Distributed Digital Data: Keeping files consistent and timely

Martijn Warnier, Frances Brazier, Martin Apistola, Anja Oskamp
Computer Science Department & Computer Law Institute
VU University Amsterdam
De Boelelaan 1081a
1081HV, Amsterdam, the Netherlands
warnier@cs.vu.nl, frances@cs.vu.nl, m.apistola@rechten.vu.nl, a.oskamp@rechten.vu.nl

Abstract

Although digitalization has made exchange of information, between governmental organizations easier, it comes at a cost. Governmental organizations need to explicitly define and implement security requirements, policies and mechanisms with which information access, processing and indexing is regulated: determining which organizations are responsible for which information, and how information, once exchanged, is kept up to date. This paper proposes the use of a Distributed Digital Dossier in combination with agent technology for this purpose. The domain of criminal dossiers prepared by the Public Prosecution and used in Courts of Law illustrates the approach.

1. Introduction

Information in large governmental organizations is inherently distributed across different physical locations and systems. As much of this information is digital, distribution of information has become easier. Distribution, however, requires governmental organizations to define, implement and enforce policies for distributed information management, stating which information can be accessed, processed, indexed and synchronized by which other organizations, when, where and how. Ensuring that the information, once distributed to other organizations, remains up to date is a real challenge. Staleness of information can have serious consequences in most organizations: especially when organizations act on the basis of information that is not only stale, but is no longer valid. Each individual organisation needs to specify the rights of regulations for all entities involved both within and between organizations.

This paper introduces the notion of a *distributed digital dossier*. Responsibility for data is delegated to authorities providing information. The digital dossier currently being evaluated in a pilot study by the Courts of Amsterdam and Rotterdam in the Netherlands [4] is an example of a dossier that contains information from multiple sources, and for which security and timeliness of data are major challenges. The Public Prosecution is responsible for dossier compilation and management – the dossiers are criminal files. Information is acquired from governmental organizations such as the Police, Prison Authorities, Municipalities, and Probation Officers. Administrative data such as name, address and marital status, for example, is provided by the Municipalities, the authority responsible for registration of this information in the Netherlands. The Municipalities have well-regulated policies with respect to information provision. Their security policies specify which information is accessible to whom and when and which information is not.

Ideally, if information changes during the course of file compilation, this will be detected and the dossier modified. Correctly propagating new information across different organizations is, however, often complex. For this reason, in practice the Public Prosecution does not always detect changes during file compilation, and -by law- cannot change data between finalization of dossier compilation and a court case. A distributed digital dossier provides the means to propagate changes in data at the source (the responsible organization) to copies of other files in which this information has been included maintained by other organizations.

A conceptual overview of the distributed digital dossier is detailed in Section 2. Section 3 provides an example of a distributed criminal file for a specific class of crimes – those committed by juvenile repeat offender. Section 4 uses this example to illustrate the role of security policies and infrastructural support. The paper ends with a discussion and conclusions.

2. Distributed Digital Dossiers

Information in a dossier comes from many different sources: the combined systems of the Courts, the Public Prosecution, the Police and all other parties together form a *semi-open system*. Not quite an open system, as the information is not publicly accessible nor a completely closed system, as several (governmental) organizations exchange information. An *open system* is a computer system that is configured to allow access to outside parties. In contrast, in a *closed system* only known users are allowed access (after authorization) to parts of a computer system. The systems used to compile a digital dossier, are neither closed nor open: the organizations involved are known, but are not allowed to access each others' systems/databases. Each individual organization is responsible for its own systems, and information provision. Requests are honored, systems trusted. Sometimes information from outside sources is used.

Each organization's computer systems can be seen as a closed systems. Only police officers can access the computer system used by the Police and only employees of the Public Prosecution can access (parts of) dossiers at the Public Prosecution. However, as all of these systems (can) exchange information, the chain of systems is no longer a closed system. The characteristics of each site can vary considerably: they can have different procedures for access, reliability and/or security. For example, a Municipal Database that contains name and address information has other goals (and hence system characteristic) than a computer system used by the Council for Child Welfare that stores information on children. Moreover, at several places in the system there are outgoing connections to public (open) computer systems, e.g. reference systems that give information on the current state of the Law. Thus the system as a whole constitutes an example of a semi-open system.

