Catalogue no. 11-633-X — No. 044
ISSN 2371-3429
ISBN 978-0-660-44782-7

Analytical Studies: Methods and References

Investigating the Use of Blockchain to Authenticate Data from the Statistics Canada Website

By Kathryn Fedchun, Lillian Klein, and Didem Demirag

Release date: September 19, 2022

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

"Do you know what a non-fungible token (NFT) is?" This question started a chain reaction that resulted in an investigation by a diverse team into how Statistics Canada (StatCan) could use blockchain, or distributed ledger technology, to authenticate a document. The question was posed as part of a more significant idea of how the Dissemination Division might use NFTs, or similar technology, to authenticate the products leaving the StatCan website. Initially, our team was composed of internal StatCan employees: Mathieu Laporte, Director of the Dissemination Division; Jacqueline Luffman, Chief of Publishing Services; and Lillian Klein, Research Librarian. These individuals discussed the idea among other StatCan staff to evaluate whether it was feasible. However, as we recognized a gap in our blockchain experience, we reached out to academics who research various aspects of blockchain technology. Through those meetings, we were connected to four blockchain experts: Dr. Florian Martin-Bariteau from the University of Ottawa, Dr. Jeremy Clark from Concordia University, Dr. Victoria Lemieux from the University of British Columbia and Dr. Tracey Lauriault from Carleton University. We met with these experts for a brainstorming session, where Jeremy Clark presented the idea of using digital signatures to authenticate StatCan documents. With this idea in mind, a team of researchers was formed to explore up-to-date cryptographic technology and applications to develop a comprehensive understanding of the technology and determine whether using this technology in StatCan's work would be meaningful. Our research team includes Kathryn Fedchun, a PhD student at Carleton University; Didem Demirag, a PhD candidate at Concordia University; and Lillian Klein, a research librarian with StatCan. This paper summarizes months of collaborative work completed by this team.

The main focus of this project is to understand more about blockchain and see whether, as StatCan expands its website, it could use blockchain technology to enable users to authenticate the data downloaded from the website. With an increased understanding of these emerging technologies, the aim of this project is to develop a process of authentication that would allow users to verify that the material downloaded from the StatCan website has not been tampered with and was produced by StatCan. This would increase overall trust in the agency as a statistical organization. By using blockchain to determine the authenticity of its data, StatCan has the ability to increase social trust with its users. It was identified that the ideal method of authenticating the data should be easy to use and available in an online and offline format to ensure users with varying degrees of Internet connection can authenticate their data.

Our research successfully defined and explained what blockchain is and identified how blockchain is currently being used in a Canadian context. We found that there has recently been a call to action for government agencies to embrace blockchain technology and take strides to implement it in their work. To create a well-rounded assessment of the technology, we included a review of concerns regarding blockchain. We focused primarily on the environmental impact, the public perception of the technology and any potential backlash our team could anticipate, the lack of regulations, and the potential to be blinded by the hype of blockchain technology. Finally, we completed a brief comparison of five blockchains that could be used in our solution. This comparison focuses on general information about each chain, along with the transactions per second, the consensus mechanism, whether it is private or public, and each blockchain's environmental impact. This analysis enabled us to decide that Avalanche is the best option for us as we move forward with our technical solution.

With the knowledge gained from this research, our team recommends that this project could be the agency's opportunity to answer the call to action. We propose that StatCan conduct a pilot project based on Jeremy Clark's idea about using digital signatures and build an application that users can download to authenticate their data. We propose using a hybrid model with a blockchain that will allow both online and offline users to authenticate their data. The technical details of this project are explained in depth below; to summarize:

> **In the hybrid solution, authentication will occur through an application that users must download. The list of hashes of files is updated periodically to contain the hashes of new StatCan products. The authentication of a file will occur as follows: the user will need to upload the file needing authentication to the app. This action will prompt the app to compute the hash of the file and compare it with the list of already existing hashes from StatCan products. The app will then inform the user whether the file is valid.**

This solution adds tremendous value to the agency's transparency and trust with users. Hosting the hash values on the blockchain creates an immutable record over time of the products the agency has released and increases users' ability to trust the information downloaded from the StatCan website. This project is an opportunity to experiment with blockchain technology without overhauling the agency's existing system.

INTRODUCTION

In the age of information, it is necessary to acknowledge the growing amount of digital information available to Canadians and their increasing distrust of digital sources (Ipsos Public Affairs for Canada's Centre for International Governance Innovation [CIGI-IPSOS], 2019). According to Ipsos Public Affairs for Canada's Centre for International Governance Innovation (2019), 36% of Canadians feel that the government contributes to their sense of distrust in the Internet. As Statistics Canada (StatCan) is the branch of government responsible for disseminating information to Canadians, it should not ignore this statistic. During the 2020/2021 fiscal year, the StatCan website had over 28 million web page visitors and 766,589 table downloads (Statistics Canada, 2021). StatCan prides itself on its transparency and accountability to the public and strives to meet the needs of its users (Statistics Canada, 2018). As an organization, StatCan advertises itself as being "a trusted source of statistics on Canada" (Statistics Canada, 2018). According to StatCan's Trust Centre, "the people of Canada can trust that information gathered from them, and about them, is done so for them—and that these activities are carried out with integrity and the highest ethical standards" (Statistics Canada, 2018). The Statistics Act guides StatCan to ensure that it promotes and develops "integrated social and economic statistics pertaining to the whole of Canada" (Statistics Act, 1985).

Users count on the agency and expect to access and download authentic, reliable data when they enter the StatCan website. But once a product has been downloaded, it is challenging to validate that it belongs to StatCan and has not been tampered with by a malicious actor. This means that users may believe they are accessing untampered data from StatCan when downloading a corrupted comma-separated value (CSV) file. Regarding the likelihood of StatCan becoming a victim of cyber threats at the hands of malicious agents, the increased number of ransomware attacks on Canadian organizations shows that the country is a potential target (Communications Security Establishment Canada, 2021). Therefore, as StatCan begins to plan the expansion and innovation of its website, it is essential that it consider how it can give users the ability to verify and authenticate the data they download from the website.

This research aims to investigate whether StatCan could respond to the authentication gap on its website by integrating emerging technologies into its existing publication methods. To find answers, we began by familiarizing ourselves with the current research surrounding blockchain and distributed ledger technology. We then considered the importance of record keeping, confidentiality, trust and authentication. We looked at multiple examples of other Canadian organizations and government agencies using blockchain and found multiple articles calling on the government to adopt this new technology. However, we also considered concerns related to these emerging technologies, including environmental impact, public image and potential backlash, a lack of regulation, and the possibility of being blinded by the "hype." We investigated five blockchains that could be used in our system design: Ethereum, Avalanche, Cardano, Hyperledger and Solana. With a better understanding of the technology available to StatCan, we worked to conceptualize a system that allows users to authenticate the data they download from the website. Our goal is that the system enables users to verify that the material downloaded from the website has not been tampered with and was produced by StatCan. We believe that our method of authenticating data should be available in online and offline formats to ensure that users with varying degrees of Internet connection can authenticate the data. Our team prioritized this component to serve all Canadian users, knowing that high-speed Internet connection is inconsistent because of the digital divide in the country (Canada's Public Policy Forum, 2014). Additionally, we prioritized usability when considering options for a solution, which needs to be as simple as possible to ensure the technology is accessible and easy to understand by users.

Before a solution can be recommended, it is necessary to introduce the technology behind it to provide the context required to understand how the technology can help StatCan accomplish its goal. The main features that need to be understood are the digital signatures and hash functions that support our concept. In addition to the introduction and literature review, Appendix A has a glossary of terms to help readers understand the more technical material.

GAPS IN THE LITERATURE

Throughout the research process, we found a few gaps in the literature. Given that blockchain is still a relatively new technology, especially for government use, it is not surprising that gaps were found. It was difficult to find any concrete Canadian government regulations or policies on how to incorporate blockchain. This means that directives on the implementation of blockchain within the government are still coming to light. This gap leaves our team with questions surrounding how policies might change in the future to simplify or complicate the implementation of this project. Another identified gap is the lack of variety in the way organizations have published their method of incorporating blockchain into their daily work. We found a lot of material about how blockchain is being used in cryptocurrency, record keeping and financial technology (fintech). However, it was difficult to determine how blockchain is used by organizations on a daily basis. We were also unable to find significant information on the legal implications of using blockchain for our purposes. For example, in the case of health records discussed below, it was difficult to determine how patient files were uploaded or tracked on the blockchain. Furthermore, it was difficult to find research on similar projects. We were unable to locate published research seeking to address the issue of how to give users the ability to authenticate data that have been downloaded from a website. We believe that our project fills some of the gaps in the literature and is a valuable step in the direction of new technology for StatCan.

METHODOLOGY

SYSTEMATIC LITERATURE REVIEW

We performed a systematic literature review for this study, which allowed us to understand “the breadth and depth of the existing body of work and identify gaps to explore” (Xiao and Watson, 2019, p. 93). A successful systematic literature review involves three stages: planning, conducting and reporting (Xiao and Watson, 2019, p. 102). The first stage, planning, is when researchers “identify the need for a review, specify research questions, and develop a review protocol” (Xiao and Watson, 2019, p. 102). In the second stage, researchers conduct the research and “identify and select primary studies, extract, analyze, and synthesize data” (Xiao and Watson, 2019, p. 102). Finally, the third stage involves researchers “writ[ing] the report to disseminate their findings” (Xiao and Watson, 2019, p. 102). For this project, we had the following three research questions in mind:

1. **What is blockchain, and how have other government agencies and organizations used this technology?**

2. **How can this technology be used to increase record keeping, confidentiality, trust and authentication on the Statistics Canada website?**

3. **How can we use this new technology in our solution to authenticate data?**

In the planning phase of this study, we compiled a list of search terms that focused on our areas of interest in this project. The list of search terms can be found in Appendix B. As Xiao and Watson (2019) described in their article on how to conduct a systematic literature review, we used these search terms to identify relevant articles. As we collected academic research articles, our team added more search terms. We then used a variety of combinations of the search terms listed in Appendix B with Boolean operators to focus our results. In total, we completed 15 unique searches.

Depending on the number of results listed in a search, we reviewed between 100 and 300 results. If the number of results listed was below 1,000, we examined the first 100. If the number of results was below 100,000, we reviewed the first 200; if there were over 100,000 results, we examined the first 300. In the review process, we assessed academic articles based on their relevance to this study using the title of the article, the abstract and the listed keywords. Overall, we collected 59 papers and entered the source information into a spreadsheet, including the title, authors, year the article or book was published, abstract, and complete citation.

