Data Quality Control:
Crowd-Sourcing Geospatial Information

Xizi Hotel, Hangzhou
24-26 May, 2012

Presented : Prof. Dr. Woosug Cho, Senior Advisor, NGII, Republic of Korea
Prepared   : Prof. Dr. Ki-Joune Lee, Senior Advisor, NGII, Republic of Korea
             Mr. Jun-Sung Kim, Deputy Director, NGII, Republic of Korea

National Geographic
Information Institute
Republic of Korea

Definition of Crowd-Sourcing

Crowd-sourcing is a process that involves outsourcing tasks to a distributed group of people. This process can occur both online and offline, and the difference between crowdsourcing and ordinary outsourcing is that a task or problem is outsourced to an undefined public rather than a specific body, such as paid employees.

Shortly, a distributed problem-solving and production model.

The term "crowdsourcing" is a portmanteau of "crowd" and "outsourcing," coined by Jeff Howe in 2006.
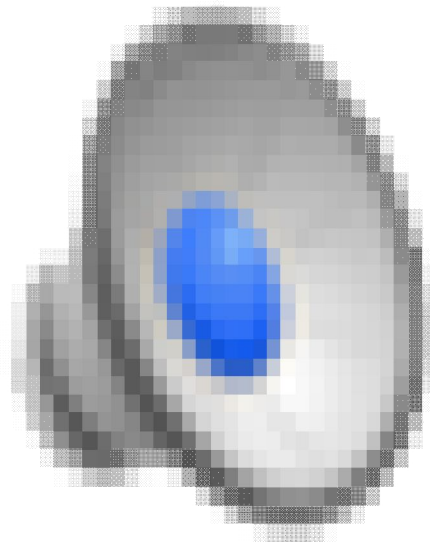
Crowd-Sourcing Geospatial Information

VGI (Volunteered Geographic Information) :

the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals (Goodchild, 2007)
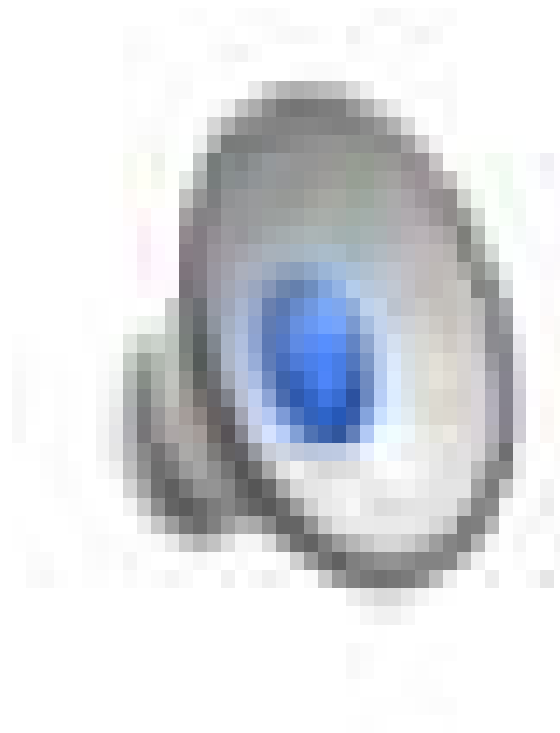
Examples on Crowd-Sourcing Geospatial Information:

1. OpenStreetMap (http://www.openstreetmap.org/)

http://www.youtube.com/watch?v=IpXk5W6b298&feature=channel&list=UL

2. Project Haiti: http://vimeo.com/9182869

3. Google Map Maker (http://www.google.com/mapmaker)
http://www.ted.com/talks/lalitesh_katragadda_making_maps_to_fight_disaster_build_economies.html

