spatial limitations: An actor must stand reasonably close to the board to acquire information, and even closer to it in order to pin up a note. The board thus has an area of influence.

To further exemplify the properties, now imagine that we place a completely different node at the same place as the board: a video monitor listing information. We can here see that the node properties differ significantly: It cannot accept actor input directly and a smaller area of influence for transmission. On the other hand it can accept information remotely via another node (in this case usually a workstation computer).

With this approach it is possible to operationalize nodes so that they can be compared and simulated.

3.3.2. Actors

The actors (humans) in the information system share many of the properties of nodes, but also have a more varied behavior, and a teleology. An actor, too, has an area of influence, storage capacity and transmission capacity.

For the salient point of interaction, we could exemplify this with two actors meeting in the coffee room. Actor A has an information that he wants to transmit to actor B. In order to do so, he has to be reasonably close to actor B, and actor B cannot be busy with another information transaction.

The actor entity is of course rather more complex than the node entity. It will be discussed in more detail in the framework chapter.

3.3.3. Arena

The arena is the playground for the information system. It is the spatial aspects of the information system. In most cases, the arena is a building within which the organization resides. There are walls, rooms, doors and potential points of interaction between actors. Throughout the arena, nodes and actors are distributed spatially. The actors move around, bounded by the limitations set by the arena.

Since both nodes and actors have areas of influence, the arena becomes an important aspect of the information system: If two actors are located so far apart that they never meet by chance, then one informal information channel is closed. Further, if a node, for example a bulletin board, is placed so that no actor ever passes it, then it will most likely not be an as active part of the information system.

It is my opinion that the spatial aspects of information systems have often been neglected. The physical layout of an information system does most likely have an effect on the efficiency of the information system. For example, if two actors are placed on opposite sides of the arena, they are less likely to speak with each other than if they sit next-door or even in the same room. Because of these aspects, the spatiality of the information system is important to capture in a model.

3.3.4. Channels

Information channels are interactions between information-carrying entities. These channels can be both between node and actor, between node and node and between actor and actor.

The channels are both explicit (formal) and implicit (informal). An explicit information channel is, for example, the connection between a workstation and a video information screen, or the procedure in which one actor always relays information to a superior. Implicit channels happen in chance encounters, for example when two actors meet in the coffee room.

Since channels are both ad-hoc and planned, a large part of them are never known at the time of planning the information system.

3.3.5. Information

The most abstract part of the information system is the information entity. This is more thoroughly discussed in the following sections rather than being treated here.

4. Information

One of the most important concepts related to information system is "information". The following sections are dedicated to defining this concept. Much of this discussion is available in article 4, which is completely targeted at defining information. The following is thus partly a summary and partly an expansion of the contents of the article.

The purpose of these sections are twofold: First, they intend to present a definition of information (in this context), and second they provide a model of the operationalization of information. Together these make up the operational definition.

Information is defined as: *an explicitly constructed message, always ultimately constructed by a sentient being*. The model, however, is not as easily summarized.

4.1. Background

From the discussion of information system, we have have its short- and long term purposes: To distribute information in order to maintain viability. This was said without the defining "information". In a simulation all entities must be explicit and quantifiable to be possible to model. Thus I will specify an operational definition of information in such a way that it will be quantifiable.

First we must make a hard distinction between *data* and *information*. Data is already quantifiable, and has been since it was operationalized by Shannon (Shannon C, 1949). However, and unfortunately, this definition was made within something called "information science", which has led to much confusion. Shannon never measured information. He measured data. Information is the abstract concepts distributed via the bits and bytes in the data transfer, it is not equal to the data transfer. Shannon cannot be used to measure information transfers other than indirectly.

However, due to the initial labeling, data and information are commonly confused. If you speak about an "information system", the association most people get is a heap of cabling, switches and routers, computer stuff, which is used to transfer *data*. The efficiency of such a system is commonly measured with terms like megabit per second, but it does not really say anything about whether the actual information was efficiently transferred.

When constructing an information system, measuring data flows is not sufficient. We must be able to know in advance whether the *information flows* are likely to be efficient. To know this we must be able to quantify information, not data. To quantify information we must know what it is, and the first step down that road is to make an operational definition of the term "information".

4.2. Other Definitions of Information

Information science and infology have spent decades discussion information, and of course there are several available definitions of "information". Why are these not applicable in this context?

First we need to recognize that information science is not really about information as such, it is about the communication of information. This has been pointed out in various summaries, such as by Holmström and Koli (Holmström J & Koli T, 2002), and by Hillman (Hillman C, 1995). Indeed, one of the most important classical bases for information science is called "A mathematical theory of *communication*" (Shannon C, 1949). Thus, it might not be entirely fruitful to discuss the information science definition or view of information as such.

However, whether defined or not, a view of information is implicit within information science. In my personal understanding, data and information are seen as equal. For example, one definition of information is "Information is any input into the system that initiates a change of state" (Vickery B & Vickery A, 1994) (p.30). The authors of this definition do, however, seem to be aware of the problem with the information definition, as they later in the text state that defining information as such would not be fruitful.

Implicitly we could understand information as a data transfer which caused an effect at the receiver. This should largely be a paraphrase of the above stated definition.

Several authors working within the information science paradigm, or close to it, approach the problem in a similar way. Cole, for example, defines information as "that which modifies knowledge structure" (Cole C, 1997a), although he himself sees this definition as not entirely clear and that the distinction between information-as-a-process and information-as-a-thing must be made clear (Cole C, 1997b).

This method of managing the information concept seems to be common within the information science literature. Information is not defined directly, rather it is specified what happens to it, what it causes and how it is treated.

One version of this the way of trying to define information as a process rather than as an entity. Losee, for example, tries his hands on a discipline-independent definition, and defines information as the following "Information may be understood as the value attached or instantiated to a characteristic or variable returned by a function or produced by a process" (Losee R, 1998).

The border between "information science" and "infology" might not be altogether obvious. In this text I shall treat the latter as the "soft" variant which concentrates on meaning, interpretation and understanding, while the former is the hard part concentrating on measurement, quantity and communication channels.

The classifications of which category authors fall into may be considered somewhat arbitrary. For example, Meadow and Yuan (Meadow C & Yuan W, 1997) discusses the distinction between data and information, from a base of mostly information science literature, but end up into definitions that I feel are infological. They end up in viewing information as partly the function of the recipient; that the interpretation process is a part of the information as such.

One of the fundaments of infology is the infological equation "I = i(D,S,t)", that information is the data (S) and the pre-knowledge (S) interpreted (i) during the time (t) (Langefors B, 1995). The important consequences of this are that information is never objective. Information only exists within a subject who has pre-knowledge, and who has performed an interpretation process of a certain piece of data. The objective components of this equation are data and time. Pre-knowledge, interpretation and information are subjective and largely uncontrollable. We can also see a hard distinction between information and data.

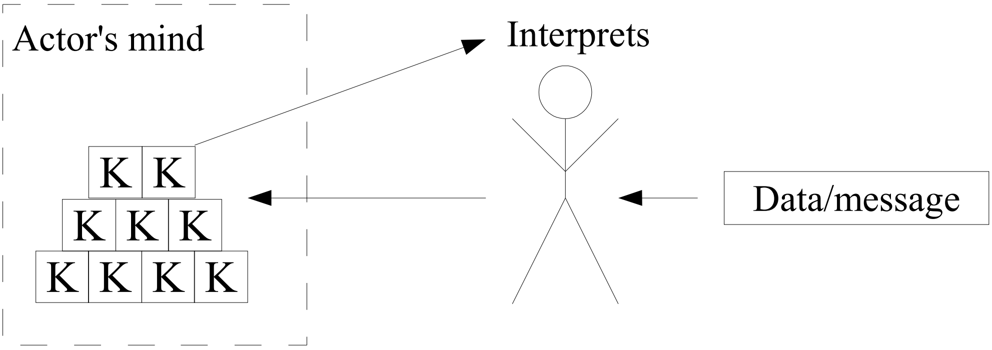
This view could be said to largely coincide with, in Buckland's wording, "information-as-knowledge". Buckland states that this is one of three major approaches for viewing information, the other two being information-as-process and information-as-thing (Buckland M, 1991).

4.3. Concepts Related to Information

In the discussion about information, there are a number of other concepts which can be considered close or related, but which themselves are not the same thing as information. In order to delimit what information in itself is, these other concepts need to be made explicit too.

The following discussion will be concentrated on how the related concepts can be formulated in relation to an operational definition of information. Thus, the approach is to operationalize these concepts too. The main discussion on the actual operationalizations can be found in article 4.

Figure 3. Human understanding

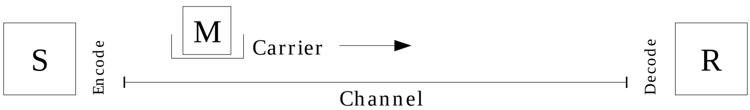


As noted in article 4, many of the concepts related to human understanding have to be reduced in order to be quantified. To start from the left in the figure, the actor has acquired a heap of "knowledge". This is a distinct entity, in practise the same as "information". This knowledge is synonymous with "pre-understanding", "informedness" and "experience". All these terms are thus reduced to being the amount of so far received information.

The interpretation process is when an actor receives new information. Interpretation is the time it takes to accept a new message, and if the message is possible to accept at all. The decision of whether it is, is dependent on if there is enough acquired information there already. Interpretation is thus reduced to time and/or reject/accept, it never transforms the incoming information.

The incoming message is an abstract construct which will be treated further down.

Figure 4. Transmission



Transmission is the entire chain from (but not including) the mind of the sender S, via encoding to a format suitable for sending, through a communication channel on a carrier, to decoding by Receiver R to (but not including) the mind of the receiver. On the route, the information might also be stored for longer or shorted durations.

The "channel" starts when information is transmitted and ends right before it is received. The "channel" consists of at least one medium, at least one carrier and at least one syntax. While information can be conceived as a somewhat abstract concept, it does not exist in an abstract void. It relies on a physical environment to be transmitted, and during the transmission it has a form and a physical representation. The environment through which the information is transmitted is the medium, and through that medium the information is transported using a carrier. For example, if two people are shouting to each other, the medium is the air and the carrier is the sound waves. If sending an email, the medium is computer cabling and the carrier the electronic signals making up the mail.

Decoding is synonymous with interpretation. It is the process in which received information is integrated with the known information. This can be equaled with time. Encoding could be said to be this process but backwards, it is the process of moving known information to a carrier on a communication channel. This, too, can be equaled with time, although not necessarily the same time as for the decoding.

To summarize, "communication" is the transmission of information in a structured way through a medium by a carrier. The term "communication" does here not include the processes required to prepare the information for transmission at the start of the channel, nor does it include the interpretation process at the end of the channel. These delimitations are chosen in order make the "communication" as such objective.

As mentioned before, with the definition of information system used in this article, there is no such thing as an information system without humans. As an implied consequence we could also question whether there is such a thing as information without humans.

As a practical postulate, I will state here that there is not. All information presupposes at least one thinking and interpreting being somewhere. However, and as a distinction against other definitions, the information does not absolutely require two ends of a communication channel.

In a model of an information system, the human entity is a part which needs to be modeled. However, the detailedness of model of this entity only need to fulfill the requirements for the information system model, it does not have to be a detailed model of human behavior. As noted by Schmidt (Schmidt B, 2000) a model does not equal a replica. Here, only a basic representation is required.

For an implementation of a simulation of human information behavior there of course remains much to be said. A model would need to include operationalizations of how humans work, look for information and communicate. However, these models are beyond the scope of this discussion, although some parts of it is addressed in the framework sections.

4.4. Production and destruction

When building a model of information flows, my experience has been that the trickiest part to specify is how and when information arises. Once information exists, making it flow through a model is usually not all that difficult.

To address this we must clearly specify the production and destruction of information. Where does it arise, and where does it disappear? In the following, I will apart from the term "production" also use aggregation as a term related to production. There is a small but important difference between the terms.

As mentioned in the discussion about human information behavior, new information can arise on a statistical basis in a human being. This is one source of information. The new information can be completely new (produced) in which case it has spontaneously arised (representatively as a result of observation, guess or inference), or it can be a combination of earlier acquired information (an aggregation). These are the only internal sources of information within an organization.

However, information systems does not exist hanging in a void. They have input and output, as all open systems. Thus, an additional source of information is external input. This represents information coming into the organization from the outside world, through for example phone calls or emails.

Concerning the destruction of information, it is never explicitly destroyed. Information exists as long as it has a carrier. If a human has received information, then the information is known as long as the human remains in the information system. Information also remains in the system if it is carried by another carrier, such as technological artifact or an archive. From these, the information can be explicitly expunged on the basis of dating.

4.5. An Operational Definition of Information

Now with the curriculum described, it is finally possible to deal with the concept as such. First a short discussion about the reformulation of the infological equation, then the definition as such.

4.5.1. Reducing and operationalizing the infological equation

As mentioned before, one classical base for defining information is Langefors' Infological Equation (Langefors B, 1995), which says that "I = i (D, S, t)", that information is the result of the interpretation process "i" operating on data "D" and pre-knowledge "S" over time "t".

As seen earlier, this definition of information is not suitable as an operational definition, since it explicitly states that information is always subjective, and thus not in practise measurable. However, with some reformulation, it is my view that it could be molded into a usable form.

In a simulation, we would need to pre-suppose a model S for the agents, thus making S implicit in our definition. In consequence with the previously stated definition of pre-knowledge, it is simply the collection of previously known information. Since it is a function of information, it cannot be a pre-requisite of information. Thus we should move S outside the equation for now: it is a question of implementation later, and does not affect the actual definition of information.

The interpretation process "i" is globally defined as the time it takes to accept information. This is according to our previously stated definition that presumes that all information in an information system is ultimately understandable by all parts of the information system. The interpretation process is thus something that *happens to* information, not a part of it. As with "S", "i" should thus be removed from the equation for now, it is a question of later implementation. Further since we already equalled "i" with time, we can also remove "t" from the equation.

That leaves us with the data transmission, which in this case is a meaningful chunk of "something". I am reluctant to call it "data" in this case, since it is the theoretical contents of a carrier transported on an information channel. A better term for D here would here be "message" M. This leaves us with a reduced "I = M", information equals "message", a self-contained package containing meaning, meta-data and other relevant properties. The term "message" implies intention, and thus as a further delimitation, a message is always explicitly constructed by a sentient being. The construction might have been done through using a technological tool, but ultimately, a "message" is always constructed by sentience.

4.6. The information model

This finally makes it possible to define the now tangible entity "information": *Information is an explicitly constructed message, always ultimately constructed by a sentient being.*

This is the base definition which enables an understanding of the thing as such. However, to fully operationalize the concept, we further need to specify its properties and qualities, and specify what it is and is not. This more detailed specification is, however, fits better in the framework section, and can thus be read there.

5. Isomorphism

One of the core concepts underlying the framework presented in the end of the thesis, is that of "isomorphism". The following sections will introduce this concept, and specify in what way it is useful for complexity management and simulations.

5.1. Background

"Isomorphism" and "isomorphic" are concepts that denote that one model is mappable onto another model without being exactly the same. Several authors have discussed this term, but the Principia Cybernetica dictionary defined the concept as: *a one-to-one correspondence between the elements of two sets such that the result of an operation on elements of one set corresponds to the result of the analogous operation on their images in the other set* (Principia Cybernetica, 2005). The definitions specified by principia cybernetica tend to draw towards the somewhat mathematical, but I feel that this is a good base for the further discussion.

For this text, I shall expand this term a bit. It is possible that after this expansion the term might mean a bit more in the way I use it, than in its original meaning. Thus, other authors who use the term might not entirely agree with my use of the term.

Let's start with a transference to a systems-theoretical language. What does isomorphism mean when applied to a system?

First we need to be aware that "system" as it is used here is a concept, a construction. It does not necessarily have a correspondence to a physical object in reality. Thus, when we apply isomorphism in this context, we are modifying or at least working with a mental model.

Secondly, and since system is an arbitrary delimitation of phenomena, it becomes interesting where to draw system borders. With one delimitation of a system, another system may be possible to treat as an isomorphism, while with a somewhat different system border it is not.

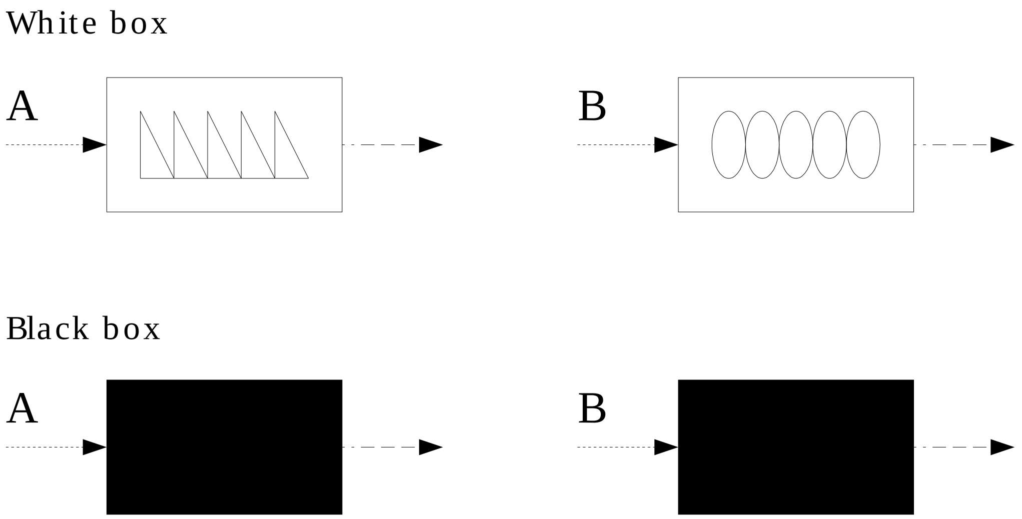
Thirdly, the "set of elements" which is mentioned in the definition above, becomes a set of system properties. The most important of these are input and output.

5.2. Isomorphically Equivalent

The point of all this is to be able to treat one system as if it was equivalent of another system. The term "isomorphism" does in itself denote some sort of equivalence, or at least that the two entities compared are analogous.

The term "isomorphically equivalent" is a construct of my own, to make the denotation of equivalence more explicit. Given that we are working with mental models and abstract constructs, two systems can be said to be isomorphically equivalent when their input and output are the same or at least analogous.

Figure 5. Isomorphic equivalence



To draw a parallel to code, let us assume that system A and system B in the figure both are sub-procedures for sorting an array of numbers, although using different sorting algorithms. They both accept the same input (an unsorted array) and produce the same output (a sorted array) in more or less the same time. In a white box perspective we see the internals, and cannot say that the two systems are equivalent: they look completely different. However, in a black box perspective, all we see of the systems lead us to conclude that they are equivalent.

Isomorphic equivalence presupposed a practical application of a black box perspective on a selected systemic level. When one system is isomorphically equivalent to another, they should in principle be interchangeable, in the same way that the choice between the two systems in the figure becomes arbitrary.

5.3. Isomorphism and Systems Hierarchies

When introducing this term into a system hierarchy, things become even more interesting. We can here imagine a tree of systems, where one node is problematic for some reason. Through the application of isomorphic equivalence, we can replace this node with another node which conforms enough to be an isomorphism.

To make this possible, each systemic level must view the immediate underlying systemic level as a set of black boxes. Each of these black boxes have a certain pattern for input and output, and this pattern is known to the systemic level in focus. With this perspective, the contents of the black box is arbitrary (this is of course the whole point with black boxes). For the systemic level in focus, it does not matter if the contents of one black box are completely replaced, as long as isomorphic equivalence is maintained: input and output should be experienced as being the same.

With this perspective, isomorphic equivalence can be maintained for all systemic levels above the node which was replaced, even when this equivalence is not maintained within the node or through underlying levels.

5.4. Practical Implementation of Isomorphism

Computer scientist would probably associate this approach to the practise used within object-orientation, although the underlying philosophy is not necessarily the same.

I feel that the combination of object-orientation practises and the approach with isomorphic equivalence is a fitting marriage.

To lend terms from object orientation, one object is equivalent of another object when their interfaces, and I/O patterns are equivalent, even when their actual implementation are radically different. Furthermore, two objects can be said to be equivalent in one aspect when they descend from the same forefather and thus share large parts of interface and I/O, despite the fact that they also extend the interface and the I/O. This is part of the whole idea with object hierarchies with objects that descend into more specialized implementations.

For the sake of practical implementation in code, the isomorphic approach can be said to be handled by good object-orientation practises.

5.5. Complexity Management through Isomorphic Equivalence

With these two aspects, the theoretical background and the practical implementation, we can begin to see a way to manage systemic complexity, namely that of encapsulation.

