DataScientia
Unity Through Diversity

datascientia.eu

DataScientia: A Human-Centered Diversity-aware AI Ecosystem

Fausto Giunchiglia
University of Trento

03 JULY 2026



© 2026 DataScientia – All rights reserved



About DataScientia

- A global community shaping AI and data for the benefit of people and society
- Combines information, education, research, and innovation in one ecosystem.
- Open to all – citizens, researchers (not only in Computer Science or AI), developers, educators, institutions
- In close collaboration with partner Universities

© 2026 DataScientia – All rights reserved

Our Mission

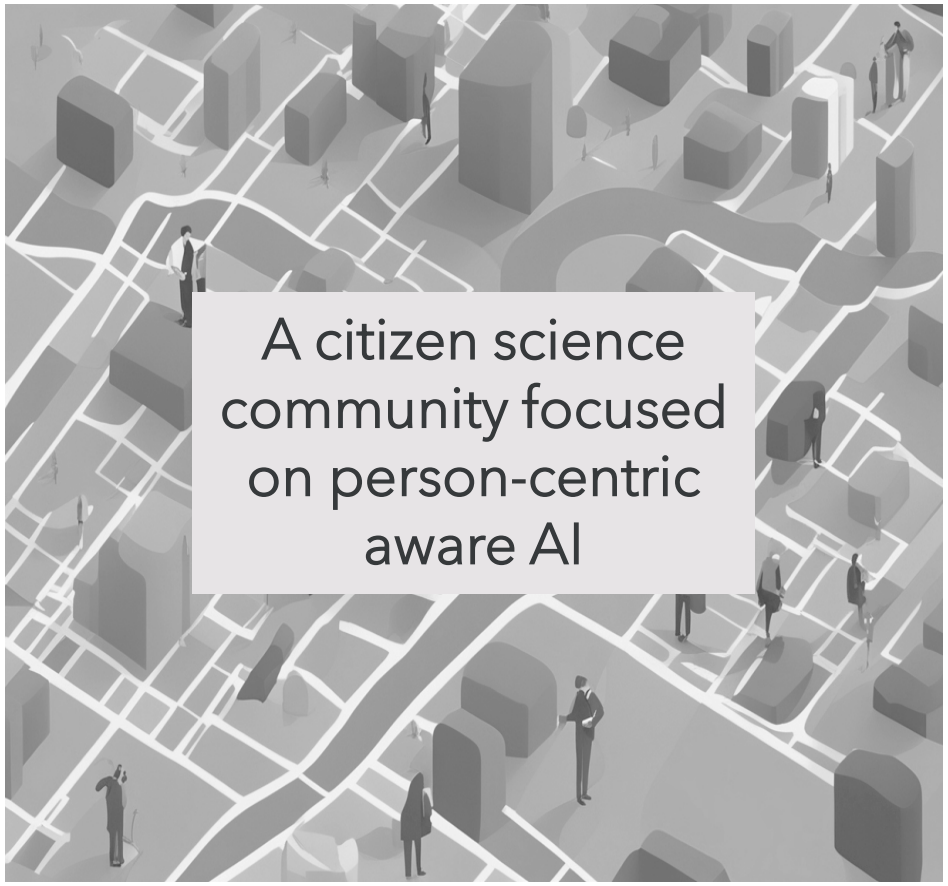


We aim to…

- Empower individuals (NOT ONLY computer scientists) 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.
- Computer scientists: Build tools, datasets, and learning paths for responsible AI innovation.
- All: validate AI, accept/ reject AI, suggest modifications (Citizen Science)

© 2026 DataScientia – All rights reserved

Solution: The DataScientia Community



The DataScientia Community

Towards an era of radical changes in the nature of knowledge

A community-driven approach is

essential

to ensure that AI and knowledge evolve with, and not apart from, the people it impacts.

AI is the means, NOT the goal!

