

Data Wrangling dalam Administrasi Pajak Indonesia: Studi Etnografi Digital

Agung Darono^α

▪ Email:
agungdarono@kemenkeu.go.id

^α Widyaiswara Ahli Utama Pusdiklat
Pajak

Riwayat artikel:

- Diterima 8 Feb 2021
- Direvisi 12 Jan 2023
- Disetujui 26 Sept 2023
- Tersedia online 2 April 2024

Abstract

The Indonesian tax administration requires data analytics for monitoring taxpayer compliance. Data wrangling is one of the stages in data analytics as a combination of multiple techniques and computing tools in such a way as to prepare data from various sources and formats to prepare data sets that can be analyzed to support organizational decision-making and actions. Using a digital ethnographic research strategy, this study found that data wrangling is part of the “digital-self” in everyday life of the organizations-actors in implementing data analytics as part of the tax administration function. This paper provides a framework to understand how the digital-self becomes part of the technological aspects of data wrangling at various levels of tax administration functions. The framework is expected to become a reference in formulating the provisions on data wrangling procedures for various levels, functions, and purposes of tax administration, or it could be applied to a broader scope of public finance management.

Keywords:

analisis, etnografi, digital-self, kepatuhan, pajak

JEL Classification :

K240, M150 , O330

1. PENDAHULUAN

Sistem administrasi perpajakan yang berlaku di Indonesia memberikan wewenang kepada Direktorat Jenderal Pajak (DJP) untuk mengumpulkan berbagai bukti dalam rangka pengawasan terhadap pelaksanaan ketentuan hak dan kewajiban wajib pajak. Perkembangan teknologi informasi menjadikan bukti ini didominasi oleh keberadaan data elektronik. Konsekuensinya adalah perlunya kemampuan DJP untuk memperoleh, mengolah, dan menganalisis data elektronik tersebut untuk melaksanakan tindakan penegakan hukum yang tepat jika berdasarkan pengawasan yang dilakukan ditemukan adanya ketidakpatuhan terhadap ketentuan perpajakan yang berlaku. Kompleksitas untuk menangani bukti dalam bentuk data elektronik ini semakin meningkat. Selain karena otomatisasi atas berbagai proses bisnis wajib pajak juga adanya berbagai aplikasi teknologi baru yang digunakan oleh wajib pajak seperti robotika, media sosial untuk bisnis, ataupun *Internet of Things (IoT)* beserta aplikasi integrasi data yang menyertainya sehingga mendorong tumbuhnya *big data* sebagai bentuk repositori data bisnis yang ukurannya (*volume*) besar, bertumbuh cepat (*velocity*), dengan format data yang bervariasi (*variety*). Situasi ini kemudian memunculkan sebuah disiplin dan *body of knowledge* yang disebut dengan analitika data perpajakan (*tax data analytics*) (ANAO 2008; Brink dan Hansen 2018; Sinnott 2016; Pijnenburg et al. 2017; EY 2017; Houser dan Sanders 2018; IOTA 2016; Microsoft dan PwC 2017; Carina dan Thompson 2019; Darono 2020).

Analitika data perpajakan, merujuk kepada uraian di atas, merupakan istilah yang menunjukkan bagaimana kerangka pengetahuan analitika data yang berlaku umum dapat diimplementasikan dalam fungsi-fungsi administrasi pajak dengan berbagai variasi tingkat kompleksitas analisis sesuai dengan proses bisnis yang ingin diselesaikan. Administrasi pajak merupakan salah area yang cukup intensif memanfaatkan kerangka pengetahuan analitika data ini. Beberapa inisiatif terkait dengan pengembangan dan implementasi analitika data perpajakan ini telah dilakukan oleh beberapa organisasi, antara lain oleh IOTA (2016) ataupun OECD (2011). Publikasi dari hasil konferensi yang dilaksanakan oleh kedua organisasi ini menunjukkan bagaimana berbagai otoritas pajak di seluruh dunia berupaya untuk menggali berbagai insight dari data yang dimilikinya untuk kepentingan pengawasan terhadap pelaksanaan peraturan perpajakan. Kajian yang dilakukan oleh Pijnenburg et al. (2017) dan juga Veit (2019) mengungkapkan berbagai aspek administrasi pajak yang dapat dibantu dengan analitika data, antara lain adalah pengawasan kepatuhan yang harus dilakukan secara masif, analisis risiko untuk pemilihan wajib pajak yang diaudit, membuktikan adanya tindak pidana perpajakan, menghadapi sengketa di pengadilan hingga menilai efektivitas suatu kebijakan.

Dalam kaitannya dengan implementasi analitika data untuk kepentingan administrasi perpajakan ini, DJP juga telah menyatakan posisinya tentang bagaimana mengadopsi analitika data ini sebagai bagian dari sistem administrasi pajak yang dijalankannya. Dalam Cetak Biru TIK DJP (DJP 2015), dengan menggunakan istilah pilar pengembangan TIK, disebutkan bahwa salah satu pilar tersebut adalah *big data analytics*. Djuniardi (2018) menyatakan bahwa DJP telah mengembangkan prototipe dan *use case* yang menggunakan *platform* teknologi *big data* sebagai perangkat untuk melaksanakan analitika data terkait dengan beberapa fungsi administrasi perpajakan, seperti aplikasi *social network analytics* untuk mengidentifikasi hubungan istimewa dan kemungkinan terjadinya transaksi tidak wajar antar wajib pajak tersebut.

Dalam praktik analitika data dikenal adanya beberapa data *analytics framework* atau metodologi (lihat misalnya Rollins 2015 atau Panoply 2019) ataupun disebut juga dengan data *analytics life cycle* (lihat misalnya: EMC 2015 atau Keller et al. 2020). Dari beberapa *framework/metodologi/life cycle* tersebut dapat dinyatakan bahwa tahapan pekerjaan dalam analitika data meliputi: identifikasi masalah bisnis yang memerlukan dukungan analitika data, data *wrangling*, penyusunan model, eksekusi model, komunikasi hasil, dan operasionalisasi. Salah satu tahap yang menarik untuk ditelaah lebih mendalam dari siklus pekerjaan analitika data ini adalah

data wrangling. Beberapa dari kajian tentang analitika data menyebutkan bahwa data *wrangling* (dikenal juga dengan berbagai istilah lain, seperti data *preparation*, *extract-transform-load/ETL*, *pre-analysis*, atau data *cleansing*) adalah bahwa ia merupakan aktivitas yang dapat memakan hingga 80 persen waktu yang diperlukan dari seluruh kegiatan analitika data (Press 2016; Rattenbury et al 2017; Horton 2019; Panoply 2019). Bahkan Rattenbury et al (2017) secara tegas menekankan sangat pentingnya kedudukan data *wrangling* dalam analitika data dengan menyatakan bahwa terdapat miskonsepsi yang lumrah ditemui, yaitu analisis data adalah sebuah proses untuk menjalankan algoritma statistik dalam lingkungan sistem komputasi berdaya besar. Pada kenyataannya, hal itu hanyalah satu langkah akhir dari serangkaian berbagai proses kompleks lain yang berlangsung sebelumnya. Analisis data dapat menghabiskan 50 s.d. 80 persen waktu kerjanya untuk melakukan data *wrangling* sehingga kemudian data tersebut dapat dianalisis. Artinya, tahapan data *wrangling* merupakan pekerjaan yang cukup memakan sumber daya dan harus dilakukan karena merupakan prasyarat agar data yang tersedia dapat diolah lebih lanjut. Hal ini perlu dilakukan agar tujuan untuk mendapatkan insight dari data yang tersedia tercapai untuk mendukung seluruh aktivitas organisasi.

Kondisi terkait dengan data *wrangling* sebagaimana yang telah dijelaskan di atas mendorong peneliti untuk lebih jauh mengungkapkan apa yang terjadi dengan data *wrangling* yang dilaksanakan sebagai bagian dari analitika data di DJP sebagai otoritas perpajakan Indonesia. Pengungkapan tersebut akan dilakukan dengan menggunakan strategi penelitian etnografi digital (Pink et al. 2016; Bakry 2017; Kautsarina 2018; Beuving 2020) dengan argumentasi bahwa pemahaman akan penerapan data *wrangling* sebagai bagian dari implementasi analitika data perpajakan tidak hanya perlu dikaji dari aspek teknis-teknologi semata. Hal ini sejalan dengan yang diungkapkan oleh hasil kajian Kling et al. (2005), Avgerou (2000), dan juga Orlikowski dan Robey (1991), bahwa implementasi teknologi informasi juga perlu didalami dari sisi sosial-kultural-organisasionalnya, termasuk dalam hal ini adalah penggunaan perspektif etnografi. Sebagai sebuah studi etnografi digital, penelitian ini berusaha untuk menyajikan pemahaman (*verstehen*, lihat misalnya Neuman 2014) yang dalam hal ini adalah tentang bagaimana terjadinya *interplay* antara "*digital-self*" (lihat misalnya Schau and Gilly 2003; Kozinets 2010) dari para aktor-organisasi di bidang analitika data perpajakan ini dengan berbagai artefak teknologi-organisasi yang ada. Pemahaman atas perilaku tersebut diharapkan dapat menjadi tambahan referensi terkait dengan implementasi analitika data perpajakan ataupun dalam skala yang lebih luas analitika data di lingkungan administrasi pemerintahan.