Project Haiti

Google Map Maker: Myanmar Case

Example in UK –Google Map & OSM

http://tools.geofabrik.de/mc

Example in Korea – Google Map & OSM

http://tools.geofabrik.de/mc

PotLatch
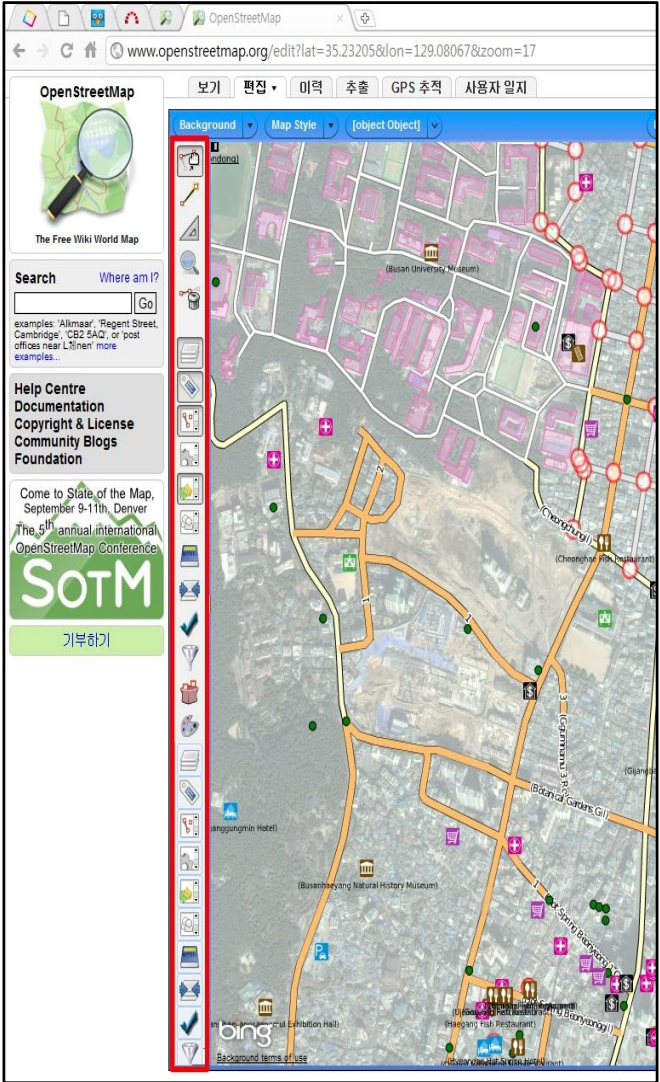
JOSM(Java OSM editor)
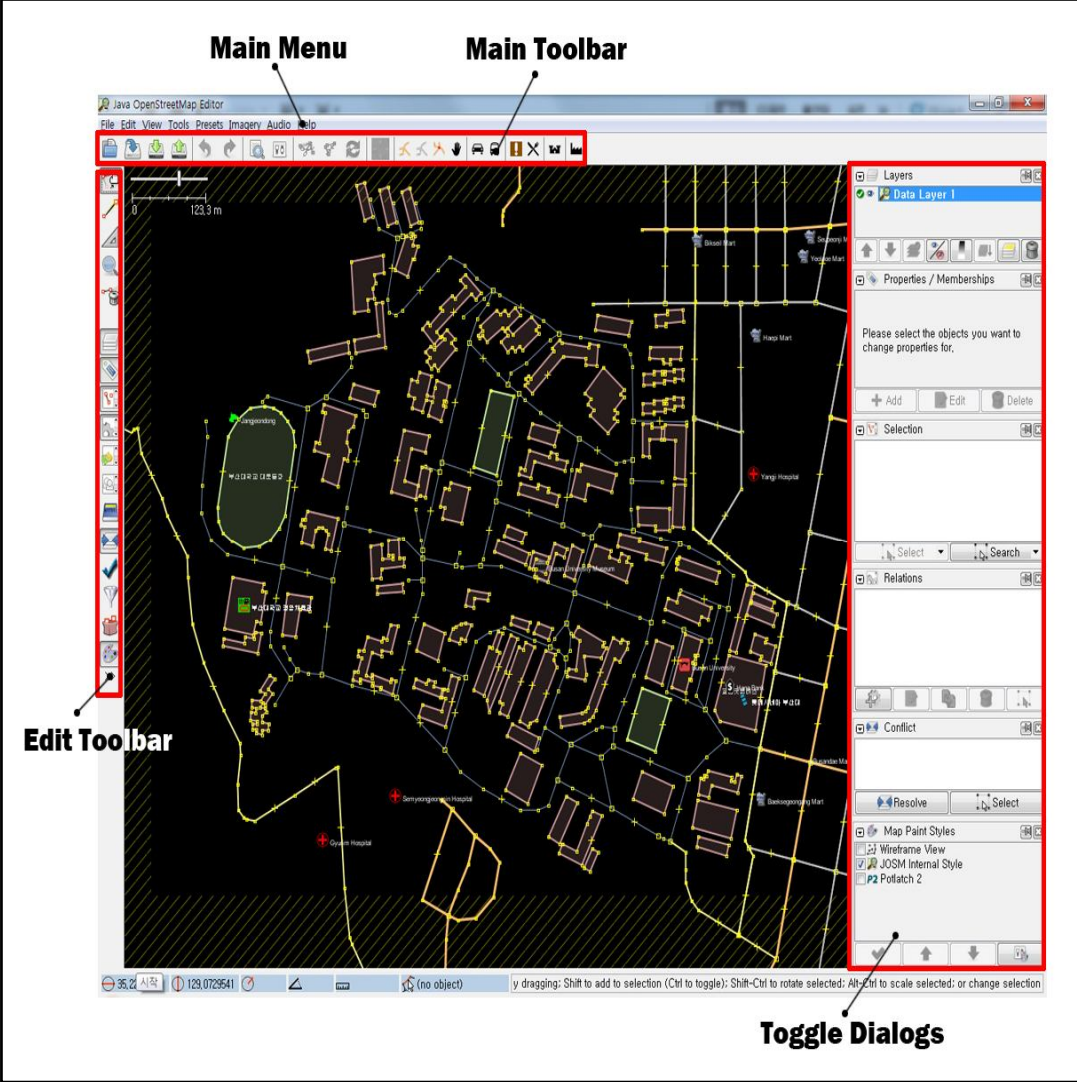
Merkaartor
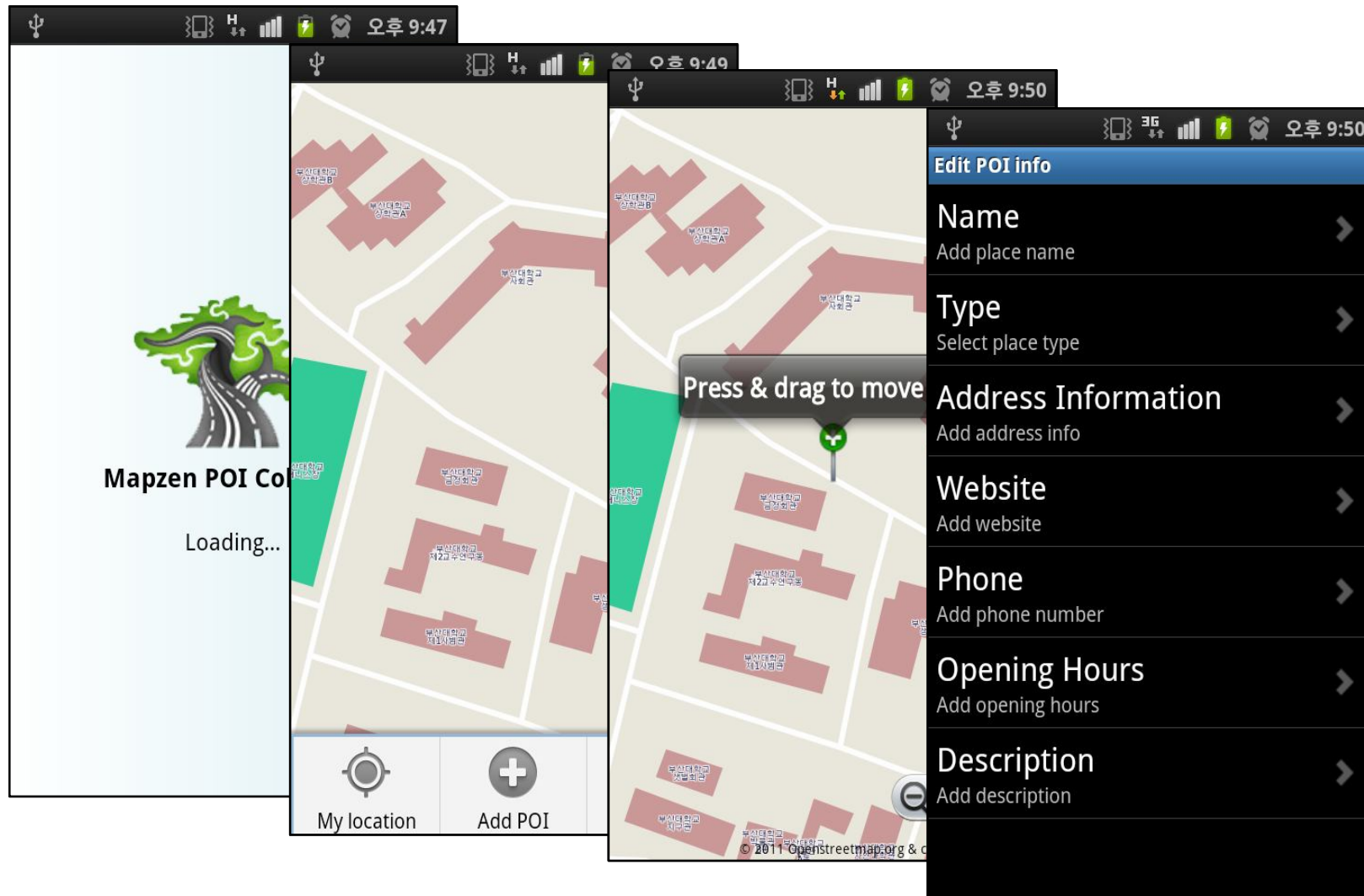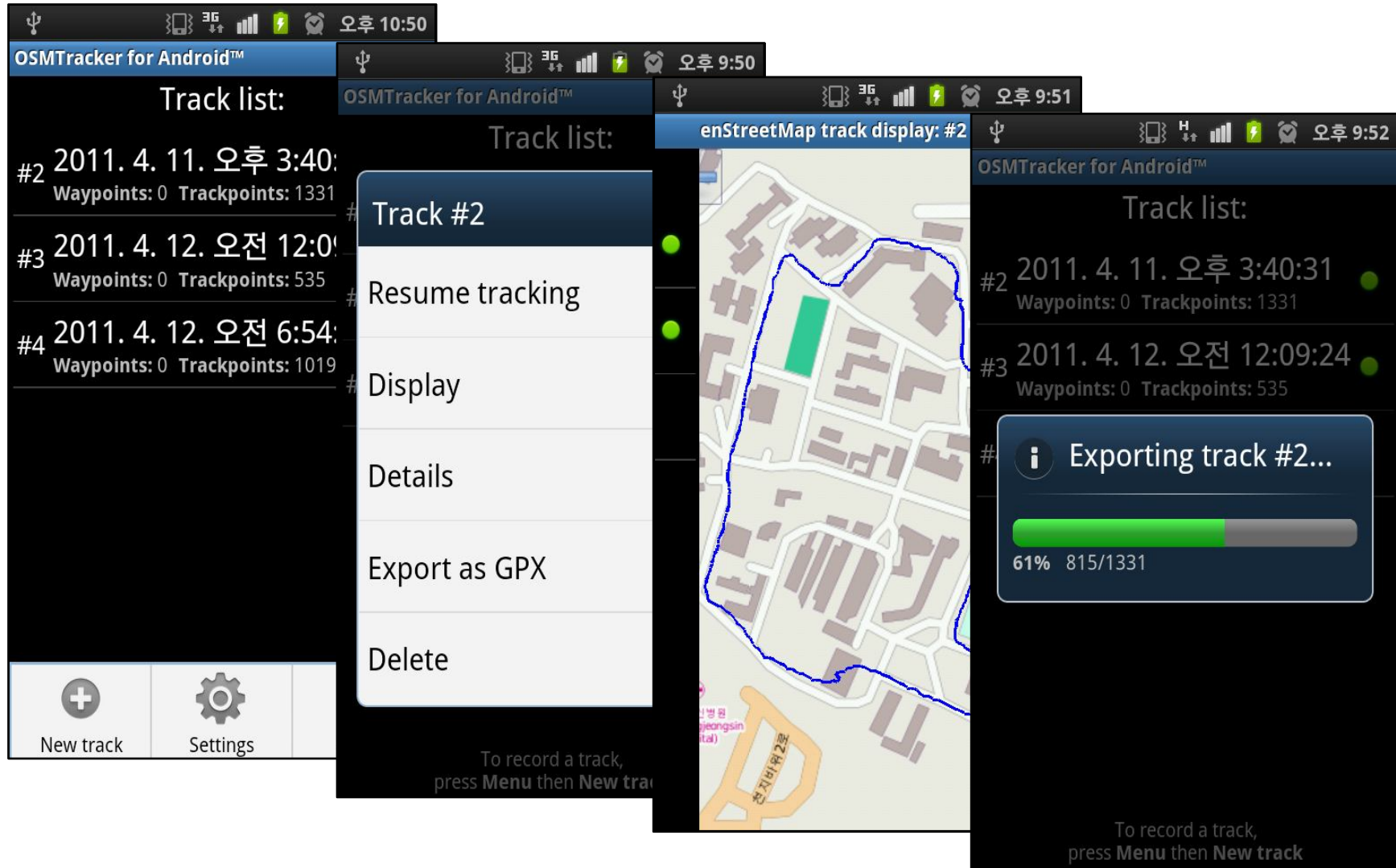
Vespucci

