

Network Analysis Formalism and the Construction of a Traceability System for Payments. A Sketch of Its Legal and Sociological Aspects

GUGLIELMO FEIS*

SUMMARY: *Introduction – Part I: Philosophical Motivations: Law & Informatics Towards a Network Analysis Formalism (NAF) – 1. Method and Organization of Part I – 2. The Relationship between the Theoretical Framework of NAF and the Applied Example of the TS for Payments – 3. The Theoretical Framework of Network Analysis Formalism (NAF) – 3.1. The Possible Objection of Killing Legal Activities through NAF and a Reply – 4. Conclusions on NAF – Part II: Introducing the Problem of the Traceability System (TS) – 5. Method and Organization of Part II – 5.1. Some Lexicon – 6. The General Principle – 6.1. Presuppositions of the General Principle – 7. The Applied Principle – 7.1. Presuppositions of the Applied Principle and its Consequences – 7.2. Going Forward Social Consequences – 7.3. Going Forward Legal Consequences – 8. Implementing the Applied Principle: Data Mining and Database Organization – 8.1. Introducing the Asymmetry – 8.2. The Mining Problem – 9. Case Studies: Recorded Future and Tassa.li – 9.1. Tassa.li – 9.2. Recorded Future – 10. Security – 11. Conclusions on the Traceability System – 12. General Conclusions*

INTRODUCTION

This paper is two-folded: in the first part, I will approach the theoretical problem of the relationships between law and computational social sciences, putting forward a framework that I call network analysis formalism (NAF); in the second part, I will discuss an applied problem concerning the technical and legal aspects of developing a full traceability system (TS) for payments and investigate this problem backward and forward on the temporal axis.

The second part can be seen as an example of the NAF framework presented in the first part that shows how the theoretical proposal of the sociologically aware weak formalism – NAF – can be used to solve some real life legal problems, in the present case, the ones related with structuring a system to track all the economical transactions.

* The Author is a PhD student of Philosophy of law at the State University of Milan (Italy).

PART I: PHILOSOPHICAL MOTIVATIONS: LAW & INFORMATICS TOWARDS A NETWORK ANALYSIS FORMALISM (NAF)

1. METHOD AND ORGANIZATION OF PART I

In this part of my paper I will present (Section 2.) how the NAF framework relates to the case of developing a traceability system (TS) that is going to be discussed in the second part. Then, after solving this methodological issue, I will sketch my NAF framework (Section 3.) and reply to a possible objection according to which NAF has nothing to do with legal science (Section 3.1.). Before moving on to part two, in Section 4., I will draw the conclusions on NAF.

2. THE RELATIONSHIP BETWEEN THE THEORETICAL FRAMEWORK OF NAF AND THE APPLIED EXAMPLE OF THE TS FOR PAYMENTS

In the second part of the paper I will investigate the possibility of developing a traceability system (TS) for payments. Besides the problem of the TS in itself – i.e., how to organize the TS and how to investigate what are its legal and social upshots – the issue is interesting because we can see the TS problem as the applied counterpart of more philosophical questions, namely: what are the relationships between informatics and the new resources of technological progress and the law? What can we gain in terms of “legal welfare” from the increasing computability power of our computers as well as from the spreading of mobile devices that can always connect us to a network?

I will first address these broad questions outlining the framework of network analysis formalism, then I will move to its application considering the TS problem.

3. THE THEORETICAL FRAMEWORK OF NETWORK ANALYSIS FORMALISM (NAF)

My thesis is that entering technological devices, procedures of acquisition of relevant legal data and tech-aware legal agents¹ can help us to reconcile two approaches to the law that, from a theoretical point of view, are

¹ I.e., agents that are active part of the legal system, that can easily achieve an average use of technologies such as credit cards, Internet banking and smartphones.

perceived and presented as opposed: the (neo)positivistic formalism and the realist antiformalism². Getting closer to the point, I think that the possibility to mine data – legal data in particular – can bring us closer both to the leibnizian *calculementus*³ dream and to the anti-formalist idea of some legal realists that stressed the importance of the sociological context in which legal action takes place.

