

Model Konseptual Platform Analitis Data Raya Di Sektor Awam

Noriha binti Abu Talib
Prof. Madya Dr. Jamaiah Yahaya

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Kajian ini bertujuan untuk mengenal pasti elemen keperluan platform analitis data raya (PADR) di sektor awam berdasarkan kepada isu dan cabaran dalam membangunkan PADR serta faktor kejayaan kritikal dalam pembangunan model konseptual PADR di sektor awam. Kajian lepas oleh IDC (2014), Alexandru et al. (2016) dan PTPA (Bil. 1) Platform aDRSA (2017) menunjukkan bahawa terdapat tiga (3) model yang berkelompokan. Justeru, kajian ini juga bertujuan untuk membangunkan model konseptual PADR yang komprehensif bagi kegunaan sektor awam. Kajian ini menggunakan kaedah yang melibatkan analisis dokumen dari dokumen tender dua organisasi serta tinjauan susastera untuk melihat pola pembangunan PADR sebelum ini. Dapatan ini diperkukuhkan dengan temubual bersama lima (5) orang informan yang menggunakan dan membangunkan PADR di sektor awam. Data daripada kajian susastera dan temubual tersebut digabungkan untuk membentuk model yang dikehendaki. Dari data yang telah dianalisis, didapati terdapat 12 isu dan cabaran dalam pembangunan PADR termasuk isu pemilihan data, pengesanan maklumat, kehilangan maklumat semasa proses transformasi dan perisian analitis. Hasil kajian juga menunjukkan bahawa faktor kejayaan kritikal terdiri daripada lapan (8) iaitu pengenalanpastian sumber data, capaian sumber data, data berkualiti tinggi, kombinasi pelbagai set data, perisian analitis yang bersesuaian, visualisasi yang interaktif, interpretasi keputusan analitis dan saluran penyampaian yang memenuhi jangkaan pengguna. Model yang terhasil menggambarkan satu proses yang linear mengandungi aspek sumber data, pemerolehan dan pengurusan data, pemprosesan maklumat dan pengetahuan. Model konsep yang dibina telah dibincangkan mengikut kegunaan dan faedah seperti keupayaannya untuk memperbaiki kedudukan Malaysia ke tahap kematangan analisis data raya yang lebih baik. Model ini juga dapat memberikan maklumat secara komprehensif tentang pembangunan platform analitis data raya sektor awam yang boleh dijadikan panduan.

1. PENGENALAN

Data raya merupakan kesan daripada perkembangan dan kepesatan teknologi digital. Melalui pelaksanaan projek analitis data raya, data dapat dimanfaatkan dengan membantu organisasi untuk lebih berdaya saing, mengoptimumkan operasi dan meningkatkan inovasi dalam penghasilan produk serta perkhidmatan. Sektor awam tidak ketinggalan dalam meneroka nilai data raya melalui pengumpulan data daripada rakyat, dan agensi sektor awam sendiri untuk membuat keputusan, merancang, melapor dan melakukan penambahbaikan terhadap kesemua proses dalam jabatan kerajaan (Fredriksson et al. 2017). Walaubagaimanapun, tidak semua projek analitis data raya dapat menghasilkan nilai yang dipersetujui semasa peringkat awal perancangan projek (Demirkan et al. 2014). Menurut Demirkan et al. (2014) lagi, antara kesilapan biasa yang dilakukan dalam projek analitis data raya adalah kegagalan untuk membangunkan keperluan data raya dalam organisasi, kurangnya siasatan ke atas produk atau perkhidmatan pembekal, dan tiada senibina integrasi data yang bersesuaian.

Berdasarkan kepada penilaian tahap kematangan analitis data raya yang telah dijalankan oleh Perbadanan Data Antarabangsa atau International Data Corporation (IDC) dan Perbadanan Ekonomi Digital Malaysia atau Malaysia Digital Economy Corporation (MDeC) pada tahun 2014, Malaysia berada pada tahap kematangan satu (1). Tahap satu (1) ini adalah tahap permulaan dan untuk kategori sektor awam, kedudukan tersebut masih kekal bagi tahun 2016.

Antara isu dan cabaran yang perlu dihadapi oleh organisasi adalah kesukaran untuk memilih teknologi PADR yang bersesuaian bagi proses mengintegrasikan, mengurus, menganalisis dan mempersembahkan nilai atau hasil analitis data raya (Demirkan et al. 2014; IDC 2014).

PADR bukan sahaja berkisar kepada pangkalan data, ia merujuk kepada teknologi keseluruhan penyelesaian analitis data raya yang kompleks (Demchenko et al. 2014). Pemilihan teknologi PADR yang bersesuaian telah memberi cabaran kepada pengurusan data dan pemprosesan maklumat (Koronios & Selle 2015; Zicari et al. 2016). PADR memerlukan data dalaman dan luaran yang diintegrasikan, model pemprosesan masa nyata, pengetahuan yang difahami dan maklumat tindakan yang perlu diambil (Chen et al. 2017). Walaubagaimanapun, memenuhi keseluruhan keperluan PADR bagi projek yang kompleks seperti analitis data raya boleh menyebabkan kegagalan dalam mencapai aspirasi projek yang ditetapkan (Infochimps 2013). Organisasi disarankan supaya mempunyai pemahaman yang jelas terhadap elemen keperluan kritikal platform analitis data raya bagi menangani isu ini (Infochimps 2013).

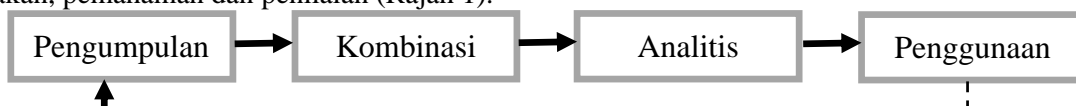
Sehingga kini belum ada model teknologi PADR yang komprehensif yang menggabungkan keseluruhan elemen keperluan kritikal pembinaan platform analitis data raya (Kuiler 2014). Walaupun wujud tiga model seperti yang dicadangkan oleh IDC (2014), Alexandru et al. (2016) dan PTPA (Bil. 1) MAMPU (2017), namun ketiga-tiga model tersebut bergerak secara berasingan dan terhad kepada set aplikasi tertentu yang tidak dapat diubah dengan mudah (Kamburugamuve et al. 2017). Setiap model tersebut mempunyai kekuatan yang boleh diambilkira dalam pembinaan platform analitis data raya.

2. ANALITIS DATA RAYA DI SEKTOR AWAM

Definisi asal data raya telah diperkenalkan melalui kajian mengenai data digital. Terdapat dua jenis data digital, data bersaiz kecil dan data bersaiz besar. Data yang bersaiz besar juga dikenali sebagai data raya. Ia mempunyai ciri-ciri 3Vs (Laney 2001) iaitu data yang banyak, diperolehi dari pelbagai sumber dan bergerak dengan kadar yang tidak dapat diuruskan oleh teknologi konvensional (Klievink et al. 2017; Raghupathi & Raghupathi 2014).

Ketiga-tiga ciri 3Vs (Laney 2001) tersebut merujuk ciri-ciri teknikal data raya dan telah dikembangkan kepada ciri-ciri lain yang dikaji oleh pengkaji terdahulu. Ciri-ciri tersebut adalah ciri-ciri tambahan kepada ciri yang wujud dalam 3Vs, termasuk nilai data kepada organisasi (Kim et al. 2014; Fredriksson et al. 2017), kebolehubahan data dalam kadar tertentu (Fox & Chang 2013), kualiti dan kepercayaan yang menentukan ketepatan dan keyakinan terhadap nilai data (Fredriksson et al. 2017), kecenderungan perubahan struktur data mengikut masa, kesesuaian data untuk digunakan serta ciri lain seperti metadata dan keterangan mengenai data (NIST Big Data Public Working Group 2015).

