

PENAMBAHBAIKAN PROSES INTEGRASI BERDASARKAN DIMENSI KETEPATAN DATA DI SEKTOR AWAM

Sharifah Norhazira binti Said Hamid, Kamsuriah binti Ahmad

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Malaysia.

p106806@siswa.ukm.edu.my, kamsuriah@ukm.edu.my

ABSTRAK

Pelbagai sistem aplikasi telah dibangunkan untuk menyokong keperluan bisnes sesebuah organisasi. Namun, kebanyakan sistem tersebut beroperasi secara silo menyebabkan terjadinya pertindihan fungsi. Bagi mengatasinya, integrasi data dilaksanakan bagi membolehkan sistem-sistem aplikasi yang berasingan digabungkan untuk perkhidmatan yang lebih bersepadu serta dilengkapi dengan maklumat yang lebih komprehensif. Cabaran utama kejayaan projek integrasi adalah memastikan data yang diperolehi melalui integrasi merupakan data yang berkualiti. Permasalahan data melibatkan data tidak tepat, duplikasi data, data tidak konsisten, kesalahan ejaan serta data tidak lengkap seringkali terjadi dan menjejaskan pelaksanaan integrasi. Ini menyebabkan lambakan data di pangkalan data, skop kerja bertambah, kos pengoperasian meningkat, kelewatan terhadap projek dan seterusnya analisis data tidak tepat menjejaskan keputusan strategik yang diambil oleh organisasi. Tujuan kajian adalah untuk menambah baik proses integrasi data dengan mengenal pasti ciri-ciri data yang berkualiti serta pengelasannya berdasarkan enam dimensi Kualiti Data. Dimensi Kualiti Data yang paling kritikal mempengaruhi kejayaan projek integrasi sektor awam di Malaysia dikenal pasti serta cadangan langkah-langkah tambah baik ke atasnya bagi meningkatkan kualiti data. Skop bagi kajian ini merangkumi tambah baik proses integrasi data untuk meningkatkan kualiti data di sektor awam. Kaedah terbaik untuk mendapatkan maklumat yang tepat adalah melalui temu bual. Oleh itu, kajian dilaksanakan secara kualitatif menerusi temu bual ahli pasukan seramai sembilan orang yang terlibat dalam pelaksanaan integrasi di sektor awam bagi mengesahkan kajian. Pengesahan kajian dilaksanakan dengan menemu bual dua orang pegawai berpengalaman di dalam projek integrasi. Kedua-dua sesi temu bual dilaksanakan secara separa berstruktur dan menggunakan teknik analisis kandungan bagi mengenal pasti maklumat penting dan dikategorikan mengikut keperluan kajian. Matlamat kajian dicapai dengan mengenal pasti dimensi utama Kualiti Data dan tambah baik proses terhadapnya dijangka akan membantu agensi sektor awam menangani isu kualiti data dalam integrasi. Hasilnya, pelaksanaan integrasi bagi menyokong perkongsian data antara agensi sektor awam dipertingkatkan dan aspirasi kerajaan ke arah kerajaan Malaysia berpacuan data dapat dicapai.

Kata kunci: integrasi data, isu integrasi, faktor kejayaan integrasi, dimensi kualiti data, ketepatan data

1. PENGENALAN

Penggunaan teknologi digital dalam membuat pesanan elektronik, urusan perbankan, dan akses kepada maklumat merupakan pilihan utama dalam menjalani kehidupan seharian. Pada tahun 2020 sahaja, teknologi digital telah berkembang pesat ekoran pandemik Covid-19 yang mewujudkan perspektif baharu dalam pelbagai sektor termasuk sektor perkhidmatan, ekonomi, pendidikan dan

sebagainya. Bagi mengelak perhubungan secara fizikal, kebanyakan sektor beralih kepada kaedah maya atau dalam talian.

Sejajar dengan perkembangan tersebut, integrasi data adalah sangat penting bagi organisasi yang memiliki sumber data yang besar bagi merangka strategi ke arah perkembangan sesuatu organisasi. Integrasi secara amnya menghubungkan sistem dan maklumat secara efisien (Kumar et al. 2018). Merujuk kepada Gray (1999) dan Mendoza et al. (2006), integrasi merupakan gabungan sistem maklumat dan pangkalan data bagi meningkatkan aliran proses kerja serta menawarkan perkhidmatan pelanggan yang lebih baik.

Menyedari kepentingan itu, organisasi abad ke-21 perlu mengurus data sebagai aset (Ladley 2020). Berdasarkan takrifan Buku Pengurusan Data Pengetahuan DAMA International (2017), Tadbir Urus Data merupakan satu bidang yang diwujudkan untuk merancang, memantau, menguatkuasa dan memutuskan sebarang keputusan melibatkan pengurusan aset data. Dengan adanya Tadbir Urus Data, maklumat yang tepat dapat disediakan kepada pihak berkepentingan organisasi samada dalaman atau luar organisasi (Kim & Cho 2018) dan membolehkan perkongsian data yang komprehensif iaitu ia boleh dirujuk setiap masa dan sentiasa dikemaskini (Liu 2020). Terdapat 10 bidang dalam Tadbir Urus Data dan kualiti data merupakan antara aspek yang sentiasa dibincangkan dalam Tadbir Urus Data kerana ia merupakan data yang menggambarkan ketepatan, lengkap, relevan dan konsisten (Cheong & Chang 2007; Fu & Easton 2017; Ladley 2020). Justeru itu, kualiti data adalah amat penting dalam memastikan kejayaan sesuatu projek integrasi.

Integrasi merupakan salah satu program ICT atau strategi bagi menggabungkan semua data keseluruhan agensi sektor awam (Abdullah et al. 2018). Sektor awam berusaha meningkatkan perkhidmatan dengan mengadaptasi teknologi maklumat bagi memastikan kelancaran aliran proses kerja dan tadbir urus yang lebih telus (Fernandez et al. 2017; Kumar et al. 2018). Kerajaan Malaysia telah mewujudkan hab perkongsian data Kerajaan Malaysia, Malaysian Government Central Data Exchange (MyGDX) sebagai peneraju utama sektor awam dalam usaha integrasi data.

Pelaksanaan integrasi ini bukan mudah kerana ianya kompleks dan mencabar. Cabaran melaksanakannya bukan sahaja dihadapi oleh sektor swasta malah juga sektor awam menghadapi isu yang sama (Hassan et al. 2020). Menurut pelaporan Pelan Strategik Pendigitalan Sektor Awam 2021-2025, penggunaan platform MyGDX untuk perkongsian data masih rendah. Oleh itu, Kerajaan Malaysia perlu merangka strategi dalam memastikan pelaksanaannya berjalan lancar dan projek berjaya memberi manfaat kepada rakyat.

Antara persoalan kajian yang dikenal pasti adalah:

RQ1: Apakah isu kualiti data semasa proses integrasi data di sektor awam?

RQ2: Apakah langkah yang perlu diambil untuk menambah baik proses integrasi data bagi meningkatkan kualiti data di sektor awam?

Oleh itu, objektif kajian berkaitan dengannya adalah mengenal pasti isu kualiti data semasa proses integrasi data di sektor awam serta mencadangkan penambahbaikan proses integrasi data untuk meningkatkan kualiti data di sektor awam. Skop kajian ini adalah tertumpu kepada sektor awam meliputi organisasi yang disasarkan menggunakan MyGDX dan memfokuskan penambahbaikan kualiti data dalam integrasi sektor awam.

2. KAJIAN KESUSASTERAAN

A. Permasalahan Isu Kualiti Data

Terdapat pelbagai cabaran yang dihadapi dalam merealisasikan integrasi sektor awam. Berdasarkan kajian Sulistyio et al. (2020) masalah kualiti data mengakibatkan penurunan secara 20% dalam produktiviti pekerja dan menjadi punca 40% inisiatif organisasi gagal mencapai matlamat. Ini menggambarkan kelemahan dalam kualiti data menjejaskan pengurusan operasi, taktikal dan perancangan strategik sesuatu organisasi. Jadual 1 menyenaraikan dapatan dari kajian lepas berkaitan permasalahan data dari pelbagai sudut dimensi kualiti data.

