

**LAPORAN PENELITIAN DOSEN**



**STUDI PENERAPAN *LINKED OPEN DATA*  
UNTUK DISEMINASI DATA *OFFICIAL STATISTICS***

**Takdir, SST, M.T.  
Farid Ridho, SST, M.T.**



**PUSAT PENELITIAN DAN PENGABDIAN MASYARAKAT  
POLITEKNIK STATISTIKA STIS  
2018**



## **LEMBAR PENGESAHAN**

Laporan Penelitian Dosen dengan judul:

### **Studi Penerapan *Linked Open Data* Untuk Diseminasi *Data Official Statistics***

**Nama Peneliti:**

Takdir, SST, M.T.

Farid Ridho, SST, M.T.

Dilaksanakan pada November 2017 sampai dengan November 2018

Telah disahkan oleh Kepala Pusat Penelitian dan Pengabdian Masyarakat (PPPM)  
Sekolah Tinggi Ilmu Statistik (STIS), pada tanggal **27 November 2018**

Menyetujui,  
Kepala UPPM

Ketua Peneliti

Setia Pramana, Ph.D.  
NIP. 197707222000031002

Takdir, SST, M.T.  
NIP. 198704142010121001

Mengetahui  
Wakil Direktur I

Dr. Hardius Usman  
NIP. 196704251989011002



# **STUDI PENERAPAN *LINKED OPEN DATA* UNTUK DISEMINASI DATA *OFFICIAL STATISTICS***

*(Study of Linked Open Data Implementation for  
Data Dissemination of Official Statistics)*

**Takdir<sup>1</sup>, Farid Ridho<sup>2</sup>**

Politeknik Statistika STIS<sup>1</sup>

Politeknik Statistika STIS<sup>2</sup>

E-mail: [takdir@stis.ac.id](mailto:takdir@stis.ac.id)<sup>1</sup>, [faridr@stis.ac.id](mailto:faridr@stis.ac.id)<sup>2</sup>

## **ABSTRAK**

Di era dimana transparansi dan integrasi data menjadi tuntutan, tiap instansi, khususnya pemerintahan, berlomba-lomba menyediakan data statistik yang dimilikinya untuk dapat diakses oleh pengguna data secara online. Dengan akses data yang melimpah, tuntutan pengguna data pun meningkat dari yang hanya sekedar availability, meluas menjadi kebutuhan akan kemudahan penggabungan/integrasi, eksplorasi, visualisasi hingga analisis dan simulasi data dari berbagai sumber. Metode diseminasi data konvensional, yakni dalam bentuk tabel, sulit mengakomodir kebutuhan tersebut dikarenakan metadata, yang mengandung makna dari data, terpisah dengan raw data. Teknologi Linked Data memungkinkan data dan metadata terhubung secara digital sehingga mesin dapat "mengetahui" informasi metadata dari sebuah data. Dengan demikian, keterkaitan antara suatu data dan data lainnya dapat diperoleh dengan mudah sehingga tahapan-tahapan pada proses pengolahan data yang melibatkan banyak sumber dapat direduksi. Penelitian ini mempelajari penerapan Linked Data pada data official statistics yang dirilis oleh BPS dengan akses terbuka melalui website BPS. Teknologi open data, berupa perangkat lunak, yang digunakan pada penelitian ini diperoleh dari berbagai komunitas pengembang software linked data yang disesuaikan dengan karakteristik data official statistics. Hasil penelitian ini diharapkan memberikan cetak biru (*blueprint*) linked open data bagi instansi pemerintahan di Indonesia.

**Kata kunci:** *linked data, open data, official statistics*

## **ABSTRACT**

*In this era when transparency and data integration become a necessity, organizations, especially governments, are competing to make their official statistics accessible by online users. Huge amount of online data accesses creates an increasing users' demand from data availability only to easier data join/integration, exploration, visualization, until analysis and simulation from various data sources. Conventional data dissemination, i.e. table format, is difficult to accommodate those requirements because the metadata, which contains the meaning of data, is separated from raw data. Linked Data technology enables data and metadata are digitally connected that makes machines could "understand" metadata information of a data. Therefore, the relationships between different data sets can be traced with little effort cutting down steps in data processing which involves large resources. This study discovered Linked Data implementations for official statistics published by BPS with open access through BPS website. Open data technologies, in form of software, used in this study are taken from several linked data communities and developers which are adjusted to meet official statistics characteristics. The results are expected to provide blueprint of linked open data for government institutions in Indonesia.*

**Keywords:** *linked data, open data, official statistics*

## PENDAHULUAN

Badan Pusat Statistik (BPS) adalah lembaga pemerintah yang bertugas menghasilkan data statistik yang akan digunakan oleh pemerintah dalam merencanakan pembangunan, selain itu data statistik juga digunakan sebagai dasar pertimbangan dalam membuat kebijakan publik. Data statistik ini juga banyak digunakan oleh masyarakat, swasta dan juga peneliti. Untuk menghasilkan data statistik ini, BPS melakukan kegiatan statistik yang diantaranya adalah pengumpulan data, pengolahan data, tabulasi data, analisis data, pembuatan publikasi dan penyebarluasan data (diseminasi data).

Dalam rangka memenuhi kebutuhan data dan informasi statistik dan amanat UU No. 16 Tahun 1997 tentang Statistik, BPS telah menerbitkan Surat Keputusan Kepala BPS Nomor 5 Tahun 2000 tentang Sistem Statistik Nasional (SSN). Salah satu upaya BPS untuk mewujudkan SSN antara lain melakukan koordinasi dan kerjasama dengan instansi pemerintah dan masyarakat, baik di pusat maupun daerah, serta dengan lembaga-lembaga internasional. Koordinasi dan kerjasama dimaksud dilaksanakan atas dasar kemitraan dengan tetap mengantisipasi serta menerapkan perkembangan ilmu pengetahuan dan teknologi, khususnya teknologi informasi dan komunikasi. Jejaring tersebut merupakan kekuatan yang terus dikembangkan dalam rangka pembangunan nasional di bidang statistik. Upaya pengembangan yang dilakukan BPS telah menghasilkan beragam data dan indikator sosial-ekonomi, yang dihasilkan melalui Sensus Penduduk (SP), Sensus Pertanian (ST), Sensus Ekonomi (SE), dan berbagai survei di bidang sosial-ekonomi dan kependudukan. Semua indikator sosial-ekonomi tersebut telah dapat dihasilkan secara reguler maupun dalam waktu tertentu.

Masyarakat menuntut ketersediaan data dan informasi statistik yang beragam, rinci, mudah dipahami, dan tepat waktu. Tuntutan kebutuhan data dan informasi statistik tersebut belum sepenuhnya terpenuhi, namun secara bertahap terus diupayakan ketersediaannya. BPS menyampaikan hasil kegiatan statistik kepada masyarakat melalui press release Berita Resmi Statistik (BRS) yang meliputi informasi statistik penting seperti inflasi, ekspor, impor, pariwisata, pertumbuhan ekonomi, produksi padi dan palawija, angka kemiskinan, dan pengangguran. Selain itu, BPS juga melakukan diseminasi data dan informasi statistik baik melalui media cetak maupun elektronik. Diseminasi statistik yang dilakukan oleh BPS menjadi tugas dari Direktorat Diseminasi Statistik yang berada di bawah Kedepuyan Metodologi dan Informasi Statistik.

Jenis diseminasi statistik yang ada di BPS sangat beragam baik dalam bentuk media cetak maupun elektronik. BPS melakukan diseminasi statistik secara berkala, ada publikasi yang dirilis setiap bulan, triwulan, tahunan dan lainnya. Salah satu media publikasi BPS yang digunakan untuk diseminasi statistik adalah melalui web [www.bps.go.id](http://www.bps.go.id). Data yang terdapat di situs web BPS cukup beragam mulai dari sosial, kependudukan, ekonomi, perdagangan sampai dengan data pertanian dan pertambangan. Format data yang terdapat di web BPS juga cukup beragam seperti

Selama beberapa tahun terakhir, beberapa inisiatif muncul untuk meningkatkan antarmuka antara warga negara dan pemerintah melalui penggunaan Teknologi Informasi dan Komunikasi (TIK) yang efektif, dan secara khusus melalui penggunaan basis standar Web. Villazón-Terrazas, dkk (2011) menyajikan ringkasan dari upaya-upaya yang dilakukan untuk membantu pemerintah dalam penggunaan teknologi dan Web untuk mengimplementasikan e-Government dengan mengelola data mereka secara transparan dan efisien.

Sejak 2008 Kegiatan e-Government W3C mempromosikan beberapa *chapter* untuk membantu pemerintah untuk mengikuti praktik terbaik dan pendekatan untuk meningkatkan penggunaan Web. Saat ini, kegiatan ini termasuk Kelompok Minat e-Government dan Kelompok Kerja Data Terkait Pemerintah.

