

# Università di Trento

Dipartimento di Ingegneria e Scienza dell'Informazione

Corso di Laurea in Informatica



## Un'infrastruttura di ricerca per lo studio di comportamenti umani

Fausto Giunchiglia

Daniele Marcolan

# Il progetto WeNet

## WeNet – Internet of Us

Una piattaforma per interazioni sociali basata su un'infrastruttura di ricerca consapevole delle diversità fra le persone.



# Approccio

## Approccio etico e partecipativo

Volontà di superare le problematiche delle piattaforme di social media esistenti, riguardo la gestione dei dati personali e delle diversità esistenti fra gli utenti.

I dati personali vengono utilizzati per la costruzione di algoritmi che rispondono alle diversità degli utenti, aiutandoli a sfruttarle reciprocamente per il loro miglioramento personale.

Gli utenti partecipano volontariamente e attivamente alla raccolta dati. La persona e la protezione della sua privacy sono al centro del progetto.

# Obiettivi

- Sviluppare una teoria sociologica sulla diversità.
- Sviluppare cultura, scienza, metodi ingegneristici, algoritmi e protocolli di comunicazione per potenziare l'interazione digitale fra le persone.
- Creare un medium che favorisca l'interazione e il mutuo supporto fra le persone e ne migliori la qualità della vita.

# Il modello standard

## Un modello basato sui dati

Il fondamento del progetto WeNet è la creazione di un'infrastruttura di ricerca, atta a raccogliere, gestire e conservare i dati personali dei partecipanti, in accordo con il GDPR.

I dati vengono raccolti in una serie di esperimenti pilota svolti in tutto il mondo con studenti universitari, direttamente dai loro smartphone, tramite iLog (un software sviluppato da WeNet), e dei questionari su LimeSurvey.

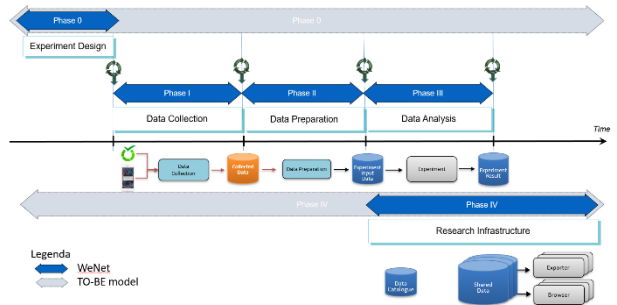
# Principi etici e legali del GDPR

- Liceità, correttezza, trasparenza.
- Limitazione delle finalità.
- Minimizzazione dei dati.
- Esattezza dei dati.
- Limitazione della conservazione dei dati.
- Integrità, protezione e riservatezza.

# Flusso e gestione dei dati

Cinque passaggi fondamentali:

- Collezione dei dati.
- Preparazione dei dati.
- Analisi dei dati in locale.
- Condivisione dei dati attraverso l'infrastruttura di ricerca.
- Riutilizzo dei dati e trasferimento alla comunità scientifica attraverso l'infrastruttura di ricerca.



# Collezione dei dati

I dati vengono collezionati nella vita quotidiana degli studenti, tramite iLog, che raccoglie dati da sensori e geo-localizzazione di uno smartphone, e questionari periodici su LimeSurvey compilati dal partecipante.

Dato che gli esperimenti pilota si svolgono in varie parti del mondo, l'aspettativa è che i dati mostrino le diverse abitudini e i diversi stili di vita degli studenti nelle varie località.



# Preparazione dei dati

Tre operazioni da attuare sui dati:

- Pulizia.
- Pseudo-anonimizzazione.
- Anonimizzazione.

L'obiettivo è operare in conformità al principio del GDPR di minimizzazione dei dati, eliminando quelli superflui raccolti e anonimizzando i restanti.

# Conservazione dei dati

## Modello FAIR

- Findable, per persone e computer, tramite open data repositories.
- Accessible, tramite licenze permissive.
- Interoperable, tramite metadati e meccanismi di integrazione.
- Reusable, tramite dataset di informazioni aggregate.