Sains data merupakan bidang berkenaan aktiviti yang berlaku dalam peringkat pemprosesan teknologi data raya ke atas data untuk mengekstrak pengetahuan daripada data mentah (NIST Big Data Public Working Group 2015). Kebesaran, kepelbagaian dan kepantasan data yang ekstrim (Klievink et al. 2017; Paul & Duggal 2016; Raghupathi & Raghupathi 2014) menyebabkan pemprosesan data menjadi kompleks bermula dengan proses pengumpulan data, kombinasi data dan analitis untuk menghasilkan nilai yang boleh digunakan bagi membuat inisiatif, kata putus, tindakan, pemahaman dan penilaian (Rajah 1).



- | | | | | |
|--------------|---------------|--------------------|----------------|----------------------|
| - Menjana | - Mengekstrak | - Menggabung | - Permodelan | - Inisiatif |
| - Mengesan | - Membersih | - Mengurus | - Menganalisa | - Membuat kata putus |
| - Memilih | - Menyediakan | - Menyimpan | - Visualisasi | - Praktikal |
| - Memperoleh | - Memproses | - Mengintegrasikan | - Interpretasi | - Pemahaman |
| - Mengumpul | | | | - Penilaian |

Rajah 1. Proses Data Raya (Klievink et al. 2017)

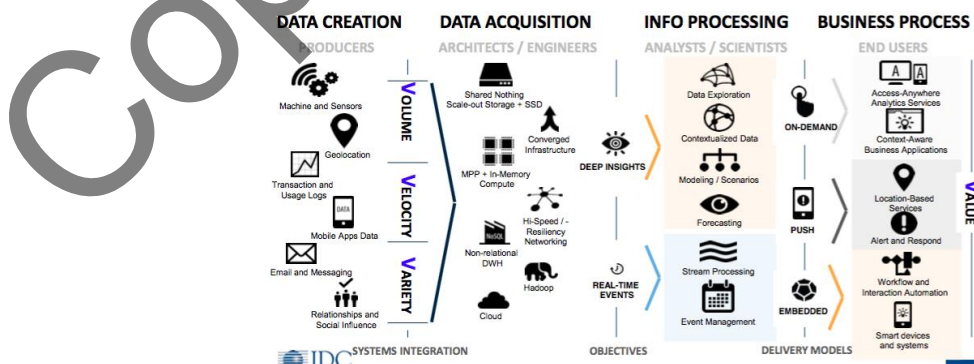
Rajah 1 menunjukkan proses-proses mentransformasi data mentah kepada pengetahuan yang boleh diambil tindakan seperti pada kitar hayat data raya iaitu pengumpulan data mentah, penyediaan maklumat, analisis untuk mensintesis pengetahuan dan tindakan untuk menghasilkan nilai (NIST Big Data Public Working Group 2015). Begoli et al. (2012) mencadangkan tiga prinsip untuk analisis yang lebih fleksibel iaitu sokongan kepada pelbagai kaedah analitis lanjutan seperti analisis statistik, pembelajaran mesin dan analisis visual, penggunaan storan yang dapat menampung jumlah data raya dan capaian data yang lebih efisien. Justeru, analitis data raya perlu disokong oleh senibina yang mampu memenuhi keperluan, prinsip, ciri-ciri, proses dan teknologi data raya dalam satu penyelesaian (IDC 2014).

3. PLATFORM ANALITIS DATA RAYA SEDIA ADA

Bagi mengenal pasti elemen keperluan platform analitis data raya yang lebih komprehensif, pengkaji telah mengkaji tiga (3) model platform analitis data raya yang bersesuaian dan sedia ada di sektor awam iaitu Ekosistem Data Raya – Aliran Kerja Teknologi serta Perkhidmatan (IDC 2014), Senibina Integrasi Ekosistem Data Raya (Alexandru et al. 2016), dan Platform aDRSA yang terkandung dalam Rangka Kerja Analitis Data Raya Sektor Awam (aDRSA) pada Pekeliling Transformasi Pentadbiran Awam 1/2017 Pelaksanaan Analitis Data Raya Sektor Awam. Ketiga-tiga model ini mengambilkira ciri-ciri data raya, pemprosesan data raya untuk menghasilkan nilai dan teknologi yang menyokong keperluan PADR di sektor awam.

3.1 Ekosistem Data Raya

Ekosistem Data Raya - Aliran Kerja Teknologi dan Perkhidmatan (IDC 2014) (Rajah 2) dipilih dalam kajian ini supaya pembangunan model konseptual PADR di sektor awam terarah kepada ciri-ciri ekosistem data raya IDC selaku badan yang menilai tahap kematangan analitis data raya yang dilaksanakan oleh sesebuah negara. Dalam masa yang sama, ekosistem ini turut dimuatkan dalam Laporan Akhir Kajian Kerangka Analitis Data Raya Kebangsaan (IDC 2014).



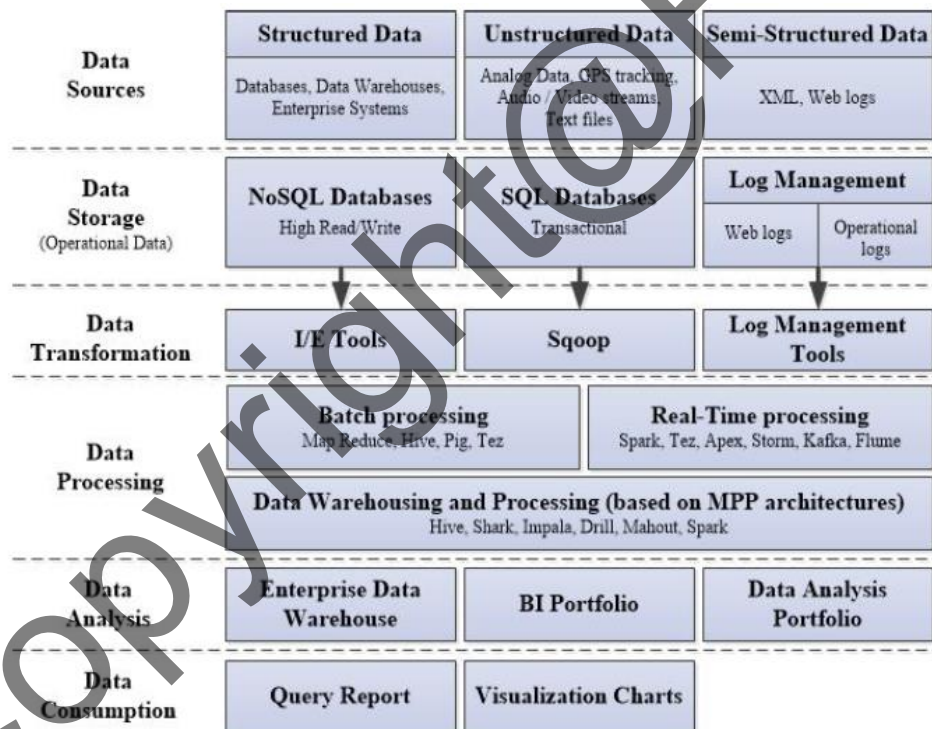
Rajah 2. Ekosistem Data Raya - Aliran Kerja Teknologi dan Perkhidmatan (IDC 2014)

