



Open Government and the Semantic Web: a field report

Enriching government gazette notices with knowledge graphs

Guido van der Wolk | Taxonic

SEMANTiCS 2019, Karlsruhe





Guido van der Wolk (PhD),
Software Engineer,
Utrecht, The Netherlands

- Model-driven development
- Data science
- Machine learning



IT consultancy company:

- Linked Data
- Dynamic Case Management
- ICT services

Partners with TopQuadrant,
PoolParty, MarkLogic, Franz
Inc., Pega and GO FAIR

Custom open government

- New Dutch legislation (starting 01-01-2021)
 - Every government organization must publish their notices in the online official gazette
 - Every 18+ citizen will receive customizable (frequency, geo-range, topics) e-mail notices about publications near their home address

Source: Dossier Bekendmakingswet, <https://zoek.officielebekendmakingen.nl/dossier/35218>

Online official gazette of The Netherlands

- 300,000+ publications a year
- Official source since July 1, 2009
- Centralized government
 - state, parliament, treaties
- Decentralized government
 - provinces (12), municipalities (355), water boards (21)
- XML, HTML, PDF, ODT, Metadata (XML)
- Search platform
- Geo-based e-mail subscription platform
- Open data (CC-0 license): search and retrieve via URL (SRU) interface
- Web address: <https://officielebekendmakinge.nl/>

Government publications in the neighbourhood

Zoek op locatie

1012AB

Zoek locatie

Onderwerpen

- Consumentenzaken
- Cultuur, sport en vrije tijd
- Familie, jeugd en gezin
- Natuur en milieu
- Overheid en democratie
- Rechtspraak en veiligheid
- Verkeer, voertuigen en wegen
- Wonen en leefomgeving
- Diverse onderwerpen
- Meerdere bekendmakingen

Aanvraag evenementenvergunning Stationsplein 1

Evenementenvergunning
Cultuur, sport en vrije tijd

12-07-2019

Lil' Cities vraagt een evenementenvergunning aan voor het evenement "Lil' Amsterdam ADE 2019" ...

Amsterdam

Overige bekendmakingen
Bekijk de bekendmakingen die voor een gebied gelden en niet op de kaart worden weergegeven.

Kaartgegevens: © CBS, Kadaster, OpenStreetMap -auteurs (CC-BY-SA).

FAIR data

FAIR Data aims to support existing communities in their attempts to enable valuable scientific data and knowledge to be published and utilized in a 'FAIR' manner.

