

# **Využití GeoKettle pro zpracování výměnného formátu Registru územní identifikace, adres a nemovitostí**

Jan Růžička

GIVS 2012 Praha  
20. - 21. 9. 2012

# Úvod

- Výměnné formáty založené na XML
- Znalost jazyků XML (případně XSD)
- Využít programátorů pro realizaci importu
- Využít ETL nástroje - pro běžné poučené uživatele

# ETL nástroje

- **Extract Transform Load**
- Načti data
- Proveď transformaci (různých druhů - např. změna struktury atributů, transformace souřadnic)
- Načti data na jiné místo (např. do databáze PostGIS)

# ETL nástroje

- HUMBOLDT Tools - Image pro VirtualBox:  
<http://gis.vsb.cz/ruzicka/down/humboldt/>
- GeoKettle
- Spatial Data Integrator
- Jazyk XSLT
- GDAL/OGR
- Jazyk SQL
- ...

# RÚIAN

- Registr územní identifikace, adres a nemovitostí
- Správa: ČÚZK a ČSÚ
- Základní registr veřejné správy

# Výměnný formát RÚIAN

- Založen na XML
- Využívá jmenný prostor GML

# Výměnný formát RÚIAN / 2

```
<vf:AdresniMisto gml:id="AD.12806471">
  <adi:Kod>12806471</adi:Kod>
  <adi:CisloDomovni>104</adi:CisloDomovni>
  <adi:Psc>25206</adi:Psc>
  <adi:StavebniObjekt>
    <soi:Kod>12717614</soi:Kod>
  </adi:StavebniObjekt>
  <adi:Ulice>
    <uli:Kod>434205</uli:Kod>
  </adi:Ulice>
  <adi:PlatiOd>2011-07-01T00:00:00</adi:PlatiOd>
  <adi:IdTransakce>0</adi:IdTransakce>
  <adi:GlobalniIdNavrhuZmeny>0</adi:GlobalniIdNavrhuZmeny>
  <adi:Geometrie>
    <adi:DefinicniBod>
      <adi:AdresniBod>
        <gml:Point srsDimension="2" gml:id="DAM.1937498" srsName=...">
          <gml:pos>747802.63 1062732.00</gml:pos>
        </gml:Point>
      </adi:AdresniBod>
    </adi:DefinicniBod>
  </adi:Geometrie>
</vf:AdresniMisto>
```

# GeoKettle - příklad

- Extrakce adresních míst z výměnného formátu systému RÚIAN
- Sestavení transformace
- Identifikace elementů s adresními místy
- Identifikace elementů se souřadnicemi
- Nastavení transformace z gml:pos do WKT MULTIPOINT



# GeoKettle - Transformace

The screenshot displays the GeoKettle GUI window titled "GeoKettle's GUI - adresy". The interface includes a menu bar with "File", "Edit", "View", "Repository", "Transformation", "Job", "Wizard", and "Help". Below the menu is a toolbar with icons for file operations and execution. The main workspace shows a transformation job with two steps: "Get data from XML 2" and "Shapefile File Output", connected by a data flow arrow. A left-hand pane lists various input and transformation steps, with "Get data from XML" highlighted. The top of the workspace shows two tabs: "parcely" and "adresy".

GeoKettle's GUI - adresy

File Edit View Repository Transformation Job Wizard Help

View Design

Steps

- Input
  - Access Input
  - CSV file input
  - CSW Input
  - De-serialize from file
  - ESRI Shapefile Reader**
  - Excel Input
  - Fixed file input
  - GML File Input
  - Generate Rows
  - Generate random values
  - Get File Names
  - Get Files Rows Count
  - Get SubFolder names
  - Get System Info
  - Get data from XML**

parcely adresy

100%

Get data from XML 2

Shapefile File Output

# GeoKettle - Input

Get XML Data

Step name

File | Content | Fields

XML source from field

XML source is defined in a field?

XML source is a filename?

