Enhancing the applicability of empirical data using legal knowledge based systems[1]

Jörgen S. Svensson
Department of Public Administration
Twente University
P.O.-box 217
7500 AE Enschede
The Netherlands
email: J.S.Svensson@bsk.utwente.nl

Abstract

Information is considered to be an important prerequisite for the process of social policy making. In the practice of policy preparation however, important problems exist in providing all the information deemed necessary.

This paper discusses the possibility of using legal knowledge-based systems to enhance the usability of data sets. As I will show, the method of Knowledge Based Microsimulation may be used to enrich data sets with information about the micro-effects of different social policies (existing or proposed).

In several experiments in the context of The Netherlands' General Assistance Act, it was found that this process of enrichment can produce very reliable results, and that there is a high correspondence between the outcomes of the simulations and the decisions made in administrative practice.

This high correspondence is also interesting in view of discussions about the role of legislation in public administration. It may well lead to a re-valuation of legislation as a factor in administrative decision making.

1. Introduction

Legislation plays an important role as a basis for administrative decision making in the Netherlands. Given this importance, it is desired that the legislator is well informed about the administrative practice and about the effects of his legislation on that practice. The legislator wants to have insight in the field for which the legislation is intended, wants to know what effects the existing legislation has, and wants to be able to predict the effects of possible changes in that legislation. Currently several problems exist in providing the legislator with this desired information requirements.

The ExpertiSZe research project (Svensson et al., 1992) started in 1987 as a joint effort of the Netherlands' Ministry of Social Affairs and Employment and Twente University. Its broad aim is to improve the quality of legislation through the application of knowledge based systems techniques. One part of this project has been dedicated to the possibility of applying knowledge based models of social security legislation to administrative data, by means of a so-called knowledge based microsimulation approach. Knowledge based microsimulation (KBMS) might help to provide better information to the process of policy development.

In the period 1991 - 1994 the possibilities of knowledge based microsimulation were examined. Several experiments were performed on the basis of a rule-based model of the Netherlands' National Standards Order, an order which specifies in much detail the level of the benefits to be provided under the General Assistance Act.

In this paper I will start with a description of the current problems in informing the legislator in relation to General Assistance Act. After having described these problems I will discuss the method of knowledge based microsimulation and determine its primary possibilities.

This discussion will be followed by a presentation of the results of several experiments which have led to the insight that KBMS offers more than these primary possibilities alone. At the end of the article conclusions are drawn about the usability of KBMS as a method for improving the quality of information available to the legislator.

2. The General Assistance Policy and the provision of information

General Assistance in The Netherlands is based on The (National) General Assistance Act. Under this act municipalities are obliged to give assistance to citizens who have themselves insufficient means of existence. The assistance by the municipalities can have several forms, of which one is the provision of a monetary benefit. Although the municipalities are responsible for

providing this benefit, it is for the larger part (90 percent) financed by the Dutch state.

Because of the financial consequences for the state, the amount of the monetary benefit is regulated on the national level by a special law: The National Standards Order. The National Standards Order contains many guidelines on how to determine the monetary benefit for a person or a family, based on characteristics of that person or family. The local municipalities have an obligation to follow these guidelines, but may deviate from them if these guidelines lead to inadequate decisions for specific cases in the specific municipal context.

In this organisational context, with a large distance between the national legislator and the local agent, the legislator wants to be informed about at least the following three topics:

1. The characteristics of the population to which the legislation is applied;
2. The (socio-economic) effects of the legislation in force;
3. The expected effects of possible changes in the legislation.

The problem of providing the information

It is the task of the Ministry of Social Affairs and Employment to provide in this desired information.

Information about the state of the social security system is partly provided on a regular basis. There are for instance several fixed dates on which certain information has to be provided to allow the (preparation of) parliamentary budget discussions. For another part the information to be provided by the ministry is determined by ad-hoc questions which are asked at the decision making level, by ministers, secretaries of state and by parliament. These additional requests for information can arise at unexpected moments and may address very specific topics. If we take the context of the Netherlands's social security system, some examples of such questions are;

- How many divorced women receive a General Assistance benefit?
- What is currently the average benefit for one-parent-families of which the parent is under the age of 21?
- What will be the average monetary effect for two-parent-families if the proposed change in the housing-benefit scheme is implemented?