Upon collecting the sources, we began reviewing each article to determine its relevance to this project. We assessed the abstracts in further detail and skimmed through the articles to assess their usefulness. Of the 59 papers, we found 18 sources that proved significantly valuable for this project. Most of the excluded papers were too technical for the purpose of this literature review. While we have attempted to make this paper relatively accessible, we have provided a list of technical terminology and definitions in Appendix A. While some of these definitions are paraphrased, they contain a fair amount of quoted material to maintain integrity.

From our 18 sources, we extracted relevant information and data and synthesized them into the literature review below. Using the research questions listed above, we provided a detailed overview of the technology; considered the significance of record keeping, confidentiality, trust and authentication; and provided a list of examples of other government organizations and agencies using blockchain. In addition, we were surprised to find multiple articles calling on the government to use these new technologies, and we also included this as a theme below.

Beyond academic articles, we reviewed multiple articles on the concerning aspects of blockchain related to environmental impact, public image and potential backlash, a lack of regulation, and the potential to be blinded by the hype of blockchain technology. We also researched five specific blockchains: Ethereum, Avalanche, Cardano, Hyperledger and Solana. The number of blockchains available grows each day, but our team chose to investigate these five. Ethereum is an extremely popular peer-to-peer blockchain that uses a fair amount of energy. Avalanche is a more environmentally friendly proof-of-stake blockchain, like Cardano, which is also a proof-of-stake blockchain that is easy on the environment compared with Ethereum. Hyperledger is an umbrella project of open-source blockchains and related tools, and Solana is a carbon-neutral, proof-of-stake blockchain. More information about these five blockchains and their differences is provided below. This systematic literature review strengthened our knowledge of this technology and supported us in creating recommended solutions and next steps for this project, found below.

LITERATURE REVIEW

This project aims to explore how technology can help users verify and authenticate data from the StatCan website. This literature review begins with a brief overview of cryptographic technology. Next, we consider the importance of record keeping, confidentiality, trust and authentication. We provide examples of organizations, agencies and companies in Canada that use this technology. Then, we list multiple sources that call on the government to move toward new technology such as blockchain. Next, we consider potential concerns with using blockchain, such as environmental impact, public image and potential backlash, a lack of regulation, and the possibility of being blinded by the hype of blockchain technology. Finally, we compare five blockchains: Ethereum, Avalanche, Cardano, Hyperledger and Solana. This project is a small step for StatCan toward new technology that can better protect its data.

OVERVIEW OF THE TECHNOLOGY

In the early 1990s, cryptographers Scott Stornetta and Stuart Haber conceived the idea of “connecting blocks via hashed data” (Treiblmaier and Clohessy, 2020, p. v). Almost 20 years later, on October 31, 2008,

> **A mysterious individual, or group of individuals, known only as Satoshi Nakamoto, posted a link to a paper entitled Bitcoin: A Peer-to-Peer Electronic Cash System to an obscure mailing list called Cryptography List. In this paper, Nakamoto proposed the creation of what would become known as a blockchain as a means of enabling an electronic payment system that did not require a trusted third-party intermediary (Urban and Pineda, 2018, p. 5).**

A blockchain is “a digital, decentralized and distributed ledger in which transactions are logged and added in chronological order with the goal of creating permanent and tamper-proof records” (Treiblmaier, 2018, p. 547). The idea of the ledger has existed for a long time—it is a permanent collection of recorded transactions, historically written in a physical book. Moving the ledger online into a digital currency is where blockchain originated. Since then, blockchain has broadened to include digital security beyond digital currency such as Bitcoin.

Much of this technology stems from cryptography. The term “cryptography” is derived from the Greek word kryptos, which is used to describe anything that is “hidden, veiled, secret, or mysterious” (Mohamed, 2020, np). Cryptography secures communication and information using technology and codes. It is well known that data are valuable and often vulnerable. “In today’s world, producing fake documents is becoming more common. As the fake ones accurately look like the originals, it is impractical for a common man to identify the real and the duplicate one” (Prathibha and Krishna, 2021, p. 71). Given this information, technology that uses cryptography and blockchain can protect the information, making it "tamper-resistant [and] exceptionally hard to change or delete" (De Filippi, 2018, p. 34–35). As people begin to recognize the significant and inherent value of data, blockchain and distributed ledger technology "may force some organizations fundamentally to rethink their relations with users and approaches to privacy" (Maull et al., 2017, p. 484). Before providing some examples of blockchain use in Canada, we will discuss the importance of record keeping, confidentiality, trust and authentication for our project.

RECORD KEEPING

Victoria Lemieux, an archival studies scholar, claims that "much of the discussion about trusted records or systems boils down to two interlinking concepts: reliability and authenticity" (2016a, p. 112). When a user accesses a record, they consider any potential risks associated with the data (Lemieux, 2016a). Users determine the reliability of data based on how they are accessing the data and on record creation, including who created the record and how (Lemieux, 2016a). Lemieux argues that "long-term preservation of information in digital form requires

that technical dangers to the longevity of authentic information be addressed" (2016a, p. 114). In our case, "the purpose of what is actually stored on chain … is not archiving but rather to establish that the original transaction record is authentic" (Lemieux, 2016b, p. 15). The aim of this project is to proactively safeguard StatCan data through the added value of blockchain

CONFIDENTIALITY

This project demonstrates that StatCan recognizes the importance of confidentiality. When dealing with data, confidentiality "refers to the protection of information, such as computer files or database elements, so that only authorized persons may access it in a controlled way" (Mohamed, 2020, np). StatCan data need to be protected from potential threats or attacks. To accomplish this, we must determine the vulnerability or weakness of the current StatCan system (Mohamed, 2020). It is possible that data on the StatCan website could be altered without the user's knowledge. This project attempts to fix the potential risk by addressing confidentiality and ensuring that information can be authenticated by the user.

TRUST

According to a chapter on how authenticity can transform social trust, Batista et al. illustrate the three most important aspects of trust: accuracy, reliability and authenticity (2021, p. 112). They argue that "accurate [and reliable] records are precise, correct, truthful … consistent, complete, and objective" (Batista et al., 2021, p. 114). To generate trust, the authors describe that authentic records need to "preserve their identity and integrity over the period of long-term preservation" (Batista et al., 2021, p. 116). In the case of digital archives, the authors describe the difficulty in maintaining trust with a digital document. For example, suppose a statistical document has been altered. In this case, it might be challenging to detect the variances between the original and the copy that has been tampered with, and this can negatively impact social trust because of what they call "uncertain authenticity" (Batista et al., 2021, p. 117). This project seeks to improve trust between StatCan and its users by providing a way to authenticate data from the StatCan website and removing uncertainty.

AUTHENTICATION

Authentication refers to the ability to determine the validity of a source. It answers the question, "How does a receiver know that [the] remote communicating entity is who it is claimed to be?" (Mohamed, 2020, np). In this project, StatCan wants to help users determine the validity of a source through an authentication process. Cryptographic algorithms support authenticated encryption, meaning that users can be sure the source is authentic (Mohamed, 2020). This verification also instills integrity—it means they can know that the information has not been modified unless StatCan employees changed it through proper authorization (Mohamed, 2020). Evidently, record keeping, confidentiality, trust and authentication are significant factors in this project. Next, we provide examples in Canada that demonstrate this technology in use.

KEY IDEAS

- Cryptography is the process of securing communication and information using technology.
- A blockchain is "a digital, decentralized and distributed ledger" (Treiblmaer, 2018, p.547).
- A blockchain can support record keeping through archiving data, increase confidentiality and decrease vulnerability, generate trust between the user and StatCan, and support users in authenticating data from the StatCan website.

EXAMPLES IN CANADA

Many examples were found in our research of the Canadian government incorporating blockchain into specific projects. In a policy book published by the Mowat Centre for Policy Innovation at the University of Toronto, Urban and Pineda (2018, p. 61–62) list many Canadian government agencies experimenting with blockchain, such as Innovation, Science and Economic Development Canada; the Treasury Board of Canada Secretariat; and the National Research Council Canada (NRCC). In January 2018, the Industrial Research Assistance Program in the NRCC used an Ethereum blockchain to "proactively publish grants and contribution data in real-time" (Industrial Research Assistance Program, 2019). This project was an experiment that ran for one year and concluded on March 1, 2019. While the experiment is not ongoing, this work has provided "constructive insight into the potential for this technology and how it may be used for more open and transparent operations for public programs" (National Research Council Canada, 2018).

Multiple levels of government have moved toward using blockchain for permits, including the Government of Ontario, the City of Toronto and the Government of British Columbia (Urban and Pineda, 2018, p. 62). One article lists a variety of ways that governments are using blockchain, including for "digital identity, the storing of judicial decisions, financing of school buildings and tracing money, marital status, e-voting, business licenses, passports, criminal records and even tax records" (Ølnes, Ubacht, and Janssen, 2017, 357). The Government of Ontario also "ran a blockchain hackathon that generated a number of ideas for other blockchain applications in government" (Urban and Pineda, 2018, p. 62). Supporting pilot projects that use blockchain is an effective way for the government to begin using these new technologies successfully (Urban and Pineda, 2018, p. 67). Governments are using blockchain in many areas, and StatCan can use this knowledge and build upon their work in this project.

In addition to government agencies implementing blockchain and distributed ledger technology, health care is moving rapidly toward blockchain and digital health care records. Storing electronic health records on a blockchain is not only improving record keeping but also "giving patients greater control over their own health and medical treatments" (Urban and Pineda, 2018, p. 42). Doctors, nurses, hospitals and other health care institutions are using blockchain to certify the health of patients (De Filippi, 2018, p. 112). It is being used to "store encoded personal health records" (Zheng, Zhu, and Si, 2019, p. 17). The blockchain can provide access to specific individuals, so a person's health records can be secure and confidential when stored in a distributed ledger (Zheng, Zhu, and Si, 2019). Lemieux writes, "the underlying conditions in Canada are particularly well-suited to leading blockchain research and implementation … Canada has a vibrant, highly active blockchain technoscape, with a diversity of start-ups and consultancies doing innovative work" (2016b, p. 5). We are excited to add to this work in our project.

KEY IDEAS

- Blockchain is being used by multiple levels of government in Canada, including the Government of Ontario, the City of Toronto and the Government of British Columbia.
- Health care in Canada has also begun to use blockchain and distributed ledger technology to securely store digital health care records.