Berdasarkan hal tersebut maka rumusan masalah penelitian adalah bagaimana memahami pelaksanaan data *wrangling* sehingga dapat menjadi bagian yang mendukung implementasi analitika data yang diimplementasikan di DJP. Kontribusi yang ingin diajukan oleh kajian ini setidaknya mencakup dua hal. Pertama, dari sisi praktik analitika data, penelitian ini merupakan sebuah upaya untuk menyajikan kerangka pemahaman tentang pelaksanaan data *wrangling* sebagai bagian dari analitika data, khususnya untuk kepentingan perpajakan di Indonesia. Kedua, dari sisi metodologi penelitian, hasil kajian ini diharapkan menjadi tambahan rujukan yang berkaitan dengan penggunaan strategi penelitian etnografi (digital) dalam penelitian administrasi perpajakan ataupun administrasi publik secara lebih luas.

Sistematika tulisan ini adalah sebagai berikut. Bagian pertama menyajikan latar belakang dan rumusan masalah penelitian. Selanjutnya, bagian kedua akan menguraikan kerangka teoritis dan penelitian terdahulu terkait dengan *data wrangling* dan analitika data perpajakan. Berikutnya, bagian ketiga akan menjelaskan metodologi penelitian. Kemudian, bagian keempat akan mendiskusikan temuan penelitian. Terakhir bagian kelima akan menyajikan kesimpulan dan saran.

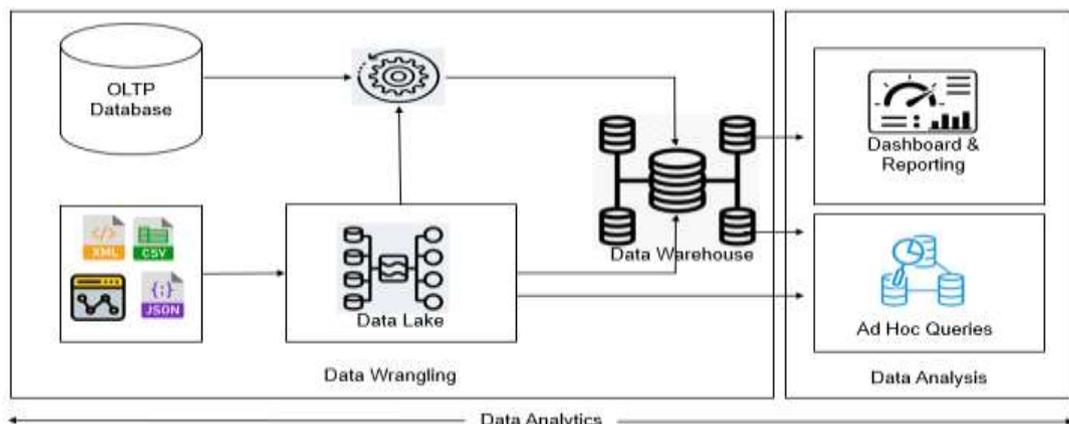
2. TINJAUAN PUSTAKA

2.1. Analitika data dalam administrasi pajak

Penjelasan tentang bagaimana sistem informasi bekerja sebagai bagian dari keseluruhan manajemen organisasi kebanyakan menggunakan dua istilah ini: *online transaction processing* (OLTP) dan *online analytical processing* (OLAP) (Rainer Jr. dan Cegielski 2011; Laudon dan Laudon 2012 ; Harrison 2015; Hall 2015). Dalam pandangan kerangka kerja pembagian ini, OLTP adalah sistem aplikasi untuk mengotomasikan pekerjaan klerikal, rutin, terstruktur, berulang, dan merupakan transaksi yang atomik dan terisolasi. Contoh: order tiket pesawat terbang, aplikasi PoS (*point-of-sale*) yang berada di meja kasir sebuah toko swalayan, ataupun penyetoran rekening tabungan yang ada di *teller* sebuah bank, serta penjualan tiket. Sementara itu, OLAP adalah sistem aplikasi yang memanfaatkan data yang telah dikumpulkan, baik melalui sistem OLTP ataupun sistem pengolahan data lainnya sehingga dapat digunakan mendukung seluruh fungsi organisasi.

Salah satu pembeda OLTP dengan OLAP adalah pendekatan dalam penyimpanan datanya. OLTP menyimpan datanya ke dalam database yang kemudian dengan proses yang disebut dengan *Extract-Transform-Load* (ETL) akan ditransformasi menjadi *data warehouse*. Perkembangan selanjutnya dari teknologi manajemen data ini kemudian juga memunculkan istilah *data lake* sebagai salah satu area dalam manajemen data yang dapat digunakan untuk memperkaya sumber daya sebagai pendukung berbagai fungsi manajemen dan pengambilan keputusan organisasi (Devlin 2018; Palanivel dan Manikandan 2019; Panoply 2019; Soddad et al. 2020;). Gambar 1 (dimodifikasi dari Devlin 2018) menjelaskan hubungan antara *database*, *data warehouse*, dan *data lake* sebagai sebuah ekosistem manajemen data yang saat ini telah diimplementasikan di berbagai organisasi.

GAMBAR-1: Hubungan antara *file*, *database*, *data lake*, *data warehouse* dan *data wrangling* untuk analitika data



sumber: dimodifikasi dari Devlin (2018)

Penjelasan atas Gambar 1 adalah sebagai berikut. *Database* merupakan area manajemen data terstruktur dalam bentuk tabular (baris dan kolom), dalam format data yang relatif terbatas (terutama huruf, angka, dan tanggal) dengan relasi antar data yang sudah didefinisikan sebagai representasi dari berbagai entitas dan *events* dalam suatu organisasi. *Database* terbentuk dari adanya aktivitas manajemen data yang bekerja secara *real-time* dan berbasis transaksi (atau dikenal dengan operasi *Create-Read-Update-Delete/CRUD*) di mana data yang akan disimpan telah melalui proses validasi secara ketat sesuai dengan proses bisnis yang telah dibakukan. Sementara itu, *data lake* terbentuk dari proses yang agak berbeda, terutama dari sisi validasi data yang cenderung longgar atau dengan kata lain “yang penting data masuk dulu, validasi nanti saja pada saat digunakan”. Baik untuk *database* atau *data lake*, data yang masuk dapat berasal dari beberapa aliran data, baik berupa *events* (peristiwa yang biasanya dipicu oleh perangkat dengan sensor), *measures* (tindakan pengguna seperti entri data dengan *keyboard* atau *mouse*), *messages* (pesan yang dipertukarkan antar sistem seperti data yang dikirim via mesin *electronic data capture/EDC*). Pada

area manajemen data yang lain, setelah data dalam yang berasal dari *database* ataupun *data lake* melalui tahapan *data wrangling* maka terbentuklah *data warehouse* yang sifatnya (merujuk Inmon dan Linstedt 2015) adalah: *non-volatile, integrated, subject-oriented*, dan *time-variant*.

Berdasarkan konfigurasi manajemen data sebagaimana telah diuraikan tersebut, para pengguna data dapat mempunyai beberapa alternatif cara untuk akses data, apakah melalui aplikasi yang *menu-driven* untuk mengakses *database*, misalnya tampilan berupa *dashboard* ataupun menggunakan perangkat yang memungkinkan *ad-hoc query* langsung ke *data warehouse* ataupun *data lake*. Tentu saja dengan tetap memperhatikan *data governance* seperti kesesuaian hak akses dengan tugas/pkerjaan pengguna data, jaminan keamanan data, ataupun bila diperlukan adalah adanya catatan tentang pemanfaatan data yang telah diaksesnya. Konfigurasi data manajemen sebagaimana yang diuraikan di atas mendorong berkembangnya disiplin dan kerangka pengetahuan analitika data (*data analytics*) sebagai proses untuk mengidentifikasi, mentransformasi, dan memodelkan data dengan tujuan untuk menemukan tren, pola, ataupun hubungan sehingga dapat menggambarkan fenomena kehidupan-nyata (Deloitte 2016; Keller et al. 2020; Cuadrado-Gallego dan Demchenko 2020). Secara umum, analitika merupakan kerangka pengetahuan (*body of knowledge*) yang berkaitan dengan penggunaan data, analisis kuantitatif, model eksplanatori dan prediktif, serta manajemen berbasis-fakta secara ekstensif untuk mempercepat keputusan dan tindakan (Davenport dan Harris 2007). *Analytics* adalah *science of analysis* (White dan Imhoff 2010). Menurut Power et al. (2018), *analytics* berbeda dengan *analysis*, adanya akhiran “-ics” tersebut menunjukkan adanya kerangka pengetahuan ataupun prinsip-prinsip yang berkaitan dengan bagaimana caranya melakukan pekerjaan *analysis*.

