A context model for generating diversity-aware data

DHHAI WORKSHOP
26th June 2023

Matteo Busso and Xiaoyue Li
University of Trento, Italy

A CONTEXT MODEL FOR GENERATING DIVERSITY-AWARE DATA

DHHAI WORKSHOP

Munich, 26th June 2023

MATTEO BUSSO and XIAOYUE LI

University of Trento, Italy

Index

DataScientia
Unitas per Varietatem

KNOW DIVE

1. **Introduction**
2. Current approaches to diversity aware data
3. The situational context: a hybrid human-AI approach
4. Methodology
5. A diversity-aware dataset
6. Use cases
7. Open issues

What do we mean by diversity-aware data?

1. Within person
   - Context
   - Point of view
2. Across people

What do we mean by diversity-aware data?

INTERPRETATION

1. Research
2. Development

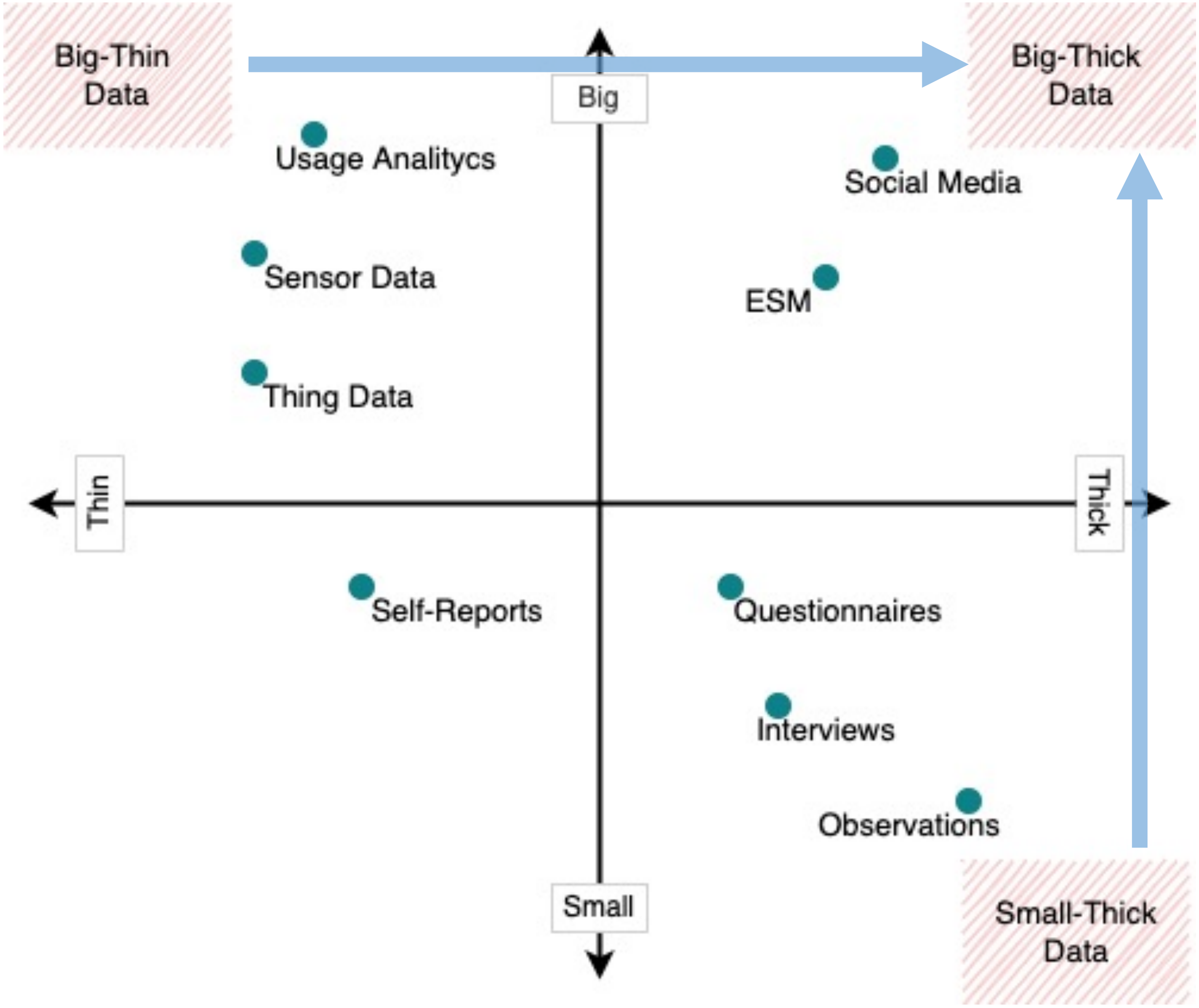
1. Introduction
2. **Current approaches to diversity aware data**
3. The situational context: a hybrid human-AI approach
4. Methodology
5. A diversity-aware dataset
6. Use cases
7. Open issues

What about Big Data?

- ‘With many interesting variables unavailable, **people are**, at best, **thinly described**’. (Blank, 2008: 540)

- Big Data are often used ‘**out of context**’, which decrease the ‘meaning and value’ (Boyd and Crawford, 2012: 670)

Big-Thick (Diversity-aware) Data



Adapted from:

- Gomez Ortega, A., et al. (2022)
- Tobias Bornakke and Brian L. Due. (2018)

Common approaches in generating Big (Thick) Data

Annotation

- HAR user diversity and of transfer learning (Fu et al, 2020),
- HITL (Wu et al., 2022),
- Context Recognition (Bontempelli et al. 2022)
- Health care (Vaizman, 2017)

Aggregation, fusion and integration

- User profiling and record linkage (Shu et al., 2017)

Blending

- Combining sensor data sources and ethnographic data (Bornakke, 2018)

