

---

# **ISWC2019報告**

## **～統計・オープンデータ関連～**

**株式会社 日立製作所**  
**松田 純一**

# 1. トピックス紹介① ～Semantic Statistics

SemStats2019\*<sup>1</sup>(ワークショップ)、ISWC2019で関連発表\*<sup>2</sup>あり

## (1)統計表だけでなく個票のRDF・LOD化の動き

- ・質の高い統計データ作成
- ・時系列個票データの名寄せに有効
- ・ニュージーランド、オーストラリアの統計局でも取り組みあり

## (2)標準化の推進

- ・既存標準語彙をもとにした統計分野の語彙策定の動き

COOS(Core Ontology for Official Statistics)、StatDCAT、XKOS(SKOSの拡張)

- ・統計LODの標準モデル

統計LODサイトは、QBモデルで作成されている場合がほとんど(欧州各国+日本)  
メタデータ(測度、次元、単位及びそれらの値の定義)に関する標準が作れないか

(\*1) SemStats2019 Program      <http://semstats.org/2019/program>

(\*2) On Modeling Linked Open Statistical Data

## 2. トピックス紹介② ～Open Data Portal

欧州オープンデータポータル(European Data Portal)について3件発表あり\*3-5

欧州各国(全28の国)のオープンデータをまとめたポータルサイトについて紹介。  
イノベーション創出やデータ増加に役立てることが目的。  
現在では、LODの技術も取り入れ改善を図っている。

概要は以下の通り。

- データへのアクセスを容易にし、相互参照性を高めるためにURIを定義し、RDFを使用
- DCAT-AP(Catalogue, Dataset, Distribution)準拠のRDF形式で公開
- メタデータはSPARQLを含むいくつかのインタフェースでアクセス可能
- SHACLツールを用いてデータチェックを実施
- VirtuosoとElasticsearchを使って検索部分を構築
- GeoDCAT-APやStatDCAT-APも視野に入れている
- チェコの発表では、DCAT-APに忠実にデータセットを公開している模様。  
なお、pdfやExcelでのデータ公開は認めていない(csv以上なら許可)とのこと。

(\*3) An Assessment of Adoption and Quality of Linked Data in European Open Government Data

(\*4) The European Data Portal: Scalable Harvesting and Management of Linked Open Data

(\*5) DCAT-AP Representation of Czech National Open Data Catalog and its Impact

### 3. 全体感想

- ◆ RDF、LODでのデータ公開がますます重要

欧州は進んでいる印象

RDF、LODならではのメリット、ユースケースの紹介が重要

- ◆ 研究色の強い発表ばかりではなく、実用面の紹介も目立った

- ◆ オープンデータポータルのさらなる充実が必要

特に、データセットをDCAT又はschema.orgベースで公開することが重要

- ◆ 大規模なSemanticデータの作成活用に向けた課題

データ作成・運用の手間を軽減させる仕組み(SHACL等のツールの活用)

大規模データに耐えうる検索スピード

(参考) 日本の政府統計LODサイト: 20億強のトリプルを公開中

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

<https://data.e-stat.go.jp/lod/sparql/> (SPARQLエンドポイント)

<https://data.e-stat.go.jp/lodw/sample/mapSearch/> (地図検索)

<https://data.e-stat.go.jp/lodw/sample/simpleSearch/> (データ簡単取得)