TAX AUDIT IN THE ERA OF BIG DATA: THE CASE OF INDONESIA

Agung Darono[1], Aldi Pratama[2,3]

Abstract

This research uses an interpretive case study strategy to investigate how big data affects tax audits in Indonesia, both with regard to tax audit management and policy, and to tax auditors' individual audit assignments. The study reveals that the impact of big data on tax audit exists in two aspects. First, at audit policy level, big data is used as part of risk analysis in order to determine which taxpayers should be audited. Second, at the individual tax audit assignment level, tax auditors must utilise big data in order to acquire and analyse data from taxpayers and other related parties. Big data has the following characteristics: it involves huge volumes of information, it is generated at a high velocity, it includes a wide array of data types, and it contains high uncertainty. Big data can be analysed in order to reinforce the results gained from risk engines as a part of a compliance risk management system at the audit policy level. Meanwhile, at the individual tax audit assignment level, empirical evidence shows that tax auditors may deal with: (1) large volumes of data (hundreds of millions of records) that originated from previous fiscal years (historical records); (2) variations in the format and sources of data acquired from taxpayers which, to some extent, may be giving an auditor the authority to request data in a format that suits their analytical tools—with an inherent risk that the data can only be acquired in its native format; (3) data veracity that requires the tax auditors to review data sources because the adopted data analysis techniques are determined by the validity of data under audit. The main benefit expected to be gained from the implementation of big data analytics in respect of tax audits is the provision of valid and reliable information that evidences that taxpayers are compliant with tax laws.

Keywords: Audit Policy, Audit Test, Big Data, Data Compatibility, Data Veracity.

1. BACKGROUND AND RELATED WORKS

Tax audit is one of the main features of tax administration. It is conducted in order to determine whether taxpayers have paid their taxes in accordance with the tax laws. As with any other type of audit, such as financial, compliance, or operational audits (see, for example, Arens et al., 2017), one of the crucial stages of a tax audit is that of evidence collection. In recent times, electronic data has become the dominant type of evidence of business transactions due to the extensive implementation of information technology (IT) in the business environment. Electronic data, in this context, has also evolved into big data. The term "big data" is used to describe a huge volume of data that has a rapid growth rate and is presented in various formats that cannot be processed using traditional data processing tools (Edery, 2016; Luisi, 2014; McAfee et al., 2012; Microsoft & PricewaterhouseCoopers [PwC], 2018; Organisation for

[1] Agung Darono, Tax Education and Training Centre, Ministry of Finance of the Republic of Indonesia, Jakarta Barat, Indonesia. Email: agungdarono@kemenkeu.go.id

[2] Aldi Pratama, Tax Education and Training Centre, Ministry of Finance of the Republic of Indonesia, Jakarta Barat, Indonesia. Email: aldi.pratama@kemenkeu.go.id

[3] The views expressed in this paper are those of the authors and do not represent the institution with which the authors are affiliated. The authors would like to thank the anonymous reviewers and Prof. Nigar Hashimzade for their invaluable comments. Any errors remaining are those of the authors.

Economic Co-operation and Development [OECD], 2016b; Schroeck et al., 2012; Vasarhelyi et al., 2015).

In global tax administration practice, big data is one of the most critical issues being discussed, particularly as it relates to the development of the digital economy (OECD, 2015). Several studies explain how tax authorities have begun to harness its powers within their tax administration business processes. Big data plays a significant role in a number of tax administration functions, including risk analysis, tax compliance monitoring, law enforcement, dispute resolution, and upstream compliance (Chen et al., 2015; Cockfield, 2016; OECD, 2016a; Veit, 2019). Moreover, Dimitropoulou et al. (2018) propose a framework in which big data would be utilised as part of the tax dispute settlement procedure using the mutual agreement procedure scheme. On the other hand, Brink and Hansen (2018), and Volvach and Solovyev (2018), emphasise how taxpayers could use big data to identify their tax risks or develop their tax planning activities.

Consequently, the existence of big data requires tax authorities to process and analyse electronic data, and to communicate the information acquired as competent and adequate audit evidence. Indonesia's tax authority, the Direktorat Jenderal Pajak (DJP), or Directorate General of Tax, is in the same situation (DJP, 2017; Djuniardi, 2016). The most critical issue is how tax auditors in Indonesia handle big data when conducting tax audits (in term of collecting and analysing audit evidence). Under the provisions of Law 6/1983 and its amendments, which concern general provisions and tax procedures (hereinafter "KUP law"), the DJP is authorised to collect evidence (including electronic data) in order to assess taxpayers' compliance. The assessment is conducted through a tax audit, in which the tax authority tests the evidence that it has collected. The result is a tax provision that states whether the tax payments that have been made are appropriate, underpayments, or overpayments, so that adjustments can be made. In its Information Technology Blueprint, the DJP states that big data analytics is one of the pillars of information and communication technology development that support the tax administration process (DJP, 2015). Accordingly, the DJP (in its internal training material) and Djuniardi (2016, 2018) explain that the DJP gradually selected several sample cases in which to utilise big data. These cases involve the identification of transactions for tax evasion purposes using transfer pricing schemes, value-added tax (VAT) invoice fraud, and asset tracing for tax arrears collection purposes (DJP, 2022; Sakti, 2021).

Based on the explanation above, it is necessary to explore how big data affects tax audits in more depth. This paper will use the Indonesian tax administration as a case study and examine how big data influences tax audits, both at audit policy level and at individual tax audit level. Indonesia was selected because of its significant economic size when compared to other G20 countries (see, for example, The Jakarta Post Editorial Board, 2019; G20 Sherpa Indonesia, 2019) and its successful tax reforms (see, for example, Lewis, 2019). Moreover, this study will help the Indonesian tax administration's stakeholders to understand how the Indonesian tax authority deals with big data issues when conducting tax audits. This study is an interpretive case study and is expected to explain the empirical situation by interpreting the authors' experiences, knowledge, and perspectives, while utilising verification procedures adopted by relevant disciplines.

Tax authorities must determine which taxpayers should be audited, based on limited resources. Consequently, they can only conduct audits in respect of a small number of registered taxpayers. The typical approach used to select taxpayers for audit is based on non-compliance risk analysis. Therefore, the tax authority requires adequate supporting data (OECD, 2004;

Pratomo, 2018). The next crucial issue faced by the tax authority is how to prepare an audit programme, particularly in respect of the acquisition and testing of audit evidence. Many factors determine an audit programme's design. One of the essential factors is the availability of data, both from taxpayers and other related parties.

The paragraphs above describe how tax authorities in many tax jurisdictions take the initiative to utilise big data in order to support their tax administration process, including tax audits. The following questions may then arise: what is big data and how can the use of it help tax authorities to perform their function?

According to Diebold (2012), the term "big data" began to be used by scholars and IT professionals, such as himself, John Mashey, Sholom M. Weiss, Nitin Indurkhya, and Douglas Laney between 1998 and 2001. The most referenced definition of big data is Gartner's[4], which is based on Laney's (2001) notion. According to Gartner, "big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation". Some scholars have tried to provide a comprehensive definition of big data. The definition proposed by Hu et al. (2014) attempts to cover several points of view:

(1) an "attributive" definition, as first noted by Laney (2001), which is popular as "the 3Vs" ("increasing volume, velocity, and variety") (Hu et al., 2014, p. 654). Gantz & Reinsel (2011) then added "veracity" (so the definition became "the 4Vs") to show that the content of big data has various degrees of validity that require data scientists to take a different approach when testing it;
(2) a "comparative" definition (see, for example, Manyika et al., 2011), i.e. big data deals with a huge volume of data that cannot be managed by common database software;
(3) an "architectural" definition (see, for example, Chang et al., 2018): big data is an efficient processing method when traditional database approaches and tools cannot be used due to data volume and velocity.

It is essential to note that big data is not a substitute for the relational databases or data warehouses that have been used to manage organisational data so far. Instead, it expands the data types, and the storage and search procedures available (Hu et al., 2014; Manyika et al., 2011; OECD, 2015). One big data characteristic that also applies to traditional databases is the considerable volume of data stored. In this context, several parties from both the industry and academic fields have proposed a solution known as the very large database (VLDB). This solution tends to take a "scale up" approach, in which the underlying relational database is continuously developed via additional hardware capacity, especially computer memory and storage systems. On the other hand, there is also a solution that tends to take a "scale out" approach. The development of data management capabilities also considers the diversity of data formats, which is not always in the form of interrelated tabular data but sometimes in the form of document networks.

