

実践！統計LOD活用セミナー 2017

統計LODのご紹介

—統計データの新しい提供方法—

独立行政法人統計センター

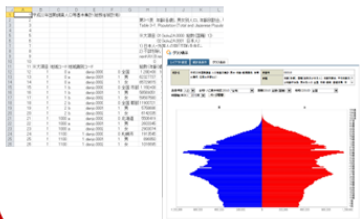
- 2008年度に整備した「政府統計の総合窓口（e-Stat）」により、各府省公表する統計データを一元的・総合的に提供。
- 基幹統計等主要な統計についてはデータベース化。



ページビュー 約1,800万件（月）

統計表ダウンロード数 約6,000万件（年）

統計表のダウンロードや、人口ピラミッドをはじめとした様々なグラフを作成できます。



地域で見る統計(統計GIS)を使うと、地域のすがたがよくわかります。



統計調査の調査表や調査項目などを詳しく調べることができます。



The figure shows a detailed survey table with multiple columns and rows of data, including headers and numerical values.

登録されている統計データ

- ・ 統計表（Excel、CSV等）
政府統計約570統計、約130万表
- ・ 統計情報データベース(XML)
99統計、約11万データセット

政府統計の総合窓口(e-Stat)

- 2008年～ e-Statの運用開始。EXCELファイル等のファイルダウンロードが中心
- 2014年～ API機能の提供開始。プログラム等から自動的にデータの取得が可能
- 2016年～ LODの提供開始。データ間のリンクが可能



-

統計表ファイル(Excelファイル)												
第1表 男女別												
Table 1. Population by Sex and Residence, Japan, 2013a, 22 Data and Performance												
国	都道府県	市町村	人口数 (人)			人口密度 (人/km ²)			人口増加率(前年比)(%)			
			総計	男	女	総計	男	女				
表例												
表頭												
国	都道府県	市町村	人口数 (人)	男	女	人口密度 (人/km ²)	男	女	人口増加率(前年比)(%)			
1	国	国	127,455,400	65,560,644	61,894,756	33.3	33.9	32.8	0.00			
2	北海道	道庁	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
3	北海道	札幌市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
4	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
5	北海道	釧路市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
6	北海道	帯広市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
7	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
8	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
9	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
10	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
11	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
12	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
13	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
14	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
15	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
16	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
17	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
18	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
19	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
20	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
21	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
22	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
23	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
24	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
25	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
26	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
27	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
28	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
29	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
30	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
31	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			
32	北海道	旭川市	1,287,400	643,700	643,700	33.3	33.9	32.8	0.00			



API機能の提供により、データのダウンロードや加工などをシステムに任せることができるようになり、利便性が向上。

しかし、似たようなデータが多数あり、データの定義も不明瞭なため、取得したいデータの検索などは、人の目で事前に確認し、対象データを指定（定義）することが必要となっている。

それぞれのデータの定義を明確にするとともに、データ間の関連性を定義することで、目的のデータを探しやすくなり、自動化がさらに進むのではないかな。

LOD (Linked Open Data)による統計データの提供

平成28年6月からLODによるデータ提供を開始。

オープンデータの5つ星レベル

Webの発明者であるティム・バーナーズ・リーがオープンデータの5つ星レベルを提案



(Source: <http://5stardata.info/>)

標準化された技術を使用

“World Wide Web Consortium(W3C)”

データの定義

Resource Description Framework(RDF)

W3C Recommendation

1.0(22 February 1999) / 1.1(25 February 2014)

データの検索

SPARQL Query Language(SPARQL)

W3C Recommendation

1.0(15 January 2008) / 1.1(21 March 2013)

統計表の定義

RDF Data Cube Vocabulary

W3C Recommendation

(16 January 2014)

RDFへの変換

RDB to RDF Mapping Language(R2RML)

W3C Recommendation

(27 September 2012)

★	オープンライセンスでデータを公開
★★	コンピュータで処理可能なデータを公開
★★★	オープンに利用できるフォーマットでデータを公開
★★★★	Web標準のフォーマット(URI)でデータを公開
★★★★★	他へのリンクを入れたデータを公開(LOD)

