



Universiteit
Leiden
The Netherlands

Urban Computing

Dr. Mitra Baratchi

31 August 2020

Leiden Institute of Advanced Computer Science - Leiden University

Table of Contents

1. Practical matters

2. Introduction to the course

- What is urban computing?
- Applications
- Data sources

3. Lessons learned

Practical matters

Teaching assistants



Daniela Gawehtns
(Teaching Assistant)

Doctoral Student



Victor Neuteboom
(Teaching Assistant)

Master Student



Anne Liebens
(Teaching Assistant)

Master Student



Lau Bannenberg
(Teaching Assistant)

Master Student

Courseware

- Courseware access: <https://urbancomputing2021.github.io/>
- Other matters (Announcements, assignment hand-in, discussion forum): [Brightspace](#)

The screenshot shows the homepage of the course website at <https://urbancomputing2021.github.io/index.html>. The top navigation bar includes links for HOME, COURSE DESCRIPTION, COURSE MATERIAL, OUR TEAM, and PROJECT MATERIAL. Below the navigation is the course title "URBAN COMPUTING 2020-2021". A central feature is a line graph titled "Positive tests" showing the number of positive tests over 100 days since the first positive test for several countries. The legend identifies the countries: Netherlands (green), Belgium (orange), Italy (red), Sweden (dark red), Norway (purple), Spain (pink), United Kingdom (yellow), Germany (light green), and France (blue). The graph shows multiple peaks in testing activity across the period.

An endeavor to solve urban problems such as traffic and pollution using computing technology. Understand the phenomena, model the patterns and predict the future of cities. Please have a look at the (provisional) schedule (last updated: 21-08-2020).

An Intelligent Journey Through Urban Modeling

Communication

Before sending emails:

- Can you ask the question during the class?
- Can you use Brightspace discussion board?

The screenshot shows a course navigation bar at the top of a web page. On the left is the University of Leiden logo. Next to it is the course title "2021-WN Urban Computing". To the right of the title are several icons: a grid, an envelope, a speech bubble, and a bell with a red dot. Below the navigation bar is a dark blue header bar containing links for "Course Home", "Content", "Grades", "Course Tools", and "Help". Underneath the header bar, the main content area has a breadcrumb trail: "Discussions List > Urban Computing discussion board". The title "Urban Computing discussion board" is displayed prominently. Below the title is a dropdown menu labeled "Urban Computing discussion board".

Course schedule

	Date	Time	Lecture
Monday	31-Aug	14:15-16	Introduction
Monday	7-Sep	14:15-16	Time-series data (assignment 1)
Monday	14-Sep	14:15-16	Spatial data
Monday	21-Sep	14:15-16	-
Monday	28-Sep	14:15-16	Spatio-temporal data processing (assignment 2)
Monday	5-Oct	14:15-16	Machine learning for urban computing
Monday	12-Oct	14:15-16	Machine learning for urban computing 2
Monday	19-Oct	14:15-16	Visualization
	26-Oct		-
	02-Nov		-
TBA	09-Nov	TBA	Meeting with teams (with appointment)
TBA	16-Nov	TBA	Meeting with teams (with appointment)
Monday	23-Nov	9:15-18	Presentations
	30-Nov		-
	07-Dec		-
	14-Dec		-
	21-Dec		-

	Date	Workgroup
Friday	4-Sep	10.30-12.15
Thursday	10-Sep	9:15-11
Friday	18-Sep	10.30-12.15
	-	-
Thursday	1 Oct	9:15-11
Friday	9 Oct	10.30-12.15

	Date	Deadline
	-	Lab Exercise I (no hand-in)
Sunday	19-Sep	-
Tuesday	22-Sep	Selecting a paper
	-	Assignment 1
Tuesday	-	-
Tuesday	13-Oct	Assignment 2
Tuesday	20-Oct	Proposal
	-	-
	-	-
	-	-
	-	-
Monday	07-Dec	Deadline draft
Monday	14-Dec	Peer review
Monday	21-Dec	Final

- Individual meeting with teams will be scheduled later

Organization of class

Course structure

- Lectures (interactive/online using Kaltura)
- Workgroup sessions for assignments (physical/online according to the schedule)
- Practical lab material (self-study)
- Feedback on projects (meeting with teams)
- Presentation by students

