GIScience Workshop on VGI for Science: Quality and Credibility

Volunteered Geographic Science?

Dr Colin Robertson

Wiflrid Laurier University
www.thespatiallab.org

LAURIER
Inspiring Lives.

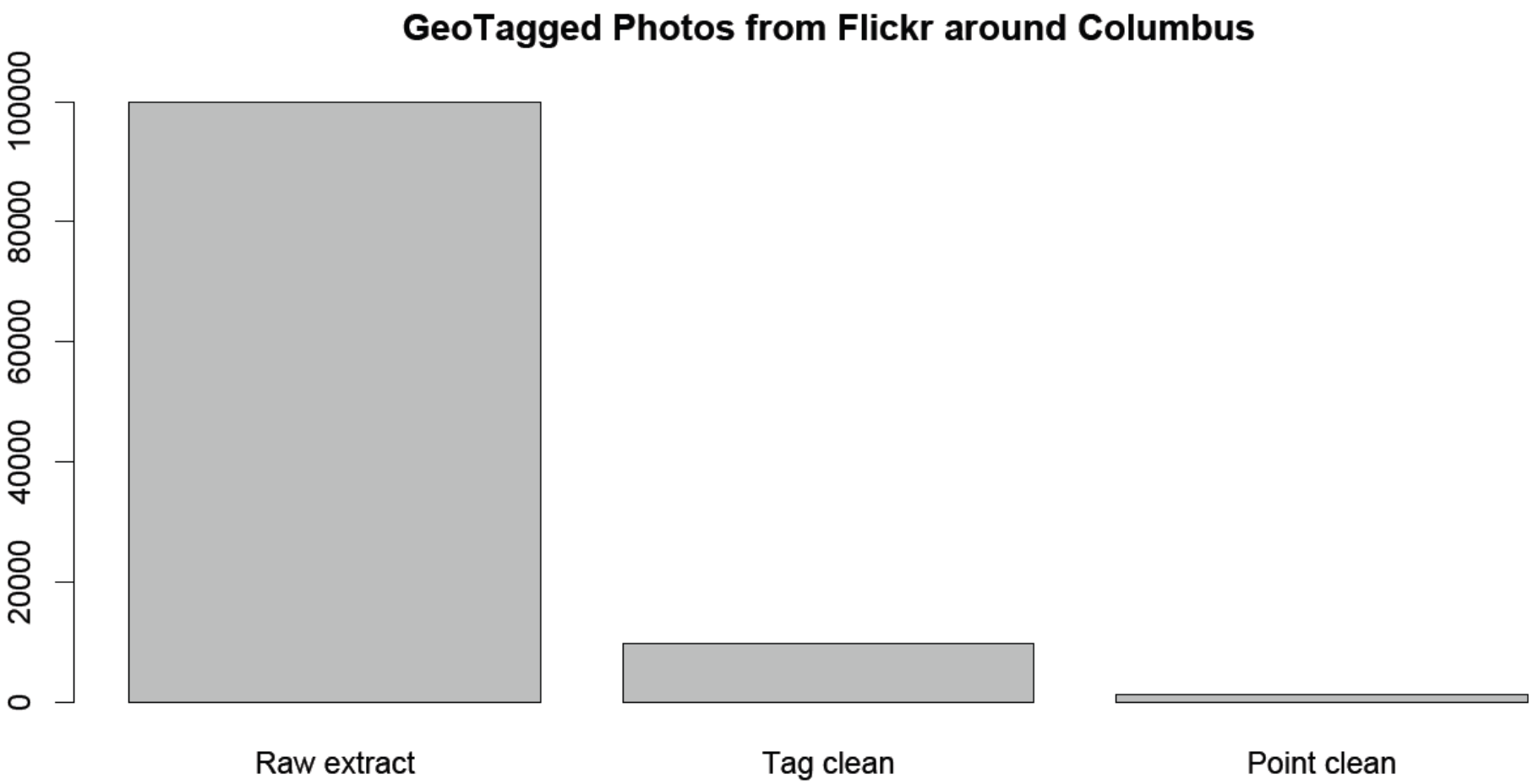
the spatial lab

Q. Do you think the volume of VGI available in your area of research and/or practice is adequate or deficient? If deficient, what can be done ~~to increase it~~?

to make use of what is available?

