



SAPIENZA
UNIVERSITÀ DI ROMA

FACOLTÀ DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA DELLE COMUNICAZIONI

Tesi di Laurea

**Evoluzione della piattaforma Telecom
Italia ITS 2.0: applicazione di strumenti
di Data Analysis e sviluppo applicazione
Android**

Laureando

Davide Palma

Relatore

Prof.ssa M.G. Di Benedetto

Correlatore

Dott. Davide Pellegrino

Anno Accademico 2014/2015

ABSTRACT

Dopo un'approfondita analisi dell' As-Is della piattaforma di Telecom Italia "ITS 2.0", a valle dei limiti emersi, si sono adottate soluzioni per la sua evoluzione. Tali soluzioni hanno richiesto l'utilizzo di strumenti di Data Analysis e lo sviluppo di un'interfaccia front-end tramite un'applicazione Android.

INDICE

Abstract	iii
Elenco delle figure	xii
Introduzione	1
1 Intelligent Transport System	3
1.1 Contesto	4
1.2 Che cosa è l'Intelligent Transport System	6
1.3 I protocolli TTI (Traffic and Travel Information)	9
1.3.1 Ertico	9
1.3.2 Traveller Information Service Association (TISA) . . .	15
1.3.3 Traffic Message Channel (TMC)	18
1.3.4 Transport Protocol Experts Group (TPEG)	20
1.4 Il modello dati DATEX II	41
1.4.1 Lista e descrizione delle principale aree di utilizzo . .	43
1.4.2 Il contesto	43
1.4.3 Caratteristiche principali	44
1.4.4 Il dizionario dei dati	46
1.4.5 Lo schema XML	46
1.4.6 Meccanismi di scambio	46

1.4.7	Profili DATEX II	47
1.4.8	I documenti di DATEX II	48
1.4.9	Lo scambio dei dati: elementi di base	49
1.4.10	UML data model	61
1.4.11	Specifiche sullo scambio di informazioni	62
1.4.12	Gli accordi sullo scambio	63
1.5	Il ruolo di Telecom Italia	65
1.5.1	MOBiNET	65
1.5.2	TEAM	66
1.5.3	COMPASS4D	67
2	Telecom Italia: la piattaforma "ITS 2.0"	69
2.1	Architettura piattaforma	70
2.2	Identity Manger	71
2.2.1	Autenticazione e autorizzazione	72
2.2.2	Ruoli	74
2.2.3	Struttura implementazione moduli IdM	76
2.2.4	Modulo di Identity Management	77
2.2.5	Autorizzazione basata su client credentials	78
2.3	Meccanismi di Validation	86
2.4	Meccanismi di Reputation	87
2.4.1	Reputation increase	88
2.4.2	Reputation decrease	88
2.5	L'aggregatore Tweet	89
2.5.1	Gli open data	90
2.5.2	Gli account Twitter dedicati all'infomobilità	91
2.5.3	La parsificazione, data cleansing e data validation	96
2.5.4	Integrazione con la piattaforma "ITS 2.0"	99
3	Analisi Dati	103
3.1	Anomalie Database	104
3.1.1	Incoerenza tra attributi e valori	104
3.1.2	Incoerenza tra i valori di uno stesso record	106

3.1.3	Incongruenza del formato dei valori di uno stesso attributo in record diversi	107
3.1.4	Disomogeneità nelle nomenclature per uno stesso oggetto	107
3.1.5	Anomalie nella georeferenziazione delle segnalazioni	108
3.2	Data Cleansing	111
3.2.1	L'approccio al Data Cleansing	111
3.2.2	L'algoritmo implementato	113
3.3	Data Mining	118
3.3.1	Tecniche di Data Mining	120
3.3.2	Strategie di Data Mining	120
3.3.3	Apprendimento supervisionato	121
3.3.4	L'algoritmo implementato	129
4	Applicazione Android	139
4.1	Casi d'uso	140
4.2	Nome e Logo Applicazione	142
4.3	Descrizione funzionalità	143
4.3.1	Login tramite credenziali fornite da Identity Manager	144
4.3.2	Visualizzazione eventi di infomobilità	144
4.3.3	Invio segnalazioni	146
4.3.4	Validazione degli eventi	150
4.3.5	Navigatore e Lista Eventi lungo il percorso	150
5	Conclusioni	155
	Bibliografia	157

ELENCO DELLE FIGURE

1.1	Modelli per la creazione di contenuto ed il relativo uso	8
1.2	Logo ERTICO	10
1.3	Logo TISA	15
1.4	TISA	16
1.5	Logo TMC	18
1.6	Logo TPEG	21
1.7	Content and Delivery Segment	25
1.8	Tabella TPEG	32
1.9	Tabella delle traduzioni	33
1.10	Struttura del messaggio TPEG	35
1.11	Rappresentazione ISO/OSI del protocollo TPEG	37
1.12	Layout idealizzato	38
1.13	Concettualizzazione del TPEG come protocollo a livelli	39
1.14	Visualizzazione TPEG a tre contenitori	39
1.15	Abbreviazioni e applicazioni citate	40
1.16	Struttura schema Datex II	47
1.17	Elementi inclusi nelle pubblicazioni	52
1.18	Architettura funzionale MOBiNET	65
1.19	Ecosistema TEAM	66
1.20	Allestimento pilota COMPASS4D	67

2.1	Architettura Funzionale della Piattaforma ITS 2.0	71
2.2	Caratteristiche autenticazione	77
2.3	Struttura piattaforma IdM	77
2.4	Moduli IdM	78
2.5	Modulo IdM	79
2.6	OAuth 2.0	82
2.7	Differenze nel formato dei tweet	92
2.8	Applicazioni verticali dell'aggregazione dei tweet	93
2.9	Esempi tweet di infomobilità	93
2.10	Struttura modello dati	94
2.11	Estensione del Modello dati	96
2.12	Tweet d'esempio	97
2.13	Tweet ANAS	98
2.14	Tweet CCISS	98
2.15	Processo di aggregazione e inserimento in ITS 2.0	100
3.1	Processo di analisi Dati	103
3.2	Incongruenze tra gli attributi city e street e i corrispondenti valori	105
3.3	Incongruenze tra l'attributo direction e i corrispondenti valori	105
3.4	Incongruenza tra l'attributo <i>road_code</i> e i corrispondenti valori	106
3.5	Incoerenze tra i valori di uno stesso record	106
3.6	Formato differente per l'oggetto A4	107
3.7	Nomenclature diverse per Luceverde Roma e Luceverde Regione Lazio	108
3.8	Rappresentazione sul piano cartesiano dei valori di latitudine e longitudine delle segnalazioni interne al perimetro italiano	109
3.9	Rappresentazione sul piano cartesiano dei valori di latitudine e longitudine delle segnalazioni	110
3.10	Record da correggere	115
3.11	Risposta Reverse Geocoding Google API	116
3.12	Risposta Geocoding Google API	117

3.13	Record corretto	118
3.14	Esempio di calcolo della matrice delle distanze	131
3.15	Selezione eventi potenziali	133
3.16	Cluster individuati in maniera iterativa	134
3.17	Individuazione del valore di h	135
3.18	Individuazione dei cluster i cui elementi distano in media di h	135
3.19	Cluster selezionati	136
3.20	Clusterizzazione degli eventi	137
4.1	Modalità d'uso previste per l'applicazione	141
4.2	Esempi di Mockup di interfacce grafiche dell'applicazione: App Home, Navigation Drawer, Modalità Navigazione . . .	142
4.3	Logo App TIMobility	143
4.4	Schermata di Login	144
4.5	Menu principale	144
4.6	Eventi di infomobilità reperiti real time dal database	145
4.7	Infowindow di un marker relativo ad un evento	145
4.8	Categorie degli eventi	146
4.9	Selezione categoria Traffico	147
4.10	Selezione categoria Pericolo	147
4.11	Selezione categoria Meteo	148
4.12	Selezione categoria Chiusura	148
4.13	Selezione categoria Cantieri	148
4.14	Caso d'uso: Chiusura-Strada-Loc. manuale	148
4.15	Invio della segnalazione tramite comandi vocali: alla pressione del tasto di aggiunta, viene mostrato il pop-up di scelta della categoria e a seguire, dopo la pressione di una delle voci previste, il tab di sotto-categoria associato alla categoria selezionata inizialmente	149
4.16	Infowindow di un evento segnalato da un utente (con i caratteristici indicatori di validazioni), dialog attivato al tap dell'infowindow, indicazione di avvenuta validazione	151

4.17 Inserimento Partenza e Arrivo con funzionalità di rilevamento della posizione	152
4.18 Indicazioni fornite in riferimento all'itinerario richiesto . . .	152
4.19 Visualizzazione del percorso calcolato	153
4.20 Lista eventi attivi lungo il percorso	153
4.21 Punto di partenza con suggerimento visualizzato	153
4.22 Dettaglio infowindow di evento lungo il percorso	153

INTRODUZIONE

Il presente lavoro di tesi si pone l'obiettivo di far evolvere la piattaforma di Telecom Italia "ITS 2.0" mediante strumenti di Data Analysis e lo sviluppo di una applicazione Android in grado di interfacciarsi con gli utenti finali.

Il *Capitolo 1* tratta del contesto ITS (Intelligent Transport System) ovvero sistemi di trasporto intelligenti con l'obiettivo, non solo di ridurre il traffico nelle reti stradali, ma anche di ridurre l'impatto ambientale e aumentare la sicurezza riducendo i tempi di soccorso in caso di necessità. Verranno, poi illustrati i protocolli TTI (Traffic and Travel Information) che regolamentano la diffusione delle informazioni circa la mobilità, in particolare TMC e TPEG. In seguito si parlerà del modello dati DATEX II che è lo standard utilizzato per la gestione della persistenza dei dati di infomobilità. Il capitolo si conclude con i progetti europei (Team, Compass4D, Mobinet) che Telecom Italia ha attuato in ambito ITS che hanno posto le basi, sia tecniche che esperienziali, da parte di Telecom Italia per poter sviluppare la piattaforma sul quale si è basato il lavoro di tesi.

Il *Capitolo 2* approfondisce lo scenario As-Is della piattaforma di Telecom Italia, analizzandola mediante un approccio di sistema.

Il *Capitolo 3* descrive le soluzioni di Data Cleansing adottate per pulire il database, rendendo fruibili le informazioni al suo interno, e quelle di Data Mining al fine di generare nuova conoscenza.

Il *Capitolo 4* è dedicato alla progettazione e sviluppo dell'applicazione Android, che permetterà agli utenti finali di fruire dei servizi della piattaforma. Il tutto andando a soddisfare i requisiti forniti da Telecom Italia come ad esempio l'utilizzo di comandi vocali volti ad agevolare la user experience dell'utente finale.

Il *Capitolo 5* espone le conclusioni del progetto nonché i benefici attesi di Telecom Italia.

CAPITOLO 1

INTELLIGENT TRANSPORT SYSTEM

È ampiamente condiviso che nel 2025 ci saranno nel mondo almeno 30 città con più di 10 milioni di abitanti e che in Europa nel 2050 la quota di popolazione con più di 65 anni di età supererà il 30%.

Tale tendenza preoccupa i governi poiché devono prevedere sempre maggiori fondi a sostegno delle iniziative sociali e aumentare l'offerta di servizi pubblici efficaci ed efficienti, migliorando la cura dell'ambiente e più in generale la qualità della vita.

Ma anche se il trasporto pubblico diverrà sempre più un elemento strategico, le automobili continueranno a circolare implicando i relativi problemi di traffico, di parcheggio, di sicurezza, di risorse energetiche, di inquinamento.

Queste problematiche devono essere affrontate beneficiando delle tecnologie ICT, in particolare quelle denominate "smart", che sono tra i principali fattori abilitanti dei futuri ITS (Sistemi di Trasporto Intelligente).

[4]

1.1 Contesto

Uno studio dell'Università di Los Angeles [1] mostra che una media del 30% dei veicoli in circolazione nel traffico cittadino è in cerca di parcheggio e, si è valutato che ciascun veicolo, nella disponibilità di parcheggi a pagamento, può arrivare a continuare nella ricerca per decine di minuti, al solo fine di trovarne uno più economico o addirittura gratuito.

Questo dato, piuttosto sorprendente, fa intuire la complessità dei problemi legati alla gestione dei trasporti nelle città più popolate. Infatti, non è sufficiente considerare i soli aspetti propriamente logistici poiché, nell'attuale struttura sociale, l'utilizzo personale dell'automobile ha una valenza di "status" che influenza il comportamento di coloro che si devono spostare.

Un primo semplice strumento messo in atto dalle Amministrazioni delle grandi città a questo scopo è stato la limitazione dell'accesso dei veicoli al centro cittadino sulla base di vari criteri (e.g. targhe pari/dispari, Classe Euro di omologazione, orario, ticket, etc.).

Tuttavia queste sono misure utili a limitare le dimensioni del problema, ma che non ne costituiscono una soluzione.

In generale quasi tutte le realtà urbane di una grandezza significativa hanno accresciuto la disponibilità di aree di sosta a pagamento (fonte non trascurabile di introiti per le Amministrazioni locali), ma quasi nulla è stato fatto per garantire funzionalità accessorie quali, ad esempio, la prenotazione del posto. Questo perché o ritenuto troppo oneroso da implementare, o di scarso valore per il gestore del parcheggio stesso (chi possiede strutture in zone molto trafficate non percepisce valore nel tenere bloccati alcuni posti in attesa dei prenotati quando questi possono praticamente essere costantemente occupati).

Il sistema di trasporto pubblico è invece un servizio particolarmente critico poiché è costoso da esercire in modo efficace ed efficiente e, in ragione della sua funzione di pubblica utilità, viene offerto tipicamente in perdita.

La dinamicità della domanda è un primo problema: esiste un

significativo divario tra la quantità di passeggeri trasportata nelle ore di punta e quella in tutto il resto della giornata.

Vi sono poi le complicazioni dovute: un percorso tipico da pendolare da una zona periferica a bassa densità abitativa fino al centro città è altamente inefficiente alla partenza (i mezzi partono pressoché vuoti) e altamente inefficiente all'arrivo (i mezzi sono talmente pieni che non riescono a caricare tutti i passeggeri). Inoltre è molto difficile una gestione efficace delle coincidenze, almeno per il trasporto su gomma, poiché la dinamica del traffico rende inservibile qualunque programmazione. Infine, i mezzi più affidabili, ovvero treni e metropolitane, sono anche quelli più vincolanti: stazioni e binari non possono essere facilmente adattati ai pattern di movimento dei passeggeri nella città che, anche se lentamente, si modificano.

Soluzioni moderne quali il bike sharing ed il car sharing, ovvero la possibilità di utilizzare in modo condiviso biciclette e automobili disponibili in appositi parcheggi distribuiti in città, sono state realizzate in molte città italiane ed hanno avuto una buona risposta, ma rimangono complementari in un sistema di trasporto pubblico locale.

Al fine di migliorare l'esperienza d'uso dei mezzi pubblici, le aziende locali fanno ricorso già da qualche tempo a tecnologie ICT ad esempio per consentire agli utenti di determinare i percorsi ottimali per spostarsi tra due punti della città, anche includendo eventuali tratti da percorrere a piedi.

Tali valutazioni sono fatte sulla base della conoscenza della posizione delle fermate, dei percorsi dei mezzi e degli orari "previsti" dei passaggi. Tuttavia questo calcolo è una stima che non tiene conto di eventuali ritardi, che possono verificarsi nel corso del viaggio, né della possibile inaccessibilità di un mezzo per sovraffollamento.

Queste variabili rendono in qualche modo non completamente affidabile la pianificazione del percorso e aleatoria anche la possibilità di mettere in atto soluzioni alternative: poter avere il controllo del mezzo di trasporto rende ancora l'automobile difficilmente sostituibile.

Se immaginiamo, come peraltro spesso già avviene, che i mezzi siano sempre localizzabili e dotati di sensori di bordo, ad esempio per rilevare

la quantità di passeggeri a bordo e il numero di persone che scendono e salgono ad ogni fermata, il sistema di trasporto pubblico potrebbe compiere un salto di qualità.

Apposite applicazioni potrebbero ad esempio determinare una variazione della frequenza delle corse sulla base sia di informazioni, sia dinamiche dei pattern di utilizzo. Tali variazioni potrebbero essere poi controllate dinamicamente sulla base dei tempi di percorrenza rilevati ed eventualmente pesate con informazioni esterne, come ad esempio le stime di percorrenza delle strade. Contestualmente apposite applicazioni su smartphone potrebbero avvisare in tempo reale i passeggeri di coincidenze a rischio e/o dirottare su percorsi e mezzi alternativi ottimali in una logica di multimodalità, mettendo così a valore la capacità di gestione della ricchezza informativa propria del concetto “smart city” e consentendo agli utilizzatori finali un maggior controllo sui termini del trasporto.

1.2 Che cosa è l’Intelligent Transport System

Con *Intelligent Transport System* (ITS) si intende l’integrazione delle conoscenze e delle tecnologie nel campo dell’ICT con l’ingegneria dei trasporti per la pianificazione, progettazione, esercizio, manutenzione, gestione e utilizzo dei sistemi di trasporto.

L’obiettivo primario dell’ITS è stato identificato dall’Europa nel “Piano di Azione ITS” [2] e consiste nel raggiungere un equilibrio tra il bisogno di assicurare una mobilità a persone e merci e quello di garantire una sostenibilità dei sistemi di trasporto e della mobilità alla luce delle scarse risorse disponibili e dei limiti da rispettare in campo di emissioni ambientali.

I tre aspetti più rilevanti dell’ITS sono: la *sicurezza dei trasporti*, l’*efficienza* e l’*ambiente*.

L’obiettivo non è tanto quello di eliminare il traffico stradale, ma quello di arrivare ad una mobilità efficiente per basata su un flusso veicolare sicuro, omogeneo e non interrotto anche in caso di situazioni congestionate.

Per raggiungere questo obiettivo è necessario che ogni Stato e, soprattutto, le sue metropoli si dotino di sistemi di raccolta e trattamento dell'informazione di alta qualità, in grado di supportare il trasporto cross-modale (mezzo privato/mezzo pubblico), grazie alla disponibilità di informazioni dinamiche ed affidabili messe a disposizione del cittadino in mobilità.

La disponibilità di queste informazioni di alta qualità deve partire dalla raccolta strutturata e dall'elaborazione di quantità sempre crescenti di dati resi disponibili da un ecosistema molto strutturato, che oggi definiamo genericamente come "smart city".

