UNIVERSITY OF TWENTE.

ITC

Mining user-generated geographic content: An interactive, crowdsourced approach to validation and supervision

Frank O. Ostermann, Gustavo A. Garcia-Chapeton,
Raul Zurita-Milla, Menno-Jan Kraak

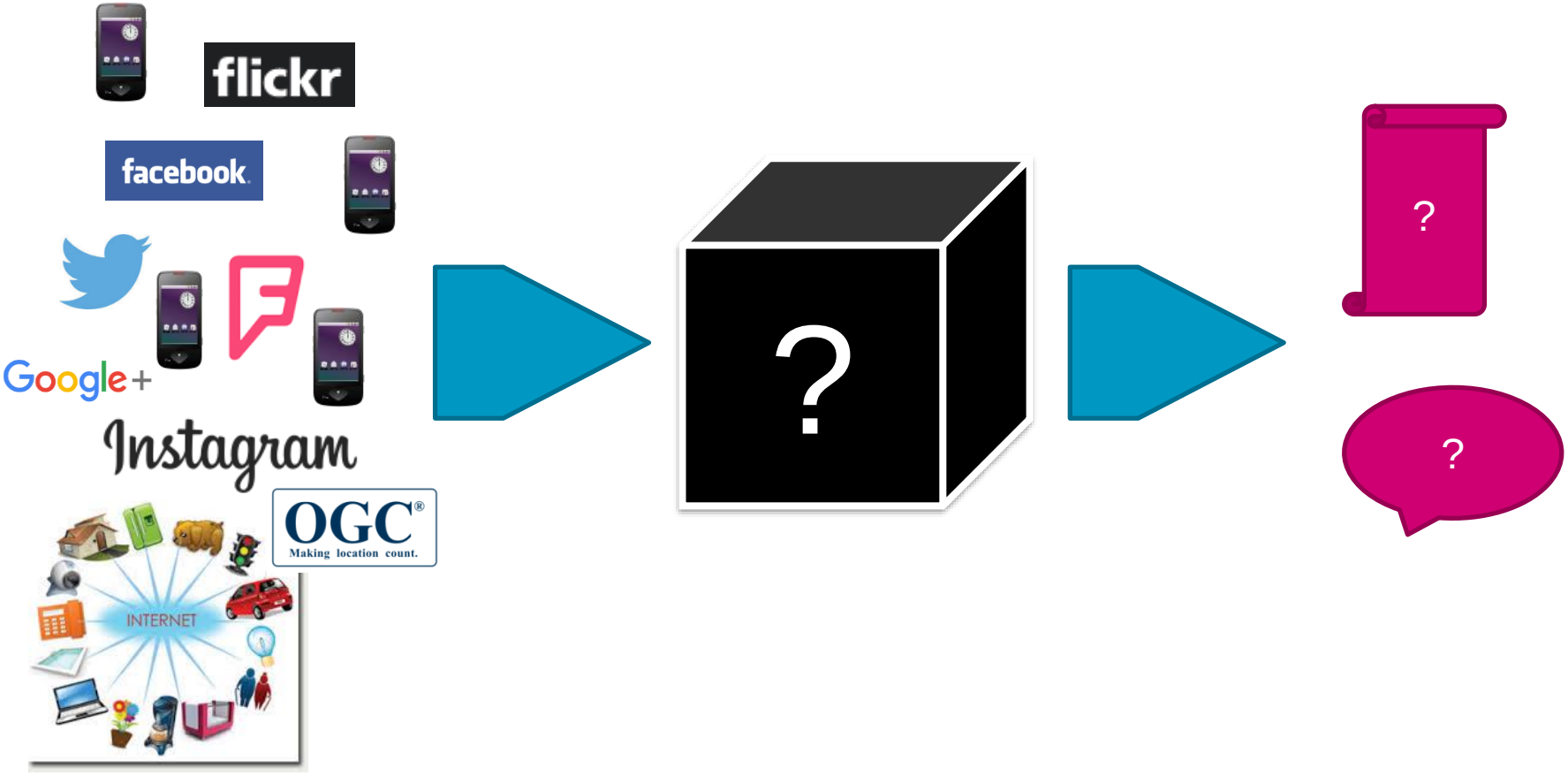
Presentation at the
20th AGILE conference on Geographic Information Science
11.05.2017, Wageningen UR

