Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 1991 Proceedings

International Conference on Information Systems (ICIS)

1991

CONCEPTUAL MODELING AND CLUSTER ANALYSIS: DESIGN STRATEGIES FOR INFORMATION ARCHITECTURES

Derk Jan Kiewiet
University of Groningen

Robert A. Stegwee
University of Groningen

CONCEPTUAL MODELING AND CLUSTER ANALYSIS: DESIGN STRATEGIES FOR INFORMATION ARCHITECTURES

Derk Jan Kiewiet
Robert A. Stegwee
Faculty of Management and Organization
University of Groningen

ABSTRACT

An information architecture (IA) specification is given for a dutch organization, the CBR (Central Bureau for certificates of driving proficiency). A discussion of the specification process shows that other issues are taken into account than those tackled by the information planning method used. However, this is done in an informal and intuitive manner. We argue that the specification of alternative information architectures can give a more solid basis to this process. In order to generate these alternative architectures, various design strategies are needed. A generalization of design strategies can be found in cluster analysis. The application of cluster analysis calls for an appropriate problem representation. This is achieved by using a model hierarchy as a conceptual model of the data being handled in an information planning process. The model hierarchy also serves as the basis of a decision support system for information architecture specification, which is described briefly. To demonstrate the feasibility of our approach, we compare the original CBR architecture with an information architecture generated by the decision support system. It turns out that the new architecture performs better with respect to the decentralization perspective of the organization.

1. INTRODUCTION

One of the activities belonging to the field of information management is Information Planning (IP). Information Planning is defined as the decision making process leading to the definition of an information systems plan for the organization, which is to be used as the basis for the analysis and design of the information systems described in it (Boersma 1989). The specification of these information systems together form the information architecture for the organization. Design strategies for information architectures are the topic addressed in this paper. It is assumed that the reader is familiar with methods for information planning in general.

The discussion starts with a description, of a specific case of information architecture specification. From this description we derive the necessity to develop various information architectures and, hence, different design strategies. In section 3, the comparison of design strategies leads to the specification of a framework for information architecture design. This framework forms the foundation of a decision support system. Conceptual modelling and cluster analysis, the main constituents of the framework, are elaborated in sections 4 and 5 respectively. Finally, an example is given of the application of the decision support system, and some conclusions are presented.

2. INFORMATION ARCHITECTURE SPECIFICATION

2.1 Short Description of the Case Organization[1]

The CBR (Central Bureau for certificates of driving proficiency) is the national institution in The Netherlands responsible for administering exams and for providing medical statements, both with respect to driving proficiency. The exams can be divided into theoretical exams and practical exams.

Theoretical exams can be taken in twenty-eight Theoretical Examination Centers (TECs), including one mobile center, which are spread over the country. These exams are administered using slide presentations of traffic situations which are accompanied by true/false questions. Practical exams are usually administered in cooperation with driving schools. The examiner of the CBR judges the candidate while (s)he is driving the driving school's car. The exam starts at one of the Practical Examination Centers (PECs). These PECs can coincide with a TEC or other location of the CBR, but sometimes a PEC is not more than an agreement with a local restaurant. A practical exam has to be applied for at the CBR. This is usually done by the driving school.

The third main activity of the CBR is providing medical statements with respect to driving proficiency. On the basis of a Personal Statement, a medic will judge the (medical) ability of a person to drive. In special cases, after consulting a general physician or the Common Medical Service, a conditional medical statement will be provided. In this case the candidate is only regarded medically fit to drive vehicles with specified adaptations.

Finally, the CBR has a special organizational unit to handle appeals, the Bureau of Further Investigation (BFI). The procedures followed for this bureau are roughly the same as for the "normal" theoretical and practical examinations. In total, approximately 900 people work for the CBR, of which 300 work for the internal service. The internal service is responsible for the administrative work regarding the exams. Taking into account 600,000 practical examinations, 480,000 theoretical examinations and 800,000 Personal Statements (1987 data), one can imagine the vastness of the administrative task.