ファイルへのリンクからデータへのリンクへ

ファイルへリンク

データへリンク

ページにアドレスを付与

(<http://www.e-stat.go.jp/pages.html>)

ファイルにアドレスを付与

(<http://www.e-stat.go.jp/xls0001.xls>)

Sex Age	Total (Sex)			Male		
	...	44 years [Person]	45 years [Person]	...	44 years [Person]	45 years [Person]
Standard area code	...	...	...	...	...	...
Saitama-city	...	16,130	19,245	...	8,293	9,938
Kawaguchi-city	...	6,582	8,022	...	3,526	4,289
...	...	...	...	...	...	...

データにアドレスを付与

(<http://data.e-stat.go.jp/lod/.../C11201>)

データにアドレスを付与

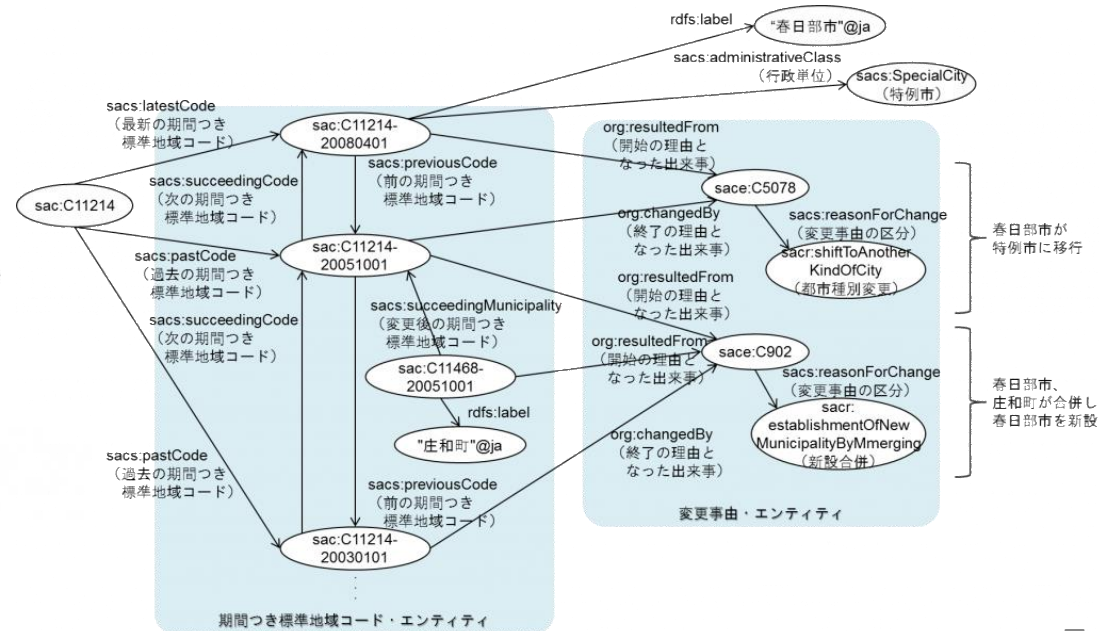
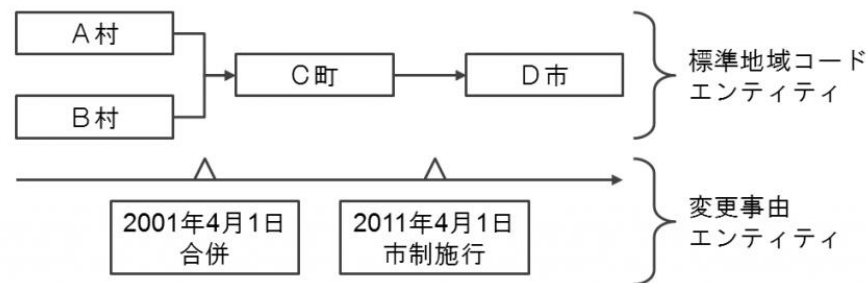
(<http://data.e-stat.go.jp/lod/.../obs00001>)

