

# PELUANG DAN TANTANGAN BIG DATA DALAM PENELITIAN ILMU SOSIAL: SEBUAH KAJIAN LITERATUR

## *THE OPPORTUNITIES AND CHALLENGES OF THE BIG DATA IMPLEMENTATION IN SOCIAL SCIENCE RESEARCH: A LITERATURE REVIEW*

**Vience Mutiara Rumata**

Puslitbang APTIKA-IKP, Kementerian Komunikasi dan Informatika, Jln. Medan Merdeka Barat  
No. 9 Jakarta Pusat 10110, 021-3800418

[vien001@kominfo.go.id](mailto:vien001@kominfo.go.id)

(Diterima: 2/8/2016, Direvisi: 1/9/2016, Disetujui terbit: 30/11/2016)

### **Abstrak**

Memasuki era digital, segala data (informasi) mudah diakses, disimpan, ditelusuri, bahkan dianalisa. *Big Data* bukan sekadar tren yang dipopulerkan oleh kalangan eksklusif, melainkan sebuah penanda perubahan paradigma untuk memahami proses-proses sosial. Data yang dihasilkan di media sosial merupakan sebagian kecil dari “longsor bola salju” di era *Big Data*. Bagi kalangan akademisi, *Big Data* telah memberikan tantangan bagi peneliti sosial dengan adanya perubahan unit analisis dari manusia menjadi algoritma. Artikel ini membahas peluang dan tantangan *Big Data* bagi peneliti sosial dari sisi literatur yang berasal dari publikasi luar negeri seperti *Taylor & Francis*, *Elsevier*, *Sage publication*. Tujuannya adalah untuk memberikan pemahaman mendasar sekaligus mengisi kesenjangan pengetahuan terkait implementasi *Big Data* dalam penelitian bagi ilmuwan non ilmu komputer. Hasil kajian ini menemukan bahwa ada tiga peluang *Big Data*: munculnya terobosan multidisiplin ilmu “*computational social science*”; menjawab keterbatasan metode pengumpulan data konvensional baik kuantitatif maupun kualitatif; serta menawarkan kerangka analisis bagi ilmuwan sosiologi. Disamping peluang, ada tiga tantangan *Big Data*: pemahaman mendasar terkait data, isu metodologi dan teori serta isu etika penelitian.

**Kata kunci:** *Big Data*, Sosial, Penelitian, Peluang, Tantangan

### **Abstract**

*In the era of digital information, data can be accessed, recorded, trajected, and analysed conveniencely. Big Data is not solely a trend amongst the exculsive group, instead, it marks the paradigm swift particularly to undersand the social processes. Data, which generated on social media, is an avalanche of Big Data era. For the academics, Big Data challenges the social researchers by the changing of unit of analysis from human to algortihms. This article discusses the opportunities and pitfalls of Big Data in social science research. This is a literature review that derived mainly from journals published in prominent scientific publications such as Taylor & Francis, Elsevier, and Sage. The main purpose is to yield a basic knowledge and fill the gap for non-computer scientists, regarding the implementation of Big Data in the research. The result that Big Data has opportunities to deliver the multidiscipline field of study (computer, social and statistic); overcome the limitation of convesional methods of data gathering (both quantitative and qualitative); and offers an analytical framework for sociology field of study. Big Data imposes challenges such as common understanding of Big Data, methodology and theory issues, and research ethic debates.*

**Key words:** *Big Data, Social, Research, Opportunities, Challenges*

## **PENDAHULUAN**

Penetrasi internet dan media jejaring sosial terus meningkat dari tahun ke tahun baik di dunia bahkan di Indonesia. Sebanyak 42 persen populasi dunia atau sekitar 3,419 miliar penduduk telah memiliki akses terhadap internet per Januari 2016 (*we are social*, 2016). Jumlah

akun aktif *Facebook* secara global mencapai 1,79 miliar akun per kuartal ketiga 2016 (statista, 2016). Dalam kuartal tahun 2015, jumlah akun aktif *Twitter* mencapai 305 juta (statista, 2016). *Facebook* dan *Twitter* merupakan 10 media jejaring sosial terbesar secara global, disamping *WhatsApp* (urutan ke-2), *Tumblr* (urutan ke-7), dan *Instagram* (urutan ke-8).

Data infografis yang sama juga ditunjukkan bagi perkembangan internet di Indonesia. Jumlah pengguna internet di Indonesia mencapai 88,1 juta atau sekitar 34,9 persen dari populasi penduduk pada tahun 2014 (APJII, 2015). Meski termasuk negara berkembang dengan kesenjangan digital, penetrasi media jejaring sosial di Indonesia sangat signifikan. Indonesia menduduki urutan ke-empat pengguna *Facebook* di dunia - setelah Amerika Serikat, India, dan Brazil - dengan jumlah akun mencapai 60,3 juta dan diprediksi terus meningkat hingga 97,5 juta pengguna di tahun 2018 (Statista, 2015). Sementara, jumlah akun *Twitter* aktif di Indonesia mencapai 14.7 juta di tahun 2015, dan angka ini terus meningkat hingga 22.8 juta akun aktif di tahun 2019 (statista, 2016). Indonesia pernah didaulat sebagai “negara *Twitter*” oleh CNN Tech pada November 2010 (Lim, 2013:636). Bahkan, pengguna *Twitter* di Jakarta menyumbang 2,4 persen dari 10.6 miliar *tweets* secara global dari Januari-Maret 2013 (techinasia, 2013).

Meningkatnya penggunaan perangkat *mobile smart devices* (e.g. *smartphones*) serta terjangkaunya paket data internet dari operator merupakan faktor pendorongnya penetrasi ini. *We are social* (2015) mencatat orang Indonesia rata-rata menghabiskan kurang lebih 5 jam per hari untuk mengakses internet dengan komputer, serta kurang lebih 3 jam dengan menggunakan ponsel. Internet dan media jejaring sosial telah mengubah proses sosial secara mendasar bagi individu dan juga masyarakat secara keseluruhan.

