Towards a Reference Architecture for Trusted Data Marketplaces

~ The Credit Scoring Perspective ~

Dumitru Roman
SINTEF
Pb. 124 Blindern, 0314
Oslo, Norway
Dumitru.Roman@sintef.no

Gatti Stefano
Cerved Group S.p.A
Via San Vigilio n.1, 20142
Milan, Italy
Stefano.Gatti@cerved.com

***Abstract*—Data sharing presents extensive opportunities and challenges in domains such as the public sector, health care and financial services. This paper introduces the concept of "trusted data marketplaces" as a mechanism for enabling trusted sharing of data. It takes *credit scoring*—an essential mechanism of the entire world-economic environment, determining access for companies and individuals to credit and the terms under which credit is provisioned—as an example for the realization of the trusted data marketplaces concept. This paper looks at credit scoring from a data perspective, analyzing current shortcomings in the use and sharing of data for credit scoring, and outlining a conceptual framework in terms of a trusted data marketplace to overcome the identified shortcomings. The contribution of this paper is two-fold: (1) identify and discuss the core data issues that hinder innovation in credit scoring; (2) propose a conceptual architecture for trusted data marketplaces for credit scoring in order to serve as a reference architecture for the implementation of future credit scoring systems. The architecture is generic and can be adopted in other domains where data sharing is of high relevance.**

Keywords—credit scoring, trusted data sharing, data marketplaces, blockchain, smart contracts, homomorphic encryption, multi-party computation

I. Towards a New Data Sharing Paradigm

The World Bank reports in its last Global Findex Database [1] that about half of the adults globally have no access to banks and financial services. For all of them there are very limited options to borrow, save, insure and spend, and therefore access the global economy. The situation remains quite uncomfortable also in countries with developed economies. Many borrowers remain unable to access credit due to "thin files" or limited credit bureau coverage. Current credit risk assessment approaches are too heavily dependent on historical financial data inputs for the consumer. For many individuals globally these inputs are simply unavailable or not enough to allow a correct evaluation and decision. In both high income and developing markets, this gap is a strong obstacle to consumer credit and one impediment to financial inclusion.

Over the last years, new alternative approaches have been developed to substitute traditional measures of credit risk with new ones that could bypass the information gap and ensure a fair and correct evaluation for everyone. "Alternative data" [2] coming from the digitalization of our society such as social networks data, phone calls information, prepaid cards transactions, are now largely used for new approaches to risk assessment. More generally, personal data has become a new currency in our digital era. The greater part of our life is now encoded into bits, data about us continue to grow day by day – data on who we are, what we do, where we live and work, our history, our friends. All these data could now be collected, managed and combined to produce a surprisingly complete picture of us.

In an overall problematic European economy, a new paradigm based on data could enable new value and lead the growth for both public and private sector organizations[1]. It seems clear that this new economic paradigm will exceed the current discussions about the monetization of personal data on players such as Google or Facebook. The data-driven economy affects our entire society and is relevant, and important for the whole economic system at global level. Areas such as financial services, the public sector, and heath could potentially realize significant value by properly leveraging the adoption of the data-driven economy.

Six main trends have been identified as key to the adoption of the new data-driven economy paradigm [3]. These include aspects related to *automation* (many of the current relationships and transactions could be improved and speeded up by integrating personal data into organizational processes and decisions); *autonomy* (users have the argued freedom being able to perform transactions autonomously, without any interaction with organizations or other people); *customization* (product and services can be tailored on particular individual's preferences and needs); *delivery* (operational insights could be easily extracted from the new data lakes enabling organizations to increasingly improve the delivery of goods and services); *data-driven R&D* (companies could leverage on a wide range of new data to continuously review their products and services); and *monetization* (organizations now recognize that

[1] https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy

978-1-5090-4054-4/16 $31.00 © 2016 IEEE
DOI 10.1109/OBD.2016.21

data they hold is valuable to other parties – if data is the new currency then it could be used in transactions).