In questo contesto, ci si attende che in un futuro prossimo numerosi sensori ed attuatori "intelligenti" saranno installati sul territorio e sui veicoli e, comunicando tra di loro e con centri servizi, saranno in grado di prendere decisioni in autonomia e/o su comando di apposite applicazioni centralizzate gestite dalle pubbliche amministrazioni dai fornitori di servizi pubblici e privati.

Il divenire attori di una città "smart" non può prescindere, quindi, da una forte valorizzazione delle reti TLC, che devono essere sempre più capillari e caratterizzate dall'utilizzo di apposite tecnologie, cosiddette M2M (Machine-to-Machine), adatte a connettere questi nuovi dispositivi.

L'informazione e la conoscenza sono sempre state le chiavi per abilitare servizi e business innovativi, ma in passato la capacità di raccogliere, elaborare, gestire dati in real-time, era impensabile. Oggi la crescente capacità tecnologica ed elaborativa di trattare e di prendere decisioni in tempo reale, basandosi su queste grandi quantità di dati, diventa un potente abilitatore alla trasformazione del business di quei soggetti che si dotano di capacità di analizzare ed utilizzare i "big data".

Dal punto di vista tecnologico, si sta sviluppando una nuova generazione di sistemi di "computing" che si basa sull'analisi di dati provenienti da sistemi embedded e che sfruttano le nuove architetture M2M in cui l'acquisizione, l'immagazzinamento e l'elaborazione del dato si sposta sempre più vicino al "sensore intelligente" che raccoglie il dato stesso.

Con l'avvento del concetto di "smart city" si sta andando verso la creazione di un ecosistema in cui l'enorme quantità di dati raccolta ed elaborata tramite un numero crescente di sensori più o meno sofisticati, così come i dati provenienti da processi di elaborazione e "fusione" dei dati da varie fonti, viene reso disponibile per usi che possono andare ben oltre le motivazioni, per cui quei dati sono stati originariamente raccolti ed elaborati.

In pratica è plausibile che concetti come quelli di M2M, "smart city", "big data" ed "open data" vadano verso una graduale convergenza ed integrazione.

Sebbene i vantaggi siano evidenti, questo obiettivo di integrazione e convergenza non è semplice da raggiungere, in quanto, oltre alla creazione di nuove infrastrutture dedicate alla raccolta ed al trattamento di nuovi dati attualmente non ancora disponibili, sarà necessaria l'armonizzazione e l'apertura di sistemi legacy già in esercizio e, affinché ciò possa avvenire, si dovrà necessariamente dimostrare di poter creare vantaggi anche a tutti quei soggetti che oggi trattano quei dati per le proprie finalità istituzionali o di business.

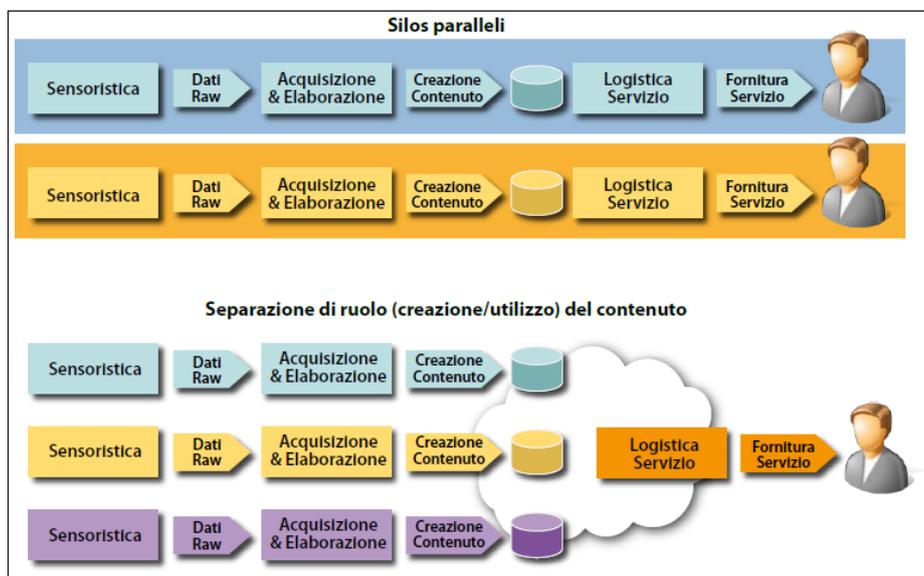


Fig. 1.1: Modelli per la creazione di contenuto ed il relativo uso

Per questo motivo, sembra ragionevole promuovere lo sviluppo di sistemi in cui la generazione del dato grezzo o del contenuto elaborato non siano più strettamente ed esclusivamente legati al processo di business per il quale il dato viene originariamente raccolto e trattato, ma diventino elementi di un processo più generale ed aperto in cui i vari contenuti diventano disponibili a più soggetti che possono utilizzarli per creare dei servizi attualmente non esistenti. [3]

1.3 I protocolli TTI (Traffic and Travel Information)

I protocolli TTI (Traffic and Travel Information) regolamentano la diffusione di informazioni inerenti agli eventi di infomobilità in essere all'interno di una rete viaria agli utenti che ne stanno usufruendo mediante procedure standardizzate e dispositivi predisposti. Tale capitolo vuole fungere da introduzione a due tra le principali tecnologie esistenti in letteratura, TMC e TPEG, contestualizzandone lo sviluppo e la diffusione all'interno delle Istituzioni sovranazionali preposte. Si esporrà pertanto un quadro generale esplicativo degli organismi europei coinvolti in tale ambito per poi presentare nel dettaglio le specifiche dei singoli protocolli.

1.3.1 Ertico

ERTICO - ITS Europe è un'associazione di circa 100 aziende e istituzioni coinvolte nella produzione di sistemi di trasporto intelligenti (ITS). Insieme, i partner ERTICO conducono una serie di attività per sviluppare e distribuire ITS al fine di aumentare la sicurezza degli utenti, proteggere l'ambiente e sostenere la mobilità nel modo più conveniente.

ERTICO - ITS Europe è stata fondata nel 1991 come piattaforma per la cooperazione di tutti i soggetti interessati a sviluppare e distribuire ITS in Europa. Il partenariato ERTICO è di tipo pubblico/privato costituito da

oltre un centinaio di soggetti attraverso 8 diversi settori, tutti impegnati a portare l'intelligenza nella mobilità di persone e merci in Europa.



Fig. 1.2: Logo ERTICO

In particolare, ERTICO prevede:

- Opportunità progettuali cofinanziate ad-hoc con un alto tasso di successo per promuovere i sistemi di trasporto intelligenti secondo specifiche priorità;
- Lavoro di cooperazione con le piattaforme dei membri coinvolti per colmare il divario tra i progetti e la distribuzione, raggiungere commesse di lungo termine e gestire la manutenzione dei servizi;
- Test di interoperabilità per supportare la distribuzione di soluzioni ITS interoperabili;
- Eventi per la condivisione delle conoscenze, il networking e la visibilità delle parti interessate e la promozione di ITS;
- Appoggio nella sensibilizzazione e discussione circa problematiche proprie con tutti i soggetti interessati.

Al fine di raggiungere obiettivi sociopolitici, rafforzare la competitività del settore e ottimizzare i servizi per l'utente, il partenariato mira a sposare gli interessi singoli degli operatori pubblici e privati.

La vision

Integrazione dell'intelligenza per una mobilità più:

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 11

- *sicura*: zero incidenti;
- *intelligente*: zero ritardi e persone costantemente informate;
- *pulita*: ridotto impatto ambientale.

La mission

- sviluppare, promuovere e implementare i sistemi e i servizi di trasporto intelligenti i quali necessitano di un coinvolgimento multi-stakeholder.
- attuare dei necessari abilitatori di distribuzione;
- valutare, adattare e utilizzare le più avanzate tecnologie relative.

I progetti – Ricerca Avanzata, Sviluppo e Distribuzione

ERTICO offre una piattaforma di networking unica per la ricerca collaborativa internazionale e lo sviluppo, in cui le idee innovative si trasformano in risultati tangibili. Viene condotto un processo di certificazione dello sviluppo e della gestione del progetto dalla generazione delle idee fino alla presentazione della proposta di progetto, e dall'esecuzione del progetto fino alla diffusione dei suoi risultati.

Il team di project management interno assicura un'alta competenza tecnica in vari settori ITS e molti anni di esperienza nella gestione dei propri progetti.

I progetti ERTICO possono essere annoverati entro le 3 seguenti aree tematiche:

- Clean Mobility (e.g. Compass 4d);
- Smart Mobility (e.g. MobiNet);
- Safe Mobility (e.g. UDrive).

Piattaforme

ERTICO sostiene in maniera crescente lo sviluppo, l'implementazione e il mantenimento di servizi nel contesto delle attività autofinanziate, come ha già fatto nel contesto delle informazioni su traffico e viaggiatori (TISA), delle mappe digitali/di navigazione legate a sistemi avanzati di assistenza alla guida (ADASIS), dell'interoperabilità delle innovazioni ICT in ambito eMobility (eMI3) e dei Dati di Georeferenziazione della Rete di Trasporti (TN-ITS, Piattaforma di diffusione dei dati di Georeferenziazione della rete di trasporti ITS).

Nel contesto delle piattaforme, il partenariato ERTICO invita stakeholder esterni per identificare e lavorare insieme su servizi/problemi prioritari in ambito ITS, dove le parti interessate (ERTICO Partners e stakeholder esterni) possono assumersi la responsabilità di coordinare e/o eseguire attività necessarie in tutta la fase di innovazione per raggiungere l'effettiva diffusione e il mantenimento dei servizi ITS conformi. Queste piattaforme garantiscono il successo dello sviluppo, l'implementazione e il mantenimento di servizi/problematiche centrali dell'ambito ITS:

- Commesse a lungo termine;
- Coerenza delle attività;
- Centri di competenza per la consulenza politica;
- Sostegno da tutte le attività pertinenti ERTICO (progetti, eventi, azioni di sensibilizzazione).

Gruppi di interesse

Gli esperti del partenariato ERTICO uniscono le proprie forze per far avanzare i fattori abilitanti nella diffusione degli ITS, come l'interoperabilità, la tecnologia, la politica, l'organizzazione, il finanziamento, l'utilizzo e la cooperazione internazionale.

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 13

I propri gruppi di interesse supportano tutte le attività ERTICO nel quadro delle loro competenze specifiche. Il primo Interest Group avviato sarà sulla interoperabilità.

L'Interest Group sulla interoperabilità ha individuato l'importanza di applicare una metodologia di sviluppo e di distribuzione guardando al successo delle industrie operanti nel settore dell'informazione e della comunicazione. Questa metodologia si basa su standard ISO ed è stata implementata con successo dalle industrie ICT.

Gli eventi

ERTICO - ITS Europe organizza una vasta gamma di eventi che affrontano temi vari e mirati.

Il loro format, nonché il livello dei relatori e dei partecipanti, cambia conseguentemente.

- **ITS congressi:** ogni anno, tutti gli stakeholder si incontrano per discutere i più recenti sviluppi attraverso un programma di cento e più sessioni, per illustrare le ultime tecnologie e servizi in appositi spazi espositivi e per sperimentarli dal vivo.
- **Forum:** sono eventi in rete in cui diverse parti interessate dibattono su argomenti specifici. Di solito partecipa anche la Commissione europea dal momento che il suo input politicamente orientato e la sua posizione è molto importante per gli stakeholder che intendono sviluppare nuovi servizi sulla base della roadmap politica.
- **Laboratori:** possono essere descritti come riunioni di lavoro molto specifiche in cui le nuove iniziative e quelle già sviluppate ed introdotte vengono discusse in presenza dei soggetti interessati. Solitamente tali incontri sono seguiti da ulteriori workshop tematici.

Supervisione

Le attività di ERTICO sviluppano, promuovono e implementano sistemi e servizi di trasporto intelligenti. Una parte fondamentale per il successo

dell'implementazione europea dei vari sistemi e servizi ITS consiste nel promuovere i loro funzionamenti e i vantaggi che apporteranno alla società.

Il partenariato ERTICO ha costruito una propria comunità online (ERTICO Network), che va aldilà della partnership stessa in quanto mira a informare/spiegare i vantaggi dei ITS a media specializzati, istituzioni europee e nazionali, congressi e delegati agli eventi, ricercatori, ingegneri, project manager, etc.

La rete ERTICO copre tutte le modalità di trasporto e la catena del valore, dai costruttori di veicoli, fornitori di servizi e fornitori per gli utenti, operatori stradali, consulenti e autorità pubbliche.

ERTICO ha inoltre creato la biblioteca ITS per assicurare che tutte le sue conoscenze raccolte attraverso progetti o iniziative rimangano a disposizione dell'intera comunità globale.

Il partenariato ERTICO ora gestisce anche IBEC, il forum internazionale che sostiene l'uso della valutazione nella pianificazione, nel processo decisionale e nel controllo della diffusione di tecnologia intelligente per la mobilità con l'industria ITS e le agenzie governative.

Da qui, la competenza valutativa di IBEC è portata vicina agli utenti e alle parti interessate nell'ambito del partenariato ERTICO.

Inoltre, il partenariato ERTICO ha stretto legami con le istituzioni dell'UE, dandosi un ruolo in iniziative politiche. ERTICO partecipa allo sviluppo della Piattaforma Cooperativa ITS (C-ITS Platform), il quadro istituito dalla Commissione europea che include autorità nazionali, C-ITS stakeholder e decisori istituzionali europei per fornire linee politiche ai fini dello sviluppo di una tabella di marcia e una strategia di distribuzione per soluzioni C-ITS in Europa.

ERTICO collabora anche con il Parlamento europeo attraverso il Forum iMobility dove gli interlocutori ITS discutono problematiche ITS ad alta priorità sotto gli auspici dei parlamentari.

Attraverso la rete di 27 Associazioni Nazionali ITS, le conoscenze e le informazioni vengono trasmesse a tutti gli attori a livello locale e nazionale facendo sì che ITS venga promosso dal basso.

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 15

ERTICO promuove la cooperazione con la comunità internazionale ITS nelle Americhe e nella regione Asia-Pacifico. La condivisione della conoscenza attraverso partenariati internazionali contribuisce allo sviluppo di soluzioni ITS e del business della stessa partnership ERTICO.

1.3.2 Traveller Information Service Association (TISA)

TISA è un'associazione di membri indirizzata al mercato di portata mondiale, definita come una società no-profit focalizzata sull'attuazione proattiva di servizi e prodotti legati alle informazioni su traffico ed infomobilità basati su standard esistenti, tra cui in primo luogo le tecnologie RDS-TMC e TPEG.

TISA sostiene il mantenimento e lo sviluppo delle tecnologie standardizzate guidandole dalla realizzazione economica alla rapida accettazione sul mercato attraverso una vasta gamma di servizi e prodotti d'informazione sulla tematica del trasporto. Oltre alle informazioni sul traffico stradale, per esempio, il trasporto pubblico, i punti di interesse, il meteo e i dati ambientali continueranno ad essere nel focus di TISA delle proprie aree di azione.



Fig. 1.3: Logo TISA

Vision

TISA concepisce un mondo dove gli standard e le politiche delle informazioni sul traffico ed infomobilità aperte (Traffic and Traveller Information, TTI) migliorano la sicurezza in condizioni di traffico e l'efficienza del viaggio.

Mission

La mission di TISA consiste nello sviluppo e la promozione di standard e policy “aperte” che:

- Facilitano una diffusione tempestiva e conveniente dei servizi e prodotti TTI che permettono agli utenti di risparmiare tempo e denaro, di aumentare la sicurezza del traffico stradale e minimizzare l’impatto ambientale;
- Migliorano la qualità e minimizzano il costo di tali servizi e prodotti massimizzando l’interoperabilità a livello mondiale.

Storia

TISA è stata fondata dalla sinergia tra organizzazioni precedentemente distinte e assicura un quadro internazionale per lo sviluppo coordinato di norme e servizi futuri.

È stata creata nel dicembre 2007 come organizzazione senza scopo di lucro ai sensi della legge belga, prendendo in consegna le attività della ex forum TMC, il forum TPEG ed il progetto tedesco Mobile.Info.

L’ufficio esecutivo TISA è ospitato da ERTICO, a Bruxelles.

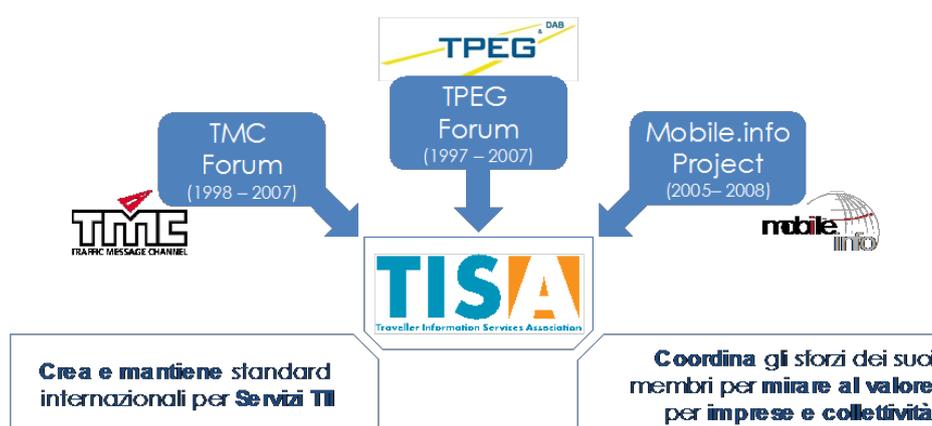


Fig. 1.4: TISA

Membri

L'adesione a TISA è aperta a tutte le organizzazioni pubbliche e private che hanno l'interesse della fornitura di informazioni sul traffico e sul viaggio e sostengono gli obiettivi dell'associazione.

L'appartenenza a TISA a livello mondiale è costantemente cresciuta contando oggi più di 100 membri ed è ampiamente rappresentata in tutta la catena del valore dei servizi informativi di infomobilità.

Cultura

TISA offre un ambiente in cui i valori e le esigenze dei servizi di infomobilità e le informazioni ai passeggeri e fornitori di contenuti, le autorità pubbliche, i produttori di auto, industria dei prodotti, le emittenti e operatori di trasmissione si incontrano.

Si tratta di un'organizzazione guidata da, e in crescita attraverso, i contributi attivi dei suoi membri.

