Citizen Data Scientists:

An interactive, crowdsourced approach to validation

and supervision

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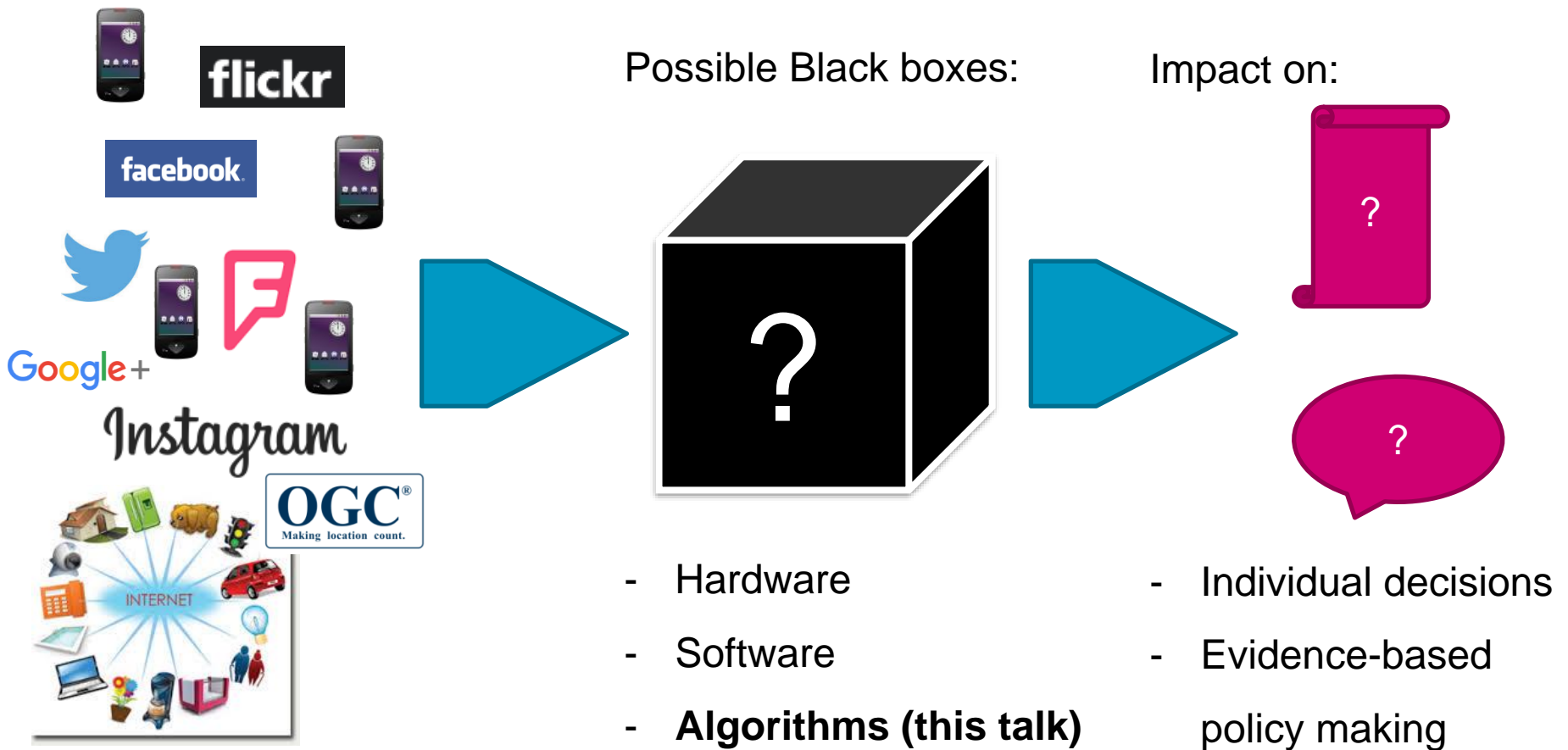
Raul Zurita-Milla, Menno-Jan Kraak