2.2 Information Planning and Information Architecture for CBR

The replacement of the central hardware and systems software, as well as the oncoming reorganization of the CBR, were reasons for the CBR management to carry out an information planning study. For this purpose, Moret Ernst & Young Management Consultants were brought in to help, which also meant a choice for the proprietary method, Arthur Young Information Engineering/Computer Systems Methodology-Planning, in short called Information Systems Planning (ISP). Based upon this method, several models have been specified for both the current situation and the desired situation with respect to information systems services. The Information Engineering Workbench (IEW) has been used to store and manipulate these models. The models include

- *Process Model.* This model divides the overall operations of the CBR into business functions and, at the highest level of detail, business processes. It gives an overview of what processes have to be carried out by the CBR in order to fulfill its goals. In this model the business functions and processes together form a tree structure of business functions with business processes at the end (the leaves), called the functional decomposition.

- *Entity Model.* The entity model specifies major categories of data which are of interest to the CBR and which are needed to perform the business processes. Also, relationships between categories of data are indicated, such as the relationship between candidates and driving schools.

- *Information Model.* The combination of business processes and entity types in one matrix forms the information model. In this matrix, it is indicated whether a business process creates data pertaining to an entity type, or whether it uses data on an entity type.

- *Responsibility Model.* This model combines business processes with organizational units in the form of a matrix, indicating the responsibility of organizational units for certain business processes.

- *Distribution Model.* This model represents the distribution of business processes over physical locations.

An information architecture has been specified as part of the study, based upon the information model. This was initially done in the well-known manner, rooted in the BSP method (IBM 1984), of moving the entity types in the matrix until the *C*'s (business process creates data pertaining to an entity type) are grouped around the diagonal. The business processes in the matrix are kept in the sequence in which they appear in the functional decomposition. Information systems areas are then specified by drawing boxes round groups of C's, which do not go beyond the boundaries of business functions. This has lead to a more or less functional division in information systems areas. Part of this information architecture is depicted in Figure 1.

2.3 Discussion

The initial outcome of this specification proved unsatisfactory. Discussion arose with respect to the responsibility for and the physical location of information systems.[2] For example, the financial processing of an application for a practical exam had always been considered the responsibility of the head office. However, the initial information architecture led to incorporation of this process in an information system for the regional offices. Therefore modifications were made to the information architecture in order to incorporate responsibility and locational considerations. This was done in a rather ad hoc and intuitive manner as the ISP method does not include procedures to handle such issues. Another, more systematic approach would have been to specify a second information architecture which explicitly takes these issues into account. Part of such an alternative architecture is given in Figure 2. Careful evaluation of the initial ISP architecture and the second architecture could have led to the specification of a final information architecture.

	Candidate	License revocation	Medical status	Medical background	Theoretical Exam	Theoretical Examination Center	Application Theoretical Exam	Practical Exam	Application Practical Exam	Slide series	Result Theoretical Exam	Result Practical Exam
Register Candidate	C											
Process Legal Information		C										
Assess Personal Statement	U		C	U								
Medical Assessment	U		U	C								
Produce Medical Statement	U		C	U								
Process Adaptations	U		C	U								
Retrieve Medical Information	U		U	C								
Set Theoretical Exam Charges					C							
Assign Capacity Theor. Exam						C	U					
Receive Application Theor. Exam	U				U	U	C					
Set Practical Exam Charges								C				
Register Pract. Exam Application									C			
Medical Acceptance	U		U						U			
Supervise Slide Series					U					C	U	
Administer Theoretical Exam	U						U			U		
Process Result Theoretical Exam	U				U					U	C	
Quality Control Practical Exam								U				U
Administer Practical Exam	U		U									
Process Result Practical Exam	U		U					U				C

Figure 1. Part of the Initial ISP Architecture

The ISP method focuses only on business functions and entity types. Therefore, information architectures specified by the ISP method emphasize a functional division of information systems areas. However, more issues are usually considered in the specification of an information architecture. Information on these issues can sometimes be found in the models specified, but is not used explicitly. Information architecture specification can be enhanced by developing alternative information architectures based upon this extra information. Hence, these alternatives focus on other issues than traditional BSP-based methods. Analysis of the merits and differences of several architectures will hopefully lead to a final information architecture which is tailored to the specific needs of the organization.