A: deficient, usually

Filtering out noise that comes with volume



- Is Flickr is an extreme case?
  - massive duplication
  - bulk tagging
  - non-geographical content
  - not intended as a mapping platform
  - multiple, conflicting objectives in how users are creating, submitting data, and variable relationships to 'geography'

1) Classification of VGI

Figure 2.5: VGI framework

Scientific knowledge VGI Type 1	Local knowledge VGI Type 2	Personal knowledge VGI Type 3
- volunteered - Objective - Structured - 2 way or n way - Digitizing, GPS, twitter - Only points or points, lines, polygons	- facilitated-VGI - Subjective - Unstructured - 1 way, 2 way or n way - Selection, Drawing, geocoding - Points, lines, polygons	- Kept private - Subjective - Unstructured - n way - Location through networks, geocoding - Only points

Deparday 2010

VGI & Syndromic Surveillance

- Syndromic surveillance: using automated data acquisition / statistical alerts to monitor disease indicators to detect outbreaks of disease earlier than would otherwise be possible with traditional public health methods

Syndromic Data Sources

EARLY DETECTION → LATER DETECTION

GOLD STANDARDS

ANIMALS

- Vets
- Zoos
- Agribusiness

Environmental

- Pollen counts
- Humidity
- Temperature
- Wind Speed/ direct.
- Allergy Index
- Pollution

HUMAN BEHAVIORS

- OTC Pharm
- Absenteeism
- Utilities
- Coughs
- Web Queries
- Traffic
- Survey
- Public Transport
- Cafeteria
- Video Surv
- Newsgroup

NON TRADITIONAL USES

- Tests ordered
- Complaints
- Poison Centers
- 911 Calls
- EMS Runs
- Nurse Calls
- ER Visits
- Prescriptions
- Radiograph Reports

