The rise of crowd-sourcing - how valuable data can we get out of VGI
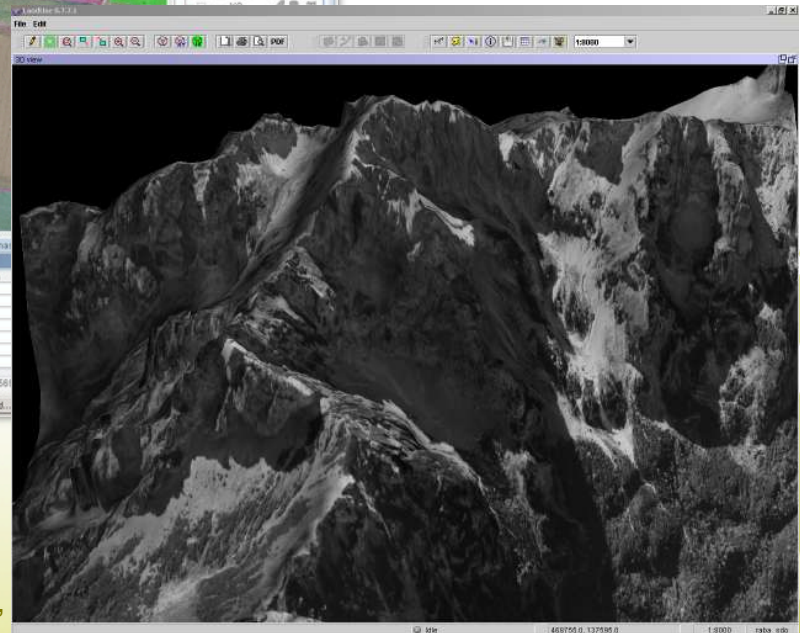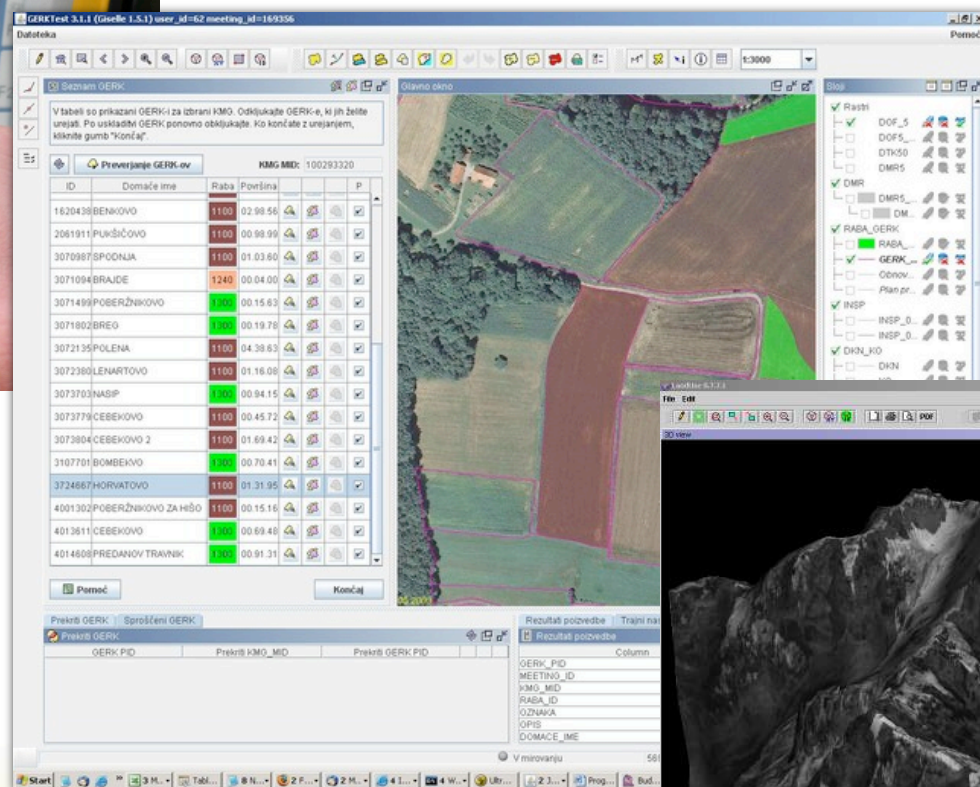
SINERGISE

Grega Milčinski

grega.milcinski@sinergise.com

Contents

- About Sinergise

- VGI trends

- Practical experiences

  - OpenStreetMap

  - Geopedia

About Sinergise

The rise of crowd-sourcing - how valuable data can we get out of VGI,

About Sinergise

About Sinergise

- 8 years in GIS

- 20 people, mostly programmers

- Governmental Solutions
  - Agriculture – IACS
  - Real-estate Management

- We run GIS in Slovenia and going outwards
  - 2008 - IACS in Croatia
  - 2009 – Real Estate in Africa
  - 2010 – IACS in Macedonia

- VGI project Geopedia.si running since 2007

Rise of Volunteered Geographic Information (VGI)

Michael Jones, Google:

The chief Internet evangelist at Google Inc., and one of the founding fathers of the Internet, says he'd like to see a geographic equivalent of Wikipedia — "Geopedia," he dubs it — where anyone could add to the world's geographic know-how.

Jack Dangerman, ESRI:

He worries that even the best-intentioned amateur could provide inaccurate data that could lead to a disaster. "Who wants to dig a hole and run into a pipe?" Dangerman asks.

(GeoWeb, summer 2007)

Short history of Wikipedia

- January 2001 – start

- October 2005 – criticism

  - "Broadly speaking, it's inaccurate and unclear."

  - "I wouldn't have thought of using Wikipedia as a serious reference source"

- December 2005 – "Wikipedia is about as good a source of accurate information as Britannica" (Nature)

- June 2008 – "Encyclopaedia Britannica To Follow Modified Wikipedia Model"

- June 2009 – "Wikipedia, with a 97% share of the online encyclopedia market, has forced Microsoft to shut down Encarta."