The systemic complexity can be said to be the number of variables and/or states the system has. With this perspective, simulating an information system becomes a laughable project. However, by carefully applying encapsulation so that each systemic level can view its subcomponents as black boxes, the systemic complexity *at each separate level* becomes manageable. For a systemic level, the systemic complexity becomes a question of communication with the higher systemic level, and communication with the black boxes which constitute the underlying systemic level. The sum complexity is thus reduced to a question of managing a number of conversations with explicitly defined interfaces. The internal complexity of each of the black boxes is not the affair of the systemic level in focus, this complexity does for all practical reasons not exist.

This view could be said to be a mix of component-based development and a bottom-up approach, although dressed in another language.

5.6. Isomorphism and Homomorphism

Some systems theoreticians might protest against the above, saying that what I are talking about is a "homomorphism", not an "isomorphism". The difference being that the homomorphism is a model of an object rather than an equivalent of it. Since I am talking about how to represent one thing through another, I should in theory be talking about a model, thus a homomorphism.

This is a sensible argument, and the difference between an isomorphism and a homomorphism becomes a bit fuzzy here. However, for the theoretical discussion I do not talk about one thing being a model of another. I am talking about two interchangeable models. One model is not a model of the other model, it is an analogue. Thus, this is not a tree of reductions, and thus the term homomorphism would be misplaced.

5.7. Conclusions Concerning Isomorphism

In the above I have presented a theoretical argumentation for how models, and submodels, in a systems hierarchy can be made interchangeable. Further I have presented arguments for how this can encapsulate complexity into bubbles that does not leak complexity into the whole. This way a very complex model can be made manageable through clearly specifying the requirements for isomorphic equivalence on each systemic level and then let the sub-components manage their own complexity. This is important, since that lets us finalize building blocks that overtly are very simple, and which thus greatly reduces the complexity necessary to maintain on each systemic level, while still retaining the necessary total system complexity.

6. Simulations

The following sections introduces the term "Simulation" in order to make explicit what is meant by the term. This is necessary in order to follow the later discussion on the framework.

6.1. Background

Before beginning a discussion about simulation technologies in more detail, we need to make explicit what is meant with the term "simulation" in this context. In the terms of this text, a simulation is the animation of a model representing a phenomenon. We have a phenomenon, whether existing or intended, we try to represent it by some means depending on our view of philosophy of science, and we use the representation to study how the model (our view of the phenomenon) reacts to various inputs such as parameter changes or modifications over time.

The purpose of a simulation is (usually) to predict and study an object reality by representing it through an operational reduction. I think most simulationists would be willing to accept this definition of purpose, since it does not enter the debate on validation approaches or how the reduction should be made.

However, I would want to add that an additional purpose of a simulation is to study an object system by proxy. Often, such as in the case of organization-level information systems, the object system is too complex to study directly. In these circumstances, one of the few viable options available is to study the object reality indirectly by using a model as a proxy.

Goldspink (Goldspink C, 2002) reviews some existing theory building concerning the purpose of simulations. He points out important keywords from the social simulation field, phrases like "prediction", "performance", "proof", "discovery", "assess stability" and "construct working systems". All these fits well with what we are after, but I shall here restrict the denotation of the term to the following: The purpose of a simulation is to *predict* and *study* a system in focus by *representing* it through an *operational reduction*.

The purpose with using a simulation is here not to replace existing methods and approaches. The purpose is to elaborate and complement them, in order to paint a more vivid picture of the studied system. The use of simulation within the social sciences is often redundant in the good sense, in the sense that it widens an already existing set of beliefs and models. Or, to cite Marney and Tarbert (Marney JP & Tarbert FE, 2000):

Thus, the purpose of simulation is not to replace traditional social science. Rather, ideally, it is to take the study of behaviour in new directions, which have not previously been possible and to contribute to the relaxation of traditional obstacles to progress in the behavioural sciences.

Traditionally, the social simulations field has been focused on "neat" equilibrium models formulated as equations. These have been targeted more at capturing a whole within the frame of one formula rather than trying to model empirically observed phenomena. However, during the 90's a shift in aim has occurred, and simulationists have become more interested in being empirically close to real-world phenomena. (Pyka A, 2001)

Agent-technologies have been developed from as early as the 70's, but it is only recently, in the end of the 90's, that the field has exploded and started to show a promising future. (Sen s, 1997)

Correlated to this, another shift has been made. Previously, most agent-based simulations were built using simple building blocks. It can be said that rather than being an approach of its own, agent technology were influencing by game theory and cellular automata. Thus, when agent technologies started to become popular during the 90's, most simulations were hybrids with an agent twist.

Lately, agent technology has started to be able to stand on its own legs. Terms like "intelligent agents" has more and more started to denote encapsulated intelligent objects, rather than abstractions.

6.2. Simulations, Models and Simulation Models

If we agree that a simulation is the animation of a model, the question about available models becomes interesting. Flood and Carson (Flood R & Carson E, 1993) present a list of basic model families: The sententical, diagrammatical, mathematical, statistical and logical model families. While we might not agree with all their conclusions about the appropriateness of each (as an example, they seem to have a bias in favor of strictly mathematical models), this seems like a good summary.

However, the problem with this list is that it might lead us to thinking that a model must be any one of these. Following the discussion in the sections on isomorphism, we can see that this is not necessarily so. Instead we can see a new model family: the "black-box container" modeling approach. In this approach, the model is broken down into black box objects, which are supposed to have enough intelligence themselves to realize what their behavior should be in each situation they find themselves in. The only thing the surrounding parts of the model need to know about are the input and the output of the container. If the container itself is built on a statistical, mathematical or logical model is uninteresting outside the container.

The basic proposition here is that when asked if we are following a statistical or mathematical approach for describing the phenomenon we are trying to simulate, we cannot answer, since we follow neither. The basic objects do, yes, but the simulation as a whole does not.

Following the idea with isomorphic equivalence, the internals of the containers do not have to have a correspondence to the phenomenon they model, as long as the I/O patterns are acceptable.

Now, with this approach, can we still be said to model reality? If there is no internal correspondence, can we still be said to have done a simulation of the reality we study? We will here lean on the words on Schmidt, where he discusses the differences between a model and a replica:

A model of the human being is always fundamentally different from the human being himself. Nevertheless, a model is useful and meaningful. Many critics [...] confuse a model with a replica and because of this become engaged in incomprehensible polemics. (Schmidt B, 2000)

In our situation, we claim that this approach is valid for all the entities we model, and not only for the representations of the humans.

6.3. Simulations and Validation

Law and Kelton (Law A & Kelton D, 2000) points out that the validity of a simulation model is related to what purpose the study has. There is no such thing as a valid model in an absolute sense. Further, they point out that there is a distinct difference between a valid model and a credible model. A credible model is a model that the owner of the project accepts as good. A valid model is a model that is a good approximation of the object reality. These two are not necessarily related.

Byrne (Byrne D, 1997) points out that a it is a mistake to treat a simulation as an equivalent to an experiment. The validity in the positivist experiment sense should not be applied to (social) simulations, since the simulation by necessity works through analogues rather than similes.

Also, the completely "correct" model of a social system is a chimera. As noted by Schmidt (Schmidt B, 2000), the human behavior is not possible to *replicate* in the foreseeable future, and that a good *model* (as opposed to a replica) do not need to conform to reality in all respects.

In some sense the goal and/or format of ImAOS could be classified as "Interpretivism" with a "highly constructivist bent" as described by Halfpenny (Halfpenny P, 1997), although I do not agree with his judgement that this approach is necessarily whimsical.

This said it is not entirely uncommon to argue that in order to ensure validity in computer models, those models must be simple and not contain too many variables, but that such a model is too reductionist to be usable in social science. There is a trade-off between validity and relevance, between reduction and complexity.

No matter which balance between complexity and reduction being chosen, a reduction must be done. All models are reductions of reality. The problem is how the reduction should be made, not if it should be done. However, when reducing reality, care must be taken that important entities are treated fairly. The reducer need to know both which entities to represent, and how to name these. We can perceive two abstract realities: The object system and the simulation system. They are abstract even if they exist, because realities are our definitions of those realities. Those definitions consist of a view of what exists (the ontology) and how we describe that (the epistemology). While such definitions exist both for the perceived reality and the simulation reality, they do currently not easily map onto each other: there is a gap between what we can see and describe in the perceived reality and what we can see and describe in the simulation reality. In order to get good sense and value out of the simulation, the ontologies and epistemologies for both realities need be made explicit and mappable onto each other: The relation between the two realities must be clear.

Partly caused by the uncertainty of the world view, the management of communication channels is a source of error. In an ontology based solely on a formal organisation structure, many of the most important actor-actor links will simply not be visible. On the other hand, in an ontology based solely on ad-hoc communication channels, such as agent meeting agent in a corridor on the way to coffee machine, the formal channels will likely be treated unfairly. A detailed world view description is necessary, and it must contain a viable integration between formal and informal information channels. As noted by Carley et al (Carley M et al, 1998), the impact of design perspective in the organisational design versus locally acting agent span, can be quite significant.

6.4. Experiences and Conclusions from the SwNDC Case

With the above as context, we can review the main practical effort which has underlied this work, namely the case study at the Swedish National Defence Center (SwNDC). The conclusions presented below are described in more detail in article 2, so the following is mainly a summary. For a more comprehensive description of the case scenario as such, see article 1.

At the start of the work at SwNDC, the model had already been decided: We were going to use Cellular Automata (CA) to perform the simulation of the information system. This is a mathematical/logical construction which inherently implies a drastic reduction.

Gradually it became apparent that this approach was not feasible when studying a complex social phenomenon such as the information system. The approach imposed several views which were incompatible with our basic view of an information system. These were that it implied geographic staticness, a very shallow systems hierarchy (indeed only one level), complete determinism with no space for teleology, and no space for content in information transfers.

In practise, this was the base for many of the conclusions presented in the previous sections and in the discussion on isomorphism. We felt at the time that if we wanted to represent a phenomenon, that phenomenon should be visible in the representation. A static grid with on/off representations of informedness did in no way make the information system phenomenon visible. Our view of the information system was that it was heterogeneous, with independent actors (both technological and human), while the CA model implied homogeneous and automatic entities.

The conclusion of this was that in order to represent a complex phenomenon, a complex model was needed. To maintain the viability of the model, it would need a deeper systems hierarchy. To construct a face validity in the model, the phenomena to represent needed to be visible as entities in the model: An actor in the reality should be represented as an actor entity in the model.

All this pointed towards the use of multi-agent simulations as basic approach, and the view in this work can be said to continue in that direction although with modifying the basic technology with a systems-theoretical view.

6.5. Simulations and this Work

Now, with the the above as background, what is meant by simulation in this work? And what characterizes a simulation?

The basic definition of simulation remains: It is an animation of a model. The purpose of the simulation is to study, measure and predict an object reality. This far, the view of simulations should not be controversial.

In this work we do, however, look at the special case of a simulation built on the principles of isomorphic equivalence, a deep systems hierarchy and credibility. There are a number of characteristics which separates this approach from "normal" simulation views.

Validity. The simulation model aims at being credible rather than valid in the positivist sense. The model is good if it has *face* validity. That is to say, that if the owner of the problem the model is representing thinks the model is good, then it is good. Good here means that the model treats the "reality" in such a way that the purpose with the model is reached. This purpose is ultimately to support design of information systems. Positivist validity is not required to reach this purpose, in the same way that a rich picture can support a design process without retaining positivist validity.

Complexity. The simulation model does not necessarily aim at removing complexity at all costs. There is no equivalence between a radically reduced model and a validatable or good model. Quite the opposite in fact: Following the principles of isomorphism, a complex behavior in the "reality" should be modeled with a complex behavior in the simulation. Complexity is necessary to keep and manage in order to retain credibility. Note here that complexity should be read as a quantitative measure like "having many subcomponents and internal relations", rather than a subjective measure like "being really complicated". It is my view that with the application of a black-box modeling approach, the model can be "complex" without being "complicated".

Correspondence. The simulation follows the principle of isomorphic equivalence. This means that there is not necessarily any internal correspondence between a simulation object and the object it models, as long as the external correspondence is viable. For example, there is no need to model human cognition as it is performed in the brain in reality, as long as the model produces reasonably viable I/O patterns.

Saying "simulation" brings several unnecessary connotation. It is a term laden with implications which sometimes goes against what is meant with the term here. In the end, I will prefer not to use the term "simulation" at all, but rather speak about an "Isomorphically Acted Organization Scenario". This term will be discussed in the framework.

7. A New Framework

The major contribution of this work is the formulation of a theoretical base for a future implementation of simulations of information systems. The following sections contain the presentation of this framework. The background is focused on some of its more philosophical aspects, while the other sections below are focused on supporting actual practise.

7.1. Background

ImAOS is a theoretical framework with the purpose of initially supporting the practical implementation of a simulation. The long-term purpose is to integrate a reference implementation and API set into the framework, but this does not take place here.

The framework, as it is here described, is a frame of reference. It includes basic definitions, basic models and a description of the basic world view. These are necessary for a concrete specification of an end product, or for the implementation of an end product. However, they should not be confused with such a specification. There is at least one step further to take first. To use life cycle terms, this could be called a change analysis or an analysis, while the design, realization and implementation still remain.

The subtitle of this thesis is "Isomorphically Acted Organization Scenarios" (from now on ImAOS). This is the intended label of the framework. The background of the name can be read in article 3. The main title of the thesis concerns simulations of information systems. Thus we have to interpret ImAOS within that context.

The name ImAOS consists of three important parts: Isomorphically, acted, and organization scenario. To start from behind, let us investigate these parts and clarify their implication for the framework.

First, the organization scenario. This denotes that what we are working with an organization. This is important, and a consequence of the previously mentioned definition of an information system. It is hardly useful to try to study an information system as an entity separate from the organization. A large part of the organization *is* the information system. The organization might consists of something more than an information system, but without an information system there is no organization, and without an organization there is no information system.

The organization scenario is our current model of the organization from the point of view of the information system. That it is a scenario implies that we are working with a theoretical construct which may or might not coincide with a current physical situation. It also implies the possibility that there may be several parallel scenarios that we want to compare.

Secondly, "acted". The denotation of acted is that we want to see the organization played out so that we can see it as it happens. We are not only interested in seeing an end result, part of the use of ImAOS is to see the organization played as in a doll house (see more on this later on). We have actors, usually human individuals, who play a role within our organization scenario. This excludes purely mathematical simulations, since we explicitly want to see behavior, not only results.

Finally, "isomorphically". This is a comment on the underlying model. The model is (to be) built as a system hierarchy of objects which are isomorphically equivalent of the phenomenon they are supposed to represent. The term isomorphically equivalent has been thoroughly discussed earlier.

Thus ImAOS is a model, built on the principle of isomorphic equivalence, for studying (a set of) ways to model an organization-level information systems, in such a way that we can see the models acted.

The roots of the motivation of the simulation model are mainly those discussed in the sections on isomorphism, and those discussed in the section on method in the introductory sections. To summarize the most important consequences and requirements relevant for the formulation of the framework, the following holds true.

An information system is a complex system. Much of what constitutes this system is its complexity. A reduction which removes this complexity will remove the phenomenon we want to study. Thus a certain level of complexity is necessary to retain in the end product.

There is a balance between complexity (implicitly relevance) and positivist validation in traditional simulation models. Either the model is so simple that it is thoroughly possible to validate, but then it is not relevant since it has been too severely reduced. Or it is complex enough to retain at least a face relevance, but then it will contain so many variables that it is virtually impossible to validate. Both these aspects needs to be handled in such a way that we can get a model which is both relevant and which can show some measure of validity.

Since we already said that a salient part of an information system is its complexity, we need to manage this, while still attaining validity. One way of managing this complexity is through black-box encapsulation. Black-box encapsulation is valid in an abstract model because of isomorphic equivalence.

The practical way to formulate the model is through defining isomorphies for important entities in such a way that they fit our end purpose (measurement and simulation). These definitions necessarily affect both the ontology in our perspective, since we redefine the world as such, and the epistemology since we need a good mapping between our terms for the newly-defined reality and the terms we use to manage it.

The philosophy of science underlying the model thus becomes a question of construction. We design the necessary concepts before designing the end product.

So far, the discussion about the simulation has not been technology driven. There is a brief discussion about implementation approaches in the section on simulations, but the model discussions have been more general.

However, the philosophy of science described so far necessarily excludes several simulation approaches. It also implies requirements for available simulation approaches. An approach here is a simulation technique, and not necessarily an underlying model.

From the discussion on models and reductions in the sections on isomorphism and simulations, I will here draw the conclusion that drastically reductionist models and/or models with a very shallow systems hierarchy, are infeasible in the current context. This excludes approaches such as Cellular Automata, Game Theory, clean stochastic models, clean mathematical models and in practise every traditional simulation model from the 80’s and earlier.

However, the approaches which have arisen since the early 90’s become the more interesting; models which do not adopt a drastic reductionism, and which allows for a deeper systems hierarchy. In practise these can be summarized as belonging to the multi-agent or intelligent-agent simulations family, within the social simulations field.

Without going so far as specifying how the implementation will actually look in practise, I can here specify some of its qualities. These are inferred mainly from the isomorphic approach, but also from the general ideas available from within the agent field.

Full isomorphism on the systemic levels within the simulation will more or less require an object-oriented approach. This is what most closely can be used to resemble a conceptual system hierarchy with encapsulated black boxes. A traditional function-oriented or process-oriented approach would be counter-intuitive with the isomorphic view. This is not to say that these approaches could not produce the same results technically, only that they would cause a severe discrepancy between model and implementation. Any of these approaches will, in the end, make up a Turing machine and will thus be at least theoretically mappable onto any of the other approaches.

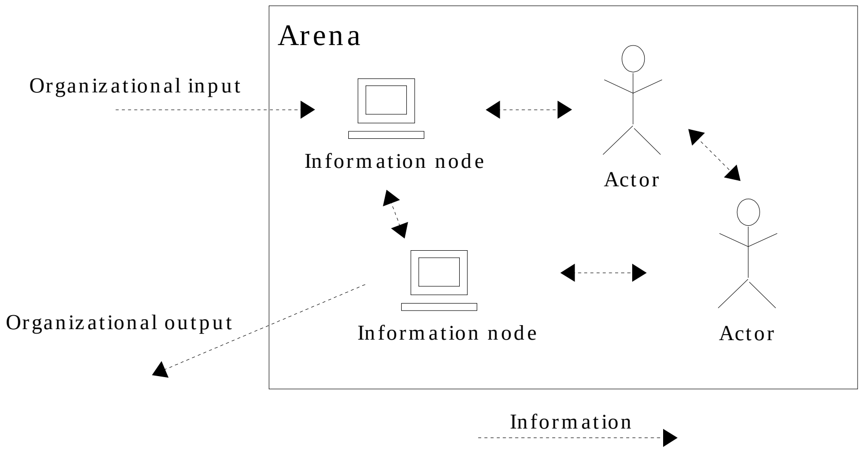
Further, an information system is in most parts asynchronous and parallel. To comply isomorphically with this, the implementation should consist of parallel processes. This is also natural from an agent-oriented view: Each agent is both an autonomous object and an autonomous process, and it interacts with other autonomous objects and processes. Conceptually (although not from a low-level perspective) this differs from the Turing machine idea, since that is usually thought of as serial and synchronous.

A third requirement is that of exchangeability. This might not be exactly required, but it is a desired consequence of full isomorphism. Since we have already specified that two objects with the same input and output are isomorphically equivalent, it is rational that one implementation of a sub-component should be possible to replace with another implementation. More concretely, we could perceive the possibility that, for example, we could have several different algorithms for indoor navigation. The navigation subcomponent of the agent could thus be possible to replace depending on context or purpose. This will by necessity require an implementation with pluggable components, or at least a coherent API.

These requirements, object-orientation, parallel processes and pluggable interfaces, excludes many platforms (programming languages) and implies others. At this stage it is unnecessary to specify such detailed decisions. Suffice to say that I am confident that the technological platform will not be difficult to find.

Now, the following sections will be dedicated to describing the entities in the framework, how they relate to each other, and how they will be used.