CLINICAL DATA

- Test Results
- Diagnosis
- Influenza isolates
- Medical Examiner

GOLD STANDARDS

- Test Results
- Sentinel MD
- Investi-gations

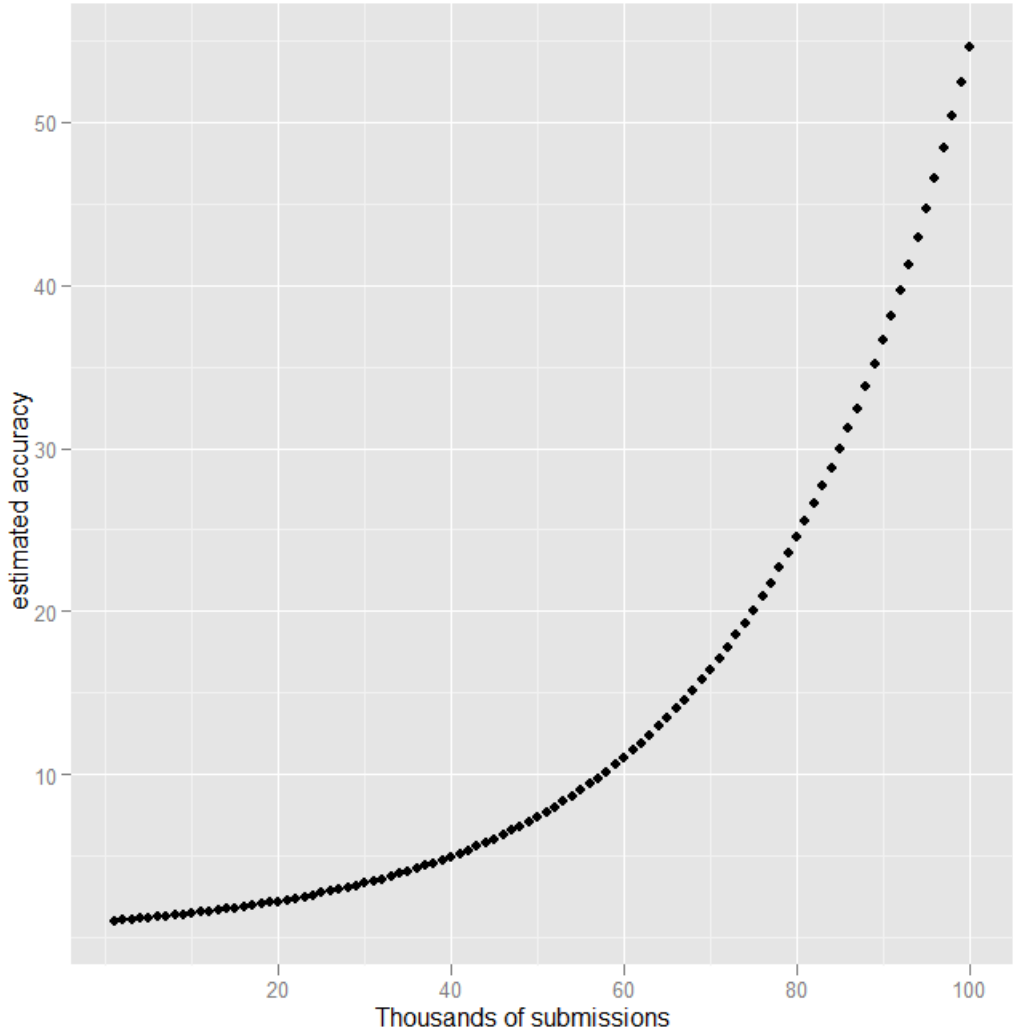
Source: Henry R. Rolka

How does VGI differ from syndromic surveillance

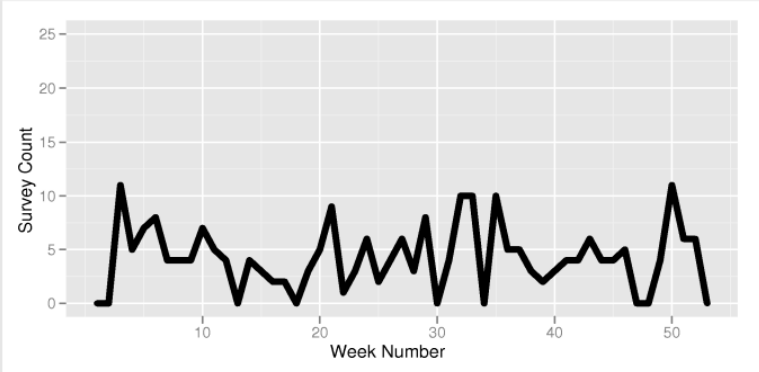
- Datasets used in syndromic surveillance differ with respect to volume:
  - behaviour oriented: large volume
  - services oriented: medium volume
  - pathogen oriented: low volume

- SS generally has a clear, narrowly defined objective

Can we establish a relationship between VGI usability and volume for a given context?