A distributed digital dossier [4] contains information from physically distributed sources, distributed across organizational boundaries, see Figure 1. Initialisation of a digital dossier is the responsibility of the Public Prosecution on the basis of information provided by the Police and other organizations. The initial dossier specifies which information is to be included on the basis of the crime committed. It also specifies an access control list (who may read and alter which parts of the dossier), dependencies between different fields in a document, the organisations responsible for the fields. This part of the dossier is stored centrally by the Public Prosecution, in their own database as meta data. The content of different fields, represented as records, are the responsibility of the organizations specified in the central document: personal information is maintained by the municipal databases, family related information for juveniles is maintained by Council for Child Welfare etc. The dossier stored by the Public Prosecution only contains references to this information. The information itself is controlled by the source (from the perspective of the Public Prosecution). If something changes at the source, this is flagged by the responsible organization, and the Public Prosecution is informed. The Public Prosecution then, on the basis of the meta-data on dependencies between fields, checks to make sure the data within the dossier is still consistent.

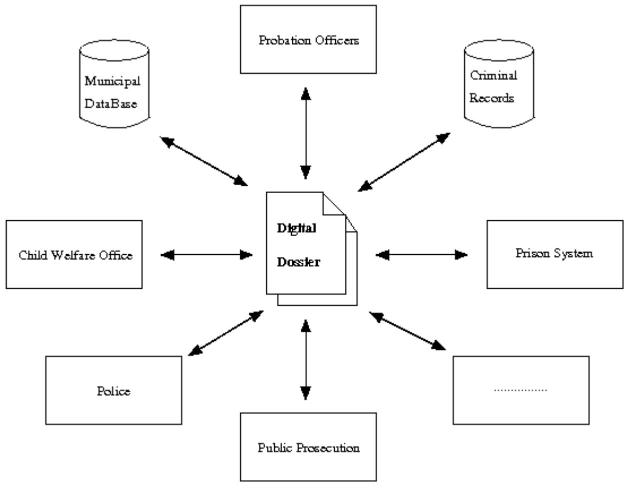

Figure 1. The Digital Dossier depends on information from numerous distributed sources.

If new information is propagated, this approach guarantees that information stays as up-to-date as possible during dossier compilation. The distributed character of the dossier, however, vanishes, by necessity, once the dossier is sent to Court. The Public Prosecutor's task is namely to compile a dossier. Once a dossier is finalized, and sent to Court and the defendant's lawyer the dossier is static. Only via a special procedure and with the judge's permission, can information be added to the dossier. This ensures that all parties have the same dossier. Note that a finalized (or frozen [4]) dossier is a local copy of the distributed file at that given point in time.

The next section presents an example of a case and some snippets of the corresponding dossier. It concerns a case involving juvenile repeat offenders. This type of case has been chosen because of the number and nature of organizations involved.

3. An example criminal prosecution chain: Juvenile Repeat Offenders

An example of a criminal prosecution chain for a juvenile repeat offender is illustrated below. This example has been previously used in [5] and is constructed from an actual case. All personal information has been anonymized and numerous details have been omitted. It illustrate the complexities involved in such cases. Note that this scenario describes the current Dutch situation, legal constraints are also analyzed in a Dutch legal setting, according to Dutch law.

The criminal prosecution chain starts when the Police arrest a juvenile suspect for vandalism. The suspect is escorted to the Police station where an assistant prosecutor questions the suspect. The Police open a new dossier specifically for this case. This dossier contains a summary of the offence for which the suspect is charged, the date and location of the incident, number of suspects, personal data of the victim, the official police report, and other relevant information. The suspect then becomes the subject of investigation: the personal data he/she has provided is cross referenced with the municipal database[1]. The Police also queries local Repeat Offender Databases to discover whether this suspect is a known repeat offender. As this case concerns a minor, a request is issued to other organizations for juvenile offenders, to provide relevant information about the minor's background. All of this information is added to the Police report. After collecting this information, the Police and the Assistant Prosecutor inform the Public Prosecutor of the case and transfer the report to the Public Prosecution.

The Public Prosecutor decides whether to press charges or, to pursue an alternative if other (minor) punishment is deemed more suitable. This decision is based both on the specific details of the current case and the (criminal) history of the suspect. A dedicated Judicial Documentation Database is used to retrieve information on the criminal past of the suspect. Typically, at this point, the Public Prosecutor will again consult municipal databases and local Juvenile Repeat Offender Systems. All information is cross referenced with the case dossier and information is updated when needed. If the Public Prosecutor decides to bring the case to court, as is the case, the Public Prosecution's distributed digital dossier is created with this information. The next mandatory step involves informing the Council for Child Welfare.