Proses pertama model ini adalah proses pewujudan data iaitu penjanaan data melalui mesin dan sensor, lokasi geografi, log transaksi dan penggunaan, data aplikasi mudah alih, permesejan

dan e-mel serta hubungan dan pengaruh sosial. Seterusnya, bagi proses memperoleh data, data yang bersaiz besar, berubah dengan pantas dan pelbagai format perlu diintegrasikan ke dalam persekitaran teknologi yang menyokong infrastruktur seperti berikut: a) adaptasi secara dinamik, b) gabungan pelayan, storan, rangkaian dan perisian (*Converged Infrastructure*), c) pemprosesan secara serentak dan besar-besaran (*Massively Parallel Processing* (MPP)) serta pengiraan dalam memori, d) rangkaian berkelajuan tinggi dan berdaya tahan, e) gudang data bukan hubungan (*Non-relational DWH*), f) hadoop, dan g) pengkomputeran awan. Proses ketiga model ini adalah proses pemprosesan maklumat yang menggunakan teknologi pemahaman mendalam dan teknologi peristiwa masa nyata. Bagi mendapatkan pemahaman yang mendalam, data perlu diterokai, dikenal pasti tahap kerelevanan mengikut konteks data, dimodelkan mengikut model tertentu seperti model ramalan, dianalisa, dipersembahkan dalam bentuk visual dan seterusnya diinterpretasi. Proses ini dikenal pasti berada pada proses analitis Klievink et al. (2017) yang terdiri daripada permodelan, menganalisa, visualisasi dan interpretasi.

3.2 Senibina Integrasi Ekosistem Data Raya

Rajah 3 menunjukkan Senibina Integrasi Ekosistem Data Raya yang diperkenalkan oleh Alexandru et al. (2016).



Rajah 3. Senibina Integrasi Ekosistem Data Raya (Alexandru et al. 2016)

Model ini menggunakan konsep dan teknologi Hadoop bagi mengatasi tiga isu utama data raya iaitu; kepelbagaian sumber data, kualiti data yang akan diintegrasikan dan visualisasi data.

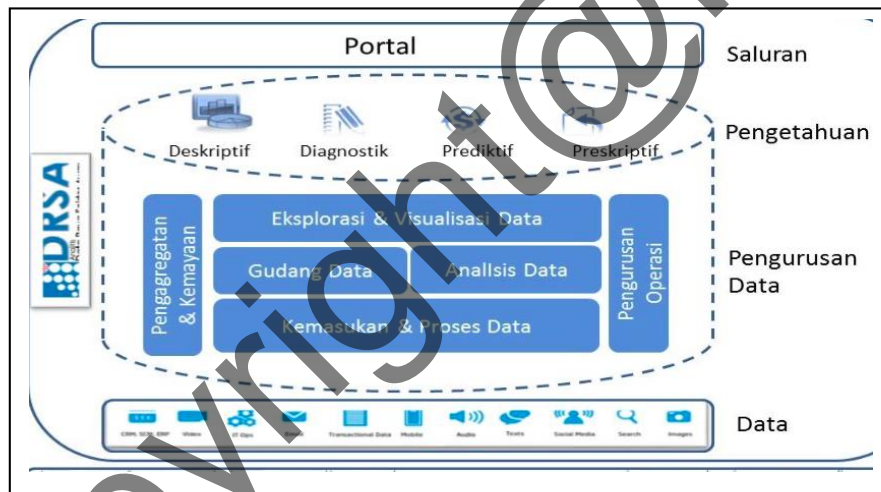
Berikut merupakan penerangan ringkas proses-proses yang terlibat dalam senibina ini.

- a. Alexandru et al. (2016) mengklasifikasikan data kepada tiga (3) jenis iaitu; data berstruktur, data tidak berstruktur dan data separa berstruktur.

- b. Proses kedua ialah storan data untuk data operasi yang disimpan dalam pangkalan data NoSQL (Non SQL) atau pangkalan data bukan hubungan (*non-relational database*), pangkalan data SQL dan pengurusan log laman web serta log operasi.
- c. Seterusnya, bagi membolehkan data diproses, data perlu ditransformasi menggunakan perisian import dan eksport tertentu bagi pangkalan data NoSQL dan SQL, perisian sqoop dari sumber data ke perisian transformasi Hadoop dan perisian pengurusan log.
- d. Pemrosesan data melibatkan kombinasi data berstruktur dan tidak berstruktur supaya pemrosesan berkelompok dan pemrosesan masa nyata dapat dilaksanakan.
- e. Proses analisis data model ini turut mengambilkira gudang data sedia ada bagi organisasi yang memerlukan fungsi baharu seperti integrasi sumber data tidak berstruktur dengan lebih baik dan memenuhi prestasi yang diperlukan dalam platform analisis.

3.3 Platform aDRSA

Rajah 4 menunjukkan platform analitis data raya sektor awam yang dibina oleh MAMPU pada tahun 2017. Model ini menekankan kepada aspek pengaliran daripada data kepada pengurusan data yang menghasilkan pengetahuan dan seterusnya bagaimana pengetahuan tersebut disalurkan.



Rajah 4. PTPA (Bil. 1) Platform aDRSA (2017)

Seperti yang telah dibincangkan sebelum ini, platform aDRSA terdiri daripada empat sub peringkat iaitu data, pengurusan data, pengetahuan dan saluran. Pengguna pengetahuan analitis data raya yang terhasil terdiri daripada pelbagai pihak termasuk sektor awam, komuniti dan sektor swasta. Pada peringkat data, sumber data adalah daripada aplikasi CRM, video, operasi IT, e-mel, data transaksi, peranti mudah alih, audio, teks, media sosial, carian dan imej. Seterusnya, pada peringkat pengurusan data, model ini tidak menyatakan proses pengumpulan data, sebaliknya terus ke proses kemasukan data dan proses data. Gudang data adalah tempat di mana berlakunya semua proses yang melibatkan pengurusan data seperti analisis, penerokaan dan visualisasi data. Pengagregatan dan kemayaan (*virtualization and caching*) serta pengurusan operasi bukanlah merupakan proses utama pada peringkat pengurusan data berdasarkan Klievink et al. (2017).

3.3 Perbincangan

Jadual 1 adalah rumusan proses ketiga-tiga model platform analitis data raya yang telah dikaji berasaskan kepada proses data raya Klievink et al. (2017).