CALL TO ACTION

Multiple papers called on governments to move toward new technology to better secure their data. Urban and Pineda argue that blockchain can "offer governments the possibility of improved transparency, efficiency, and effectiveness" (2018, p. 42). While blockchain is not a new technology, its use in government is relatively new, so "the level of blockchain expertise and capacity within Canadian governments and regulators is currently limited" (Urban and Pineda, 2018, p. 61). They claim that one of the first things the Canadian government should do is what we are doing currently in this project: "building up groups of technologists and policymakers within government who understand the technology, its implications, and the potential opportunities and challenges that flow from it" (Urban and Pineda, 2018, p. 61). While Urban and Pineda (2018) are pushing for more blockchain in government, Ølnes, Ubacht and Janssen emphasize that the government should "shift from a technology-driven to need driven approach with blockchain applications" (2017, p. 355). They argue that blockchain "will lead to innovation and transformation of governmental processes" (Ølnes, Ubacht, and Janssen, 2017, p. 355). Considering "the ease with which digital files can be altered" (Bell et al., 2019, p. 6), we argue that this project is driven by a need for authentication on the StatCan website.

According to De Filippi, “governments have established and stewarded a variety of systems and institutions designed to enhance social welfare and provide the foundational infrastructure for economic and political growth” throughout history (2018, p. 107). In an article on cryptography and government, Aljeaid et al. argue that “e-government … acts as a communication bridge … between government to citizen, or government to government, or government to business in efficient and reliable ways” (2014, p. 581). The authors emphasize the importance of data security in government related to potential vulnerability if left unsecured. They claim that “end users need robust security solutions to achieve assurance when dealing with e-government systems” (Aljeaid et al., 2014, p. 581). Creating a “tamper-resistant and resilient repository for public records” (De Filippi, 2018, p. 107–108) using cryptography and blockchain can help the government avoid data leaks, data loss and other vulnerabilities. We agree with this call to action and believe that this project will improve public trust in StatCan and the Government of Canada.

KEY IDEAS

- Multiple authors are calling on the Canadian governments to move toward using new technology, such as blockchain, to better secure their data.
- This technology can help the government avoid data vulnerabilities and improve transparency, security, efficiency and effectiveness.

CONCERNS

While the call to action is significant, we also want to take the time to investigate any potential concerns regarding blockchain. We have summarized our findings into four categories: environmental impact, public image and potential backlash, a lack of regulation, and the potential to be blinded by the hype of blockchain technology.

ENVIRONMENTAL IMPACT

There have been many claims about the environmental impact of new blockchain technology. In November 2021, a blockchain project called Solana contracted Robert Murphy, a climate and energy advisor, to publish an energy use report (Solana, 2021). They compared common activities that involve energy consumption with one Solana transaction, one Ethereum transaction and one Bitcoin transaction (Solana, 2021). While they did not include all of the blockchain options that we have chosen to investigate, it is helpful to consider how blockchain transactions compare with everyday activities. Conducting a single Google search uses 1,080 joules of energy, working on a computer with a monitor for an hour uses 46,800 joules, and using one gallon of gasoline uses 121,320,000 joules (Solana, 2021). By comparison, one Solana transaction uses 1,837 joules of energy, one Ethereum transaction uses 692,820,000 joules, and one Bitcoin transaction uses 6,995,592,000 joules (Solana, 2021). According to Huang, O’Neill, and Tabuchi for The New York Times, “the process of creating Bitcoin to spend or trade consumes around 91 terawatt-hours of electricity annually, more than is used by Finland, a nation of about 5.5 million” (2021). While we are not using Bitcoin for our project, these numbers are staggering.

Many of the big players in blockchain, including Ethereum, are using an astonishing amount of energy because of their proof-of-work (PoW) consensus mechanism. PoW requires network participants on the blockchain “to expend large amounts of computational resources and energy on generating new valid blocks” (Chandler, 2021). In comparison, proof of stake (PoS) requires network participants on the blockchain “to stake cryptocurrency as collateral in favor of the new block they believe should be added to the chain” (Chandler, 2021). Chandler argues that PoW, such as Ethereum, can be more secure and decentralized, but also uses an immense amount of electricity, is slower and is less scalable (Chandler, 2021). By contrast, PoS, such as Avalanche, Cardano and Solana, has a smaller environmental impact and allows for faster transactions and better scaleability, but it is a newer form of technology and “may not be as secure or tamper-resistant as proof of work” (Chandler, 2021). Evidently, both PoS and PoW have advantages and disadvantages, and we consider the specific environmental impact of five blockchains (Ethereum, Avalanche, Cardano, Hyperledger and Solana) in the chart below.

PUBLIC IMAGE AND POTENTIAL BACKLASH

There have been multiple examples of companies and organizations that received backlash when attempting to use blockchain. In December 2021, Kickstarter announced that it was moving to blockchain (Plunkett, 2021). The blog post, titled “Let’s Build What’s Next for Crowdfunding Creative Projects,” received many critiques and complaints from creators (Plunkett, 2021). Kickstarter responded by providing a frequently asked questions section, where it claims it is “confident that a crowdfunding protocol built on top of Celo will not significantly negatively impact our carbon emissions given its underlying architecture" (Kickstarter, 2022). Still, many creators and backers have claimed that they will no longer be using Kickstarter, given this information (Morse, 2021).

Similar to Kickstarter, the digital communication platform Discord tweeted about integrating Ethereum into its platform in November 2021 (Pearson, 2021). The founder and chief executive officer of Discord, Jason Citron, quickly backed off the project two days later, after public backlash (Pearson, 2021). Pearson states that people in the game industry hate blockchain "either because of the environmental impact of proof-of-work tokens on Ethereum, the idea that blockchain collectibles are a grift based on mythical thinking, or both" (2021). Many users unsubscribed from the platform's premium "Nitro" paid service or threatened to do so (Jiang, 2021). Given that both of these examples took place recently, in November and December 2021, it is difficult to consider what the public opinion might be regarding StatCan and this project. However, it is important to be aware of these examples and recognize that backlash is a potential outcome.

LACK OF REGULATION

Another concern is the decentralized and unregulated nature of blockchain. Given that control and decision making about the blockchain is not conducted by a single entity, this is an area of concern for StatCan. Rather than putting trust in one entity, trust is put in mathematical algorithms. Given that there have been other blockchain projects by Canadian governments, they should be used as a guide for StatCan policies regarding this project. Between the five blockchains we look at below, each has different regulations, goals and abilities. It can also be difficult to scale, depending on the blockchain chosen. This may be a concern because it has not yet been decided how many StatCan products will be available for authentication. Since we looked at trust and confidentiality earlier in this literature review, the lack of regulation is less worrisome than the impact on the environment and public image. In fact, this project is an opportunity to be an early example and leader in blockchain implementation regulations, and we hope that we will be able to incorporate new policies into our project.

BLINDED BY THE HYPE

The overall hype of blockchain technology needs to be addressed. According to Victoria Lemieux, we need to "address the shortcomings in designs and implementations of blockchain record keeping so as to be better able to realize the worthy vision of blockchains" (Lemieux, 2019). She writes, "claims associated with use of blockchain technology for recordkeeping are, in a number of cases, overhyped. As an example, blockchain solutions that claim to provide 'archival' solutions do not actually preserve or provide for long-term accessibility of records" (Lemieux, 2016b, p. 4). She claims that the biggest danger in blockchain comes from blindly trusting it (2016b, p. 23). However, critically investigating these limitations is the "key to successfully leveraging technological innovations like the blockchain for the benefit of all Canadians" (Lemieux, 2016b, p. 8). While blockchain technology does not solve every problem that it has been claimed to, it is a useful technology that "will continue to be used in industry and is deserving of further research and experimentation" (Ruoti et al., 2020, p. 53). While this relatively new technology is exciting, and considering risks can bring up "fears of stifling innovation" (Lemieux, 2016b, p. 5), it is imperative that we are critical of the potential limitations and concerns about blockchain technology to have the best possible outcome in this project.

KEY IDEAS

- There are four concerns regarding the use of blockchain:
  1. environmental impact related to energy consumption, with different blockchains using different amounts of energy
  2. potential backlash based on the experience of companies that tried to move toward blockchain and were critiqued by their users
  3. general lack of regulation because of the decentralized nature of blockchain
  4. becoming blinded by the hype of blockchain technology.

FIVE BLOCKCHAINS: ETHEREUM, AVALANCHE, CARDANO, HYPERLEDGER AND SOLANA

For this project, we chose to evaluate and compare five different blockchains, with specific considerations. We decided to look at Ethereum, Avalanche, Cardano, Hyperledger and Solana. Ethereum is one of the most popular blockchains, yet it conducts the fewest transactions per second and has significant energy consumption compared with other options because it uses proof of work (PoW). PoW means that a majority of users need to vote on each new blockchain, and this takes more time and effort than proof-of-stake (PoS) blockchains. We also included Avalanche and Cardano, which are both PoS public blockchains. While Avalanche's environmental impact is carbon neutral, its transaction rate per second is the highest, compared with the other four blockchains we analyzed. Meanwhile, Cardano is less energy efficient and slower than Avalanche. We also chose to include Hyperledger, as it is a private blockchain that uses Practical Byzantine Fault Tolerance as its consensus mechanism. It is a private blockchain, which means that it is centralized. This potentially impacts trust, as fewer nodes can make the network less secure. Finally, we included Solana because it is carbon neutral, uses PoS and has provided a report on energy consumption in comparison with blockchains such as Ethereum. All of the blockchains outlined below have advantages and disadvantages. Upon reviewing them, we have decided to use Avalanche for this project. Avalanche is an open-source PoS blockchain with the highest transaction rate per second, at 4,500. Additionally, it is a public network that is carbon neutral, an important consideration for us.

KEY IDEAS

- Our project is moving forward with Avalanche, which is an open-source proof-of-stake blockchain with the highest transaction rate per second, at 4,500.
- It is a public network and carbon neutral, resolving one of our concerns listed above, regarding environmental impact.