Weakness

1. A domain expert is always required
2. Domain specific datasets, which are hardly reusable

DataScientia
Unitas per Varietatem

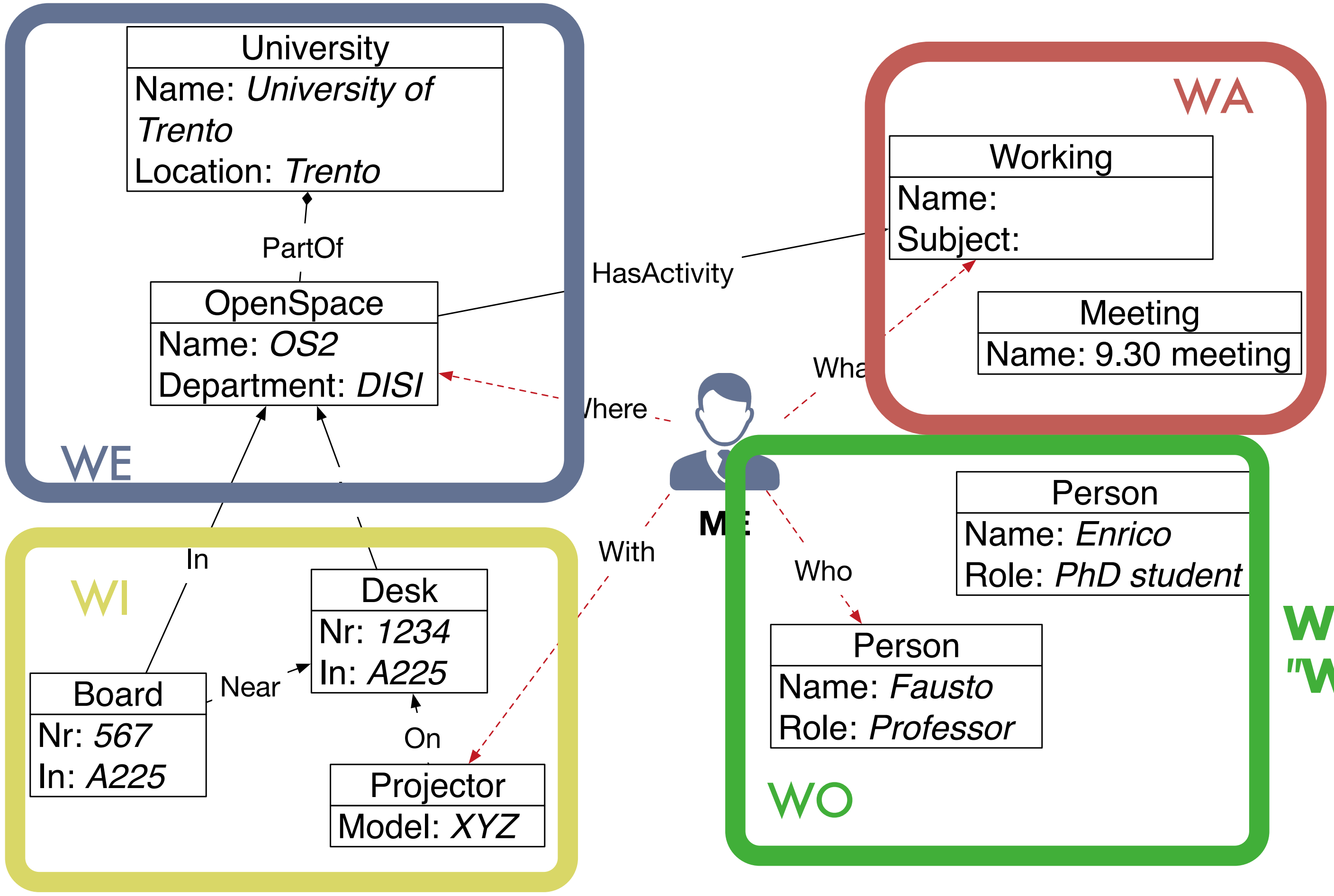
KNOW DIVE

1. Introduction
2. Current approaches to diversity aware data
3. **The situational context: a hybrid human-AI approach**
4. Methodology
5. A diversity-aware dataset
6. Use cases
7. Open issues

The situational context

Spatial cxt =
"WhEre are you?"