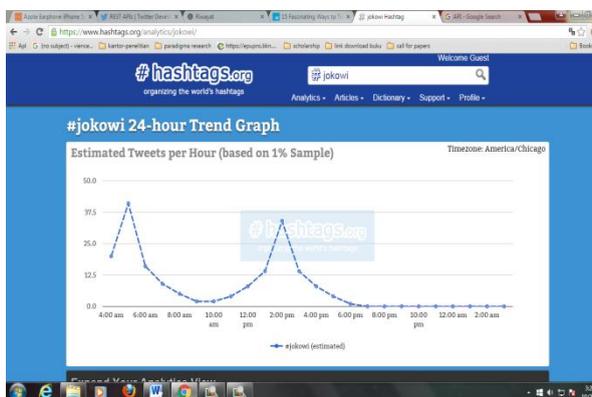
Interaksi antar manusia yang didominasi pada interaksi manusia dan komputer (*computer mediated communication/CMC*) ini telah mengubah tata cara perilaku (interaksi) sosial serta pemaknaannya. Memasuki transisi era informasi menuju era “*internet of things*”, segala perilaku manusia direkam, disimpan, bahkan dianalisa dalam bentuk data. Saat ini, *Facebook* menjadi sumber informasi berharga bagi lembaga komersil maupun non komersil, riset, serta

pemerintah untuk mempelajari pola-pola sosial (interaksi, hubungan, hingga mobilitas sosial). Data informasi seperti demografi, perilaku share dan memberikan tanda *like* pada *fanpage* tertentu, merupakan sebagian kecil dari jutaan informasi berharga lainnya yang dihasilkan *Facebook*. Tidak jarang, data yang dihasilkan *Facebook* dijadikan dasar kebijakan ataupun strategi dari lembaga yang berkepentingan. .

Kita memasuki tren yang disebut *Big Data*, era di mana ketersediaan data sosial yang terekam secara digital semakin berlimpah. Bagi kalangan industri atau praktisi, *big data* telah membuka peluang untuk menetapkan strategi bisnis serta inovasi dalam hal memproses, menganalisa dan menyimpan data dengan *volume* serta tingkat vitalitas yang tinggi secara cepat dan efektif. Bagi kalangan akademisi, *Big Data* telah menobrak tradisi lama penelitian ilmu sosial. *Big data* memberikan solusi bagi penelitian sosial konvensional, khususnya untuk menangkap realita seperti pola jaringan komunikasi, diseminasi informasi, atau bahkan memprediksi pola gerakan sosial atau politik berdasarkan perilaku secara *online*. Istilah-istilah seperti *webometrics*, *social network analysis*, *digital social research*, *web social science* atau *computational social science* menandakan transisi penelitian sosial konvensional penelitian sosial *cyber* atau “*e-research*”, yaitu di mana transisi unit analisis dari manusia menuju algoritma (Lupton, 2015:17).

Banyak pengembang aplikasi memanfaatkan *big data*. *Hastags.org*, misalnya, mampu menampilkan grafis topik *Twitter* yang diambil dari *Application Program Interface (API) streaming twitter* dalam kurun waktu tertentu (gambar 1.). *Wolframalpha*, salah satu aplikasi berbasis web yang menghasilkan laporan berbentuk *cluster* ataupun *mapping* jaringan teman tingkat satu maupun tingkat dua (jaringan temannya teman di *facebook*) hingga tempat-tempat yang pernah dikunjungi (*check-in* di *Facebook*).

Gambar 1. Grafis tren #Jokowi dalam kurun waktu 24 jam (29 Oktober 2015)



Sumber:

<https://www.hashtags.org/analytics/jokowi/>

Berlimpahnya data secara kuantitas serta beragam, menuntut sebuah mekanisme analisa data secara menyeluruh, atau setidaknya mewakili populasi *big data* tersebut. Meski aplikasi untuk menganalisa *big data* banyak tersedia di internet, tetapi akses terhadap *big data* itu sendiri sangat terbatas. *Facebook*, misalnya, telah mengubah kebijakan akses API-nya pada pertengahan tahun 2015 lalu. Hal ini dilakukan untuk melindungi data pengguna *Facebook* dari para pengembang aplikasi yang meraup keuntungan secara ekonomi (*techcrunch*, 2015).

Penelitian yang memanfaatkan *big data* masih belum populer di Indonesia, khususnya bagi kalangan peneliti sosial. Keterbatasan piranti lunak yang masih bergantung pada produksi luar negeri serta keterbatasan akses terhadap publikasi ilmiah *big data* dianggap menjadi salah satu penyebabnya. Kajian literatur ini diharapkan dapat memberikan pemahaman mendasar – khususnya topik yang menjadi perdebatan diantara kalangan akademisi sosial di luar negeri – bagi para ilmuwan sosial dalam negeri sebelum mereka mengaplikasikan *big data* di dalam penelitiannya. Untuk itu, pertanyaan penelitian pada artikel ini adalah “Apa peluang dan tantangan *Big Data* dalam penelitian sosial?” Karena cakupan *Big Data* sangat luas – tidak terbatas pada data

media sosial, tetapi juga mencakup bidang kesehatan, keuangan, dan sebagainya – maka kajian literatur ini ini memfokuskan pada *big data* yang bersumber pada media sosial. Makalah ini tidak membahas hal teknis yang berkaitan dengan teknologi pendukung *big data*, melainkan memberikan gambaran peluang dan tantangan untuk mengaplikasikan *big data* ke dalam prosedur ilmiah bidang sosial melalui sebuah kajian literatur dari jurnal-jurnal yang telah dipublikasikan dari sumber yang dapat diandalkan seperti *Taylor & Francis*; *Elsevier*; ataupun *Sage publication*.

Penelitian ini merupakan kajian literatur dengan pendekatan konseptual-tradisional. Kajian literatur adalah sebuah produk tulisan yang mengupas sebuah topik atau hasil penelitian yang telah dipublikasikan, tanpa ada gambaran metodologi ilmiah (Jesson *et al.*, 2011:10). Setidaknya ada dua pendekatan dalam menulis kajian literatur: pendekatan tradisional dan pendekatan sistematis (p.14). Pendekatan tradisional di dalam kajian literatur memiliki beberapa pendekatan seperti *critical approach*, *a conceptual review*, *a state-of-art review*, *an expert review*, *a scoping review* (p.15). Artikel ini menggunakan pendekatan konseptual-tradisional di mana mensitesiskan pengetahuan atau hasil penelitian di bidang atau topik tertentu, yang dalam ini adalah *Big Data*, dengan tujuan untuk memberikan pemahaman mendalam tentang topik tersebut.