Given the usual pace in politics, there is seldom much time to answer such questions. Answers needed for a debate are often required within a few days or even a few hours.

Therefore, the ministry has to prepare itself to be able to answer all the questions that may logically be expected. How can it do that?

Answering questions: the data base model

At first, preparing for all questions that can be formulated, may seem a hopeless task; One fool may ask more questions than ten wise men may answer. A closer look at the factual questions which are asked, reveals however that a lot of these questions are in fact similar in structure. Many of the factual questions about social security legislation typically concern aggregations on three types of data:

- data about the case itself;
- data about the consequences of the existing legislation;
- data about the consequences of proposed changes in the legislation.

For example: the question "What will be the average monetary effect for two-parent-families if the proposed change in the housing benefit scheme is implemented?" asks for:

- an aggregation (in the form of an average);
- concerning the difference of two legal effects (the housing benefits under the new and under the old legislation);
- for cases with specific characteristics (making them two-parent families).

In this perspective, it becomes evident that the ministry may (theoretically) prepare itself for all of these questions by constructing a huge representative data base concerning *individual* beneficiaries, which should contain all relevant case-characteristics, all relevant legal consequences under the current legislation and all relevant legal consequences under different legislative alternatives.

The factual questions concerning combinations of client characteristics and legal effects could then be answered by simple selection and aggregation, in the form of data base queries, on this set.

Current problems concerning data collection and forecasting

The data base model, as described above, forms the rationale behind many of the current efforts in data collection. The Ministry of Social Affairs and Employment spends millions of guilders each year in order to get data sets which should be useful in answering policy questions. The results reached sofar are however rather disappointing.

With respect to the three types of data wanted, three major problems exist. Firstly, it proves very difficult to collect information about all the relevant client characteristics. Apart from the general problem of collecting reliable information about individuals, a big problem is that it is actually unclear what "all relevant characteristics" are. There is a limited insight in the topics parliament will discuss in the future and therefore data sets which are constructed in advance often prove inapplicable at the moment when they are to be used. Items addressed in surveys are often selected on the basis of existing political discussions. When the survey data become available eventually the topic of the discussions may have changed and other information will be required. This new desired information may very well be

information nobody had thought about at the time that the survey was set up.

Secondly, it is very difficult to collect all relevant consequences of the legislation. Although it is possible to collect data about the total benefit in a specific case, data about the specific sub-decisions, which have resulted in this total benefit, are hard to get. The only place such detailed data can be found currently is in the administrations of the local social security offices. Retrieving this information from these offices is however time consuming and costly, and moreover the possibilities of getting the desired data depend on the willingness and possibilities (practical and juridical) of the local offices to collaborate in such surveys.

Thirdly, it is currently impossible to get the necessary micro-data on the effects of new policy amendments. The effects of amendments can generally only be determined by domain experts, who are very scarce and who often need quite some time themselves to determine the exact effects of a change for a single individual, let alone for a whole data base of individuals.

All three problems: the problem of collecting the right data, the problem of collecting detailed data about legislation in force and the problem of forecasting the consequences of amendments to the regulations, are considered very serious. There have been several occasions in which information delivered by the ministry later proved to be incorrect. This has blemished the reputation of the ministry in the eyes of the general public and in the eyes of the political decision makers. The limited ability of the ministry to give accurate information is regarded especially serious with regard to the enormous investments in data collection and information systems in recent years.

3. Knowledge-based microsimulation

In order to find a solution for especially the second and third problem, we looked at the possibilities of two developments in information technology. On the one hand, there is a development going on in the field of simulation, partly owing to the availability of faster and faster computers, towards an increasing level of detail in modelling. This development has led to a shift from simulations at macro- and meso-level towards micro-analytical simulations. In micro-analytical simulations, use is made of data and models at the level of the units of which a whole is composed (Orcutt et al., 1986). Developments in poverty for example may be investigated by simulating the development in incomes and expenses of individual households in a population.