統計LODの概要（市区町村コード情報）

- ・ 都道府県・市区町村コード情報（統計に用いる標準地域コード）をLODで公開（2016年3月）
- ・ JIS X0401及びX0402で規格化されている5桁のコード(例:01000(北海道))に関する情報を提供。1970年からの廃置分合情報も含まれている。
- ・ 統計の結果を市区町村などで表す時には、原則当該コードが使用されているが、統計以外の情報についても当該コードを使用している事例が多数あるため、当該コード情報をLODで提供することにより、各種データのリンクが期待できる。



共通語彙基盤と連携



イタリア国家統計局などの事例を参考しつつ、
RDF Data Cube Vocabulary(W3C勧告) をベースに作成

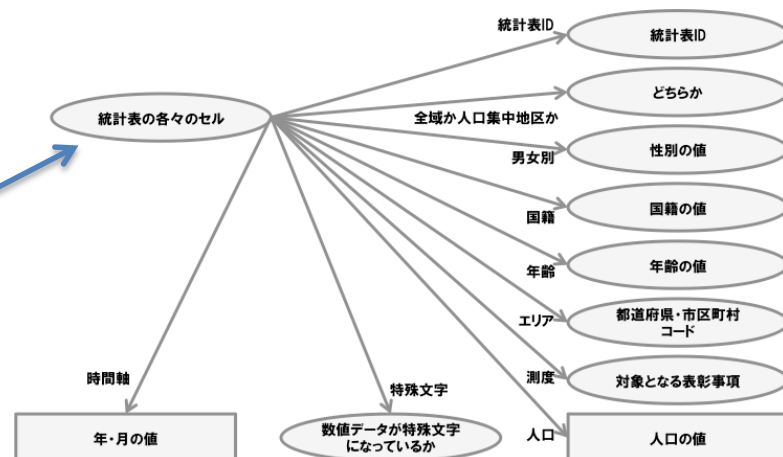
RDF Data Cube Vocabularyは、統計データとメタデータの交換規約であるSDMX(Statistical Data and Metadata eXchange)をベースとしており、統計データを以下の要素で表現している。

- ・次元(Dimension): 何で分類しているか(地域、性別、産業分類など)
- ・測度(Measure): 何で集計等しているか(人、世帯、価格など)
- ・観測値(Observation): 実際のデータ(セルに該当)
- ・属性(Attribute): データの属性(単位、状況(暫定値か)など)

イメージ

例：2010年国勢調査の人口

次元	総数(男女別)		男		女	
	44歳	45歳	44歳	45歳	44歳	45歳
さいたま市	...	16,130	...	8,293	...	8,293
川口市	...	6,582	...	3,526	...	3,526
...	...	...	...	...	...	...



1 データ単位に、複数の次元（分類）や測度（人口）などの情報を定義する。

統計データを表すメタ情報（次元など）を定義。

調査共通(Cross Domain)