Before going on, I have to say a few words on the two conflicting thesis – the formalist and the anti-formalist – and on the idea of an instrument I would like to use to reduce the gap between the two positions, i.e., informatics in general and data mining in particular.

I propose to call my attempt to rejoin formalism and antiformalism “network analysis formalism” (NAF) and to classify this approach as a “socially aware formalism”. You may object that once you let society, the empirical world and some other kind of context or some other facets of the *Sein* enter the picture of formalism you lose a pure formalist approach. By “socially aware weak formalism” I mean that the abstract types of legal situations are constantly monitored and checked (confirmed or disconfirmed) by the data. These data tell us how and in which way, if any, the law is affecting our daily life.

The idea of data mining comes from informatics and is the way of organizing huge amounts of data in order to discover patterns inside them. The aim of data mining is somehow pretentious: it is not only to describe a structure, but also to use the described structure to make further predictions. The formalists’ idea is that the legal system is complete, that is, it can handle every possible situation: all we need to know is the fact, then the judges need only to be the voice of the law, that is, find the right law that fits that fact and apply it. This is enough to sentence.

On the other hand, the antiformalists deny this dictatorship of the law, saying that we need to be realist and stick to what happens in real trials: the judges have a lot of power; there is no way to predict a sentence just by way of knowing what happened (the facts) and what should have happened (the relevant laws), legal interpretation cannot be avoided. The anti-formalists stressed this point by saying that what you need, in order to know the law,

² This is probably an oversimplified contraposition that does not bring into account the distinctions between American legal realists and Scandinavian legal realists. Despite this inaccuracy, I think it helps to get my point in characterizing NAF.

³ G.W. LEIBNIZ, *The Art of Discovery*, 1685.

is predicting judges' behaviors and decisions (this led to the provocative idea of "digestive jurisprudence" held by some American legal realists).

Despite these conflicting views, I claim that, on the descriptive side, gathering data by way of data mining can help you both to discover what the relevant laws are (getting closer to Leibniz's *calculemus*) and what the past histories of similar cases were (giving the legal realists and the anti-formalists much more data and statistics they imagined when they talk about predicting judges behavior upon the dishes they have eaten).

Thus, on this very general level, I claim that the introduction of tools that can help us in providing a better description of legal practices will help both the formalists and the anti-formalists.

The prospective used to bring closer these different paradigms is empirical and sociological. In the context of legal philosophy, my proposal can be described as Hartian⁴, if you interpret the Hartian rule of recognition and the external point of view as a way to bring sociology into play.

My proposal is the following: I rely on the Hartian idea of an empirical recognition of the elements of a legal system, then I extend it by way of mining that part of legal data that is relevant for the legal system (financial data in the case of the TS, data concerning judges, sentences and litigation in case of the formalists vs. anti-formalists dispute concerning trials and legal interpretation and so on).

This is where data mining and informatics come in handy. I think modeling the data using statistics and network analysis⁵ could be really helpful: you can discover patterns and graphs of tax evasion in the case of the TS or have statistics concerning how different judges and different courts solve similar controversies around the world. Collecting and extracting such information will enable us to access a lot of relevant information concerning prejudice and specific courts ideology: are we more severe in punishing a theft committed by an immigrant with no VISA and low income rather than a theft committed by a young citizen? Such data will give us an opportunity

⁴ H.L.A. HART, *The Concept of Law*, Oxford, Oxford University Press, 1961, introduced the rule of recognition as a secondary rule to empirically recognize what is the *corpus* of a positive legal system.

⁵ (Social) network analysis studies the interaction of individuals, groups and organizations with the tools of graph theory. For an introduction on its tools see D. EASLEY, J. KLEINBERG, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Oxford, Cambridge University Press, 2010. Considering the issue of the TS, network analysis can study the relevant behaviors related to tax payments and goods exchanges and provide models and patterns suggesting possible violation of taxation laws.

to find out if really “the law is the same for everybody” – as it is written in the Italian courts – no matter what is the color of your skin, your income, or how many degrees of separation you are away from a political leader.