On the other hand, there are the results of research in A.I. and Law. In recent years much insight has been gained in how to develop models of legislation. It has been proved that, at least in some domains, it is possible to develop models of legislation and regulation which can advice on decisions with respect to individual cases.
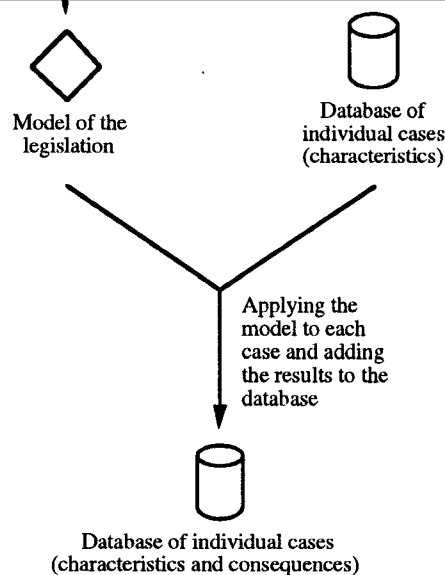
Our own experiences with the development of the Twente Expert system for Social SECurity (TESSEC) are of great importance in this connection. In TESSEC we used a rule base approach to model written knowledge sources (especially laws and royal and ministerial decisions). By applying what we call a method of one-to-one representation[2], in which separate, relevant, elements in a text, are represented in corresponding elements (a rule or a set of rules) in the knowledge base. We got very favourable results. In a laboratory research (Nieuwenhuis, 1989) it was proved that for the Dutch General Assistance Act our one-to-one approach led to a reliable model of the legislation in question, and that this model could be applied to accurately determine detailed legal effects for individual cases.

Based on the developments in simulation and knowledge-based technology a new method for ascertaining socio-economic consequences of regulation seems possible, namely that of knowledge based micro simulation (or KBMS).

In knowledge based micro simulation the application of legislation is simulated on the level of individual cases (generally persons or households). A knowledge based model of the legislation in question is used to determine for each individual case in a database of cases the consequences of this legislation.

In its most simple form, this knowledge based micro-simulation method consists of three steps (see figure 1):

1. Developing a knowledge based model of the legislation in question (this may be existing legislation which is already in force or a proposal for alternative legislation which is the subject of political debate). This model allows the automatic determination of conclusions about the impact of this legislation on a specific case.

2. Collecting for a large representative sample of individual cases the data which are needed to apply this model, and administering these data in a database. (Or acquiring a similar data base already existing).

3. Automatically applying the model to each case in the data base (thus simulating the application of the legislation to each individual) and adding the individual results of this application to the original database in order to get a new database which contains both the original data as well as the inferred consequences of the legislation.

Model of the legislation

Database of individual cases (characteristics)

Applying the model to each case and adding the results to the database

Database of individual cases (characteristics and consequences)

Foreseen applications of the KBMS method and scepticism

The idea is that knowledge based micro simulation may be applied to determine the effects of legislation. With respect to our described problems we firstly think that knowledge based models may be used to determine the detailed effects of existing regulation, and thus may reduce the need for empirical data about these effects. Secondly, we think it possible to use KBMS to determine the effects of policy amendments by using models of the amended regulation.

If both these applications are possible, we may in the future use surveys in which only data about case characteristics are collected and transform them, through the application of our models, in complete databases which can be used in answering policy questions.

However, many people are very sceptical about this knowledge based micro simulation approach. In many arguments, it is claimed that the actual process of determining social security benefits is much more than just applying the rules in the legislation. Or, in the words of Moles and Dayal (1993): there is more to life

with statements made by Hunter et al. (1993) and Bench-Capon (1994) our argument is: We are aware that our method of modelling will not capture the full richness of real juridical decision making, but this does not necessarily mean that the rule based approach can not be a very powerful method to predict the outcomes of social security decisions.

We also think that the effects of case law and local policy may be limited (they may only apply to a limited number of people or they may have only marginal effects on the actual decisions). However, we observe there is clearly a serious doubt about the possibilities of simulating the actual application of legislation by means of rule based models of that legislation. In order to determine the value of our method, empirical research is necessary. The level of correspondence between the results of our model and the actual decisions in practice has to be determined objectively on the basis of experiments.

4. Experiments with KBMS

In order to determine if our method can provide valid information about the socio-economic effects of legislation, we have set up several experiments. The model we use in these experiments is a model of the National Standards Order. It is a rule-based model which contains 45 questions about characteristics of the case, several of which may be asked for more than one person relevant to that case (e.g. the one question about the age of a child will be asked for each child present). The answers to the questions are used by the model to determine, using up to 95 intermediate parameters, the level of the monthly General Assistance Benefit.

