Position papers of the 2014 Federated Conference on Computer Science and Information Systems pp. 259–264
DOI: 10.15439/2014F444
ACSIS, Vol. 3

Music Information Retrieval as a Key Framework to Explore Legal Issues Linked to Personal Data Computation

Pierre Saurel
Université Paris-Sorbonne, SND
Email: pierre.saurel@paris-sorbonne.fr

Francis Rousseaux
Université Reims Champagne, CReSTIC
Email: francis.rousseaux@univ-reims.fr

Marc Danger
ADAMI
Email: mdanger@adami.fr

***Abstract*—The forthcoming European regulation on data privacy penalizes violations by a fine of up to one hundred million euros: European Music Information Retrieval researchers must be compliant with any personal data processes. They are not allowed to transfer personal data to the rest of the world, excepted by using so-called "Safe Harbors".**

Detection of any personal data is mandatory, and "whether a person is identifiable, account should be taken of all the means reasonably likely to be used to identify or single out the individual directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification".

The paper proposes a roadmap for ISMIR involving:

- **Methodology (Is "Privacy by Design" a universal solution?);**
- **Epistemology (Are all authorship attribution algorithms separable into data and processes?);**
- **Science (What characterizes a maximal subset from the big data that could not ever be computed by any Turing machine to identify a natural person with any algorithm?);**

Politics (How can ISMIR influence data privacy policies? Is it possible through some metadata standardization activities?).

I. Introduction

THE Music Information Retrieval (MIR) community addresses a wide range of scientific, technical and social challenges, dealing with processing, searching, organizing and accessing music-related data and digital sounds through many aspects, considering real scale use-cases and designing innovative applications, exceeding its academic-only initiatory aims.

Recent Music Information Retrieval tools and algorithms aim to attribute authorship and to characterize the structure of style, to reproduce the user's style and to manipulate one's style as a content [1], [7]. They deal for instance with active listening, authoring or personalised reflexive feedback. These tools will allow identification of users in the big data: authors, listeners, performers [2], [10].

As the emerging MIR scientific community leads to industrial applications of interest to the international business (start-up, Majors, content providers, platforms) and to experimentations involving many users in living labs (for MIR teaching, for multicultural emotion comparisons, or for MIR user requirement purposes) the identification of legal issues becomes essential or strategic.

The MIR community already seized the technical challenge of Digital Right Management. This challenge was one of identified legal issue related to copyright and Intellectual Property. The MIR community seized the challenge related to Information Access. This challenge was connected to security, business models and right to access [11]. Privacy is another legal challenge. A classification of personal data and processes is necessary to address this challenge precisely. A naive classification appears when you quickly look at the kind of personal data MIR deals with:

User's evaluation, comments, annotation and music recommendations are obvious personal data as long as they are published under their name or pseudo;

Internet Protocol (IP) addresses, Media Access Control (MAC) addresses and addresses allowing identification of a device or an instrument, are linked to personal data;

Any information allowing identification of a natural person, as some MIR processes do, shall be qualified as personal data and processing of personal data.

But the legal professionals do not unanimously approve this classification. For instance the Court of Appeal in Paris judged in two decisions (2007/04/27 and 2007/05/15) that the IP address is not a personal data.

II. Regulation of Personal Data Processes

A specific classification of MIR personal data processes must consider the applicable law of personal data and take the diverging international regulations into account.

A. Taking the divergence between European and American legal approaches into account

Europe regulates data protection through one of the highest State Regulations in the world (two Directives and a Regulation of the European Parliament and of the Council to come) when the United States lets contractors organize data protection through agreements supported by consideration and entered into voluntarily by the parties. These two legal approaches are deeply divergent. United States lets companies specify their own rules with their consumers while Europe enforces a unique regulated framework on all companies providing services to European citizens. For instance any company in the United States can define how long they keep the personal data, when the regulations in Europe would specify a maximum length of time the per-

sonal data is to be stored. And this applies to any company offering the same service.

The European Commission's Directive on Data Protection (95/46/CE – The Directive) prohibits any transfer of personal data to non-European Union countries that do not meet the European Union adequacy standard for privacy protection is strictly forbidden. The divergent legal approaches and this prohibition alone would outlaw the proposal by American companies of many of their IT services to European citizens. In response the U.S. Department of Commerce and the European Commission developed the Safe Harbor Framework (SHF) [8]. Any non-European organization is free to self-certify with the SHF and join.