The NFA approach tries to take into account every single case in its individuality and, in this regard, it is closer to the realists’ perspective. On the other hand, the data acquired can be used to reinforce a weak formalist perspective: in order to know what will happen in the legal system, you need to know the law and how the law is statistically applied. Thus, data and computation provide an *a posteriori* pattern – close to the Hartian tradition but compatible with the realists’ approach – and not an *a priori* formalist ideal of the *Sollen* or the Ought of the normative world.

3.1. *The Possible Objection of Killing Legal Activities through NAF and a Reply*

One of the first reactions to my proposal could be this: “this is no longer jurisprudence! You are killing legal activity! There is no room left for interpretation in your NFA model!”. You may go even further by saying that NAF is not a weak formalism but, indeed, is the strongest form of formalism ever, whose aim is to replace judges with machines.

Nonetheless, this picture is wrong. Judges are still part of the model: they provide and shape the data that will be the input of the NFA system. Judges are still making decisions, deciding controversies, sentencing and so on. Network analysis is used only to grasp what is going on in legal activities trying to have a larger scale view that cannot be obtained only by way of citing some paradigmatic case. After years, the data collected through a NFA approach might be relevant for legal decision making but, this is something that needs to be stressed, the legal decisions are still realized by bone and flesh human beings, not by siliceous computers.

Computers are useful only in sorting out how the law is used by humans and how a particular law is performing in the achievement of its goals. If we expect some benefits for a society due to the introduction of certain new laws, we can track whether we are gaining such benefits or not (I will show this aspect later, while describing the positive outcomes of the implementation of the TS). The evaluation is done by way of machines, nonetheless, every time the law is applied, there is a human behind it and not a computer. The computer is there just to keep a record of that particular decision.

4. CONCLUSIONS ON NAF

I presented NAF as a methodological proposal that enables us to find something like a third way between formalism and anti-formalism. The idea is that computers are not to be feared and that there is no gap between humanities and computer sciences or between the law in books and its creation, evaluation and application.

Computational social sciences can provide a great load of useful information to evaluate and monitor how the good intentions and purposes that motivate a law are doing in the world both in mapping the consequences of a law (think about a law designed to improve the assumption of young workers) and in mapping how a law has been applied (think about criminal laws related with cultural influenced murders or high debated and vagueness-affected laws concerning issues of bioethics that are highly influenced by the judges' personal beliefs).

PART II: INTRODUCING THE PROBLEM OF THE TRACEABILITY SYSTEM (TS)

In this second part I will present a TS for payment as an example of applied NAF. I want to show how the philosophical considerations exposed in part one can be used as a framework that can be applied to solve problems of our everyday life. I think the issue of payments and tax is a fairly good example: we deal with payments and money in quite a familiar and easy way, despite the fact that there are many problems and legal consequences attached to such phenomena that are by far less intuitive (e.g.: tax regulations).

I will engage an interdisciplinary approach trying to handle the philosophical and legal problems as well as the economical and technical problems related to the security and the possibility of realizing the TS sketched in the paper.

5. METHOD AND ORGANIZATION OF PART II

Here I will discuss the technical and legal aspects of developing a full traceability system (TS) for payments and then investigate it backward and forward on the temporal axis.

Going backward, I will consider what should be the conditions that allow us to build such a system and the benefits given to the society by the

implementation of such a TS. These benefits, I claim, will be both social and economical.

Going forward, I will concentrate on legal and technical aspects of detecting and mining personal data concerning earnings and expenses. The most relevant issues, I think, are related to privacy, especially concerning the (non) anonymity of the collected data.

The two questions that guide my research are the following:

1. What are going to be the legal upshots of a TS for every payment made in a community?
2. What might be the practical consequences of such a TS?

I think the main reasons to justify the development of a TS for payments can be the following:

1. it provides a solid base to collect taxes⁶;
2. if the data are somehow publically available, it may also provide good data to drive financial economy (trading and investing) avoiding financial bubbles.