Jadual 1. Rumusan Proses PADR di Sektor Awam

| Proses Data Raya (Klievink et al. 2017) | Proses IDC (2014) | Proses aDRSA (2017) | Proses Alexandru et al. (2016) | Proses PADR | Rujukan |
|---|-------------------------|--------------------------|--------------------------------|---|--|
| -Pengumpulan Menjana, Mengesan, Memilih | Pewujudan Data | Data | Sumber Data | 1. Sumber Data | IDC (2014); Klievink et al. (2017) |
| -Pengumpulan Memperoleh | Pemerolehan Data | Pengurusan Data | | 2. Pemerolehan & Pengurusan Data | IDC (2014); Klievink et al. (2017); aDRSA (2017) |
| -Pengumpulan Mengumpul | | | | - Pengumpulan | Klievink et al. (2017) |
| -Kombinasi Mengekstrak, Membersih, Menyediakan | | -Kemasukan data | Transformasi Data | - Kombinasi Transformasi Data | Alexandru et al. (2016); Klievink et al. (2017) |
| -Kombinasi Memproses, Menggabung, Mengurus, Menyimpan, Mengintegrasikan | | -Proses Data | Pemprosesan Data | Pemprosesan Data | Alexandru et al. (2016); aDRSA (2017); Klievink et al. (2017); |
| | Pemprosesan Maklumat | | | 3. Pemprosesan Maklumat | IDC (2014) |
| -Analitis | | Analisa | Analisa | - Analitis | aDRSA (2017); Klievink et al. (2017) |
| | Eksplorasi Konteks Data | Eksplorasi & Visualisasi | | Eksplorasi & Visualisasi | IDC (2014); aDRSA 2017 |
| Permodelan, | Permodelan | | | Konteks Data | IDC (2014); |
| Analisa, Visualisasi, | | | | Permodelan | Klievink et al. (2017) |
| | | Pengetahuan | | Analisa | Klievink et al. (2017) |
| -Interpretasi | | | | Visualisasi | (2017) |
| | | | | 4. Pengetahuan | aDRSA (2017); |
| | | | | - Interpretasi | Klievink et al. (2017) |
| -Penggunaan | Penggunaan | Saluran | Penggunaan | - Saluran Penyampaian | IDC (2014); aDRSA(2017) |

Ketiga-tiga model yang dikaji iaitu IDC (2014), Alexandru et al. (2016) dan aDRSA (2017) tidak menyatakan proses dan subproses analitis data raya dengan jelas yang merupakan asas

kepada keperluan pembangunan PADR (Demchenko et al. 2014). Justeru, proses data raya Klievink et al. (2017) dijadikan panduan dalam menentukan kesesuaian kedudukan teknologi, isu dan cabaran serta faktor kejayaan kritikal pembangunan PADR. Empat (4) proses utama yang dikaji oleh Klievink et al. (2017) adalah pengumpulan data, kombinasi data, analitis dan penggunaan hasil analitis data raya. Walaubagaimanapun, empat (4) proses utama ini bukanlah merupakan proses utama elemen keperluan PADR yang dibangunkan. Justeru, terdapat proses dan subproses daripada ketiga-tiga model ini yang turut dijadikan sebagai rujukan dan panduan dalam membangunkan proses-proses pada model konseptual PADR di sektor awam.

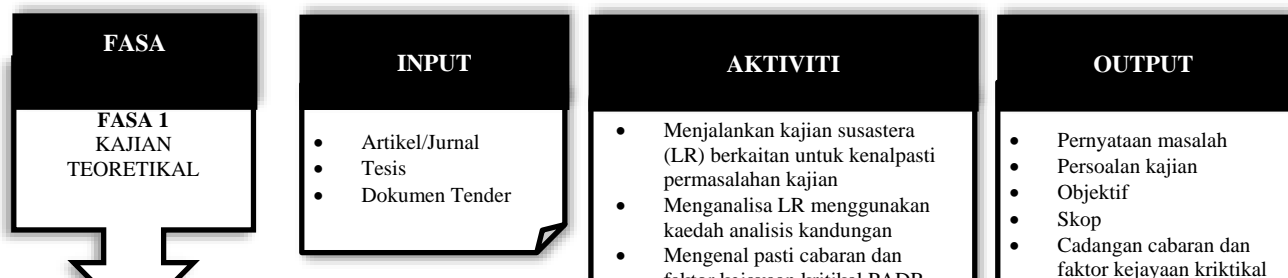
Proses pertama adalah melibatkan proses pra pengumpulan sumber data iaitu menjana data, mengesan dan memilih daripada data berstruktur, data tidak berstruktur dan data separa berstruktur (IDC 2014; Klievink et al. 2017). Proses kedua adalah proses pemerolehan data yang melibatkan subproses pengumpulan data ke dalam storan data iaitu pangkalan data NoSQL, pangkalan data SQL dan pengurusan log (Klievink et al. 2017; IDC 2014; Alexandru et al. 2016). Sumber data yang berbeza ini kemudiannya dikombinasi dan perlu melalui subproses transformasi data yang bertujuan untuk mengekstrakan dan pembersihan data (Alexandru et al. 2016; Klievink et al. 2017). Seterusnya, baharulah subproses pemprosesan data dapat dilaksanakan sama ada secara berkelompok atau secara masa nyata (Alexandru et al. 2016; G. Fox & Chang 2013; Zicari et al. 2016; Kamburugamuve et al. 2017). Dalam pada itu, set data yang telah dikombinasi disimpan dalam gudang data berasaskan semibina MPP (Alexandru et al 2016; aDRSA 2017; Koronios & Selle 2015).

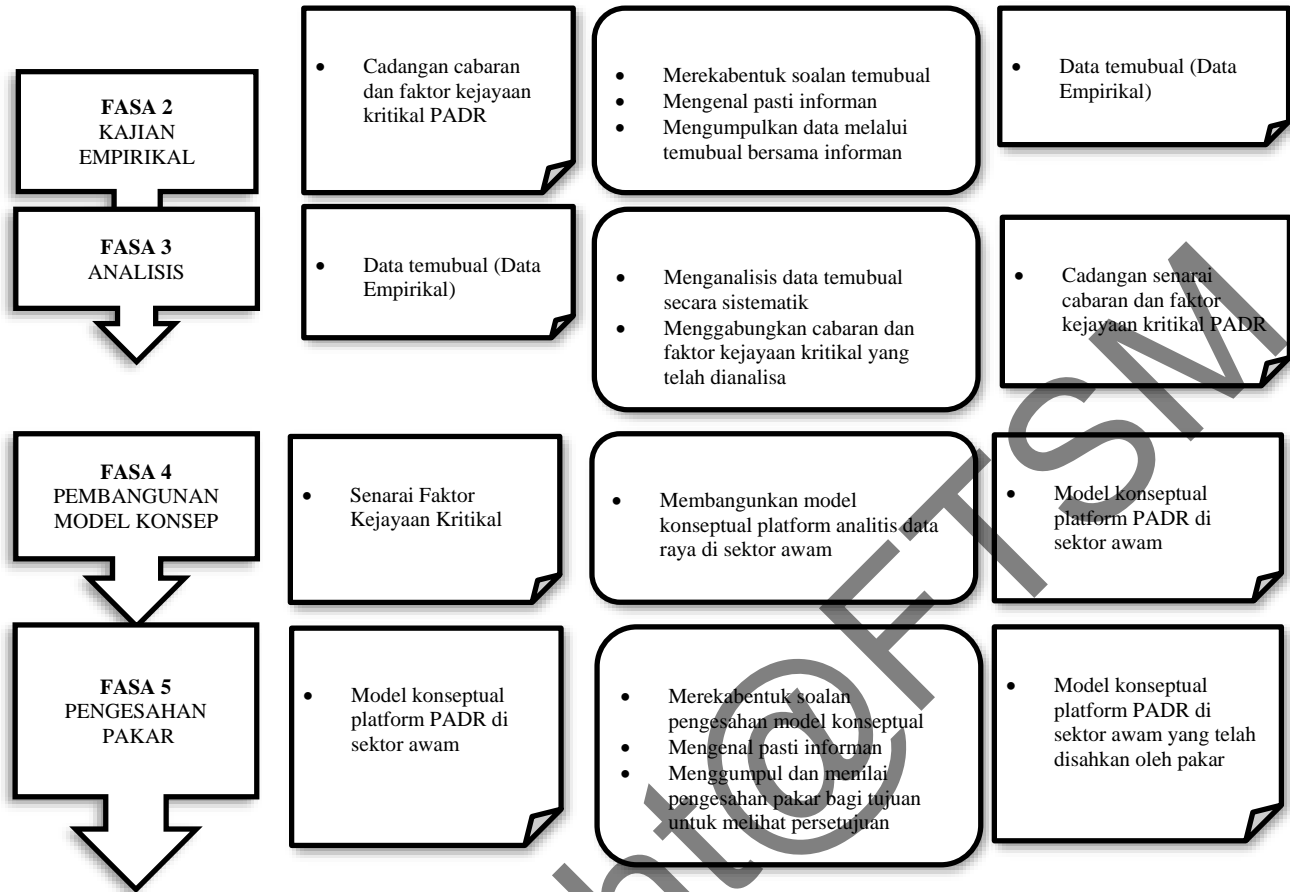
Pada peringkat pemprosesan maklumat, maklumat pada Gudang Data MPP dan Gudang Data Organisasi (Alexandru et al. (2016); Hortonworks (2014) kemudiannya dianalisis bagi menyokong keperluan peristiwa masa nyata dan pemahaman mendalam (IDC 2014; Alexandru et al. 2016; aDRSA 2017). Pada peringkat analitis, subproses eksplorasi dan visualisasi data, konteks data, permodelan, analisa dan visualisasi dilaksanakan. Permodelan analitis data raya melibatkan algorithma, statistik, pembelajaran mesin dan ramalan (Begoli & Horey 2012). Visualisasi akan diinterpretasi dan pengetahuan yang digabungkan dengan wawasan (aDRSA 2017) akan disalurkan mengikut model saluran tertentu untuk dimanfaatkan oleh pengguna (IDC 2014).

