THE DATASCIENTIA ECOSYSTEM

Fostering sharing, learning, research and innovation at local level

DataScientia

ICTAS – DataScientia Symposium
22.07.2025
Matteo Busso

8,000+ evolving languages
Scientific experts and local experts
(expertise so deep that few people know it)

How can we scale?
How can we engage people in a long-lasting experience?

INDEX

CITIZEN SCIENCE

DataScientia

WHAT IS CITIZEN SCIENCE?

- Citizen science refers to the **involvement of non-professional scientists** (ordinary members of the public) in scientific research.
- It enables **large groups** of people to contribute to **data collection**, analysis, and discovery, often through online platforms or local projects.
- This approach **helps researchers gather and process vast amounts of data** that would be impossible to handle alone.

Design:

- Scientists design a project with clear tasks that can be completed by volunteers.

Recruitment:

- People are recruited, often via the internet, schools, or community groups.

Training:

- Participants are provided with tutorials or guidelines to perform tasks correctly.

Contribution:

- Volunteers analyze data, report observations, or perform classifications (e.g., identifying species, counting objects).

Validation:

- Data is checked for accuracy using methods like redundancy (multiple people doing the same task).

Outcome:

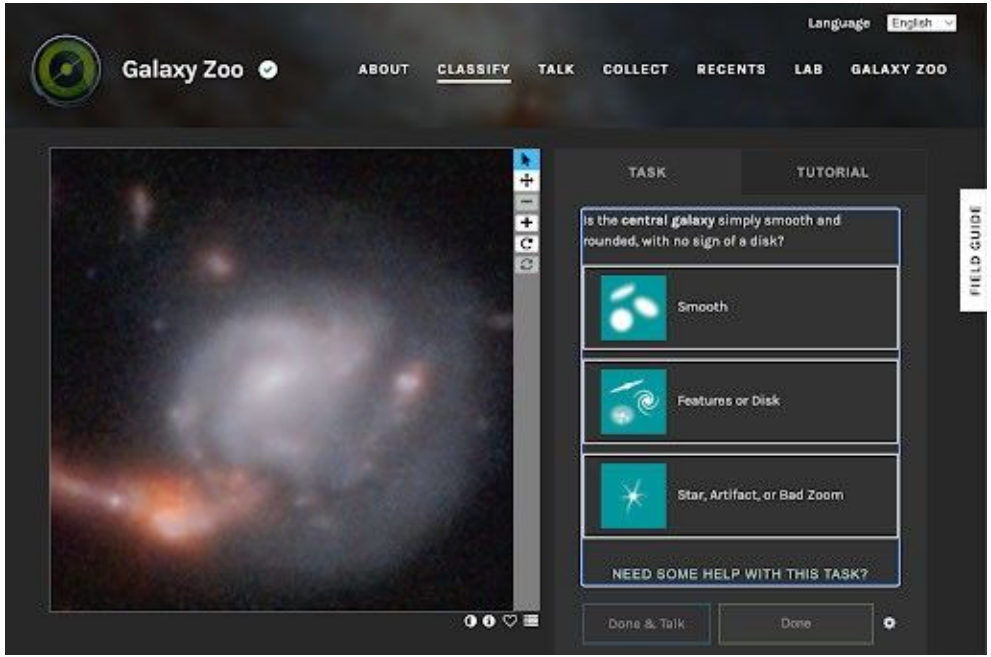
- Results contribute to scientific publications, conservation efforts, or technological advancements.

EXAMPLE: GALAXY ZOO

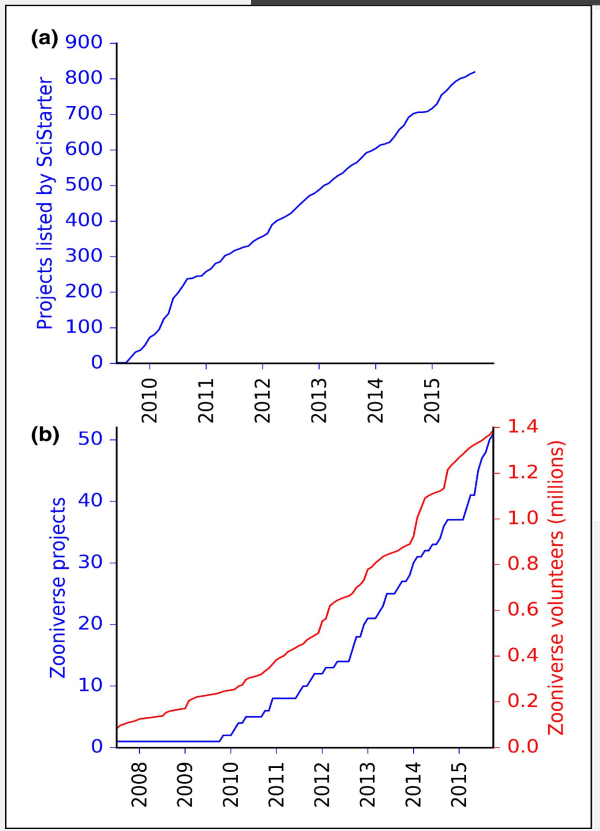
Background: Galaxy Zoo (launched in 2007) asked volunteers to classify images of galaxies from the Sloan Digital Sky Survey (SDSS).

Task: Participants determined galaxy shapes (spiral, elliptical, etc.) to help understand galactic evolution.

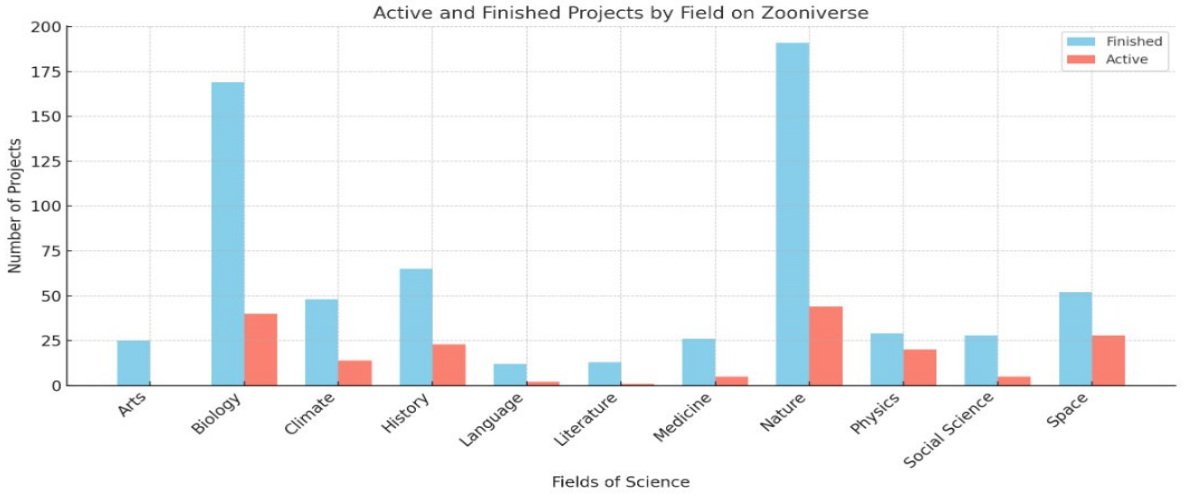
Impact: Millions of classifications were made, leading to numerous scientific papers and discoveries (including new



EXPONENTIAL GROWTH IN POPULARITY



Increase in citizen-science projects and volunteers
~ McKinley et al. [2015]



iNaturalist
[iNaturalist](#)

eu citizen science
[eu-citizen.science](#)