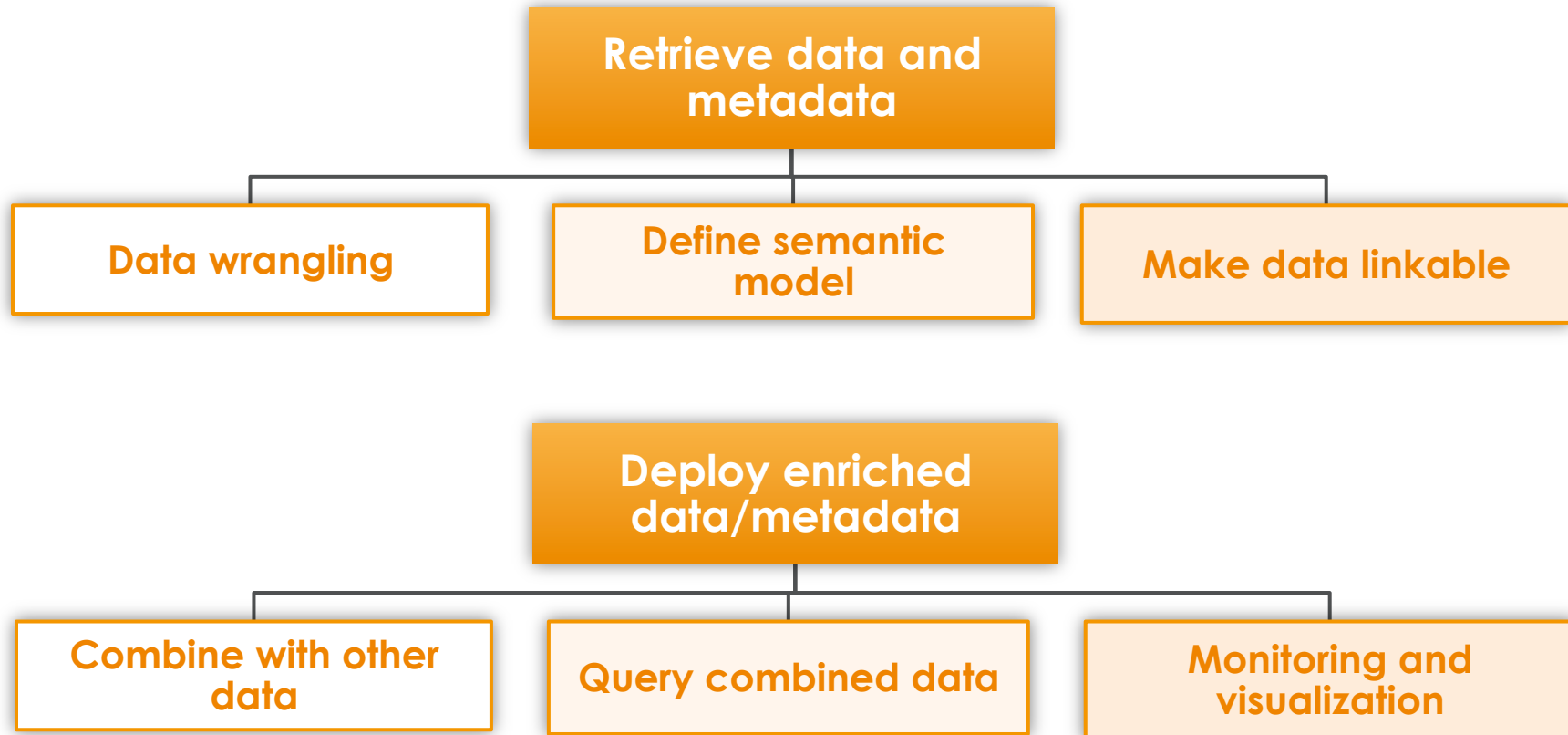
Findable - (meta)data is uniquely and persistently identifiable. Should have basic machine-readable descriptive metadata.

Accessible - data is reachable and accessible by humans and machines using standard formats and protocols.

Interoperable - (meta)data is machine readable and annotated with resolvable vocabularies/ontologies.

Reusable - (meta)data is sufficiently well-described to allow (semi)automated integration with other compatible data sources.

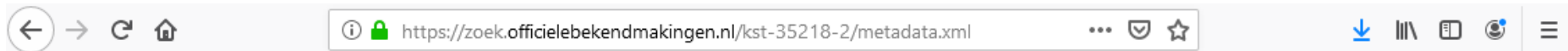
Enriching government gazette notices with knowledge graphs: a data hub



Data wrangling done FAIR

- Understanding the data
- Unique and persistent
- Data domain and range
- Unravel concatenation
- Merging fields
- Provenance
- XML to Turtle