各種統計調査共通で利用される要素は、共通語彙(Cross Domain)として定義。

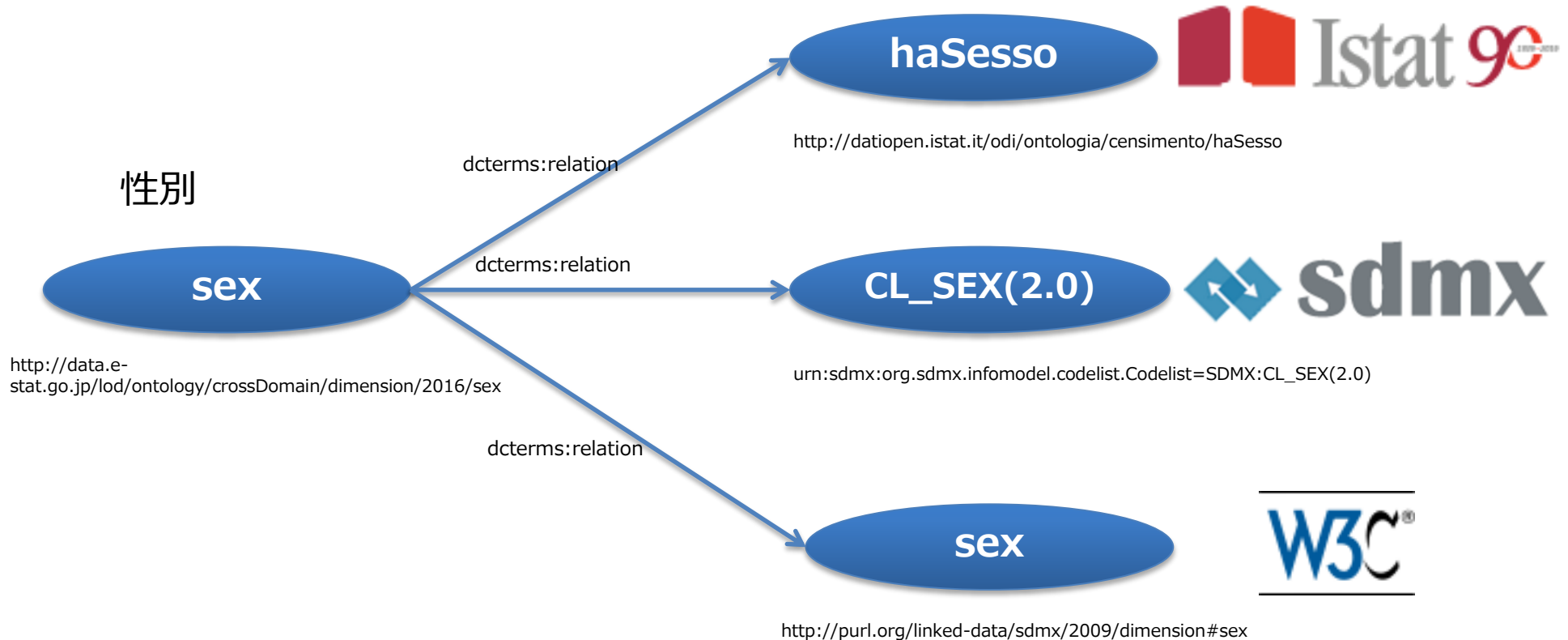
- ・性別
 - ・年齢
 - ・日本標準産業分類（大・中分類）
 - ・日本標準職業分類（大分類）
- など