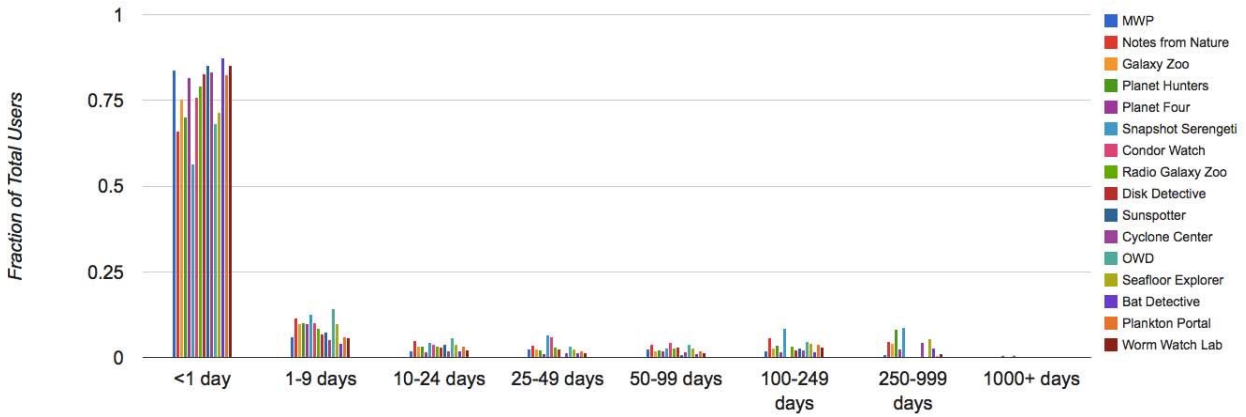
[zooniverse](#)

CitSci.org
helping you do great science
[CitSci](#)

scistarter
Science we can do together.
[SciStarter](#)

and many more …

DROP OUT IN 16 CITIZEN SCIENCE PROJECTS



Source: Segal, A., Gal, Y. A., Simpson, R. J., Victoria Homsy, V., Hartswood, M., Page, K. R., & Jirotka, M. (2015, May). Improving productivity in citizen science through controlled intervention. In *Proceedings of the 24th international conference on world wide web* (pp. 331-337).

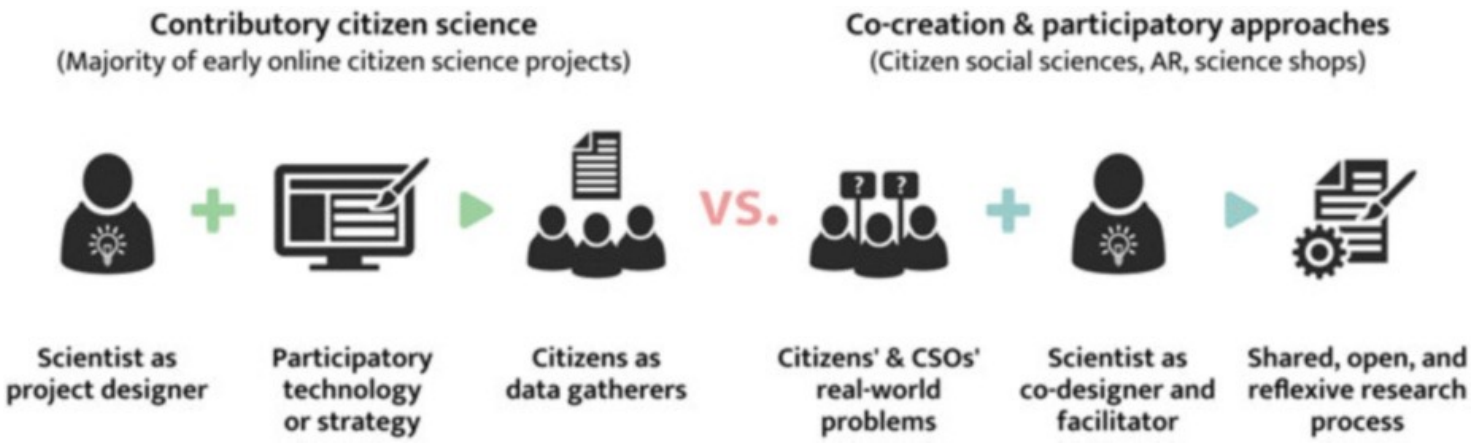
LIMITS OF CITIZEN SCIENCE APPROACH

Insight	Summary
High dropout rates	Over half of volunteers leave within a year, 60–83% don't return after the first visit
Core contributors	A small group supplies the majority of work: top 10% = ~80% of contributions
Motivation matters	Learning, to be part of something, s
Learning vs doing	Classifying objects doesn't necessa structured education is needed
Recognition improves quality	Switching from competitive leaderboards to egalitarian acknowledgment enhanced data integrity
Engagement diversity	Volunteer types vary—understanding this helps tailor retention strategies

There is no AI community!

Sources: Swanson, H., & Gazzoni, C. (2013). Crowdsourcing: user contribution patterns and their implications. *Proceedings of the national academy of sciences*, *112*(3), 679-684.
Shinbrot, X. A., et al. (2023). Why citizen scientists volunteer: the influence of motivations, barriers, and perceived project relevancy on volunteer participation and retention from a novel engagement. *Journal of Environmental Planning and Management*, *66*(1), 122-142.

TOWARDS PARTICIPATORY APPROACHES



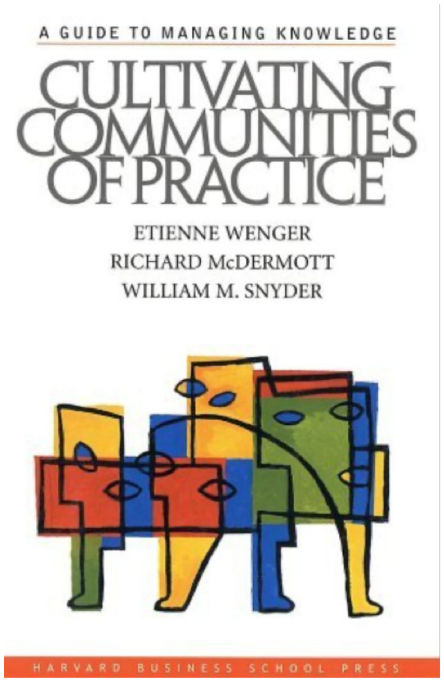
COMMUNITY OF PRACTICES

- How to scale: ideas from Citizen Science
- **Learning experience: Community of practices**
- Requirements for a growing community
- The DataScientia Community
- The Community in action
- Conclusion

DataScientia

COMMUNITIES OF PRACTICE

- These are groups of people who meet regularly to learn, share experiences, solve problems, and grow collective knowledge. They exist everywhere—in both professional and social settings—and foster the development of tacit skills and shared practices.



DataScientia

CHRYSLER CASE STUDY

- In the 1990s, Chrysler drastically improved its product development cycle by shifting from functional structures to product-oriented platforms.

- However, this created problems of knowledge fragmentation. The solution was the creation of **Tech Clubs**, informal cross-functional communities that facilitated knowledge sharing and led to the creation of an **Engineering Book of Knowledge (EBoK)**.

THE DOUBLE-KNIT

Manages living knowledge

- communities capture and update expertise even when projects end or company structures change.

Accelerates problem solving

- employees can quickly access distributed knowledge and experience within the organization.