With the new data-driven economy paradigm recognized as the way forward for both organizations and individuals there are still many unaddressed concerns about data usages. Various studies highlight that very few individuals are fully aware and in control of their data and many of them are worried about how this information can be used by companies. For example, the Eurobarometer 2015[2] reports that 81% of Europeans feel that they do not have complete control over their personal data online and the majority (69%) would like to give their explicit approval before the collection and processing of their personal data. Two key concepts are emerging as essential to enable the data-driven economy to keep data flowing as the new currency [3]:

- *Awareness*: The more awareness individuals have regarding their data, the more they are willing to share them and the more they require in terms of associated benefits. Individuals with higher awareness require 26% more benefit in return for sharing their data[3]. Additionally, awareness is key to ensure organizations will be able to leverage on data avoiding usage that is not fully recognized and approved.
- *Control*: Privacy, security and a full and easy control of data increase the willingness of individuals to share them. Individuals who are confident that they have full control over their data and a secure means to preserve their privacy are more willing to share their data with third parties.

As data are increasingly becoming the fuel of algorithms that are key in almost every process of companies, and play a role even in individuals' everyday life, a new awareness of the importance of how these data are managed and used is growing all over the world. The widespread opinion that "we are experiencing the largest transformation since the end of the Second World War" carries a clear message, as described in [4]: "after the automation of production and the creation of the self-driving cars the automation of society is next. With this, society is at a crossroads, which promises great opportunities, but also considerable risks. If we take the wrong decisions it could threaten our greatest historical achievements". In this landscape, one of the main legislations that is trying to address these issues is the new (May 2016) European data protection directive[4], which proposed a comprehensive reform on how data protection is managed in all European countries. The objective of this new set of rules is to give back citizens control over their personal data and to simplify the regulatory environment for business. Key points from the directive relevant to our context include:

- The "right to be forgotten": when it is no longer necessary to store personal data and there are no legitimate grounds for retaining it, the data should be deleted by the data owner with an easy access to them. Only in a trusted framework could this goal be easily reached, by turning back to the individuals the keys of their data, rather than leaving redundant copies in various unregulated data centers all over the world.
- The two principles of "data protection by design" and "data protection by default" come also in the center of this reform. This means that especially during the early stages of development of a service or product, the protection framework has to be taken into account in an explicit way (including the introduction of a data protection officer in companies) to ensure that, by default, in the initial settings, the most restrictive data-policies are applied, thus giving citizens the possibility to actively lift the restrictions later.
- The principle of "data-portability", which means the freedom to transfer easily personal data from one service provider to another.
- The principle of "one single law" applicable across the EU, but also to be respected by all the companies located outside the EU who market services and products to EU citizens. This should improve the data and algorithms market that is dominated worldwide by a few large corporations.

Data and their power to enable a new economic paradigm represent a great opportunity when properly and carefully managed. At the same time, there are huge risks in case the involved stakeholders are not able to create a fair and trusted environment to enable the use of data as the new currency. To unlock the value of data, a new business model needs to be set up, with a formal transactional process across the entities or individuals who own the data and the ones who want to use it. Organisations have to make clear and formalise each specific type of data usage and the benefits that they are ready to offer to the data owner as counterpart of the transaction. Organisations need to communicate and be open about how exactly they acquire and use data. The new approach needs to be supported by a trusted environment recognized by all users as the place to enable transactions within the new economic model. This environment will need to ensure privacy, security and a robust control of each transaction agreement. To help organisations participate in such environment several aspects need to be addressed, including:

- *Support the economic paradigm shift with a cultural one*: organisations and individuals need to cooperate to establish new standards and adopt them, including aspects related to clear and formalized benefits for data owners, how and where data can be used, etc.
- *Create a common trusted data marketplace*: a trusted and safe environment where data could be used as a currency to close transactions across multiple entities as is, for example, the current setup in traditional currencies. Every business and public organization should cooperate in identifying guiding principles and standards for how the new environment should work. Rules and principles that need to be fully recognized

[2] http://ec.europa.eu/justice/data-protection/files/data-protection-big-data_factsheet_web_en.pdf
[3] https://www.ftc.gov/reports/section-319-fair-accurate-credit-transactions-act-2003-fifth-interim-federal-trade
[4] http://ec.europa.eu/justice/data-protection

and accepted by the market taking commitment for the creation of a new standard.

