ELSEVIER

Artificial Intelligence in Medicine 26 (2002) 25–36

Artificial Intelligence in Medicine

www.elsevier.com/locate/artmed

Confidentiality issues for medical data miners

Jules J. Berman*

Pathology Informatics Cancer Diagnosis Program, DCTD, NCI, NIH, EPN-Room 6028, 6130 Executive Bluilding, Rockville, MD 20892, USA

Received 5 March 2002; accepted 11 March 2002

Abstract

The first task in any medical data mining effort is ensuring patient confidentiality. In the past, most data mining efforts ensured confidentiality by the dubious policy of witholding their raw data from colleagues and the public. A cursory review of medical informatics literature in the past decade reveals that much of what we have ''learned'' consists of assertions derived from confidential datasets unavailable for anyone's review. Without access to the original data, it is impossible to validate or improve upon a researcher's conclusions. Without access to research data, we are asked to accept findings as an act of faith, rather than as a scientific conclusion.

This special issue of *Artificial Intelligence in Medicine* is devoted to medical data mining. The medical data miner has an obligation to conduct valid research in a way that protects human subjects. Today, data miners have the technical tools to merge large data collections and to distribute queries over disparate databases. In order to include patient-related data in shared databases, data miners will need methods to anonymize and deidentify data. This article reviews the human subject risks associated with medical data mining. This article also describes some of the innovative computational remedies that will permit researchers to conduct research AND share their data without risk to patient or institution.

Keywords: Data mining; Confidentiality; Security; Encryption; HIPAA; IRB

1. Introduction: medical research risks

Patient risks imposed by medical research generally fall into one or more categories which are given in the following sections.

1.1. The risk to life and health as a direct result of a medical intervention

Two recent examples are the case of 18-year-old Jesse Gelsinger, who died of liver failure 17 September 1999, 4 days following after receiving a viral inoculation in a gene

* Tel.: +1-301-496-7147; fax: +1-301-402-7819.
E-mail address: bermanj@mail.nih.gov (J.J. Berman).

0933-3657/02/$ – see front matter
PII: S0933-3657(02)00050-7

transfer experiment conducted at the University of Pennsylvania [11]; and Ellen Roche, a 24-year-old healthy volunteer who died from lung failure on 2 June 2001, several weeks after inhaling hexamethonium as part of a Johns Hopkins asthma study [6]. Morbidity and mortality resulting from interventional medical research is the most serious of human subject risks, and special measures are taken to ensure that interventional studies are conducted with rigid standards. For interventional studies, patients must give written consent for participation in the experiment. The consent document must inform the patient of all the risks (adverse consequences) that might result from participation in the study. The patient signs the consent form to indicate that she understands the risks and consents to participate in the study despite the risks [4,5].

1.2. The risk of loss of privacy resulting from participation in a medical study

Loss of privacy relates to adverse or unwanted intrusions into the private life of a person who shared personal information in the course of a medical study.

Examples. Members belonging to an extended family consent to have their blood samples used in the study of a familial disease. In the course of research, a husband is informed by the research team that an examination of blood samples had determined that one of his children is not his biological progeny. Unless he had given specific consent to have this kind of information brought to his attention, this would be an action that violates the patient's privacy.

Another example of loss of privacy may occur when a person who has participated in a completed study is pursued by the medical researcher for permission to obtain additional samples of blood for another project.

Both examples describe unanticipated intrusions into a person's life that occurred, because the researcher was given private information about the research subject. In general, privacy issues arise from studies that create new patient data. Privacy issues are almost always pertinent to heritable genetic studies, even when the study does not involve the creation of new experimental data, because studies that review data records of many members of a family may produce new information related to the likelihood of disease in individual members of the family. In general, studies that have privacy risks will require the individual consent from all of the people whose data records are used in the study. Assuming that the only risk for the study participants would be loss of privacy, the consent form would need to clearly delineate the conditions under which study participants would be contacted by the researchers.