The aim of the experiments is:
To determine how the outcomes of our model of the National Standards Order correspond with the actual decisions made by municipal social security offices based on this Order.

The first experiment which was performed will be discussed in some detail in order to explain the method and because it serves as an anchor point for the results of two subsidiary investigations, of which I will only discuss the outcomes.

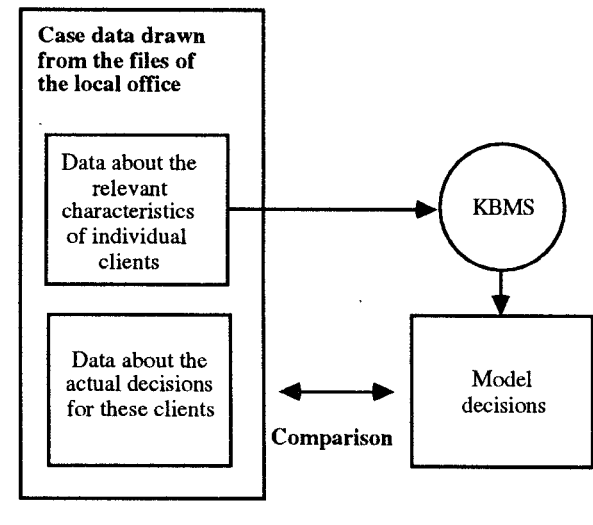## Design of the first experiment

For the first experiment, which took place in 1992, we have drawn a random sample of 101 cases from the client population of one municipal social security office. For each of these cases we have determined, using the files of the office, two kinds of data:

- the actual decisions taken by the local office on each of these cases;
- the client characteristics that are relevant to make these kind of decisions based on the existing legislation.

We then used our method of knowledge based micro simulation, in which we applied our model of the legislation in power at that moment in order to determine for each case the model-decision, based on the administered client characteristics. These model decisions were next compared with the actual decisions already administered. The experimental design is shown in figure 2.

Figure 2

Experimental design used to compare the outcomes of our model with the actual decisions made by a local office.

Results of the first experiment

As a result of the experiment both the benefit according to our model of the legislation and the actual benefit, as administered by the local security office, are known. A comparison of these amounts shows us that the actual correspondence between both amounts is high. In the 101 cases only 16 differences are found, half of which are under 100 guilders (see table 1).

Table 1

Overview of the discrepancies between the outcomes of our model and the actual decisions made by the local office (n=101).

| Discrepancy	Frequency
No difference	85
difference of less than 1 guilder	4 *
difference between 1 and 10 guilders	0
difference between 10 and 50 guilders	2
difference between 50 and 100 guilders	2
difference between 100 and 250 guilders	4
difference between 250 and 500 guilders	2
difference of more than 500 guilders	2
Total	101

* among which 3 discrepancies of less than 1 cent due to rounding errors

On an aggregated level the results of the simulation are thus highly compatible with the results of the data collection. The difference in average benefit is Dfl. 23.88 (the model average being less than 2% higher than the average benefit of Dfl. 1248.77 per month). The correspondence between both outcomes is probably best illustrated by their linear correlation which amounts to .95. (see figure 3).

Figure 3

Model decisions (axis Y) plotted against the decisions as implemented in practice (axis X)
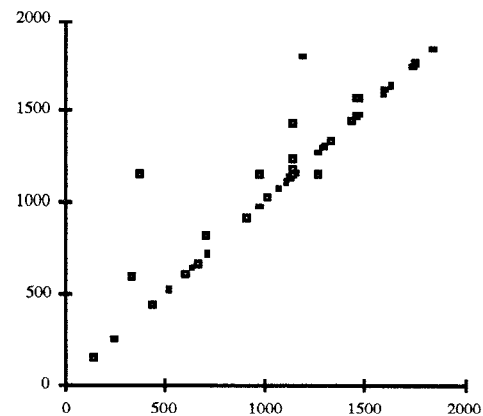
Results of two subsidiary experiments

Because the first experiment was limited to only one social security office of one municipality two other experiments have been performed. These experiments were carried out using two data sets from nation wide surveys which were already available.