Stimolando l'innovazione e facilitando la collaborazione, TISA fornisce un ambiente, in cui:

- Gli standard e le politiche esistenti vengono mantenute e, se necessario, aggiornate per adattarsi alla rapida evoluzione del mondo ITS;
- Vengono forniti processi e strumenti efficienti per supportare tempestivamente la standardizzazione o la pubblicazione dei risultati del lavoro prodotto all'interno del TISA;
- Viene mantenuto un livello di qualità alto per tutto il lavoro creato da TISA attraverso un processo di revisione e approvazione rigoroso;
- Viene valorizzata la proprietà intellettuale e TISA supporta la sua protezione;
- Legami attivi con organizzazioni adiacenti/correlate, come GENIVI, WorldDMB, Forum RDS, e attività DATEX-correlate, facilitano le attività trasversali.

TISA si sforza di utilizzare le sinergie sviluppate dall'interazione dei suoi membri e di sfruttare l'esperienza raggiunta di tutti i membri TISA per favorire business legati ai servizi TTI così come benefici per la società.

1.3.3 Traffic Message Channel (TMC)

TMC è un'applicazione specifica dello standard FM RDS usato per trasmettere informazioni su traffico e meteo in tempo reale. I dati vengono ricevuti silenziosamente, decodificati da un sistema di navigazione TMC-attrezzato e presentati al conducente, in genere offrendo un'esperienza di navigazione dinamica - allertando il conducente circa un problema sul percorso pianificato e calcolando un percorso alternativo per evitare l'incidente.



Fig. 1.5: Logo TMC

Funzionamento

Il fornitore di servizi invia i messaggi in codice verso l'emittente radio FM appropriata come segnale RDS (Radio Data System) all'interno di normali trasmissioni radio FM.

I dati TMC vengono ricevuti dalla radio e dall'antenna del veicolo e decodificati da un decoder TMC.

Questo dispositivo ricostruisce il messaggio originale, utilizzando un database di codici, e il luogo dell'evento, e tale informazione ricostruita viene presentata al conducente come messaggio visivo o parlato.

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 19

I messaggi TMC standard forniscono cinque elementi di base per la trasmissione dell'informazione: descrizione evento, posizione, direzione ed estensione, durata prevista, suggerimento di deviazione.

Vantaggi per gli utenti

- Informazioni aggiornate sul traffico, consegnate in tempo reale;
- Conoscenza immediata di incidenti, lavori in corso e ingorghi;
- Informazioni filtrate solo per il percorso immediato;
- Informazioni nella lingua dell'utente;
- Trasmissione digitale di alta qualità;
- Compatibilità standardizzata dei ricevitori;
- Fornitura di servizi diffusa in tutta Europa, negli Stati Uniti e in molti altri paesi in tutto il mondo.

Database delle località a supporto dei sistemi RDS TMC

Il database TMC (Traffic Message Channel) è un elenco di tabelle che descrivono una rete stradale, tramite un insieme di punti e strade.

Un punto del TMC corrisponde ad un punto ben definito sulla rete stradale, rispetto al quale vengono riferiti gli eventi che accadono sulle strade vicine.

Maggiore è il numero di punti e strade inserite nel TMC, maggiore è il dettaglio con cui possono essere fornite le informazioni sulla mobilità.

Il TMC è unico per ogni paese dell'Unione Europea: tutti i TIC (Traffic Information Center) e TCC (Traffic Control Center), nonché tutti i servizi di infomobilità (ad esempio l'RDS-TMC trasmesso dalla RAI nel Nord-Italia su Radio 1 in FM) di uno stato, debbono usare lo stesso TMC.

I TMC degli stati dell'Unione sono stati realizzati seguendo gli standard CEN: ogni apparato che impieghi il database TMC può indifferentemente impiegare quello di qualunque stato (Italia, Germania,

ecc.), questo permette, ad esempio, ad un ricevitore RDS-TMC abilitato, semplicemente impiegando la chip card (contenente il database delle località notevoli) del paese in cui si trova, di ricevere le informazioni stradali attinenti al paese in cui l'automobilista si trova in quel momento.

Poichè il database è unico per ogni paese, l'aggiornamento deve essere effettuato da un Ente che provvede sia alle modifiche necessarie, sia alla distribuzione. Per l'Italia il gestore del database TMC Italia è il Ministero delle Infrastrutture e dei Trasporti che provvede periodicamente agli aggiornamenti ed alla distribuzione.

Il database RDS-TMC è finalizzato a:

- scambio dati tra i centri di informazione sul traffico (i.e. tra il CCISS e le concessionarie autostradali);
- scambio dati tra i centri Italiani ed esteri (i.e. tra il CCISS e la Francia, la Baviera, e tra i centri delle concessionarie autostradali e i centri esteri);
- erogazione del servizio RDS-TMC. È il servizio erogato dalla RAI che consiste nell'invio automatico di informazioni di traffico codificate di fonte CCISS tramite frequenze radio FM - RadioRai agli autoveicoli con radio o navigatori di bordo dotati di questa funzionalità.

Il provider del servizio TMC in Italia è l'azienda Infoblu sulle frequenze di RTL 102.5 e le informazioni possono essere ricevute sull'80% del territorio.

Per ricevere il servizio TMC.IT è necessario un accordo tra il produttore del navigatore e Infoblu – RTL 102.5 affinché il dispositivo sia abilitato alla ricezione del segnale TMC.

1.3.4 Transport Protocol Experts Group (TPEG)

Gli esperti EBU (European Broadcasting Union) erano originariamente decisi a lavorare su sistemi idonei a trasmettere informazioni agli utenti

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 21

finali e hanno avuto una particolare attenzione nei confronti della viabilità e della tematica del viaggio.

L'idea della sigla TPEG è stata discussa ed è stato concordato, abbastanza facilmente, che fosse atta a definire il Transport Protocol Experts Group.

Il termine Trasporto è stato scelto con un doppio significato ovvero come termine inserito nel contesto dell'Infomobilità ma anche nel contesto di trasferimento delle informazioni (i dati) da un fornitore di servizi (service provider) ad un utente finale.

È stato previsto che la tecnologia TPEG dovesse essere in grado di gestire la distribuzione delle informazioni sia all'interno del dominio tradizionale TTI, ma anche molto efficacemente al di fuori da esso.



Fig. 1.6: Logo TPEG

Storia dello sviluppo

Nelle prime fasi dello sviluppo della tecnologia TPEG (biennio 1997-1998), è stato infatti previsto lo sviluppo di Applicazioni in grado di estendere i servizi di informazione multimodali ben oltre rispetto a quanto fino ad allora tentato di ottenere con tecnologie come RDS-TMC.

Nel momento in cui la tecnologia TPEG si è gradualmente evoluta, l'applicazione Road Traffic Message (prima implementazione del protocollo volta alla comunicazione di eventi stradali) è stata integrata con l'applicazione di Public Transport Information (per la diffusione di informazioni sul trasporto pubblico) ed entrambe hanno potuto sfruttare un nativo metodo di Local Referencing.

La prima generazione dell'applicazione TPEG (TPEG "generazione 1") forniva soltanto una codifica binaria, avendo in alcuni casi una specifica

separata per il mapping verso una codifica XML. La seconda generazione dell'applicazione TPEG (TPEG2 applications) è al momento descritta con un modello UML dal quale automaticamente sia la codifica binaria che quella XML vengono derivate.

Ad oggi, la tecnologia TPEG viene considerata in grado di fornire un "toolkit" per la consegna di diversi tipologie di contenuti (con diversi metodi di referenziazione della location).

Alcune delle applicazioni TPEG standardizzate includono ad oggi la gestione informativa del traffico, dei parcheggi, dei prezzi del carburante e delle condizioni meteorologiche.

In futuro, ci si aspetta che potranno essere fornite anche le informazioni sugli arrivi e le partenze del trasporto pubblico utilizzando la tecnologia TPEG.

Client non specifici

Dato che le specifiche TPEG corrispondono, a tutti gli effetti, all'interfaccia on-air, il concetto abbraccia inoltre l'idea che verranno attuate nel prossimo periodo una serie di diverse implementazioni di dispositivi client TPEG utilizzabili da utenti finali.

Un altro importante aspetto della filosofia costitutiva del TPEG consisteva nell'andare a considerare fondamentali sia la scalabilità dei contenuti sia i dispositivi client.

Per tale ragione, un fornitore di servizi avrebbe potuto scegliere dal toolkit TPEG di implementare servizi sia altamente sia poco dettagliati, i quali rispettivamente avrebbero usato più o meno banda e richiesto dispositivi più o meno complessi.

Le specifiche TPEG offrono un metodo per la trasmissione di informazioni di infomobilità multimodale, indipendentemente dal tipo di client, la posizione o il canale di consegna richiesto (e.g. DAB, radio HD, Internet, DVB-x, DMB, GPRS, Wi-Fi...).

L'indipendenza dalla lingua ha rappresentato un principio fondante nella sua progettazione.

Idea di Funzionamento

A differenza del TMC (informazioni sul traffico stradale basato su eventi), TPEG fa riferimento a tutta una serie di strumenti di specifiche, per offrire una gamma più ampia di servizi ad una vasta gamma di utenti e dispositivi.

Volendo tracciare una prima caratterizzazione di base del protocollo, si può affermare che i servizi TPEG sono definiti in forma modulare e possono quindi variare in una serie di "direzioni":

- Applicazione - per esempio Messaggi sul Traffico stradale, Trasporto Pubblico o Informazioni sul parcheggio. Ogni applicazione è identificata in modo univoco da un ID di applicazione (AID) che viene assegnato dal Gruppo TPEG Application Working (TAWG) del Tisa;
- Metodo di trasmissione - per esempio Radio digitale DAB, DMB, internet;
- Metodo di referenziazione delle Località - basato su tabella (utilizzando ad esempio le tabelle di localizzazione TMC) oppure on-the-fly (utilizzando un metodo che fornisce un riferimento della posizione funzionante con o senza mappe e che non richiede una tabella di look-up per la decodifica nel ricevitore);
- Dispositivi client - ad esempio sistemi di navigazione del veicolo, browser internet o dispositivi mobili;
- Conditional Access - se i dati vengono inviati gratuitamente o solo per gli utenti/dispositivi che hanno il diritto a riceverlo, per esempio pagando un abbonamento. La crittografia dei dati TPEG è possibile per mezzo di indicatori standardizzati di crittografia che sono stati assegnati dal Working Group TPEG Application (TAWG) del TISA.

Il termine profilo è utilizzato per individuare una combinazione di quanto definito sopra che, insieme, vanno a costituire quello che potrebbe essere individuato come un unico servizio TPEG.

Per esempio:

- Visualizzazione di eventi legati al traffico su una mappa grafica e possibilità di ricalcolare il percorso oppure ottimizzarlo;
- Visualizzazione delle informazioni sullo stato del trasporto pubblico su uno schermo del telefono cellulare.

Il dominio TTI

Il contenuto TTI viene fornito agli utenti finali attraverso molti meccanismi, in particolare dai trasmettitori che consegnano annunci parlati, usano dati RDS-TMC, teletext e internet per consegnare tali informazioni.

Ma, chiaramente, il contenuto deve essere reperito ed editato in conformità a rigorosi standard per assicurare che sia tempestivo ed accurato.

Il servizio di supporto TTI si occupa quindi completamente del reperimento, editing e consegna delle informazioni.

Per favorire una buona comprensione dei processi, originariamente la EBU ha escogitato l'idea consistente nei 2 segmenti visualizzati in Figura 1.7.

Il content segment copre tutte le possibili sorgenti di informazioni che devono essere reperite e processate prima che il delivery segment venga adottato per mandare le informazioni all'utente finale.

Negli ultimi anni, il Gruppo di Lavoro TISA per la direttiva ITS è stato molto attivo e ha ampliato l'idea di una catena del valore TTI.

Questo è in realtà un insieme complesso di interfacce e connessioni che ammontano a circa 30 attività identificabili, che sono effettivamente disposte in serie dall'accadimento di un evento TTI fino alla messa a disposizione delle informazioni sull'evento ad un dispositivo client di un utente finale.

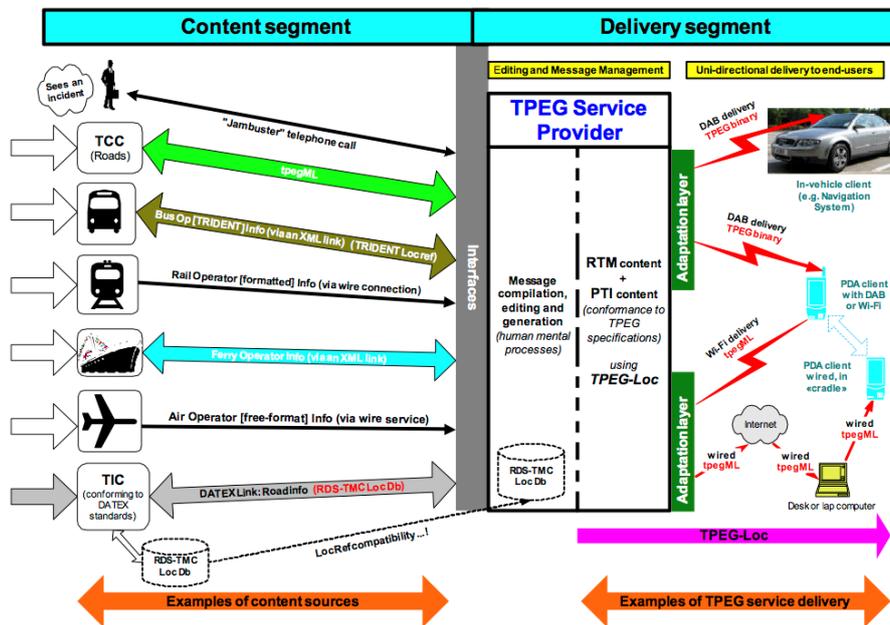


Fig. 1.7: Content and Delivery Segment

Filosofia di progettazione TPEG

I seguenti principi vengono considerati essenziali nello sviluppo del protocollo, della struttura e della semantica TPEG (referenza ISO/TS 18234-2):

- è unidirezionale;
- è orientato ai byte;
- fornisce una struttura di protocollo che impiega un framing asincrono;
- include un sistema di rilevazione dell'errore CRC applicabile ad una serie di livelli differenti;
- assume l'uso di un canale dati trasparente;
- assume che i sistemi sottostanti abbiano un appropriato livello di affidabilità;

- assume che i sottostanti sistemi possono adottare una correzione di errori;
- ha una struttura dati gerarchica;
- è usato per trasferire informazioni da Database a Database;
- fornisce il nome del service provider, il nome del servizio e le informazioni di rete;
- permette l'uso di meccanismi di crittografia se richiesti da un'applicazione;

Inoltre, TPEG:

- ha due segmenti significativi: Content e Delivery;
- permette ai servizi di essere liberi dal sincronismo del database della location con i client;
- estende le applicazioni ai servizi multimodali;
- offre un contenuto molto ricco indipendente dalla lingua;
- favorisce ad un dispositivo client di filtrare le opzioni per gli utenti finali;
- è progettato per dispositivi client spessi (sistemi di navigazione) e sottili (smartphone);
- non ha bisogno di un database di location precaricato;
- contiene un messaggio scalabile;
- si adatta ai portatori di consegna è semplice.

N.B. In TPEG i dispositivi degli utenti finali vengono definiti come client mentre nel protocollo RDS-TMC i dispositivi degli utenti finali sono noti come terminali.

TPEG non necessita di un database di location precaricato I service providers raccolgono e forniscono contenuto multimodale di ampio spettro ma la possibilità di consegna dei dati fornita dalla prima tecnologia dati Europea in ambito TTI, denominata RDS-TMC, aveva significative limitazioni. Quel sistema è essenzialmente limitato agli eventi sulle strade extraurbane ed ogni terminale decoder deve avere un database di location atto ad interpretare qualsiasi messaggio ricevuto. Tale aspetto ha contribuito alla costituzione di una situazione complessa per tutti gli utenti finali e questo inconveniente è stato risolto da diversi approcci B2B che risolvono i problemi di sottoscrizione e aggiornamento/rinnovo del database. La tecnologia TPEG supera questa limitazione mediante l'introduzione di diversi metodi di rilevazione della location i quali possono fornire un'informazione di localizzazione molto ricca tale da non rendere necessario un database offline. Il più grande vantaggio, tuttavia, consiste nella possibilità di consegna di TTI anche per le aree urbane, perfino quelle densamente popolate. I sistemi di navigazione con mappe digitali possono leggere automaticamente il contenuto legato alla location e localizzare un evento direttamente sul display della mappa. Un dispositivo client puramente testuale (e.g. un PDA) è in grado di mostrare i luoghi rilevati localmente, come stazioni ferroviarie o il nome di un binario, direttamente all'utente finale attraverso un messaggio di testo.

Indipendenza dalla Lingua RDS-TMC ha presentato la modalità ai fini della consegna di dati informativi volti all'aggiornamento di un utente finale mobile che desidera ottenere contenuti utilizzando una lingua diversa dalla propria nel momento in cui si trova in una certa località. Il concetto è implementato affinché il dispositivo client presenti informazioni nella lingua prescelta dall'utente. RDS-TMC è limitato perché si basa su frasi predeterminate, spesso non esatte, per fornire ciò che il service provider desidera esprimere. La tecnologia TPEG si è spinta un passo avanti decomponendo l'informazione essenzialmente in singole parole che possono essere tradotte in varie lingue più velocemente. Inoltre, il concetto di costruzione del messaggio TPEG consiste nella possibilità di assemblare

l'informazione disponibile circa un evento in messaggi potenzialmente molto ricchi ed informativi, esattamente come vorrebbe il service provider.