© 2026 DataScientia – All rights reserved

University Joint Labs as global Anchors of "Unity through Diversity"

Joint Labs are the foundational partners driving sharing, education, research, and innovation

SHARING

Disseminate knowledge across local communities and students

EDUCATION

Joint curricula, MOOCs, mobility programs, and training

RESEARCH

Lead Research initiative — data collection, language science, AI

INNOVATION

Local AI ecosystems, apps, and socio-economic value creation

Why Universities Are Essential

- **Guarantee for the proper evolution of knowledge, culture and society (AI is means, NOT the goal!)**
- Maintain local linguistic and cultural diversity
- Connect regional knowledge globally
- Prevent domination of single cultural worldview in AI
- Ensure long-tail, minority-language, knowledge, behavior and culture is preserved

Partnership Growth Path

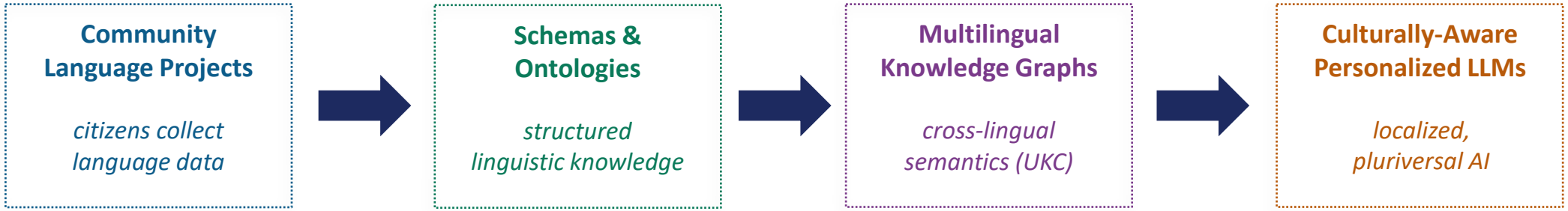
- Joint sharing & education projects
- Multilingual data collection research
- Local AI innovation & engineering

→ **DataScientia Local Node (DataScientia Intermedriaty DSC)**

© 2026 DataScientia – All rights reserved

From **Language Preservation** to **Personalized LLMs**

Why preserving linguistic diversity matters for the future of AI



LiveLanguage Catalog

datascientiafoundation.github.io/LiveLanguage

- Open-access datasets on linguistic diversity
- Cross-lingual lexical semantics (UKC)
- Monolingual & multilingual lexicons
- NLP services & lexicon development tools

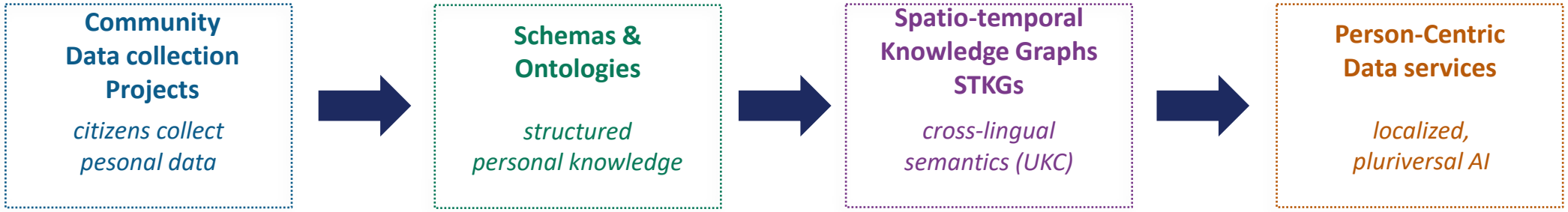
Language Communities as Co-Designers of AI

- Knowledge holders of cultural intelligence
- Contributors to personalized, local AI models
- Long-term stewards of pluriversal knowledge
- Instead of one universal LLM: many culturally-grounded, community-trained models