Fosters innovation

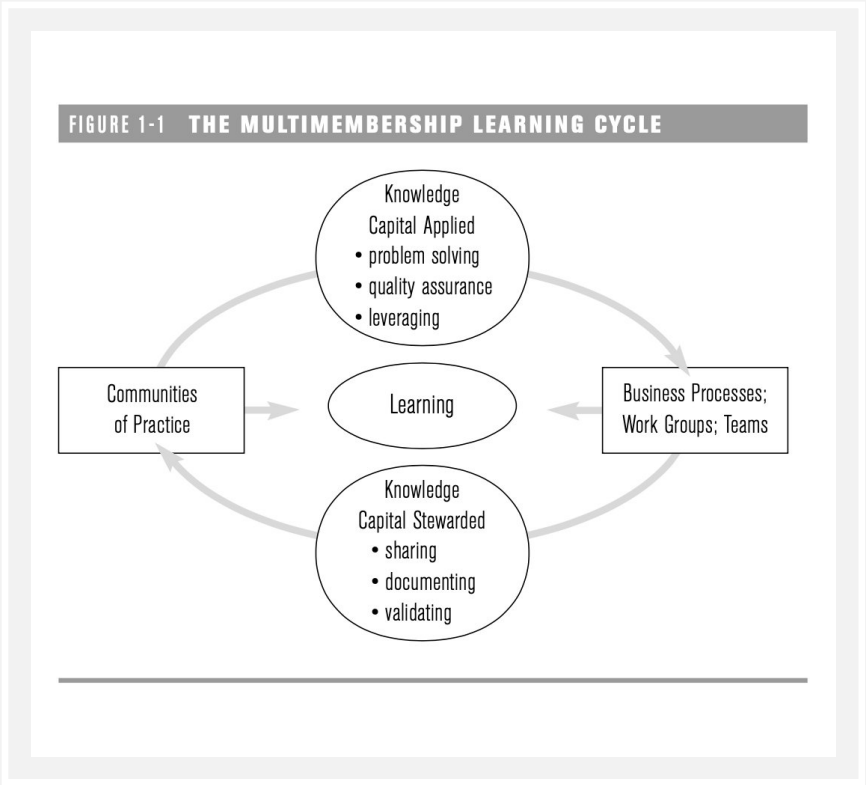
- informal exchanges between experts stimulate new ideas and solutions.

Makes the organization more resilient

- even during reorganizations or market shifts, knowledge stays cohesive thanks to the communities.

Strengthens identity and belonging

- employees maintain a “professional home” despite changes in organizational structures.



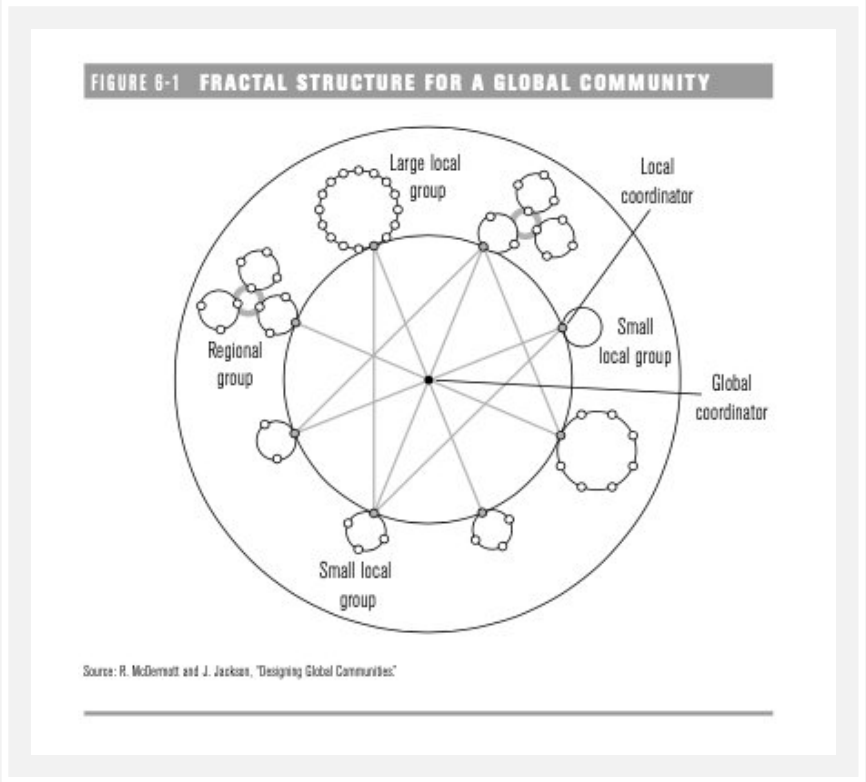
FRACTAL STRUCTURE (SHELL PILOTS)

Structured as a Network of Local "Cells":

- Local communities (cells) were formed inside business units.
- Each local cell had **autonomy** to structure itself (e.g., weekly meetings, informal chats).
- A **local coordinator** connected each cell to the global community.

A Coordinators' Network:

- Coordinators regularly met (virtually and sometimes in-person).
- They became the backbone of the global network, facilitating both local and global knowledge sharing.



REQUIREMENTS

IN A NUTSHELL: REQUIREMENTS FOR A GROWING COMMUNITY

Citizen science

- Focus on data
- Diversity of projects
- Focus on the relationship between researcher and participant
- Platform for connecting people through data collections

Community of practice

- Focus on knowledge and learning
- Focus on sharing
- Path to get involved and become an expert
- Path to foster local knowledge in a global setting

THE DATASCIENTIA COMMUNITY

How to scale: ideas from Citizen Science

Learning experience: Community of practices

Requirements for a growing community

The DataScientia Community

The Community in action

Conclusion

ABOUT DATASCIENTIA

A global Community shaping AI and data for the benefit of people and society

Combines sharing, education, research, and innovation in one ecosystem.

Open to all – citizens, researchers, developers, educators, institutions

In close collaboration with partner Universities

THE DATASCIENTIA DOUBLE-KNIT

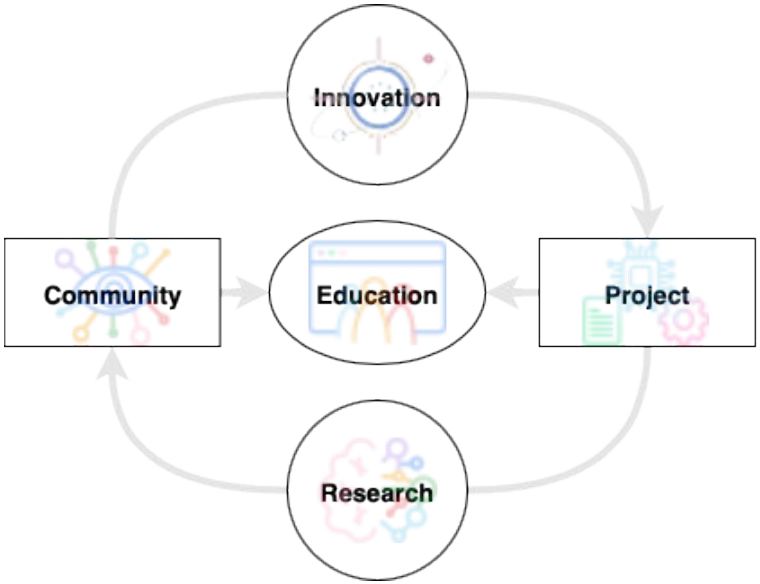
Community: Explore AI & data's real-life relevance with the help of others

Projects: Collaborate or lead ethical, inclusive scientific projects.

Education: Learn via multilingual courses and local learning paths.

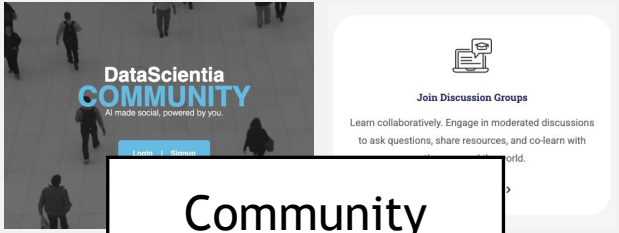
Research: Develop and share new insights and knowledge

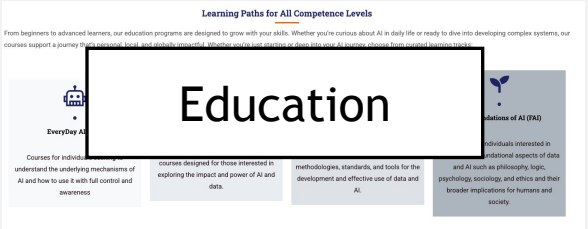
Innovation: Build real-world applications grounded in shared knowledge.