Applicazioni multimodali TPEG è la prima applicazione TTI che copre tutte le modalità di trasporto all'interno del panorama dei trasporti. Può infatti servire l'automobilista all'interno di un'area urbana così come il passeggero di un autobus, il viaggiatore pendolare e il guidatore sulle lunghe distanze. All'inizio dello sviluppo del TPEG, la prima applicazione (generazione 1) denominata TPEG1-RTM fu progettata per occuparsi dei Messaggi sul Traffico Stradale (RTM, road traffic messages) indipendentemente dalla location. Essa mirava alle informazioni urbane in virtù della ricchezza di contenuti che poteva offrire. Ora il RTM è stato superato dall'applicazione afferente alla generazione 2 denominata TPEG-TEC (Traffic Event Compact) che offre un più ampio potenziale per questa area di applicazione dal momento che presenta sia cause che gli effetti nei messaggi. TPEG-TEC è tuttavia compatibile anche per il vecchio protocollo. TPEG2-TEC, insieme a TPEG2-TFP (Traffic Flow and Prediction) consente l'effettiva ed efficiente consegna di info su tutti i tipi di condizioni di traffico stradale. Inoltre, la tecnologia TPEG è progettata per facilitare molte più applicazioni che trattano molti altri aspetti del dominio TTI. Di seguito si elencano alcune applicazioni TPEG2 che valgono da esempio:

- TPEG2-RMR: (Road and Multi-modal Routes) consente ad un service provider di fornire avvisi stradali mixando varie modalità di trasporti, includendo quindi il trasporto pubblico per una parte del percorso o la gestione del cosiddetto "ultimo miglio";
- TPEG2-WEA (Weather Information) consente ai service provider di fornire report su previsioni meteo globali, comprese previsioni di più giorni;
- TPEG2-PKI (Parking Information) consente ai service providers di fornire informazioni globali su parcheggi attrezzati e tariffe;

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 29

- TPEG2-FPI (Fuel Price Information) consente ai service provider di fornire informazioni globali su stazioni di rifornimento e prezzi del carburante;
- TPEG2-ADT, in fase di sviluppo, permetterà ad un service provider di fornire informazioni sul trasporto pubblico globali in merito agli arrivi e le partenze di aerei, autobus, traghetti, tram e treni.

Pertanto, la tecnologia TPEG estende i servizi di informazione multimodali ben oltre qualsiasi tentativo finora portato avanti da altre tecnologie come RDS-TMC e pone la consegna di TTI di nuovo in primo piano nell'essere una fonte onnipresente di informazioni.

Filtraggio La tecnologia TPEG è stata sviluppata nel contesto della fornitura di servizi di trasmissione nel quale i messaggi vengono recapitati a molti dispositivi client. In un determinato momento solo alcuni degli utenti finali vorrebbero ricevere certe informazioni (ad esempio: le informazioni su ingorghi in una città a più di 200 km di distanza non sono utili). Per consentire che grandi quantità di informazioni vengano fornite senza sovraccaricare gli utenti finali con dati di scarsa utilità per loro, la filosofia di progettazione del TPEG, attraverso la codifica esplicita, si basa sull'idea di un filtro lato client. Ciò consente agli utenti finali di scegliere i messaggi in base a un certo numero di criteri, come il tipo di evento, luogo, modalità di trasporto, direzione di marcia, ecc. Originariamente, quando la tecnologia TPEG è stata pianificata, la filosofia progettuale è stata incentrata sul concetto di fornitura di informazioni all'utente finale in modo tale da non rendere necessario alcun set-up preventivo in occasione del primo accesso al servizio e da permettere al filtro di essere applicato al contenuto per presentare solo le informazioni che l'utente finale ha richiesto. L'aspettativa iniziale dei progettisti era quella in cui i servizi TPEG sarebbero poi diventati prolifici e che i dispositivi client TPEG sarebbero stati in numero elevato e in varie tipologie comprendenti i cosiddetti thick client quali i sistemi di navigazione a bordo dei veicoli fino ai thin client, come PDA con connettività wireless o cablata. Su questa base

ci si aspettava che un utilizzatore volesse accedere ai servizi sia nel suo territorio nazionale sia in viaggio, incluse le situazioni in cui si trova lontano da casa in un territorio in cui viene parlata un'altra lingua. Perciò, la tecnologia TPEG ha due principali esigenze da soddisfare: la mobilità dell'accesso e l'indipendenza della lingua. Con i dispositivi client mobili fu subito chiaro che le informazioni indicate all'utente finale legate alla posizione occupata dovesse essere sia comprensibile alle persone sia leggibili dalla macchina. TPEG ha originariamente soddisfatto questa esigenza con la metodologia TPEG-LOC, fornendo entrambe le tipologie di contenuto per tutti i messaggi. I successivi sviluppi hanno creato il concetto di un "contenitore di riferimenti delle località" affinché i messaggi TPEG potessero indicare una certa posizione sulla base di una scelta tra più metodi di riferimento della localizzazione (metodi on-the-fly e/o metodi TMC basati su tabelle di location), per soddisfare sia i clienti "thin" che quelli "thick". Questa tecnica ha superato un'altra problematica rilevata nel sistema RDS-TMC nel quale vi era la necessità per ogni dispositivo client di contenere un database delle posizioni che deve essere mantenuto sia dal fornitore di servizi che da tutti i dispositivi client. Utilizzando la tecnologia TPEG, non è obbligatorio per l'interpretazione del messaggio avere alcun database delle posizioni nel dispositivo client sebbene sia preferibile tenere aggiornate le mappe nel dispositivo client, come nel caso dei sistemi di navigazione, utilizzando i dati forniti dinamicamente da uno dei servizio TPEG.

Le tabelle TPEG

La tecnologia TPEG soddisfa facilmente il requisito di indipendenza dalla lingua usando una tabella di valori di codice attraverso l'interfaccia "on-air" (in onda) per consegnare gran parte del contenuto. In questo contesto l'interfaccia "on-air" può essere considerata appropriata a descrivere l'interfaccia "service provider to end-user" indipendentemente dal tipo di connessione (che potrebbe ad esempio essere una trasmissione wireless o internet cablata). Le tabelle sono estendibili, con compatibilità

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 31

legacy, e non necessitano di essere completamente aggiornate in un dispositivo client per essere consultate. Per sintetizzare i vantaggi legati all'adozione delle tabelle si elencano di seguito i principali elementi che le caratterizzano:

- Tabelle TPEG estendibili;
- Contenuto del messaggio ricco con le tabelle TPEG;
- Indipendenza dalla lingua attraverso le tabelle TPEG.

Nel core della tecnologia TPEG risiede il concetto che gli eventi saranno descritti tramite riferimenti ad un set di tabelle che sono predeterminate e volte a coprire tutte le possibili situazioni che possono necessitare di essere descritte da una particolare applicazione. Una delle ragioni chiave che risiedono dietro tale concetto è il fatto che il valore del codice per una "parola" è consegnato al dispositivo client ed essa viene poi tradotta in informazioni visualizzabili dall'utente finale nella lingua da esso scelta. Ogni tabella può contenere fino a 255 "termini" con un uso comune del codice '0' per segnalare la parola "sconosciuto". Prendiamo come esempio l'applicazione TPEG2-TEC. Gli esperti hanno trovato che esistevano 31 cosiddette "cause principali" e l'applicazione richiedeva 45 tabelle in totale per descrivere ogni aspetto delle informazioni sugli eventi a partire dal tipo di veicolo fino allo stato di performance della rete viaria (e.g. traffico scorrevole). Il concetto di termine (o parola) è fondamentale: non c'è nessuna intenzione di costruire frasi predeterminate nella tecnologia TPEG ma la frase è composta al volo da un service provider nella compilazione dei messaggi dando così una grande libertà nell'esprimere un evento e favorendo l'opportunità di trasmettere ricchi contenuti. Ciononostante, il concetto di parola è lato per consentire a singole entità come largo veicolo a 4 ruote motrici di utilizzare un unico valore di codice.

Tabelle TPEG – struttura e 3 colonne Nelle specifiche della tecnologia TPEG si rende necessario assegnare valori di codice ad ogni parola (sia in senso stretto che lato) e determinare come le tabelle vengono usate nella

struttura gerarchica di ogni messaggio. Ogni tabella è caratterizzata da un titolo e a seguire le parole sono elencate con i loro valori di codice, il termine in Inglese di riferimento e 2 colonne addizionali per commenti ed esempi. Queste ultime 2 colonne sono scarsamente popolate al momento e durante la traduzione e i processi di testing operativi è d'aiuto aggiungere ulteriori commenti ed esempi al fine di supportare altri soggetti a comprendere il concetto dettagliato in ciascuna parola (Figura 1.8).

Code	Reference-English 'Word'	Comments	Examples
1	security alert	A security alert has been declared, possibly resulting in restrictions on traffic movement and delays.	A.1.1.1.1 An area is closed due to a hostage situation.
2	contagious disease	A contagious disease is affecting the location potentially resulting in a movement restriction being applied.	A disease of animals such as foot and mouth or blue tongue.
3	environmental	An environmental incident has occurred, potentially resulting in a movement restriction being applied.	An area is closed due to pollution.
4	smog alert	A smog alert has been declared due to pollution levels in the air.	
5	batch service in progress	Traffic may flow in only one direction at a time. Commonly known as 'Single Alternate Line Traffic'.	One side of a single carriageway road is closed; traffic in one direction uses the remaining open side for two minutes, then traffic in the other direction uses it for two minutes; alternately.
6	road closed by the regulatory authorities	The regulatory authorities have closed the road for whatever reason.	

Fig. 1.8: Tabella TPEG

La tabella maggiormente popolata in TPEG2-TEC, al momento, utilizza 32 istanze, quindi c'è spazio in abbondanza per la sua espansione nel caso di nuove situazioni da descrivere. L'idea della tabella di parole default consiste nel definire una parola che un dispositivo client dovrebbe esprimere se non avesse nel proprio dataset di parole il valore di codice ricevuto. Questa situazione può verificarsi se è stata estesa la tabella dal service provider ma il dispositivo client non è stato aggiornato.

Riferimento in Inglese Nella specifiche TPEG, le tabelle TPEG sono scritte nel cosiddetto Reference English (Inglese di riferimento); si tratta di una sorta di uso dell'Inglese tecnicamente accettabile che possiede definizioni tecniche comprese dagli esperti (la maggior parte contenute in un dizionario dati) e che dovrebbe consentire la traduzione a qualsiasi

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 33

lingua. In realtà, sarà necessario tradurre in Inglese Britannico perché alcune parole non vengono usate nel linguaggio quotidiano; ad esempio, “road-rail crossing” (“incrocio strada-rotaia”) corrisponde alla terminologia Inglese di riferimento, mentre “level-crossing” (passaggio a livello) è la formula comunemente usata dai madrelingua Inglese così come negli Stati Uniti si dovrebbe piuttosto parlare di un “railroad crossing”.

Tabella delle traduzioni Le tabelle TPEG descrivono singole entità, codificando una parola o un’espressione, ad esempio “piccola auto” o “largo veicolo a 4 ruote motrici”. Il compito della traduzione è inoltre quello di trovare l’espressione rilevante nel linguaggio in cui l’utente finale si ritiene a proprio agio; ad esempio, in Tedesco: “auto” o “all-radfahrzeug” (ugualmente “auto” in tedesco). Questo concetto assume che sarà possibile tradurre massicciamente senza il riferimento al contesto nel quale la parola sarà usata. Una procedura per gestire questa attività è stata stabilita dal TISA come segue:

write Use Case Proposal (from any TISA member)	UCP registered	evaluation of UCP and allocation of work	Translation work (by e-mail working of TISA experts)	TISA website publication – public domain
to TISA Executive Office	TISA Executive Office	BAWG meeting	Translations TF	by TISA Executive Office

TPEG tables – new words proposals and translations procedures

Fig. 1.9: Tabella delle traduzioni

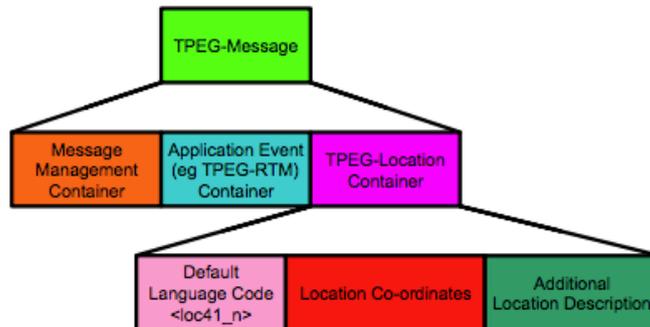
Dispositivi client non specifici Nel caso di dispositivi client non specifici (cioè dispositivi non costruiti specificatamente per servizi TPEG e quindi non dotati internamente di tabelle TPEG) è reso possibile il download della tabella (della lingua appropriata) al momento dell’uso, ad esempio quando si accede ad internet utilizzando un browser standard. Specifici dispositivi client TPEG saranno prodotti con le tabelle TPEG già installate, adeguate al mercato nel quale i dispositivi vengono venduti.

TPEG – Posizione di Riferimento

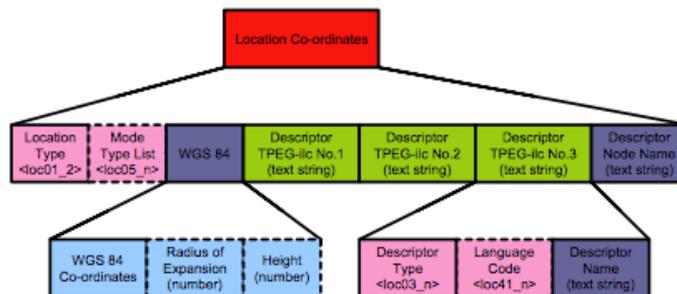
Metodo comune per il riferimento della posizione per tutte le applicazioni Le applicazioni TPEG in genere richiedono un metodo per identificare informazioni sulla posizione in riferimento agli eventi che vi si svolgono, ad esempio la chiusura di un ponte stradale o la cancellazione del servizio di autobus. Ciò ha portato alla definizione di un unico approccio ai fini dello sviluppo del Riferimento alla Posizione per le applicazioni TPEG e il metodo si chiama TPEG-Loc; è un metodo che viene condiviso anche per le diverse esigenze delle applicazioni TPEG-RTM e TPEG-PTI. TPEG-Loc consente l'erogazione di una informazione del riferimento alla posizione molto ricca relativamente ad ogni messaggio, consentendo a tutti i tipi di dispositivi client di localizzare un evento e visualizzare il suo posizionamento qualora sia utile per il cliente. Ogni messaggio TPEG si compone di tre parti principali: Gestione del Messaggio, Application Event e Posizione di riferimento. La posizione di riferimento è a sua volta divisa in molti sotto-componenti che sono raggruppati in tre categorizzazioni di alto livello (Figura 1.10).

Descrizioni dei riferimenti alle posizioni indipendenti dalla lingua

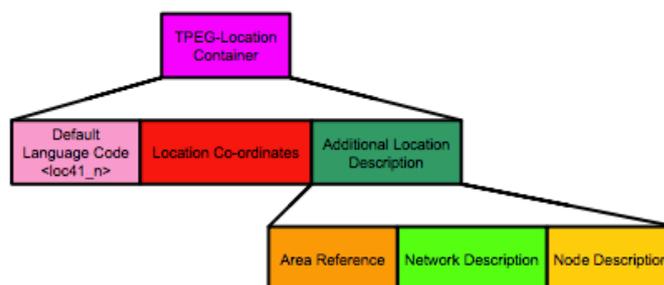
Un concetto chiave per TPEG-Loc consiste nel descrivere luoghi più similmente possibile ai nomi che l'utente associa ad essi. Questo inevitabilmente significa che occorre adottare un approccio indipendente dalla lingua, soprattutto nelle città bi-lingue, come Bruxelles/Brussel e Newport/Casnewydd, dove possono esserci service provider che indirizzano i loro servizi a una particolare comunità demografica/linguistica. Così TPEG- Loc permette di indicare un codice di lingua predefinita per tutte le descrizioni delle stringhe di testo di localizzazione e possono anche essere indicate singole stringhe dove per esempio due nomi delle strade sono stati descritti in due lingue diverse (ad esempio, rue Berkman e Berkmanstraat).



The TPEG-Loc container - related to overall message structure



Location Co-ordinates - the elements of a Nodal Area (i.e. radius < 5km)



Additional Location elements: Area reference, Network and Node descriptions

Fig. 1.10: Struttura del messaggio TPEG

Posizione Coordinate Le TPEG-Loc Location Coordinates (Coordinate di Posizione) offre la possibilità di descrivere le posizioni in un certo numero di modi diversi, ciascuno con specifici scopi. Questi sono definiti da una tabella nota come la tabella Location_Type, che definisce attualmente sette tipologie di location: Grande Area, Area Nodale, Segmento, Punto di Intersezione, Punto Racchiuso, Punto Non Legato e Punto Connesso. Un esame più dettagliato dell'Area Nodale mostra i singoli elementi che possono essere inclusi. Alcuni di questi sono chiaramente leggibili dalla macchina, alcuni sono chiaramente leggibili dall'uomo in modo da soddisfare un altro requisito fondamentale della tecnologia TPEG ovvero la sua utilizzabilità da tutti i tipi di dispositivi client per gli utenti finali.

Ulteriore descrizione della posizione La Descrizione della Posizione Addizionale del TPEG-Loc permette inoltre al service provider di fornire ulteriori strumenti importanti, molto utili in alcuni dispositivi client. L'area di riferimento (Area Reference) è uno strumento potente che permette alle posizioni di essere legate in una struttura ad albero le une alle altre, in modo che il filtraggio di messaggi del dispositivo client con luoghi ben definiti venga effettivamente presentato all'utente finale. La descrizione di rete (Network Description) è un altro strumento di filtro utile: un utente finale può limitare i messaggi visualizzando soltanto quelli, per esempio, legati ad una rete di tram. Il Nodo Descrizione (Node Description) consente a un fornitore di servizi di descrivere molti aspetti di un nodo, come un aeroporto, dove hanno bisogno di essere esposti molti piani, gli ingressi e le uscite.

TPEG in relazione al modello a livelli ISO/OSI

ISO/OSI Layer Model dalle specifiche TISA Il modello ISO/OSI favorisce una miglior comprensione delle connessioni per la comunicazione di dati bi-direzionale e i ruoli giocati nel stabilire una connessione. TPEG, originariamente, fu sviluppato soltanto per canali di

1.3. I PROTOCOLLI TTI (TRAFFIC AND TRAVEL INFORMATION) 37

consegna unidirezionali, ciononostante si adattava bene all'interno del modello ISO/OSI presentandosi nella descrizione mostrata dalla Figura 1.11.