Figure 6. Major entities



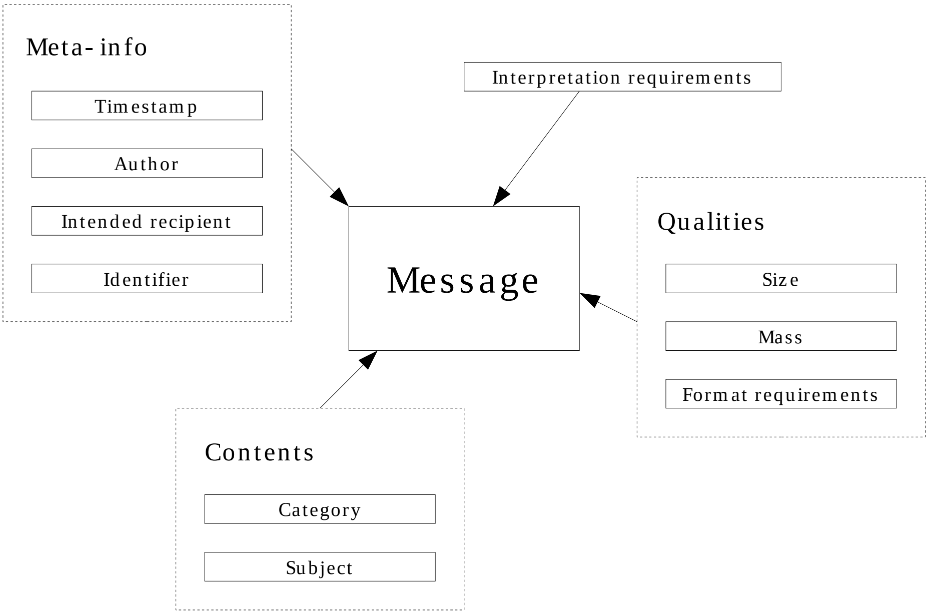
There are four major entities: The arena which represents the physical environment, the node which represents the information technology, the actor which represents the individual in the organization, and information which flows between actors and nodes. Information arise either in the actors, or in the organizational input.

7.2. Information

The information entity has been defined conceptually in an earlier section. While the entity as such is thus defined, we will not be helped until its properties are described. While these properties do not necessarily take part of the operational definition, they are part of the operationalization.

These properties do not necessarily have to be general across implementations where the operational definition is used. The use of them depends on what a model of an information system is intended to be used for. This, said, the following are properties which will likely be relevant for most implementations.

Figure 7. Information



The meta-data properties of "an information", are those things that describes it curriculum without touching its contents. Examples useful for tracing the paths of information through an information system are things like *creation date*, *author*, and *intended recipient*. As noted, since information according to the definition always arise ultimately as caused by an actor, these properties should be possible to automatically assign in a model. Most aspects of information is, as seen above, possible to reduce to time. Thus most of the behavior of information is reducible to time. Thus we can specify the *size* and *mass* of information as the times it takes to interpret it and to send it respectively. Apart from this we also need to specify specific requirements. These are things such as *format requirements*, such as whether this current information is not possible or feasible to transform to certain channel formats. Further we have *interpretation requirements*, which operationally is a list of other information which is necessary to have acquired before this current information can be integrated. As mentioned, we cannot fully model the contents of information: It would void the purpose of the model. However, we can *represent* the information contents through marking information with *category* and *subject*. These are arbitrary labels, which are case-scenario specific. To acquire them we need to study the information of the information system to be modeled. We can then likely find a few large categories (such as "marketing", "sales" and "environment") within each we can find a number of subjects. These subjects does not necessarily have to be named. Each instance of information has a category and a subject, but several instances can have the same category and subject while still being unique.

The category and subject properties are the base for the routing of information: The actors produce information of a certain classification, and look for information of a certain classification.

It is important to make the distinction between the representation of information, which is what we have in the simulation, and the measurement of information in the modelled system. What is called 'information' here, is the *representation* of information. What we are quantifying and tracing, is this representation. This is an abstract entity which is supposed to be a correspondent, in the isomorphical sense, of information, not a replica or a one-to-one representation of an existing information.

Quantification thus becomes the rather trivial issue of keeping track of which "informations" exist, and where they are. The quantification is simply an enumeration. The traceability of information is a implementation-specific issue. Examples of this would be tagging each "information" that has travelled through the information system with a history of the points it has passed.

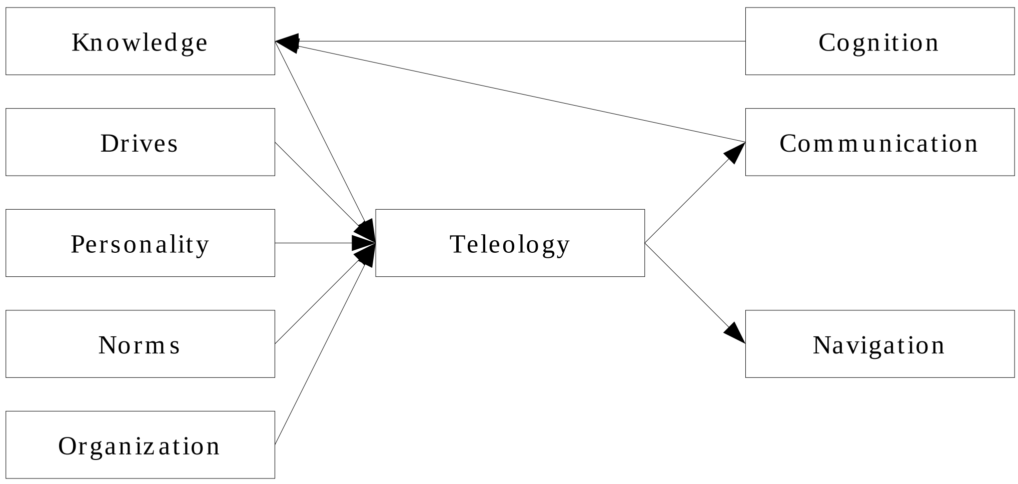
Information arise on a statistical basis at designated points. These points are in actors, as a result of work, and on information points representing input into the organization from the outside world. At the point of modeling the information representation, the designer needs to specify how often an information of a certain type will arise, and where. The information that has been introduced does not vanish unless all actors who had gotten it vanish.

7.3. Actors

The actor is an individual participating in the information system we want to study. As such, he takes an active part in both producing and relaying information.

On the face level, the actor moves around, interacts with technology, interacts with other actors, produces information for output, and accepts information for input. The actor is, however, the by far most complex part of the information system. As such I will decompose it one systemic level more, and discuss the subsystems of the actor.

Figure 8. Subsystems of the Actor



As per the definition of knowledge (from the discussion about the definition of information), knowledge is the number of so far acquired informations. This is thus synonymous with *memory*. Each actor has a memory subsystem which manages the stack of so far acquired information.

Closely related to this is *cognition*. This is the subsystem where incoming information is handled and compared with the memory in order to see if the new information should be merged with the current knowledge. This is also the subsystem where production of new information takes place.

Incoming information comes through *communication*, a subsystem which manages both social communication (interaction with other actors) and technological communication (interaction with a technological artifact).

The actor is exposed to (or actively searches for) points to communicate with by moving around in the organization. The basic functionality for moving around is in the *navigation* subsystem. This is simply the algorithms for moving from point A to point B. Several subsystems participate in compromising around where the target destination actually is.

The *drives* subsystem manages all basic needs, such as hunger or the need to go to the bathroom. The *teleology* system manages the longer-term goals of the actor, such as a plan on how to reach desired information. The *norms* subsystem defines common behavior within the organization (such as how often and when it is relevant to have a coffee break). The *personality* subsystem modifies the behavior slightly, for example in deciding how social the actor is.

The awareness of goals and tasks is managed by the *organization* subsystem. This manages the awareness of the actor’s duties, where he is supposed to be and what he is supposed to do. Thus, this is the formulation of the goals, while the teleology subsystem manages the actual implementation or execution of those goals.

The reason for breaking down actor behavior into subsystems is partly, again, complexity management. By encapsulating complexity in strict compartments and defining slim interfaces, the top-level system complexity can be managed.

Further, the break-down of subsystems does in many cases coincide with areas for which both implementations, good models and algorithms exist. That way, the work of making the actual implementation of one of the subsystems becomes the work of fitting previously made algorithm into the rest of the frame-work. To mention a few examples, there exists research and theory both for agent cognition (Moss S, 1998) (Dautenhahn K, 2000) (Carley M et al, 1998), for agent communication (Lepperhof N, 2002) (Vaucher J & Ambroise N, 2004).

An indication of the availability of models and research concerning these subsystems can be given by Google. At the time of writing this, in summer 2005, the following number of hits was found: For "agent communication" 89000, for "agent cognition" 419, for "agent norms" 121, for "agent personality" 580, and for "agent navigation" 573.

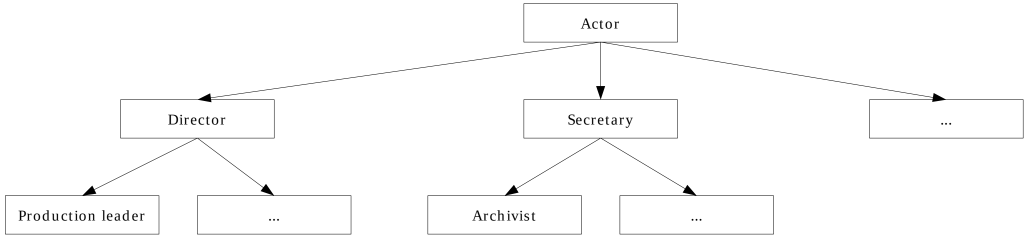
As noted previously, one of the major problems with a simulation like this is the huge task of data entry. In order to describe the information system in a reasonably believable way, all the participants needs to be modelled in such a way that they move around and behave as they are supposed to do. To specifically model every actor from the basic building blocks (the actor subsystems) would not be viable, because both data collection and data entry would cost too much.

To address this, the simulation needs to specify a descendant hierarchy of basic types, of archetypes.

At the top is the basic individual. All individual shares most of their basic behavior, and thus the average behavior should be modeled only once and then either modified or extended in sub types. Below the basic individual, are archetypes for the work situations we are going to model: Secretary, intelligence gatherer, director and so on.

These are further descended into specific functions. An archivist might be a subclass of secretary, and production leader might be a subclass of director.

Figure 9. Actor archetypes



In the same way, archetypes for personality can be constructed. This is implementation-dependent, and may vary, but for example one could consider a Jungian personality classification as a base.

The idea with the archetype/descendant infrastructure is to be able to construct a class library of basic behavior. With a coherent and reasonably complete class library, the data entry of a new simulation could be limited to "this person is an X, but with Y modifications to the basic behavior". This should drastically reduce the cost and effort required to set up the simulation.

The decision whether the behavior should be deterministic or based on stochastic principles does not have a clear-cut answer. This is dependent on the implementation of the subsystems of the actor. It is entirely possible to mix determinist subsystems with stochastic subsystems (which of course would make the top model as such stochastic).

At this point, I feel that a certain level of statistical distributions in behavior is relevant. The archetypes are not exact, nor are their descendants. They do not even attempt to be. However, by specifying a span within which behavior can move on a statistical basis, the archetypes should be able to include most normal behaviors.

Thus, the archetype for a secretary role will not specify that he spends 37.5 minutes reading email and then 5.5 minutes speaking in phone before going on a coffee break. It is more relevant to specify that the role spends between 30 and 50 minutes on email with a mean around 40, and a specified deviation.

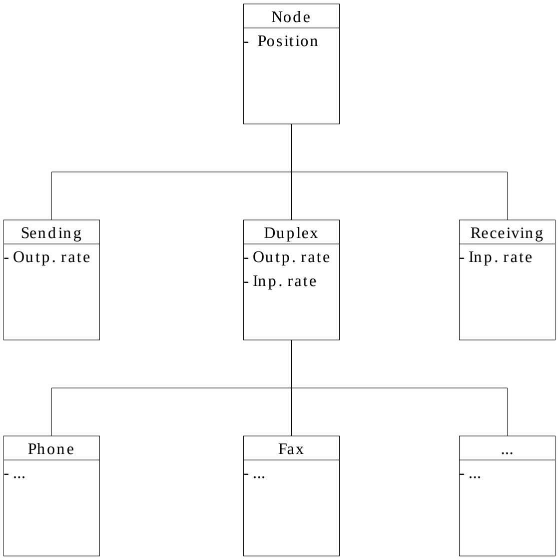
7.4. Information technology

Information technology is all, except the actors, that contribute to information distribution within the information system. This can be artifacts ranging from a physical bulletin board to a mail sever. In the simulation the artifact is represented as a ’node’, a point where information is collected and/or distributed.

As with actors, the nodes are represented through a hierarchy of descending archetypes, which is at least in part implementation-specific. At the top of the hierarchy is the abstract node, which in itself does not have many properties. One of the few that are general for all information nodes is their physical existence in the arena, or in other word ’position’.

Further down in the hierarchy we find more concrete nodes, which have properties related to their information distributing character. From the discussion on information technology in the sections on information systems earlier we can find a few examples of properties that these nodes have: speed for accepting information, speed for transmitting information, storage capacity, range of interaction, and so on.

Figure 10. Node archetype hierarchy



The nodes behave differently according to their implementation-specific characteristics. For example, one class of nodes might only transmit information (nodes like loudspeakers, video screens..) while other might not be able to store information (phones, fax machines..).

7.5. Arena

The arena is the physical enclosure within which the organization scenario is acted. It could be likened with the physical system border of the organization.

When building the framework, the target has always been a geographically enclosed information system, such as one office building. This is still the focus, and the following sections are formulated from this focus.

However, as long as the arena fulfills some basic requirements (that there is a physical enclosure, that actors have contact with each other and that input and output from the system are defined), there is nothing saying that the arena could not be something else. For example, a tent camp in the middle of the forest could work equally well.

But for now, the language used here is that used to describe a building. Thus we have the normal enclosure that we expect to find in a building. The movement of actors is limited and steered by walls, doors and stairs. The walls enclose areas which constitute rooms. Rooms have target functions, such as being a coffee room, an office or a corridor. Apart from that they have no properties.

In the simulation, the actors are thought to have a complete knowledge of the physical enclosure. When they are hungry, they will therefore not need to search for the lunch room. The spot that is the lunch room, and the intermediate areas, are predefined by the arena.

Similarly, actors are tied to offices, or home spots, in the arena. These are the places where the actors normally conduct their work.

7.6. Measurements

So far, the components of ImAOS have been described. The following sections are dedicated to describing the output of the simulation. The output, from now on called "measurement", takes several forms, both quantitative and qualitative. These complement each other.

The main qualitative measurement is the visual impression of the simulation. The intention is that by "acted" in ImAOS we shall be able to see the information play out, as if we were looking into an animated doll house.

The idea with this is that we shall be able to see actors move around, meet, share information, work and make decisions. The use of this is that the developer shall be able to see points where things work out-right wrong, or where they could be improved. This cannot be detected automatically by the simulation, here the simulation is simply a visual support for the developer.

In order to support visual measurement, the system needs to provide a full graphical user interface, able to visualize the scenario both concretely and with abstract layers. The concrete visualization is where the actors and the arena is viewed as such. The abstract support layers consist of functions for tracing information flows. For example, we could color all actors who have been reached by a certain information with a special code. This way we can visually see how an information propagates through the organization.

Further tentative abstract visualizations are sociographs where actor communication are mapped, propagation from an organization-diagram point of view, and coloration of frequently visited areas in the arena.

The main quantitative measurement is "efficiency". This concerns information flow between two (or more) explicitly designated end points. There is no efficiency measurement for the system as a whole. Efficiency is defined as the time it takes for an information to flow from the designated starting point to all the designated end points. In itself this does not say much, but it provides a base for comparison between alternative organization scenarios.

To use this measurement, the developer needs to specify a number of important information routes and then test them with the simulation. This makes it necessary for the developer to in advance be aware of which information and which routes are interesting to study. Some of these can be found through an initial study of the "doll house" described above.

Another quantitative measurement is that of "congestion". This is something which can be discovered automatically by the simulation. As specified in the definition of information, it takes time for the actor to parse. Thus, an actor can only handle to much information within a given time frame. If an actor gets too much information so that it queues up, this is congestion, and it decreases the efficiency of an information route.

Congestion can also happen in technological artifacts, either in the nodes themselves, or in the channels between the nodes. If a node cannot accept more information and thus have to reject incoming information, or if a send or receive queue fills up so that information is temporarily halted awaiting parsing by the nodes, this too is congestion.

As congestion simply is a matter of numbers of informations in queue, this can be easily detected by the simulation.

The complement of congestion is "underfeed". Underfeed is when an actor is not reached by information he wants, or when he is idle awaiting information. Since in the best of worlds, everyone would want the right information at the right time, underfeed is a token of a suboptimal system.

Underfeed, too, can be quantitatively detected by the simulation. An actor which is idling is underfed. Underfeed can also be applied on nodes and channels. A channel or a node which is never or almost never used to transport or hold information is possibly superfluous.

7.7. The use of the framework

The use of the framework must be separated into two questions: What is the immediate use of the framework, and what is the intended long term use of the simulation. The first question will be given an explicit answer below. The second question has been addressed previously, see for example the introductory sections.

The most salient use of the framework is of course its use as a base for building a functional implementation of a simulation of the kind described in this text. For this, the framework is just that: a framework. It specifies the rough shape of the simulation model. It further specifies a frame within which a more detailed specification could be made. For the case of several developers working on the implementation of the simulation, it serves as a common base of understanding of the purpose and goals with the simulation.

Related to the implementation is the identification of subsystems that need collected models, reformulated to fit within the frame of the simulation. The framework implicitly identifies areas where algorithms should be searched for or developed: Areas such as agent cognition, information node behavior and so forth. Many of these algorithms already exist within the broader community of agent systems and social simulations.

More related to the use of the simulation is the framework's relation to information systems development (ISD) practises. Some ideas in the framework are foreign to current practises, but with the framework it is possible to see how these practises could be slightly modified or extended to accommodate the possible existence of a simulation. With the framework, it is possible to identify positions in the ISD process where a simulation could be used.

8. Conclusions

The following sections summarize the main findings of the thesis, and contains a discussion about the way forward.

The purpose of the thesis was to provide a) a theoretical base for simulations of information systems, and b) an operational framework providing a base for a later practical implementation.

The theoretical base has been presented in four sections, namely the ones on information systems, information, isomorphism and simulations. In these, the theoretical base has been specified in the form of operational definitions, intended to support a quantitative simulation model of an information system.

The information systems section contained a definition of an information system, where it is made clear that the human actors are part of the information system, rather than users of the information system. The parts of the information system were specified and made explicit: the concepts of nodes and actors were introduced.

The information section specified an operational definition of information, in the form of a definition and a model. The definition equalled information with an explicitly constructed message, having certain properties, which were described in a collected model.

The isomorphism section presented the theoretical base for how a very complex model, such as one of an information system, could be viable. This through the introduction of black box hierarchies and isomorphic equivalence.

The simulation section provided a base for how the isomorphical approach could be utilized to turn the model into a viable simulation. This was essentially through combining the isomorphism approaches with established object-orientation practises.

All in all these theoretical sections provided the grounds for the formulation of a collected framework which modeled the entities in an explicit and quantifiable manner, and provided a base for how the models could be turned into a practical simulation implementation.

While not being a final design specification, the framework still provides the necessary means for building such a design specification. A later implementation of the simulation now feels possible. As such, I feel that the purpose of the thesis has been fulfilled.

The method utilized through the thesis and most of the work underlying that of the articles has been a constructivist approach, or to use informatics terms, a design approach. This has made it possible to construct a new functional whole, but also brings the downside of the problems with discussions around validity: Before the findings in this thesis have been put into practise, it is not entirely certain that they are valid or good. I believe they are, and I believe my argumentation shows that they are, but I also admit that it is not proved that they are.

This thesis is a step in a process which is expected to continue. So far, the theoretical base has been provided. With this done it is now possible to continue with a practical implementation.

In practise this means that the framework should be implemented in a reference API, and that that API should be used to construct a running simulation. In order to do this, algorithms for all sub-components specified needs to be constructed or imported from existing implementations. In order to iteratively test the simulation application, a case scenario is needed. This is what I expect to do for the PhD thesis.

Another area which is open for continuation, is that of the implications of the isomorphic approach to complexity management. This needs a thorough systems-theoretical analysis, and a more formal ontology and epistemology. What has been specified in this thesis are the rudiments necessary for formalizing the framework.

Once there is a functional simulation available, several interesting studies are possible to conduct in an easy manner. For example, the effects of spatial positioning of staff inside an information system could be studied, something which I expect could yield interesting results.

With the existing results, and what can be seen of the way forward, I have good hopes that one day a functional simulation be made available.

