

Exploring data science in sustainability and DIY projects

Liubov Tupikina, LPI, France

*In collaboration with Muki Haklay, Professeur et
Coordinateur de projet, UCL, London , UK
Cléa Montanari, Project manager au sein du projet ECS
Zahra Farook, Doctorante*



Université
Paris Cité



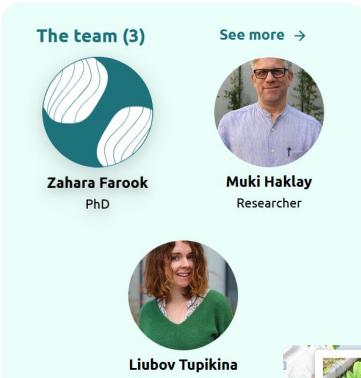
ecs
academy

Maker's Asylum, 2024

Research projects (Liubov Tupikina, LPI)

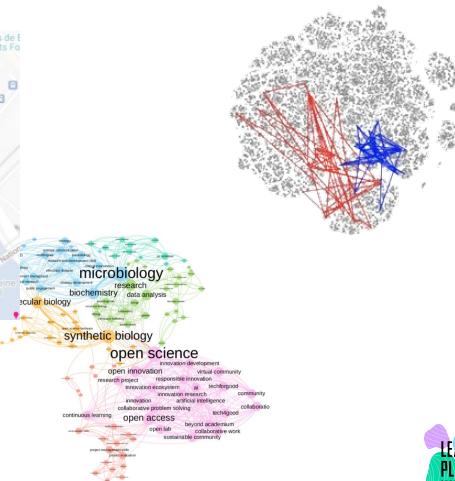
Citizen science

[Link to citizen science project](#)



Innovation research

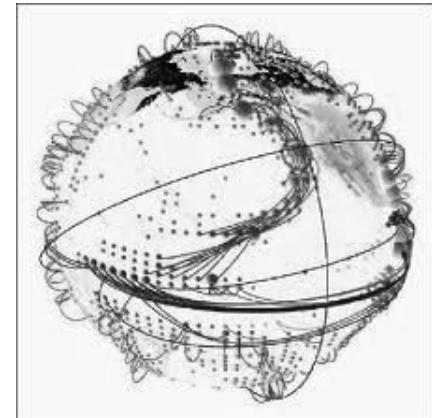
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0293289>



Teaching, participatory learning

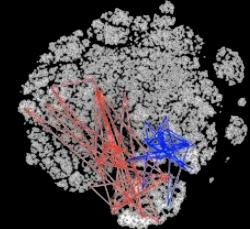
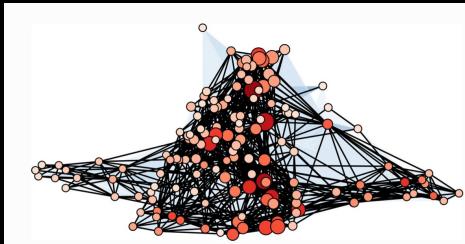
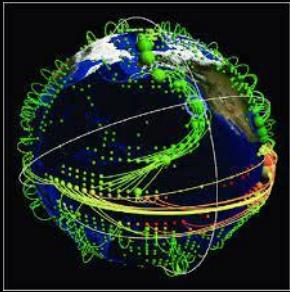
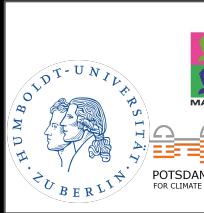
Course on digital data, AI, citizen science, sustainability

<https://github.com/Big-data-a-course-CRI>



fondation
BOTNAR

 Université
Paris Cité



Sustainability
research

Network
science

Data science,
science impact
on society

Innovation
research

Background

Innovation research:

How far is it from DIY medical devices to building satellites?

Quantifying the rise and fall of scientific fields

Chakresh Kumar Singh, Emma Barrie, Robert Ward, Liubov Tupikina, Marc Santolini 

Published: June 23, 2022 • <https://doi.org/10.1371/journal.pone.0270131>

Physics > Physics and Society

[Submitted on 25 Feb 2023]

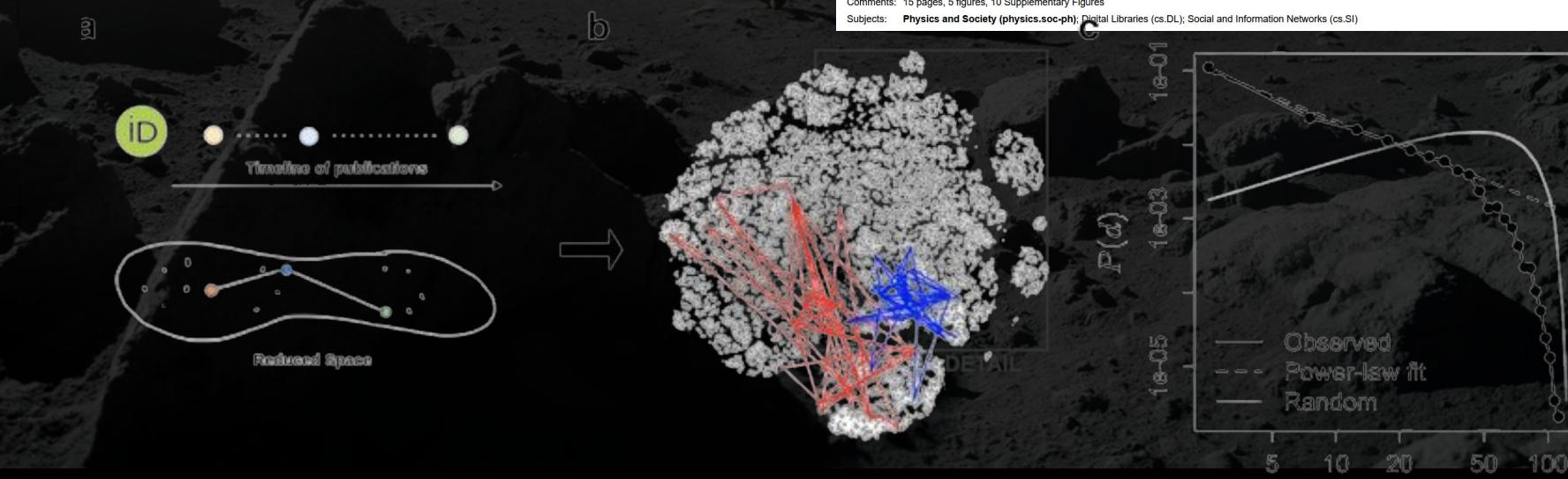
Charting mobility patterns in the scientific knowledge landscape

Chakresh Kumar Singh, Liubov Tupikina, Fabrice Lécuyer, Michele Starnini, Marc Santolini

From small steps to great leaps, metaphors of spatial mobility abound to describe discovery processes. Here, we ground these ideas in formal terms by systematically studying scientific knowledge mobility patterns. We use low-dimensional embedding techniques to create a knowledge space made up of 1.5 million articles from the fields of physics, computer science, and mathematics. By analyzing the publication histories of individual researchers, we discover patterns of knowledge mobility that closely resemble physical mobility. In aggregate, the trajectories form mobility flows that can be described by a gravity model, with jumps more likely to occur in areas of high density and less likely to occur over longer distances. We identify two types of researchers from their individual mobility patterns: interdisciplinary explorers who pioneer new fields, and exploiters who are more likely to stay within their specific areas of expertise. Our results suggest that spatial mobility analysis is a valuable tool for understanding knowledge evolution.

Comments: 15 pages, 5 figures, 10 Supplementary Figures

Subjects: Physics and Society (physics.soc-ph); Digital Libraries (cs.DL); Social and Information Networks (cs.SI)



Your background

DIY

Computer science

Data science

Citizen science?

11AM - 12PM Data science introduction

Notebook example

Break

12PM - 1PM Citizen science, DIY projects data

Project example

Working with spatial data

Materials you may need to use for data analysis

https://github.com/Liyubov/materials_big_data_cri_2019

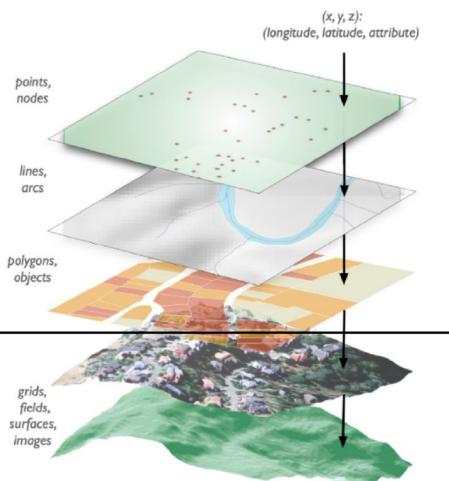
Spatial data

<https://github.com/mszell/geospatialdatascience>

Image Maggi Kelly

Vector: Geometric objects
.shp, .svg

Raster: Grid of pixels
.tif, .jpg, .png, .bmp



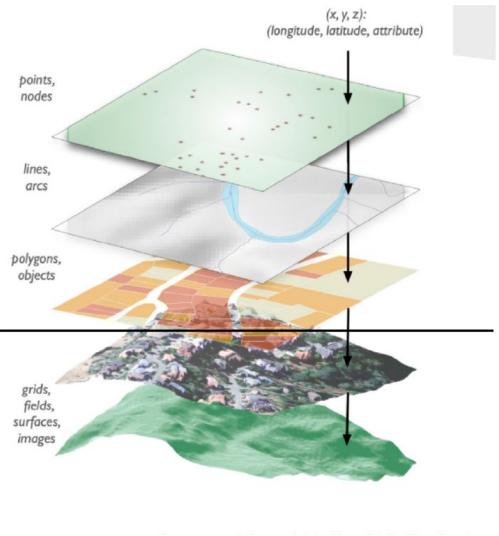
Working with spatial data

Shapefiles main objects

Shapefiles - they are usually groups of files containing geospatial vector data, with the same name and extensions like .shp, .shx, .dbf. Here we only care about the .shp file which contains the geometric features.