1.3. The risk of loss of database security

In general, security issues relate to unauthorized intrusions into hospital-based or other institutional databases where large caches of data are kept on many individuals. In the US, recent Health Insurance Portability and Accountability Act (HIPAA) regulations detail the general steps that hospitals must take in order to meet minimal requirements for securing patient-based electronic data [4]. In most instances, when researchers access patient data

for research purposes, they will be following the same security procedures designed for accessing patient data for patient care purposes. These might include obtaining permission (authorization) from institutional authorities to access the data, obtaining and guarding valid user passwords, using institutional computer equipment in a prescribed manner, etc. A special case will arise when a data miner seeks to expand or modify the hospital information system with attributes specific for their research. For example, a data miner may want be to modify the server to function as a peer node in a peer-to-peer network that receives and exchanges data records with other institutions participating in a data mining project. The data miner would probably be expected to provide a proposal to the institutional security office detailing a safe and secure research plan.

1.4. The risk of loss of confidentiality

Privacy, security and confidentiality are related but distinct conditions. Confidentiality is broken when researcher (a person who holds the patient's confidence) conveys private information about a patient to another party unauthorized to receive the information. This is very different from privacy issues, manifested when subjects in a research effort face unanticipated intrusions in their personal life. Confidentiality issues are also distinct from security issues. Security is broken when an unauthorized individual gains access to patient records. Although there have been reported instances of hospital employees releasing patient records for malicious or mischievous purposes, the authors could find no instances where this occurred during a research study. Based on many conversations with scientists and scientific administrators with expertise in the area of human subject protections, it would appear that researchers have done a remarkably good job at guarding patient confidentiality.

Some would-be data miners hold the opinion that their research should not be classified as human subjects research, because they use patient records, not patients. In the US, the Common Rule (45CFR46) clarifies that this is not so [5]. When a patient's record is used in a research project, the patient is considered a participant in the research, and the research project is classified as human subject research. Federally funded human subjects research is regulated by the Common Rule. The Common Rule provides some relief to data miners, however, by permitting the free use of patient records that contain no information that could associate a particular person to the patient record. Another way of looking at it is that if a record has no patient, then it is not a "patient record". In addition, the law suggests several ways of disassociating patients from their records: anonymization and deidentification. These two processes of rendering patient data harmless to patients can be discussed as computational algorithms, and fall under the purview of the Artificial Intelligence expert, as described herein.

2. The responsibilities of data miners to human subjects

For most medical data mining efforts, issues of direct interventional risk will not apply. Most data mining work will be done using anonymized and deidentified patient records that have been obtained from institutional databases via authorized data transfers.

Researchers will not need to obtain patient consent for the use of each patient record in the dataset. Federal laws applying to the research uses of medical data [4,5] permit the unlimited and unburdened use of patient records provided that the records are anonymized (as described by the Common Rule) [5] and deidentified (as described by HIPAA) [4].

In the US, the responsibility of the would-be data miner will often come down to this:

(1) Demonstrating to the hospital's Institutional Review Board (IRB, the committee that ensures ethical and legal conduct in human subject research) that the data miner's chosen methodology for anonymizing and deidentifying records are reliable, and meet the requirements specified under the Common Rule and HIPAA [4,5].
(2) Proving to the hospital's IRB and to the hospital's Security Review Board (the committee that safeguards patient records) that any transfer of data between researchers can be performed without producing system security breaches and meets the requirements specified by law [4].
(3) Knowing and obeying local laws related to the use of patient records. Individual states have written regulations in this area. Some of these regulations may apply to the transfer of patient records into or out of a state, and these regulations may have interstate scope. In general, it is always a good policy to use anonymized and deidentified data when transferring data electronically.

3. Patient record anonymization

The term anonymization is an unofficial term popularly used in the US to embody the language contained in Exemption 4 of 45CFR46, the Common Rule [5].

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Anonymization usually involves stripping patient identifiers (name, address, social security number, hospital record number, etc.) from records or substituting a false identifier for the real identifier. Once anonymized, even the researcher has no way of determining the patient based on inspection of the patient record. The researcher is *not* permitted to create a table that maps the real identifiers to the false identifiers.

Research conducted with E4-exempted material is considered harmless under Common Rule and can be used ad libitum without obtaining patient consent. There are several points that the data miner must understand.

1. In most institutions, the IRB will need to review any anticipated E4-exempted research to determine whether the research actually satisfies the E4 exemption. This requirement irks researchers who complain that if their research is exempted from IRB approval, they should not need any IRB review. Most institutional assurance documents stipulate that IRBs review all institutional research, if only to determine that the research satisfies the E4 exemption and does not need to undergo the IRB approval process [1].