Sejak awal ditemukannya teknologi web, yakni world wide web, oleh Tim Berners Lee pada tahun 1989 sebagai media untuk menyebarkan informasi melalui internet, penggunaannya sangat masif dan memberikan pengaruh signifikan baik di kalangan bisnis maupun pemerintahan. Berbagai layanan dapat diberikan secara online seperti akses terhadap berita dalam berbagai macam format (teks, audio, video), layanan pengiriman pesan, dan layanan penyempadan file pada web. Dampak dari hal tersebut adalah saat ini terdapat sejumlah besar data dan informasi yang tersimpan pada web. Web adalah sebuah ruang digital yang terbentuk dari sejumlah perangkat komunikasi yang saling terhubung melalui internet yang memuat sejumlah data dalam format digital.

Salah satu kelemahan teknologi web konvensional, yang juga dikemukakan oleh Tim Berners Lee, adalah belum adanya mekanisme dan format standard yang menjadi rujukan untuk mendefinisikan metadata dari data yang tersimpan pada web. Akibatnya, informasi relasi antar data yang berguna untuk penelusuran data yang memiliki keterkaitan juga tidak tersedia (Bizer, Heath, & Berners-Lee, 2009). Teknologi Linked Data yang prinsip-prinsipnya diperkenalkan pada tahun 2006 oleh Tim Berners Lee ditujukan untuk mengakomodir hal tersebut (Berners-Lee, 2006). Linked Data memberikan panduan untuk membangun sistem metadata dari data yang tersimpan pada web dengan format yang standar sehingga dapat diproses oleh mesin yang akan memudahkan penelusuran data terkait. Linked Data menjadikan web sebuah penyimpanan data yang saling terhubung satu sama lain oleh informasi metadata yang kemudian dikenal dengan semantic web (Herman, 2001).

Struktur informasi pada web yang berupa *hypertext* dan *hyperlink* yang saling terkait menjadi dasar pembentukan teknologi linked data. Oleh karena itu, sejumlah panduan yang ada mengacu pada data struktur tersebut. Di sisi lain, data statistik memiliki ciri khas tersendiri dimana sebagian besar berupa numerik dan kategorik dengan dimensi data yang tinggi (Kämpgen & Harth, 2011) (Hausenblas, Halb, Raimond, Feigenbaum, & Ayers, 2009). Terlebih lagi, pengguna data statistik memiliki kebutuhan untuk melakukan operasi agregasi terhadap data, seperti penjumlahan dan rata-rata. Oleh karena itu, perlu pendekatan khusus untuk mendiseminasikan data statistik agar dapat memenuhi kebutuhan dasar pengguna data dan tetap kompatibel dengan teknologi linked data. Data statistik akan memberikan lebih banyak manfaat apabila dipublikasikan dalam bentuk linked data yang memungkinkan pengguna untuk menggabungkan data dari sumber yang berbeda untuk menggali informasi baru serta menyajikan data yang memiliki kualitas, kelengkapan, dan tingkat rincian yang lebih baik (Do et al., 2014).

Berbagai inisiatif linked data untuk statistik (linked statistical data) telah dimulai oleh sejumlah peneliti dan telah melahirkan berbagai macam *prototype* seperti SDMX (Capadislis, Auer, & Ngonga Ngomo, 2015) dan Scovo (Hausenblas et al., 2009). Di Indonesia sendiri, BPS menempati peringkat 27 dunia dan peringkat 2 Asia Tenggara pada The Open Data Inventory (<http://odin.opendatawatch.com>) 2017 dengan cakupan/coverage menempati peringkat 14 dan openness (<http://opendefinition.org/od/2.1/en/>) pada peringkat 43. Teknologi linked data memiliki peran penting untuk meningkatkan openness suatu open data. Tim Berners-Lee juga mendefinisikan 5 level open data dimana peringkat 2 peringkat tertinggi (4 dan 5) adalah penerapan linked data (Berners-Lee, 2006). Penelitian ini mengelaborasi hasil penelitian terkait linked open data serta melakukan kustomisasi yang diperlukan untuk diterapkan pada open government data di Indonesia. Hasil penelitian

diharapkan dapat menjadi cetak biru penerapan linked open data di Indonesia untuk meningkatkan level open data Indonesia.

## TINJAUAN PUSTAKA

### Open Data

Opendatahandbook.org mendefinisikan open data sebagai data yang dapat digunakan secara bebas, dapat digunakan kembali, dan didistribusikan ulang oleh siapa saja dan harus mengikuti aturan atau persyaratan ketika akan digunakan untuk berbagi.

Definisi ringkas yang paling penting dari kata terbuka adalah sebagai berikut:

*Ketersediaan dan Akses:* data harus tersedia secara keseluruhan dan biaya yang diperlukan untuk produksi ulang wajar, data dapat diunduh melalui internet. Data juga harus tersedia dalam bentuk yang nyaman dan dapat dimodifikasi.

*Penggunaan Kembali dan Redistribusi:* data harus disediakan menurut ketentuan yang mengizinkan penggunaan ulang dan redistribusi termasuk penggabungan dengan dataset lain.

*Partisipasi Universal:* setiap orang harus dapat menggunakan, menggunakan kembali dan mendistribusikan kembali - tidak boleh ada diskriminasi terhadap bidang usaha atau terhadap orang atau kelompok. Misalnya, pembatasan 'non-komersial' yang akan mencegah penggunaan 'komersial', atau pembatasan penggunaan untuk tujuan tertentu (misalnya hanya dalam pendidikan), tidak diizinkan.

### Linked Open Data (LOD)

Teknologi web memungkinkan kita untuk menghubungkan suatu dokumen dengan dokumen lainnya. Dengan menerapkan teknologi web kita dapat menghubungkan data atau dokumen yang memiliki keterkaitan. Istilah linked data merujuk pada cara yang digunakan untuk memublikasikan dan menghubungkan data yang terstruktur pada web. Wikipedia mendefinisikan *linked data* sebagai istilah yang digunakan untuk menjelaskan cara terbaik untuk membagikan dan menghubungkan bagian-bagian dari data, informasi dan pengetahuan pada web semantik dengan menggunakan URIs dan RDF.

Berners-Lee dalam Bizer (2009) menjelaskan tentang aturan yang digunakan untuk memublikasikan data di web agar semua data yang dipublikasikan dapat menjadi bagian dari satu ruang data global, aturan tersebut adalah:

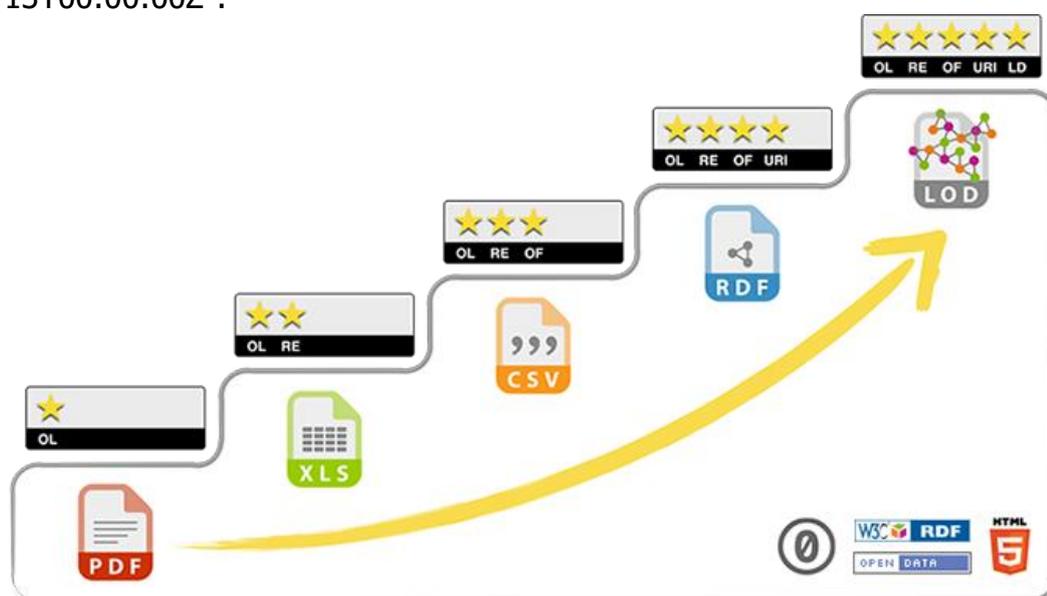
1. Gunakan URI sebagai nama untuk berbagai hal
2. Gunakan HTTP URI sehingga orang dapat mencari nama-nama itu
3. Ketika seseorang mencari URI, berikan informasi yang berguna, menggunakan standar (RDF, SPARQL)
4. Sertakan tautan ke URI lain, sehingga mereka dapat menemukan lebih banyak hal