Data primer dalam kajian literatur ini adalah jurnal-jurnal yang membahas peluang dan tantangan implementasi *big data*, yang dipilih secara subyektif oleh peneliti. Pemilihan jurnal tidak saja berdasarkan kesesuaian topik, tetapi juga publikasi serta sitasi dari jurnal-jurnal tersebut. Penulis merangkai kembali gagasan-gagasan penting dari jurnal-jurnal yang terpilih tersebut dengan cara yang logis, sistematis, serta argumentatif yang kritis. Argumentatif dimaksud adalah

mengedepankan pro dan kontra di kalangan akademisi di luar negeri terkait implementasi *Big Data* dalam penelitian sosial. Salah satunya, diskusi terkait prosedur penelitian *Big Data* yang masih menjadi perdebatan hingga sekarang.

**PEMBAHASAN**  
**“Big Data” secara Konseptual**

Pada tahun 2007, Yayasan Keilmuan Nasional Amerika Serikat (*the National Science Foundation*) mengeluarkan sebuah rencana aksi bertajuk “*Cyberinfrastructure Vision for 21<sup>st</sup> Century Discovery*” di mana salah satu rencana aksi tersebut adalah memanfaatkan *big data* dalam segala upaya mengembangkan ilmu, termasuk ilmu eksakta dan ilmu sosial (Hesse et al., 2015:17). Inilah yang menjadi titik awal perkembangan dan pemanfaatan *big data*. Kata “big data” secara harafiah berarti volume kuantitas data yang berlimpah (*terabytes* ataupun *petabytes*).

Namun definisi “*big data*” belum ada kesepakatan di antara akademisi maupun praktisi. Ada beragam definisi “*big data*”. Berman (2013:p.xx) berpendapat bahwa *big data* harus memenuhi “3Vs”: *volume* (jumlah kuantitas); *variety* (beragam bentuk: dokumen, rekaman suara, gambar, video, dsb); *velocity* (perubahan data yang cepat karena sifatnya berasal dari *multiple sources*). Hashem dan kolega (2015:100) berpendapat bahwa *Big Data* harus memenuhi unsur “4’s” dengan menambahkan “value” yang merupakan faktor penting dari *big data*, yaitu proses menemukan makna dibalik sekumpulan data. Sementara, produsen perangkat keras dan lunak komputer global, IBM menambahkan kategori “*veracity*” merujuk pada kepastian keakuratan data yang tersedia (IBM, 2016).

Secara garis besar, *big data* tersebar di ruang elektronik (salah satunya: internet) yang berasal dari beragam sumber yang kompleks di mana proses pengerjaannya dilakukan secara bertahap (pengumpulan, pengorganisasian, penyimpanan hingga

analisa) serta membutuhkan sumber daya yang tidak murah. Karena begitu besar skala dan kompleks prosedurnya, *big data* dikelompokkan menjadi lima kategori, yaitu: 1) *data sources*; 2) *content format*; 3) *data stores*; 4) *data staging*; 5) *data processing* (tabel 1.). Media sosial diakategorikan *big data* karena merupakan sumber data atau informasi yang dapat dibagikan (*share*) atau ditukarkan (*exchange*) antar individu atau kelompok (*communities*) melalui URL. Khusus data media sosial, Batrinca dan Treleven (2015:93) mengklasifikasikannya menjadi dua jenis: *historic data sets* (data yang sebelumnya telah diakumulasikan dan tersimpan) dan *real-time data* (data yang diambil secara langsung tanpa ada jeda waktu tertentu, *e.g. live feeds*).

Tabel 1. Kategori Big Data (Hashem et al., 2015:102)

Klasifikasi	Varian	Deskripsi
<i>Data source</i>	a. Media sosial	Sumber informasi melalui laman URL yang dihasilkan dari pertukaran informasi dan gagasan dalam jaringan komunitas virtual (Facebook, Twitter, Blogs, dsb)
	b. <i>Machine-generated data</i>	Data/ informasi yang dihasilkan dari perangkat keras maupun lunak
	c. <i>Sensing</i>	Alat pendeteksi dan pengukur kuantitas fisik dan mengubahnya ke dalam sinyal
	d. Transaksi	Data transaksi seperti transaksi keuangan atau data kerja dengan melibatkan dimensi waktu
	e. IoT	Data atau informasi yang berisi obyek-obyek dalam internet
<i>Content format</i>	a. Data terstruktur	Data berbasis format SQL ( <i>Structured Query Language</i> ) yang siap diolah dan disimpan.
	b. Data semi terstruktur	Data yang terstruktur, tetapi tidak diorganisir di dalam standar model database konvensional
	c. Data tidak terstruktur	Data yang berisi informasi yang tidak ter kategorisasi seperti pesan teks, lokasi, video, bahkan konten

		yang ada di dalam media sosial.
<b>Data Stores</b>	a. Data dokumen	Penyimpanan dan pengambilan data berupa data dokumen seperti JSON, XML ataupun format PDF atau MS Words.
	b. Data kolom	Penyimpanan data berdasarkan kategorisasi kolom yang sama. Data berorientasi kolom ini berbeda dari penyimpanan data <i>BigTable</i>
	c. Infografis database	Penyimpanan dan pengambilan data grafis yang terdiri dari <i>nodes</i> dan <i>edges</i> (e.g. Neo4j)
	d. <i>Key-value</i>	Alternatif model database untuk menyimpan dan mengakses data <i>key-value</i> dengan skala yang tidak terbatas. Model ini biasanya mendukung transaksi multi-akses dengan single-akses (e.g. <i>Apache Hbase</i> , <i>Apache Cassandra</i> ataupun <i>Voldemort</i> )
<b>Data Staging</b>	a. Pembersihan ( <i>cleaning</i> )	proses pemisahan data lengkap dengan data yang tidak lengkap atau data yang tidak masuk akal
	b. Transformasi ( <i>transform</i> )	proses transformasi bentuk data yang siap diolah, misalnya dari aplikasi <i>RStudio</i> yang fungsinya <i>scraping</i> data twitter dan mengubahnya ke dalam <i>wordcloud</i> , ataupun data berbasis <i>.csv excel</i>
	c. Normalisasi ( <i>normalization</i> )	proses penstrukturan data untuk menghindari redundansi data
<b>Data Processing</b>	a. <i>Batch</i>	Sistem berbasis MapReduce yang memudahkan penskalaan kluster-kluster yang berisi ribuan nodes.
	b. <i>Real-time</i>	S4 merupakan contoh pemrosesan big data secara real-time di mana teknologi ini dapat memudahkan