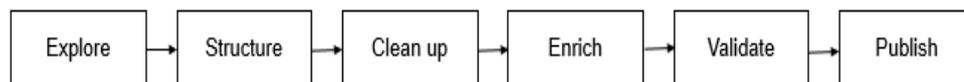
Jika uraian di atas digunakan untuk memahami sistem administrasi pajak, maka OLTP untuk menangani data detail yang terkait dengan registrasi, pelaporan (surat pemberitahuan, faktur, bukti pungut/potong), pembayaran (surat setoran, pemindahbukuan, kelebihan bayar), keberatan, banding, ataupun permohonan layanan lainnya oleh wajib pajak. Sementara itu, area OLAP dapat berangkat dari tahap yang paling sederhana, seperti menguji kepatuhan pajak secara formal yang dapat dilaksanakan secara masal karena kriterianya sudah terukur seperti menghitung denda karena keterlambatan pembayaran atau penyampaian surat pemberitahuan. Dari OLAP yang demikian itulah kemudian dapat dikembangkan menjadi analitika data perpajakan (*tax data analytics*). Jadi analitika data perpajakan adalah kombinasi antara pengetahuan perpajakan dengan berbagai perangkat dan teknik analisis data untuk menyajikan insight dan pengetahuan dari kumpulan data untuk kepentingan perpajakan (Regan et al. 2017; Microsoft dan PwC 2017, 2018; Tansey 2019). Sebagaimana analitika data pada umumnya, analitika data perpajakan mempunyai tahapan pekerjaan yang kurang lebih sama, ciri utamanya terletak pada sumber data yang menitikberatkan pada berbagai data ekonomi, keuangan, dan akuntansi. Dari sisi pemanfaatannya, analitika data perpajakan tentu saja akan spesifik berkaitan dengan berbagai fungsi administrasi pajak, seperti registrasi, pelaporan, pembayaran, penetapan, keberatan, ataupun penagihan (lihat misalnya Darono 2015; Tansey 2019).

2.2. Data wrangling

Merujuk kepada beberapa *framework/metodologi/life cycle* analitika data, salah satu tahapan yang harus dilakukan sebelum melakukan analisis data adalah *data wrangling* (atau dengan berbagai sebutan lainnya seperti *data pre-processing, data cleansing, data preparation, data extraction and transformation*, dll.). Tulisan ini memilih menggunakan istilah *data wrangling* dengan argumentasi, sebagaimana pendapat yang diajukan oleh Endel dan Piringner 2015 ataupun Azeroual (2020), bahwa istilah ini lebih secara komprehensif mencakup semua pekerjaan yang harus dilalui sebelum analisis data dengan melakukan *cleansing*, validasi, konversi, transformasi, dan kemudian *me-load* data tersebut ke dalam repositori data. Terdapat cukup banyak definisi dan penjelasan tentang *data wrangling* ini.

Pierson (2015) mendefinisikan *data wrangling* dengan proses dan prosedur yang digunakan untuk membersihkan dan mengkonversi data dari suatu format dan struktur menjadi format dan struktur sesuai dengan proses dan teknik analisis data yang akan dijalankan. Di sisi lain, Rattenbury et al. (2017) menjelaskan bahwa *data wrangling* mencakup aktivitas untuk memahami data, memahami relasi antar data, dan memutuskan data yang akan dianalisis. Dalam istilah lain dapat juga dinyatakan sebagai *access, profiling, transform, dan publish*. Sementara itu, Azeroual (2020) menyebutkan bahwa *data wrangling* terdiri dari langkah-langkah untuk *explore, structure, clean up, enriching, validation, dan publish*. Di sisi lain, McKinney (2018) menyatakan bahwa *data wrangling* merupakan istilah yang menggambarkan tentang keseluruhan proses untuk mengubah data yang tidak beraturan sehingga dapat digunakan untuk pekerjaan analisis. Berdasarkan beberapa definisi yang telah disebutkan di atas dan merujuk kembali Gambar 1, *data wrangling* dapat diuraikan lagi ke dalam beberapa langkah yang lebih detail. Mengacu Azeroual (2020), Gambar 2 adalah tahapan yang dilakukan dalam *data wrangling*. Tahapan ini dapat dilakukan secara iteratif dengan tujuan untuk mendapatkan data yang siap untuk dianalisis.

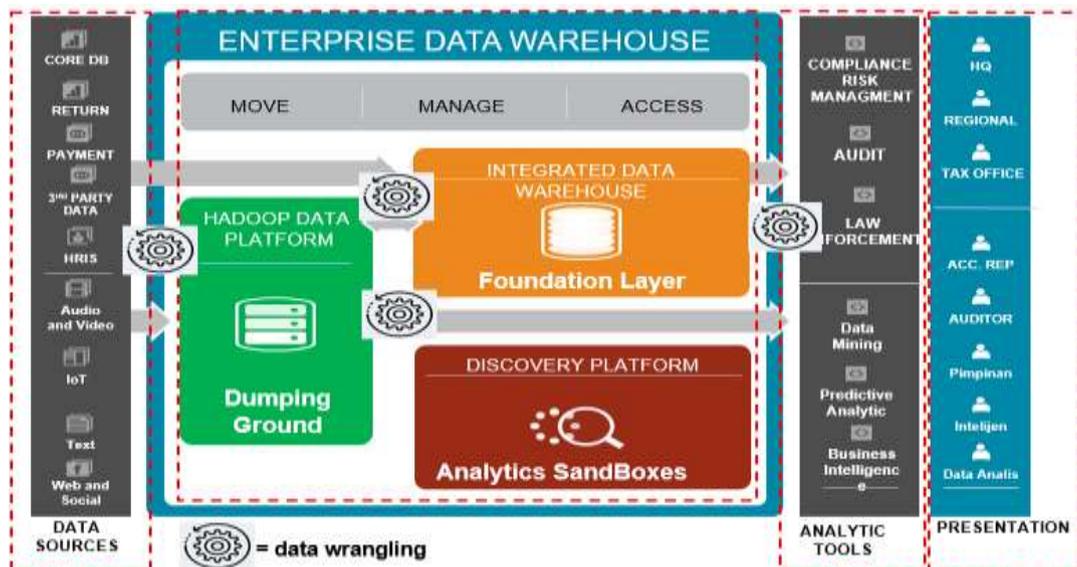
GAMBAR-2: Urutan pekerjaan dalam *data wrangling*



sumber: diadaptasi dari Azeroual (2020)

Jika uraian tentang data wrangling tersebut diterapkan untuk memahami sistem administrasi pajak, DJP telah menetapkan konfigurasi analitika data perpajakannya, terutama dikaitkan dengan adanya inisiatif untuk big data analytics (Djuniardi 2016; 2018). Gambar 3 menyajikan konfigurasi analitika data perpajakan yang telah diimplementasikan saat ini. Secara umum, konfigurasi ini mengikuti framework/metodologi yang telah dikembangkan oleh kalangan praktisi-industri teknologi informasi sebagaimana yang telah dijelaskan sebelumnya. Dari konfigurasi yang diimplementasikan terlihat bahwa tahapan data wrangling juga mempunyai fungsi yang krusial karena sebagaimana uraian sebelumnya, ia menjadi penghubung antara sistem OLTP (dengan berbagai varian aliran sumber datanya) ataupun data *lake* dengan *enterprise data warehouse*.

GAMBAR-3: Konfigurasi analitika data perpajakan di DJP



sumber: dimodifikasi dari Djuniardi (2018)

3. METODE PENELITIAN

Creswell (2013) menyatakan bahwa suatu penelitian harus menetapkan pendekatan penelitiannya. Hal ini meliputi tiga komponen, yaitu paradigma dan klaim pengetahuan, strategi penelitian, dan metode penelitian. Berdasarkan konsep tersebut maka paradigma yang dipilih adalah interpretif dan klaim pengetahuan yang diajukan adalah pemahaman mendalam dan empatik (*verstehen*) dengan sudut pandang dari para aktor yang terlibat dalam analitika data perpajakan (Berntsen et al. 2004; Bhattacharya 2008; Bakker 2010). Untuk mencapai hal tersebut, strategi penelitian yang dipilih adalah etnografi digital. Metode penelitian menggunakan teknik pengumpulan dan juga analisis data. Penelitian ini menggunakan teknik pengumpulan data berupa studi dokumentasi dan wawancara. Dokumen yang digunakan mencakup pengertian yang luas termasuk prosedur operasi standar, peraturan, manual sistem aplikasi, laporan dan juga hasil rekaman pelatihan yang bisa diakses oleh penulis. Wawancara dilakukan kepada para pegawai DJP yang sedang ditugaskan untuk bermitra dengan peneliti sebagai pengajar dalam pelatihan untuk mata pelajaran yang berkaitan dengan teknologi informasi dan analisis data perpajakan (misalnya pengajar untuk pelatihan administrator sistem, forensik digital, audit berbantuan komputer, dan analitika data perpajakan). Penulis, dengan merujuk metode yang diajukan oleh Bolls (2012), juga mengamati keseharian dari para narasumber selama berinteraksi dengan menekan perhatian pada berbagai ide ataupun praktik yang berkaitan dengan *data wrangling* dalam kaitannya dengan analisis data perpajakan.