Mapen POI COllector

OSMTracker

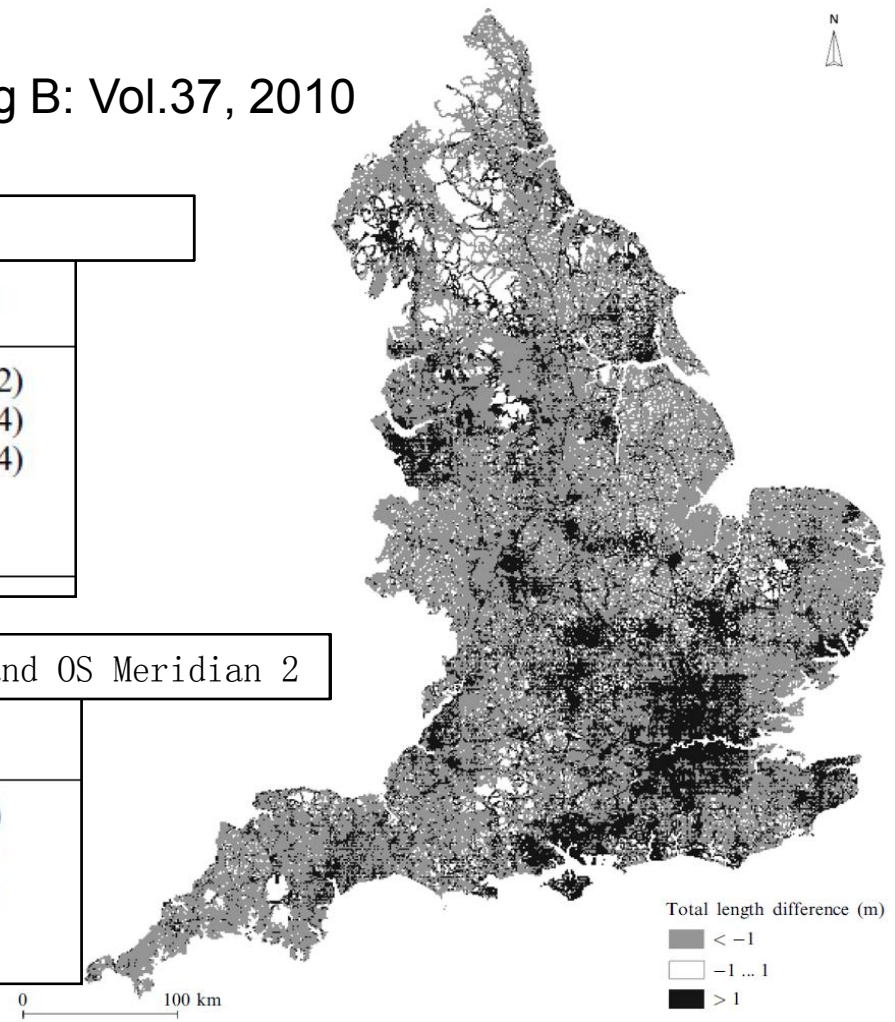# A Comparative Study of Crowd-Sourcing GI

OSM vs. OS datasets
- M. Haklay, Environment and Planning B: Vol.37, 2010

Length comparison between OSM and OS Meridian 2

Cells	Area (km²)
Empty cells	16 300 (13.2)
Meridian 2 more detailed than OSM	75 977 (61.4)
OSM more detailed than Meridian 2	31 437 (25.4)
Total	123 714

Length comparison with attributes between OSM and OS Meridian 2

Cells	Area (km²)
Empty cells	17 632[a] (14.3)
Meridian 2 more detailed than OSM	80 041 (64.7)
OSM more detailed than Meridian 2	26 041 (21.0)
Total	123 714

Total length difference (m)
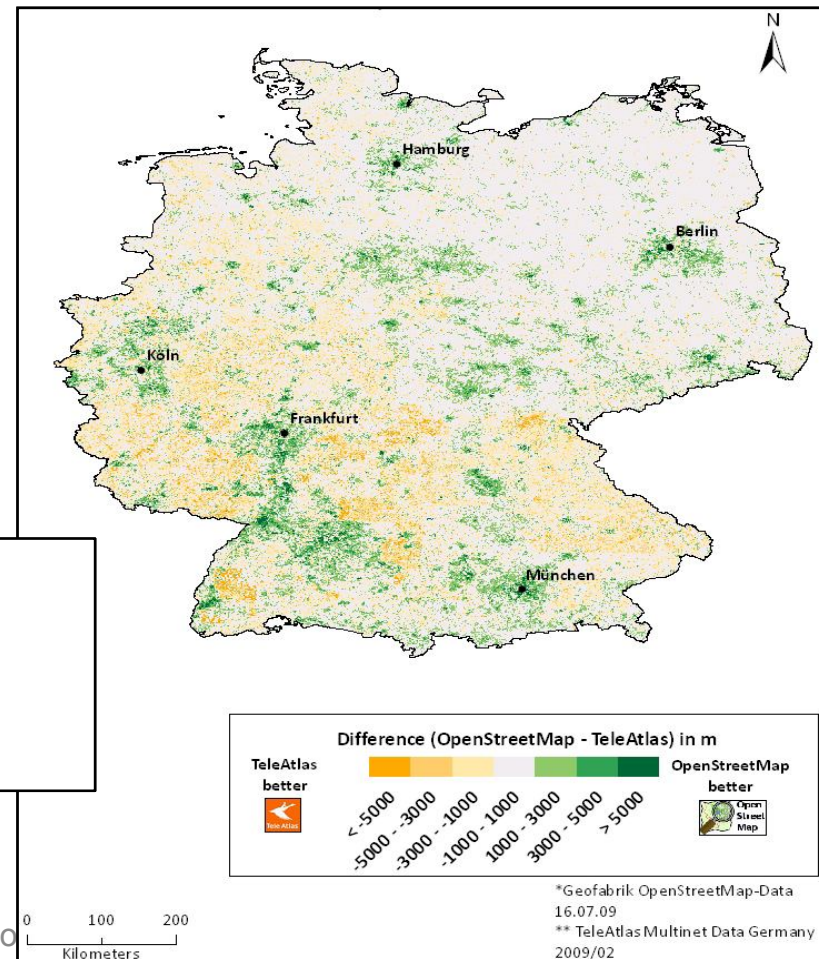- < −1
- −1 ... 1
- > 1

A Comparative Study of Crowd-Sourcing GI

OSM vs. Tele Atlas in Germany
  - D. Zielstra and A. Zipf, Agile Conf., 2010

Total Length Differences of
all Street Network Data between
OpenStreetMap and Tele Atlas per
Km^2