		programer untuk mengembangkan aplikasi
--	--	--

Disamping berbagai potensi yang luar biasa yang dimiliki *big data* untuk kegiatan penelitian sosial, berbagai kritikan juga datang dari kalangan akademisi. Di dalam jurnalnya, Boyd dan Crawford (2012) mengajukan beberapa pertanyaan yang menjadi bahan pertimbangan bagi para peneliti sosial terkait *big data*. Apakah *big data* dan aplikasi analisisnya dapat membuat kita memahami fenomena sosial di suatu masyarakat, atau justru mendorong sebuah cara baru untuk ‘menyerang’ privasi? Apakah *big data* mentransformasikan cara kita mempelajari pola komunikasi dan budaya manusia, atau justru mempersempit makna penelitian itu sendiri? Boyd dan Crawford (2012) berpendapat bahwa *big data* hanyalah sebuah fenomena budaya, teknologi dan ilmiah (akademis) yang merupakan interaksi timbal balik antara teknologi, analisis (identifikasi pola jaringan), serta mitos. Mitos di sini adalah sebuah kepercayaan yang diterima oleh masyarakat luas bahwa *big data* merupakan produk pengetahuan intelijen yang sebelumnya tidak mungkin dicapai oleh masyarakat global (p.663).

### **Peluang Big Data bagi Penelitian Sosial**

Dalam simposium “*A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*”, Boyd dan Crawford (2012) berpendapat bahwa *big data* telah menciptakan sebuah sistem pengetahuan baru yang mengubah obyek pengetahuan. Tidak hanya itu, big data mampu menganalisa, menggambarkan serta memprediksikan pola komunikasi, perilaku bahkan isu-isu sosial dan non-sosial seperti kejahatan, pola penyebaran penyakit, dan sebagainya (Tene dan Polonetsky, 2013).

Big data telah mentransformasi paradigma konvensional ilmu sosial menuju apa yang disebut “*computational*

*social science*” yang memampukan manusia mengamati dunia yang begitu kompleks serta lingkungan yang dinamis seperti saat ini (Chang et al., 2014:71). Bagi Mason dan koleganya (2014) *Computational social science* bukan sekadar paradigma, melainkan sebuah bidang penelitian yang meneliti persinggungan antara ilmu komputer, ilmu statistik, dan ilmu sosial, di mana metode komputasi digunakan untuk menjawab permasalahan yang berkaitan dengan masyarakat (p.257). Berbeda dengan *Computational social science*, kata lain yang juga populer di kalangan akademisi adalah “*social computing*” yang lebih menekankan pada desain sistem pendukung untuk memfasilitasi interaksi sosial manusia, termasuk mempelajari bagaimana dan alasan manusia memproduksi konten (Mason et al., 2014).

Secara metafora, Berman (2013:p.xxvi) berpendapat bahwa *big data* telah menjadi obyek pusat rotasi bagi kalangan akademisi, peneliti, korporat (bisnis), bahkan regulator karena informasi yang disediakan oleh penyedia *big data* bersifat permanen dan berkembang dari segi kuantitas dan kualitas. Dengan menggunakan data yang sama, siapa saja dapat mengkritisi serta mengkaji teknik analisis yang sudah ada sehingga menemukan sebuah formulasi yang efektif dan pada akhirnya menjadi sebuah pengetahuan baru. Senada dengan Berman, Hesse dan koleganya (2015) menyakini bahwa *big data* dapat ditransformasikan menjadi pengetahuan bagi para peneliti sosial setidaknya dalam dua cara. Pertama, metode komputasi yang digunakan untuk mengintegrasikan serta menganalisa *big data* merupakan fondasi pengetahuan dalam ilmu sosial. Kedua, peneliti ilmu sosial harus memahami bagaimana mengakses, mengolah, menginterpretasikan hingga memproduksi pengetahuan dari *big data* demi perkembangan tidak saja ilmu sosial itu semata tetapi multi disiplin ilmu (p.18). Karenanya, komunitas peneliti sosial harus bekerja sama dengan peneliti

dari disiplin keilmuan lain untuk meninjau kembali peran teori-teori, publikasi-publikasi ilmiah, data yang dibagikan (*data sharing*), tantangan hak cipta, akumulasi pengetahuan serta akuntabilitas publik untuk menjawab tantangan metodologi dan epistemologi dari penelitian *big data*.

*Big data* membuka peluang yang baik bagi peneliti sosial di kala metode penelitian konvensional, baik itu kualitatif maupun kuantitatif (*survei*), memiliki keterbatasan dalam memahami fenomena sosial secara luas atau menjawab tantangan-tantangan sosial yang memberikan dampak yang luas bagi masyarakat. Anderson (2008) secara ekstrim mengklaim bahwa era *big data* – atau apa yang disebutnya sebagai *The Petabyte Age* – telah membuat prosedur penelitian konvensional menjadi tidak relevan.

*“This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear ... Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.”* ([www.wired.com/](http://www.wired.com/) diakses pada 4 Mei 2016)

Chang dan koleganya (2014:73) berpendapat bahwa *big data* dapat memberikan hasil generalisasi yang jauh lebih besar dibandingkan pada hasil survei semata yang mengandalkan sampel populasi. Survei, lebih lanjut, merupakan pelengkap data empiris yaitu berupa respons dari para partisipan. Jadi, peneliti sosial dapat memperkaya nilai temuan surveinya (atau metode pengumpulan data konvensional lainnya) dengan analisa dari

ribuan bahkan jutaan *tweet* atau status *post Facebook* yang dihasilkan setiap hari. Contoh kasus, studi yang dilakukan oleh Kim dan koleganya (2016) ketika mempelajari sentimen publik dengan menganalisa hampir 30 juta konten *tweet* yang mengandung kata “nuklir” (per 1 Oktober 2010 hingga 30 September 2013). Untuk menganalisanya, Kim dan koleganya menggunakan piranti lunak konten analisis non parametrik milik perusahaan swasta Crimson Hexagon, *Forsight* (p.432). Secara kuantitatif, angka 30 juta data *Tweet* jauh lebih berharga dibandingkan ribuan data sampel populasi sebuah survei. Akan tetapi, apakah besaran kuantitatif ini mencerminkan opini publik keseluruhan populasi sebenarnya? Hal ini akan secara lebih lengkap di bahas pada subab “tantangan *big data* bagi peneliti sosial”.