5stardata.info membuat contoh dari penerapan level open data dengan menampilkan data tentang ramalan cuaca 3 hari ke depan untuk Kota Galway yang berada di Irlandia. Pada level 1 contoh data yang ditampilkan dapat diunduh dalam format PDF. Pada level 2 data yang ditampilkan sudah dalam bentuk xls, format data proprietary yang digunakan pada Microsoft Excel. Pada level 3 data sudah menggunakan format yang lebih umum dan tidak terkait pada format proprietary 5stardata.info memberikan contoh dalam format csv. Pada level 4 data sudah ditampilkan ke dalam dokumen web dan memiliki struktur URIs

yang dapat berisi hirarki data. Apabila kita melihat kode sumber halaman ini akan terlihat dokumen tersebut memiliki struktur hirarki.

```
<div id="data" about="#Galway" typeof="meteo:Place">
<tr rel="meteo:forecast" resource="#forecast20101113">
<div about="#forecast20101113">
    <span    property="meteo:predicted"    content="2010-11-13T00:00:00Z"
datatype="xsd:dateTime" >Saturday, 13 November 2010</span>
</div>
```

Dari struktur kode sumber ini kita dapat mengetahui beberapa fakta seperti "#Galway" adalah suatu tempat (Place), terdapat prediksi cuaca untuk kota "#Galway" yang terdapat pada tag "#forecast20101113", dan "#forecast20101113" adalah prediksi cuaca untuk "2010-11-13T00:00:00Z".



**Gambar 1** Level open data menurut Tim Berners-Lee  
 Sumber: [www.5stardata.info](http://www.5stardata.info)

## Official Statistic LOD

Official Statistics menjadi data publik suatu negara yang sangat penting. Data ini merupakan bukti empiris yang dibutuhkan dalam pengambilan keputusan untuk membuat kebijakan publik atau dalam melakukan suatu riset di bidang ekonomi. Statistik menjadi bagian yang tidak terpisahkan dari dunia modern saat ini. Beberapa indikator statistik di bidang sosial dan ekonomi dapat menggambarkan besar populasi, inflasi, perdagangan dan pengangguran.

Saat ini statistik dihasilkan dalam format basis data relasional. Data mentah yang dikumpulkan dari responden ketika disimpan sudah dalam kondisi bersih setelah melalui proses validasi. Data ini disimpan dalam tabel ke dalam bentuk data agregat untuk menjamin data yang dipublikasikan tidak menampilkan data individu dari responden yang bersifat konfidensial. Data agregat disimpan dan biasanya disebar dalam *data cubes*.

Statistik untuk publikasi dapat disebar menggunakan berbagai media. Publikasi yang masih umum adalah dengan menggunakan media cetak berupa kertas, walaupun secara de facto publikasi menggunakan web lebih tepat menjadi sarana diseminasi. Data dapat

disebarkan menggunakan halaman HTML dan dokumen PDF. Format data yang ada di web biasanya menggunakan tabel, grafik, visualisasi, teks deskriptif dan kubus.

Tabel 1 Contoh format open data bintang-5 untuk diseminasi statistik

| Rating | Makna                                       | Format             |
|--------|---|--------------------|
| ★      | Ada di web, lisensi terbuka                 | Dokumen PDF        |
| ★★     | Sda, Dalam format terstruktur               | Sheet Excel        |
| ★★★    | Sda, Tidak dalam format proprietary         | CSV, SDMX          |
| ★★★★   | Sda, Dapat diobservasi dan memiliki URIs    | RDF/XML            |
| ★★★★★  | Sda, Terdapat tautan menuju dataset lainnya | RDFa dengan tautan |

Tabel2 menunjukkan upaya Cyganiak dkk (2011) dalam mengidentifikasi level diseminasi data untuk berbagai lembaga di dunia. Daftar ini memuat cakupan data dan bagaimana cara mengakses suatu data yang disebarluaskan.

Tabel 2. Penyedia utama data statistik di dunia

| Penerbit   | Cakupan data   | Cara Akses          |
|------------|--|---------------------|
| Eurostat   | Ekonomi, populasi, industri, transportasi, dll               | Jelajah, unduh      |
| Factbook   | Sejarah, manusia, pemerintah, geografi, dll                  | Jelajah, unduh      |
| IMF        | Keuangan   | Jelajah, unduh      |
| OECD       | Bantuan, ekonomi, pendidikan kesehatan, pengangguran, dll.   | Jelajah, unduh      |
| UN         | Kriminal, pendidikan, energi, lingkungan, tenaga kerja, dll. | Jelajah, unduh      |
| WHO        | Kesehatan  | Jelajah             |
| World Bank | Pertanian, infrastruktur, tenaga kerja, pendidikan, dll.     | Jelajah, unduh, API |

Sumber: Cyganiak, R., dkk. (2011). *Official Statistics and the Practice of Data Fidelity*.

### Tantangan penerapan LOD untuk official statistics

LOD untuk official statistics memiliki karakteristik tersendiri karena selain ditujukan untuk diakses mesin/komputer, juga diperuntukkan untuk diakses oleh manusia atau pengguna data (Capadisli, 2012). Selain itu, pada umumnya publikasi official statistik didiseminasikan dalam bentuk raw data berupa tabel dan kolom pada file comma separated value (CSV) ataupun microsoft excel (xls) yang minim akan informasi metadata. Struktur data official statistik yang berdimensi tinggi serta dapat memiliki nilai turunan merupakan tantangan tersendiri untuk dipublikasikan dalam bentuk LOD yang terstruktur. **Gambar 1** berikut menunjukkan contoh format publikasi data official statistics.

Angka Partisipasi Murni (APM) menurut Tipe Daerah, Jenis Kelamin dan Jenjang Pendidikan, 2009-2013

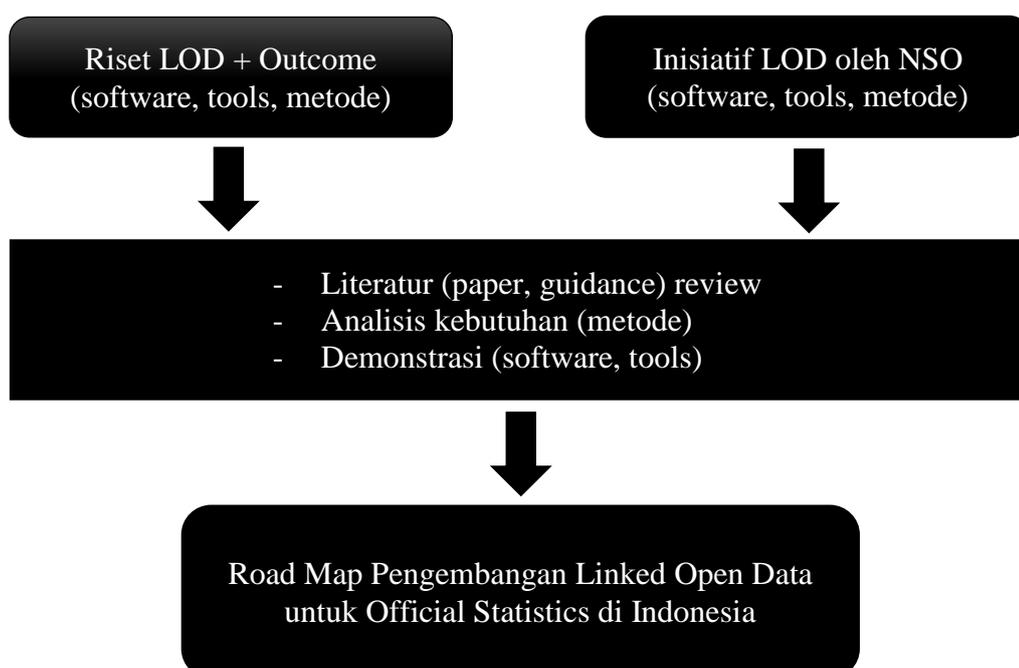
| Daerah tempat<br>Tinggal | Jenis Kelamin         | SD    |       |       |       |       |                    |       |       |       |       |       |       |
|--------------------------|-----------------------|-------|-------|-------|-------|-------|--------------------|-------|-------|-------|-------|-------|-------|
|                          |                       | APM   |       |       |       |       | APM+ <sup>*)</sup> |       |       |       |       |       |       |
|                          |                       | 2009  | 2010  | 2011  | 2012  | 2013  | 2009               | 2010  | 2011  | 2012  | 2013  | 2009  | 2010  |
| Perkotaan                | Laki-laki             | 94,70 | 95,64 | 91,62 | 92,39 | 96,11 | 94,70              | 95,67 | 91,70 | 92,46 | 96,22 | 69,32 | 69,85 |
|                          | Perempuan             | 94,07 | 94,28 | 90,09 | 92,17 | 95,58 | 94,07              | 94,33 | 90,19 | 92,23 | 95,65 | 68,68 | 69,30 |
|                          | Laki-laki + Perempuan | 94,40 | 94,98 | 90,89 | 92,28 | 95,85 | 94,40              | 95,02 | 90,97 | 92,35 | 95,94 | 69,00 | 69,58 |
| Perdesaan                | Laki-laki             | 94,32 | 94,06 | 91,35 | 92,60 | 95,22 | 94,32              | 94,10 | 91,44 | 92,66 | 95,25 | 64,64 | 64,54 |
|                          | Perempuan             | 94,38 | 94,99 | 90,62 | 92,50 | 95,25 | 94,38              | 95,02 | 90,71 | 92,56 | 95,30 | 67,59 | 67,43 |
|                          | Laki-laki + Perempuan | 94,35 | 94,50 | 91,00 | 92,55 | 95,23 | 94,35              | 94,54 | 91,09 | 92,61 | 95,28 | 66,04 | 65,90 |
| Perkotaan+<br>Perdesaan  | Laki-laki             | 94,49 | 94,79 | 91,48 | 92,50 | 95,63 | 94,49              | 94,83 | 91,56 | 92,57 | 95,71 | 66,75 | 66,96 |
|                          | Perempuan             | 94,24 | 94,65 | 90,37 | 92,34 | 95,41 | 94,24              | 94,69 | 90,46 | 92,41 | 95,47 | 68,10 | 68,33 |
|                          | Laki-laki + Perempuan | 94,37 | 94,72 | 90,95 | 92,43 | 95,52 | 94,37              | 94,76 | 91,03 | 92,49 | 95,59 | 67,40 | 67,62 |