FIGURE 1:
An overview of Ethereum, Avalanche, Cardano, Hyperledger and Solana

Consideration	Ethereum	Avalanche	Cardano	Hyperledger	Solana
General information	"Ethereum is a technology that lets you send cryptocurrency to anyone for a small fee. It also powers applications that everyone can use and no one can take down."	"Avalanche is an open, programmable smart contracts platform for decentralized applications."	"Cardano is a proof-of-stake blockchain platform: the first to be founded on peer-reviewed research and developed through evidence-based methods."	"Hyperledger is a global collaboration, hosted by The Linux Foundation, and includes leaders in finance, banking, Internet of Things, supply chains, manufacturing and Technology."	"Solana is a decentralized blockchain built to enable scalable, user-friendly apps for the world."
Transaction per second	14	4,500	257	3,000	2295
Consensus mechanism	Proof of work (PoW)	Proof of stake (PoS)	Proof of stake (PoS)	Practical Byzantine Fault Tolerance (PBFT)	Proof of stake (PoS) and proof of history (PoH)
Public or private	Public	Public	Public	Private	Private
Environmental impact	Significant energy consumption	Net zero CO_2 (carbon neutral)	Energy efficient	Energy efficient	Net zero CO_2 (carbon neutral)
Website	www.ethereum.org	www.avax.network	www.cardano.org	www.hyperledger.org	www.solana.com

THE TECHNICAL SOLUTION

Our research team has designed a solution that incorporates blockchain technology using the knowledge gained from our literature review and pre-existing technical experience. This section outlines system details and the recommended solution for enabling users to authenticate documents downloaded from the StatCan website. We will begin by introducing three technical elements that are the pillars of our solution: digital signatures, hash functions and secure tunnels. These three technical elements interact as follows: a hash computed over the file that belongs to StatCan is used to make sure the file has not been tampered with; a digital signature over this hash proves that the file is owned by StatCan, and the secure tunnel ensures secure communication between the user and the StatCan website. In this section, we explain how these building blocks work and how they are integrated into our proposed solutions.

DIGITAL SIGNATURES

When users download a file from the StatCan website, there are two questions that they may have. First, does the data actually belong to StatCan? And second, has the data been tampered with?

QUESTION 1: DOES THE DATA ACTUALLY BELONG TO STATCAN?

To address this question, we propose using a digital signature. The idea is similar to signing a document with a pen—if you receive a signed letter or document from “x,” you can check whether the signature on the document belongs to “x” and consequently whether the document is theirs. In a digital signature scheme, a private-public key pair is used to sign a document and verify the signature over a document’s hash. There are three steps to a digital signature scheme: StatCan needs to (1) generate the public-private key pair, so that (2) it can sign the hash of the document with its private key, and (3) any user with the public key can verify the signature.

Step 1: Key generation

Using a function that generates keys, StatCan can obtain a public-private key pair. The public key is shared on the website for users to download and use during the signature verification. StatCan would not share the private key, as it might lead to a malicious actor using the private key to forge StatCan’s signature on documents. It is important to note that key generation is a one-way function, which means that it is infeasible to compute the private key, given the public key. StatCan would use its private key to generate the signature over a document’s hash rather than the document itself, as it is faster and more efficient, and the resulting signature is shorter. Consider the signature generation as a function that asks the user to provide their private key and hash of the document and generates a file that contains the signature.

Step 2: Signing the hash of a document

To create the signature, StatCan needs its private key and the hash of the document. It is infeasible to compute a signature on the hash of a document if the private key is not known. The resulting signature is kept in a separate file. StatCan would upload the signature file and its public key on its website, so that users can download (1) the file they want to use, (2) the signature file created over the hash of that document and (3) StatCan’s public key. Consider the signature verification as a function that asks the user to provide the three files that they downloaded from the website.

Step 3: Verifying a signature

Any user can verify the validity of the signature by providing (1) the file they want to check, (2) the signature file created over the hash of that document and (3) StatCan’s public key. If the signature is verified, the user can be sure that the file actually belongs to StatCan.

KEY IDEAS

- A digital signature is an electronic signature that is generated and verified by public key encryption.
- In a digital signature scheme, StatCan needs to (1) generate the public-private key pair, so that (2) it can sign the hash of the document with its private key, and (3) any user with the public key can verify the signature.

HOW CAN USERS MAKE SURE THAT THEY USE STATCAN’S KEY?

Public key infrastructure binds public keys with identities. This is done through a registration process where a certification authority (CA) issues certificates by signing StatCan’s public key. As a result, a CA verifies that the public key really belongs to StatCan. CAs are entities that issue certificates used to verify the ownership of a public key. Any user with access to the CA's public key can verify the certificate issued over StatCan's public key. The certificates are valid for a specific amount of time.

HASH FUNCTIONS

QUESTION 2: HAS THE DATA BEEN TAMPERED WITH?

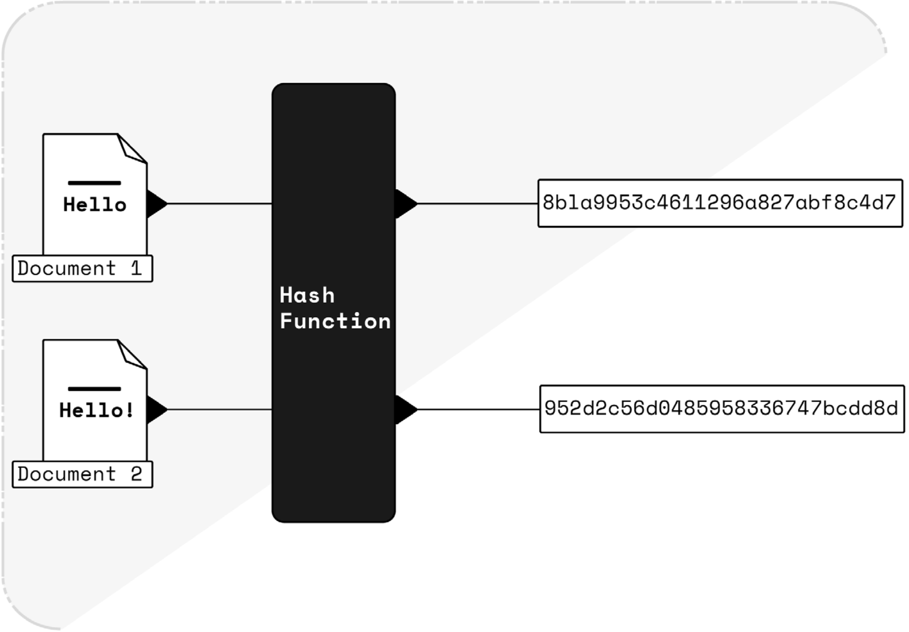
Hash functions are used to create a unique fingerprint for the input message. This technology gives StatCan the ability to hash a document (such as a CSV file) and create a unique "fingerprint" of it in the form of a fixed-size hash. Once StatCan computes the hash of the file, it uploads it to the website. When users download a file, the document is hashed. The resulting hash is compared with the uploaded value to make sure that the file has not been tampered with. This part of the process is handled by the application itself. We will explain this in more detail in the proposed solutions.

KEY IDEAS

- Hashing is a term that refers to the use of an algorithm, called a hash function, to convert a piece of information into an alphanumeric string of characters.

To solve users' concern about the authentication of their downloaded file, along with digital signatures, we must use hash functions in our solution. This is common practice in cryptography, as hash functions are known to be secure (Al-Kuwari, Davenport, and Bradford, 2011). They are used against malicious parties that may try to change data deliberately. Using hash functions fills a demand in our proposed system, because an attacker should not be able to create a file with a particular hash and replace it with a file from StatCan. For the hash functions to operate effectively, they require certain properties. For example, when two people hash the same document using the same hash function, they get the same hash value. The hash function produces the same output for a given input (which is also called "pre-image"); this means that hash functions are deterministic. Even if a single letter is added to a single cell in the document, the resulting hash will be different (see Figure 2). The determinism property is relevant in the context of guessing the pre-image. Input to the hash function cannot be computed by just looking at the hash value. However, one can try to guess the pre-image, hash it and compare it with the hash value. Consider user authentication—passwords are generally stored as hashes. If an attacker can access this database of hashes, they can pick a password (for example, one of the most commonly used passwords), hash it and compare it against the database to see whether there is a match.

FIGURE 2: An illustration of how hashing works



Note: This image illustrates how hashing works. Document 1 contains the word "Hello", and the hash function creates "Hash 1" over this document. The second document differs from Document 1 by one character: "Hello!" The hash function creates "Hash 2" over Document 2. Hash 1 and Hash 2 have different values, as Document 1 and Document 2 are different. Hash 1 and Hash 2 are the same size, as the hash function produces fixed-size outputs.

Most relevant to our project, it is imperative to note that we expect a hash function to have the collision resistance property, meaning that it is infeasible to find any two different messages that have the same hash. In other words, an adversary cannot find another CSV file with different content that has the same hash as the original document and cannot replace the original document with another one.

For the sake of a comprehensive understanding, we must also mention the other two properties that a hash function should have. To ensure clarity, note that a message to be hashed is known as the pre-image, and the resulting hash is known as an image. “Pre-image resistance” implies that given the hash of a message, it is infeasible to find a corresponding message. “Weak collision resistance” states that given a message, it is infeasible to find another message with the same hash. As previously mentioned, the hash function is also needed for the signing operation. StatCan signs the hash of the document, rather than the document itself, to have a shorter signature. This increases efficiency, as signing the hash is much faster. Since the hash is used in the signature function, we need the collision resistance property.

There are well-known hash functions, such as MD 5, SHA1, SHA2 and SHA3. However, not all are secure. MD 5 and SHA1 are proven to be insecure, as they do not have the collision resistance property. While it takes longer to attack SHA1 than MD 5, both are currently considered weak. Hash functions can break over time, but they get replaced with secure ones. For now, we know that SHA2 and SHA3 are secure (National Institute of Standards and Technology, 2015). As SHA3 is more secure than SHA2, we propose using SHA3 in our solution.

WHY NOT CHECKSUMS AND ERROR-CORRECTING CODES?

- Should checksums be used instead of hash functions? Checksums are truncated hashes, and they are not primarily secure; checksums are used to detect random faults. An adversary can manipulate checksums to change the data while ensuring the checksum does not change. Unlike hash, creating data (such as a file) with a particular checksum is not difficult. This property prevents the detection of errors. For these two reasons, checksums cannot protect against malicious adversaries.

- Error-correcting code is used to detect errors during the transmission of data over an unreliable channel. The message is encoded with redundant information. The receiver uses this redundant information to detect a limited number of errors. These errors can be corrected within certain limits. The errors can be corrected on the receiver side (i.e., retransmission of data is unnecessary).

SECURE TUNNELS

The proposed solutions require a secure tunnel between the user and the StatCan website for communication. In both the offline and hybrid solutions found below, the user has to download an application from the StatCan website. The user has to make sure they get the actual application, and a secure tunnel is needed between the user and StatCan for that purpose. Also, in the online solution, the user communicates with the StatCan website using the secure tunnel. “Https” provides a secure tunnel, meaning that if an attacker observes the traffic in the tunnel, they will not know the content of the message being transmitted. All an attacker can observe is that there is traffic between two parties.