Strategi penelitian yang dipilih adalah etnografi digital. Alasan penggunaan strategi ini adalah karena penggunaan berbagai perangkat digital merupakan hal yang tidak dapat lagi dipisahkan dari kehidupan sosial manusia (termasuk di dalamnya tentu saja lingkungan organisasi-pemerintahan) saat ini. Etnografi digital ini merupakan pengembangan dari etnografi konvensional yang pada dasarnya merupakan cara untuk melihat kebiasaan dan tingkah laku masyarakat dalam konteks sosial (CIPG 2012; McKerchar dalam Bolls 2012; Björklund Larsen 2013; Johannesson dan Perjons 2014; Sugiyanto dan Winardi 2018). Etnografi digital dikembangkan sebagai cara untuk memahami berbagai aspek kehidupan sosial yang “berada dalam serta menggunakan perangkat digital” (Schau dan Gilly 2003). Hal yang ingin diungkapkan dalam etnografi digital adalah:

- *experience* (apa yang orang rasakan);
- *practice* (apa yang orang lakukan);
- *things* (objek yang menjadi bagian dari hidup kita);
- *relationship* (lingkungan sosial yang diakrabi);
- *social world* (kelompok dan konfigurasi sosial yang lebih luas di mana orang berhubungan satu sama lain);
- *locality* (konteks bersama secara fisik aktual yang seseorang tempati);
- *event* (pertemuan berbagai hal dalam konteks publik)

yang berkaitan dengan keberadaan berbagai aspek digital dalam sebuah situasi sosial yang sedang diteliti (Pink et al. 2016).

Etnografi digital (ada yang menyebutnya dengan *virtual ethnography* atau *netnography*) merupakan sebuah upaya untuk menggali bagaimana aktor-sosial menjadi dan menggunakan *digital-self* (Schau dan Gilly 2003; Kozinets 2010; Pink et al. 2016). Tulisan ini mengajukan definisi operasional tentang *digital-self* ini, dengan mengacu pada hasil kajian Schau dan Gilly (2003), Kozinets (2010) dan juga Pink et al. (2016). Namun demikian, agak sedikit berbeda dari cakupan yang diungkapkan Schau dan Gilly (2003) yang cenderung membatasi *digital-self* dengan aktivitas online di internet, studi ini berupaya untuk memperluas pengertian tersebut dengan memasukkan keberadaan manusia sebagai makhluk sosial, di dunia yang sudah terdigitalisasi sedemikian rupa, untuk kemudian berusaha mengelaborasi berbagai kemungkinan dalam dunia-digital itu menjadi bagian dari tindakan kesehariannya dan bahkan mungkin mengubah identitasnya. Dalam konteks penelitian ini, hal yang ingin disajikan adalah pemahaman yang berasal dari hasil refleksi penulis

(lihat misalnya Pink et al. 2016; Johannesson dan Perjons 2014) tentang keberadaan *digital-self* ini menjadi bagian dari tindakan keseharian para aktor-organisasi dalam kaitannya dengan *data wrangling* dan analitika data.

4. HASIL ANALISIS DAN PEMBAHASAN

Secara ringkas, penelitian ini mengajukan tiga temuan penting. Pertama, bagaimana *data wrangling* (merujuk Pink et al. 2016) menjadi *practices* yang keberadaannya menjadi aspek yang tidak terpisahkan dari implementasi analitika data untuk kepentingan administrasi perpajakan di DJP saat ini. Kedua, *data wrangling* dengan segala aspek dan pola relasinya telah membawa beberapa aktor-organisasi yang sebenarnya adalah *end-user* namun karena berbagai faktor menggunakan kemampuannya untuk juga melakukan *data wrangling* untuk memenuhi kebutuhan data dalam pelaksanaan pekerjaan. Hal ini ternyata mengarah pada apa yang disebut oleh Schau dan Gilly (2003) dan juga Kozinets (2010) sebagai “diri-yang-digital” (*digital-self*). Ketiga, penulis mengajukan kerangka yang diharapkan dapat digunakan untuk memahami praktik *data wrangling* sebagai bagian dari analitika data dalam administrasi perpajakan.

4.1. *Data wrangling* sebagai *organizational-practices*

Misi utama otoritas pajak adalah memastikan bahwa semua wajib pajak telah melaksanakan seluruh kewajiban dan haknya sesuai dengan ketentuan hukum yang berlaku. Untuk itu, otoritas pajak memerlukan *tax policy*, *tax law*, dan *tax administration* (Pistone et al. 2019). Implementasi dari administrasi perpajakan adalah adanya berbagai prosedur (atau populer dengan istilah proses bisnis) yang menjadi panduan bagi wajib pajak untuk melaksanakan hak dan kewajibannya. Di sisi lain, proses bisnis ini juga menjadi acuan bagi para petugas pajak untuk menjalankan tugas mengawasi kepatuhan wajib pajak.

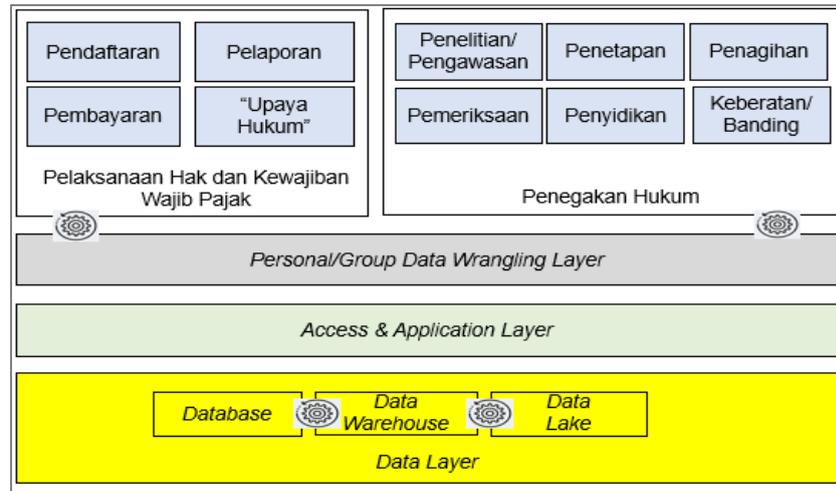
Secara garis besar proses bisnis dalam sistem administrasi perpajakan yang dikelola oleh pemerintah pusat Indonesia dapat digambarkan dalam Gambar 4. Proses bisnis pada saat ini tidak dapat dilepaskan dari dukungan teknologi informasi terutama manajemen data. Oleh karena itu, Gambar 4 menyertakan pula kaitan antara prosedur dengan data beserta mekanisme untuk menggunakan data tersebut. *Data layer* terdiri repositori data beserta tata kelolanya. Untuk menghasilkan data yang dapat digunakan manajemen data melaksanakan, menurut istilah penulis sebagai *enterprise-wide data wrangling*. Istilah yang demikian ini penulis ajukan karena area ini merupakan agregasi dari semua data DJP pada level *enterprise* yang berasal dari berbagai sumber data disimpan. *Data wrangling* (dengan semua tahapan sebagaimana dipaparkan dalam Gambar 2) dilaksanakan dengan menggunakan tata kelola (termasuk penjaminan mutu data) sehingga kualifikasi data dapat dibedakan untuk disimpan dalam *database*, *data warehouse*, ataupun *data lake*.

Access & application layer merupakan area yang mengandung berbagai aplikasi yang dikembangkan dan diimplementasi oleh Direktorat Teknologi Informasi dan Komunikasi. Beberapa aplikasi, misalnya SIDJP, Approweb, Apportal (DDIP 2020) akan menampilkan data dengan menggunakan yang disimpan dalam *database*. Tampilannya bersifat detil (*fine-grained*) untuk setiap *record* wajib pajak, seperti tampilan tentang profil wajib pajak yang menguraikan seluruh data yang terkait wajib pajak yang dipilih. Sementara itu, *data warehouse* dan *data lake* lebih banyak dimanfaatkan berbagai aplikasi analitika yang memerlukan data dalam bentuk agregat. Contoh penggunaannya adalah untuk berbagai kepentingan kebijakan ataupun pengembangan dan implementasi *compliance risk management* yang terutama dikaitkan dengan penentuan wajib pajak yang diperiksa.