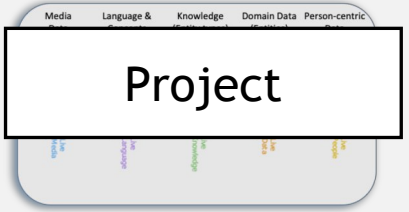


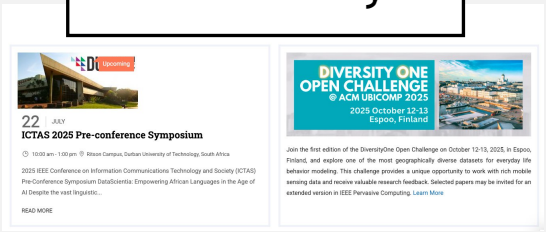
The DataScientia Ecosystem

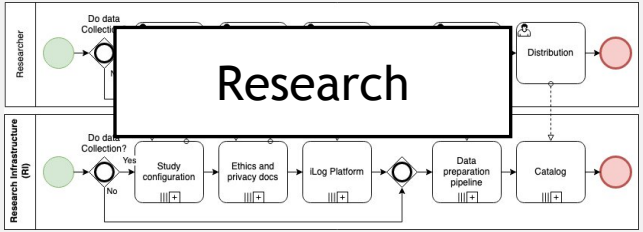
DataScientia











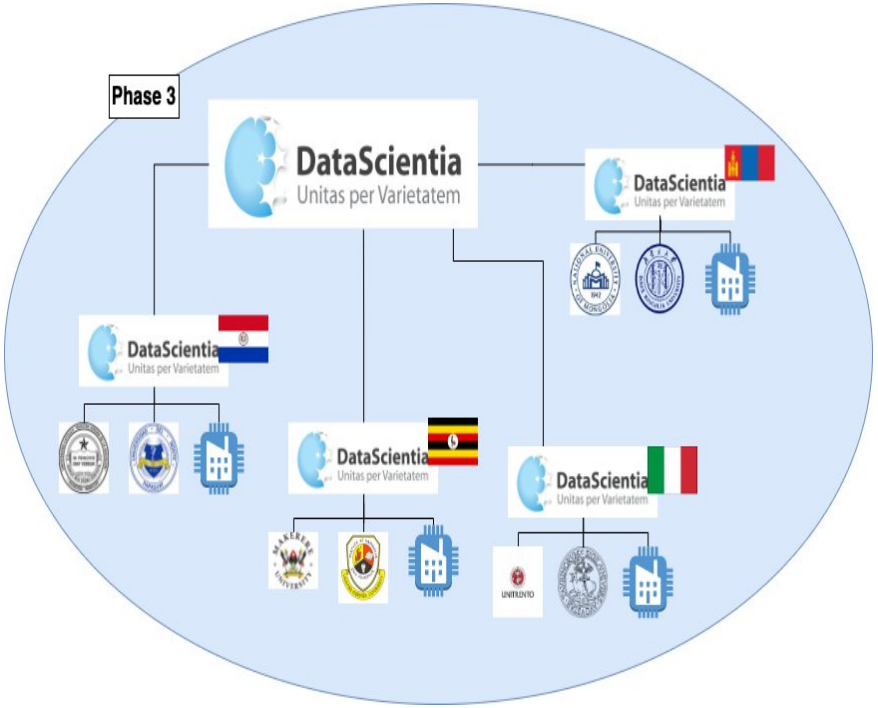
Phase 1

UNITRENTO

Phase 2

DataScientia
Unitas per Varietatem

Phase 3



DataScientia

DataScientia

DataScientia

HOW TO GET INVOLVED?

There are many ways in which you can contribute. You can choose the one or more activities that best fit your competence, skills, and interests. Contribute to DataScientia by supporting data collection, educating others, programming AI tools, or managing local communities. You can create or participate in experiments, develop educational materials, share experiences, teach classes, code open-source AI tools, or lead local community efforts. Dive into diverse opportunities tailored to your skills and interests.

Contribute to collecting person-centric data by creating or participating in projects.

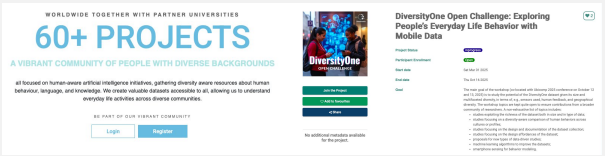
Contribute to education: materials, experiences, teaching.

Contribute in development of software and AI tools.

Help manage local DataScientia communities.

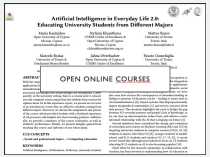
THE COMMUNITY IN ACTION

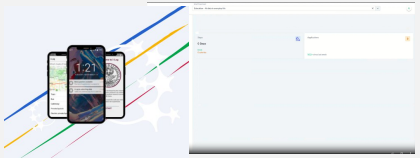
EXAMPLE: THE PARTICIPANT EXPERIENCE



Incentives

- Monetary
- Badges
- Recognition





DataScientia

WORLDWIDE TOGETHER WITH PARTNER UNIVERSITIES

60+ PROJECTS

A VIBRANT COMMUNITY OF PEOPLE WITH DIVERSE BACKGROUNDS

all focused on human-aware artificial intelligence initiatives, gathering diversity aware resources about human behaviour, language, and knowledge. We create valuable datasets accessible to all, allowing us to understand everyday life activities across diverse communities.

BE PART OF OUR VIBRANT COMMUNITY

Login Register

DataScientia

Join the Project

♡ Add to favourites

Share

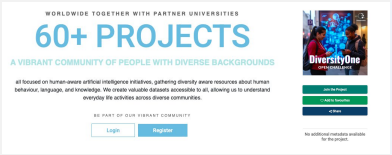
No additional metadata available for the project.

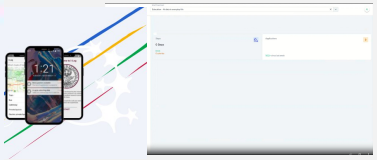
DiversityOne Open Challenge: Exploring People's Everyday Life Behavior with Mobile Data

♥ 2

Project Status	inprogress
Participant Enrollment	Open
Start date	Sat Mar 01 2025
End date	Thu Oct 16 2025
Goal	The main goal of the workshop (co-located with Ubicomp 2025 conference on October 12 and 13, 2025) is to study the potential of the DiversityOne dataset given its size and multifaceted diversity, in terms of, e.g., sensors used, human feedback, and geographical diversity. The workshop topics are kept quite open to ensure contributions from a broader community of researchers. A non-exhaustive list of topics includes: • studies exploiting the richness of the dataset both in size and in type of data; • studies focusing on a diversity-aware comparison of human behaviors across cultures or profiles; • studies focusing on the design and documentation of the dataset collection; • studies focusing on the design affordances of the dataset; • proposals for new types of data-driven studies; • machine learning algorithms to improve the datasets; • smartphone sensing for behavior modeling.