Creating your own VGI system

- idea

- infrastructure

- technology

- basic content
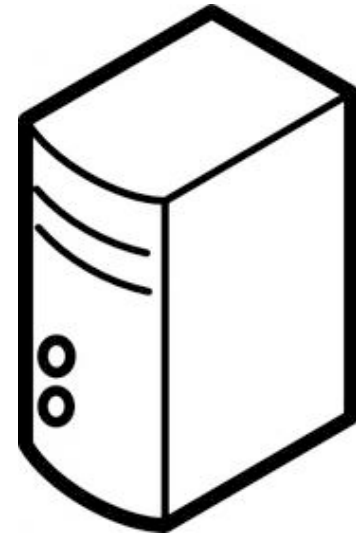
- rules

- community

- users

Idea

- What would we like to achieve?

- Which data are we collecting?

- More focused the idea is, better possibilities for a success.

Infrastructure

- Requirements

  - Data storage (raster and vector data)

  - Application servers

  - Network

- Possibilities

  - Set of own servers

  - Cloud (Google Maps, SimpleGeo,...) + own servers

  - Cloud (Google MyMaps, Amazon, Geopedia.si,..)

Technology

- Open source
  - Geoserver, OpenLayers, Drupal
  - PostGis, MySQL

- Minimum development
  - ESRI
  - Oracle, MS SQL

- Custom software

Basic content

- Orientation

  - Topographic maps, satellite/aerial imagery

    ▸ Google Maps (restrictions!) and similar, open data (NASA), government

- Geo-location

  - Municipalities, street numbers

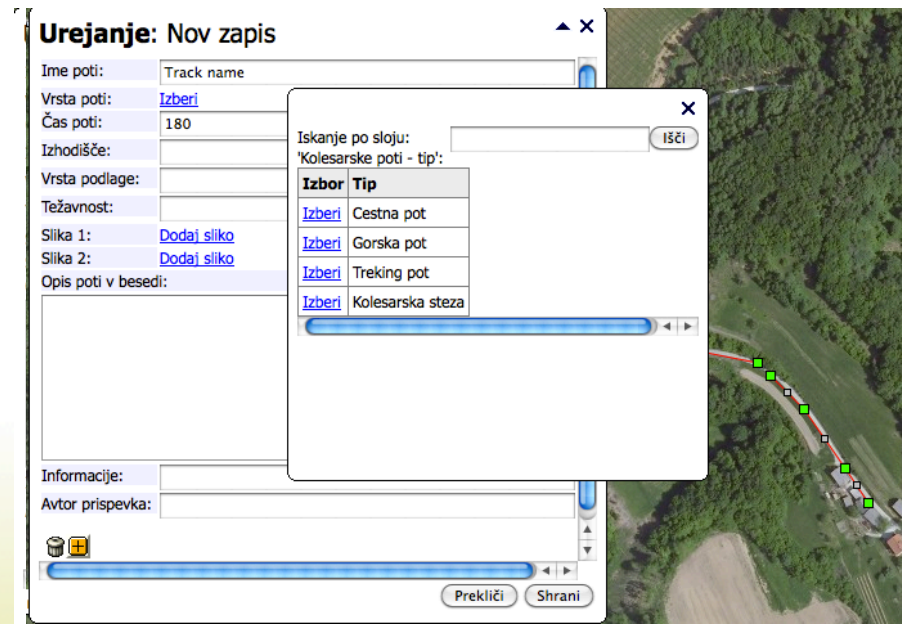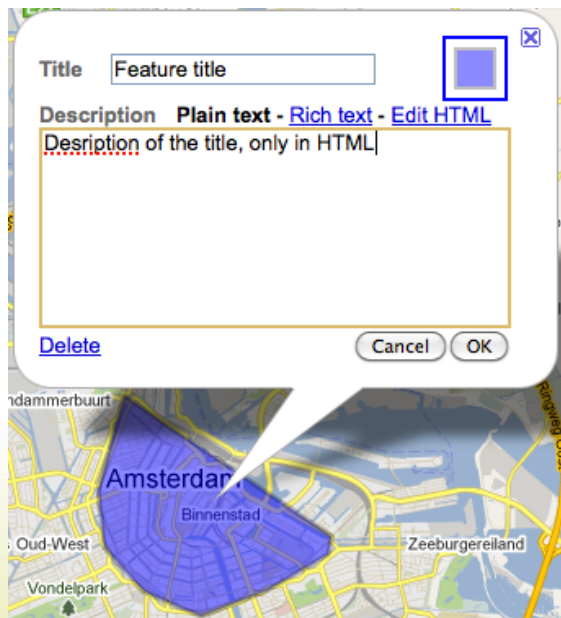   ▸ Google Maps (restrictions!), open data (Natural earth), government

- Topic data

  - Depends on the idea

Rules

- **Lesson from Wikipedia**

  - "Nature said its reviewers found that Wikipedia entries were often poorly structured" (June 2005)

- **Maps vs GIS**

The rise of crowd-sourcing - how valuable data can we get out of VGI, CAPIGI 2011

Community

- Wikipedia

  - 365 MIO visitors (2010)

  - 500.000 contributors

  - 2000 power users (mothly activity)

- Geopedia.si

  - 620.000 visitors (2010)

  - 15.000 contributors

  - 55 power users

Users

- Who will make use of the data collected

- How?

