Amo, D. [et al.]. Personal data broker instead of blockchain for students' data privacy assurance. "Advances in intelligent systems and computing", 2019, vol. 932, p. 371-380.
The final authenticated version is available online at https://doi.org/10.1007/978-3-030-16187-3_36

Personal Data Broker instead of Blockchain for students' data privacy assurance

Daniel Amo[1], David Fonseca[1], Marc Alier[2], Francisco José García-Peñalvo[3], María José Casañ[2]

[1] La Salle, Universitat Ramón Llull, Spain
[2] Universitat Politècnica de Catalunya, Spain
[3] Universidad de Salamanca, Spain
daniel.amo@salle.url.edu, fonsi@salle.url.edu, marc.alier@upc.edu, fgarcia@usal.es, mjcasany@essi.upc.edu

Abstract. Data logs about learning activities are being recorded at a growing pace due to the adoption and evolution of educational technologies (Edtech). Data analytics has entered the field of education under the name of learning analytics. Data analytics can provide insights that can be used to enhance learning activities for educational stakeholders, as well as helping online learning applications providers to enhance their services. However, despite the goodwill in the use of Edtech, some service providers use it as a means to collect private data about the students for their own interests and benefits. This is showcased in recent cases seen in media of bad use of students' personal information. This growth in cases is due to the recent tightening in data privacy regulations, especially in the EU. The students or their parents should be the owners of the information about them and their learning activities online. Thus they should have the right tools to control how their information is accessed and for what purposes. Currently, there is no technological solution to prevent leaks or the misuse of data about the students or their activity. It seems appropriate to try to solve it from an automation technology perspective. In this paper, we consider the use of Blockchain technologies as a possible basis for a solution to this problem. Our analysis indicates that the Blockchain is not a suitable solution. Finally, we propose a cloud-based solution with a central personal point of management that we have called Personal Data Broker.

Keywords: Blockchain, Smart Contracts, Learning Analytics, Educational Data Mining, Academic Analytics, Data Privacy, Digital Identity, Moodle.

1 Introduction

Learning Analytics has become a key tool for assessment [1]. The students' interactions within online learning environments are collected, processed by statistical models, and finally presented to teachers and other stakeholders. These results are growing in detail and complexity. The student's personal data is generated from interactions in Learning Management Systems (LMS) [2, 3]. That data is analyzed by

algorithms and afterwards visualized by stakeholders in dashboards [4]. This helps to provide insights about behavior and learning needs [5]. These enhanced results enable teachers to better understand the progress of students and other related educational context aspects [6], such as validity of resources or even assignments. Hence, analytics have become essential to understand the students, get actionable recommendations [7] or even predict patterns of learning [8].

This paper is structured in four further sections. In section 2, we introduce the objectives, the research question and the methodology used. In section 3 we introduce the fundamentals and argument that Blockchain is a novel technology which still has some security flaws that makes it an unreliable protocol in terms of data privacy. In Section 4, we introduce an alternative solution to ensure students' data privacy and its architecture in the cloud, which we call Personal Data Broker. Section 5 concludes and closes the paper and describes future works.

2 Methodology

In the Learning Analytics process, different tools are used to collect private, personal and highly sensitive data from students, even from minors. This data collection generates privacy issues related to data leakages and misuses. The collected data can be stored in unknown servers, making it possible for administrators without legal permissions to access them. Moreover, even if the logs could be stored in the same students' institution, the data inside the logs cannot be trusted due to possible alterations by system administrators. This context generates fears against the use of educational analytical approximations. Some data misuses such as inBloom schools case [9] highlight the importance of respecting students' privacy. They also manifest the urgency of finding a definitive and global solution to student data protection.

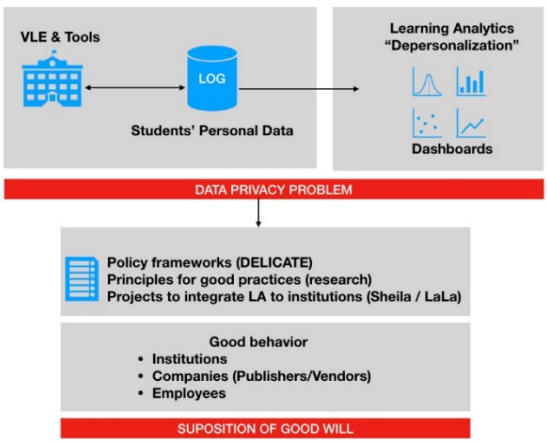

Fig. 1. Data privacy problem treated by law under supposition of good will.

