

Development, Implementation and Use of a Judicial Data Space System

Jan van Dijk

Ministry of Security and Justice
Research and Documentation Centre
The Hague, The Netherlands

j.j.van.dijk@minvenj.nl

Sandra Kalidien

Ministry of Security and Justice
Research and Documentation Centre
The Hague, The Netherlands

s.n.kalidien@minvenj.nl

Sunil Choenni

Ministry of Security and Justice
Research and Documentation Centre
The Hague and Rotterdam University
of Applied Sciences, The Netherlands

r.choenni@minvenj.nl

ABSTRACT

For the purpose of good and trustworthy management information for the government, our research institute has built a data space system that makes it possible to periodically monitor data flows within as well as between organizations of the justice domain. The aim of this system is to get insight into the success and possible bottlenecks in the criminal justice domain. Data from heterogeneous sources are extracted and related by means of domain knowledge. The system consists of three layers: a data space, a space manager and an interface layer. The data space layer consists of a set of heterogeneous databases. The space manager layer – implemented as a variable database and a relationship module – consists of domain knowledge. The relationship module decides which databases in the data space layer are required to answer a query. The interface layer consists of a user friendly web interface, which facilitates policy makers to interpret data flows within the justice domain.

Categories and Subject Descriptors

H.2.4 [Database Management]: Systems – *rule-based databases*.

General Terms

Management, Measurement, Performance, Design, Reliability, Experimentation, Standardization.

Keywords

Data space system, heterogeneous data, knowledge exploitation, chain management.

1. INTRODUCTION

In the field of the criminal justice domain, many organizations are involved that can be considered partners in a chain of events (i.e. criminal cases, offenders or victims). This implies that the input of one organization is dependent on the output of another organization in the chain (e.g. [1-2]). From the perspective of the government responsible for this domain, it is desirable to have a good and reliable understanding of the flows within as well as between the organizations in this domain [3]. Having a good understanding of these flows may give insight in possible bottlenecks in the justice domain. This way, the government can

actively stimulate organizations to solve potential problems at an early stage, which in turn may lead to a system that acts more efficiently.

Current means, such as statistical yearbooks may help the government to get insight in the criminal justice chain. For instance, the Dutch publication ‘Crime and Law enforcement’ describes developments of crimes experienced by victims, crimes registered by the police and the reaction of the criminal justice system on crime [4]. The book is updated annually and is valuable to a broad public, such as policy makers, criminologists and media. However, for the means of monitoring data flows within and between organizations the book is less practical, as it is not easy to compare and combine the data of different organizations [1].

Furthermore, the information need of the government is dependent on the policy that is applicable on a certain moment in time. For instance, the current Dutch government’s policy on the justice domain is to reduce crimes such as street robberies, burglaries and violent crime in general [5]. Detailed information for specifically these types of crimes are of interest to the government at this time. So, the information need of the government is in some sense ‘ad hoc’, depending on the topics in daily politics. Yearbooks often have static information available, with limited possibilities to drill down information. Moreover, to be able to act on the criminal justice system on possible bottlenecks in an early stage, the government requires information rather quickly and as recent as possible. So more frequent updates than on a yearly basis are a must.

But books are not just less practical. Until recently, policy makers and experts from the justice domain periodically discussed developments of data flows in the justice domain through meetings with some organizations in the justice domain, assisted by a manually composed collection of written reports describing main indicators for each of these organizations. These meetings were less practical and useful, since the focus often remained on one organization (in particular the Public Prosecution, having a central role in the Dutch Criminal Justice System), and a clear understanding of what happens with data flows between organizations was missing. If there was a comparison between organizations this happened in an unstructured way. Moreover, only a few organizations were involved in these meetings.

To improve the management information in the justice domain, the government recently requested our research institute to develop and implement a system that could periodically provide reliable and valid management information on the criminal justice domain. The system should among other things, be able to monitor flows within and between organizations, periodically

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Conference '10, Month 1–2, 2010, City, State, Country.

Copyright 2010 ACM 1-58113-000-0/00/0010 ...\$15.00.

show the data in a way that makes it easy for experts to draw conclusions and have the possibility to produce graphs and reports. To realize such a system, data from different heterogeneous sources need to be extracted and related, see amongst others [2, 6-9]. While in many fields data from different databases are related via automated data integration concepts and techniques – such as schema mapping or tuple mapping approaches (see amongst others [10-12]), data in the field of justice are manually or semi-manually related. The fact that data are related in a (semi-)manual manner in this specific field has to do with the complexity that is required in understanding and interpreting the data ([13-14], [2]). In our case relating data means to bring (aggregated) data together that pertain to a same real world entity and define how the data should be interpreted in the context of other real world entities. Furthermore, a concept of primary/foreign key relation [15] between different databases is often lacking. If such a relation does exist, privacy law and regulations may not allow us to use it. Therefore, relating data from different information systems (semi-)manually is an error-prone and tedious process.