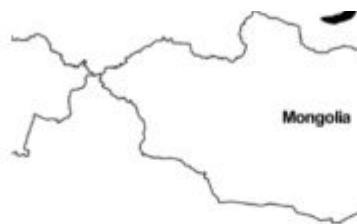
Vector: Geometric objects
.shp, .svg

Raster: Grid of pixels
.tif, .jpg, .png, .bmp



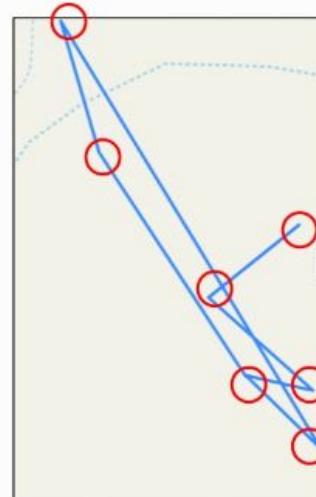
Examples of spatial data

Typhoons



Movement of one User during an Earthquake

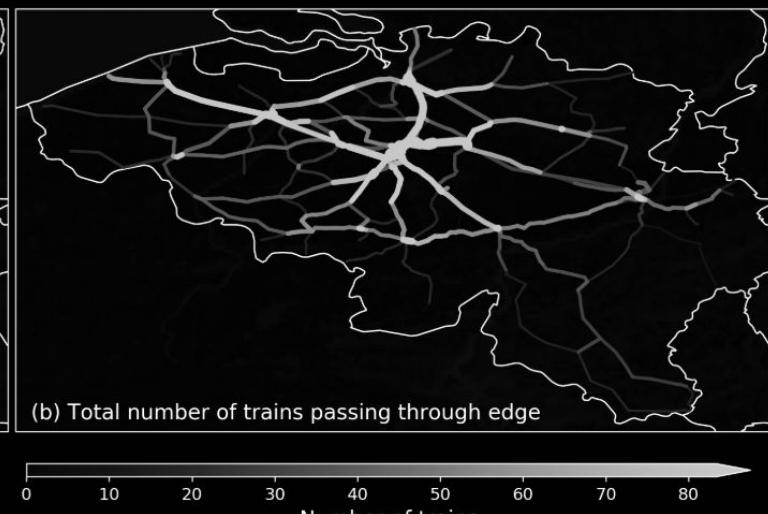
Earthquake
in Iquique
Chile



GeoJson File

```
[-70.144336, -17.975391, 0, 1395904540],  
[-70.146159, -17.976769, 0, 1395996037],  
[-70.144062, -17.978553, 0, 1396482110],  
[-70.145409, -17.978258, 0, 1396653147],  
[-70.143879, -17.979667, 0, 1396864856],  
[-70.149145, -17.971148, 0, 1397655993],  
[-70.148378, -17.974008, 0, 1397656479],  
[-70.145405, -17.978278, 0, 1395655322],  
[-70.145405, -17.978278, 0, 1395655428]
```





Examples of data for social good projects

Transport in networks Nature Comm. (2020)

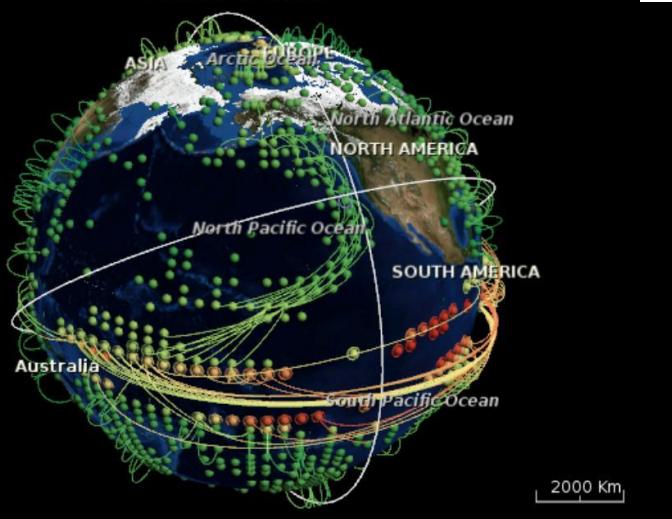
Modelling railway delay

<https://arxiv.org/abs/2105.06111>

C.Lagesse, L.Tupikina, P.Bauman

Shared bike systems

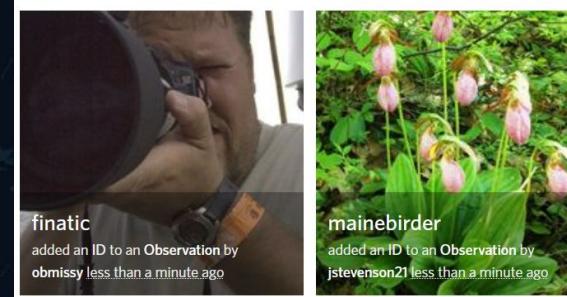
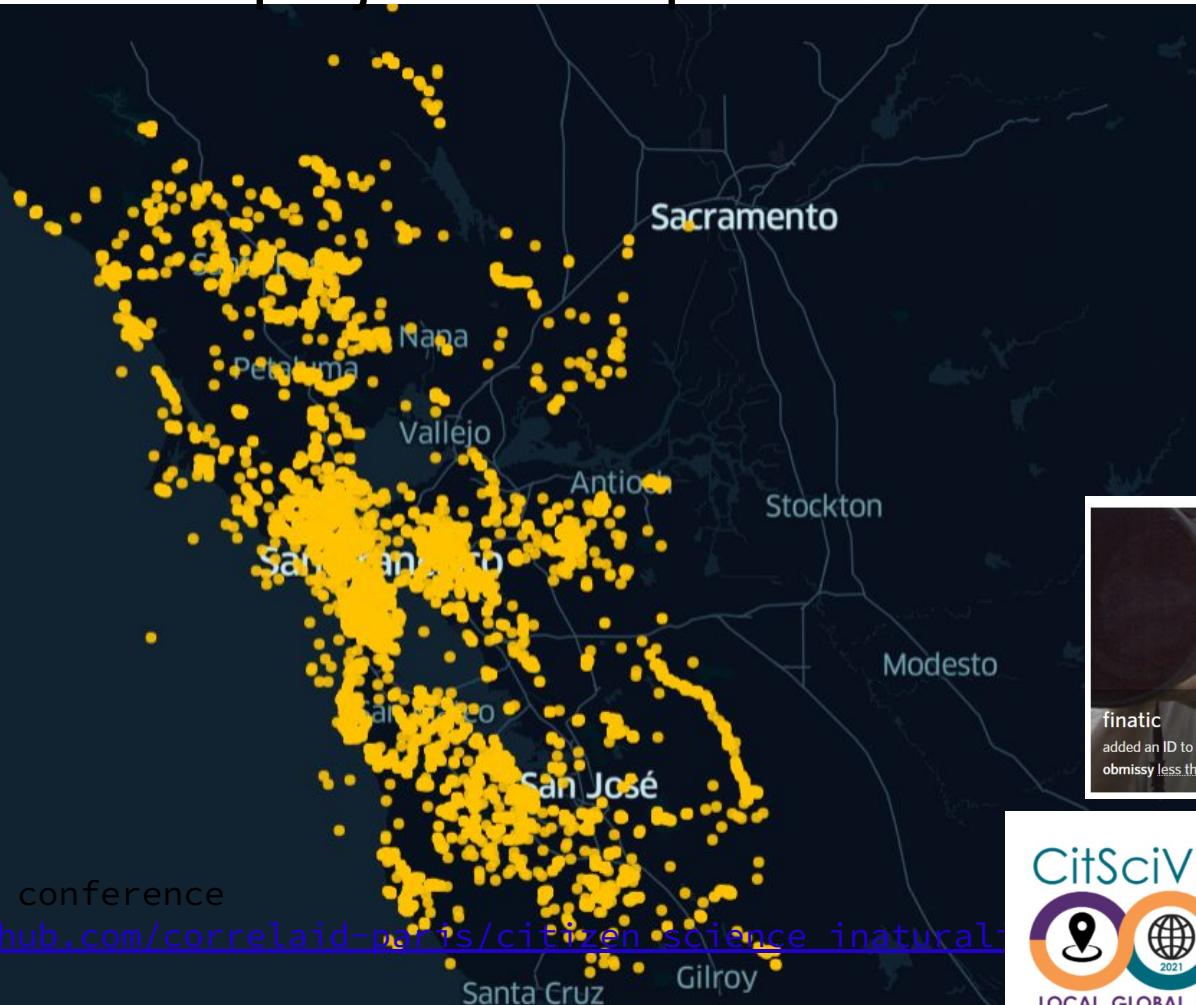
<https://github.com/Liyubov/bike-sharing>



Closed cities ZATOs

<https://theterraforming.strelka.com/>

Citizen science project example



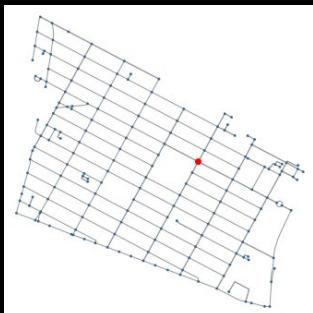
Data, code, conference

https://github.com/correlaid-parts/citizen_science_inaturalist



Coming

Registration Op



<https://github.com/Big-data-course-CRI/materials big data cri 2019/tree/master/resources%20Python>

https://github.com/correlaid-paris/open_inaturalist_code

Outline

11AM - 12PM Data science introduction

Break

12PM - 1PM Citizen science, DIY projects data

A photograph showing a group of people, including children and adults, gathered around a white cloth spread on the ground in a grassy field. They are examining a collection of small, green, leafy plants or seedlings. The scene suggests a field study or a citizen science activity. The image has a slightly grainy, outdoor quality.

What is citizen science?

No PhDs needed: how citizen science is transforming research

Projects that recruit the public are getting more ambitious and diverse, but the field faces some growing pains.

Aisling Irwin



ZOOiverse - 1 mln users volunteers, 2014
Openstreetmaps > 7 mln users 2021

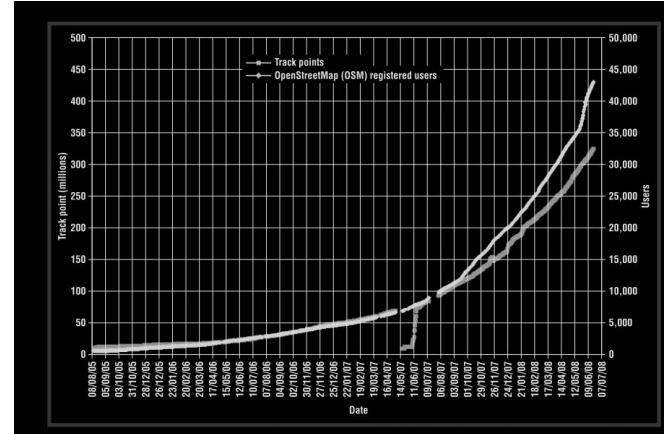


Figure 1. Graph of user and contributions growth to OSM on a monthly basis. The graph shows the accelerating growth in number of users and the rapid increase in data entry measured in track points (source: <http://wiki.openstreetmap.org>).

NEWS FEATURE | 23 October 2018

No PhDs needed: how citizen science is transforming research

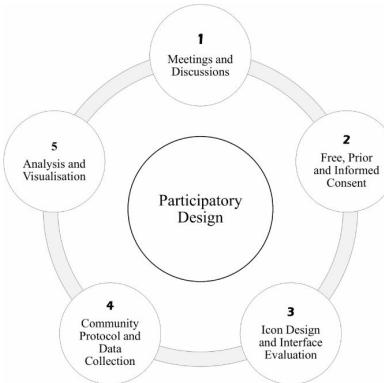
The Science of Citizen Science



cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

OPEN ACCESS

 Springer



party of openstreetmaps

Example of participatory projects

Disaster mapping,
Humanitarian openstreetmaps

Expanding knowledge across partners
and communities

HOT enables communities, NGOs, international organizations, and government partners to use and contribute to OpenStreetMap for locally-relevant challenges through provision of training, equipment, knowledge exchange, and field projects.