Situasi yang menarik untuk diungkapkan sebagai temuan penelitian ini adalah apa yang disajikan dalam Gambar 4, yaitu keberadaan dari hal yang penulis sebut sebagai *personal/group data wrangling* sebagai kontras terhadap *enterprise-wide data wrangling*. Kedua jenis *data wrangling* ini perlu

penulis refleksikan (merujuk Pink et al. 2016) sebagai *practices* dalam organisasi dalam perspektif etnografi digital. Sebagai *practices*, *data wrangling* merupakan sebuah atrefak sosio-teknikal yang keberadaannya sengaja dibuat oleh para aktor-organisasi yang memerlukannya. *Data wrangling* dari sisi perwujudan fisik tetap saja *intangible* karena ia tetap berupa kombinasi dari serangkaian prosedur dan perangkat lunak komputasi yang digunakan untuk mengubah struktur data, keberadaannya sebagai *practices* lebih merupakan penanda bahwa implementasi analitika data memerlukan tahapan yang mungkin saja belum terpenuhi oleh prosedur perolahan data sebagaimana yang telah ditetapkan.

GAMBAR-4: *Data wrangling* sebagai bagian dari administrasi dan analitika data perpajakan



sumber: hasil analisis penulis

Terdapat fakta bahwa dalam beberapa situasi, para aktor-organisasi yang bertindak sebagai *end-user* dari data dan aplikasi masih memerlukan beberapa langkah *data wrangling* agar data yang mereka peroleh dari tampilan aplikasi standar dapat digunakan untuk mendukung tugas mereka. Contohnya adalah jika seorang pemeriksa pajak sebagai *end-user* memerlukan data tentang faktur pajak yang terkait dengan seluruh transaksi yang dilakukan oleh wajib pajak, baik dari sisi pembelian maupun penjualan yang dilakukannya, maka data tersebut belum tersedia dalam bentuk yang langsung dapat disandingkan *record-by-record*. *End user* perlu melakukan beberapa tahapan *data wrangling* sehingga dia mendapatkan data dalam format yang sesuai dengan kepentingan pekerjaannya. Situasi lain sebagai contoh untuk penjelasan tambahan tentang perlunya aktivitas *data wrangling* yang masih dilakukan pada level personal, seorang *account representative* yang sedang melakukan penggalan potensi sektor *e-commerce* maka dia akan melakukan *data wrangling* dari berbagai laman web (internet maupun intranet) sehingga kebutuhan data yang dia perlukan untuk melaksanakan tugasnya dapat terpenuhi.

Aktivitas *personal-* (hanya untuk kepentingan personal, seorang diri saja) ataupun *group-* (untuk sekelompok pengguna dalam sebuah unit yang sama atau rumpun pekerjaan yang sejenis) *data wrangling* ini pun pendekatan teknisnya juga bervariasi, yaitu: (1) *ad hocs* dengan menggunakan perangkat *query* data tertentu seperti Navicat, SQLWorkbench, Excel, ataupun Power BI; (2) menjadi sebuah "sub-aplikasi" yang berupa *add-in* dari aplikasi Excel (misalnya: Apiseta, e-Audit Utilities); (3) aplikasi yang otonom (misalnya MPN Info). Ketiga pendekatan teknis tersebut penggunaannya bergantung pada kebutuhan data, kompleksitas struktur data yang akan diambil, dan tentu saja kemampuan teknis dari setiap *end-user* yang akan melakukan *data wrangling*. Keberadaan *personal/group data wrangling* ini akan mengalami dinamika yang dipengaruhi oleh bagaimana respon organisasi. Dalam pandangan penulis, respon ini sebaiknya tetap

mempertimbangkan kemampuan organisasi memenuhi kebutuhan data yang diperlukan untuk pelaksanaan pekerjaan yang ada, jaminan kualitas data, dan kompetensi para pengguna dalam melakukan *personal/group data wrangling*.

4.2. *Data wrangling* dan *digital-self*

Data wrangling, walaupun tidak secara eksplisit bukan istilah yang digunakan, secara substantif telah menjadi bagian dari keseharian para aktor-organisasi di DJP tidak hanya terbatas pada fungsi yang mempunyai tugas sebagai pengelola data ataupun teknologi informasi namun sampai dengan tingkat tertentu juga pada fungsi yang lain, seperti pengawasan kepatuhan ataupun penegakan hukum. Dari data penelitian yang berhasil dikumpulkan, *data wrangling* merupakan aktivitas yang sudah menjadi kegiatan pekerjaan keseharian (*in everyday life*) para pegawai di DJP. Dengan mengadopsi frasa “... *in everyday life*...” yang dikemukakan oleh Geofmann dalam Silva (2013) sebagai bagian dari *impression management*. Istilah ini digunakan untuk menggambarkan bahwa setiap individu terus-menerus menyesuaikan kehadiran dirinya (*self-presentation*) agar sesuai dengan konteks sosial di mana dia sedang berada, dan ini kemudian berlangsung terus-menerus sehingga terbentuk sebagai hal wajar dalam aktivitas sehari-hari dan sampai dengan tingkat tertentu bahkan akan menjadi identitas dari diri-pelakunya.

Situasi yang demikian itulah kurang lebih yang terjadi dengan aktivitas *data wrangling* sebagai bagian dari analitika data perpajakan atau bahkan sistem administrasi perpajakan dalam skala yang lebih luas. Aktivitas *data wrangling* menjadi hal yang berlangsung setiap hari, dianggap sebagai hal yang biasa saja walaupun sebenarnya itu berawal dari sebuah aktivitas yang sebelumnya tidak pernah dilakukan sama sekali. Dalam istilah Kozinetsn (2010) dan juga Pink et al. (2016), situasi ini mengarah pada apa yang disebut sebagai *locality*. Bahwa, *data wrangling* adalah tindakan dalam keseharian para petugas pajak sebagai bagian dari pelaksanaan tugas dan fungsi mereka.

Sampai dengan tingkat tertentu, terdapat fenomena bahwa kemampuan *data wrangling* ini akan menjadi identitas dari pegawai yang melakukannya. Pengembangan aplikasi MPN-Info oleh *end-user* pada awalnya untuk memenuhi kebutuhan data personal saja, untuk melengkapi aplikasi MPN Harian Rekon (Kemenkeu 2015). Kemudian aplikasi ini dibagikan secara bebas sehingga bisa dimanfaatkan oleh banyak pengguna. Pengembang aplikasi ini pun secara suka rela kemudian menjadikan pemeliharaan aplikasi MPN-Info menjadi bagian kesehariannya ataupun bahkan mungkin telah menjadi identitas tambahan bagi dirinya.

Data wrangling juga dapat membentuk *digital-self* dari cara yang berbeda. Para pengguna data di beberapa fungsi perpajakan seperti pemeriksa pajak karena ketentuan yang mengatur pekerjaannya akan menghadapi kondisi di mana mereka harus mengolah data elektronik yang mereka dapat dari wajib pajak sebagai konsekuensi dari ketentuan yang mengatur pelaksanaan tugas mereka. Sesuai dengan Surat Edaran DJP SE-10/PJ/2017 tentang Petunjuk Teknis Pemeriksaan Lapangan dalam rangka Pemeriksaan untuk Menguji Kepatuhan Pemenuhan Kewajiban Perpajakan, pemeriksa pajak diminta untuk memahami konfigurasi sistem informasi yang diterapkan oleh wajib pajak dan mendapatkan data elektronik dari sistem informasi atau perangkat komputer wajib pajak untuk dapat diolah menjadi bukti pemeriksaan pajak yang ditugaskan kepadanya. Untuk mengolah berbagai berbagai jenis data yang diterima dari wajib pajak tersebut, beberapa pemeriksa pajak kemudian atas inisiatif mereka sendiri mengembangkan aplikasi *data wrangling* dalam bentuk Excel *add-in* untuk melakukan berbagai pekerjaan *data wrangling* (lihat kembali Gambar 2). Hal ini dilakukan agar data yang diperolehnya dari wajib pajak dapat diolah untuk menghasilkan kertas kerja pemeriksaan dan juga laporan hasil pemeriksaan. Berdasarkan uraian ini, *data wrangling* dapat dikatakan telah menjadi bagian yang tidak terpisahkan dari kegiatan pemeriksaan pajak. Kemampuan *data wrangling* dalam konteks ini bahkan sudah menjadi bagian identitas dari personal yang melakukannya.