3. GENERALIZATION OF DESIGN STRATEGIES

In this section, the structure of different design strategies is investigated. Based upon similarities in structure, a framework is specified as the foundation of a decision support system for information architecture specification.

3.1 The Structure of Design Strategies

The ISP design strategy leads to an information architecture for the CBR which contains an information systems area called Medical Information System. This information systems area supports the business processes Assess Personal Statement, Medical Assessment, Produce Medical Statement, Retrieve Medical Information, and Process Adaptations. Also, the entity types Medical Status and Medical Background are part of this information systems area. The design strategy can be specified by of the following rules for forming the information systems areas.

- Processes that belong to the same function should be supported by one information systems area; e.g., all the business processes mentioned above are contained in the business function Medical Administration, which is why they form one information systems area.

- Entity types should be assigned to the information systems area in which data pertaining to the entity type is created; e.g., Medical Status and Medical Background are the entity types created by the business processes in this information systems area.

	Slide series	License revocation	Theoretical Exam	Practical Exam	Application Theoretical Exam	Candidate	Theoretical Examination Center	Medical status	Medical background	Result Theoretical Exam	Result Practical Exam	Application Practical Exam
Supervise Slide Series	C		U							U		
Quality Control Practical Exam				U							U	
Process Legal Information		C										
Set Theoretical Exam Charges			C									
Set Practical Exam Charges				C								
Reserve Application Theor. Exam			U		C	U	U					
Administer Theoretical Exam	U				U	U						
Register Candidate						C						
Assign Capacity Theor. Exam					U		C					
Administer Practical Exam						U		U				
Assess Personal Statement						U		C	U			
Medical Assessment						U		U	C			
Produce Medical Statement						U		C	U			
Process Adaptations						U		C	U			
Retrieve Medical Information						U		U	C			
Process Result Theoretical Exam	U		U			U				C		
Medical Acceptance						U		U				U
Process Result Practical Exam				U		U		U			C	
Register Pract. Exam Application												C

Figure 2. Part of an Alternative Architecture

The information used for applying these rules is drawn from the Process Model and the Information Model.

The alternative architecture, which focusses on the responsibility structure, contains an information systems area called Head Office System. This information systems area includes the business processes Supervise Slide Series, Quality Control Practical Exams, Set Practical Exam Charges, Set Theoretical Exam Charges, and Process Legal Information. The entity types Slide Series, License Revocation, Theoretical Exam, and Practical Exam are also included. Rules that make up the design strategy followed here are:

- Processes that are the responsibility of the same organizational unit should be grouped into one information systems area; e.g., the business processes mentioned above are all the responsibility of the Head Office, and hence are grouped together.

- Entity types should be assigned to the information systems area in which data pertaining to the entity type is created.

In this case information from the Responsibility Model and the Information Model is used to design the information architecture.

Although the two design strategies described above produce different architectures, they do have the same structure. They differ only in the specific rules that have to be applied in order to form information systems areas, and hence in the information needed to carry out this process. The similarity in structure allows us to specify a generic framework for design strategies.

3.2 A Framework for Design Strategies

In general, design strategies for information architectures use a selection of the available models to arrive at a specific information architecture. The focus of a particular design strategy depends on the selection of the models that are used. With respect to a framework for design strategies, it is useful to view the set of models as one overall model. Such a comprehensive model specifies the way in which information flows through the enterprise. For the CBR, this model is given in terms of entity types, processes, organizational units, physical locations, and the interactions they have with one another. Other organizations may have specified different models, and hence their comprehensive model may contain different aspects.

The rules used in design strategies can be generalized as follows:

- Processes that are in some way similar to one another should be grouped together.
- Entity types that have a certain relation to a business process should be assigned to the group in which the business process belongs.
- A group of processes and entity types forms an information systems area.

Alternatively, rules can take the entity types as a starting point, leading to three similar rules:

- Entity types which are in some way similar to one another should be grouped together.
- Processes that have a relation to an entity type should be assigned to the group which the entity type belongs to.
- A group of entity types and business processes forms an information systems area.