References

G Egonsdotter and LM Öberg (2002): *AMSIDO 2: Validation. . Unpublished report, http://www.palmius.com/joel/lic/a2validation.pdf*

G Egonsdotter, J Palmius, and LM Öberg (2002): *AMSIDO 2: Main Report. . Unpublished report, http://www.palmius.com/joel/lic/a2mainreport.pdf*

J Palmius (2002): *AMSIDO 2: Social Simulations. . Unpublished report, http://www.palmius.com/joel/lic/a2socsim.pdf*

G Egonsdotter (2002): *AMSIDO 2: Social Simulations. . Unpublished report, http://www.palmius.com/joel/lic/a2socsim.pdf*

M Buckland (1991): *Information as Thing.* Journal of the american society for information science vol 42 no 5

D Byrne (1997): *Simulation - A Way Forward?.* Sociological Research Online vol 2 no 2. *http://www.socresonline.org.uk/socresonline/2/2/4.html*

M Carley, M Prietula, and Z Lin (1998): *Design versus Cognition: The interaction of agent cognition and organizational design on organizational performance.* Journal of Artificial Societies and Social Simulation vol 1 no 3. *http://jasss.soc.surrey.ac.uk/JASSS/1/3/4.html*

P Checkland and J Scholes (1997): *Soft Systems Methodology in Action.* Wiley, USA

C Cole (1997a): *Calculating the information content of an information process for a domain expert using Shannon's mathematical theory of communication: A preliminary analysis.* Information processing & management vol 33 no 6

C Cole (1997b): *Information as a process: the difference between corroborating evidence and "information" in humanistic research domains.* Information processing & management vol 33 no 1

J Cullberg (1975): *Kris och utveckling: En psykoanalytisk och socialpsykiatrisk studie.* Natur och kultur, Sweden

K Dautenhahn (2000): *Human Cognition and Social Agent Technology. Advances in Consciousness Research*John Benjamins Publishing Co, Holland

R Espejo and R Harnden (1989): *The Viable System Model.* John Wiley and Sons Ltd, England

K Ewusi-mensah (1997): *Critical issues in abandoned information systems development projects.* Communications of the ACM, ACM Press, USA vol 40 no 9

R Flood and E Carson (1993): *Dealing With Complexity, 2nd ed.* Plenum Press, USA

C Goldspink (2002): *Methodological Implications of Complex Systems Approaches to Sociality: Simulation as a Foundation for Knowledge.* Journal of Artificial Societies and Social Simulation vol 5 no 1. *http://jasss.soc.surrey.ac.uk/JASSS/5/1/3.html*

P Halfpenny (1997): *Situating Simulation in Sociology.* Sociological Research Online vol 2 no 3. *http://www.socresonline.org.uk/socresonline/2/3/9.html*

C Hillman (1995): *What is information?.* [ej fullständig referens]. *Pappersform. "Mathematics subject classification, primary 94a17, secondary 94a15*

J Bergkvist and M Kristoferson (2002): *Information node characteristics for model construction.* Department of Information Technology and Media, Mid Sweden University, Sweden. *Master thesis in informatics http://www.palmius.com/joel/lic/infonode.pdf*

J Holmström and T Koli (2002): *Making the Concept of Information Operational*. Department of Information Technology and Media, Mid Sweden University, Sweden. *Master thesis in informatics, http://www.palmius.com/joel/lic/infoop.pdf*

B Langefors (1995): *Essays on Infology.* Studentlitteratur, Sweden

A Law and D Kelton (2000): *Simulation Modeling and Analysis.* McGraw-Hill Higher Education, USA

N Lepperhof (2002): *SAM - Simulation of computer-mediated negotiations.* Journal of Artificial Societies and Social Simulation vol 5 no 4. *http://jasss.soc.surrey.ac.uk/JASSS/5/4/2.html*

R Losee (1998): *A discipline independent definition of information.* Journal of the american society for information science vol 48 no 3

P Kruchten (2003): *The Rational Unified Process: An Introduction, Third Edition.* Addison-wesley professional, USA

JP Marney and FE Tarbert (2000): *Why do Simulation? Towards a Working Epistemology for Practitioners of the Dark Arts.* Journal of Artificial Societies and Social Simulation vol 3 no 4. *http://jasss.soc.surrey.ac.uk/JASSS/3/4/4.html*

R Martin (2002): *Agile Software Development, Principles, Patterns, and Practices.* Prentice Hall, USA

C Meadow and W Yuan (1997): *Measuring the impact of information: Defining the concepts.* Information processing & management vol 33 no 6

S Moss (1998): *Critical incident management: An empirically derived computational model.* Journal of Artificial Societies and Social Simulation vol 1 no 4. *http://jasss.soc.surrey.ac.uk/JASSS/1/4/1.html*

J Palmius, G Egonsdotter, and V Asproth (2003): *Simulations of Highly Complex Social Systems as a Tool for Designing Information Systems.* D Dubois (2004): *IJCAS. vol 14*

J Palmius and J Silvergran (2003): *An Evaluation of Ideal Route Models of Indoor Navigation.* Department of Social Sciences, Mid Sweden University, Sweden. *Bachelor Thesis in Psychology*

Principia Cybernetica (2005): *Web Dictionary of Cybernetics and Systems - Isomorphism.* Principia Cybernetica Web. *http://pespmc1.vub.ac.be/ASC/ISOMORPHISM.html (visited july 2005)*

A Pyka (2001): *Special Issue on Applied Simulation Analysis.* Journal of Artificial Societies and Social Simulation vol 4 no 3. *http://jasss.soc.surrey.ac.uk/JASSS/4/3/0.html*

B Schmidt (2000): *The Modelling of Human Behaviour.* SCS-Europe, Belgium

S Sen (1997): *Multiagent systems: Milestones and new horizons.* Trends in Cognitive Sciences vol 1 no 9. *Elsevier Science Ltd*

C Shannon (1949): *A mathematical theory of communication.* University of Illinois Press. *[Endast som papperskopia, hitta fullständig referens]*

J Vaucher and N Ambroise (2004): *JADE Tutorial and Primer.* Dep. d'informatique, Université de Montréal. *http://www.iro.umontreal.ca/~vaucher/Agents/Jade/JadePrimer.html*

B Vickery and A Vickery (1994): *Information science in theory and practise.* Bowker-saur, UK

Article 1

Article 1

This article was first submitted to the UKSS’02 conference, and then printed in the conference proceedings. The full formal reference is thus:

Egonsdotter G and Palmius J (2002): *A Base for Simulating Information Distribution*. From Ragsdell R et al (ed) (2002): *Systems Theory and Practice in the Knowledge Age*. Kluwer Academic / Plenum Publishers. USA.

The article has been reformatted to fit into the current work. Further, the email addresses to the authors were changed to what is currently correct.

A Base for Simulating Information Distribution

G. Egonsdotter (gunilla.egonsdotter@miun.se)
J. Palmius (joel.palmius@miun.se)

1. Introduction

A common problem with implementation of new communications hardware is that it is difficult to predict their effects on the information flow in an organisation. This leads to hardware investments, which are at best based on qualified guesses that they will improve the situation, and at worst based on half-pronounced promises from the local computer salesman. To help this problem it would be desirable to be able to simulate the information flow in an organisation in such a way that the effects of new hardware implementations could be studied. As the information flow in an organisation can be said to be a complex system, a Cellular Automata (CA) approach is appropriate to build such a simulation. CA approaches have been used for several simulations of complex situations before, as an example, models of combat (Dockery J & Woodcock A, 1993) and models of traffic flow (Wahle et al, 1999). In order to create a simulation, behaviour and information flows must be analysed and formalised using a computer-parsable notation language.

During early fall -01, the Cellular Automata Based Simulation of Information Distribution project (CASID) was started as a joint effort between Mid Sweden University (MSU) and Swedish National Defence College (SwNDC). The main participants were doctoral students from the centre for Research on Anticipatory Systems (RAS) at MSU, and from the AQUA group at SwNDC. The intent with the project is to construct software with the ability to simulate how information distributes in an organisation.

The whole project spanned over the entire fall -01, but the first part of the project consisted of charting, analysing and making operational observed information flows in two exercise organisations. In other words, the first part consisted of an exploratory phase with the purpose of deciding whether it was reasonable to think a simulation could be built. This paper describes the preliminary results of this first part.

The need to be able to study how information distributes through organisations has lately been a topic for some discussion in connection with ROLF 2010 and AQUA at SwNDC (see below for explanations of ROLF and AQUA). It was felt that new solutions for information distribution were implemented even when the knowledge about their effects was incomplete or lacking. When formulating the project plan for CASID it was thought that one way to get a better base for decisions was through analysing information flows, behaviour and communication and build a simulation thereof.

Both the Swedish Defence Research Agency (FOI) and SwNDC have been studying information behaviour, and have even made formalised descriptions of communication (Wikberg P & Modeér B, 1999) (Albinsson PA, 2000), but have as far as we know not made a computer-implementable description. Thus one major aim with the first exploratory phase of CASID to make an operational implementable description.

The project described in this paper is formally a part of AQUA, which in its turn was started as a complement to (or part of) ROLF 2010. The Mobile Joint Command and Control System 2010 (ROLF 2010) has the purpose to: "test alternative Command and Control (C2) concepts, work forms, decision-making support and C2 organisations adapted to mobile or movable joint C2 systems for the year 2010" (Sundin C & Friman H, 2000).

AQUA (which is a name, not an acronym) is a research and development project with the "VisionariumTM", a specially designed room with technical infrastructure for supporting dialogue and decision-making, as its main product. See as an example (Artman H & Persson M, 2000) for description. AQUA has over time developed into an umbrella covering several sub-projects concerning specific issues within the science of Command and Control. One of the planned applications of the results of the CASID project is to be able to study the effects of different configurations of information infrastructure, as an example the implementation of a VisionariumTM.

2. The Case

SwNDC annually arranges exercises in strategic command as a part of the Programme for Advanced Command Studies. In these exercises, headquarters of command are formed in each of the two groups "battling" each other. The exercise tries to represent a crisis during a wartime situation. The staffs are modelled according to current North Atlantic Treaty (NATO) standards, something that makes analysing the organisation an easier task. This since the above type of organisation has clearly defined goals, clearly defined resources and is strictly hierarchical. To give a comprehensible description of this organisation here would require far too much space. However, a short summary could be as follows:

As said the staffs are organised strictly hierarchical. During the exercises there are operational and tactical command levels. The operational level is responsible for defining the long-term goals, while the tactical level implements the short-term decisions to reach this goal. Information is gathered "on the floor" and where it is sent upwards in the organisation for aggregation and analysis.

The exercises try implement the organisation during a situation of crisis, such as during a war, but in a somewhat smaller scale. As an example, sometimes a few persons will represent a whole staff.

The people participating in the exercise are a very homogeneous group. They have the same type of education, the same military language and they have good knowledge of the said hierarchical organisation. Thus they can easily understand and communicate with each other, something which might be different during a real situation.

The staffs communicate internally via an information system consisting of a combination of IT tools (email, phones?), meetings and manual routines (messengers, bulletin boards...). One important part of the exercises is to use and test new methods and tools for staff meetings.

3. Theory and World View

During the last decades, several attempts have been made to simulate human behaviour, and many of these with a specific Artificial Intelligence approach (Vickery BC & Vickery A, 1994). While trying to simulate behaviour and strategies for finding information it has, however, not been the intention of this project to claim that it will implement AI in the traditional sense. The focus has rather been on the information distribution per se, within which the human information-moving behaviours claim a significant role.

It should be noted that when the term "information system" is used in this paper, the humans using the physical infrastructure are considered part of that system. The authors believe that only concentrating on the technical infrastructure of an information system is too limited, since significant parts of the information distribution in an organisation takes places during unplanned encounters in corridors and coffee rooms.

One of the fundamental background theories used in the project is Shannon's Information Theory. It was decided that the traditional model with transmitter, channel and receiver would well describe what the project was trying to simulate. The participants of the project are well aware that it is virtually impossible to handle all the problems of semantics, technology and efficiency described within Information Theory. The military communications are, however, strictly formal, explicit and redundant, properties which might somewhat balance these problems (Fiske J, 1997).

Another fundamental theory, or rather technology, which is the base for the whole project, is Cellular Automata (CA). The choice of CA was made since it had previously been applied successfully in battle simulations (Dockery J & Woodcock A, 1993).

The use of CA necessarily limits the possible variation in worldview when analysing the case situation. First, it is inherent in the technology that systems are described as being very mechanical. Secondly, it is assumed that everything that should be described can be analysed down to explicits without losing anything important on the way: Synergy and relational effects are described as properties of the "cells" rather than modelled separately. Thus, as an example, a group is never described per se. Instead, the individual behaviour when being in the presence of other automatons is modelled.

While this "all contained in the individual" approach might be discussed in relation to its relevance for descriptions of human behaviour, it is thought in the project that it will be a practical approach to construct a good representation of the said behaviour. It might not be entirely exact, but it could paint a usable picture.

4. Method

4.1. Scope

Falling outside the scope of the project (and therefore outside this paper), is the actual content of information. To model the significance and consequences of information content would prove far too complex to be possible to handle. Therefore, all material has been studied with the intention, as far as possible, to exclude all information content. The authors realise that this delimitation is a drastic reduction of the situation.

Another aspect is the question about how general the results can be said to be. As mentioned, the modelled situation is a military organisation in crisis (in this case a war situation). The authors cannot, because of this, claim that the results are representative for "normal" organisations.

4.2. Methodological Steps

The first step in the project was to do an inventory of existing material about the behaviour and organisation of the said exercises, and in the practise about what the exercises were built to emulate. As SwNDC has significant amounts of documentation, it was necessary to limit the search somewhat.

While having formulated a project vision, a purpose and a goal during the project planning, the first actual step of the project was to formulate an operative picture in the sense of Löwgren and Stolterman (Löwgren J & Stolterman E, 1998). In practise this consisted of an exploratory modelling of the different entities identified within the situation. This could be viewed as a kind of exploratory research in the sense of Patton (Patton MQ, 1990).

The information points and the behaviour of the observed individuals in the organisation (this material having been fetched primarily from videotapes of the exercises), were analysed with a base in Shannon's Information Theory.

After having constructed a basic operative picture of the situation, the third step consisted of complementing the material with structured deep interviews of exercise leaders, and back-comparisons with the existing documentation. Further, the results have been continually discussed with a senior representative from SwNDC in order to ensure basic reasonableness.

The results of the back-comparison were fed into a new modelling, which were then compared to the documentation, to form iterative steps in an evolutionary design of the models.

5. Making an Operational Description

Based on the initial exploratory modelling of the situation, it was agreed that descriptions of four entities could provide a sufficient basis for simulating an information system. These were the physical structure (walls, doors, areas?), the information points (bulletin boards, phones, computers?), the person automatons (with their roles and behaviour) and the actual information.

5.1. Information Points

The first and most obvious part of an information system is the physical infrastructure, the points through which the information spreads. In this study the information points are defined as "physical objects with the ability to receive, store and transmit information packets".

The information points were found to be possible to describe using a few general properties. The most notable of these were "send rate", "receive rate", "storage capacity", "medium" and "area of influence".

The send and receive rates are measured in packets per second and define the speed of information transmission between the point and a person automaton. These rates, as well as many of the other properties, are described as a mean and a standard deviation. After discussions with statisticians at MSU, it was in the project group decided that a randomisation from a normal distribution would better describe the values than if the values were chosen from a square distribution. The randomisation function taking the mean and the standard deviation as parameters, and in which r1 = [0,1] and r2 = [0,1], is described as:

$$Nf\left(\mu,\sigma\right) = \mu * \sigma * \sqrt{-1 * \ln(r_1)} * \cos(2\pi * r_2)$$

The storage capacity decides how many information packets the point can contain before being full. The medium is which type of information packets the point can handle. As an example, a bulletin board can accept pictures and text, but not sound.

The area of influence is two values: one describing how close an automaton must be to receive an information packet from the point, and one how close it must be to add an information packet to the storage of the point.

5.2. The Physical Structure

The physical structure describes the "arena" within which the person automatons move and operate. It was observed that a significant part of the information distribution was done in chance encounters within corridors and in the coffee room. Therefore it was decided that the simulation needed to recapture the movements of the individuals, even when they did not explicitly work with information distribution.

In order to provide a basis for the behaviour schemata (described below), key areas were implemented. These areas represent localities where the individuals within the organisation spend time, such as "lunch room", "personal office", "corridor" and "bathroom".

The description of the arena will (when these descriptions are implemented) be modelled in a special editor. In practise this work will consist of trying to represent a plan sketch of the modelled building.

Within the arena, the information points described above will be placed according to the modelled situation.

5.3. Information Packets

The definition of an information package is in this study "a discreet quantity of data". While in the simulation being completely devoid of actual content, they represent "information" in the modelled situation. Therefore the term is "information package" rather than "data package".

The information packages have four main parameters: A serial number, a type, a priority and a class. The serial number is plainly a unique identifier to enable tracing of specific packages. The type is the media through which the information can be spread, at this stage "visual", "text" or "audio". This decides through which information points the package can be spread. The priority is "normal", "medium" or "high", and provides a base for selection when the automatons have several information packages to choose between.

The class is primarily a way of grouping a stream of information packages, in order to provide a base for statistics after the simulation run. The class also decides whether the information package exists when the simulation starts, or if it has to be constructed through an aggregation of other information packages.

5.4. Person Automatons and their Behavior

The by far most complex modelling was that of the behaviour of the person automatons. At the time of writing this modelling is only partly finished, and therefore what is offered below is the preliminary results of this modelling.

The person automatons represent the individuals working within the organisation. As a simulation might contain several hundred person-automatons, it was early decided that it would be impossible to describe all the represented automatons individually. Instead, the preliminary modelling was directed at finding behaviour patterns possible group into a number of distinct classes. As these groups do not necessarily equate with a formal function within the organisation, the term "behaviour group" was chosen rather than "role" in order to avoid confusion when discussing with personnel at SwNDC.

A person automaton consists of several properties and a general behaviour. Most of the properties, and the whole behaviour, are inherited from the definition of the behaviour group to which the automaton belongs.

The behaviour is an aggregate of a number of schemata (such as a schema for "take coffee break"), and a meta-schema controlling the overall strategy of the automaton. The schemata will be triggered through events in the surrounding ("interact with automaton met in corridor"), according to schedule ("take lunch break", "sleep") and according to the overall strategy ("look for information through email in personal office").

The schemata consist of "instructions" or in other words a low-level code language with atomic functions such as "go to area", "wait" and "interact with point". At the moment of writing, this language is at best at a conceptual stage. Apart from the behaviour, the automatons will have properties such as "position in organisation" and "home area".

5.5. XML:ifying the Entities

As a basis for a simulation the models are pointless unless they are possible to implement in code. Therefore, all the modelling were made with the possible ways for implementation as a constant points of reference. Parallel to the conceptual modelling, the project members have worked with describing the results of the models in XML code.

At the moment of writing, preliminary document type definitions (DTD) exist for the physical structure and for the information points. Initial attempts at describing behaviour and schemata have been made.

6. Summary and Conclusions

The above material describes the preliminary results of the first exploratory phase of the CASID project, a material which will serve as a base for the further work. As such, the models and operational descriptions can be seen as somewhat sketchy. However the purpose with the described phase was to collect information and do preliminary exploratory modelling in order to decide whether it could be believed that a simulation of the mentioned kind could be built. With the existing material, the authors of this paper are quite confident that believing so is reasonable.

The further work will be concentrated at polishing the models, and to implement them as Java-based software. Work with formulating DTDs for the description language used in the applications has, as mentioned, already begun. As material of this kind start to mature, it will be made available on the project homepage (Egonsdotter G & Palmius J, 2001).

The initial attempts at trying to describe information distribution through modelling individual behaviour and modelling the structure the individuals interact with, indicates interesting possibilities. The approach to describe human behaviour as a number of schemata constructed by low-level instructions, and aggregating into an overall behaviour, seems promising at this stage.

In spite of the above operational description being a drastic reduction of human behaviour, it is the belief of the authors that a simulation based on the material can serve as useful tool for studying information distribution, and for being a decision support for implementations of new technological structures.

References

PA Albinsson (2000): *Visualisering av loggfiler - ett hjälpmedel för utvärdering av verksamhetsmodeller.* FOA, Sweden

H Artman and M Persson (2000): *Old Practices - new technology: Observation of how established practices meet new technology.* C Sundin and H Friman (2002): *ROLF 2010 the way Ahead and The First Step*The Swedish National Defence Collage, Department of Operational Studies and the Centre for Command and Control Science, Sweden

J Dockery and A Woodcock (1993): *The Military Landscape - Mathematical Models of Combat.* Woodhead Publishing Limited, England

G Egonsdotter and J Palmius (2001): *CASID - Homepage.* . *http://gathering.itm.mh.se/casid*

J Fiske (1997): *Kommunikationsteorier: en introduktion.* Centrala tryckeriet AB, Sweden

J Löwgren and E Stolterman (1998): *Design av Informationsteknik - Materialet Utan Egenskaper.* Studentlitteratur, Sweden

MQ Patton (1990): *Qualitative Evaluation And Research Methods.* Sage Publications inc, USA

C Sundin and H Friman (2000): *ROLF 2010 the way Ahead and The First Step.* The Swedish National Defence Collage, Department of Operational Studies and the Centre for Command and Control Science, Sweden

BC Vickery and A Vickery (1994): *Information Science in Theory and Practice.* Bowker-saur, England

J Wahle, L Neubert, J Esser, and M Schreckenberg (1999): *A Cellular Automaton Traffic Flow Model for Online Simulation of Traffic.* Parallel Computing vol 27. *pp 719-735*

P Wikberg and B Modeér (1999): *Modellering av operativ ledningsenhet - Visuell modellering som metod för problemanalys och operationalisering av begrepp i komplexa system.* FOA, Sweden

Article 2

Article 2

This article was first submitted to the CASYS'03 conference, and then printed in the international journal of computing anticipatory systems. The full formal reference is thus:

Palmius J, Egonsdotter G and Asproth V (2003) *Simulations of Highly Complex Social Systems as a Tool for Designing Information Systems* From Dubois D (ed) (2004): *International Journal of Computing Anticipatory Systems*. Vol 14.

