The Geodata Warehouse of the Swiss Federal SDI

Dr. Patrick Gamma
Federal Office of Topography swisstopo
patrick.gamma@swisstopo.ch

Esri European User Conference
Salzburg, 15 October 2015
Agenda

• Introduction
• Legal Basis, Terms & Definitions
• Life Cycle of Geodata
• Architecture of the Geodata Warehouse
• Technology
• Next Steps & Conclusions
What is the value of your Geodata?

Imagine that your organisation has lost an important dataset…

• Could you recreate it at all?
• Possibly with some information loss?
• Price for recollecting / reconstructing?
• What is the impact of not having the data?
• Good strategies (besides classic IT approach):
  – Use the data
  – Open data
• Our strategy: the above, plus a Geodata Warehouse, a «living archive»
Could you archive a geodataset from 1995 without any loss?

Imagine that your organisation should file a geodataset, that was used or created 20 ago…

• Could you still read it?
• Is the expertise available to convert it into the required formats?
• Would you be able to provide the necessary information, e.g. Documentation, Metadata, Data Model, Representation Model?
Just some thoughts about data loss...

- Data loss is one of the greatest risks for a NMA on the enterprise level, caused by e.g.
  - Technical problems, or even on purpose
  - Deleting the wrong data
- Some losses could only be detected after years, at the time when someone tries to access the data
- One can not keep all data → limited storage costs
- Even if you keep all data, you probably won’t have the software to read it in the future
- Decide about which datasets should be kept and for how long, before being sent to the archives
- You should provide the documentation about a dataset while the know-how (people!) is still available
- Future effort for preparing & sending the data to the archives is minimal (ROI)
Legal Basis, Terms and Definitions

• The *Federal Act on Geoinformation* (GeoIA, 2008) and the *Ordinance on Geoinformation* (GeoIO, 2008) essentially define 2 states / conditions in which a geodata set can be:

  «in use»

  • Long-term availability  ➔ GIS Community
    The authorities responsible for the collection, updating and management of official geodata guarantee the long-term availability of this geodata.

  «not in use»

  • Archived  ➔ No GIS
    The Federal Archives conserves the geodata
The «Life Cycle» of Geodata (1)
The «Life Cycle» of Geodata (2)

Long-term Availability in the FSDI

Geodata Warehouse

Production

Release

Continuous update

Integration

Storage

A
A_{2010}
A_{2005}
A_{1999}

Delivery

Publication

Archiving

Archives

Federal Office of Topography swisstopo

Esri European User Conference, Salzburg
15 October 2015
The Aim of Long-term Availability…

…is to *conserve* official geodata for a *limited period*…

…in such a way that their quantity and quality are maintained…

…and they are available for *continuous active use*.

• Online availability should extend not just to the data fonds that are current at a given time but also to defined older snapshots (in the sense of time series).

### Long-term Availability vs. Archiving of Geodata

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Long-term availability</th>
<th>Archiving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preservation period</td>
<td>Limited in time</td>
<td>Unlimited in time</td>
</tr>
<tr>
<td>GIS functionality</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Formats</td>
<td>Current formats, ready for use</td>
<td>Archivable formats</td>
</tr>
<tr>
<td>Principle</td>
<td>Keep the data accessible, both current and older releases</td>
<td>Keep permanently and safely</td>
</tr>
<tr>
<td>Frequency of use</td>
<td>Frequently</td>
<td>Seldom</td>
</tr>
</tbody>
</table>

«Hot pizza from pizza service»  «Frozen pizza from home delivery»
Long-Term Availability
Conceptual View of the FSDI

Production
- Metadata
- Data models
- Representation models
- Geodata
- Documentation

Preservation
- FSDI Data Dictionary
  Catalogue, Status, Cross Reference Preservation Planning
- Model Repository
  Data & Representation models
- Geodata Warehouse
  Geodata
- Document Management
  Accompanying Documentation

Utilisation
- Geoservices
- Web
  Mapviewer, geocat.ch, Model Publication, Shop
- Federal Administration GIS Platform
- Federal Archives SFA
- Other Publication Channels

