Spatial Data Infrastructures (SDI)
COMMIT TO PUTTING OPEN IN ACTION

OPEN ACCESS WEEK 2016
Where are we going?
The DUPC2 proposal defines a particular goal for SDI related activities:

Support and promote the Spatial Data Infrastructure (SDI). The SDI aims to make spatial data generated in DUPC funded and other projects available for strategic partners and wider audiences.

In DUPC2 the SDI prototype (started in DUPC1) will be further developed with and shared with strategic partners.

**Targets:**
- 75% of the spatial datasets produced under DUPC2 available in the SDI
- > 50% of appr. 20 strategic partners contributing to the SDI
What is SDI?

- Enabling environment that supports easy access to and utilization of geospatial data.

- SDIs are more than just data repositories:
  - **Discovery**
    - catalogues, search engine
  - **Visualization**
    - interactive maps, infographics, real time data, etc.
  - **Evaluation**
    - quality, metadata
  - **Access** to geospatial data and information
    - web-based, apps, services, GIS
Benefits of SDI

- to organizations by more efficient data collection and processing and **reduce duplicate efforts**
- to scientists through better research when **more data** is available and can be **linked with models and tools**
- to public administration, citizens, and businesses through **better services**
- through **innovation** for new services and business
- greater **democratic accountability**
- to society through **better management** of the common environment
Why is data not easily shared?

- Data collection costs relatively high compared with the GIS hardware and software
- GIS users tend to develop their own data sets, even if there are existing geospatial data sets available for them, because
  - they may not know available existing data sets
  - access to these data sets is difficult
  - they are not used to sharing data sets with other sectors and/or organisations
  - existing geospatial data sets stored in a certain GIS system may not be easily exported to another system.

→ Duplicate efforts in geospatial data development, which sometimes hinders further dissemination of GIS applications in local, national, regional and global circumstances.
Why is sharing so difficult?

- a priori suspicion of the quality of third party data is common.

- a priori presumption that the institutions’ own data (generally deemed of high quality by the latter) may be “wrongly” used if shared with a third party, or even that ownership thereof may be lost.

- fear that other users discover the poor quality of their data by sharing them.

Source: UNECA, The SDI Handbook for Africa
What is the value of data?

Value of information is the amount a decision maker would be willing to pay for information prior to making a decision.

Source: European Commission
A piece of data is **open** if anyone is free to **use**, **reuse**, and **redistribute** it — subject only, at most, to the requirement to **attribute** and/or **share-alike**.

Publicly funded **data** are a public good, produced in the public interest and thus **should be freely available** to the maximum extent possible.
Types of licenses

• Libre licenses
  
  Share-alike or Copyleft:
  - copies and modifications of the original work must be available under the same or similar license
    - GNU General Public license
    - Creative Commons Attribution-ShareAlike
  
  Permissive or copyfree:
  - do not require derivative works to be licensed under the same license as the original work.
    - Creative Commons Attribution Alone
  
  Public Domain:
  - are those whose intellectual property rights have expired, have been forfeited, or are inapplicable
    - Creative Commons Public Domain
CC-NC licenses

- Creative Commons Non-Commercial licenses: **non-libre license**
  - Attribution + Noncommercial
    - E.g. adopted by UNESCO-IHE
    
    ![CC BY NC license](image)

  - Attribution + Noncommercial + ShareAlike
    
    ![CC BY ND license](image)

  - Attribution + NoDerivatives
    
    ![CC BY NC SA license](image)
A sound data policy should look carefully at ways to remove the potential risks so that the data producers are happy and confident in sharing their data.

Commitment of all contributing stakeholders is needed.

Data in an SDI doesn’t always have to be Open Data, you can choose the license.

Data policy is linked to business model.
Costs of selling data

Overestimation of return on investment of selling the data:
- legal costs of creating and enforcing restrictive licenses
- development costs of restricting access and use of data
- administrative costs of issuing licences
- sales and marketing costs to promote the data

Figure 3: Economic Benefits of Open Data

<table>
<thead>
<tr>
<th>Benefit to Government</th>
<th>Drive Revenue through multiple areas</th>
<th>Cut Costs and Drive Efficiency</th>
<th>Generate Employment and develop future-proof skills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>▪ Increased tax revenues through increased economic activity</td>
<td>▪ Reduction in transactional costs</td>
<td>▪ Create jobs in current challenging times</td>
</tr>
<tr>
<td></td>
<td>▪ Revenues through selling high value added information for a price</td>
<td>▪ Increased service efficiency through linked data</td>
<td>▪ Encourage entrepreneurship</td>
</tr>
<tr>
<td>Benefit to Private Sector</td>
<td>▪ Drive new business opportunities</td>
<td>▪ Reduced cost by not having to invest in conversion of raw government data</td>
<td>▪ Gain skilled workforce</td>
</tr>
<tr>
<td></td>
<td>▪ Better decision making based on accurate information</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Capgemini Consulting Analysis
Strategic risks of selling data

- **Low willingness to pay** (like with music, movies, etc.). People are becoming less prepared to pay for digital products that can easily be copied and shared with others.