Weakness of Crowd-Sourcing GI

NO Quality Control on GI and NO Metadata

– Data quality depends on each individual volunteer

– No metadata
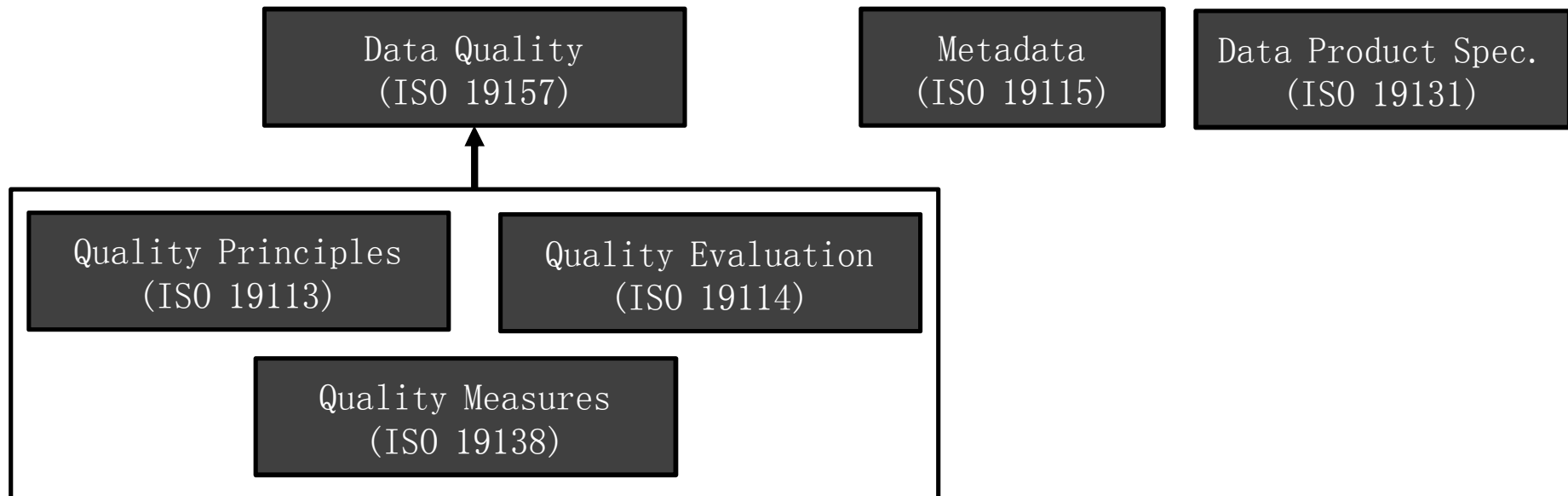
– Only some lineage data (change-sets) is provided

Data Quality Control in ISO 191xx

Data Quality Evaluation – ISO

Quality Evaluation Method – ISO 19114
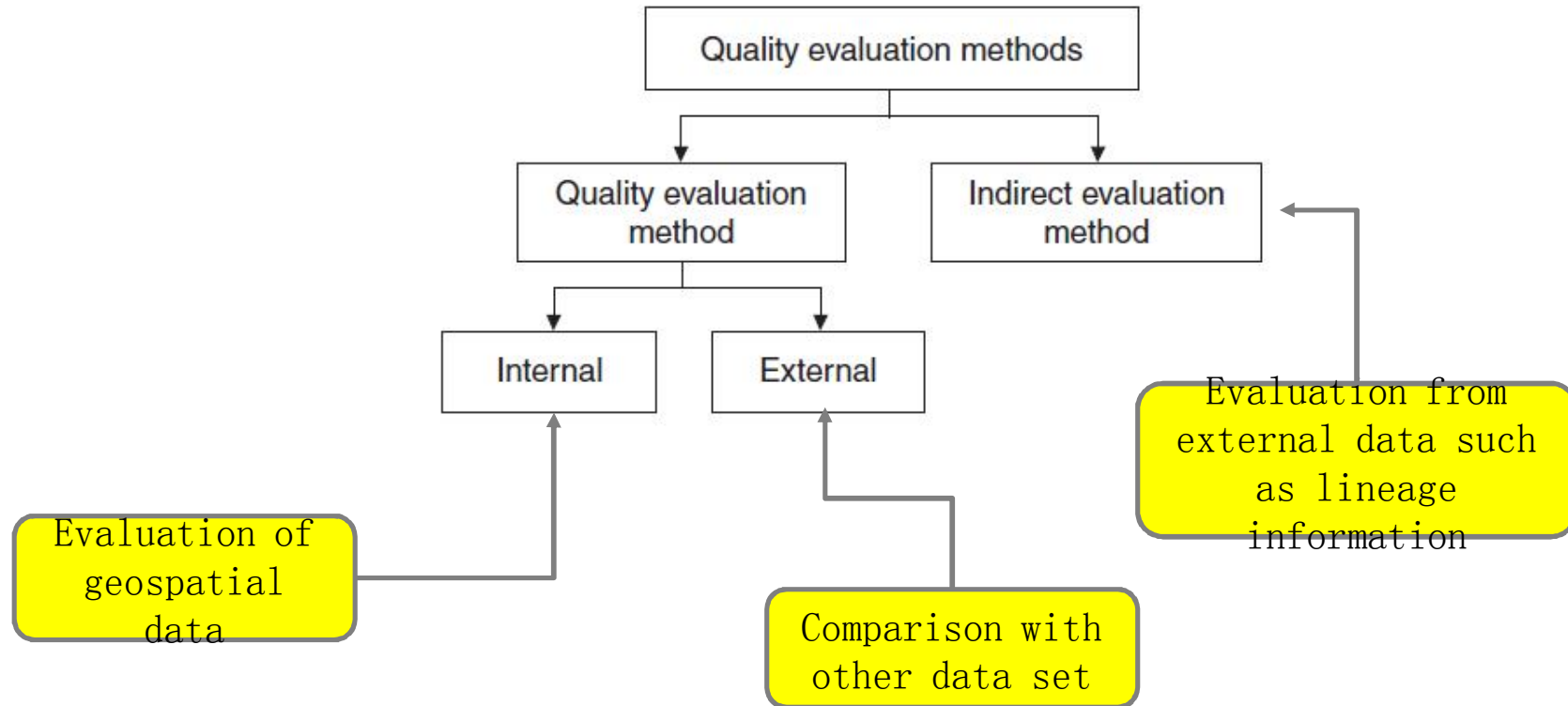
Data Quality Control for Crowd-Sourcing GI

Impossible to enforce data quality control process
to volunteer map makers

- Possibly by governmental or semi-governmental institutes

ISO Quality Control

- As defined by ISO 19113, 19114, and 19138
- Direct Evaluation
  - Comparison with other map data such as national maps
- Indirect Evaluation, such as
  - Number of participants at given area,
  - Profile of each participant, and
  - Frequency of updates and last update date