The secure tunnel provides

1. message confidentiality: as the messages being transmitted are encrypted, an attacker cannot decipher them to read the content (they know only that there is a message transmission between two parties)

2. message integrity: an attacker on the path is not able to modify the traffic

3. server authentication: it is known where the tunnel ends, and it does not lead to the adversary.

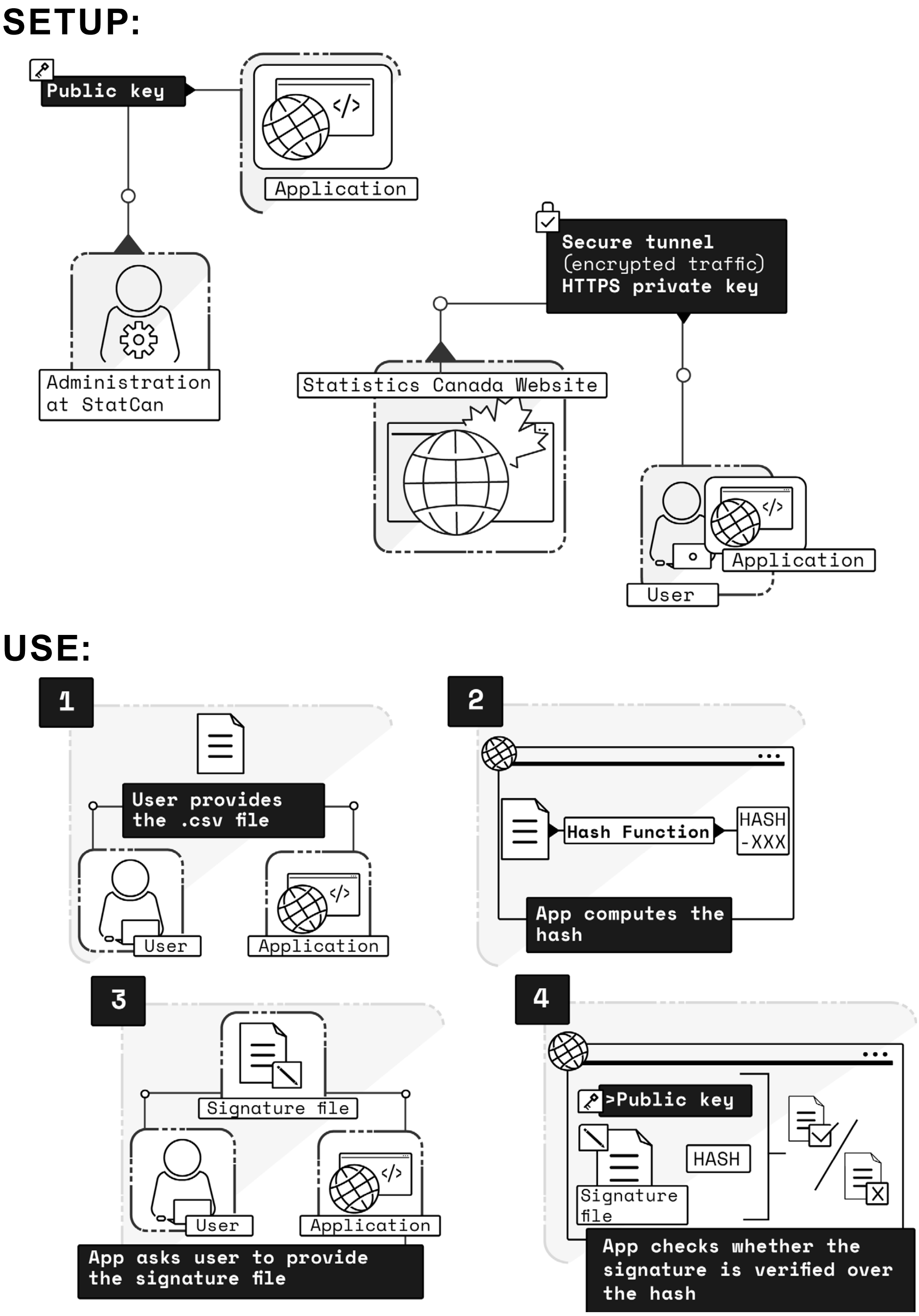
THREE POTENTIAL SOLUTIONS

There are three potential solutions that could be implemented using the previously mentioned technology to resolve user needs to authenticate a StatCan document. Offline and hybrid solutions require the creation of an application that is downloaded by the user. In these solutions, the user interacts with the application to check the validity of a document.

1. OFFLINE SOLUTION

In this solution, the user downloads an application from the StatCan website through the secure tunnel. This enables the user to ensure that the application they download belongs to StatCan. The application checks the validity of the user's document. The user takes the CSV file and signature file they downloaded from the website together, then drags and drops the CSV file into the app. The app computes the hash over the file, then prompts the user to provide the corresponding signature file computed over the hash of the CSV file. The application checks whether the signature is verified over the hash. To do so, the StatCan keys must be hard-coded into the app (setup phase in Figure 3). The key is needed to verify the signature over a file.

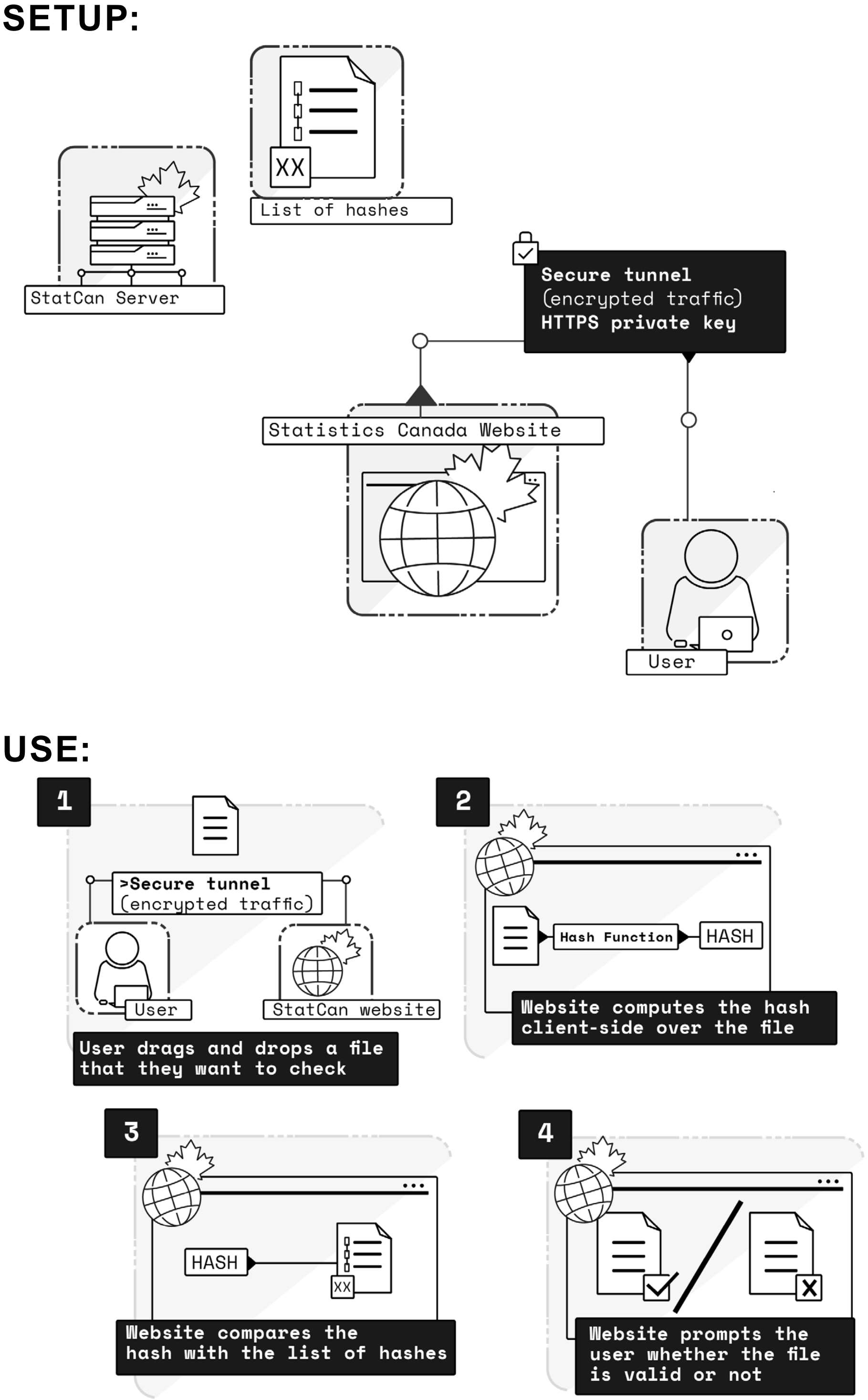
FIGURE 3: An illustration of our offline solution



2. ONLINE SOLUTION

In this solution, StatCan maintains a page on its website for the user to check document validity. The user communicates with the StatCan website using the secure tunnel, and they drag and drop a file that they want to check. Since the website knows the list of hashes of all files, it can compute the hash client side over the file provided by the user and compare it with the list; StatCan maintains a server where the list of hashes is kept. The user then learns whether the file they uploaded is valid. If valid, the file has not been tampered with and belongs to StatCan. Compared with the offline solution, this approach offers a more straightforward experience for the user, as they only have to provide the product's file. However, this solution requires the user to be online, unlike the previous application that runs offline.