In the Dutch context the Council for Child Welfare has the task to investigate all crimes of minors. In addition to the criminal offences of the minor, the family situation and other relevant social factors are taken into account. This results in a motivated advice for suitable punishment of the suspect. This advice is added to the dossier. The prosecutor then serves a summons and a lawyer is assigned to the juvenile suspect. Adding the summons to the dossier finalizes the dossier at this point. A copy of the dossier is sent to the Court and to the lawyer of the suspect. At the court session the information in the dossier is used by all parties involved. The Public Prosecutor demands a suitable sentence, the lawyer presents the defence and ultimately the judge comes to a verdict. The suspect is sentenced and all information regarding the court session is added to the dossier. The dossier itself is filed in Judicial Documentation Database for future reference.

4. System Architecture

Distributed digital dossiers can be implemented in several ways. A distributed database [2] is an obvious option. As the digital dossier, however, not only needs to store information, but to actively monitor all fields continuously, to autonomously act when changes occur, across different organisations each with their own security policies, to guarantee up-to-dateness, consistency, completeness and security of the data a slightly different approach has been chosen: an agent-based approach. Agents provide the means for decentralised autonomous monitoring and processing of data,

[1] In the Dutch setting municipals are obliged by Law to maintain databases with administrative information on citizens. This information is then used by other governmental organizations, such as tax organizations and the police, to check the correctness of their own records.

coordination mechanisms, and interaction support. A light weight `skeleton' framework using XML for the dossiers themselves in combination with agent technology [1] ensures the more complex and dynamic properties such as completeness etc.

4.1. Digital Dossier Skeleton

As stated above the dossier created by the Public Prosecution specifies which information is to be included and its status (mandatory, optional), which dependencies between information exist, which organisations are responsible for which information, which access rights are assigned to whom/which organisations, etc. This meta-data depends on the crime committed. This paper assumes that standard XML templates exist for each class of problems, and that these templates are used by the Public Prosecution to structure the meta-data in a dossier. Figure 2 depicts an example of part of a dossier with both meta-data and data provided by the Public Prosecution. Meta-data includes the dossier number, creation date, type of offence and access control lists. The example depicted in Figure 2 specifies that in this case the dossier has been opened by the Public Prosecution of Amsterdam on the first of July 2007. The offence committed is vandalism, and the suspect is a juvenile repeat suspect. This dossier also specifies that four named employees of the Public Prosecution have been assigned permission to read this document (identified in this example by numbers `1234, 2345, 3456, 4567`) and that only two of these four employees have permission to edit parts of the dossier (i.e. write permission).

```
<Dossier>
  <MetaData>
    <Ref>PubProsAmsterdam-00001</Ref>
    <CreationDate>1-7-2007</CreationDate>
   <Offence>
    <Main>Vandelism</Main>
    <Category>JuvenileRepeatOffender</Category>
   </Offence>
    <Access>
       <Read>1234,2345,3456,4567</Read>
       <Write>1234,4567</Write>
    </Access>
    …
  </MetaData>
  <Records>
    <MandatoryInfo>
      <PersonalInfo>ref:MuniDatAmsterdam-123456</PersonalInfo>
      <ReportCCW>ref:CCW-234567</ReportCCW>
      …
    </MandatoryInfo>
    <OptionalInfo>
      <StatementSupect>ref:PP-00001statement.pdf</StatementSuspect>
      <Media>
            ref:PP-00001-movie1.avi,
            ref:PP-00001-movie2.avi
      </Media>
      …
    </OptionalInfo>
  </Records>
</Dossier>
```

Figure 2. Example Digital Dossier structure.

