

Open Data Node

Platform and Methodology



@

 itapa

Peter Hanečák <peter.hanecak@eea.sk>, EEA s.r.o.

ITAPA, 4 November 2015



Who am I

- Peter Hanečák <peter.hanecak@eea.sk>
- member of COMSODE project
 - architecture, design and implementation of ODN
- enthusiast in many things “Open”,
active in NGOs and other communities
 - member of OpenData.sk and SOIT
 - Fedora Linux packager



<https://www.linkedin.com/in/peterhanecak>
<https://twitter.com/PHanecak>



<http://www.eea.sk/>



COMSODE



Agenda

- What is COMSODE
- What is COMSODE Methodology
- **What is Open Data Node (ODN)**
- **How can ODN help you**
- **Examples of usage**
- Few more details about ODN



COMSODE

COMSODE

Components Supporting the Open Data Exploitation

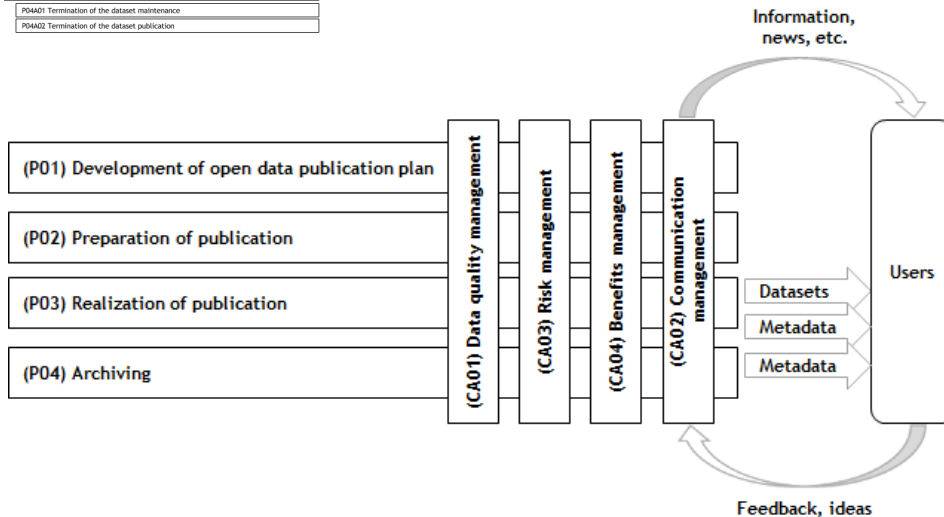
- main target: **publication platform for Open Data**
 - software tool
- supplemental goal: **methodology for publication of Open Data**
 - mainly for those with little or no experience with Open Data
 - because software as of itself is useless for such people, organizations
- validation: **pilots**
 - pilots by 3rd parties
 - pilot by COMSODE itself: 150 datasets + 3rd party-like Search app by Spinque



COMSODE

COMSODE Methodology

Phase P01 Development of open data publication plan
P01A01 Analysis of data sources
P01A02 Identification of datasets for opening up
P01A03 Determination of target level of openness
P01A04 Effort estimation
P01A05 Definition of the open data publication plan
Phase P02 Preparation of publication
P02A01 Data sources access configuration
P02A02 Definition of the catalogue record schema and the target data catalogues
P02A03 Description of the datasets
P02A04 Selection and implementation of the software tools
P02A05 Definition of the approach to the dataset publication
P02A06 Design and implementation of the ETL procedures
P02A07 Testing of the ETL procedures
P02A08 Licensing
Phase P03 Realization of publication
P03A01 Initial publication of the dataset
P03A02 Data cataloguing
P03A03 Dataset maintenance
Phase P04 Archiving
P04A01 Termination of the dataset maintenance
P04A02 Termination of the dataset publication



- publication plan
- preparation of publication
- realization of publication
- archiving

reference: <http://opendatanode.org/product/methodology-for-od-publishing/>



Open Data Node - what is is?



- **Open Source platform for publication of Open Data**
 - allowing those having some data to publish/share it
 - allowing users of data to get it, understand it and use it
- **accompanied with Methodology**, to give wider perspective

Open Data Node - how can it help?

problem:

How to publish/share
data from internal systems
as Open Data
in sustainable way?



Open Data Node - how can it help?

- step 1: think about what and how to publish
- step 2: create the dataset manually
- step 3: create the dataset manually again
- ...
- step N: **create the dataset manually yet again (having difficulties remembering „how I did it previously?“)**
- ...