DataScientia

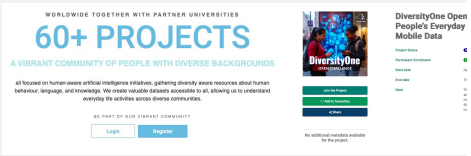


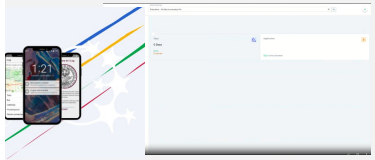


DataScientia

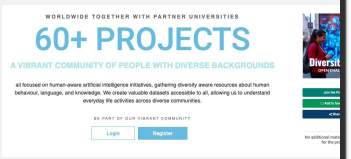
Incentives

- Monetary
- Badges
- Recognition





DataScientia



Artificial Intelligence in Everyday Life 2.0: Educating University Students from Different Majors

Maria Kasinidou
Open University of Cyprus
Nicosia, Cyprus
maria.kasinidou@ouc.ac.cy

Styliani Kleanthous
CYENS Centre of Excellence &
Open University of Cyprus
Nicosia, Cyprus
s.kleanthous@cyens.org.cy

Matteo Busso
University of Trento
Trento, Italy
matteo.busso@unitn.it

Marcelo Rodas
University of Trento &
Fondazione Bruno Kessler

Jahna Otterbacher
Open University of Cyprus &
CYENS Centre of Excellence

Fausto Giunchiglia
University of Trento
Trento, Italy

OPEN ONLINE COURSES

ABSTRAC

With the surge… recognition to applications h… he integration misunderstan… v applications associated advantages and disadvantages are widespread. Consequently, in the university setting, there is a crucial need to educate not only computer science majors but also students from various disciplines about AI. In this experience report, we present an overview of an introductory course that we offered to students coming from different majors. Moreover, we discuss the assignments and quizzes of the course, which provided students with a firsthand experience of AI processes and insights into their learning patterns. Additionally, we provide a summary of the course evaluation, as well as students' performance. Finally, we present insights gained from teaching this course and elaborate on our future plans.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**.

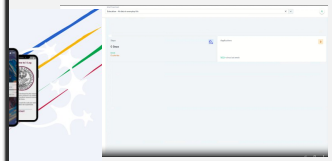
KEYWORDS

Artificial Intelligence, AI Education, AI literacy, university students

…nments, social …nade, we have also seen how serious the consequences of misunderstanding or failing to question AI decisions can be – leading to issues such as viral misinformation [8], biased systems that disproportionately impact marginalized communities [1], and serious concerns about data privacy. This situation highlights the need to bridge the gap between AI's everyday presence and people's lack of knowledge, so we can clear up misconceptions, reduce fears, and embrace a more informed relationship with the AI that is shaping our future [4].

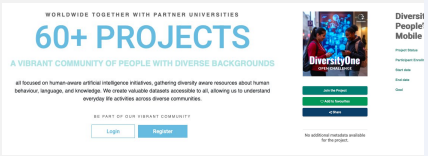
Several initiatives have considered the design and evaluation of curriculum and courses on Machine Learning (ML) and AI [14], targeting university students in computer science (CS) [6, 22, 24], students in majors other than CS [25], younger students in middle-school, and K-12 students and teachers [15, 17, 18, 21]. A recent review found that although most initiatives focus on CS students, educating K-12 students on AI is also becoming popular [19].

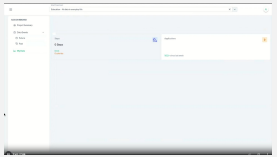
Much effort by the research community, in collaboration with teachers, has been devoted to understanding how AI education at



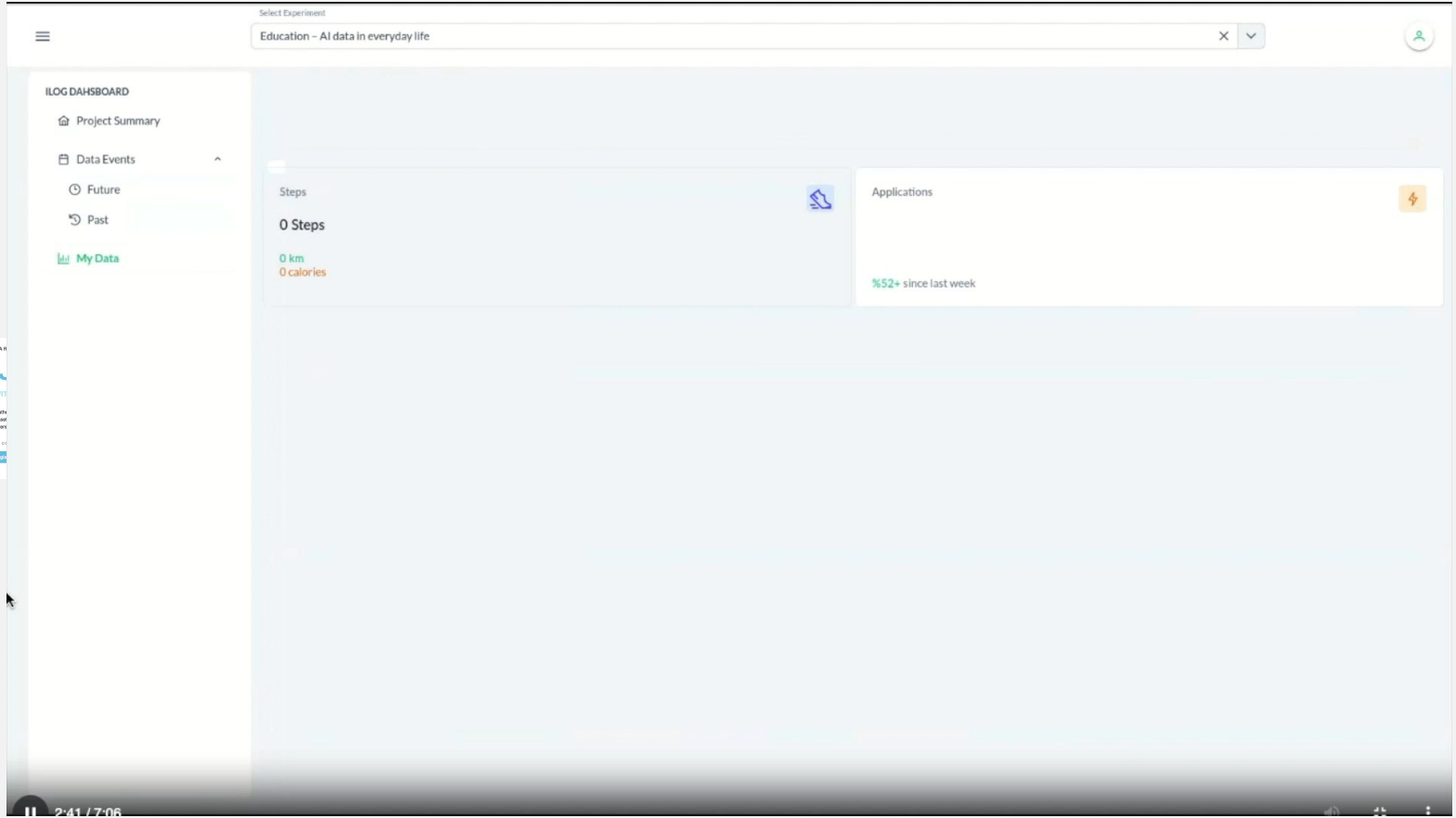

UNIVERSITÀ DI TRENTO
DataScientia
Everyday AI:
Introduzione ai dati dell'Intelligenza Artificiale
TRAILER

DataScientia





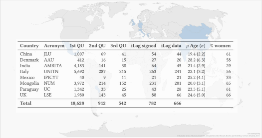
DataScientia

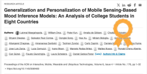


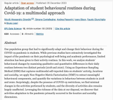
THE MAKERERE UNIVERSITY EXPERIENCE (UGANDA)