Projects to check out

Supporting decentralization in Liberian cities: LEGIT >



Mapping our world together



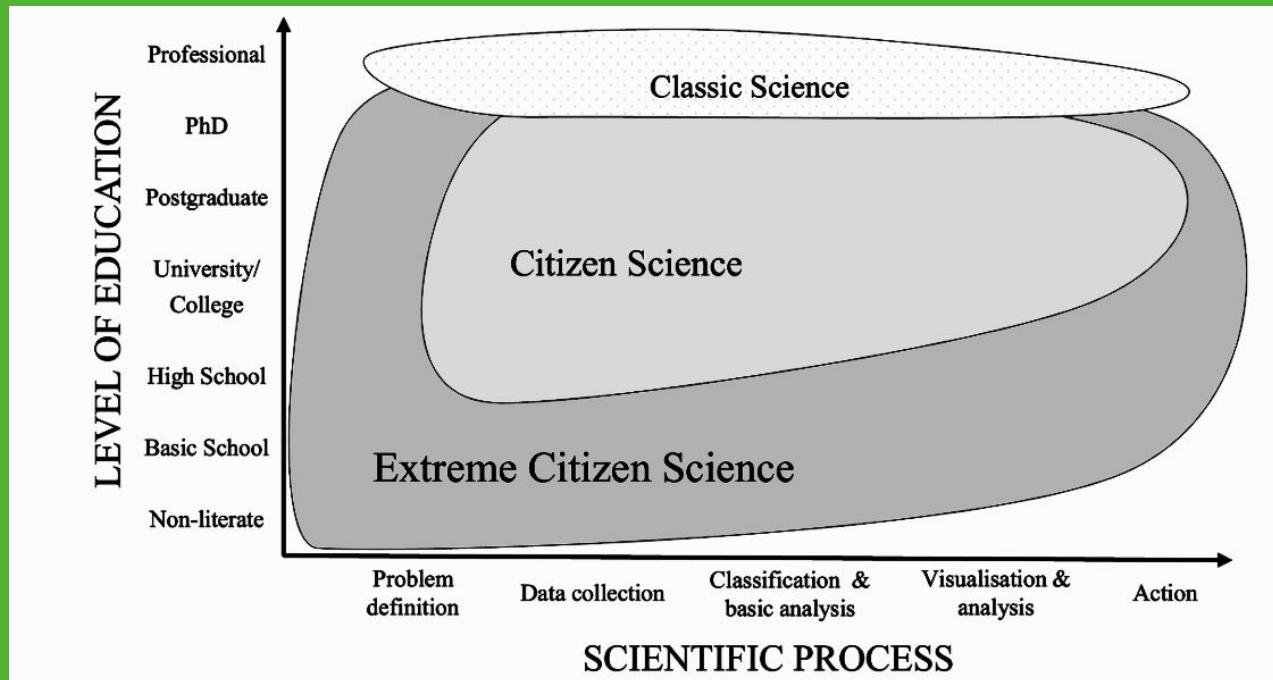
WHAT WE DO OUR WORK TOOLS & DATA NEWS

TOOLS & DATA

What is extreme citizen science?

Involving communities in the loop

[M.Haklay, Sapelli citizen science app]



What is participatory/crowdsourced data?

What are the advantages of the participatory data?

Caractéristiques différentes de la collecte de donnée.

Rarity

- Investment in training
- Maximize the return on each action
- Top-down and standardized methods to guarantee "once and for all" - optimization
- Standard equipment and software

Abundance

- **Assumption of variable skills and training**
- **Ensure micro tasks are enjoyable and rewarding**
- **Multiplicity of procedures and interactions to guarantee commitment**
- **Multiplicity of equipment with limited information on characteristics**

Characteristics of crowdsourced data

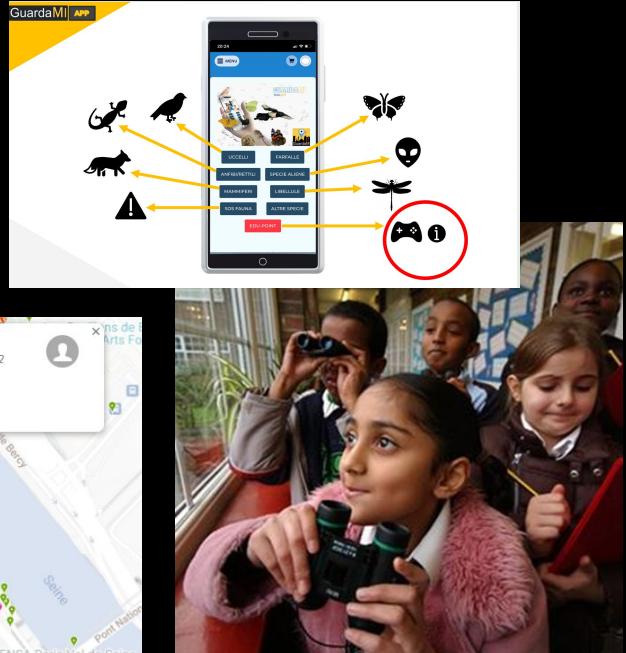
Observation, identifications, photos, etc.

Structured/unstructured

Large or small databases

Feature-rich data (metadata)

Missing data (low data quality)



Comment assurer des données fiables?

Comment Castagnyrol et al., (2021) définissent-ils une une donnée fiable?

“Data reliability concerns the uncertainty in the value of the data provided by the observer.” Know the uncertainty of this data “It is therefore not so much the quality of the data that is critical in the context of PRS as the knowledge that scientists have about this quality, and what they do with it.”

Facteurs qui influencent la qualité des données

- Protocole standardisé
- Volontaire gagnant de l'expérience
- Verifications des données
- Traitement statistique approprié

Castagneyrol, B., Botella, C., & Fontaine, B. (2021). Science citoyenne et qualité des données sur la biodiversité: un faux problème?. NOV'AE.

Other ways to ensure data quality

- Crowdsourcing – the number of people who edited the information
- Social - Controller and moderators
- Geographic - broader geographic knowledge
- Domain knowledge - knowledge about the information domain
- Instrumental observation - Calibration assisted by technology
- Process oriented - follow a procedure, with successive steps

Haklay, M., 2017. Volunteered geographic information, quality assurance. in D. Richardson, N. Castree, M. Goodchild, W. Liu, A. Kobayashi, & R. Marston (eds.) *The International Encyclopedia of Geography: People, the Earth, Environment, and Technology*. Hoboken, NJ: Wiley/AAG

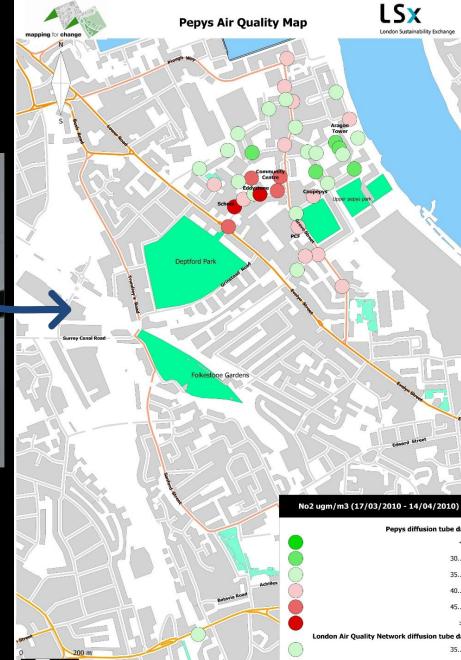
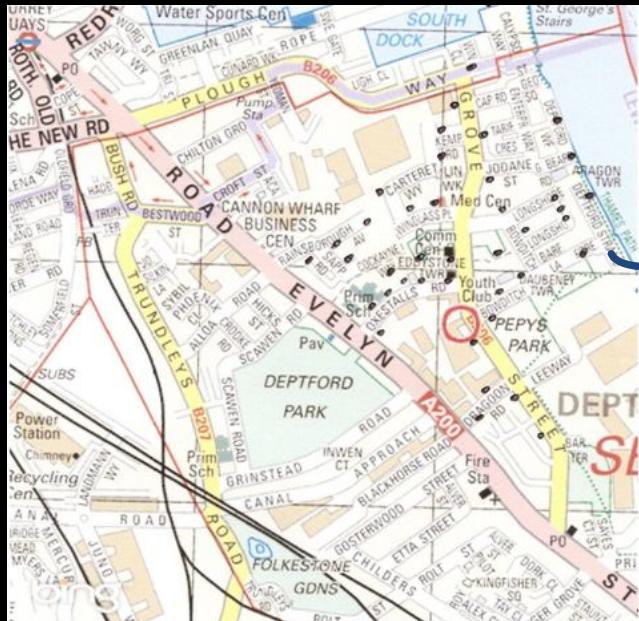
Outline

11AM - 12PM Data science introduction

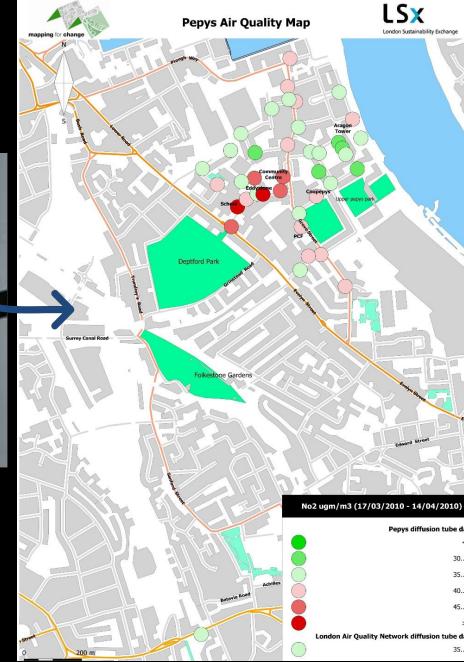
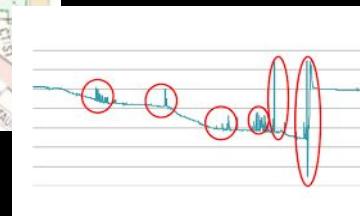
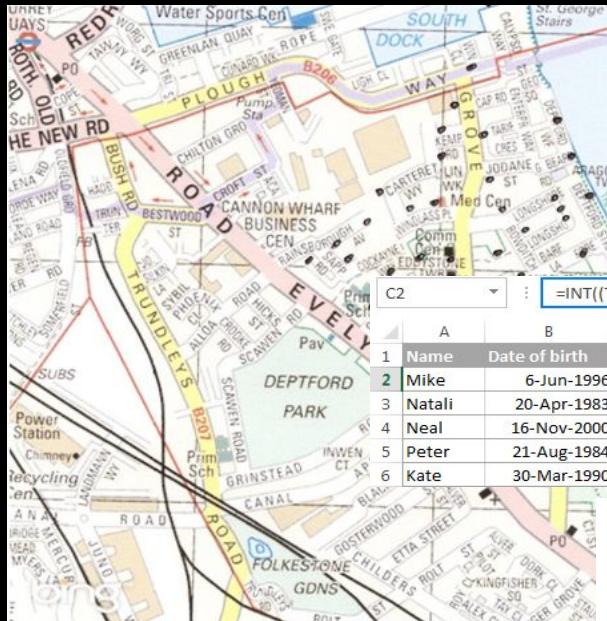
Citizen science, DIY projects data

Project: describe your data you are working with, sources of open data (potentially), crowdsourcing campaigns, describe potential biases in analysis, data (AI blackbox etc.)