- typically in public bodies



- **expensive:** cost growing with frequency of updates
- **error prone:** mistakes creeping in, getting fixed but then occurring again

Open Data Node - how can it help?

- step 1: think about what and how to publish
- step 2: create automated publication process
- step 3: **let server run the publication periodically**
- step 4: from time to time check whether it still works



solution: automation

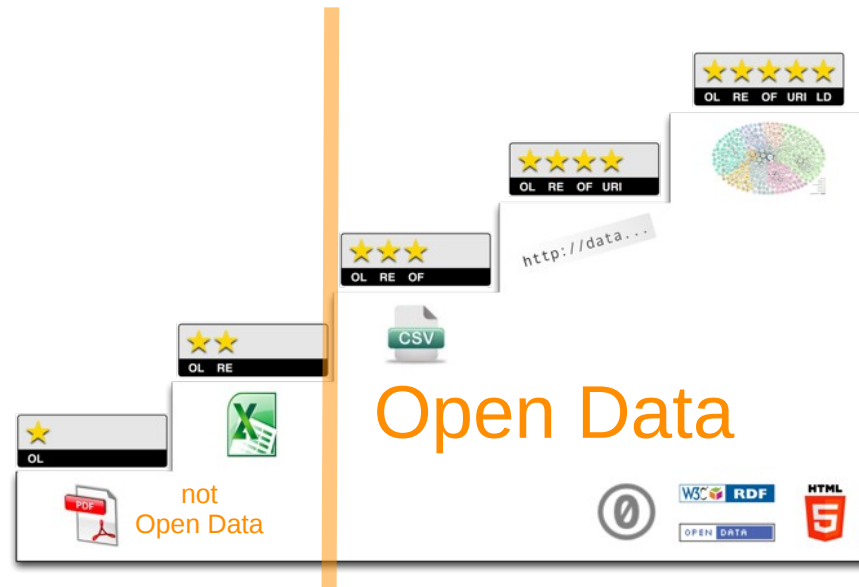
have your published data always up to date, but ...

... avoid repeated manual publication

Open Data Node - how can it help?

most common use-cases: 2* -> 3*+

- **input:** XLS, SQL DB, ...
- **transformations:**
XLS, SQL -> CSV, „bad CSV“ -> CSV,
CSV -> Linked Data
- **output:**
 - tabular/relational data: CSV, REST API
 - Linked Data: RDF, SPARQL endpoint



getting through the border
into the “Open” realm

Examples of ODN usage

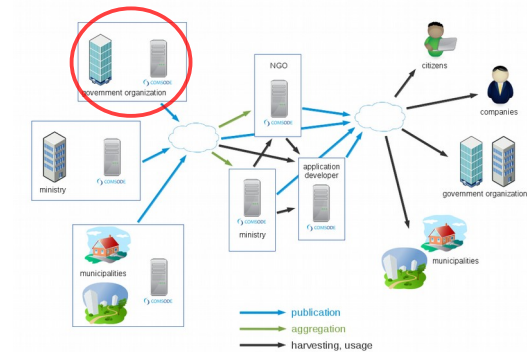
- SNK
- SAZP
- COI
- eDemokracia

Example: SNK

SNK uses ODN to:

- get the data from old bibliographic information system
- convert it from „little know“ format into Open Data form:
 - for quick overview: CSV
 - for more complicated uses: modern Linked Open Data form

visit <http://data.snk.sk/>



Open Data Node

Log In

Datasets Organizations

/ Organizations / snk / Slovak National Library ...

Slovak National Library bibliographic data

Followers
0

Organization

snk
There is no description for this organization

Social

Slovak National Library bibliographic data

Experimental dataset subsetting Slovak National Library's bibliographic data

Data and Resources

CSV dump [Explore](#)

RDF dump [Explore](#)