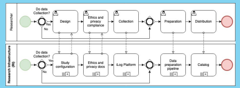
DataScientia Community GLOBAL

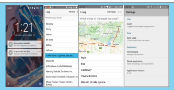
DS MAK LOCAL

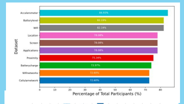


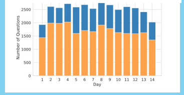






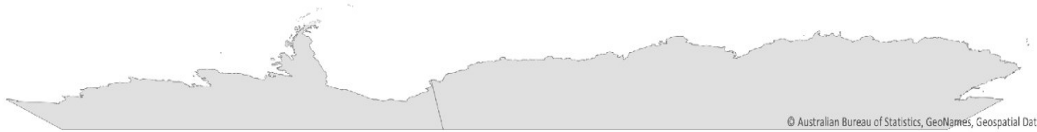






Mak2 Oct 2025

Country	Acronym	1st QU	2nd QU	3rd QU	iLog signed	iLog data	μ Age (σ)	% women
China	JLU	1,007	69	41	54	44	19.4 (2.2)	61
Denmark	AAU	412	16	15	27	20	28.2 (6.3)	58
India	AMRITA	4,183	141	38	64	45	21.4 (2.9)	29
Italy	UNITN	5,692	287	215	263	241	22.1 (3.2)	56
Mexico	IPICYT	40	9	11	21	21	25.2 (4.1)	33
Mongolia	NUM	3,972	214	152	231	201	20.0 (3.1)	65
Paraguay	UC	1,342	33	25	43	28	23.3 (5.1)	61
UK	LSE	1,980	143	45	88	66	24.6 (5.0)	66
Total		**18,628**	**912**	**542**	**782**	**666**		



DataScientia

RESEARCH-ARTICLE

Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries

Authors: Lakmal Meegahapola, William Droz, Peter Kun, Amalia de Götzen, Chaitanya Nutakki, Shyam Diwakar, Salvador Ruiz Correa, Donglei Song, Hao Xu, Miriam Bidoglia, George Gaskell, Altangerel Chagnaa, Amarsanaa Ganbold, Tsolmon Zundui, Carlo Caprini, Daniele Miorandi, Alethia Hume, Jose Luis Zarza, Luca Cernuzzi, Ivano Bison, Marcelo Rodas Britez, Matteo Busso, Ronald Chenu-Abente, Can Günel, Fausto Giunchiglia, Laura Schelenz, Daniel Gatica-Perez (Less) Authors Info & Claims

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Volume 6, Issue 4 • Article No.: 176, pp 1–32 • https://doi.org/10.1145/3569483

Regular article | Open access | Published: 05 December 2023

Adaptation of student behavioural routines during Covid-19: a multimodal approach

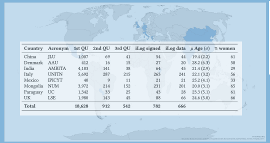
Nicolò Alessandro Girardini, Simone Centellegher, Andrea Passerini, Ivano Bison, Fausto Giunchiglia & Bruno Lepri

EPJ Data Science **12**, Article number: 55 (2023) | Cite this article

Abstract

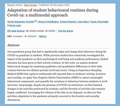
One population group that had to significantly adapt and change their behaviour during the COVID-19 pandemic is students. While previous studies have extensively investigated the impact of the pandemic on their psychological well-being and academic performance, limited attention has been given to their activity routines. In this work, we analyze students' behavioural changes by examining qualitative and quantitative differences in their daily routines between two distinct periods (2018 and 2020). Using an Experience Sampling Method (ESM) that captures multimodal self-reported data on students' *activity*, *locations* and *sociality*, we apply Non-Negative Matrix Factorization (NMF) to extract meaningful behavioural components, and quantify the variations in behaviour between students in 2018 and 2020. Surprisingly, despite the presence of COVID-19 restrictions, we find minimal changes in the activities performed by students, and the diversity of activities also remains largely unaffected. Leveraging the richness of the data at our disposal, we discover that activities adaptation to the pandemic primarily occurred in the *location* and *sociality* dimensions.

DataScientia Community GLO

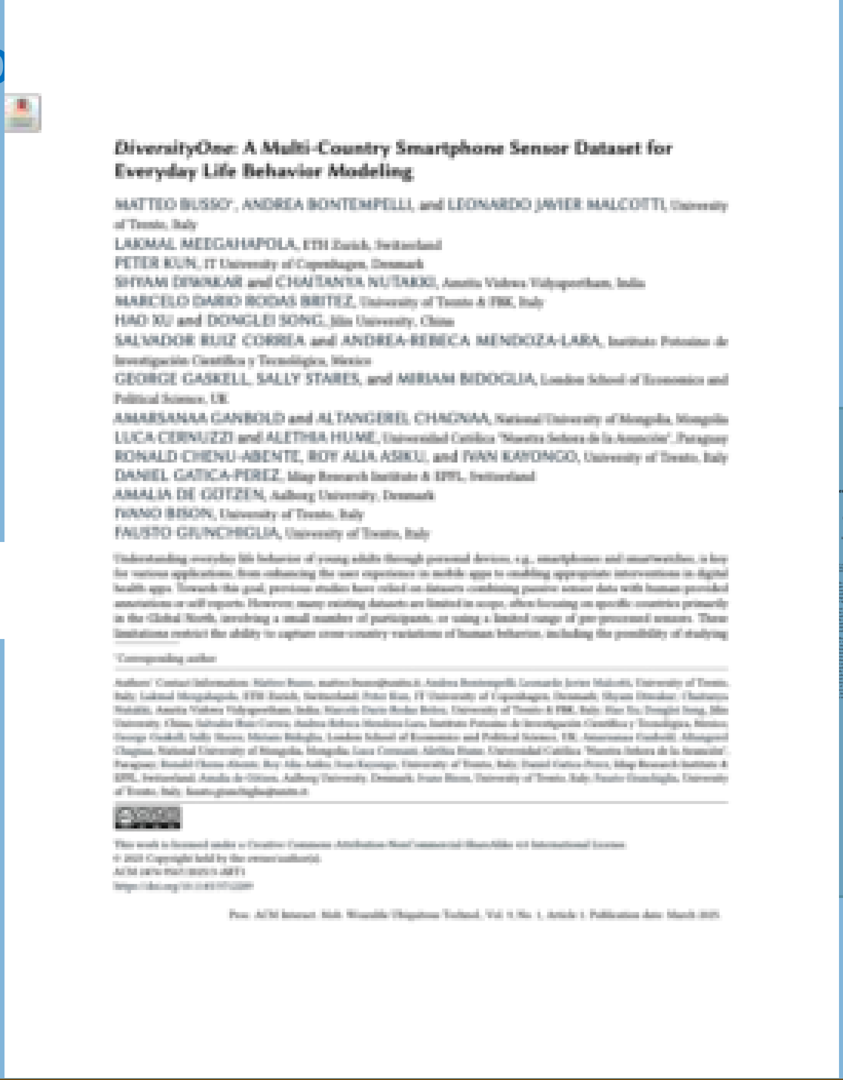


Country	Acronym	1st QU	2nd QU	3rd QU	iLog signed	iLog data	μ Age (σ)	% women
China	JLU	1,007	69	41	54	44	19.4 (2.2)	61
Denmark	AAU	412	16	15	27	20	28.2 (6.3)	58
India	AMRITA	4,183	141	38	64	45	21.4 (2.9)	29
Italy	UNITN	5,692	287	215	263	241	22.1 (3.2)	56
Mexico	IPICYT	40	9	11	21	21	25.2 (4.1)	33
Mongolia	NUM	3,972	214	152	231	201	20.0 (3.1)	65
Paraguay	UC	1,342	33	25	43	28	23.3 (5.1)	61
UK	LSE	1,980	143	45	88	66	24.6 (5.0)	66
Total		**18,628**	**912**	**542**	**782**	**666**		

Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries



Adaptation of student behavioural routines during Covid-19: a multimodal approach



DiversityOne: A Multi-Country Smartphone Sensor Dataset for Everyday Life Behavior Modeling