Indeed, in the IT field, the terms "volume" and "speed" become relative. Diebold (2012) states that, in the field of econometrics, any more than 200 gigabytes (GB) of data is considered to be a large data set. However, in physics, experimental data sets usually contain much larger amounts of data. For example, according to Gaillard (2017), during Higgs boson particle

[4] https://www.gartner.com/en/information-technology/glossary/big-data

research, particle collisions produce about 1 petabyte (1 petabyte = 1,000,000 GB) of data within one second. Helskyaho (2017) even claims that data is considered to be big "when traditional processing with traditional tools is not possible due to the amount or the complexity of the data" (Slide 21). Moreover, the term "big" should not only be interpreted as "large in volume", which would be relative in this circumstance because it would depend upon the computing environment being used.

In the authors' view, the use of the term big data is more about the capability of data management tools to deal with more data types or formats in much larger sizes, while maintaining processing speed and ensuring data validity. Previously, organisations stored structured data within relational databases in tabular form and only included limited data types, such as numbers, text, and dates. Big data repository tools are able to manage advanced data formats, such as spatial data, global positioning systems (GPS), clickstreams, sensor devices, log files, images, audio files, videos, and other forms of unstructured data, as well as traditional data formats (Hu et al., 2014; Podesta et al., 2014; Vasarhelyi et al., 2015).

At the same time, the speed of data growth makes data volumes even larger. The increasing variety of inputs is not limited to traditional hardware, such as keyboards, mice, and barcode readers, but also includes sensory equipment connected to the network (known as the "Internet of Things" or the IoT). A couple of years ago, data input was performed only by the end user (i.e. human action). Various types of new hardware, such as sensors, generate data, thus data size increases exponentially. Consequently, IT specialists cannot rely on the traditional relational database management system software that has been standard for organisational data management since the 1980s. More non-relational database models are now available, such as the graph model, columnar data, document, and multi-model databases (Helskyaho, 2017; Lu & Holubová, 2019).

The following is a short illustration of how a typical transaction in our daily life, realised or not, involves the extensive use of big data. Someone uses a social media platform. In their timeline, there is an offer for a coffee maker from an online store that sells products via an e-marketplace. The user presses the "like" button and visits an online store using the link provided. A few moments later, after clicking a few times to see photos of, and videos about, the coffee maker, the person decides to buy it. They use a delivery service from a ride-hailing courier provided by the e-marketplace partner. The purchaser then pays for the coffee maker using a mobile banking application that is integrated with the e-commerce platform. About five hours later, the courier delivers the coffee maker to the buyer's front door. The buyer checks their new coffee maker to make sure that it works properly, then uses their smartphone to confirm that they have received the goods. That simple transaction involves the creation of hundreds (or even thousands) of data records by each party involved.

This type of transaction can happen thousands or even tens of thousands of times in one day on just one e-commerce platform. Each transaction may create dozens or hundreds of rows of data that are stored by the various parties associated with the transaction, including e-marketplaces, banks, ride-hailing couriers and mobile phone operators. In short, that is how big data works and grows. Data proliferates in various formats, including customer profile photos, product videos, clickstreams, the buyer's GPS location, and details of the delivery route taken. Data validation also happens in many ways: payment transaction data must be precise, for example, while ride-hailing couriers may receive delivery route suggestions via their smartphones. This illustration depicts how the digital economy works. Tax authorities must also find ways by which to capture the data generated from every single part of each

transaction. By collecting the data from the transaction above, for example, tax authorities can assess how every party involved complied with the tax rules for any combination of the size, type, and validity of the data.

Hence, from the authors' perspective, big data is a collection of data that grows very fast because it is generated from many sources. It comes from known input devices (e.g., mice, keyboards, scanners, QR-code/barcode readers, electronic data capture machines, and so on) and interconnected sensory devices within communication networks. The development of the data processing technology landscape has enabled the development of what we called big data when it was first introduced a decade ago. Big data is the data that we currently generate every day; data itself is now big data. Most people have used it whether they realised it at the time or not. Today, the main challenges arising from the use of big data involve how to facilitate the convergence and standardisation of devices and tools relating to it, so that the data can be analysed quickly (Helskyaho, 2017; Hu et al., 2014; Lu & Holubová, 2019).

As a result of big data's characteristics, tools and frameworks are required in order to prepare, process, and analyse it. Therefore, some studies discuss big data analytics as part of the wider business process (Alles & Gray, 2016; Davenport & Harris, 2007; Houser & Sanders, 2017; Pijnenburg et al., 2017; Santos et al., 2018; Tian et al., 2017). Big data analytics, in principle, also uses frameworks that are already available: (1) descriptive (in order to obtain data distribution); (2) diagnostic (in order to integrate data and identify relationships between data); (3) predictive (in order to make predictions based on existing data); and (4) prescriptive (in order to present suggested actions based on data collected (Richardson et al., 2019). The main difference in the way in which traditional databases (structured data) and big data are analysed stems from the capabilities of the tools used for the analysis. Big data analytics requires tools or devices that can deal with large volumes of data. For example, Microsoft Excel can only access up to one million rows of structured data. Therefore, we cannot use common spreadsheet tools like Microsoft Excel to process big data. Instead, we are required to utilise big data tools, such as R, Stata, Tableau, Power BI, or SAS (OECD, 2016b; Richardson et al., 2019).

Tax authorities can use big data analytics to conduct tax audits in order to assess taxpayers' compliance. Some experts also point out that data analytics, whether descriptive, diagnostic, predictive, or prescriptive, can be used at all stages of audits (selecting the auditees, preparing audit programmes, and performing both tests of controls and substantive audit tests) (Alles & Gray, 2016; Intra-European Organisation of Tax Administrations [IOTA], 2016; Kundu & Kundu, 2016; Mehta et al., 2019). In the same vein, the OECD (2016a) shows that using advanced analytics with large data sets will improve accuracy when selecting taxpayers for audit.

In the first part of this paper, the authors have discussed the background, related works, and purpose of the study. In the second part, they will explain the research methodology. The third section describes the audit function in the Indonesian tax administration system in order to give context to the discussed case. The fourth section includes the analysis and discussion of the research findings. In the last part of the paper, the authors present their conclusions and discuss the contribution made by the study to the literature.

2. METHODOLOGY

This study utilises a qualitative research method. It includes interpretive case studies, and the authors used documentation and interviews to collect the data. It explains the meaning of social-organisational phenomena in the context of their environment and complements the researchers' emic (insider) perspectives (Bakker, 2010; Creswell, 2013; Hartley, 2004; Johannesson & Perjons, 2014; McKerchar, 2008; Yazan, 2015; Yin, 2018). The use of interpretive case studies allows the researchers to explore the various meanings of the investigated cases in order to acquire a deep understanding of them. The researchers can then use their experience and knowledge to interpret these cases. Oates (2006) clarifies that interpretivism in IT-related research is concerned with the social setting of innovation, how IT is built by individuals and, moreover, how IT impacts individuals. This social focal point recognises that all conclusions are epistemological by nature. This type of research will enable us to acquire knowledge that is structured in such a way and can, to some extent, be applied to different cases by considering any surrounding contexts (Howcroft & Trauth, 2004; Klein & Myers, 1999; Walsham, 2006).

This study uses data acquired from interviews and documentation. Informants were selected using the interview approach taken by Salijeni et al. (2019). The authors used the snowballing method based on the recommendation of a key person (key informant) who participated in specific focussed group discussions organised by the DJP's Tax Audit Directorate. The group discussed how electronic data affects tax audits. The author was invited to the discussion and obtained permission to interview the informants for the study. Using the snowballing approach, the authors identified informants who had adequate credibility and capacity in respect of the study's topic. The authors approached potential informants and asked them whether they were willing to be interviewed. Once they agreed, the interviews were conducted. Due to some limitations (primarily related to the informants' time availability), the authors were required to contact some informants later for additional information via online communication channels, such as email and the WhatsApp instant messaging application. Table 1 consists of a list of research informants. The interviews were conducted in compliance with data confidentially provisions as regulated by Article 34 of KUP law. On the other hand, the documentation studies were conducted using tax administration documents in a broad sense, including legal rules, standard operating procedures, annual reports, training materials, and system application manuals (see Bowen, 2009; Coffey, 2014; Olson, 2010).

Table 1: List of Interviewees as Research Informants

Code	Role
TA1	Senior tax auditor; e-auditor, i.e. a tax auditor who is assigned by the DJP (Decree Number KEP 20/PJ/2019) to gain understanding of the taxpayers' information systems, and to acquire and convert electronic data in order to carry out tax audits.
TA2	Former tax auditor and e-auditor. Now a lecturer at the Indonesian Ministry of Finance's tax academy.
TA3	Tax officer assigned as a digital forensics specialist within a preliminary tax crime investigation (*pemeriksaan bukti permulaan*) task force.
TA4	Section head in the area of tax audit policy.
TA5	Head of a tax office.