調査単位

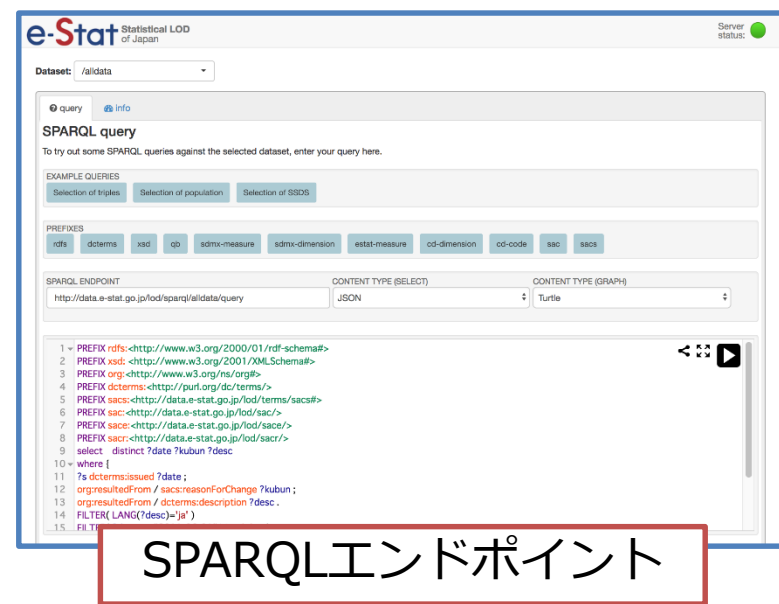
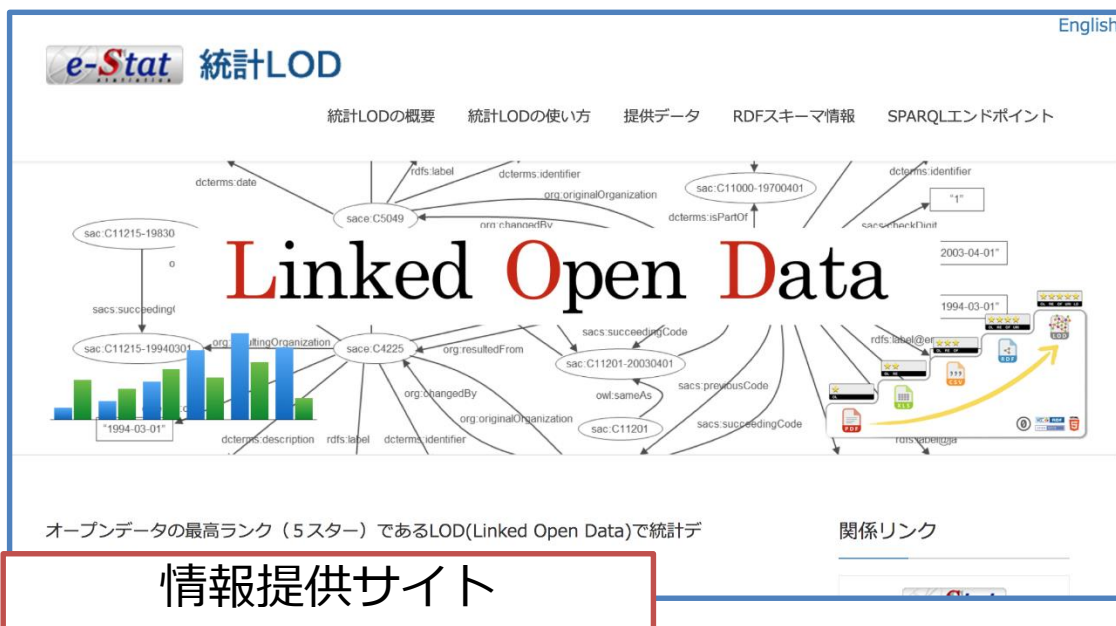
統計調査ごとに利用する要素は、調査ごと定義。

- ・家族類型（国勢調査）
 - ・従業上の地位（国勢調査、経済センサス）
 - ・産業分類（国勢調査、経済センサス）→ 日本標準産業分類とリンク
 - ・職業分類（国勢調査）→ 日本標準職業分類とリンク
- など

他の定義情報とリンク



<http://data.e-stat.go.jp/lodw/>



- ・ 英語ページを追加
- ・ 統計LODの使い方や統計LODの考え方などを取りまとめたなどガイドを拡充
- ・ URIアクセスに対応



統計LODのデータの検索にはSPARQLを使用

複数のデータセットを横断的に検索することが可能

```
select ?pop ?estb ?birth
where {
  ?s estat-measure:population ?pop ;
  cd-dimension:sex cd-code:sex-all ;
  cd-dimension:nationality cd-code:nationality-japan ;
  g00200521-dimension-2010:area g00200521-code-2010:area-all ;
  cd-dimension:age cd-code:age-all ;
  cd-dimension:timePeriod "2015"^^xsd:gYear ;
  sdmx-dimension:refArea ?areacd.
  ?sc estat-measure:establishments ?estb ;
  g00200552-dimension-2014:industryClassification g00200552-code-2014:industryClassification-A ;
  g00200552-dimension-2014:legalOrganization g00200552-code-2014:legalOrganization-all ;
  cd-dimension:timePeriod "2014"^^xsd:gYear ;
  sdmx-dimension:refArea ?areacd.
  ?o2 sdmx-measure:obsValue ?birth;
  g00200502-dimension:indicator g00200502-code:indicator-A4101;
  cd-dimension:timePeriod "2015"^^xsd:gYear ;
  sdmx-dimension:refArea ?areacd.
  ?areacd dcterms:identifier "13000".
}
```

横断的な検索

Population census (2015)

Population- prefectures

	Total	Male	Female
...	...	...	...
Tokyo-to	12,948,463	6,388,206	...
Kanagawa-ken	8,887,304	4,437,047	...
...	...	...	...