© 2026 DataScientia – All rights reserved

From **personal data** to **big thick data services**

Why preserving behavioural diversity matters for the future of AI



LivePeople Catalog

datascientiafoundation.github.io/LiveLanguage

- Open-access to behavioral datasets
- Cross-cultural behavioral datasets
- Culture-aware data driven services
- Dataset composition and exploitation tools

local Communities as Co-Designers of AI

- Knowledge holders of cultural intelligence
- Contributors to personalized, local STKGs
- Long-term stewards of pluriversal knowledge
- Instead of one universal STKG: many culturally-grounded, community-developed STKGs

© 2026 DataScientia – All rights reserved

University Joint Labs as Anchors of "Unity through Diversity"

Foundational partners driving sharing, education, research, and innovation

SHARING

Disseminate knowledge across local communities and students

EDUCATION

Joint curricula, MOOCs, mobility programs, and training

RESEARCH

Lead Research initiative — data collection, language science, AI

INNOVATION

Local AI ecosystems, apps, and socio-economic value creation

Why Universities Are Essential

- Maintain local linguistic and cultural diversity
- Connect regional knowledge globally
- Prevent domination of single cultural worldview in AI
- Ensure long-tail, minority-language, knowledge, behavior andculture is preserved

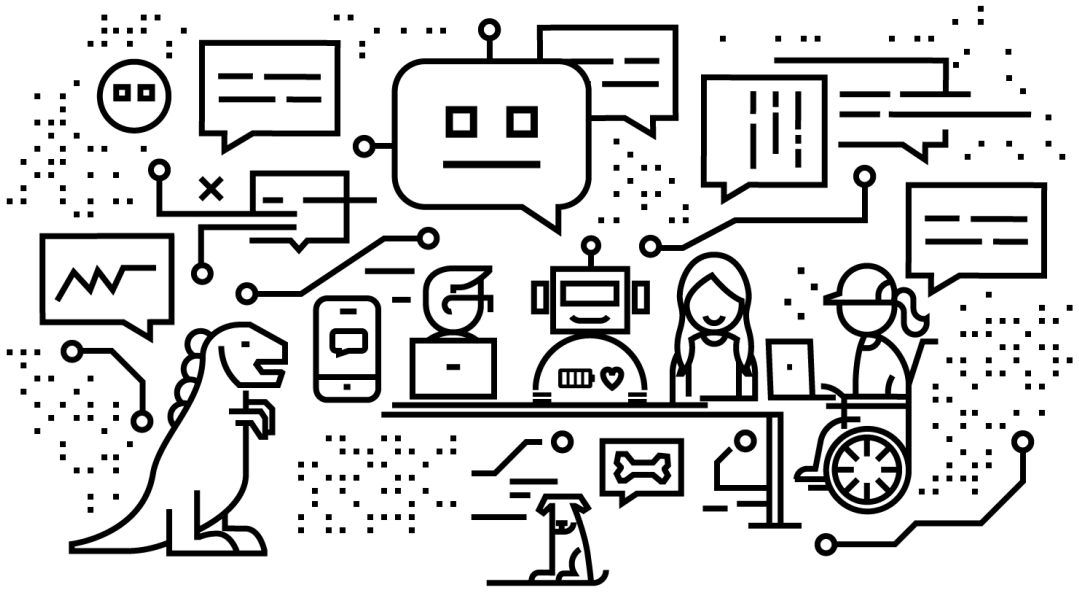
Partnership Growth Path

- Joint sharing & education projects
- Multilingual data collection research
- Local AI innovation & engineering

→ **DataScientia Local Node (DataScientia Intermedriaty DSC)**

© 2026 DataScientia – All rights reserved

Sharing knowledge



[Image Source]

THE KNOWLEDGE COMMONS

- Subscribe to a (moderated) discussion group
- See the most recent presentation at a DataScientia conference or event
- Share short essays about what you have learned in a course
- Ask questions about a topic you consider of importance
- Propose a topic of discussion
- Share what you have learned and teach others (for instance the sourse students of next years)
- Help others by providing explanations (for instance computer science students to sociology / humanities students AND vice versa

PARTICIPATE IN BUILDING A COMMUNITY OF KNOWLEDGE!!!