Tiga model platform analitis data raya yang dikaji melalui analisis dokumen menunjukkan bahawa terdapat kelemahan-kelemahan seperti tidak mengambilkira teknologi pengurusan data sedia, sumber data yang tidak dikategorikan mengikut kepelbagaian struktur data dan tidak menerangkan teknologi yang digunakan. Terdapat juga model yang tidak menyatakan proses-proses asas analitis data raya dengan jelas dan lengkap. Ketiga-tiga model tersebut bukan sahaja tidak lengkap, malah tidak mengandungi faktor kejayaan kritikal yang merupakan keperluan dalam membangunkan platform analitis data raya di sektor awam. Bagaimanapun, model yang dibangunkan akan mengambilkira kelemahan-kelemahan tersebut.

4. KAEDAH KAJIAN

Kajian ini terdiri daripada empat fasa iaitu kajian teoretikal, kajian empirikal, analisis, pembangunan model konsep dan pengesahan pakar. Secara umumnya, kajian ini dilakukan untuk mendapatkan data yang asli dalam bentuk penerokaan. Data berbentuk berayat atau “string data” dihasilkan bagi menerangkan sesuatu fenomena yang dikaji dengan lebih terperinci. Data yang dikumpulkan pula adalah melalui analisis dokumen, temubual mendalam, temubual fokus atau pun melalui pemerhatian. Kajian yang dijalankan ini mempunyai metodologi kajian yang dibahagi kepada lima (5) fasa seperti yang ditunjukkan dalam Rajah 5.





Rajah 5. Metodologi Kajian

Fasa 1 telah dilaksanakan menggunakan kaedah analisis kandungan yang melibatkan pengkaji menganalisis beberapa dokumen berkaitan dengan fenomena yang dikaji. Menurut Hsieh dan Shanon (2005), terdapat tiga pendekatan yang berbeza dalam menjalankan analisis kandungan iaitu: pendekatan konvensional, diarahkan, atau summatif. Manakala menurut Creswell (2014), analisis kandungan adalah satu kaedah penyelidikan yang memberikan tumpuan kepada isi kandungan yang dipaparkan di dalam sesuatu media (termasuk media cetak atau media elektronik). Untuk menjalankan analisis kandungan menggunakan teks, pengkaji telah membina satu set skema kod atau tema berdasarkan pembacaan berkaitan dengan fenomena yang dikaji. Kod skema tersebut digunakan untuk mengenalpasti teks-teks yang dapat menerangkan faktor-faktor kejayaan kritikal yang telah dicadangkan dalam model konseptual PADR.

Di dalam fasa kedua, pengkaji telah menggunakan data yang dikumpulkan dari fasa satu (1) untuk membina kajian empirikal. Bagaimanapun, sebelum kajian empirikal dapat dijalankan, pengkaji telah mengenalpasti isu dan cabaran serta faktor kejayaan kritikal dalam memastikan sesuatu projek data raya berjaya dan ini dianggap sebagai data teoretikal fasa dua (2). Data tersebut juga telah digunakan oleh pengkaji untuk membentuk soalan temubual, mengenal pasti informan dan seterusnya menemubual informan yang telah dipilih untuk mendapatkan data yang lebih mendalam bagi mengkaji isu dan cabaran serta faktor kejayaan kritikal pembangunan PADR di sektor awam. Senarai informan yang ditemubual adalah seperti di Jadual 2 di bawah:

Jadual 2. Senarai Profil Informan

| Kod Informan | Jawatan | Pengalaman Perkhidmatan (Tahun) | Pengalaman Projek Analitis Data Raya (Tahun) | Domain / Bidang |
|--------------|--|---------------------------------|--|---|
| IP 1 | Ketua Penolong Pengarah Kanan (Gred F52) | 15 | 3 | Analisis Sentimen |
| IP 2 | Pegawai Kesihatan Persekitaran (Gred U41) | 10 | 2 | Kesihatan - Ramalan Penyakit Berjangkit |
| IP 3 | Penolong Pengarah Kanan (Gred F44) | 10 | 2 | Ekonomi - Pemantauan Harga |
| IP 4 | Penolong Penguasa Polis (Assistant Superintendent of Police (ASP)) | 15 | 2 | Polis - Pencegahan Jenayah |
| IP 5 | Pengarah Projek | 15 | 6 | Pelbagai |

Data temubual yang dikumpulkan adalah berbentuk empirikal. Data tersebut telah ditranskripsikan secara verbatim ayat demi ayat mengikut perbualan dalam temubual tersebut dan telah dikategorikan mengikut isu dan cabaran serta faktor kejayaan kritikal yang telah ditetapkan dalam model konseptual yang dibina pada fasa satu (1).

5. DAPATAN KAJIAN

Wujud pelbagai cabaran dalam pembangunan PADR. Kajian ini telah dapat mengenalpasti 12 cabaran tersebut melalui analisis dokumen dengan menggunakan sumber-sumber dari kajian susastera dan dokumen-dokumen yang digunakan sebagai salah satu sumber data dan perlu diambil kira oleh sesebuah organisasi yang ingin membina PADR.

5.1 Isu dan Cabaran PADR Di Sektor Awam

Jadual 3 menerangkan isu dan cabaran tambahan berdasarkan kepada empat (4) proses utama platform analitis data raya iaitu sumber data, pemerolehan dan pengurusan data, pengurusan maklumat dan pengetahuan.