Economic census for business frame (2014)

numberOfEstablishments- prefectures

	A	A01	A02
...	...	...	...
Tokyo-to	491	449	...
Kanagawa-ken	667	601	...
...	...	...	...

System of social and demographic statistics (2015)

numberOfBirth- prefectures

	Birth	...	...
...	...	...	...
Tokyo-to	113,194	...	...
Kanagawa-ken	73,475	...	...
...	...	...	...

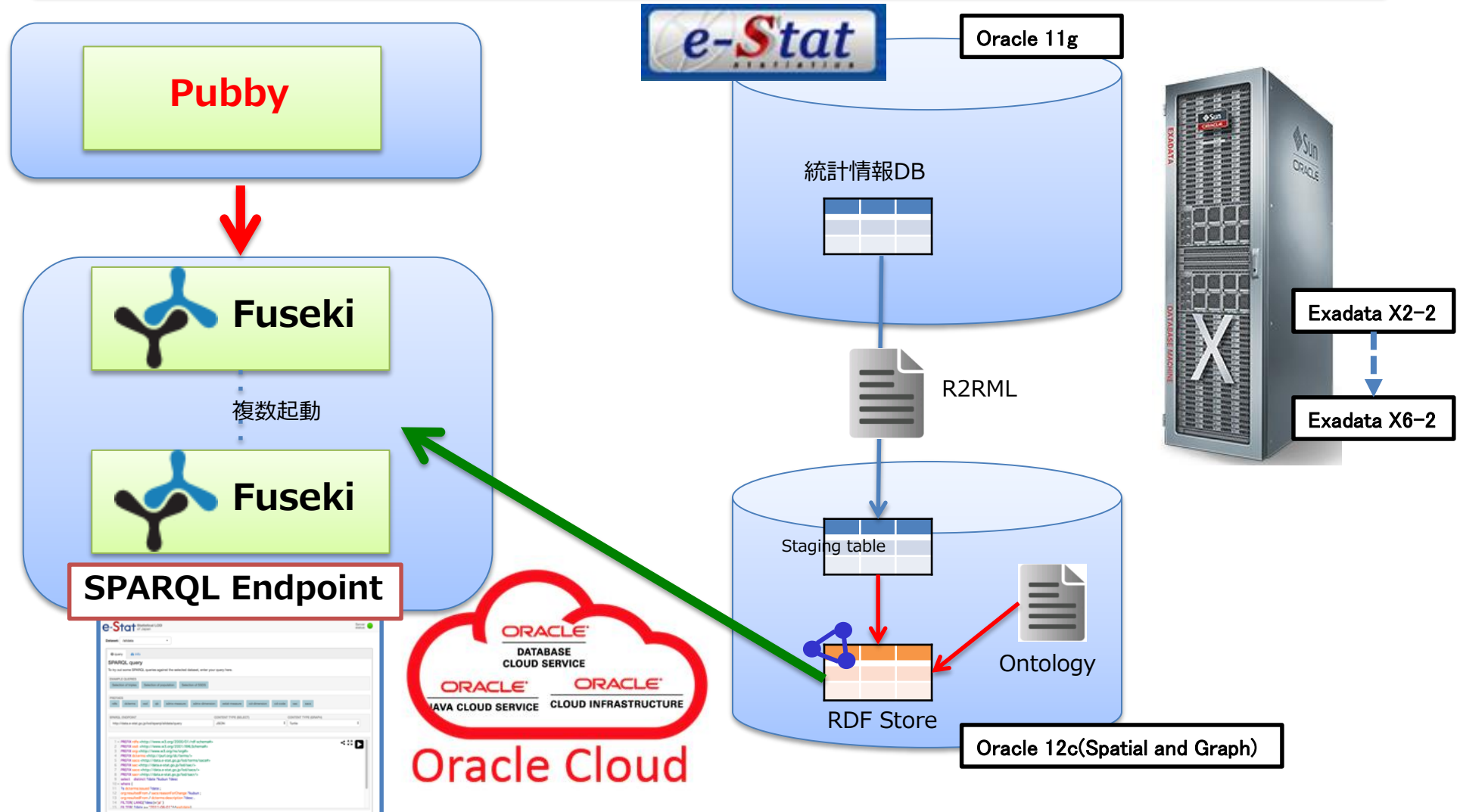
dataset

dataset

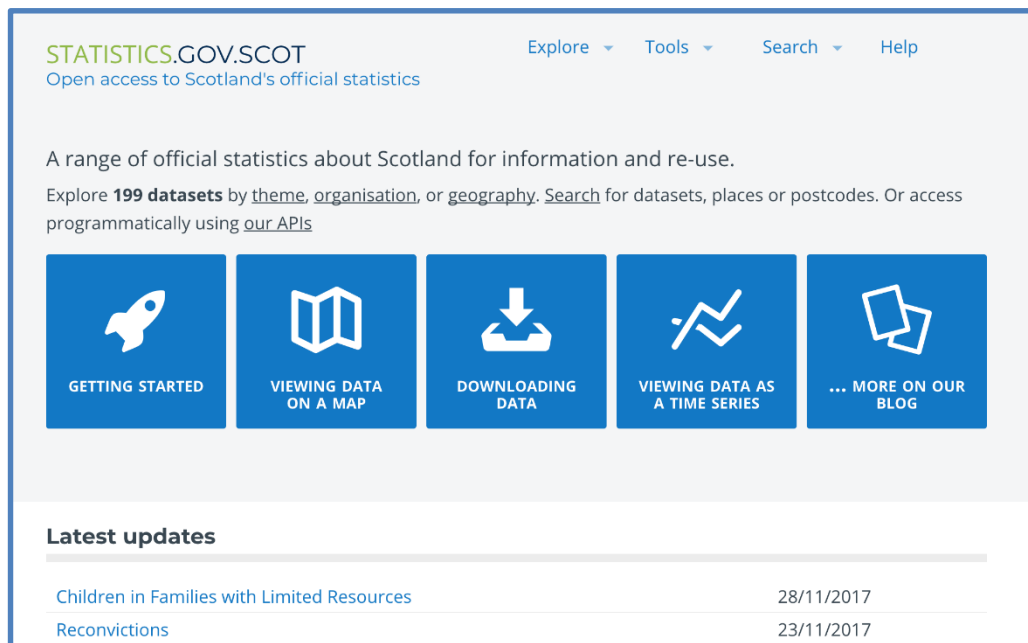
dataset

統計LODの概要（システム構成）

e-Statの統計情報DBから「R2RML」を利用してRDFに変換
SPARQLエンドポイントは「Fuseki」を使用し、
URIでのアクセスには「Pubby」を使用。



スコットランド



STATISTICS.GOV.SCOT
Open access to Scotland's official statistics

Explore ▾ Tools ▾ Search ▾ Help

A range of official statistics about Scotland for information and re-use.
Explore **199 datasets** by [theme](#), [organisation](#), or [geography](#). Search for datasets, places or postcodes. Or access programmatically using [our APIs](#)