2. If the research records are ''public'' (e.g. found in a published book or excerpted from a website), exemption is virtually non-contestable under statute. However, an institution may have compelling ethical reasons to suspend research on publicly available data that had been obtained illegally (e.g. Nazi medical experiments on concentration camp prisoners).
3. If the data is to be rendered anonymous through a computational algorithm exercised over each record, you may need to prove that the algorithm actually works and that the process does not require human review. If the process removes a patient's name from an identifier field but neglects to remove the name from a section of text buried in a surgical pathology report, then the anonymization algorithm has failed. If the anonymization algorithm needs to be developed by taking a set of identified reports (training set) and visually comparing them with an output (typically repeating the process with test sets), then someone needs to have direct access to identified patient reports, and this access violates exemption.
4. As commonly used, the anonymization process precludes the ability to identify patients, even when the research uncovers information that may have critical importance to the patient. Consequently, anonymization often removes the opportunity to check data integrity by reviewing the primary patient data record and of adding data later accrued to the patient record (i.e. follow-up clinical data is disallowed).
5. Protocols have been proposed that functionally anonymize data using an encryption broker. The broker receives from an institution medical data with encrypted identifiers. The broker encrypts the identifiers one more time and passes the data (now with doubly encrypted identifiers) to the researcher. The data has been made functionally anonymized in that no single entity can link the medical data to a patient. However, if the original institution, and the researcher and the broker all agree to re-identify the data, they may do so. Presumably, this would only be done with IRB approval [2]. Most IRBs have had no experience with brokered double-encryption protocols, and it remains to be seen whether they will be accepted as legitimate anonymization methods.
6. Even when a study has been exempted under the Common Rule, the use of patient data may still be restricted under other federal regulations (e.g. HIPAA). Or it may fall victim to restrictions in applicable State laws.

4. Patient record deidentification

HIPAA permits unrestricted research use of electronic patient records that have been deidentified [4]. Deidentification is subtly different from anonymization. In deidentification, a record may actually contain an encrypted patient identifier with which authorized persons may re-link a patient with her research record. However, a deidentified record must contain no information that will allow an unauthorized person to infer a patient's identity using clues from the data elements. For instance, if a data record contains zip code, gender, data of birth, and ethnicity of a patient (common demographic elements), along with diagnostic information, a malicious individual could identify the patient using public records (such as birth records, the telephone book, or any lists of residents of an area

along with non-medical demographic data) [9,10]. For instance, if a medical recordset indicates that a person with a particular disease is an Hispanic male of a certain age living in a specified zipcode, then obtaining the individual's indentity might be easy for someone with a list of area residents along with non-medical demographic data (age, address, ethnicity).

5. Computational algorithms that achieve patient confidentiality

A patient record may be anonymized and deidentified, a combination of the two, or neither. Data miners should develop strong algorithms that anonymize AND deidentify patient records.

5.1. Anonymization algorithms

5.1.1. Stripping identifiers

An anonymization algorithm removes all patient identifiers in a record. Most hospital information systems have a well-defined set of patient identifiers (name, social security number, medical record, etc.). These usually are accessed from reserved data dictionary fields. Once these reserved fields are accessed, all other fields (including free text fields) can be parsed and deleted of any identifier matches. So, if the patient's name (as found in the ''name'' field) is Thomas Patterson, then any mention of Thomas Patterson should be deleted from free-text patient fields (such as admission notes, history notes, discharge notes). In addition, a robust algorithm should search and destroy any reference to Tom Patterson, or just plain Tom or the word Mr., Miss, Mrs., or Ms. followed by any name. It may be beneficial to have a list of all the names of patients registered in the hospital system and delete any free-text match (even if the name is not the name of the person attached to the record). It might be useful to remove all mention of staff doctor names or of any surname following the Dr. title. A similar process might be applied to the medical record number or social security number and would include all variations of presentation of a social security number (with or without hyphenation).

5.1.2. One-way hashing algorithms

A one-way hash is an algorithm that transforms a string into another string is such a way that the original string cannot be calculated by operations on the hash value (hence the term ''one-way'' hash). Examples of public domain one-way hash algorithms are MD5 and Standard Hash Algorithm (SHA) [8,12]. These differ from encryption protocols that produce an output that can be decrypted by a second computation on the encrypted string.