Presentation at the

Citizen Science Association's International Conference

19.05.2017, St. Paul

WHAT THIS TALK IS ABOUT…

- **Smart Cities need Smart Citizens**

- Hybrid (human & machine) geoprocessing

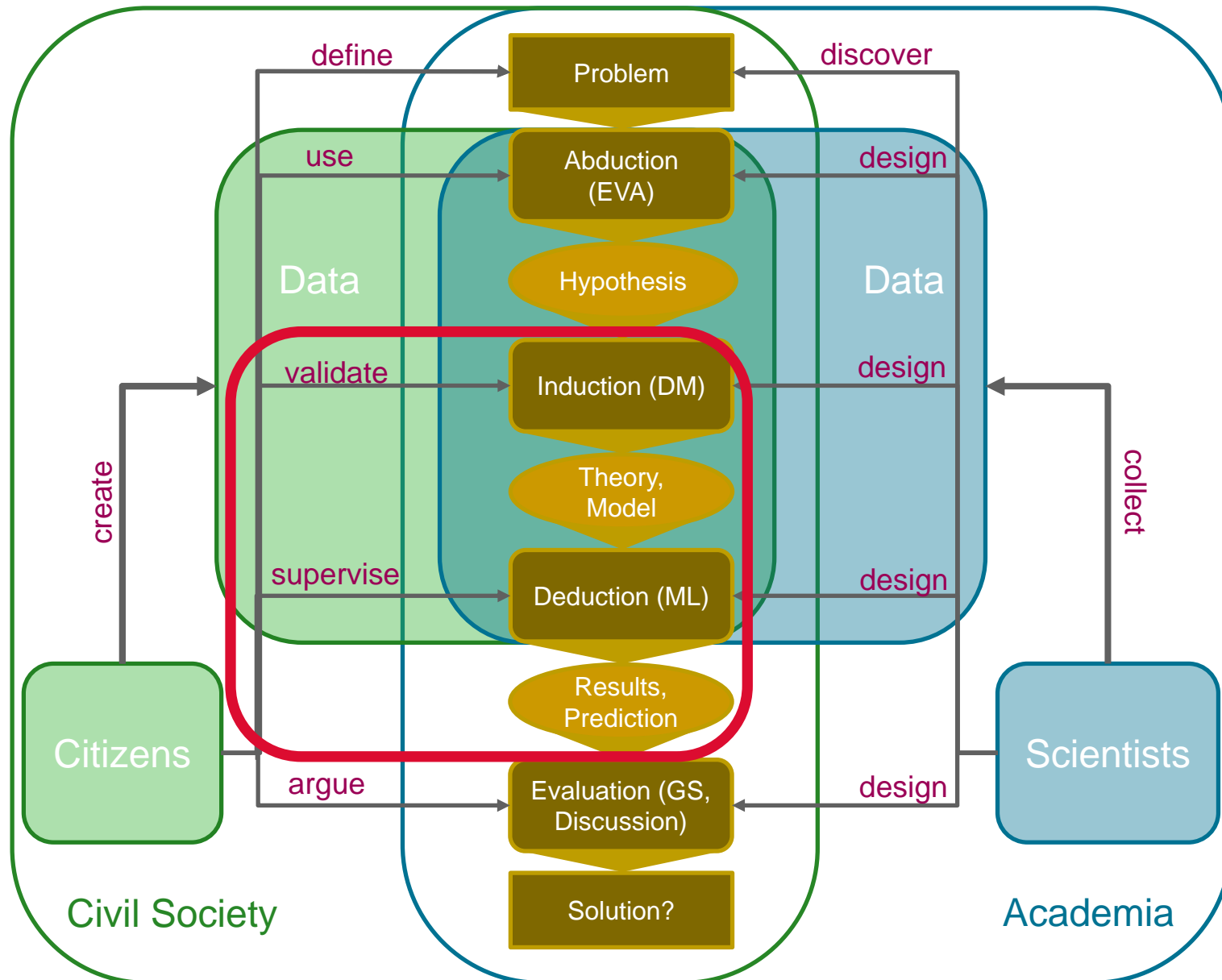- Pilot study on finding places from Flickr metadata

- Moving forward

OPENING UP THE BLACK BOX
SMART CITIES NEED SMART CITIZENS

Possible Black boxes:

Impact on:

- Hardware
- Software
- **Algorithms (this talk)**

- Individual decisions
- Evidence-based

  policy making

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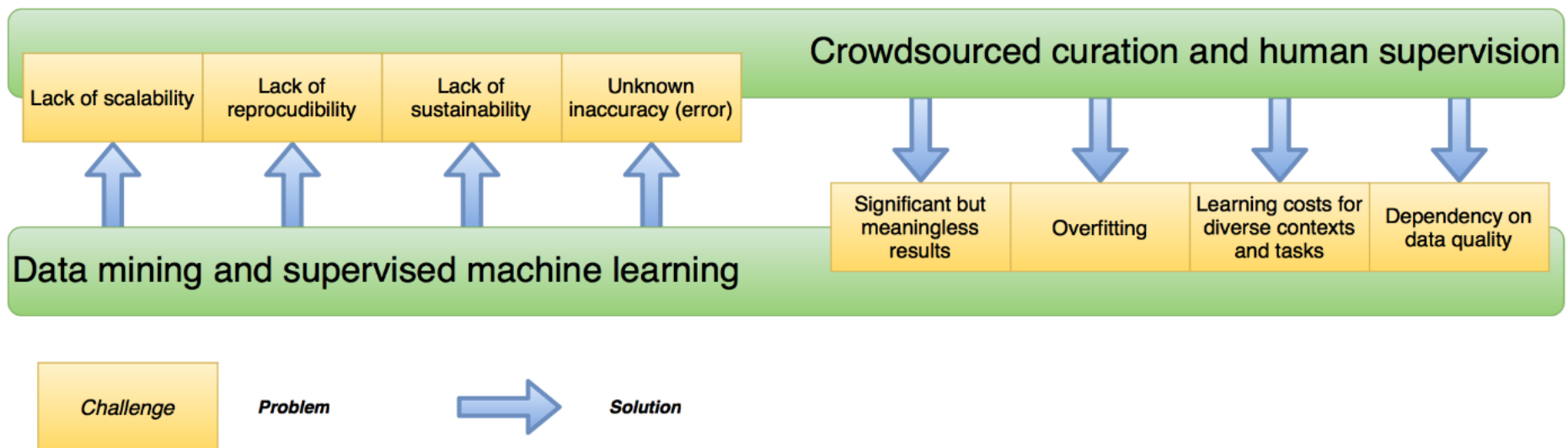## CITIZEN DATA SCIENTISTS TO THE RESCUE?