In this paper, we discuss the design, implementation, and use of a data space system for generating management information from heterogeneous databases in the criminal justice field. To build such a system we had to face several challenges. The major challenges were that the different sources appeared to be independent and therefore technically hard to relate to each other in an automatic way. Second, some of the sources are “person” oriented, while others are “case” oriented. For example, the police record data about suspects in their information systems. These suspects are delivered to the public prosecutor. However, the information systems of the public prosecutor pertains to cases. The mapping of persons to cases is not always clear from the conceptual schemata and documentation of different information systems. To tackle these problems, we have developed a system that exploits domain knowledge from experts and combines this with the concepts of data space systems [16].

Our system consists of a number of modules. An important module is the so-called relationship module. In this module, domain knowledge of experts is modeled. We have used this module to encode the relationships between different databases.

Our experience is that in the field of safety and justice, domain experts have a lot of knowledge about the relationships between different databases and the semantics of the data in the databases [9]. Our data space system exploits this knowledge to streamline the content of the different databases and to combine data from the different databases that pertain to the same entity. Another module of the data space system is a variable database, which contains a large set of variables. A variable is defined as a meaningful notion to the end user, for example the elapsed time of a case. Several attributes are required to compute elapsed time. Finally, the system is equipped with a web interface, which include capabilities to facilitate the interpretation of the data in a structured way. The web interface facilitates officials of the government and related partners in monitoring the functioning of the entire Dutch criminal justice domain.

The remainder of this paper is organized as follows: Section 2 provides some background knowledge with regard to the criminal justice system. Sections 3 is devoted to the design and implementation of our data space system, while Section 4 describes the usage of our system to support data/information management in the criminal justice chain. Finally, Section 5 concludes the paper.

2. BACKGROUND

In the criminal justice domain many organizations are involved, such as the police, the public prosecution and organizations involved in the execution of penalties, for example the correctional institute of the ministry of Security and Justice. Each organization has its own information system and its own definitions of variables [1]. As the output of one organization is the input of another organization in the justice domain, the data of the organizations are related to each other.

To get a good insight in the efficiency of the criminal justice domain, a good understanding of flows within as well as between organizations is needed. Figure 1 shows an example of a simplified schema of flows in a specific part of the criminal justice system. The schema shows the number of crimes registered by the police and how the police act on these crimes. After a suspect is found, the police have several ways to deal with a suspect, for instance by imposing a police fine. This means that the suspect has to pay a sum of money.

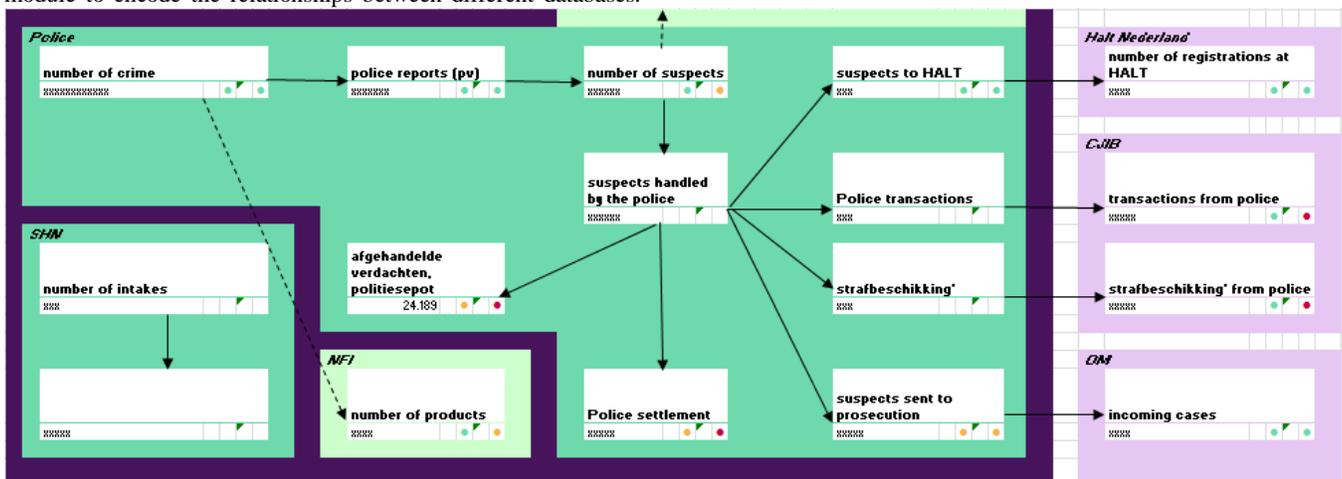


Figure 1. Schema of flows in a part of the criminal justice system. (Note that exact numbers are hidden, because it concerns non-public information.)

In case of a police transaction, the police send a notification to the organization responsible for the execution of transactions (‘CJIB’

in Dutch). The police may also decide to give a warning and to dismiss a suspect. In case of minors, the police may decide to send

a suspect to HALT. This bureau gives minors the opportunity to prevent getting a crime record. Also, the police may decide to send a suspect to the Public Prosecution ('OM' in Dutch). Note that from the police there are also connections to several other organizations, for instance to the National Forensic Institute ('NFI' in Dutch) and the bureau for victims ('SHN' in Dutch). So, the criminal justice domain is not necessarily a linear process, as there are different flows to and from several organizations possible.