The generalizations given above for both models and design strategies lead to a generic framework as depicted in Figure 3. The discussion of the framework is structured along the marked digits in the figure.

1. The application of specific design strategies to the contents of models leads to an information architecture.

2. Specific design strategies are formed depending on the specific models available. For instance, in case of the CBR, the availability of a Responsibility Model and an Information Model leads to the formulation of a specific design strategy which focusses on the responsibility for information systems areas. Application of this specific design strategy produces, among others, an information systems area for the Head Office as part of the information architecture.

3. In order to give a generic specification of design strategies, a generic structure of the available models has to be given.

4. Conceptual modelling is used in the next section to develop a model hierarchy for information architecture specification, ranging from the generic structure of models to the actual contents of models and the resulting information architecture.

5. The generic design strategies, of which impressions are given above, are formulated in terms of cluster analysis. Section 5 provides a brief discussion of cluster analysis, indicating the issues on which choices have to be made to transform generic design strategies into specific design strategies.

3.3 Decision Support for Information Architecture Specification

The generic framework for design strategies is used as the basis of a decision support system for information architecture specification, called GOSSIP (Groningen On-line Simulation Support for Information Planning). The GOSSIP system enables the information planning team to design several information architectures, each highlighting different issues that have to be taken into account. The expression "decision support system" reflects the fact that GOSSIP has not been designed to provide an optimal information architecture by itself. Rather, by providing several alternatives, it is meant to support and enhance the decision making process. However, the provision of alternatives is hardly useful if it is not combined with instruments for the evaluation of alternative information architectures.

Therefore, GOSSIP implements a set of evaluation criteria for information architectures. As each design strategies focusses on a distinct issue to be taken into account in the specification of an information architecture, such issues are also the foundation of the evaluation criteria. For example, one of the aforementioned specific design strategy focusses on the issue of responsibility for information systems areas. An evaluation criterion based upon this issue can produce a ranking of all information architectures specified. The information architecture that has the highest ranking will follow the responsibility structure better than the other ones. Section 5 will show that, as design strategies can be formulated in terms of cluster analysis, evaluation criteria can be expressed through the notion of cluster validity.

4. CONCEPTUAL MODELLING

In our view, the framework, and hence the DSS, should be independent of a specific organization and IP method. In order to ensure this independence, conceptual modelling is used. Several levels of modelling are presented, which together form a hierarchy of models.