- Competitors or communities might also be able to undermine your data business by releasing their data as **open data**.
Open Data Benefits

€ 325 billion
direct market size
EU28+ for 2016-2020

25,000
jobs created

Open Data jobs

75,000 100,000
2016 2020

€ 1.7 bn
government
cost savings

Source: European Commission
SDI: Spatial Data Infrastructure
Added value, available for the water sector

Development of apps and services

Models

Decision support systems (DSS)

SDI

Entrepreneurs/Researchers

Agriculture

Health

Drinking water

Sanitation

Economy

Added value, available for the water sector
Spatial Data Infrastructures

Sharing of:
- Spatial data
- Metadata
- Documents (e.g. Open Access papers)
- Profiles (social media)
Analogous to a road:
- Reliable environment allowing the movement of data
- Maximize the reuse of data: open standards
- All about reuse: data, capabilities, skills, investments,…
- Sharing: data, knowledge, …
- Learning from others: collaboration and co-operation

Working smarter not harder!
Caution: speed bumps ahead!
Hydrological data are difficult to integrate

- Incompatibilities (formats, models)
- Missing documentation (metadata)
- Data fragmentation and replication
- Data policies
Technical
- Machine to machine communication
- Software module interaction
- APIs, formats, schemas

Semantic
- Common understanding
- Common concepts, terms, ...
- Interdisciplinary special vocabularies

Legal
- Data rights
- Ownership
- Responsibility
- Copyright

Human
- Cooperation
- Collaboration
- Training

Interoperable System Aspects
Intangibles more important than tangibles!
Digitization is necessary!

Data Nakumatt!
SDI: a universal adapter
Open Geospatial Consortium

Geospatial and location standards for:

- Aviation
- Built Environment & 3D
- Business Intelligence
- Defense & Intelligence
- Emergency Response & Disaster Management
- Geosciences & Environment
- Government & Spatial Data Infrastructure
- Mobile Internet & Location Services
- Sensor Webs
- University & Research

http://www.opengeospatial.org
OGC Services

Data

Web Mapping Services (WMS)
Web Feature Services (WFS)
Web Coverage Service (WCS)

Metadata

Catalogue Services for the Web (CSW)

Processing

Web Processing Service (WPS)
Interoperable standards
The globalised SDI standards allow exchange of data between different scales.
International SDIs

- GEOSS: Global Earth Observation System of Systems
- GMES: Global Monitoring of Environment and Security
- TIGER: Technology Informatics Guiding Education Reform
- UNSDI: United Nations SDI
- INSPIRE: Infrastructure for Spatial Information in the European Community
- Africa SDI
- GMES for Africa
- ...
Metadata should clarify terms of use of data
The user should be able to determine if the data is suitable for his/her purpose
SDI Metadata

- What information do you need in order to choose the right dataset?
  - Area coverage
  - Temporal resolution
  - Raster/vector
  - Projection
  - How to get the data
  - License
  - Raw data? Processed? How?
  - Contact
  - …
SDI Metadata

- Metadata is a detailed description of a dataset that includes information about the content, quality, structure, accessibility and other characteristics of the data.
- Detailed and robust metadata must document the project, datasets, and services in such a way that data can be transmitted, interpreted, reused, and understood.
- Use of metadata standards is critical for data discovery, integration, and sharing.
Choice for open source software

- Open source software often provide better **interoperability** between internal and external components
- Open source software often use **international standards**
- Proprietary tools impose limits to the user; it is difficult to make improvements, complicated to change supplier (**lock in**)
- Open source software is improved continuously thanks to the participation of the **user community**
  - Quick implementation of new developments (at the forefront of technology)
  - Opportunity for **innovation**
Software

We seek an open source software stack that can serve our needs today.

We do embrace new technology, but will not allow it to delay us from proper data management now.
SDI Software

- PostGIS/PostgreSQL for storing data
- Geoserver for publishing data
- OpenLayers for visualisation of data
- pycsw for storing and searching metadata
• Open Source Geospatial Content Management System.
• GeoNode is a web-based application and platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructures (SDI).
• It is designed to be extended and modified, and can be integrated into existing platforms.

➤ Browse and search for geospatial data and web services
➤ Upload, manage, and share geospatial data and documents
➤ Create and share interactive maps
➤ Collaborate and interact with other users
Challenges

• Limited access to internet
• Limited capacity to:
  – Install SDI systems
  – Maintain SDI systems
  – Data treatment
  – Quality assurance
• Business plan
  – return on investment/valorisation
• Data Policy: Open Access / Restricted Access
• Standardisation
• Quantity of metadata needed
• User friendly GUI
• Time…
Challenges