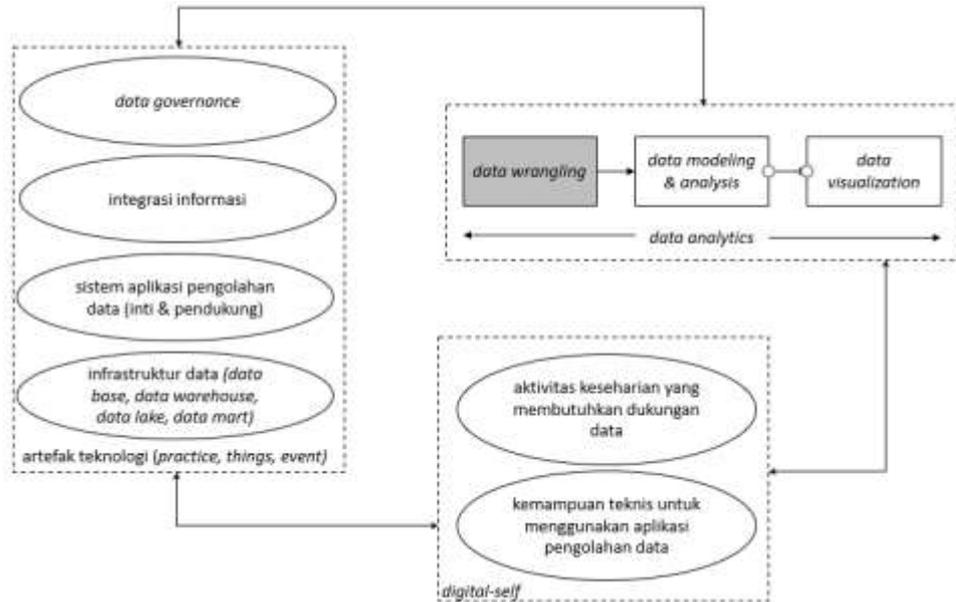
4.3. *Digital-self* dan *interplay* antar artefak: refleksi etnografis dan kerangka pemahaman

Artefak teknologi secara umum dicirikan secara sempit sebagai objek material yang dibuat oleh agen-manusia (*human-agent*) sebagai sarana untuk mencapai tujuan praktis. Merujuk Aristoteles, artefak teknologi adalah objek yang tidak dibuat oleh alam (Verbeek and Vermaas 2009). Atau, dalam pandangan digital etnografi yang dikemukakan oleh Pink et al. (2016) hal ini disebut dengan *thing*. Berada dalam konteks yang demikian ini, *data wrangling* merupakan salah satu *thing* (atau artefak teknologi) dalam administrasi perpajakan Indonesia. Diskusi pada bagian sebelumnya menunjukkan bagaimana *data wrangling* telah menjadi bagian dari praktik organisasional dan sampai dengan tahap tertentu, bahkan secara personal telah menumbuhkan sisi *digital-self* dari para aktor-organisasinya. Pada sisi lain, kajian etnografi digital juga menekankan pentingnya mengungkapkan dan memahami *event* sebagai kombinasi antara berbagai isu dan situasi kontemporer dalam sebuah situasi sosial-organisasional yang memengaruhi perilaku aktor-organisasi. Situasi yang demikian ini dalam hemat penulis mengarahkan penulis untuk mengungkapkan bagaimana situasi dan isu kontemporer organisasi ber-*interplay* dengan *digital-self* yang terbentuk dalam aktivitas *data wrangling* dalam administrasi perpajakan. Beberapa artefak DJP dalam situasi kontemporer organisasional DJP yang dalam catatan penulis seharusnya diperhatikan saling-pengaruhnya dengan *digital-self* yang terbentuk dari aktivitas *data wrangling* ini adalah:

- (1) munculnya fungsi penjaminan kualitas data di kantor pelayanan pajak sebagaimana ditetapkan dalam Peraturan Menteri Keuangan Nomor 184/PMK.01/2020 tentang Perubahan Atas Peraturan Menteri Keuangan Nomor 210/PMK.01/2017 tentang Organisasi dan Tata Kerja Instansi Vertikal DJP. Hal ini sebenarnya secara implisit menunjukkan meningkatnya peran *data wrangling*. Sebagaimana substansi aktivitas *data wrangling* adalah berupa *data preparation* untuk menjamin kualitas data sehingga analisis data dilakukan terhadap data yang valid dan reliabel.
- (2) adanya inisiatif untuk menerapkan *cooperative compliance* sebagaimana diatur dalam Keputusan Direktur Jenderal Pajak Nomor KEP-251/PJ/2020 tentang Pembentukan Tim Integrasi Data Wajib Pajak Pada Direktorat Jenderal Pajak Tahun 2020. Menurut KEP-251 *cooperative compliance* adalah hubungan kerja sama berbasis transparansi dan kepercayaan antara Wajib Pajak dengan DJP berdasarkan kepatuhan pajak sukarela yang bertujuan untuk menciptakan pemenuhan kewajiban perpajakan wajib pajak, baik secara formal maupun material, sesuai dengan ketentuan peraturan perpajakan yang berlaku. Salah satu komponen dari implemmentasi *cooperative compliance* ini adalah adanya *compliance arrangements*, yaitu perjanjian tertulis antara wajib pajak dan DJP untuk menyepakati kriteria-kriteria tertentu atau akun objek pelaporan dalam mekanisme integrasi data perpajakan.
- (3) pengembangan sistem inti administrasi (*core-system*) (sebagaimana diatur dalam Keputusan Menteri Keuangan Nomor 130/KMK.03/2020 tentang Tim Pembaruan Sistem Inti Administrasi Perpajakan) sebagai *enterprise information system* yang akan menangani seluruh (*end-to-end*) proses administrasi perpajakan. Pembaruan sistem yang diawali dengan rancang ulang seluruh proses bisnis perpajakan. Modernisasi sistem dan redesign proses bisnis ini diharapkan akan meningkatkan kapabilitas dan kinerja DJP. Dengan demikian, sistem ini diharapkan mampu menyediakan layanan dan pengawasan pajak yang lebih mudah, berkeadilan, dan andal. Implementasi sistem baru tersebut diharapkan bisa selesai pada 2024 (DDTC 2020). Intensitas *data wrangling* akan dipengaruhi oleh seberapa komprehensif *core-system* yang dikembangkan tersebut dapat memenuhi kebutuhan informasi para pengguna yang selama ini telah menjadikan aktivitas *data wrangling* tidak saja menjadi bagian pekerjaannya namun sampai dengan tahap tertentu telah menjadi bagian dari identitas dirinya yang cenderung menjadi *digital-self*.

Interplay antar berbagai artefak dalam situasi kontemporer sebagaimana diuraikan di atas membawa penulis kepada sebuah pemahaman yang dapat diungkapkan dari perspektif etnografi digital sebagai kerangka pemahaman tentang bagaimana *data wrangling* menjadi bagian dari proses organisasional administrasi perpajakan. Gambar 5 dengan menggunakan artefak TIK DJP dalam situasi kontemporer memaparkan bagaimana saling-pengaruh (*interplay*) antar-artefak terjadi kemudian memengaruhi aktor-organisasi mengembangkan praktik *data wrangling* sebagai bagian dari pelaksanaan tugas dan fungsi mereka, terutama yang berkaitan dengan analisis data perpajakan.

GAMBAR-5: Kerangka pemahaman *data wrangling* dalam analitika data perpajakan: perspektif etnografi digital



sumber: hasil analisis penulis

5. KESIMPULAN DAN REKOMENDASI

Data wrangling, dari praktik yang ada dapat disimpulkan sebagai kombinasi teknik dan perangkat pengolahan data tertentu sedemikian rupa untuk mempersiapkan data yang berasal dari berbagai sumber dan format. Tujuannya adalah menyediakan data yang siap untuk dianalisis sehingga dapat menyajikan berbagai wawasan (*insights*) sebagai pendukung pengambilan keputusan dan tindakan organisasi. Dari perspektif etnografi digital, tulisan ini menemukan dua hal yang perlu diperhatikan dari keberadaan dan praktik *data wrangling* sebagai bagian dari sistem administrasi perpajakan. Pertama, *data wrangling* dilakukan untuk kepentingan *enterprise-wide* dan menghasilkan data yang akan dikonsumsi oleh aplikasi standar. Namun demikian, sampai dengan derajat tertentu, terdapat pula praktik di mana para aktor-organisasi yang bertindak sebagai *end-user* dari data dan aplikasi masih memerlukan beberapa langkah *data wrangling* agar data yang mereka peroleh dari tampilan aplikasi standar dapat digunakan untuk mendukung tugas mereka. Pada area tugas tertentu, seperti pemeriksaan pajak, bahkan *data wrangling* ini dapat disebut sebagai *personal data wrangling* karena sifat aktivitasnya yang sangat spesifik sebagai akibat dari data yang harus diolah mempunyai karakteristik yang berbeda antara satu tugas pemeriksaan dengan tugas pemeriksaan yang lainnya. Kedua, dalam konteks analitika data yang demikian ini, pada level individual *end-user* ternyata masih diperlukan beberapa langkah untuk melakukan *personal data wrangling* agar data yang dia mereka peroleh dari berbagai aplikasi standar atau sumber data lain dapat digunakan sesuai dengan tugas mereka masing-masing. Situasi *data wrangling* yang sangat spesifik ditemui untuk beberapa area tugas, misalnya pemeriksaan pajak, yang secara *native* sifat