This template distinguishes between mandatory and optional information for a specific crime. This information represents the knowledge used by the Public Prosecution to check completeness of a dossier - and provides the structure needed for automated completeness checks [5]. Certain information must always be included in a dossier, such as the suspect's personal data, and the original police report concerning an incident. Other information will only be mandatory for certain types of offences and/or suspects. In this particular example, independent of the crime for which the suspect has been accused, the suspect is a juvenile: according to Dutch law, a report by the Council for Child Welfare is mandatory. Other information, however, is optional. In this example, footage of the incident on the basis of which the suspect has been charged, recorded by a video surveillance camera, may be included. Note that a court case can only commence if all mandatory information is included in a dossier. The document also contains a number of references, indicated by the '`ref:`' keyword, to other documents and their source. In this example references local to the Public Prosecution (PP) are the suspect's original statement (`ref:PP-00001statement.pdf`) and the video evidence (`ref:PP-00001-movie1.avi` and `ref:PP-00001-movie2.avi`). References to XML documents to be provided by other organizations, in this example, include the (national) Council for Child Welfare (`ref:CCW-234567`) and the Municipal database of Amsterdam (`ref:MuniDatAmsterdam-123456`). Figure 3 depicts the data to which the ref `ref:MuniDatAmsterdam-123456` refers, the suspect's personal data as stored and maintained by the Municipality – in this case the Municipality of Amsterdam. For the sake of simplicity, this information is assumed to be stored in the same format as the XML document stored by the Public Prosecution. (Note also that if data is not provided in this format, wrapper agents can perform the necessary mappings).

The data depicted in Figure 3 states that this record may be read by *all* Dutch governmental organizations, but that only one employee, the employee of the Municipality of Amsterdam with identifier `09876`, can alter the file. This is a policy that is controlled by the local organisation (the Municipality of Amsterdam). Note that although organizations are independent entities, they do not always have complete freedom in setting their own security policies. In certain cases they are legally obliged (by Dutch Law) to provide information to specific authorities on request. In other cases Dutch Law forbids organisations to give other organizations access to their databases, e.g. due to privacy regulations. The ability to add/modify information is typically controlled by the organizations themselves. Note also that the information provided by the Municipality in the document depicted in Figure 3 again makes the distinction between mandatory and optional information without more detail. It could, although this is not shown in Figure 3, also have contained references to other documents, stored either locally or at another organization.

```
<PersonalInfo>
  <MetaData>
    <Ref>MuniDataAmsterdam-123456</Ref>
    <CreationDate>05-08-1993</CreationDate>
    <Access>
       <Read>All</Read>
       <Write>09876</Write>
    </Access>
    …
  </MetaData>
  <Records>
    <MandatoryInfo>
      <Name>
        <Surname>Jan</Surname>
        <FamilyName>Jansen</FamilyName>
      </Name>
      <DateOfBirth>02-08-1993</DateOfBirth>
      <Sex>Male</Sex>
      <Adress>…</Adress>
      …
    </MandatoryInfo>
    <OptionalInfo></OptionalInfo>
  </Records>
</Dossier>
```

Figure 3. Example of personal information stored at the municipal database.

A distributed digital dossier thus consists of a number of documents in a networked structure distributed over physically distributed locations with a central coordinator. Note that digital dossiers always form the root of this network, i.e. digital dossiers are not included in (referenced from) other documents, unless the referring document is also a digital dossier. This ensures that digital dossiers themselves (e.g. as shown in Figure 2) are always controlled by the Public Prosecution, an important security requirement. The next section discusses the infrastructure used to link these documents together.

4.2. Multi Agent Systems

As described above, a distributed digital dossier is, in fact, a distributed XML document. The Public Prosecution maintains the defining document, based on a crime specific template. Multiple other organisations maintain the data in the records for which they are responsible. Each of these organisations is also responsible for its own security policies. An infrastructure is needed to link the organisations together, making information exchange possible.

The *multi-agent paradigm* provides a conceptual modelling framework for distributed systems. Agents are pro-active, autonomous, possibly mobile, systems that can interact with other systems, for example other agents or (web) services, and can adapt to a changing world [6]. Specific tasks within the paradigm can be implemented by dedicated agents, allowing for a clear separation of concerns and straightforward integration of new functionality as new agents.

Agent technology provides a means to implement large scale secure distributed autonomous systems [1]. Agents interact through message passing. Agents run on agent platforms. An agent platform is a dedicated middleware layer that provides the infrastructure, such as secure communication, resource management, mobility[2] and access control, for software agents. The agent platform middleware also ensures that multiple hosts within a location can be viewed as one logical unit, each with their own security policies.