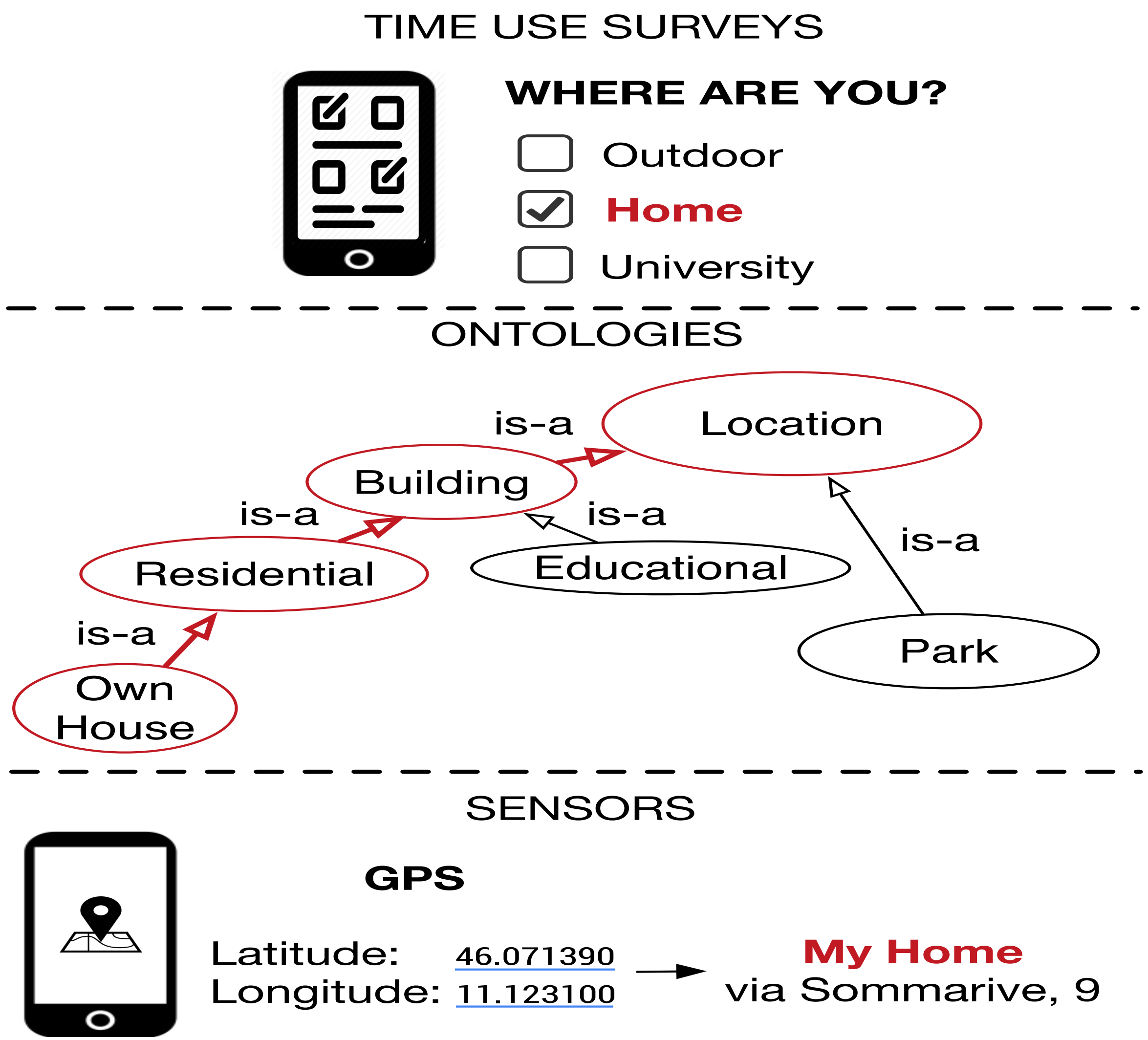
Object cxt = "What are you wIth?"



Temporal cxt =
"WhAt are you doing?"

WO = Social cxt =
"WhO are you with?"

A hybrid human-AI approach



User feedback

Knowledge Representation

Statistical analysis

+

machine learning

DataScientia
Unitas per Varietatem

KNOW
DIVE

Main issues in generating diversity-aware data

Design Human Machine interactions

- Avoiding bias and respondent burden
- Considering ethics & GDPR
- Avoiding Self-Selection Bias
- Maintaining motivation

Data collection

- Standards
- Collect active & passive data
- Consider Back-End & Front-End
- Protect Privacy

Data Sharing

- Findability, Accessibility, Interoperability, Reusability

Overall methodology

Design

1. Profiling questionnaire
2. Time Diaries
3. Sensor data
4. Sampling strategy & incentives

Ethics & Privacy

1. Data minimization process

Collection

1. App iLog
2. Timing
3. Helpdesk & Monitoring

Preparation & Distribution

1. Data preparation pipeline

Profiling Questionnaire

Psycho-social Profile

- Personality traits (MiniIPIP-10)
- Perceived stress scale (PSS)
- Irrational procrastination scale (IPS)
- Smartphone Addition Scale (SAV-SV)

Daily Routines

- Academic, Transport

DataScientia
Unitas per Varietatem

KNOW DIVE

Time diaries

A3. What are you doing?

Sleeping
Self-care
Eating
Study
Lesson
Social life
Watching YouTube Tv-shows etc.
Social media (Facebook Instagram etc.)
Travelling (*go to A3a*)
Coffee break cigarette beer etc.
Phone calling; in chat WhatsApp
Reading a book; listening to music
Movie Theatre Concert Exhibit ...
Housework
Shopping
Sport
Rest/nap
Hobbies
Work

A3a. How are you moving?

By subway
By car
By foot
By bike
By bus
By train
By motorbike
Other

A4. Where are you?

Home Apartment Room
Relatives Home
House (friends others)
Classroom / Laboratory
Classroom / Study hall
University Library
Other university place
Canteen
Other Library
Gym
Shop supermarket
Pizzeria pub bar restaurant
Movie Theatre Museum
Workplace
Other place
Outdoors

A5. With whom are you?

Alone
Friend(s)
Relative(s)
Classmate(s)
Roommate(s)
Colleague(s)
Partner
Other

A6. What is your mood?