3. TAX AUDIT POLICY AND TAX DATA MANAGEMENT: THE CONTEXT OF THE CASE

The Indonesian tax administration embraces a self-assessment system. The system is applied to taxes administered by the central government, such as income tax, VAT, stamp duty, and land and buildings taxes for the forestry, fishery, and mining sectors. KUP law mandates that taxpayers can calculate the amount of tax that they owe. Simultaneously, the law also authorises the DJP to conduct tax audits in order to verify the accuracy of taxpayers' calculations and ensure tax compliance.

In this section, the authors will highlight the interplay between tax audit and tax data management policies in order to gain a comprehensive understanding of the context. Tax audit is the responsibility of the Minister of Finance, who delegates it to the DJP. The DJP is responsible for establishing tax audit procedures, as stipulated in Article 31 of KUP law. The delegation is stated in Regulation of the Minister of Finance Number 17/PMK.03/2013 and its amendments (hereinafter PMK-17), which concern audit procedures. PMK-17 includes:

(1) policies for determining the criteria used when selecting taxpayers for audit;
(2) tax audit standards, including general standards (relating to tax auditor qualifications), audit implementation standards (relating to tax audit plans, audit programmes, and audit supervision), and reporting standards (relating to the preparation of audit reports);
(3) the obligations, rights, and authorities of both taxpayers and tax auditors when tax audits are being conducted;
(4) various procedures relating to tax audits, including document borrowing and electronic data acquisition procedures.

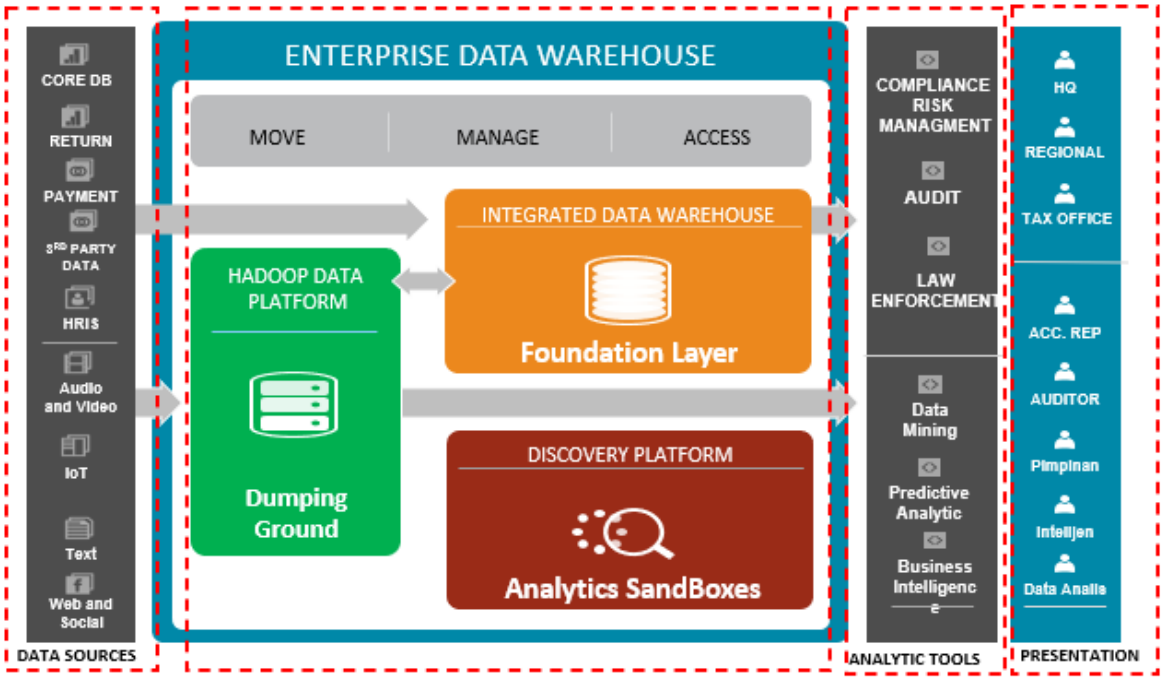
Subsequently, the DJP issued some circular letters (tax audit policies and technical guidelines) that act as implementation guides for PMK-17. For example, Circular Letter No. SE-15/PJ/2018 was issued in order to renew the existing tax audit policies. Generally, tax audit policy in respect of the selection of taxpayers for audit requires tax offices to establish a list of prioritised taxpayers. Each list is then reviewed and validated by the audit planning committees at regional tax office and DJP head office levels. The validation criteria are non-compliance indication and mode, tax potential, and tax debt collectability level. Meanwhile, the DJP published Circular Letter No. SE-24/PJ/2019, which concerned the implementation of compliance risk management (CRM) as part of the wider tax administration system. The output generated from CRM can be used for tax audit policy purposes.

Tax data management is a crucial part of the data collection process, helping to ensure that the selection of taxpayers for audit is implemented according to PMK-17 and Circular Letter No. SE-15/PJ/2018. Figure 1, as modified from a presentation by Djuniardi (2018), illustrates how tax data management supports the implementation of tax audit policy. The figure depicts the data flow and the functions that use the flow.

As illustrated in Figure 1, the tax data management system in Indonesia has several layers. The first is the data sources layer, which contains data from all sources. This layer is often called a data lake (see, for example, Devlin, 2018). Most of the data in the layer is acquired from taxpayers in the form of tax administration reports (such as taxpayer registration forms, tax returns, tax payment receipts), complaints (such as tax objection letters), and other tax administration service applications (such as tax clearance certificate requests). In addition, KUP law also requires other institutions, such as land administration offices, and banks and

other financial institutions, to send data to the DJP on a regular basis. Meanwhile, the DJP is also developing its own database by collecting data from other sources, such as social media platforms and tax intelligence practices.

Figure 1: Tax Data Management



Source: Adapted from Djuniardi, 2018. Used with permission.

The second layer is the enterprise data warehouse layer, where the results from the data ingestion and transformation processes are managed. This layer also applies data governance to ensure that access to the data is granted according to the information security policies. The third is the analytic tools layer, which consists of analytical tools used for tax administration functions. This layer uses some (big) data features in combination with data that has been processed through the enterprise data warehouse layer in order to conduct taxpayer compliance risk analysis. This analysis creates a list of taxpayers to be audited and provides details of the rationale for their selection. There can also be interplay between compliance risk analysis and tax audit policy. On one side, tax audit policy determines how compliance risk analysis algorithms work. On the other side, tax audit policy uses compliance risk analysis's output for policy formulation purposes.

The fourth presentation layer is the data visualisation layer. This consists of information dashboards, in aggregate and in detail, for each tax administration function, so that the information can be directly executed (for example, in order to issue tax audit assignments or form data marts (analytical cube or data models). Alternatively, the visualisation layer can be reprocessed to suit the needs of end users working in tax audit management.

The DJP has also published several circular letters containing technical guidance for tax audit fieldwork, as exhibited in Table 2. These circular letters show that electronic data has become part of the tax audit ecosystem, so tax auditors must also be prepared to handle big data. Consequently, tax auditors are required to have specific skills in order to prepare, process, and analyse big data as part of audit assignments.

Table 2: List of DJP Circular Letters Related to Tax Audit Policy Implementation

Number	Concerning
SE-10/PJ/2020	The use of the audit desktop application for detailed arrangements of the tax audit's implementation stages. The audit desktop application is the audit application used by the DJP's tax auditors for audit assignments, including preparation, implementation, and reporting activities. The implementation of the audit desktop application can, indirectly, be considered to be tax audit business process engineering because the application, in some ways, integrates tax audit workflows.
SE-10/PJ/2017	Tax compliance audit guidelines which regulate: (1) how to understand taxpayers' information systems; (2) how to get authorisation from taxpayers in order to access their electronic data processing equipment; and (3) how to obtain image files[5] and hash values[6] of audit evidence in the form of electronic data acquired from taxpayers.
SE-25/PJ/2013	e-audit guidelines which regulate: (1) the assignment of an e-auditor as part of a tax audit engagement; (2) the e-auditor's responsibility to process any electronic data acquired from taxpayers or provided by the other tax audit team members, and to provide such processed data in the format requested by the tax auditors.

4. CASE ANALYSIS AND DISCUSSION

This section will discuss the research findings obtained from the interviews. Five cases will be detailed using the presentation and analysis techniques described by Yin (2018). In this context, "case" refers to things scrutinised, such as organisations, departments, information systems, discussion forums, system developers, development projects, decisions, and so on (Oates, 2006). Cases will be studied thoroughly using various data collection methods (interviews, observation, document analysis, and/or questionnaires). The goal is to obtain rich and detailed insights into how each case became part of human activity.