Grading

- Assignments (individual) (35 %)
- Peer review (individual) (10%)
- Presentation (team) (15 %)
- Project (team) (40 %)
 - Novelty of the idea
 - Maturity of experiments (considering availability of data sources, or source codes)
 - Results and evaluation approach
 - Documentation

Course rules

- To pass the course both the combined grade and the assignments grade should be over 5.5
- Deadlines are fixed → no late submission
- Late submission is considered as resit
- Resit option for assignments and project has maximum grade of 7
- No resit option for presentations
- Submitting project proposal is compulsory (no submission implies deduction of 20% of the project grade)

Project

- Select a reference paper as a starting point
 - Bid on the paper bidding list available (select top 5 papers)
 - Alternatively you can bid for papers accepted to [Urban Computing workshop](#) any year
 - Register your bid (form is available through the website)
- Write a brief proposal including
 - problem statement, research question, methodology, evaluation approach, data sources
- Proceed with the project
- Write a report (6-8 pages (ACM-proceedings Latex template, sample-sigplan)) [[download](#)]

Introduction to the course

Table of content

1. Practical matters

2. Introduction to the course

- What is urban computing?
- Applications
- Data sources

3. Lessons learned

What does Urban Computing Mean?



Figure 1: Urban Computing

¹image source: <http://uctutorial.chinacloudsites.cn>

- A bit of history
- When was it mentioned first?
- The term urban computing was first introduced by Eric Paulos at the 2004 UbiComp (**Ubiquitous and pervasive computing**) conference

My own story as a member of Pervasive Systems

"Pervasive and ubiquitous computing (or "ubicomp") is a concept in software engineering and computer science where computing is made to appear anytime and everywhere. In contrast to desktop computing, ubiquitous computing can occur using any device, in any location, and in any format."²

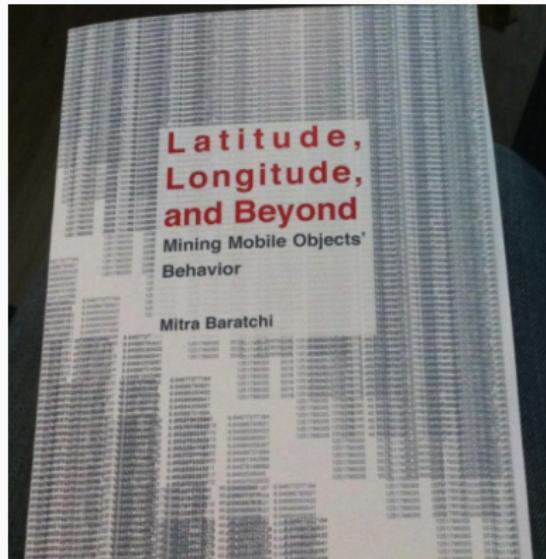


3

²source: <https://en.wikipedia.org/wiki/Ubiquitouscomputing>

³image source: <https://www.parc.com/wp-content/uploads/2010/03/ubicompvenn.jpg>

Ubiquitous Computing research with the focus on mobility data



Back to Urban Computing

Urban Computing is a process of acquisition, integration, and analysis of big and heterogeneous data generated by diverse sources in urban spaces, such as sensors, devices, vehicles, buildings, and humans, to tackle the major issues that cities face (e.g., air pollution, increased energy consumption, and traffic congestion). Urban computing connects unobtrusive and ubiquitous sensing technologies, advanced data management and analytic models, and novel visualization methods to create win-win-win solutions that improve urban environment, human life quality, and city operation systems.⁴

⁴Yu Zheng et al. "Urban computing: concepts, methodologies, and applications". In: *ACM TIST* 5.3 (2014), p. 38.

Table of content

1. Practical matters

2. Introduction to the course

- What is urban computing?