The resultant one-way hash values for text strings consist of near-random strings of characters, and the length of the strings (e.g. the strength of the one-way hash) can be made arbitrarily long. Therefore, name spaces for one-way hashes can be so large that the chance of hash collisions (two different names or identifiers hashing to the same value) is negligible. For the fussy among us, protocols can be implemented guaranteeing a dataset free of hash-collisions, but such protocols may place restrictions upon the design of the dataset (e.g. precluding the accrual of records to the dataset after a certain moment).

In theory, one-way hashes can be used to anonymize patient records while still permitting researchers to accrue data over time to a specific patient's record. If a patient returns to the hospital and has an additional procedure performed, the record identifier, when hashed, will produce the same hash value held by the original dataset record. The investigator simply adds the data to the "anonymous" dataset record containing the same one-way hash value. Since no identifier in the experimental dataset record can be used to link back to the patient, the requirements for anonymization, as stipulated in the E4 exemption are satisfied (vida supra).

The use of one-way hashes to anonymize patient records has been employed and promoted in France. Bouzelat et al. [3,7] have standardized a protocol for coding names using SHA one-way hashes. There is no practical algorithm that can take an SHA hash and determine the name (or the social security number or the hospital identifier, or any combination of the above) that was used to produce the hash string. In France, the name-hashed files are merged with files from many different hospitals and used in epidemiologic research. They use the hash-codes to link patient-data across hospitals. Their methods have been registered with Service Central de la Securite des Systemes d'information (SCSSI), http://www.hbroussais.fr/Broussais/InforMed/InforSante/Volume9/34.html.

Implementation of one-way hashes carry certain practical problems. Attacks on one-way hash data may take the form of hashing a list of names and looking for matching hash values in the dataset (dictionary attacks). This can be solved by encrypting the hash or by hashing a secret combination of identifier elements or both. Issues arise related to the multiple ways that a person may be identified within a hospital system (Tom Peterson on Monday, Thomas Peterson on Tuesday), all resulting on inconsistent hashes on a single person. Resolving these problems is an interesting area for further research.

5.2. Deidentification algorithms

One of the keystones of deidentification process is the creation of datasets that contain no unique records (with the exception of the unique identifier code). If every record has at least one additional record to which it is identical, then it becomes logically impossible to distinguish any one individual's record from any of the other individuals whose records contain the same data elements. Guaranteeing that any medical dataset contains only ambiguous records (i.e. records with multiple identical instances) is a feasible computational task. It may involve constantly revising the scope of certain data elements (such as using only the first few digits of zip codes, or using only the State to mark a patient's address) or even adding fake (ambiguating) records to the data. The method of deidentification would depend largely on the purpose of the data mining effort.

6. The consequences (in the US) to the data miner and to the sponsoring institution when research is conducted in violation of human subject regulations

The following three sections are the opinions of the author, who is not a lawyer and are not provided as legal advice or authoritative legal opinion.

6.1. Violations against the common rule

The Common Rule applies to federally funded agencies. Violations of the Common Rule may result in:

1. The loss to the institution of its funding for the grant in question.
2. The loss to the institution of its Federal Assurance.

   The Office of Human Research Protections issues Assurances (currently called Worldwide Federal Assurances or WFAs) to institutions that have in-place processes for IRB reviews of research and for maintaining research standards. An institution must have an assurance registered with OHRP in order to receive federal funding for human subjects research.
3. An institution-wide suspension of human subject research efforts.
4. The imposition of grant-related restrictions imposed on the investigators (e.g. a prohibition from applying for federal grant funding).

Needless to say, violating the Common Rule is not recommended as a wise career move for would-be data miners.

6.2. Violations against HIPAA

HIPAA is the 1996 Health Insurance Portability and Accountability Act. As the result of that act, the Health and Human Services (HHS) has issued 45 CFR Parts 160 and 162, the Final Rule of the Health Insurance Reform: Standards for Electronic Transactions. Although the HHS Final Rule is actually separate from HIPAA, common parlance refers to the Final Rule as ''HIPAA''. The HIPAA guidelines place certain requirements on entities that hold and transfer electronic medical records. Specifically HIPAA applies to health plans, health care clearinghouses and to any health care provider who transmits health information in electronic form. Entities that are not one of the above are not subject to HIPAA regulations. HIPAA contains language describing penalties for the misuse of identified patient information.