FIGURE 4: An illustration of our online solution



1 User drags and drops a file that they want to check

2 Website computes the hash client-side over the file

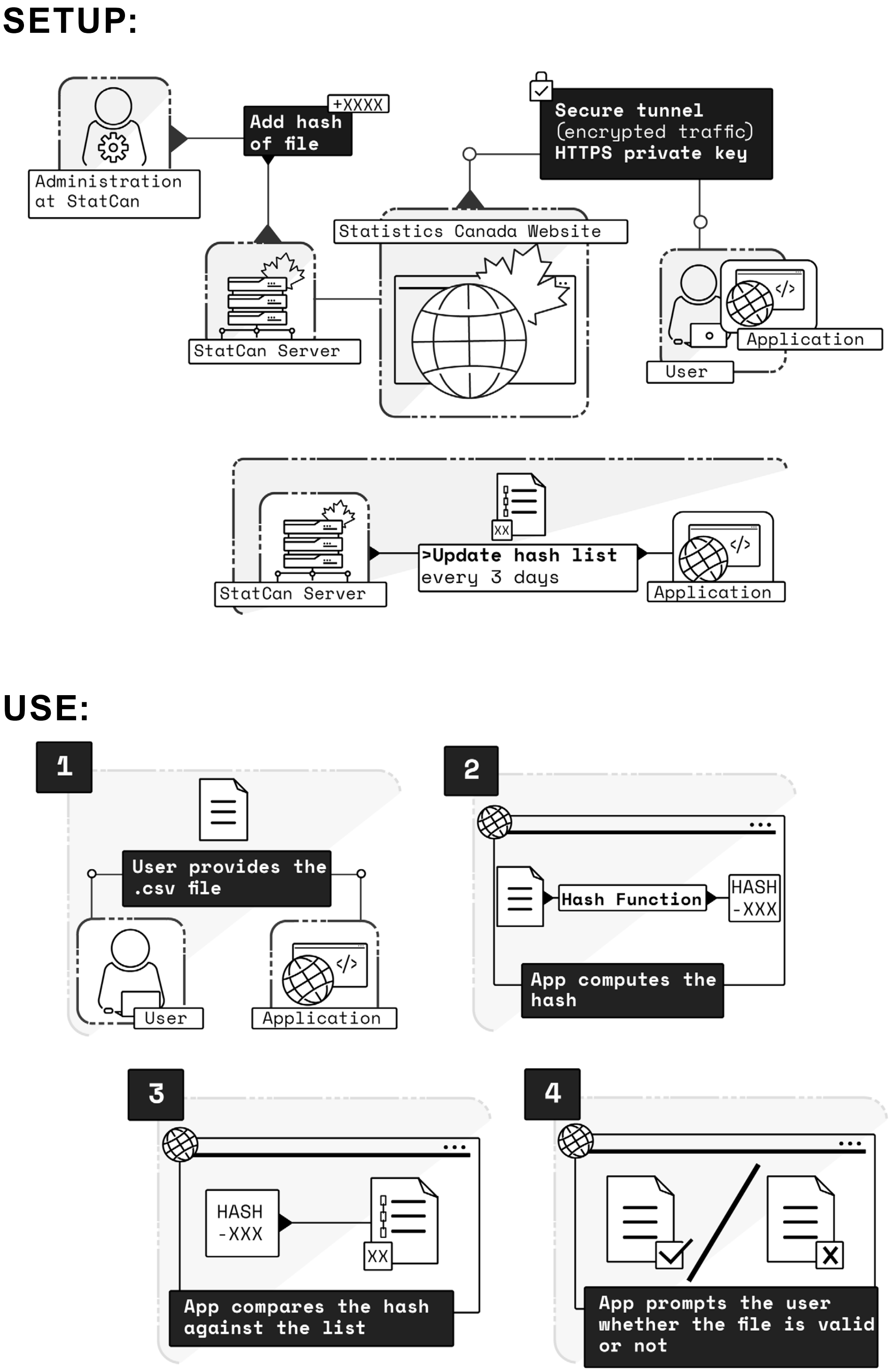
3 Website compares the hash with the list of hashes

4 Website prompts the user whether the file is valid or not

3. HYBRID SOLUTION

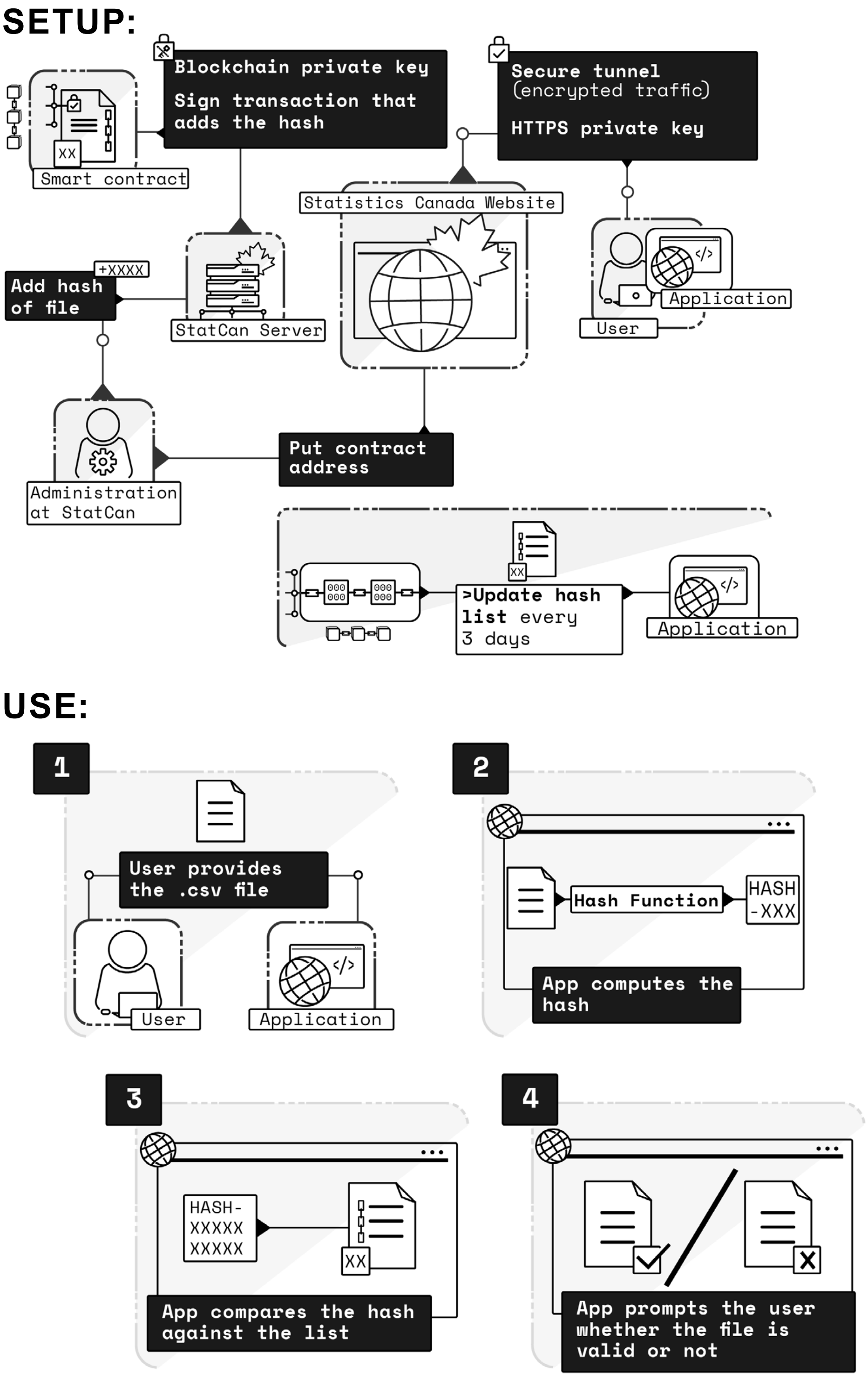
In the hybrid solution, the user must download an application (similar to the offline solution) over the secure tunnel. The app has a list of hashes of files that belong to StatCan. To authenticate the document, the user uploads the file to the app, which computes the hash and compares it with the list. Then, the app informs the user whether the file is valid. The app occasionally connects to the StatCan website to update the list of hashes; StatCan maintains a server where the list of hashes is kept. While we suggest that the app connect every three days, the duration can be greater or shorter, depending on how frequently StatCan shares files. Every three days, the app receives the updated list of hashes that is kept on the server to have the most recent list. A signature over a hash proves ownership. Receiving the list of hashes over the secure connection means that StatCan is the owner of the hashes. This solution eliminates the step of providing the signature file, if the hash of the file that the user offers appears in the list of hashes. If the hash is not in the list, the app prompts the user to provide the signature file over the hash, so the app can compute the hash and verify the signature over the file. This situation might occur if a user tries to authenticate a file before the app has the opportunity to connect to the StatCan website and update the list of hashes.

FIGURE 5: An illustration of our hybrid solution



RECOMMENDED SOLUTION: HYBRID+ BLOCKCHAIN

FIGURE 6: An illustration of our recommended solution


SETUP:
Blockchain private key
Sign transaction that adds the hash
Smart contract
XX
Secure tunnel (encrypted traffic)
HTTPS private key
Statistics Canada Website
+XXXX
Add hash of file
StatCan Server
User
Application
Administration at StatCan
Put contract address
>Update hash list every 3 days
Application
USE:
1
User provides the .csv file
User
Application
2
Hash Function
HASH -XXX
App computes the hash
3
HASH- XXXXX XXXXX
XX
App compares the hash against the list
4
App prompts the user whether the file is valid or not

All three solutions offer users the opportunity to authenticate data from the StatCan website. However, they do not all equally meet the standards we set in our objectives for the project. While the offline solution meets our objective of allowing users across the digital divide to authenticate data, it requires the user to submit the corresponding signature file to the app. With regard to the online solution, the user only needs to provide the CSV file, minimizing the number of downloads for the user. Therefore, the online solution offers better usability compared with the offline solution. However, the online solution does not meet the requirement to provide an accessible method of authentication, regardless of the user's access to the Internet.

For these reasons, we have decided that the hybrid solution is ideal because it provides a usability level comparable to the online solution and does not require the user to be online to check the file they have. This solution addresses the barriers discussed above regarding consistent access to the Internet. Adding blockchain to the hybrid solution provides improvements; it affects a subcomponent of the proposed solution—the way the hash of a file is stored. StatCan creates the hash of a file and logs this hash on the ledger. When compared with Figure 5, the hashes are logged on the blockchain, and the app receives the updated list of hashes from it. The added element of blockchain increases trust between StatCan and the public: StatCan cannot change the data once it is posted. If StatCan changes the data, a history of that change is recorded. Another benefit of including blockchain is that hashes can still be reached if the StatCan website is down, as they are recorded on the blockchain. Blockchain also offers better archival properties, as it ensures that the recorded data are reachable over a longer period than if the data are stored on a server. The server may go down or may not be continuously maintained, making the data unreachable. Blockchain provides provenance over the data (hash of the file) for a long time, but does not actually archive files. A possible drawback of incorporating blockchain into the hybrid solution is that if the ledger nodes manipulate the list of hashes, StatCan cannot do anything about it—a global network has control over the data. Ledger nodes are the entities in this network that accept or reject a block of transactions based on their validity; they broadcast these transactions so all of the nodes stay up to date. However, in the hybrid solution without blockchain, StatCan maintains exclusive control.