There is a huge amount of data available from all the organizations together. To have a good understanding of how the criminal justice system acts, the government is not only interested in information of the organizations separately, but also in flows between organizations. Having information about flows within as well as between organizations may give the government insight in, for instance, workloads of organizations and the time needed to handle a case compared to the elapsed time.

As the criminal justice domain can be considered as a chain, one can imagine that what goes out from one organization, goes in to another. For instance, say the police send 100 suspects to the prosecution, then one may expect the input of the prosecution to be a 100 as well. Although this may seem logical, in practice this is not the case. In reality, out of these 100 suspects there may remain 50 cases in the input for the prosecution. This 'loss' of suspects may have several reasons. Firstly, there is a definition issue, namely, the police register suspects, while the prosecution registers cases. So the unit changes from person to cases. A person may have more than one case. Secondly, due to a new information system at the public prosecution, not all input from the police is registered automatically. The prosecution may calculate beforehand that a case will be dropped for reasons of feasibility and therefore decide not to register the input. So the output of the police does not necessarily mirrors the whole input at the prosecution. Moreover, in practice it seems there is also an unexplainable 'loss' of suspects [17]. So, there is no 'one to one' connection between organizations in the criminal justice domain.

As organizations have several means to handle cases the data reflects a hierarchy. For instance, if we consider the prosecution, then an officer has the possibility to handle a case in different ways, for instance by means of a transaction, a settlement or he can send the case to court. A transaction in turn can be a community service or the payment of a sum of money. Furthermore, a community service in turn can be on probation or not etc. (see Figure 2).

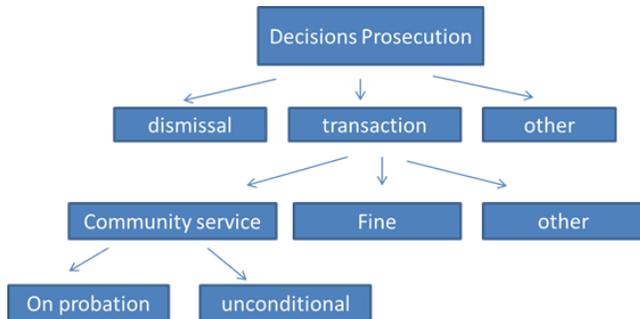


Figure 2. Example of the hierarchy of the data.

So there is a clear hierarchy in the data of the criminal justice domain we have to take into account. Presenting this hierarchy may facilitate the interpretation of the data flows.

3. DESIGN AND IMPLEMENTATION

For the development and implementation of the system we used the data space concept. In this concept, we distinguish three layers: a data space layer, a space manager and an interface layer [1-3, 18]. The data space layer consists of a set of databases. The space manager handles the information need of users and communicates the results. The interface layer consists of a set of user interfaces, which communicates on the one side with the space manager and on the other side with its users. In our system, we have elaborated on these three layers. In Figure 3, the overall architecture of the system is depicted.

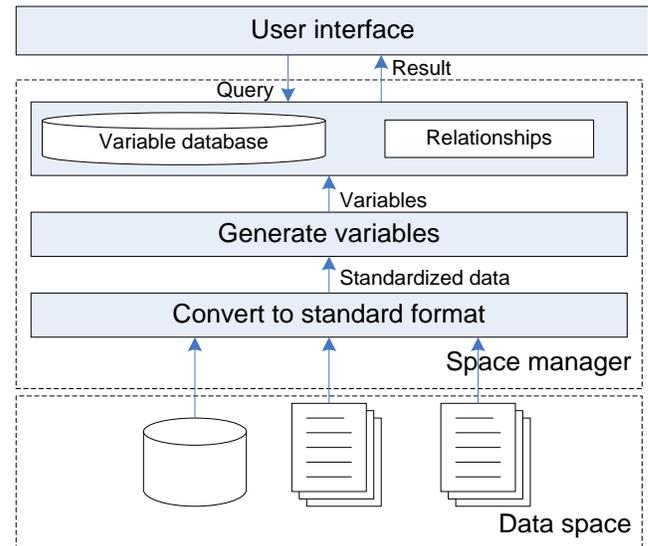


Figure 3. Architecture of the monitor.

The data space contains data sources of different types, ranging from comma-separated files containing aggregated data to a data warehouse with detailed information about a part of the justice system. The space manager consists of different components. An important component is the variable database, containing a large set of variables. A variable is defined as a meaningful notion to the end users, which is obtained by manipulation of a set of attributes. Attributes are fields/records that can be found in a database, and are stored in the data space. For example, the elapsed time of a case is a variable that is obtained by looking up the date that a case flowed in and the date that the case flowed out of the system. Subtracting the latter date from the former one results into the elapsed time of a case.