Furthermore, the authors use an analysis technique that Diop and Liu (2020) describe as a "single setting case with multiple sub-cases" (p. 10). The single setting case is the implication for big data in the tax audit, while the multiple sub-cases are activities in the tax audit that are affected by the existence of big data. These sub-cases explore the use of big data, both at tax audit policy level and individual audit assignment level, by discussing:
(1) risk analysis and its relationship with taxpayer selection criteria for tax audit;
(2) how auditors extract audited taxpayers' data sets when these are large and in multiple formats;

[5] An image file is an exact duplicate, or bit-by-bit copy, of electronic data that contains all artefacts, i.e. information or data created as a result of the use of electronic devices that show past activity, such as time of access, deleted data, data fragments, hidden files, and unused or unallocated space (Goldstein, 2019; McKemmish, 2008; Montasari, 2017).

[6] A hash value is a value in the form of a combination of numbers, letters, or other characters with a fixed number of symbols generated by specific logical sequences and calculations (algorithms) for a set of electronic data. Hash values are used to maintain the integrity of the electronic data (Pethe & Pande, 2016). The DJP's Circular Letter No. SE-10/PJ/2017 states that, in order for it to be used for audit evidence, electronic data acquired from taxpayers needs to be in image file format and a hash value must be generated for it to maintain its integrity.

(3) audit tests and data analysis workflows;
(4) use of spatial data;
(5) web and social media data extraction.

The following subsections (4.1 to 4.5) describe the situations and the authors' interpretations of the problems. In subsection 4.6, the authors use an interpretive point of view and, based on the various facts that have been described above, attempt to construct guidelines that can be used to gain understanding of how the Indonesian tax authorities deal with big data for tax audit purposes.

4.1. Case 1: Risk Analysis in Audit Selection

Tax audit policy emphasises the importance of selecting taxpayers for audit using all data from the DJP's information system and facts obtained from observation and supervision activities. The utilisation of internal and external data when selecting taxpayers for audit generates a list of taxpayers. These can then be plotted on a two-dimensional graph with level of compliance and collectability on the axes. Therefore, data analysis should be performed to ensure that selected taxpayers meet the selection criteria.

In this context, the risk-based audit selection system can use an enterprise data warehouse that has been developed by the data and information management function at the DJP. As previously described, the enterprise data warehouse aggregates and transforms tax compliance data, such as information obtained from tax returns and VAT invoices, and comparative data acquired from third parties. The DJP already configures big data as part of its data management processes for its administrative functions. Next, as mandated by Circular Letter No. SE-24/PJ/2019 (which concerns compliance risk management), the system will generate a list of taxpayers to be audited based on a risk analysis conducted using various types of data stored within the enterprise data warehouse.

Djuniardi (2018) reported that the use of big data for compliance risk monitoring is still ongoing, with some use cases[7] being seen as "quick wins" for tax audit purposes and other law enforcement actions, including:

(1) data matching analysis (data equalisation) between third party data and data from tax returns and tax payment data, including data relating to income tax and VAT;
(2) network analytics to assess ownership and distribution relationships, so as to identify any related party or insider transactions;
(3) deep analytics for data matching and pattern identification in order to detect VAT fraud. One method that fraudulent taxpayers often use when attempting to reduce payable VAT is to credit VAT input using void VAT invoices.

The DJP's ability to use the three processes mentioned above results from its hard work over a long period to ensure that all taxation data is submitted in electronic format. Its efforts in this area have been recorded since 2000 (see, for example, Darono & Irawati, 2015). Tax return (Surat Pemberitahuan, or SPT) data has begun to include e-payment, e-tax return, e-filing, and e-invoice initiatives, which are being implemented gradually as the process takes a long time.

[7] A use case is series of intra-organisational or inter-organisational activities that are presented in a diagram (known as a use case diagram) in order to explain the relationships between business processes, procedures, application systems, and users that, if executed, will produce specific outputs (see, for example, Booch et al., 1998).

On the other hand, additional data from third parties (for example, banks, stock exchanges, and other financial institutions) became available as a result of the enactment of law number 9/2017, which concerns access to financial information for tax purposes. This law was introduced in order to fulfil Indonesia's obligation to carry out automatic exchange of financial information with fellow G20 members. This combination of factors provided the DJP with a large amount of data that covered various types of transactions, allowing it to carry out more in-depth tax audit risk analysis. The DJP immediately took steps to boost many of its organisational components, including its technology and human resource capabilities, in order to utilise big data for more in-depth risk analysis so as to determine tax audit targets more precisely.

4.2. Case 2: Extracting Large Data Sets Containing Multiple Data Formats

Interviews with TA-1 and TA-2 revealed that they had acquired all accounting data entries in the form of report files, spool files, or PRN files. These files are the electronic form of printed financial statements, so the structure, content, and file layouts are the same as the printed versions. The essential documents are general and supporting journals, general and subsidiary ledgers, and other relevant information that can be used to check taxpayers' tax returns. Data extraction and analysis techniques are used to check these returns (Cascarino, 2017; Hunton et al., 2004; ISACA, 2011; Zuca & Tinta, 2018).

TA-1, a senior tax auditor and e-auditor team member with more than twenty years of experience, stated that his most challenging audit assignment was when he received general ledger data consisting of more than ten million records. Meanwhile, TA-2, who worked as a tax auditor for more than ten years, related his experiences of dealing with several sources containing huge volumes of data. He was also required to handle various types of data source, including database management systems (known as DBMS), spreadsheets, and HyperText Markup Language (known as HTML) files. In order to deal with many types of data formats, auditors need to use a combination of data processing applications (for example, EmEditor, Power BI, and Python).

From another perspective, TA-3 noted that he faced a further challenge: the need to consolidate reports generated by many branches because the audited taxpayer did not report the details of the consolidation procedure used. There was only a summary of the transactions report available in the taxpayer's head office. Thus, TA-3 had to carry out a detailed data consolidation procedure. Approximately 110 million records from the company's branches needed to be analysed and tested. In another assignment, TA-3 stated that processing and analysing large report files also presented its own challenges if the data layouts were inconsistent. Moreover, when working with a report file that had a complicated layout pattern, the tax auditor is required to spend more time creating data extraction orders, so the execution of the data extraction query itself is slower. TA-3 mentioned that he obtained general ledger data with 16 million records in a report file format with a complicated layout that needed to be converted into a tabular or structured data format.

These audit assignments reveal several of the challenges faced by tax auditors:

(1) auditors need to use audit software that requires more powerful hardware support in order to perform various audit tests and data analysis tasks within acceptable timeframes;

(2) some data types and formats cannot be recognised and processed by specific audit tools, so tax auditors need to determine the audit tools that they can use while considering any other resources that they have.

In order to anticipate possible device constraints (hardware and software) and tax auditors' ability to acquire and process data, the DJP issued Circular Letter No. SE-25/PJ/2013, which stipulates that a tax auditor may request assistance from an e-auditor so that they can acquire and analyse electronic data as evidence for the tax audit that they are conducting.

TA-1 needed to use the Audit Command Language (ACL) audit software package instead of a standard spreadsheet application because the financial data acquired from taxpayers was relatively large. TA-2 was unsuccessful in performing data extraction using a Power BI Desktop software package with 8GB of memory. The auditor finally managed to perform data extraction procedures using a more powerful computer with 32GB of memory. Meanwhile, TA-3 had to use the pandas software library and Python on a computer with 8GB of RAM in order to extract the data. Previously, TA-3 had used the same computer with the Power BI Desktop application installed, yet had not managed to complete the extraction process.

The audit assignments described above illustrate how tax auditors face challenges relating to big data. Most of the time, data obtained from the taxpayer application system is in a ubiquitous format (such as report files). However, it is still important that the auditors understand the data layout. Moreover, if the data comes in various formats, the auditors should equip themselves with various tools suitable for use with the data.

The DJP issued Decree KEP-251/PJ/2020, concerning the establishment of the taxpayers' data integration team, in response to the evolution of taxpayer data management. It regulates cooperative compliance, which has been incorporated in the Indonesian tax system, in a broader scope. This provision also regulates tax audit functions relevant to cooperative tax compliance, such as general ledger tax mapping. However, it does not describe the technical procedures and data formats used to implement general ledger tax mapping in detail. In comparison, some authorities use the Standard Audit File for Taxation (SAF-T) approach, as proposed by the OECD (2004), or eXtensible Business Report Language (XBRL) (see, for example, Mousa, 2011). Understanding how the tax authorities set the standards and data formats used for audit purposes in this era of big data will make it easier for tax auditors to map the tools and techniques needed in order to acquire and process electronic data obtained from taxpayers.