After providing some lexicon (Section 5.1.), I will first (Section 6.) state a general principle (“all economic transactions should be traced”) and then (Section 7.) I will move on to consider an applied principle (“economic transactions should be traceable and the data should be accessible in order to prompt the development and provide resources to combat tax evasion”). For both principles, I will analyze the presuppositions for the principle to work and the practical upshots of its application. In dealing with the general principle, I will address the problems only at an abstract level.

I will then go deeper into the development of a TS starting from the applied principle (Section 8.), present some case studies concerning already existing traceability systems (Section 9.) and, after a brief consideration on security (Section 10.), draw the general conclusions.

5.1. Some Lexicon

Here I will introduce some terminology that I will adopt during the paper. I will refer to the ones that have to pay the taxes as *tax payer* (TP) while, on the other hand, I will call *tax collector* (TC) the other part of the model that has to collect the data concerning expenses and payments. These will be the two agents of the traceability system (TS).

⁶ Establishing the principles of the taxation system is not a part of the present paper.

Tax collectors also have the burden of directly or indirectly collecting and organizing the data: their final goal is to use these data to increase everybody's welfare and eventually make the data publically available⁷.

6. THE GENERAL PRINCIPLE

The general principle, which expresses the idea behind the TS proposal in the most naïve way is the following:

[*general principle*]: all economic transactions should be traced.

From the point of view of the motivations and goals for adopting a traceability system (see Section 5.1.), the general principle works mainly on the first one: it provides a solid base in order to apply taxation. As it does not say anything concerning the accessibility of the data, it cannot be used to obtain the second goal⁸.

6.1. *Presuppositions of the General Principle*

The conceptual presuppositions of the general principle are at least the following:

1. There should be – on TCs part – a standard model to register and keep the data of all the transactions;
2. There should be – on TPs part – a standard way to enter the data, so that it will be easy for every TP to help TCs contributing to the “traceability game”.

The most problematic presupposition, i.e., the one that is most likely to end up being inefficient, is (2). As the principle is general, we do not have to investigate the effective requirements the society needs to have in order to embrace the principle.

7. THE APPLIED PRINCIPLE

It is now time to move to a more concrete scenario and consider the “applied principle”:

⁷ I leave open whether the TC is the state or some more specific institution.

⁸ Note that, stated in this way, the general principle is still compatible with the circulation of paper money. The general principle does not imply that paper money is to be dismissed or declared illegal.

[*applied principle*]: economic transactions should be traceable and the data should be accessible in order to prompt the development and provide resources to fight tax evasion.

7.1. Presuppositions of the Applied Principle and its Consequences

The conceptual presuppositions of this second principle are the same as the general principle, i.e.:

1. There should be – on TCs part – a standard model to register and keep the data of all the transactions;
2. There should be – on TPs part – a standard way to enter the data, so that it will be easy for every TP to help TCs contributing to the “traceability game”.

Then, we also need to presuppose the two following things:

1. the realization of an archive for the collected data available to the public;
2. a policy stating who can access what kind of information.

The going backwards presupposition related to the society is a well developed technological infrastructure (wi-fi and Internet access) that is needed in order to not exclude people from accessing the data⁹.

Considering the going forward analysis, the upshots of an implementation of the applied principle will result in a better accessibility to the Internet and, inside the European Union, it may help to meet the Digital Agenda¹⁰.

7.2. Going Forward Social Consequences

A TS for payments will help us to detect tax evasion and presupposes high level services to the citizens who need to be given at least:

1. some sort of bank account, credit card or mean of payment almost for free: TPs join the traceability game and should suffer no major

⁹ Of course, there should also be the possibility to request a paper version of the data (for a fair price or for free).

¹⁰ What I have in mind are the following points of the Digital Agenda: 2.1.1. Opening up access to content; 2.1.3. Building digital confidence; 2.2.1. Improving ICT standard-setting; 2.3. Trust and security; 2.4.1. Guarantee universal broadband coverage with increasing speeds; 2.4.3. Open and neutral Internet; 2.6.1. Digital literacy and skills; 2.7.4. eGovernment. See *A Digital Agenda for Europe* 2010, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF>.