Sumber: BPS RI - Susenas, 2009-2012

Gambar 1. Contoh publikasi data official statistics

### Kerangka pikir

Riset dan inisiatif LOD oleh National Statistical Office (NSO) yang menyediakan outcome berupa software, tools, ataupun metode merupakan objek dari penelitian ini. Outcome tersebut akan dielaborasi melalui literatur review untuk memahami konsep yang ditawarkan, kemudian metodologi yang diusulkan dianalisis untuk menyesuaikan kebutuhan LOD untuk perstatistikan di Indonesia. Selain itu, demonstrasi teknis software dan tools untuk menerapkan LOD dilakukan untuk melihat maturity dan visibilitasnya dalam menrancang roadmap pengembangan LOD untuk Official Statistics di Indonesia. Visualisasi kerangka pikir dapat dilihat pada **Gambar 2** berikut.

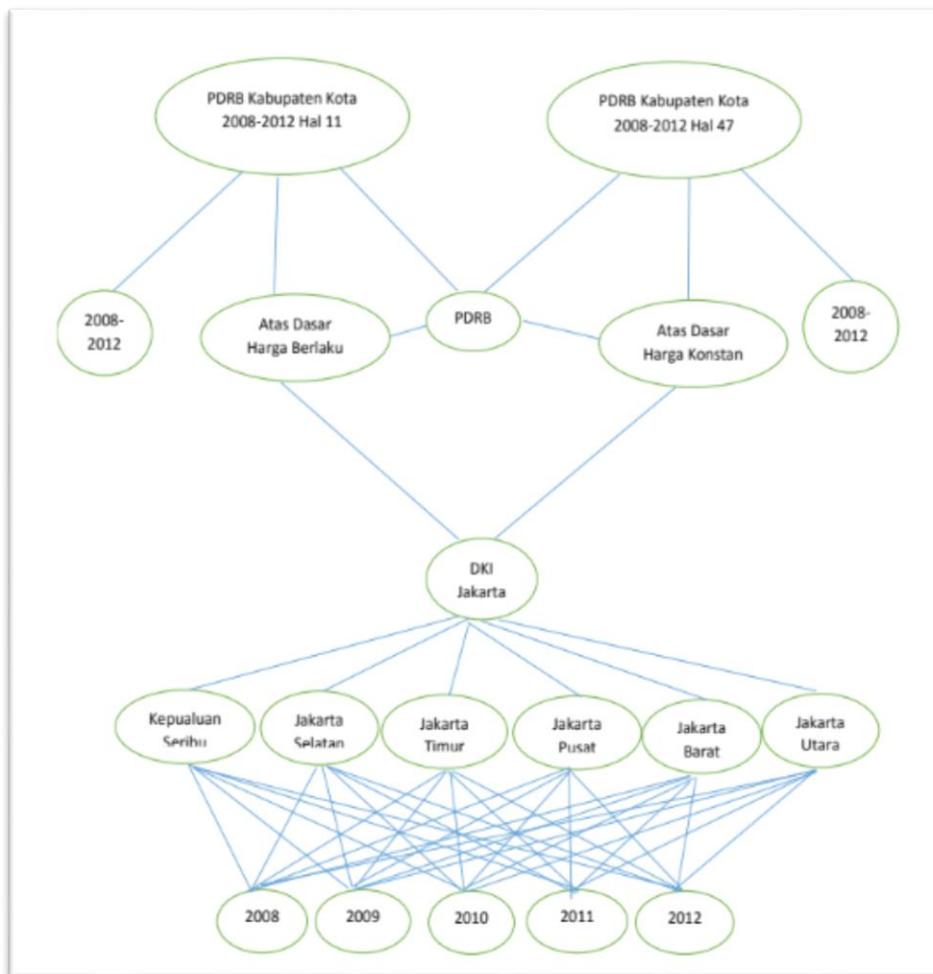


Gambar 2. Bagan Kerangka Pikir

### Penelitian Terkait

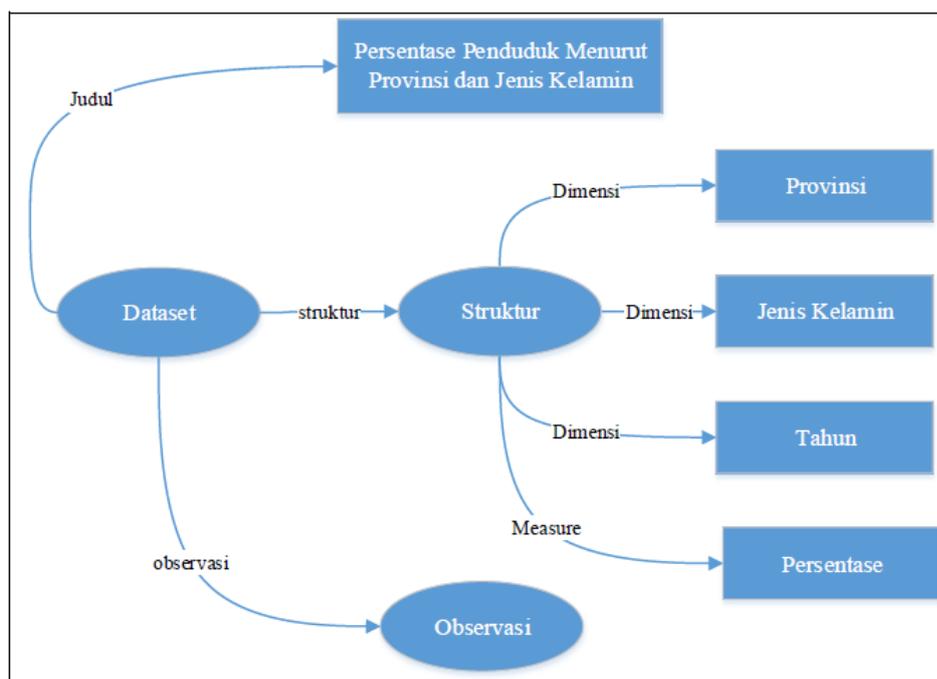
Meskipun panduan tentang bagaimana membuat Linked Data sudah diinisiasi oleh beberapa pihak, namun panduan yang ada belum mencakup keseluruhan proses mulai dari sumber data sampai pemilihan publikasi. Penelitian yang ada juga belum memberikan penjelasan yang rinci tentang semua langkah ini dan teknologi yang digunakan. Villazón-Terrazas, dkk (2011) menyusun panduan metodologi untuk menerbitkan Linked Data pada pemerintah. Penelitian ini mengusulkan panduan metodologi untuk semua aktifitas yang ada pada proses pembuatan Linked Data pada pemerintah. Panduan ini adalah hasil dari pengalaman mereka dalam menghasilkan Linked Data di beberapa instansi pemerintah.

Di Politeknik Statistika STIS juga telah terdapat penelitian berupa skripsi yang memperkaya literasi desain linked open data untuk official statistics di Indonesia, diantaranya adalah struktur data BPS yang diusulkan oleh Fakhriyanto tahun 2015 dimana peneliti mencoba membuat hirarki struktur publikasi di BPS (lihat **Gambar 3**).