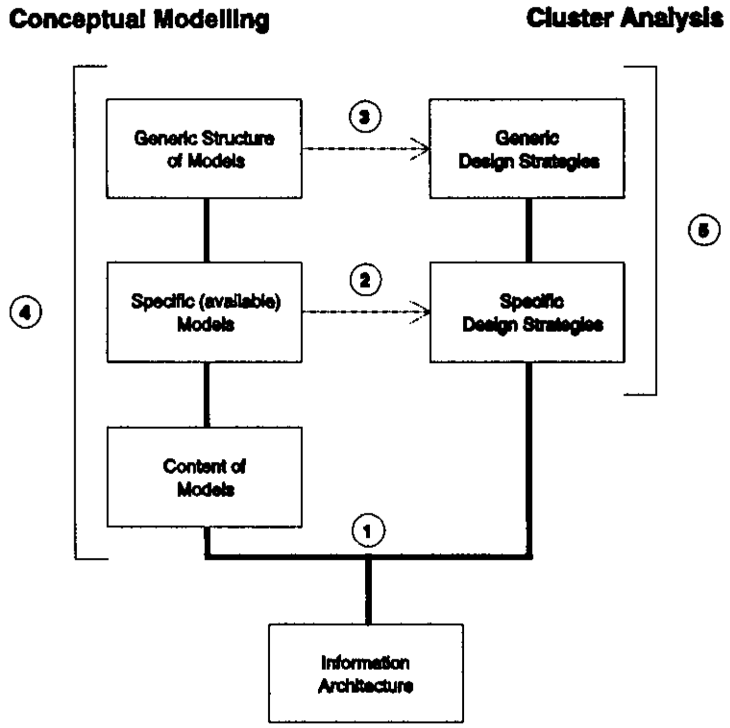

Figure 3. A Framework for Design Strategies

4.1 Generic Structure of Models: A Meta Model

The generic structure of models that are used in an information planning study can be described as follows. Each model covers one or more dimensions, which contain a number of elements. Between a pair of dimensions a relationship type can exist, specifying the kind of relationship that elements of these dimensions can have with one another.[3] Furthermore, because it is the intention to specify an information architecture, groups and group relationships are embodied in the generic structure. A group consists of elements belonging to several dimensions. The relationships between elements of different groups give rise to relationships between groups. This generic structure is called a meta model for information architecture specification.

4.2 Specific Models: A Model Type

A specific method for information planning defines the models to be specified. For example, the ISP method, as used at the CBR, calls for a Process Model, an Entity Model, an Information Model, a Responsibility Model and a Distribution Model. In terms of dimensions and relationship types, as specified in the meta model, the model type used by the ISP method is described in Table 1. It follows that a model type can be defined by giving substantial specifications of the dimensions and relationship types that are covered by the prescribed models.

4.3 Content of Models: The Business Information Model

The Business Information Model includes substantial specifications of the elements and relationships present in the models. Examples of elements from the CBR models are Head Office (an Organizational Unit), Medical Administration (a Business Function), and Medical Status (an Entity Type). An example of a relationship is "Medical Administration consists of Produce Medical Statement." This relationship belongs to the relationship type "Function consists of Process."

4.4 Information Architecture

An information architecture consists of groups of business processes and entity types. Examples of such groups were given in section 2, where parts of information architectures were given. An example of a group relationship can be found in the ISP architecture, where the Medical Information System uses information from the Candidate System.

Table 1. Description of Model Type Used by the ISP Method

Dimensions	Relationship types
Business Function	Function consists of Function
Business Process	Function consists of Process
Entity Type	Process involves Entity Type
Physical Location	Process is performed at Physical Location
Organizational Unit	Organizational Unit is responsible for Process

group	element	dimension	MODEL	relationship type	relationship	group relationship
▲	▲	▲	MM	▲	▲	▲
▲	▲	■	MT	■	▲	▲
▲	■	■	BIM	■	■	▲
■	■	■	IA	■	■	■

▲: model includes structural specifications in given category
■: model includes substantial specifications in given category

Figure 4. Model Hierarchy for Information Architecture Specification

4.5 A Model Hierarchy

The models specified above can be summarized by a model hierarchy as depicted in Figure 4. Each model is defined at a different level of abstraction. The meta model (MM) only contains structural (generic) specifications. The model type (MT) is defined at a methodological level, where only the information planning method used is known. The business information model (BIM) is specific for a certain organization. Finally, the information architecture (IA) forms the result of applying design strategies to the business information model.

5. DESIGN STRATEGIES

In specifying an information architecture, a specific design strategy is used for grouping processes or entity types. The groups are formed in such a way that processes (or entity types) that are in some way similar to each other belong to the same group. Apparently, keywords in applying a specific design strategy are "grouping" and "similar." A generic design strategy which incorporates these keywords is cluster analysis.

5.1 Generic Design Strategies: Cluster Analysis

Cluster analysis can be defined as the dividing of a set of objects into several subsets with the use of similarity measurements. Each subset is called a cluster, and all clusters together form a clustering. The similarity measurements are the result of the comparison of pairs of objects. Given similarity measurements, a cluster analysis can result in different clusterings. These differences originate in the exact order in which objects and already existing clusters are joined to form new clusters.

In general, one can say that a clustering of objects depends on

- how the similarity measurements are defined.
- which cluster technique is used: how the objects and already existing clusters are joined to form new clusters.

The similarity of objects can be established in several ways. For example, the objects can be compared directly by asking people their opinion on the similarity of the objects. However, this direct determination of similarity measurements has several methodological difficulties. It is also possible to compare the objects indirectly by means of the scores on a number of attributes. These scores are then combined into one single measurement. The combining of attribute scores into one measurement can be done in a lot of different ways. This results in different similarity measurements. In the following, we will only apply indirect similarity measurements.