SPARQL endpoint [Explore](#)

library

Additional Info

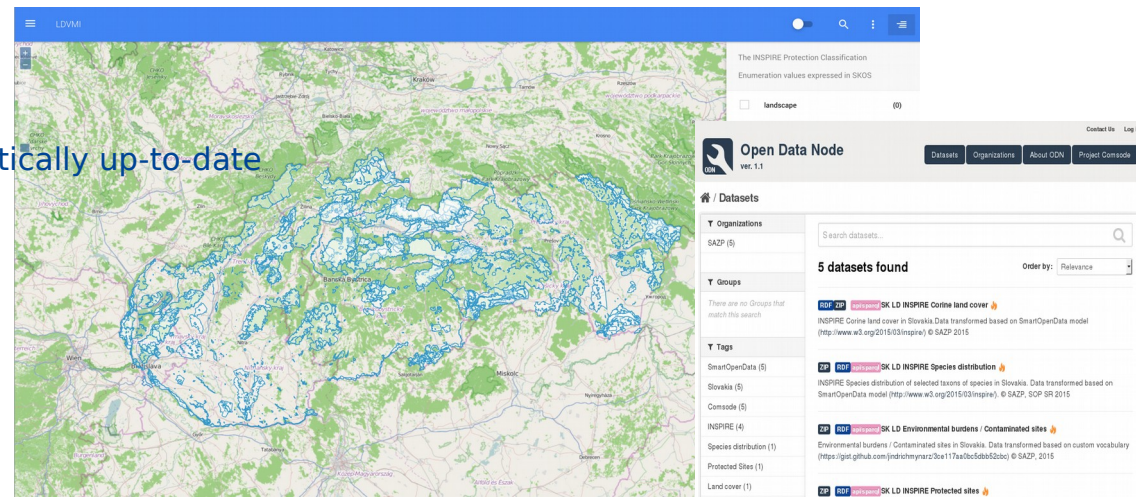
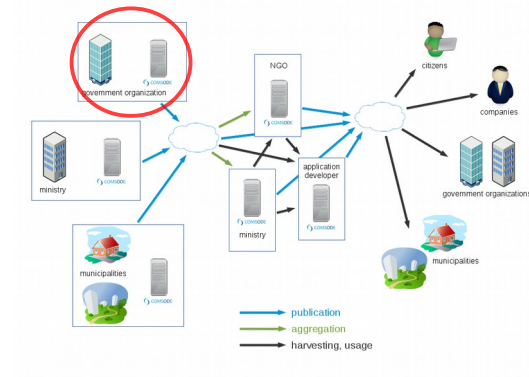
Field	Value
Source	http://www.snk.sk
Last Updated	October 13, 2015, 13:26

Example: SAZP

SAZP uses ODN to:

- tap internal sources of data
 - SAZP staff works with and updates internal sources
- have the data automatically published as Linked Open Data
 - published data is kept automatically up-to-date
- create visualizations

visit <http://data.sazp.sk/>



Example: ČOI

ČOI uses ODN to:

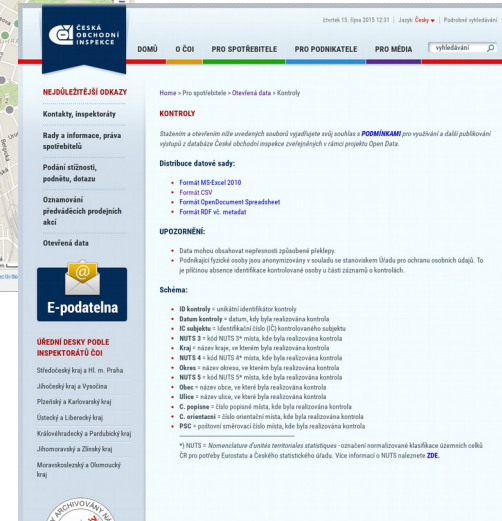
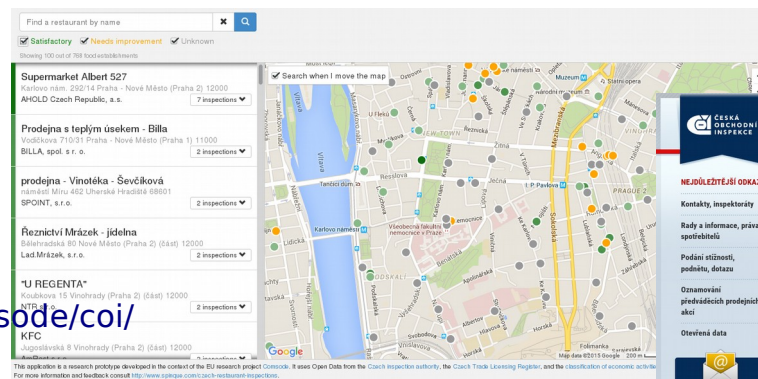
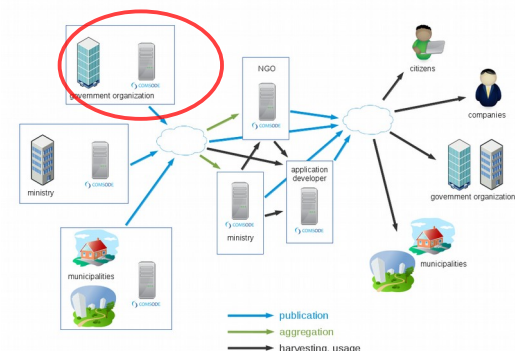
- tap internal sources of data
- have the data automatically published in multiple formats (XLS, CSV, ODS, Linked Data)