Data Quality Control for Crowd-Sourcing GI

(-) May discourage volunteer map production
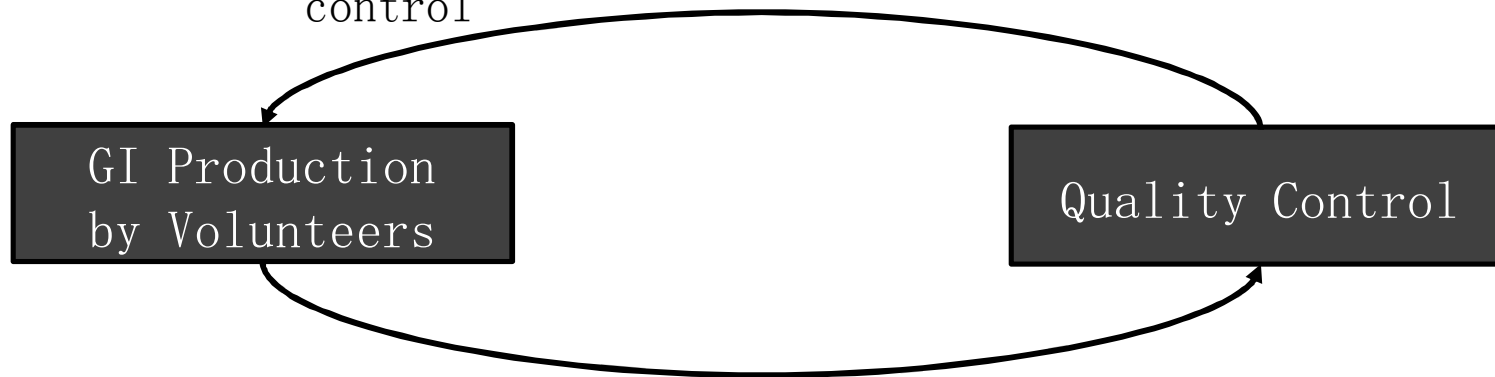
(-) Poor environment to implement quality control

GI Production by Volunteers

Quality Control

(-) Limited use of data due to the lack of quality control and metadata

Data Quality Control for Crowd-Sourcing GI : How to break the vicious circle

(−) May discourage volunteer map production

(−) Poor environment to implement quality control

Improving VGI environment of quality control

(+) Quality control by means of map production environments

GI Production by Volunteers

Quality Control

(+) Providing data quality information (metadata)