The European Parliament voted on 12 March 2014 a new Proposal for a Regulation on the protection of individuals with regard to the processing of personal data. The Directive allows adjustments from one European country to another and therefore diversity of implementation in Europe when the regulation is directly enforceable and should therefore be implemented directly and in the same way in all countries of the European Union. This regulation should apply in 2016. This regulation enhances data protection and sanctions to anyone who does not comply with the obligations laid down in the Regulation. For instance (Article 79) the supervisory authority will impose, as a possible sanction, a fine of up to 100 million euros or up to 5% of the annual worldwide turnover in case of an enterprise, whichever is higher.

B. Data protection applies to any information concerning an identifiable natural person

Under French Law were personal data only defined considering sets of data containing the name of a natural person. This State of the Law changed with the application in France of the 95/46/CE European Directive. The definition of personal data has been extended; the 95/46/CE European Directive (ED) defines 'personal data' (Article 2) as: "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity".

For instance the identification of an author through the structure of his style as depending on his mental, cultural or social identity is a process that must comply with the European data privacy principles.

C. Safe Harbor is an identified Framework allowing to avoid to pay the financial fine up

Compliance with Safe Harbor is an issue for an organization using MIR processing to fulfill the high level European standard about personal data, to operate abroad and to be confident in avoiding prosecution regarding personal data. An American organization may decide to enter the US-EU SHF's requirement. This Company has to design a data privacy policy complying the seven Principles (SHP).

First of all organizations must identify personal data and personal data processes. Then they apply the SHP to these data and processes. By joining the SHF, organizations must implement procedures and modify their own information system whether paper or electronic.

Organizations must notify (P1) individuals about the purposes for which they collect and use information about them, to whom the information can be disclosed and the choices and means offered for limiting its disclosure. Organizations must explain how they can be contacted with any complaints. Individuals should have the choice (P2) (opt out) whether their personal information is disclosed or not to a third party. In case of sensitive information explicit choice (opt in) must be given. A transfer to a third party (P3) is only possible if the individual made a choice and if the third party subscribed to the SHP or was subject to any adequacy finding regarding to the ED. Individuals must have access (P4) to personal information about them and be able to correct, amend or delete this information. Organizations must take reasonable precautions (P5) to prevent loss, misuse, disclosure, alteration or destruction of the personal information. Personal information collected must be relevant (P6: data integrity) for the purpose for which it is to be used. Sanctions (P7 enforcement) ensure compliance by the organization. There must be a procedure for verifying the implementation of the SHP and the obligation to remedy problems arising out of a failure to comply with the SHP.

III. Classification for Music Information Retrieval Personal Data Processing

Considering the legal definition of personal data we can now propose a less naive classification of MIR processes and data into three sets: (i) nominative data, (ii) data leading to an easy identification of a natural person and (iii) data leading indirectly to the identification of a natural person through a complex process.

A. Nominative data and data leading easily to the identification of a natural person

We first consider two sets of processes. The first set aggregates the information systems directly containing the name of a natural person. The second set aggregates the cases allowing a direct or an indirect identification easily done for instance through devices.

In these two sets we find that the most obvious set of data concerns the "Personal Music Libraries" and "recommendations". As long as one recommends music to a user or analyze their personal library, he certainly deals with his privacy?

B. Data leading to the identification of a natural person through a complex process

The third set of personal data aggregates the information systems when a natural person is indirectly identifiable using a complex process, like some of the MIR processes.

Can one work on Machine Learning and especially on Categorization with no consideration about the taste or the

style of the consumers or of the users? These processes belong for the most part to this third set. Looking directly at the data without any sophisticated tool does not allow any identification of the natural person. In contrast, MIR nonlinear algorithms or machine learning leads frequently to an indirect identification [7].

Usually do MIR algorithms use inputs to build new data which are outputs or data stored inside the algorithm, like weights for instance in a neural net [6].

C. The legal criteria of the costs and the amount of time required for identification

This third set of personal data is not as homogeneous as it seems to be at first glance. Can we compare sets of data that lead to an identification of a natural person through a complex process?