  - on-line viewer

  - web-services

  - data export

- Where the data will be maintained?
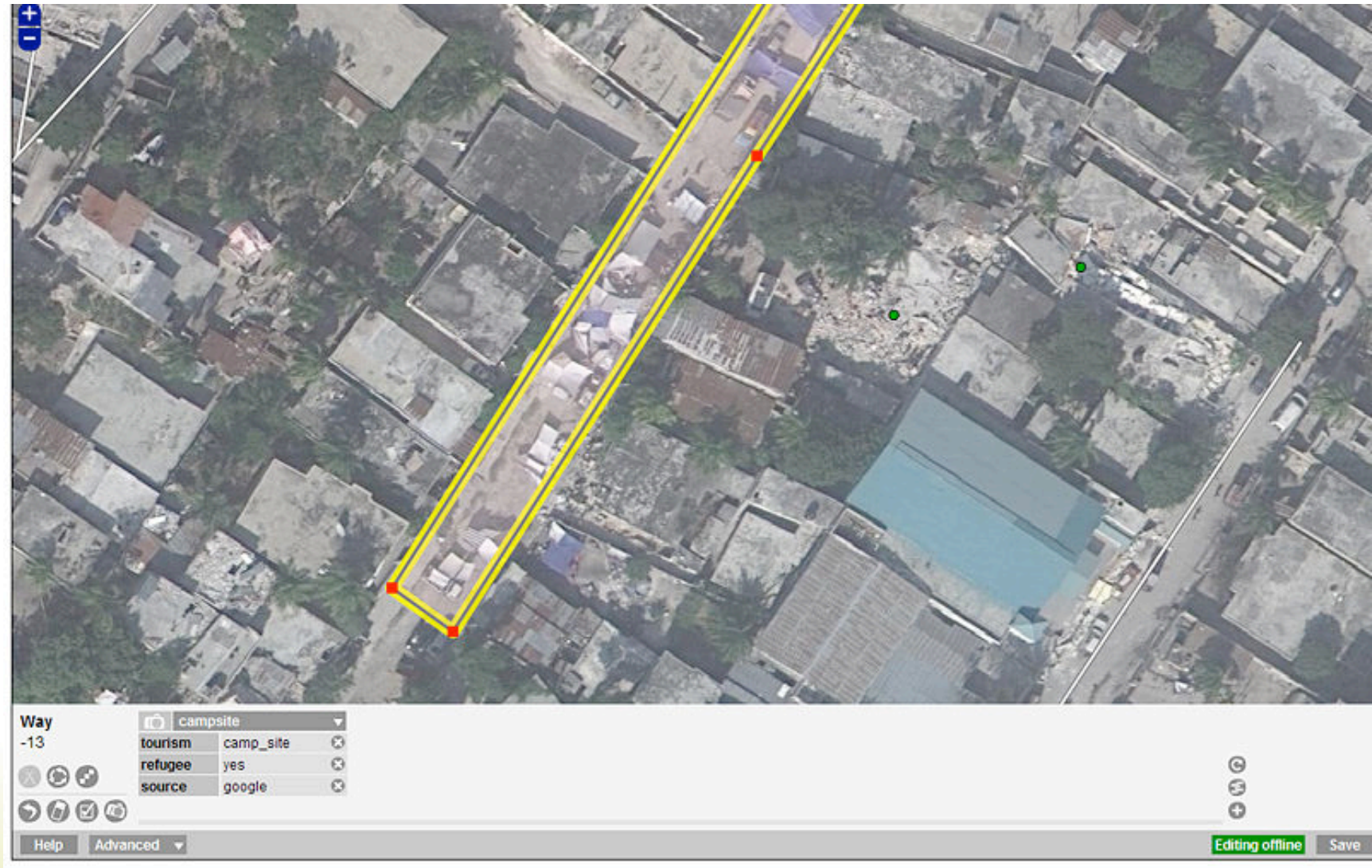
Real-life experiences

- OpenStreetMap

- Geopedia.si

Open-street map

Idea	Creation of a free, editable map of the whole world (roads, buildings, etc.)
Infrastructure	Own servers
Technology	OpenLayers PostgreSQL Other software (Potlatch, Maplink...)
Basic content	Yahoo! Aerial Imagery Some government's data (US Gov, AND Holland, Ordnance Survey, ...)
Rules	Well-defined feature list
Community	>200.000 (end of 2009) 10 % contributing the majority of the data
Users	Haiti Earthquake, large number of web-sites (web-services, tiles)