WHAT THIS TALK IS ABOUT…

- Smart Cities need Smart Citizens

- **Hybrid (human & machine) geoprocessing**

- Pilot study on finding places from Flickr metadata
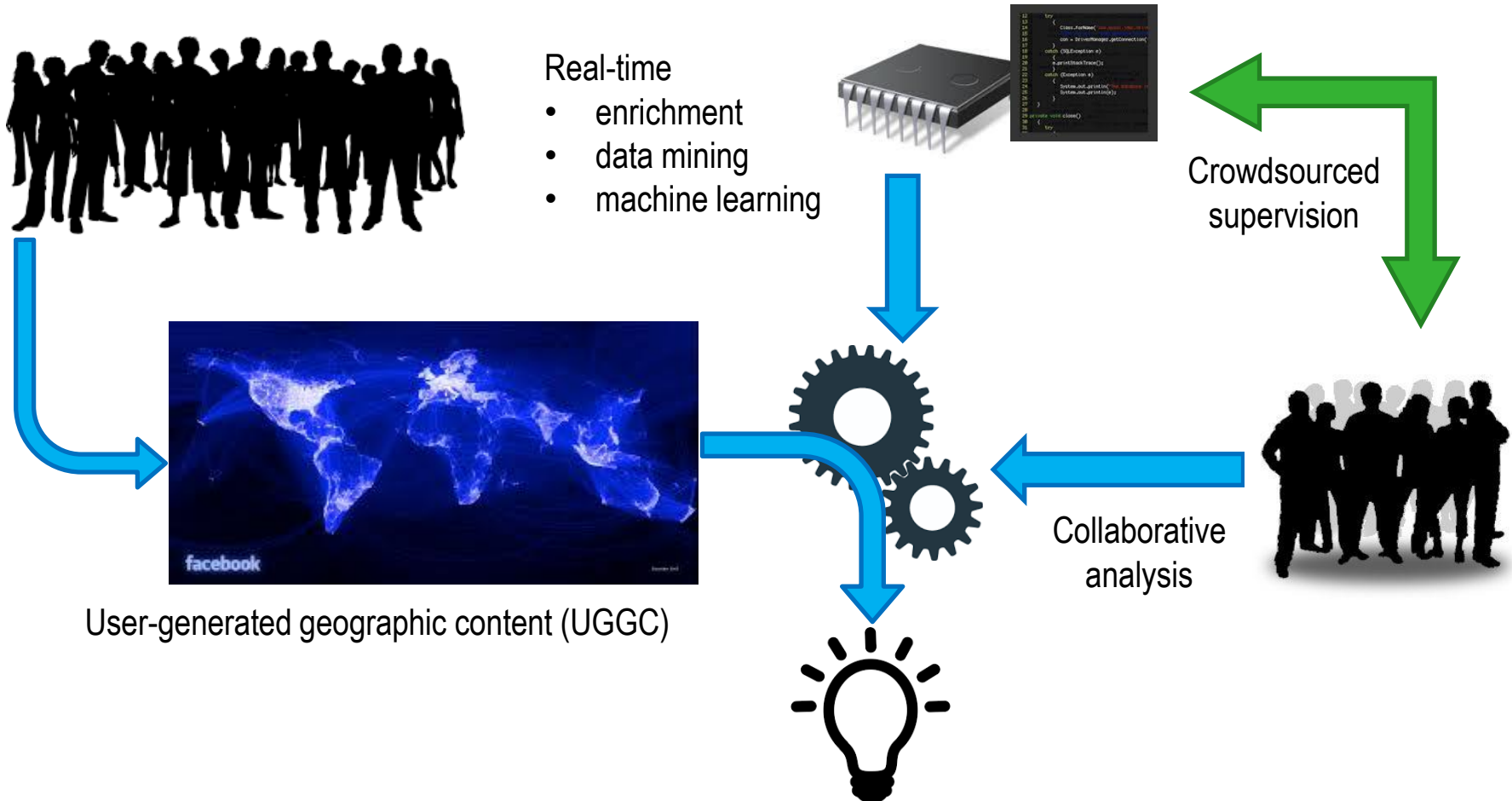
- Moving forward

COMPLEMENTARY STRENGTHS
HYBRID (HUMAN & MACHINE) GEOPROCESSING

AN EXAMPLE HYBRID WORKFLOW
HYBRID (HUMAN & MACHINE) GEOPROCESSING

Real-time
- enrichment
- data mining
- machine learning

Crowdsourced supervision

Collaborative analysis

User-generated geographic content (UGGC)