The European Proposal for a Regulation designs the concept of identifiability. It gives a legal criteria to determine if an identifiable set of data is or is not personal data. It regards the identification process (Recital 23) as a relative one that would change according to the effectiveness of the identification: *"To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person to identify or single out the individual directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development."*

How can we define, as MIR practitioners, when a set of data will result in an easy identification and should then be classified into the second set or, on the other hand, is especially uncertain so that it cannot be considered and categorized as personal data? New criteria are necessary to answer these new challenges and interrogations about MIR processes.

IV. What is the Third Set about

"Identifiability", in our classification, is a potentiality of a set of data. This set should be qualified as being personal data if the cost and the amount of time required for identification are reasonable. These new criteria are a step forward since the legal qualification is not an absolute one anymore and depends specifically on the state of the art [13].

A. Available technology and technological development to take into account at this present moment

The internet is one of the high level technologies that modifies the identifiability of a set of data. All over the world, people publishes data and personal data without heeding the potentialities of these data. They are usually not completely aware about the ways these data can be used to describe their personal habits. When a listener tags music or recommends an item, he publishes information allowing to paint a portrait of his personality. If a company exploits this user data without integrating strong privacy protections, he can encounter legal issues and extensive disaffection from his customers.

The volume of data have increased faster than "Moore's law": This is the "Big Data". New data is generally unstructured and traditional database systems such as Relational Database Management Systems cannot handle the volume of data produced by users and by machines & sensors. This challenge was the main drive for Google to define a new technology: the Apache Hadoop File System. Within this framework, data and computational activities are distributed on a very large number of servers. Data is not loaded to be computed, and the result stored. Here, the algorithm is close to the data. This situation leads to the epistemological problem of separability into the field of MIR personal data processing: are all MIR algorithms (and for instance the authorship attribution algorithms) separable into data and processes. An answer to this question is required for any algorithm to be able to identify the set of personal data it deals with.

Databases of personal data are no more clearly identified. We can identify new scientific challenges about MIR personal data processing. These are the result of five complementary sides of the situation.

Data Sources Profusion. Many new databases and datawarehouses are developed every day allowing to trace and recognize many kind of information. Software, as Spotify for example, become new kind of live and on-line data sources providing a flow of music consumption information from numerous users. These data sources will shortly integrate new devices under the Internet of Things. Not all of these data are of high-quality. These new data and datasources dot not always preserve the legal rights of the users. Taking this into account will be of great help to shape reliable and sustainable systems.

Crossing & Reconciling Data. Data sources are not separated and independent one of the other. The aggregation of the data will first allow a more precise identification through user id, cookies or emails and then make technically possible to bring closer, combine and blend data that were earlier incommensurate.

Temporal Aspects. The memory of the web, such as information systems in general, is beyond all what people can usually imagine. The Public Status of Data changes frequently. Most of the users do not trace their own personal data. A video posted during a party could reappear suddenly when one is seeking a job. All the traces collected one day with a given purpose could be technically exploited later with a different or an opposite purpose.

Permanent Changes. The general instability of the data sources, technical formats and flows, applications and use is another strong characteristic of the situation. The impact on personal data is very likely. If the architecture of the systems changes a lot and frequently, the social norms also change. Users today publicly share information that they

would have considered totally private a few years earlier. And the opposite could be the case.

User Understandability and Control. These situations are less and less easy to understand for normal users. The complexity of the systems and of the interactions between humans and machines results in non-linear causality which may leads to confusing situations. The case of the private Facebook posts displayed all at once on the timeline (Sept. 2012) is full of meaning. Facebook announced that there was no bug. Those information were old personal posts which became more publically visible with the new timeline. This situation is due to two simultaneous factors: the misunderstanding of a human user combined with the rate of change of an information system.

Changes in the Information Technology result to a shift in the approach of data management: from computational to data exploration. The main question is "What to look for?" Many companies design new systems and processes to "make the data speak". Direct marketing is one of these players: dataflow of personal data are produced through the big data.

One of the solution could be a stabilization of the dataflow through a universal design of metadata. This could be a way to speed up a specific classification of MIR processing of personal data into identifying and non-identifying processes.

This situation results into a new scientific challenge: What could be an absolute criterion about the identifiability of personal data extracted from a set of data with a MIR process? How could we define into the big data, a maximal subset that could not ever be computed by any Turing machine to identify a natural person with any algorithm?