- *Ensure security and privacy in order to safeguard data*: data breaches have a direct cost for both individuals and organizations. While on the one side individuals foresee the risk to not have authorized access to their personal data, conversely, organizations are worried about operational, legal and reputational risks. A trusted data marketplace will be enabled not just by technologies – it will need a legal framework and a set of policies to define the highest level of security.
- *Build a data-driven society using data in every interaction*: data could drastically change our entire social model – not only relationships across individuals and organizations, but also interactions across business partnerships will be affected. Human dynamics and company exchanges are already strongly affected by data and the use of them to enable or improve the relationships.

Based on the above discussion on opportunities and challenges in sharing of data, and outline of the areas of relevance for a new data sharing paradigm, in this paper we look in more detail into concrete issues related to the use and sharing of data for credit rating purposes (Section II). Furthermore, we propose a reference architecture for a technical infrastructure for trusted sharing and access to relevant data in terms of a "trusted data marketplace" for credit scoring (Section III). Finally, we outline directions for realization of such a data marketplace (Section IV).

II. CREDIT SCORING: A DATA PERSPECTIVE

Credit scoring[5] is a critical process in the entire world-economic environment because it regulates:

- the way (either in quality or quantity) banks and financial services lend money to companies and individuals, and
- the level of trust that causes companies to do more or less commercial transactions with each other.

The accuracy of credit scoring, and of the ensuing trust level, is crucial in speeding up the economic system and in supporting its growth while making it more stable and reliable. Finding new ways to enable better and more objective credit scoring will be helpful for all the key stakeholders: banks, financial services, companies, individuals, and policy makers too.

The state of the art in credit scoring is based on various sources of data influenced by different factors such as the item scored (companies or individuals), national laws on mandatory data publication, and local privacy laws. For example, there are countries that impose legal obligations on companies to make the balance sheets publicly available (and even among the companies the practice widely varies), but other countries do not require this. For instance, in the UK all balance sheets are updated daily and published in an open data format on the site of the UK public administration. On the other hand, in the US, except for the Securities and Exchange Commission (SEC), it is not easy to have access to these data. In Switzerland, with the exception of banks, insurance companies and companies traded on the stock exchange, there is no obligation to file any information at all. In addition, the access to contextual data (e.g., cadastral data on corporate properties, geo data, personal data about directors and shareholders, etc.), which are the foundation of credit scoring, depends strongly on national laws.

In particular, the history and kind of payments – the most important data for credit scoring – have different levels of access for different stakeholders of the credit scoring process. In the Anglo-Saxon countries, there is a stronger tendency to share the data from the scored subject to stakeholder and to credit scoring providers. Recently, successful start-up companies in the US operating in the domain of credit scoring (e.g., CreditKarma[6]) bring evidence that sharing their historical payments data helps individuals monitor and improve their credit score and gain access to better services (financial and others). In Western Europe, the privacy laws are stricter and the access to the historical payments of individuals, is reserved only to a few authorized stakeholders for special purpose. Furthermore, the data of economic transactions between companies are rarely shared for the credit scoring process, either for cultural reasons, or for fear of spreading information about customer bases of companies.

Typically, the credit score computation is event-driven with algorithms that have to be approved by regulatory organizations, and rarely take advantage of data from unauthoritative sources.

Given the importance of credit scoring and the current state of the art, in the following paragraphs we list issues that could be overcome by the new technology stream.