MATTEO BUSSO*, ANDREA BONTEMPELLI, and LEONARDO JAVIER MALCOTTI, University of Trento, Italy
LAKMAL MEEGAHAPOLA, ETH Zurich, Switzerland
PETER KUN, IT University of Copenhagen, Denmark
SHYAM DIWAKAR and CHAITANYA NUTAKKI, Amrita Vishwa Vidyapeetham, India
MARCELO DARIO RODAS BRITEZ, University of Trento & FBK, Italy
HAO XU and DONGLEI SONG, Jilin University, China
SALVADOR RUIZ CORREA and ANDREA-REBECA MENDOZA-LARA, Instituto Potosino de Investigación Científica y Tecnológica, Mexico
GEORGE GASKELL, SALLY STARES, and MIRIAM BIDOGLIA, London School of Economics and Political Science, UK
AMARSANAA GANBOLD and ALTANGEREL CHAGNAA, National University of Mongolia, Mongolia
LUCA CERNUZZI and ALETHIA HUME, Universidad Católica "Nuestra Señora de la Asunción", Paraguay
RONALD CHENU-ABENTE, ROY ALIA ASIKU, and IVAN KAYONGO, University of Trento, Italy
DANIEL GATICA-PEREZ, Idiap Research Institute & EPFL, Switzerland
AMALIA DE GÖTZEN, Aalborg University, Denmark
IVANO BISON, University of Trento, Italy
FAUSTO GIUNCHIGLIA, University of Trento, Italy

Understanding everyday life behavior of young adults through personal devices, e.g., smartphones and smartwatches, is key for various applications, from enhancing the user experience in mobile apps to enabling appropriate interventions in digital health apps. Towards this goal, previous studies have relied on datasets combining passive sensor data with human-provided annotations or self-reports. However, many existing datasets are limited in scope, often focusing on specific countries primarily in the Global North, involving a small number of participants, or using a limited range of pre-processed sensors. These limitations restrict the ability to capture cross-country variations of human behavior, including the possibility of studying

*Corresponding author

Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., Vol. 9, No. 1, Article 1. Publication date: March 2025.