The agent based architecture designed for the distributed digital dossier assumes that each that the agent locations are the organisations involved. Thus the Public Prosecution, the Municipal databases and the Council for Child Welfare (and all other organizations) are locations, each running their own agent platform middleware supporting their own organisational policies. Figure 4 shows schematically how the agent platforms are deployed:

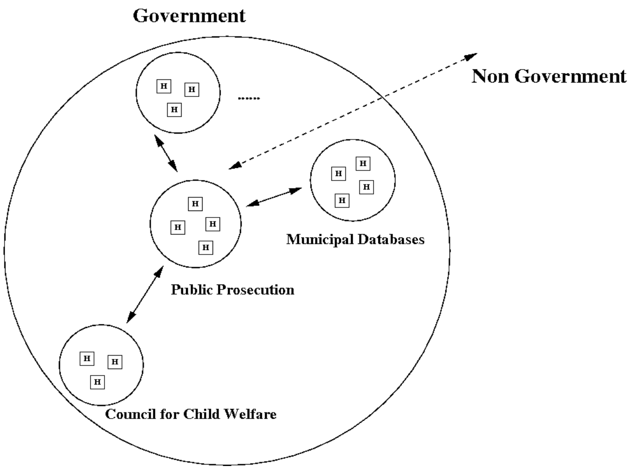

Figure 4. Schematic overview of agent middleware deployment

[2] Not all agent systems allow migration of agents between hosts.

The agent platform middleware provides a uniform software layer across locations. Each location represents a separate organization. Access to the digital dossier is also facilitated and controlled by this layer. Dedicated agents per organisation access their parts of the dossier and keep information up-to-date. The Public Prosecution has dedicated agents for more complicated tasks such as completeness and consistency checking. With regards to consistency checking, dedicated software agents perform such tasks as guarding consistency both within a source used in an organization and sources used across organizations within the digital dossier, and notifying appropriate (human) parties when needed. With respect to completeness dedicated agents monitor the availability of mandatory documents in the digital dossier. For instance, a trial cannot start if a copy of the original police report is not in the digital dossier. Software agents guard such completeness issues [5].

In a future design the mapping of a specific ontology used within one organization to the shared ontology of the digital dossier will be done automatically. However, this is not the main focus of this paper and for now it is assumed that all (governmental) organizations share one common ontology.

4.3. Guarding Timeliness

The design of the distributed digital dossier ensures that information will stay as up to date as possible. Agents, however, also perform many other tasks such as checking consistency or completeness, but also enforcing access control or handling automatic backups. An example of a functional task for which a dedicated agent provides a transparent implementation is that of guarding expiration dates. Specific to the domain of a Court of Law is the example of expiration of criminal files for juvenile offenders after they reach the age of 18. By Dutch law information provided by the Council for Child Welfare may not be used by the Public Prosecution (or any other organisation for that matter) after the suspect turns 18, or after a period of 5 years, if no other offence has been committed.

Dedicated agents at the Public Prosecution guard the timeliness of information in the dossier. In the example of the juvenile repeat offender, the dedicated agent can remove (independently) the *reference* to the information stored at the Council of Child Welfare without human intervention, once the suspect turns 18 if all other constraints are fulfilled. The Public Prosecutor of the case and the Council for Child Welfare are notified that the information in the dossier can no longer be used. This ensures that the dossier stays up to date (does not contain information that can no longer be used), without actually deleting information (as only the reference is removed). The Council for Child Welfare can have its own (security) policy that states what needs to be done once information can no longer be used. This may well differ from the policies implemented by the Public Prosecution

4.4. Security Policy

Semi-open systems are challenging for numerous reasons. The security policies in such a system cannot be too restrictive, because they also contain public information. Yet it is clear that the privacy sensitive information concerning criminal records and other personal information needs to be guarded well [3]. Privacy sensitive data needs to be well protected against malicious intent and at the same time less sensitive information needs to be made available to a large public

The sensitive nature of the information in a dossier makes security very important. Ensuring that only authorized persons are allow to read/write/alter the dossier is clearly important, as are other security issues such as integrity of data, transfer, accountability, privacy, reliability and persistence. *Security policies* explicitly state the restrictions posed on a system in order to make it secure. The security architecture proposed for the distributed digital dossier is based on the design in which (1) agents control all access to the digital dossier and (2) security policies are defined for each organization (each organization can use its own security policies) to regulate all other aspects.