Data delay. With the increase in speed and complexity of the world commercial transactions, timely access to relevant data can make credit scoring more effective. Waiting for official data that is slowed down by bureaucracy or other unwanted delays of public administration, makes the credit scoring ineffective and unrepresentative of the real world. For example, balance sheets contain highly relevant data for credit scoring, but they display the economic situation with a delay of several months.

Distrust in proprietary data sharing. The fear of sharing complete control over the data for credit scoring, unclear benefits of sharing the data, and, very often, the opacity of credit score algorithms, are the main reasons that have brought many companies and individuals to the situation of being unwilling to share their data unless obliged by law.

Lack of data (particularly on small and micro-enterprises). In the SMEs segment, there are not enough economic official data available, and often the best data are unofficial or approximated based on statistical dimensions (such as number

[5] For the purpose of this paper we use "credit scoring" as a general term for capturing the creditworthiness of both businesses and individuals (traditionally referred to as credit rating and credit score, respectively).

[6] https://www.creditkarma.com/

of employees, share capital and so on). For example, in Italy only one million active companies (less than 20% of all active companies) have public balance sheet data available, while the others, about five millions, have no official economic figures (including, for instance, data on the total yearly turnover).

Overcoming the above data-related issues could open up various opportunities in many aspects of the credit scoring chain, in particular:

- Streamline and speed-up many processes where credit scoring is crucial in transactions, such as home mortgage and insurance policy.
- Free more credit money from bank and financial services to companies, on the one side lowering the risk of non-performing loans (NPL), and on the other, improving transparency in lending. The NPL is an important issue, especially in Western Europe where the rate of NPL over the total bank credit has dramatically increased over the last 10 years with some countries like Italy, Portugal, and Greece going over 10%[7].
- Enable innovative credit access service, especially in the micro-finance sector, see for example [5] – a 2015 study by the European Microfinance Network – which argues that that new technologies and "Big Data can assist lenders by using social network data, psychometric principles and mobile phone payment history to easily create credit identities and customer scores in real time". A 2014 research work [6][8] calls for sharing of social data as a mechanism to assess and improve credit-worthiness of individuals who, due to the bureaucracy and the financial system and old rules that are simply not adapting to the new environment, have no good mechanisms to communicate credit-worthiness information to the banking system. New credit-risk models for the unbanked could be developed as underlined in a McKinsey study from 2012 [7], which, besides pointing out six new data sources for credit scoring (telecom providers, utilities, wholesale suppliers, retailers, government, financial institutions' own previously overlooked data), argues for strong privacy frameworks for lenders and borrowers as a key enabler for innovation in credit scoring.

An ecosystem for provisioning technical infrastructure for trusted sharing and access to relevant data for credit scoring, mechanisms for reducing the delay to access data, and more transparent credit scoring algorithms are essential aspects in creating an environment where stakeholders are willing to share data in a timely manner. This will, in turn, provide incentives for more availability of data, making the credit scoring process more objective, effective, and ultimately more helpful for all its stakeholders. We propose a reference architecture for enabling such an ecosystem, by arguing for the need of a "trusted data marketplace" for data relevant to credit scoring – a concept which we define in the next section.

III. Reference Architecture for Trusted Data Marketplaces

A. *Data Marketplaces for Credit Scoring*

Data have increasingly become valuable assets that can be sold and bought as a commodity. Data marketplaces [8,9] have recently emerged as virtual places where data providers and data consumers trade data. Data marketplaces can offer a wider range of capabilities, ranging from data gathering, aggregation, integration, processing, enrichment, to buying and selling the data. Data marketplaces have become popular in domains such as trading of personal data (i.e., personal data markets [10], such as Datacoup[9]), and open data (i.e., open data marketplaces [11] such as proDataMarket[10]).