B. Personal data produced on the fly: the Gamelan Project as a case study

IRCAM supervised the Gamelan Project (2009-2013) in partnership with EMI, INA and UTC [3] [4].

Gamelan is a software environment, built upon the production ecosystem, to address the reconstitution issue of digital music production, by combining trace engineering, knowledge modelling and knowledge engineering. Reconstitution is usually relegated afterwards. Gamelan reconstructs the composer-system interactions that have led to the creation of a work of art that is about to leave the production studio. The purposes concerns long-term preservation, repurposing, versioning, evolution of the work of art and the disclosure of the contingencies of its initial outcome.

A creator finishing his or her work in a studio marks the end of the production process: the long-awaited object is finally there, thus the creator, the producer, the sound engineer and all the people involved are happy or at least relieved; the goal is reached and the story reaches its end. However, at this very moment, because the final object is there, no one wonders about its reconstitution.

But —say ten years later— when "back-catalog" teams of music companies want to edit some easy-to-sell Greatest Hits in up-to-date audio formats, delving into the musical archives is no longer easy. Returning to the reachable-recorded digital files, it may be difficult to figure out which one of the bunch of files left is the one needed. File dates and file names cannot be trusted.

Closer in time —say two months after the production— the simple task of collecting vital information on the contributors who actually worked on the project may turn into a real problem. There is a whole set of information on contributions (name, role, time spent, etc.) necessary to manage salaries, rights and royalties that regularly proves hard to collect afterwards. Evidently, this kind of information would be far easier to collect directly at production time.

On the surface this is nothing to do with privacy and personal data! But in fact, and it is typically the case as soon as a complex person-software device is involved, this type of project invites us to rethink classical approaches and qualifications of privacy issues. The Gamelan project exemplifies several of the many R&D emerging questions that are raised in the digital audio processing domain.

First of all reconstitution requires to collect traces during the production process itself. Automatically-collected software traces differ from human-entered traces. The former can be seamlessly collected through automatic watching components, with interfaces traces and logs as heuristic material, while the latter inevitably requires a human contributor for information that cannot be automatically retrieved or inferred from automatic traces. A full-production tracking environment would resemble Living Labs, towards a Living Studio.

Secondly, these traces call for an appropriate knowledge model. To remain as uninvasive as possible, such a model should provide means to determine which information is worth asking people during the production or not compared to the cost of disturbing the creativity. Without a knowledge model, it would not be possible to interpret the traces or to determine the kind of traces worth retrieving. To achieve this model, professional knowledge must be identified, listed and characterized with experts, defining a digital music production Knowledge Level.

Within Gamelan, traces from used operating system and from used professional applications are extracted. Semantic networks dealing with typical digital audio composition acts are involved towards some abstraction of those traces, but personal data is nowhere considered: some real time digital audio flow is involved, transformed on the fly by creative acts that assign the composers particular style and their artistic singularity.

The composers style, as part of built up personal data, often not even named, is computed to support the Gamelan reconstruction process: what is to be reconstituted has to do with the abstract truth of the piece of art and its stylistic genesis. To understand that "the composer is currently testing a sample within the whole piece framework" is more efficient than being aware of a succession of cut-paste-listen actions that has to be generalized.

Thus some personal data, like artistic style, is built up on the fly, relating to processing algorithms, knowledge bases and title repositories [5], evolving from the system experience itself, and only known by the system. The ultimate target is clearly the style-recognition [1] of the creators, viewed as the correlation between their practice and the character of their work of art.

C. What if the Gamelan Project is called in front of a law court?

The Gamelan Project is a case study where machines produce personal data on the fly. Under our classification it is a case study which produces a kind of personal data from the third set. In that kind of case, one must be reactive and cannot just apply the Safe Harbor Principles at the end of the project. How could we apply the Safe Harbor Framework from the beginning and avoid any prosecution for breaking the data privacy laws since we cannot decide at first which set of data is personal?

A special methodology must be applied taking into account the possibility that, during the project, new data and processes could be qualified as personal data.

V. Privacy by Design: A Methodology for MIR Project Management?

Privacy by Design (PbD) was developed in the 1990s. This methodology has become an international reference about privacy. It now evolves taking the big data into account.