Mapping the unmapped, sensors data

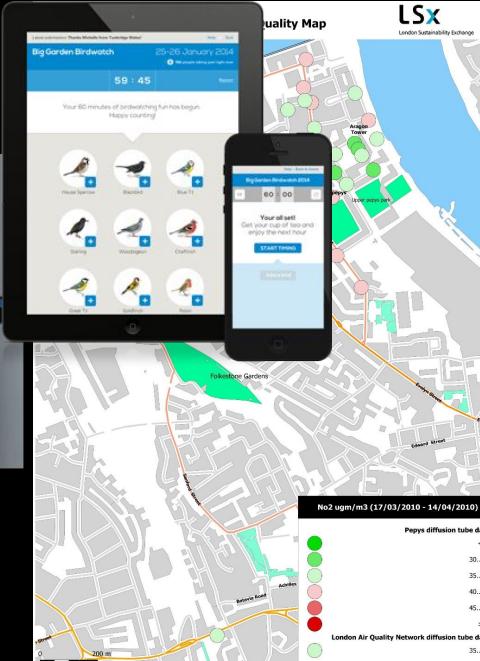
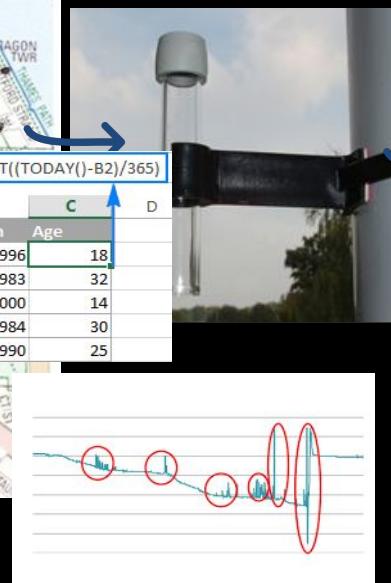
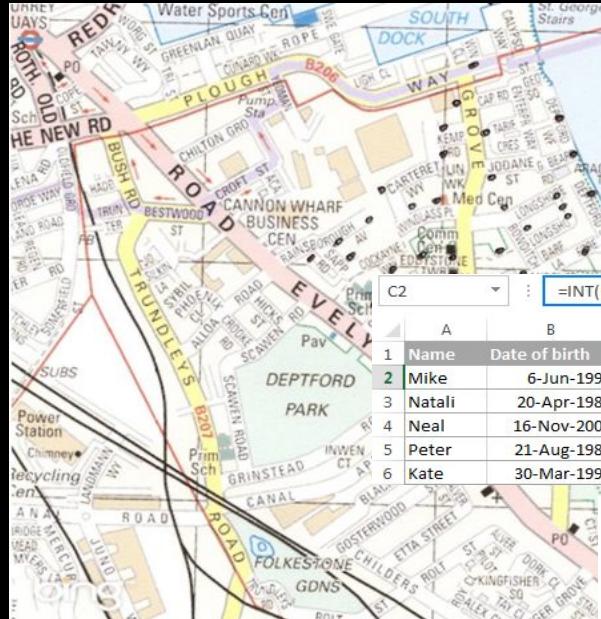


Mapping the unmapped, sensors data



Mapping the unmapped, sensors data

Scarcity of data, abundance





Jim Lemon

Joined: Jan 26, 2017

Last Active: Jan 26, 2024

iNaturalist

Entomologist, Volunteer Naturalist, Retired Systems Person

Following 20 People



dougbeln

176



chrisbehnfeldt

65



lundbergj

2501



bmmrlemon

66



sallypsandpiper

4371



ricknirschl

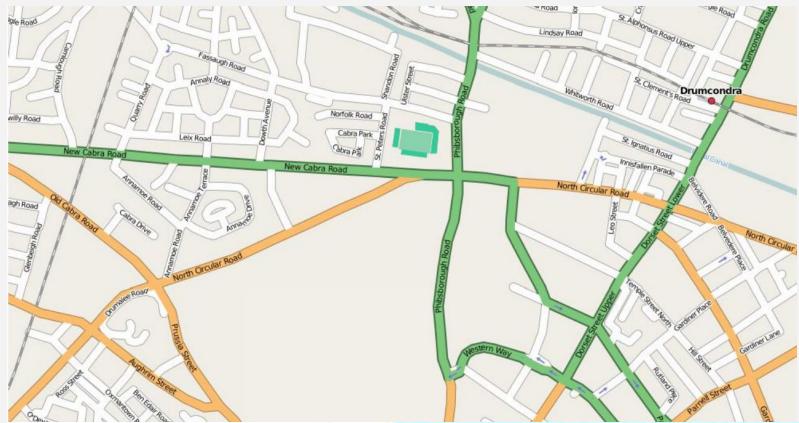
16441

[View All](#)

Main question:

How to work with the data?

1. **Collect the data**
2. **Meta-data:** to know who collected the data, where it comes from?
What is the data?
3. **Processing:** cleaning
4. **Statistical analysis:**
Statistical tests for hypothesis, to visualise some datasets



C2

=INT((TODAY()-B2)/3)

	A	B	C
1	Name	Date of birth	Age
2	Mike	6-Jun-1996	18
3	Natali	20-Apr-1983	32
4	Neal	16-Nov-2000	14
5	Peter	21-Aug-1984	30
6	Kate	30-Mar-1990	25

Main question:

Why citizen science data is important?

"Your map is wrong", say people in India. And it is not because the map was outdated or data is not collected on OSM, it just may have changed by then.

Hence the concept of citizens as observers may be an important one to consider for this case [Goodchild et al.
"Citizens as sensors"]

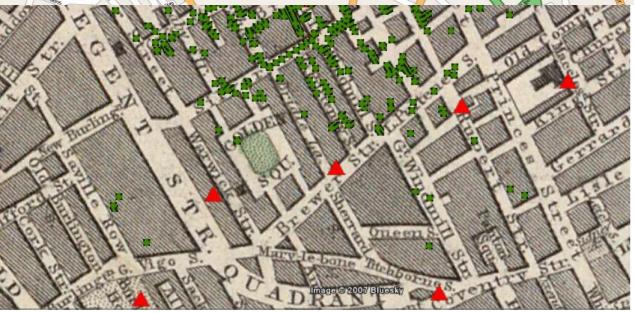
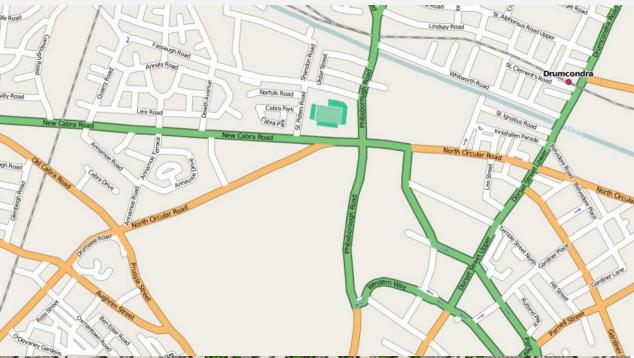
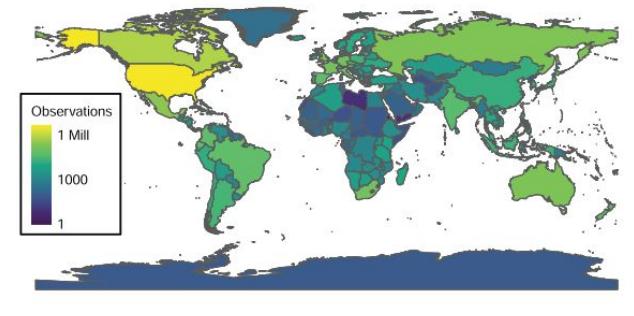


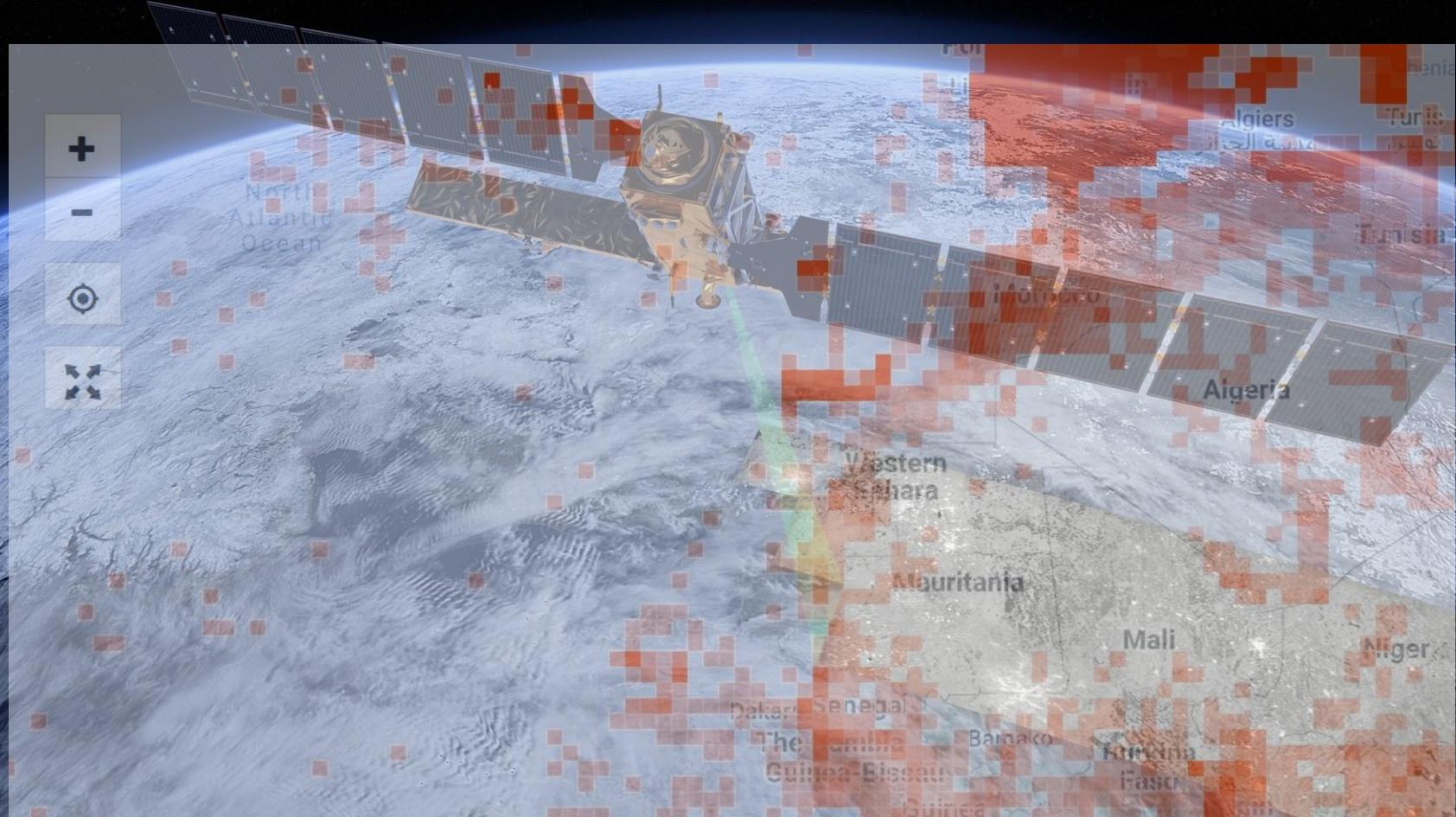
Figure 5. A Google Earth mash-up of the area of Soho, London. The contemporary imagery base has been obscured by an 1843 map from the David Rumsey collection.