Jadual 1 Isu elemen dimensi kualiti data tahun 2017-2021

No	Sumber	Dimensi Kualiti Data			
		Ketepatan Data	Konsisten	Masa	Lengkap
1.	Xiao et al. (2017)				√
2.	Kim & Cho (2017)		√	√	
3.	DAMA International (2017)		√		
4.	Kusumasari & Fitria (2017)	√	√		
5.	Fu & Easton (2017)	√			
6.	Pinto et al. (2017)	√		√	
7.	Dagade & Mali (2017)	√			
8.	Ahmed (2018)	√	√		
9.	Kaur et al. (2018)	√			
10.	Calabrese (2018)	√			
11.	Muthee et al. (2018)				√
12.	Juneja & Das (2019)		√		

bersambung...

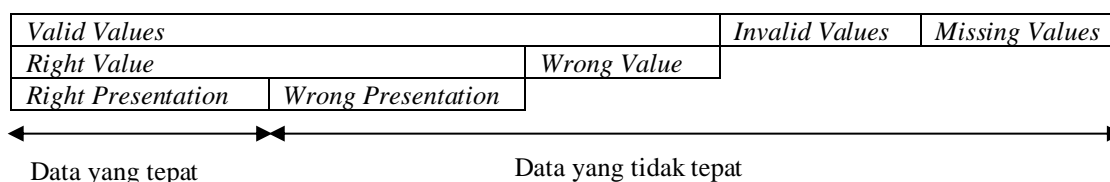
...sambungan

No	Sumber	Dimensi Kualiti Data			
		Ketepatan Data	Konsisten	Masa	Lengkap
13.	Dong et al. (2019)	√	√		√
14.	Ridzuan & Wan Zainon (2019)	√	√		
15.	Lucas (2019)				
16.	Sulistyo et al. (2020)				√
17.	Liu (2020)	√			√
18.	Juddoo & George (2020)	√			√
19.	Njuguna et al. (2020)	√	√		
20.	Campbell et al. (2020)	√			√
21.	Wang & Wang (2020)				
22.	Ibrahim et al. (2021)	√		√	√
23.	McGilvray (2021)	√			
Jumlah		15	8	3	8

Didapati sebanyak 15 kertas kerja hasil penyelidik menyatakan dimensi ketepatan memberi pengaruh besar keatas kelemahan kualiti data manakala hasil kajian dari lapan kertas kerja menyatakan data yang tidak konsisten dan data tidak lengkap memberi impak besar kepada kualiti data. Dimensi masa dikaitkan oleh tiga kertas kerja penyelidik serta mana-mana penyelidik tidak membincangkan elemen dimensi sah dan integriti sebagai punca utama kelemahan kualiti data. Merujuk kepada kajian-kajian tersebut jelas tidak mengaitkan dimensi integriti serta kesahihan data sebagai punca utama kelemahan kualiti data. Oleh itu, kesimpulan awal dari kajian sedia ada menunjukkan data yang dibekalkan sekitar lima tahun kebelakangan ini menepati kualiti data dari segi integriti data dan kesahihan data.

B. Ketepatan Data

Ketepatan data merujuk kepada data yang bebas daripada ralat yang boleh digunakan sebagai sumber maklumat yang dipercayai. Di dalam pengurusan data, ketepatan data merupakan satu elemen kritikal yang sering diperlukan di dalam sistem maklumat.



Rajah 1 Pecahan Data

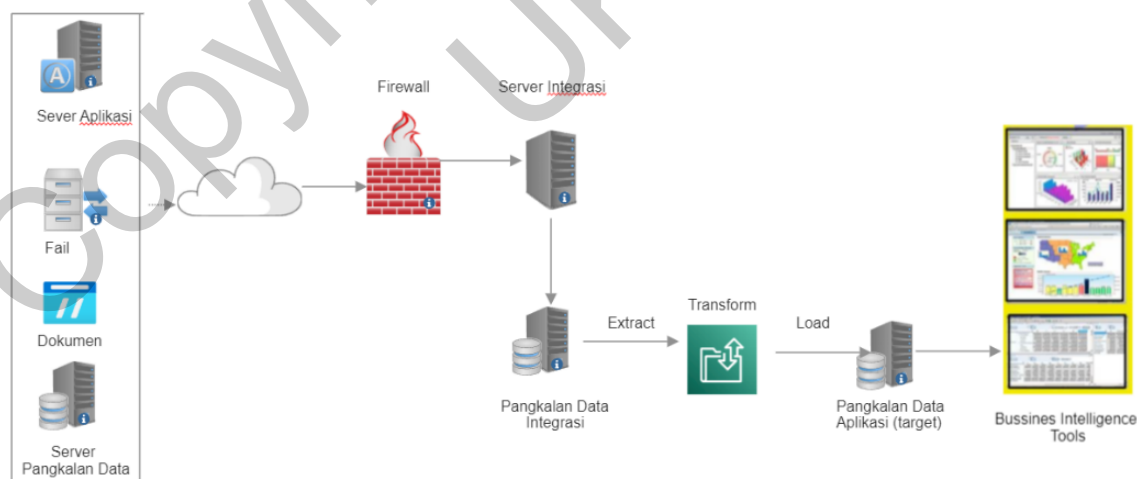
Sumber: Olson (2003)

Berdasarkan Rajah 1, data yang tepat merujuk kepada data yang sah yang mana ia merupakan nilai yang betul dan dipamerkan dalam format yang ditetapkan. Manakala, jika data yang dipaparkan mempunyai ciri-ciri selain daripada itu, ia merupakan data yang tidak tepat.

Di dalam pengurusan data, ketepatan data merupakan satu elemen kritikal yang sering diperlukan di dalam sistem maklumat. Namun, kebanyakan sistem kini masih lagi mempunyai data yang tidak tepat yang terhasil dari input data ke medan yang salah, kesalahan dalam penukaran format pengukuran (Abdullah et al. 2015), kesalahan ejaan, singkatan perkataan yang tidak memberi maksud serta data yang meragukan. Selain itu, terdapat faktor lain yang mempengaruhi ketepatan data, salah satu darinya adalah kesilapan teknikal dari sistem legasi seperti pengaturcaraan yang salah dan kesilapan yang terjadi semasa kemasukan data oleh pekerja kurang mahir atau pelanggan.

Lazimnya, isu data tidak tepat telah dialami sekian lama di sesuatu organisasi namun ia tidak dibincangkan kerana ia hanya disimpan sebagai maklumat sokongan. Berdasarkan penerangan tersebut, tidak dapat disangkal masih banyak sistem maklumat yang mengandungi data yang tidak tepat. Namun, kebanyakan organisasi tidak menyedarinya kerana kurang kefahaman asas tentang konsep kualiti data dan masalah ini hanya disedari apabila data berkenaan digunakan dalam pelaporan atau sebagai maklumat untuk menyokong pihak pengurusan bagi membuat keputusan di dalam organisasi.

C. Pelaksanaan Proses Integrasi Sedia Ada



Rajah 2 Arkitektur Integrasi

Rajah 2 menerangkan arkitektur yang seringkali digunakan oleh agensi kerajaan dalam membangunkan integrasi. Proses integrasi bermula dengan proses pengumpulan sumber data (*source*) seperti pangkalan data, *flat file*, aplikasi dan dokumen iaitu pada proses ini dilakukan integrasi data dari pelbagai sumber. Data yang dihantar akan melalui *firewall* organisasi serta ditempatkan di server

integrasi. Antara kaedah integrasi yang sentiasa digunakan adalah sama ada kaedah *Web Service*, *Secure File Transfer Protocol* (SFTP) atau sambungan terus ke pangkalan data (*database direct access*) samada *database view*, *database replication* atau *database link*.