- expenses. It is up to the TCs to enable the TPs to enter the game (see the *tassa.li* case study below in Section 9.1.);
2. some new technological instruments that will make it easy for the TPs and TCs to enter and track the data: e.g., mobile phones with some sort of paying system connected to some sort of governmental expenses and tax tracking network;
 3. the possibility of accessing the Internet almost for free: the collected information are a value not only for the TCs, but for the whole society, so every TP needs to have a quick access to (some of) them in the fastest way. Until now, the best way to give data accessibility is the Internet¹¹.

7.3. Going Forward Legal Consequences

The legal upshots of these preconditions may be:

1. a clear statement of the right to access the Internet: it emerged from the previous point 3 in Section 7.2.;
2. new problems concerning privacy and the accessibility of financial data: the TS knows everything of the economy of a state. This means a great deal of information that leads to the following question: what is or should be accessible and what is not? What are the criteria to choose what has to be accessible and what is not;
3. new transparency standards for corporations, political parties, governments: TPs will be given the power of raising their voice when a government wastes public money or when corporations make speculations or fraudulent tax declarations.

Determining *a priori* the full spectrum of legal consequences is not possible and, in the rest of the paper, I will just hint at some of the points.

8. IMPLEMENTING THE APPLIED PRINCIPLE: DATA MINING AND DATABASE ORGANIZATION

The third presupposition of the applied principle is the one that will shape the organization of the TS. We can imagine the TS as a big spread-

¹¹ The physical access to the place where the data gained through the TS are stored is by far less efficient. It would be like obliging every people to go to the bank instead of using online services for paying.

sheet or excel file, accessible through the Internet. TPs send all the data to TCs, then TCs track everything.

The main problem, as the TS is able to provide a full database of the economic transactions, is: *who may access (and mine) what?* This requires to state better what a TP is and what information are associated to each TP's profile.

TPs are going to be all the agents of an economic environment: from individuals buying goods to corporations paying salaries or buying other corporations.

Every TP profile associates to the specific data (i.e., the money paid or received) any relevant data concerning who the TP is (proper name or corporation name), what are the earnings of the corporation or of the individual, education and profession of the individual (founding statute of the corporation and list of the employees for the corporation) and possibly other data.

These data will help the TCs to:

1. mine the data and recognize patterns of illicit activities (some corporations or workers could be more likely to evade taxes, we may find out what are the most used accounting tricks to make money disappear, etc.);
2. improve the general economy providing accurate data concerning (un)employment rate, average salaries compared with age or geographical area or type of job and, last but not least, balances of the corporations.

This will avoid a lot of "numbers wars" disputes between politicians and worker unions or between different political parties and, hopefully, may avoid financial bubbles.

A lot of problems arise at this point: does this mean that every company can see the names of its customers and all the items of the company that a particular customer has purchased? Does this mean that a company can know all the purchases of a TP as well as its name? Does it mean that Apple and Microsoft can see how much Google pays for its server maintenance¹²?

¹² If you read S. LEVY, *In the Plex. How Google Thinks, Works and Shapes Out Lives*, New York, Simon & Shuster, 2001, you will find out that this is one of the strategic data that Google cares the most not to reveal.

8.1. *Introducing the Asymmetry*

I think these problems can be solved by making the TS *asymmetrical* both in storing the information related to the different subjects (individual TP or corporation TP) and in giving access to the information collected by the TCs.

Concerning the asymmetry in storing the data, individual TPs contributions will always be anonymously registered by the other part of the TCs - i.e., by the other TP who sold the good or the service¹³. On one hand, the TC is the only kind of subject entitled to visualize the whole profile of all the transactions of any TP (full name, age and so on) for (tax) investigative purposes only.

The TCs can also elaborate statistics concerning relevant TP information and ought to produce the official statistics (unemployment rate, salaries by age, income compared to education and so on). These elaborations will be made publically available to everyone and the TCs will grant that, in elaborating the data, there will be no privacy infringement.