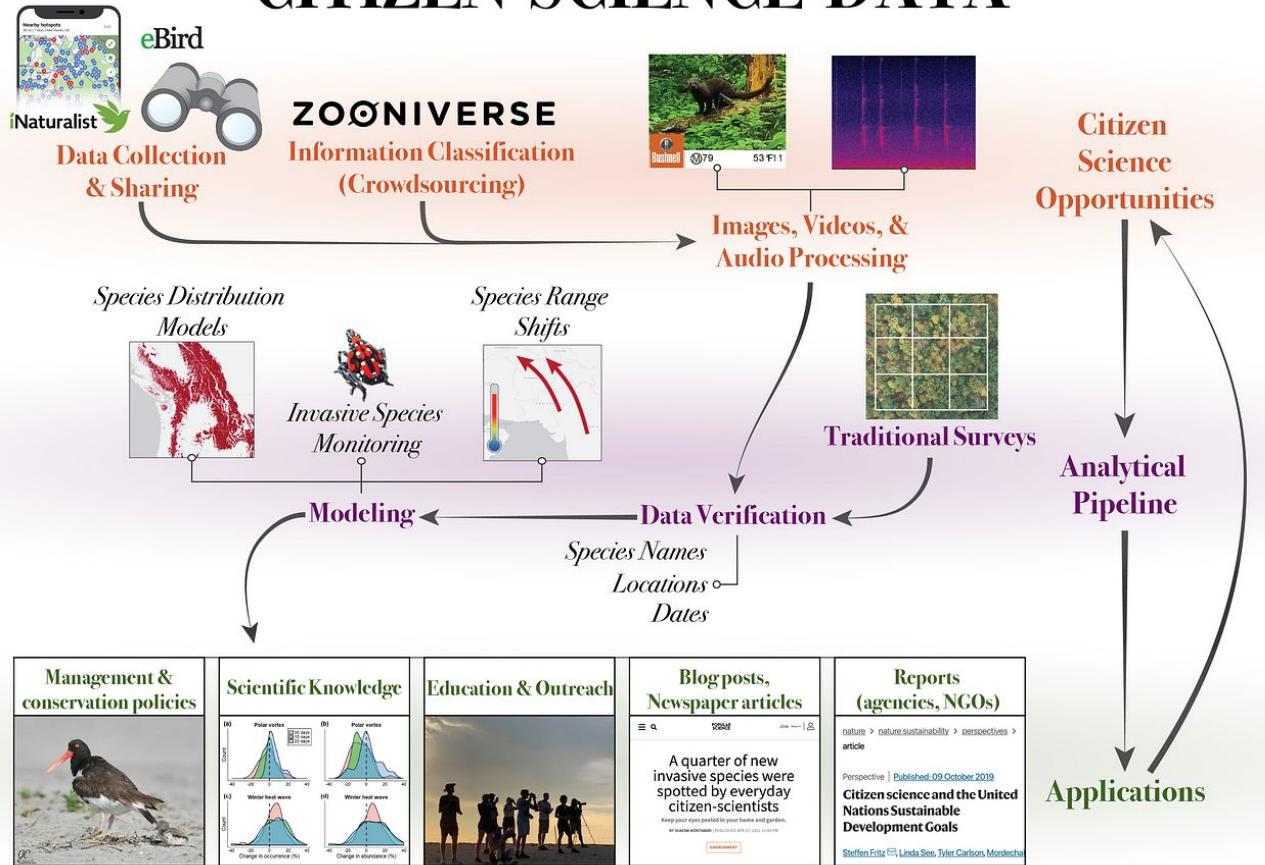
How to combine citizen science data and existing Earth observations?



How to combine citizen science data and existing Earth observations?

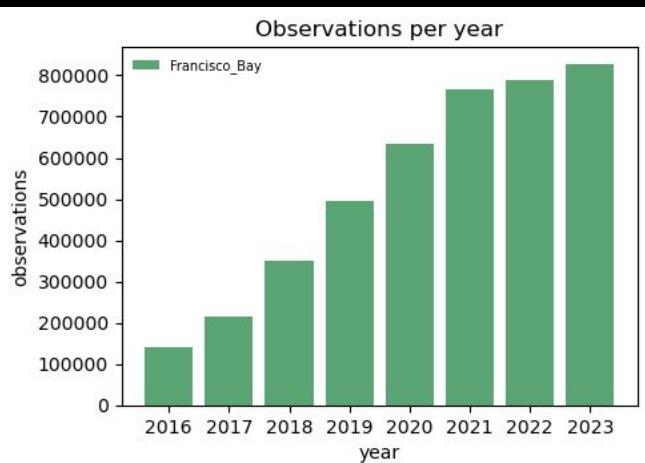


The Life Cycle of CITIZEN SCIENCE DATA



How to enhance quality of crowdsourced data?

- **Environmental evidence-based policies** require a large amount of data, and citizen science is one efficient strategy to achieve this knowledge challenge.



Synthese
DOI 10.1007/s11229-016-1238-2

CrossMark

Crowdsourced science: sociotechnical epistemology in the e-research paradigm

David Watson¹ · Luciano Floridi²

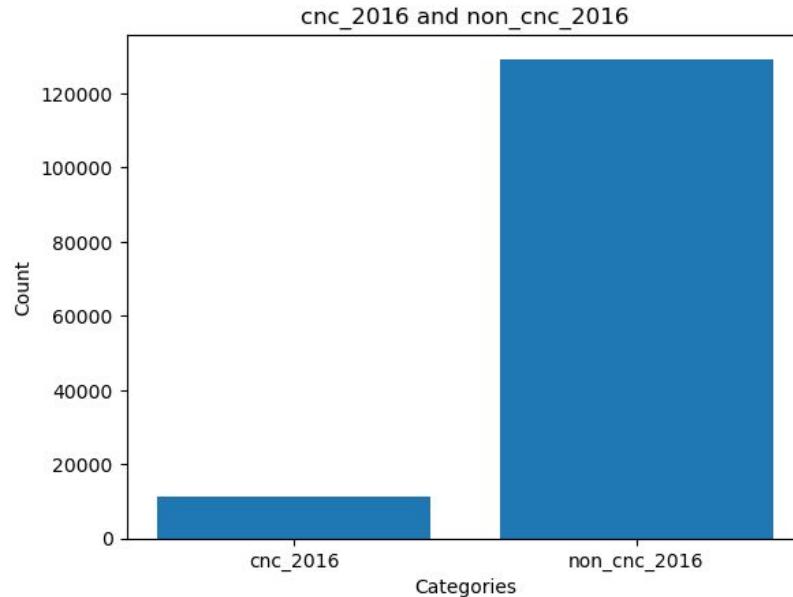
Received: 30 April 2016 / Accepted: 29 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Recent years have seen a surge in online collaboration between experts and amateurs on scientific research. In this article, we analyse the epistemological implications of these crowdsourced projects, with a focus on Zooniverse, the world's largest citizen science web portal. We use quantitative methods to evaluate the platform's success in producing large volumes of observation statements and high impact scientific discoveries relative to more conventional means of data processing. Through empirical evidence, Bayesian reasoning, and conceptual analysis, we show how information and communication technologies enhance the *reliability*, *scalability*, and *connectivity* of crowdsourced e-research, giving online citizen science projects powerful epistemic advantages over more traditional modes of scientific investigation. These results highlight the essential role played by technologically mediated social interaction in contemporary knowledge production. We conclude by calling for an explicitly sociotechnical turn in the philosophy of science that combines insights from statistics and logic to analyse the latest developments in scientific research.

Example 2

- Quel est le profil de contribution pendant la CNC et la non CNC ?
- Classer les observations par catégories (CNC et non CNC)

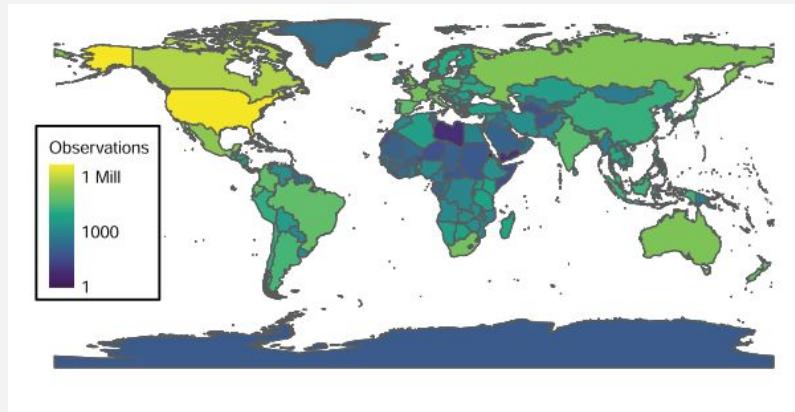
Par exemple, les dates du CNC en 2016 : 14-21 avril 2016.



Main question:

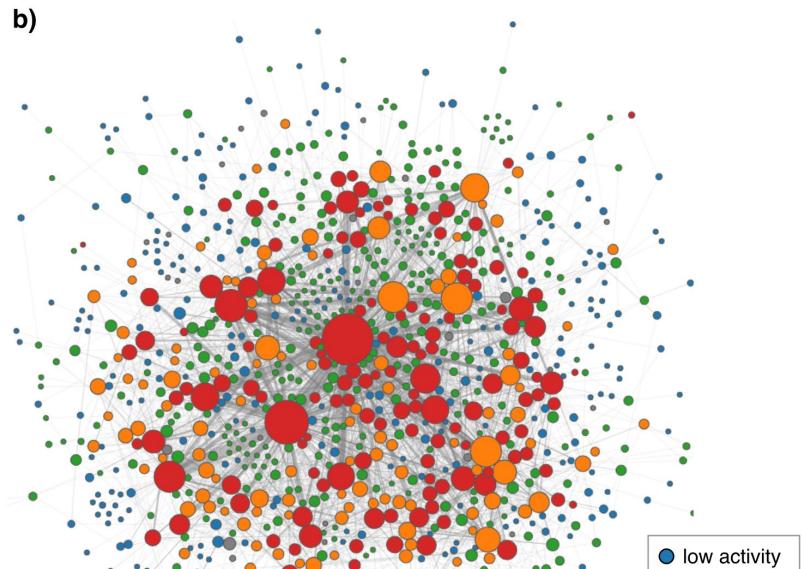
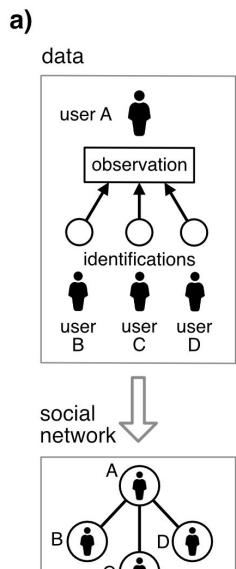
How to use citizen science data?

iNaturalist records have been used to build species distribution models of plants, reptiles (Chapman et al. 2019) etc.



Citizen science community analysis

The research is a continuation of the previous research on “iNaturalist citizen science community during City Nature Challenge: New computational approach to user analysis” carried out in 2021.