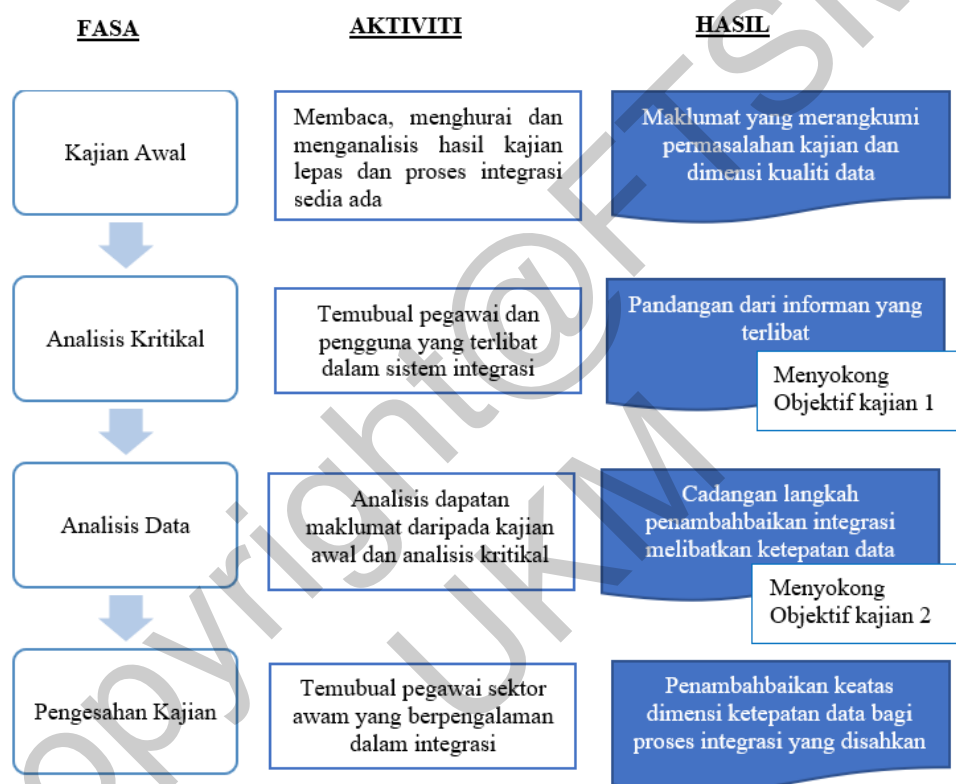
Data yang dikumpulkan dari pelbagai sumber mempunyai format dan jenis data yang berbeza. Oleh itu, kesemua data tersebut diproses menggunakan proses *Extract, Transform & Load (ETL)* sebelum dipam masuk ke pangkalan data sistem. Kekerapan proses ETL samada secara harian, mingguan atau bulanan adalah bergantung kepada persetujuan kedua-dua pihak antara pemberi data dan penerima data. Proses pengekstrakan data (*extract*) merupakan proses mengimport data dari punca data serta menyalin data tersebut ke pangkalan data berlainan (Sreemathy et al. 2021). Pengekstrakan data boleh dilaksanakan dengan beberapa pendekatan bergantung kepada sistem aplikasi itu sendiri samada data tersebut dihantar secara pukal (*bulk*) atau data yang berubah sahaja akan dihantar samada pertambahan atau perubahan data (*incremental update*) (Biswas et al. 2020).

Data yang diekstrak dari sumber data merupakan data mentah dan kebanyakannya tidak boleh terus digunakan sejurus selepas ia dimuat naik ke pangkalan data sasaran (Lokaadinugroho et al. 2021). Oleh itu, ia perlu dibersihkan, dipetakan dan diubah terlebih dahulu. Proses transformasi (*transform*) melaksanakan pembersihan data iaitu membuang mana-mana nilai data yang tidak sepadan, ralat atau sebarang anomali dan menukar format data yang berpadanan dengan struktur skema pangkalan data baharu. Selain itu, proses pengagregatan dan pemetaan data juga berlaku disini. Transformasi data merupakan langkah utama di ETL di mana data diubah suai mengikut struktur pangkalan data sasaran sebelum ia dimuat naik samada disatukan, dipecahkan kepada beberapa bahagian, ditafsirkan berdasarkan kod dan lain-lain proses (Theodorou et al. 2016; Souibgui et al. 2019 ; Galici et al. 2020).

Langkah terakhir dalam ETL adalah proses muat naik (*load*) data iaitu memasukkan data ke pangkalan data sasaran (*target*) sistem tersebut. Memandangkan ia melibatkan pangkalan data sasaran, proses muat naik perlu dilaksanakan secara teliti (Biswas et al. 2020). Kebiasaannya, proses muat naik terutamanya muat naik secara pukal dibuat semasa pengguna jarang menggunakan sistem pada masa tersebut, contohnya pada waktu tengah malam atau awal pagi. Setelah data telah dimuat naik, ia akan disemak terlebih dahulu dan hasilnya, data ini akan digunakan sebagai informasi, laporan, pertanyaan dan perlombongan data (*data mining*) untuk membuat pertimbangan dalam keputusan.

3. METODOLOGI KAJIAN

Kajian ini dilaksanakan secara kualitatif menerusi empat fasa iaitu kajian awal, analisis kritikal, analisis data dan pengesahan kajian. Metodologi kualitatif dipilih bagi mendapat kefahaman mengenai realiti sebenar tentang pelaksanaan integrasi di sektor awam. Selain itu, kaedah ini merupakan kaedah yang paling berkesan dalam mendapatkan input yang tepat terhadap kajian yang sedang dilaksanakan (Almeida et al. 2017). Aktiviti terperinci bagi setiap fasa kajian dan hasil akhir bagi setiap fasa diterangkan seperti dalam Rajah 3.



Rajah 3 Fasa Kajian

A. Fasa 1 : Kajian Awal

Kajian awal merupakan peringkat utama dalam mengenal pasti permasalahan yang menjurus kepada kajian yang dilaksanakan. Pelaksanaan kajian awal adalah dengan mendapatkan maklumat daripada penyelidikan terdahulu untuk dijadikan asas kepada kajian yang dijalankan. Aktiviti yang terlibat meliputi pembacaan, penghuraian dan penganalisan proses integrasi sedia ada serta hasil kajian penyelidikan lima tahun kebelakangan bertujuan bagi mengenal pasti jurang dalam kajian lepas dan cadangan-cadangan yang telah dikemukakan oleh penyelidik bagi menyokong kajian yang dilakukan.

Dapatan daripada fasa kajian awal iaitu dimensi ketepatan data dikenal pasti sebagai elemen utama kualiti data yang perlu ditambahbaik dalam pelaksanaan projek integrasi.

B. Fasa 2 : Analisis Kritikal

Hasil dari Kajian Awal mendapati ketepatan data adalah dimensi paling kritikal dalam menentukan data berkualiti. Fasa Analisis Kritikal dilaksanakan bagi menyokong hasil daripada Kajian Awal. Fasa ini akan mengesahkan samada dapatan tersebut seiring dengan fenomena yang sering berlaku di sektor awam Malaysia serta mengenal pasti amalan biasa yang sering dilakukan oleh agensi yang menjadi punca kepada masalah ketidaktepatan data. Pengesahan diperoleh menerusi kaedah temubual daripada informan terdiri daripada sembilan individu yang merupakan penjawat awam terdiri dari pegawai kerajaan serta pasukan pelaksana iaitu ahli pejabat pengurusan projek, ketua pasukan, ahli pasukan dan pengguna sistem dalam projek integrasi di sektor awam.

Soalan temubual dirangka dan soalan tersebut dikemukakan untuk mendapat pengesahan sebelum aktiviti temubual dijalankan. Pengesahan soalan temubual adalah diperlukan bagi memastikan soalan kajian yang diajukan kepada informan adalah munasabah serta berkaitan dengan perkembangan projek integrasi semasa. Temubual dibuat ke atas beberapa individu yang terdiri daripada kumpulan pegawai (gred F44 sehingga F52) dan pelaksana (N32) terlibat secara langsung dalam projek integrasi di sektor awam merangkumi ketua pasukan, ahli pasukan, pembangun sistem, *Projek Management Office (PMO)* serta pengguna sistem. Informan yang dipilih adalah bekerja di sekitar Lembah Kelang iaitu Wilayah Persekutuan Putrajaya dan Cyberjaya memandangkan sistem aplikasi integrasi dibangunkan oleh agensi pusat iaitu lokasinya bertempat di Putrajaya dan Cyberjaya. Agensi yang dipilih adalah Kementerian Kewangan (MOF), Unit Pemodenan Tadbiran dan Perancangan Pengurusan Malaysia (MAMPU), Kementerian Kesihatan (MOH) dan Bahagian Hal Ehwal Undang-Undang (BHEUU JPM). Pelantikan informan untuk ditemubual adalah daripada pelbagai kategori ahli pasukan seperti di Jadual 2.

Jadual 2 Pemilihan Profil Informan

Jawatan Dalam Pasukan Projek Integrasi	Kuantiti	Pengalaman dalam Projek Integrasi	Bilangan Projek Yang Terlibat
A, Pasukan Projek			
Pengarah Projek	1	> 6 tahun	>7
Ketua Pasukan	2	> 10 tahun	>3
Ahli Pasukan	2	> 7 tahun	>2
Pembangun Sistem	2	> 9 tahun	>3
PMO	1	> 5 tahun	>1
B. Pengguna			

bersambung...