WHAT THIS TALK IS ABOUT…

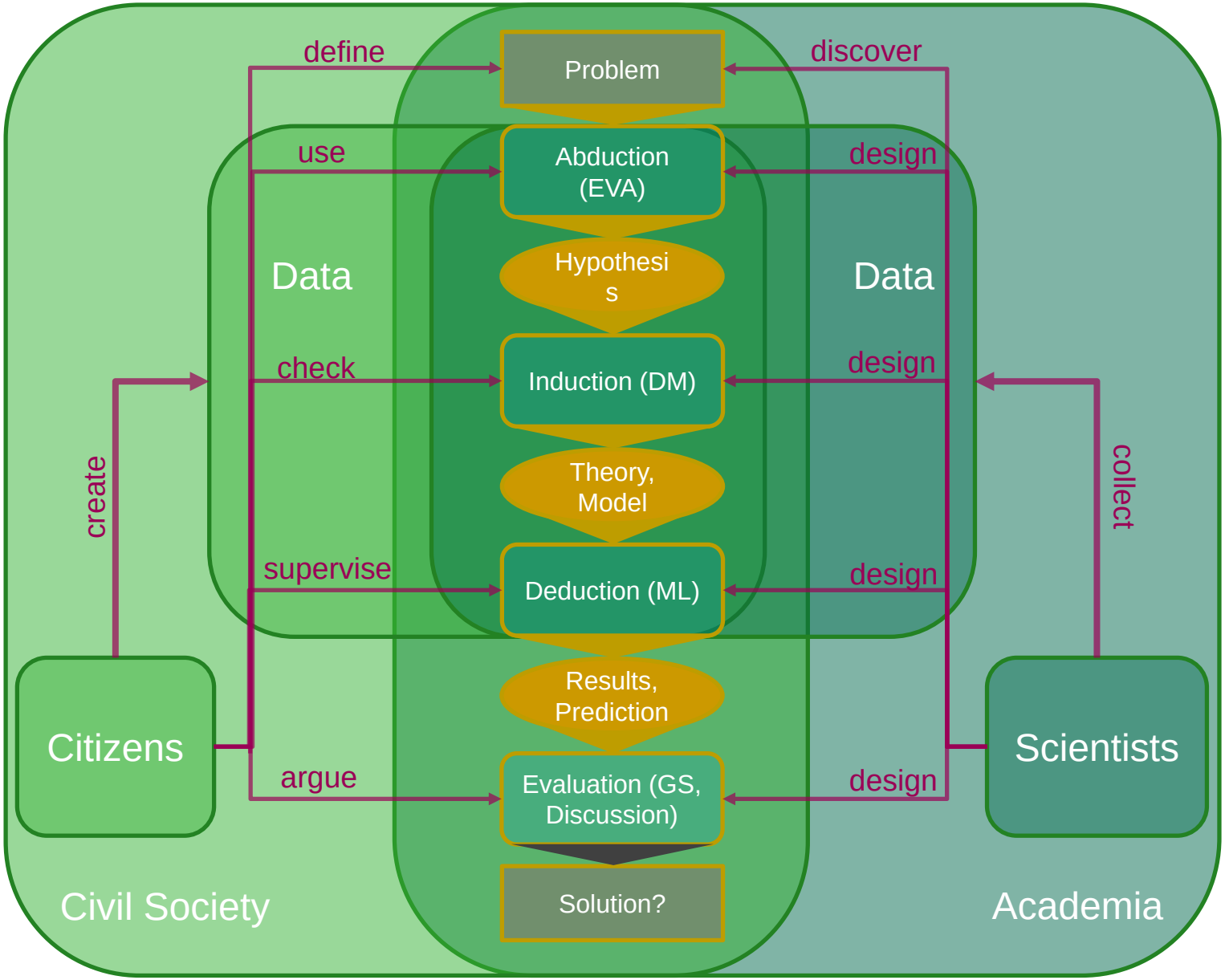
- **Smart Cities need Smart Citizens**
- Hybrid (human & machine) geoprocessing
- Mining geosocial media for place-related information
- Moving forward

OPENING UP THE BLACK BOX

SMART CITIES NEED SMART CITIZENS



SMART CITIZEN SCIENCE FRAMEWORK:
CITIZEN DATA SCIENTISTS TO THE RESCUE?



PILOT STUDY: PLACE AS THE GREAT CONNECTOR

SMART CITIES NEED SMART CITIZENS

- User-generated geographic content offers rich and multi-faceted view on
  - perception of geographic places by users
  - shared semantics of geographic places
- Knowledge about place semantics
  - improves interoperability between datasets
  - information retrieval for future streams of geodata
- Place links citizen's needs and infrastructure affordances