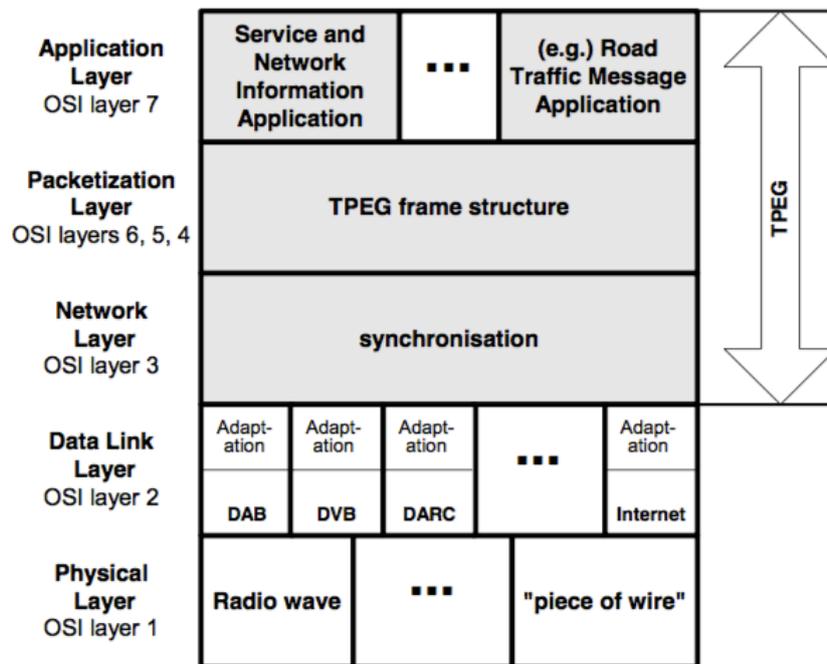


Fig. 1.11: Rappresentazione ISO/OSI del protocollo TPEG

TPEG in relazione al modello ISO/OSI tramite diversi canali portanti (da ISO 18.234-2) Oggi quasi tutti i sistemi di trasmissione digitali richiedono che TPEG venga trasmesso entro il loro livello di applicazione; per superare l'implicita dipendenza dal trasporto sui layers 3-6, tutti i livelli TPEG (livelli ISO 3-7) sono inclusi all'interno dello strato di applicazione dei portatori (strato 7 ISO). Il layout idealizzato viene esposto dalla Figura 1.12.

Modello TPEG a Strati espanso, il concetto dei 3 contenitori e il framing Un'altra modalità di concettualizzazione del TPEG come un protocollo a livelli è quella mostrata nella Figura 1.13. La tecnologia TPEG è stata essenzialmente progettata per una consegna unidirezionale "uno a

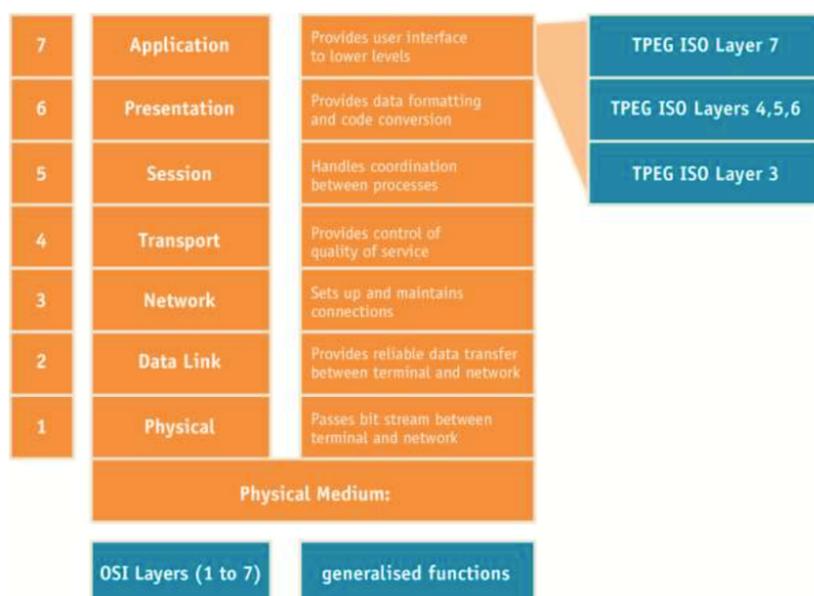
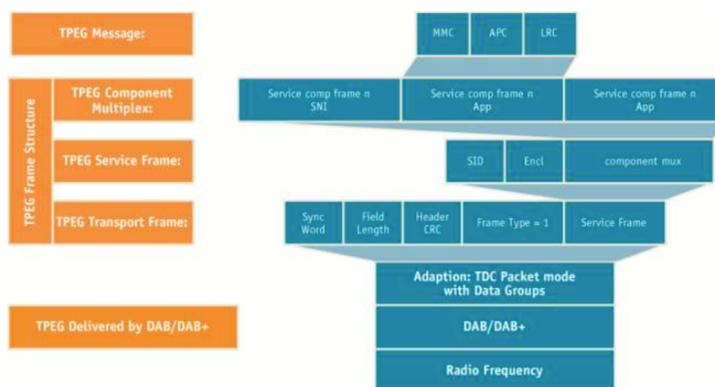


Fig. 1.12: Layout idealizzato

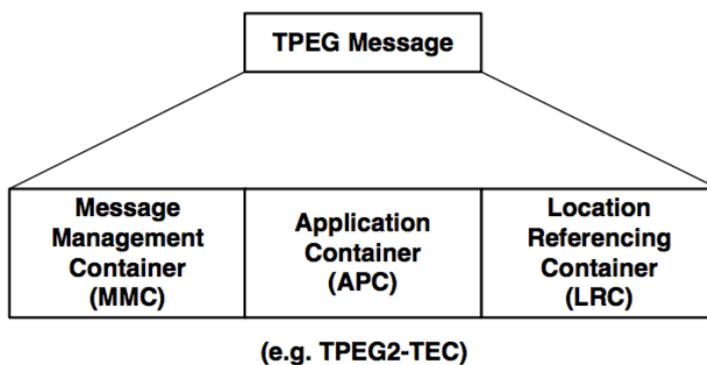
multi”; tuttavia, esso si adatta largamente all’interno del protocollo a 7 livelli ISO risultando quindi compatibile con molti futuri livelli di modificazione che verranno sviluppati nel momento in cui si creeranno determinate esigenze. Inoltre è possibile visualizzare il costrutto del messaggio TPEG usato frequentemente – benché disegnato dall’alto verso il basso – utile a comprendere come è composta un’applicazione TPEG usando una visualizzazione a 3 contenitori (Figura 1.14).

In conclusione, si tenga presente la seguente tabella per la descrizione degli acronimi afferenti alla tecnologia esaminata:



TPEG using Digital Radio adaption/delivery

Fig. 1.13: Concettualizzazione del TPEG come protocollo a livelli



TPEG Container concept visualisation - frequently used in Specifications

Fig. 1.14: Visualizzazione TPEG a tre contenitori

Abbreviation	Explanation
ADT	Arrival and Departure Time information – a TPEG2 application
APC	Application Container
CEN	Comité Européen de Normalisation
DAB	Digital Audio Broadcasting (aka Digital Radio)
DATEX	Data Exchange - protocol for exchanging traffic management information
EBU	European Broadcasting Union
EC	European Commission
FPI	Fuel Price Information – a TPEG2 application
IPR	Intellectual Property Right
ISO	International Standards Organisation
JPEG	Joint Picture Expert Group
LRC	Location Reference Container
MPEG	Motion Picture Expert group
MMC	Message Management Container
OSI	Open Systems Interconnection - an ISO concept
PDA	Personal Digital Assistant
PKI	ParKing Information – a TPEG2 application
RDS	Radio Data System
RDS-TMC	RDS-Traffic Message Channel - a feature of RDS
RMR	Road and Multi-modal Routes – a TPEG2 application
TEC	Traffic Event Compact – a TPEG2 application
TFP	Traffic Flow and Prediction – a TPEG2 application
TPEG	Transport Protocol Experts Group
TPEG1	TPEG Generation 1
TPEG2	TPEG Generation 2
TPEG1-LOC	Location referencing - a TPEG1 Binary application
TPEG1-RTM	Road Traffic Message - a TPEG1 Binary application
TPEG-SNI	Service and Network Information - a TPEG1 and TPEG2 application
TS	Technical Specification – the status of an ISO standard
TTI	Traffic and Travel Information
WEA	WEAther information – a TPEG2 application

Fig. 1.15: Abbreviazioni e applicazioni citate

1.4 Il modello dati DATEX II

Al fine di sostenere la mobilità sostenibile in Europa, la Commissione europea ha supportato lo sviluppo di mezzi che hanno l'obiettivo di armonizzare e standardizzare le strutture dei dati e dei servizi di scambio delle informazioni tra gli attori del traffico stradale.

È stato sviluppato, pertanto, lo standard **DATEX** per lo scambio di informazioni tra i centri di gestione del traffico. Poiché ciò permette uno scambio istantaneo delle informazioni, una gestione più veloce delle reti stradali e, quindi, una massimizzazione delle capacità della rete stradale di ridurre gli effetti negativi della congestione DATEX gioca un ruolo fondamentale nell'implementazione delle tecnologie ITS in Europa.

Esso costituisce il riferimento per le applicazioni che sono state sviluppate negli ultimi 10 anni. È proprio l'ultima decade che, in Europa, è protagonista di molti investimenti sia nei centri di controllo del traffico e di informazione che nel monitoraggio della rete transeuropea. In particolare, Datex consiste in una serie di specifiche per lo scambio di informazioni sul traffico in un formato standard tra sistemi di diverso tipo. È stato messo a punto con l'obiettivo di armonizzare lo scambio di informazioni sul traffico e sui viaggi a tutti i livelli, sia tra gli operatori addetti alla circolazione stradale (urbana e non urbana) e che tra fornitori di servizi.

Datex II è un modello di dati strutturati di nuova generazione che utilizza UML. Offre un modello indipendente dalla piattaforma che può essere associato a molteplici piattaforme di implementazione. DATEX è, dunque, uno standard sviluppato per garantire lo scambio di informazioni tra i centri di gestione del traffico, i centri di informazione sul traffico e i fornitori di servizi.

La rete DATEX esistente è costituita da 50 a 60 nodi operativi in tutta Europa. La maggior parte dei nodi sono utilizzati per lo scambio di dati nazionali, altri per lo scambio internazionale.

La documentazione tecnica di DATEX ovvero il dizionario di dati (pre-standard ENV13106: 2000) e le specifiche DATEX-net per lo scambio di dati (ENV13777 pre-standard: 2000) necessitano di evoluzioni che riflettono

quelle tecnologiche raggiunte, l'esperienza acquisita nelle implementazioni che permettono lo scambio dei dati nei paesi europei e le nuove esigenze individuate dal mercato.

Il MoU, Supervisory Management Committee, Technical Committee and User Forum, conferma in modo formale che lo sviluppo dello scambio di dati di traffico internazionale dovrebbe basarsi sulle specifiche tecniche DATEX e stabilisce un quadro organizzativo che ha permesso agli utenti di influenzare e partecipare agli sviluppi.

Lo sviluppo di DATEX II è iniziato alla fine del 2003 ed è stato sostenuto e parzialmente finanziato dalla Commissione Europea in quanto giocatore di un ruolo fondamentale nel settore ITS all'interno degli Stati europei.

DATEX II si rivolge ad una base di utenti molto più ampia rispetto a quella del primordiale DATEX. Si è passati, infatti, dall'uso dei centri di controllo del traffico (enti relativi) all'uso di tutti i fornitori di servizi nel settore ITS. Il dominio dei dati è stato esteso dalle reti stradale transeuropea alle informazioni sulla rete urbana.

Le specifiche originali di DATEX presentavano una serie di carenze che hanno reso improbabile interoperabilità tra i nodi DATEX di diversi produttori. L'aggiornamento tecnologico, in questi termini, è il fattore chiave per lo sviluppo di DATEX II.

Una prima versione (1.0) è stata realizzata alla fine del 2006 per poi diffondersi rapidamente tra i paesi. Tuttavia, poiché sono state riscontrate una serie di errori e sono state sollevate necessità di cambiamento, dopo più di un anno di lavoro, si è resa disponibile la versione 2.0. La stessa si rivela abbastanza stabile. Non si escludono, comunque, ulteriori modifiche.

1.4.1 Lista e descrizione delle principale aree di utilizzo

Le principali aree di uso di DATEX II sono:

- Pianificazione di rerouting (reistradamento), gestione del traffico e della rete;
- Controllo dei ritardi e delle corsie e relative applicazioni quali misurazione delle rampe, dei limiti di velocità e controllo dei sorpassi;
- Legame tra i sistemi di gestione del traffico e gestione delle informazioni;
- Applicazioni che considerano fondamentale lo scambio di informazione tra i singoli veicoli e i gestori del traffico;
- Applicazioni che considerano fondamentale lo scambio di informazione tra modelli differenti;
- Prestazioni di servizi fortemente legati alla sicurezza stradale.

DATEX II presta particolare attenzione ai problemi di interoperabilità dovuti alla presenza di molteplici operatori coinvolti e, quindi, allo scambio senza ostacoli dei dati o delle informazioni.

1.4.2 Il contesto

DATEX II, ad oggi, tratta un ampio range di concetti relativi al trasporto e al traffico stradale al fine di diventare il modello leader per lo scambio delle informazioni sul trasporto stradale in Europa.

In particolare, tratta di informazioni relative a:

- Stato delle infrastrutture
- Lavori stradali
- Incidenti

- Tempi di viaggi
- Interruzioni, ostruzioni e ingorghi
- Misure connesse al traffico
- Eventi pubblici che impattano sul traffico
- Impostazioni sui pannelli stradali
- Impatto delle condizioni metereologiche sulla strada

1.4.3 Caratteristiche principali

La descrizione tecnica distingue gli aspetti di modellazione indipendenti della piattaforma (PIM) e gli aspetti di modellazione specifici della piattaforma (PSM).

Dunque, è stata introdotta una chiara distinzione tra la modellazione dei dati (dominio del traffico) e le specifiche per lo scambio di dati (dominio dell'ICT).

Tale approccio presenta i seguenti vantaggi:

- Separare i diversi aspetti rende più semplice la comprensione e l'applicazione degli standard;
- Gli aspetti della piattaforma indipendente possono essere considerati più stabili di quelli della piattaforma principale;
- Sono stati prodotti modelli in un linguaggio indipendente dai meccanismi di scambio e dalle tecnologie implementate

La prima versione di DATEX era "chiusa" nel senso che non permetteva di estendere lo standard. Ciò ha comportato il non utilizzo dello stesso e il ricorso a modelli non interattivi. Superare questo problema è uno dei principali requisiti di DATEX II.

Il modello di livello A

DateX II mette a disposizione un modello, chiamato “modello di livello A”, risultato di studi su dati che sono stati condivisi da molti utilizzatori in Europa.

Poiché il detto modello può mancare di caratteristiche richieste da un particolare Utente, DateX II permette di estendere lo standard (modello di livello A) introducendo i concetti mancanti (modello di livello B).

Gli utenti, tuttavia, sono autorizzati ad applicare una serie limitata di meccanismi UML ben definite al fine di mantenere l'interoperabilità tecnica con i sistemi standard DATEX II.

Ciò significa che il modello standard sarà ancora in grado di elaborare le pubblicazioni generate da un modello esteso, naturalmente senza essere in grado di elaborare il contenuto di estensione.

Il principio fondamentale del livello B è permettere agli utenti di trovare un modello appropriato per gran parte delle loro pubblicazioni. Qualora venissero introdotti concetti che vanno al di fuori del campo di applicazione del modello di livello A si introduce un nuovo modello chiamato Modello di livello C che può non essere interoperabile con il livello A.

Il modello di livello B

Nonostante la completezza del modello A, si presentano casi in cui è necessario introdurre nuove applicazioni, sia a livello nazionale che internazionale, al modello esistente.

Il modello B risponde a tali richieste. Esso arricchisce il modello A con nuove specifiche applicazioni, mantenendo l'interoperabilità con lo stesso. Al fine di estendere il modello è necessario effettuare una procedura di registrazione, qualora il modello esteso raggiunge un importante grado di consenso può essere assorbito da una nuova versione del livello A.

Il modello di livello C

Il modello C presenta caratteristiche molto diverse dal modello A, in coerenza con il motivo per cui è stato creato. Non è conforme ai modelli A e B, tuttavia, lo è nei confronti di tutte le specifiche di DATEX II.

1.4.4 Il dizionario dei dati

I precedenti standard di DATEX fornivano definizioni comprensibili solo da ingegneri del traffico ed esperti simili. Oggi, è disponibile un dizionario di dati leggibile da non professionisti dell'ambito IT tramite un tool software che estrae automaticamente le definizioni del modello UML tramite il sito web DATEX II (<http://www.datex2.eu>).

1.4.5 Lo schema XML

Lo schema XML è lo strumento utile per comprendere il contenuto dei dati scambiati. Il modello di dati UML DATEX II viene convertito per mezzo di un tool in un XML SCHEMA.

Innanzitutto il modello UML viene esportato dallo strumento di modellazione UML in un file XMI, che lo strumento di conversione converte poi a uno schema XML.

XMI (XML Metadata Interchange) è uno standard OMG utilizzato per lo scambio di informazioni tramite XML. Può essere utilizzato per qualsiasi metadato il cui metamodello può essere espresso in MOF (Meta-Object Facility). XMI è comunemente usato come formato di interscambio per modelli UML, anche se può anche essere utilizzato per la serializzazione dei modelli di altre lingue (metamodelli).

1.4.6 Meccanismi di scambio

DATEX II permette di operare in modalità push e pull nello scambio delle informazioni. La modalità push (come in DATEX) consente al Supplier di inviare informazioni al Client, mentre la modalità pull permette al Client

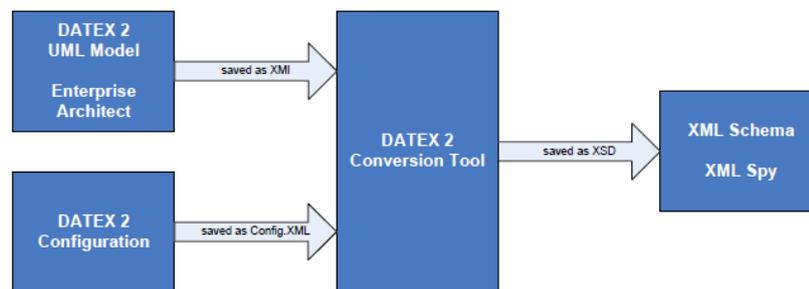


Fig. 1.16: Struttura schema Datex II

di richiedere il download di informazioni al Supplier. Lo scambio delle informazioni può avvenire secondo tre modalità operative:

- Publisher Push on occurrence: i dati vengono trasmessi quando si modificano;
- Publisher push periodic: i dati vengono trasmessi periodicamente;
- Client pull: i dati vengono trasmessi su richiesta del Client.

Ogni modalità di funzionamento può essere usata sia on che offline.