A second important component in the space manager layer is the relationship module. In the relationship module a set of rules is implemented to ensure the quality of the data. These rules are based on domain knowledge available from experts who are working in the field of justice. Most of the rules pertain to the handling of missing values, referred to as rules for imputation [3], and rules that establish relationships between different databases. Suppose that we need the date that a crime is committed but for some reason this date is not available or unreliable. We know in practice that the date the crime is committed, this crime is usually also reported at the same day. Therefore, we may use the latter date as substitute for the former one. Such a rule is included in the relationship module.

Besides the variable database and the relationship module, the space manager contains two auxiliary modules to create the variables ("Convert to standard format" and "Generate variables"). Data that are required to create a variable are extracted from the

data space layer and converted to a standard format. Then, the converted data are manipulated in order to create the value of a variable and passed to the variable database.

The user interface, third layer in this concept, communicates with the space manager and gets its variables directly from there. The interface layer is responsible for the presentation of the data to end users. It returns the information to the user visualized in the form of a table or, when requested, a graph. Since the user interface communicates with the data space layer through the space manager, users can generate the information needed by means of this interface without seeing the data space layer.

Our architecture has two major advantages. First, the architecture is flexible and extendable. New data sources can be included in the data space without changing existing data sources. Also new variables can be defined on the basis of the new data sources and passed to the interface layer. Whenever new relationships between databases are made explicit, they can be included in the relationship manager. Secondly, our architecture supports the independency of layers. The way that variables are presented to the end user may be changed without changing anything at the space manager and the data space layer. Also, we may change the definition of variables or including/ deleting/ updating rules in the relationship module at the space manager, while leaving the organization of the data space as it is.

In the next subsections we will focus on the challenges we faced by elaborating on the data space system. In subsections 3.1 and 3.2, we elaborate respectively on the relationship module and the variable database in the space manager layer. The user interface layer is described in section 3.3.

3.1 The Relationship Module: Knowledge Exploitation

Similar properties of a real-world event may be stored differently in the information systems of organizations involved in the criminal justice chain. For example, an event may be labeled by the police as a burglary with violence. However, the public prosecutor may come to the conclusion that violence will be hard to prove, and therefore the same event is labeled as a burglary. Also, chain partners may use their own definition for seemingly similar data, depending on their core processes and management goals. Furthermore, chain partners register their data at different points in time, namely at the time when their involvement is required, and also at different precision levels, since detailed information relevant to one chain partner may not be relevant to another. For example, courts register the number of sentences imposed while the executing organizations register the number of sentences executed. Another example is that courts register the type of crime as it is relevant for the sentence, but executing organizations may not record it as it has no relevance for the execution. Finally, we note that organizations at the beginning of the chain, such as the police, and at the end of the chain (the executing organizations) are primarily interested in data that pertain to individuals, while the remainder of the chain partners is rather interested in data about a case than an individual.

In our system we take advantage of above-mentioned properties of judicial data to relate (entities in) different databases. We have modeled the domain knowledge in two types of relationships. The first type deals with missing data. If the value of an attribute is (temporarily) unavailable, for example due to technical problems, it may be replaced with the value of a similar attribute that is measured at a slightly earlier or later point in time, but more or less covers the same notion. This is called imputation. The

imputed value may come from either the same or a different source. For instance, if the number of summonses is unknown, but we do know when the first court session took place, we may insert the value of the number of first sessions as an approximation for the number of summonses. In most cases these two attributes are highly correlated, because each summons eventually has to lead to a court session. Defining such relationships can help us to choose which data need to be selected when the preferred data are not available.

The second type of relationship deals with differences between seemingly similar but actually very different data. Different chain partners may label their data using similar or even the same names, but because these data are measured on different points in time and based on their own criteria, the outcomes for these data will vary. Knowing the relationship will help us compare the data correctly. This problem is generally known as a semantic problem, and is difficult to detect [6]. To illustrate the second type of relationship, consider the number of community services agreed with the public prosecutor. These data come from two chain partners, namely the public prosecutor and the execution organization. Thus, the relationship rule for these data is that the number of community services registered by the executing organization should be lower or equal than the number registered by the public prosecutor. For example, convicts may die before execution or cannot be found. Historical data confirm this rule.

3.2 The Variable Database

Variables are generated from attributes that are stored in the data space. To control the complexity of our system, we studied the scope of the variables, i.e. what data from which phases in the criminal justice chain are required for a variable. In the data space layer, the databases are conceptually mapped on the main phases of the criminal justice system. For example, information needs that pertain to investigation require attributes from the police databases, while information needs that pertain to trials require attributes from the prosecution and the courts.

We have organized variables in hierarchical trees. Variables are constructed dynamically by using naming conventions in the variable names. For example, the database holds the tree as depicted in Figure 4.

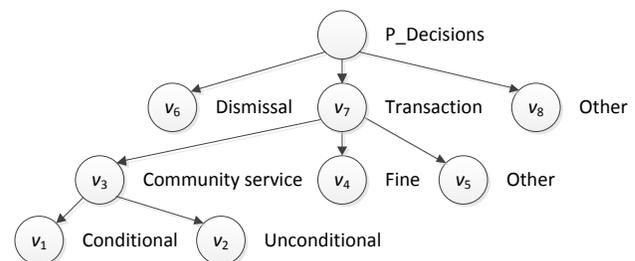


Figure 4. Tree of variables that pertain to prosecutor decisions.

