



Virtual Event 15-18 June
2020
**2020 Asia-Pacific
Statistics Week**

Leaving no one and nowhere behind

Linked Open Data Implementation for Integrated Dissemination

Action Area A. Engaging users and investing in statistics (SA2)

Innovations in meeting the demand for development data

Presenter:

Sulisetyo Puji Widodo¹; Wa Ode Zuhayeni Madjida¹; I

Komang Yudi Hardiyanta¹; Brilian Surya Budi¹

¹ Badan Pusat Statistik

#apstatsweek2020



#apstatsweek2020

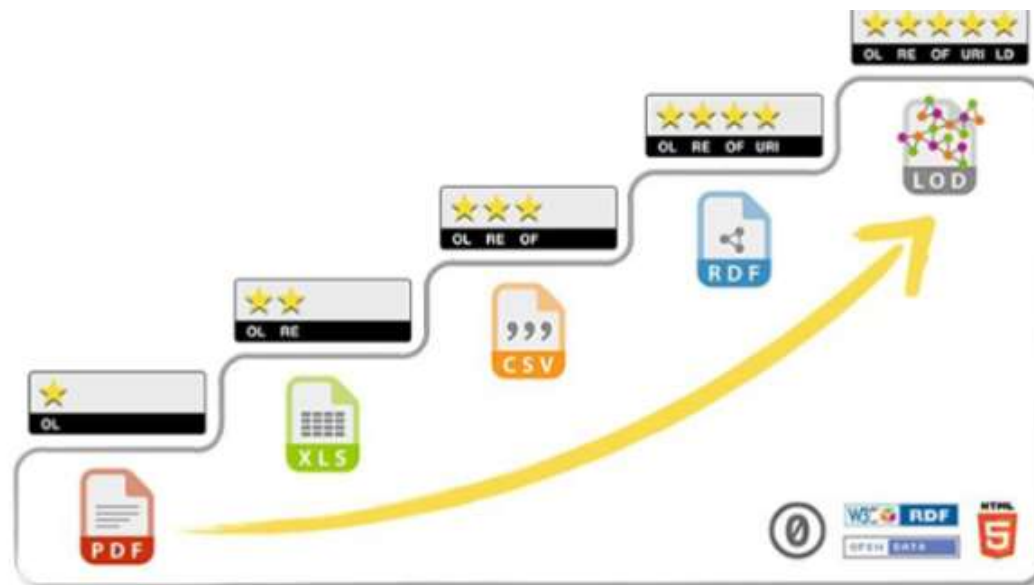


Virtual Event 15-18 June 2020
2020 Asia-Pacific
Statistics Week

Leaving no one and nowhere behind

Introduction

- **Linked Open Data (LOD)** is a way to utilize Open Data and similarity of metadata so that data can be interconnected (Blaney. J, 2017). Tim Berners-Lee established five level in applying open data levels.



Picture 1. Five Levels of Open Data



Virtual Event 15-18 June 2020
2020 Asia-Pacific
Statistics Week

Leaving no one and nowhere behind

Introduction (2)

- BPS - Statistics Indonesia disseminate official statistics on a web portal with various **formats such as tables, Microsoft excel, e-books, and Application Programming Interface (APIs).**
- This condition is not bad, where BPS is at level 3 according to Tim Berners-Lee. But BPS currently has not implemented LOD in disseminating its data. **This paper provides recommendations on how to implement LOD on BPS web portal.**

The screenshot shows a web interface with a table and a dropdown menu for data download options. The table has columns for income quintiles and years 2015-2019. The dropdown menu includes options for Print Table, XLS File, XML File, CSV File, and PDF File.

Kelompok Pengeluaran	Proporsi Populasi Penduduk Yang Memiliki Akses Terhadap Layanan Sanitasi Layak Dan Berkelanjutan Menurut Kelompok Pendapatan (Persen)				
	2015	2016	2017	2018	2019
Kuintil 1	41.76	48.51	48.61	50.72	62.77
Kuintil 2	51.30	57.95	58.12	59.92	70.08
Kuintil 3	58.89	64.89	65.30	68.14	76.85
Kuintil 4	68.47	74.35	73.32	74.28	82.45
Kuintil 5	81.14	85.57	85.09	86.20	89.22

Picture 2. Open Data in BPS Web Portal



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Methodology

The Japanese Statistics Bureau published the LOD website in 2016. In his writings, (Asano. Y., et al, 2016) explains the steps of Japanese Statistics Bureau in implementing LOD. The steps are as follows:

- **Data selection:** In this paper, we take one example of data "number of poor people by province in 2007-2019". This data is one of the data that can be used as support for the achievement of SDGs indicators.
- **Ontology preparation:** Lists items needed to declare target data as LOD. When the standard vocabulary is available, the vocabulary can be used. If it is not available, it is necessary to define vocabulary as an ontology. Some of the vocabularies that have been available are purl.org, w3.org.
- **Make a definition of dataset and data structure definition (DSD).**
- **Conversion of observations into RDF.** In this case, we convert a table from BPS web portal into RDF form.