From HIPAA (Final Rule) ''Section 1177'' of the Act established penalties for any person that knowingly misuses a unique health identifier, or obtains or discloses individually identifiable health information in violation of this part. The penalties include: (1) a fine of not more than US$ 50,000 and/or imprisonment of not more than 1 year; (2) if the offense is ''under false pretenses,'' a fine of not more than US$ 100,000 and/or imprisonment of not more than 5 years; (3) if the offense is with intent to sell, transfer, or use individually identifiable health information for commercial advantage, personal gain, or malicious harm, a fine of not more than US$ 250,000 and/or imprisonment of not more than 10 years. We note that these penalties do not affect any other penalties that may be imposed by other federal programs''.

6.3. Violations against individuals

Violations of either the Common Rule or of HIPAA do not preclude (and may well enhance) claims against institutions and researchers arising from individuals (or groups) who may have suffered as a consequence of the violation.

7. SPIN, a real-life example of data mining

The Shared Pathology Informatics Network (SPIN) is a 5-year research project funded by the US National Cancer Institute involving about 20 medical institutions located throughout the United States. The purpose of SPIN is to develop a network of institutions with heterogeneous medical data structures that can respond to client queries distributed via the Internet with data extracted from each institution's patient records. The general protocols, algorithms and software developed by the SPIN members will be made available for public use. This system is intended to serve the needs of cancer researchers who need pathologic and clinical data related to tissue.

Beginning in April 2001, the SPIN consortium institutions will develop protocols for initiating a query from an Internet client, distributing the query to all of the SPIN members, authenticating the query at each institution, accepting the authenticated query at institutional databases, conducting the query, producing a query response, transforming the query response into a standard output that contains no information that can be linked to the patient, merging query responses from all of the Network institutions, and sending a query response to the Internet client.

The human subject aspects of the SPIN project will focus on developing the methodology that can accomplish these activities in a way that minimizes or eliminates patient risks.

8. Recommendations

In the past, many studies that analyzed confidential patient information were simply ''done''. Investigators felt entitled to study any data that they could access, sublimely insouciant of the IRB review process, just so long as they limited data access to the members of the research team. This era has ended. Researchers have no automatic right to review patient data. Besides developing strategies for minimizing patient risk, as described herein, investigators should take simple steps to characterize their compliance with human subjects requirements.

1. Whenever an investigator submits a work for publication, where the data is derived from patient records, the Methods section should include a description of the steps taken to minimize patient risks, and should document that that the IRB reviewed the research proposal. When these items are missing from a paper, editors and reviewers should feel free to ask authors to supply this information.
2. When an investigator submits a grant application (particularly an application to a US Federal Agency), a detailed strategy for protecting human subjects from human subject risks is required. Investigators should be aware that research using patient records is human subject research. Investigators should also be aware that current US Federal guidelines call for the inclusion of minorities, women and children in clinical studies, unless there is a good reason for excluding them from the study population. For the purpose of satisfying federal inclusion guidelines, most agencies consider studies based on patient records to be clinical studies. A statement describing the inclusion of

minorities, women and children will be, in most circumstances, a requirement for would-be data miners who seek federal funding.

8.1. Human subject issues are a legitimate area of research for the medical data miners

Novel protocols for achieving confidentiality and security are urgently needed by the data mining community. The kinds of studies requiring confidentiality protocols include: distributing network queries across disparate databases, extending the patient's record to collect rich data from an expanding electronic medical record, linking patient records to the records of relatives or probands, and using peer-to-peer nodes to exchange medical data. Data miners would do well to stay abreast of regulations controlling the use of medical data so that they can develop regulation-compliant protocols for data mining activities.

8.2. Anonymized data, by definition, cannot be linked to patients

There is no legal or ethical reason to withhold anonymized datasets from the public. Quite the opposite. Anonymized datsets have enormous value to other researchers who can merge your data with theirs, derive new ways of analyzing your data, or develop new questions that can be addressed by your dataset. Researchers who have created anonymized datasets should seriously consider publishing their data as a primary resource or as a secondary resource attached to any publication that results from the research project. Many journals and on-line publication services (such as PubMed Central and BioMed Central) encourage authors to submit their datasets as publication attachments. Issues of intellectual property impacting on the investigator and the institution (e.g. ownership, licensing of data, derivative work ''reach-through'') have accumulated very little legal precedent.