The definition of similarity measurements depends on

- the attributes that are used and the scores on these attributes.
- the similarity measure that is used: in what way the several attribute scores are combined into one single measurement.

A cluster technique determines the eventual structure of the clustering. A clustering structure refers to the shape of the clusters; for example, clusters can have a compact shape which means that every object in such a cluster has a minimal degree of similarity to all the other objects in that same cluster. Another possibility is that clusters have a chain shape, which means that every object in such a cluster has a minimal degree of similarity to only a limited number of other objects in that cluster.

One can conclude that cluster techniques do not necessarily retrieve that division that is "naturally" present in the objects (Everitt 1974). Instead, cluster techniques impose a clustering structure upon the objects that is inherent to the technique, although one technique can be more compulsory than others. Therefore, if one is looking for a clustering that comes as close as possible to the "natural" clustering of the objects, then one has to carefully choose the cluster technique that has to be used. In such cases one really encounters the problem of what comes first: to find a clustering one has to already know the clustering structure.

Alternatively, this property of cluster techniques can be used to advantage. Suppose one wants to use the clustering for a specific purpose. If in the light of this purpose a certain shape of clusters is desired (the clustering structure), one can enforce this shape by using the "right" cluster technique. For example, suppose that one wants to have compact clusters, then one should not use cluster techniques that impose a chain shape upon the clustering.

When we view a generic design strategy in terms of the meta model, cluster analysis groups elements of a certain dimension. This grouping is based upon relationships between these elements and elements of other dimensions. These relationships are transformed into similarity measurements, which form the input of a cluster technique.[4]

5.2 Generic Evaluation Criteria: Cluster Validity

Cluster analysis can produce different clusterings, meaning that objects that belong to one cluster according to a particular cluster technique, belong to different clusters according to another cluster technique. Because of this, one can ask oneself the question: which clustering is the best. An achieved clustering can be judged from two different points of view (Mezzich and Solomon 1980):

- A cluster technical point of view.
- A substantial point of view.

From the first point of view, it is determined to what degree an object belongs to a certain cluster and not to another cluster. So it is evaluated how clearly clusters are separated from each other. Several indices have been developed to measure this (see Jain and Dubes 1988). From the second point of view, the clustering is judged according to the question: how good is the clustering with respect to its future use. Using this criterion, the achieved clusterings are judged according to the objectives of the cluster analysis.

5.3 Specific Design Strategies and Evaluation Criteria

The choices for (1) the objects to be clustered, (2) the attributes (elements of other dimensions), (3) the similarity measure, and (4) the cluster technique together form a specific design strategy. For example, the specific design strategy which focusses on the responsibility for business processes (as described in sections 2 and 3) has been modelled by means of the following choices:

1. Business processes are the objects that have to be clustered.
2. The organizational units are the attributes for the cluster analysis, with attribute scores consisting of a binary score "organizational unit is responsible for the process."
3. The similarity measure is the Russell-Rao coefficient.
4. The cluster technique is Complete Linkage.

These choices have been based upon an investigation of the properties of several similarity measures and cluster techniques. A thorough discussion of these topics will not be given here, but can be found in Anderberg (1973).

An example of a specific evaluation criterion is what is the extent to which the information architecture differentiates between departmental and interdepartmental information systems areas. This evaluation criterion can be modelled along similar lines:

1. Business processes are the objects.
2. Organizational units are the attributes.
3. The similarity measure is the Russell-Rao coefficient.
4. The criterion index is a modified Davies-Bouldin index.

6. APPLICATION OF THE GOSSIP SYSTEM

The GOSSIP system, which implements the concepts introduced above, has been applied to the CBR case. The input for the system has been created by means of an interface with IEW. Although several alternative information architectures have been developed by using GOSSIP, only one alternative is elaborated in this section.