...sambungan

Pengguna Sistem	1	> 5 tahun	>3
Jumlah	9		

Sesi temu bual secara separa berstruktur dilaksanakan dalam tempoh 14-21 hari bagi mendapatkan pandangan daripada informan berkaitan kualiti data yang diperolehi dari agensi-agensi dan pada masa yang sama mengesahkan dapatan awal semasa fasa Kajian Awal. Dapatan dari fasa ini mengesahkan pernyataan masalah bahawa dimensi ketepatan data merupakan dimensi paling kritikal dalam menentukan kejayaan sesuatu projek integrasi. Menerusi temubual, informan mendedahkan amalan biasa yang sentiasa dilakukan semasa mengendalikan sistem yang mana terdapat beberapa amalan tidak efisien menyumbang kepada ketidaktepatan data.

C. Fasa 3 : Analisis Data

Fasa Analisis Data melibatkan aktiviti menganalisis segala input daripada fasa Kajian Awal dan Analisis Kritikal berkaitan elemen kawalan kualiti data. Melalui aktiviti ini, potensi penyelesaian masalah dicadangkan sebagai membantu menambah baik integrasi di sektor awam. Hasil temubual informan akan dianalisis menggunakan teknik analisis kandungan iaitu ia merupakan kaedah menganalisis data kualitatif yang terdiri dari data teks (Graneheim et al. 2017). Menerusi fasa ini, faktor terpenting dalam kualiti data yang menyumbang kepada kejayaan integrasi dikenal pasti. Dapatan ini kemudiannya dianalisis dan cadangan mekanisma penambahbaikan integrasi dihasilkan.

D. Fasa 4 : Pengesahan Kajian

Pengesahan kajian merupakan fasa terakhir dalam kajian ini dan penting dilaksanakan bagi mendapatkan pandangan dan persetujuan pegawai terlibat berkenaan cadangan penambahbaikan proses integrasi data.

Fasa ini melibatkan dua orang pegawai yang dikenal pasti dari segi pengalamannya dalam projek integrasi. Pandangan lebih daripada seorang pegawai dan pemilihan pegawai yang tepat diperlukan bagi mendapatkan hasil kajian yang berkualiti. Kriteria pemilihan pegawai adalah pengalaman melebihi daripada 7 tahun dalam projek integrasi sektor awam dan pernah terlibat beberapa pelaksanaan projek integrasi. Kaedah bagi aktiviti dalam fasa ini dilaksanakan menerusi sesi temu bual secara separa berstruktur. Hasil analisis data yang diperolehi daripada Fasa Ketiga akan dimaklumkan terlebih dahulu kepada pegawai untuk penelitian oleh mereka sebelum soalan dikemukakan. Walau bagaimanapun, pandangan tambahan daripada pegawai tetap diterima bagi penambahbaikan hasil kajian.

Kajian ini penting kerana hasil dan cadangan tersebut mampu membantu organisasi mengesan kelemahan data serta cara mengatasinya. Ia seterusnya boleh dijadikan panduan pelaksanaan projek integrasi sektor awam pada masa akan datang.

4. ANALISIS DATA

A. Isu dalam Pelaksanaan Integrasi

Jadual 3 merumuskan dapatan hasil temuramah pandangan informan berkaitan dimensi paling penting mempengaruhi kualiti data semasa pelaksanaan integrasi di sektor awam. Berdasarkan temuramah, terdapat 15 isu telah dibangkitkan oleh informan berkaitan permasalahan integrasi yang menyumbang kepada kelemahan kualiti data di mana 13 isu menjadi punca ketidaktepatan data dan dua isu memberi impak dari segi data tidak konsisten.

Jadual 3 Isu dalam Integrasi Sistem Aplikasi.

Kategori	Isu	Bilangan Informan
Ketepatan Data	i. Kesilapan kemasukan data oleh pengguna	7
	ii. Kesilapan semasa pepadanan data	5
	iii. Kesilapan semasa perubahan data	4
	iv. Kesalahan ejaan	6
	v. Salah tafsir data elemen di IDD	3
	vi. Jenis data (<i>data type</i>) yang berbeza antara pangkalan data agensi pembekal dan pangkalan data agensi pengguna	3
	vii. Input data yang diterima tidak menepati standard IDD yang dipersetujui	3
	viii. Data bukan dari sumber data yang sebenar	2
	ix. Data terkini yang telah dikemaskini tidak dihantar oleh sumber data	3
	x. Perubahan tanpa disedari pada data di fail data yang dihantar oleh sumber data	2
	xi. Perubahan pada <i>master data</i> tidak dimaklumkan oleh agensi pembekal	2
	xii. Perubahan pada struktur pangkalan data sumber	4
	xiii. Penggunaan teknologi atau platform integrasi yang berbeza antara agensi pembekal data dan agensi pengguna data menjejaskan proses integrasi	2
Konsisten Data	xvi. Duplikasi Data	2
	xvii. Data tidak dinormalisasi	2

Berdasarkan hasil temubual dengan informan, dimensi ketepatan data telah disahkan merupakan faktor paling kritikal melibatkan kualiti data dalam pelaksanaan integrasi di sektor awal. Penerangan isu permasalahan ketepatan data diterangkan di perenggan berikutnya.

a. Kesilapan Kemasukan Data oleh Pengguna

Isu paling kerap terjadi yang memberi kesan kepada ketidaktepatan data adalah kesilapan kemasukan data ke dalam sistem oleh pengguna. Seramai tujuh informan menyatakan isu tersebut kerap terjadi di agensi masing-masing. Kesilapan semasa pemilihan kod yang merupakan kod rujukan terhadap sesuatu item, kesilapan menaip IC atau nilai melibatkan kos dan kewangan akan memberi implikasi kepada sistem. Selain itu, kebiasaannya kesilapan ini terjadi kerana terdapat pengguna yang salah faham terhadap kefungsiian di sistem tersebut seterusnya mendorong melakukan ralat tersebut.

b. Kesalahan Ejaan

Selain itu, kesalahan ejaan merupakan amalan yang sering terjadi yang menyumbang kepada masalah data tidak tepat. Seramai enam informan bersetuju menyatakan permasalahan ini kerap terjadi di agensi masing-masing.

c. Kesilapan Pemadanan Data

Isu ketiga tertinggi yang dinyatakan berkaitan ketepatan data adalah kesilapan semasa pemadanan data di proses ETL. Pemadanan yang salah semasa proses klasifikasi data menyebabkan data dipaparkan tidak tepat di agensi pengguna data. Kurang kefahaman tentang *bussines rules* data tersebut mendatangkan implikasi buruk kepada pelaporan, mengelirukan pengguna serta mengambil masa untuk melaksanakan pembersihan semula data.

d. Kesilapan Perubahan Data

Kesilapan semasa perubahan data juga memberi impak kepada kualiti data. Contoh perubahan data kepada data terbitan (data baharu yang terhasil dari aktiviti pemprosesan dan analisis data) adalah penukaran titik perpuluhan, proses *decode* atau proses melibatkan pengiraan lebih kompleks seperti susut nilai aset, nilai perolehan keseluruhan dan cukai yang perlu dikenakan untuk setiap barangan.

e. Perubahan pada Struktur Pangkalan Data Sumber

Terdapat situasi iaitu perubahan pada struktur pangkalan data tidak dimaklumkan kembali kepada agensi pengguna data. Contohnya, perubahan pada struktur kod iaitu kod terdahulu mempunyai panjang aksara 12 telah diubah kepada 15 aksara bagi menampung pertambahan kod item yang meningkat atau penukaran struktur nombor kod pekerja yang berbeza format berbanding sebelumnya. Situasi tersebut boleh menyebabkan ketepatan data pada sistem terjejas ia merupakan kod yang

dirujuk untuk mendapatkan informasi. Sebarang perubahan yang terjadi kepada kod boleh menyebabkan maklumat berkaitan dengannya tidak dapat dipaparkan atau paparan maklumat yang tidak tepat.

f. Salah tafsir Data Elemen di IDD

Kesalahan dalam mentafsir data elemen semasa penetapan IDD berlaku apabila agensi pengguna tidak mengkaji dengan lebih mendalam tentang data input yang bakal diterima oleh mereka semasa integrasi. Ini menyebabkan medan data tidak dipadankan dengan betul seterusnya menyebabkan data yang dipaparkan oleh sistem aplikasi agensi pengguna tidak tepat.