A. Foundations Principals (FP) of Privacy by Design

PbD is grounded on on seven FP1: PbD *"*is an approach to protect privacy by embedding it into the design specifications of technologies, business practices, and physical infrastructures. That means building in privacy up front – right into the design specifications and architecture of new systems and processes. PbD is predicated on the idea that, at the outset, technology is inherently neutral. As much as it can be used to chip away at privacy, it can also be enlisted to protect privacy. The same is true of processes and physical infrastructure*"*:

- Proactive not Reactive (FP1): the PbD method is grounded not on reactive but on proactive parameters anticipating and preventing privacy invasive events before they occur;
- Privacy as the Default Setting (FP2): the default parameters attempt to fix the maximum degree of privacy;
- Privacy embedded into Design (FP3): the architecture of IT systems takes privacy into account;
- Full Functionality – Positive Sum, not Zero-Sum (FP4): PbD results into a "win-win" solution considering the different interests and objectives;
- End-to-End Security – Full Lifecycle Protection (FP5): the design of the solution regarding to privacy and security measures is defined in a complete Lifecycle;
- Visibility and Transparency — Keep it Open (FP6): PbD is an open process and the quality of the solution can be asserted by an external audit. Transparency is a key success factor;
- Respect for User Privacy — Keep it User-Centric (FP7): as they apply PbD will SI designers take the user's personal interest into account especially concerning his privacy and personal data.

PbD is a key-concept in legacy [9] when you have to design numerical and digital processes. The European Union[2] affirms that *"PbD means that privacy and data protection* are embedded throughout the entire life cycle of technologies, from the early design stage to their deployment, use *and ultimate disposal"*. Europe [12] took the Canadian experiments into account when it decided to use PbD as a key-concept in the heart of the legal data protection.

B. Prospects for a MIR Privacy by Design

PbD is a standard for designing systems and processing involving personal data. PbD was enforced by the new european proposal for a Regulation (Article 23). It becomes a method for these designs whereby it includes signal analysis methods and may interest MIR developers.

This proposal leads to new questions as for instance: Is PbD a universal methodological solution about personal data for all MIR projects? Most of ISMIR contributions are still research oriented, in the sense of Article 83 of the "Safeguarding Privacy in a Connected World". To say more about that intersection, we need to survey the ISMIR scientific production, throughout the main FP.

FP6 (transparency) and FP7 (user-centric) are usually respected among the MIR community as source code and processing are often (i) delivered under GNU like licensing allowing audit and traceability (ii) user-friendly. However, as long as PbD is not embedded, FP3 cannot be fulfilled and accordingly FP2 (default setting), FP5 (end-to-end), FP4 (full functionality) and FP1 (proactive) cannot be fulfilled even. Without any PbD embedded into Design, there are no default settings (FP2), you cannot follow an end-to-end approach (FP5), you cannot define full functionality regarding to personal data (FP4) nor be proactive. Principle of pro-activity (FP1) is the key. Fulfilling FP1 you define the default settings (FP2), be fully functional (FP4) and define an end-to-end process (FP5).

In brief is PbD useful to MIR developers even if PbD is not the definitive martingale!

VI. Structuring the Data Bases heeding Privacy by Design: The case of the Gamelan Project

The three sets of personal data designed considering legal rules relative to data privacy should be tuned and man-

[1] http://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

[2] "Safeguarding Privacy in a Connected World – A European Data Protection Framework for the 21st Century" COM(2012) 9 final.

aged to be able to be stored into different tables or servers.

The Gamelan project [3] is a used case allowing the design of structuration of the personal data. The Gamelan project was designed into three layers to:

- Track production process,
- Interpret collected traces according to a domain ontology,
- Help querying and visualizing to foster production understanding.

These levels are closed to the classical layers: the data level, the information level and the knowledge level. The classification of the three sets could deal with these three levels.

The data level is more or less specifically relative to the two first sets. The information and knowledge level are connected to the third set.

Concerning the Gamelan project, the learning processes at the information and knowledge level can be tuned depending on the goals and aims. Depending on this tuning and on the learning time will the processed data become personal data or not meaning that they have allowed the identification of a person in the data base.

VI. Conclusion and Future Work

Methodological Recommendations. This classification leads to methodological recommendations for MIR researchers. The first step is to audit the used algorithm and the data. Could the algorithm identify a natural person in the sense of the two first sets of our classification? In that case, the researcher should use the SHF. To use SHF is not as simple as it seems to be. In some cases it can lead to huge industrial challenge for instance regarding Cyber Security (P5).