pekerjaannya memang akan menerima data dalam berbagai ukuran, sumber, dan format karena menghadapi perubahan sistem aplikasi sebagai sumber data dari wajib pajak yang sedang diperiksa. Apakah *data wrangling* hanya dapat dilakukan pada level *enterprise* saja agar kualitas dan validitas data akan lebih terjamin? Dalam pandangan penulis, jawabannya cenderung tidak sepenuhnya demikian, sekali lagi tentu dengan konteks bahwa sampai dengan saat ini masih terdapat kebutuhan data dari para *end-user* yang tidak secara final dapat sepenuhnya terpenuhi dari sistem aplikasi standar yang tersedia. Jadi, sampai dengan tingkat tertentu, *personal data wrangling* masih tetap dibutuhkan. Hal yang perlu dipersiapkan untuk kondisi tersebut adalah adanya *data governance* yang secara jelas menata aktivitas *data wrangling* sehingga tercapai sebuah titik optimal (ekuilibrium) antara pemenuhan kebutuhan data yang sesuai dengan tugas *end-user*, di satu sisi, dengan penjaminan kualitas dan keamanan data, di sisi yang lain.

Tulisan ini dengan menggunakan perspektif etnografi digital mengungkapkan *data wrangling* terutama sebagai artefak teknologi yang berasal dari *practice, things, relationship* dan *event* dari situasi sosial-organisasional administrasi perpajakan di DJP. Berdasarkan perspektif tersebut, kajian ini ingin berkontribusi kepada disiplin administrasi perpajakan, atau bahkan mungkin administrasi keuangan negara dan juga publik secara lebih luas. Kerangka ini diharapkan dapat membantu pemahaman tentang bagaimana aktivitas *data wrangling* sebagai bagian dari analitika data saling-memengaruhi (*interplay*) dengan pelbagai artefak TIK dalam lingkup administrasi perpajakan. Studi ini mengajukan kerangka pemahaman yang berkaitan dengan keberadaan *data wrangling* sebagai bagian dari administrasi perpajakan dari perspektif etnografi digital melalui tiga langkah, yaitu: (1) memahami keberadaan artefak teknologi sebagai *practice, thing* dan juga *event*; (2) memahami keberadaan *digital-self* yang terjadi pada aktor organisasi; (3) memahami kerangka *data analytics* yang diimplementasikan dalam suatu organisasi. Dari sisi metodologi penelitian, hasil kajian ini diharapkan berkontribusi dengan menyajikan sebuah studi kasus berkaitan dengan penggunaan strategi penelitian etnografi digital dalam penelitian administrasi perpajakan, yang kemungkinan juga dapat diterapkan untuk manajemen keuangan negara atau bahkan kajian administrasi publik dalam skala yang lebih luas.

Beberapa perkembangan situasi terakhir implementasi TIK DJP seperti inisiatif *cooperative compliance*, perubahan struktur organisasi yang memunculkan fungsi penjaminan kualitas data pada unit kantor pelayanan pajak, serta perancangan dan pengembangan kembali sistem informasi utama (*core-system*) administrasi perpajakan. Untuk itu, tulisan ini mengajukan rekomendasi agar manajemen TIK DJP dapat mengukuhkan sebuah tata kelola TIK DJP yang *robust* sehingga dapat meletakkan kedudukan *data wrangling* (pada level *enterprise* ataupun *personal/individual*) sejalan (*aligned*) dengan strategi sistem informasi organisasi tersebut secara keseluruhan.

6. DAFTAR PUSTAKA

- ANAO. 2008. "The Australian Taxation Office's Use of Data Matching and Analytics in Tax Administration." ANAO Audit Report No.30 2007-08. The Auditor-General Australian National Audit Office (ANAO).
- Avgerou, Chrisanthi. 2000. "IT and Organizational Change: An Institutional Perspective." *Information Technology and People* 13: 234–62.
- Azeroual, Otmane. 2020. "Data Wrangling in Database Systems: Purging of Dirty Data." *Data* 5 (2): 50. <http://dx.doi.org/10.3390/data5020050>.
- Bakker, Andrew. 2010. "Interpretivism." dalam *Encyclopedia of Case Study Research*, editor Albert J. Mills, Gabrielle Durepos, and Elden Wiebe, 486–92. Thousand Oaks, California: SAGE Publications, Inc.
- Bakry, Umar. 2017. "Pemanfaatan Metode Etnografi Dan Netnografi Dalam Penelitian Hubungan Internasional." *Jurnal Global & Strategis* 11: 15. <https://doi.org/10.20473/jgs.11.1.2017.15-26>.
- Berntsen, Kirsti E., Jennifer Sampson, and Thomas Østerlie. 2004. "Interpretive Research Methods

- in Computer Science.”
- Beuving, J. Joost. 2020. “Ethnography’s Future in the Big Data Era.” *Information, Communication & Society* 23 (11): 1625–39. <https://doi.org/10.1080/1369118X.2019.1602664>.
- Bhattacharya, Himika. 2008. “Interpretive Research.” dalam *The Sage Encyclopedia of Qualitative Research Methods*, editor Lisa M. Given, 4674–467. Thousand Oaks, California: SAGE Publications, Inc.
- Björklund Larsen, Lotta. 2013. “Moulding Knowledge into a Legal Complex: Para-Ethnography at the Swedish Tax Agency.” *Journal of Business Anthropology* 2: 209. <https://doi.org/10.22439/jba.v2i2.4159>.
- Bolls, Karen. 2012. “Ethnography and Tax Compliance.” dalam *Taxation A Fieldwork Research Handbook*, editor Lynne Oats, 50–58.
- Brink, William, dan Victoria Hansen. 2018. “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>.
- Carina, Federico, dan Travis Thompson. 2019. “Do IRS Computers Dream About Tax Cheats? Artificial Intelligence and Big Data in Tax Enforcement and Compliance.” *Journal of Tax Practice & Procedure*, no. Feb-Mar: 43–47. <https://www.crowell.com/files/2019-Feb-March-Do-IRS-Computers-Dream-About-Tax-Cheats-Federico.pdf>.
- CIPG. 2012. *Tentang Riset*. Center for Innovation Policy and Governance (CIPG). <http://cipg.or.id/wp-content/uploads/2015/05/CREAME-1.-Tentang-Riset-2012.pdf>.
- Creswell, John W. 2013. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, California: SAGE Publications.
- Cuadrado-Gallego, Juan J., dan Yuri Demchenko. 2020. “Introduction to the Data Science Framework.” In , 1–7. Springer.
- Darono, Agung. 2015. “Pengawasan Terhadap Kepatuhan Pajak Sebagai Praktik-Diskursif.” In *Konferensi Nasional Perpajakan UGM. Magister Akuntansi FEB UGM*.
- . 2020. “Data Analytics dalam Administrasi Pajak Di Indonesia: Kajian Institutional Arrangement.” *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)* 6: 195–211. <https://doi.org/10.35957/jatisi.v6i2.194>.
- Davenport, Thomas H., dan Jeanne G. Harris. 2007. *Competing on Analytics: The New Science of Winning*. Harvard Business Review Press.
- DDIP. 2020. “Overview Data Internal dan Eksternal.” Materi Pelatihan PJJ Data Dan Informasi. Jakarta: Direktorat Data dan Informasi Perpajakan - Direktorat Jenderal Pajak (DDIP).
- DDTC. 2020. “Kontrak Pengadaan System Integrator Pembaruan Sistem Inti DJP Diteken.” <https://news.ddtc.co.id/kontrak-pengadaan-system-integrator-pembaruan-sistem-inti-djp-diteken-26214>.
- Deloitte. 2016. “Big Data Analytics for Policy Making.” European Commission, Directorate-General for Informatics.
- Devlin, Barry. 2018. “The EDW Lives On The Beating Heart of the Data Lake.” 9sight Consulting. <http://www.asprom.com/technologie/9sight.pdf>.
- DJP. 2015. “Cetak Biru Teknologi Informasi Dan Komunikasi Direktorat Jenderal Pajak Kementerian Keuangan Republik Indonesia Tahun 2015-2019.” Direktorat Jenderal Pajak.
- Djuniardi, Iwan. 2016. “Next Generation Data Analysis The Implementation of Big Data in Directorate General of Taxes Republic of Indonesia.” dipresentasikan dalam the 13th ATAIC Annual Technical Conference, Melaka, Malaysia. <http://ataic13.hasil.gov.my/SLOT3/IndonesiaDGTBigData.pdf>.
- . 2018. “Journey to Big Data.” Indonesian Ministry of Finance.
- EMC. 2015. “Data Analytics Lifecycle.”, 25–62. EMC Service. <https://doi.org/10.1002/9781119183686.ch2>.
- Endel, Florian, dan Harald Piringer. 2015. “Data Wrangling: Making Data Useful Again.” *IFAC-Papers On Line* 48: 111–12. <https://doi.org/10.1016/j.ifacol.2015.05.197>.
- EY. 2017. “The Tax Authority of the Future: How Tax Authorities Are Using Analytics to Deliver New Levels of Value.” UK: Ernst & Young Global Limited.
- Hall, James A. 2015. *Accounting Information Systems*. Cengage Learning.