#apstatsweek2020



Virtual Event 15-18 June 2020
**2020 Asia-Pacific
 Statistics Week**

Leaving no one and nowhere behind

Result

Number of Poor People By Province In 2007-2019

Tabel Unduh Data

Provinsi									
	2007	2008	2009	2010	2011	2012		2013	
	Tahunan	Tahunan	Tahunan	Tahunan	Tahunan	Semester 1 (Maret)	Semester 2 (September)	Semester 1 (Maret)	Semester 2 (September)
ACEH	218.80	195.80	182.19	173.37	176.02	171.80	165.43	156.37	156.37
SUMATERA UTARA	833.50	761.70	688.04	689	691.13	669.25	669.36	654.04	654.04
SUMATERA BARAT	149.20	127.30	115.78	106.18	140.49	127.81	124.25	119.53	119.53
RIAU	246.40	245.10	225.60	208.92	141.92	148.17	156.41	146.30	146.30
JAMBI	137.20	120.10	117.29	110.82	108.17	103.48	105.35	100	100
SUMATERA SELATAN	545.90	514.70	470.03	471.22	409.15	388.65	367.64	384.77	384.77
BENGGULU	135.60	131.80	117.60	117.21	95.28	93.67	92.67	91.91	91.91
LAMPUNG	366	365.60	349.31	301.73	241.94	239.07	237.94	233.01	233.01
KEP. BANGKA BELITUNG	38.60	36.50	28.78	21.85	25.32	25.13	24.01	22.73	22.73

Picture 3. Sample Table in BPS Web Portal

#apstatsweek2020





Virtual Event 15-18 June 2020
**2020 Asia-Pacific
Statistics Week**

Leaving no one and nowhere behind

Result (2)

- There is a vocabulary that has been provided by external BPS and can be used immediately.
- BPS with other NSOs can have different concepts or code lists, so each NSO needs to build its own vocabulary. Code lists standardization is very important to facilitate the integration of statistical data from different sources.

Table 1. Standard Vocabulary from External

URI

http://purl.org/linked-data/sdmx/2009/metadata#STAT_POP
http://purl.org/linked-data/sdmx/2009/metadata#OBS_VALUE
http://purl.org/linked-data/sdmx/2009/dimension#REF_AREA
http://purl.org/linked-data/sdmx/2009/metadata#UNIT_MEASURE
http://purl.org/linked-data/sdmx/2009/dimension#REF_PERIOD
http://reference.data.gov.uk/doc/gregorian-year/2007 - 2019

Table 2. Internal Defined Vocabulary

URI

http://bps.go.id/codelist/area/province/
http://bps.go.id/codelist/survey_period/
http://bps.go.id/dataset/dynamictable/2016/01/18/1119/jumlah-penduduk-miskin-menurut-provinsi-2007-2019/thousand

#apstatsweek2020





Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Result (3)

```

- <rdf:Description rdf:about="http://bps.go.id/dataset/dynamictable/2016/01/18/1119/jumlah-penduduk-miskin-
menurut-provinsi-2007-2019/thousand/AC/Perdesaan/Prov/36/Period/1S/Year/2016/SemesterPeriod/2">
  <ns2:UNIT_MEASURE rdf:resource="http://bps.go.id/dataset/dynamictable/2016/01/18/1119/jumlah-penduduk-
miskin-menurut-provinsi-2007-2019/thousand"/>
  <ns1:REF_AREA rdf:resource="http://bps.go.id/dataset/dynamictable/2016/01/18/1119/jumlah-penduduk-miskin-
menurut-provinsi-2007-2019/thousand/rural"/>
  <ns2:STAT_POP rdf:datatype="http://www.w3.org/2001/XMLSchema#double">277.58</ns2:STAT_POP>
  <ns1:REF_PERIOD rdf:resource="http://bps.go.id/codelist/survey_period/1S"/>
  <rdf:type rdf:resource="http://purl.org/dc/elements/1.1/Dataset"/>
  <ns1:REF_PERIOD rdf:resource="http://reference.data.gov.uk/doc/gregorian-year/2016"/>
  <ns1:REF_PERIOD rdf:resource="http://bps.go.id/codelist/survey_period/1S/2"/>
  <ns1:REF_AREA rdf:resource="http://bps.go.id/codelist/area/province/36"/>
</rdf:Description>

```

Picture 4. Sample RDF of One Cell in One Table

```

: # sample query
import rdflib

sq = rdflib.Graph()

# ... add some triples to g somehow ...
sq.parse("province.xml")
#print(sq.serialize(format="turtle").decode("utf-8"))

qres = sq.query(
    """
    SELECT ?sLiteral ?sLabel
    where
    {
    ?s ns1:notation ?sLiteral.
    ?s ns1:prefLabel ?sLabel.
    }
    """
)

for row in qres:
    print("%s | %s" % row)

```

```

82 MALUKU UTARA
14 RIAU
53 NUSA TENGGARA TIMUR
15 JAMBI
12 SUMATERA UTARA
11 ACEH
61 KALIMANTAN BARAT
33 JAWA TENGAH
65 KALIMANTAN UTARA
71 SULAWESI UTARA
72 SULAWESI TENGAH
13 SUMATERA BARAT
81 MALUKU
76 SULAWESI BARAT
21 KEPULAUAN RIAU

```

Picture 5. Sample Query SPARQL to Test RDF Form



#apstatsweek2020



Virtual Event 15-18 June 2020

2020 Asia-Pacific Statistics Week

Leaving no one and nowhere behind

Challenges

- The construction of ontology is also strongly influenced by the dimensions and measures available in the cube data. **Multiple measures will make the ontology development become more complex.**
- Another challenge in LOD implementation is **building a user-friendly interface** for users, because not all users are familiar with the use of queries.



#apstatsweek2020