The name of a variable is built up from different parts, distinguished by an underscore. Each part has its own meaning. So, the variable meaning can be constructed from its parts. The first and the second part of the variables (P_Decisions) refer to all variables that pertain to prosecutor decisions. The variable v_1 pertains to decisions that are conditional community service transactions (see also Figure 2), while v_2 also pertains to community service transactions, but without conditions. The total amount of community service transactions v_3 is defined by the sum of variables v_1 and v_2 . The remainder of the tree can be read analogously.

The variable trees are constructed bottom-up: the lowest (most specific) variables in the tree are generated directly from attributes, and upper variables are generated from lower variables.

The sum operator can not always be applied in a straightforward manner in a tree. This is the case whenever the attributes that are used to compute a variable are not disjunct. For instance, a transaction can have more than one ‘transaction type’; a community service for one part of the crime, and a fine for another part. As a result, the total number of transactions has to be extracted separately from the data space.

3.3 The User Interface

The third layer of the system is the user interface layer. As the system is meant for end users as policy makers, an important feature of the user interface was to present the information such that it facilitates policy makers to interpret the data flows in the criminal justice domain. To meet these goals we ran into several challenges. At first, as the criminal justice domain is quite complex given different flows and a lot of heterogeneous data, we faced the challenge to present the data as such that misinterpretations of the data flows are minimized. To prevent misinterpretation we created a fixed set of variables that reflects the hierarchy of the data. So end users do not have the possibility to create variables by means of mash ups themselves. From this fixed set of variables end users are able to drill down the data along several dimensions. These dimensions are region, sex, crime type and age.

To visualize the variables in a user friendly manner and facilitate the interpretations of data flows, the easiest way to get a first glance on the variables would be with tables and charts. Furthermore, the user interface needs to deal with regular updates in changes in requirements and variables. For now a choice has been made to show the data for the last 7 periods with the possibility to switch between twelve-month and three-month periods. The table is automatically refreshed when a new update is added.

We solved the visualization of the data and its hierarchy by creating a main results table where links indicate the existence of a next level of information. Clicking on them would get the user to a yet deeper level of the data, while the table shown would not change its visible structure. A breadcrumbs trail of links right above the table gives the user both an indication of the present level, and the opportunity to move back to the previous level.

To drill down on the desired dimensions, a separate selection panel was added to the interface, on the left side of the main results table. Filtering the initial results of the main results table on dimensions can be seen by selecting one or more items from the left dimension panel, and then by pushing the Select-button to initiate the filtering. After this selection, only figures are shown that correspond to the selected dimensions.

To show the user unexpected deviations in the data, a special feature was developed that is easily noticeable and traceable to its origins. While no clear definition of ‘unexpected deviations’ were available, we had to consult domain experts in order to build this alert function. In a few meetings with the organizations and government we discussed the norm for this alert feature. It became clear that the alert should not be too sensitive but sensitive enough to capture possible bottlenecks. These meetings with experts from the field and the government led to a decision to build the alert feature in the form of a traffic light, showing a green light with deviations of 5 percent and lower, orange between 5 and 10 percent, and a red light for deviations of 10

percent and more. The red and orange ‘traffic lights’ alert the user to unexpected fluctuations in trends. In the web interface, users have the ability to analyze the cause of an unexpected fluctuation by using the drill down functionality. Two types of alerts were developed, as the interest was to take into account seasonal effect or not. Figure 5 shows an example of these alerts. The first alert compares the last twelve-month period with the year before (2012 IV versus 2011 IV), while the second alert compares the last period with 3 months earlier (2012 IV versus 2012 III).

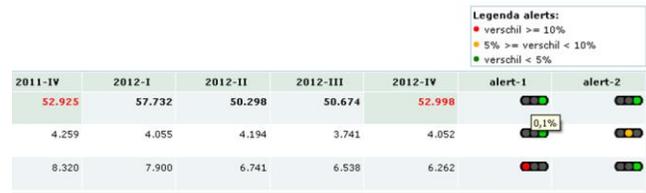


Figure 5. Example of the alert feature in the web interface.

To select data flows between or within organizations we added an option in the main navigation menu. Here, the user can select from a list of external interfaces (‘Externe koppelvlakken’ in Dutch) or internal interfaces (‘Interne koppelvlakken’ in Dutch), the latter being the list of data flows inside organizations and the former the flows between two organizations. After selection of one of these options, a table of results is shown with again the same layout and possibilities of seeing more details within the chosen selection.

3.4 Implementation

Currently, the data space consists of 21 data sources coming from 18 different organizations. The data sources are of different types, from comma-separated files containing aggregated data to a data warehouse with detailed information about a part of the justice system. The data from these data sources are stored in the data space layer. From the data space 682 attributes are extracted. Currently, the system contains more than 1300 variables in total and the number of variables is being extended at a regular basis.