An important, though expected, problem in using these two existing data sets was the availability of input data. It was found that in the two sets information about several relevant case characteristics were missing on both an incidental and structural basis (thereby showing that the question of what the relevant case characteristics are in the context of the social policy is indeed a difficult task).

This problem was solved by adding this missing information in advance of the actual simulation. By inferring the missing data from the information which was available and by using the Monte-Carlo technique[3] the data sets were made complete with respect to the input-requirements of our model of the National Standards Order.

Using the completed data sets in our simulations, we again reached high levels of correspondence between our model and the available information about the benefits (table 2).

Although the level of correspondence between the actual benefits and the outcomes of our model was not quite as high as in our first experiment, the results of these experiments show that KBMS may also be a method to enhance the value of existing data sets.

Table 2

Results of two subsidiary investigations in which the model of the National Standards Order was applied to existing data collections.

Data set	N of cases	correlation
SEP*	353	89
PES4**	449	78

*   Socio-Economic Panel
**   Project Evaluation Systems reform (PES4)

Another interesting conclusion reached in these experiments was that the deviation between our model outcomes and the data about the actual benefits was not so much the effect of a poor correspondence between our model and the social security practice, but much more the result of clear defects in the original case data. Moreover, it was discovered that in several cases the data about the actual benefits were clearly faulty.

> For example: in one of the surveys, couples under certain conditions are to be treated as two single individuals. In these cases the question about the level of the benefit has however led to unclarities for the people filling in the questionnaires. Some of them have filled in the total benefit for both spouses together, while others have only given the individual benefit for the spouse intended in the questionnaire. A comparison of the outcome of our model with this collected benefit

information shows us that the model results are probably more in correspondence with the actual decision being taken by the municipality than the 'real' data!

5. Conclusions and discussion

Legal knowledge based systems may be used to support the process of policy preparation in several ways (Allen, 1980; Bench-Capon, 1991; Svensson et al. 1993).

This paper has focused on a very specific type of application in this field, namely using knowledge based microsimulation to enhance the usability of empirical data sets.

The conclusion of our research is that it is possible to use models of legislation to add information about the likely effects of legislation to databases containing micro-level data off beneficiaries. With regard to the National Standards Order in The Netherlands, it has been proved that the information our model of that legislation yields is reliable in the sense that there is a high correspondence between the generated information and the information which can be acquired by administering the actual decisions (where this last type of information can only be obtained with great effort). In fact, our study suggests that the reliability of the information derived with the KBMS method in practice depends more on the reliability of the case data which are available for the simulation than on the adequacy of the rule-based model. In some cases our KBMS-method may even produce information of better quality than the current method of collecting this information in time an money consuming surveys.

In an investigation which was not discussed in this paper the method of KBMS was also used to predict, in a similar way, the effects of certain amendments to the legislation. These experiments also led to very useful results, in terms of reliable estimates of the direct effects of the proposed legislative change[4].

We conclude that the method of KBMS has indeed a potential to solve the two existing problems in generating policy information for which is was intended. Moreover, as a side-result of our research, we found our model may have interesting possibilities in view of the problem of collecting information about the relevant client characteristics. It was found that in using our simulation method we could point out several defects in the existing data bases. These defects had to do with both the type of information available in these data sets (not all relevant characteristics were present) as well as the quality of the information (some of the information proved unreliable). This conclusion has led us to the belief that we may use knowledge based micro simulation in the process of collecting future data. In a recent research project (Van Ginneken and Svensson, 1995) we have used the KBMS method in the process of data collection to get more reliable data.

One last question: Did we win the debate?
One last warning is probably in place. This research does not prove that critics of A.I. and Law, like Moles and Dayal, are wrong. There are undoubtedly discrepancies between *the manner* in which our model determines a benefit and the way this is done in administrative practice. What we proved however was that *the results* of both ways of reasoning, human and automatic, are very similar, and that the model thus can be used for the intended purpose. Moreover we have tried to show that it is rather pointless to discuss the validity of A.I.-models before-hand. The validity of models of law is not a general question which can be answered in a debate. It is dependant on characteristics of the field in question and of the intended use of the model. Insight in this validity may only be gained if steps are taken to test this validity of models empirically. Our results removed much scepticism about the possibilities of this kind of models in the social security context in the Netherlands. I suggest that in other fields and in other countries these kind of experiments have to be carried out as well, in order to really determine the current possibilities of A.I. and Law. As Hunter at al. (1993) argue, the results from such experiments may provide very useful insights which may in time lead to truly intelligent systems.