1. 😃
2. 🙂
3. 😑
4. 😕
5. 🙁

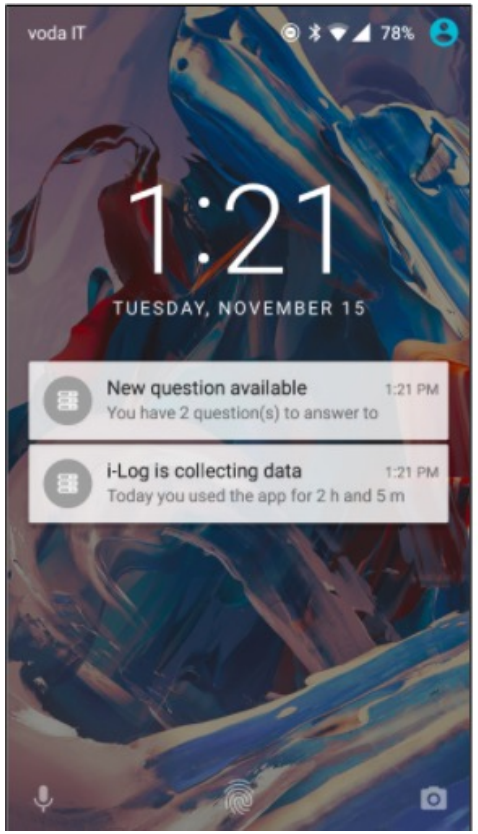
DataScientia
Unitas per Varietatem

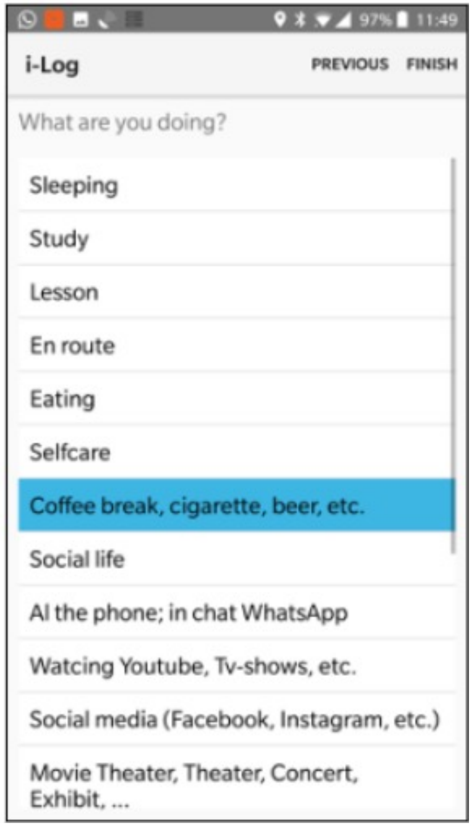
KNOW DIVE

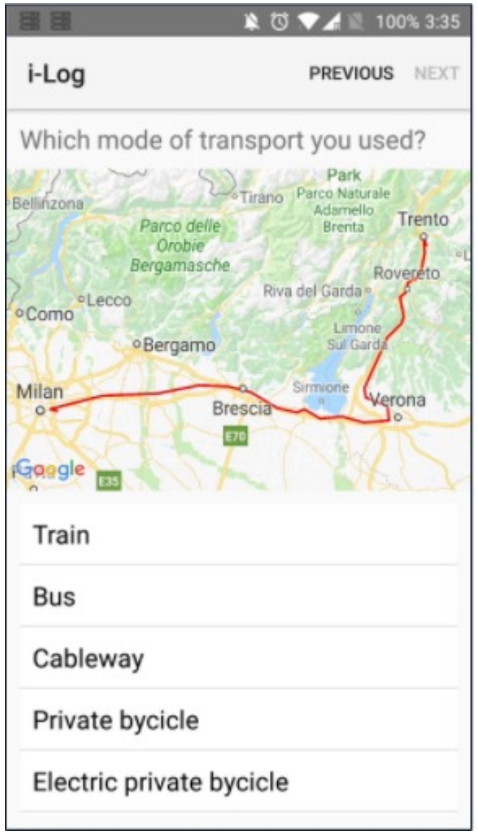
iLog App

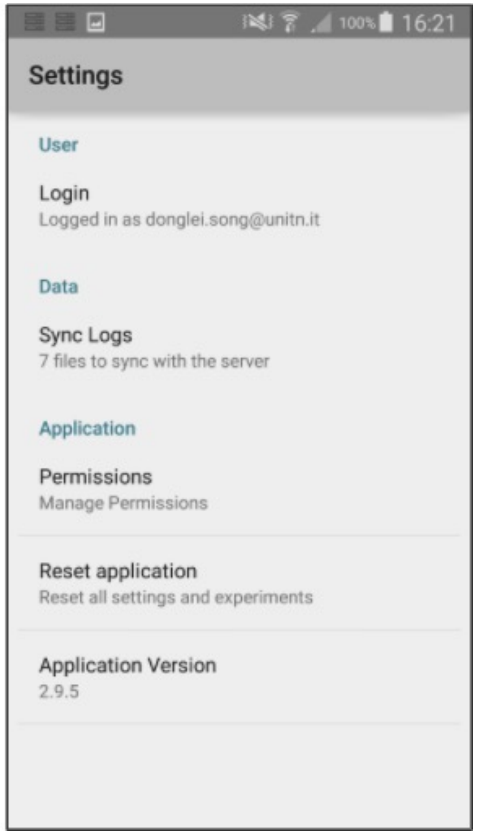
Available on the App Store

GET IT ON Google Play









Sensor data


Accelerometer
Linear Acceleration
Gyroscope
Gravity
Rotation Vector
Magnetic Field
Orientation
Ambient Temperature
Pressure
Relative Humidity
Proximity
Location
Wi-Fi
Bluetooth
Running Applications
Screen Status
Airplane Mode
Battery
Doze Mode
Headset Status
Ring Mode
Music Playback
Notifications
Touch Event
Cellular Network
Movement Activity
Step Counter
Light

Sampling strategy & incentives

10.000+ invitation questionnaire

350 selected participants

≈ 250 app download

≈ 158 active participants

20€ every 2 weeks

5€*5 participants everyday

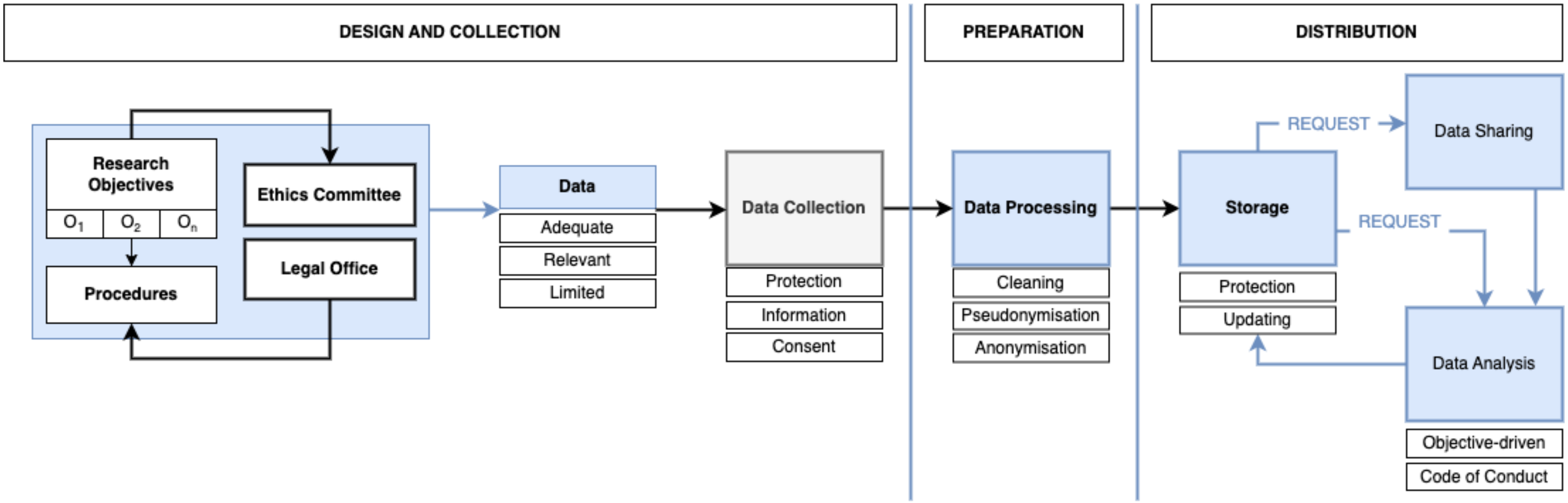
100*3 best participants (1^{st} weeks)