We see the data marketplace paradigm as an appropriate model for credit scoring related data because it has the potential to provide:

- A common point of entry for both providers and consumers to discover and compare credit scoring related data, along with indicators of quality and scope.
- A uniform way to handle the access and quality of credit scoring related data and make it ready for use in algorithms for computing credit scores.
- An environment where data providers can disseminate their data at a larger scale and enable them to monetize their data.
- An economic model for broad access to credit scoring related data.

B. *Enabling Trust in Data Marketplaces for Credit Scoring*

All of the existing data marketplaces are essentially centralized systems, where participants in the marketplace (data providers and consumers) have to trust a third party – the data marketplace provider/operator – with managing their data. Typically access to data on a marketplace is safeguarded by a set of terms and conditions and privacy policies, often rather vague, unclear, and difficult to understand by data providers and consumers, leaving data providers with little control over their own data (e.g., with whom the data are shared, under which conditions, how it is monetized, etc.). The guarantees current data marketplace players receive give them little confidence that data recipients will really treat the data they receive in the way they promise. For these reasons, trust in data marketplaces in their current form is a major issue of concern[11].

Recently, some technical approaches have emerged on how to address this issue. For example, Maguire et al. [13] proposed to embed more trust and accountability in the data-sharing ecosystem, by using a metadata based architecture for user-

[7] https://www.forexinfo.it/Crediti-deteriorati-il-problema (in Italian)
[8] http://knowledge.wharton.upenn.edu/article/using-social-media-for-credit-scoring/

[9] https://datacoup.com/
[10] http://prodatamarket.eu/
[11] For example, a recent study [12] suggests that precisely because of the trust issues on how data are shared, personal data markets in their current form will have difficulties to find acceptance among people.

centered data accountability, binding policies and permissions negotiated with users to the data that are being collected from them. Such approaches represent a step forward in addressing the trust issue; however, they fall short of addressing the core trust aspect related to the centralised nature of existing data marketplaces. Credit scoring related data can both be highly sensitive and economically valuable data. A solution that does not ensure full control over the data for the owners and strong guarantees for how the data are used, cannot be accepted in a credit-scoring ecosystem.

A set of emerging technologies in the context of cryptography, cloud, and decentralized computing, such as Blockchain, Smart Contracts, homomorphic encryption, and multi-party computation, offer a unique opportunity for the creation of a trusted ecosystem where large-scale data sharing can be enabled between credit scoring stakeholders. Below we present the reference architecture for Trusted Data Marketplaces and discuss the underlying technologies that could address the issues we identified in Section II.

C. Reference Architecture for Trusted Data Marketplace: The Credit Scoring Perspective

The underlying principle in the design of the reference architecture is the ability of different parties to jointly store and run computations on data while keeping the data completely private.

Figure 1 shows the overall reference architecture, which, together with the underlying enabling technologies, is described below. The key stakeholders in the marketplace are as follows:

- Companies and Individuals (left side of the figure) that own credit scoring relevant data and need access to credit.
- Information Providers (top part of the figure) representing organizations such as credit rating and scoring companies that want to run algorithms on data provided by companies and individuals, and provide the results of the algorithms to third parties (e.g., banks or insurance companies).
- Credit Scoring Users (right side of the figure) representing credit providers (e.g., banks), which in addition to accessing credit rating information may run various analytics algorithms (e.g., for decision making purposes) on the data owned by companies and individuals.

The *Orchestrator* component provides the companies and individuals with the ability to enable access to their data in an encrypted form (using homomorphic encryption). In the general case, this component will consist of mechanisms to (1) connect/integrate their data (e.g., in the case of companies data from internal systems such as ERP systems, via APIs); and (2) encrypt the data before it is made available for processing to third parties on the Encrypted Cloud (this is typically done through ETL processes).