Open-street map – Haiti earthquake

Open-street map – Kibera slum
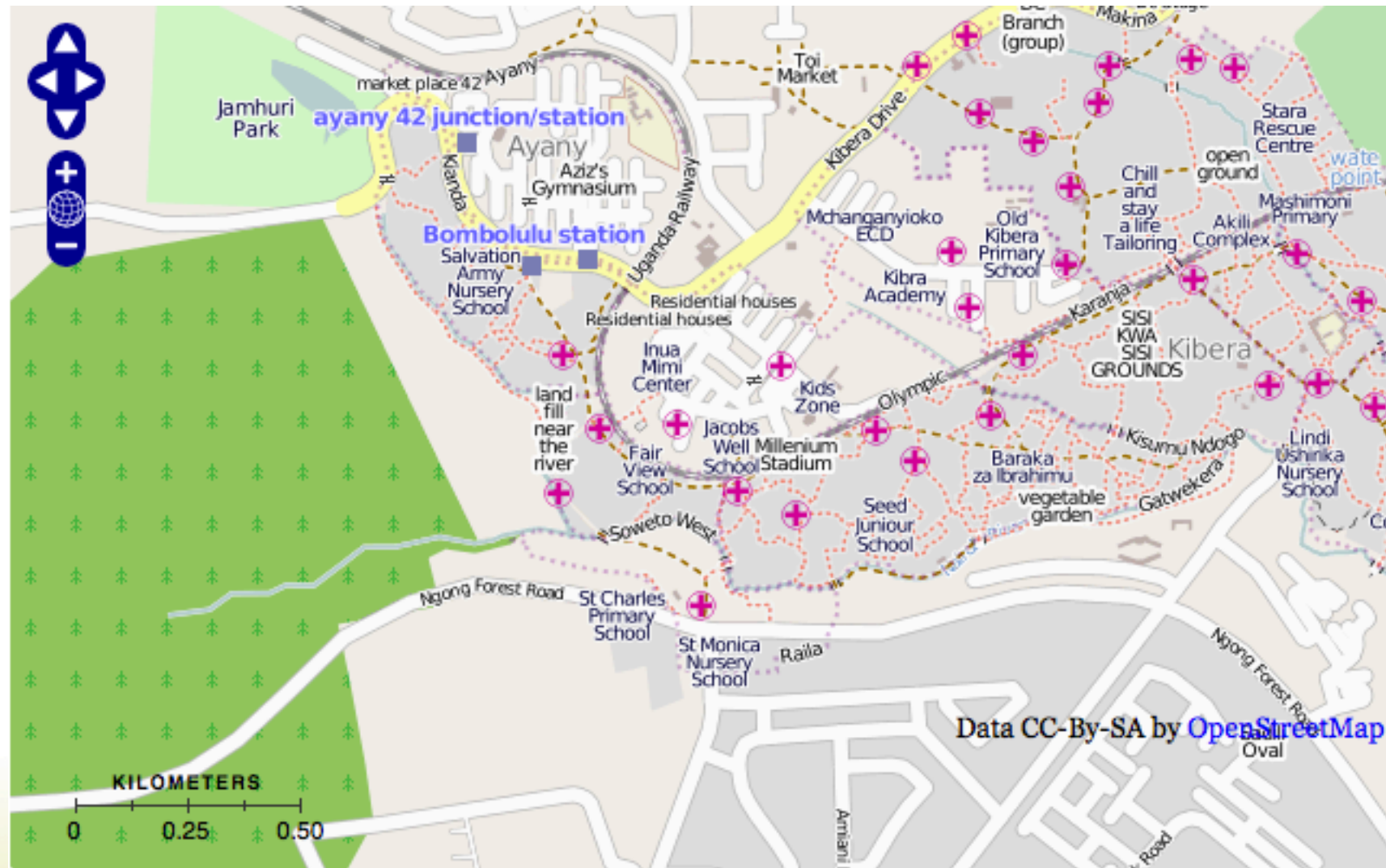
Geopedia.si

Idea	Infrastructure for VGI systems, collection of all spatial data in Slovenia
Infrastructure	Own servers
Technology	Own software (Giselle) MySQL / PostgreSQL
Basic content	Topographic maps, aerial imagery, DMR (Slovenia) Street numbers, other government data
Rules	• single-theme applications with strict business rules • editable spatial layers with pre-defined structure • creation of own layers
Community	20.000 5% contributing the majority of the data
Users	Illegal dump-sites, Cyclists, Mountaineering organization, Encyclopedia of natural and cultural heritage, Energy agency, ...