On the other hand, corporations and single TPs are not entitled to earn for free a full non anonymous customer profile for every TP, knowing that TP John Smith bought such and such from them in a year. A transaction between John Smith and a (Big) Corporation will be thus registered as “JS to BC” for John Smith but for the Big Corporation it will be a “General Customer to BC”¹⁴.

Inside the TS, accessible only to TCs, there is also the nominal transaction made by John Smith to the Big Corporation. This is needed in case John Smith complains about the services and wants his money back: before giving him his money, the Big Corporation needs to know that it was TP John Smith to buy the good and not somebody else, otherwise the Big Corporations will have to give money away automatically for every complaint they received.

¹³ Here I am assuming goods selling as the paradigm. Of course, when positive law requires something different (such as in some kinds of contracts) it is possible for both the seller and the buyer to register the transaction with the full name of the buying TP. A supermarket works and sells in a different way than a lawyer.

¹⁴ The same goes for shops or other commercial activities. This is to ensure that business activities are not given by the TS a full customer's profile for free: they have promotion cards for it.

Concerning the exchanges between corporations or shops, I think that they should be symmetrically recorded: the owner of shop A knows that shop B bought something from him and the other way round.

Concerning the asymmetry in giving access to the information collected by the TCs, I think full reports of units sold during a year by corporations should be public and accessible to all the citizens as well as corporate balances. This will provide reliable data for the economy and let everybody interested in the performances of a corporation know how many units were sold by a company, helping citizen and financial markets to have a better understanding of how firms are performing; I hope that knowing these information will lessen the frequency of financial bubbles. Collecting these data will also enable the state and the market to have more instruments to review corporations' budgets.

I think that, in order to keep free market going, the relevant information concerning strategic issues and for market competition between corporations have to be removed from these budgets (I think operative cost of infrastructure maintenance might be omitted, but here is where the problems will rise)¹⁵.

8.2. *The Mining Problem*

Another relevant legal issue is how far the different subjects (TCs, corporations and individual TPs) can go into mining the public available data. Should data be profiled by age, race, education, geolocalized and so on?

As I told before, I think that it is up to the TCs to produce the relevant statistical analysis of the collected data, granting both TPs' privacy and data accessibility.

I also think that TCs should make a public register of all the exchanges of a certain TP (both individuals and corporations) that shows how much money they have spent and received without any specification concerning the item bought or the sum of money received. I propose to call this open access database, the *anonymous in/out database* (the "anonymous" refers to the fact that, even if we know the name of the TP we are considering, we do

¹⁵ One of the most complicated issue concerns the (non) publicity of private deals such as the price of acquisition of a society or the sum received by an individual TP after suing and winning against a corporation or another TP. Hopefully, this will be ruled somewhere in the legal system the TS will be embedded into. Concerning companies, I think they can omit how many employees are working on a single project or department, but not their total employees' number.

not know the name of the other part in the selling and buying process, but just the sum exchanged)¹⁶.

Here problems arise: what other relevant information should be included and attached to the name of the single TP in this *anonymous in/out database*? This is a crucial point because what we will include here will also determine the information that could be mined.

My idea is that researches for every demographic category should be possible. What becomes problematic is to add more information concerning the TP status and to allow for more complicated searches. An example might help: find out what is the average monthly expense of a man or woman seems to be a question worth answering and it is positive to have a statistical answer to it. The statistic given by the TS is the result of the access to all the relevant data, so it is a very good one. It becomes problematic if you ask the same question but add to the data of the average expense an X for the age of the person, a Y for the region, a Z for education or, even worst, if you add a certain type of goods.

I think that this is a great problem for the development of a TS and it has to be ruled by both TCs and TPs. Whatever the criteria are going to be – the discussion is going to be on what should be the different types and combinations of anonymous information, trying to find the limit of whether anonymity is no longer warranted by the data¹⁷ – I think that there should be a public accessible database giving non-anonymous information on who run searches on the *anonymous in/out database*¹⁸. This will enable both TCs and TPs to spot any suspect query in the system and to locate possible queries resulting in non-anonymous results (as the foreign student example in note 17). These data will be used to have an evolution of the system and to keep controlling the TS.