- Applications

- Data sources

3. Lessons learned

Mention some urban computing applications

Applications



5

Figure 2: Traffic management

⁵ source: <https://www.autoevolution.com/news/the-longest-traffic-jam-in-history-12-days-62-mile-long-47237.html>

Applications

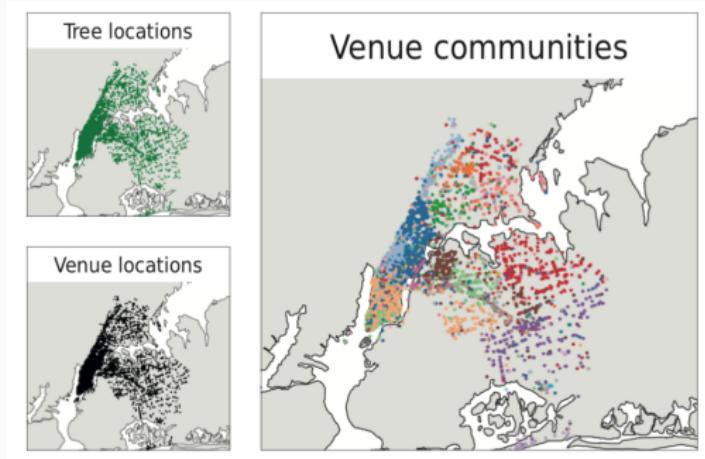


Figure 3: Urban planning

6

⁶source: (JH van Staalanduin et al. "An Intelligent Tree Planning Approach Using Location-based Social Networks Data". In: *Foursquare, future cities challenge, NetMob 2019* [])

Applications



Figure 4: Event management

Applications

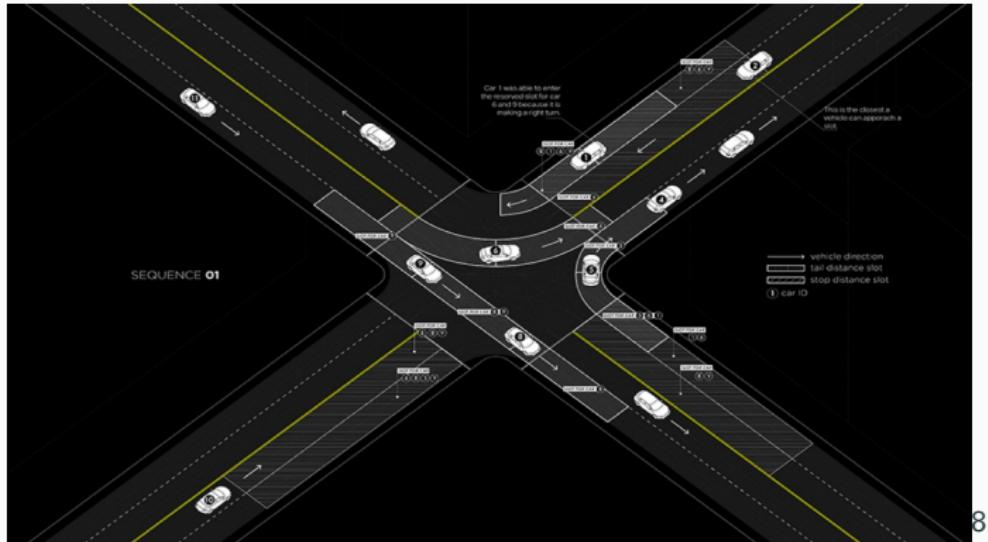
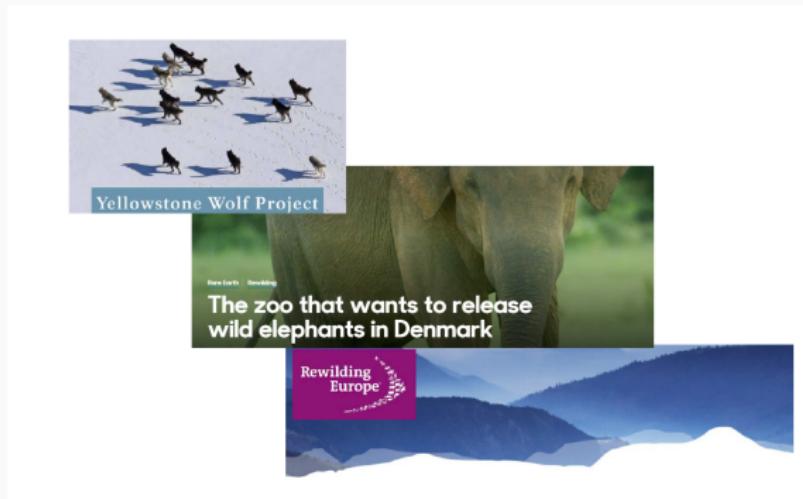