In some cases the MIR community develops new personal data on the fly. It can be so when researchers use all the known data algorithms and data analysis especially relative to machine learning. The PbD methodology should then be applied. This methodology frames a design that preserves personal data and avoids any unintentional loss of data.

But the time when data (on the one hand) and processing (on the other hand) were functionally independent, formally and semantically separated, has ended. Nowadays, MIR researchers currently use algorithms that support effective decision, supervised or not, without introducing 'pure' data or 'pure' processing, but building up acceptable solutions together with machine learning or heuristic knowledge that cannot be reduced to data or processing: The third set of personal data may appear, and raise theoretical scientific problems.

Political Opportunities. The MIR community computes algorithms dealing with style description since a long time. The MIR community could join expert groups dealing the legal aspects of personal data. It could explain to the lawyers these algorithms and the ones relative to machine learning. This could be of great benefit for both parties.

Future Scientific Works. The three sets and the new definition of personal data leads to new pure scientific challenges. These challenges constitute our research program for future works. What are the conditions to specify a set of data resulting into an identification in the sense of the second set. When can we assure that an algorithm allows a too hazardous recognition? In that case we would say that the data are not personal in a legal sense? How can we design or carve a maximal subset from the big data that could not lead to the identification of a natural person by any Turing machine and with any known or forthcoming algorithm? These are some of the new scientific challenges we are now dealing with.

References

[1] S. Argamon, K. Burns, S. Dubnov (Eds): The Structure of Style, Springer-Verlag, 2010. DOI 10.1007/978-3-642-12337-5

[2] K. Barkati, A. Bonardi, A. Vincent, F. Rousseaux: "GAMELAN: A Knowledge Management Approach for Digital Audio Production Workflows", Proceedings of the European Conference on Artificial Intelligence, Workshop "Artificial Intelligence for Knowledge Management", 2012.

[3] C. Barlas: "Beating Babel - Identification, Metadata and Rights", Invited Talk, Proceedings of the International Symposium on Music Information Retrieval, 2002.

[4] T. Bertin-Mahieux, D.P. Ellis, B. Whitman, P. Lamere: "The Million Song Dataset", Proceedings of the International Symposium on Music Information Retrieval, 2011.

[5] J.S. Downie, J. Futrelle, D. Tcheng: "The International Music Information Retrieval Systems Evaluation Laboratory: Governance, Access and Security", Proceedings of the International Symposium on Music Information Retrieval, 2004.

[6] A. Gkoulalas-Divanis, Y. Saygin, Vassilios S. Verykios: "Special Issue on Privacy and Security Issues in Data Mining and Machine Learning", Transactions on Data Privacy, Vol. 4, Issue 3, pp. 127-187, December 2011.

[7] D. Greer: "Safe Harbor - A Framework that Works", International Data Privacy Law, Vol.1, Issue 3, pp. 143-148, 2011.

[8] M. Levering: "Intellectual Property Rights in Musical Works: Overview, Digital Library Issues and Related Initiatives", Invited Talk, Proceedings of the International Symposium on Music Information Retrieval, 2000.

[9] F. Pachet, P. Roy: "Hit Song Science is Not Yet a Science", Proceedings of the International Symposium on Music Information Retrieval, 2008.

[10] V. Reding: "The European Data Protection Framework for the Twenty-first century", International Data Privacy Law, volume 2, issue 3, pp.119-129, 2012. DOI 10.1093/idpl/ips015

[11] A. Seeger: "I Found It, How Can I Use It? - Dealing With the Ethical and Legal Constraints of Information Access", Proceedings of the International Symposium on Music Information Retrieval, 2003.

[12] A.B. Slavkovic, A. Smith: "Special Issue on Statistical and Learning-Theoretic Challenges in Data Privacy", Journal of Privacy and Confidentiality, Vol. 4, Issue 1, pp. 1-243, 2012.

[13] P. Symeonidis, M. Ruxanda, A. Nanopoulos, Y. Manolopoulos: "Ternary Semantic Analysis of Social Tags for Personalized Music Recommendation", Proceedings of the International Symposium on Music Information Retrieval, 2008.