**Gambar 3.** Struktur data pada publikasi BPS. Sumber: Fakhriyanto, 2015

Selain itu, adapula penelitian dari Adrian yang telah mencoba mengimplementasikan LOD untuk data BPS dan menyusun struktur metadata berdasarkan beberapa publikasi yang diamati pada website BPS (lihat **Gambar 4**). Struktur disusun dengan mengacu pada Statistical Core Vocabulary (SKOVO).



**Gambar 4.** Struktur LOD BPS berbasis SKOVO. Sumber: Adrian, 2016

Kedua penelitian tersebut dapat menjadi dasar desain struktur metadata LOD untuk statistik pemerintahan di Indonesia. Namun, dengan kemunculan dan perkembangan inisiatif LOD dari berbagai riset dan NSO, perlu dilakukan eksplorasi yang mendalam, baik untuk mengupdate desain yang telah ada, maupun untuk melihat metode dan tools yang dapat diadopsi dengan modifikasi dan penyesuaian yang minim. Penelitian ini akan fokus pada ekplorasi tersebut.

## METODOLOGI

Penelitian ini melakukan pengamatan pada sejumlah inisiatif linked open data, khususnya yang memuat data official statistics. Objek penelitian difokuskan pada sistem aplikasi/perangkat lunak yang digunakan atau dikembangkan untuk mempublikasikan data official statistics. Terdapat berbagai pihak yang mengembangkan sistem aplikasi berbasis linked data, baik dari private sector maupun public sector.

Untuk menemukan hasil riset yang sesuai dengan karakteristik di atas, beberapa kata kunci (keywords) digunakan dalam menelusuri literatur maupun produk/tools linked statistical data yang dapat diakses melalui internet, seperti "linked statistical data", "linked data for statistics", dan "RDF statistical data", dimana. Pada umumnya outcome riset yang berupa aplikasi dapat diunduh melalui website yang disediakan atau pada web repository source code publik seperti Github. Halaman demonstrasi juga disediakan untuk dicoba oleh pengunjung.

### Metode Analisis

Literatur review dilakukan dengan target utama adalah penelitian yang memiliki deliverable berupa perangkat lunak yang mendukung linked data yang bersifat opensource sehingga dimodifikasi. Data official statistics diperoleh dari website BPS, kementerian dan lembaga di Indonesia, serta inisiator open data, seperti portal data.id yang diinisiasi oleh Sekretariat Negara, serta sumber data online dari organisasi internasional, seperti World Bank, United Nations Economic Commision for Europe (UNECE), dan United Nations Statistics Division (UNSD).

Selain mendalami dokumentasi riset, outcome riset yang berupa aplikasi diujicoba untuk memahami secara teknis fitur yang disediakan dan visibilitasnya untuk digunakan dengan data statistik pemerintahan di Indonesia. Ujicoba dilakukan melalui halaman live demo yang disediakan ataupun mengunduh dan menginstall aplikasi jika live demo tidak disediakan. Adapun tools yang tidak menyediakan live demo dan installer untuk diujicoba penggunaannya, eksplorasi dilakukan dengan mengamati user manual serta video demo aplikasi.

Berikut adalah variable yang menjadi pertimbangan utama dalam mengevaluasi software linked statistical data pada penelitian ini adalah:

1. Dukungan terhadap format standard linked data
2. Skalabilitas
3. Kemudahan penggunaan
4. Tampilan (Graphical User Interface)
5. Reliabilitas

Dukungan terhadap format standard linked data yang ditetapkan oleh World Wide Web Consortium (W3C) menjadi prioritas pertama karena merupakan syarat utama agar interoperabilitas antar-sistem yang meng-hosting linked data tetap ada. Hal ini untuk menghindari setiap sistem memiliki metadata yang tidak standard yang dapat mengakibatkan silo system yang sangat dihindari pada penerapan linked data.

Pertimbangan kedua adalah skalabilitas dari aplikasi ataupun skalabilitas dari teknologi yang digunakan oleh aplikasi. Sebuah server linked data akan memuat beragam jenis data yang jumlahnya besar serta melayani query yang banyak karena akan dirujuk oleh system lain yang terkait. Oleh karena itu, skalabilitas merupakan hal yang penting jika sistem ditujukan untuk digunakan pada data official statistics yang memiliki banyak target pengakses.

SPARQL merupakan mekanisme query standard untuk mengakses linked data. Struktur syntax SPARQL yang kompleks sulit dipahami pengguna data untuk mengeksplorasi data set dari berbagai sumber linked data yang saling terkoneksi. Oleh karena itu, kemudahan penggunaan, yang juga erat kaitannya dengan tampilan (graphical user interface), menjadi variable selanjutnya yang dievaluasi. Sedangkan realibilitas sistem ditempatkan pada pertimbangan terakhir dengan tujuan akan diakomodir pada roadmap pengembangan sistem.

## **HASIL DAN PEMBAHASAN**

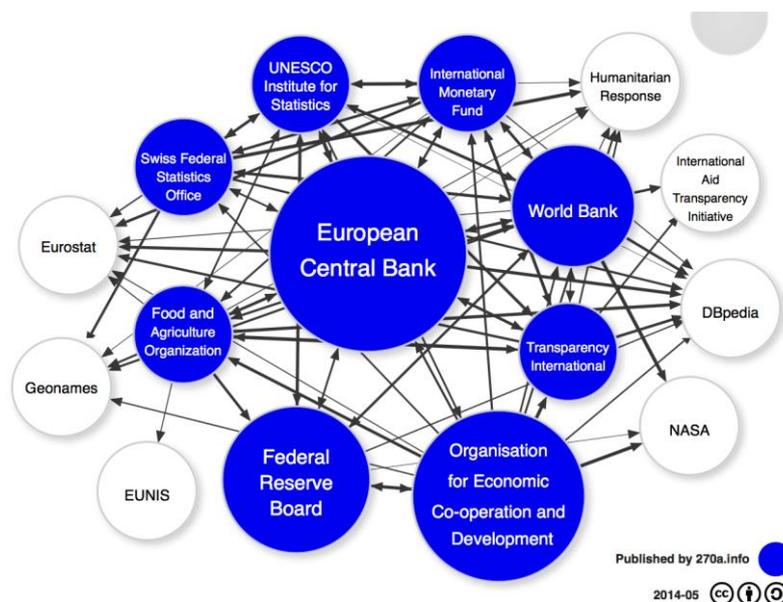
Dari sejumlah riset dan inisiatif linked data yang diamati, pada bab ini dipaparkan beberapa diantaranya yang relelvan terhadap tujuan penelitian.

### **Inisiatif-inisiatif Linked Statistical Data**

#### **270a.info**

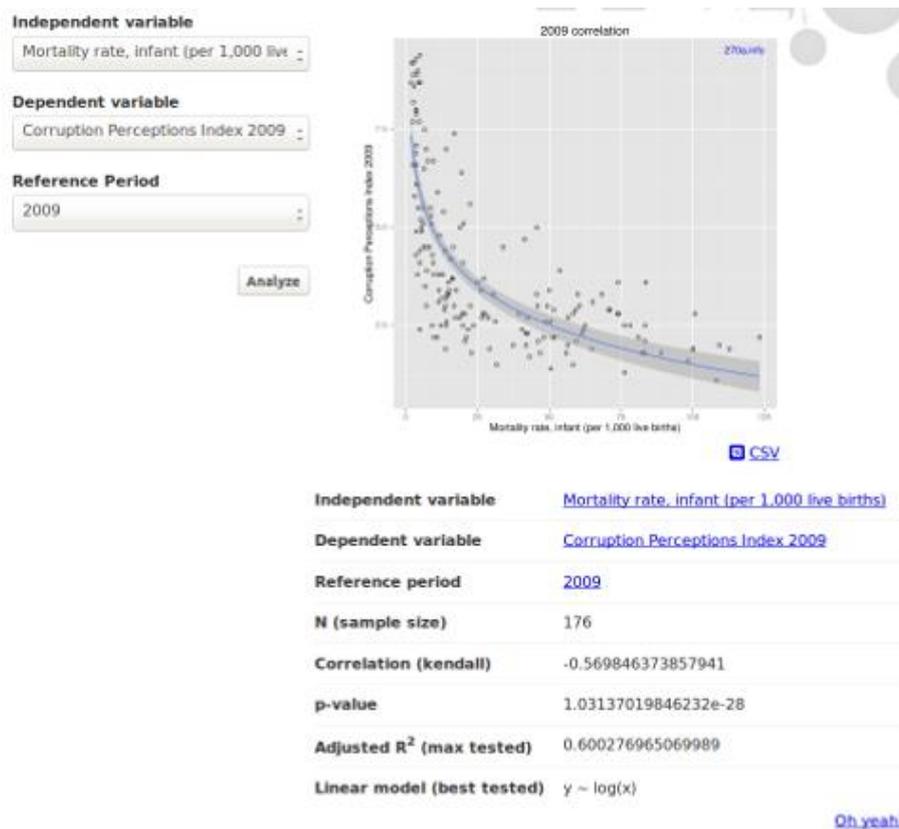
Merupakan portal linked data yang dikhususkan untuk data statistik. 270a.info dikembangkan oleh Sarven Capadisli melalui sejumlah riset yang dikhususkan untuk data official statistik dari berbagai organisasi internasional, seperti World Bank, IMF, dan FAO. Sekumpulan data statistik yang dihimpun dikenal dengan istilah "Dataspaces". Riset ini melakukan transformasi dataspaces kedalam bentuk linked data yang dikenal dengan istilah "Linked Dataspaces". Data untuk organisasi statistik yang terkait pemerintahan dikenal dengan "Statistical Linked Dataspaces".