150€*3 best participants (2^{nd} weeks)

Helpdesk & Monitoring

- iLog app instructions
- Daily contact with participants (especially during registration period)
- Daily reports
- Target: non-missing data

GDPR & Data Minimization[1]



[1]The data minimisation principle is expressed in **Article 5(1)(c) of the GDPR and Article 4(1)(c) of Regulation (EU) 2018/1725**, which provide that personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed".

DataScientia
Unitas per Varietatem

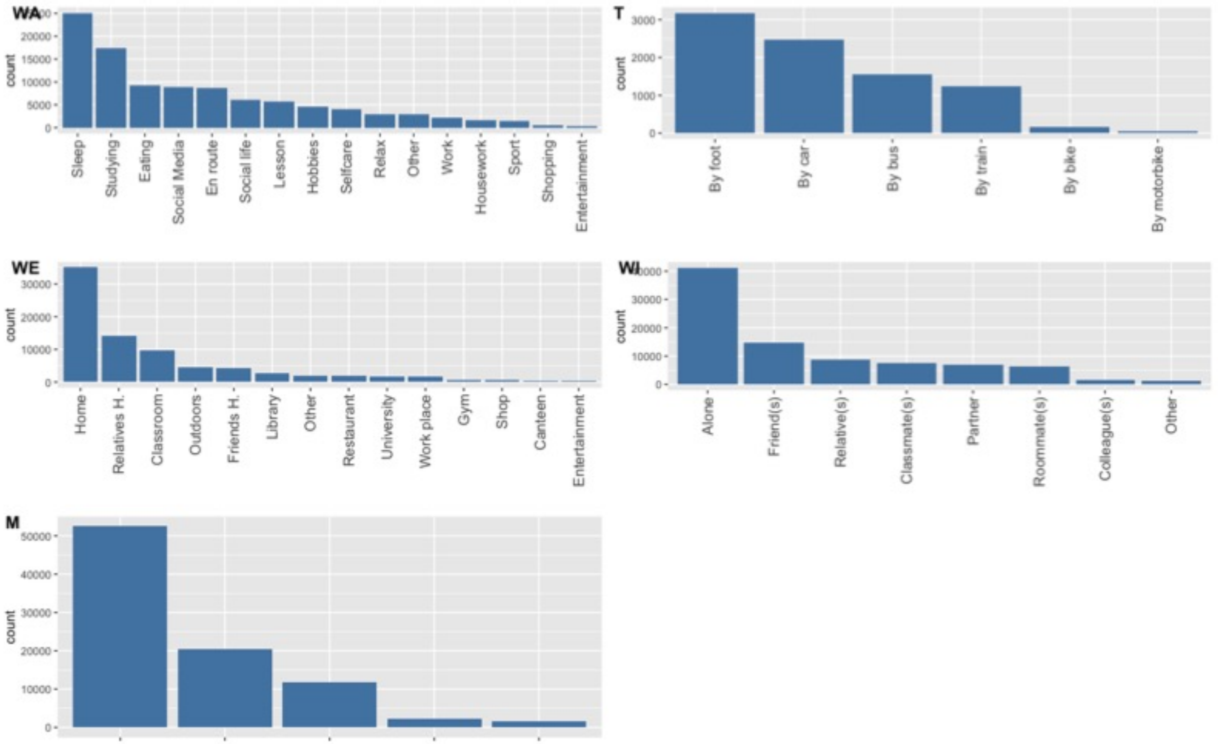
KNOW
DIVE

Questionnaire data

	%
Gender	
Female	48.7
Male	51.3
Age	
<22	47.5
22-26	52.5
Departments	
Hard Sciences	37.3
Soft Sciences	33.5
Humanities	29.2
Total	100.0 (N=158)
iLog Obs.	396-932

	mean	sd	range
Agreeableness	6.7	1.74	2-10
Conscientiousness	7.3	1.76	2-10
Extraversion	6.1	1.94	2-10
Neuroticism	6.6	2.17	2-10
Openness	6.9	1.92	2-10
Procrastination	22.7	6.25	10-40
Smartphone Addiction	27.5	17.6	0-93.3
Perceived Stress	43.3	17.4	7.5-85

SU2 – iLog data



No	HW Sensor	N. Obs.	Estimated Frequency	U.M.
1	Position	4.144.214	Once every minute	degrees, minutes, seconds
2	POI	1.583.389	Once every 5 minutes	Unitless