WHAT THIS TALK IS ABOUT…

- Smart Cities need Smart Citizens

- Hybrid (human & machine) geoprocessing

- **Pilot study on finding places from Flickr metadata**

- Moving forward

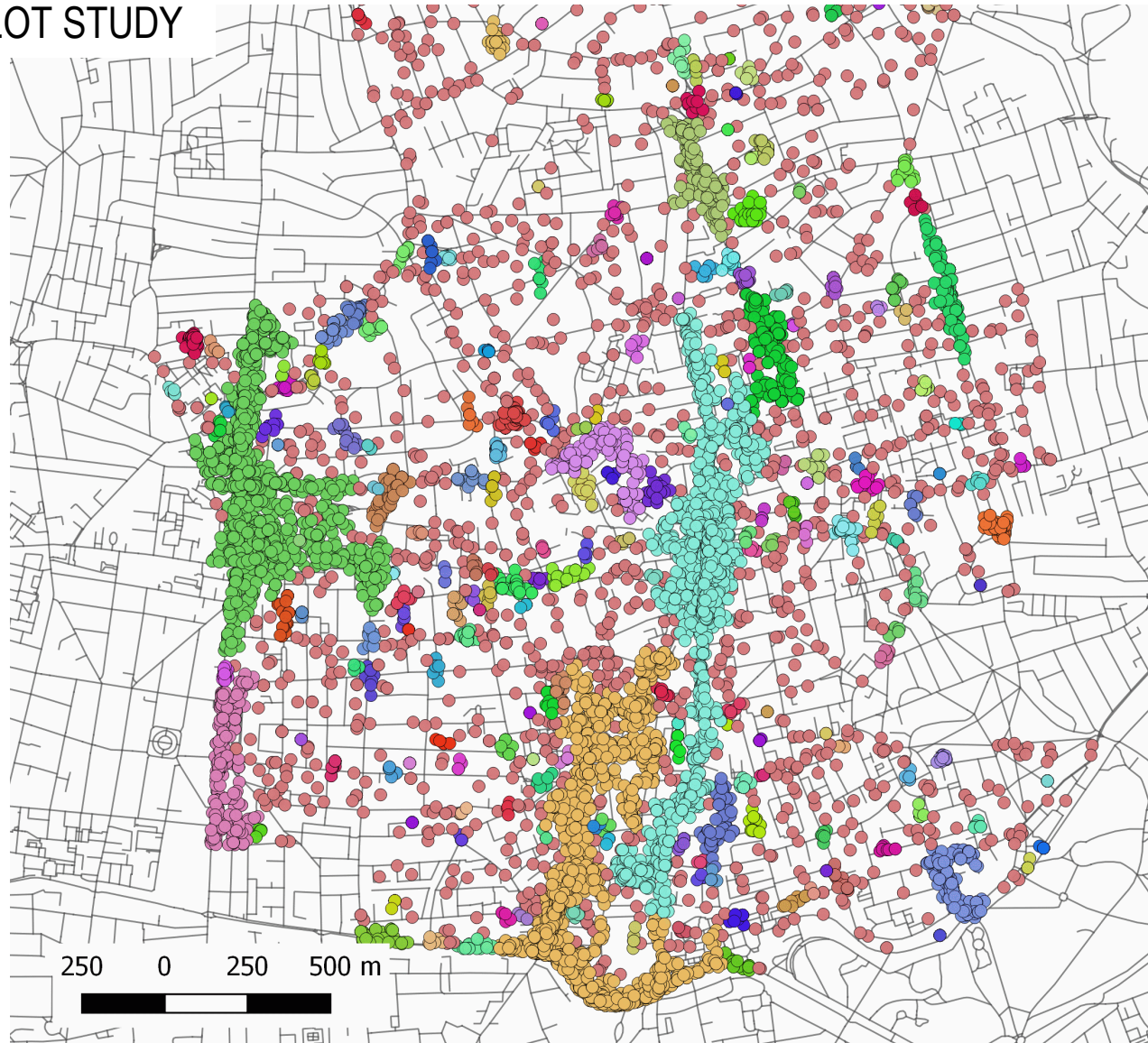# CONCEPTUALIZING & OPERATIONALIZING PLACES
PILOT STUDY

- Agnew (1987):

  - Specific location (where?)

  - Locale (properties)

  - Sense of place (attachment)

- Winter and Freksa (2012):