The space manager layer is implemented on top of an Oracle DBMS. The layer contains two databases: an attribute database and a variable database. The attribute database is filled using a common ETL-process: data are extracted from the data sources in the data space, then transformed into standardized data and finally loaded in the attribute database. The generation of variables is a specific data space process, because the relationships between variables have to be taken into account during this process.

The implementation of the web interface is done on a Microsoft.Net platform, as this is a well-known and powerful technology for developing complicated websites that deal with high data traffic. For output purposes, the data are stored in a SQL Server database. This database contains the data from the variable database, and is optimized for web usage before building webpages with ASP.Net. For now only organizations who are connected to the internal web environment of Justice department will be granted access to the web interface, and only if they are authorized by the government.

4. THE DATA SPACE SYSTEM IN PRACTICE

In this section, we describe the use of the data space system in practice. In subsection 4.1 an example is given of how the system is currently used. Subsection 4.2 describes how we evaluate the system. The system is currently used to periodically monitor the

data flows within and between organizations. Every three months an update of the system takes place.

4.1 An Example

Figure 6 shows a screenshot of the web interface. In this screenshot the output of the Court in first instance is presented for seven time periods. The total number of output of the Court is presented in bold. This output is built from three other variables, from bottom to top respectively: “acquittals”, “convictions”, and “other verdicts”. Note that only this middle variable is clickable, and by clicking it the user asks the interface to descend in the variable hierarchy and show the child variables of this parent variable. Note that in this example only court region is available (the option for police region is disabled, as there is no data

available). Only the prosecution region can be selected (‘Arrondissement’ in Dutch), as only data of this region is available.

As the total output of the Court is green for the second alert, there is no reason to think there is a problem in the total output at the Court, but as the first alert is orange (deviation between 5 and 10%), policy makers may decide to drill down along several levels and dimensions to see whether there is a possible bottleneck somewhere. We stress the web interface serves as a tool to help a policy maker to visualize where there may be possible bottlenecks and facilitate the interpretation of the data. The actual interpretation of the data by experts is still necessary.