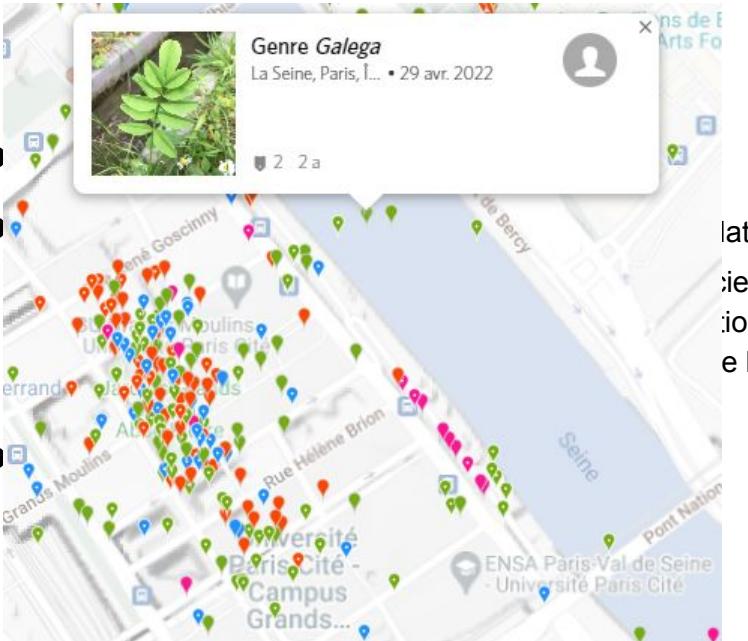


patterns
of
engagement



How to work with crowdsourced data?

Notebook example



In [4]:

```
path = '/Users/zahrafarook/Desktop/DataAnalysis/data'  
dir_list = os.listdir(path)  
print(dir_list)  
  
['downloaded_dataset', 'cnc', '.DS_Store', 'San_Francisco_Bay_2020.  
', 'census_ethnicities', 'San_Francisco_Bay_2023.csv', 'San_Francis  
Bay_2019.csv', 'San_Francisco_Bay_2018.csv', 'non_cnc', 'San_Franc  
, 'San_Francisco_Bay_2017.csv', 'old data']
```

Load Data

In [5]:

```
dfs = defaultdict(dict)  
dfall = []
```

Introduction/ Objective of study of citizen science

How citizen science community evolved across cities: what are the common patterns across the globe, and what are the drivers for CS communities?

- Citizen science communities evolve over time. Using iNaturalist platform, City Nature Challenge as a showcase (CNC) we aim to identify common patterns and drivers that shape communities
- Understand and compare citizen science communities in different years based on the City Nature Challenge
- Analyze patterns of engagement of users in London, San Francisco (SF), and Los Angeles (LA), with a focus on how user behaviors evolve over time (2018-2020)
- We look into the interaction network representing each community using the network analysis, combining this with statistical data analysis

H: Participants' activity types in cities differs from year to year with main variations in the frequency and types of contributions made by users during the City Nature Challenges.

Data from iNaturalist

Data : <https://www.inaturalist.org/>



Format: CSV and JSON files (we used the already available CSV files)

Size and length: 9 files in total for each city (London, San Francisco and Los Angels) /year (2018,2019 and 2020)

Information: 200,000 + rows , 34 columns which are indexed:

```
Index(['id', 'observed_on_string', 'time_observed_at', 'created_time_zone',
       'created_at', 'updated_at', 'description', 'user_id', 'user_login',
       'quality_grade', 'reviewed_by', 'faves_count',
       'num_identification_agreements', 'num_identification_disagreements',
       'identifications_most_agree', 'identifications_most_disagree',
       'captive', 'place_guess', 'place_ids', 'longitude', 'latitude',
       'positional_accuracy', 'geoprivacy', 'taxon_geoprivacy', 'obscured',
       'species_guess', 'taxon_id', 'taxon_name', 'preferred_common_name',
       'iconic_taxon_name', 'taxon_rank', 'taxon_parent_id', 'taxon_native',
       'taxon_endemic', 'taxon_threatened', 'taxon_search_rank',
       'taxon_observations', 'identifications'],
      dtype='object')
```

Data from other citizen science

Data : zooniverse <https://help.zooniverse.org/next-steps/data-exports/>

Format: CSV and JSON files (we used the already available CSV files)

Notebook_01_Exploratory_Data_Analysis

https://github.com/zahrafarook/participatory_methods_and_citizen_science/blob/main/Notebook_01_Exploratory_Data_Analysis.ipynb

Data from iNaturalist

Data : <https://www.inaturalist.org/>

	<code>id</code>	<code>observed_on_string</code>	<code>observed_on</code>	<code>time_observed_at</code>	<code>time_zone</code>	<code>user_id</code>	<code>user_login</code>	<code>created_at</code>
0	11479007	2018-04-27 7:52:34 am BST	2018-04-27	2018-04-27 07:52:34 UTC	UTC	159021	muki	2018-04-27 07:06:17 UTC
1	11480212	2018-04-27 8:39:41 am BST	2018-04-27	2018-04-27 07:39:41 UTC	London	664459	lucyrobinsonnhm	2018-04-27 08:02:52 UTC
2	11480902	Fri Apr 27 2018 08:34:31 GMT+0100 (GMT+1)	2018-04-27	2018-04-27 06:34:31 UTC	Amsterdam	908315	tess26	2018-04-27 08:42:01 UTC
3	11481097	Fri Apr 27 2018 09:54:35 GMT+0100 (GMT+1)	2018-04-27	2018-04-27 07:54:35 UTC	Amsterdam	796473	bryonycross	2018-04-27 08:54:45 UTC
4	11482247	2018-04-27 8:42:21 am BST	2018-04-27	2018-04-27 07:42:21 UTC	London	664459	lucyrobinsonnhm	2018-04-27 10:09:55 UTC

Activity 1

Download the citizen science data from website : <https://www.inaturalist.org/>

The screenshot shows the iNaturalist NZ website. At the top, there is a navigation bar with the iNaturalist logo, a search icon, and links for 'Explore', 'Community', and 'More'. Below the navigation bar, the main content area has a title 'Download iNaturalist NZ data'. To the left of the main content, there is a sidebar with several blue links: 'Help with iNaturalist NZ', 'Add an observation', 'There's an app for us', 'Here's our pamphlet', and 'Add more info with observation fields'. The main content area contains text explaining that users can do amazing things right on the website but suggests downloading data for scientists. It also includes a section titled 'First, a word about copyright'.

Download iNaturalist NZ data

Help with iNaturalist NZ
Add an observation
There's an app for us
Here's our pamphlet
Add more info with observation fields

You can do some amazing things with iNaturalist NZ data right here on our website. However, if you're someone with data skills, like a scientist, then you'll be able to do a lot more if you can download the data. Here we summarise the many ways to do this.

First, a word about copyright

Step 01

Create an iNaturalist Account if you haven't already.

The screenshot shows the iNaturalist sign-up page. At the top left is the iNaturalist logo with a magnifying glass icon. The top navigation bar includes 'Explore', 'Community', and 'More' dropdowns. A large banner on the right side reads 'Join the largest group of naturalists in the world!'. Below the banner are four input fields: 'Email' (with placeholder 'zahra_farook'), 'Username' (placeholder '.....'), 'Password' (placeholder '.....'), and 'Password confirmation'. There is a 'reCAPTCHA' checkbox labeled 'I'm not a robot'. Below the form are two additional checkboxes: one checked for 'Yes, license my photos, sounds, and observations so scientists can use my data (recommended). Learn More' and another unchecked for 'I consent to allow iNaturalist to store and process limited kinds of personal information'. At the bottom left, there is a small image of a bird and the text 'Eastern Phoebe observed by zahra_farook'.

iNaturalist

Explore Community More

Join the largest group of naturalists in the world!

Email: zahra_farook

.....

.....

Password confirmation

I'm not a robot

reCAPTCHA

Privacy - Terms

Yes, license my photos, sounds, and observations so scientists can use my data (recommended). [Learn More](#)

I consent to allow iNaturalist to store and process limited kinds of personal information

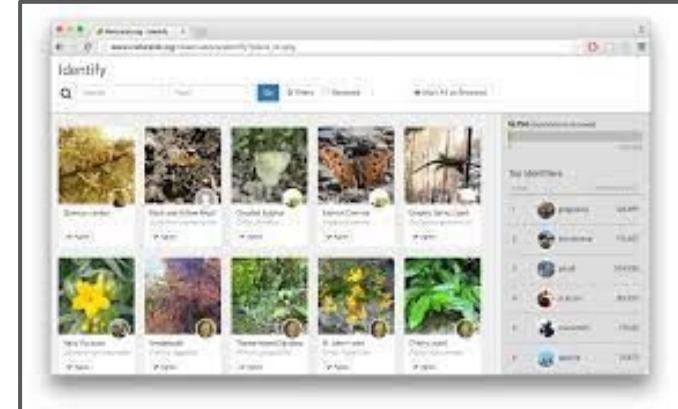
Eastern Phoebe observed by zahra_farook

Step 02

Think about the you want to download from iNaturalist

Examples :

- All observations made in Paris in 2023
- All insects observed in London from year 2015 - 2016
- Observations made in India in July 2020



Step 03

Go to : <https://www.inaturalist.org/observations/export>

The screenshot shows the iNaturalist website interface. At the top, there is a navigation bar with links for 'Search', 'Explore', 'Your Observations', 'Community', 'Identify', 'More', 'Upload' (with a file icon), and user notifications (0 messages, 0 comments). A profile picture is also visible.

The main content area has a yellow background. It features a section titled 'Other Options' which contains a note about large exports slowing down infrastructure and provides links to alternative export methods:

- **GBIF:** GBIF ingests most of our Research Grade, licensed observations, and their export tools are excellent. GBIF exports also come with a [DOI](#) you can cite if you publish research based on the data you export, and these citations are enormously helpful to us as they allow us to see how iNat data gets used.
- **DarwinCore Archive (DwC-A) for GBIF:** this is the [DwC-A](#) file we generate for GBIF to ingest, so it should not contain anything that isn't in GBIF, but if you want the whole archive, you can download it. It is a very large zip file containing records in CSV. Updated weekly.
- **iNaturalist Challenge at FGVC 2017:** links to 675,000 licensed iNaturalist photos of 5,089 species for use in computer vision training. Created June 2017, not updated.

Below this, there is a section titled 'Export Observations' with a sub-section '1 Create a Query'. It includes instructions and a text input field:

Create an observation query just like you would elsewhere on the site. You can also cut and paste an observations URL from another part of the site. Your query should return **no more than 200,000 observations**.

Choose query parameters below or paste in an observations URL

Step 4

Scroll down and “Create a Query” and specify the data you want or paste the URL

Export Observations

1 Create a Query

Create an observation query just like you would elsewhere on the site. You can also cut and paste an observations URL from another part of the site. Your query should return **no more than 200,000 observations**.

Choose query parameters below or paste in an observations URL

Search search all fields

Filter

w/ photos w/ sounds Quality grade Any Research Needs ID Reviewed Any Yes No
Identifications any most agree some agree most disagree Captive / Cultivated Any Yes No

Geoprivacy Taxon Geoprivacy

Show only

Place Start typing place name... SW Lat. SW Lon. NE Lat. NE Lon. clear

Not in place Start typing place name...

Taxon Start typing taxon name... Observed on Day Month Year

Not in taxon Start typing taxon name...