Bagi ilmuwan Sosiologi Budaya, *big data* dapat menganalisa hubungan antara elemen-elemen budaya ataupun mekanisme pembentuk makna (*meaning*) dalam tingkat makro ataupun meso yang menjadi pertanyaan yang belum terjawab hingga saat ini (Bail, 2014:468-469). Analisa tingkat makro ataupun meso ini merupakan strategi yang diperlukan dalam studi-studi yang menggunakan pendekatan *field theory*, yang banyak diadopsi dalam studi Sosiologi Budaya. Pendekatan teori ini sulit untuk mengidentifikasi interelasi ataupun “ruang sosial” yang tercipta di antara pelaku-pelaku sosial dan elemen-elemen budaya ketika menghasilkan sebuah struktur atau pola budaya tertentu. Teknologi *big data* tentu mampu menyerap serta memetakan seluruh teks yang ada di ruang digital lebih mudah. Semisal, perubahan pola perilaku penetrasi sosial dari kelompok sesama jenis dapat dianalisa dari konten *Twitter* dari waktu ke waktu. Dengan teknologi *big data* memudahkan peneliti untuk menganalisa hubungan antara pelaku-pelaku sosial dan jaringannya tanpa perlu menghabiskan waktu yang lama. Hal ini tentu menimbulkan perdebatan. Di satu sisi, para peneliti etnografi menghabiskan waktu

yang lama untuk mempelajari sebuah komunitas yang kecil. Sementara di sisi lain, peneliti ilmu komputer hanya menghabiskan waktunya di depan komputer untuk mempelajari jejak digital manusia di internet. Manovich (2012:467) berpendapat bahwa jutaan data *Tweet* yang dikumpulkan oleh peneliti ilmu komputer tidak berarti tidak sedalam data kualitatif peneliti etnografi. Yang membedakan kedua peneliti ini adalah akses terhadap data serta desain riset (pertanyaan penelitian, kerangka pemikiran, metode dan sebagainya). Jadi dengan *big data*, peneliti tidak perlu menyoalkan jumlah data ataupun kedalaman data.

*Big data* juga memunculkan cabang ilmu baru, salah satunya “Digital Humanities”. Meski belum dikategorikan sebagai disiplin ilmu, *Digital Humanities* merupakan sebuah kajian baru yang tengah berkembang untuk mengeksplorasi keterlibatan manusia dengan perkembangan teknologi, media dan metode komputasi (Svensson, 2010). Dalam presentasinya di konferensi *Modern Language Association Convention* pada tahun 2011, Stephen Ramsay (2013:240-241) berpendapat bahwa seorang pakar Digital Humanis adalah orang yang harus memiliki kemampuan koding.

### **Tantangan *Big Data* bagi Penelitian Sosial**

Disamping peluang, *big data* memberikan tantangan yang cukup signifikan bagi para peneliti sosial. *Big data* cenderung menyebabkan “apophenia”, yaitu kondisi di mana peneliti melihat adanya pola, atau hubungan dari dari sesuatu yang sebenarnya tidak ada (Boyd dan Crawford, 2011). Setidaknya beberapa pertanyaan kritis muncul di kalangan peneliti sosial maupun studi media, seperti: Apa arti data yang tergolong *big data*? Siapa yang memiliki akses terhadap *big data* tersebut? Bagaimana data tersebut dianalisa? (Boyd dan Crawford, 2012:664). Berdasarkan penelusuran jurnal-jurnal yang secara khusus membahas *big data* untuk

penelitian sosial, peneliti menyimpulkan setidaknya tiga tantangan yang mengemuka yakni pemahaman mengenai *big data*; isu metodologi dan landasan teori; serta isu etika penelitian.

Hal yang paling mendasar dan menjadi perdebatan di kalangan akademisi adalah kategori *big data* yang dibutuhkan oleh peneliti sosial. Pemahaman kita mengenai “data” berubah dengan hadirnya TIK. Dulu, data dikumpulkan untuk tujuan yang jelas dan oleh institusi yang resmi (e.g. sensus penduduk, data pemilih tetap untuk pemilihan umum, dsb). Sekarang, data adalah sekumpulan data orang-orang di dalam sebuah jaringan yang dapat kita akses dengan mudah di dunia maya. Bahkan, data tersebut dimiliki atau dapat diakses oleh sedikit perusahaan komersil atau korporat berskala global yang berbasis di Amerika Serikat (e.g. *Facebook*, *Google* dan sebagainya).

Bila data tersebut diklasifikasikan sebagai *public goods* artinya tersedia secara publik, mengapa akses terhadap data tersebut sangat terbatas bagi kalangan akademisi maupun pemerintah, tetapi tidak bagi perusahaan komersil seperti *Gnip* (perusahaan agregat data media sosial yang berbasis di Colorado, AS)? Gitelman dan Jackson (2013) berpendapat bahwa adanya kecenderungan “data” dimaknai sebagai fakta yang tanpa disadari dan telah diterima (oleh kita) menjadi sebuah kebenaran tanpa perlu dipertanyakan atau dicari penjelasannya. Perlu dikritisi bahwa data atau informasi yang tersaji dalam kerangka media sosial seperti *Facebook* (e.g. data sociodemografi atau interaksi sosial) merupakan hasil rekayasa publik yang bersifat “imajinatif” (Boyd, 2010).

Perubahan radikal yang dilakukan *big data* dalam ilmu pengetahuan ataupun prosedur riset (Boyd dan Crawford, 2012:665) menimbulkan isu-isu dalam prosedur penelitian, khususnya metodologi. Jumlah data yang tidak terbatas bukan berarti *big data* menyajikan data yang terbaik dibandingkan data yang didapat dari teknik pengumpulan

konvensional. Apakah orang yang menggunakan *twitter* dapat mewakili publik tertentu? Apakah akun-akun aktif dalam media sosial serta konten yang dihasilkan dari akun-akun tersebut dapat mewakili ‘orang’?