Big data's presence in tax audits is a natural symptom of the entire data management constellation for business purposes. Computer hardware, mobile devices, various types of IoT sensors, data communication devices, and computer networks—especially those developed by emerging economies—are becoming increasingly affordable. The dominance of open source software (including operating systems, programming languages, and database management systems) within all communities makes data flow even faster. As a consequence, computing applications and data processing are ubiquitous and, ultimately, generate big data. The result of this situation for tax administration functions, including tax audit, is that it is necessary to acquire and process large amounts of data incorporating various data types. A further implication, especially for tax audits, is that while, in the 1990s, only big business entities had the capacity to implement end-to-end IT systems, today even a start-up company can choose to use the most sophisticated information management system—one that ultimately stores its data in a big data configuration. Every tax auditor must be alert to situations like this.

4.3. Case 3: Audit Test and Data Analysis Workflow

From the audit techniques perspective, data extraction and analysis (DEA) is the most suitable technique to use in order to test electronic data in the Indonesian tax administration system (Darono & Ardianto, 2016). DEA is a workflow that starts with data extraction and continues with data analysis. It is performed using various audit procedures designed in the audit plan and programme. Data extraction and analysis is a reliable tax audit technique that can, when used with the right audit tool to manage the size and variance of the data, be utilised to handle big data.

The DEA workflow can be summarised as follows:

(1) understanding the taxpayer system configuration and data processing applications;
(2) acquiring and extracting data using appropriate audit tools;
(3) performing data analysis using various audit tests, as stipulated in the audit plan.

Tax audit standards, which mainly relate to the audit test, suggest that tax auditors perform duplicate and gap detection, data relation, and data range validation tests. Furthermore, specific audit tests can be performed to test audited taxpayers' formal compliance, such as examining the timeliness and amounts of their withholding income tax payments, ascertaining whether they have met their VAT obligations (including crediting VAT input and under/overpayments), and confirming that their financial statements for commercial purposes and for tax purposes can be reconciled.

One of the objectives of a tax audit is to assess whether each transaction complies with the applicable tax regulations. The process by which tax auditors extract data from taxpayers' financial statements so that they can be analysed is as follows. The first step is to identify the data format. If it is structured data, it can be processed directly. Second, the data is summarised in a trial balance format according to the available chart of accounts. The auditor should be able to confirm all balances in the financial statement. Third, the auditor performs the validity test (such as a gap detection, a duplicate detection, or a validation of data range test). The purpose of this step is to ensure that the auditor is sufficiently confident that the data is valid.

Next, the auditor will run the tax compliance test for every transaction. One procedure used by auditors is the keyword search. This extracts every transaction record that includes specific keywords. Subsequently, those transactions are checked to ensure that they meet the applicable tax provisions. For instance, a tax auditor might search for all records that contain the words "rent" or "lease". Every record that matches the criteria will be examined, whether the applicable VAT and income taxes have been paid or not.

Other relevant keywords can be added in order to generate more results. The issue that arises in respect of big data is how to extract unstructured data and transform it into structured data in order to facilitate data matching. Tax auditors must also deal with data veracity issues. These require auditors to prepare audit tests in order to ensure data validity.

Several authors specifically address how big data analytics is related to audit programmes and tests. First, reconciliation between financial statements for commercial purposes and for taxation purposes can be carried out using diagnostics analytics techniques (see, for example, Richardson et al., 2019). This type of analysis will reduce the amount of time that it takes to complete a tax audit assignment. Our own observations show that it still takes a long time for

the tax auditor to produce a confirmed trial balance in order to reconcile the amounts recorded in each account detailed in the taxpayer's financial statements and confirm that these conform with the relevant tax provisions. Second, data veracity issues can be overcome by implementing machine learning. Of course, this is not a task for the individual tax auditor. It is the DJP's obligation to provide auditors with a tool that can ensure the veracity of the data obtained when producing audit evidence. Both practices will be easier to carry out following the implementation of big data analytics for tax audit purposes.

4.4. Case 4: Spatial Data

TA-5 discussed his experiences of the relationship between big data and the use of spatial data. Regardless of its size, spatial data can be generated by drones that conduct taxation data searches. The data can be used in tax administration functions, including tax audits. John Villasenor (2012, as cited in Fox, 2017, p. 82), states that a drone is "an unmanned aircraft that can fly autonomously". TA-5 added that it is legal to use drones to collect data for taxation purposes, as regulated in Article 35A of KUP law. According to a report by the Indonesian Ministry of Finance (2015), drones are seen as an innovation that can assist with the management of state finances. Spatial data can be collected using aerial photography techniques. The use of drones can be advantageous for spatial data collection because they can create maps effectively and efficiently, and at a relatively low cost. In addition, the data obtained can be used for comparison because it is relatively accurate and consists of detailed pictures of an object's current condition. It is also helpful when establishing benchmarks for data from the previous fiscal year. Moreover, the data obtained by drones can be enriched using Google Maps data in order to highlight specific issues or objects. Spatial data can be a significant factor when auditing certain taxpayers, such as oil palm plantation or chicken farm owners.

Spatial data, in principle, is commonly used to measure plantation or farm areas. It is combined with plant-level investment data released by official authorities in the relevant fields and compared with the information provided in taxpayers' tax returns. The critical reason for using spatial data is its veracity. Tax auditors need to prove that the data is valid in order to use it as an anchor or reference. Furthermore, tax auditors are required to have the capacity to convert and combine the spatial data with Google Maps data, so that it can be matched with data obtained from other sources.

In the authors' view, the use of spatial data for tax audit purposes will, in certain situations, be more complicated but it may also be more supportive for taxpayers whose businesses involve spatial areas, such as those in the agriculture, mining, forestry, or real estate sectors. Spatial data analysis still requires the use of some remote sensing analysis techniques because of its ambiguity (see, for example, Jain, 2008), making its utilisation as audit evidence a winding road. Spatial data analysis and interpretation techniques also require expertise that some tax auditors may not have. In this case, the proposed solution is to assign a tax appraiser who has knowledge of geomatics to the audit team.

4.5. Case 5: Web-Based and Social Media Data Extraction

The Internet has changed rapidly since it was first founded, with both private sector (e-commerce) websites and government (e-government) websites moving from being online brochures to becoming vehicles for interactive transactions (Drula, 2014; The World Bank, 2002). Recently, there have been initiatives to make the Internet a social interaction, social media, or new media platform (Edosomwan et al., 2011; Menke & Schwarzenegger, 2019). This suggests that the Internet itself is big data, as it shares the characteristics known as the 4Vs (described at the beginning of the paper). One thing that has exponentially driven changes in the way that the Internet is used is the convergence between two technologies that initially developed independently: the Internet and the mobile phone. This has allowed new businesses, such as e-commerce and social media platforms, to multiply. Transactions can be completed quickly using e-commerce (or e-marketplace) platforms. Some of these platforms are standalone businesses while others are integrated with other social media applications. The coffee maker purchase example detailed above illustrates the convergence of these technologies and the resulting data integration.

These changes in Internet use functions and patterns have resulted in online interactions taking place between tax authorities and taxpayers. Taxpayers can use the Internet in order to conduct business transactions and social interactions. The tax authorities can take advantage of various forms of information that appear on the Internet, including e-commerce and social media data, in order to monitor taxpayer compliance. The critical challenge for tax authorities is to level the playing field between online commerce (via e-marketplaces or social media platforms) and conventional commerce. If the tax treatments for these environments are uneven, it will cause inequality. If they are to take advantage of the boom in e-commerce, tax authorities should develop a data search and collection method that can be performed online. Technically, one method that can be used to search and collect data from online environments is web data extraction or web scraping (Chaudhari & Paikrao, 2012; Krotov & Silva, 2018). Web data extraction for taxation purposes in Indonesia is a legitimate technique that is regulated in Article 35A of KUP law.

TA-5, for example, was able to estimate a YouTube content creator's income utilising the web data extraction technique. The auditor used the extracted or scraped data for the risk analysis process before performing audit tests. Next, he familiarised himself with the monetisation process so that he was able to estimate the content creator's income by multiplying the number of content viewers and subscribers by the Google AdSense tariff. Finally, the auditor compared the estimated income with the income reported in the tax return to see if the amounts were reasonably close.

Web-based data and, in particular, social media data must have very high veracity levels in order to be of use to the tax authority. For example, one store in an online marketplace says that it sells fashion products, and shows the price of the items and the number that have been sold. Can the information immediately be stated as the sales value of the store for VAT or income tax purposes? It is, of course, not that easy to confirm the store's income solely based on that data. It still takes some work to establish the veracity of the data obtained from the Internet, including that acquired from social media platforms. Web and social media data is currently still used as supporting information (see, for example, OECD, 2017) for tax audit purposes such as profiling or gaining an understanding of a taxpayer's business network (i.e. ownership, consumers, and suppliers).