Format data yang berbeda-beda dari berbagai organisasi, misalnya the World Bank Linked Dataspace yang berbasis custom XML dengan menyediakan API yang mengimplementasikan XSL Templates dan The Transparency International Linked Dataspace yang berbasis Comma Separated Value (CSV) file dan menyediakan transformasi data melalui Google Refine dan RDF Extension, pertama-tama diubah kedalam bentuk Statistical Data and Metadata eXchange (SDMX). Setelah itu, data dan metadata dalam bentuk SDMX kemudian dikoneksikan satu sama lain yang kemudian dikenal dengan "Linked SDMX Data" yang visualisasinya disajikan pada **Gambar 5**.



**Gambar 5.** Visualisasi linked data pada 270a.info. Sumber: 270a.info

Selain menyediakan Statistical Linked Dataspace, tool untuk analisis juga disediakan untuk melakukan analisis regresi keterkaitan antara 2 variabel dari sekumpulan data. Fitur analisis tersebut diberi nama Linked Statistical Data Analysis. Tool analisis akan memberikan output berupa grafik model regresi terbaik dari 2 variabel yang diproses.



**Gambar 6.** Output Linked Statistical Data Analysis pada 270a.info. Sumber: 2701.info

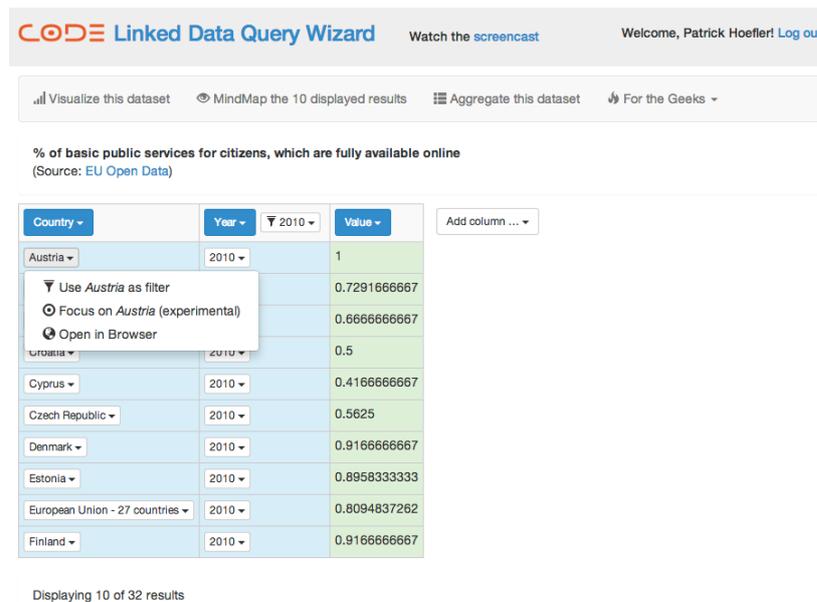
Pada **Gambar 6** dapat dilihat bahwa model regresi terbaik untuk 2 data yang diproses adalah linear model  $y = \log(x)$ .

Untuk mengeksplorasi data, SPARQL endpoint disediakan melalui <http://stats.270a.info/sparql> sehingga pengguna perlu menuliskan query pada editor yang disediakan, ataupun menggunakan aplikasi SPARQL client, sedangkan untuk tool analisis tersedia Graphical User Interface sehingga lebih mudah digunakan. SPARQL Endpoint untuk masing-masing statistical linked dataspaces yang diimplementasikan menggunakan Fuseki server yang disediakan oleh Apache pada project Apache Jena.

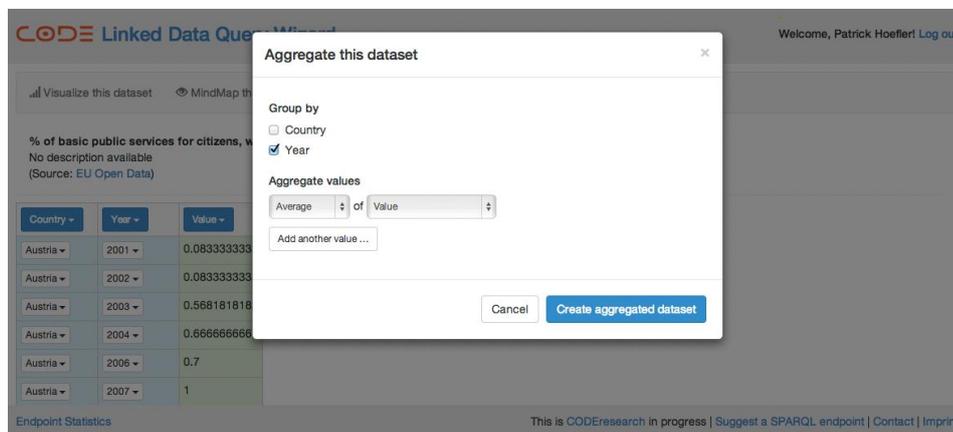
### **CODE Linked Data Query Wizard**

Untuk memudahkan pengguna data melakukan query terhadap linked data (RDF), riset ini mengembangkan sebuah tools yang diberi nama CODE Linked Data Query Wizard pada project yang diberi nama CODE project di Know-Center, Graz, Austria. GUI yang disediakan memungkinkan pengguna tanpa latar belakang ilmu komputer ataupun SPARQL query mengeksplorasi data pada linked dataset yang direpresentasikan sebagai RDF Datacube. Raw query SPARQL diautomatisasi dengan tampilan grafis yang mudah dioperasikan seperti pada **Gambar 7 (a)**.

Query wizard juga dilengkapi dengan full text search yang dapat menelusuri data dan metadata berdasarkan kata kunci yang dimasukkan. Selain data statistik, riset ini juga mengoneksikan query wizard ke DB Pedia dimana kata kunci dapat diquery dari sekumpulan linked data dari ensiklopedi DB Pedia. Khusus untuk data numerik, fitur untuk memperoleh nilai turunan dari operasi aggregasi seperti pada **Gambar 7 (b)**.



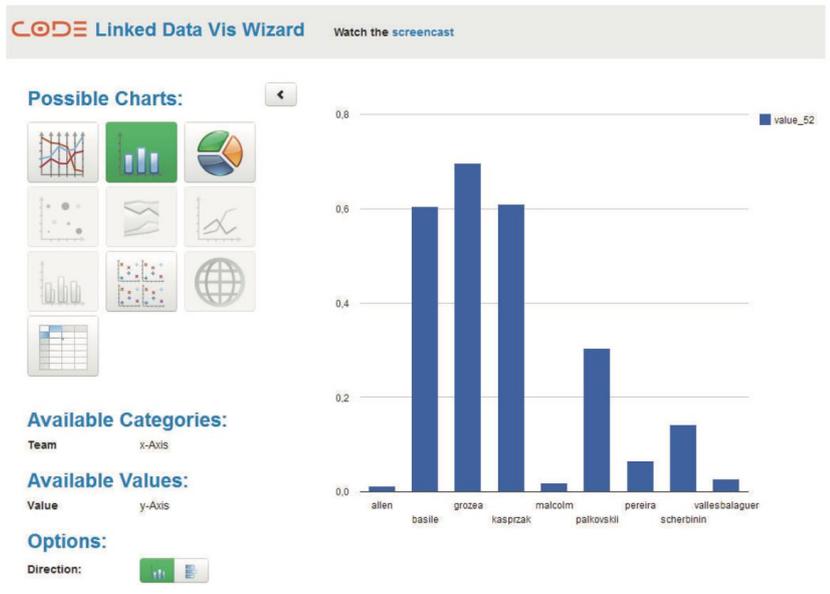
(a)



(b)