Per la modalità operativa Client Pull sono state definite due profili per operare sul web: l'uso diretto del protocollo HTTP / 1.1 o tramite Web Services su http. Per le restanti è stata definita una piattaforma mediante Web Services su HTTP.

1.4.7 Profili DATEX II

DATEX II permette di scambiare diversi tipi di dati tramite diverse modalità operative. Non tutte queste modalità devono necessariamente essere implementate in ogni sistema DATEX II. Così DATEX II consente di definire un sottoinsieme personalizzato di opzioni offerte da uno standard per una particolare esigenza.

Che cos'è un profilo? Un sistema DATEX II è costituito da diverse pubblicazioni. Ogni utente può definire un profilo in base alle proprie

esigenze, pur mantenendo l'interoperabilità su parti comuni (pubblicazioni, Modalità operative) con altri utenti, al fine di facilitare il processo di pubblicazione in quanto permette di personalizzarne le implementazioni. La realizzazione di un profilo deve seguire ad una fase di analisi dei costi – benefici, in particolare relativi a:

- Caratteristiche standard;
- Implementazione della piattaforma;
- Livello da raggiungere.

1.4.8 I documenti di DATEX II

I documenti di riferimento di DATEX II sono:

- Data model;
- Modelling methodology;
- XML schema generation tool;
- XML schema;
- Exchange PSM.

I documenti informativi che vengono utilizzati per capire DATEX II e prendere decisioni principali sono:

- Guida per l'utente
- Dizionario
- Guida per sviluppatori di software

1.4.9 Lo scambio dei dati: elementi di base

Le informazioni scambiate da DATEX II sono costituiti da elementi di base:

- Eventi relativi alle strade e al traffico, chiamati "Traffic elements";
- Azioni degli operatori;
- Impatti;
- Informazioni su eventi non relativi alla strada;
- Dati elaborati (lo stato del traffico, il tempo di viaggio);
- Dati rilevati (da stazioni).

Vengono scambiate anche informazioni relative a Location predefinite e siti.

Traffic elements

Sotto tale nomenclatura si indicano tutti gli eventi non dovuti ad azioni degli operatori del traffico ma che richiedono, tuttavia, il loro intervento.

Sono classificati in:

- Traffico irregolare;
- Incidenti;
- Attività (eventi pubblici, etc.);
- Rotture di apparecchiature (non funzionamenti dei ventilatori nei tunnel, non funzionamenti dei telefoni di emergenza, etc.);
- Condizioni: condizioni di guida relative al meteo (ghiaccio, neve, etc.) o no precipitazioni, vento, etc.);
- Ostruzioni:

- Presenza di animali;
- Presenza di veicoli;
- Ostruzioni dovuti a condizioni ambientali (alberi caduti, cadute di massi, etc.);
- Ostruzioni dovute alle infrastrutture (cadute di cavi, etc.);
- Altre ostruzioni.

Azioni degli operatori

Le azioni degli operatori sono classificate in:

- Gestione della rete: chiusura di una strada, etc.;
- Controllo del traffico: reindirizzamenti, limiti temporanei;
- Lavori stradali: asfaltamento, potatura;
- Assistenza stradale: riparazione veicoli, elicottero di salvataggio, consegna alimenti, etc.;
- Riparazione segnali.

Impatti

Contiene informazioni su corsie disponibili e ritardi.

Informazioni su eventi non relativi alla strada

Contiene informazioni relative ad eventi non collegati direttamente ad informazioni stradali: parcheggi, informazioni sui servizi di transito, interruzione dei servizi degli operatori stradali.

Dati elaborati

Questi insiemi di dati sono normalmente derivati periodicamente da dati di input sul traffico per specifiche location:

- Tempi di percorrenza: tempo di percorrenza elaborato, tempo di percorrenza senza traffico, tempo di percorrenza normalmente previsto;
- Stato del traffico, 5 possibili valori:
 - flusso libero
 - pesante
 - congestionato
 - impossibile
 - sconosciuto
- Valori sul traffico (normalmente sono rilevati da stazioni apposite) : flusso, velocità, andamento, concentrazione e misure sui singoli veicoli;
- Dati sul meteo (normalmente sono rilevati da stazioni apposite): precipitazioni, vento, temperatura, inquinamento, condizioni della superficie stradale e visibilità.

Possono essere valori stimati.

Dati rilevati

Si tratta di dati derivati da dati in input ricevuti con frequenza regolare da stazioni localizzate in particolari siti:

- Valori sul traffico: flusso, velocità, andamento, concentrazione e misure sui singoli veicoli;
- Dati sul meteo: precipitazioni, vento, temperatura, inquinamento, condizioni della superficie stradale e visibilità;

- Tempi di percorrenza (normalmente sono dati elaborati tuttavia possono essere rilevati da alcune stazioni): tempo di percorrenza elaborato, tempo di percorrenza senza traffico, tempo di percorrenza normalmente previsto;
- Stato del traffico (normalmente sono dati elaborati tuttavia possono essere rilevati da alcune stazioni): flusso libero, pesante, congestionato, impossibile, sconosciuto.

Pubblicazione degli elementi di base

Le precedenti informazioni possono essere scambiate singolarmente o in gruppo. In questi casi si utilizza la nozione di pubblicazione.

Ci sono quattro tipi di pubblicazioni principali:

- Pubblicazioni delle situazioni;
- Pubblicazioni dei dati elaborati;
- Pubblicazioni dei dati rilevati;
- Pubblicazioni sul traffico.

Nella seguente tabella si riporta cosa include ogni pubblicazione:

	Pubblicazioni delle situazioni	Pubblicazioni dei dati elaborati	Pubblicazioni dei dati rilevati	Pubblicazioni sul traffico
Traffic elements	✓			✓
Azioni degli operatori	✓			✓
Impatti	✓			
Informazioni su eventi non relativi alla strada	✓			
Stato del traffico		✓	✓	✓
Tempi di percorrenza		✓	✓	✓
Valori sul traffico		✓	✓	✓
Meteo		✓	✓	✓

Fig. 1.17: Elementi inclusi nelle pubblicazioni

Esistono altri due tipi di pubblicazioni che definiscono le caratteristiche di quelle appena citate (Pubblicazioni dei dati su location predefinite e Pubblicazione delle tabelle) Nel seguito si forniscono dettagli sulle stesse. La location, richiamata in ogni pubblicazione, verrà approfondita successivamente.

Pubblicazione delle situazioni La pubblicazione delle situazioni contiene differenti situation.

Una situation rappresenta una situazione sul traffico / viaggi e comprende una o più circostanze relative al traffico / viaggio collegate da una o più relazioni causali. Ogni circostanza relativa al traffico / viaggio è rappresentata da una Situation Record.

Ogni situation ha un unico identificatore stabilito quando viene creata nel sistema.

Se una situation è collegata con un'altra, tale collegamento può essere indicata con "relatedSituation" che diventa così l'identificatore della corrispondente situation.

Una Situation Record viene aggiornata in seguito al cambiamento dei valori contenuti, ha un unico identificatore stabilito quando è creato la prima volta.

Può contenere informazioni relative agli impatto sul traffico (quando il record è "TrafficElement" o "Azione dell'operatore"):

- Ritardi (codice, range di valori, valore)
- Altri dettagli (capacità restante, effetto sul traffico, tipo di restrizione)

Un situation record è composto da:

- Un momento iniziale (obbligatorio);
- Un momento finale (opzionale);
- Diversi momenti che attestano la validità del record (ora/giorno/settimana/mese);

- Diversi momenti che attestano la non validità del record (ora/giorno/settimana/mese);
- La definizione opzionale dello stato di “attivo” o “sospeso”;
- La conferma opzionale della prosecuzione dello stato di “attivo”.

Previsione: record la cui data di inizio è futura e pertanto non certo. Nel momento in cui il record diventa attivo il Supplier può aggiornare lo stato, se non viene aggiornato il client non può assumere che è attivo ma solo che si stima essere attivo.

Ci sono tre categorie principale di situation record:

- Informazioni sul traffico: eventi non previsti ma che hanno ripercussioni sul traffico;
- Azioni degli operatori: azioni che un operatore di traffico può decidere di applicare per prevenire o correggere le condizioni di guida pericolose, tra cui la manutenzione delle infrastrutture stradali;
- Informazioni su eventi non relativi alla strada: informazioni su un evento che non avviene sulla strada, ma che può influenzare il comportamento dei conducenti e quindi le caratteristiche del flusso di traffico.

Informazioni sul traffico Le informazioni sul traffico si dividono in sei tipi:

- **Ostruzione:** qualsiasi ostacolo fisso o in movimento di natura fisica (e.g. ostacoli o veicoli provenienti da un incidente precedente, caduta di massi, trasporti eccezionali o pericolosi, animali, etc.) che potrebbero interrompere o mettere in pericolo il traffico. Per ogni tipo di ostruzione, può essere indicato il numero di ostacoli e il tipo di mobilità(mobile/fisso).

Ci sono 5 tipi di ostacoli: presenza di animali, ostruzione ambientale, ostruzione di infrastrutture, ostruzione generali (tipo), ostruzione del veicolo;

- **Traffico irregolare:** una condizione del traffico non regolare. Possono essere dati dettagli relativi a il tipo di irregolarità, il numero di veicoli in attesa, la lunghezza della coda, flusso del traffico, le caratteristiche del flusso del traffico, trend del traffico;
- **Incidente:** eventi in cui uno o più veicoli perdono il controllo e non lo recuperano. Includono collisioni tra veicolo i , tra il veicolo s e qualche ostacolo s , o fuoriuscita dalla strada di un veicolo. Si possono fornire informazioni su causa, tipo di incidente, visione delle persone coinvolte (numero, danno, ruolo, tipo), visione dei veicoli coinvolti in base alla tipologia (numero, status, tipo, uso), dettagli dei veicoli coinvolti (tipo e singole caratteristiche del veicolo);
- **Guasto delle apparecchiature o del sistema:** rottura o malfunzionamento di un apparecchiatura un sistema che interessa gli utenti della strada;
- **Attività:** azioni umane intenzionali esterne al flusso del traffico o alla strada che potrebbero disturbarlo;
- **Condizioni:** qualsiasi condizione che può disturbare il traffico. Ci sono tre categorie di condizioni: condizioni del fondo stradale dovute a condizioni meteorologiche avverse (ghiaccio, neve, acqua), condizioni del fondo stradale non legate alle condizioni meteorologiche ma che possono influenzare le condizioni di guida (e.g. fango, olio, etc.), condizioni ambientali che possono influenzare le condizioni di guida su strada (precipitazioni, visibilità, inquinamento, temperatura, vento).

Azione degli operatori Per ciascun azione dell'operatore possono essere forniti dettagli su:

- l'origine: interno/esterno
- lo stato: richiesta, approvata, in corso di attuazione, implementata, respinta, richiesto il termine, in fase di chiusura.

Ci sono quattro tipi di azioni:

- **Lavori stradali** - attività di manutenzione, installazione e costruzione che possono interessare le operazioni di traffico. Si possono dare dettagli su durata, scala (grande, medio, piccolo), traffico, urgenza, mobilità (mobile o fissa), costruzione (brillamento, costruzione, demolizione, ampliamento stradale, etc.), manutenzione stradale, oggetto del lavoro (ponte, guardrail, cavalletto, tunnel stradale, etc.) e veicoli di manutenzione associati.
- **Modifica segnali** - provvede ad individuare indicazioni e l'informazione attualmente visualizzata.
- **Gestione della rete** - modifiche alla configurazione o utilizzabilità della rete stradale di ordine giuridico o operativo. Comprende la chiusura di strade e corsie, limiti di peso e dimensionali, restrizioni per i veicoli e le operazioni di instradamento.

Ci sono 7 tipi di gestione della rete: reinstradamento, velocità, strada/carreggiata/corsia, guida, istruzioni generali, rete generale.

L'assistenza stradale che può essere fornita riguarda la distribuzione di cibo, la riparazione dei veicoli, il soccorso con l'elicottero.

- **Assistenza stradale** - ci sono quattro tipi di eventi: disponibilità dei servizi di trasporto e le loro partenze, interruzioni di servizi (telefono per le emergenze fuori servizio, etc.), informazioni concernenti la perturbazione di servizi (mancanza di carburante, area di servizio chiusa, area commerciale chiusa, etc.), informazioni sui parcheggi (numero di parcheggi liberi, etc.).

Pubblicazioni dei dati rilevati

Questa pubblicazione viene utilizzata per inviare periodicamente dati derivanti da una stazione di traffico e relativa ad una posizione specificata. La posizione può essere esplicitamente specificata oppure può essere individuata mediante riferimenti a posizioni definite in apposite pubblicazioni.

Presentano una lista di valori che si riferiscono a valori base relativi a:

- Tempo di percorrenza;
- Stato del traffico;
- Valori del traffico;
- Condizioni metereologiche;

Pubblicazioni dei dati elaborati

Tipicamente queste pubblicazioni sono realizzate da centri di traffico. Vengono utilizzate per inviare i dati elaborati periodicamente relativamente a siti specifici identificati con riferimenti ad una voce di una tabella che per ogni sito fornisce i dettagli della posizione e i parametri associati ai diversi tipi di misurazioni che possono essere effettuati.

Contengono uno o più data set i cui attributi dipendono dalle richieste del Client. Gli item elaborati presentano un unico identificatore. La fonte può essere specificata.

Si possono fornire dati circa:

- Un momento iniziale (obbligatorio);
- Un momento finale (opzionale);
- Diversi momenti che attestano la validità del record (ora/giorno/settimana/mese);
- Diversi momenti che attestano la non validità del record (ora/giorno/settimana/mese).

Ogni dato elaborato comprende valori di base:

- Accuratezza;
- Metodo computazionale;
- Errori e ragione degli errori;
- Numero di dati in input usati;
- Periodo;
- Fattori di livellamento;
- Deviazione standard;
- Qualità dei dati;
- Momento in cui il dato è stato calcolato.

Ogni valore di base può riferirsi ad una zona ovvero ad un gruppo di location identificate tramite riferimenti.

Possono essere valori che si riferiscono a:

- Tempo di percorrenza
 - Tempo di viaggio in secondi
 - Trend del tempo di percorrenza: stabile, crescente, decrescente
 - Tipo: stimato, istantaneo, migliore, ricostituito
 - Tempo di percorrenza in assenza di traffico
 - Velocità in assenza di traffico
 - Tempo di viaggio medio
 - Tipo di veicolo associato al tempo di viaggio
- Stato del traffico
 - Stato del traffico: impossibile, congestionato, libero, sconosciuto, pesante

- Trend: traffico stabile, crescente, decrescente, sconosciuto
- Condizioni climatiche (informazioni sulle precipitazioni, vento, temperatura, inquinamento, condizioni stradali, visibilità, umidità)
- Valori sul traffico
 - Avanzamento
 - Flusso
 - Velocità
 - Concentrazione
 - Misure sui singoli veicoli.

Publicazioni dei dati su location predefinite

Queste pubblicazioni vengono usate per predefinire un set di location cui ci si può riferire nelle altre pubblicazioni al fine di semplificarne la dimensione.

Esistono tre tipi di location:

- **Point location**, è utilizzato per localizzare fenomeni la cui lunghezza è trascurabile, come ad esempio incidenti, ostacoli, etc.;
- **Linear location**, è utilizzato per localizzare determinati fenomeni o azioni dell'operatore la cui lunghezza è rilevante, come ad esempio per il flusso anomalo di traffico (code, etc.), il cattivo stato delle strade (e.g. superficie stradale), etc.;
- **Area location**, è utilizzato per localizzare alcuni fenomeni superficiali quali fenomeni meteorologici o fornire indicazioni su una destinazione.

Publicazione delle tabelle

Questa pubblicazione contiene uno o più tabelle costituite a sua volta da record che presentano i seguenti attributi riferiti ad una singola location:

- Metodo computazionale;
- Riferimento allo strumento di misurazione;
- Tipo di strumento di misurazione;
- Nome del sito di misurazione;
- Numero di corsie;
- Riferimento al sito di misurazione;
- Lato misurato;
- Versione del record;
- Lista delle caratteristiche di misurazione:
 - Accuratezza
 - Periodo
 - Fattori di livellamento
 - Corsia
 - Tipo dei valori rilevati
 - Caratteristiche del veicolo

Pubblicazione sul traffico

La pubblicazione sul traffico contiene una snapshot di ciò che accade su un percorso, in una direzione, in un dato momento. Contiene tutte le informazioni necessarie per capire cosa sta succedendo in un itinerario in un dato istante.

L'aggiornamento del record comporta l'eliminazione di quello precedente, dunque non si tiene traccia del passato.

Si può riferire ad un set di location e contiene uno o più traffic view, ognuna delle quali riporta:

- Elementi del traffico;

- Azioni degli operatori;
- Dati elaborati (stato del traffico, tempo di percorrenza, condizioni sul traffico e sul meteo);
- Fonte.

1.4.10 UML data model

Il modello si compone di diversi diagrammi:

- Schemi di analisi;
- Diagrammi dinamici;
- Schemi funzionali;
- Schemi logici.

Questi schemi sono stati progettati quando è stato costruito il Modello Indipendente dalla Piattaforma.

Schemi logici

Gli schemi logici sono suddivisi in 4 pacchetti:

- **Exchange:** contiene gli schemi di scambio e le definizioni sulle enumerazioni di scambio;
- **Payload:** contiene le descrizioni di dati di pubblicazione da scambiare;
- **General:** contiene le "General Classes" (tipi di dati, le classi di localizzazione, Payload enumeration, classi riutilizzabili);
- **Management:** contiene la gestione della modalità "Publisher Push on occurrence".

D2DLogicalModel

Il D2DLogicalModel comprende due componenti: Exchange e Payload. Exchange è usato per scambiare i parametri associati ai dati consegnati. Payload comprende le pubblicazioni che includono dati relativi al traffico e ai percorsi.

Ogni pubblicazione deve mostrare quando è stata creata, il linguaggio di default utilizzato e chi l'ha creata. Quest'ultimo può essere differente da chi la pubblica.

1.4.11 Specifiche sullo scambio di informazioni

Lo scambio di informazioni tra il Supplier e il Client è composto da due sottosistemi principali:

- Un sottosistema Publisher, che rende i dati disponibili e crea le pubblicazioni payload (richieste);
- Un sottosistema Delivery, che aggiunge informazioni specifiche sullo scambio ed esegue la consegna fisica.

Si può definire una sottoscrizione che definisce cosa scambiare dal Supplier o dal Client. DATEX II consente agli utenti di avere la libertà di svilupparla con le raffinatezze di cui hanno bisogno.