Current solutions are delivered as policies, such as General Data Protection Regulation (GDPR) laws and regulations [10]. There are also research results in form of frameworks, good practice guides or principles [11–14]. Policies, regulations or guidelines do not correct the problem from its root because they do not prevent data leakages or misuses in real time. They are only applied when the problem has occurred and someone has reported it. Moreover, policies assume that the intentions of each of the stakeholder are in the interests of the students –see figure 1-, but reality shows that this is not always the case [5, 9]. Laws, regulations, and principles are not enough to assure data privacy in any kind of educational data log created by Learning Analytics, Academic Analytics, Educational Data Mining -or even other no analytical tools-. Therefore, there is a fragility in educational data privacy that needs to be addressed.

This situation can be solved by support the legislation with a layer of automation technology to enforce their policies and regulations in real time. This new approach could safeguard the privacy of students' data in its current and future uses. Some emerging technologies such as Blockchain seem to be strong candidates to ensure privacy and secure sensible data of students. The use of Smart Contracts inside Blockchain allows the automation of legal actions. They could be executed as soon as irregularities are detected in the use or collection of data. However, there is an ongoing academic debate and growing research on the security and privacy of the Blockchain [15–18]. The Blockchain approximation might not be ready to ensure data privacy in educational contexts.

Our hypothesis is that Blockchain is not the best technology to protect and manage safe access to personal data in the educational context. Consequently, the main purpose of this manuscript is double. On the one hand, we show the limitations of Blockchain as a possible solution to the problem posed. On the other hand, we present an alternative solution to Blockchain.

To achieve the objectives and validate the hypothesis, we propose to expose the weaknesses of Blockchain related to the privacy and data security dimensions in the educational context. As a guide to formulating a suitable methodology, we launch the research question: Is the Blockchain a suitable technology to ensure data privacy for students? We find publications that are motivated by the upward trend of Blockchain's use, advocate the use of it in education [19]. However, there are studies based on the very foundations of Blockchain that question it as a private and secure technology [20–24]. All of them repeat a series of security flaws that can be included in different categories: Identity filtering, Lack of transactional privacy, Private Key insecurity, Vulnerabilities in Smart Contracts, Discovery of user identities and Quantum computing.

3 Blockchain fundamentals

Satoshi Nakamoto [25] announced a cryptographic solution to a number of game theory problems. This would also allow the creation of a peer-to-peer electronic cash

system. After that, Nakamoto open sourced the implementation of the solution for the cryptocurrency Bitcoin introducing Blockchain as a cryptographic networked platform. Since then, Blockchain has attracted different opinions in education. This platform implemented a distributed ledger managed by consensus with a four core characteristics [26, 27]. Such characteristics are both the strengths and weaknesses of Blockchain:

1. **Immutability**: When a block is added, it cannot be altered. This data persistence and irreversibility creates trust in the transaction ledger.
2. **Decentralized**: Each user of the network has a copy of all the transactions. All the data is distributed through the network and decentralized.
3. **Consensus Driven**: Data is validated through cryptographic proof-of-work. Hence, no central authority is needed nor trusted.
4. **Transparent**: All transactions are public and history records are accessed by anyone in the network. Hence, users' data is public.

3.1 Consensus and permission-less

Blockchain is permission-less by default. No one needs a special permission, nor a central authority, to access the Blockchain users' network. In a context with untrusted users, where data is decentralized and no central authority is required, a solution was needed to assure data integrity. Blockchain solved this problem. Each transaction in the Blockchain must be validated by consensus [25]. A network of peers that perform cryptographic calculations achieves this consensus when each peer validates the transaction by proof-of-work. The proof-of-work consists in solving a cryptographic algorithm using intensive computation power. If the majority of the network that has completed the proof of work says a transaction is valid, then it is. The peers validating transactions are called "miners" because they get monetary rewards depending on the computing power they commit – linked to the energy they consume – and randomness.