Hargittai (2015:64) menemukan bahwa penelitian *big data* memiliki keterbatasan dalam metode kerangka sampling (*sampling frames*) baik variabel dependen (variabel yang diukur) maupun independen (variabel yang dapat dimanipulasi atau dikendalikan oleh peneliti). Metode sampling dalam *big data* seperti media sosial sulit diterapkan karena ketidakpastian sumber data itu sendiri. Perlu dipahami bahwa akses publik terhadap *big data* sangat terbatas. Hanya beberapa perusahaan komersil seperti *Gnip* dan *Data Sift* yang memiliki akses yang penuh terhadap data mentah perusahaan media sosial raksasa seperti *Facebook* maupun *Twitter* (Batrinca dan Treleaven, 2015:95).

Perusahaan penghasil *big data* seperti *Facebook* dan *Twitter* telah membatasi publik untuk mengakses *Application Programming Interface* (API) mereka. *Twitter* pernah membuka akses terhadap peneliti melalui *Twitter Data Grants Program*, tetapi program tersebut telah ditutup. Saat ini, *Twitter* hanya membuka akses kira-kira sebesar 1% dari jutaan *tweets* tiap harinya secara gratis. Meskipun berbayar, seseorang bisa mengakses API *twitter* sebesar 10% secara berbayar melalui ‘*Gardenhose*’. Karena teknik pengumpulan *big data* mengandalkan aplikasi piranti lunak, peneliti tidak memiliki daya untuk mengatur data yang terjaring secara otomatis dengan aplikasi tersebut. Tentu, berapapun jumlah *big data* yang dijadikan sampel dalam penelitian, tentu tidak dapat mewakili seluruh populasi. Karena itu, hasil penelitian *big data* tidak dianjurkan untuk digeneralisasikan.

Disamping kelemahan metode sampling, tantangan prosedur penelitian *big data* lainnya adalah belum memadainya

teori atau teknik untuk mempelajari perilaku manusia dan memaknai teks melalui *big data* (Bail, 2014:467). Peneliti *big data* cenderung kehilangan informasi penting ketika menganalisa perilaku sosial dari penggunaan media sosial (Hargittai, 2015:65). Ketika peneliti mencari hubungan antara komunikasi dan ikatan/hubungan pengguna *Facebook* dengan peluang mendapatkan pekerjaan baru, ada kemungkinan hasil penelitiannya menjadi bias akibat peneliti tidak mempertimbangkan kemungkinan pengguna *Facebook* menggunakan media sosial lainnya (e.g. *LinkedIn*) atau faktor internal lainnya seperti motivasi yang tidak diperhitungkan di dalam penelitian tersebut (p.66). Karenanya, peneliti harus berhati-hati dalam menarik kesimpulan dari temuannya. Manovich (2012:465) berpendapat bahwa informasi (konten) yang disajikan melalui media sosial bukanlah seperti jendela yang transparan yang betul-betul mencerminkan di dalam benak atau diri penggunanya. Tetapi, informasi tersebut merupakan sebuah hasil konstruksi presentasi publik yang sifatnya rekayasa.

Ada yang bisa dilakukan peneliti dengan teknologi *big data*, tetapi ada yang tidak bisa dilakukan dengan *big data*. Pola interaksi atau jaringan sosial atau pertumbuhan kelompok sosial mungkin bisa dihasilkan oleh teknologi *big data*, tetapi tidak bisa menjelaskan lebih dalam faktor penyebab pola tersebut. White dan Breckenridge (2014) berpendapat bahwa *big data* menawarkan 'trade-off' dalam penelitian ilmu sosial, yaitu sebatas alat yang berguna bagi ilmu, tetapi tidak dapat menjawab topik atau pertanyaan penelitian sosial (p.336).

*Google Flu Trends* (GFT), sebagai contoh, merupakan *web service* milik *Google* yang berisi informasi penyebaran virus influenza di lebih dari 25 negara. Sebagai penyedia mesin pencarian (search engine) berbasis internet berskala global, *Google* berupaya untuk memprediksikan penyebaran virus influenza (dan juga

demam berdarah) berdasarkan perilaku pencarian informasi mengenai flu tersebut di *Google* (Lazer dan Kennedy/ wired.com, 2015). GFT ini bahkan diklaim dapat melampaui kemampuan prediksi Pusat Pengendalian dan Pencegahan Penyakit milik pemerintah Amerika Serikat (Centers for Disease Control and Prevention/ CDC). Akan tetapi, GFT ini gagal memprediksikan wabah flu pada tahun 2013. Apa yang dapat dipelajari dari kasus ini adalah bahwa GFT hanya mampu menyajikan data statistik trend kata atau informasi yang dicari mengenai influenza, tetapi tidak mampu menjelaskan hubungan yang terkait antara pencarian informasi tersebut dengan penyebaran dari flu itu sendiri (Harford, 2014). Hal yang biasa terjadi dalam analisa *big data* adalah minim teori serta mencari korelasi data kuantitatif, dimana kemungkinan bias sangat bisa terjadi.

Tantangan ketiga yang dihadapi oleh peneliti sosial bahwa *big data* membawa perubahan radikal mengenai etika penelitian (Boyd dan Crawford 2012:665). Etika sangat diperlukan dalam penelitian yang melibatkan manusia sebagai obyek penelitinya. Namun, prosedur untuk mendapatkan persetujuan dari pemilik data cenderung diabaikan dalam penelitian *big data*. Karena pengumpulan data dilakukan secara otomatis melalui piranti lunak, maka data apa saja yang tersimpan di dunia maya dapat dengan mudah ditarik (e.g. transaksi belanja online, email, bahkan GPS). Tentu ini berpotensi melanggar privasi si pemilik data.

Tene dan Polonetsky (2013:242) berpendapat bahwa *Big Data* menantang prinsip-prinsip cakupan dari 'data pribadi' atau apa yang disebut dengan 'personally identifiable information' (PII); konsep 'data minimization'; persetujuan (kontrol individu); serta hak akses individu. Dalam pengumpulan data konvensional, peneliti diwajibkan untuk melampirkan lembar persetujuan dari para responden (e.g. surat persetujuan/*consent letter*). Karena

prosedur *big data* mengandalkan aplikasi teknologi, surat pernyataan kesediaan atau persetujuan dari pemilik data tampaknya bukan hal yang penting.