KEY IDEAS

- Our recommended solution is a hybrid solution with blockchain.
- In the hybrid solution with blockchain, authentication will occur through an application that users must download, like in the hybrid solution. The difference is that the list of hashes of files will be logged on the blockchain. This list will be updated periodically with the hashes of new StatCan products. The authentication of a file will occur as follows: the user will need to upload the file that requires authentication to the app. This action will prompt the app to compute the hash of this file and compare it with the list of existing hashes from StatCan products. The app receives this list of hashes from the blockchain. The app will then inform the user whether the file is valid.

APPENDIX A: TERMINOLOGY

BLOCKCHAIN

- "Distributed, peer-to-peer system for validating, time-stamping, and permanently storing transactions on a distributed ledger that uses cryptography to authenticate digital asset ownership and asset authenticity, and consensus algorithms to add validated transactions to the ledger and to ensure the ongoing integrity of the ledger's complete history" (Lacity, 2018, p. 41).
- Blockchain is digital and decentralized, and its goal is to create "permanent and tamper-proof records" (Treiblmaier, 2018, p. 547).
- According to A. Welfare (2019), blockchain has five important characteristics: truth and trust, transparency, security, certainty, and efficiency.
- In computing, a blockchain is "a sequence of digital records or 'blocks' linked using cryptography so that each block is verifiable and virtually unchangeable, which is distributed and managed typically in a peer-to-peer network" (OED, n.d.a).

CERTIFICATION AUTHORITY

- The Computer Security Resource Center (CSRC) defines a certification authority as "an entity authorized to create, sign, issue, and revoke public key certificates" (n.d.a).

CHECKSUM

- The Oxford English Dictionary defines checksum as "a sum calculated from the digits in a set of data and transmitted or stored with the data to provide a means of automatic checking for any subsequent corruption." (n.d.b).

COLLISION

- The CSRC explains that in the context of hash functions, collision is when "two or more distinct inputs produce the same output" (n.d.b).

COLLISION RESISTANCE PROPERTY

- The CSRC states that collision resistance property is "an expected property of a cryptographic hash function whereby it is computationally infeasible to find a collision" (n.d.c).

DETERMINISM PROPERTY

- Hash functions are always deterministic in nature, meaning the hash value is "purely determined by its inputs, where no randomness is involved in the model" (Techopedia, 2019). Hash functions "will always come up with the same result given the same input" (Techopedia, 2019).

DIGITAL SIGNATURE

- According to the Oxford English Dictionary, a digital signature is "an electronic signature … one generated and verified by public key encryption" (OED, n.d.c).

DISTRIBUTED LEDGER TECHNOLOGY (DLT)

- A ledger is "a record-book; a book that lies permanently in some place; ordinarily employed for recording mercantile transactions" (OED, n.d.d).
- "DLT-based platforms refer to a set of technologies and different fields of science (like algebra and statistics) that, grouped together, enable the storage and exchange of data in a decentralized and secure manner for which no central regulator is required" (Lesmes, 2019, np).
- They "facilitate the storage, processing, and exchange of data, bringing higher speed and greater transparency and reliability than traditional legacy systems" (Lesmes, 2019, np).

ERROR-CORRECTING CODE (ECC)

- Katz and Dash describe ECCs as "an encoding scheme that transmits messages as binary numbers, in such a way that the message can be recovered even if some bits are erroneously flipped. They are used in practically all cases of message transmission, especially in data storage where ECCs defend against data corruption" (n.d.).

HASH OR HASHING

- "Hashing is a term that refers to the use of an algorithm – called a hash function – to convert a piece of information into an alphanumeric string of characters like this: e9ffc424b79f4f6ab42d11c81156d3a17228d-6b1edf4139be78e948a9332d7d8. Hashing is a commonly used technique in computer science and cryptography" (Urban and Pineda, 2018, p. 16).
- "The manipulation of log files can take many different expressions, but the possibilities to recognize such manipulations also vary greatly. 'File verification' mechanisms seek to ensure that a file has not been changed. For example, checksum or hash techniques can be used to verify content, authors or digital ownership" (Radinger-Peer and Kolm, 2020, p. 134).
- In computing, hashing is "the assignment of a numeric or alphanumeric string to a piece of data via the application of a function whose output values are all the same number of bits in length" (OED, n.d.e).

KEY GENERATION

- The process of key generation can be completed "either as a single process using a random bit generator and an approved set of rules, or as created during key agreement or key derivation" (CSRC, n.d.d).

MESSAGE INTEGRITY

- Message integrity refers to "the validity of a transmitted message. Message integrity means that a message has not been tampered with or altered" (PCmag Encyclopedia, n.d.).

MESSAGE DIGEST

- Message digest can be understood as "the fixed-length bit string produced by a hash function," according to the CSRC (n.d.e). Synonyms of this term are "digital fingerprint, hash output, or hash value" (CSRC, n.d.e).

NODE

- The CSRC defines a node as "an individual system within the blockchain network" (n.d.f).

NON-FUNGIBLE TOKEN (NFT)

- According to Merriam-Webster, NFTs are "a unique digital identifier that cannot be copied, substituted, or subdivided, that is recorded in a blockchain, and that is used to certify authenticity and ownership (as of a specific digital asset and specific rights relating to it)" (n.d.).

PRE-IMAGE

- The CSRC defines a pre-image as "a message X that produces a given message digest when it is processed by a hash function" (n.d.g).

PRE-IMAGE RESISTANCE

- This property of hash functions is defined by the CSRC as being "an expected property of a cryptographic hash function such that, given a randomly chosen message digest, message_ digest, it is computationally infeasible to find a preimage of the message_digest" (n.d.h). Please see message digest and pre-image if further clarification is necessary.

PROOF OF WORK (POW)

- Chowdhury explains that "PoW is an economic measure to deter denial-of-service attacks and other service abuses such as spam on a network by requiring some work from the service requester, usually meaning processing time by a computer" (2019, p. 56).

PROOF OF STAKE (POS)

- Chowdhury explains PoS by comparing it with PoW, saying, “instead of making operations expensive by consuming electricity, this system requires miners to deposit a wealth that bad actors will lose if [they] try to bend the rules” (2019, p. 19).

PROVENANCE

- According to the Oxford English Dictionary, provenance is “the fact of coming from some particular source or quarter; origin, derivation,” and in relation to the notion established by the arts field, “the history of the ownership of a work of art or an antique, used as a guide to authenticity or quality; a documented record of this.” (n.d.f)
- “Data provenance tracks the origin of information with the goal of improving trust among interested parties. Data provenance is an important requirement for a range of applications such as food safety, supply chains, and tracking of epidemic outbreaks. Many of these applications are inherently distributed and require high levels of privacy and trust” (Lautert, Pigatto, and Gomes 2020, p. 1).
- “Provenance is the process or techniques utilized to track the origin, authorship and history of any given object. It was originally used in the context of works of art to make sure that an object was created by the claimed author” (Lautert, Pigatto, and Gomes 2020, p. 1).

PUBLIC KEY

- The Oxford English Dictionary defines a public key as “a cryptographic key that can be obtained and used by anyone to encrypt messages in such a way that the encrypted messages can be deciphered only by using a second ‘private’ key known only to the recipient.” (n.d.g).

PUBLIC KEY INFRASTRUCTURE

- The CSRC explains that public key infrastructure is “the architecture, organization, techniques, practices, and procedures that collectively support the implementation and operation of a certificate-based public key cryptographic system. Framework established to issue, maintain, and revoke public key certificates” (n.d.i).

WEAK COLLISION RESISTANCE

- Hirose defines weak collision resistance as “the probability of failure to find a collision is not negligible” (2004, p. 88).

APPENDIX B: METHODS

KEYWORDS

- Authentication
- Certification
- Verification
- Applied
- Distributed ledger technology
- DLT
- Cryptography
- Data
- Government
- Blockchain
- Cryptographic key
- Hash function
- Digital signature
- Data flow
- Tool
- Checksum
- Data provenance
- Digital fingerprint
- Data security
- Unique identifiers
- Error-correcting codes

SEARCHES

1. (“Distributed ledger technology”) OR (DLT)
2. (“Blockchain”) AND (“Authent*) OR (Certif*) OR (Verif*)
3. (“Digital signature”)
4. (“Hash function”) AND (“Cryptography”)
5. (“Government”) AND (“Applied”) AND (“Cryptographic key”)
6. (“Government”) AND (“Applied”) AND (“Distributed ledger technology”) OR (“DLT”)
7. (“Government”) AND (“Applied”) AND (“Digital fingerprint”)
8. (“Error-correcting codes”) AND (“Applied”) NOT (“Algebra”)
9. (“Error-correcting codes”) AND (“Applied”) AND (“Government”)
10. (“Checksum”) AND (“Verif*) AND (“Applied”)
11. (“Data provenance”) AND (“Blockchain”)
12. (“Data flow”) AND (“Data provenance”)
13. (“Digital signature”) OR (“Digital fingerprint”) AND (“Data security”)
14. (“Unique identifiers”) AND (“Data security”)
15. (“Error-correcting codes”)

REFERENCES

Al-Kuwari, S., Davenport, J. H., and Bradford, R. J. (2011). Cryptographic Hash Functions: Recent Design Trends and Security Notions. Short Paper Proceedings of Inscrypt '10, 1–37. Retrieved 2022, from https://eprint.iacr.org/2011/565.pdf.

Aljeaid, D., Ma, X., and Langensiepen, C. (2014). Biometric identity-based cryptography for e-Gov ernment environment. Proceedings of 2014 Science and Information Conference, SAI 2014. 581-588. https://doi.org/10.1109/SAI.2014.6918245.

Batista, D., Kim, H., Lemieux, V. L., Stancic, H., and Unnithan, C. (2021). Block and provenance: How a technical system for tracing origins, ownership and authenticity can transform social trust. In V. Lemieux and C. Feng (eds.) Building decentralized trust: Multidisciplinary perspectives on the design of blockchains and distributed ledgers (First ed., pp. 111–128). Springer. https://doi.org/10.1007/978-3- 030-54414-0.