Geopedia.si - SDI
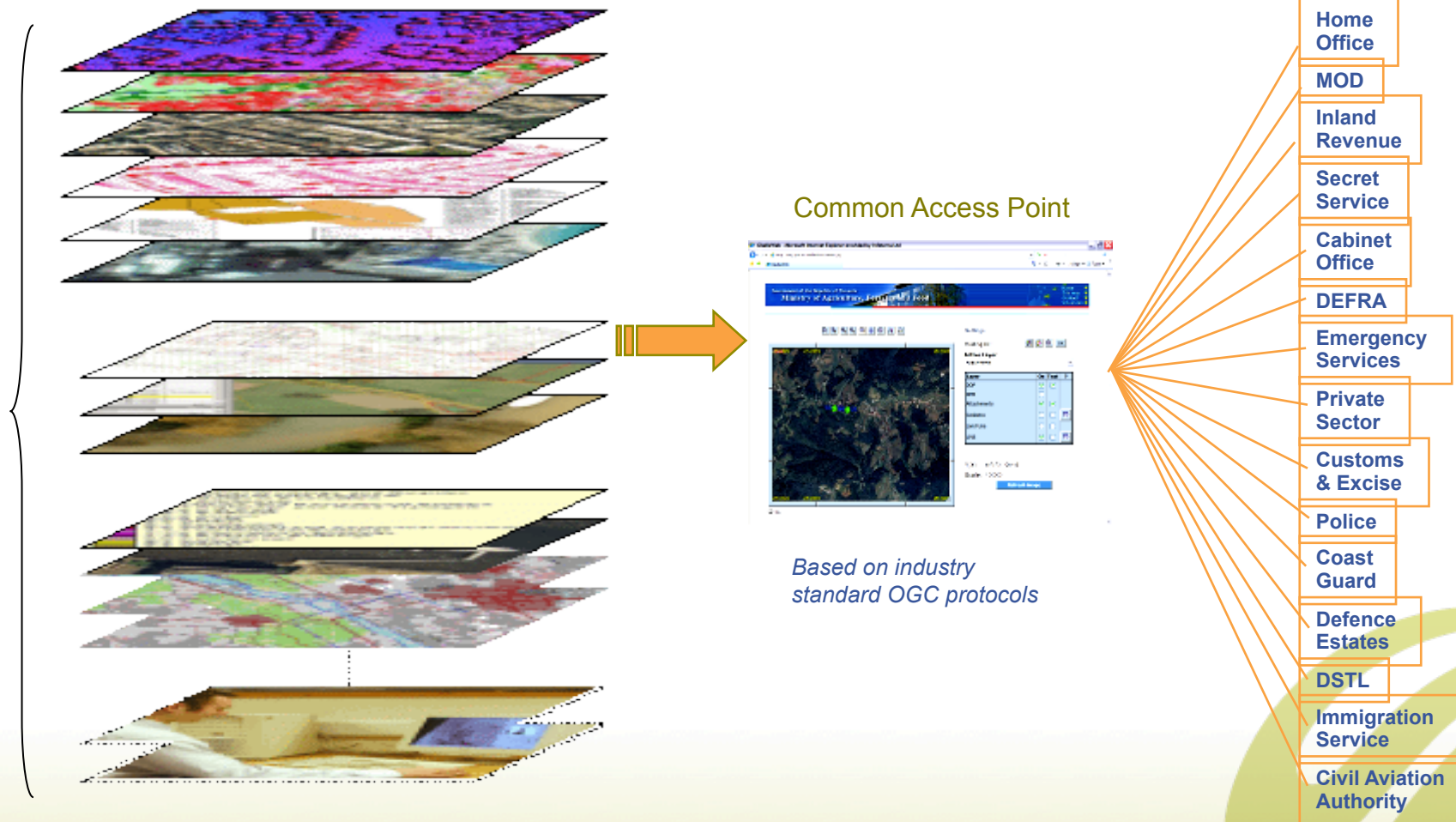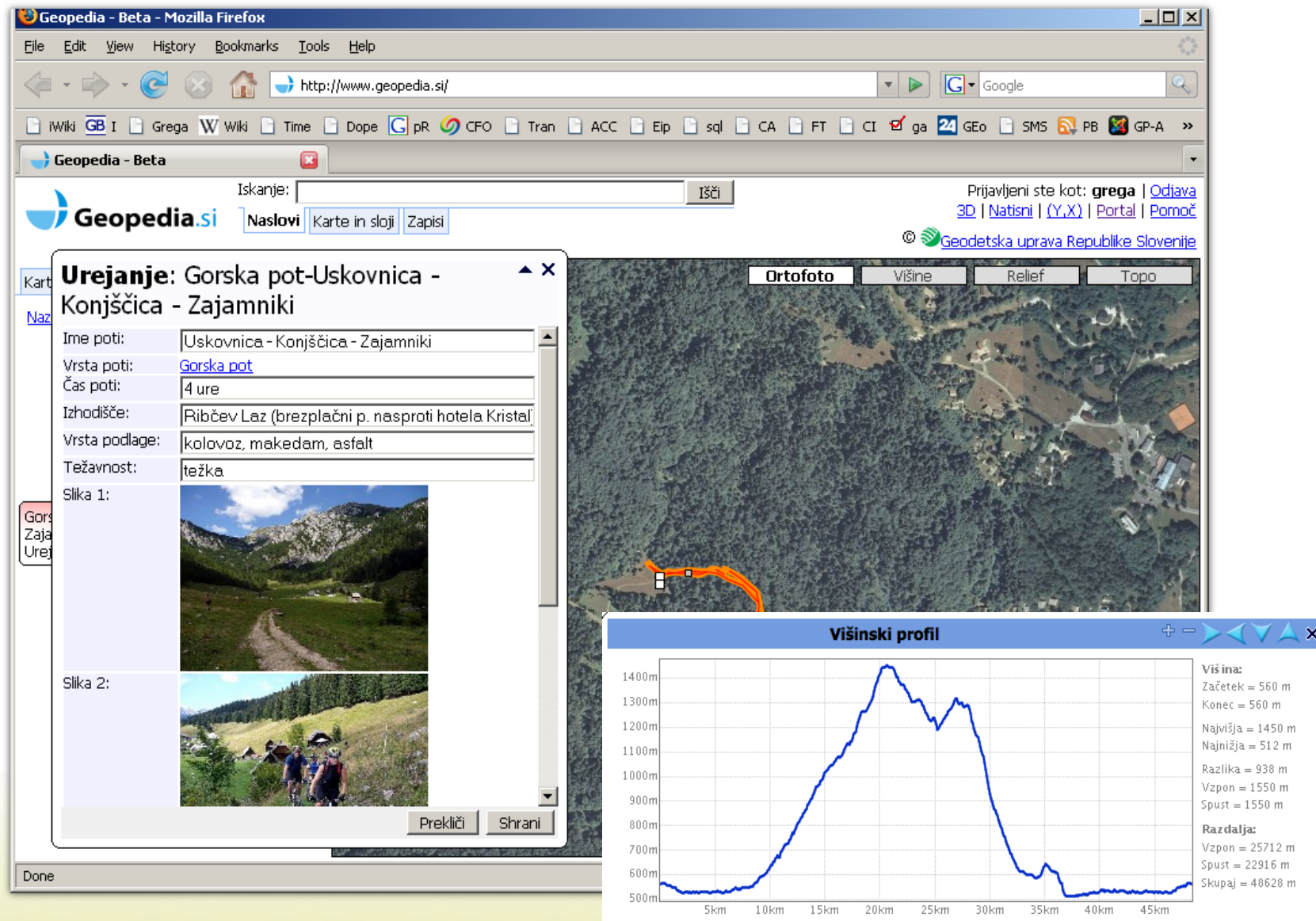
Common Access Point