Figure 5: Autonomous driving

⁸ source: <http://senseable.mit.edu/light-traffic/>

Applications



9

Figure 6: Rewilding

⁹ source:<https://www.ark.eu/gebieden/buitenland/rewilding-europe>

Example of rewilding in the Netherlands

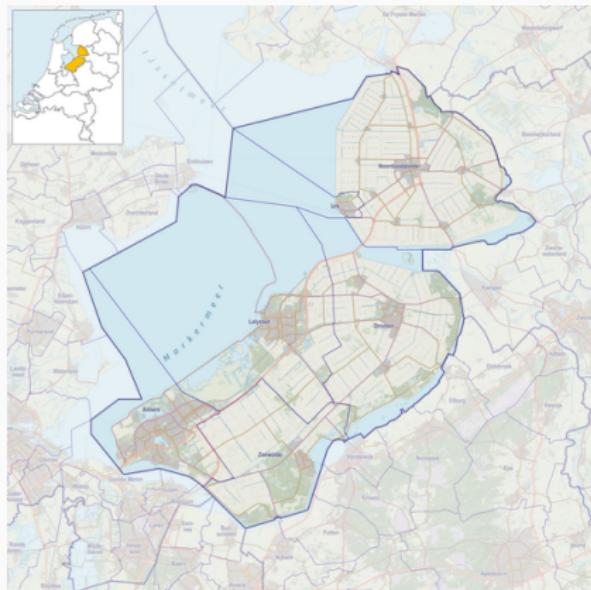
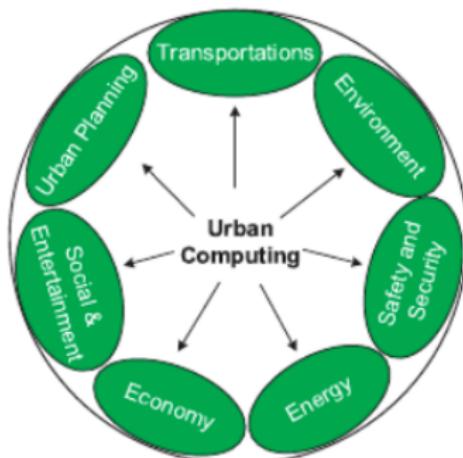


Figure 7: Rewilding

Main categories of applications in Urban Computing



10

¹⁰image source: (Yu Zheng et al. "Urban computing: concepts, methodologies, and applications". In: *ACM TIST* 5.3 [2014], p. 38)

Topics

- Data sources for Urban Computing research
- Processing time-series data
- Processing spatial data
- Processing spatio-temporal data
- Visualization techniques for Urban Computing research
- Machine learning algorithms for Urban Computing research
- Deep learning for Urban Computing research

Why Urban Computing as a new field?

Thinking about urban problems is not new, people have collected data to solve these problems since a long time ago....

Table of content

1. Practical matters

2. Introduction to the course

- What is urban computing?
- Applications
- Data sources

3. Lessons learned

Data used for solving urban problems¹¹

- Old data sources
- Modern data sources

¹¹Dani Arribas-Bel. *Geographic Data Science'17*. 2018. DOI: 10.5281/zenodo.1135210. URL: <http://darribas.org/gds17> (visited on 01/04/2018).

Old data sources

- Calling people on phones for calculating origin destination matrices (traffic engineering)
- Questionnaires
- Census
- Observations by social scientists

Modern data sources¹²

Modern data sources are categorized into the following three categories based on the origin of the data:

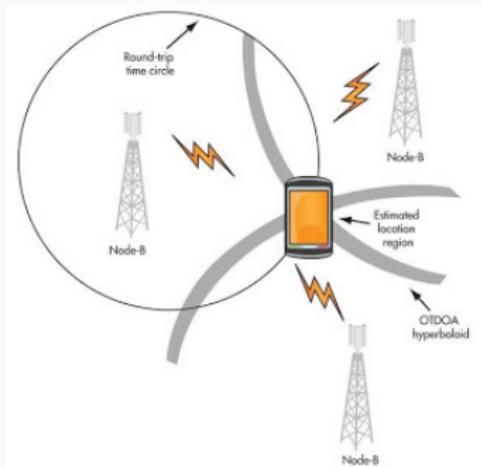
- **Bottom up:** Citizens
- **Intermediate:** Digital companies
- **Top down:** Government

¹²Daniel Arribas-Bel. "Accidental, open and everywhere: Emerging data sources for the understanding of cities". In: *Applied Geography* 49 (2014). The New Urban World, pp. 45–53; Arribas-Bel, *Geographic Data Science'17*.