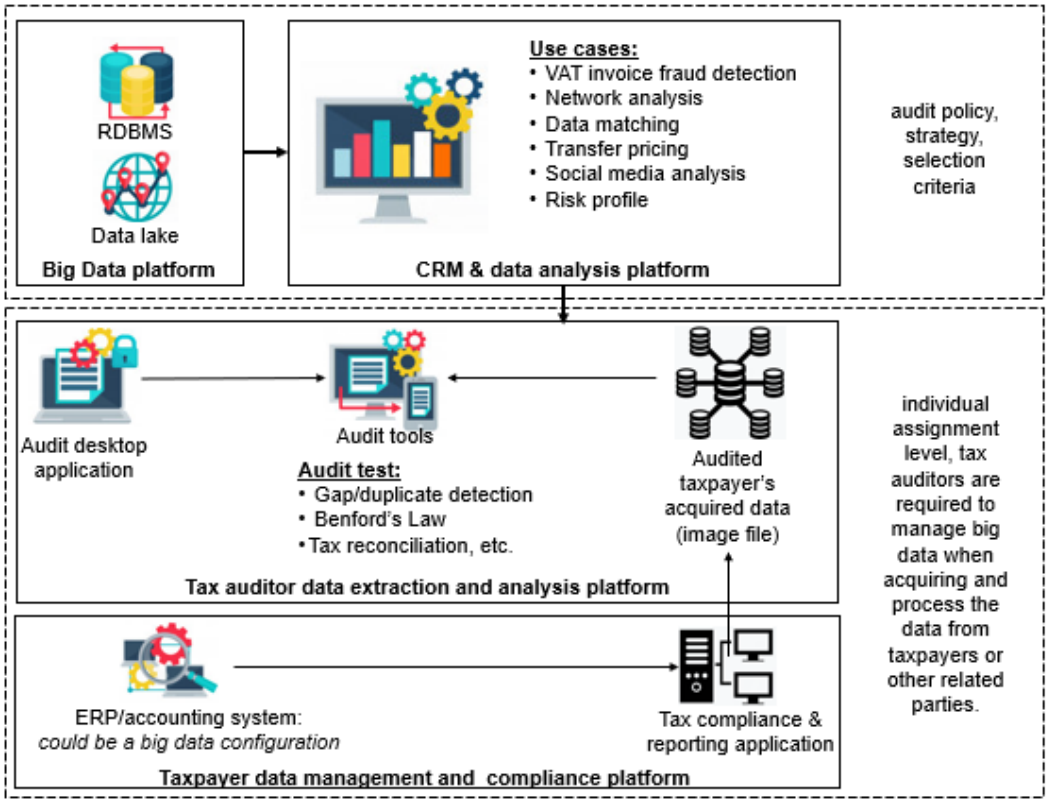
At present, when conducting tax audits, tax authorities tend to use big data analytics in order to pursue quick wins. In the authors' view, big data analytics for tax audits should be developed continuously in order to implement machine learning more comprehensively.

4.6. Multiple Sub-Case Analysis: Some Interpretive Insights

These cases have some similarities, which can be analysed further in order to formulate a framework that will help us to understand the impact of big data on tax audits. This section assembles some constructs from each sub-case so that a more systematic framework can be developed.

The authors observed that the influence of big data on tax audits—as one of the Indonesian taxation system's functions, as illustrated in Figure 2—exists in two aspects.

Figure 2: Influence of Big Data Deployment in Indonesian Tax Audits



Source: Designed by Freepik. This table was created by the authors using royalty-free images from www.Freepik.com

First, at the audit policy level, big data integrates structured and unstructured data in a data lake. The integrated data is obtained either from internal or external DJP data sources. Data is analysed using various analytical techniques, and fed to CRM and tax audit management systems for tax audit policy and audit selection purposes. Moreover, at this level, big data is considered to be a large volume of data that multiplies rapidly, consists of various data formats, and has a high level of uncertainty. It can be processed in a way that supports the DJP's CRM risk engine as part of the whole compliance risk management procedure. The goal is to find and manage taxpayers with high compliance risk profiles so that tax auditors can focus their

resources in order to supervise and audit these taxpayers. This should, ultimately, enable them to establish sustainable tax compliance. The use of big data for tax audit purposes will, of course, cause issues related to data security and privacy to arise. As a result, the DJP issued a series of provisions relating to procedures for maintaining data integrity, security, and availability, as detailed in DJP Regulation PER-41/PJ/2010, which concerns information security management policy, and DJP Circular Letter No. SE-30/PJ/2019, which concerns tax data access authority policy.

Second, at the individual tax audit engagement (or assignment) level, the assigned tax auditor can access various data relating to the audited taxpayer through the audit desktop application. The data includes taxpayers' full profiles, including tax returns, tax payments, VAT invoices, and other withholding tax slips. Tax auditors are not allowed to directly access the data lake that contains unstructured data relating to the taxpayer. They can only use data that has been processed and presented in the audit desktop application. Subsequently, the auditors design audit plans and programmes using the data provided. They extract the data using data extraction and analysis audit techniques, and perform audit tests. At audit test level, empirical evidence shows that:

(1) Tax auditors may need to handle large volumes of data (hundreds of million records) from previous fiscal years (historical records).
(2) They may also need to deal with taxpayer data format and source variations. This issue may, to some extent, be resolved by giving an auditor the authority to request data in a format that is compatible with their analytical tools. However, there is still a possibility that the data will only be available in its original format.
(3) The need to establish data veracity requires the auditors to review data sources. The validity of a data source will determine how the data obtained from it should be processed and analysed.

Based on the interviews with TA-2 and TA-3, it seems that tax auditors may be exposed to big data during the data extraction process. Alternatively, tax auditors can access the data lake directly. Therefore, tax auditors must have the competencies and skills needed to acquire data from taxpayers whose systems incorporate the use of big data. Meanwhile, from a data governance perspective, it is necessary to consider giving tax auditors access to data from established data lakes. In order to promote the idea of continuous improvement for tax auditors, the DJP, in collaboration with the Ministry of Finance's Tax Education and Training Centre, has organised a series of capacity-building activities in the form of certification training concerning business models related to the digital economy, tax data analytics, web data extraction and analysis, and digital forensics for tax law enforcement. However, in the authors' view, these measures are not sufficient to deal with current issues related to advanced big data analytics, such as cloud computing, containerisation, blockchain-based data, or deep learning for tax risk analysis. The DJP and the Tax Education and Training Centre should develop additional training programmes that address such issues.

This section ends with the authors' interpretation of how tax data should be understood, in the context of, and using data related to, tax audits as part of tax administration in Indonesia. With all its technical aspects, big data is a technological phenomenon and has become an organisational artefact. Therefore, it is critical that big data analytics becomes part of the tax administration's functions. Based on the description above, the authors argue that the main outcome expected from the implementation of big data analytics in tax audits is that it will provide tax auditors with information that can be used as evidence that taxpayers are either complying or failing to comply with tax laws.

5. CONCLUDING REMARKS

It can be concluded that the use of data, both in businesses and government institutions, is now always related to big data. Many vendors and technology developers are struggling to converge and standardise the available technological frameworks and applications. One day, big data platforms will be much easier to use for many purposes, including tax administration. This study has shown that big data is an integral part of the tax administration business process. It plays a critical role in all tax administration functions, including risk analysis, taxpayer supervision, and law enforcement. Publications produced by several tax consultancy firms have also noted that big data is now part of the tax planning and management procedures developed by taxpayers (Deloitte, 2016; EY Americas, 2019; PwC, 2015).

This study set out to determine how big data influences tax audits in the context of the current Indonesian tax administration practices. The findings suggest that there are two indicators of its influence:

(1) at the audit policy level, big data is used for risk analysis in order to identify taxpayers with high compliance risks who should be audited; thus, it can help tax administrations to achieve sustainable tax compliance;
(2) at the individual audit assignment level, tax auditors are required to design audit programmes that demonstrate how they acquire, process, and analyse audit evidence in big data formats.

This paper proposes two recommendations relating to the impact of big data on tax audit practices in Indonesia. First, it is crucial to improve tax auditors' capacity to acquire, process, and analyse big data. Second, it is necessary to have a data governance policy that allows tax auditors to use a data lake in order to obtain data that complements the structured data that is readily available through standard applications.

In terms of contributing to the broader tax administration literature, the results of this study show how tax authorities incorporate big data in tax administration tasks. Subsequent research could explore this further, investigating, for example, the characteristics of taxpayers' business processes that drive them to utilise big data as part of their data management procedures, how big data affects corporate tax planning and management processes (including data exchange between taxpayers and their tax advisers), or the need for tax authorities to adjust their provisions for data exchange in terms of cooperative tax compliance schemes, tax audit technical guidelines, and improved standard audit file structures and formats.