Read source as Url

get XML source from a field

File or directory

Regular Expression

Selected files:

#	File/Directory	Wi
1	/data/Vyuka/SkoleniDMVSZK/Metadata/6_TS_ETL/RUIAN/data/20120131_OB_539490_UZSZ.xml	

# GeoKettle - Input

```
//vf:AdresniMisto[count(./adi:Geometrie) > '0']
```

The screenshot shows the 'Get XML Data' dialog box in GeoKettle. The window title is 'Get XML Data'. The 'Step name' field contains 'Get data from XML 2'. The 'Content' tab is selected, and the 'Fields' sub-tab is active. The 'Settings' section includes:

- Loop XPath: `//vf:AdresniMisto[count(./adi:Geometrie) > '0']` (with a 'Get XPath nodes' button)
- Encoding: UTF-8 (dropdown menu)
- Namespace aware?
- Ignore comments?
- Validate XML?
- Use token?
- Ignore empty file?
- Do not raise an error if no files?
- Limit: 0
- Prune path to handle large files: (empty field)

The 'Additional fields' section includes:

- Include filename in output?  Filename fieldname: (empty field)
- Rownum in output?  Rownum fieldname: (empty field)

The 'Add to result filename' section includes:

- Add files to result filename?

At the bottom, there are three buttons: 'OK', 'Preview rows', and 'Cancel'.

# GeoKettle - Input

```
concat('MULTIPOINT ((',  
string(adi:Geometrie/adi:DefinicniBod/adi:AdresniBod/gml:Point/gml:pos), ')))'
```

Step name:

File | Content | Fields

#	Name	XPath	Element	Type	Format
1	id	@gml:id	Node	String	
2	Kod	adi:Kod	Node	Integer	
3	CisloDomovni	adi:CisloDomovni	Node	Integer	
4	Psc	adi:Psc	Node	Integer	
5	StavebniObjekt	adi:StavebniObjekt/soi:Kod	Node	String	
6	Ulice	adi:Ulice/uli:Kod	Node	String	
7	PlatiOd	adi:PlatiOd	Node	String	
8	IdTransakce	adi:IdTransakce	Node	Integer	
9	GlobalniIdNavrhuZmeny	adi:GlobalniIdNavrhuZmeny	Node	Integer	
10	Geometrie	concat('MULTIPOINT ((' , string(adi:Geometrie/adi:DefinicniBod/adi:AdresniBod/gml:Point/gml:pos), ')))'	Node	Geometry	

# GeoKettle - Run

## Execute a transformation

Local, remote or clustered execution

Local execution

Execute remotely

Execute clustered

Remote host

Pass export to remote server

Post transformation

Prepare execution

Start execution

Show transformations

Details

Enable safe mode

Clear the log before execution

Log level

Debugging

Replay date (yyyy/MM/dd HH:mm:ss)

Parameters

#	Parameter	Value	Default value
1			

Variables

#	Variable	Value
1	Internal.Job.Filename.Directory	Parent Job File Directory
2	Internal.Job.Filename.Name	Parent Job Filename
3	Internal.Job.Name	Parent Job Name
4	Internal.Transformation.Repository.Directory	/

Arguments

#	Argument	Value
1		

Launch

Cancel

# GeoKettle - Output

The screenshot displays the GeoKettle GUI for a job named 'adresy'. The interface includes a menu bar (File, Edit, View, Repository, Transformation, Job, Wizard, Help), a toolbar with icons for file operations and execution, and a central workspace showing a workflow diagram. The workflow consists of two steps: 'Get data from XML 2' (represented by an XML icon) and 'Shapefile File Output' (represented by a shapefile icon), connected by a data flow arrow. On the left, a 'Steps' panel lists various input and processing options, with 'Get data from XML' highlighted. Below the workspace, the 'Execution Results' panel is open, showing a table of execution metrics for the two steps.

#	Stepname	Copynr	Read	Written	Input	Output	Updated
1	Get data from XML 2	0	0	504	504	0	0
2	Shapefile File Output	0	504	504	0	1	0

# GeoKettle - Další

- Živé ukázky - když bude zájem, prostor a podmínky
  - Extrakce adresního místa do ESRI Shapefile
  - Extrakce parcely do GML

# Děkuji za pozornost

## Využití GeoKettle pro zpracování výměnného formátu Registru územní identifikace, adres a nemovitostí

<http://gis.vsb.cz/ruzicka/>  
[jan.ruzicka@vsb.cz](mailto:jan.ruzicka@vsb.cz)  
Jan Růžička

GIVS 2012 Praha  
20. - 21. 9. 2012