«Ingest once, use anywhere»
Requirements for the GDWH

• Ensure the long-term availability of official geodata
• Import, preserve and export the geodata for use within the FSDI
• Reduce or limit redundancy
• Minimise software manufacturer dependencies
• For swisstopo and for other federal agencies (replace swisstopo legacy application, make it useable for others)
• Fit the system in the overall FSDI framework and the corresponding processes
OAIS Open Archival Information System
Reference Model ISO 14721:2012

Submission Information Package
Archival Information Package
Dissemination Information Package
Geodata Warehouse FSDI System Architecture

**Business Logic Tier**
- GDWH Data Dictionary Service
- Geodata Delivery
- Geodata Validation & Ingest
- Geodata Transformation

**Data Tier**
- DB-Access Layer
- Transformation Functions
- Consistency Checks

**Aggregation Layer**
- Generate Transform Aggregate

**Filesystem Layer**
- Standardised Directory Structure
- Standardised Data Formats
- Consistency Checks
- Accompanying documents

**Services**
- Interfaces to FSDI
- Local Catalogue
- Current GIS technology
- «Product dependent»
- Replacable/exchangeable
- Generated from Filesystem Layer
- «Product independent»
- Original-/Master Data
- Archivable
Filesystem & Aggregation Layer Demo
Geodata Warehouse FSDI
Migration Strategy for Software and Data

- **Business Logic Tier**
  - 2015: Services, Software 2015
  - 2020: Services, Software 2020
  - Migrate

- **Aggregation Layer**
  - 2015: Software 2015, GIS Data 2015
  - 2020: Software 2020, GIS Data 2020
  - Migrate
  - Convert

- **Filesystem Layer**
  - 2015: Original Data 2015
  - 2020: Original Data 2020
  - No change

- **Services**
  - Interfaces to FSDI
  - Local Catalogue
  - New/adapted functionality
  - Current GIS technology
  - «Product dependent»
  - Replacable/exchangeable
  - Generated from Filesystem Layer
  - ...or converted from Aggregation layer
  - «Product independent»
  - Original-/Master Data
  - Archivable
  - Untouched
Geodata Warehouse FSDI
Migrating over Time

2015

Business Logic Tier
- Services
  - Software 2015

Aggregation Layer
- Software 2015
- GIS Data 2015

Filesystem Layer
- Original Data 2015

Convert

2020

Business Logic Tier
- Services
  - Software 2020

Aggregation Layer
- Software 2020
- GIS Data 2020

Filesystem Layer
- Original Data 2020

Generate

2025

Business Logic Tier
- Services
  - Software 2025

Aggregation Layer
- Software 2025
- GIS Data 2025

Filesystem Layer
- Original Data 2025

Generate

2030

Business Logic Tier
- Services
  - Software 2030

Aggregation Layer
- Software 2030
- GIS Data 2030

Filesystem Layer
- Original Data 2030

No change
Consistency Checks / Validation for Raster and Vector Data (1)

Geodata Warehouse Raster

Last nightly maintenance module finished with Severity-Level Error

**Warnings: 3**

<table>
<thead>
<tr>
<th>Warning</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOP25_LV95</td>
<td>08:10 00:00:07 WARN: [DOP25_LV95] has Inject-Operation [Injector] checked in since [7 hours, 34 minutes, 37 seconds]</td>
</tr>
<tr>
<td>SMR25_LV95</td>
<td>08:10 00:00:36 WARN: [SMR25_LV95] has newer obj.xml than the timestamp</td>
</tr>
<tr>
<td></td>
<td>08:10 00:00:59 WARN: [SMR25_LV95] has obj.xml that is newer than Geodatabase (obj.xml: 07.10.2015-20.38 timestamp: 01.01.0000-00.00)</td>
</tr>
</tbody>
</table>

**Errors: 5**

<table>
<thead>
<tr>
<th>Error</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMR25_LV95</td>
<td>08:10 00:00:08 ERROR: !!! Errors in rebuilding Geodatabase for [SMR25_LV95]</td>
</tr>
<tr>
<td></td>
<td>08:10 00:00:34 ERROR: -&gt; [SMR25_LV95] has updated obj.xml file(s)</td>
</tr>
<tr>
<td></td>
<td>08:10 00:00:34 ERROR: -&gt; [SMR25_LV95] has updated grid.xml file</td>
</tr>
<tr>
<td></td>
<td>08:10 00:00:34 ERROR: -&gt; [SMR25_LV95] has updated release file</td>
</tr>
</tbody>
</table>