BIBLIOGRAPHY

Alles, M., & Gray. G. L. (2016). Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors. *International Journal of Accounting Information Systems, 22*, 44-59. https://doi.org/10.1016/j.accinf.2016.07.004

Arens, A. A., Elder, R. J., Beasley, M. S., & Hogan, C. E. (2017). *Auditing and assurance services: An integrated approach* (16th ed.). Harlow, England: Pearson Education Limited.

Bakker, J. I. (H.). (2010). Interpretivism. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.) *Encyclopedia of Case Study Research* (Volume 1, pp. 486-492). Thousand Oaks, CA: SAGE Publications, Inc.

Booch, G., Rumbaugh, J., & Jacobson, I. (1998). *Unified modeling language user guide*. Reading, MA: Addison-Wesley.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40. https://doi.org/10.3316/QRJ0902027

Brink, W., & Hansen, V. (2018, 1 May). Using big data to identify tax risk. *The Tax Adviser*. https://www.thetaxadviser.com/issues/2018/may/use-big-data-to-identify-tax-risk.html

Cascarino, R. E. (2017). *Data analytics for internal auditors*. Boca Raton, FL: CRC Press.

Chang, W. L., Grady, N., & the National Institute of Standards and Technology Big Data Public Working Group (2018). *NIST big data interoperability framework: Volume 1, Big data definitions* (Special Publication (NIST SP) - 1500-1 Version 2). https://doi.org/10.6028/NIST.SP.1500-1r1

Chaudhari, P. A., & Paikrao. R. L. (2012). Web data extraction. *IJCA Proceedings on Emerging Trends in Computer Science and Information Technology-2012(ETCSIT2012) ETCSIT (4)*, 13-17.

Chen, S.-C., Wu, C.-C., & Miau, S. (2015). Constructing an integrated e-invoice system: The Taiwan experience. *Transforming Government: People, Process and Policy*, *9*(3), 370-383. https://doi.org/10.1108/TG-09-2014-0043

Cockfield, A. J. (2016). Big data and tax haven secrecy. *Florida Tax Review*, *18*(8), 483-539.

Coffey, A. (2014). Analysing documents. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 367-379). London, England: Sage Publications Ltd.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). London, England: SAGE Publications Ltd.

Darano, A., & Ardianto, D. (2016). The use of CAATTs in tax audits—lessons from some international practices. *eJournal of Tax Research*, *14*(2), 506-526.

Darano, A., & Irawati, D. (2015). Service innovation in the complex environment of tax administration: The Indonesian public service perspective. *International Journal of Innovation and Regional Development*, *6*(1), 102-123.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business Review Press.

Deloitte. (2016). *Tax data analytics: A new era for tax planning and compliance*. n.p.: Deloitte.

Devlin, B. (2018). *The EDW lives on: The beating heart of the data lake* [White paper]. 9sight Consulting. http://www.asprom.com/technologie/9sight.pdf

Diebold, F. X. (2012). *On the origin(s) and development of the term "big data*" (PIER Working Paper 12-037). Philadelphia, PA: Penn Institute for Economic Research, University of Pennsylvania.

Dimitropoulou, C., Govind, S., & Turcan, L. (2018). Applying modern, disruptive technologies to improve the effectiveness of tax treaty dispute resolution: Part 1. *Intertax*, *46*(11), 856-970.

Diop, K. A. S., & Liu, E. (2020). Categorization of case in case study research method: New approach. *Knowledge and Performance Management*, *4*(1), 1-14. https://doi.org/10.21511/kpm.04(1).2020.01

Direktorat Jenderal Pajak. (2015). *Peraturan Direktur Jenderal Pajak Nomor PER-46/PJ/2015 tentang Cetak Biru Teknologi Informasi Dan Komunikasi Direktorat Jenderal Pajak Kementerian Keuangan Republik Indonesia Tahun 2015-2019*. [Regulation of the Director General of Taxes, No. 46/PJ/2015: Information and Communications Blueprint of the Director General of Taxes 2015-2019]. https://datacenter.ortax.org/ortax/aturan/show/16755

Direktorat Jenderal Pajak. (2017). *Annual report 2017*. Jakarta, Indonesia: Direktorat Jenderal Pajak.

Direktorat Jenderal Pajak. (2022). *CRM BI: Langkah awal menuju data driven organization*. [CRM: BI: First steps towards a data driven organization]. Jakarta, Indonesia: Direktorat Jenderal Pajak.

Djuniardi, I. (2016, September). *Next generation data analysis: The implementation of big data in Directorate General of Taxes Republic of Indonesia* [Conference presentation]. 13th Association of Tax Authorities of Islamic Countries Annual Technical Conference, Melaka, Malaysia.

Djuniardi, I. (2018). *Journey to big data*. Jakarta, India: Direktorat Jenderal Pajak.

Drula, G. (2014). Media convergence and mobile technology. *Journal of Media Research*, *7*(3), 47-71.

Edery, C. (2016). Big data serving tax compliance. In Intra-European Organisation of Tax Administrations, *Data-driven tax administration* (pp. 48-50). Budapest, Hungary: IOTA.

Edosomwan, S., Prakasan, S. K., Kouamé, D., Watson, J., & Seymour, T. (2011). The history of social media and its impact on business. *Journal of Applied Management and Entrepreneurship*, *16*(3), 79-91.

EY Americas. (2019, June 27). How tax and finance departments can deliver value in the digital era. *EY*. https://www.ey.com/en_us/digital/how-can-tax-and-finance-departments-deliver-value-in-the-digital

Fox, S. J. (2017). The rise of the drones: Framework and governance—Why risk it! *Journal of Air Law and Commerce*, *82*(4), 683-715.

Gaillard, M. (2017, July 6). CERN data centre passes the 200-petabyte milestone. *CERN*. https://home.cern/news/news/computing/cern-data-centre-passes-200-petabyte-milestone

Gantz, J., & Reinsel, D. (2011). *Extracting value from chaos*. Framingham, MA: IDC.

Goldstein, S. (2019, 24 July). *Two key differences between digital forensic imaging and digital forensic clone and how they can affect your legal case*. Capsicum Group. https://capsicumgroup.com/2-key-differences-between-digital-forensic-imaging-and-digital-forensic-clone-and-how-they-can-affect-your-legal-case/

G20 Sherpa Indonesia. (2019). *History of the G20*. G20 Indonesia Secretariat, Coordinating Ministry of Economic Affairs, Government of Indonesia. https://sherpag20indonesia.ekon.go.id/public/en/history-of-the-g20

Hartley, J. (2004). Case study research. In C. Cassell & G. Symon (Eds.), *Essential guide to qualitative methods in organizational research* (pp. 323-33). London, England: SAGE Publications Ltd. https://www.doi.org/10.4135/9781446280119.n26

Helskyaho, H. (2017). *Big data and the multi-model database* [Presentation]. The Nordic ACE Tour 2017.

Houser, K, A., & Sanders, D. (2017). The use of big data analytics by the IRS: Efficient solutions or the end of privacy as we know it? *Vanderbilt Journal of Entertainment & Technology Law*, *19*(4), 817-872.

Howcroft, D., & Trauth, E. M. (2004). The choice of critical information systems research. In B. Kaplan, D. P. Truex III, D, Wastell, A. T. Wood-Harper, & J. I. DeGross (Eds.), *Information systems research: Relevant theory and informed practice* (IFIP International Federation for Information Processing, Vol. 143, pp. 195-211). Boston, MA: Springer.

Hu, H., Wen, Y., Chua, T.-S., & Li, X. (2014). Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, *2*, 652-87. https://www.doi.org/10.1109/ACCESS.2014.2332453

Hunton, J. E., Bryant, S. M., & Bragranoff, N. A. (2004). *Core concepts of information technology auditing*. Hoboken, NJ: John Wiley & Sons, Inc.

InfoDev & the Center for Democracy & Technology. (2002). *The e-government handbook for developing countries*. Washington, D.C.: The Center for Democracy and Technology.

Intra-European Organisation of Tax Administrations. (2016). *IOTA good practice guide: Applying data and analytics in tax administrations*. Budapest, Hungary: IOTA.

ISACA. (2011). *Data analytics—A practical approach*. Schaumburg, IL: ISACA.

Jain, S. (2008). Remote sensing application for property tax evaluation. *International Journal of Applied Earth Observation and Geoinformation*, *10*(1), 109-121. https://doi.org/10.1016/j.jag.2007.10.008

The Jakarta Post Editorial Board. (2019, June 27). We are G20. *The Jakarta Post*. https://www.thejakartapost.com/academia/2019/06/27/we-are-g20.html#:~:text=Indonesia%20is%20in%20the%20Group,that%20the%20country%20stands%20out

Johannesson, P., & Perjons, E. (2014). *An introduction to design science.* Cham, Switzerland: Springer International Publishing.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67-93. https://doi.org/10.2307/249410

Krotov, V., & Silva, L. (2018). Legality and ethics of web scraping. *Twenty-fourth Americas conference on information systems, New Orleans, 2018*, 1-5.