**Gambar 7.** CODE Linked Data Query Wizard. Sumber: code.know-center.tugraz.at

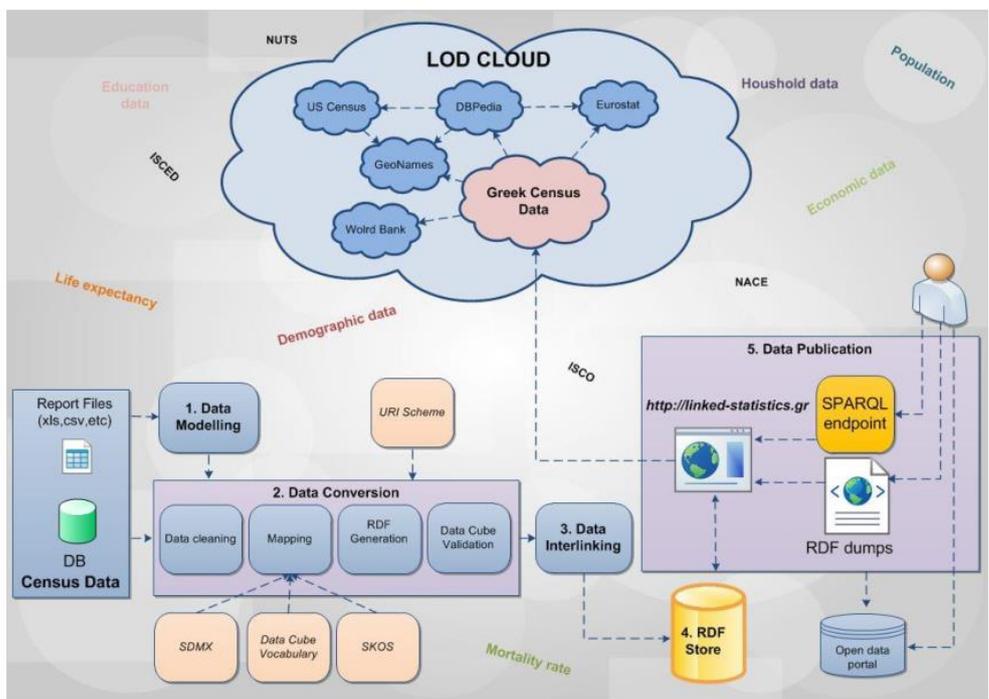
CODE project juga menghasilkan beberapa modul lain, misalnya linked data visualization untuk menampilkan linked data dalam bentuk berbagai jenis visualisasi data yang sesuai dengan karakteristik data yang dipilih untuk divisualisasikan (lihat **Gambar 8**).



Gambar 8. CODE Linked Data Vizualization Wizard. Sumber: code.know-center.tugraz.at

**linked-statistics.gr**

Kantor statistik nasional Yunani (Greece), Hellenic Statistical Authority (ELSTAT), mempublikasikan hasil sensus mereka pada tahun 2011 dalam bentuk linked open data. Inisiatif tersebut disajikan dalam sebuah portal web linked-statistics.gr yang memuat data sosial-ekonomi dan sosial-demografi dalam bentuk linked open data. Sistem aplikasi yang digunakan dibangun dari teknologi open source berupa software tools (Protégé-OWL editor and Google Refine), struktur data dan metadata (Data Cube Vocabulary, SDMX dan SKOS) serta teknologi penyimpanan LOD (OpenLink Virtuoso).



Gambar 9. Arsitektur Linked Open Data (LOD) ELSTAT, Greece. Sumber: linked-statistics.gr

Untuk mempublikasikan data dalam format LOD, berikut adalah metode yang diusulkan alurnya ditunjukkan pada **Gambar 9**:

- Data modelling: mengidentifikasi dan memodelkan ontologi semua konsep dan index yang spesifik pada sesus yang belum didefinisikan sebelumnya atau sumber metadata lainnya.
- Data RDF-ization: membersihkan data dan pemilihan skema URI yang sesuai untuk tiap jenis resources (datasets, dimensi, obeservasi, dan lainnya) serta melakukan mapping tiap konsep yang ada pada sumber data (misalnya nama kolom pada file ms.excel) ke komponen yang sesuai pada repository metadata.
- Data interlinking: data yangtelah ditransformasi menjadi RDF dikoneksikan satu sama lain.
- Data storage: RDF yang dihasilkan selanjutnya diunggah ke LOD triple store, yakni OpenLink Virtuoso.
- Data publication: data tersedia di server LOD dan dapat diekplorasi melalui SPARQL endpoint

Dengan langkah tersebut, project ini menghasilkan ontologi official statistics yang standard dan teratur sehingga data official statistik lainnya dapat dengan mudah dikonversi menjadi LOD mengikuti ontologi tersebut. **Gambar 10** menunjukkan codelist yang telah disusun dalam pembentukan ontologi official statistics ELSTAT.

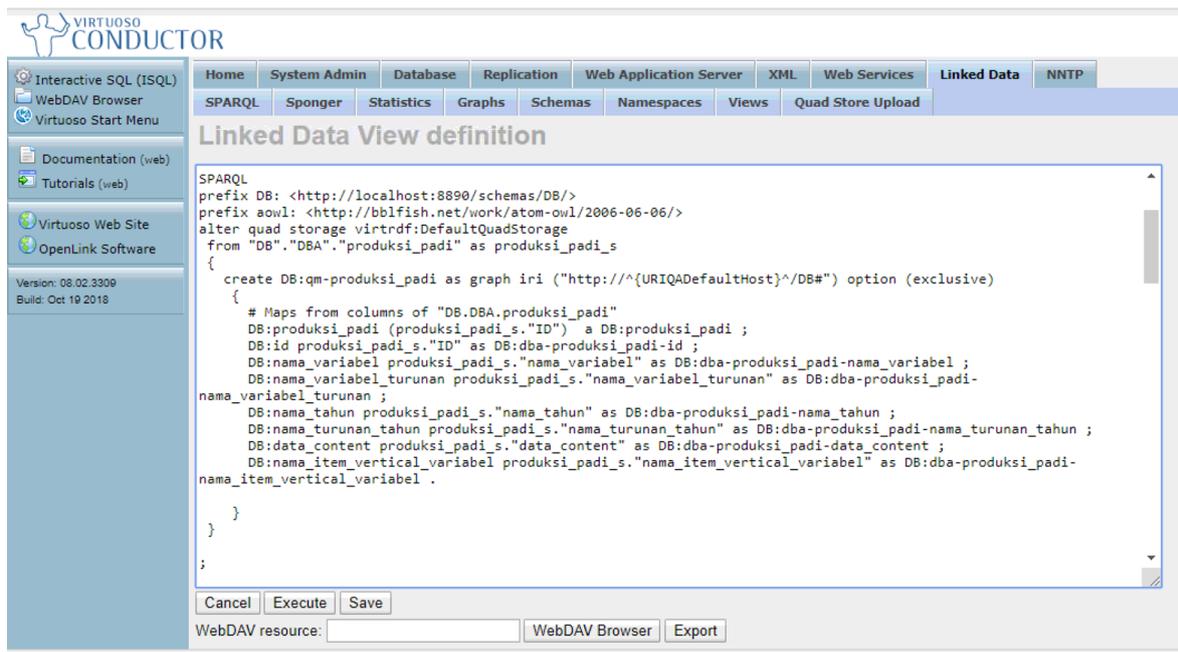
| Codelists  |                                       |
|--|---------------------------------------|
| Codelist Name  | Codelist Name                         |
| <a href="#">Age</a>                                  | <a href="#">Age of oldest child</a>   |
| <a href="#">Sex</a>                                  | <a href="#">Age of youngest child</a> |
| <a href="#">Household size</a>                       | <a href="#">Employment status</a>     |
| <a href="#">Marital status</a>                       | <a href="#">Activity status</a>       |
| <a href="#">Members less than 15 years</a>           | <a href="#">Number of children</a>    |
| <a href="#">Members over 15 years</a>                | <a href="#">Residential density</a>   |
| <a href="#">Citizenship</a>                          | <a href="#">Tenure status</a>         |
| <a href="#">Population type</a>                      | <a href="#">ISCO-08</a>               |
| <a href="#">Education Level</a>                      | <a href="#">Category</a>              |
| <a href="#">Place of birth (Regions, Abroad)</a>     |                                       |
| <a href="#">Place of residence (Regions, Abroad)</a> |                                       |
| <a href="#">Unit of measure</a>                      |                                       |
| <a href="#">Indicator type</a>                       |                                       |

**Gambar 10.** Codelist pada Ontologi LOD ELSTAT, Greece. Sumber: linked-statistics.gr

### Ujicoba Tools LOD

Selain melakukan eksplorasi dan demonstrasi beberapa tools LOD melalui portal yang dapat diakses publik, pada penelitian ini juga dilakukan ujicoba instalasi, konfigurasi, serta pengoperasian tools LOD yang digunakan pada riset rujukan yang telah dipaparkan di atas. Data diperoleh dari website BPS, yakni data produksi padi. OpenLink Virtuoso digunakan

sebagai LOD storage dimana data produksi pada berupa file csv dikonversi menjadi RDF. SPARQL endpoint yang dihasilkan dapat dilihat pada **Gambar 11** berikut.