The article has been reformatted to fit into the current work. The style of some of the references was changed. A few spelling mistakes were corrected.

Simulations of Highly Complex Social Systems as a Tool for Designing Information Systems

J. Palmius (joel.palmius@miun.se)
G. Egonsdotter (gunilla.egonsdotter@miun.se)
V. Asproth (viveca.asproth@miun.se)

When designing information systems, it would be good to be able to compare alternatives. However, information systems are complex phenomena as they encompass the humans involved in distributing the information. One possible way of making comparisons would be through simulation. Having constructed a prototype for such a simulation we have seen that the traditional approaches, such as Cellular Automata, utilized within the social simulations field are usable but not sufficient. However, the newer agent-based approaches show more promise. We conclude that in order to make simulations of our kind possible, the new technologies, such as multi-agent systems, need be adapted and extended. One of the pieces missing is an agent-based infrastructure building on anticipatory principles for agent information behavior.

Keywords: Information System, Simulation, Social System, Multi-agent Systems, Social Simulation

1. Introduction

When developing information systems, there is a multitude of available systems development methods to use. In these, there is often functionality provided for ascertaining the systemic desirability of possible actions before they are taken. However, while being based on structured and proven approaches, these methods can only provide an indicator of whether the direction of the development is likely to be good. These predictions often fail, which is shown by the number of systems development projects which have failed on the border of implementation, because it was then realized they were not systemically desirable. Most systems development methods have in common that they are prescriptive and normative; they do not contain prognostic or predictive capabilities.

One possible reason why these predictions do not work as well as one would want is the overwhelming complexity of even small information systems. An information system does not only include cables connecting a collection of more or less intelligent boxes stuffed with electronics, but also the humans using the technical infrastructure. An information system should be seen as the sum total of all that contributes to the propagation of information and all that has an effect on that propagation, within an organization: Humans, routines, policies, geographical locations and technological artefacts to mention a few examples. In order to handle this complexity, a computer-based tool such as a simulation is useful. In combination with the traditional strengths systems development methods provide, simulation could provide capabilities for prognosis and comparison.

For clarity, it should here be pointed out that the kind of simulation we are talking about here is indicative, a study of patterns and probabilities. While sharing the same roots, this simulation is in its approach often very different from simulations of, say, electronic circuits.

Simulation as method is well known and used in social science. In different ways and techniques, simulations had been made to describe phenomenon in the field of social science since early 60's. The simulations are different and include the whole area from macro sociological theories to simulation about human ability to learn. (Halpin B, 1999) (Morreti S, 2002)

Another argument in favor of simulation as a method for the study of societies is the expanded number of studies which use simulation as method. Further another argument is that the number of disciplines involved in simulation as method has grown to encompass fields like cognitive science, biology, and neuroscience. Still another argument is the increase of number of theoretical perspectives involved. (Conte R & Gilbert N, 1995)

Of course there are also opponents against simulation as a method for studying social systems. One of the critics is Byrne (Byrne D, 1997) who discussed one of the main problems with simulation of complex system, namely that of setting up the initial parameters. Because of the difficulties with this, he claimed that the result of a simulation is uncertain. Gilbert and Troitzsch (Gilbert N & Troitzsch KG, 1999) maintain that simulation is an acceptable experimental methodology and that it is possible to change parameters and execute the simulation many times.

Another problem with simulation as a method is to isolate specific phenomena. However, Goldspink (Goldspink C, 2002) claims that even if there are problems with isolating specific phenomena from its environment, simulation can show interesting results. Therefore it is important to not view simulation as a single method. It should rather be seen as a complement to other methods.(Goldspink C, 2002)

The basic idea is so far that simulation is a useful method, perhaps combined with other methods, to study a complex phenomenon, in this case information propagation in an organization.

The field of social simulations is well-developed and contains many useful methods for studying well-defined and well-delimited phenomena. We shall in the following examine a few of these methods, compare them with our experience of trying to construct a simulation of the above mentioned kind, and suggest areas where development need be conducted.

2. The Historical Development of Simulating Social Systems

"Social systems" can be defined in many ways, for example through the definition by Luhmann (Luhmann N, 1995). In his work he maintains that social system is characterized by autopoiesis and have clear and definite boundary. Our definition of "Social system" in this paper is a system with a clearly defined boundary, an input and an output, and it includes individuals together and other elements (such as computers) that interact to form a whole.

As mentioned above, simulation as a method to study social systems has been used for at least 40 years. In the beginning the simulations were not computer based, but rather a kind of social game. Further, statistics and simulation has a long tradition of working together, and some of that kind of simulations had been used to study sociological problems (Halpin B, 1999). One example of that are Monte Carlo Simulations (Miller JG, 1978).

2.1. Different Approaches in Simulation of Social Systems

When studying the different sorts of simulations and techniques it is possible to see two main streams. These two main streams came from different research fields: different researchers have approached various phenomena in social science from different angles. One way to attack social phenomena was through the top-down approach. Researchers have studied macro-sociological phenomena through a couple of parameters. These kinds of simulations produce a comprehensive result from an overview and do not focus on the particular parts: they try to represent the reality through a few important aspects.

The other approach is the bottom-up approach, which in practice implements the simulation through different sorts of agents that often include some sort of artificial intelligence. This kind of simulation tries to represent some sort of human behavior or part of behavior and those qualities are implemented in the agents. The result of the simulation is the aggregated result from the agents' interactions. This sort of simulation has lately become more in use and is now a functional complement to the macro-sociological perspective.

2.2. Macro-Perspective Social Simulation

One example of a macro approach is System Dynamics, with its roots in cybernetics and system theory, an approach which has been useful for describing macro-sociological theories. Forrester (Forrester JW, 1973) worked in the field System Dynamics, and one of his first works was a model about worldwide growth, pollution and population. Another example of this kind of simulations is Jacobsen and Vanki (Jacobsen C & Vanki T, 1996) who study norms through the use of System Dynamics.

Some critics against models and simulation in System Dynamics claim that such models often have a high level of aggregation, some subjective assumptions and a weak empirical base. System Dynamics is still a common way of creating simulations, even though it is not living up to its initial promise. It is a possible way to study sociological phenomena if there is a very stylized model and a careful manipulation of parameters. Then it can be possible to draw some interesting relationship. (Halpin B, 1999)

Another major tradition in social science is Game Theory, which has been used as base for simulations. One common example of game theory is the "prisoner's dilemma", which has been used to simulate to study the emergence of cooperation phenomena. In Game Theory, techniques like cellular automata and genetic algorithm have been used. With these techniques, it has been possible to study individual-level parameters and their effects on the overall outcome.

These kinds of simulations do not attempt to include a complete picture of human behavior; they do rather aim at studying a specific phenomenon, like cooperation, and the effects of parameters on the agents or automata. (Axelrod R, 1987) (Epstein JM, 1997) (Hegelsmann R, 1996) (Kirchkamp O, 1996)

2.3. Micro-Perspective Social Simulations

The other extreme approach for creating social simulation is to create every individual as an agent and from that point of view study the emergence of a pattern. This approach has gained acceptance during the later part of the 90's. Within this approach, the challenge of modeling human behavior is inherent. Even if there are a lot of problems with representing human behavior in a simulation, many researches aim to construct such models. (Schmidt B, 2000) (Moffat D, 1997) (Sloman A, 1997)

In the discussion about whether it is possible to model human behavior, it is important to distinguish between a model and a replica. "A replica is an identical copy of an original. A replica is completely indistinguishable from the original. It appears to be impossible, at least for the foreseeable future, to produce an artificial replica of a human being." (Schmidt B, 2000)

It is not necessary, probably not even possible, that a model of human behavior includes everything in the human nature. Even a quite simple model with focus on the dominant facts concerning the problem could give excellent result. (Schmidt B, 2000)

In the bottom-up approach, techniques like Distributed Artificial Intelligence and Multi Agent Systems are very important. According to Sen (Sen S, 1997) there is ongoing research about problem solving in the context of a group of agents using these techniques. He describes how cooperative agents work jointly on achieving a common goal. According to him the most important parts in the construction of the simulation are:

- How agents decompose goal into sub goal.
- How to solve the organization of the agent and the agent's problem solving protocol that gives the agents the ability to share results and knowledge.
- How do agents keep coherence and problem solving focus?

A lot of research is conducted within the area of agent system. Wooldridge and Jennings (Woolridge M & Jennings NR, 1995) for example describe an agent as hardware or software- based computer system with these properties:

- "autonomy: agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- social ability: agents interact with other agents (and possibly humans) via some kind of agent-communication language;

- reactivity: agents perceive their environment, (which may be the physical world, a user via a graphical user interface, a collection of other agents, the INTERNET, or perhaps all of these combined), and respond in a timely fashion to changes that occurs in it;
- pro-activeness: agents do not simply act in response to their environment; they are able to exhibit goal-directed behavior by taking the initiative."

Wooldridge and Jennings (Woolridge M & Jennings NR, 1995) points out that there are researchers who include more in the agent term. Especially researchers within the field Artificial Intelligence use concepts that are more applied to humans, such as knowledge, beliefs, intention obligation and emotion.

2.4. The Span Between the Different Approaches

The above shows that there are many different ways of representing phenomena in a simulation today. We identify the main difference between the various approaches as that of either describing the phenomenon on an overall level with a top-down approach or through implement the behavior in every single part and let the pattern grow from the parts with an bottom-up approach. We further conclude that there is a significant span between the two perspectives and that a researcher should consider which technique he is using, as it will implicitly place him within one of the streams.

2.5. Things Done in Agent-Based Simulations Relevant for Simulations of IS

Research close to one of our questions has been done by Lepperhof (Lepperhof N, 2002), who has studied negotiations via email as communication media. In his simulation he demonstrates which parameters are important during such a negotiation. He created his simulation using Multi Agent technology and bases his models upon theories such as Herrmann's negotiation principle.

An import aspect in an information system is the communication and interaction between individuals. Models hereof include some sort of interaction between agents or at least some form of interaction between the agent and his environment. This interaction could involve some sort of passing information to each other like in the case of negotiation of contracts. Agents representing humans also need some sort of 'language' to be able to communicate. There is a considerable amount of literature in the field computer languages for communication between agents but it is also obvious that it is a very difficult and problematic area. One way to avoid this problem is to take for granted that messages pass straight between agents. This could be a possible way to solve the problem dependent on the object of the simulation (Gilbert N & Troitzsch KG, 1999)

Loyall (Loyall AB, 1997) has earlier built simulations and created agents that could both act and generate language. This is an ongoing work for development of what he thinks is a key question for creating believable agents.

"Human language provides, among other things, a mechanism for distinguishing between relevant objects in the natural environment. This mechanism is made up of two components -forms and meanings- which must be shared by the community of language users." (Hutchins E & Hazlehurst B, 1995)

Hutchins and Hazlehurst (Hutchins E & Hazlehurst B, 1995) have developed a sort of language for interaction between agents. They use a model based on interacting artificial neural networks. This language consists of shared symbols of form and meaning pairs.

3. Experiences of Trying to Simulate an Information System

During the last two years the AMSIDO project1, aiming at constructing better methods for comparing different potential information systems, has been working on a framework for the simulation of information systems.

3.1. The Case Scenario

The project has so far been conducted as joint project between the Mid Sweden University (MSU) and the Swedish National Defense College (SwNDC), with the bulk of the financing coming from the latter. The incitement for the study from the view of SwNDC has been the evaluation of their "AQUA lab", a laboratory built for the purpose of developing a new technology-aided form for Command and Control (C2).

SwNDC annually conducts exercises, OBS/OPS (Operative Strategic Decision game), as a part of a course in applied war science. The exercise is a decision game for students at the school for experts in political science and students in strategic command, and is a part of the Program for Advanced Command.

The purpose of the decision game exercise is that two "states" (North and South) are confronted in a conflict situation. The decision-making within the "states" are organized in three levels: the political, the military strategic and the operative. Different decisions have to be made by the participants depending on the information they receive. The calculated intention of the opponent has to be taken into account and compared to the own "state’s" goals and resources in every specific situation.

As a way to measure the projected efficiency of the AQUA lab, the AMSIDO project has been trying to build a simulation of the exercise situation. It has been intended that the simulation should be able to quantify and compare the information distribution efficiency of the two "states", or in other words compare two information systems organized slightly differently.

3.2. Things That were Given But Which had to be Rejected

Before the project even begun, there were a few things that were considered as given. The most important of these was that the simulation was to be constructed using a cellular automata (CA) approach and perspective. The project was even initially called "Cellular Automata Simulation?". The reason that this was given was that SwNDC had previously constructed battle simulations using CA, and thus had experience with the approach.

The CA way of thinking implies a drastic reductionist view of the phenomena to simulate. In consequence, the components of the information distribution phenomenon were initially reduced to consist of transference of an inherently abstract token ("information") without any properties, between homogenous "actors" viewed as automatons without will or intention.

It soon became apparent that a model as reduced as this could have no use whatsoever in order to describe a concrete phenomenon: It became too abstract to be possible to map back on the reality which was supposed to be modeled and simulated.

Eventually, the CA view had to be dropped in favor of a more iso-morphic agent-based approach. This was one of the major shifts within the projects. These are described in more detail later in the article.

3.3. Tangible Results so Far

A functional prototype for the simulation software has been developed, consisting of 20000 lines of Java code. This software is capable of interpreting and run a simulation of a model of an organization, consisting of humans ("agents"), information points, information and physical/geographical constraints ("arena").

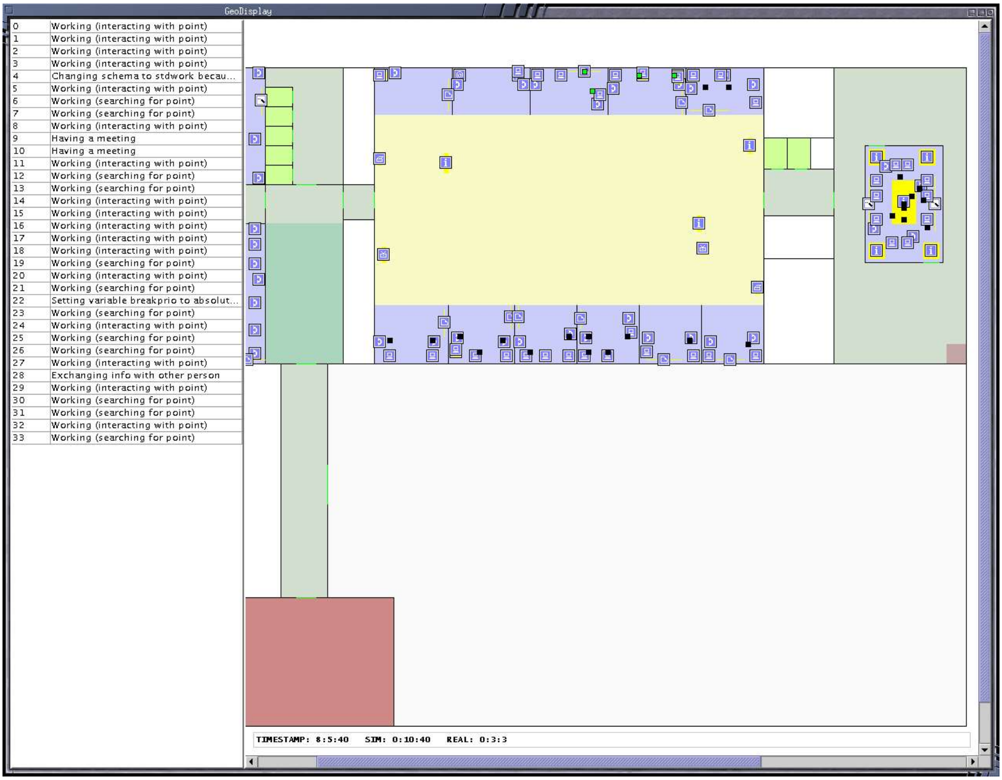


Figure 1: Screenshot of a concept prototype of the IS simulator. A model of a building is visible with humans (the black dots) walking around searching for information.

It should be noted that this is the results after having dropped the CA approach. The prototype is thus built using an agent-approach with intentions and goals, something which became quite different from what a CA approach would have looked like.

In the simulation it is possible to inject information through an information point (such as a computer, a telephone or an information screen) and watch it propagate through the organization via mouth-to-mouth distribution and transmissions between information points.

While this is possible, the simulation has so far not undergone a validation above the level of face validity. In other words, it shows internal consistency and experts of the modeled activity say the underlying models look believable2.

4. Experienced Paradigmatic Shifts

The work during the project cannot be said to have been completely straight-forward. During several periods of the work, the crew has been forced to rethink fundamental concepts. The below is a summary of four of the major shifts in underlying assumptions.

4.1. Cellular Automata to Multi Agent System

The project started with the explicit intention to utilize a Cellular Automata (CA) approach for the simulation. The project was even named "Cellular Automata Simulation of...". The CA base was taken as a given, since much previous work on battle simulation had been successfully conducted with a CA perspective.

However, after the first survey of literature, and after the first exploratory modeling of the system to be simulated, it became quite obvious that CA is too limited to describe such a complex phenomenon as an information system. First and foremost, the CA technology implies homogeneity over the whole grid. This was not feasible since our actors were of several different kinds: Human agents with different tasks and goals, and technological agents in various forms such as computers, phones and bulletin boards. To forcibly reduce all this variety into a synthetic form characterized by a few quantitative or logical parameters would simply completely remove the phenomenon we wanted to study.

Secondly, CA implies geographical immobility. Simulations such as Conway's Game of Life does not really exhibit movement, the "moving" patterns are rather propagation of grid cell properties. This was contradictory to out design goals. One of the things we intended as different in our simulation was that information should be allowed to propagate through chance encounters between humans in corridors and lunch rooms. This would not be possible if the entity that represented the humans was geographically static.

Thirdly and somewhat related to the first point, we found it difficult to represent our most important entity in the simulation. The most important aspect of information propagation is the information that propagates. In a classical CA model, this would be on/off property of the grid cells, and not a traceable entity in itself. We felt this was unsatisfactory.

In the end, it was decided that the focus should be directed towards the construction of a simulation utilizing Multi Agent technology. Instead of trying to fit all entities into a cell grid matrix model, the entities were modeled as entities. The perspective now became that of a game arena on which various agent entities were placed with the task of shuffling packets of the information entity between them.

To summarize this, our experience was that in a simulation of an inherently heterogeneous collection of active objects, a CA approach with its implication of reductive forced homogeneity does not function well. The solution is to move away from CA to a model more supportive of heterogeneity: Multi-agent simulations.

4.2. Mechanic to Teleological

Partly caused by the initial focus on CA, the components of the information system were seen as automatons, or in other words completely deterministic and mechanical. This included the "human" components.

It was, however, soon discovered that the attempt to represent the human as a passive information shuffler, somewhat akin to a network switch, was not feasible. Information in an organization is not directed through static routes.

First and foremost, this view would implement a strict "push" approach to information distribution. Individual A pushes information to B who pushes it to C. In practice things do not work this way. The humans do not only sit silently waiting for an information packet to arrive, they also go about searching for wanted information.

Secondly, and as a consequence of the first point, the humans must have a goal with their information collecting behavior. If a "pull" approach is to be used for describing information collection, it must be known which information it is that the human searches for.

Because of the above it was necessary to start searching for a way to describe a limited teleology in the human agents. It should be noted that the produced code has so far not succeeded in implementing this, mainly because of the lack of a structured model for containing it.

The experienced problem here was that in a heterogeneous collection of supposedly active objects, a way to describe intentionality is required. The solution to this is, again, to move away from CA's forced passivity and aim at a form of teleology implemented in the agent structure.

4.3. Information to Decision

In the beginning, information was viewed as something static that floated around in the information system without any inherent properties. It was then never question where the information came from, something which is partly a problem of the selected case scenario. In the case scenario, the information gathering took place outside the system in focus and arrived to be processed and sent upwards in the organization.