Monitoring quality

(−) Limited use of data due to the lack of quality control and metadata
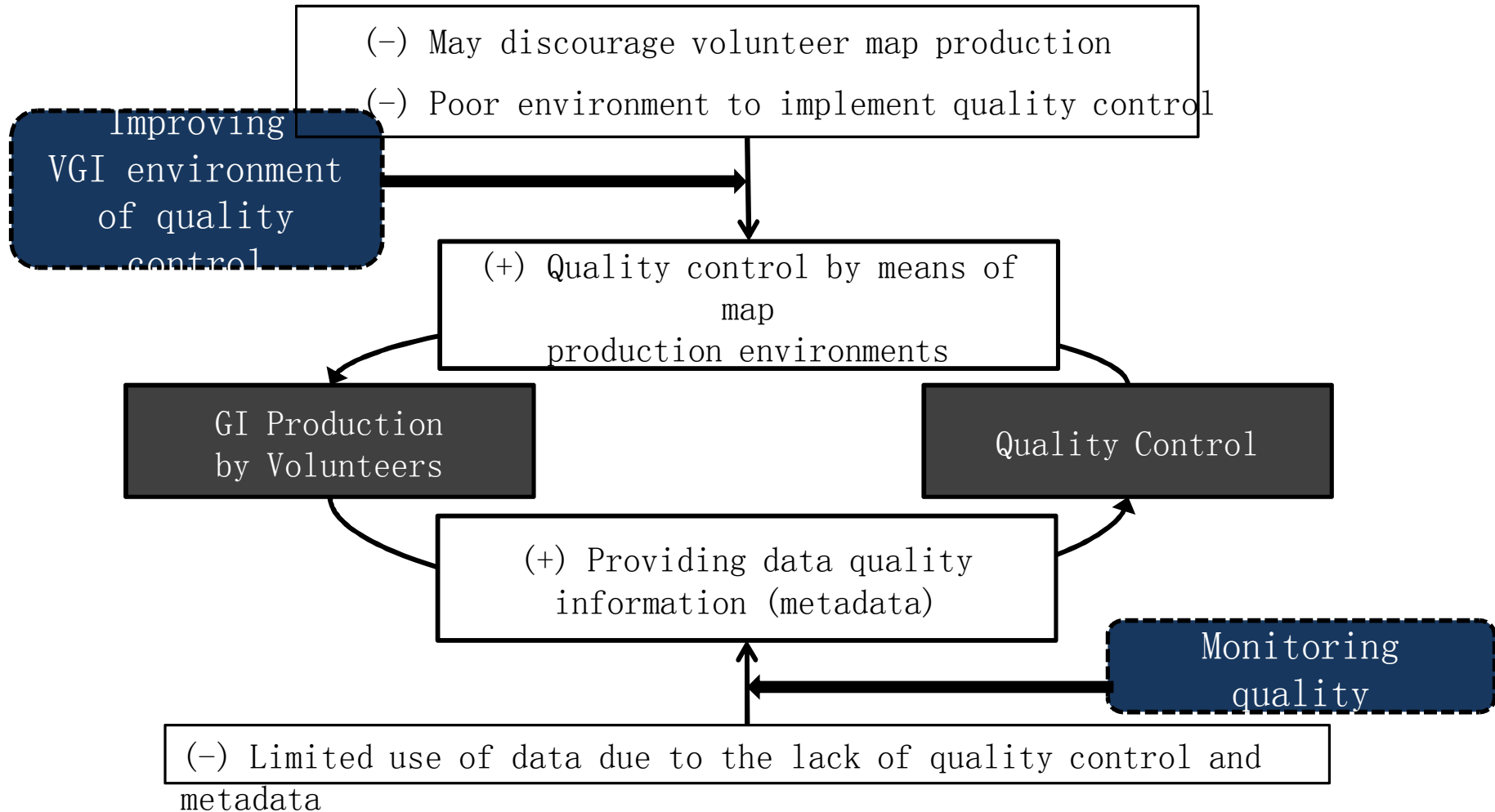
Data Quality Control for Crowd-Sourcing GI : Improving Map Production Environment

```
┌─────────────────────────────────────────┐
│  ┌─────────────────────────────┐         │
│  │   Map Production by         │         │
│  │   Volunteers                │         │
│  └─────────────────────────────┘         │
│                                           │
│  Map Production Tools for VGI             │
└─────────────────────────────────────────┘
                    │
                    ▼
┌──────────────────────┐   ┌──────────────────────┐   ┌──────────────────┐
│   Raw Test Data      │──▶│   Report on Data     │──▶│    Metadata      │