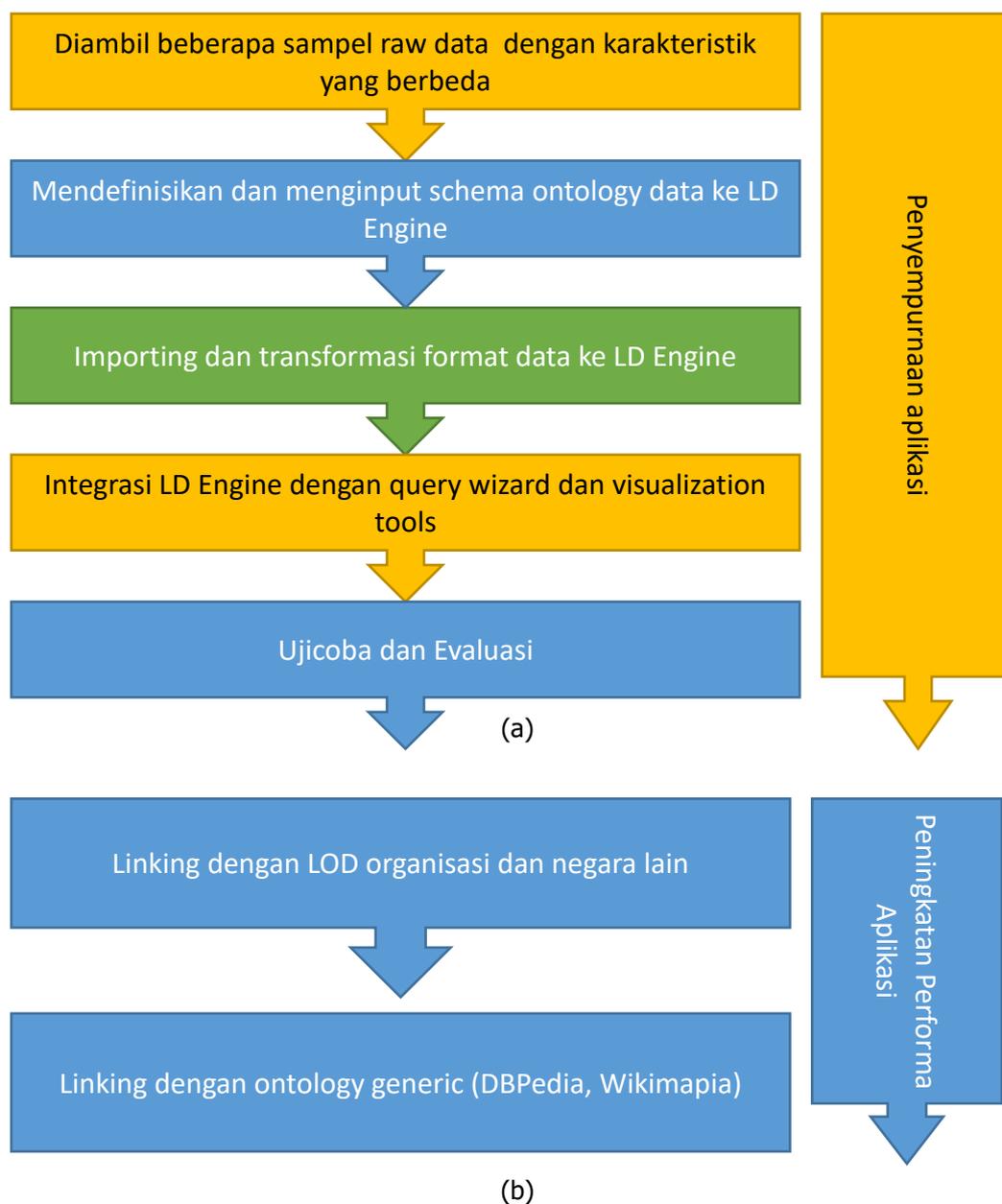


**Gambar 11.** Tampilan Ujicoba Aplikasi OpenLink Virtuoso

Deployment CODE Linked Data Query Wizard juga dilakukan pada penelitian ini dan telah berhasil melakukan query LOD dengan menggunakan GUI yang *user friendly*. Hal ini penting bagi tahap awal penerapan linked data di Indonesia mengingat saat ini belum tersedianya portal LOD di Indonesia dan SPARQL query merupakan hal yang belum populer di kalangan pengguna data ataupun penyedia data.

### Road Map Pengembangan LOD

Setelah melakukan serangkaian analisis dan ujicoba, penelitian ini mengusulkan suatu workflow dan roadmap (**Gambar 12**) pengembangan yang dapat dijadikan acuan dalam mengembangkan LOD official statistics. Pada tahap awal pengembangan sistem, LOD dikembangkan dengan kerangka metadata internal hingga nasional. Sedangkan pada pengembangan jangka panjang, interkoneksi LOD diperluas pada level antar-negara hingga interkoneksi dengan metadata pada repository linked data terbesar di internet, yakni seperti DBpedia dan Wikimapia. Usulan workflow dan roadmap ini mengacu pada perkembangan riset dan kemunculan inisiatif LOD di bidang official statistics saat ini serta pertimbangan untuk menjadikan LOD dapat menjangkau pengguna data yang memerlukan interaksi ke repository yang mudah dipahami dan digunakan.



- : telah selesai dilakukan pada penelitian ini
- : telah dilakukan dan belum selesai pada penelitian ini
- : belum dilakukan pada penelitian ini

**Gambar 12.** Workflow (a) dan Road Map (b) usulan Pengembangan Sistem LOD

## KESIMPULAN

Setelah melakukan penelitian ini, dapat disimpulkan bahwa:

- 1) Perkembangan linked open data untuk official statistics telah mengalami berbagai kemajuan, baik dari segi metodologi, serta teknologi pendukungnya.
- 2) Dari berbagai hasil riset yang diamati, tujuan utama linked open data official statistics adalah interkoneksi antara semua data yang memiliki keterkaitan ontologi. Namun, kompleksitas dan banyaknya variasi metadata pada official statistics membuat pengembangan struktur metadata diinisiasi oleh masing-masing inisiator untuk membentuk metadata yang standard yang lebih generic.
- 3) Outcome riset dan inisiatif LOD dapat dimanfaatkan dan akan mempercepat implementasi LOD. Perlu dilakukan pemilihan dan kombinasi metode, software, serta tools yang akan digunakan untuk implementasi sesuai dengan karakteristik data dan metadata.
- 4) Query wizard memungkinkan pengguna data yang tidak paham teknis teknologi LOD dan SPARQL untuk melakukan eksplorasi data secara mandiri pada LOD.

## DAFTAR PUSTAKA

- Adrian, Y.S., "Perancangan Linked Statistical Data Untuk Meningkatkan Level Open Data Pada Publikasi Data BPS", Skripsi, Politeknik Statistika STIS, 2015.
- Berners-Lee, T. (2006). Linked Data. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <http://doi.org/10.4018/jswis.2009081901>
- Cyganiak, R., dkk. (2011). Official Statistics and the Practice of Data Fidelity. Linking Government Data. Springer. New York, USA.
- Capadisli, S., Auer, S., & Ngonga Ngomo, A. C. (2015). Linked SDMX data:: Path to high fidelity statistical linked data. *Semantic Web*, 6(2), 105–112. <http://doi.org/10.3233/SW-130123>
- Do, B.-L., Trinh, T.-D. T.-D., Wetz, P., Anjomshoaa, A., Kiesling, E., & Tjoa, A. M. (2014). Widget-based exploration of linked statistical data spaces. *3rd International Conference on Data Management Technologies and Applications, DATA 2014*, (August 2014), 282–290. <http://doi.org/10.5220/0005110102820290>
- Fakhriyanto, "Perancangan Semantic Search Engine dan Penjawab Otomatis Berbasis Twitter untuk Permintaan Data Publikasi BPS", Skripsi, Politeknik Statistika STIS, 2014.
- Hausenblas, M., Halb, W., Raimond, Y., Feigenbaum, L., & Ayers, D. (2009). SCOVO: Using Statistics on the Web of Data. *6th European Semantic Web Conference (ESWC2009), Semantic Web in Use Track*, 708–722.
- Herman, I. (2001). The Semantic Web. Retrieved from <https://www.w3.org/2001/sw/>
- Kämpgen, B., & Harth, A. (2011). Transforming statistical linked data for use in OLAP systems. *Proceedings of the 7th International Conference on Semantic Systems - I-Semantics '11*, 33–40. <http://doi.org/10.1145/2063518.2063523>
- Othman, A., & Gloaguen, R. (2013). *River Courses Affected by Landslides and Implications for Hazard Assessment: A High Resolution Remote Sensing Case Study in NE Iraq–W Iran. Remote Sensing*, 5(3), 1024–1044.