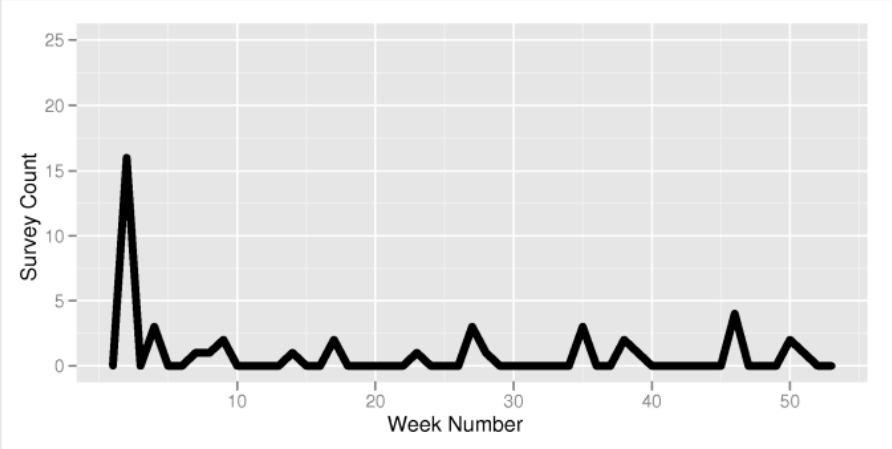


Vet ID 22



Survey Date	Species	Reported Symptoms	Suspected Diagnoses
01/12/09	cattle	lameness	Footrot
01/12/09	cattle	decreased feed intake/milk production	OTHER diagnosis
01/12/09	cattle	decreased feed intake/milk production	Bloat
01/12/09	cattle	decreased feed intake/milk production	Bacteremia/Septicemia
01/13/09	cattle	respiratory signs	[illegible]
01/13/09	cattle	decreased feed intake/milk production	[illegible]
01/13/09	cattle	decreased feed intake/milk production	[illegible]
01/16/09	cattle	decreased feed intake/milk production	[illegible]
01/16/09	cattle	decreased feed intake/milk production	[illegible]
01/16/09	cattle	decreased feed intake/milk production	[illegible]
01/16/09	cattle	decreased feed intake/milk production	[illegible]
01/19/09	cattle	decreased feed intake/milk production	[illegible]
01/19/09	cattle	decreased feed intake/milk production	[illegible]
01/19/09	cattle	decreased feed intake/milk production	[illegible]
01/22/09	cattle	decreased feed intake/milk production	[illegible]
01/22/09	cattle	decreased feed intake/milk production	[illegible]
01/26/09	cattle	decreased feed intake/milk production	[illegible]
01/27/09	cattle	decreased feed intake/milk production	[illegible]
01/27/09	cattle	decreased feed intake/milk production	[illegible]
01/28/09	cattle	lameness	[illegible]

Vet ID 26



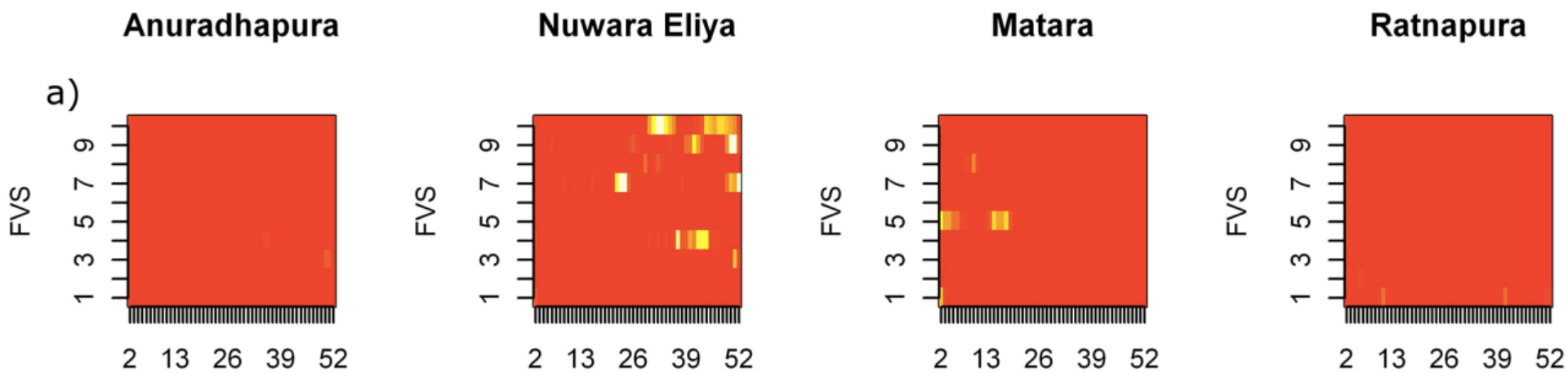
Survey Date	Species	Reported Symptoms	Suspected Diagnoses
01/08/09	buffalo	decreased feed intake/milk production	OTHER diagnosis
01/08/09	buffalo	decreased feed intake/milk production	Malnutrition
01/08/09	cattle	decreased feed intake/milk production	Mastitis
01/08/09	cattle	decreased feed intake/milk production	Bloat
01/08/09	cattle	decreased feed intake/milk production	Bloat
01/08/09	poultry	decreased egg production/weight gain/ appetite	IBD/gumboro
01/08/09	poultry	decreased egg production/weight gain/ appetite	Heat stress
01/08/09	poultry	decreased egg production/weight gain/ appetite	Coccidiosis
01/09/09	buffalo	decreased feed intake/milk production	OTHER diagnosis
01/09/09	cattle	decreased feed intake/milk production	Mastitis
01/09/09	cattle	decreased feed intake/milk production	Malnutrition
01/09/09	cattle	decreased feed intake/milk production	Bloat