Data pribadi dan privasi di ruang internet menjadi perdebatan di era *big data*. Apakah data pribadi yang tersedia dalam ruang publik seperti internet dikategorikan sebagai ‘data publik’? Jika ya, mengapa akses data tersebut terbatas bagi perusahaan komersil semata? Jika tidak, apakah seorang peneliti berhak mengakses dan mengumpulkan data pribadi di internet demi perkembangan ilmu pengetahuan? Siapa yang akan bertanggung jawab bila tindakan tersebut dapat menciderai hak privasi seseorang?

*Big data* membutuhkan sebuah prosedur yang memadai untuk mengurangi potensi pelanggaran privasi tersebut. Untuk meminimalisir pelanggaran hak privasi, seseorang dapat menggunakan cara “de-identifikasi” atau menyamarkan identitas pemilik data sehingga tidak dapat dilacak atau ditentukan subyeknya (e.g. nama samaran, singkatan, enkripsi, koding, dan sebagainya). Akan tetapi, cara ini belum memiliki prosedur yang baku serta tidak berkekuatan hukum.

Crawford dan Schultz (2014) mengusulkan “*procedural data due process*” yakni sebuah prosedur pengumpulan dan pengolahan data yang menegakkan prinsip keadilan atau “the right of due process” seperti dalam sistem hukum. Bila prinsip ini diterapkan dalam proses *big data*, maka peneliti harus tata cara atau prosedur yang memiliki kekuatan hukum yang tetap untuk menjaga keadilan (*fairness*) bagi pemilik data pribadi dalam penelitiannya.

Untuk menjalankan prosedur “*due process*” ini, peran regulator sangat dibutuhkan untuk menghasilkan sebuah produk hukum yang melindungi data dan pemanfaatannya. Di Amerika Serikat, prinsip *Fair Information Practice Principles* (FIPPs) merupakan pedoman yang dikembangkan oleh Komisi Perdagangan Amerika Serikat (Fair Trade

Commission - FTC) sejak 1970an. Saat ini, Kementerian Komunikasi dan Informatika Republik Indonesia tengah menyusun draft RUU Perlindungan Data Pribadi sejak 2015 dan akan menjadi agenda Program Legislasi Nasional DPR 2016. RUU ini merupakan payung hukum untuk perlindungan data pengguna internet. Perlu diingat bahwa “*Due Process*” dalam sistem hukum membatasi wewenang para regulator yang terlibat dalam menegakkan prinsip supremasi hukum (e.g. DPR berwenang mengesahkan peraturan, hakim berwenang untuk menegakkan peraturan dan sebagainya). Di dalam sistem *big data*, “*Due Process*” ini sulit diterapkan karena ketiadaan regulasi yang mengatur interaksi antara designer aplikasi, algoritma, output dari komputer itu sendiri, ataupun pemilik data. Ini yang akan menjadi tantangan Kemenkominfo didalam penyusunan RUU tersebut.

## **KESIMPULAN DAN SARAN**

Sebagaimana yang diutarakan oleh Manovich (2012:464) bahwa siapa saja yang mengandalkan alat atau metode komputasi untuk mempelajari isu-isu sosial dan budaya manusia, harus paham betul bahwa ada hal-hal yang mungkin bisa dipraktikkan tetapi tidak sejalan dengan prinsip, atau sebaliknya. *Big Data* memberikan peluang sekaligus tantangan bagi peneliti sosial. Peluang tersebut diantaranya (1) *Big Data* memiliki peluang melahirkan paradigma pengetahuan baru serta terobosan penelitian multidisiplin ilmu - ilmu komputer, ilmu statistik, dan ilmu sosial atau dikenal dengan sebutan “*computational social science*”; (2) *Big data* menjawab keterbatasan metode pengumpulan data konvensional baik itu kualitatif dan kuantitatif. Jadi, peneliti sosial dapat memperkaya nilai temuan surveinya (atau metode pengumpulan data konvensional lainnya) dengan analisa dari *Big Data*; (3) *Big Data* menawarkan solusi bagi ilmuwan sosiologi budaya, khususnya, ketika menganalisa tingkat makro maupun meso antara pelaku-pelaku sosial dan

elemen-elemen budaya yang selama ini sulit dilakukan dengan metode konvensional.

*Big Data* memiliki tantangan yang harus diperhatikan bagi peneliti sosial. *Big Data* cenderung menyebabkan “apophenia”, yaitu kondisi di mana peneliti melihat adanya pola, atau hubungan dari dari sesuatu yang sebenarnya tidak ada (Boyd dan Crawford, 2011). *Big Data* masih menjadi perdebatan yang sengit dikalangan ilmuwan. Setidaknya ada tiga topik perdebatan terkait *big data*. Pertama, tantangan terkait definisi terminologi *big data* itu sendiri. Jika definisi big data itu menekankan pada kuantitas (jumlah), maka seberapa banyak data yang dibutuhkan untuk menjawab isu-isu sosial? Tantangan kedua adalah metodologi dan teori. White dan Breckenridge (2014) berpendapat bahwa *Big Data* menawarkan ‘trade-off’ dalam penelitian ilmu sosial. *Big data* dapat menunjukkan pola komunikasi atau interaksi sosial secara online, lintas geografis, kultur dan sosial. Tetapi *Big Data* tidak bisa menjelaskan latar belakang – motivasi misalnya – dari interaksi tersebut. Diperlukan sebuah terobosan untuk mengintegrasikan metode konvensional dan *Big Data* untuk menjawab jurang metodologi antara keduanya. Yang terakhir, *Big Data* masih terganjal isu etika penelitian. Big Data telah meniadakan prosedur pengumpulan data ilmiah konvensional. Dengan menggunakan alat komputasi, data dikumpulkan melalui internet dengan cepat dan mudah. Akan tetapi, hal ini mengundang perdebatan terkait data sebagai *public goods*. Bila data yang tersimpan di internet dikategorikan sebagai *public goods*, mengapa akses terhadap data tersebut sangat terbatas dan hanya dimiliki bagi perusahaan komersil berskala global. Isu selanjutnya terkait hal ini adalah privasi dan kepemilikan data itu sendiri. Apakah data *public goods* yang tersedia di internet (termasuk informasi pribadi) dapat diakses, dikumpulkan dan dianalisa oleh pihak

tertentu? Bukankah ini sebagai bentuk pelanggaran privasi?