g. Jenis Data (*Data Type*) Yang Berbeza

Tetapan jenis data (*data type*) yang berbeza antara pangkalan data agensi pembekal dan pangkalan data agensi pengguna turut memberi kesan kepada ketetapan data. Format jenis data yang berbeza contohnya memasukkan tarikh ke dalam pangkalan data agensi pengguna tanpa menyemak terlebih dahulu format standard samada ddmmYYYY atau mmddYYYY di peringkat transformasi data menyebabkan paparan di sistem tidak tepat, mengelirukan pengguna atau paling kritikal ia tidak dapat dimasukkan kepada pangkalan data sama sekali kerana tidak menepati format standard yang digunapakai oleh sistem berkenaan.

h. Input Data Tidak Menepati IDD

Terdapat input data yang diterima tidak menepati standard IDD yang dipersetujui kerana agensi pembekal data menghantar data input yang berbeza berbanding yang telah dipersetujui di IDD. Terdapat juga input yang dihantar tidak menepati panjang aksara yang dipersetujui oleh IDD.

i. Data Terkini Tidak Dihantar Oleh Sumber Data

Data terkini tidak dihantar oleh agensi pembekal kepada agensi pengguna data kerana data dihantar sekali sahaja semasa data diwujudkan dan aplikasi integrasi di agensi pembekal tidak mengesan sebarang perubahan data (samada kemaskini atau padam) yang telah diperbaiki oleh pengguna dan seterusnya tidak menghantar data tersebut ke agensi pengguna data. Ini memberi implikasi data di agensi pengguna tidak dikemaskini kepada data terkini seterusnya maklumat yang dipaparkan kepada pengguna tidak tepat.

j. Data bukan dari Sumber Data Sebenar

Terdapat agensi yang keberatan untuk berkongsi maklumat mereka dengan agensi berlainan disebabkan polisi atau keselamatan. Memandangkan maklumat yang ingin diperolehi itu sukar diperolehi dari agensi sumber data, agensi pengguna data membuat keputusan untuk mendapatkan data berkenaan dari agensi sekunder lain yang mempunyai maklumat tersebut. Pendekatan ini sangat berisiko kerana maklumat terkini bagi data tersebut hanya disimpan di agensi sumber data dan tidak dikongsikan dengan agensi lain. Ini seterusnya menyumbang kepada ketidaktepatan data dalam aktiviti integrasi.

k. Perubahan Data Pada Fail Data Tanpa Disedari

Isu ketepatan data juga melibatkan perubahan data pada fail data tanpa disedari oleh agensi. Hal ini kerap terjadi melibatkan data yang berbentuk nombor yang bermula dengan aksara 0. Sesetengah sistem mengabaikan aksara 0 pada permulaan data contohnya no telefon '014-2445683' dan data no telefon dimasukkan ke pangkalan data ialah '14-2445683' yang menggambarkan data direkodkan secara tidak tepat.

Selain itu terdapat juga situasi yang mana nombor berbentuk desimal diringkaskan tanpa disedari kepada bentuk eksponen dan seterusnya disimpan ke pangkalan data. Situasi ini mendorong kemasukan data yang tidak tepat seterusnya menyebabkan kekeliruan kepada pengguna sistem.

l. Perubahan pada *Master Data*

Data integrasi terbahagi kepada *master data* dan data transaksi. Terdapat situasi agensi pembekal data tidak menyenaraikan keseluruhan *master data* yang perlu diterima oleh agensi pengguna data semasa integrasi atau tidak memaklumkan sebarang perubahan yang berlaku di *master data* berkenaan. Hal ini seterusnya menyebabkan terdapat beberapa tambahan nilai pada *master data* tersebut yang boleh menyebabkan data tidak dapat diproses sebaiknya dan ia dipaparkan tidak tepat atau memberi implikasi yang lebih teruk iaitu ralat pada sistem.

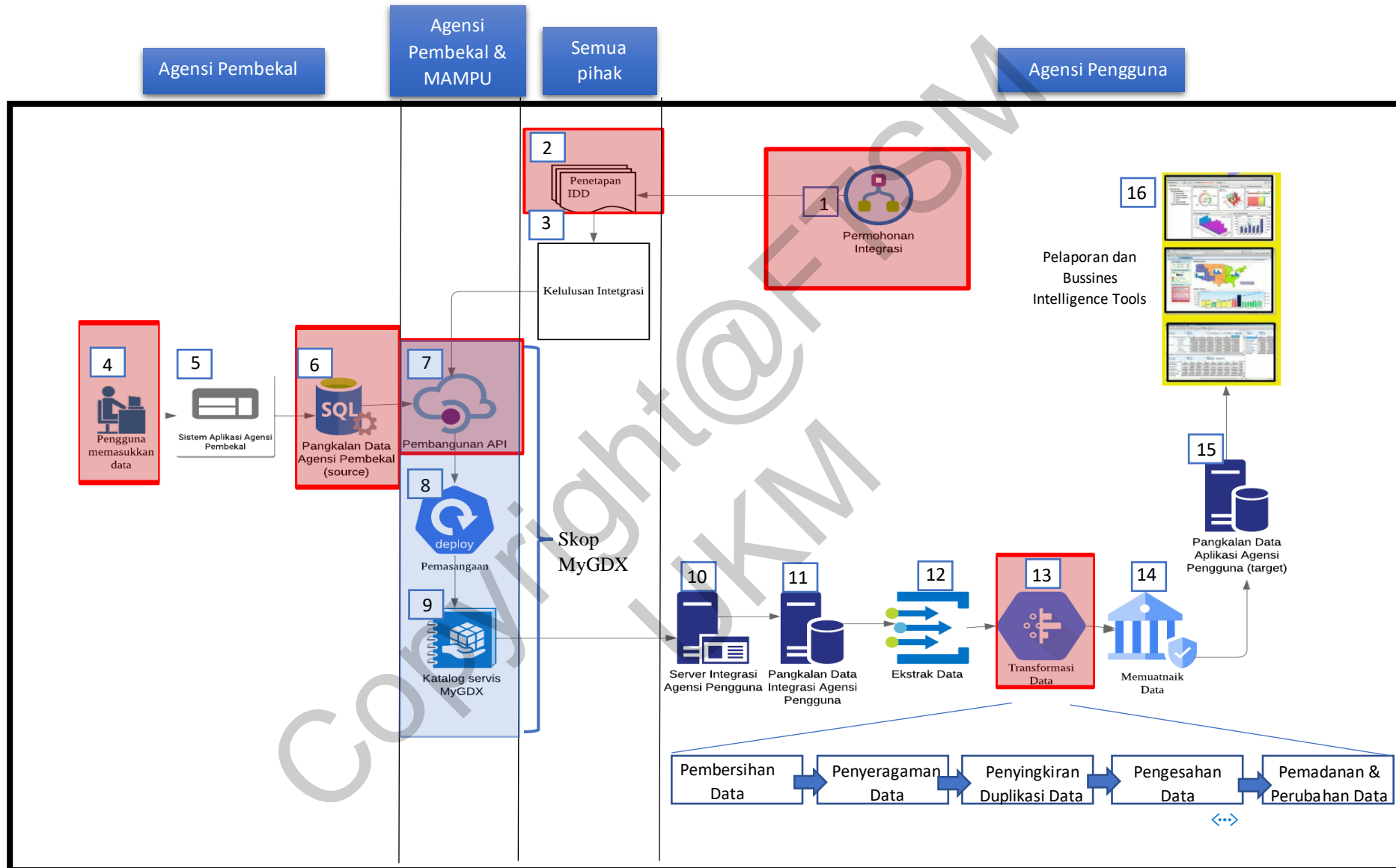
m. Penggunaan Teknologi Atau Platform Integrasi Yang Berbeza

Permasalahan terakhir mengaitkan isu ketepatan data adalah penggunaan teknologi atau platform integrasi yang berbeza antara agensi pembekal dan agensi pengguna terutamanya bagi keadah integrasi SFTP mempengaruhi ketepatan data. Ini kerana, fail data yang dihantar ke server integrasi agensi pengguna berkemungkinan diusik dan menyebabkan proses pengekstrakan data terjejas.

Berdasarkan permasalahan tersebut, ia dipopulasikan dan dipetakan dengan langkah-langkah aktiviti integrasi seperti Jadual 4 dan ia turut diilustrasikan seperti Rajah 4.