With the base in the CA approach, information was implemented as a serial number without any other properties. The agents in the system either had the information, or they did not. It was, however, discovered that this was not a feasible model of information, and that it could not be used to represent information flows.

Firstly, information is not a collection of opaque homogenous building blocks. Information has properties which determine what will happen with it. As an example, it is difficult to model an information flow without knowing the sender and the intended recipient of the information.

Secondly, information is simply not just distributed and collected. The purpose of the information system is to enable humans to acquire information so they can do something with it. In the case scenario, it soon became apparent that it was not feasible to model the system as simply a shuffling of information upwards in the organization.

To solve these problems, a model for aggregation was implemented. The view now became that humans look for different pieces of information in order to puzzle it together into other pieces of information, something we took to be a good emulation of "decision". This further led to the expansion of the information concept. Information now had to have a few properties to enable humans to decide if it fit with the other information they had, and whether they were a recipient at all.

The problem here is the decision what information actually symbolizes in an organization: What is it there to do? In essence this is a problem of operationalization. The solution, one solution, is to implement the information as a decision rather than as a somewhat abstract token. A decision here is an action leading to a difference in behavior. In other words, the information should be noticed in the sense that it makes a difference somewhere.

5. Discussion

So far this project has suffered from a conflict which is central to all simulations of complex systems, namely that of richness, validatability and usefulness. To phrase the conflict drastically, we can summarize the options as:

The richness option - we can include pretty much everything relevant in the model, and thus capture the richness which is inherent in a social system. With this approach, the human behavior should be modeled fully with, for example, language, teleology and perception. However, such a model would be almost impossible to validate since there are too many independent variables. Further the simulation would be rendered more or less useless, since the cost of filling it with input data would be greater then the gain with the result of the study.

The validity option - we can reduce the model to only include a very limited number of independent variables, so that we can ensure that those variables are correctly described and simulated. With this approach, we should not model a human teleology, and information should probably not be more than a serial number, a homogenous building block. This would, however, be rendered useless since such few variables cannot describe a social system in a satisfactory way. The richness inherent in the system would have been cut off.

The tricky part would be to find a usefulness option. To simply aim at somewhere in between richness and validatability is not likely to be a success since chances are both would become unsatisfactorily handled that way.

As a parenthesis we can also mention that the paraphrased triangular model (Klir GJ, 1988) for the trade-off is not very fair. It implies that as long as the model is neither complex nor exact, it must be useful, which is of course not the case.

Another problem so far has been the staticness of human behavior in the simulation. So far the teleological parts have not been satisfactorily modeled. So far the model has been completely build on mechanical reaction patterns. Events occur which makes the agents change their behavior.

Real humans plan their work in advance and decide on a procedure for reaching a personal goal. They anticipate their needs and act accordingly. This planning occurs both in the long-term perspective (when formulating a strategy for, in this case, acquiring information) and in the short-term perspective (when doing immediate choices about what the next course of action should optimally be).

During the later part of the 90's, much has happened within the field of social simulation. New tools have been made available, drastically extending the scope of what simulations are possible to construct and validate. The social scientist now has access to approaches such as multi-agent simulations, multi-level simulations, BDI architectures and easy-to-use toolsets and APIs for very rich simulations.

We believe that our experiences with trying to apply both CA and agent-based approaches to a complex social system demonstrate that these new tools can extend and better formulate models of social systems, of systems that should encompass a heterogeneity of agents and organizational levels.

The CA approaches are useful and sufficient for many simulations, of course also for quite complex social systems. However, by adopting a more iso-morphic approach, a closer connection to "reality" can be achieved: It is simply easier to compare results between model and reality when the model encompassed the recognizable elements of the modeled reality.

The newer agent-based technologies, as seen in the overview of recent method development, are very useful, but for simulations of highly complex social systems, of clusters of heterogeneous and teleological components, some additions need to be made.

In order to do truly iso-morphic simulations of social systems, the agent architectures would need to be extended with, among other things, weak anticipation. The anticipatory subsystems would drive human behavior in order more closely resemble how humans act in a work situation.

6. Conclusions

In the above, we have seen that the field of social simulation has developed new interesting technologies lately. It is our experience that in simulations of highly complex social systems, such as the simulation of how information is distributed as a base for decisions within an organization, the older technologies, such as CA, do not function well since they are too reductive. The newer technologies, such as multi-agent systems, can provide a more iso-morphic simulatory approach. These new technologies do, however, have to be extended further, for example with weak anticipation. The agent technologies fit well for such extensions, as anticipation and teleology can then be encapsulated in a collection of more intelligent objects, rather than being a subroutine of a complex whole.

We suggest that a coherent framework for anticipation-based iso-morphic simulations be developed. This framework, a framework for Simulations of Highly Complex Social Systems, would need to include models for validation and comparisons with the object reality. In practise this would mean developing not only theories and models, but also collecting consistent APIs and class hierarchies building on validated and reliable models. Such APIs exist to some extent, but the real work would be fitting them all together into a larger whole and adding the pieces missing for the real possibility of simulating highly complex social systems.

References

R Axelrod (1987): *The evolution of strategies in iterated prisoner's dilemma.* L Davis*Genetic algorithms and simulated annealing*Morgan Kaufman, England

D Byrne (1997): *Simulation - A way forward?.* Sociological research online vol 2 no 2. *http://www.socresonline.org.uk/2/2/4.html*

R Conte and N Gilbert (1995): *Computer simulation for social theory.* N Gilbert and R Conte*Artificial societies: the computer simulation of social life*UCL press, England

JM Epstein (1997): *Zones of cooperation in demographic prisoner's dilemma.* . *http://www.santafe.edu/sfi/publications/wplist/1997*

JW Forrester (1973): *World dynamics.* Wright-allen, USA

N Gilbert and KG Troitzsch (1999): *Simulation for the social scientist.* Biddles limited, England

C Goldspink (2002): *Methodological Implications of Complex Systems Approaches to Sociality: Simulation as a Foundation for Knowledge.* Journal of Artificial Societies and Social Simulation vol 5 no 1. *http://jasss.soc.surrey.ac.uk/JASSS/5/1/3.html*

B Halpin (1999): *Simulation in sociology.* American behavioral scientist vol 42 no 10. *pp 1488-1508*

R Hegelsmann (1996): *Understanding social dynamics: the cellular automata approach.* KG Troitzsch, U Mueller, N Gilbert, and JE Doran*Social science microsimulation*Springer, Germany

E Hutchins and B Hazlehurst (1995): *How to invent a lexicon: the development of shared symbols in interaction.* N Gilbert and R Conte*Artificial societies: the computer simulation of social life*UCL press, England

C Jacobsen and T Vanki (1997): *Violating an occupational sex-stereotype: Israeli women earning engineering degrees.* Sociological research online vol 1 no 4. *http://www.socresonline.org.uk/1/4/3.html*

O Kirchkamp (1996): *Spatial evolution of automata in the prisoner's dilemma.* KG Troitzsch, U Mueller, N Gilbert, and JE Doran*Social science microsimulation*Springer, Germany

GJ Klir (1988): *Fuzzy sets, uncertainty, and information.* Prentice-hall, USA

N Lepperhof (2002): *SAM - Simulation of computer-mediated negotiations.* Journal of Artificial Societies and Social Simulation vol 5 no 4. *http://jasss.soc.surrey.ac.uk/JASSS/5/4/2.html*

N Luhmann (1995): *Social Systems.* Stanford university press, USA

AB Loyall (1997): *Some Requirements and Approaches for natural Language in a believable Agent.* R Trappl and P Petta*Creating personalities for synthetic actors*Springer-verlag, Germany

JG Miller (1978): *Living Systems.* McGraw-hill, USA

D Moffat (1997): *Personality Parameters and Programs.* R Trappl and P Petta*Creating personalities for synthetic actors*Springer-verlag, Germany

S Morreti (2002): *Computer simulation in sociology.* Social science computer review vol 20 no 1. *pp 43-57*

B Schmidt (2000): *The Modelling of Human Behaviour.* SCS-Europe, Belgium

S Sen (1997): *Multiagent systems: Milestones and new horizons.* Trends in Cognitive Sciences vol 1 no 9. *Elsevier Science Ltd*

A Sloman (1997): *What Sort of Control System Is Able to Have a Personality?.* R Trappl and P Petta*Creating personalities for synthetic actors*Springer-verlag, Germany

M Woolridge and NR Jennings (1995): *Agent Theories, Architectures, and Languages - A Survey.* M Woolridge and NR Jennings*Intelligent agents*Springer-verlag, Germany

Article 3

Article 3

This article was first submitted to Systems Research and Behavioral Science, and was accepted for publication there in 2005. At the time of writing it is still in print, but is expected to appeach in the number 22 issue.

Palmius J (2005): *A Theoretical Base for High-level Simulation of Organisational Information Systems*. From *Systems Research and Behavioral Science*. No 22. [In print]

The article has been reformatted to fit into the current work. The style of some of the references was changed. A few spelling mistakes were corrected.

A Theoretical Base for High-level Simulation of Organisational Information Systems

J. Palmius (joel.palmius@miun.se)

The design of organisational information systems is a study of the complex. An organisational information system in the Human Activity System sense, is the sum total of all that contributes to distributing information within an organisation. This includes the human actors participating in the system. Managing information about such a system is a difficult task. Technical support for measuring organisational information systems and comparing different ideas about organisational information systems would thus be useful. A high-level simulation of information systems could be such a technical support. There are a number of issues making such a simulation infeasible today. These mainly arise from the complexity of the simulation model, and in consequence of the data input problem. However, by introducing isomorphism and object templates, it is possible to solve some of these problems. In order to utilise these approaches, a framework for isomorphic simulations need to be developed.

KEYWORDS: Information system, social system, social simulation, systems design, ImAOS

1. Introduction

In the words of this article, an information system is all that within an organisation contributing to the distribution of information. The information system consists of nodes capable of holding information, channels able to distribute information and actors acting and re-acting upon the information. Thus the actors - the humans within the organisation - are parts of the information system rather than users of the information system. With this definition, there is no such thing as an information system without people.

Since the actors are parts of the system, and since "the system" is our arbitrary grouping of the activities and components we study, the information system can be characterised as a human activity system in the sense of Checkland and Scholes (Checkland P & Scholes J, 1997). The system is an aggregate function of the organisation, its people and its technological artefacts. The higher-order object, the human activity system, the "information system" does not exist as a separate tangible object. It is rather an abstraction, a holon, to which it is convenient to attach properties and behaviours. When further mentioned, the term "information system" shall denote this abstraction as described above.

To clearly make the distinction between the information system in the mentioned sense, and things like software systems or simple computer systems, the term used in the following will be "organisational information system" (OIS). Further, the purpose of the OIS is to keep all parts of the organisation supplied with the information needed to keep the organisation running, and thus not only to supply decision-maker with bases for decisions. Please note that the term OIS has been used in other contexts, but that no reference is made to those contexts here: The meaning of OIS is the one stated here.

With the above definition of OIS, information systems design (ISD) is a study of the complex. By following a set of recommendations and rules, the designer is supposed to collect information about an intended system, construct a feasible compromise between users and management, and produce a technically coherent setup corresponding to all that was wished and wanted. In the best of cases, the setup can be incrementally tested through prototyping, but in the case of larger information systems on the organisational level, this is often too expensive.

To design organisational information systems this way is the commonly accepted approach today, since there simply is no better way to do it. The alternative is ad hoc development which, while not uncommon, often results in badly organised and poorly documented organisational information systems which do not support the organisation in an efficient manner.

The current situation, with complex realities and systems development methods of varying applicability, is usually not questioned unless it leads to an economic disaster in the case of a malformed organisational information system. It is not questioned why we only have normative and prescriptive methods, while there is a distinct lack of ways to do pre-fact evaluation, comparison and prediction of IS concept sketches.

There may of course be reason to question whether the problem with inefficient organisational information systems arise from deficiencies in the development methods, in lack of resources or perhaps even in the inexpert application of the said methods. I do not address these questions. My aim with this article is to demonstrate that there is, at least in theory, another solution available. Namely, that current-day OIS development methods be complemented with another layer, namely that of isomorphic social simulation on an organisational level (isomorphism will be discussed in more detail below).

Until the last few years, the simulation technologies available for use have been too limited to be even considered for massively complex simulations such as an isomorphic simulation of a complete organisational information system including both the human actors and the technological used by these actors. However, simulations of a less complex kind have been widely accepted within a number of social sciences fields, such as economics, for many years (Hanneman R & Patrick S, 1997).

Goldspink (Goldspink C, 2002) reviews existing theory concerning tentative purposes of simulations and points out keywords like "prediction", "performance", "proof", "discovery", "assess stability" and "construct working systems". Intuitively this has a correspondence with ISD.

Traditionally, the social simulations field has been focused on "neat" equilibrium models formulated as equations. These have been targeted more at capturing a whole within the frame of one formula rather than trying to model empirically observed phenomena. However, during the 90’s a shift in aim has occurred, and simulationists have become more interested in being empirically close to real-world phenomena (Pyka A, 2001). Further, agent-based approaches have recently gained momentum, and show a promising future (Sen s, 1997).

In the game industry, life-like simulations have been available for some time. The inspiration for the kind of simulation presented below did initially come from "The Sims" (Electronic Arts, 2004), a game which simulates social relations with an overtly isomorphic focus. The inspiration was that something similar should be possible to do for organisational information systems.

It should be noted that the discussion is mainly based on literature, but also partly on experiences from previous work in close relation to a case scenario at the Swedish National Defense College (Egonsdotter G & Palmius J, 2002) (Palmius J & Egonsdotter G & Asproth V, 2003). Because of this, some interpretations may be skewed towards military practices.

In the following, I will first specify how a simulation could be used and why. Then I will point out some problems that make simulations infeasible today. Finally I will suggest how these problems could be addressed.

2. The purpose and use of a simulation

The application of simulations in the context of this article is to support organisational information systems development through adding additional points of evaluation and elaboration during the development project.

Organisational information system sketches and models can be evaluated through the use of the simulation approach. The evaluation produces measurements relevant for comparing different options. It can also detect congestions and underfeeds. The evaluation is quantitative. Examples of outputs from the evaluation is "information takes in average five hours in propagating from point A to point B", "person C has a projected overload of 125% information messages and can therefore be counted as a point of congestion" and "information directed to person D underfeeds with a projected propagation factor of 43%".

The development can also be elaborated through the use of a simulation approach. This is not an exact output, rather a qualitative study of effects. Information transmissions and actor behaviour can be studied real-time. Examples of output from this can be "look, person A and person B never meets in person despite working close organisationally speaking. Perhaps they should have offices next-door?" and "What happens if we move the main bulletin board from this corridor to the coffee room?". The elaboration takes place in the manner that pieces of a system sketch can be moved around to gain an understanding in options and effects.

Before starting to apply the simulation approach, it is assumed that the user (the systems developer) has an organisational information system he is interested in studying. This organisational information system may be an existing or a planned system, but there must be at least an idea about an organisational information system.

It is further assumed that the developer has a goal associated with working with the organisational information system. This goal might be "to improve efficiency in information propagation" or "to support organisation viability through providing accessible information".

The simulation approach is not necessarily dependent on any specific systems development model and/or method. In this article, it is assumed that the developer approaches the development with a soft (SSM) view, and that there are distinct moments in time when an intervention into the development is feasible.

It is assumed, but not necessary, that the developer retains a Soft Systems Methodology (SSM) perspective (Checkland P & Scholes J, 1997) designing and developing systems. The language in this article builds on a SSM worldview. For example, "the system" is an abstraction, not a physical object. Another example is that the purpose of design is to produce "culturally feasible" and "systemically desirable" systems.

The simulation approach presupposes reasonably large organisations contained within one geographically delimited area, such as a building. There is no upward limit in size, as long as the geographically delimited criterion is met. However, small organisational information systems (say, an organisation with less than 20 employees) will be pointless to simulate. The simulation approach depends on overall human behaviour models and in small samples, the individual-dependent variables will have too much impact for the simulation to perform. Further, in a small-scale IS, the complexity level will approach being low enough to overview, and thus a good developer will perform better intuitively than the simulation can ever hope to contribute.

3. What makes simulations infeasible today?

There are several issues which make the use of simulations as a support for ISD infeasible today. A short summary of these issues are:

3.1. Complexity

The greatest problem with simulating all kinds of social systems is that of complexity. The social system is inherently complex, since the actors (humans) display a very wide variety of behaviour and seldom follow an easily reproducible pattern. The problem boils down to a fundamental conflict: Do we reduce the system in focus to the extent that it is possible to overview, even though this causes the simulation to no longer bear even the closest resemblance to the system in focus? Or do we accept the richness of the social system and try to implement this in the simulation, even though this causes "bloat" in the software and we no longer have detailed understanding of the higher-order systemic level construed of the relation between all the processes and events?

The second problem is in part caused by the first. If we try to describe a very complex reality, we have a huge task to address in data input. For example, if we are to simulate organisations with several actors, each of which have their own agenda, designated role, geographical turf, informal relations and pre-knowledge, then we would of course need to design how the simulation should represent all this. Depending on the chosen balance between complexity and reduction, we would need to enter the detailed information for each new organisation to simulate. This would be a very time-consuming task, even disregarding the fact that we would have to gain the knowledge about the studied phenomenon first (possibly by observing the organisation and actors we wish to simulate).

As a consequence of the data input problem, the cost becomes an unreasonable factor. The cost of performing the observation needed for the data input and for entering the data itself would be many times greater than the perceived gain in having performed the intended simulation.

3.2. Measurements, ontology and epistemology

No matter which balance between complexity and reduction being chosen, a reduction must be done. All models are reductions of reality. The problem is how the reduction should be made, not if it should be done. However, when reducing reality, care must be taken that important entities are treated fairly. The reducer need to know both which entities to represent, and how to name these. We can perceive two abstract realities: The object system and the simulation system. They are abstract even if they exist, because realities are our definitions of those realities. Those definitions consist of a view of what exists (the ontology) and how we describe that (the epistemology). While such definitions exist both for the perceived reality and the simulation reality, they do currently not easily map onto each other: there is a gap between what we can see and describe in the perceived reality and what we can see and describe in the simulation reality. In order to get good sense and value out of the simulation, the ontologies and epistemologies for both realities need be made explicit and mapable onto each other: The relation between the two realities must be clear.

Partly caused by the uncertainty of the world view, the management of communication channels is a source of error. In an ontology based solely on a formal organisation structure, many of the most important actor-actor links will simply not be visible. On the other hand, in an ontology based solely on ad-hoc communication channels, such as agent meeting agent in a corridor on the way to coffee machine, the formal channels will likely be treated unfairly. A detailed world view description is necessary, and it must contain a viable integration between formal and informal information channels. As noted by Carley et al (Carley M et al, 1998), the impact of design perspective in the organisational design versus locally acting agent span, can be quite significant.

As a further consequence of the ontology/ epistemology issue, it is not altogether clear what, more exactly, it is that the simulation is supposed to measure, nor how. Obviously it is "efficiency" somehow, but this term is not yet defined in this context.

It would be difficult and likely pointless to conceive a single-value result for the simulation as a whole. For what would an efficiency of "42" tell us?

3.3. The organisation around ISD

In the end, the problem which is the most likely to stop simulations for systems design, is the context of systems development projects of today. The project life spans tend to become shorter, and the results they produce become quickly outdated. In this stressful environment, adding another time-consuming phase will simply not be an option. Doing things quickly although not entirely optimal will likely be preferred over doing things more carefully but taking longer time.

4. Making simulations feasible

The first task here is to assess just how complex the system in focus need be. Fliedner (Fliedner D, 2001) discusses non-equilibrium systems in hierarchical levels (in a social systems context) and lists a number of population types, their basic tasks and their basic institutions. Paraphrasing him, we could perceive the modeled organization as the sum population. It should be noted that there is a difference between the theoretical sum-total systemic complexity and the actual complexity at one system level. Klir (Klir G, 2001) states that "the first general principle" of a system's complexity is the amount of information required to describe the system. He further states that our ability to understand a system decreases with the number of entities and relations within the system. I shall call this perspective the theoretical complexity. This view is of course relevant if the whole system is perceived as one systemic level. However, when taking encapsulation into account and perceiving the system as several encapsulated systemic levels, the practical complexity of any given systemic level is considerably less overwhelming than the sum theoretical complexity.

When we start to discuss practical complexity rather than theoretical complexity, we can see that one way of handling sum total complexity is through a deep systems hierarchy. This approach is influenced by object orientation and has also been called "embeddedness" (Hanneman R & Patrick S, 1997). The idea is that complexity can be contained in lower systemic levels, in order to leave the higher systemic levels uncluttered, that each functionality, such as human navigation, can be considered a self-contained sub entity. The higher systemic level (in this case the actor object) does not need to know about the details, as long as there is a well-defined interface for input and output. From the point of view of the actor, the navigation sub system is a black box. Taking the step further, to the simulation systemic level, the actors are considered black boxes.