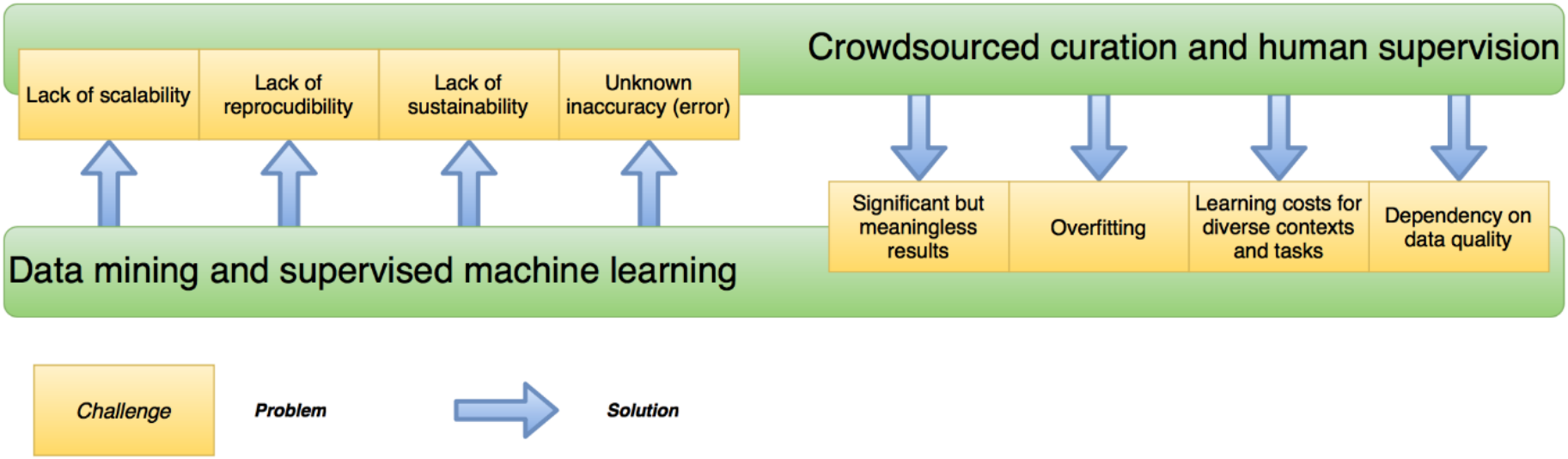
Mind the (semantic) gap!

WHAT THIS TALK IS ABOUT…

- Smart Cities need Smart Citizens
- **Hybrid (human & machine) geoprocessing**
- Mining geosocial media for place-related information
- Moving forward

COMPLEMENTARY STRENGTHS

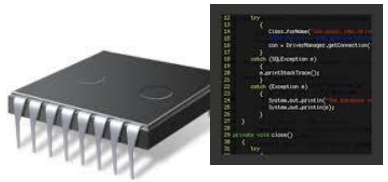
HYBRID (HUMAN & MACHINE) GEOPROCESSING



AN EXAMPLE HYBRID WORKFLOW

HYBRID (HUMAN & MACHINE) GEOPROCESSING

Real-time
- enrichment
- data mining
- machine learning



Crowdsourced supervision

User-generated geographic content (UGGC)

Collaborative analysis

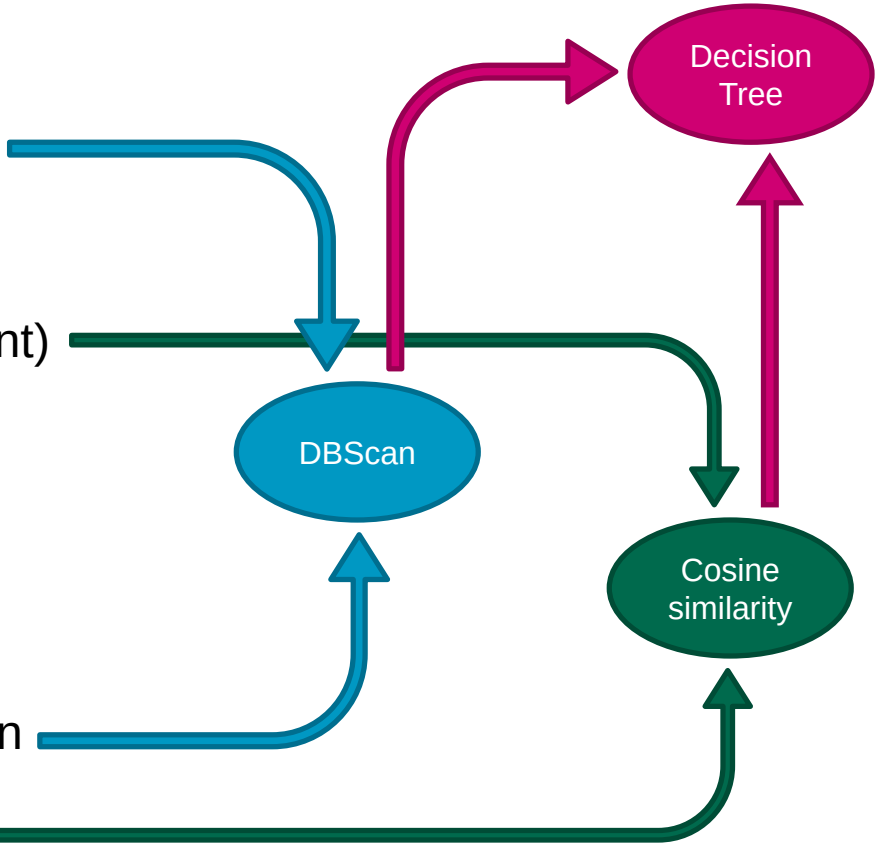
WHAT THIS TALK IS ABOUT…

- Smart Cities need Smart Citizens
- Hybrid (human & machine) geoprocessing
- **Mining geosocial media for place-related information**
- Moving forward