Ci sono 3 possibili modalità operative per la trasmissione dei dati:

- **Publisher Push on occurrence:** il Publisher aggiorna i dati in simultanea al cambiamento;
- **Publisher push periodic:** il Publisher aggiorna i dati periodicamente;
- **Client pull:** il Client chiede i dati.

L'aggiornamento dei dati può avvenire secondo tre modalità:

- **singleElementUpdate**: l'aggiornamento riguarda solo il singolo dato che ha subito modifiche;
- **allElementUpdate**: l'aggiornamento riguarda tutta la stringa legata al dato che ha subito modifiche;
- **snapshot**: si fornisce una nuova pubblicazione.

Con la modalità Client pull si usa solo la modalità snapshot. In base alle sue esigenze, lo stakeholder decide le modalità da implementare.

1.4.12 Gli accordi sullo scambio

L'accordo sullo scambio è un documento nel quale si descrivono bilateralmente gli accordi presi dal Supplier e dal Client nello scambio delle informazioni. Si assume che lo scambio segua le specifiche di base di DATEX II. In particolare si stabilisce:

- Il periodo di validità dell'accordo;
- Le condizioni da verificarsi al fine di considerare il contratto terminato prima della scadenza;
- I dettagli sul nome, indirizzo, telefono, fax e e-mail delle compagnie/autorità;
- L'IP address, il nome e la password del Client e del Supplier;
- Le scelte tecniche necessarie per lo scambio dei dati e la configurazione;
- La qualità dei dati forniti dal Supplier;
- Le azioni che il Supplier e il Client devono intraprendere quando il sistema mostra problematiche;
- I meccanismi di pagamento;
- Le responsabilità in caso di trasmissione di informazioni errate;

- Le regole per l'ulteriore diffusione da parte del Client delle informazioni ricevute tramite DATEX II.

1.5 Il ruolo di Telecom Italia

Telecom Italia, dal 2013, è coinvolta in tre progetti Europei (COMPASS4D, TEAM e MOBiNET) che, in varia misura, sviluppano e testano sul campo soluzioni per l'acquisizione, il trattamento e la valorizzazione di dati per il mondo ITS.

1.5.1 MOBiNET

Il progetto ha l'obiettivo di finalizzare lo sviluppo di un framework pre-operativo distribuito a livello EU (in tecnologia cloud) in grado di supportare un marketplace ITS aperto per l'offerta di dati (contenuti) e nuovi servizi (applicazioni) interfacciandosi anche con soluzioni ed infrastrutture legacy. Il modello di collaborazione tra gli stakeholder è primariamente di tipo B2B2C. Nel progetto si intende anche testare pre-operativamente le possibili regole di governance tra i membri del framework MOBiNET e delle modalità di esercizio operativo del framework. Il processo di test e validazione prevede un'interazione con le amministrazioni locali e/o i loro enti strumentali responsabili per la gestione della mobilità nelle aree metropolitane e le città coinvolte nel pilota pre-operativo. In Italia, la sperimentazione si svolgerà a Torino con funzionalità e complessità crescente a partire dalla metà del 2014. [3]

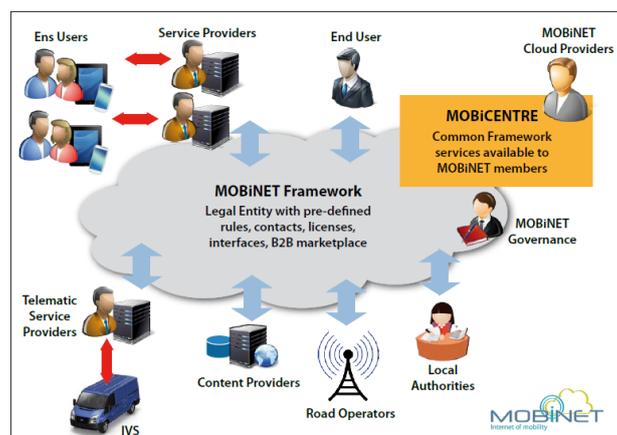


Fig. 1.18: Architettura funzionale MOBiNET

1.5.2 TEAM

L'obiettivo generale del progetto TEAM è di creare, testare, dimostrare e valutare un sistema di gestione della mobilità dinamica e collaborativa. In dettaglio, gli obiettivi tecnici sono:

- Sviluppare e testare processi decisionale/collaborativo e algoritmi di ottimizzazione;
- Sviluppare elementi abilitanti di una Cloud per servizi automotive;
- Sviluppare elementi per la gestione dinamica della mobilità;
- Creare meccanismi di partecipazione di conducenti e viaggiatori;
- Quantificare le prestazioni tecniche e gli impatti operativi;
- Promuovere la mobilità collaborativa.

In questo progetto, Telecom Italia intende mettere a disposizione e testare la propria soluzione prototipale per l'acquisizione di dati dinamici sulla mobilità per mezzo di meccanismi social (ITS2.0) e la relativa creazione di contenuto informativo ITS che possa essere utilizzato come abilitatore per la creazione di nuovi servizi a supporto della mobilità. In Italia, la sperimentazione si svolgerà a Torino ed a Trento a partire dalla fine del 2014. [3]

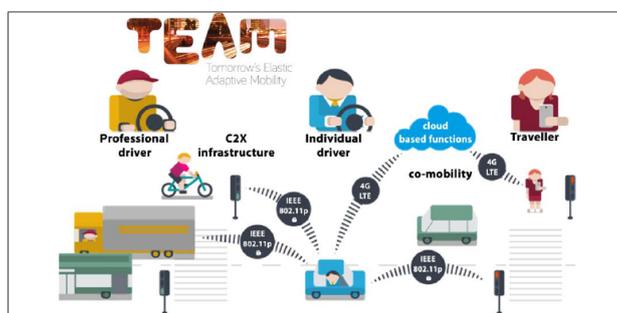


Fig. 1.19: Ecosistema TEAM

1.5.3 COMPASS4D

Il progetto COMPASS4D ha l'obiettivo primario di coinvolgere direttamente le Amministrazioni Locali nella fase di messa in esercizio pre-operativa e nel testing delle soluzioni innovative di ITS cooperative (C-ITS), in cui tre elementi funzionali: l'infrastruttura stradale, il veicolo ed i centri servizio si scambiano dati ed informazioni utilizzando modalità di comunicazione sia a corto raggio (comunicazione V2I a 5.9GHz) che a lungo raggio (comunicazione su rete mobile).

Il progetto è organizzato attorno ad una serie di Pilota nazionali che coinvolgono direttamente le città.

Il pilota italiano si svolge a Verona ed è mirato a valutare in condizioni realistiche le performance di comunicazione, acquisizione dati e distribuzione di informazioni ITS sia utilizzando comunicazione a 5.9GHz che LTE. [3]



Fig. 1.20: Allestimento pilota COMPASS4D

CAPITOLO 2

TELECOM ITALIA: LA PIATTAFORMA "ITS 2.0"

La piattaforma "ITS 2.0" di Telecom Italia rientra nell'ambito delle piattaforme *social* per la mobilità, ovvero quelle infrastrutture informatiche capaci di raccogliere dati di infomobilità in tempo reale dai membri di una comunità virtuale mediante un approccio social di *crowdsourcing*; tali informazioni vengono poi rese fruibili attraverso un software (nella fattispecie, una mobile app per smartphone) al fine di condividere notizie connesse alle condizioni del traffico e, più in generale, alla mobilità per rendere omogeneo il flusso degli utenti delle reti di viabilità ed evitare preventivamente il verificarsi di fenomeni di congestionamento.

Il concetto che è alla base del funzionamento di simili soluzioni consiste nell'assegnazione di un valore di *reputation* (reputazione) all'utente responsabile della segnalazione; tale valore rappresenta l'indice della veridicità dell'utente: più alto è il valore della sua *reputation*, maggiore è l'aderenza alla realtà delle segnalazioni da esso inviate.

La determinazione del valore di reputazione associata ad un certo membro avviene per mezzo del processo di *validation* (validazione) delle

segnalazioni che coinvolge gli altri utenti con l'invio di conferme/smentite dell'evento segnalato sulla base di quanto si è riscontrato dal vivo nella zona oggetto del report.

La situazione As-Is della piattaforma in esame evidenzia una architettura nella quale singole componenti interagiscono per perseguire gli scopi prefissi nella concettualizzazione dell'infrastruttura.

In tale contesto, è stato richiesto un miglioramento della stessa tenendo in considerazione la struttura generale del backend attualmente in essere (la parte applicativa residente nel server e deputata alla gestione della persistenza dei dati e degli accessi) ma migliorandola attraverso l'integrazione con Twitter; dal social network in questione vengono reperiti tweet di infomobilità da prestabiliti canali istituzionali per mezzo di un aggregatore già sviluppato, tali segnalazioni vengono poi inserite nella banca dati dei report parallelamente ai contributi degli utenti base.

2.1 Architettura piattaforma

Lo schema sottostante mostra l'architettura generale del sistema. Distinguiamo principalmente quattro differenti componenti:

- **Componente Data Ingestion** che gestisce l'inserimento delle segnalazioni degli user e, a valle dell'integrazione con Twitter, dei tweet di infomobilità parsificati opportunamente all'interno del database della piattaforma. A tale scopo è deputata un'apposita API di Data Ingestion trattata nell'appendice del presente documento.
- **Componente User** che memorizza le informazioni sugli user, esegue l'aggiornamento della loro reputation e gestisce gli eventuali meccanismi di rewarding. La gestione degli accessi viene effettuata per mezzo dell'IdM.
- **Componente Data Validation**, facente parte della componente Core, che si occupa di validare le segnalazioni ricevute dagli user.

- **Componente Lifetime**, facente parte della componente Core, che regola il ciclo di vita relativo ad ogni segnalazione; la segnalazione viene eliminata dalla banca dati al termine del tempo medio di persistenza della categoria di evento segnalato (noto a priori).

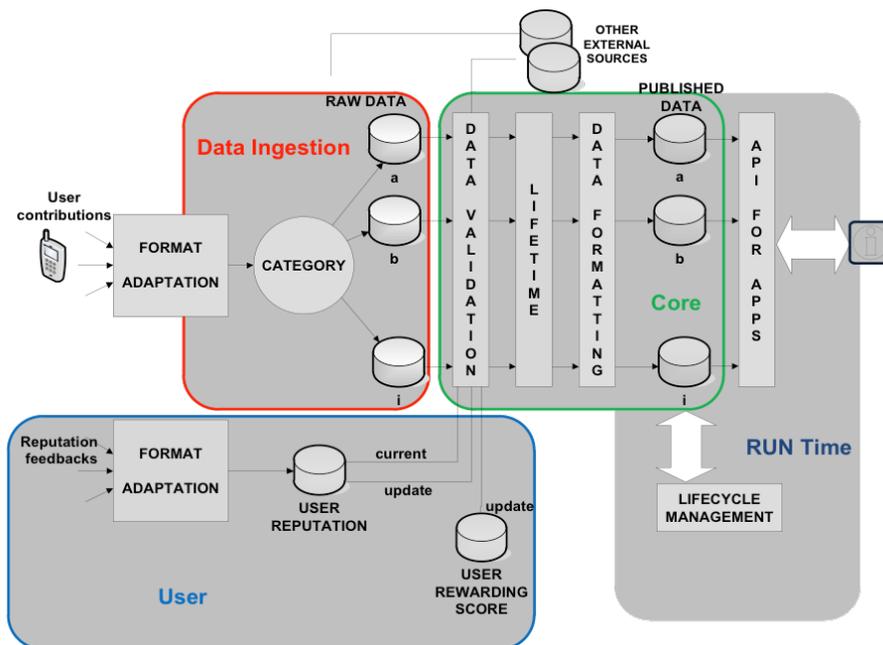


Fig. 2.1: Architettura Funzionale della Piattaforma ITS 2.0

2.2 Identity Manger

L'*Identity Manager* (IdM) descrive la gestione dei singoli utenti, l'autenticazione, l'autorizzazione e dei privilegi all'interno del sistema con l'obiettivo di aumentare la sicurezza e la produttività riducendo i costi, i tempi di inattività e azioni ripetitive. L'IdM si occupa di come gli utenti acquisiscono un'identità, della protezione di tale identità e delle tecnologie di supporto alla protezione (e.g. protocolli di rete, certificati digitali, password, etc.)

2.2.1 Autenticazione e autorizzazione

Nell'IdM sono state implementate non solo le funzioni di autorizzazione, ma anche quelle di autenticazione, anche se non sono state originariamente previste nel sottosistema IdM.

In generale, le funzioni di controllo accessi coinvolgono sia gli utenti che le applicazioni.

Da un lato, un utente specifico può operare in modo diverso all'interno della piattaforma e utilizzare dei servizi diversi a seconda dei suoi ruoli e/o di alcuni attributi.

Dall'altro, un'applicazione specifica deve essere registrata dal suo sviluppatore per essere in grado di utilizzare i servizi forniti dalla piattaforma che richiedono autorizzazione. I servizi sono definiti nella Service Directory e l'IdM fornisce sia un server di autorizzazione (e.g. OAuth 2.0 - Client Credential grant) sia un sito di registrazione del cliente per essere utilizzato dagli sviluppatori per registrare i nuovi clienti.

Il controllo accessi a specifici servizi interni potrebbe essere role-based, determinato dalla piattaforma o dal proprietario dell'entità: queste informazioni sul profilo possono essere fornite anche durante la fase di autenticazione OpenID.

L'Identity Manager, per ogni entità, entity user, deve prendere in considerazione i seguenti tipi di dati del profilo:

- **Identity Profile:** chi è? Gli attributi comunemente presi in considerazione da parte di un authentication server sono nickname, nome, cognome, email, data di nascita, sesso, CAP, paese, lingua e fuso orario. Tutti questi dati possono essere gestiti dal server LDAP e il server OpenID, che è in grado di leggere ed esporli in un messaggio di risposta nell'ambito della procedura di autenticazione standard. L'interfaccia EntityManagement permette di creare/modificare/cancellare gli attributi;
- **Permission Profile for a user:** che cosa può fare? Questo è gestito dall'EntityAuthorizationManager attraverso l'interfaccia

EntityAuthorizationManagement. A seconda del ruolo (e.g. PlatformAdmin, PartyAdmin, PartyUser con attributi specifici di accesso, DeveloperUser, per l'utente finale) e/o altri attributi, le autorizzazioni potrebbero essere fornite per l'utente;

- **Permission Profile for a client:** che cosa può fare l'applicazione? Questo è gestito dal ClientManager attraverso l'interfaccia ClientAuthorizationManagement. A seconda dello sviluppatore, i permission token potrebbero essere previsti per quel tipo di cliente che può essere identificato dall'Authorization Server;
- **Application Profile:** questo profilo è specifico per ogni applicazione ed è rappresentato dagli attributi i-scope di un Party che possiede quella applicazione. Essi sono memorizzati nel modo in cui l'applicazione trova più opportuno, eventualmente (ma non necessariamente) in un preference database dell'IdM o nel dispositivo mobile.

Per dare l'autorizzazione a qualsiasi applicazione che richiede di utilizzare i servizi della piattaforma, vi è la necessità di fornire un modo per registrare le applicazioni e quindi un processo di autorizzazione in fase di esecuzione: in particolare, per tale scopo, è stato implementato uno dei casi d'uso del protocollo *OAuth* (Client Credentials Grant).

Qualsiasi applicazione che vuole usare dei servizi, che richiede l'autorizzazione del cliente, deve essere registrato dallo sviluppatore al fine di utilizzare i servizi necessari.

Questa esigenza ha bisogno di una registrazione di tale domanda verso un Authorization Provider (utilizzando il sottomodulo Client Manager), che darà allo sviluppatore un clientID e un clientSecret (per gli scopi richiesti), che sarà inserito nel codice di tale domanda.

Poi, quando l'applicazione vuole utilizzare uno dei servizi per cui l'applicazione è stata registrata (cioè gli scope), chiede all'Authorization Server un token da utilizzare in alcuni degli scope per cui l'applicazione è stata registrata.

Questo dovrebbe essere il token che verrà inserito, per esempio, nella chiamata del REST Service: il servizio potrebbe quindi verificare la validità del token contattando l'Authorization Server stesso, e quindi fornire il servizio richiesto qualora il token venisse validato.

2.2.2 Ruoli

Si distinguono i seguenti due tipi di ruoli:

- I **service role**, ovvero i ruoli che sono specifici di un singolo servizio e che sono utilizzati dal servizio per personalizzare le sue funzionalità. Un service role è significativo solo nel settore IT/Service nel quale è specificato;
- I **business role**, ovvero i ruoli che sono specifici di una singola attività (party). I business role sono utilizzati dall'azienda per distribuire le responsabilità. Un business role è significativo solo nel business domain nel quale è specificato. Le aziende possono decidere di mantenere una corrispondenza tra i loro business role e service role esistenti, quindi attivare o disattivare i loro business officer per eseguire funzionalità relative ai service role. Si noti che l'azienda può cambiare tali mappature a volontà, senza richiedere modifiche nei servizi (software).

In questa trattazione verranno approfonditi unicamente i service role.

Service role

Nella piattaforma sono stati identificati i seguenti service role:

- **PlatformAdmin**: unico amministratore della piattaforma in grado di gestire tutti gli utenti. In particolare, può creare PartyAdmin;
- **PartyAdmin**: uno o più amministratori del dominio di un determinato Party (e.g. Service Provider). In particolare, può creare DeveloperUsers e PartyUsers;

- **DeveloperUser**: può registrare le sue applicazioni al fine di utilizzare i servizi della piattaforma, utilizzando i servizi della piattaforma;
- **PartyUser**: utente di un Party;
- **EndUser**: cliente generale.

PlatformAdmin C'è solo un utente con questo ruolo all'interno del sistema ed il suo account è cablato nella piattaforma. Pertanto, per l'avvio della piattaforma è presente un utente predefinito con il ruolo di PlatformAdmin. La sua password di default dovrebbe essere cambiata da lui stesso dalla sua pagina di profilo. Non è ancora chiaro se ci dovrebbe essere un requisito, per cui un eventuale accesso simultaneo di questo PlatformAdministrator dovrebbe essere consentito o meno.

Il PlatformAdmin ha i permessi per:

- Modificare il suo profilo e la password.
- Gestire tutti gli utenti (ad esempio, abilitare/disabilitare e creare/modificare/eliminare utenti PartyAdmin e forse altri, come gli EndUser).
- Creare utenti PartyAdmin, almeno uno per ciascuna Party (Service Provider).