Based on industry standard OGC protocols

Home Office

MOD

Inland Revenue

Secret Service

Cabinet Office

DEFRA

Emergency Services

Private Sector

Customs & Excise

Police

Coast Guard

Defence Estates

DSTL

Immigration Service

Civil Aviation Authority

Geopedia.si – cycling

Geopedia.si – cycling

- 800 paths

- 25.000 km
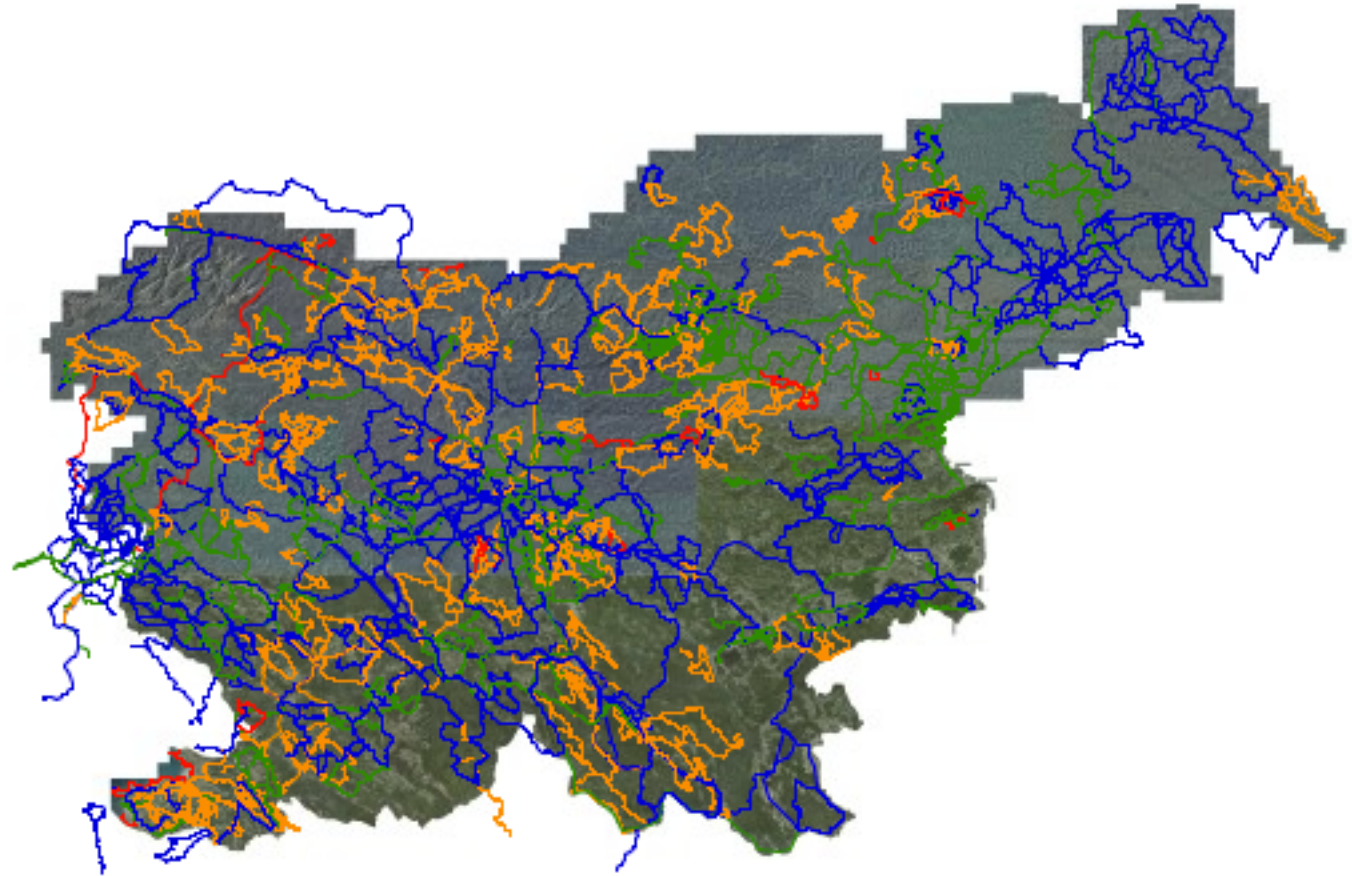
Geopedia.si – illegal dump-site registry

The rise of crowd-sourcing - how valuable data can w

Geopedia.si - floods
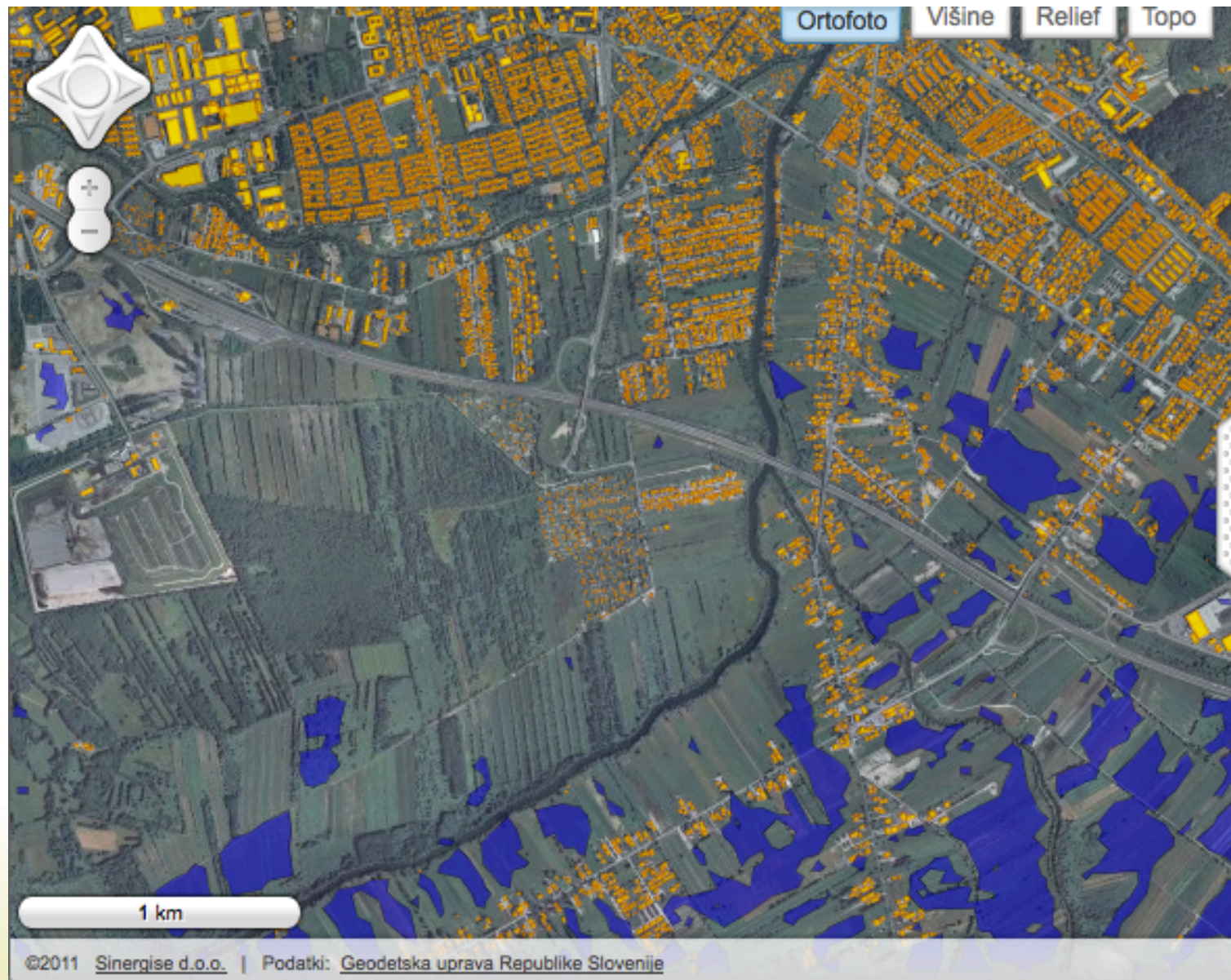
Geopedia.si – stats