Jadual 4 Pemetaan Antara Isu Ketepatan Data dan Proses Integrasi

No	Proses Integrasi	Isu	Peringkat
1	Permohonan Integrasi	i. Data bukan dari sumber data yang sebenar	Langkah 1
2	Penetapan IDD	i. Salah tafsir data elemen di IDD	Langkah 2
3	Pengguna memasukkan data	i. Kesilapan kemasukan data oleh pengguna ii. Kesalahan ejaan	Langkah 4
4	Pangkalan Data Agensi Pembekal	i. Input data yang diterima tidak menepati standard IDD yang dipersetujui ii. Data terkini yang telah dikemaskini tidak dihantar oleh sumber data iii. Perubahan pada <i>master data</i> tidak dimaklumkan oleh agensi pembekal iv. Perubahan pada struktur pangkalan data sumber	Langkah 6
5	Pembangunan API	i. Perubahan tanpa disedari pada data di fail data yang dihantar oleh sumber data ii. Penggunaan teknologi atau platform integrasi yang berbeza antara agensi pembekal data dan agensi pengguna data menjejaskan proses integrasi	Langkah 7
6	Transformasi Data	i. Kesilapan semasa pepadanan data ii. Kesilapan semasa perubahan data iii. Jenis data (<i>data type</i>) yang berbeza antara pangkalan data agensi pembekal dan pangkalan data agensi pengguna	Langkah 13



Rajah Error! No text of specified style in document. Pemetaan Punca Ketepatan data berdasarkan Carta Alir Integrasi Data

Rumusan dari hasil penemuan dalam fasa analisis kritikal didapati ketidakketepatan data kerap terjadi semasa peringkat pangkalan data agensi pembekal iaitu merangkumi empat isu, diikuti peringkat transformasi data iaitu tiga isu, diikuti peringkat kemasukan data oleh pengguna dan pembangunan API iaitu masing-masing terdapat dua isu serta peringkat lain-lain terdapat masing-masing satu.

B. Cadangan Mekanisme Penambahbaikan

Hasil dapatan kajian di peringkat pertama akan dipopulasikan mengikut peringkat, kemudiannya prosedur standard berkaitan integrasi di peringkat tersebut akan dikenal pasti. Seterusnya, mekanisme penambahbaikan akan dicadangkan untuk menambah baik prosedur kerja di peringkat berkenaan. Cadangan dibuat berdasarkan turutan proses aktiviti integrasi.

i. Langkah 1: Permohonan Integrasi

Berintegrasi dengan organisasi yang merupakan punca kepada data merupakan prosedur standard yang dipraktikkan bagi mendapatkan data yang tepat dan terkini, Namun, ia menjadi sebaliknya apabila organisasi pembekal data berkeberatan berbuat demikian kerana faktor polisi, akta dan kerahsiaan data. Beberapa pendekatan yang perlu diperhalusi bagi memastikan organisasi tersebut bersetuju mempertimbangkan perkongsian data. Antara faktor yang perlu diambil kira adalah akta organisasi yang membenarkan data dikongsi dengan agensi pengguna data sahaja tanpa penglibatan pihak ketiga dan faktor keselamatan serta kerahsiaan maklumat bagi menjamin tiada ketirisan data berlaku melibatkan pihak yang berkepentingan yang tidak mempunyai kebenaran akses kepada maklumat tersebut.

- Akta Perkongsian Data Tanpa Penglibatan Pihak Ketiga

Terdapat organisasi yang terikat dengan akta iaitu menghalang perkongsian data melibatkan pihak ketiga. Kad pengenalan, nama penuh, alamat, gambar, nombor passport dan nombor telefon merupakan data peribadi yang berisiko untuk dikongsikan dan pada masa yang sama banyak pihak berkepentingan menginginkan data tersebut. Jabatan Pendaftaran Negara (JPN) merupakan salah satu organisasi yang tergolong dalam situasi tersebut iaitu ia terikat dengan Akta Pendaftaran Negara 1959 (Akta 78) Dan Peraturan-Peraturan Pendaftaran Negara 1990 (Pindaan 2017). Peraturan tersebut menyatakan mana-mana orang yang selaku penjawat awam, kecuali demi kepentingan awam dan dengan kebenaran Ketua Pengarah, atau bagi maksud prosiding jenayah, menerbitkan atau menyampaikan kepada sesiapa sebarang maklumat yang terkandung dalam daftar atau mana-mana buku atau rekod yang dibuat di bawah atau bagi maksud melaksanakan peruntukan-peruntukan Peraturan-Peraturan ini.

Pelaksanaan platform perkongsian data sektor awam melalui MyGDX merangkumi dua kaedah iaitu Registri di Agensi Pembekal Data dan Registri di MyGDX. Struktur arkitektur integrasi semasa MyGDX yang menempatkan perisian MyGDX di antara agensi pengguna dan agensi pembekal data bagi kedua-dua kaedah tersebut. Kedua-dua kaedah ini tidak menepati akta yang dikuatkuasa oleh sesetengah agensi termasuk JPN. Oleh itu, kaedah perkongsian data tanpa diketahui oleh pasukan perkongsian data perlu diperhalusi.

Antara platform perkongsian data yang selamat dan beroperasi secara *de-centralized* adalah Unified eXchange Platform (UXP) yang dibangunkan oleh Cybernetica dan X-road. Kedua-dua perisian ini menjamin pertukaran data yang selamat dan pada masa yang sama penyedia perkhidmatan tidak mengetahui butiran maklumat yang dikongsi kerana ia dijalankan di persekitaran *decentralized data exchange network*. Perisian berkenaan telah digunakan oleh beberapa negara termasuk Jepun, Ukraine, Amerika Syarikat, Estonia, Finland dan Kyrgyzstan merangkumi perbankan, perubatan serta perkhidmatan e-kerajaan secara menyeluruh.

- **Faktor Keselamatan Data Bagi Menjamin Kerahsiaan Maklumat**

Prinsip perkongsian data secara selamat sangat ditekankan oleh agensi sektor awam. Pengendalian data rahsia rasmi dalam persekitaran ICT hendaklah mematuhi tatacara yang ditetapkan di dalam Arahan Keselamatan (Semakan dan Pindaan 2017). Polisi perkongsian data perlu diperkukuhkan dengan memperkenalkan pekeliling yang perlu dipatuhi oleh agensi dalam memastikan pengendalian data rahsia rasmi semasa integrasi terpelihara. Ciri-ciri keselamatan bagi infrastruktur dan data perlu ditambahbaik seperti penggunaan *Secure Socket Layer (SSL)*, *Secure Posture Assessment & Penetration Testing (SPA & PEN-TEST)*, *Digital Time Stamping*, *Token/roaming certificate*, *authentication certificate & authentication key*, *data encryption*, *mutual certificate*, antivirus dan *backup* perlu disediakan bagi menjamin keselamatan perkongsian data merentasi agensi.

ii. **Langkah 2: Penetapan IDD**

Setelah mendapat kelulusan perkongsian data, Pelan integrasi perlu dibangunkan sebagai panduan dan rujukan bagi keseluruhan pelaksanaan integrasi sistem yang menggariskan kaedah, strategi dan jadual pelaksanaan integrasi yang perlu dipatuhi. Maklumat lengkap struktur data termasuk *field name*, *field attributes* (seperti *data type* dan *length*) dan *field properties* (*mandatory*, *optional*, *conditional*) perlu difahami terlebih dahulu sebelum diperincikan untuk penetapan IDD di dalam dokumen Pelan Integrasi.

Bagi mengelakkan salah tafsir data elemen semasa penetapan IDD, contoh data perlu dimohon terlebih dahulu untuk memastikan data yang diinginkan menepati format struktur data yang dirancang. Ini kerana kefahaman mengenai data tersebut adalah berbeza antara organisasi dan menerusi sampel data yang diperolehi, ia dapat memberi gambaran sebenar kepada agensi pengguna data tentang data yang bakal mereka perolehi setelah integrasi itu dilaksanakan kelak.

iii. Langkah 4: Pengguna memasukkan data

Maklumat yang diperolehi dan dijana oleh sistem adalah berasal dari kemasukan data oleh pengguna. Secara amnya, data bermula dari pengguna sistem itu sendiri sebelum ia diproses untuk memenuhi kehendak organisasi. Data yang tepat menyebabkan proses di dalam sistem berjalan lancar tanpa ralat serta menghasilkan pelaporan jitu. Beberapa langkah boleh dipraktikkan untuk mengatasi kesilapan pengguna semasa kemasukan data ke dalam sistem seperti berikut:

- Verifikasi Data

Proses verifikasi boleh dilakukan secara manual atau menggunakan teknologi.