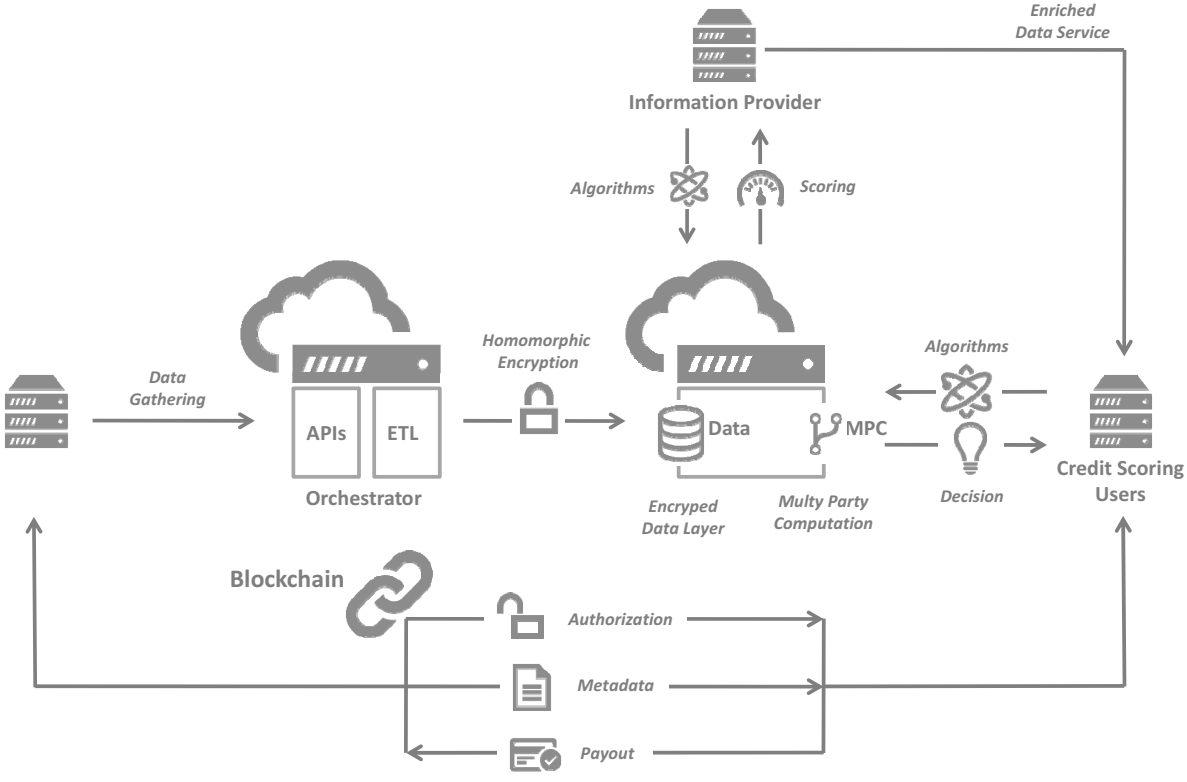

Figure 1. Reference Architecture for Trusted Data Marketplaces – the Credit Scoring Perspective

The *Encrypted Cloud* component stores the data in encrypted form. It offers the data "seller" a safe space for data where unauthorised parties are not able to access it. Algorithms run on the encrypted data, and due to the multi-party computation engine part of this component, the algorithms cannot be reverse-engineered.

Companies and individuals will "sell/share" data signing Smart Contracts on the *Blockchain* with third parties (banks, utilities, insurances, etc.) interested to run algorithms on their data. The payout could be direct, such as a payment, or a service – e.g., access to credit or better conditions for products and services.

A typical example to be supported by such an architecture would be a company that wants to access credit and for that purpose it shares its ERP data with a credit scoring company and a bank. The data will be made available in the data marketplace in an encrypted form; Smart Contracts will be signed with the credit score provider and the bank specifying that only they can have access to those data and only for obtaining a given credit. With the Smart Contracts being executed on the Blockchain it is easier to enforce the contract and monitor its execution.

This architecture is made possible by a set of enabling technologies briefly described in the following.