- Harrison, Guy. 2015. *Next Generation Databases: NoSQL, NewSQL, and Big Data*. California: Apress. <https://doi.org/10.1007/978-1-4842-1329-2>.
- Horton, Greg. 2019. "Reversing the 80-20 Rule in Data Wrangling for AI and Machine Learning." *Blog.Timextender.Com* (blog). 2019. <https://blog.timextender.com/reversing-the-80-20-rule-in-data-wrangling>.
- Houser, Kimberly A., dan Debra Sanders. 2018. "The Use of Big Data Analytics by the IRS: What Tax Practitioners Need to Know." *Journal of Taxation* 128 (2).
- Inmon, W. H., dan Daniel Linstedt. 2015. "The Operational Data Store." dalam *Data Architecture: A Primer for the Data Scientist*, editor by W. H. Inmon and Daniel Linstedt, 121–26. Boston: Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-802044-9.00019-2>.
- IOTA. 2016. "Applying Data and Analytics in Tax Administration." *IOTA Good Practice Guide*. Intra-European Organisation of Tax Administrations (IOTA). <https://www.iota-tax.org/good-practice-guide-applying-data-and-analytics-tax-administrations>.
- Johannesson, Paul, dan Erik Perjons. 2014. *An Introduction to Design Science*. Heidelberg: Springer.
- Kautsarina, Kautsarina. 2018. "Perkembangan Riset Etnografi Di Era Siber: Tinjauan Metode Etnografi Pada Dark Web." *Masyarakat Telematika Dan Informasi: Jurnal Penelitian Teknologi Informasi Dan Komunikasi* 8: 145. <https://doi.org/10.17933/mti.v8i2.109>.
- Keller, Sallie Ann, Stephanie S. Shipp, Aaron D. Schroeder, dan Gizem Korkmaz. 2020. "Doing Data Science: A Framework and Case Study." *Harvard Data Science Review* 2 (1). <https://doi.org/10.1162/99608f92.2d83f7f5>.
- Kemenkeu. 2015. *Catatan Inspirasi Transformasi Inovasi Dan Terobosan Untuk Inspirasi Perubahan*. Balai Pustaka.
- Kling, Rob, Howard Rosenbaum, dan Steve Sawyer. 2005. *Understanding and Communicating Social Informatics: A Framework for Studying and Teaching the Human Contexts of Information and Communication Technologies*. Medford, New Jersey: Information Today.
- Kozinets, Robert. 2010. *Netnography: Doing Ethnographic Research Online*. SAGE Publications Ltd.
- Laudon, Kenneth C., dan Jane P. Laudon. 2012. *Management Information Systems: Managing the Digital Firm*. Pearson Prentice Hall.
- McKinney, Wes. 2018. *Python for Data Analysis Data Wrangling with Pandas, Numpy, and Ipython*. O'Reilly.
- Microsoft, dan PwC. 2017. "Digital Transformation of Tax Administration." Netherland: Microsoft and PricewaterhouseCoopers Belastingadviseurs N.V.
- . 2018. "The Data Intelligent Tax Administration - Meeting the Challenges of Big Tax Data and Analytics." Netherland: Microsoft and PricewaterhouseCoopers Belastingadviseurs N.V. <https://www.pwc.nl/nl/assets/documents/the-data-intelligent-tax-administration-whitepaper.pdf>.
- Neuman, W. Lawrence. 2014. *Social Research Methods: Qualitative and Quantitative Approaches*. 7th ed. Essex, England: Pearson Education Limited.
- OECD. 2011. "Summary Report." In *Conference on The Use of Advanced Analytics in Tax Administrations*. Dublin: OECD.
- Orlikowski, Wanda J., and Daniel Robey. 1991. "Information Technology and the Structuring of Organizations." *Information Systems Research* 2 (2): 143–69.
- Palanivel, K, dan J Manikandan. 2019. "Business Analytics Architecture Stack to Modern Business Organizations." *International Journal of Computer Sciences and Engineering* 7 (8): 275–87.
- Panoply. 2019. "Modern Data Management - Next Generation Data Tools Eliminate Data Maintenance." Panoply Ltd. 2019.
- Pierson, Lillian. 2015. *Data Science For Dummies*. John Wiley & Sons, Inc.
- Pijnenburg, Mark, Wojtek Kowalczyk, dan van der Hel-van Dijk Lissette. 2017. "A Roadmap for Analytics in Taxpayer Supervision." *The Electronic Journal of E-Government* 15 (1): 19–32.
- Pink, Sarah, Heather Horst, John Postill, Larissa Hjorth, Tania Lewis, dan Jo Tacchi. 2016. *Digital Ethnography: Principles and Practices*.

- Pistone, Pasquale, Jennifer Roeleveld, Johann Hattingh, João Félix Pinto, dan Nogueira Craig West. 2019. *Fundamentals of Taxation: An Introduction to Tax Policy, Tax Law and Tax Administration*. Amsterdam The Netherlands: IBFD.
- Press, Gil. 2016. "Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says." *Forbes*, 2016. <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#72dc87c76f63>.
- Rainer Jr., R. Kelly, dan Casey G. Cegielski. 2011. *Introduction to Information Systems*. New Delhi: John Wiley & Sons, Inc.
- Rattenbury, Tye, Joseph M. Hellerstein, Jeffrey Heer, Sean Kandel, dan Connor Carreras. 2017. *Principles of Data Wrangling Practical Techniques for Data Preparation*. Sebastopol, CA: O'Reilly Media, Inc.,.
- Regan, David, George Krasadakis, John Baxter, dan Yaz Yazicioglu. 2017. "The New Data Reality for Revenue Agencies." *Accenture*. https://www.accenture.com/_acnmedia/pdf-48/accenture-the-new-data-reality-for-revenue-agencies.pdf.
- Rollins, John B. 2015. "Foundational Methodology for Data Science." *Route 100 Somers, NY 10589: IBM Analytics*.
- Saddad, Emad, Ali El-Bastawissy, Hoda M. O. Mokhtar, dan Maryam Hazman. 2020. "Lake Data Warehouse Architecture for Big Data Solutions." *International Journal of Advanced Computer Science and Applications* 11 (8). <https://doi.org/10.14569/IJACSA.2020.0110854>.
- Schau, Hope, dan Mary Gilly. 2003. "We Are What We Post? Self-Presentation in Personal Web Space." *Journal of Consumer Research* 30. <https://doi.org/10.1086/378616>.
- Silva, Suen de Andrade E. 2013. "The Presentation of Self in Everyday Play: On Actor-Networks and Identity Performance in Facebook Games." *Master Thesis, Utrecht: Faculty of Humanities Theses - Utrecht University*.
- Sinnott, Daniel. 2016. "Advanced Analytics for Tax Administration: New Opportunities, Old Challenges." dalam *Data-Driven Tax Administration*, 6–11. Budapest: IOTA.
- Sugiyanto, Catur, dan Rijadh Winardi. 2018. "Strategi Penelitian Etnografi." dalam *Strategi Penelitian Bisnis*, editor Jogiyanto Hartono M. Yogyakarta: Andi Publisher.
- Tansey, David. 2019. "Tax Administration Information Systems Concept, Design, and Implementation." 36. *The Governance Brief*. ADB.
- Veit, Amanda. 2019. "Swimming Upstream: Leveraging Data and Analytics for Taxpayer Engagement – an Australian and International Perspective." *EJournal of Tax Research* 16 (3): 474–99.
- Verbeek, Peter-Paul, dan Pieter E. Vermaas. 2009. "Technological Artifacts." dalam *A Companion to the Philosophy of Technology*, editor Jan Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks, 165–71. Blackwell Publishing Ltd.
- White, Colin, dan Claudia Imhoff. 2010. "Advanced Analytics and Business Intelligence: Term Abuse?" 2010. <http://www.b-eye-network.com/view/13797>.