CONCEPTUALIZING & OPERATIONALIZING PLACES

MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION

- Agnew (1987):
  - Specific location (where?)
  - Locale (properties)
  - Sense of place (attachment)
- Winter and Freksa (2012):
  - place through contrast
    - Spatio-temporal location
    - Semantics



FLICKR PHOTO METADATA AS INPUT

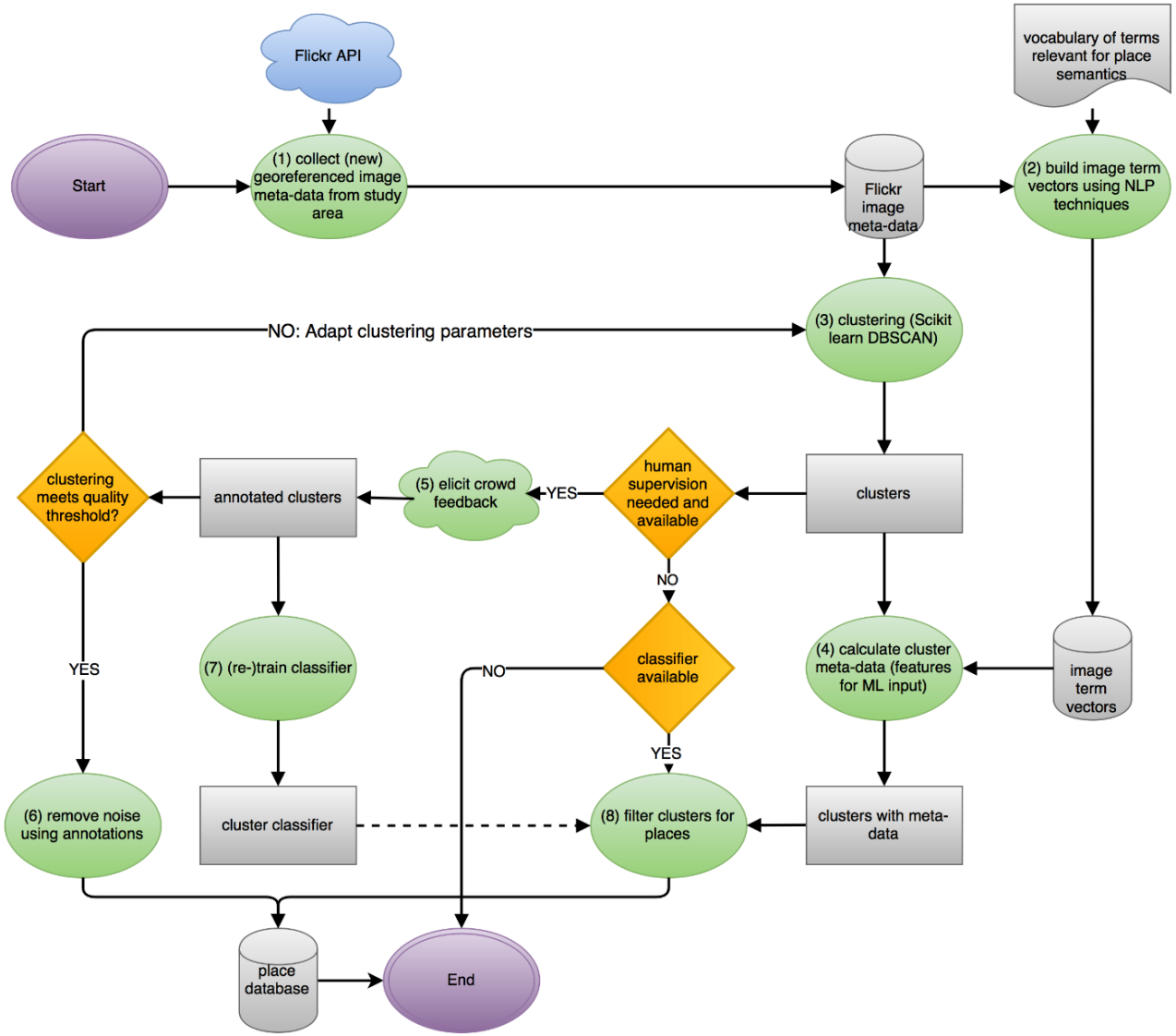
MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION

- Flickr Search API:
  - Rich metadata (tags, description, title)
  - Relatively stable
  - Flexible API
  - Many (~20%) geolocated
- What: All geo-referenced Flickr images
- Where: Greater London Area bounding box
- When: until November 2014

➢ More than five million images.