#### DAFTAR PUSTAKA

- Anderson, C. (2008) The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. <http://www.wired.com/2008/06/pb-theory/> (diakses pada 4 Mei 2016)
- APJII (2015) *Profil Pengguna Internet Indonesia 2014*. Asosiasi Penyelenggara Jasa Internet Indonesia. Jakarta
- Bail, C. A. (2014) The Cultural Environment: Measuring Culture with Big Data. *Theor Soc.* 43(2014):465-482
- Batrinca, B. dan Treleaven, P. C. (2015) Social Media Analytic: A Survey of Techniques, Tools and Platforms. *AI & Soc.* 30 (2015): 89-116.
- Berman, J.J. (2013). *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Morgan Kaufmann. USA.
- Boyd, D. (2010) *Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications dalam Networked Self: Identity, Community, and Culture on Social Network Sites* (ed. Zizi Papacharissi). Pp.39-58. Routledge
- Boyd, D., dan Crawford, K. (2011) Six Provocations for Big Data. Presentasi paper pada Oxford Internet Institute simposium “A decade in Internet Time: Symposium on the Dynamic’s of the Internet and Society, 21 September 2011.
- Boyd, D., dan Crawford, K. (2012) Critical Questions for Big Data. *Information, Communication & Society.* 15 (5) : 662-679. DOI: 10.1080/1369118X.2012.678878
- Chang, Ray M., Kauffman, Robert J., dan Kwon, YoungOk (2014) Understanding the Paradigm Shift to Computational Social Science in the Presence of Big Data. *Decision Support Systems.* 63(2014):67-80.

- Crawford, K., dan Schultz, J. (2014) Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms. *Boston College Law Review*. vol. 55, pp.93-1287
- Gitelman, L. dan Jackson, V. (2013) Introduction dalam *Raw Data is an Oxymoron* (editor Lisa Gitelman). Massachusetts Institute of Technology. Cambridge. Amerika Serikat.
- Harford, T. (2014) Big data: are we making a big mistake? Diakses pada 5 Juni 2016 dari website Financial Times <http://www.ft.com/intl/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html/>
- Hargittai, E. (2015) Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *ANNALS, AAPSS*. 659, May 2015: 63-76
- Hashem, Ibrahim A.T., Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, dan Samee Ullah Khan (2015) The Rise of “Big Data” on Cloud Computing: Review and Open Research Issues. *Information Systems*. 47 (2015):98-115
- Hesse, Bradford W., Moser, Richard P. dan Riley, William T. (2015) From Big Data to Knowledge in the Social Science. *ANNALS, AAPSS*. 659:May 2015.
- IBM (2016) The Four V's of Big Data. Diakses pada 27 April 2016 dari website IBM. <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>.
- Jesson, J. K., Matheson, L, dan Lacey F.M. (2011) *Doing Your Literature Review: Traditional and Systematic Techniques*. SAGE publication Ltd. California. USA.
- Kim, J.; D. Brossard; D. A. Scheufele; dan M. Xenos. “Shared” Information in the Age of Big Data: Exploring Sentiment Expression Related to Nuclear Energy on Twitter. *Journalism & Mass Communication Quarterly*. 93(2):430-445.
- Lazer, D. dan Kennedy, R. (2015) What We Can Learn From the Epic Failure of Google Flu Trends. Diakses pada 5 Juni 2016. <http://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/>
- Lupton, D. (2015) *Introduction: Life is Digital dalam Digital Sociology*. Routledge. New York.
- Mason, Wi.; Vaughan, J.W.; dan Wallach, H. (2014) Computational Social Science and Social Science. *Mach Learn*. 95 (2014):257-260. DOI 10.1007/s10994-013-2456-8
- Manovich, L. (2012) Trending: The Promises and the Challenges of Big Social Data dalam *Debates in the Digital Humanities* (ed. Matthew K. Gold). University of Minnesota Press. Minneapolis.
- Ramsay, S. (2013) Who’s in and Who ‘s out dalam *Defining Digital Humanities: A Reader* (ed.Melissa Terras, Julianne Nyhan, dan Edward Vanhoutte). Ashgate Publishing Limited. England
- Statista (2016) Number of Monthly Active Facebook Users Worldwide as of 3<sup>rd</sup> Quarter 2016. Diakses pada 11 November 2016 dari website Statista. <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- Statista (2015) Number of Facebook users in Indonesia from 2012 to 2018 (in millions). Diakses pada 28 Oktober 2015 dari website Statista. <http://www.statista.com/statistics/304829/number-of-facebook-users-in-indonesia/>
- Statista (2016) Number of monthly active Facebook users worldwide as of 4th quarter 2015 (in millions) Diakses pada 26 April 2016 dari website Statista. <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>

- Statista (2016) Statistics and facts about Twitter. Diakses pada 26 April 2016 dari website Statista. <http://www.statista.com/topics/737/twitter/>
- Statista (2016) Number of Twitter users in Indonesia from 2014 to 2019 (in millions). Diakses pada 26 April 2016 dari website Statista. <http://www.statista.com/statistics/490548/twitter-users-indonesia/>
- Svensson, P (2010) The Landscape of Digital Humanities. Volume 4 Number 1. Diakses pada 19 Juli 2016. <http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html#>
- Techcrunch (2015) 'Facebook Is Shutting Down Its API For Giving Your Friends' Data To Apps'. Diakses pada 30 Oktober 2015 <http://techcrunch.com/2015/04/28/facebook-api-shut-down/#.teofye:nael>
- Techinasia (2013) Indonesia is Social: 2.4% of World's Twitter Posts Come From Jakarta. Diakses pada 26 April 2016 <https://www.techinasia.com/indonesia-social-jakarta-infographic>
- Tene, O., dan Polonetsky, J. (2013) Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern University School of Law*, 11(5):239-272
- We Are Social (2016) *Digital in 2016 report*. <http://wearesocial.com/sg/special-reports/digital-2016> (diakses pada 12 November 2016)
- White, P., dan Breckenridge, R. S. (2014) Trade-Offs, Limitations, and Promises of Big Data in Social Science Research. *Review of Policy Research*, 31(4):331-338
- Wolfram|Alpha Personal Analytics for Facebook. Diakses pada 30 Oktober 2015. <http://www.wolframalpha.com/facebook/>