Although Blockchain seems to be a secure and private technology, there are some factors that support the criticism and skepticism it has received in education that make it unsuitable as a form of safeguarding data privacy. Blockchain is adequate for data and transaction trustworthiness, certification and validation by consensus. However, data privacy in Blockchain is not fully assured by design. Blockchain, as stated by Nakamoto, is only anonymous until the public key is shared. Therefore, Blockchain is pseudonymous and arises some insecurities about data privacy in it.

Considering that it is possible to link the public key to a real person, any user in Blockchain could discover real identities. Although users' data can be encrypted in Blockchain, users' real identity can be discovered through different attack techniques [28–30]. Hence, Blockchain is a novel technology that needs a new design approach and definition to assure users' data privacy in Blockchain.

3.2 Limitations in permission-less architecture

Blockchain has a permission-less design. Such designs presents some limitations [28]:

1. **Sequential execution**: Blockchain platforms, such as Ethereum, can execute Smart Contracts. For each transaction, each Smart Contract is executed sequentially in each and every node of the network. In some cases, it can provoke a Denial of Service when the execution takes too long.
2. **Non-deterministic execution**: Non-deterministic execution does not assure the same results. In a Blockchain network, this can lead to ambiguous results.
3. **Execution on all nodes**: This is at odds with confidentiality, given that each and every node has to execute the smart contracts. These should only be executed in the granted contracts.
4. **Privacy-invasive**: Blockchain transactions and data are public. The ledger is distributed to all nodes, so all users have the same data. This situation is clearly privacy-invasive for many use-cases. Although data in blocks is encrypted, each and every user of the Blockchain network has the same data. This data can be exported and decrypted in the future where decrypting conditions could be optimal.
5. **Hard-coded consensus**: The consensus protocol is hard-coded to any Blockchain service. Thus, is very difficult, if not impossible, to change without any recoding or code refactoring.

These limitations could affect the performance and efficiency of possible automatic legal enforcements in the Blockchain. This automation will be executed in every node by smart contracts, which could result in a potential bottleneck. Moreover, the hard-coded consensus cannot be reprogrammed quickly enough to assure data trustworthiness. This could lead to a situation where data in Blockchain is not reliable. Hence, Blockchain is not suitable to automatically enforcing legal controls, principles and good practices for data privacy agreements.

3.3 Data immutability

Data immutability in Blockchain is a problem due to:

1. **Privacy-invasion**: If each user of the Blockchain network has the same data and this data can be exported, better decryption conditions can release real data and expose it to non-permitted users.
2. **Bureaucracy in itself**: The users' network of Blockchain is the central bureaucracy in itself. The Blockchain exists if there are enough users to validate data, as in the networks of centered authorities
3. **Need of a database**: If the word "Blockchain" can be substituted by "database" inside a text without losing any sense, maybe what is needed is a database instead of a Blockchain. Future needs may arise to change the data inside the Blockchain. These could be solved using links to data storage services such as Drobox or Google drive. However, links to data may change in the future. New services could appear with new and stronger encryption algorithms or current outdated data storage services y disappear.

Blockchain is about data immutability and unbreakable encryption algorithms. What will happen when ultrahigh computing systems are able to break encryption of

such encrypted data? Standard cryptographic systems are known to be vulnerable [31]. Eventually, it will be impossible to ensure data privacy. Such a consequence will be unacceptable for any Blockchain user. The data will be immutable as long as we live in a computational power era in equilibrium with the hard-coded consensus protocol. Hence, data privacy is not assured.

4 Solution and architecture

In the area of learning analytics, both digital artifacts and real devices are involved. Digital artifacts involve all virtual learning environments. Real devices involve all those that allow to take biometric data or measure body actions such as face detection, sweating, body position or even "graphological". Both could potentially collect students' private data. All collected data is stored in logs. These logs may be physically stored in the educational institution in which they are generated. However, depending on the platforms used, the data may be stored on the servers of the companies that provide them. These storages keep personal data of students, including minors. Therefore, data privacy in learning analytics is a sensitive issue –the same for other contexts where educational data is traced, collected, stored and used in some manners.