│   About Internal Test│   │   Quality by Internal│   │                  │
│                      │   │   Test               │   │                  │
└──────────────────────┘   └──────────────────────┘   └──────────────────┘
```
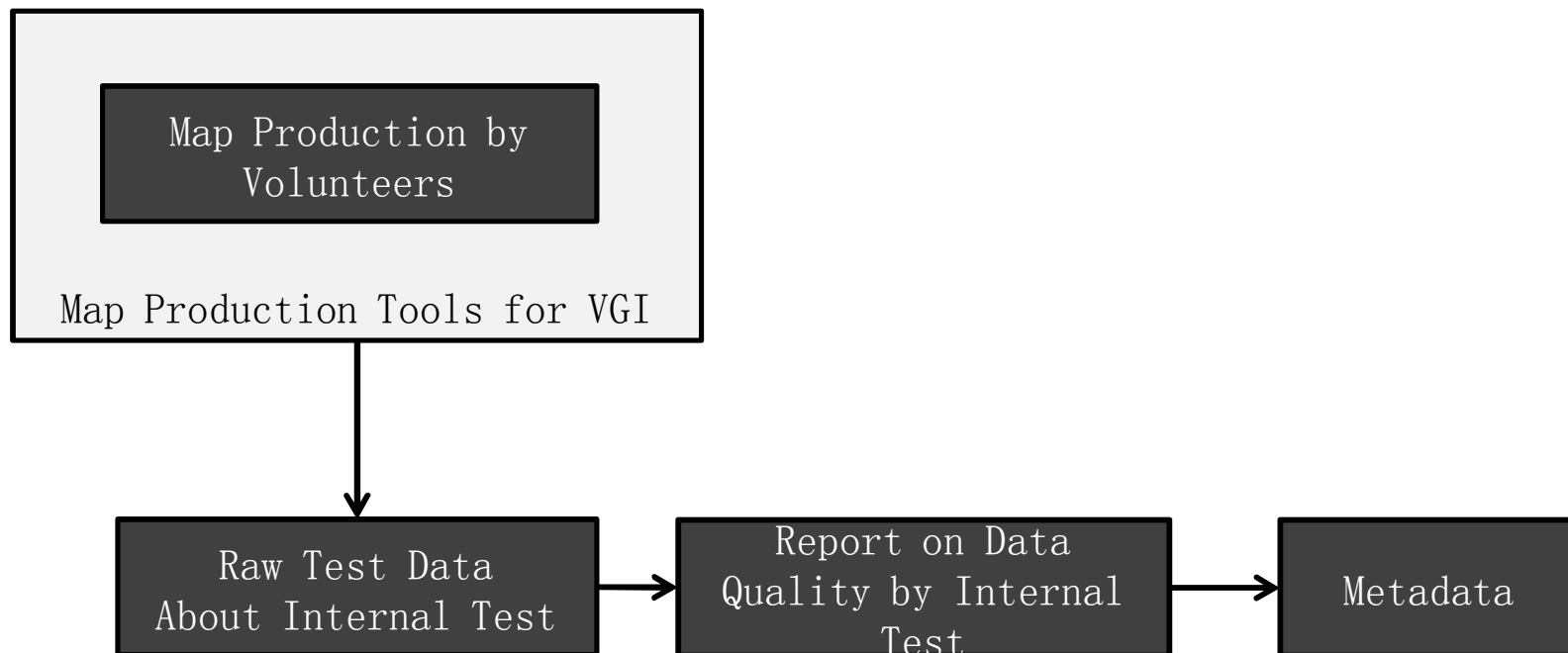
Data Quality Control for Crowd-Sourcing GI : Monitoring Data Quality
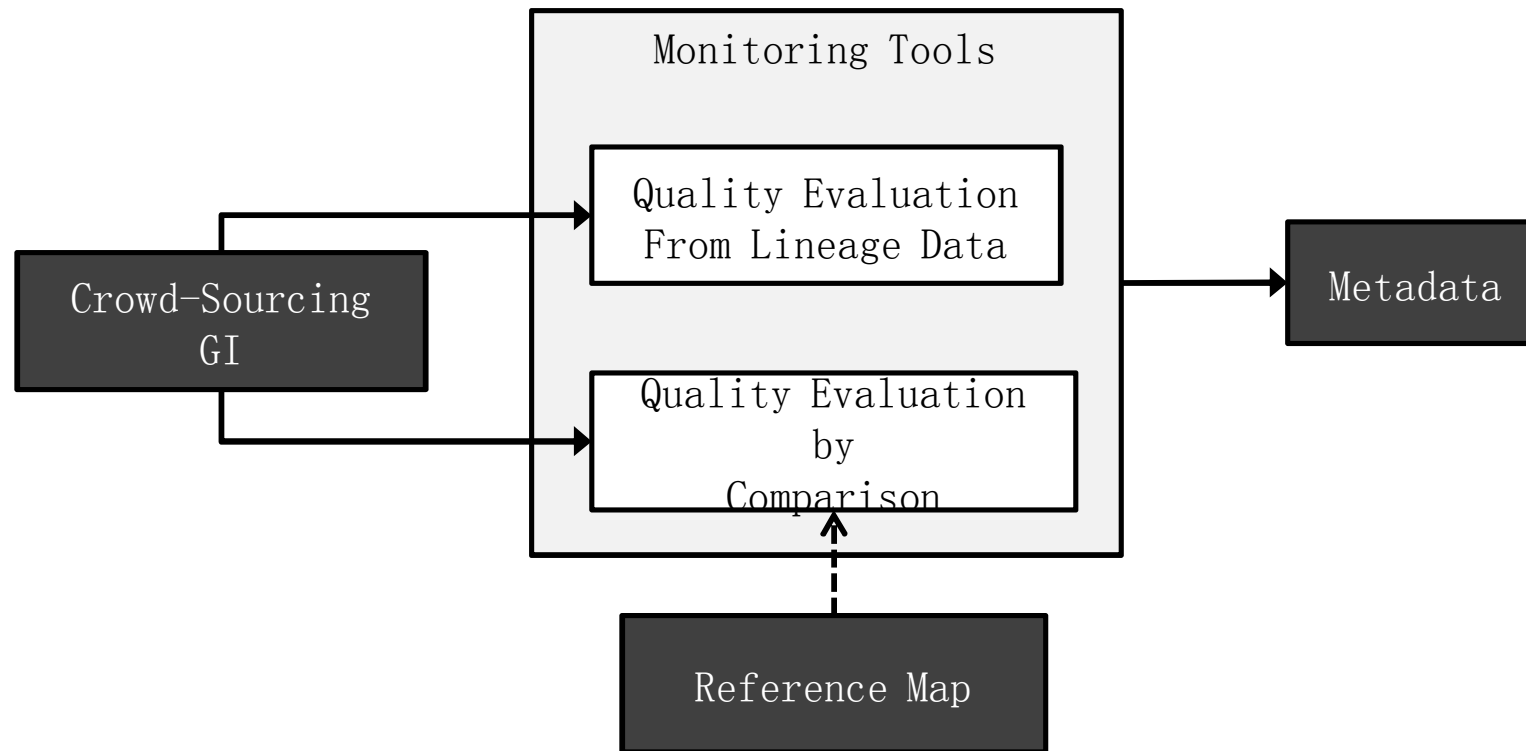
Data Quality Control for Crowd-Sourcing GI : Proposed Flow
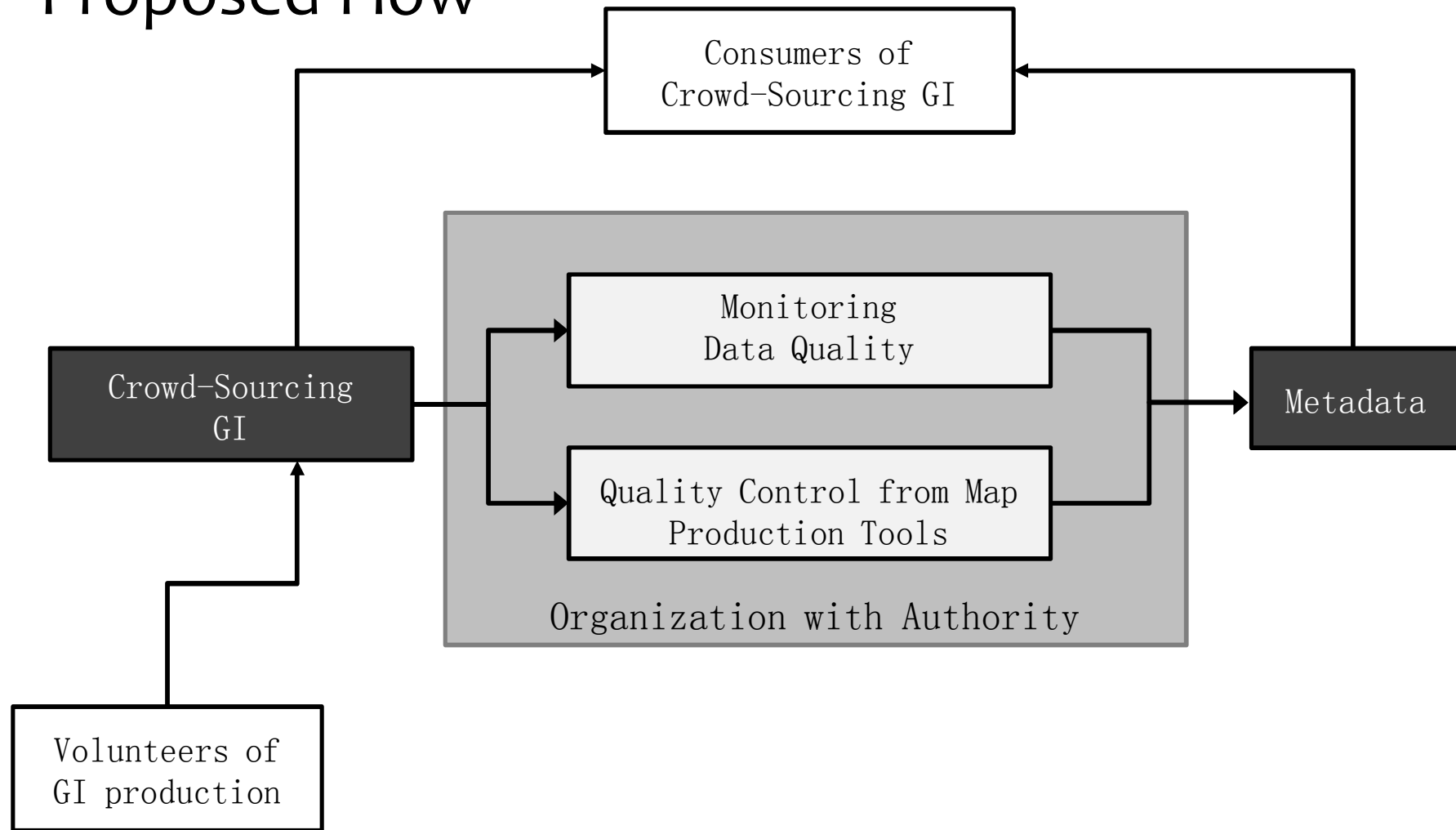
Summary

Crowd-Sourcing GI

- – An Important Source of GI
- – Very helpful for developing countries

Difficult to enforce rigorous quality control

Quality Control of Crowd-Sourcing GI

- – Map Production Tool with QC functions
- – Monitoring QC of Crowd-Sourcing GI