Bottom-up: citizens as sensors

- Data collected through sensing phones (in some manner)
- Data generated as a result of using Apps
- Participatory sensing: (communities (or other groups of people) contributing sensory information to form a body of knowledge)

Ways to collect data by localizing phones



13

Figure 8: Sensing movement using cellular networks

¹³source: <http://unbonmotgroundswell.blogspot.com/2013/07/hybrid-location-technologies-gps.html>

Wifi sensing

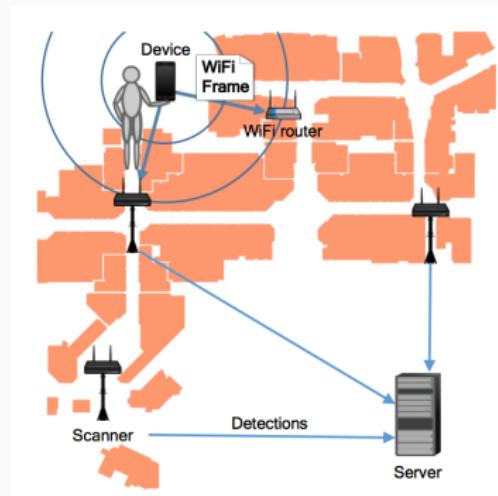


Figure 9: Sensing movement using WiFi networks¹⁴

¹⁴Petre et al., "WiFi Tracking of Pedestrian Behavior, in Smart Sensors Networks".

Wifi sensing, and privacy



15

Figure 10: Wifi sensing

¹⁵source: <https://obj.ca/article/techopia-ottawas-edgewater-wireless-unveils-wi-fi-location-tracking-tech>

Intermediate: digital companies

- Free services provided by companies through Internet
- Data generated as a result of the side activity of a digital company
- Companies that aggregate data from local brokers (e.g., Funda)

https://www.funda.nl

Op vijf minuutjes van de klas
van juf Rianne

Koop Huur Nieuwbouw Recreatie Europa

Plaats, buurt, adres, etc. + 0 km Van € 0 Tot Geen maximum Zoek

Laatste zoekopdracht: Jan Lievensstraat, Leiden >

> Op kaart zoeken > NVM makelaar zoeken

Woning in beeld

Alles over wonen

Figure 11: Funda

Foursquare

The screenshot shows the Foursquare mobile application interface. At the top left is a banner for "FOURSQUARE CITY GUIDE" featuring a photo of a meal. Below it is a search bar with the placeholder "What are you looking for?". To the right of the search bar are six category icons: Breakfast (eggs), Lunch (sandwich), Dinner (steak), Coffee & Tea (cup), Nightlife (drinks), and Things to do (person). On the right side of the screen, there is a vertical sidebar titled "Lists" with sections for "Featured" and "Discover more". A large banner for "FOURSQUARE TRENDING THIS WEEK" is displayed, with a sub-section for "Trending This Week: New York City" showing a map of Manhattan and a list of trending places. Below this are two sections: "My Saved Places" (435 places) and "My Liked Places" (214 places), each with a grid of food-related images.