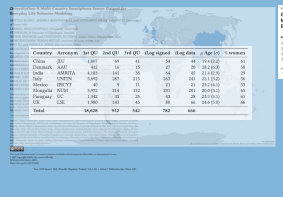
DataScientia

DS MAK LOCAL

Mak2 Oct 2025

DataScientia Community GLOBAL

DataScientia

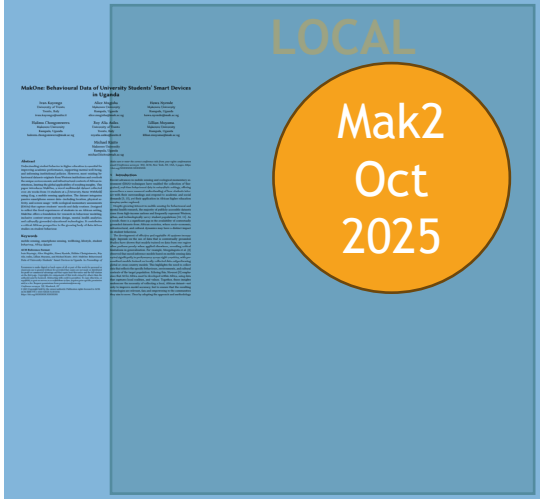


WE BUILD FOR THE FUTURE

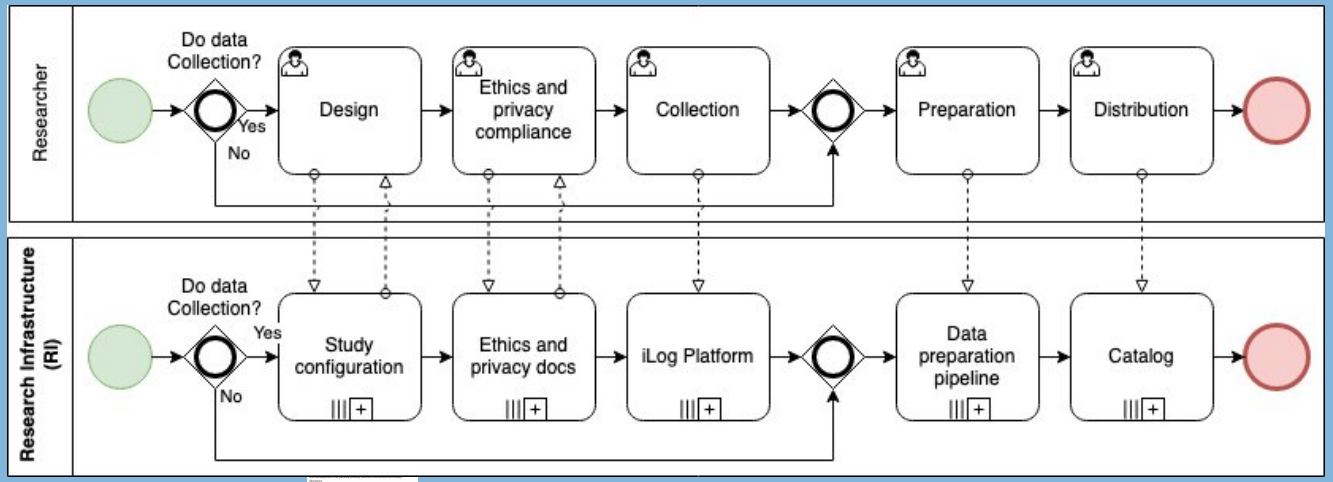
MAKERERE UNIVERSITY

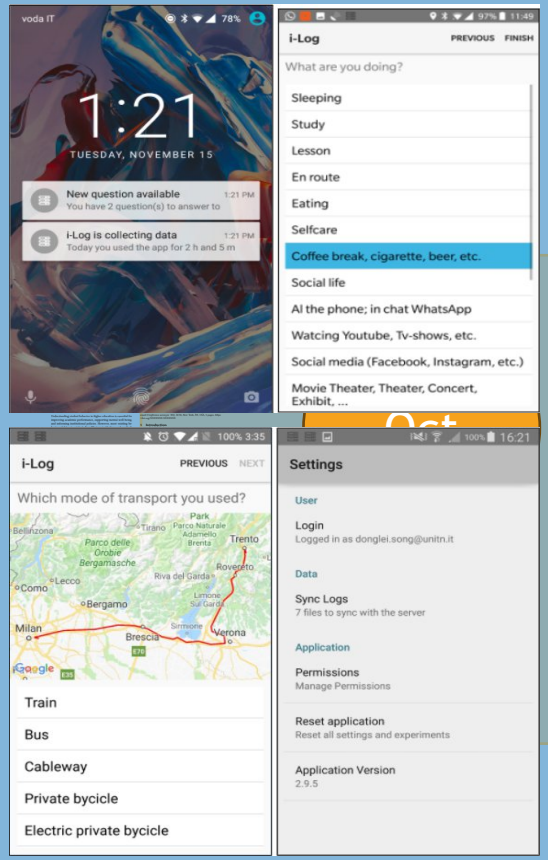
DS MAK LOCAL

Mak2 Oct 2025

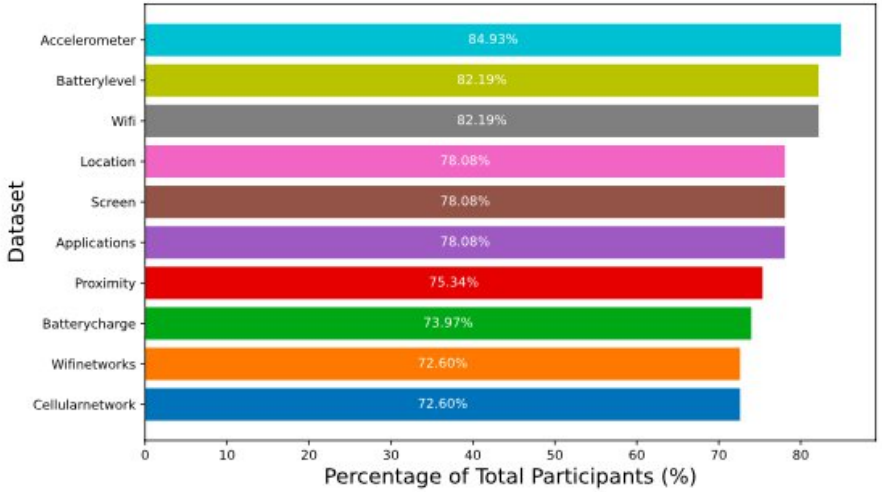


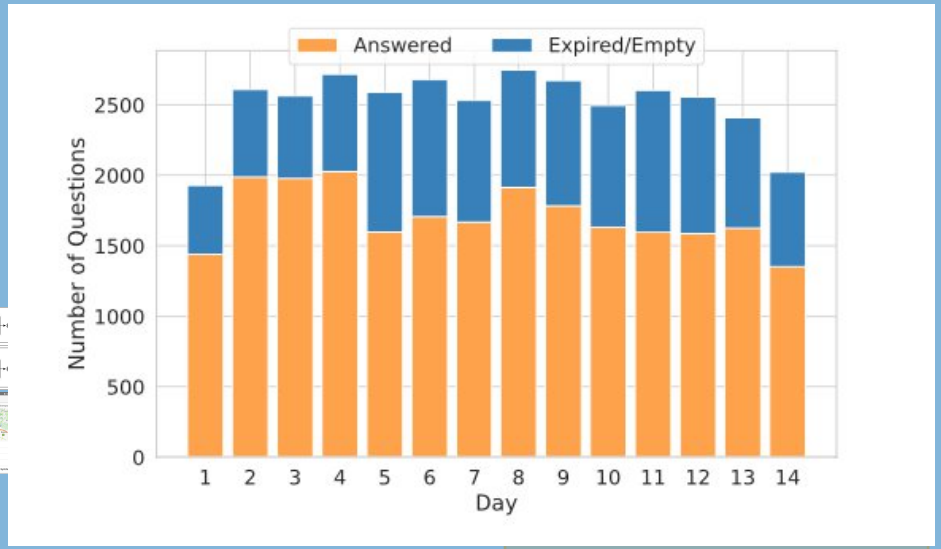
DataScientia Community GLOBAL





DataScientia Community GLOBAL



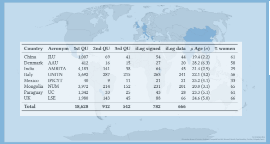


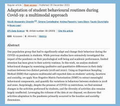
DataScientia Community GLOBAL

DS MAK LOCAL

DataScientia

Scientific paper at ACM Ubicomp Conference





MakOne: Behavioural Data of University Students' Smart Devices in Uganda

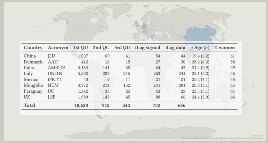
Ivan Kayongo, University of Trento, Trento, Italy, ivan.kayongo@unitn.it
Alice Mugisha, Makerere University, Kampala, Uganda, alice.mugisha@mak.ac.ug
Hawa Nyende, Makerere University, Kampala, Uganda, hawa.nyende@mak.ac.ug
Halimu Chongomweru, Makerere University, Kampala, Uganda, halimu.chongomweru@mak.ac.ug
Roy Alia Asiku, University of Trento, Trento, Italy, royalia.asiku@unitn.it
Lillian Muyama, Makerere University, Kampala, Uganda, lillian.muyama@mak.ac.ug
Michael Kizito, Makerere University, Kampala, Uganda, michael.kizito@mak.ac.ug

Mak2
Oct
2025

DS MAK LOCAL

DataScientia

DataS… …nity GLOBAL

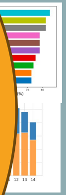


Country	Acronym	1st QU	2nd QU	3rd QU	iLog signed	iLog data	μ Age (σ)	% women
China	JLU	1,007	69	41	54	44	19.4 (2.2)	61
Denmark	AAU	412	16	15	27	20	28.2 (6.3)	58
India	AMRITA	4,183	141	38	64	45	21.4 (2.9)	29
Italy	UNITN	5,692	287	215	263	241	22.1 (3.2)	56
Mexico	IPICYT	40	9	11	21	21	25.2 (4.1)	33
Mongolia	NUM	3,972	214	152	231	201	20.0 (3.1)	65
Paraguay	UC	1,342	33	25	43	28	23.3 (5.1)	61
UK	LSE	1,980	143	45	88	66	24.6 (5.0)	66
Total		**18,628**	**912**	**542**	**782**	**666**		

Generalization and Personalization of Mobile Sensing-Based Mood Inference Models: An Analysis of College Students in Eight Countries

Adaptation of student behavioural routines during Covid-19: a multimodal approach

DiversityOne: A Multi-Country Smartphone Sensor Dataset for Everyday Life Behavior Modeling



MakOne: Behavioural Data of University Students' Smart Devices in Uganda

Mak2
Oct
2025

DataScientia

OUR FIRST DATASCIENTIA JOINT LAB @ZHUHAI (CINA)

Focus on

Human Behavior Data Collection

Everyday AI Education

CONCLUSION

- How to scale: ideas from Citizen Science
- Learning experience: Community of practices
- Requirements for a growing community
- The DataScientia Community
- The Community in action
- **Conclusion**

DataScientia

CONCLUSION: DATASCIENTIA MISSION

- Empower individuals to understand, create, and govern AI.
- Promote cultural and contextual diversity in data and AI systems.
- Ensure all participants are active contributors, not passive subjects.
- Build tools, datasets, and learning paths for responsible innovation.

THANK YOU!!!

Don’t forget our next DS Global Event

03-04 October in TRENTO (Italy)!!!

DataScientia

DataScientia
Unitas per Varietatem

SCAN ME

DataScientia

COMMUNITY

- News & Events
- Monthly Newsletter
- Data Marketplace
- Interest Groups

EDUCATION

Introductory courses:

- Everyday AI series
  - Open University of Cyprus Interactive Annual Course
  - 4 MOOCs (EduGain CFU)

Vertical courses:

- AI & Society
  - Knowledge Graph Engineering (3 Universities – UNITN, NUM, JLU)
  - Foundation Model
  - Studies on human behavior

We see education as a service to the community.

- Each course is available to everyone, adapted to everyone's diversity, anywhere in the world.
- Each course is accessible from the site and comprises individual modules that can be enjoyed separately.
- Each module is based on easily usable and adaptable resources depending on the context's needs.

End-to-end process

LingoGap - Task Creation

Design

Data collection

Data processing

Data distribution

DataScientia

PROJECTS

TYPE OF RESOURCES HANDLED

- Data collection
- Data curation
- Data quality enhancement
- Feature exctraction
- ...

TYPE OF ACTIVITIES (AS OF TODAY)

Media	• that depicts how the world appears to us, with all its diversity.
Data	• that describes how the world appears to us, i.e., what is true and false.
Personal	• data that describes how we are.
Languages	• that we use to describe how the world appears to us.
Knowledge	• that we use to provide a unitary view of how the world appears to us is the key to the purpose-driven composition of data.

INNOVATION (WITHIN DS INITIATIVE)

- User-oriented and context-aware applications

For data collection, management, and sharing, e.g.:

- iLog app,
- Dashboard,
- iTelog methodology,
- DataScientia Catalogs

- Processes and services

SU2OSM: integrating personal GPS location with OpenStreetMap

SKEL: AI assistant for cleaning of your own data

…