Jadual 3. Isu dan Cabaran PADR

| Proses | Cabaran (C) | Rujukan |
|---|-------------------------------|---|
| Sumber Data | C1. Pemilihan Data | Koronios & Selle (2015) |
| | C2. Pengesanan Maklumat | Dokumen Tender NRE (2016 & 2017); Khan et al. (2014); Infochimps (2013) |
| Pemerolehan & Pengurusan Data - Pengumpulan | C3. Masa Mendapatkan Maklumat | Dokumen Tender NRE (2016 & |

| | | |
|----------------------------------|--|--|
| - Kombinasi Transformasi Data | C4. Kualiti Data | 2017) Koronios & Selle (2015); Demirkan et al. (2014); Khan et al. (2014) |
| Pemrosesan Data | C5. Kehilangan Maklumat | Koronios & Selle (2015) |
| | C6. Ketekalan Dalam Penyimpanan Data | Khan et al. (2014); Gudipati et al. (2013) |
| | C7. Data Yang Tiada Had | Fox (2017) |
| | C8. <i>Internet of Things</i> (IoT) | Zicari et al. (2016); Koronios & Selle (2015); Khan et al. (2014); Ekanayake & Fox (2010); NIST Big Data Public Working Group (2015) |
| Pemrosesan Maklumat | | |
| -Analitis Visualisasi | C9. Perisian Analitis C10. Visualisasi | NRE (2016); IDC (2014) Alexandru et al. (2016); Khan et al. (2014) |
| Pengetahuan | | |
| Interpretasi Saluran Penyampaian | C11. Hasil dan Tafsiran C12. Penyaluran Pengetahuan | Paul & Duggal (2016) IDC (2014) |

Berikut merupakan isu dan cabaran PADR di sektor awam.

- a. Koronios dan Selle (2015) telah mengenal pasti bahawa organisasi sering menghadapi masalah dalam membuat keputusan sama ada untuk menyimpan kesemua data atau menyimpan sebahagian daripadanya sahaja
- b. Cabaran kedua adalah pengesanan maklumat untuk memenuhi keperluan kes bisnes analitis data raya yang dibangunkan oleh agensi (Dokumen Tender NRE 2016 & 2017; Khan et al 2014; Infochimps 2013).
- c. Cabaran ketiga berlaku pada peringkat pengumpulan data iaitu masa yang lama untuk mendapatkan maklumat (Dokumen Tender NRE 2016 & 2017).
- d. Cabaran keempat adalah masalah kualiti data apabila berlakunya proses transformasi data daripada pelbagai sumber data untuk dianalisis oleh pasukan analisis data (Koronios & Selle 2015; Demirkan et al. 2014; Khan et al. 2014).
- e. Semasa proses transformasi data juga, sangatlah mencabar untuk mengekstrak data mentah yang tidak berstruktur ke bentuk yang berstruktur di mana berkemungkinan akan berlaku kehilangan maklumat semasa proses transformasi data (Koronios & Selle 2015).
- f. Cabaran seterusnya adalah tiada ketekalan dalam penyimpanan data (Khan et al. 2014; Gudipati et al. 2013).
- g. Cabaran ketujuh adalah data yang tiada had memberikan isu kepada pembangunan platform analitis data raya.

- h. Internet of Things atau IoT turut memberikan cabaran dalam pemprosesan data (Zicari et al. 2016; Koronios & Selle 2015; Khan et al. 2014; Ekanayake & Fox 2010; NIST Big Data Public Working Group 2015).
- i. Pemilihan perisian analitis turut dikenalpasti memberikan cabaran kepada pembangunan platform analitis data raya (NRE 2016; IDC 2014).
- j. Dalam pada itu, visualisasi turut memberikan cabaran dari segi persembahan hasil analitis data raya (Alexandru et al. 2016; Khan et al. 2014).
- k. Cabaran pada proses akhir iaitu pengetahuan adalah mentafsirkan hasil yang diperoleh iaitu kepada wawasan yang boleh diambil tindakan (Paul & Duggal 2016).
- l. Manakala isu dan cabaran terakhir pembangunan PADR adalah menyalurkan pengetahuan yang diperoleh (IDC 2014).

5.2 Faktor Kejayaan Kritikal PADR

Berdasarkan isu dan cabaran yang merupakan elemen keperluan PADR yang telah dikenal pasti melalui dapatan kajian susastera dan analisis temubual yang dijalankan, terdapat lapan (8) faktor kejayaan kritikal yang telah dikenal pasti iaitu pengenalpastian sumber data, capaian sumber data, data berkualiti tinggi, kombinasi pelbagai set data, perisian analitis yang bersesuaian, visualisasi yang interaktif, interpretasi keputusan analitis dan saluran penyampaian yang memenuhi jangkaan pengguna.

5.3 Isu dan Cabaran Serta Faktor Kejayaan Kritikal PADR

Isu dan cabaran utama membangunkan PADR pada proses pengumpulan sumber data adalah pemilihan data dan pengesanan maklumat. Cabaran ini dapat diatasi dengan memastikan sumber data berjaya dikenal pasti. Seterusnya, cabaran pembangunan PADR pada proses pemerolehan dan pengurusan data adalah masa untuk mendapatkan maklumat, kualiti data, kehilangan maklumat, ketekalan dalam penyimpanan data, data yang tiada had dan Internet of Things (IoT). Faktor kejayaan kritikal bagi cabaran-cabaran tersebut adalah capaian sumber data, data berkualiti tinggi dan kombinasi pelbagai set data. Pada proses utama yang seterusnya iaitu pemprosesan maklumat, dua (2) cabaran dalam membangunkan PADR telah dikenal pasti iaitu perisian analitis dan visualisasi. Perisian analitis yang bersesuaian dan visualisasi yang interaktif merupakan faktor kejayaan kritikal dalam membangunkan platform analitis data raya di sektor awam. Proses terakhir pada PADR adalah menginterpretasi dan menyampaikan pengetahuan yang diperoleh juga merupakan cabaran yang perlu ditangani. Dua (2) faktor kejayaan kritikal bagi cabaran tersebut adalah menginterpretasi keputusan analitis dan saluran penyampaian yang memenuhi jangkaan pengguna.

Dalam menghasilkan model konseptual PADR yang komprehensif, ketiga-tiga model yang dikaji dibanding dan dirumus berdasarkan kepada padanan proses, teknologi, isu dan cabaran serta faktor kejayaan kritikal (rujuk Jadual 4).