Current metadata scheme



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
-<metadata_gegevens>
  <metadata name="OVERHEIDop.documentStatus" scheme="" content="Opgemaakt na onopgemaakt"/>
  <metadata name="DC.title" scheme="" content="Wijziging van de Bekendmakingwet en andere wetten in verband met de elektronische publicatie van algemene bekendmakingen, mededelingen en kennisgevingen (Wet elektronische publicaties); Voorstel van wet; Voorstel van wet"/>
  <metadata name="OVERHEIDop.documenttitel" scheme="" content="Voorstel van wet"/>
  <metadata name="OVERHEIDop.dossiernummer" scheme="" content="35218"/>
  <metadata name="OVERHEIDop.dossiertitel" scheme="" content="Wijziging van de Bekendmakingwet en andere wetten in verband met de elektronische publicatie van algemene bekendmakingen, mededelingen en kennisgevingen (Wet elektronische publicaties)"/>
  <metadata name="OVERHEIDop.indiener" scheme="" content="R. W. Knops"/>
  <metadata name="OVERHEIDop.ondernummer" scheme="" content="2"/>
  <metadata name="OVERHEIDop.publicationName" scheme="" content="Kamerstuk"/>
  <metadata name="OVERHEIDop.vergaderjaar" scheme="" content="2018-2019"/>
  <metadata name="DCTERMS.language" scheme="DCTERMS.RFC4646" content="nl"/>
  <metadata name="DCTERMS.available" scheme="DCTERMS.W3CDTF" content="2019-06-11"/>
  <metadata name="DCTERMS.issued" scheme="DCTERMS.W3CDTF" content="2019-06-06"/>
  <metadata name="DC.type" scheme="OVERHEID.Informatietype" content="officiële publicatie"/>
  <metadata name="OVERHEID.organisationType" scheme="OVERHEID.Organisatietype" content="staten generaal"/>
  <metadata name="DC.creator" scheme="OVERHEID.StatenGeneraal" content="Tweede Kamer der Staten-Generaal"/>
  <metadata name="OVERHEID.category" scheme="OVERHEID.TaxonomieBeleidsagenda" content="Bestuur | Organisatie en beleid"/>
  <metadata name="DC.type" scheme="OVERHEIDop.KamerstukTypen" content="Voorstel van wet"/>
  <metadata name="DC.type" scheme="OVERHEIDop.Parlementair" content="Kamerstuk"/>
  <metadata name="DC.identifier" scheme="OVERHEIDop.ParlID" content="kst-35218-2"/>
</metadata_gegevens>
```

Semantic enrichment

- dcterms
- dcam
- dcat
- foaf
- geo
- prov
- rdf/rdfs
- skos

- bwb
- ecli
- lido

- oep
- overheid
- overheidop

```
oep:prb-2019-781
  dcterms:title "Kennisgeving besluit op een aanvraag van een vergunning Wet
  natuurbescherming voor de locatie Mossendamsdwarsweg 5 in Goor";
  dcterms:identificer "prb-2019-781";
  dcterms:type [
    dcam:memberOf overheidop:Provinciaal_blad;
    prov:wasDerivedFrom "Beschikkingen | afhandeling";
    skos:broader overheid:Beschikkingen;
    skos:prefLabel "afhandeling";
  ];
  dcterms:creator [
    dcam:memberOf overheid:Provincie;
    skos:prefLabel "Overijssel";
  ];
  dcterms:spatial [
    prov:wasDerivedFrom "238040.86 474118.01"^^xs:string;
    rdf:type geo:Point;
    geo:lon "238040.86"^^xs:decimal;
    geo:lat "474118.01"^^xs:decimal;
  ];
  dcterms:modified "2019-02-04"^^xs:date;
  dcterms:available "2019-02-04"^^xs:date;
  dcterms:date "2019-02-04"^^xs:date;
  dcterms:subject [
    dcam:memberOf overheid:TaxonomieBeleidsagenda;
    prov:wasDerivedFrom "Ruimte en infrastructuur | Organisatie en beleid";
    skos:broader overheid:Ruimte_en_infrastructuur;
    skos:prefLabel "Organisatie en beleid";
  ];
  overheidop:publicationname "Provinciaal blad";
  overheidop:jaargang "2019";
  overheidop:publicationissue "781";
  foaf:page "https://zoek.officielebekendmakingen.nl/prb-2019-781.html";
  .
```

Semantic enrichment (type/subject)