The fragility in data privacy is a problem within learning analytics. It generates fears and feelings against its use. In this sense, different proposals have been developed to ensure the privacy of data and ensure the digital identity of students. We find policy frameworks such as DELICATE, principles and rules of good practice such as Sheila Project or LALA Community, or even European laws such as the General Regulation of Data Protection and other specific ones depending on the country such as the Organic Law on the Protection of Personal Data in Spain. These privacy policies and agreements are contingent upon good-doing. Consequently, the approach of ensuring the privacy of data bylaws or policies is dysfunctional, since it assumes that institutions, companies or workers will behave in a correct manner. Reality shows that there are misuses of personal data collected in processes of learning analytics. The most dramatic case was the closure of inBloom schools, which shared personal data of students to third parties without parental consent. Even in other areas, there are data leaks such as the Cambridge Analytics from Facebook.

It is necessary to develop an approach that addresses the problem from its design and creation. It is necessary for laws to be applied before or at the time of data misuse. Consequently, the study and generation of a technological framework of architecture and concrete rules to allow the automation of policies, laws, principles and good practices in a safe and interoperable environment is required. The solution has to be:

1. **Centrally managed**: Distributed data in a public ledger is not suitable to assure data privacy and real anonymity. A central platform required to enable students to manage their data.
2. **Data Link based**: Private and personal data do not have to be stored in the solution. The solution has to be a gate to the data where students can use any platform.

3. **Interoperable**: Has to be able to communicate with different platforms to collect data and perform CRUD (Create, Read, Update, and Delete) actions.
4. **Secure and reliable**: The architecture has to be reliable and the data behind it have to be secured and not accessible without any consent. The slightest failure can raise the same fears and angst as those that appeared in Learning Analytics.
5. **Scalable**: Private and personal data of so many students from all connected platforms could be considered as Big Data. The architecture of the solution has to provide mechanisms to scale in time with interoperability with different platforms.
6. **Automated agreements**: The data privacy agreements between students and entities have to be enforced automatically to provide real time checking. In addition, the laws, the regulations and all the policies involved have to be checked automatically in real time.

Hence, the solution has to be a technological mechanism to automatically enforce policy and private data agreements.

4.1 Personal Data Broker

The idea of the Personal Data Broker (PDB) is to mediate between entities -such as students and universities-. This figure 2 activates different abilities for the students: A) Store the data wherever they want (logs, portfolios, certificates, qualifications ...) and in whichever technology they want (relational databases, xAPI Learning Record Stores, cloud storage services, etc.); B) Automate privacy agreements between entities; C) Decide the CRUD actions to be taken and how long to activate them for certain entities; D) Allow to read encrypted or open data. Our narrative of the PDB mechanism reads: “An entity that links to student data which incorporates different privacy agreements related to the relationships students have with different entities. It is able to allow the access, modification or storage of their data. These actions are determined by the privacy agreements."

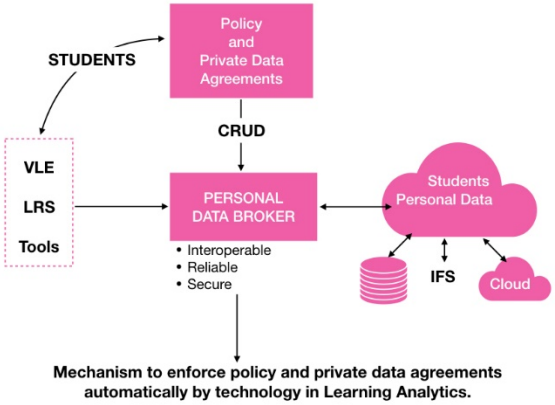

Fig. 2. Personal Data Broker Scheme.

4.2 Smart Contracts for automation and Blockchain for trustable context

Digital automation technologies such as Smart Contracts are candidates for study. Their use can make the interoperable and secure framework of automation of policies, laws, and principles of good practices a reality. Smart Contracts automate business rules between two entities. This technology is based on short programs whose activation depends on specific conditions. These conditions are found in the privacy agreements or policies between entities, such as agreements between universities and students. Hence, Smart Contracts are applicable in this solution.