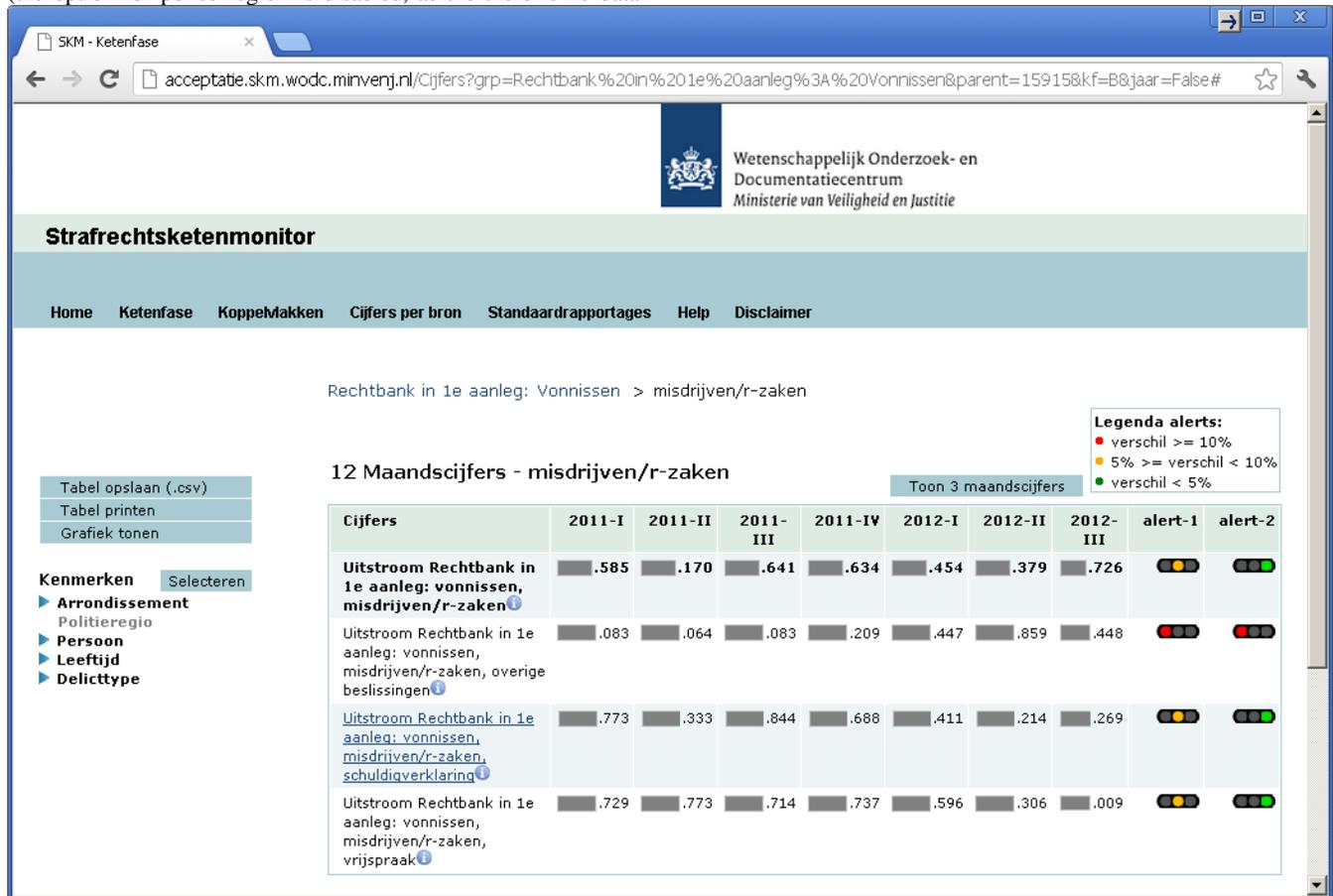


Figure 6. Screenshot from the web interface. (Note that exact numbers are hidden, because it concerns non-public information.)

Besides possibilities to view the information need at different levels, the web interface also has the possibility to print (standardized) PDF reports. These reports are the basis for periodically meetings with experts who interpret the developments. They contain a selection of the main variables of the web interface. In the reports no drill-down situations are available, but for some main variables a dimensional subset of the parent variable (‘split variables’) is presented. For instance, for the variable ‘total prison population’ the split variables ‘total prison population [males]’ and ‘total prison population [females]’ are presented as well. If during the meetings there is a need to drill down even further the web interface can be used.

Functionalities such as presentation of variables in graphs as well as printing options are also present. For new users there is a

manual available on how to use the web interface. For an extensive description to the web interface, we refer to [19].

4.2 Evaluation of the Data Space System

To evaluate the usability and user-friendliness of the data space system, we closely interact with the end users. Therefore we organized a few sessions with the end users. From the feedback we learned that the end users were quite content with the result of the first release of the system, however new wishes on usability and accessibility soon came up.

As the usability concerns, we learned from the end users that the web interface works as an adequate tool to help end users to indicate possible bottlenecks in the criminal justice chain. However, we stress the interface is not a tool on its own. Getting insight in the criminal justice domain is one thing, for the

interpretation of the data flows experts from the criminal justice domain are still needed given their knowledge and expertise. The web interface serves to facilitate the interpretation of data flows by means of a fixed set of variables that reflects the data flows and its hierarchy and minimizes misinterpretation, and also visualizes possible bottlenecks by means of traffic lights. The actual interpretation of the data flows is done in so called Early Warning meetings where experts in the criminal justice domain and the government take place. The web interface is used as a tool to visualize possible bottlenecks in the chain in these meetings.

As the accessibility concerns, new wishes soon came up, for instance, one wish was to narrow down the crime type dimension to a few main categories, as this dimension existed of over 30 different crime types. By concentrating this dimension, users can focus on a main category and drill down from this category, instead of scrolling through the whole list of crime types.

As the new wishes are on the usability as well as accessibility and new information need, we stress that different developers and data experts from the justice domain closely interact. Our experience is that all these experts together should be intensively involved. Experts and developers can indicate what is possible and technically feasible. Given the wishes of the end users, the developers and researchers of the data space system decide where adaptations or extensions in the system should be made, e.g. at the interface layer or/and at the space manager layer or data space layer.

Meanwhile, the system has been updated a couple of times after each past three-month period. After each update a user group meets in person to give feedback on the system. Questions as to what works and what not etc. are central in these meetings. With this feedback we are able to evaluate our system and implement new wishes.

5. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

To generate management information in the field of justice, data from different heterogeneous sources has to be extracted and processed. Given the complexity entailed in understanding and interpreting these data, the processing occurs in a (semi-)manual manner. At best, this is a tedious and error-prone process. Therefore, we have designed and implemented a data space system to improve the generation of management information. The system provides insight into the success and possible bottlenecks in the criminal justice domain. Our system makes it possible to periodically monitor data flows within as well as between organizations in the judicial domain. The system exploits domain knowledge, which is encoded in the system in the form of relationships and variables. We distinguish two types of relationships. The first type deals with imputation issues, i.e. missing values for attributes. If the value of an attribute is (temporarily) unavailable, for example due to technical problems, it may be replaced with the value of a similar attribute that is measured at a slightly earlier or later point in time, but more or less covers the same notion. The second type of relationship deals with differences between seemingly similar but actually very different data. Different chain partners may label the same entities differently or different entities get the same label. This is due to the different classification systems that organizations use and because data are measured on different points in time. Knowing the relationship between the labels and the data help for an adequate interpretation of the data. Variables are defined such that they are meaningful notions to the end user and their values are

computed from a set of attributes. These attributes are coming from different databases in the data space layer.

Another important feature of the data space system is the interface. The interface is a user friendly tool that facilitates policy makers to interpret the data flows in the justice domain. To prevent the misinterpretation of data we created a fixed set of variables that may be accessed by policy makers. Creation of new variables should be done in consultation with our institute.