```
oep:wsb-2019-1141
  dcterms:title "Losser, dempen waterloop WL00161";
  dcterms:identifier "wsb-2019-1141";
  dcterms:type [
    dcam:memberOf overheidop:Waterschapsblad;
    prov:wasDerivedFrom "Beschikkingen | afhandeling";
    skos:broader overheid:Beschikkingen;
    skos:prefLabel "afhandeling";
  ];
  dcterms:subject [
    dcam:memberOf overheid:TaxonomieBeleidsagenda;
    prov:wasDerivedFrom "Natuur en milieu | Organisatie en beleid";
    skos:broader overheid:Natuur_en_milieu;
    skos:prefLabel "Organisatie en beleid";
  ];
```

Semantic enrichment (spatial)

```
oep:stcrt-2019-6120
  dcterms:title "Yokohamadreef 136 Intrekking gehandicaptenparkeerplaats";
  dcterms:identifiser "stcrt-2019-6120";
  dcterms:spatial [
    prov:wasDerivedFrom "135545 459918"^^xs:string;
    rdf:type geo:Point;
    geo:lon "135545"^^xs:integer;
    geo:lat "459918"^^xs:integer;
  ];
  dcterms:spatial [
    prov:wasDerivedFrom "Utrecht"^^xs:string;
    dcam:memberOf overheid:Gemeente;
    dcterms:identifiser "overheid:Utrecht_(gemeente)";
  ];
```

Coordinates in EPSG28992 format

Semantic enrichment (legal source)