In educational contexts, smart contracts can be used to enforce regulations, laws, principles and good practices to ensure students' data privacy. The use of Blockchain as a certification mechanism between entities could prove interesting as it could generate trust in the educational context while preserving private and personal data with PDB at the same time. In no case would Blockchain be used to protect the privacy of students.

5 Conclusions

Analytics in education is a growing aspect in educational technology. The results are very useful for stakeholders and tools providers and be used to enhance education and learning processes in many different ways, as other studies have demonstrated [32–34]. Despite of improvements, students' personal and private data is being exposed, shared and traded by the same service and tools providers. Regulations, ethics, laws, principles and good practices are not enough to stop data leaks and misuse. The personal and private data collected from students also includes data from minors. Therefore, it is a sensitive issue. There is an urgent need to develop a solution that can leverage the technology to automate those regulations and principles to enable real-time detection. The solution has to assure students' data privacy and respect the privacy agreements between students and educational entities.

Emergent technologies such as Blockchain seemed to be a potential solution. In this paper, we argued that Blockchain is a novel technology with some security flaws by design that turns it an unreliable protocol in terms of data privacy. Hence, is not a suitable technology to assure data privacy in the educational context.

Smart contracts can automate data privacy agreements between entities, such as students and educational institutions. We propose an alternate solution to Blockchain that uses smart contracts, is cloud based, enables law automation and assures the data privacy of students and respects privacy agreements. We call it Personal Data Broker (PDB), which enables students to take control and manage their own data and decide who and when can make create, read, update, and delete actions.

We are now working on the implementation of PDB in Moodle to save logs outside the LMS, in real time, and to provide a layer of security to users' private data. The results of this ongoing and other future work will be published in article format.

Acknowledgment. To the support of the Secretaria d'Universitats i Recerca of the Department of Business and Knowledge of the Generalitat de Catalunya for the help regarding 2017 SGR 934.

References

1. Filvà, D.A., Forment, M.A., García-Peñalvo, F.J., Escudero, D.F., Casañ, M.J.: Clickstream for learning analytics to assess students' behavior with Scratch. Futur. Gener. Comput. Syst. 93, 673–686 (2019).
2. Gros, B., García-Peñalvo, F.J.: Future Trends in the Design Strategies and Technological Affordances of E-Learning. In: Learning, Design, and Technology. pp. 1–23. Springer International Publishing, Cham (2016).
3. Conde, M.Á., García-Peñalvo, F.J., Rodríguez-Conde, M.J., Alier, M., Casany, M.J., Piguillem, J.: An evolving Learning Management System for new educational environments using 2.0 tools. Interact. Learn. Environ. 22, 188–204 (2014).
4. Amo, D., Alier, M., Casañ, M.J.: The Student's Progress Snapshot a Hybrid Text and Visual Learning Analytics Dashboard. Int. J. Eng. Educ. 34–3, 990–1000 (2018).
5. Lupton, D., Williamson, B.: The datafied child: The dataveillance of children and implications for their rights. New Media Soc. 19, 780–794 (2017).
6. Conde, M.Á., Hérnandez-García, Á., J. García-Peñalvo, F., Séin-Echaluce, M.L.: Exploring Student Interactions: Learning Analytics Tools for Student Tracking. Presented at the (2015).
7. Chatti, M., Dyckhoff, A., Schroeder, U.: A Reference Model for Learning Analytics. Int. J. Technol. Enhanc. Learn.
8. Papamitsiou, Z., Economides, A.A.: Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. (2014).
9. Herold, B.: inBloom to Shut Down Amid Growing Data-Privacy Concerns - Digital Education - Education Week, http://blogs.edweek.org/edweek/DigitalEducation/2014/04/inbloom_to_shut_down_amid_growing_data_privacy_concerns.html.
10. Hoel, T., Chen, W.: Implications of the European Data Protection Regulations for Learning Analytics Design. (2016).
11. Drachsler, H., Greller, W.: Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. In: Proceedings of the sixth international conference on learning analytics & knowledge. pp. 89–98 (2016).
12. Tsai, Y.-S., Moreno-Marcos, P.M., Tammets, K., Kollom, K., Gašević, D.: SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In: Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18. pp. 320–329. ACM Press, New York, New York, USA (2018).
13. Sclater, N., Biley, P.: Code of practice for learning analytics | Jisc, https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics.
14. Pardo, A., Siemens, G.: Ethical and privacy principles for learning analytics. Br. J. Educ. Technol. 45, 438–450 (2014).