Kundu, A., & Kundu, S. G. (2016). Big data analytics & its applications in the tax domain. *BIMS International Journal of Social Science Research*, *1*(2), 117-133.

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety* (Application Delivery Strategies, File No. 949). Stamford, CT: META Group, Inc.

Laney, D. (2012). *Deja vvvu: Others claiming Gartner's construct for big data* [Blog post]. https://ctovision.com/deja-vvvu-others-claiming-gartners-construct-for-big-data/

Lewis, C. (2019). *Raising more public revenue in Indonesia in a growth- and equity-friendly way* (OECD Economics Department Working Papers No. 1534). Paris, France: Organisation for Economic Co-operation and Development.

Lu, J., & Holubová, I. (2019). Multi-model databases: A new journey to handle the variety of data. *ACM Computing Surveys*, *52*(3), 1-38. https://doi.org/10.1145/3323214

Luisi, J. V. (2014). *Pragmatic enterprise architecture - Strategies to transform information systems in the era of big data*. Waltham, MA: Morgan Kaufmann.

Manyika, J., Chui, M., Brown, B., Bughin J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York, NY: McKinsey Global Institute.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 60-68.

McKemmish, R. (2008). When is digital evidence forensically sound? In I. Ray & S. Shenoi (Eds.), *Advances in Digital Forensics IV* (pp. 3-15). New York, NY: Springer.

McKerchar, M. (2008). Philosophical paradigms, inquiry strategies and knowledge claims: Applying the principles of research design and conduct to taxation. *eJournal of Tax Research*, *6*(1), 5-22.

Mehta, P., Mathews, J., Kumar, S., Suryamukhi, K., Sobhan Babu, Ch., Kasi Visweswara Rao, S. V., Shivapujimath, V., & Bisht, D. (2019). Big data analytics for tax administration. In A. Kö, E. Francesconi, G. Anderst-Kotsis, A. Tjoa, & I. Khalil (Eds.), *Electronic Government and the Information Systems Perspective: 8th International Conference, EGOVIS 2019, Linz, Austria, August 26–29, 2019, Proceedings* (pp. 47-57). Cham, Switzerland: Springer Nature Switzerland AG.

Menke, M., & Schwarzenegger, C. (2019). On the relativity of old and new media: A lifeworld perspective. *Convergence: The International Journal of Research into New Media Technologies*, *25*(4), 657-672. https://doi.org/10.1177/1354856519834480

Microsoft & PricewaterhouseCoopers. (2018). *The data intelligent tax administration: Meeting the challenges of big tax data and analytics* [White paper]. Microsoft and PwC. https://www.pwc.nl/nl/assets/documents/the-data-intelligent-tax-administration-whitepaper.pdf

Ministry of Finance. (2015). Use of drone and aerial in tax potential search activities (in Bahasa Indonesia). In *Notes on innovation and transformation* (pp. 62-67). Jakarta, Indonesia: Balai Pustaka & Ministry of Finance (MoF).

Montasari, R. (2017). A standardised data acquisition process model for digital forensic investigations. *International Journal of Information and Computer Security*, *9*(3), 229-249. https://doi.org/10.1504/IJICS.2017.085139

Mousa, R. (2011). *E-government adoption process: XBRL adoption in HM revenue and customs and companies house*. [Doctoral dissertation, University of Birmingham]. UBIRA E Theses. https://etheses.bham.ac.uk/id/eprint/1752/

Oates, B. J. (2006). *Researching information system and computing*. London, England: SAGE Publications.

Oats, L. (2012). On methods and methodology. In L. Oats (Ed.), *Taxation: A fieldwork research handbook* (pp. 9-18). Abingdon, England: Routledge.

Olson, M. (2010). Document analysis. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of Case Study Research* (Volume 1, pp. 319-320). Thousand Oaks, CA: SAGE Publications, Inc.

Organisation for Economic Co-operation and Development. (2004). *Forum on tax administration: Guidance note: Guidance for the standard audit file - Tax version 2.0*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2015). *Addressing the tax challenges of the digital economy, Action 1: 2015 final report*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2016a). *Advanced analytics for better tax administration: Putting data to work*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2016b). *Technologies for better tax administration: A practical guide for revenue bodies*. Paris, France: OECD Publishing.

Organisation for Economic Co-operation and Development. (2017). *Technology tools to tackle tax evasion and tax fraud*. Paris, France: OECD Publishing.

Pethe, H. B., & Pande, D. S. (2016). An overview of cryptographic hash functions MD-5 and SHA. *IOSR Journal of Computer Engineering: National Conference on Recent Trends in Computer Science and Information Technology*, *5*, 37-42.

Pijnenburg, M., Kowalczyk, W., & van der Hel-van Dijk, L. (2017). A roadmap for analytics in taxpayer supervision. *The Electronic Journal of E-Government*, *15*(1), 19-32.

Podesta, J., Pritzker, P., Moniz, E. J., Holdren, J., & Zients, J. (2014). *Big data: Seizing opportunities, preserving values*. Washington, D.C.: Executive Office of the President, The White House. https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Pratomo, M. H. (2018). *Investigating tax compliance risks of large businesses in Indonesia* [Doctoral thesis, RMIT University]. RMIT University Research Repository. https://researchrepository.rmit.edu.au/esploro/outputs/doctoral/Investigating-tax-compliance-risks-of-large-businesses-in-Indonesia/9921864066501341

PwC. (2015). Tax policy. *PwC*. http://www.pwc.com/gx/en/services/tax/tax-policy-administration/what-is-tax-policy.html

Richardson, V. J., Teeter, R. A., & Terrell, K. L. (2019). *Data analytics for accounting*. New York, NY: McGraw-Hill Education.

Rogers, H., & Oats, L. (2012). Case studies. In L. Oats (Ed.), *Taxation: A fieldwork research handbook* (pp. 26-33). Abingdon, England: Routledge.

Sakti, N. W. (2021, September 7-9). *Implementation of data analytics in Directorate General of Taxes* [Conference presentation]. The 2nd Conference of Belt and Road Initiative Tax Administration Cooperation Forum - Digitalization of Tax Administration, Online. https://www.britacom.org/news/2ndOnlineMeeting/

Salijeni, G., Samsonova-Taddei, A., & Turley, S. (2019). Big data and changes in audit technology: Contemplating a research agenda. *Accounting and Business*, *49*(1), 95-119. https://doi.org/10.1080/00014788.2018.1459458

Santos, M. R. C., Laureano, R. M. S., & Albino, C. E. R. (2018). How tax audit and tax advisory can benefit from big data analytics tools data analysis and processing in relational databases using SQL Server and Power Pivot & Power View in Excel. *2018 13th Iberian conference on information systems and technologies* (CISTI), 1-6. https://doi.org/10.23919/CISTI.2018.8399472

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano (2012). *Analytics: The real-world use of big data - How innovative enterprises extract value from uncertain data*. Somers, NY: IBM Corporation.

Tian, F., Lan, T., Zheng, Q., Chao, K-.M., Godwin, N., Shah, N., & Zhang, F. (2017). Mining suspicious tax evasion groups in big data (Extended abstract). *Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 25-26. https://doi.org/10.1109/ICDE.2017.19

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: An overview. *Accounting Horizons*, *29*(2), 381-96. https://doi.org/10.2308/acch-51071

Veit, A. (2019). Swimming upstream: Leveraging data and analytics for taxpayer engagement – An Australian and international perspective. *eJournal of Tax Research*, *16*(3), 474-499.

Volvach, D., & Solovyev., M. (2018). Tax administration in the digital era: The FTS of Russia Approach. In Intra-European Organisation of Tax Administrations (Ed.), *Impact of digitalisation on the transformation of tax administrations* (pp.13-15). Budapest, Hungary: IOTA.

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320-330. https://doi.org/10.1057/palgrave.ejis.3000589

The World Bank. (2002). *The e-government handbook for developing countries: A project of InfoDev and the Center for Democracy and Technology*. Washington D.C.: The World Bank Group.

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. *The Qualitative Report*, *20*(2), 134-152. https://doi.org/10.46743/2160-3715/2015.2102

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Los Angeles, CA: SAGE Publications, Inc.

Zuca, M., & Tinta, A. (2018). The contribution of computer assisted auditing techniques (CAAT) and of the business intelligence instruments in financial audit. *Academic Journal of Economic Studies*, *4*(1), 183-191.