```
oep:stcrt-2019-6236
  dcterms:title "Besluit van het college van burgemeester en wethouders en van de
  burgemeester van de gemeente Oegstgeest houdende regels omtrent het
  Aanwijzingsbesluit toezichthouders Haags Economisch Interventie Team Gemeente
  Oegstgeest";
  dcterms:identificer "stcrt-2019-6236";
  dcterms:source [
    prov:wasDerivedFrom "artikel 92 van de Woningwet;1.
    0:c:BWBR0005181&artikel=92&g=2018-06-13";
    dcterms:identificer "1.0:c:BWBR0005181&artikel=92&g=2018-06-13";
    skos:prefLabel "artikel 92 van de Woningwet";
  ];
  dcterms:source [
    prov:wasDerivedFrom "artikel 33, eerste lid, van de Huisvestingswet 2014;1.
    0:c:BWBR0035303&artikel=33&lid=1&g=2017-07-01";
    dcterms:identificer "1.0:c:BWBR0035303&artikel=33&lid=1&g=2017-07-01";
    skos:prefLabel "artikel 33, eerste lid, van de Huisvestingswet 2014";
  ];
];
```

A new resource identifier strategy

Type	Example
Register	{prefix}/id/gemeente
Register item	{prefix}/id/gemeente/gm0004
Page	{prefix}/doc/gemeente/gm0004
Concept	{prefix}/def/concept/{identifier}
Class	{prefix}/def/class/gemeente