No	SW Sensor	N. Obs.	Estimated Frequency	U.M.
3	Audio mode [Silent/Normal]	2.042.901	On change	Unitless
4	Battery Charge [ON/OFF]	42.664	On change	0/1
5	Battery Level	26.934	On change	%
6	Doze Modality [ON/OFF]	11.914	On change	0/1
7	Flight Mode [ON/OFF]	2.567	On change	0/1
8	Headset plugged in [ON/OFF]	171.677	On change	0/1
9	Music Playback (no track information) [ON/OFF]	92.510	On change	0/1
10	Notifications received	3.224.577	On change	Unitless
11	Proximity	11.405.724	On change	0/1
12	Running Application	35.184.768	Once every 5 seconds	Unitless
13	Screen Status [ON/OFF]	1.252.576	On change	0/1
14	WIFI Network Connected to	747.366	On change	Unitless
15	WIFI Networks Available	2.859.187	Once every minute	Unitless

The dataset catalog

Extending the dataset

LiveData Trentino

1. Introduction
2. Current approaches to diversity aware data
3. The situational context: a hybrid human-AI approach
4. Methodology
5. A diversity-aware dataset
6. **Use cases**
7. Open issues

DataScientia
Unitas per Varietatem

KNOW
DIVE

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

ELSEVIER

Full length article

Mobile social media usage and academic performance

Fausto Giunchiglia [a], Mattia Zeni [a], Elisa Gobbi [b], Enrico Bignotti [a,*], Ivano Bison [b]

[a] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive 9, 38123, Trento, Italy
[b] Department of Sociology and Social Research, University of Trento, Via Verdi 26, 38123, Trento, Italy

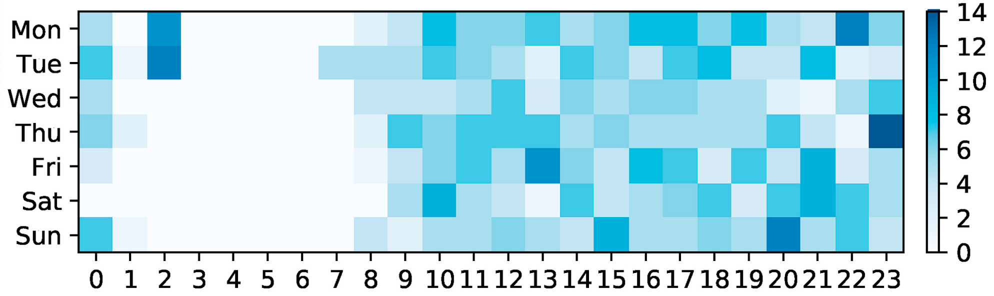
A R T I C L E I N F O

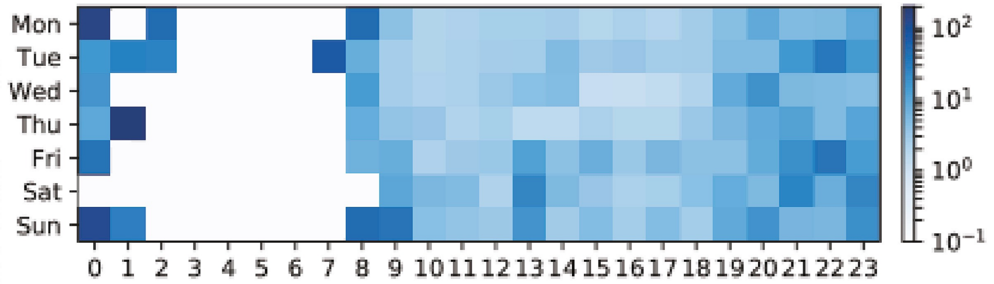
Article history:
Received 5 October 2017
Received in revised form
19 December 2017
Accepted 29 December 2017
Available online 5 January 2018

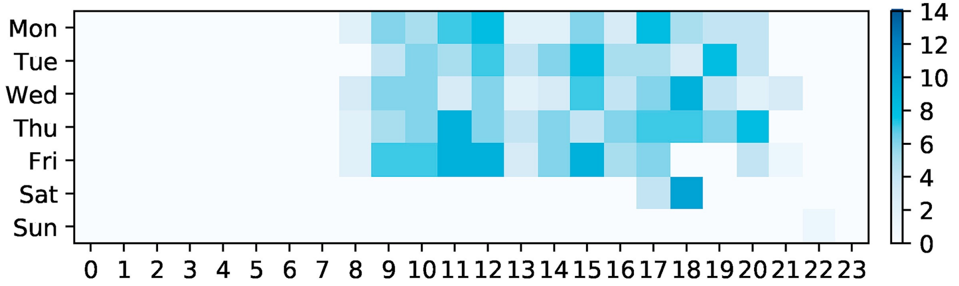
Keywords:
Social media
Academic performance
Smartphone
Time diaries

A B S T R A C T

Among the general population, students are especially sensitive to social media and smartphones because of their pervasiveness. Several studies have shown that there is a negative correlation between social media and academic performance since they can lead to behaviors that hurt students' careers, e.g., addictedness. However, these studies either focus on smartphones and social media addictedness or rely on surveys, which only provide approximate estimates. We propose to bridge this gap by *i*) parametrizing social media usage and academic performance, and *ii*) combining smartphones and time diaries to keep track of users' activities and their smartphone interaction. We apply our solution on the 72 students participating in the SmartUnitn project, which investigates students' time management and their academic performance. By analyzing the logs of social media apps on students' smartphones and by comparing them to students' credits and grades, we can provide a quantitative and qualitative estimate of negative and positive correlations. Our results show the negative impact of social media usage, distinguishing different influence patterns of social media on academic activities and also underline the need to control the smartphone usage in academic settings.