¹⁶ The difference with what is inside the TC database is that they know both the seller, the buyer, the value of the transaction and the good exchanged. In the *anonymous in/out database* we know just one of the two between the seller and the buyer (it depends on the role that our considerer TP had in the transaction) and the value.

¹⁷ Imagine that you know that 1 student out of 10 comes from a foreign country. If a professor puts in his evaluation form a local/foreign box, filling the evaluation form will no longer be anonymous for the foreign student. The TS should avoid such situations.

¹⁸ I.e., there is a list of all the searches done by a TP on the TCs *anonymous in/out database*. This will enable you to know that the shampoo seller John Smith did some research on the wages of the TPs near his shop trying to sort out all the relevant data. I think it is useful to have such an archive in order to spot mining activities that may go against the privacy standards set inside the TS.

9. CASE STUDIES: RECORDED FUTURE AND TASSA.LI

Here I will present two case studies that are interesting to see how traceability systems have been employed and the issues they arose. The first regards money: I will present the “Tassa.li” smartphone application that was designed to construct a map of evaded taxes in Italy.

The second regards human behavior in general, but has relevant financial application. The software is called “Recorded Future” and its mission is to predict the future mining the data available on the Internet. This latter case study is important both for its financial applications (it seems that you can have bigger profits in trading if you just mine the data from Twitter) as well as for the issue of how far you can go in mining public data.

9.1. *Tassa.li*

Tassa.li¹⁹ is a smartphone app(lication) developed in Italy to cope with the problem of tax evasion (cafes not making the bills, workers performing jobs without giving you any receipt and so on). It employs a volunteer and non automatic system for tracing the money that is evaded. Tassa.li users voluntarily download the app and, whenever they do not receive the receipt or the bill, they – again voluntarily – submit a record of the evaded sum that is geolocalized and stored anonymously and displayed on the tassa.li website. As the site claims: “the data that we collect will be used to draw a tax evasion map, available to everybody on our website”.

Concerning anonymity, they say that “your name will never appear on the map, neither in any area of the website nor in any field of the application. The name of the tax evader will not be reported as well”²⁰.

Another important thing concerns the chance of other people knowing that we are participating into the “tassa.li game”. Instead of what happens inside the TS, “nobody will ever know that you are using tassa.li, unless you connected the tassa.li activity with Facebook or Twitter”. Even in this case, the app will never reveal private information and will only post on your social network wall a message saying “I reported a tax evasion case using tassa.li – take a look at the tax evasion map”.

¹⁹ See <http://tassa.li>.

²⁰ This and the following quotes are translations of the Frequently Asked Questions of tassa.li, <http://tassa.li/faq>.

This case shows that we need no tremendous computational power to obtain results, all we need is a way of integrating the processes of legal data acquisition into our daily life.

9.2. Recorded Future

Recorded Future - RF²¹ is a company that produces forecasts using public data mined from the web. You pay money to use the service and can ask their algorithm to perform a big variety of searches. RF idea is that the web has enough information to answer our concerns and enables us to make predictions about many different subjects.

Recorded Future is already doing something that the TCs in the model will do: they are mining publically available data, i.e., the whole web.

This is different from what the TCs will do because TCs will mine only some selected data (i.e., the one contributed by the TPs) but investigating RF may give us insights on the problems related with the mining of all the data of the *anonymous in/out database* produced by the TCs.

Before listing possible problems, there are differences between the two databases to be outlined:

1. information on the RF database are less controlled: not all the Internet users are well aware of the consequences of publically sharing information and the laws concerning the Internet are still to be fully figured out (think about the right to online oblivion or the right to access the Internet for free);
2. in the *anonymous in/out database* compiled by the TCs, all the users have a clear awareness of the kind of data they are contributing to and there is a clear positive law concerning its organization (there are different forbidden mining combination, the TCs will be responsible of privacy infringements when issuing their reports and so on).

With these differences in mind, we can get more information on how RF works by reading their blank paper on data mining²². I think these are the questions raised by the RF actions:

²¹ See <http://www.recordedfuture.com>.