Jadual 4. Elemen Keperluan PADR

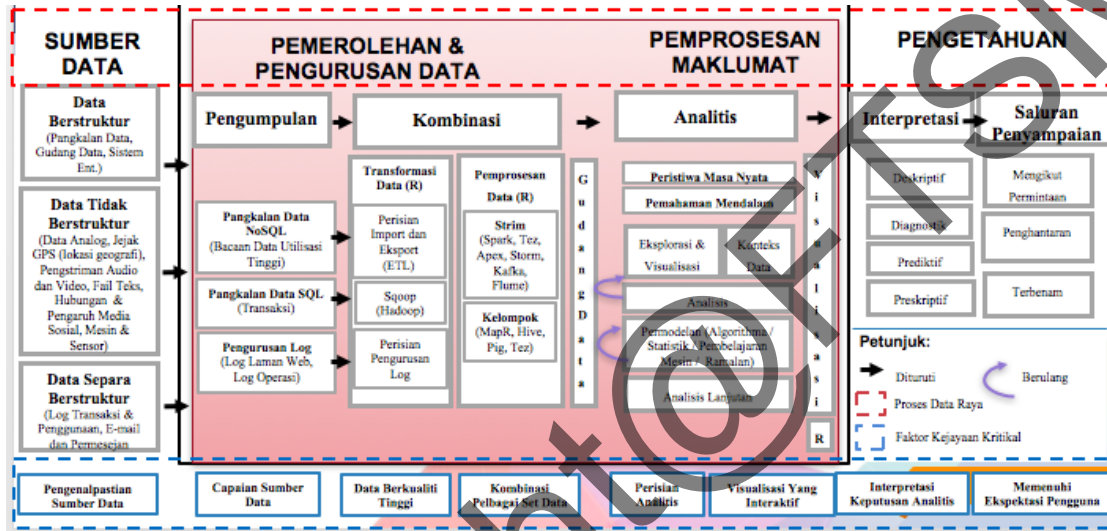
| Proses (P) | Teknologi | Isu dan Cabaran (C) | Faktor Kejayaan Kritikal (FKK) |
|--|--|---|--|
| P1. Sumber Data | Data Berstruktur Data Tidak Berstruktur Data Separa Berstruktur | C1. Pemilihan Data C2. Pengesanan Maklumat | FKK1. Pengenalpastian Sumber Data |
| P2. Pemerolehan & Pengurusan Data | | | |
| -Pengumpulan Data | Storan Data (Pangkalan Data NoSQL, Pangkalan Data SQL, Pengurusan Log) | C3. Masa Mendapatkan Maklumat | FKK2. Capaian Sumber Data |
| -Kombinasi Transformasi Data | Perisian Import/Eksport, Sqoop (Hadoop), Perisian Pengurusan Log | C4. Kualiti Data C5. Kehilangan Maklumat | FKK3. Data Berkualiti Tinggi |
| -Kombinasi Pemprosesan Data | Berkelompok Masa Nyata | C6. Ketekalan Dalam Penyimpanan Data C7. Data Yang Tiada Had | FKK4. Kombinasi pelbagai set data |
| -Kombinasi | Gudang Data | C8. <i>Internet of Things</i> (IoT) | |
| P3. Pemprosesan Maklumat | Peristiwa Masa Nyata & Pemahaman Mendalam | | |
| -Analitis | | C9. Perisian Analitis | FKK5. Perisian Analitis Yang Bersesuaian |
| -Eksplorasi & Visualisasi Konteks Data | | C10. Visualisasi | FKK6. Visualisasi Yang Interaktif |
| Permodelan | Algorithma / Statistik / Pembelajaran Mesin / Ramalan | | |
| Analisa | | | |
| Visualisasi | | | |
| P4. Pengetahuan | | | |
| Interpretasi | Deskriptif, Diagnostik, Prediktif, Preskriptif | C11. Hasil dan Tafsiran | FKK7. Interpretasi Keputusan Analitis |
| -Saluran Penyampaian | Model Saluran (Mengikut Permintaan, Penghantaran, Terbenam) | C12. Penyaluran Pengetahuan | FKK8. Memenuhi Jangkaan Pengguna |

Ketiga-tiga model yang dikaji iaitu IDC (2014), Alexandru et al. (2016) dan aDRSA (2017) tidak menyatakan proses dan subproses analitis data raya dengan jelas yang merupakan asas kepada keperluan pembangunan PADR (Demchenko et al. 2014). Justeru, proses data raya Klievink et al. (2017) dijadikan panduan dalam menentukan kesesuaian kedudukan teknologi, isu dan cabaran serta faktor kejayaan kritikal pembangunan PADR. Empat (4) proses utama yang dikaji oleh Klievink et al. (2017) adalah pengumpulan data, kombinasi data, analitis dan penggunaan hasil analitis data raya. Walaubagaimanapun, empat (4) proses utama ini bukanlah merupakan proses utama elemen keperluan PADR yang dibangunkan. Justeru, terdapat proses

dan subproses daripada ketiga-tiga model ini yang turut dijadikan sebagai rujukan dan panduan dalam membangunkan proses-proses pada model konseptual PADR di sektor awam.

6. RUMUSAN

Pada asasnya data dikumpul dan diproses untuk menghasilkan nilai. Nilai tersebut digunakan untuk membantu organisasi membuat keputusan dan melakukan penambahbaikan dalam sistem pengurusan. Rajah 6 menunjukkan Model konseptual PADR di Sektor Awam yang turut mengambilkira penerangan informan-informan yang telah ditemubual.



Rajah 6. Model Konseptual PADR di Sektor Awam

Rajah di atas menerangkan hasil gabungan data, proses dan teknologi melalui dapatan kajian susastera dan analisis temubual yang dilaksanakan bersama informan MAMPU yang membangunkan platform aDRSA. Berdasarkan jadual tersebut, peringkat terdiri daripada proses yang terlibat bagi menghasilkan nilai daripada sumber data raya. Peringkat pertama adalah sumber data atau asal pewujudan data dalam bentuk data berstruktur, data tidak berstruktur dan data separa struktur yang terhasil daripada teknologi digital terkini. Peringkat kedua adalah pengurusan data yang merangkumi proses pemerolehan data bermula dengan pengumpulan dan kombinasi data melibatkan teknologi pengurusan data sedia ada. Peringkat seterusnya adalah peringkat bermulanya pemprosesan maklumat dengan menggunakan analisis data sedia ada di organisasi dan teknologi analitis lanjutan bagi mendapatkan pemahaman yang lebih mendalam. Proses analitis data raya melibatkan eksplorasi dan visualisasi, konteks data, analisis, pemodelan, analisis lanjutan dan visualisasi. Seterusnya, pada peringkat pengetahuan, interpretasi dibuat berdasarkan deskriptif, diagnostik, prediktif dan preskriptif. Kesemua pengetahuan yang diperolehi digunakan untuk membuat inisiatif, kata putus, praktikal, pemahaman, penilaian melalui model saluran penyampaian tertentu.

Pakar telah diberikan instrumen soalan pengesahan untuk melihat persetujuan mereka ke atas ke atas elemen-elemen model konseptual yang telah dibina. Empat (4) pakar telah dipilih untuk melakukan pengesahan. Mereka terdiri daripada:

- a. Pakar Pertama – Pensyarah Jabatan Kepimpinan dan Polisi Pendidikan, Fakulti Pendidikan, Universiti Kebangsaan Malaysia
- b. Pakar Kedua - Chief Technology Officer MyKadProSolutions (Teknologi Maklumat, Portal Negara dan Pengkomputeran Awan)
- c. Pakar Ketiga - Pengarah Urusan CloudConnect Asia (Data Raya & Artificial Intelligence)
- d. Pakar Keempat - Pengarah Urusan SongketMail Sdn. Bhd (Infrastruktur Data Raya dan Visualisasi Data dengan Kibana (ELK Stack))

Pengesahan pakar telah dilakukan melalui proses persetujuan untuk menilai elemen-elemen yang dicadangkan dalam model konseptual. Hasil analisis Cohen Kappa (K) yang dilakukan menunjukkan bahawa semua pakar bersetuju elemen-elemen yang dicadangkan adalah merupakan keperluan yang perlu diberi pertimbangan apabila membina platform data raya yang berkesan. Nilai K yang diperolehi adalah 1.00. Ini menunjukkan bahawa nilai persetujuan itu adalah sangat tinggi. Denzin dan Lincoln (2003) menyatakan bahawa, jika nilai persetujuan Kappa mencecah 1.00, ini bermakna bahawa persetujuan yang diperolehi adalah sangat kuat dan jika persetujuan ini berkaitan dengan sesuatu pembinaan model, ini menunjukkan bahawa model tersebut adalah model yang mantap.