Under normal circumstances[3], all interactions with a dossier are performed via agents. Agents are bound to specific users/employees of the organisations involved. Thus, individual users have their own agent(s) to act on their behalf. A highly ranked legal clerk employed by the Public Prosecution, for example, may have an agent that has the right to change all information maintained by the Public Prosecution in the central digital dossier. Another clerk, employed by the Municipality of Amsterdam, may have an agent whose only task is to flag and communicate modifications to a suspect's data to the dossier's consistency agent at the Public Prosecution. This use of agents for access to the dossier is implicit, users are not continually aware of their agent's state. They interact with their agents through standard interfaces and perform the tasks they have been assigned. The access control token of a user, such as the legal clerk, is a

[3] System administrators can access the system manually, when necessary. Sometimes this is required, for example, when a backup copy needs to be replaced, or an operating system has to be upgraded.

combination of a password and public key (X509) certificate. The agent uses the password and certificate to authenticate itself to the system, thereby also authenticating the user on whose behalf the agent acts. Agents can also be linked to organizations. The Public Prosecution can thus use its own agents. As but users and organizations can be linked to agents, it is clear who is accountable for mistakes and/or malicious behaviour. Agents are bound to a legal entity by signing the agents' code with the private key that corresponds with the users' public key certificates.

The other main security attribute is that each organization uses its own security policy. Security policies state rules for all security related issues. Some of these can be global, that is, for all organizations that provide information to the digital dossier. An example is the authentication described above: all users of the system have to use a combination of a password and certificate to authenticate themselves. Another example is the use of one shared lookup service that can be used to find individual agents within an agent system. However, most rules in the security policy are local: the local organization controls who has access to its files, but also which backup procedures is used, what the privacy policy regarding its data is, etc. This ensures that if the computer system of one organization is comprised the other organizations in the semi-open system are not at risk. In particular, as long as the computer system of the Public Prosecution is not compromised, the digital dossier can still be used. The computer systems of the Public Prosecution need to be trusted completely (for 100%). The lookup service, PKI service and authentication mechanism are all hosted by the Public Prosecution. If other systems are compromised the digital dossier is not necessarily effected. Automatic consistency checking, performed each time part of a dossier is altered, detects modifications. If these modifications are unwarranted, the original data may be restored. See [4] for a more detailed discussion of the security architecture for the distributed digital dossier.

5. Discussion

Distributed digital dossiers have clear benefits for keeping data up to date and consistent across multiple organizations organized in semi-open environments. Organizations determine their own (local) security policies while still exchanging information in a transparent and efficient way. The added advantages of using agent systems in this setting include the option to transparently implement complex functional tasks such as monitoring, consistency, completeness and security with dedicated agents per organisation.

Current work includes the implementation of a prototype system using the Public Prosecution's simulation environment, implementing different security policies for different organisations. From a legal perspective research focuses on analysis of legal constraints with respect to integration of the individual systems as determined by (Dutch) law. More research is needed to answer these and other questions.

6. Acknowledgements

This research is supported by the NLnet Foundation, http://www.nlnet.nl, and is conducted as part of the ACCESS project, http://www.iids.org/access funded by the NWO TOKEN program, grant number 634.000.431.

7. References

[1] Luck, M. and McBurney, P. and Preist, C., *Agent Technology: Enabling Next Generation Computing (A Roadmap for Agent Based Computing)*, AgentLink, 2003.

[2] Ozsu, M.T. and Valduriez, P., *Principles of distributed database systems*, Prentice-Hall, 1991.

[3] Stone, E. and Stone, D. "Privacy in organizations: Theoretical issues, research findings, and protection mechanisms", *Research in Personnel and Human Resources Management,* 8(3):349--411, 1990.

[4] Warnier M., Brazier F.M.T., Apistola M., and Oskamp A., "Secure Distributed Dossier Management in the Legal Domain", in *Proceedings of the 2nd IEEE International Workshop Dependability and Security in e-Government (DeSeGov 2007)*, Vienna, Austria, April 10-13 2007.

[5] Warnier M., Brazier F.M.T., Apistola M., and Oskamp A., "Towards Automatic Identification of Completeness and Consistency in Digital Dossiers", *in Proceedings of the 11th ACM International Conference on Artificial Intelligence and Law (ICAIL'07)*, Palo Alto, CA, June 4-8 2007.

[6] Wooldridge, M. and Jennings, N.R, "Intelligent Agents: Theory and Practice", *The Knowledge Engineering Review*, 10(2), p 115-152, 1995.