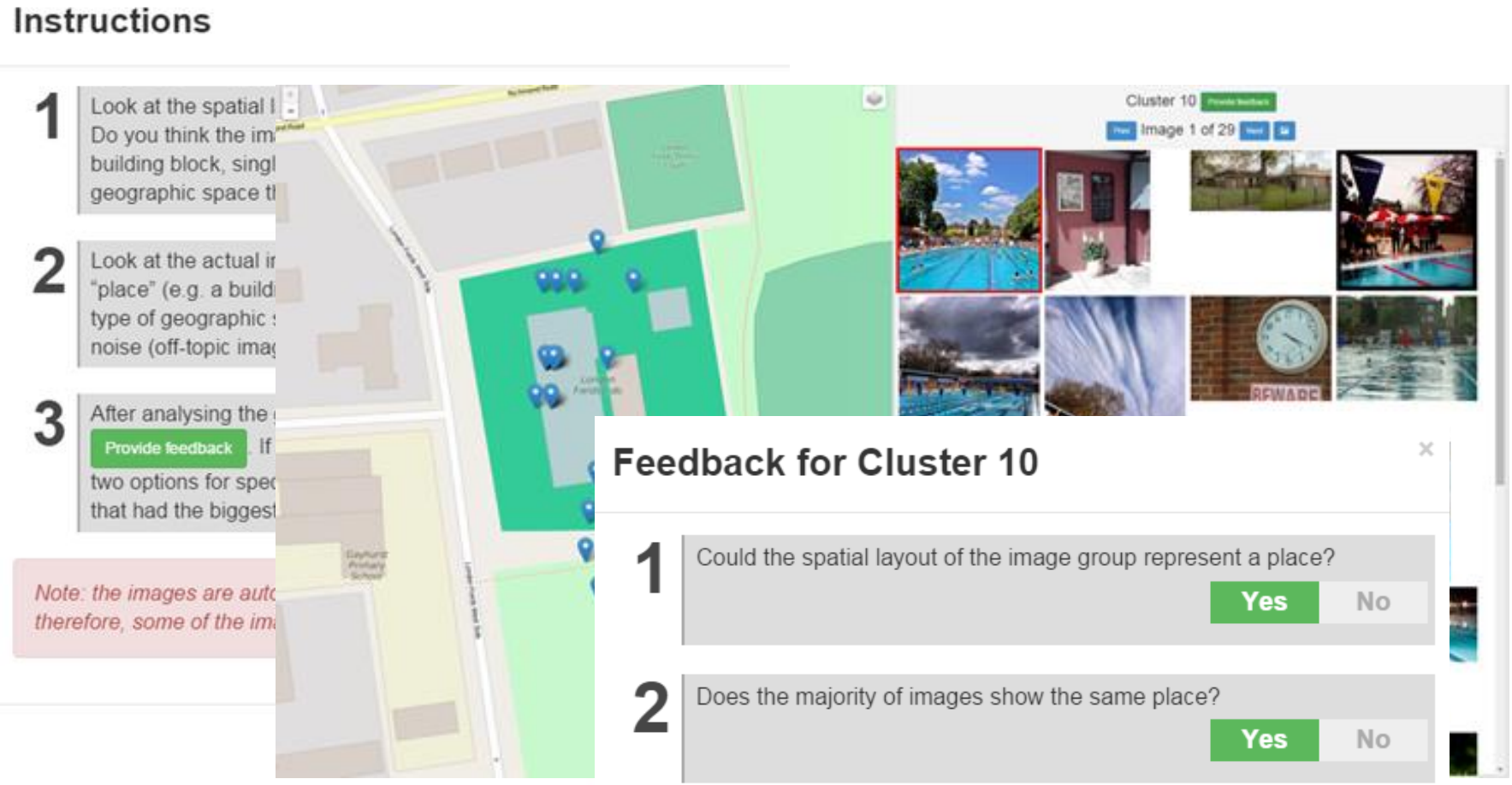
IMPLEMENTATION OF WORKFLOW

MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION



DEMO OF INTERFACE

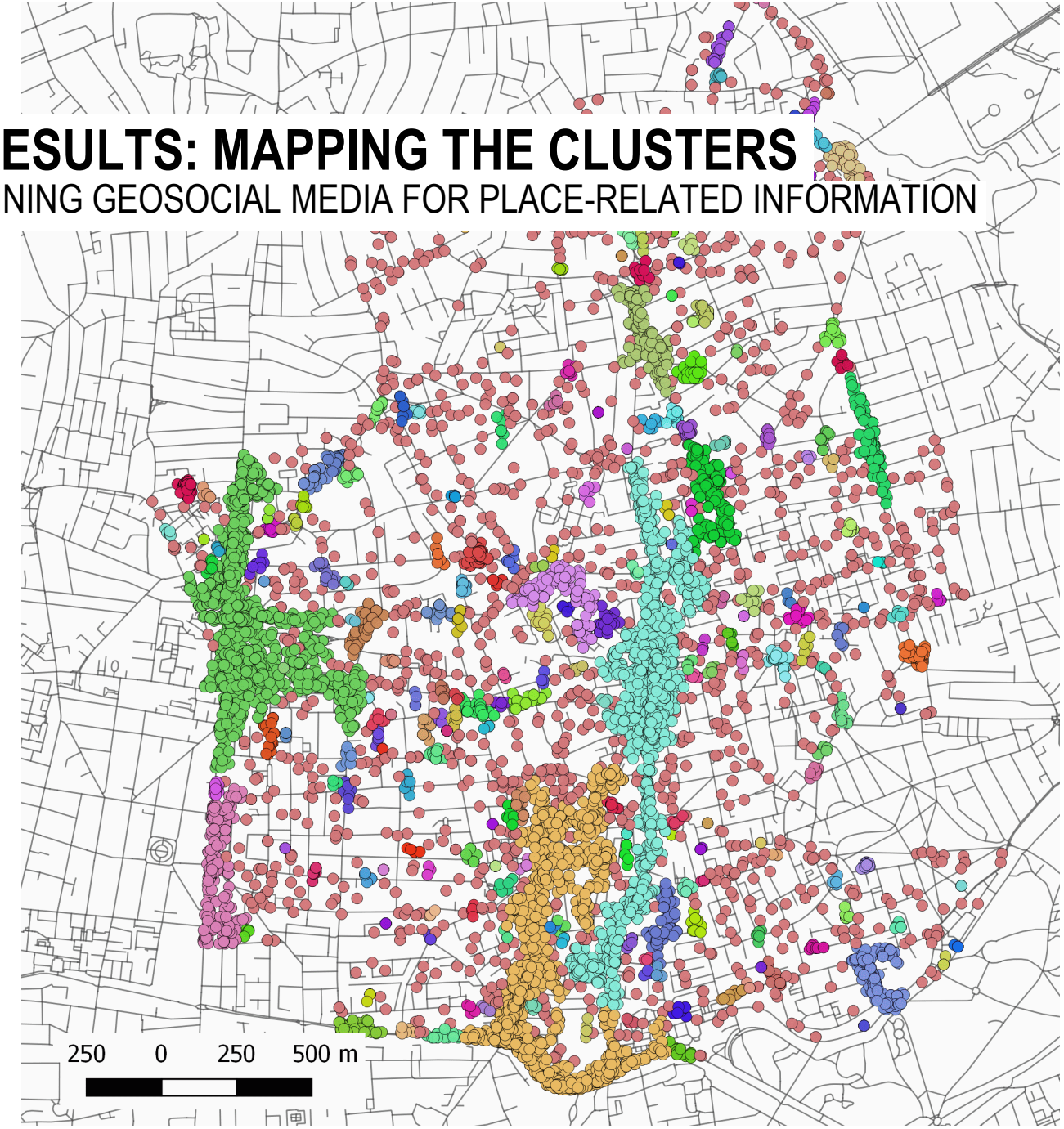
MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION



http://gip.itc.nl/resources/ggarcia/cluster_validator

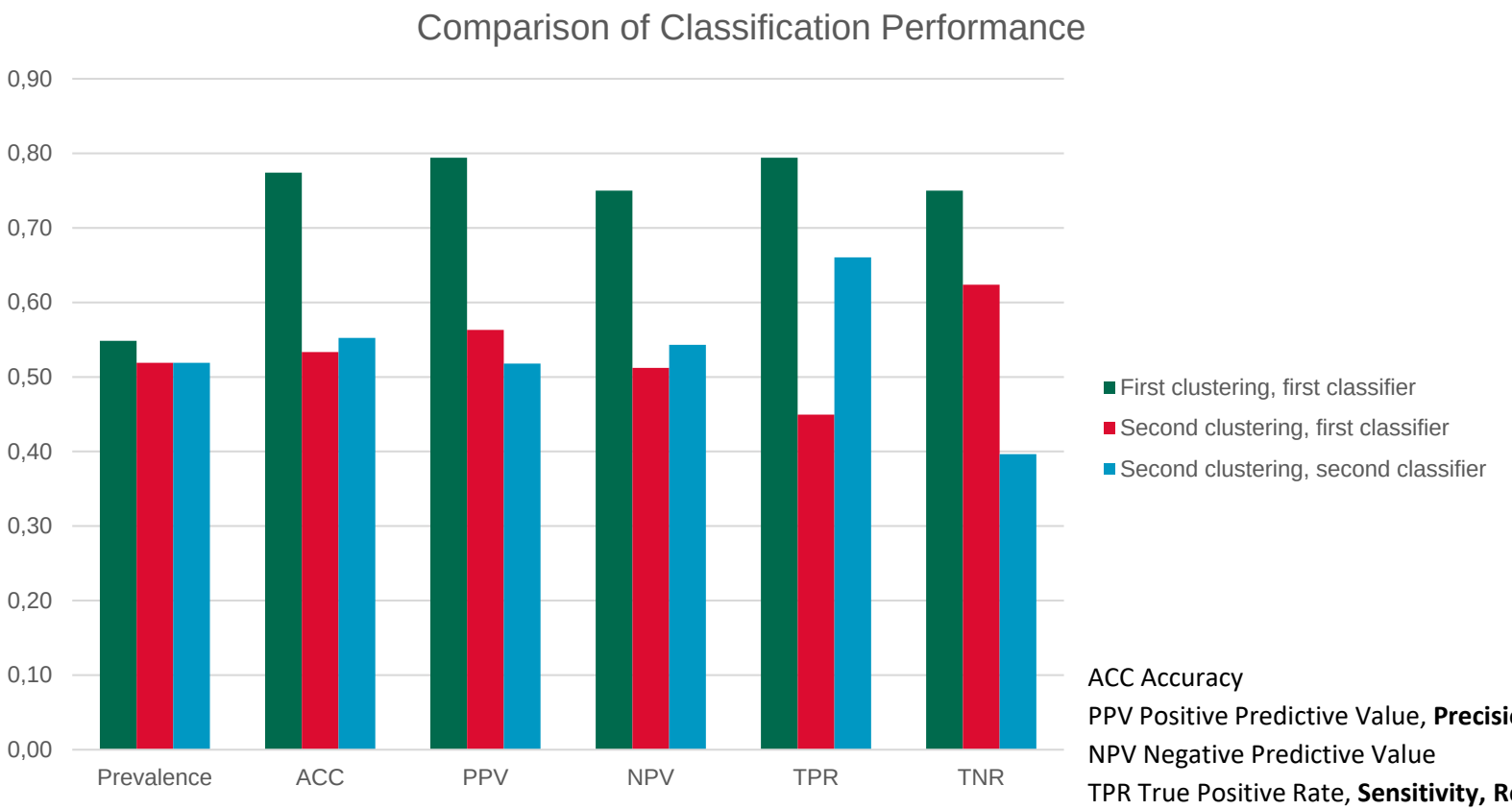
RESULTS: MAPPING THE CLUSTERS

MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION



RESULTS: CLASSIFYING THE CLUSTERS

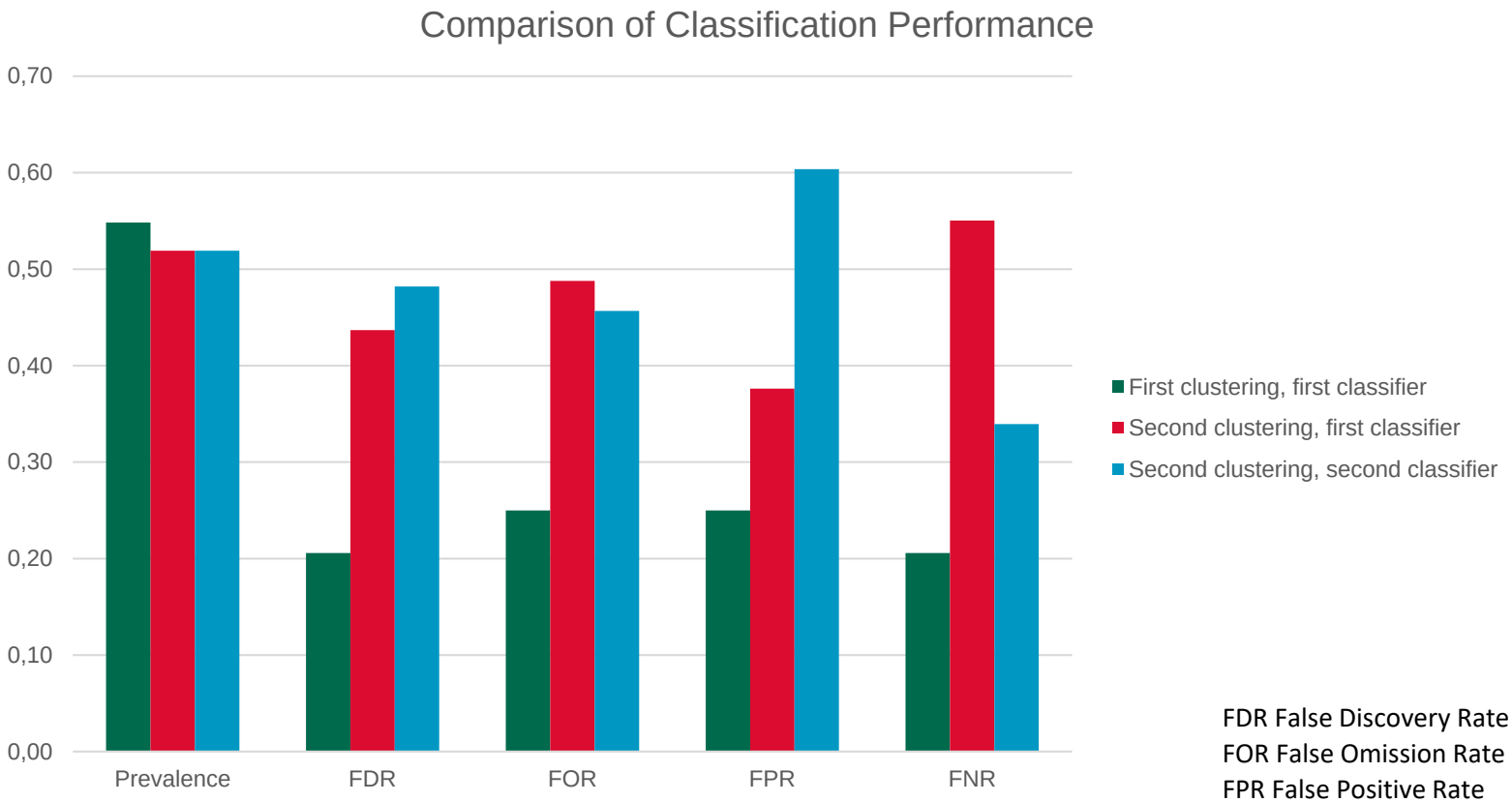
MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION



ACC Accuracy
PPV Positive Predictive Value, **Precision**
NPV Negative Predictive Value
TPR True Positive Rate, **Sensitivity, Recall**
TNR True Negative Rate, **Specificity**

RESULTS: CLASSIFYING THE CLUSTERS

MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION



FDR False Discovery Rate
FOR False Omission Rate
FPR False Positive Rate
FNR False Negative Rate

LESSONS LEARNED

MINING GEOSOCIAL MEDIA FOR PLACE-RELATED INFORMATION

- A single source of user-generated geographic content might not be rich enough to capture places.
- Inter-rater agreement surprisingly high considering lack of local knowledge.
- Difficult to recruit people with local knowledge.
- Iterative workflow requires good coordination and orchestration with crowdsourced efforts.
- Adjustment of clustering hyper-parameters requires retraining of classifier.

WHAT THIS TALK IS ABOUT…

- Smart Cities need Smart Citizens
- Hybrid (human & machine) geoprocessing
- Mining geosocial media for place-related information
- **Moving forward**

MORE DATA?
MOVING FORWARD

- More geosocial data sources