7. RUMUSAN

Kesimpulannya, ketiga-tiga objektif kajian ini telah berjaya dicapai. Objektif pertama adalah untuk mengenal pasti isu dan cabaran dalam membangunkan platform analitis data raya di sektor awam telah diperolehi melalui kajian teoritikal dan kajian empirikal. Objektif kedua iaitu mengenal pasti faktor kejayaan kritikal pembangunan platform analitis data raya turut dicapai melalui metod yang sama. Manakala objektif terakhir iaitu pembangunan model konseptual platform analitis data raya di sektor awam turut menggunakan metod yang sama. Pengesahan pakar ke atas model konseptual yang dibina juga menunjukkan bahawa model tersebut adalah satu model yang komprehensif. Setiap elemen yang dicadangkan dalam model tersebut untuk membina platform data raya telah dipersetujui sepenuhnya oleh kesemua pakar yang terlibat. Justeru, model tersebut boleh digunakan untuk membina platform data raya baik di sektor awam atau swasta.

Kajian ini boleh menyumbang kepada pengetahuan dalam bidang sains data serta amalan berhubung dengan pembangunan platform analitis data raya. Walaubagaimanapun, kajian yang dijalankan mempunyai beberapa kekangan. Justeru, hasil dapatan kajian ini, tidak boleh digeneralisasikan secara umum dan kajian yang dijalankan merupakan kajian awal untuk melihat pembangunan PADR secara menyeluruh.

RUJUKAN

- Alexandru, A., Alexandru, C.A., Coardos, D. & Tudora, E. 2016. Big Data : Concepts, Technologies and applications in the public sector. *International Journal of Computer and Information Engineering* 10(10): 1751-1757.
- Begoli, E. & Horey, J. 2012. Design principles for effective knowledge discovery from big data.

2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture, 215–218.

Chen, H.-M., Schütz, R., Kazman, R. & Matthes, F. 2017. How lufthansa capitalized on big data for business model renovation. *MIS Quarterly Executive* 1615(14): 299–320.

Cohen J. 1960. A coefficient of agreement for nominal scale. *Educational and Psychological Measurement* 20(2): 25-31.

Creswell, J.W. 2014. *Research design qualitative, quantitative and mixed methods approaches*, Ed. Ke-4. Thousand Oaks: CA Sage.

Demchenko, Y., De Laat, C. & Membrey, P. 2014. Defining architecture components of the Big Data Ecosystem. *2014 International Conference on Collaboration Technologies and Systems, CTS 2014* 104–112.

Demirkan, H. & Dal, B. 2014. Why do so many analytics projects still fail? Key considerations for deep analytics on big data, learning and insights. *INFORM Analytics*, July-August, 44-52.

Denzin N.K. & Lincoln, Y. 2003. *Collecting and interpreting qualitative material*, Thousand Oaks: Sage Pub.

Dokumen Tender MAMPU. 2014. Request for proposal. Analisis sentimen (sentiment analysis) terhadap pelaksanaan cukai barang dan perkhidmatan (CBP) Malaysia.

Dokumen Tender NRE. 2016. T03/2016 - Pembangunan rangka kerja dan projek rintis big data kementerian sumber asli dan alam sekitar (NRE).

Dokumen Tender NRE. 2017. T02/2017 - Membekal, Menghantar, memasang, mereka bentuk, membangun, menguji dan mentauliah perkakasan, perisian dan sistem aplikasi bagi big data analytics kementerian sumber asli dan alam sekitar (BDA NRE).

Ekanayake, J. & Fox, G. 2010. Architecture and performance of runtime environments for data intensive scalable computing. Disertasi Ph.D. Indiana University.

Fox, G.C. 2017. Twister2 : Design of a big data toolkit twister2 : Design of a big data toolkit (November). Disertasi Ph.D. Indiana University.

Fox, G. & Chang, W.O. 2013. Big data use cases and requirements.

Fredriksson, C., Mubarak, F. & Tuohimaa, M. 2017. Big data in the public sector : *A Systematic Literature Review* 21(3): 39–61.

Gudipati, B.M., Rao, S., Mohan, N.D. & Gajja, N.K. 2013. Big data : *Testing approach to overcome quality challenges structured testing technique* 11(1): 65–73.

Hortonworks. 2014. The public sector improves security and efficiency with hadoop @ a modern data architecture for government agencies (December) 1–12.

Hsieh, H.F. & Shannon, S.E. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research* 15(9): 1277-1288.

- IDC. 2014. Laporan akhir kajian kerangka analitis data raya kebangsaan. *Slaid* 1-129.
- Infochimps. 2013. CIOs & Big data: What your IT team wants you to know. *Infochimps Whitepaper* 1-10.
- Khan, N., Yaqoob, I., Hashem, I.A.T., Inayat, Z., Mahmoud Ali, W.K., Alam, M. & Shiraz, M. 2014. Big data: Survey, technologies, opportunities, and challenges. *The Scientific World Journal* 1–18.
- Kamburugamuve, S., Govindarajan, K., Wickramasinghe, P., Abeykoon, V. & Fox, G. 2017. Twister2: Design of a big data toolkit. Disertasi Ph.D. Indiana University, 1–18.
- Kim, G.-H., Trimi, S. & Chung, J.-H. 2014. Big-data applications in the government sector. *Communications of the ACM* 57(3): 78–85.
- Klievink, B., Romijn, B., Cunningham, S. & de Bruijn, H. 2017. Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers* 19(2): 267–283.
- Koronios, A. & Selle, S. 2015. Towards a process view on critical success factors in big data analytics projects. *Twenty-first Americas Conference on Information Systems* 1–14.
- Kuiler, W. Erik. 2014. From Big Data to Knowledge: An Ontological Approach to Big Data Analytics. *Review of Policy Research* 311–318.
- Laney, D. 2005. META Delta (February 2001).
- MAMPU. 2016. Analitis Data Raya (*Big Data Analytics*) Melonjak Kreativiti Dan Inovasi Penyampaian Perkhidmatan Kerajaan, 1-3.
- MAMPU. 2017. Mentransformasi penyampaian perkhidmatan : Program berpacuan data - Analitis Data Raya. *Seminar Kerajaan Digital Sarawak*. September *Slaid* 1-43.
- Analysis Model. *Journal of Advances in Information Technology* 7(3): 208–213.
- NIST Big Data Public Working Group. 2015. NIST special publication 1500-1 - NIST big data interoperability framework: Vol. 1, Definitions. *NIST Special Publication* 1: 32.
- NIST Big Data Public Working Group: Use Cases and Requirements Subgroup. 2015. NIST big data interoperability framework: Vol. 3, Use Cases and General Requirements 3: 260.
- Paul, S. & Duggal, P. S. 2016. Big data analysis : Challenges and solutions big data analysis : Challenges and solutions (December 2013), *International Conference on Cloud, Big Data and Trust 2013* 269-276.
- Pekeliling Transformasi Pentadbiran Awam (PTPA) Bil. 1 Tahun 2017 Pelaksanaan Analitis Data Raya Sektor Awam (aDRSA). MAMPU, 2017, 1-16.
- Raghupathi, W. & Raghupathi, V. 2014. Big data analytics in healthcare: promise and potential. *Health Information Science and Systems* 2(1): 1-10.
- Platform Analitis Data Raya Sektor Awam (aDRSA). 2017. Rangka kerja analitis data raya sektor

awam (aDRSA). Pekeliling transformasi pentadbiran awam (PTPA) Bil. 1 Tahun 2017
Pelaksanaan analitis data raya sektor awam (aDRSA). MAMPU, 2017, 5.

Zicari, R.V., Rosselli, M., Ivanov, T., Korfiatis, N., Tolle, K., Niemann, R. & Reichenbach, C.
2016. Setting up a big data project: Challenges, opportunities, technologies and
optimization, *big data optimization: Recent Developments and Challenges* 18: 17-47.

Copyright@FTSM