  - place through contrast

    - Spatio-temporal location

    - Semantics

Decision Tree

DBScan

Co si ty
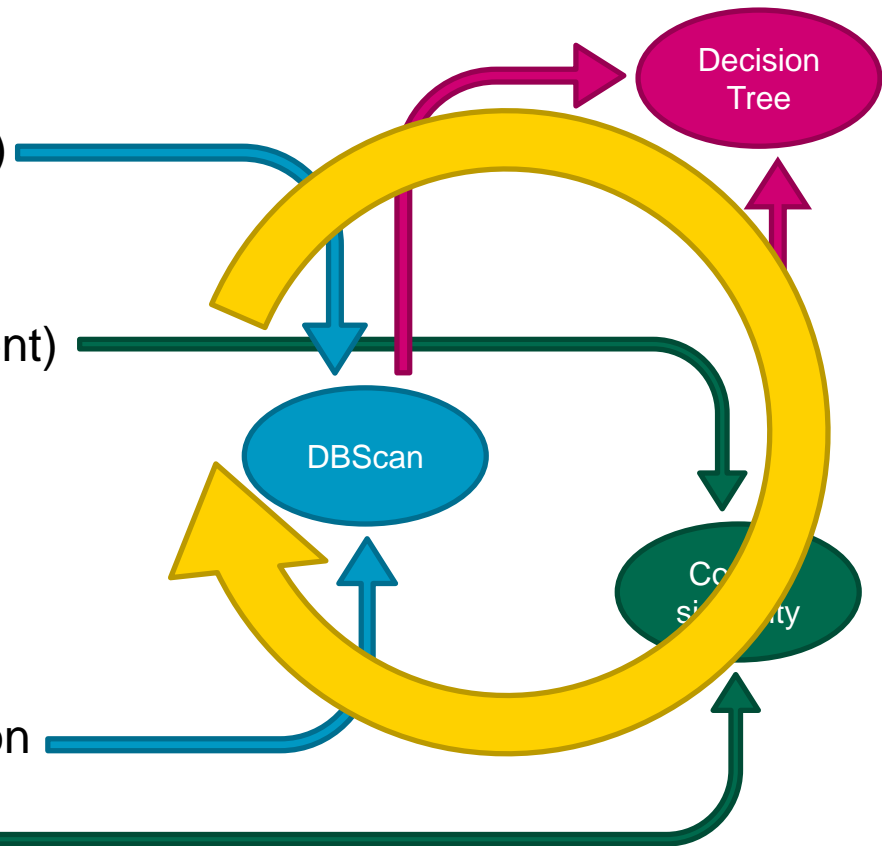
RESULTS: MAPPING INITIAL CLUSTERS IN LONDON
PILOT STUDY

250   0   250   500 m

DEMO OF VALIDATION AND SUPERVISION INTERFACE
PILOT STUDY

Instructions

1 Look at the spatial l
   Do you think the im
   building block, singl
   geographic space tl

2 Look at the actual i
   "place" (e.g. a build
   type of geographic :
   noise (off-topic imag

3 After analysing the
   Provide feedback If
   two options for spec
   that had the biggest

Note: the images are aut
therefore, some of the im

Cluster 10 **Provide feedback**
Image 1 of 29

Feedback for Cluster 10   ×

1 Could the spatial layout of the image group represent a place?
   Yes   No

2 Does the majority of images show the same place?
   Yes   No

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

Submit   Skip and show next   Close

LESSONS LEARNED
PILOT STUDY

- Inter-rater agreement surprisingly high

- Difficult to recruit people with local knowledge.

- Iterative workflow requires good coordination and orchestration.

- Adjustment of clustering parameters required retraining of classifier.

- Classifier performance not great: A single content source might not be rich enough

WHAT THIS TALK IS ABOUT…

- Smart Cities need Smart Citizens

- Hybrid (human & machine) geoprocessing

- Pilot study on finding places from Flickr metadata

- **Moving forward**

MORE DATA AND IMPROVED ANALYSIS?
MOVING FORWARD

- More data sources (geosocial data and authoritative)

- Improve

  - data mining techniques (ST-DBSCAN, spectral clustering)

  - **machine learning techniques (SVM, random forests, online and active learning, more features)**

  - semantics (other vocabularies, ontologies)

- But

  - Balance with accessibility

  - Ensure reproducibility and avoid Black Box

TENSORFLOW AND WQ
MOVING FORWARD

- Great Create Together Day, big thanks to S. Andrew Sheppard and Marco Willi (and Lucy Fortson, …)

- Data collection with WQ-Framework

- Tensorflow learner to remove obvious noise (food, etc.) based on **image content**

- Trained on UMN campus data only

- Geographic transferability? Re-training or overfitting?

- https://tensorflow.wq.io

- https://insideoutside.wq.io/

TENSORFLOW AND WQ
MOVING FORWARD

TENSORFLOW AND WQ
MOVING FORWARD

MORE LOCAL KNOWLEDGE AND USER NEEDS?
MOVING FORWARD

- Bigger or entirely different case study?

- Recruit locals for supervision and validation

- To gamify or not?

- 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

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

Thank you for your attention!

Questions & Comments:

f.o.ostermann@utwente.nl