- **Verifikasi secara manual** dilakukan oleh individu yang melaksanakan proses kerja berkaitan penghasilan data tersebut. Oleh itu, bagi memastikan data adalah tepat tanpa kesilapan kemasukan data serta mengelak kesalahan ejaan, proses verifikasi oleh penyelia atau pihak atasan sebelum data tersebut disimpan ke pangkalan data perlu ditambah di dalam fungsi sistem.
- **Verifikasi menggunakan teknologi** boleh dilaksanakan menerusi proses semakan data secara integrasi. Contohnya, bagi memastikan no kad pengenalan adalah tepat, semakan data secara *webservice* boleh dilaksanakan dengan JPN sebelum sebarang transaksi dibuat.

Selain dari verifikasi, penggunaan teknologi seperti imbasan RFID, kod *Quick Response* (QR) dan *barcode* juga membantu dalam mengatasi masalah kesalahan ejaan atau kesilapan kemasukan data oleh pengguna. Ia meningkatkan ketepatan data dan pada masa yang sama menjimatkan masa pelaksanaan.

- Pembudayaan Perkongsian Data

Selain itu, agensi juga hendaklah melaksanakan usaha pembudayaan perkongsian data. Pembudayaan ini boleh diterapkan melalui aktiviti kesedaran dan kefahaman seperti sisipan

tentang implikasi yang bakal terjadi serta menjejaskan kredibiliti penjawat awam dan organisasi akibat ketidaktepatan data semasa sesi taklimat dan bengkel mengenai sistem berkenaan.

Ia juga boleh dilaksanakan dengan program pembangunan kompetensi pengguna sistem yang terlibat meliputi prosedur dan garis panduan tatacara penggunaan sistem seperti penganjuran kursus, latihan teknikal serta sesi *Transfer Of Technology* (TOT) oleh pengguna yang berpengalaman menggunakan sistem berkenaan. Program ini harus diadakan secara berkala bagi memastikan pengetahuan pengguna terhadap sistem berkenaan sentiasa diperbaharui.

iv. **Langkah 6: Pangkalan Data Agensi Pembekal (*source*)**

Terdapat situasi apabila data yang dihantar kepada agensi pembekal tidak menepati standard yang telah didokumenkan di dalam IDD, pertambahan atau perubahan terhadap master data yang merupakan rujukan kepada data transaksi atau perubahan pada struktur pangkalan data bagi memenuhi keperluan bisnes agensi pembekal data. Sebarang perubahan terhadap struktur dan data integrasi perlu dimaklumkan kepada agensi pengguna dan pindaan terhadap dokumen IDD perlu dibuat dan ditandatangani semula kerana ia akan mempengaruhi jadual pelaksanaan dan skop integrasi. Reka bentuk terhadap integrasi data akan diubah suai berdasarkan pindaan Pelan Integrasi yang telah dipersetujui bersama.

v. **Langkah 7: Pembangunan API**

Bagi mengatasi masalah ketepatan data berkaitan nombor, pasukan pembangun sistem integrasi perlu memahami struktur pangkalan data serta jenis data (*data type*) yang terlibat terlebih dahulu. Jika input data yang diterima merupakan *varchar*, pentadbir pangkalan data sistem di agensi pengguna boleh menetapkan jenis data tersebut kepada *char*. Jenis data *char* merupakan medan *fixed length* berbanding *varchar* yang merupakan medan *variable length*. Dengan tindakan ini, sistem akan menyimpan input tersebut secara keseluruhan tanpa mengabaikan mana-mana aksara.

Tidak dinafikan penggunaan teknologi atau platform integrasi yang berbeza berkemungkinan akan menjejaskan ketepatan data. Namun, jika pengujian integrasi dijalankan secara terperinci, perkara ini boleh diatasi. Pengujian integrasi iaitu *System Integration Testing* (SIT) lazimnya dilaksanakan setelah pembangunan integrasi selesai. Ini bagi menguji tahap kesediaan data yang telah diintegrasikan dan digunakan di sistem agensi pengguna data. Adalah disarankan data yang akan digunakan semasa SIT adalah mirip kepada data sebenar yang bakal diperolehi bagi menguji kefungsi keseluruhan sistem integrasi termasuk ETL. Data yang berjaya dipam masuk ke pangkalan data sistem aplikasi sistem pengguna akan disemak bagi memastikan ia dipaparkan secara tepat. Hasil

dapatan pengujian akan direkodkan di dalam skrip pengujian SIT dan hasil keseluruhan SIT akan didokumenkan di dalam Laporan SIT yang perlu ditandatangani oleh kedua-dua belah pihak iaitu agensi pembekal data dan agensi pengguna data.

vi. Langkah 13 : Transformasi Data

Berdasarkan dapatan temuramah informan, terdapat tiga isu melibatkan proses transformasi data iaitu jenis data (*data type*) yang berbeza, kesilapan semasa pepadanan data dan kesilapan semasa perubahan data.

Kunci penyelesaian kepada isu-isu tersebut adalah pasukan pembangunan integrasi haruslah memahami dengan mendalam tentang struktur pangkalan data *source* (agensi pembekal data) dan struktur pangkalan data *target* (agensi pengguna data). Kefahaman tentang struktur pangkalan data serta data boleh didapati dengan membuat kajian tentang elemen data dan hal-hal berkaitan dengannya. Sesi perbincangan bersama SME dari sistem agensi pembekal dan agensi pengguna perlu diatur dengan kerap bagi mendapatkan *tacit knowledge* berkenaan proses cara kerja dan data terlibat. Selain itu, isu-isu itu juga boleh dikesan semasa pengujian SIT dilaksanakan. Oleh itu, pelaksanaan pengujian yang teliti dan menyeluruh dapat membantu bagi mengatasi permasalahan ini.

5. KEPUTUSAN DAN PERBINCANGAN

Cadangan mekanisme penambahbaikan dirumuskan bagi mendapatkan pandangan pegawai yang berpengalaman melibatkan integrasi. Pemilihan pegawai tersebut adalah berdasarkan pengalaman melebihi tujuh tahun dalam projek integrasi, pernah terlibat dalam pengurusan dan teknikal bagi pelaksanaan projek. Dua pegawai tersebut telah dikenal pasti dan dikodkan seperti di Jadual 5 ditemubual secara separa berstruktur bagi memberi pandangan berkaitan kajian.

Jadual 5 Maklumat Pegawai Berpengalaman

Kod	Pengalaman (Tahun)		Bilangan Projek
	Dalam Perkhidmatan	Dalam Pelaksanaan Projek	
P01	18 tahun	13 tahun	8
P02	13 tahun	7 tahun	3

Pengesahan dari pegawai telah dilakukan di mana setiap pegawai menyatakan persetujuan mereka. Didapati pegawai P01 dan P02 mencapai **persetujuan menyeluruh** terhadap cadangan penambahbaikan tersebut.

6. RUMUSAN

Kajian ini dilaksanakan bagi mengenal pasti isu kualiti data semasa proses integrasi data di sektor awam dan mencadangkan penambahbaikan proses integrasi data untuk meningkatkan kualiti data di sektor awam. Dimensi Ketepatan data telah dikenal pasti sebagai dimensi paling banyak mempengaruhi kejayaan projek integrasi sektor awam di Malaysia. Sebanyak 13 isu berkaitan ketepatan data telah dikenal pasti di dalam kajian. Cadangan mekanisme penambahbaikan telah dibentangkan merujuk kepada permasalahan di semua fasa yang terlibat dan kesemua cadangan tersebut telah dipersetujui dan disahkan oleh pegawai berpengalaman. Cadangan ini dapat membantu pasukan projek integrasi menjayakan pelaksanaan projek integrasi di organisasi masing-masing. Ia juga boleh diaplikasikan menerusi panduan khusus atau diterapkan sebagai salah satu standard prosedur bagi pelaksanaan integrasi.

Pada masa hadapan, kajian ini boleh diteruskan lagi agar lebih berkembang dan dipertingkatkan dari sudut perspektif yang berbeza. Ia boleh dipanjangkan ke agensi sektor awam yang lain dengan mendapatkan jumlah informan yang lebih besar bagi mendapatkan sudut perspektif yang berbeza untuk mengenal pasti sama ada terdapat elemen dimensi kualiti data lain yang mempengaruhi kejayaan proses integrasi data di sektor awam.