Zhang et al. *EPJ Data Science* (2021) 10:42
https://doi.org/10.1140/epjds/s13688-021-00299-2
EPJ.org
EPJ Data Science
a SpringerOpen Journal
REGULAR ARTICLE Open Access

Putting human behavior predictability in context

Wanyi Zhang[1*], Qiang Shen[3], Stefano Teso[1], Bruno Lepri[4], Andrea Passerini[1], Ivano Bison[2] and Fausto Giunchiglia[1,3]

*Correspondence: wanyi.zhang@unitn.it
[1]Department of Information Engineering and Computer Science, University of Trento, Trento, Italy
Full list of author information is available at the end of the article

Abstract

Various studies have investigated the predictability of different aspects of human behavior such as mobility patterns, social interactions, and shopping and online behaviors. However, the existing researches have been often limited to a single or to the combination of few behavioral dimensions, and they have adopted the perspective of an outside observer who is unaware of the motivations behind the specific behaviors or activities of a given individual. The key assumption of this work is that human behavior is deliberated based on an individual's own perception of the situation that s/he is in, and that therefore it should also be studied under the same perspective. Taking inspiration from works in ubiquitous and context-aware computing, we investigate the role played by four contextual dimensions (or modalities), namely time, location, activity being carried out, and social ties, on the predictability of individuals' behaviors, using a month of collected mobile phone sensor readings and self-reported annotations about these contextual modalities from more than two hundred study participants. Our analysis shows that any target modality (e.g. location) becomes substantially more predictable when information about the other modalities (time, activity, social ties) is made available. Multi-modality turns out to be in some sense fundamental, as some values (e.g. specific activities like "shopping") are nearly impossible to guess correctly unless the other modalities are known. Subjectivity also has a substantial impact on predictability. A location recognition experiment suggests that subjective location annotations convey more information about activity and social ties than objective information derived from GPS measurements. We conclude the paper by analyzing how the identified contextual modalities allow to compute the diversity of personal behavior, where we show that individuals are more easily identified by rarer, rather than frequent, context annotations. These results offer support in favor of developing innovative computational models of human behaviors enriched by a characterization of the context of a given behavior.

Keywords: Human behavior; Personal context; Predictability; Diversity

KNOW DIVE

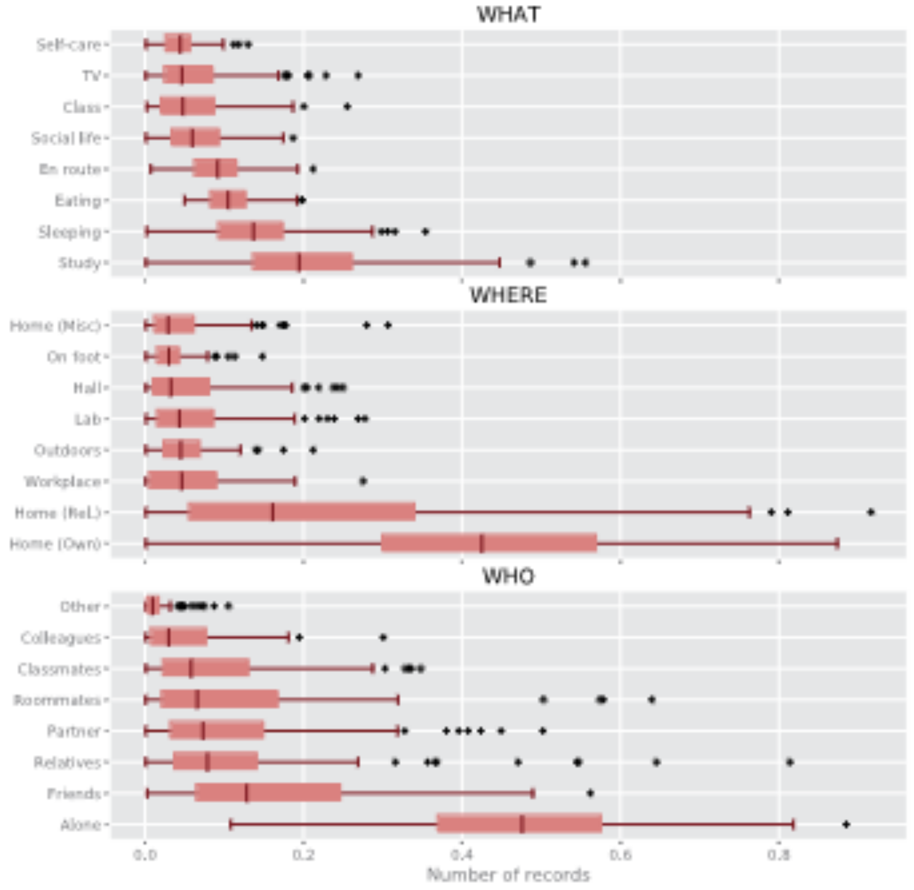


Figure 4 Value distribution of different aspects. From top to bottom: WA, WE, and WO. Only the eight most frequent annotations are shown for each aspect. The boxes extend from the 1st to the 3rd quartiles, while the bars extend to ±1.5 inter-quartile range from the median. Study participants with very high annotation frequency (i.e. outliers) are denoted by crosses

Selected publications

1. Zeni, M., Zaihrayeu, I., Giunchiglia, F.: Multi-device activity logging. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. pp. 299–302 (2014)
2. Giunchiglia, F., Bignotti, E., Zeni, M.: Human-like context sensing for robot surveillance. International Journa“l of Semantic Computing 12(01), 129–148 (2017)
3. Giunchiglia, F., Bignotti, E., Zeni, M.: Personal context modelling and annotation. In: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 117–122. IEEE (2017)
4. Giunchiglia, F., Zeni, M., Gobbi, E., Bignotti, E., Bison, I.: Mobile social media and academic performance. In: International conference on social informatics. pp. 3–13. Springer, Cham (2017)
5. Giunchiglia, F., Zeni, M., Big, E.: Personal context recognition via reliable human-machine collaboration. In: 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). pp. 379–384. IEEE (2018)
6. Zhang, Wanyi, Qiang Shen, Stefano Teso, Bruno Lepri, Andrea Passerini, Ivano Bison, and Fausto Giunchiglia. "Putting human behavior predictability in context." EPJ Data Science 10, no. 1 (2021)
7. Schelenz, Laura, Ivano Bison, Matteo Busso, Amalia De Götzen, Daniel Gatica-Perez, Fausto Giunchiglia, Lakmal Meegahapola, and Salvador Ruiz-Correa. "The theory, practice, and ethical challenges of designing a diversity-aware platform for social relations." In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 905-915. (2021)
8. Li, Xiaoyue, Marcelo Rodas-Britez, Matteo Busso, and Fausto Giunchiglia. "Representing Habits as Streams of Situational Contexts." In International Conference on Advanced Information Systems Engineering, pp. 86-92. Springer, Cham, (2022)

DataScientia
Unitas per Varietatem

KNOW
DIVE

1. Introduction
2. Current approaches to diversity aware data
3. The situational context: a hybrid human-AI approach
4. Methodology
5. A diversity-aware dataset
6. Use cases
7. **Open issues**

Open issues

1. What could be other diversity aware experiments?
2. What about interaction with sensor data analytics in streaming?
3. What about mapping time sensitive information within dataset extension?
4. Other possible extension?