© 2026 DataScientia – All rights reserved

University Joint Labs as Anchors of "Unity through Diversity"

Foundational partners driving sharing, education, research, and innovation

SHARING

Disseminate knowledge across local communities and students

EDUCATION

Joint curricula, MOOCs, mobility programs, and training

RESEARCH

Lead Research initiative — data collection, language science, AI

INNOVATION

Local AI ecosystems, apps, and socio-economic value creation

Why Universities Are Essential

- Maintain local linguistic and cultural diversity
- Connect regional knowledge globally
- Prevent domination of single cultural worldview in AI
- Ensure long-tail, minority-language, knowledge, behavior andculture is preserved

Partnership Growth Path

- Joint sharing & education projects
- Multilingual data collection research
- Local AI innovation & engineering

→ **DataScientia Local Node (DataScientia Intermedriaty DSC)**

© 2026 DataScientia – All rights reserved

Education – an example (Moocs for everybody)

- **“EveryDay AI (0)” – learn about the data of your own *personal life***
- **“EveryDay AI (1)” – learn about the data of your own *social life***
- **“EveryDay AI (2)” – learn about how to *collect* your own data**
- **“EveryDay AI (3)” – learn about to *use* your own data in your own *personal interest***
- **EveryDay AI (4) – join the DataScientia community as an educator**
- ***… and much more to come***

© 2026 DataScientia – All rights reserved

University Joint Labs as Anchors of "Unity through Diversity"

Foundational partners driving sharing, education, research, and innovation

SHARING

Disseminate knowledge across local communities and students

EDUCATION

Joint curricula, MOOCs, mobility programs, and training

RESEARCH

Lead Research initiative — data collection, language science, AI

INNOVATION

Local AI ecosystems, apps, and socio-economic value creation

Why Universities Are Essential

- Maintain local linguistic and cultural diversity
- Connect regional knowledge globally
- Prevent domination of single cultural worldview in AI
- Ensure long-tail, minority-language, knowledge, behavior andculture is preserved

Partnership Growth Path

- Joint sharing & education projects
- Multilingual data collection research
- Local AI innovation & engineering

→ **DataScientia Local Node (DataScientia Intermedriaty DSC)**

© 2026 DataScientia – All rights reserved

Data collection – an example (Europe, China, India, …)

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

© 2026 DataS

Data collection – another example (Uganda)

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
haleem.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

© 2026 DataScientia – All rights reserved

Data collection – another example (South Africa)

Hybrid Transformer-Based Large Language Models for Word Sense Disambiguation in the Low-Resource Sesotho sa Leboa Language.

Hlaudi Daniel Masethe, Mosima Anna Masethe, Sunday O Ojo, Pius A Owolawi, Fausto Giunchiglia

This study addresses a lexical ambiguity issue in *Sesotho sa Leboa* that arises from terms with various meanings, often known as homonyms or polysemous words. When compared to, for instance, European languages, this lexical ambiguity in *Sesotho sa Leboa* causes computational semantic problems in NLP when trying to identify the lexicon of a language. In other words, it is challenging to determine the proper lexical category and sense of words due to this ambiguity problem. In order to address the issue of polysemy in the *Sesotho sa Leboa* language, this study set out to create a word sense discrimination (WSD) scheme using a corpus-based hybrid transformer-based architecture and deep learning models. Additionally, the performance of baseline and improved machine learning models for a sequence-based natural language processing (NLP) task was assessed and compared. The baseline models …

© 2026 DataScientia – All rights reserved

University Joint Labs as Anchors of "Unity through Diversity"

Foundational partners driving sharing, education, research, and innovation

SHARING

Disseminate knowledge across local communities and students

EDUCATION

Joint curricula, MOOCs, mobility programs, and training

RESEARCH

Lead Research initiative — data collection, language science, AI

INNOVATION

Local AI ecosystems, apps, and socio-economic value creation

Why Universities Are Essential

- Maintain local linguistic and cultural diversity
- Connect regional knowledge globally
- Prevent domination of single cultural worldview in AI
- Ensure long-tail, minority-language, knowledge, behavior andculture is preserved