This approach is already employed within software industry, in component-based development. Once a component functions as to specification, there is no longer any need to consider its internals from the point of view of the top-level software. The component can be "closed" into a black box, and only its interface considered in the development. To use OO terms, this is called complexity containment through encapsulation.

The number of systemic levels are arbitrary: It is a question of practicality. A very complex system will typically have a deeper systems hierarchy, with many layers of black boxes, developed and validated individually. A simple system will be shallow, with as little as two systemic levels.

The idea with all the above is to provide a finished set of black box hierarchies, of archetypical role blocks. The positive outcome of this would be that much of the data input is already done: the things that are archetypically general across many organisations can be inherited rather than re-specified. With a hierarchy of specified objects, much of the data input problem can be solved. While individuals have individual aspects of their behaviour, there is also a great part of the behaviour which is common for the individual’s class. Thus, specifying for example the archetype "secretary" provides us with the means of assigning the same basic behaviour to a group of actors, in order to later on either specify the individual variety, or specify that the variety should be emulated through statistical distributions.

Because the greater part of the cost is related to the data input problem, reducing the data input through providing general templates will also drastically reduce the cost.

The measurement problem aspect of the output does not have a clear-cut answer. The system in focus is complex. The goals of the system in focus are complex. The definitions of "good" and "bad" in the system in focus are complex. Thus the measurements need be complex.

Efficiency measurements are of course one necessary point of output, but they depend on what part of the system we are studying. A single efficiency measurement variable for the sum total of the organisational information system is not usable, but a measurement on the level of a few-hop channel might be.

Other viable measurement outputs are overview graphs of how the system functions. One example could be the output of social network graphs as proposed by Klovdahl

(Klovdahl A, 1998). Through these it could be possible to get an overview of the aggregate of low-level operations.

The problem with the span between formal and informal channels is in part addressed simply keeping an isomorphic focus in the simulation.

The acting of a designed role is a conjunction of the assigned specific role, and the generic (or at least organisation-wide) human behaviour. All modelled agents move around, speak when meeting each other in lunch rooms or corridors, and have schedules making them available or not available in their rooms.

The idea is that the informal channels arise spontaneously when the assigned roles are acted: If one actor often meets another actor in a corridor, and they exchange information during these chance encounters, then an informal channel has been formed.

It would be counter-intuitive to explicitly model informal channels. However, an organisational design should take the available opportunities for the arisal of informal channels into account. If it is desired that some people speak with each other outside the formal organisation structure, then chances are better that this will happen if they sit close to each other geographically than if they sit in opposite parts of the building.

5. Introducing isomorphism

One important difference between traditional simulatory technologies such as Cellular Automata (CA) and Game Theory (GT) and the development of the last few years is the emergence of isomorphic approaches. In the isomorphic view, the object of reality should be represented as closely as possible in the simulation, preferably as objects and entities of their own. This has been made possible by multi-level simulation and multi-agent systems. In practice this means that, when simulating a human activity system with technological artefacts, the simulation should contain both a "person" object and, for example, a "computer" object. They should in all likelihood be placed within a spatial model of the building containing the human activity system. The output of the simulation should ideally display something looking like a top-down view of people moving around in the spatial model, doing what humans do, spreading information, going to the coffee machine, talking, working and meeting each other at the lunch table.

In an isomorphic approach, we try to capture reality through the whole process: Input, process and output should have an object-to-object correspondence with the perceived object reality.

It should be noted that the isomorphic view is far from uncontroversial. Traditionalists within the field criticise (and rightly so) the difficulties that arise from a philosophy of science point of view: The variables, both dependent and independent, become so numerous that is practically impossible to do any kind of validation of the simulation model in the traditional sense of the word. The only validity possible here is "face validity": if experts and initiates of the modelled system look at the inputs and outputs, and feel that this gives a believable impression, then some small manner of validation has been reached.

One headache the simulationist using an isomorphic approach soon encounters, is the relation between perceived objects (which are isomorphically modelled) and aggregate or higher-order objects (which are known, but not perceived per se).

In the best of worlds, a simulation would be built only using objects on one systemic level. For example, a simulation of an organisation with an actor perspective would only model actors. The higher-order entities (such as "group", "organisation" and "power structure") would be left to appear spontaneously as a synergetic effect of the interaction between the atomic objects. This would be true isomorphism, as this is how we perceive the system in focus.

However, as all models are reductions, we can not depend on this to necessarily happen. One of the things we in practice are forced to reduce (not exclude, but reduce) in the model is the overwhelming complexity of the relations and interactions between the atomic objects.

In practice, an isomorphic simulation of a complex system will need to implement some aspects of multilevel simulations. If we perceive that "organisation" is an entity, albeit abstract, we will need to model it as an entity with properties and behaviour. The actual implementation is simulation-dependent, but the relation between what we perceive as the atomic objects and the higher-order object needs be made explicit.

As a side note, all isomorphism is a matter of perspective. We choose to perceive certain entities as atomic. They are not really. They are simply one application of Occam’s razor, a selection of which systemic level to put in focus. For example, an actor is made up of sub-systems (cognition, movement, needs...) which in another simulation could be perceived as the atomic objects.

6. A new framework

One general problem with the word "simulation" is that it is a term used in many fields, and that it there denotes very different things. Simply saying "simulation" is bound to cause confusion. It can mean as different things as studying the result of an equation when changing its base parameters, via deterministic models of electrical circuits when adding or removing components, to massively complex multi-agent and multi-level studies of social contexts.

One way of getting around this is by defining "simulation" in the context of an article, and then blame the reader if he did not read or understand that definition. Another way is to use another, more explicit, term. In the latter option, the problem would of course be to find a term which is not already heavily laden with implications from already existing areas.

I shall in the following use the term ImAOS (Isomorphically Acted Organisation Scenario) for describing the kind of simulation I am after. Having googled for the term, I think the only field it might be already defined in is psychiatry (IMAOS is a class of anti-depressant drugs). I am certain that there will be no confusion between the two fields.

Think of ImAOS as a top-down view (as in geography) of an organisation with actor. When the simulation is started, a view of the building containing the organisation will be displayed. In this, it will be possible to see the actors (humans), computers and other information transmission equipment. The actors walk around as they would do in the real simulation, go to the lunch room and have a cup of coffee, sit at the computer and send a mail, talk in the corridors and attend meetings. While this happens, the person running the simulation can study information channels and actors to see where information piles up, or if there are people who sit idle because they do not get the required information.

As of now ImAOS is mainly a vision of how an isomorphic simulation should look and behave. It is not an existing framework.

7. Validity and/or credibility

Law and Kelton (Law A & Kelton D, 2000) points out that the validity of a simulation model is related to what purpose the study has. There is no such thing as a valid model in an absolute sense. Further, they point out that there is a distinct difference between a valid model and a credible model. A credible model is a model that the owner of the project accepts as good. A valid model is a model that is a good approximation of the object reality. These two are not necessarily related.

Byrne (Byrne D, 1997) points out that a it is a mistake to treat a simulation as an equivalent to an experiment. The validity in the positivist experiment sense should not be applied to (social) simulations, since the simulation by necessity works through analogues rather than similes.

Let me here state that the primary objective of ImAOS would not be validity in the positivist sense. While being important and something to strive towards, positivist validity is secondary to usefulness, or credibility. A model is good when it is usable for studying a system. The model does not necessarily have to be completely correct in the positivist sense to be usable. Abstractions and metaphors may enhance usefulness while still in the strict sense be deficient to positivist validity.

Also, the completely "correct" model of a social system is a chimera. As noted by Schmidt (Schmidt B, 2000), human behaviour is not possible to replicate in the foreseeable future, and that a good model (as opposed to a replica) does not need to conform to reality in all respects.

In some sense the goal and/or format of ImAOS could be classified as "Interpretivism" with a "highly constructivist bent" as described by Halfpenny (Halfpenny P, 1997), although I do not agree with his judgment that this approach is necessarily whimsical.

8. Conclusions

Trying to build a simulation of a whole organisational information system in order to use it for design and modifications, presents significant difficulties, mainly stemming from the complexity of specifying the simulation model and its contents.

In the above we have seen how the introduction of isomorphism and object templates could solve some of the problems that make simulations of organisational information systems infeasible today. However, this is as of now a theoretical construct, and it needs to be formalised into a coherent framework - ImAOS. It is yet early to predict the full work, but some issues need to be resolved. The theoretical approach need to be made explicit: There should be full and comprehensible ontology and epistemology for ImAOS models in general. The balance between complexity and reduction in the model needs to be made clear.

To make an implementation of ImAOS practically usable, low-level models for agent behaviour, such as communication and navigation, need to be integrated into a comprehensive template structure.

References

D Byrne (1997): *Simulation - A way forward?.* Sociological research online vol 2 no 2. *http://www.socresonline.org.uk/2/2/4.html*

M Carley, M Prietula, and Z Lin (1998): *Design versus Cognition: The interaction of agent cognition and organizational design on organizational performance.* Journal of Artificial Societies and Social Simulation vol 1 no 3. *http://jasss.soc.surrey.ac.uk/JASSS/1/3/4.html*

P Checkland and J Scholes (1997): *Soft Systems Methodology in Action.* Wiley, USA

G Egonsdotter and J Palmius (2002): *A Base for Simulating Information Distribution.* R Ragsdell (2002): *Systems Theory and Practice in the Knowledge Age*Kluwer Academic / Plenum Publishers, USA

Electronic Arts (2004): *The Sims II. . http://thesims2.ea.com/ [2004-11-08]*

D Fliedner (2001): *Six levels of complexity; A typology of processes and systems.* Journal of Artificial Societies and Social Simulation vol 4 no 1. *http://jasss.soc.surrey.ac.uk/JASSS/4/1/4.html*

C Goldspink (2002): *Methodological Implications of Complex Systems Approaches to Sociality: Simulation as a Foundation for Knowledge.* Journal of Artificial Societies and Social Simulation vol 5 no 1. *http://jasss.soc.surrey.ac.uk/JASSS/5/1/3.html*

P Halfpenny (1997): *Situating Simulation in Sociology.* Sociological Research Online vol 2 no 3. *http://www.socresonline.org.uk/socresonline/2/3/9.html*

R Hanneman and S Patrick (1997): *On the Uses of Computer-Assisted Simulation Modeling in the Social Sciences.* Sociological Research Online vol 2 no 2. *http://www.socresonline.org.uk/socresonline/2/3/5.html*

G Klir (2001): *Facets of Systems Science, 2ed. IFSR International Series on Systems Science and Engineering* vol 15Kluwer Academic / Plenum Publishers, USA

A Law and D Kelton (2000): *Simulation Modeling and Analysis.* McGraw-Hill Higher Education, USA

A Klovdahl (1998): *A Picture is Worth...: Interacting Visually with Complex Network Data.* W Liebrand, A Nowak, and R Hegelsmann (1998): *Computer Modeling of Social Processes*Sage Publications, UK

J Palmius, G Egonsdotter, and V Asproth (2003): *Simulations of Highly Complex Social Systems as a Tool for Designing Information Systems.* D Dubois (2004): *IJCAS. vol 14*

A Pyka (2001): *Special Issue on Applied Simulation Analysis.* Journal of Artificial Societies and Social Simulation vol 4 no 3. *http://jasss.soc.surrey.ac.uk/JASSS/4/3/0.html*

B Schmidt (2000): *The Modelling of Human Behaviour.* SCS-Europe, Belgium

S Sen (1997): *Multiagent systems: Milestones and new horizons.* Trends in Cognitive Sciences vol 1 no 9. *Elsevier Science Ltd*

Article 4

Article 4

This article was first submitted to IRIS 2005, and accepted for publication in its proceedings.

Palmius J (2005): *Defining the "information" part of "information system" - A Base for Simulation*. From *Proceedings of IRIS 2005*

The article has been reformatted to fit into the current work. The style of some of the references was changed. A few spelling mistakes were corrected.

Defining the "Information" part of "Information System" - A Base for Simulation

J. Palmius (joel.palmius@miun.se)

When building a simulation of an information system, a quantitative model of an information system is required. In order to construct a quantitative model of an information system, an operational definition of "information" is needed. This operational definition needs to mold the concept into a quantitative entity. There are many differend definitions of information already, but they have been formulated with other purposes. In this article the concept of operation is operationally defined with the long-term purpose of building simulation model of an information system where the human actors are considered part of the system. The concept is discussed from the viewpoint of earlier definitions, related concepts, things that happen to it and its various aspects. The curriculum of information is reformulated into quantitative terms, in order to finally arrive in a new definition of information within the current context: Information is an explicitly constructed message, always ultimately constructed by a sentient being. Along with describing the properties assigned to this entity, it is believed that it would be possible to construct a quantitative model of a complete information system.

1. Background

In order to build a simulation of an information system, a quantitative model of the information flows within the information system is required. In order to make a quantitative model of an information system we need an operational definition of information. The operational definition presented in this article is a part of the way towards building such a simulation. The concept "information" needs to be properly defined and put into its context. The definition need to be concise and operational (see below), since its purpose is to make information possible to represent quantitatively. The use of the operational definition is, in the extension, to make it possible to quantitatively compare the efficiency of information systems.

In the following I will concentrate on "information" and view "information system" as an extension to it. Further, as a catch 22, I will for now use the term "information" in the discussion without defining it, because the actual definition is the whole purpose of the article. I am aware that this may feel confusing at first.

1.1. An operational definition?

The difference between a definition in the common sense and an operational definition is that the operational definition needs to be concise and exact enough to precisely fit into the purpose for which the operational definition was made. It is not only a description of the thing as such, but also a way to mold a concept into something which possible to use in practice.

An operational definition can here be viewed as a design: We construct and mold concept in order to use it as a tool in the following work. A design is a specification of something new or changed, not a summary of something that exists. Thus it does not disturb me in any horrible manner that the definition does not altogether agree with what earlier authorities within the area has said, although it is still important that the definition relates to those authorities in some way.

1.2. The problem with "information"

Now, there are several available definitions of "information". Why are these not usable within the current context? There are a number of problems with them, and with the everyday use of the concept. First we must make a hard distinction between data and information. Data is already measurable, and has been since it was operationalized by Shannon (Shannon C, 1949). However, and unfortunately, this definition was made within something called "information science", which has led to much confusion. Shannon never measured information. He measured data. Information implies abstract concepts distributed via the bits and bytes in the data transfer, it is not equal to the data transfer. Shannon cannot be used to measure information transfers other than indirectly.

However, due to the initial labeling, data and information are commonly confused. If you speak about an "information system", the association most people get is a heap of cabling, switches and routers, computer stuff, which is used to transfer data. The efficiency of such a system is commonly measured with terms like megabit per second, but it does not really say anything about whether the actual information was efficiently transferred. When constructing an information system, measuring data flows is not sufficient. We must be able to know in advance whether the information flows are likely to be efficient. To know this we must be able to construct a quantitative model of information, not data. To make this quantification we must know exactly what it is we want to represent and the first step down that road is to make an operational definition of the term "information".

1.3. This article

The following article is intended to specify such an operational definition. It should be noted, that the definition is operational within the current context: The simulation of information systems. It should not be viewed as a generic definition of the term, and it is possible that it does not entirely overlap the commonly used sense of the word.

The approach and method for conducting the operational design definition, is to a) study earlier definitions, b) discuss related concepts, c) discuss aspects of the concept as such, in order to d) conclude a summarized operational definition.

2. Information system

Before beginning to define the concept "information", we need to know within what scope the definition is to be applied. As mentioned, the target area is "information systems", but this term needs to be further clarified. In common speech it can mean everything from cabling and routers to an organizational abstraction of how communication channels are organized.

2.1. What is an Information System?

In the words of this article, an information system is all that within an organization contributing to the distribution of information. The information system consists of nodes capable of holding information, channels able to distribute information and actors acting and re-acting upon the information. Thus the actors - the humans within the organization - are parts of the information system rather than users of the information system. With this definition, there is no such thing as an information system without people.

Since the actors are parts of the system, and since "the system" is our arbitrary grouping of the activities and components we study, the information system can be characterized as a human activity system in the sense of Checkland and Scholes (Checkland P & Scholes J, 1997). The system is an aggregate function of the organization, its people and its technological artifacts. The higher-order object, the human activity system, the "information system" does not exist as a separate tangible object. It is rather an abstraction, a holon, to which it is convenient to attach properties and behaviors. When further mentioned, the term "information system" shall denote this abstraction as described above. The short-term purpose of the information system is to distribute the right piece of information to the right actor in the right time. The long-term purpose of the information system is to support the viability of the organization through providing a sound base for decision, information merging and organizational development.

To make certain that no confusion arises upon this subject: let me state explicitly that an information system here is not the same thing as a software system. A software system, such as a customer database, may certainly be a part of the information system, but the information system is an aggregate phenomenon on the organizational level. There is no such thing as an information system without humans. However, it is very possible to have an information system without a single computer and even without electricity.

To summarize, an information system is all that, abstract or not, within an organization that contributes to and shapes the distribution of information: Humans, routines, policies, information nodes, spatial design, location of the coffee machine and the size of the tables in the lunch room.

3. Information

The concept "information" is somewhat tricky to define, as it is usually not a tangible object. It is rather an abstract phenomenon which arises between tangible objects.

Thus, rather than trying to immediately define the "information" concept, I will in the following try to define the context of the concept: What relates to it, what its aspects are and what happens to it. When this is done, a collected definition will be presented.

3.1. Survey of literature

The first thing we need to be aware of is that "information" is not one thing. It is an abstraction which differs significantly across different disciplines.

The next few sections will present a brief discussion about the concept from the viewpoint of literature within information science and infology. The fields can be said to be somewhat arbitrarily separated, as the current field can be said to be a conjunction of several. First we need to recognize that information science is not really about information as such, it is about the communication of information. This has been pointed out in various summaries, such as by Holmström and Koli (Holmström J & Koli T, 2002). Indeed, one of the most important classical bases for information science is called "A mathematical theory of communication" (Shannon C, 1949). Thus, it might not be entirely fruitful to discuss the information science definition or view of information as such.

However, whether defined or not, a view of information is implicit within information science. In my personal understanding, data and information are seen as equal. For example, one definition of information is "Information is any input into the system that initiates a change of state" (Vickery B & Vickery A, 1994). The authors of this definition do, however, seem to be aware of the problem with the information definition, as they later in the text state that defining information as such would not be fruitful. Implicitly we could understand information as a data transfer which caused an effect at the receiver. This should largely be a paraphrase of the above stated definition.

Several authors working within the information science paradigm, or close to it, approach the problem in a similar way. Cole, for example, defines information as "that which modifies knowledge structure" (Cole C, 1997a), although he himself sees this definition as not entirely clear and that the distinction between information-as-a-process and information-as-a-thing must be made clear (Cole C, 1997b).

This method of managing the information concept seems to be common within the information science literature. Information is not defined directly, rather it is specified what happens to it, what it causes and how it is treated.

One version of this is trying to define information as a process rather than as an entity. Losee, for example, tries his hands on a discipline- independent definition, and defines information as the following "Information may be understood as the value attached or instantiated to a characteristic or variable returned by a function or produced by a process" (Losee R, 1998).

The border between "information science" and "infology" might not be altogether obvious. In this text I shall treat the latter as the "soft" variant which concentrates on meaning, interpretation and understanding, while the former is the hard part concentrating on measurement, quantity and communication channels.

The classifications of which category authors fall into may be considered somewhat arbitrary. For example, Meadow and Yuan (Meadow C & Yuan W, 1997) discusses the distinction between data and information, from a base of mostly information science literature, but end up into definitions that I feel are infological. They end up in viewing information as partly the function of the recipient; that the interpretation process is a part of the information as such.

One of the fundaments of infology is the infological equation "I = i(D,S,t)", that information is the data (S) and the preknowledge (S) interpreted (i) during the time (t) (Langefors B, 1995). The important consequences of this are that information is never objective. Information only exists within a subject who has pre-knowledge, and who has performed an interpretation process of a certain piece of data. The objective components of this equation are data and time. Pre-knowledge, interpretation and information are subjective and largely uncontrollable. We can also see a hard distinction between information and data. This view could be said to largely coincide with, in Buckland's wording, "information-as-knowledge". Buckland states that this is one of three major approaches for viewing information, the other two being information-as-process and information-as-thing (Buckland M, 1991).

Stonier (Stonier T, 1996) discusses information as an entity or concept which is inter-convertible with entropy. In essence, the view is that information equals energy, which performs a work to change the state of a system, which in its turn is an organization of existing entropy. To drive it even further, Stonier claims that "order is considered to be a manifestation of information" (Stonier T, 1996). A closer analysis of Stonier also reveals that he treats "information" in largely the same way that infology treats "data", and what infology would call "information" he calls "message".