← Rubirosa

FOURSQUARE CITY GUIDE

Save Rate Leave a tip Check-in

Based on 2,078 ratings

9.6

Rubirosa

235 Mulberry St (btwn Prince & Spring St)
\$\$\$\$ - Open until 11:00 PM LOWER MANHATTAN
Pizza Place, Italian Restaurant
(212) 965-0500

View Menu Directions

You, Emily, and 83 others have been to this place

My Saved Places 435 places

My Liked Places 214 places

Figure 12: Foursquare

A dataset search engine (to be tested)

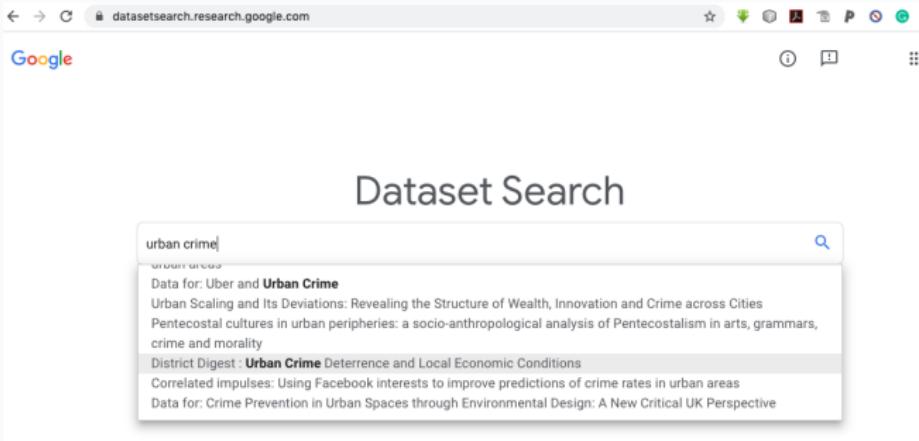


Figure 13: Google dataset search

Top-down: government (Open Data)

- Open data is the idea that some data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control
- Government institutions release (part of) their internal data in open format.

Dutch open data portal

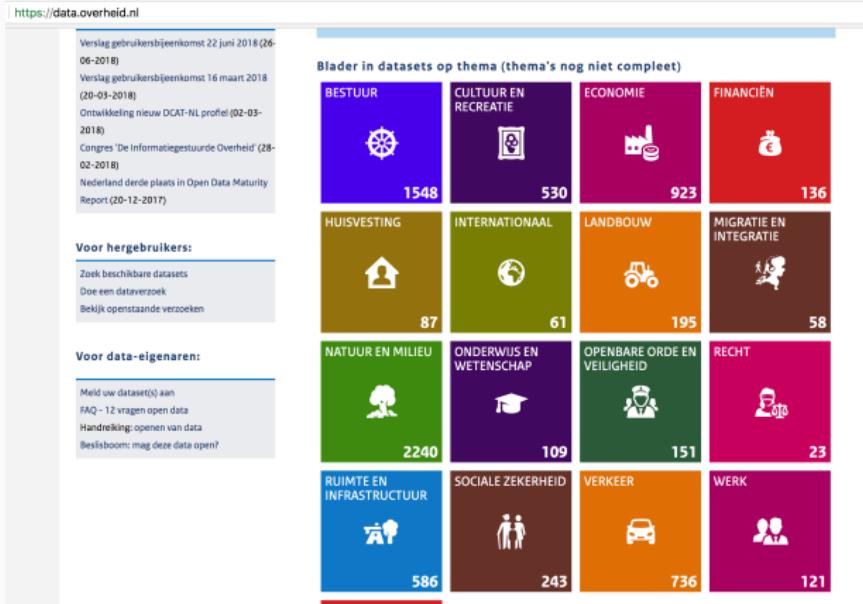


Figure 14: Dutch open data portal.

Check also [CBS data portal](#)