D. Enabling Technologies

Homomorphic Encryption (HE). "Homomorphic" denotes a system's ability to perform computations on the cipher-text without decrypting it first. The idea behind homomorphic encryption is relatively simple. With most encryption schemes, the encrypted data has to be decrypted entirely before any significant work – e.g., mathematical or programming operations – can be done on it. HE, on the other hand, enables math directly on the encrypted data and the results of that math reflect the underlying data. Recent scientific achievements in this field have shown that homomorphic encryption is actually possible and feasible.

Multi Party Computing (MPC). MPC deals with computational methods in which different parties jointly compute a function over their inputs while keeping those inputs private. It is most widely applied in the area of privacy-preserving data mining: running data mining algorithms on confidential data that is not supposed to be revealed – even to the party running the algorithm. There are two classic settings for privacy-preserving data mining, although these are by no means the only ones. In the first, the data are divided among two or more different parties, the aim being to run a data-mining algorithm on the union of the parties' databases without allowing any party to view another individual's private data. In the second, some statistical data that is to be released so that it can be used for research using statistics and/or data mining. It may contain confidential data, hence, it is first modified so that (a) the data does not compromise anyone's privacy, and (b) it is still possible to obtain meaningful results by running data mining algorithms on the modified data set.

Smart Contracts. Smart Contracts are self-executing contractual states, stored on the Blockchain. Smart Contracts are computer programs that can automatically execute the terms of a contract. In the future, these programs may replace lawyers and banks for handling certain common financial transactions. The potential for Smart Contracts goes way beyond simple transfers of funds. The door of a car or a house could be unlocked by connecting Smart Contracts to the Internet of Things. Smart contract technology is nowadays being built on top of Bitcoin and other virtual currencies. Ethereum is an example of entirely new currency with Smart Contracts baked into its payment system.[12]

The use of Blockchains for credit rating was advocated in [14], where the author of that paper argues that credit risk modelling can be improved through increased trust and better timing of accounting data releases.

IV. OUTLOOK

This paper discussed various challenges and opportunities related to sharing of data, and specifically concrete issues related to use and access to data for credit scoring purposes. We identified main data-related challenges that hinder innovation in this space, namely delayed access to data, distrust in data sharing, and as a consequence, lack of data, which all together contribute to the inefficiencies in the current credit scoring process. To address those issues, we argued for the development of a Trusted Data Marketplace for credit scoring, and outlined a conceptual architecture, together with the enabling technologies for such a marketplace that can be used as a reference for future implementation of credit scoring ecosystems.

Next steps in the realization of such an architecture include:

- *Integration of enabling technologies*. The concept proposed in this paper relies on the integration of existing, and some emerging technologies. How to best fit them together is not a trivial task that requires further research, development and integration work.
- *Understanding the need for cultural change in decision making for credit scoring stakeholders*. The emergence of a technology set to enable trusted data sharing may disrupt existing decision-making processes, and therefore some cultural changes will be necessary. Understanding the social aspects of the need for such a change and how existing processes will have to adapt in the future is an area that requires further investigation.
- *Identifying concrete technologies to work with.* Evaluating concrete technologies that could fit in the proposed architecture is a time consuming task. We are currently investigating for example Enigma [15] – a decentralized computation platform with focus on privacy, and ZeroDB [16] as potential concrete technologies that could support the realization of the architecture.

It is worth noting that whereas the architecture was developed specifically for dealing with data that is both highly

[12] https://ethereum.org/

sensitive and of high economic value in the domain of credit scoring, the architecture is generic and is applicable to other domains that have strong requirements for data privacy and timely access to data. For example, business lead generation – the process of finding new customers – is another domain where such an architecture could be of high relevance, gaining better efficiency by leveraging the trusted sharing and use of Web and personal data.

V. Disclaimer and Acknowledgment

The views in this paper are those of the authors and do not necessarily reflect the official policy or position of their employers. This work is partly supported by the proDataMarket project (H2020 644497).