8.3. Most data mining studies will not require signed patient consent for each individual record included in the dataset

However, data miners will encounter situations where acquiring consent is unavoidable. This will be pertain to cases in which the analysis of data may result in the creation of new data elements to be included in the patient record (e.g. likelihood of response to treatment, predicted time to relapse, relative risk for certain diseases) and studies in which identified patient records must be reviewed to find additional information necessary for the success of the study. The issues surrounding patient consent are complex and beyond the scope of this article. Data miners should recognize that the process of tracking patient consent is a data miner's dream (or nightmare). Each patient record in a consented study needs to contain highly accurate information regarding consent status. Some of the complex consent questions that need to be accessible to the data miner are:

8.3.1. What consents does the patient have on record?

A patient may have consented to many studies over many visits to the hospital. The consents may apply to certain specimens/data and not to others and to certain uses of specimens/data and not to others.

8.3.2. Does each consent form have an identifier and a locator, a study number, and a data element indicating that the consent form itself was approved by an IRB?

If needed, could you put your hands on the physical consent document? Does your database indicate the study for which consent was approved? Was the consent form complex, allowing the patient to approve certain uses of specimens/data and decline other uses?

8.3.3. Is each consent tagged with tracking data?

Was the consent approved or declined? What day was consent signed? For children and challenged subjects, was the consent signed by a surrogate? Who was the surrogate? Did the subject change her mind and withdraw consent after consent had been approved? If so, what date did this occur? Did she withdraw consent for a particular use of a specimen/data, or for all purposes described by the consent document? Does her withdrawal of consent apply to more than one consent form?

8.3.4. The costs of consented research often exceed the expectations of investigators and funding agents

It is the personal opinion of the author that consent activities inflate grant budgets without producing any societal benefit. A data miner clever enough to devise a strategy that avoids consent while protecting human subjects from research risk, has done a very good thing.

9. Conclusion

Medical data miners can perform their work in an open environment in which their data can be reviewed and shared with colleagues if they anonymize and deidentify their datasets. Every medical data mining effort, should be considered human subjects research until the IRB determines that it falls in an exempt category. As data mining projects draw on more and more collected data held in hospital information systems, IRBs are likely to need help from Institutional Security Committees before granting approvals.

References

[1] Behlen FM, Johnson SB. Multicenter patient records research security policies and tools. J Am Med Inform Assoc 1999;6(6):435–43.
[2] Berman JJ, Moore GW, Hutchins GM. Maintaining patient confidentiality in the public domain internet autopsy database. J Am Med Inform Assoc (JAMIA), Symp. Suppl. 1996;328–32.
[3] Bouzelat H, Quantin C, Dusserre L. Extraction and anonymity protocol of medical file. Proc AMIA Annu Fall Symp 1996;323–27.
[4] Department of Health and Human Services. 45 CFR (Code of Federal Regulations), Parts 160 through 164. Standards for Privacy of Individually Identifiable Health Information (Final Rule). Federal Register: vol. 65, number 250, 28 December 2000. p. 82461–510.
[5] Department of Health and Human Services. 45 CFR (Code of Federal Regulations), 46. Protection of Human Subjects (Common Rule). Federal Register, vol. 56, 18 June 1991. p. 28003.
[6] Kennedy D. Death at Johns Hopkins. Science 2001;293(5532):1013.

[7] Quantin C, Bouzelat H, Allaert FA, Benhamische AM, Faivre J, Dussere L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. Meth Inf Med 1998;37:271–7.
[8] Rivest R. Request for Comments: 1321, The MD5 Message-Digest Algorithm.
[9] Sweeney L. Guaranteeing anonymity when sharing medical data, the Datafly system. Proc AMIA Annu Fall Symp 1997;51–5.
[10] Sweeney L. Computational disclosure control, a primer on data privacy protection. http://www.swiss.ai.mit.edu/classes/6.805/articles/privacy/sweeney-thesis-draft.pdf.
[11] Teichler Zallen D. US gene therapy in crisis. Trends Genet 2000;16(6):272–5.
[12] World Wide Web Consortium. SHA-1 Digest. http://www.w3.org/TR/1998/REC-Dsig-label/SHAI-1_0.