6.1 Reasons to Consider GOSSIP for the CBR

The information plan for the CBR contains an information architecture, which specifies the information systems to be developed. The CBR architecture is a modified version of the initial ISP architecture. Considerations not covered by the ISP method led to these modifications, as described in section 2. However, it is not clear whether the CBR architecture can cope well with issues, such as the consequences of the reorganization.

One of the major aims of the proposed reorganization of the CBR is to arrive at large-scale decentralization and deconcentration of tasks and responsibilities. For this purpose, regional offices are to be established throughout the country, and be responsible for tactical and operational decisions. One would expect the information plan to reflect this intended decentralization. This means that the proposed regional offices should be able to decide upon their information systems services themselves or, at least, have control over their own hardware and software systems. However, if we look at the distribution of information systems over head office, regional offices, and other locations (TEC/PEC/BFI), we get a picture like the one presented in Figure 5. This picture tells usthat, apart from the Medical System, the Head Office is a participant in all information systems specified. Therefore, it is hard to establish local responsibility for information systems development and use, as implied by the decentralization concept.

Moreover, when using the CBR architecture for the design of the future hardware configuration, one may conclude that it is ahead of the technological possibilities. The combination of a deconcentrated hardware concept and general use of common data calls for the application of a distributed database management system (DDBMS). Since a stable implementation of this technology is not yet on the market, the CBR has chosen a central hardware solution, combined with a national terminal network. It is obvious that this solution does not compare well with the intentions of decentralization and deconcentration, as strived for by the reorganization process. This prompted us to apply the GOSSIP system to come up with an alternative which focusses explicitly on the geographical aspect.

6.2 Specification of an Information Architecture Using GOSSIP

By employing the GOSSIP system, it becomes possible to develop alternative information architectures. In this example, we have chosen to develop an alternative architecture that pays more attention to the decentralization and deconcentration aspects. The cluster techniques and similarity measures offered by GOSSIP provide the possibility of clustering processes on the basis of the location at which they are performed. As described in section 5.3, this can be implemented by making choices about the objects, the attributes, the similarity measure, and the cluster technique. In this case, it has been realized by choosing processes as the objects to be clustered, the locations being the attributes, and the attribute score consisting of a binary score "process is performed at location." The similarity measure employed here is the so-called Simple Matching coefficient, which is combined with Ward's cluster technique. The interpretation of this strategy for information architecture specification could be "Specify information systems that are homogeneous in terms of location(s) of use, with respect to the processes these systems are designed to support." The information systems that are identified by following this strategy are

- Central control system
- Local control system
- Coordination system Region/Central
- Application Processing
- Registration Practical Exams
- Article Administration
- Counter system TEC/PEC

The consequences for the distribution of information systems over locations are depicted in Figure 6. In contrast to Figure 5, only three of the seven proposed information systems have potential Head Office/Regional Office conflicts. The general philosophy of the GOSSIP specified information architecture is that locally performed tasks are supported by information systems for which the Regional Office itself is responsible.

6.3 Comparison of Alternative Architectures

The above clearly shows the ability of the GOSSIP system to come up with truly different information architectures. The utility of these alternatives can be investigated in two ways, as was described in section 5.2. The proposed evaluation criteria, as implemented in GOSSIP, provide a judgement of alternative information architectures on technical grounds. The scores of the two architectures on some of the evaluation criteria are given in Table 2. In general these evaluation criteria are defined in analogy to the Davies-Bouldin index as described in Jain and Dubes (1988, p. 185-186). The criteria evaluate the architectures with respect to

- the (lack of) spread of information systems over geographical locations.

	Candidate System	Medical System	Theoretical Application System	Practical Application System	Practical Scheduling System	Theoretical Examination System	Practical Examination System	General Ledger System	Payroll System	Accounts System	Facilities System	Article System	Personnel System	Relation System	Meta Information System
Head Office	√		√	√	√	√	√	√	√	√	√	√	√	√	√
Regional Office	√	√	√	√	√	√	√	√	√		√	√	√	√	
Theoretical Examination Centre	√		√			√						√	√		
Practical Examination Centre							√					√			
Bureau Further Investigation				√	√		√					√	√	√	