²² S. TRUVÉ, *Big Data for the Future: Unlocking the Predictive Power of the Web*, 2011, <https://www.recordedfuture.com/assets/RF-White-Paper.pdf>.

1. what do we have to do if the data processed and stored through the algorithm are no longer on the Internet?²³;
2. what if RF algorithm comes to mine information that are secret or somehow highly protected (by this I mean something more than a book copyright, such as a patent of a product or a military secret)?;
3. how can RF impact on censorship?

I think that RF can claim that “what is public is public” and they have no limits in mining it. Concerning 3, RF may claim that the fact that they store information once online and kept them stored, no matter what happens to the original source, is a way to contrast censorship.

The upshots of debating the RF case may end up being relevant to questions such as: should there be someone able to control the public access of the data (for RF the Internet, for TS the accessibility of the anonymous database compiled by TCs and the log of the activities on it)?²⁴

10. SECURITY

The problem of security for a TS has to be faced on all the levels of security: physical security of the infrastructures, security in the code lines of the TS and security of its content (the TP data). I think that, concerning physical and software level, this enterprise has to face the same danger of storing anagraphical data or credit card data and it will be up to the law to provide extra warrants for the contents (and we have seen that, in sketching the asymmetries of the TS, we had the law granting some extra legal protection to the TPs’ privacy).

One of the biggest problem in realizing an effective TS is sociological: at least in Italy you need to win the fear of the web (people prefer to give their credit card to a waiter, giving him all the time to copy the data – credit card’s number and code –, then typing them into an online form, and make people

²³ It is not quite clear to me from the RF paper what they do in analyzing the different daily slices of the data. It seems that they record and track day by day. So, if you remove on day 2 what you wrote on day 1, they will keep having and tracing also the information of what you wrote on day 1 also on day 3, day 4, ..., day n.

²⁴ This relates to the questions of the web access as a human right and having UNO controlling the web. In his talk at Mobile World Congress 2012, Google’s CEO Eric Schmidt denies the need of United Nations to safeguard the Internet (<http://www.youtube.com/watch?4DKLSO8wYzk> at minute 36.20). I do not think that the introduction of TS and the *anonymous in/out* minable TC database will give reasons for a central control. On the contrary, it should give extra reasons for a transparent access to it.

more aware of what it is like to share information or access the Internet and what is it like to be a victim of social engineering)²⁵.

11. CONCLUSIONS ON THE TRACEABILITY SYSTEM

Having all transactions archived and available to the tax collectors will give an enormous data set that will enable the creation of a model that can be improved through time (an integration of *tassa.li* into RF may well be a first good hypothesis).

Of course, this will not solve the problem of evasion if money transactions are not registered. Nonetheless, provided the full implementations of the applied principle plus the elimination of paper money, the only way to have evaded taxation will be going back to barter or introducing parallel money²⁶.

The TS can be seen as a passive form of e-democracy: it teaches all the good practice of e-democracy (transparency and participation to the taxation game) but does not ask us to make a political choice (direct democracy: voting and proposing law).

The last question to be asked is: how far are we from the adoption of a TS for payment in our ordinary life? Unfortunately, this question deserves another paper to be addressed properly.

12. GENERAL CONCLUSIONS

In this paper I provided a methodological framework – NAF – in which computational social science can help legal science to be more effective and obtain control on how the law is performing in the world.

In the second part I investigated a possible application of the framework dealing with the development of a traceability system for payments. I think that the NAF framework can be further extended with the implementation of tools and conceptual resources from sociology, psychology and economics.

²⁵ For an introduction of how a social engineering works you can see K.D. MITNICK, *The Art of Deception*, Indiana, Wiley, 2002.

²⁶ There may be parallel money used to escape the TS, but it will be difficult for the parallel money to spread without being recognized and introduced into the TS. Maybe non durable goods may be used as such money. On this see T. KAWAGOE, *Can Chocolate Be Money As a Medium of Exchange? Belief Learning vs. Reinforcement*, in “Learning”, 2009, n. 5, pp. 279-292.