Bell, M., Green, A., Sheridan, J., Collomosse, J., Cooper, D., Bui, T., Thereaux, O., & Higgins, J. (2019). Underscoring archival authenticity with blockchain technology. Insights: The UKSG Journal, 32. https://link.gale.com/apps/doc/A594619025/AONE?u=anon~eafb9693&sid=googleScholar&x-id=e1544e71

Canada's Public Policy Forum. (2014). Northern Connections: Broadband and Canada's Digital Divide. Public Policy Forum: Reports. Retrieved January 27, 2022, from https://ucarecdn.com/68b98fff-32c9-4904-904c-09b1d98cdd2e/

Chandler, S. (2021, December 22). Proof of stake vs. proof of work: Key differences between these methods of verifying cryptocurrency transactions. Business Insider. https://www.businessinsider.com/personal-finance/proof-of-stake-vs-proof-of-work

Chowdhury, N. (2019). Inside blockchain, Bitcoin, and cryptocurrencies. Auerbach.

Communications Security Establishment Canada. (2021, December 6). Ministers urge Canadian organizations to take action against Ransomware. Canada.ca. Retrieved January 26, 2022, https://www.canada.ca/en/communications-security/news/2021/12/ministers-urge-canadian-organizations-to-take-action-against-ransomware.html

Computer Security Resource Center. (n.d.a). Certification authority. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/certification_authority

Computer Security Resource Center. (n.d.b). Collision. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/collision

Computer Security Resource Center. (n.d.c). Collision resistance. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/collision_resistance

Computer Security Resource Center. (n.d.d). Key generation. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/key_generation

Computer Security Resource Center. (n.d.e). Message digest. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/message_digest

Computer Security Resource Center. (n.d.f). Node. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/node

Computer Security Resource Center. (n.d.g). Preimage. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/preimage

Computer Security Resource Center. (n.d.h). Preimage resistance. In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/preimage_resistance

Computer Security Resource Center. (n.d.i). Public key infrastructure (PKI). In Computer Security Resource Center: Glossary. Retrieved from, https://csrc.nist.gov/glossary/term/public_key_infrastructure.

De Filippi, P. D. F. (2018). Blockchain and the law: The rule of code. Harvard University Press. https://doi.org/10.4159/9780674985933

Hirose, S. (2004). Yet another definition of weak collision resistance and its analysis. In Lim JI., Lee DH. (eds) Information Security and Cryptology - ICISC 2003. ICISC 2003. Lecture Notes in Computer Science, vol. 2971. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24691-6_8

Huang, J., O'Neill, C., and Tabuchi, H. (2021, September 3). Bitcoin uses more electricity than many countries. How is that possible? The New York Times. https://www.nytimes.com/interactive/2021/09/03/climate/bitcoin-carbon-footprint-electricity.html

Industrial Research Assistance Program. (2019). Blockchain publishing prototype. National Research Council of Canada. Government of Canada. https://nrc-cnrc.explorecatena.com/en

Ipsos Public Affairs for Canada's Centre for International Governance Innovation. (2019). Global Survey Internet Security & Trust. CIGI-Ipsos Global Survey Internet Security Trust Part I & II: Internet security, online privacy & trust. Retrieved from, https://www.cigionline.org/sites/default/files/documents/2019%20CIGI-Ipsos%20Global%20Survey%20-%20Part%201%20%26%202%20Internet%20Security%2C%20Online%20Privacy%20%26%20Trust.pdf

Jiang, S. (2021, November 9). Discord's hints about crypto, NFTs are tearing its community apart. Kotaku. https://kotaku.com/discords-hints-about-crypto-nfts-are-tearing-its-commu-1848023955

Katz, A., & Dash, S. (n.d.). Error correcting codes. Brilliant Math & Science Wiki. Retrieved January 27, 2022, from https://brilliant.org/wiki/error-correcting-codes/

Kickstarter. (2022). Let's build what's next for crowdfunding creative projects. Kickstarter. https://www.kickstarter.com/articles/lets-build-whats-next-for-crowdfunding-creative-projects?ref=section-homepage-promo-the-future-of-crowdfunding-creative-projects

Lacity, M. (2018). A manager's guide to blockchains for business: From knowing what to knowing how. SB Publishing.

Lautert, F., Pigatto, D. F., and Gomes, L. (2020). A fog architecture for privacy-preserving data provenance using blockchains. Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC). https://doi.org/10.1109/ISCC50000.2020.9219724

Lemieux, V. (2016a). Trusting records: Is Blockchain technology the answer? Records Management Journal, 26(2), 110–139. https://doi.org/10.1108/RMJ-12-2015-0042

Lemieux, V. (2016b). Blockchain for recordkeeping: Help or hype? SSHRC Technical Report, p. 1–31.

Lemieux, V. (2019). Blockchain and public record keeping: Of temples, prisons, and the (re) configuration of power. Frontiers Blockchain, 2, n.p. https://doi.org/10.3389/fbloc.2019.00005

Lesmes, J. (2019). The internet of value: How distributed ledger technologies will reshape the financial services industry (First ed.). O'Reilly Media.

Maull, R., Godsiff, P., Mulligan, C., Brown, A., & Kewell, B. (2017). Distributed ledger technology: Applications and implications. Strategic Change, 26(5), 481–489. https://doi.org/10.1002/jsc.2148

Merriam-Webster. (n.d.). Non-fungible token. In Merriam-Webster.com dictionary. Retrieved January 26, 2022, from https://www.merriam-webster.com/dictionary/non-fungible%20token

Mohamed, K. S. (2020). New frontiers in cryptography: Quantum, blockchain, lightweight, chaotic and DNA (1st ed.). Springer. https://doi.org/10.1007/978-3-030-58996-7

Morse, J. (2021, December 16). Kickstarter said it's moving to the blockchain, and creators are pissed: Decentralized frustration. Mashable. https://mashable.com/article/kickstarter-protocol-blockchain-creator-reaction

National Institute of Standards and Technology. (2015, August 5). NIST releases SHA-3 Cryptographic Hash Standard. NIST: News. Retrieved January 27, 2022, from https://www.nist.gov/news-events/news/2015/08/nist-releases-sha-3-cryptographic-hash-standard

National Research Council Canada. (2018). Exploring blockchain for better business. Government of Canada. https://nrc.canada.ca/en/stories/exploring-blockchain-better-business

Ølnes, S., Ubacht, J., & Janssen, M. (2017). Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. Government Information Quarterly, 34(3), 355–364. https://doi.org/10.1016/j.giq.2017.09.007

Oxford English Dictionary. (n.d.a). Blockchain. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

Oxford English Dictionary. (n.d.b). Checksum. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

Oxford English Dictionary. (n.d.c). Digital Signature. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

Oxford English Dictionary. (n.d.d). Distributed Ledger Technology. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

Oxford English Dictionary. (n.d.e). Hashing. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

Oxford English Dictionary. (n.d.f). Provenance. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

Oxford English Dictionary. (n.d.g). Public Key. In Oxford English Dictionary. Retrieved from https://www-oed-com.proxy.library.carleton.ca

PCmag Encyclopedia. (n.d.). Message Integrity. PCmag. Retrieved January 27, 2022, from https://www.pcmag.com/encyclopedia/term/message-integrity#:~:text=Message%20integrity%20means%20that%20a,been%20tampered%20with%20or%20altered.&amp; text=Integrity%20checking%20is%20one%20component, Parkerian%20Hexad%20and%20data%20integrity.

Pearson, J. (2021, November 11). Discord backs off of crypto after entire internet yells at CEO. Vice. https://www.vice.com/en/article/7kb9dg/discord-backs-off-of-crypto-after-entire-internet-yells-at-ceo

Plunkett, L. (2021, December 17). Kickstarter announces blockchain future, doubles down after users say “no thank you.” Kotaku. https://kotaku.com/kickstarter-announces-blockchain-future-doubles-down-a-1848231993

Prathibha, Sona, T. R., & Krishna Priya, J. (2021). Secured Storage and Verification of Documents Using Blockchain Technology. In Transforming Cybersecurity Solutions using Blockchain (pp. 71–90). Springer Singapore. https://doi.org/10.1007/978-981-33-6858-3_5

Radinger-Peer, W. and Kolm, B. (2020). A blockchain-driven approach to fulfill the GDPR recording requirements. In Treiblmaier, H. and Clohessy, T. (Eds.), Blockchain and distributed ledger technology use cases: Applications and lessons learned (pp. 133-148). Springer. https://doi.org/10.1007/978-3-030-44337-5

Ruoti, S., Kaiser, B., Yerukhimovich, A., Clark, J., and Cunningham, R. (2020). Blockchain Technology: What is it good for? Communications of the ACM, 63(1), 46–53. https://doi.org/10.1145/3369752

Solana. (2021, November 24). Solana’s energy use report: November 2021. Solana. https://solana.com/news/solana-energy-usage-report-november-2021

Statistics Act, Revised Statutes of Canada (1985, c. S-19). Retrieved from the Justice Laws website: https://laws.justice.gc.ca/eng/acts/S-19/

Statistics Canada. (2018, October 5). Acts and regulations. Statistics Canada. Retrieved January 26, 2022, from https://www.statcan.gc.ca/en/about/frp/frp?MM=as

Statistics Canada. (2021). Departmental Results Report (Catalogue no. 11-628-X). Retrieved from, https://www.statcan.gc.ca/en/about/drr/2020-2021/index

Techopedia. (2019, August 29). What is deterministic algorithm?—definition from Techopedia. Techopedia.com. Retrieved January 28, 2022, from https://www.techopedia.com/definition/18830/deterministic-algorithm

Treiblmaier, H. (2018). The impact of the blockchain on the supply chain: A theory-based research framework and a call for action. Supply Chain Management, 23(6), 545–559. https://doi.org/10.1108/SCM-01-2018-0029

Treiblmaier, H., & Clohessy, T. (2020). Preface. In H. Treiblmaier and T. Clohessy (eds.), Blockchain and distributed ledger technology use cases: Applications and lessons learned (1st ed., pp. v-vii). Springer International Publishing. https://doi.org/10.1007/978-3-030-44337-5

Urban, M. C. and Pineda, D. (2018). Inside the black blocks: A policymaker’s introduction to blockchain, distributed ledger technology and the “internet of value.” Mowat Centre for Policy Innovation, University of Toronto.

Welfare, A. (2019). Commercializing blockchain: Strategic applications in the real world. Wiley.

Xiao, Y. and Watson, M. (2019). Guidance on conducting a systematic literature review. Journal of Planning Education and Research, 39(1), 93–112. https://doi.org/10.1177/0739456X17723971

Zheng, X., Zhu, Y., & Si, X. (2019). A survey on challenges and progresses in blockchain technologies: A performance and security perspective. Applied Sciences, 9(22), 1–24. https://doi.org/10.3390/app9224731