- 150.000 monthly visits (3 MIO page-views)

- average time on site – 8 minutes

- 6.000 spatial layers (1000 marked as "good quality")

- 10 MIO of spatial entries

- (only in Slovenia)

Comparison to non-VGI

Michael Jones, Google:

The chief Internet evangelist at Google Inc., and one of the founding fathers of the Internet, says he'd like to see a geographic equivalent of Wikipedia — "Geopedia," he dubs it — where anyone could add to the world's geographic know-how.

Jack Dangerman, ESRI:

He worries that even the best-intentioned amateur could provide inaccurate data that could lead to a disaster. "Who wants to dig a hole and run into a pipe?" Dangerman asks.
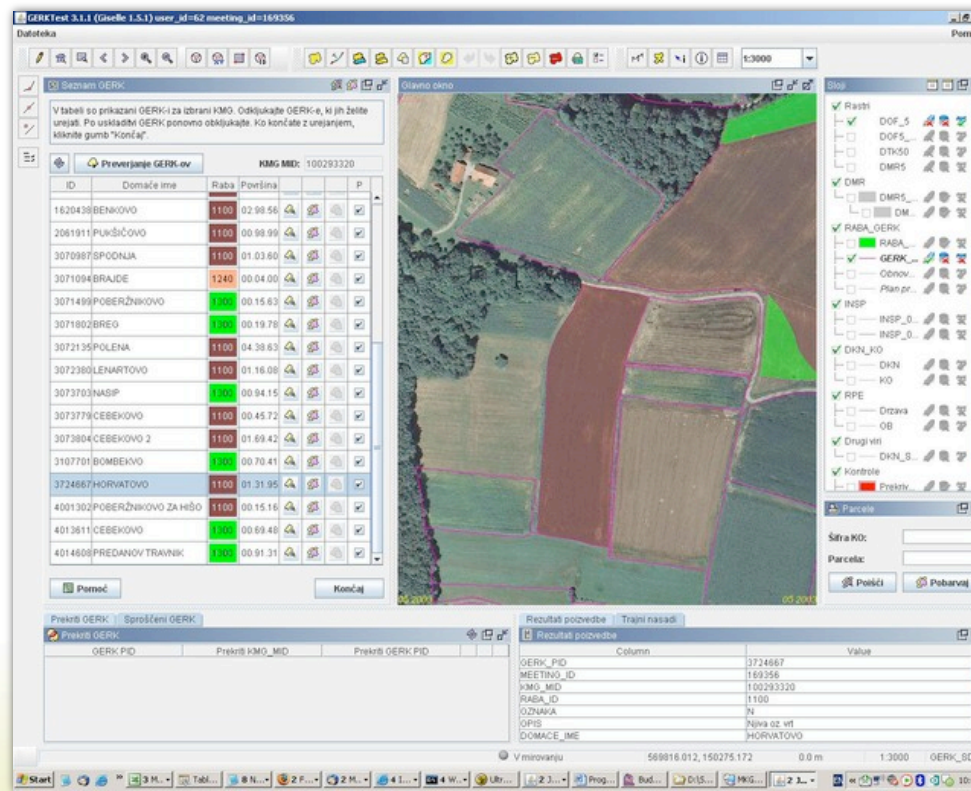
(GeoWeb, summer 2007)