6. Acknowledgements

7. Notes

1. This paper is partly based on my dissertation (Svensson, 1993), which I have defended success-fully at Twente University.
2. The method of one-to-one representation (Nieuwenhuis, 1989) is a method similar to the ideas presented by Bench-Capon and Coenen (1992), in that it provides guidelines to maintain correspondence between the knowledge base and the source texts on which it is based.
3. The Monte-Carlo technique is a technique in which missing values on certain attributes are added to an individual case by randomly drawing a value from an existing distribution which represents the distribution of values for that attribute in the population concerned (e.g. if for a certain case the monthly rent is unknown, we may fill in a value for this rent by drawing that value from a distribution which represents the distribution of rents for similar cases).
4. Note that there are also indirect effects of legislation which are not considered in our project. Behavioural response, the fact that people may react on changes in legislation, is such an indirect effect. It may in time lead to changes in the population to which the legislation is to be applied.
   For example: recently there has been a change in the existing policy towards the income of the client. This income is now fully deducted from the benefit, where previously only a certain percentage of the income was regarded. This change will lead to the direct effect that people with income will receive lower benefits. Indirectly however, it may lead to a future situation in which these people will have a lower income because the financial incentive to get more income (by acquiring more or better paid work or by successfully demanding a raise) is taken away.

8. References

Allen, L.E. (1980). Language, law and logic: plain legal drafting for the electronic age, in Niblett, B. (ed.), *Computer Science and Law*. Cambridge, Cambridge University Press, p.75-100.

Bench-Capon, T.J.M. (1991, ed.). *Knowledge-based systems and legal applications*. Academic Press, London.

Bench-Capon, T.J.M. and F.P. Coenen (1992). Isomorphism and Legal Knowledge Based Systems. In: *Artificial Intelligence and Law*. Vol. 1, No. 1

Bench-Capon, T.J.M. (1994). Legal Theory and Legal KBS: a computer scientists perspective. In: H. Prakken, A.J. Muntjewerff, A. Soeteman en R.G.F. Winkels (eds.). *Legal Knowledge Based Systems: the Relation with Legal Theory*. Koninklijke Vermande, Lelystad.

Hunter, D., A. Tyree and J. Zeleznikow (1993). There is less to this argument than meets the eye. In: *Journal of law and information science*. Vol 4. Nr. 1.

Moles, R.N. and S. Dayal (1992). There is more to life than logic. In: *Journal of law and information science*. Vol 3. Nr. 2.

Nieuwenhuis, M.A. (1989), *TESSEC: een expertsysteem voor de Algemene Bijstandswet*. Deventer, Kluwer.

Orcutt, G.H., J. Merz en H. Quinke (1986), *Microanalytic Simulation Models to support Social and Financial Policy*. Amsterdam, Elseviers Science Publishers B.V.

Svensson, J.S., P.J.M. Kordelaar, J.G.J. Wassink and G.J. van 't Eind (1992). ExpertiSZe, a Tool for Determining the Effects of Social Security Legislation. In: Grütters, C.A.F.M., J.A.P.J. Breuker, H.J. van den Herik, A.H.J. Schmidt and C.N.J. de Vey Mestdagh (eds), *Legal Knowledge Based Systems: Information Technology & Law, JURIX'92*. Lelystad, Koninklijke Vermande.

Svensson, J.S. (1993). *Kennisgebaseerde Microsimulatie: een nieuwe methode voor het voorspellen van de*

gevolgen van wet- en regelgeving. (Phd. thesis). Enschede, Universiteit Twente.

Svensson, J.S., J.G.J. Wassink and B. van Buggenhout (eds.) (1993). *Legal Knowledge Based Systems: Jurix '93: Intelligent Tools for Drafting Legislation, Computer-Supported Comparison of Law.* Lelystad, Koninklijke Vermande.

Van Ginneken, J. and J.S. Svensson (1995, prep). *Improving the quality of information in the social security domain.*

.

.