Property	{prefix}/def/property/gemeentecode
Dataset	{prefix}/set/gemeente
Dataset item	{prefix}/set/gemeente/20190101

Based on international guidelines:

- Gov.UK: <http://ukgovld.github.io/ukgovldwg/recommendations/uri-patterns.html>
- ISA: <https://joinup.ec.europa.eu/sites/default/files/document/2013-02/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf>

Linkable data

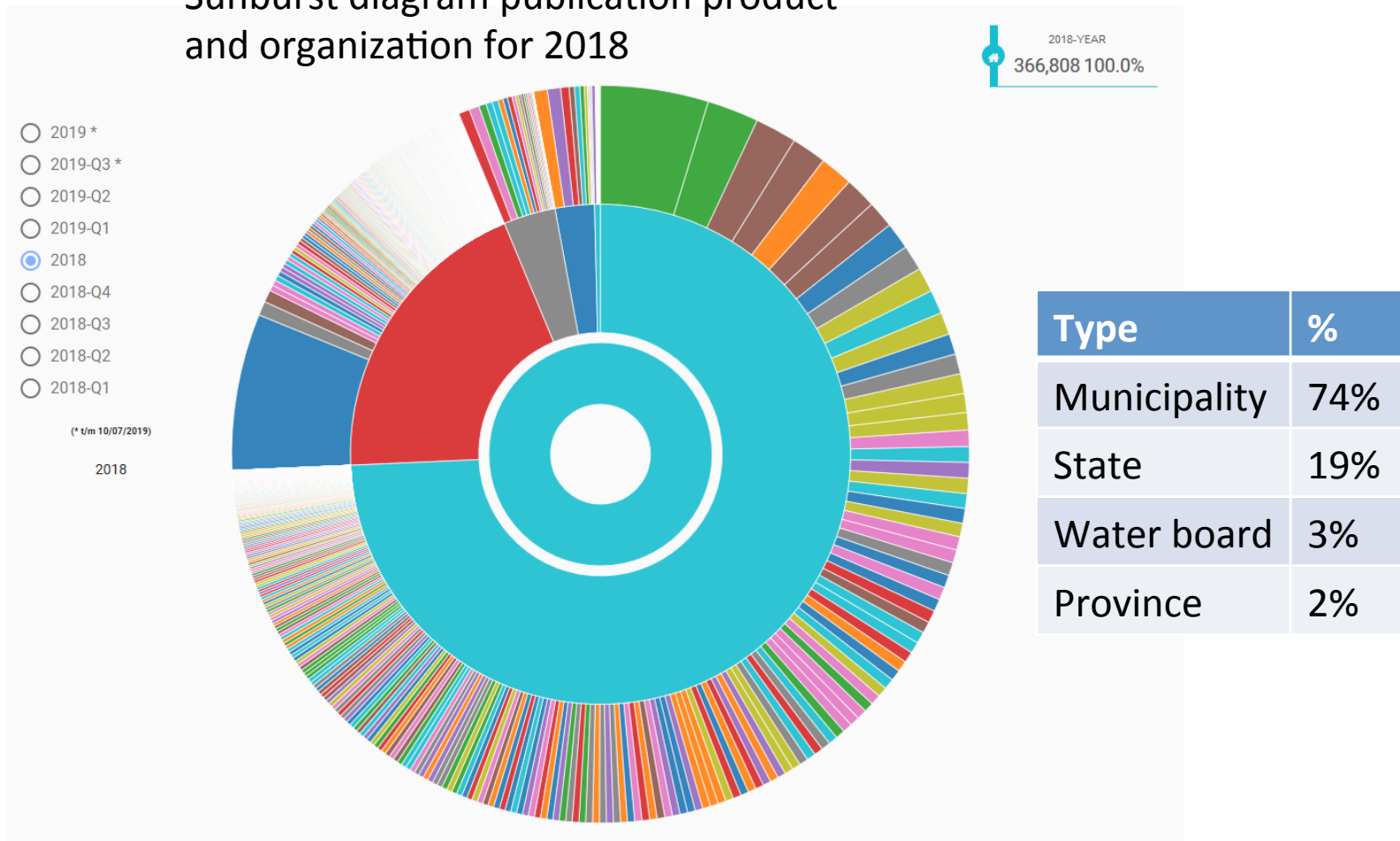
- Apply the semantic model to the original data to make data linkable
- Deploy to data hub
 - Search on data and metadata
 - Machine can discover the data
 - Inferencing: deriving new facts from existing data
 - Federated queries still a challenge
 - Serviceable SPARQL
 - Other datasets incomplete for certain time slices
 - Other datasets not FAIR

Monitoring data quality

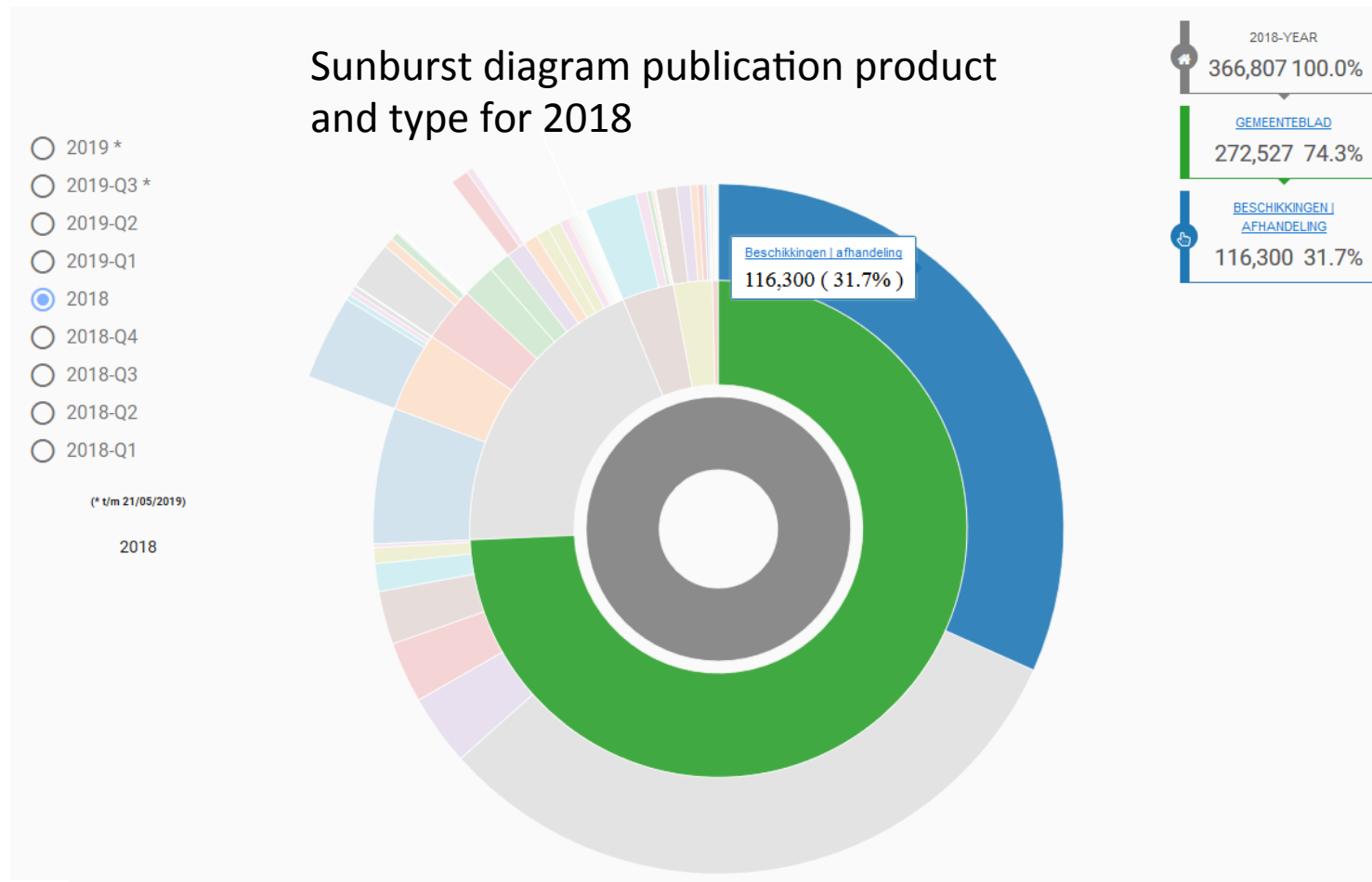
- Discover publications with
 - address in title but without geo:Point
 - 100+ geo:Point's
 - dcterms:type "other" but type in title
 - article of law in title without reference to law
 - legal source "unknown"

Monitoring and visualization (1)

Sunburst diagram publication product and organization for 2018