Comparison to non-VGI

- LPIS

- Land Cadastre

LPIS

- Government operated, based on farmer's input

- Used for distribution of agriculture subsidies

LPIS – strict control
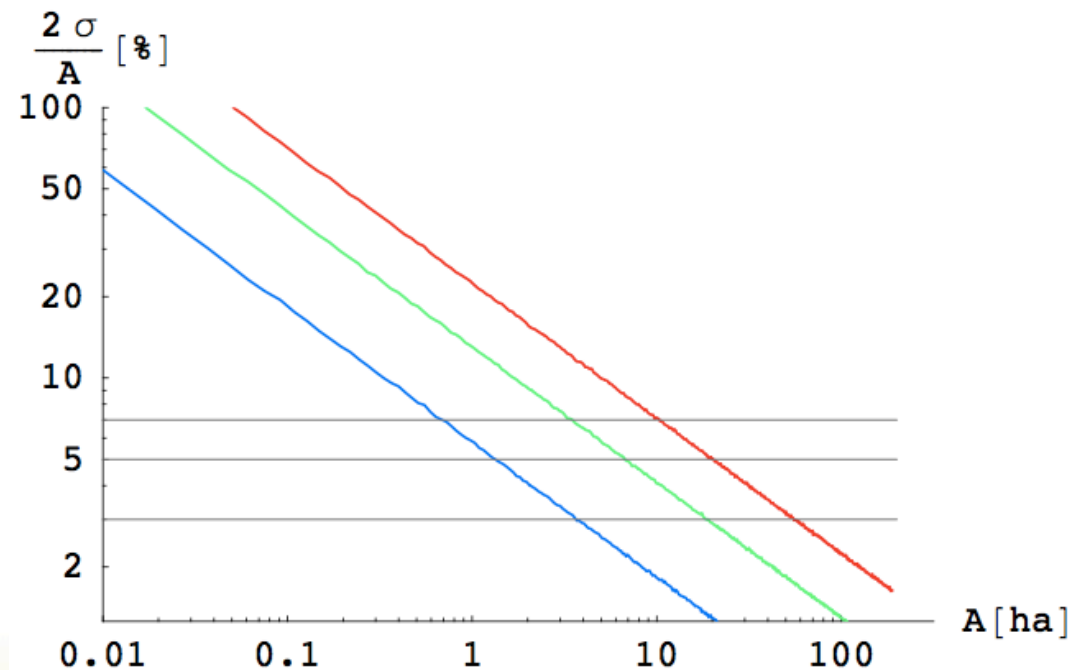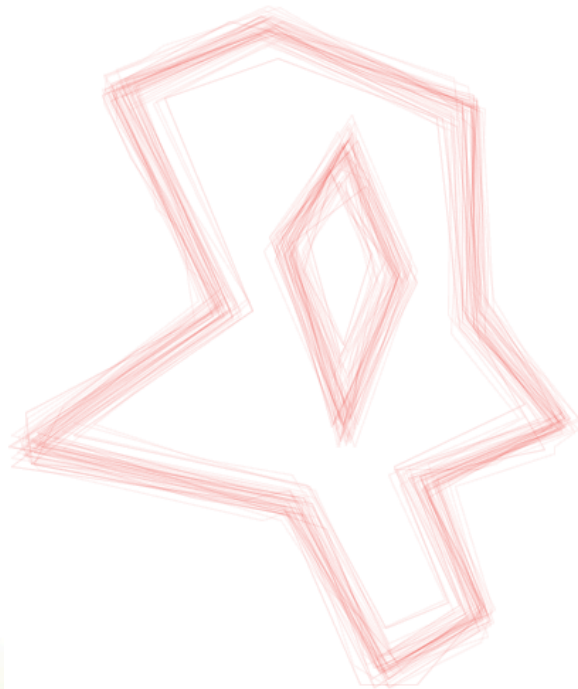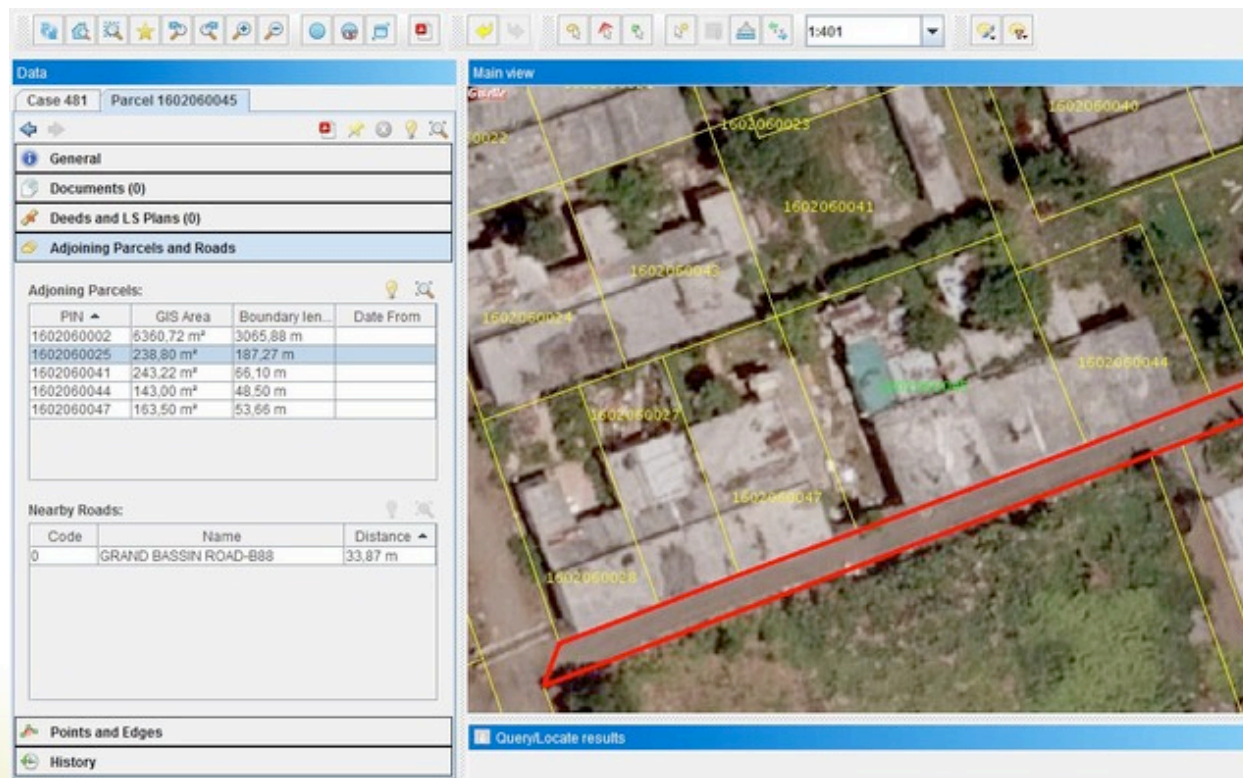
LPIS - Quality

- Theoretically not possible to meet EU requirements in some cases (max 3% of area difference).

Land Cadastre

- Government operated, based on field survey

- Used for land administration, taxation

Land Cadastre – strict control

Land Cadastre – quality

- Not good enough as a base for LPIS (in Slovenia and some other cases)

Conclusion

- VGI does not produce perfect results

  Neither do "professional" systems

- Good enough for practical purpose

- Improving in the future (mobile GIS, etc.)

Thank you

Questions?