DataScientia
Unitas per Varietatem

KNOW DIVE

Links and contacts

- http://knowdive.disi.unitn.it/
- http://datascientia.disi.unitn.it/
- https://lp.datascientia.eu

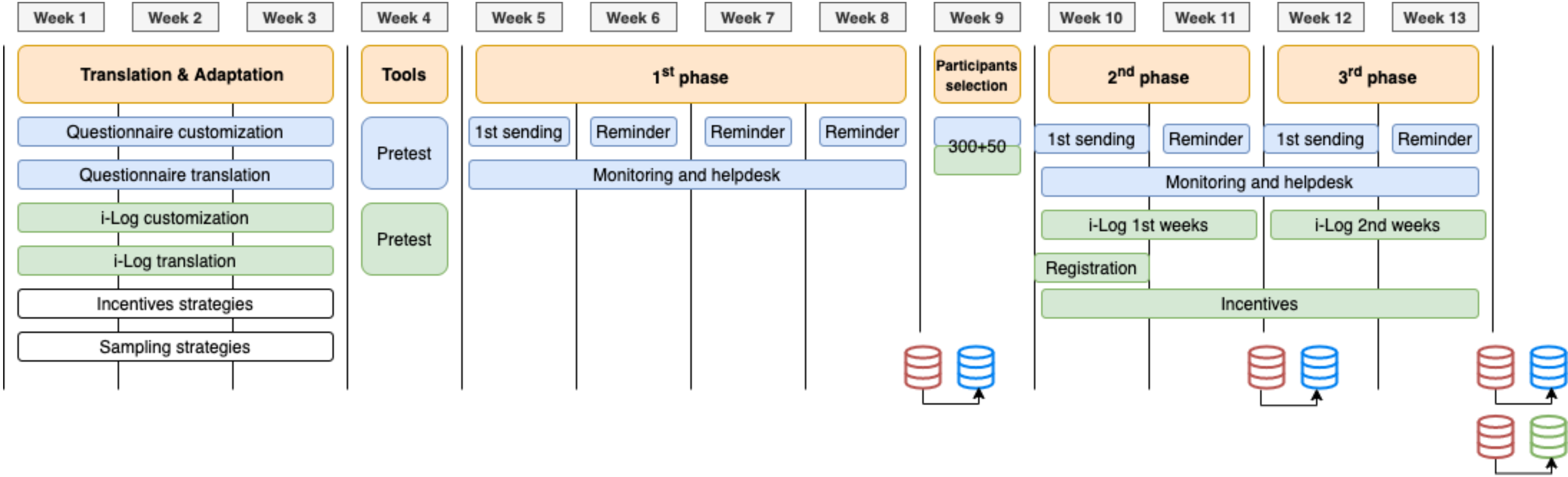
- @knowdive

- matteo.busso@unitn.it

THANK YOU!

Any questions?

Timing



Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Week 11
Week 12
Week 13
Translation & Adaptation
Questionnaire customization
Questionnaire translation
i-Log customization
i-Log translation
Incentives strategies
Sampling strategies
Tools
Pretest
Pretest
1st phase
1st sending
Reminder
Reminder
Reminder
Monitoring and helpdesk
Participants selection
300+50
2nd phase
1st sending
Reminder
3rd phase
1st sending
Reminder
Monitoring and helpdesk
i-Log 1st weeks
i-Log 2nd weeks
Registration
Incentives

DataScientia
Unitas per Varietatem

KNOW DIVE

Year	Project	Data Collection	Tools	Total Participants	iLog Participants	Country	Measurement
2011-2015	iLog testing	.	.	.	.	.	.
2016	Smart University	SU1	LS, iLog	72	72	1	Questionnaire, Time Diaries, Sensors
2018		SU2	LS, iLog	184	158	1	Questionnaire, Time Diaries, Sensors
2019	QROWD	QR '19	iLog	40	21	1	Images, Sensors
2019	EUROSTAT	EUR'19	iLog	100	100	1	Questionnaire, Time Diaries, Sensors
2020	DOXA	DX'20	iLog	10	10	1	Questionnaire, Time Diaries, Sensors
2020	KnowDive	Indoor HAR	iLog, SW	5	5	1	Questions, Sensors
2020	WeNet	DIV1	LS, iLog	21476	784	8	Questionnaire, Time Diaries, Sensors
2022		DIV2	LS, iLog	586	248	3	Questionnaire, Time Diaries, Sensors
2021		CH1	LS, Chat, iLog	.	195	5	Questionnaire, Chatbot, Sensors
2022		CH2	LS, Chat, iLog	158	158	5	Questionnaire, Chatbot, Sensors
2023		CH3	Chat, iLog	.	50	4	Quest., Chatbot, Time Diaries, Sensors
2022	WeNet Open Calls	UTH	LS, iLog	310	141	1	Questionnaire, Time Diaries, Sensors
2022		FTP	LS, iLog	117	117	1	Questionnaire, Time Diaries, Sensors
2023	DatiMeteoX	DMX	iLog	12	12	1	Images, Questions, Sensors
MAY 2023	KnowDive	SKEL	iLog	100	100	1	AR, Questions, Sensors
SEPT 2023	Makerere University	MAK	iLog	100	100	1	Questionnaire, Time Diaries, Sensors
TOTAL				**23270**	**2271**	**12**	