Exact Rank Any Highest Rank Any Lowest Rank Any

Verifiable Any Yes No

Threatened Any Yes No

Introduced Any Yes No

Step 5

To specify the data you want, fill in the menus below

Export Observations

1 Create a Query

Create an observation query just like you would elsewhere on the site. You can also cut and paste an observations URL from another part of the site. Your query should return **no more than 200,000 observations**.

Choose query parameters below or paste in an observations URL

Search search all fields

Filter
 w/ photos w/ sounds Quality grade Any Research Needs ID Reviewed Any Yes No
Identifications any most agree some agree most disagree Captive / Cultivated Any Yes No
Geoprivacy Taxon Geoprivacy

Show only Select All; None

Place Start typing place name... SW Lat. SW Lon. NE Lat. NE Lon. clear

Not in place Start typing place name...

Taxon Observed on

Not in taxon

Exact Rank Highest Rank Lowest Rank

Verifiable Any Yes No

Threatened Any Yes No

Introduced Anv Yes Nn

Step 6

Check the “Preview” to make sure that you are getting the observations

2 Preview		1 - 30 of 79312			
Photos / Sounds	Species / Taxon Name	Observer	Date observed	Place	
	 Genus <i>Xylaria</i>	mcsvt	November 11, 2023 12:22 PM CET	📍 Bois de Vincennes, Paris, Île-de-France, FR (Google, OSM)	1 ID Needs ID View »
	 Something...	mcsvt	November 12, 2023 10:47 AM CET	📍 Bois de Vincennes, Paris, Île-de-France, FR (Google, OSM)	Needs ID View »

Step 7

Choose the columns

Each observation is one row, with several columns of data about.

3 Choose columns

Choose the columns you want to export

Basic (All | None)

<input checked="" type="checkbox"/> id	<input checked="" type="checkbox"/> observed_on_string	<input checked="" type="checkbox"/> observed_on	<input checked="" type="checkbox"/> time_observed_at
<input checked="" type="checkbox"/> time_zone	<input checked="" type="checkbox"/> user_id	<input checked="" type="checkbox"/> user_login	<input checked="" type="checkbox"/> user_name
<input checked="" type="checkbox"/> created_at	<input checked="" type="checkbox"/> updated_at	<input checked="" type="checkbox"/> quality_grade	<input checked="" type="checkbox"/> license
<input checked="" type="checkbox"/> url	<input checked="" type="checkbox"/> image_url	<input checked="" type="checkbox"/> sound_url	<input checked="" type="checkbox"/> tag_list
<input checked="" type="checkbox"/> description	<input checked="" type="checkbox"/> num_identification_agreements	<input checked="" type="checkbox"/> num_identification_disagreements	<input checked="" type="checkbox"/> captive_cultivated
<input checked="" type="checkbox"/> oauth_application_id			

Geo (All | None)

Town, county, state, country fields based on coordinates. Analogous administrative levels will be used for observations where these place types do not apply. The private_* fields will only be included when using the user filter and you are the user, or when using the project filter and you curate the project.

<input checked="" type="checkbox"/> place_guess	<input checked="" type="checkbox"/> latitude	<input checked="" type="checkbox"/> longitude	<input checked="" type="checkbox"/> positional_accuracy
<input checked="" type="checkbox"/> private_place_guess	<input checked="" type="checkbox"/> private_latitude	<input checked="" type="checkbox"/> private_longitude	<input checked="" type="checkbox"/> public_positional_accuracy
<input checked="" type="checkbox"/> geoprivacy	<input checked="" type="checkbox"/> taxon_geoprivacy	<input checked="" type="checkbox"/> coordinates_obsured	<input checked="" type="checkbox"/> positioning_method
<input checked="" type="checkbox"/> positioning_device	<input type="checkbox"/> place_town_name	<input type="checkbox"/> place_county_name	<input type="checkbox"/> place_state_name
<input type="checkbox"/> place_country_name	<input type="checkbox"/> place_admin1_name	<input type="checkbox"/> place_admin2_name	

Step 8

Click “Create Export.”

The screenshot shows a list of taxonomic fields with checkboxes: taxon_suborder_name, taxon_superfamily_name, taxon_family_name, taxon_subfamily_name, taxon_supertribe_name, taxon_tribe_name, taxon_subtribe_name, taxon_genus_name, taxon_genushybrid_name, taxon_species_name, taxon_hybrid_name, taxon_subspecies_name, taxon_variety_name, and taxon_form_name. Below this is a section titled "Observation Fields" with links for "All" and "None". At the bottom, there is a large red box around a blue "Create Export" button. A small number "4" is located in the top left corner of this red box.

The screenshot shows a modal dialog box titled "Exporting...". It contains the text "Please be patient, this can take a few minutes. If this is taking too long, you can check back soon!" and "Receive an email with your data". It also lists several checked checkboxes for export fields: time_observed_at, user_name, license, tag_list, num_identification_agreements, num_identification_disagreements, and captive_cultivated. The status bar at the top indicates the date and location: "January 25, 2024 10:24" and "Paris Ct, Royal Oaks, CA, US (Google)".

You can wait for it to finish or get an email with your data.

Step 9

Once the download is done, “Export Complete” box will appear. Click “Download.” Save the file.

Export complete						
	File size	Queued	Started	Finished	Inputs	
Download	11 KB	03 Jun 15:20	03 Jun 15:21	03 Jun 15:21	Query quality_grade=any&identifications=any&taxon_id=71258&verifiable=true& user_id=sambiology,bouteloua,wdvanhem&d1=2021-01-01 Columns id, observed_on_string, observed_on, time_observed_at, time_zone, user_id, user_login, created_at, updated_at, quality_grade, license, url, image_url, sound_url, tag_list, description, num_identification_agreements, num_identification_disagreements, captive_cultivated, oauth_application_id, place_guess, latitude, longitude, positional_accuracy, private_place_guess, private_latitude, private_longitude, public_positional_accuracy, geoprivacy, taxon_geoprivacy, coordinates_obscured, positioning_method, positioning_device, species_guess, scientific_name, common_name, iconic_taxon_name, taxon_id	Delete



Hooray! You now have iNaturalist data.

Challenges and benefits of iNaturalist data

Data from <https://www.inaturalist.org/>

Well structured: records have structured features

Well related to other datasets

Open API (everyone can download it and analyze it)

For additional analysis of the data one needs to do specific ethical approvals

Activities during the class

- Activity 1 to download iNaturalist data together.
- Presentation of how this data can be analysed
- Saying what are the challenges of working with this set of data compared to traditional data. What are the benefits?
- What are other open access citizen science data sets? Zooniverse, SciStarter
- ~~- Activity 2: types of datasets that can come from citizen science project- another way to understand what is citizen science or not~~
- Activity 3 : set of question of data sets to ask.

Appendix: show-case of further analysis

Lecturers without borders

Educational NGO hosted in Learning Planet Institute in 2018

Mission: connecting traveling scientists with schools, universities

Main nominations: Botnar Foundation, Falling Walls Foundation,

Connections: >400 scientists, Marie-Curie science network, EPSC,

www.lewibo.org



fondation
BOTNAR



**Share travel
plans with us**

Share with us your
travel plans:
approximate time and
place per email, slack



**Make lecture
/seminar**

We connect you to local
coordinator and help
you to come to
school/university



**We find
schools**

We find you local
schools/universities,
to make
lecture/seminar



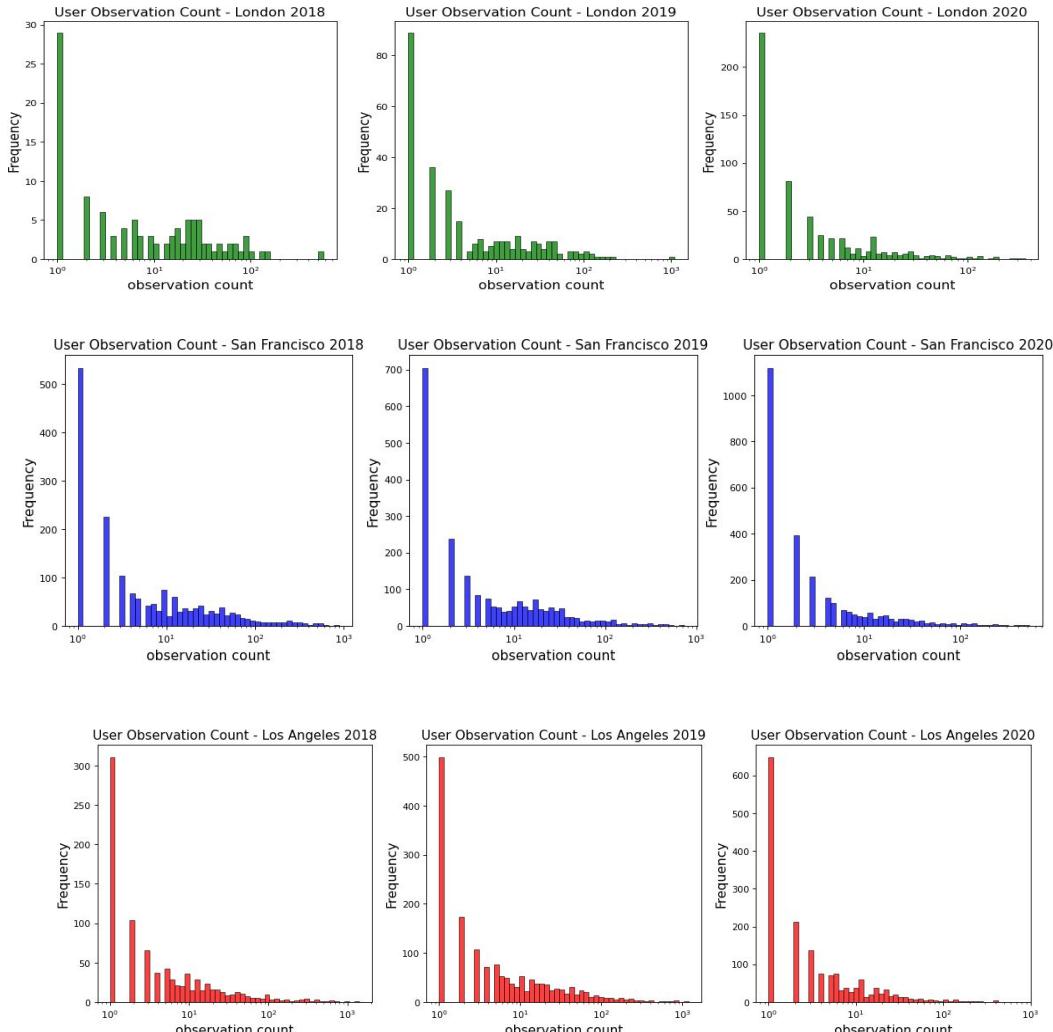
**We connect
you**

We match you with the
local organiser and help
with connecting to
lecture place

Statistical Analysis

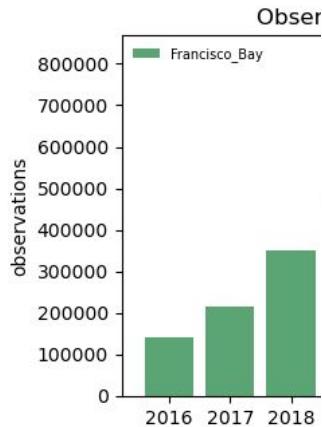
1. General statistics for London, San Francisco, and Los Angeles 2018-2020.