15. Forment, M.A., Filvà, D.A., García-Peñalvo, F.J., Escudero, D.F., Casañ, M.J.: Learning Analytics' Privacy on the Blockchain. In: Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM'18. pp. 294–298. ACM Press, New York, New York, USA (2018).
16. Filvà, D.A., García-Peñalvo, F.J., Forment, M.A., Escudero, D.F., Casañ, M.J.: Privacy and identity management in Learning Analytics processes with Blockchain. In: Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM'18. pp. 997–1003. ACM Press, New York, New York, USA (2018).
17. Bartolomé Pina, A.R., Bellver Torlà, C., Castañeda Quintero, L., Adell Segura, J.: Blockchain en Educación: introducción y crítica al estado de la cuestión. Edutec. Rev. Electrónica Tecnol. Educ. 0, 363 (2017).
18. Grech, A., Camilleri, A.F.: Blockchain in Education. JRC Sci. Policy Rep. (2017).
19. Sun, H., Wang, X., Wang, X.: Application of Blockchain Technology in Online Education. Int. J. Emerg. Technol. Learn. 13, 252 (2018).
20. Henry, R., Herzberg, A., Kate, A.: Blockchain Access Privacy: Challenges and Directions. IEEE Secur. Priv. 16, 38–45 (2018).
21. Karame, G., Capkun, S.: Blockchain Security and Privacy. IEEE Secur. Priv. 16, 11–12 (2018).
22. Casino, F., Dasaklis, T.K., Patsakis, C.: A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telemat. Informatics. 36, 55–81 (2019).
23. Feng, Q., He, D., Zeadally, S., Khan, M.K., Kumar, N.: A survey on privacy protection in blockchain system. J. Netw. Comput. Appl. 126, 45–58 (2019).
24. Li, X., Jiang, P., Chen, T., Luo, X., Wen, Q.: A survey on the security of blockchain systems. Futur. Gener. Comput. Syst. (2017).
25. Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System. (2008).
26. Sultan, K., Ruhi, U., Lakhani, R.: Conceptualizing blockchains: characteristics & applications. (2018).
27. Puthal, D., Malik, N., Mohanty, S.P., Kougianos, E., Das, G.: Everything You Wanted to Know About the Blockchain: Its Promise, Components, Processes, and Problems. IEEE Consum. Electron. Mag. 7, 6–14 (2018).
28. Vukolić, M., Marko: Rethinking Permissioned Blockchains. In: Proceedings of the ACM Workshop on Blockchain, Cryptocurrencies and Contracts - BCC '17. pp. 3–7. ACM Press, New York, New York, USA (2017).
29. Wan, Z., Lo, D., Xia, X., Cai, L.: Bug Characteristics in Blockchain Systems: A Large-Scale Empirical Study. In: 2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR). pp. 413–424. IEEE (2017).
30. Halpin, H., Piekarska, M.: Introduction to Security and Privacy on the Blockchain. In: 2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). pp. 1–3. IEEE (2017).
31. Ikeda, K.: Security and Privacy of Blockchain and Quantum Computation. Adv. Comput. 111, 199–228 (2018).
32. Pinto, M., Rodrigues, A., Varajão, J., Gonçalves, R.: Model of Funcionalities for the Development of B2B E-Commerce Solutions. In: Cruz-Cunha, M.M. and Varajão, J.

(eds.) Innovations in SMEs and Conducting E-Business: Technologies, Trends and Solutions. p. 26. IGI-Global (2011).

33. Pereira, J., Martins, J., Santos, V., Gonçalves, R.: CRUDi framework proposal: Financial industry application, (2014).

34. Pires, J.A., Gonçalves, R.: Constrains associated to e-business evolution. In: E-business issues, challenges and opportunities for SMEs: driving Competitiveness. pp. 335–349. IGI-Global (2011).