References

[1] Demirgüç-Kunt, A., Klapper, L. F., Singer, D., & Van Oudheusden, P. (2015). The Global Findex Database 2014: measuring financial inclusion around the world. World Bank Policy Research Working Paper, (7255).

[2] Michael J. McEvoy. Enabling financial inclusion through "alternative data". Available via http://compendium.mastercard.com/app/SKU_pdfs/alternativeData.pdf.

[3] Rose, J., Rehse, O., & Röber, B. (2012). The value of our digital identity. Boston Cons. Gr. Available via http://www.libertyglobal.com/PDF/public-policy/The-Value-of-Our-Dig ital-Identity.pdf.

[4] D. Helbing, B. S. Frey, G. Gigerenzer, E. Hafen, M. Hagner, Y. Hofstetter, J. v. d. Hoven, R. V. Zicari, A. Zwitter. Behavioural Control or Digital Democracy? - A Digital Manifesto. June 2016. Available via https://www.researchgate.net/publication/303813069_Behavioural_Control_or_Digital_Democracy_-_A_Digital_Manifesto.

[5] Credit Scoring in the European (Micro)finance Sector, by the European Microfinance Network, April 2015. Available via http://www.european-microfinance.org/docs/activities/think_tanks_idea_labs/SCORING_EMN2015_LINKS_2.pdf.

[6] Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit Scoring with Social Network Data. Marketing Science. Available via http://ssrn.com/abstract=2475265.

[7] Baer, T., Goland, T., & Schiff, R. (2012). New credit-risk models for the unbanked. 2013-04]. Available via https://www.mckinsey.de/sites/mck_files/files/30_New_credit_risk_models_for_the_unbanked.pdf.

[8] Stahl, Florian; Schomm, Fabian; Vossen, Gottfried (2014) : The data marketplace survey revisited, Working Papers, ERCIS - European Research Center for Information Systems, No. 18. Available via https://www.econstor.eu/dspace/bitstream/10419/94187/1/779859502.pdf.

[9] Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). A classification framework for data marketplaces. Vietnam Journal of Computer Science, 1-7. Available via http://link.springer.com/article/10.1007%2Fs40595-016-0064-2.

[10] Spiekermann, S., Böhme, R., Acquisti, A., & Hui, K. L. (2015). Personal data markets. Electronic Markets, 25(2), 91-93.

[11] Zuiderwijk, A., Loukis, E., Alexopoulos, C., Janssen, M., & Jeffery, K. (2014, May). Elements for the development of an open data marketplace. In Conference for E-Democracy and Open Governement (p. 309). Available via http://duweb.donau-uni.ac.at/imperia/md/content/department/gpa/zeg/bilder/cedem/cedem14/cedem14_proceedings.pdf#page=310.

[12] Roeber, B., Rehse, O., Knorrek, R., & Thomsen, B. (2015). Personal data: how context shapes consumers' data sharing with organizations from various sectors. Electronic Markets, 25 (2). doi: 10.1007/s12525-015-0183-0.

[13] Maguire, S., Friedberg, J., Nguyen, M.-H. C., & Haynes, P. (2015). A metadata-based architecture for user-centered data accountability. Electronic Markets, 25 (2). doi: 10.1007/s12525-015-0184-z.

[14] Byström, H. (2016). Blockchains, Real-Time Accounting and the Future of Credit Risk Modeling. Working Paper/Department of Economics, School of Economics and Management, Lund University, (4). Available via http://project.nek.lu.se/publications/workpap/papers/wp16_4.pdf.

[15] Zyskind, G., Nathan, O. and Pentland, A., 2015. Enigma: Decentralized Computation Platform with Guaranteed Privacy. arXiv preprint arXiv:1506.03471. Available via http://arxiv.org/abs/1506.03471.

[16] Egorov, M., Wilkison, M., 2016. ZeroDB white paper. arXiv preprint arXiv:1602.07168. Available via http://arxiv.org/abs/1602.0716.