*** Executing E:\\v4s\Server\ServerComponents\GeodatabaseBatch.exe *** Arguments -op syncdb -g SMR25_LV95 -e P *** Executing E:\\v4s\Server\ServerComponents\GeodatabaseBatch.exe *** Arguments -op syncdb -g SMR25_LV95 -e P The Database
## Validation Cockpit (Integration)

### Interval

<table>
<thead>
<tr>
<th>Date</th>
<th>Offset</th>
<th>LOAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>01.10.2015</td>
<td>1 Week</td>
<td></td>
</tr>
</tbody>
</table>

The validations in the interval 24.9.2015 - 1.10.2015 will be loaded.

<table>
<thead>
<tr>
<th>State</th>
<th>GdsKey ▲</th>
<th>Title</th>
<th>RunDate</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Daily_DHM25</td>
<td>01.10.2015 05:08:20</td>
</tr>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Daily_DHM25</td>
<td>30.09.2015 05:09:51</td>
</tr>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Daily_DHM25</td>
<td>29.09.2015 05:08:40</td>
</tr>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Daily_DHM25</td>
<td>28.09.2015 05:08:11</td>
</tr>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Daily_DHM25</td>
<td>27.09.2015 05:08:38</td>
</tr>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Daily_DHM25</td>
<td>24.09.2015 05:09:39</td>
</tr>
<tr>
<td>OK</td>
<td>DHM25</td>
<td>Weekly_DHM25</td>
<td>25.09.2015 23:00:12</td>
</tr>
<tr>
<td>OK</td>
<td>DOM</td>
<td>Weekly_DOM</td>
<td>25.09.2015 23:00:12</td>
</tr>
<tr>
<td>OK</td>
<td>DOM</td>
<td>Daily_DOM</td>
<td>01.10.2015 05:00:05</td>
</tr>
</tbody>
</table>
Fundamental Factors supporting LTA in the Geodata Warehouse FSDI

Archivable Data Formats:
- Standardised
- Open
- Documented

Consistency Checks:
- Checksums
- File Dates
- Structures
- Mapping
- Daily, Weekly, Monthly

System Architecture:
- Dual Data Tier
- Human Readable Structures
- Service oriented

Long-term Availability in the Geodata Warehouse:
- Backup
- Mirroring
- Access Control
- Inhouse Servers

Classic IT Measures:
- Dual Data Tier
- Human Readable Structures
- Service oriented
Geodata Warehouse FSDI
Tools and Technologies

Services
- ArcGIS Server
- FME Server
- C# .Net
- Python
- JavaScript

Aggregation Layer
- ArcGIS Desktop
- FME Desktop
- GDAL Library
- REFRAME Library
- INTERLIS Tools
- LAS Tools
- ESRI FGDB
- TIFF
- TXT, CSV
- PDF/A
- ILI, ITF
- XML

Business Logic Tier
- GDWH Data Dictionary Service
- Geodata Delivery
- Geodata Validation & Ingest
- Geodata Transformation

Data Tier
- DB-Access Layer
- Transformation functions
- Consistency Checks

Filesystem Layer
- Standardised Directory Structure
- Standardised Data Formats, e.g. model based
- Consistency Checks
- Accompanying documents

Represent Transform
Geodata Warehouse FSDI
Service oriented and modular Elements (1)
Additional software components to support specific data types can be added if necessary, e.g. LAS or 3D.
Next Steps and Conclusions

• The new Geodata Warehouse enables the Swiss Federal Spatial Data Infrastructure to provide an environment to preserve geodata (sustainable / long-term availability), both for swisstopo itself and other federal agencies

• Swisstopo is currently adapting the remaining systems of the FSDI to achieve this goal, including the processes

• A geodataset becomes more interesting with increasing age → no need to archive
The End