Robertson, C., Sawford, K., Gunawardana, W.S.N., Nelson, T.A., Nathoo, F., & Stephen, C. (2011). A hidden markov model for analysis of frontline veterinary data for emerging zoonotic disease surveillance. PLoS One.

2) Statistical modelling for usability

Parameter	Posterior mean (95% credible interval)	Standard deviation
μ_1	0.34 (0.28–0.41)	0.10
μ_2	6.65 (6.05–7.29)	0.05
Training	1.19 (1.08–1.32)	0.05
Years	0.59 (0.55–0.63)	0.03
Male	0.90 (0.84–0.96)	0.03

doi:10.1371/journal.pone.0024833.t003



Robertson, C., Sawford, K., Gunawardana, W.S.N., Nelson, T.A., Nathoo, F., & Stephen, C. (2011). A hidden markov model for analysis of frontline veterinary data for emerging zoonotic disease surveillance. PLoS One.

3) System-level Contextual Factors

- **Scale** - The spatial and temporal extent of the system (e.g., local/regional/national/ international)
- **Scope** - The intended target of the system (e.g., single disease/multiple disease, single host/multiple host, known pathogens/unknown pathogens)
- **Function** - The objective(s) of the systems (outbreak detection, outbreak characterization, outbreak control, case detection, situational awareness, biosecurity and preparedness
- **Disease characteristics** - Is the pathogen infectious? Is this a chronic disease? How does it spread? What is known about the epidemiology of the pathogen?
- **Technical** - The level of technological sophistication in the design of the system and its users (data type and quality, algorithm performance, computing infrastructure and/or reliability, user expertise)

Robertson, C., T.A. Nelson, Y.C. MacNab, and A.B. Lawson. 2010. Review of Methods for Space-time Disease Surveillance. Spatial and Spatio-temporal Epidemiology.

2. How easy is it to collect VGI for scientific experiments? What can be done to ensure that VGI is as useful as possible in science?

- What is the relationship between data volume, access, and intrinsic quality?
  - easier to access = ↑ volume = ↓ quality?? (flickr)
  - difficult to access = ↑ volume = ↑ quality?? (proprietary Big Data sets)
- How can classification of VGI sources pre-condition our understanding of these relationships? How can we automatically classify a new source?
- What systemic factors of the research need need to be considered in the evaluation of VGI for scientific research?
  - space / time scale (& extent)
  - objectives: ground truth / validation, characterization / description
  - method: observational, experimental
- What methods are there for meeting our needs based on these criteria?
  - statistical modelling
  - quality metrics and aggregate indices
  - visualization