- encourage usage:

— demonstration app:

<http://devel.spinque.com/comsode/coi/>

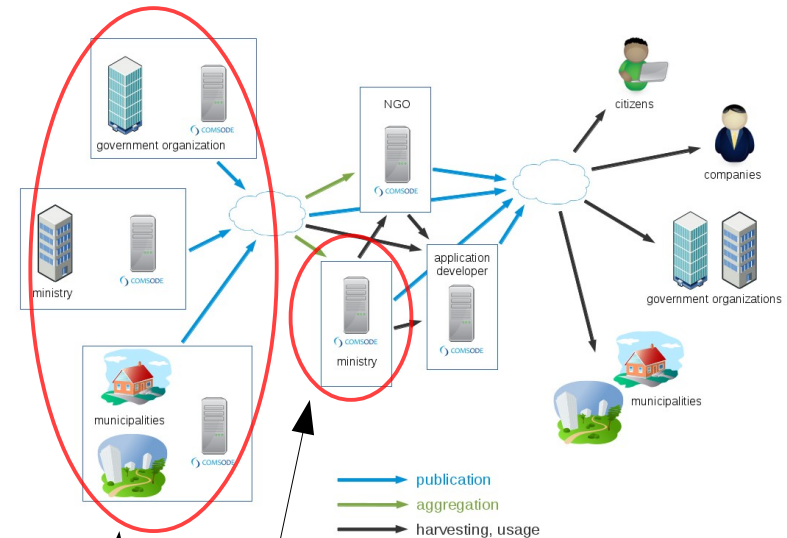
visit <http://www.coi.cz/cz/spotrebiteel/otevrena-data/>



Example: eDemokracia

In eDemokracia project, ODN is used as:

- **centralized component**
- **de-centralized component**



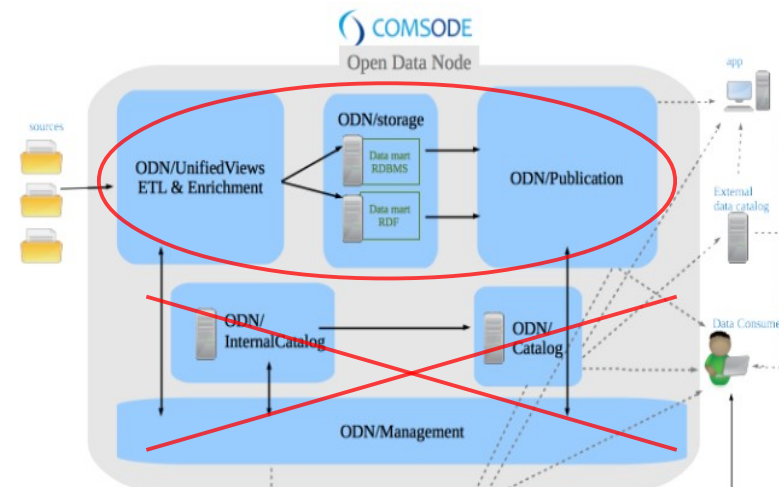
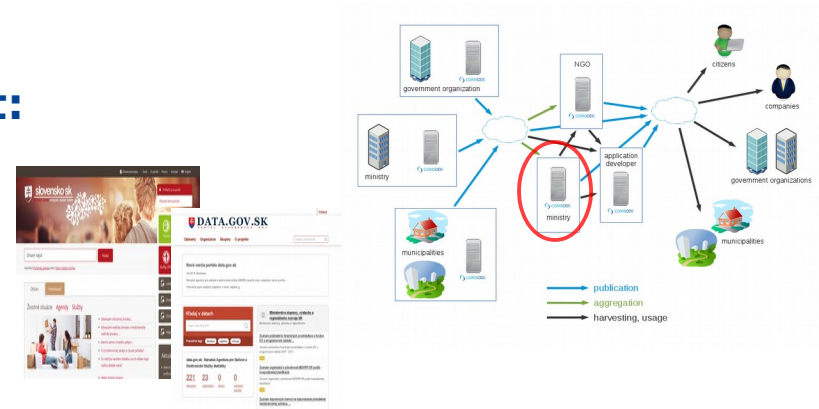
centralized component

de-centralized component

Example: eDemokracia

ODN as part of centralized component:

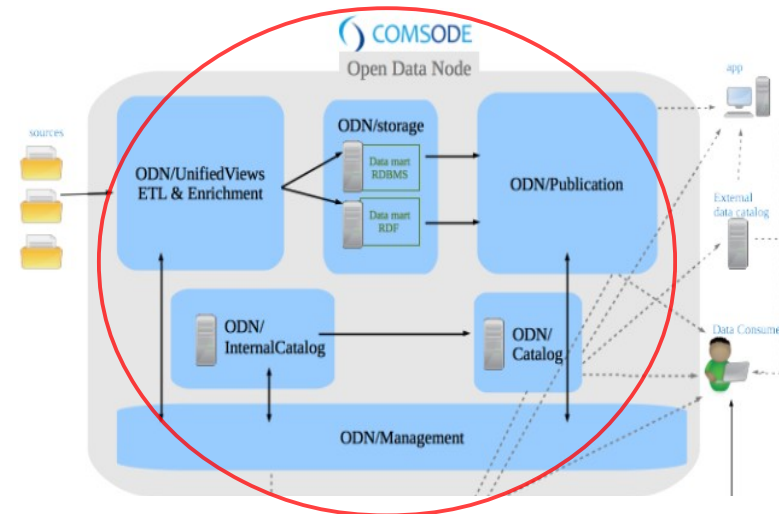
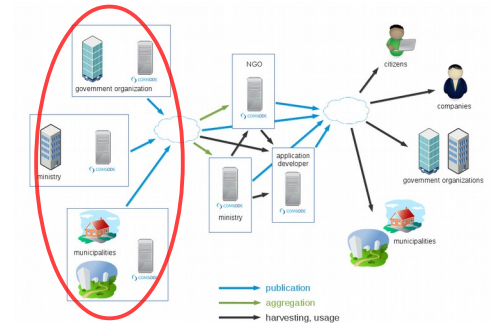
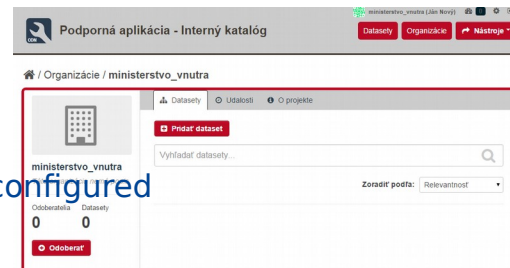
- heavily customized ODN to serve as:
 - data.gov.sk
 - central Open Data storage for whole SR
 - (i.e. Open Data extension for slovensko.sk)
- ETL and other functionality aimed at:
 - help with publication of data from other central systems
 - unification, integration and enrichment of partial datasets published by various Slovak organizations



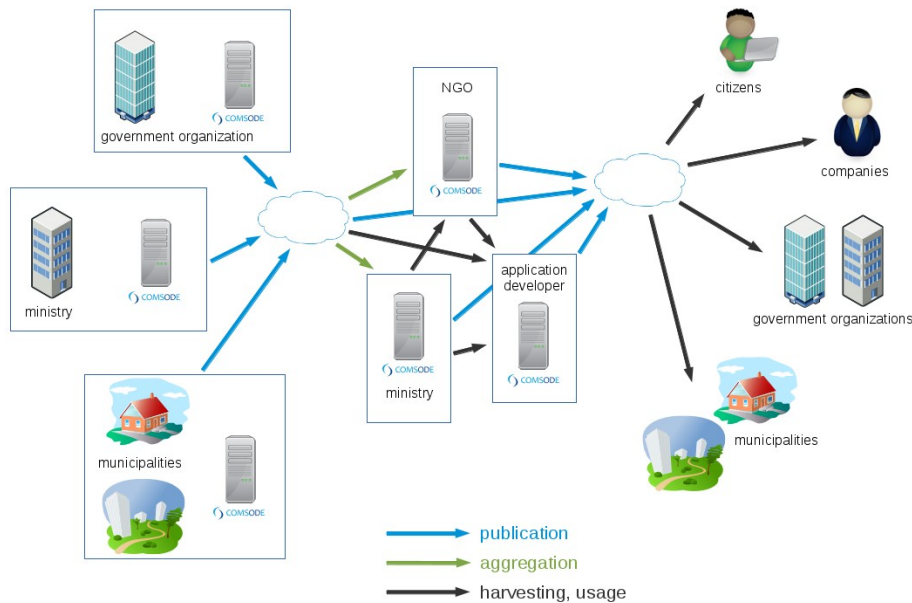
Example: eDemokracia

ODN as de-centralized component:

- ODN with little customization
 - central catalog and storage pre-configured
 - etc.
- helping with cases where central solution is not possible or desired:
 - need to anonymize data before publication
 - volume of data is too big even for fast LAN/WAN
 - etc.



Few more details about ODN



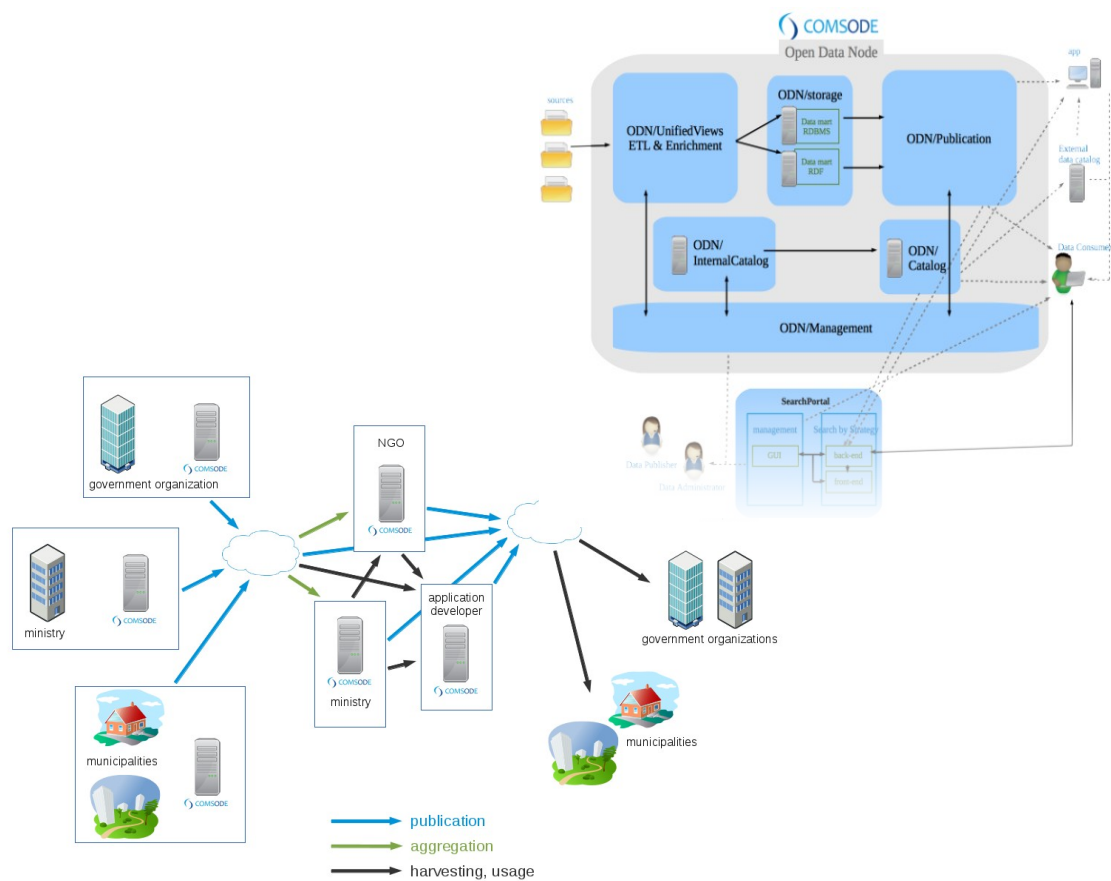
ODN can be used by:

- data publishers
- data users

Many publishers are also users, thus the data ecosystem is quite complex.

ODN can be used in many roles within that ecosystem.

Few more details about ODN



- platform supporting whole OD publishing process
- modular design
- allowing to create distributed network of nodes
- able to be integrated to existing infrastructure

UnifiedViews



midPoint



[*www.comsode.eu*](http://www.comsode.eu)