The system proves to be successful in managing data from different information systems that are involved in the criminal justice chain. Currently, the system is used by several chain partners as well as by policy makers at the Dutch Ministry of Security and Justice to generate management information about the criminal justice chain. Although an effect evaluation of the system might be desired, we note that setting up such an evaluation will be a hard task. This is due to the fact that we introduce the system gradually at the Ministry and there is no "ex ante" description before the system. Furthermore no metrics are defined that can be used to compare the situation before and after the introduction of the data space system.

To our best knowledge, SKM is the first implementation of the data space approach targeting the criminal justice chain. Topics for further research are the customization of the interface to the individual needs of the users, based on their roles, and the generalization of the data space concept by applying it in other parts of the judicial domain. Also the application of our approach to different contexts, such as Open Data, is a topic for further research.

6. REFERENCES

- [1] Kalidien, S.N., Choenni, R. and Meijer, R. 2009. Towards a Monitoring Tool for Crime and Law Enforcement. In: *Proceedings of ECIME 2009, 3rd European Conf. on Information Management and Evaluation*. Gothenburg, Sweden, September 17-18, Academic Publishing Limited, UK.
- [2] Kalidien, S.N., Choenni, R. and Meijer, R.F. 2010. Crime statistics online: potentials and challenges. In: *Proceedings of 11th Annual International Digital Government Research Conference on Public Administration Online: Challenges and Opportunities (dg.o '10)*, pp. 131-137. Digital Government Society of North America.
- [3] Choenni, R., Kalidien, S.N., Ariel, A. and Moolenaar, D.E.G. 2011. A framework to monitor public safety based on a data space approach. In: *Electronic Government and Electronic Participation, Joint Proceedings of Ongoing Research and Projects, EGOV and ePart 2011*. Delft, the Netherlands (pp. 196-202). Linz: Trauner Verlag. Schriftenreihe Informatik, 37.
- [4] Kalidien, S.N. and de Heer-de Lange, N.E (eds.). 2013. In Dutch: 'Criminaliteit en rechtshandhaving 2012'. Justitie in statistiek, nr. 2. The Hague, Boom Juridische uitgevers, in press.
- [5] Rutte II. 2012. Coalition Agreement: Building Bridges. Dutch government, October 2012. The Hague.
- [6] Graves, A. 2011. A case study for integrating public safety data using semantic technologies. In: *Information Polity 16(3)*, pp. 261-275. IOS Press.
- [7] Mena, J. 2003. Investigative Data mining for Security and Criminal Detection. Elsevier Science, USA.

- [8] Choenni, S. and Leertouwer, E. 2010. Public safety mashups to support policy makers. In *Electronic Government and the Information Systems Perspective* (pp. 234-248). Springer Berlin Heidelberg.
- [9] Choenni, S., van Dijk, J. and Leeuw, F. 2010. Preserving privacy whilst integrating data: Applied to criminal justice. *Information Polity*, 15(1), 125-138.
- [10] Sheth, A.P. and Larson, J.A. 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys*, vol. 22, no. 3, pp. 183-236. ACM New York, NY, USA.
- [11] Cabibbo, L. and Torlone, R. 2005. Integrating heterogeneous multidimensional databases. In: *Proc. of 17th Int. Conference on Scientific and Statistical Database Management (SSDBM'05)*, pp. 205-214.
- [12] Winkler, W.E. 1995. Matching and record linkage. In *Business Survey Methods*, B.G. Cox, D.A. Binder and B.N. Chinnappa, eds, Wiley, New York, NY, pp. 355-384.
- [13] Wilkins, D. and Pillaipakkamatt, K. 1997. The effectiveness of machine learning techniques for predicting time to case disposition. In: *6th International Conference on Artificial intelligence and Law*, pp. 106-113. ACM New York, NY, USA.
- [14] Van den Braak, S., Choenni, S. and Verwer, S. 2013. Combining and Analyzing Judicial Databases. In: Custers, B., Calders, T., Schermer, B., Zarsky, T. (eds.) *Discrimination and Privacy in the Information Society. SAPERE 3*, pp. 191-206. Springer, Heidelberg.
- [15] Elmasri, R. and Navathe, S. 2004. *Fundamentals of Database Systems*, 4th edition. Addison-Wesley.
- [16] Dijk, J.J. van, Choenni, S., Leertouwer, E., Spruit, M. and Brinkkemper, S. 2013. A Data Space System for the Criminal Justice Chain. In: *12th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE)*, in press.
- [17] Court of Audit. 2012. In Dutch: 'Prestaties in de strafrechtsketen'. The Hague.
- [18] Franklin, M., Halevy, A. and Maier, D. 2005. From databases to dataspace, a new abstraction for information management. *SIGMOD Record*, vol. 34, issue 4, pp. 27-33.
- [19] Kalidien, S.N., Witzenburg, R. van and Choenni, S. 2013. Developing, Implementing and Evaluating a Web Interface in the field of e-government. In: *Evaluating Websites and Web Services: Interdisciplinary Perspectives on User Satisfaction*, D. Yannacopoulos, P. Manolitzas, N. Matsatsinis, E. Grigoroudis, Ed. IGI Global, 2013, in press.