- We conducted a general statistical analysis on contributions across cities.
- The plots capture only the static perspective on data (snapshots per one city one year).
- Overall the observation made by users has drastically reduced from 2018 to 2020.

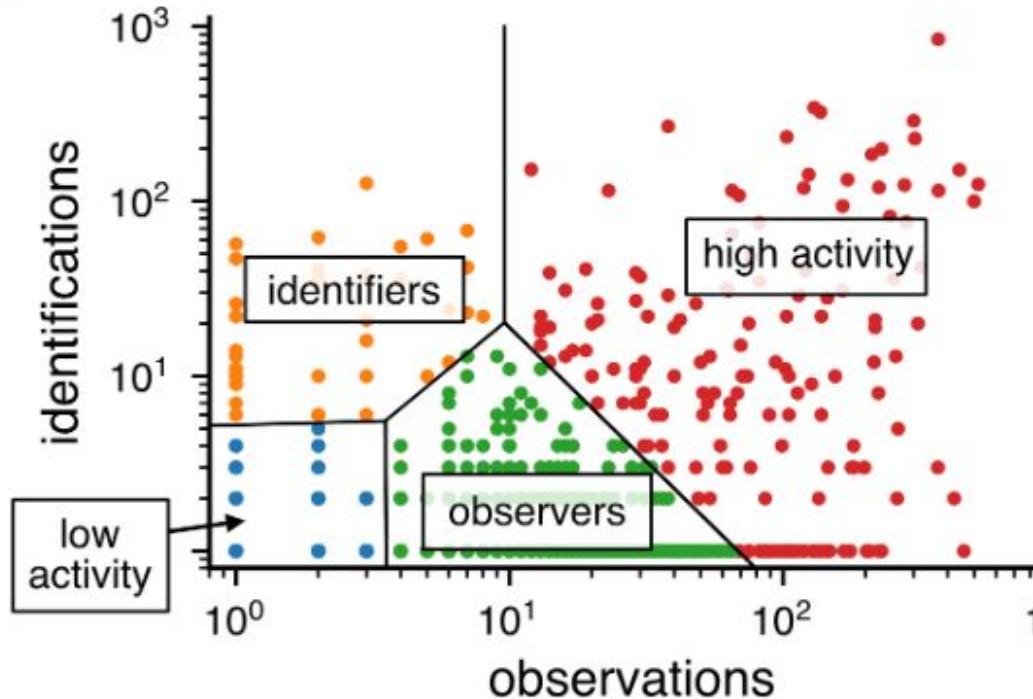


Citizen scier

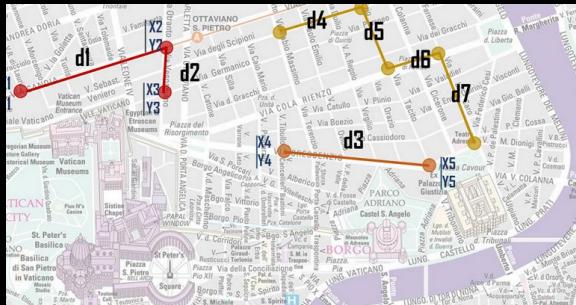
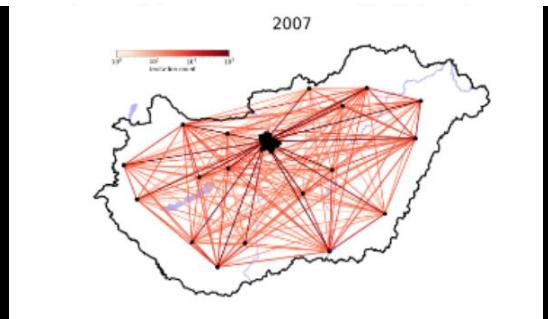
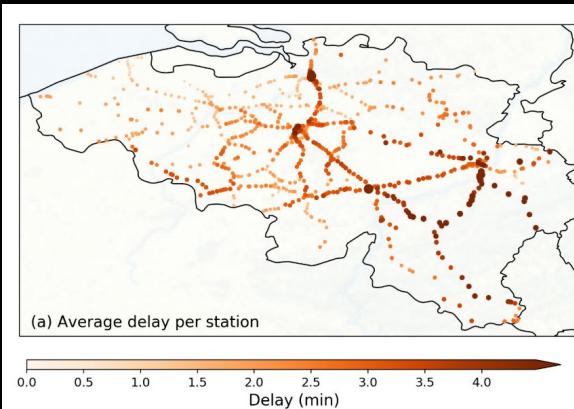
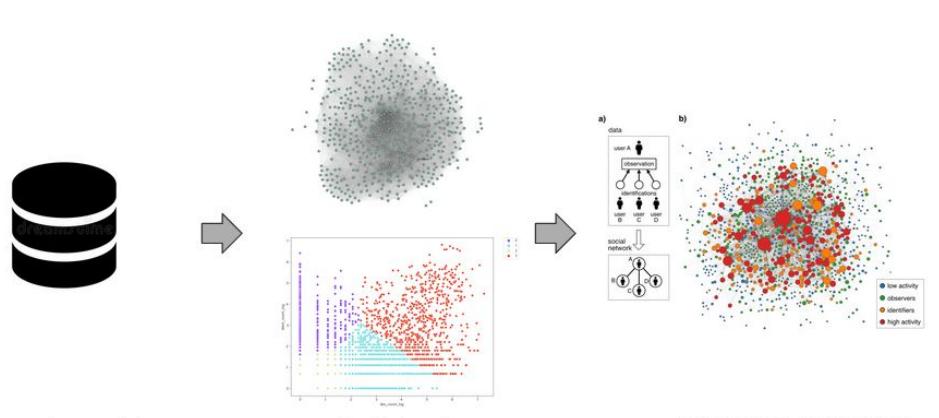
Example of a



a)



Citizens, cities and networks



Projects which can enable online collaboration

www.aino.world



AI for spatial data analysis Save hundreds of hours in data work

The screenshot shows a map of Paris with street networks highlighted in blue. A search bar at the top left says "Show streets typology in Paris" with a "Go!" button. A modal window titled "Streets typology in Paris" displays the following data:

Category	Value
Total length	2540154 meters
Residential roads	764508 m
Service roads	473970 m

Below the modal is a pie chart showing the distribution of road types. On the right side of the map, there are two circular markers labeled "O" and "A" with arrows pointing towards them, and a "Viewer" button.

Thank you for your attention

- Connect @liyubov

Projects with iNaturalist, cri-paris.org and correlaid.org

Blog posts in LPI [Paris](#)

Social media: <https://twitter.com/luyibov>

L.Tupikina, M.Haklay “Participatory projects in urbanism”, “Platform urbanism” book chapter (2024)

Session “Levaraging data in citizen science“ ECSA (2024)

Computer Science > Social and Information Networks

[Submitted on 5 Dec 2021]

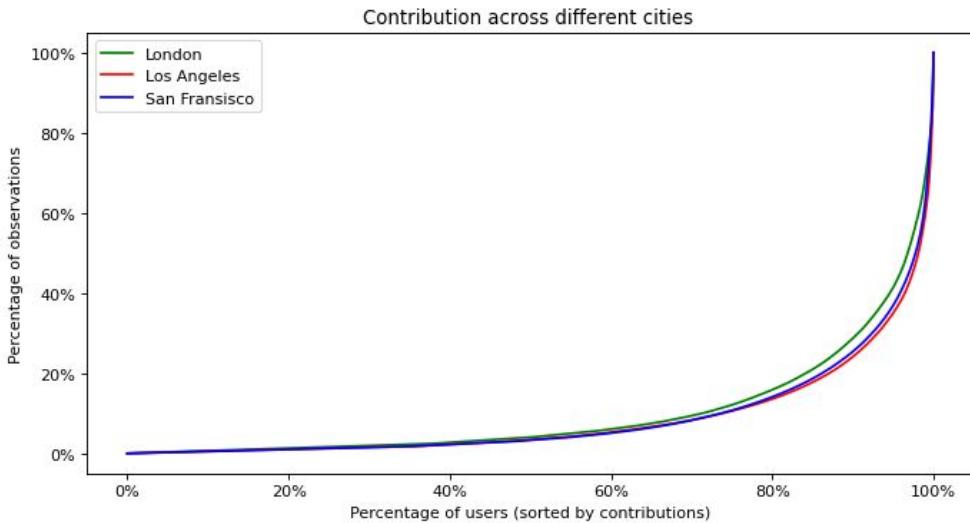
iNaturalist citizen science community during City Nature Challenge: new computational approach for analysis of user activity

Lubov Tupikina, Frank Schlosser, Vadim Voskresenski, Katharina Kloppenborg, Florence Lopez, Albrecht Mariz, Anna Mogilevskaja, Mu Haklay, Bastian Greshake Tzovaras

Analysing patterns of engagement among citizen science participants can provide important insights into the organisation and practice of individual citizen science projects. In particular, methods from statistics and network science can be used to understand different types of user behaviour and user interactions to help the further implementation and organization of community efforts. Using publicly available data from the iNaturalist community and their yearly City Nature Challenges (CNC) from 2017-2020 as an example, we showcase computational methods to explore the spatio-temporal evolution of this citizen science community that typically interacts in a hybrid offline-online way. In particular, we investigate the user types present in the community along with their interactions, finding significant differences in usage-behavior on both the level of engagement and the types of community tasks/roles and how they interact with the network of contributors. We expect that these computational analysis strategies will be useful to gain further understanding of other citizen science communities and projects.

General statistics for London, San Francisco, and Los Angeles 2018-2020.

The total number of observation made by users were sorted by their observations in ascending order and we could see that the top 50% of users contribute 97% of observations, 10% of users contribute 75% of observations and 1% of users contribute 37% of observations

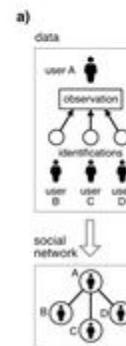
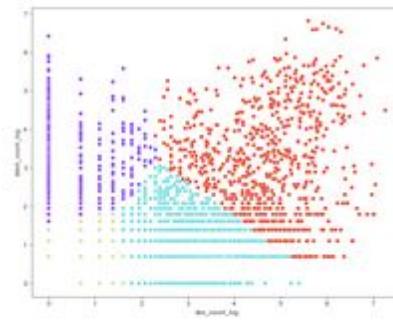
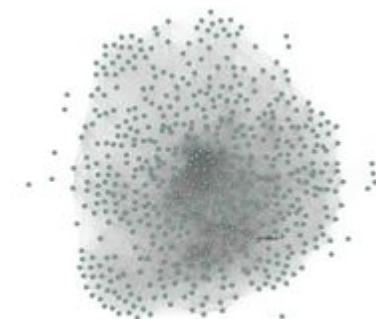


3. Temporal evolution of participants' activity types in cities

Clustering : We used the k-means algorithm to group the data into different clusters based on user contributions and assigned the clusters into four categories: **Observers, Identifiers, High activity users, and Low activity users.**

High activity users (Red): both identified and observed in high quantity, identifiers (blue)– who helped to identify, observers(green) - who made more observation, and low activity users (yellow) - least contribution/identification. The plots use the same colors through the different years to observe any shifts in user behavior over time. Eg: The plot in green always used people in 2018/2019 who were observers in 2018. Similarly, for other colors too.

Social Network Analysis



b)