Monitoring and visualization (2)

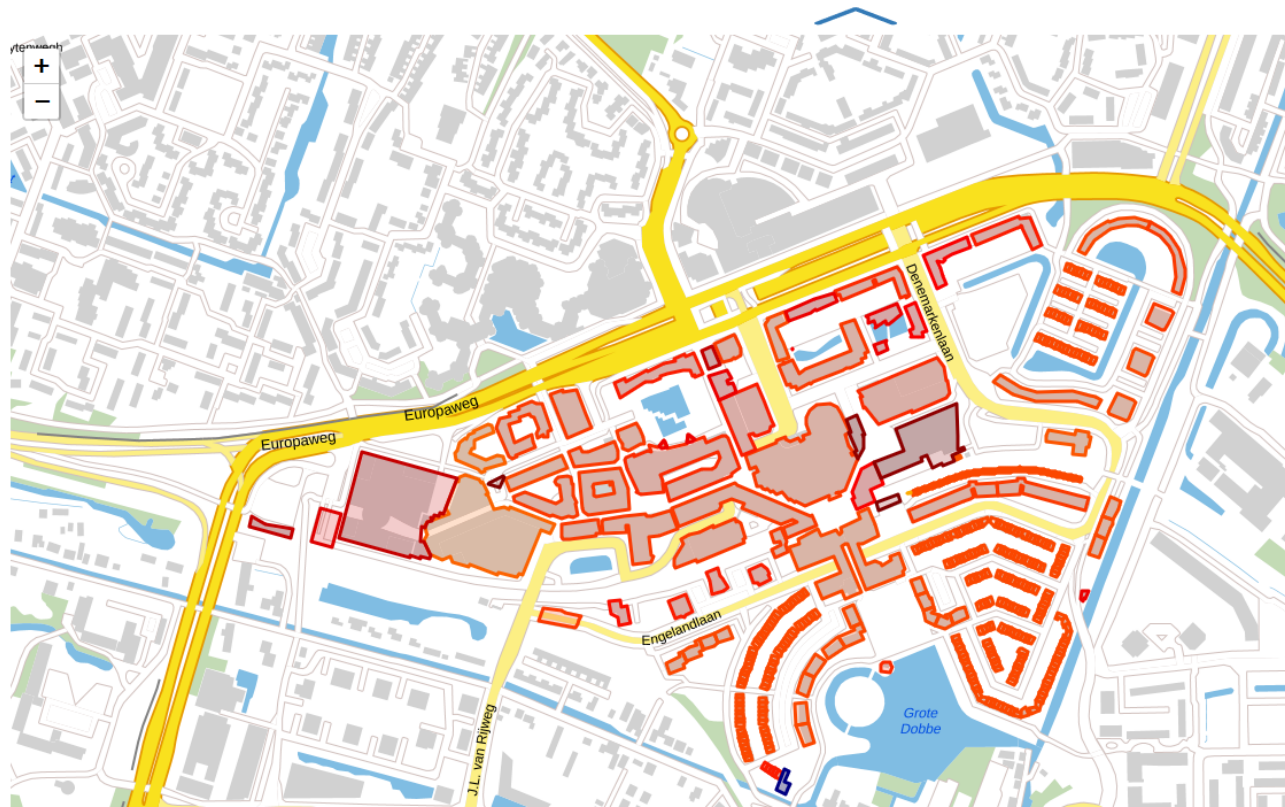


Combine with other data

- Registry of government organizations
- Geo- and demographic and social-economic information
- Judicial information: laws, article of law, court judgments, treaties, guidelines

Query combined data

Buildings by age inside a district (geosparql)



CBS and BAG data, <https://data.labs.pdok.nl/yasgui>

The road ahead

- Which districts are under- or overpublishing?
- What are good geo-ranges to offer citizens in notification service? District, street, radius?
- Offer search interface for
 - publications of similar type in street/city/province;
 - publications related to same legal source;
 - older publications for the same address;
 - publication types (higher granularity)

Open government and the Semantic Web: a field report

- Official gazette currently almost FAIR
- Enriching publications with knowledge graphs
 - enables data quality monitoring/analysis
 - enables custom open government
 - prepares for machine learning enrichment

Open government and the Semantic Web: a field report

Q&A