  - Paint richer picture
  - Have more legal constraints and terms of service
  - Stream of data, not batches or packages
- More authoritative data sources
  - Add value
  - Hide/wrap queries of portals to lower entrance barrier

IMPROVED ANALYSIS?
MOVING FORWARD

- Improved data mining techniques (ST-DBSCAN, spectral clustering)
- Improved machine learning techniques (SVM, random forests, online and active learning, more features)
- Improved semantics (other vocabularies, ontologies)
- Geographic transferability (see yesterday's talk): Overfitting vs. specific models for specific (local) data
- Balance with accessibility
- Ensure reproducibility and avoid Black Box

MORE LOCAL KNOWLEDGE?
MOVING FORWARD

- Bigger case study?
- Different case study?
- Managing the crowd
  - Recruit locals for supervision and validation
  - Retaining contributors
  - Cater for dabblers and dropouts
- To gamify or not?

MORE USER NEEDS?

MOVING FORWARD

- Asks locals (not tourists) about interests
- What makes a citizen smarter?
  - Finding a shop?
  - Participate in evidence-based policy-making?

LAST STOP: ADVERTISEMENT
MOVING FORWARD

Thank you for your attention!

Questions & Comments:

f.o.ostermann@utwente.nl

Workshop on

Semantic Information of Geosocial Data for the Mobile Age

September 21 – 22, Zürich

Call for Papers open now

More info: http://www.geo.uzh.ch/~rsp/semGeoSoc/index.html