3.2. Narrowing down the field

Clearly, "information" can mean many different things, some of them incompatible. To delimit the aim of this work, I shall begin with a root definition (Checkland P & Scholes J, 1997). The system in focus (the system we will study through the use of the simulation, not the simulation itself) is "an actor-owned and manned system to distribute and treat information, through the use of technological and social information channels, in keeping with the overall organizational goals, in order to support the viability of the organization through keeping all its parts up to date with necessary information." (This is a paraphrase of a root definition by Checkland & Scholes 1997).

3.3. Related concepts

In the discussion about information, there are a number of other concepts which can be considered close or related, but which themselves are not the same thing as information. In order to delimit what information in itself is, these other concepts need to be made explicit too. The following discussion will be concentrated on how the related concepts can be formulated in relation to an operational definition of information. Thus, the approach is to operationalize these concepts too.

3.3.1. Data

As previously mentioned, one key concept is that of data. It is important since it is often confused with information. However, while data is something tangible, information is not, at least not yet. "Information" implies meaning, while "data" only implies the possibility of meaning (Meadow C, 1996). In the following I shall view "data" as synonymous with the carrier in an information channel (see below).

3.3.2. Concepts related to human understanding

One major thing which differentiates information from data is the concept that information contains some form of meaning. Meaning is implicitly something humanly attributed: It exists only in the relation between data and a conscient, interpreting entity. For all practical reasons we can for now equal this entity with a human being. While it is conceivable to manage some kind of AI or communicating with animals, this is not the common use of an information system today, and it is therefore for now disregarded. From Langefors (Langefors B, 1995) we acquire two key concepts here: interpretation and pre-understanding. Pre-understanding can be expanded into knowledge and experience. For completeness we can also include the terms wisdom and understanding here.

In the current context, the problem with all these concepts is that they are highly relative, or at least subjective. They are a function of a human being, and not representable in any easy manner. This is a problem with the current context, not with Langefors' definition: It is correctly observed that information is relative and subjective. This does not help us right now though. It is for us to find a work-around for this. Thus, to work around the problem, the concepts need to be operationalized. If they can be formulated in a quantifiable manner, then quantifying the thing they relate too will be easier too.

Thus, let us begin with "knowledge" and "experience". These can be readily operationalized if we presuppose that there is a quantitative and objective definition of information. Essentially they are synonymous: knowledge and experience is the amount of information that an individual has successfully managed to receive. Knowledge and experience within a certain topic is the amount of information received concerning that topic.

Of course, we have to take into account the situation where a person is exposed to information, but does not manage to receive it. Or to use more general terms, where someone got the information but did not understand it. This is where the term "interpretation" comes into focus. To make things easier, I will presuppose that all information in an information system is, in the end, theoretically possible to understand. It might require other information in order to be understandable, and it might take longer or shorter time to understand, but theoretically all individuals in the information system can, with some effort, understand all information available.

With this delimitation in mind, "interpretation" becomes the process in which information is integrated with the already acquired knowledge and experience. The result of an interpretation process is either that the amount of known information is expanded, or that the information in focus is rejected because there was not enough knowledge and experience available to be able to accept it. With this delimitation, the term "understanding" can be left out of the equation, since all information is understood once it has passed the interpretation process.

The existing knowledge and wisdom forms a frame within which new information is interpreted. This frame can be said to be synonymous with "pre-understanding". Thus, pre-understanding is operationalized to denote the availability of knowledge and experience. Or, in consequence, pre- understanding is the already acquired relevant information.

This leaves us with the term "wisdom". Essentially, it is possible to continue the discussion without taking this term into account, but for completeness, let's define it. And operational definition of "wisdom" in this context is that it equals the speed and success factor of the interpretation process. A person is wise if he has much knowledge and speedily can acquire new information.

Essentially, the above leaves us with the opportunity to make a collected reformulation of Langefors' infological equation. This is an interesting opportunity which, as we will see, will end up being useful.

3.3.3. Concepts related to communication

As the term information system implies communication, it is relevant to think that communication has an impact on the definition of information in this context. Descriptions of communication usually involve the actors and the path between the actors. Leaving the actors themselves aside for now (they will be discussed below in the section on human behavior), we can list a number of important concepts related to the path between two actors in a communication.

First we have the path itself, the channel. This is the complete path, with all transformations, from actor A (sender) to actor B (receiver). It does not include the actors themselves, and does thus not include the interpretation process mentioned above. The "channel" starts when information is transmitted and ends right before it is received. The "channel" consists of at least one medium, at least one carrier and at least one syntax. While information can be conceived as a somewhat abstract concept, it does not exist in an abstract void. It relies on a physical environment to be transmitted, and during the transmission it has a form and a physical representation. The environment through which the information is transmitted is the medium, and through that medium the information is transported using a carrier. For example, if two people are shouting to each other, the medium is the air and the carrier is the sound waves. If sending an email, the medium is computer cabling and the carrier the electronic signals making up the mail.

The thing that separates information from line noise (apart from the meaning and the content which we disregard for now), is the "syntax". The syntax is the structure of the transmission, the protocol agreed upon by both sides of the channel. Without the syntax, the transmission can not be properly received at the end of the channel.

To summarize, "communication" is the transmission of information in a structured way through a medium by a carrier. The term "communication" does here not include the processes required to prepare the information for transmission at the start of the channel, nor does it include the interpretation process at the end of the channel. These delimitations are chosen in order make the "communication" as such objective. The subjective activities related to the communication are discussed later.

3.3.4. Concepts related to decision-making and human behavior

As mentioned before, with the definition of information system used in this article, there is no such thing as an information system without humans. As an implied consequence we could also question whether there is such a thing as information without humans.

As a practical postulate, I will state here that there is not. All information presupposes at least one thinking and interpreting being somewhere. However, and as a distinction against other definitions, the information does not absolutely require two ends of a communication channel.

Having accepted that the definition of information requires at least one human, we need to specify how. We might need to take into account things such as information behavior, how people behave in relation to information. The first aspect to point out here is how the phenomenon of information arises. As a continuation of the previously stated practical postulate, I will claim that information is exclusively caused by human beings, although possibly through their behavior and interaction with their environment. Completely new information can arise through observation. However, usually new information is dependent on existing information. Partially new information will arise through the processing of other information, in human activities such as a guess, through inference or through deduction. The experiences of a human form how that human processes further information. The human has formed schemata and scripts for information processing.

Finally, there is an output in behavior (of which non-action is an example). The human has made a decision, usually after having come to a conclusion.

These concepts, related to construction, processing and output of information, need to be made operational to make an operational definition of information possible.

In a model of an information system, the human entity is a part which needs to be modeled. However, the detailedness of model of this entity only need to fulfill the requirements for the information system model, it does not have to be a detailed model of human behavior. As noted by Schmidt (Schmidt B, 2000) a model does not equal a replica. Here, only a basic representation is required.

So, as for the representation of the production of information. As we cannot model the detailed contents of information, we must represent it through a reduction, such as an amount of categorized information. Since the information is roughly represented, its arisal should likely be equally roughly represented. Thus, in a simulation, completely new information can arise on a statistical basis in the human entities. An observation, a guess or an inference is thus the event when new information arises in a human entity based on a statistical model. To discuss exactly how this statistical model should look is beyond the scope of this discussion. For the processing of information we have the terms "schemata" and "script", which are essentially terms fetched from cognitive psychology. The operationalization of both these terms are the allowed combinations of already acquired and incoming information. The complete reduction is a list of which categories of information is allowed to be combined with which other categories to produce new categories of information. These combination and categories are implementation-dependent.

Finally, decision and conclusion can be operationalized to denote output (or non-output) of information. A visible decision or a conclusion is when two chunks of information have been combined and result in the human entity transmitting the resulting information. An invisible ditto is then the combination has happened, but the human does not transmit the result, and simply keep it to himself for later.

For an implementation of a simulation of human information behavior there of course remains much to be said. A model would need to include operationalizations of how humans work, look for information and communicate. However, these models are beyond the scope of this discussion.

3.3.5. Other important concepts

There are also a number of random other concepts which relate closely to the concept of information. Especially of note in this discussion are terms such as fact and context for content and interpretation of information, and document and message for its transmission.

To begin with "fact", it is essentially left without meaning in an operationalization. Information is information, when it comes to quantitatively representing information flows. Whether the information is true or not is not and does not have to be measurable. Thus, concepts such as disinformation, too, lose their meanings. All information according to the coming operationalized definition is informative whether it is true or not.

In a simulation, some things will be too fuzzy to take into account. One of those things is the context in which information is received. In reality, the context has a heavy influence on interpretation and evaluation of received information. However, since we do not take into account the truthfulness of information, and since it is infeasible to cover the detailed contents of an information transmission, the notion of context is also dropped: the prerequisites for the influence of context of the interpretation process does not exist, and thus the context cannot be taken into account.

Tangible concepts such as "message" and "document" become all the more interesting though. I shall here view these terms as more or less synonymous. A document is merely a formalized message. All documents are thus messages, although not all messages are documents. The notion of message fulfills many of the criteria of quantification we want, and as we will see below, it can be useful for the final operational definition of information.

3.3.6. Excluded concepts

This discussion has necessarily excluded several concepts. The obvious synonyms have been left out, but also things that might be considered key concepts by some.

Psychologists of various brands might be put off by the fact that I have left out concepts such as action, reaction, collective consciousness and subconcsciousness. The first two were left outside, since I felt they were already covered by the section on human behavior. The last two were left outside since I did not think them relevant for the discussion: The terms are fuzzy from the start, and would not contribute to the operationalization of the information concept.

Infologists, information scientists, computer scientists and AI researchers might react against the exclusion of terms such as data type, scale, denotation, proof, capta and elementary message. Scale and data type are left outside because they are too lowlevel to be relevant for the discussion. Denotation and proof are left outside, because they concern the contents of information rather than the information itself. Capta and elementary message are theoretical constructs concerning the interaction between data and interpretation and are too abstract to be considered important.

3.4. Related activities

Having discussed what other entities relate to information, we can now discuss what happens to our elusive concept itself. Some of these activities should not be very controversially defined, but some may seem a bit strange at first. As before, the goal with the definitions here is to make the information concept as explicit and quantifiable as possible.

3.4.1. Transmission and storage

Transmission is the entire chain from (but not including) the mind of individual A, via encoding to a format suitable for sending, through a communication channel on a carrier, to decoding by individual B to (but not including) the mind of individual B. On the route, the information might also be stored for longer or shorted durations.

The communication channel and the carrier have been discussed earlier, and they are thus left outside of the discussion here.

Decoding is synonymous with interpretation. It is the process in which received information is integrated with the known information. This can be equaled with time. Encoding could be said to be this process but backwards, it is the process of moving known information to a carrier on a communication channel. This, too, can be equaled with time, although not necessarily the same time as for the decoding.

Information can be stored for longer or shorter durations. A short duration is when a technological artifact holds the information during the process of transmitting it along a communication channel. A longer duration is what is usually understood by the term storage: It can be put in an archive. This, too, is done in some form of technological artifact (which may, of course, be an analogous traditional archive).

3.4.2. Production and destruction

When building a model of information flows, my experience has been that the trickiest part to specify is how and when information arises. Once information exists, making it flow through a model is usually not all that difficult.

To address this we must clearly specify the production and destruction of information. Where does it arise, and where does it disappear? In the following, I will apart from the term "production" also use aggregation as a term related to production. There is a small but important difference between the terms.

As mentioned in the discussion about human information behavior, new information can arise on a statistical basis in a human being. This is one source of information. The new information can be completely new (produced) in which case it has spontaneously arised (representatively as a result of observation, guess or inference), or it can be a combination of earlier acquired information (an aggregation). These are the only internal sources of information within an organization.

However, information systems does not exist hanging in a void. They have input and output, as all open systems. Thus, an additional source of information is external input. This represents information coming into the organization from the outside world, through for example phone calls or emails.

Concerning the destruction of information, it is never explicitly destroyed. Information exists as long as it has a carrier. If a human has received information, then the information is known as long as the human remains in the information system. Information also remains in the system if it is carried by another carrier, such as technological artifact or an archive. From these, the information can be explicitly expunged on the basis of dating.

3.5. Aspects of information

We have now almost surrounded the information concept by defining things close to it, and what happens with it. Before stating clearly what information actually is, I will address some of its aspects. The purpose with this is mainly to specify what can be specified in the collected definition, and why. The major outcome of this section is a list of things that cannot be considered in order for there to be an operational definition.

3.5.1. Aspects concerning content

The thing that makes information difficult to nail down is its content, and the terms relating to it, such as relevance, integrity, value, reliability and authenticity.

A model of an information system needs to specify information content somehow. A flow model need to take into account which information should end up where. This causes a severe conflict with the need to make the same model simple and possible to overview. On one extreme of the simplicity-relevance axis here we have the complete copy: We know the actual contents of the information, and we specify that content for each information transmission. This makes the model huge and cumbersome and, say, a computerized implementation would be difficult to make. Further, knowing all the information in the information system in advance is likely not possible. On the other extreme, we have the complete reduction of information into a meaningless token. We know nothing about the contents of the information, and the flow model could look exactly the same as if we shuffled coins or water around. This makes the model simple, but hardly relevant. Some notion of content is necessary in order to build a model of an information system, since the paths of information distribution is highly dependent on the contents of the information. Now, a compromise would be ending up somewhere around the middle of the axis. We could specify a representation or classification of the information without specifying the contents fully.

3.5.2. Aspects concerning quantification

The whole purpose of the information definition is to make information quantifiable. Thus, terms such as quantity, mass and size need to be specified.

In the end the information entity must just that: an entity. It should therefore be possible to count, and therefore be possible to come in a quantity. In common language, information is not countable. You can have "a piece of information", but not (at least not grammatically correct) "two informations". However, with information as an entity, you can have two informations. "One information" is one instance of the entity we call information. We must, however, recognize that not all information is equal in size. The operational definition of size is here equaled with time for interpretation and/or for encoding. A large information takes time to accept. A small information is quickly accepted. Thus, the measurement of the size of an information is specified in average number of seconds it takes for a recipient to accept it.

To make a difference between interpretation and transmission, I will also specify the "mass" of an information. This is here defined as the time in seconds it takes to transmit it via a communication channel. In implication, it can be viewed as the amount of data required for the carrier in the communication channel.

3.6. Operationalization and Definition

Now with the entire curriculum described, it is finally possible to deal with the concept as such. First a short discussion about the reformulation of the infological equation, then the definition as such.

3.6.1. Reducing and operationalizing the infological equation

As mentioned before, one classical base for defining information is Langefors' Infological Equation (Langefors B, 1995), which says that "I = i (D, S, t)", that information is the result of the interpretation process "i" operating on data "D" and pre-knowledge "S" over time "t". As seen earlier, this definition of information is not suitable as an operational definition, since it explicitly states that information is always subjective, and thus not in practice quantifiable. However, with some reformulation, it is my view that it could be molded into a usable form.

In a simulation, we would need to pre-suppose a model S for the agents, thus making S implicit in our definition. In consequence with the previously stated definition of pre-knowledge, it is simply the collection of previously known information. Since it is a function of information, it cannot be a pre-requisite of information. Thus we should move S outside the equation for now: it is a question of implementation later, and does not affect the actual definition of information.

The interpretation process "i" is globally defined as the time it takes to accept information. This is according to our previously stated definition that presumes that all information in an information system is ultimately understandable by all parts of the information system. The interpretation process is thus something that happens to information, not a part of it. As with "S", "i" should thus be removed from the equation for now, it is a question of later implementation. Further since we already equaled "i" with time, we can also remove "t" from the equation.

That leaves us with the data transmission, which in this case is a meaningful chunk of "something". I am reluctant to call it "data" in this case, since it is the theoretical contents of a carrier transported on an information channel. A better term for D here would here be "message" M. This leaves us with a reduced "I = M", information equals "message", a self- contained entity containing meaning, meta-data and other relevant properties. The term "message" implies intention, and thus as a further delimitation, a message is always explicitly constructed by a sentient being. The construction might have been done through using a technological tool, but ultimately, a "message" is always constructed by sentience.

The reduced parts, time (interpretation process) and pre-knowledge remain hanging. These are re-introduced as time and behavior further down.

3.6.2. The new definition

This finally makes it possible to define the now tangible entity "information": *Information is an explicitly constructed message, always ultimately constructed by a sentient being.*

This is the base definition which enables an understanding of the thing as such. However, to fully operationalize the concept, we further need to specify its properties and qualities, and specify what it is and is not. This will be done in the following sections.

3.7. Properties of the newly-defined entity

While the entity as such is thus defined, we will not be helped until its properties are described. While these properties do not necessarily take part of the operational definition, they are part of the operationalization.

These properties do not necessarily have to be general across implementations where the operational definition is used. The use of them depends on what a model of an information system is intended to be used for. This, said, the following are properties which will likely be relevant for most implementations.

3.7.1. Automatically assigned properties

The automatically assigned properties of "an information", are those things that describes it curriculum without touching its contents. Examples useful for tracing the paths of information through an information system are things like creation date, author, and intended recipient. As noted, since information according to the definition always arises ultimately as caused by an actor, these properties should be possible to automatically assign in a model.

3.7.2. Behavior

Most aspects of information are, as seen above, possible to reduce to time. Thus most of the behavior of information is reducible to time. Thus we can specify the size and mass of information as the times it takes to interpret it and to send it respectively. Apart from this we also need to specify specific requirements. These are things such as format requirements, such as whether this current information is not possible or feasible to transform to certain channel formats. Further we have interpretation requirements, which operationally is a list of other information which is necessary to have acquired before this current information can be integrated.

3.7.3. Category, subject and instances of information

As mentioned, we cannot fully model the contents of information: It would void the purpose of the model. However, we can represent the information contents through marking information with category and subject. These are arbitrary labels, which are case-scenario specific. We can then likely find a few large categories (such as "marketing", "sales" and "environment") within each we can find a number of subjects. These subjects do not necessarily have to be named. Each instance of information has a category and a subject, but several instances can have the same category and subject while still being unique.

The category and subject properties are the base for the routing of information: The actors produce information of a certain classification, and look for information of a certain classification.

4. Summary and conclusions

Information is an explicitly constructed message, always ultimately constructed by a sentient being. This is the basis for the operational definition which is the result of this discussion. With the discussion around the definition as a complement, it should be possible to use this definition for building a model of an information system, in which the information flows are quantifiable.

Implicitly, the discussion above has formulated both an ontology and an epistemology surrounding the concept of "information". The ontology has, consciously, been focused on a reduction of reality in order to mold the concept into something tangible. The epistemology has been focused on quantification and measurability, the re-formulation of qualitative concepts into distinct quantitative entities.

To iterate, the definition is not to be considered a generally applicable truth. It is a construction, a design, of a tool. It could look differently, and for another context it probably must look different. The definition and the curriculum is a theoretical base for a model of an information system, it is not the model as such. An actual functional model needs to implement further operationalizations which are case-specific.

References

M Buckland (1991): *Information as Thing.* Journal of the american society for information science vol 42 no 5

P Checkland and J Scholes (1997): *Soft Systems Methodology in Action.* Wiley, USA

C Cole (1997a): *Calculating the information content of an information process for a domain expert using Shannon's mathematical theory of communication: A preliminary analysis.* Information processing & management vol 33 no 6

C Cole (1997b): *Information as a process: the difference between corroborating evidence and "information" in humanistic research domains.* Information processing & management vol 33 no 1

J Holmström and T Koli (2002): *Making the Concept of Information Operational.* Department of Information Technology and Media, Mid Sweden University, Sweden. *Master thesis in informatics, http://www.itm.mh.se/~vivasp/term8/Exarbeten/02D-008.pdf*

B Langefors (1995): *Essays on Infology.* Studentlitteratur, Sweden

R Losee (1998): *A discipline independent definition of information.* Journal of the american society for information science vol 48 no 3

C Meadow (1996): *A Proposed Method of Measuring the Utility of Individual Information Retrieval Tools.* Canadian Journal of Information and Library Science vol 21 no 1

C Meadow and W Yuan (1997): *Measuring the impact of information: Defining the concepts.* Information processing & management vol 33 no 6

B Schmidt (2000): *The Modelling of Human Behaviour.* SCS-Europe, Belgium

C Shannon (1949): *A mathematical theory of communication.* University of Illinois Press. *[Endast som papperskopia, hitta fullständig referens]*

T Stonier (1996): *Information as a basic property of the universe.* BioSystems vol 38. *Elsevier Science Ltd*

B Vickery and A Vickery (1994): *Information science in theory and practise.* Bowker-saur, UK