# Esperimenti pilota

## Obiettivo

Osservare come la gestione e l'organizzazione del tempo nella vita quotidiana di studenti influenzi le loro prestazioni universitarie; l'ipotesi è che abbia un impatto significativo.

Gli studenti appartengono a diversi anni accademici, dipartimenti, generi e classi sociali.

## Processo

Campagna di raccolta dati lunga due settimane.

Dati da sensori dello smartphone costantemente raccolti in background da iLog, questionari di LimeSurvey una volta ogni 30 minuti.

Firma del consenso informato da parte dei partecipanti.

# Raccolta dati con iLog e LimeSurvey



The screenshot shows the iLog application interface on a mobile device. At the top, there's a status bar with icons for USB, Wi-Fi, and battery level (11%), along with the time 16:26. Below the status bar is the iLog logo and a table with four columns: Source, Memory, DB, and Server. The table is divided into several sections: Motion Sensors, Position Sensors, Environment Sensors, Location Detectors, Ambience sensors, Meta log, and TOTALS. Each section lists various sensors and their corresponding data values, mostly marked as N/A.

Source	Memory	DB	Server
<b>Motion Sensors</b>			
Accelerometer	N/A	N/A	N/A
Gravity	N/A	N/A	N/A
Gyroscope	N/A	N/A	N/A
Linear Accel.	N/A	N/A	N/A
Rotation Vec.	N/A	N/A	N/A
<b>Position Sensors</b>			
Magnetic field	N/A	N/A	N/A
Orientation	N/A	N/A	N/A
Proximity	N/A	N/A	N/A
<b>Environment Sensors</b>			
Ambient temp.	N/A	N/A	N/A
Light	N/A	N/A	N/A
Pressure	N/A	N/A	N/A
Humidity	N/A	N/A	N/A
Device temp.	N/A	N/A	N/A
<b>Location Detectors</b>			
GPS locations	N/A	N/A	N/A
Network locations	N/A	N/A	N/A
<b>Ambience sensors</b>			
WiFi networks	N/A	N/A	N/A
<b>Meta log</b>			
Log monitor	0	N/A	N/A
Battery monitor	0	N/A	N/A
<b>TOTALS</b>	<b>N/A</b>	<b>N/A</b>	<b>N/A</b>
Log files to sync:		N/A	
Size of log files to sync:		N/A	

What are you doing?	Where are you?	Who is with you?
Lesson	Class	Alone
Study	Study Hall	Classmate(s)
Eating	Library	Friend(s)
Personal Care	Other University place	Roommate(s)
En route (*)	Canteen	Partner(s)
Social life	Bar/ Pub/etc...	Relative(s)
Social media & internet	Home	Colleague(s)
Cultural Activity	Other Home	Other
Sport	Workplace	
Shopping	Outdoors	
Hobbies	Gym	
Other Free Time	Shop	
Work	Other Place	
Housework	(*) How are you travelling?	
Volunteering	By Foot	
Other	By Bus	
	By Train	
	By Car	
	By Motorbike	
	By Bike	
	Other	

# Anonimizzazione dei dati raccolti

## User ID

1. Tramite una funzione hash computazionalmente difficile da invertire.
2. Tramite cifratura con chiave segreta, poi rimossa dai datasets costruiti.

## GPS

1. Tramite rounddown, troncando la precisione di latitudine e longitudine.
2. Tramite trasformazione in Point of Interest, corrispondente ad un'area generica e le sue vicinanze.

# Gestione dei datasets

Vengono generati due datasets principali: uno completamente anonimizzato e uno contenente tutti i dati personali.

I dati personali sono mantenuti per 5 anni e non vengono condivisi o resi accessibili tramite open repositories, mentre quelli anonimizzati sono resi disponibili all'interno dell'infrastruttura di ricerca.

Ogni partecipante può richiedere la totale eliminazione dei suoi dati personali in qualunque momento, in accordo con il GDPR.

# Università di Trento

Dipartimento di Ingegneria e Scienza dell'Informazione

Corso di Laurea in Informatica



## Grazie per l'attenzione.

Fausto Giunchiglia

Daniele Marcolan