Earth observation data

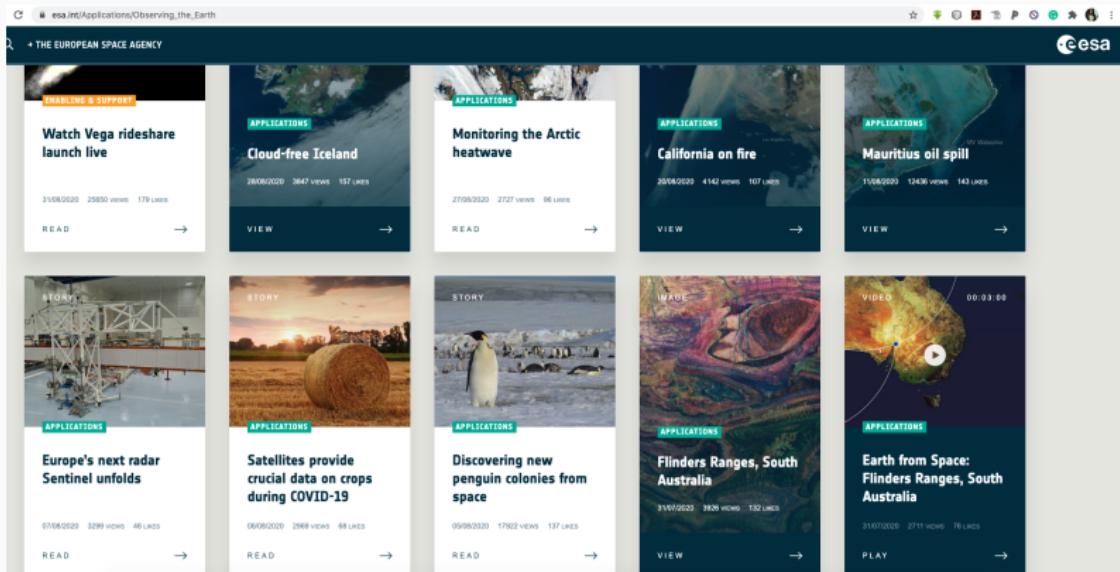


Figure 15: ESA

Check also Google Earth Engine

Old and modern sources (comparison)

Where does each approach for addressing urban problems start from?

- Collecting data for a specific research question (old)
- Finding a research question based on available data (new)

Machine learning/data mining versus statistics

Current approaches to spatio-temporal data handling:

- Statisticians approach
- Machine learning approach

Statisticians approach

Statistics is a branch of mathematics dealing with **data collection, organization, analysis, interpretation and presentation of data** [Wikipedia]

- Hypothesis testing
- T-test
- Permutation test
- ...

Machine learning/data mining approach

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to **effectively perform a specific task** without using explicit instructions, relying on models and inference instead [Wikipedia]

- Design algorithms
- Measure the performance of the algorithm to baselines
 - Classification, clustering accuracy
 - Error metrics

Let's see an example study...

Case: You are designing an algorithm to find periodic patterns from people's trajectory data..

- A periodic activity is an activity that repeats with a specific temporal period
 - Sleeping (period 24)
 - Going to the gym (every 3 days)
 - Going to market (every 7 days)
- Where to get data for this task?
- How to validate that your algorithm is effectively performing this task?

Where to get data?

- Collect data by deploying some sensing technology
 - GPS trackers, Wifi scanning, proximity sensing
- Search for an alternative solution, collect data from a different source
 - Available datasets on the web, use APIs

How to validate your algorithm?

- **The recurring issues of ground-truth**
- Let's see a number of approaches in dealing with these issues

Dealing with ground truth issues

- **Solution 1:**
 - Ask data collectors to label their data (e.g., Lausanne Data Collection Campaign)

Dealing with ground truth issues (real examples)

- **Solution 2:**

- Ask other people to label your data (e.g., Mechanical Turk, Amazon SageMaker Ground Truth)

Dealing with ground truth issues (real examples)

- **Solution 3:**
 - Validation through defining a higher level machine learning task (e.g., does finding periodic patterns help me predict the future of trajectories better?)

Dealing with ground truth issues (real examples)

- **Solution 4:**

- Validation using additional data which is considered highly correlated with the pattern you are looking for (people who work 5 days a week have a strong periodic pattern. Can we distinguish them better from people who do not work?)

Dealing with ground truth issues (real examples)

- **Solution 5:**
 - Synthetic data generator (e.g., data simulated based on known patterns)
 - Make synthetic data as close as possible to actual data (add noise, missing data, random patterns,...)
 - Mess with data in all possible ways to make sure your algorithm works all the time

Lessons learned

Lessons learned

1. Urban Computing field was born due to the adoption ubiquitous and pervasive systems
2. Urban Computing concerns the design of computational methods for analysis of new data sources for old urban problems
3. Old and new data sources
 - Old → questionnaire, census, survey, ...
 - New → citizen generated data, open data, digital businesses data, ...
4. Adoption of machine learning methods for new data sources requires careful experiment design and identifying validation approaches (ground truth, labeling, connecting data sources, synthetic data)