Partnership Growth Path

- Joint sharing & education projects
- Multilingual data collection research
- Local AI innovation & engineering

→ **DataScientia Local Node (DataScientia Intermedriaty DSC)**

© 2026 DataScientia – All rights reserved

Data catalogs – two examples

The LivePeople data catalog

https://datascientiafoundation.github.io/LivePeople/

The LiveLanguage data catalog

https://datascientiafoundation.github.io/LiveLanguage/

© 2026 DataScientia – All rights reserved

DataScientia (past) projects – an example

The LiveLanguage data catalog projects

https://ukc.datascientia.eu/projects

© 2026 DataScientia – All rights reserved

Hacathon – an example (continued)

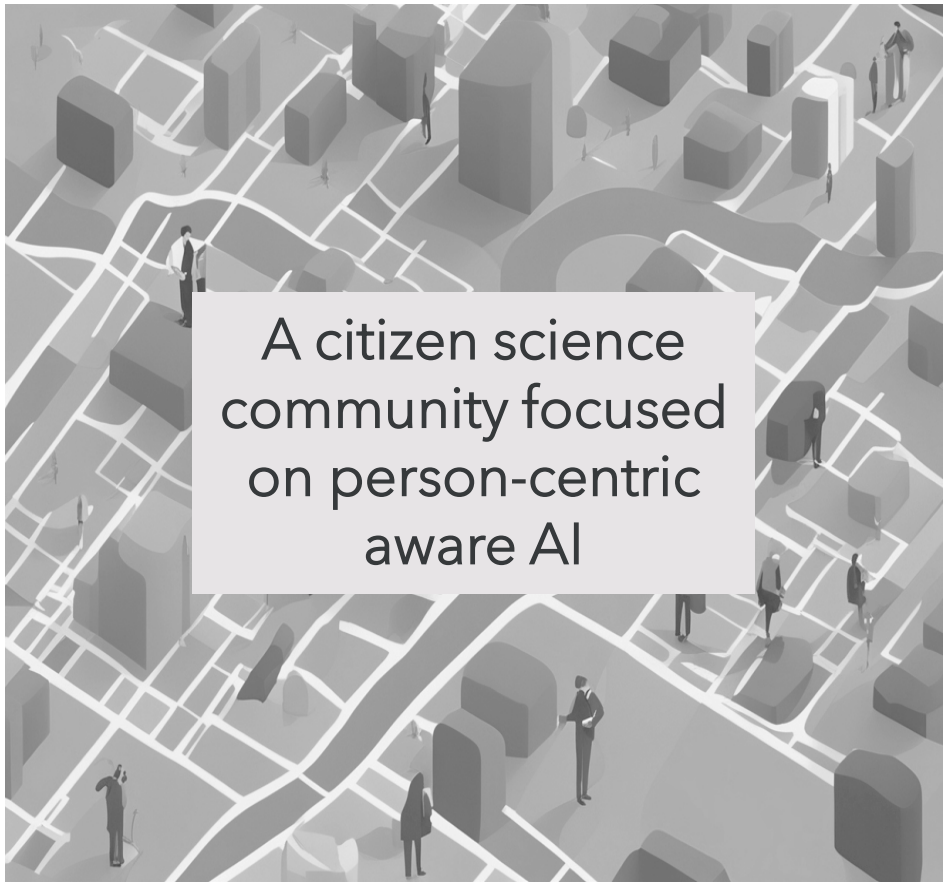
The DiversityOne ACM workshop / Hacathon web page

https://datascientiafoundation.github.io/diversityone-2025/

November 2026 @ Shanghai

© 2026 DataScientia – All rights reserved

Solution: The DataScientia Community



The DataScientia Community

A community-driven approach is essential to ensure that AI evolves with, not apart from, the people it impacts.

Why not joining the joint lab?

© 2026 DataScientia – All rights reserved

DataScientia
Unity Through Diversity

Thank you!

Questions?



© 2026 DataScientia – All rights reserved