PENAKUAN DAN PENGHARGAAN

Penulis ingin mengucapkan ribuan terima kasih kepada Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia kerana memberi peluang kepada penulis untuk menjalankan penyelidikan ini

RUJUKAN

- Abdullah, A., Zuhoor, A. K., Kraiem, N. & Jamoussi, Y. al. 2018. Enhanced eGovernment integration framework for higher interoperability in eGovernment initiatives. *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2017* 2018-Janua: 1322–1329. doi:10.1109/ICICICT1.2017.8342761
- Abdullah, N., Ismail, S. A., Sophiyati, S. & Sam, S. M. 2015. Data quality in big data: A review. *International Journal of Advances in Soft Computing and its Applications* 7(Specialissue3).
- Almeida, F., Queirós, A. & Faria, D. 2017. Strengths and Limitations of Qualitative and Quantitative Research Methods Innovation and Entrepreneurship. *European Journal of Education Studies* 369–387. Retrieved from <https://www.researchgate.net/publication/319852576>
- Biswas, N., Sarkar, A. & Mondal, K. C. 2020. Efficient incremental loading in ETL processing for real-time data integration. *Innovations in Systems and Software Engineering* 16(1). doi:10.1007/s11334-019-00344-4

- Campbell, J. R., Falzon, D., Mirzayev, F., Jaramillo, E., Migliori, G. B., Mitnick, C. D., Ndjeka, N., et al. 2020. Improving quality of patient data for treatment of multidrug- or rifampin-resistant tuberculosis. *Emerging Infectious Diseases* 26(3). doi:10.3201/eid2603.190997
- Cheong, L. K. & Chang, V. 2007. The need for data governance: A case study. *ACIS 2007 Proceedings - 18th Australasian Conference on Information Systems*.
- Dagade, A. & Mali, M. 2017. De-duplication framework to reduce the record linkage problem. *2017 International Conference on Computer Communication and Informatics, ICCCI 2017*. doi:10.1109/ICCCI.2017.8117732
- DAMA International. 2017. DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition) July 2017. *Technics Publications, LLC 14 Elm St Denville NJ United States*, hlm. Vol. 44.
- Fernandez, D., Zainol, Z. & Ahmad, H. 2017. The impacts of ERP systems on public sector organizations. *Procedia Computer Science* 111: 31–36. doi:10.1016/j.procs.2017.06.006
- Fu, Q. & Easton, J. M. 2017. Understanding data quality: Ensuring data quality by design in the rail industry. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, hlm. Vol. 2018-January. doi:10.1109/BigData.2017.8258380
- Galici, R., Ordile, L., Marchesi, M., Pinna, A. & Tonelli, R. 2020. Applying the ETL process to blockchain data. Prospect and findings. *Information (Switzerland)* 11(4). doi:10.3390/INFO11040204
- Graneheim, U. H., Lindgren, B. M. & Lundman, B. 2017. Methodological challenges in qualitative content analysis: A discussion paper. *Nurse Education Today* 56. doi:10.1016/j.nedt.2017.06.002
- Gray, P. 1999. Welcome to the Communications of the Association for Information Systems. *Communications of the Association for Information Systems* 1. doi:10.17705/1cais.00106
- Hassan, N. H. M., Ahmad, K. & Salehuddin, H. 2020. Diagnosing the issues and challenges in data integration implementation in public sector. *International Journal on Advanced Science, Engineering and Information Technology* 10(2). doi:10.18517/ijaseit.10.2.10271
- Ibrahim, A., Mohamed, I. & Satar, N. S. M. 2021. Factors Influencing Master Data Quality: A Systematic Review. *International Journal of Advanced Computer Science and Applications* 12(2): 181–192. doi:10.14569/IJACSA.2021.0120224
- John Ladley. 2020. Data Governance How to Design, Deploy and Sustain an Effective Data Governance Program.
- Juneja, A. & Das, N. N. 2019. Big Data Quality Framework: Pre-Processing Data in Weather Monitoring Application. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects, COMITCon 2019*. doi:10.1109/COMITCon.2019.8862267
- Kim, H. Y. & Cho, J. S. 2018. Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of Business and Retail Management Research* 12(3). doi:10.24052/jbrmr/v12is03/art-04
- Kumar, P., Dixit, S. & Routray, S. 2018. Integration of Robotic Process Automation with E-Governance. *International Journal of Innovative Science and Research Technology* 3(12): 315–319.

- Kusumasari, T. F. & Fitria. 2017. Data profiling for data quality improvement with OpenRefine. *2016 International Conference on Information Technology Systems and Innovation, ICITSI 2016 - Proceedings*. doi:10.1109/ICITSI.2016.7858197
- Liu, Z. 2020. Practice Research on University Data Center Construction from the Perspective of Data Governance. *Proceedings - 2020 2nd International Conference on Information Technology and Computer Application, ITCA 2020* 489–492. doi:10.1109/ITCA52113.2020.00108
- Lokaadinugroho, I., Girsang, A. S. & Burhanudin, B. 2021. Tableau Business Intelligence Using the 9 Steps of Kimball's Data Warehouse & Extract Transform Loading of the Pentaho Data Integration Process Approach in Higher Education. *Engineering, MAThematics and Computer Science (EMACS) Journal* 3(1): 1–11. doi:10.21512/emacsjournal.v3i1.6816
- Lucas, A. 2019. Critical Success Factors for Corporate Data Quality Management. *Advances in Intelligent Systems and Computing*, hlm. Vol. 930. doi:10.1007/978-3-030-16181-1_60
- McGilvray, D. 2021. Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM). *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information (TM)*. Elsevier. doi:10.1016/B978-0-12-818015-0.09991-6
- Mendoza, L. E., Pérez, M. & Grimán, A. 2006. Critical success factors for managing systems integration. *Information Systems Management* 23(2). doi:10.1201/1078.10580530/45925.23.2.20060301/92674.7
- Muthee, V., Bochner, A. F., Osterman, A., Liku, N., Akhwale, W., Kwach, J., Prachi, M., et al. 2018. The impact of routine data quality assessments on electronic medical record data quality in Kenya. *PLoS ONE* 13(4). doi:10.1371/journal.pone.0195362
- Njuguna, C., Vandi, M., Mugagga, M., Kanu, J., Liyosi, E., Chimbaru, A., Gachari, W., et al. 2020. Institutionalized data quality assessments: A critical pathway to improving the accuracy of integrated disease surveillance data in Sierra Leone. *BMC Health Services Research* 20(1). doi:10.1186/s12913-020-05591-x
- Olson, J. E. 2003. Data Quality: The Accuracy Dimension. *Data Quality: The Accuracy Dimension*. doi:10.1016/B978-1-55860-891-7.X5000-8
- Pinto, C., Pita, R., Barbosa, G., Araujo, B., Bertoldo, J., Sena, S., Reis, S., et al. 2017. Probabilistic Integration of Large Brazilian Socioeconomic and Clinical Databases. *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, hlm. Vol. 2017-June. doi:10.1109/CBMS.2017.64
- Ridzuan, F. & Wan Zainon, W. M. N. 2019. A review on data cleansing methods for big data. *Procedia Computer Science* 161: 731–738. doi:10.1016/j.procs.2019.11.177
- Souibgui, M., Atigui, F., Zammali, S., Cherfi, S. & Yahia, S. ben. 2019. Data quality in ETL process: A preliminary study. *Procedia Computer Science*, hlm. Vol. 159. doi:10.1016/j.procs.2019.09.223
- Sreemathy, J., Naveen Durai, K., Lakshmi Priya, E., Deebika, R., Suganthi, K. & Aishwarya, P. T. 2021. Data Integration and ETL: A Theoretical Perspective. *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*. doi:10.1109/ICACCS51430.2021.9441997
- Sulistyo, H. A., Kusumasari, T. F. & Alam, E. N. 2020. Implementation of Data Cleansing Null Method for Data Quality Management Dashboard using Pentaho Data Integration. *2020 3rd*

International Conference on Information and Communications Technology, ICOIACT 2020.
doi:10.1109/ICOIACT50329.2020.9332030

Theodorou, V., Abelló, A., Lehner, W. & Thiele, M. 2016. Quality measures for ETL processes: from goals to implementation. *Concurrency and Computation: Practice and Experience*, hlm. Vol. 28. doi:10.1002/cpe.3729

Wang, X. & Wang, C. 2020. Time Series Data Cleaning: A Survey. *IEEE Access* 8.
doi:10.1109/ACCESS.2019.2962152

Copyright@FTSM
UKM