Figure 5. Distribution Matrix for CBR Architecture

	Central control system	Local control system	Coordination system Region/Central	Application Administration	Registration Practical Exams	Article Administration	Counter system TEC/PEC
Head Office	√	√	√			√	
Regional Office		√	√	√	√	√	√
Theoretical Examination Centre						√	√
Practical Examination Centre						√	√
Bureau Further Investigation		√			√	√	

Figure 6. Distribution Matrix for GOSSIP Specified Architecture

Table 2. Ranking on Evaluation Criteria

	location	function	entity type	organizational unit
GOSSIP	better			better
ISP		better	better	

- the coincidence of information systems with functional areas.
- the (lack of) common use of data on entity types by different information systems.
- the coincidence of information systems with organizational units.

The scores in Table 2 tell us that the GOSSIP information architecture can be ranked higher, on cluster technical grounds, than the CBR architecture, when viewed from the location and organizational angles. On the other hand, with respect to the compliance with functional boundaries and the amount of data exchange between information systems, the GOSSIP information architecture has to be judged inferior to the CBR architecture. It is up to the information planning team to assign relative importance to these criteria and choose one of the architectures as the final architecture. However, it is also possible to incorporate features of one architecture into the other, which renders a new alternative architecture, which might outperform both of them with respect to these evaluation criteria.

7. CONCLUSION

We have examined the applicability of cluster analysis to information architecture specification. The development of conceptual models has rendered a suitable problem representation. The implementation of cluster techniques in a decision support system enables us to apply these techniques to a practical problem of information architecture specification. The results of applying the decision support system show that alternative architectures can be generated, based on objectives that are recognized in practice, but have not been implemented by known information planning methods. Moreover, the employment of an abstract problem representation and cluster techniques enables us to generate evaluation scores for alternative architectures. In our view, this is an improvement on the usual informal and intuitive manner of comparison. Whether the decision support system that implements the cluster techniques is regarded useful by information planning practitioners is subject to further research. Other proposed research topics include investigating the relevant characteristics of the problem representation for information architecture specification and the transferability of this approach to other decision making situations.

8. ACKNOWLEDGMENTS

The authors wish to thank Jacques Boersma, John Simons, Enno Ebels, Pieter Terlouw, Anja Koopman and the anonymous reviewers for their helpful comments.

9. REFERENCES

Anderberg, M. R. *Cluster Analysis for Applications*. New York: Academic Press, 1973.

Boersma, S. K. Th. *Beslissingsondersteunende Systemen: Een Praktijkgerichte Benadering (Decision Support Systems: A Practice Based Approach)*. Schoonhoven, The Netherlands: Academic Service, 1989.

Everitt, B. S. *Cluster Analysis*. London: Heinemann Educational Books Ltd, 1974.

International Business Machines. *Business Systems Planning — Information Systems Planning Guide*. GE20-0527, Fourth Edition. June 1984.

Jain, A. K., and Dubes, R. C. *Algorithms for Clustering Data*. Englewood Cliffs, New Jersey: Prentice Hall, 1988.

Mezzich, J. E., and Solomon, H. *Taxonomy and Behavioral Science*. London: Academic Press, 1980.

10. ENDNOTES

1. The case organization described here is a Dutch organization, therefore most of the terminology was originally in Dutch. An attempt has been made to translate the organization's terminology, but by lack of comparable English terms, some of these translations may seem rather awkward.

2. The terms information system and information systems area are sometimes used interchangeably. In theory, one information systems are could lead to the development of several information systems.

3. Mathematically, this can be stated as:

$$dimension_m = \{element_1,\ldots,element_M\}$$
$$V_n = \{value_1,\ldots,value_N\}$$
$$element_i \in dimension_{m_1},\quad element_j \in dim$$
$$value_k \in V_n$$
$$(element_i,element_j,value_k) = relationship_r$$
$$relationship_r \in relationshiptype_R$$
$$= dimension_{m_1} \times dimension_{m}$$

4. Affinity analysis as provided by tools for Information Planning such as IEW limits the choice to the dimensions and the relationships. It can be shown that affinity analysis is equivalent to a specific combination of a cluster technique and a similarity measure.