GETTING STARTED VIEWING DATA ON A MAP DOWNLOADING DATA VIEWING DATA AS A TIME SERIES ... MORE ON OUR BLOG

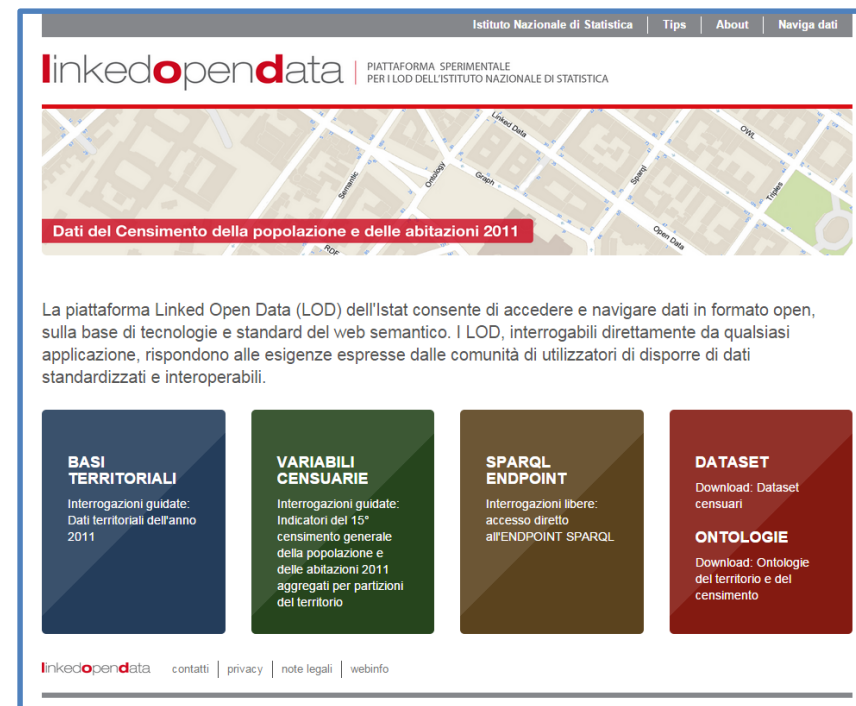
Latest updates

Children in Families with Limited Resources	28/11/2017
Reconvictions	23/11/2017

17のテーマの119のデータセットを提供

<http://statistics.gov.scot/>

イタリア国家統計局(ISTAT)



Istituto Nazionale di Statistica | Tips | About | Naviga dati

linkedopendata | PIATTAFORMA SPERIMENTALE PER I LOD DELL'ISTITUTO NAZIONALE DI STATISTICA

Dati del Censimento della popolazione e delle abitazioni 2011

La piattaforma Linked Open Data (LOD) dell'Istat consente di accedere e navigare dati in formato open, sulla base di tecnologie e standard del web semantico. I LOD, interrogabili direttamente da qualsiasi applicazione, rispondono alle esigenze espresse dalle comunità di utilizzatori di disporre di dati standardizzati e interoperabili.

BASI TERRITORIALI
Interrogazioni guidate: Dati territoriali dell'anno 2011

VARIABILI CENSUARIE
Interrogazioni guidate: Indicatori del 15° censimento generale della popolazione e delle abitazioni 2011 aggregati per partizioni del territorio

SPARQL ENDPOINT
Interrogazioni libere: accesso diretto all'ENDPOINT SPARQL

DATASET
Download: Dataset censuari

ONTOLOGIE
Download: Ontologie del territorio e del censimento

linkedopendata | contatti | privacy | note legali | webinfo

2011年の国勢調査のデータを提供。

<http://datiopen.istat.it>

国連欧州経済委員会

公的統計の近代化のためハイレベルグループにおいて「Linked Open Metadata」のプロジェクトで検討を行っている(2015-2016)



High-Level Group for the Modernisation of Official Statistics

<http://www1.unece.org/stat/platform/display/hlgbas/>

Linked Open Metadata(Project output)

<http://www1.unece.org/stat/platform/display/IMS/Project+Output>

ユーロスタット

ESS Vision 2020のDIGICOMプロジェクトで検討を行っている(2015-2019)



ESS Vision 2020

<http://ec.europa.eu/eurostat/web/ess/about-us/ess-vision-2020>



DIGICOM

DIGICOM(DIGITAL DISSEMINATION AND COMMUNICATION)

<http://ec.europa.eu/eurostat/web/ess/digicom>

提供環境

- ・ SPARQLのレスポンス向上
- ・ 提供データの拡充（メタデータの整備）
- ・ 統計LOD作成のシステム化

関係機関等の連携

- ・ 統計LODを作成・検討している機関や有識者等との意見交換を引き続き実施
→設計等の見直しや今後の取り組み方法等を検討