PartyAdmin Viene creato dal PlatformAdmin. Ogni Party può avere uno o più utenti con questo ruolo. Deve cambiare la sua password di default dalla pagina del profilo.

Ha i permessi per:

- Modificare il suo profilo e la password;
- Creare DeveloperUser e/o PartyUser e gestire (ad esempio, abilitazione/disabilitazione, forse modificare/cancellare).

PartyUser/DeveloperUser Un PartyAdmin della propria Party è in grado di creare sia un utente con il ruolo di PartyUser che di DeveloperUser. Per il momento, non vi è alcuna differenza nella gestione di questi due ruoli utente, anche se il DeveloperUser, in futuro, dovrebbe avere degli attributi e privilegi diversi (e.g. poter registrare i servizi e/o app). Ci sono diversi utenti con questi ruoli per ogni Party.

EndUser Un EndUser può essere registrato:

- **Direttamente dall'EndUser stesso:** l'EndUser può unire l'account con quello di uno più partner della piattaforma e/o social network (e.g. Google, Facebook, Microsoft). Ma questa caratteristica non è stata ancora implementata.
- **Da un Party:** un PartyAdmin è in grado di registrare alcuni (o tutti) dei suoi clienti, affinché possano avvantaggiarsi con l'utilizzo di alcune delle caratteristiche fornite, ma questa caratteristica non è stata ancora implementata.

2.2.3 Struttura implementazione moduli IdM

Dopo aver definito una struttura di ruolo utente, è stata implementata sia un'entità (e.g. user) di gestione che un provider OpenID per l'autenticazione: per gestire i dati del profilo identità rilevati, viene utilizzato un server LDAP, mentre per contenere ulteriori dati profilo relativi sia ad identità che a servizi viene utilizzato un postgres DB.

L'Identity Manager gestisce tutte le informazioni che sono memorizzate nel sottosistema IdM, ma quest'ultimo non si assume alcuna responsabilità per il significato dei dati di Uniform Resource Identifier (URI) che sono memorizzati. Il responsabile è il partner che ha registrato tali dati. Inoltre, l'IdM non cancella nessun dato che è stato registrato, questo dovrebbe essere fatto da parte dell'ente che ha registrato i dati con una sola eccezione. Quando un Party lascia l'ecosistema, tutti i dati registrati, per cui è stato responsabile, verranno eliminati. Le seguenti figure, che mostrano i tre

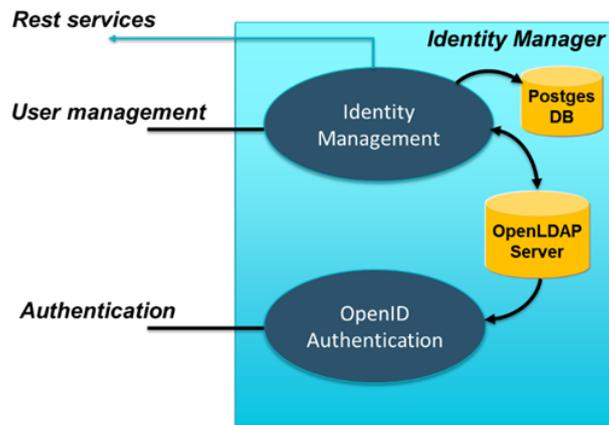


Fig. 2.2: Caratteristiche autenticazione

livelli dell'architettura Java EE6, dà una panoramica della struttura della piattaforma utilizzata per sviluppare i moduli IdM.

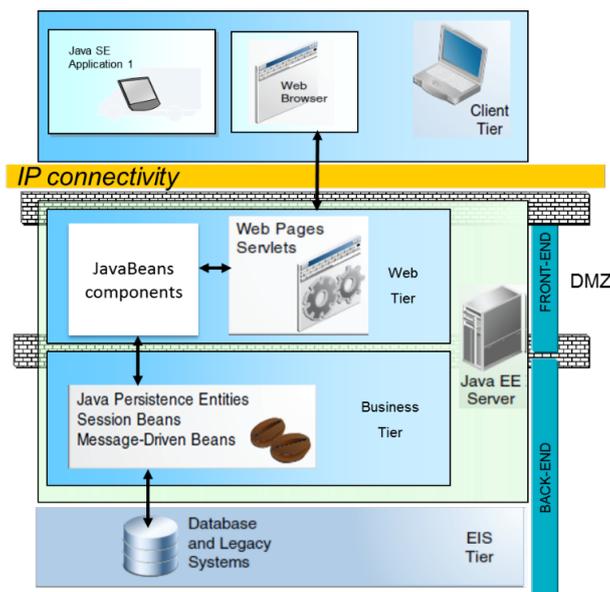


Fig. 2.3: Struttura piattaforma IdM

2.2.4 Modulo di Identity Management

Il modulo di Identity Management prende in considerazione, per ora, solo le entità utente e quelle TSP/IP.

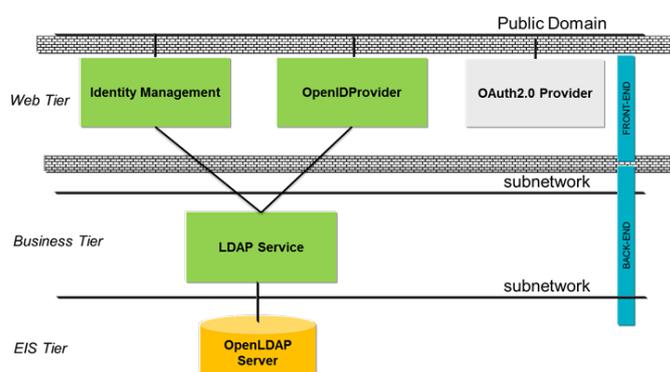


Fig. 2.4: Moduli IdM

Gestisce i profili di identità per i ruoli definiti supervisionati nella piattaforma: PlatformAdmin, PartyAdmin, DeveloperUser, PartyUser e EndUser.

Dopo il login al sito, saranno disponibili caratteristiche differenti a seconda del ruolo dell'utente connesso.

In generale, il sito permette di creare nuovi utenti (e poi eventualmente disabilitare alcuni di essi) a seguito di una procedura gerarchica top-down: PlatformAdmin può creare utenti PartyAdmin, PartyAdmin può creare DeveloperUsers e PartyUsers. Gli EndUser possono solamente registrare loro stessi.

Il livello Web è implementato sulla base di JSF 2.0 servlet, pagine XHTML e managed JavaBeans. Il livello di business si basa su un EJB 3.1 singleton (ApplicationConfig) e un EJB 3.1 Session (servizio LDAP).

2.2.5 Autorizzazione basata su client credentials

L'autenticazione con "chi sei?", e quindi la gestione del profilo Entity Identity permette, per esempio, che un utente possa essere autenticato da alcuni siti partner (Relying Parties o RP) utilizzando un servizio di terze parti (Identity Provider o IdP), eliminando la necessità di fornire dei sistemi ad hoc. Permette, inoltre, agli utenti di consolidare le loro identità digitali. Dall'altra parte l'autorizzazione con "che cosa puoi fare?" definisce i diritti di accesso alle risorse:

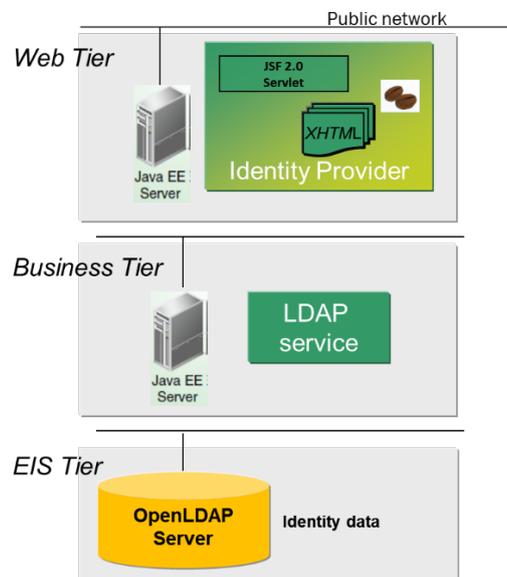


Fig. 2.5: Modulo IdM

- Può fornire un metodo ai clienti per accedere alle risorse del server per conto di un proprietario della risorsa (e.g. un client diverso o un EndUser).
- Può fornire, anche, un processo agli EndUser per autorizzare l'accesso di terzi alle risorse del server senza condividere le proprie credenziali (in genere un nome utente e una password), utilizzando un reindirizzamento di user-agent.

Il protocollo **OAuth 2.0** è stato scelto per fornire alcune funzionalità di autorizzazione all'interno dell'IdM, così nel seguito vi è una breve analisi delle di tale protocollo e di come nasce.

OAuth

OAuth, Open Authorization, è un protocollo standard di comunicazione mediante il quale è possibile gestire l'accesso di applicazioni terze alle risorse di un utente senza conoscerne le credenziali.

Sono due le parti coinvolte nel funzionamento di OAuth: il Service Provider, vale a dire il Web Service (quello cioè nel quale risiede l'account

Utente) che fornisce l'accesso tramite il protocollo OAuth, ed il Consumer, ovvero sia l'applicazione o il sito web che utilizza le specifiche OAuth per ottenere accesso presso il Service Provider.

Il protocollo OAuth permette lato utente di condividere alcune sue risorse (e.g. foto, video, elenchi, report) memorizzate in un Service Provider da un Consumer.

Tale comunicazione avviene senza dover scambiare le reciproche credenziali, tipicamente il nome dell'utente e la sua password; l'utente così ha la possibilità di non divulgare la sua identità digitale a terzi.

Il protocollo OAuth permette al Service Provider di pubblicare ed interagire con dati protetti e gestire la granularità delle risorse che vengono messe a disposizione dell'applicazione ed eventualmente revocare permessi; il tutto senza la necessità di cambiare le credenziali degli utenti.

OAuth 2.0

Analizziamo, ora, il comportamento di OAuth 2.0, descrivendo prima gli attori coinvolti e poi come questi interagiscono tra loro. La gestione degli access token sarà approfondita nel paragrafo successivo.

Gli attori coinvolti:

- **Protected resource:** la risorsa che sta sul server per la quale si richiede l'autorizzazione;
- **Resource Owner:** colui che autorizza o meno l'accesso alla risorsa protetta;
- **Client:** l'applicazione che chiede di accedere alla risorsa protetta;
- **Server:** server che protegge la risorsa. Esistono due tipi di server, quello che autorizza il client, Authorization Server, e quello che detiene la risorsa, Resource Server. Questi possono essere sia due entità diverse sia entità che coesistono sulla stessa macchina.

In Figura 2.6 è possibile osservare la dinamica delle iterazioni tra gli attori coinvolti, descritta nella seguente lista di passi successivi:

1. Il Client, registrato presso l'Authorization Server, richiede l'autorizzazione al Resource Owner. L'autorizzazione può essere richiesta direttamente al Resource Owner o preferibilmente, come si vede dalla Figura 3, indirettamente attraverso l'Authorization Server.
2. Nel caso in cui Resource Owner concede l'autorizzazione, il Client riceverà l'Authorization Grant che rappresenta l'autorizzazione del Resource Owner. La specifica definisce quattro possibili Grant: authorization code, implicit, resource owner password credentials e client credentials. Questi dipendono dalle modalità usate dal Client per richiedere l'autorizzazione e dai tipi di autorizzazione supportati dall' Authorization Server. Gli Authorization Grant saranno analizzati nel paragrafo successivo.
3. Il Client richiede un access token all'Authorization Server, autenticandosi e presentando l'Authorization Grant.
4. L'Authorization Server autentica il Client, valida l'Authorization Grant in caso positivo, restituisce un access token ed eventualmente un refresh token.
5. Il Client richiede la risorsa protetta al Resource Server presentando l'access token.
6. Il Resource Server valida l'access token e serve la richiesta.

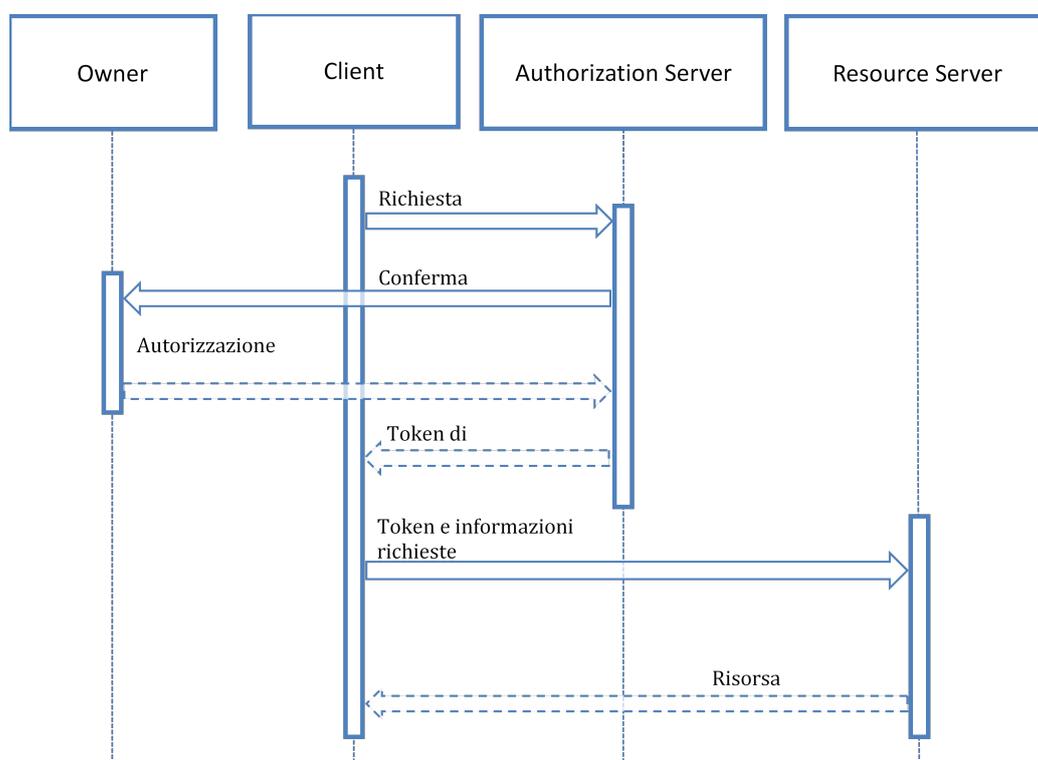


Fig. 2.6: OAuth 2.0

Authorization Grant

L'Authorization Grant sono le credenziali utili per acquisire un access token. Queste vengono richieste al Resource Owner dal Client, ne esistono 4 tipi diversi di seguito descritti:

- **Authorization Code**

Per ottenere un Authorization Code viene usato un Authorization Server come intermediario, in questo modo il Client e il Resource Owner non dovranno scambiarsi le proprie credenziali.

- **Implicit**

L'Implicit Grant è un metodo di autorizzazione semplificata rispetto all' Authorization Code, in questo caso il Client ottiene direttamente l'access token, senza avere, come nel caso precedente, un Authorization code. In questo caso il Client non si autentica presso l'Authorization Server, ma l'identità viene verificata attraverso l'URI di reindirizzamento dove verrà fornito l'access token del Client.

- **Resource Owner e Password Credential**

Le credenziali del Resource Owner possono essere usate direttamente per ottenere un access token. Usato quando gli altri metodi di autorizzazione non sono disponibili o quando vi è un alta affidabilità fra il Client e il Resource Owner.

- **Client Credential**

Le credenziali del client vengono usate quando il Client è anche il Resource Owner o se la richiesta di autorizzazione era già stata fatta precedentemente dall'Authorization Server. In questi casi le credenziali del client possono essere usate direttamente per l'autorizzazione e ottenere l'access token.

Authorization Code Grant. L'Authorization Code Grant può essere usato per ottenere sia access token che refresh token. Il Client deve essere

capace di interagire con l'User-Agent del Resource Owner (tipicamente un web browser) e capace di ricevere richieste in ingresso attraverso il reindirizzamento dell'Authorization Server. Le comunicazioni dovrebbero essere fatte attraverso un canale sicuro come TLS. L'Authorization Code deve avere un life-time breve ed è possibile usarlo una sola volta. Se l'Authorization Server riceve più volte uno stesso Authorization Code per avere un access token, allora dovrebbe revocare tutti gli access token che sono stati rilasciati con quell'Authorization Code.

Implicit Grant. È la semplificazione della modalità precedente, ottimizzata per Client implementati in browser che usano script o dispositivi mobili. È necessario laddove il Client non abbia modo di poter conservare un access token o un secret. Si è cioè nel caso, ad esempio, di applicazioni che si riducono a codice JavaScript che viene interpretato in un browser. In un tale contesto, il flusso precedente, con tutti i suoi scambi di codici e refresh token, diventa inutile e ridondante. Per questi Client si adotta allora il flusso implicito. Infatti, in questo flusso il token viene fornito al client senza alcuna richiesta esplicita. Pertanto questa volta il Client invia meno parametri al server in quanto non è richiesta l'autenticazione del Client. Anziché ritornare un authorization code, il server riceve direttamente l'access token, nel fragment hash dell'URI. A questo punto, il resto della logica avviene nello JavaScript all'interno del browser del client, che infatti svolge il ruolo di user agent. Questo esegue un reindirizzamento al server senza inviare il fragment contenente il token. Il server, a questo punto, risponde con una pagina web (HTML5) in cui c'è il link ad una libreria JavaScript che consente allo user agent, cioè al browser, di estrarre l'access token dal fragment hash. Il client adesso dispone del token per accedere alle risorse del provider.

Resource Owner Password Flow. Il Client fornisce le sue credenziali e riceve in cambio un access token ed un refresh token. Tale modalità è da implementare solo quando l'applicazione Client è a tutti gli effetti fidata. Ciò detto, da un punto di vista funzionale, il flusso è molto semplice. Il

Client chiederà subito le credenziali e l'utente dovrà inserirle direttamente nell'applicazione. A questo punto è necessario che l'utente sia certo che l'applicazione client sia davvero quella ufficiale. L'importante è che sia certo di potersi fidare dell'applicazione client e di potervi inserire le credenziali.

Client Credentials Grant. Questo flusso è assai simile al precedente. In sostanza il flusso consiste di una singola "request and response" come il precedente. Solo che in questo caso si inviano solo *clientId* e *clientSecret*. L'authorization server risponderà direttamente fornendo l'access token ma non un refresh token. Ciò implica che, allo scadere dell'access token, il client dovrà semplicemente richiederne uno nuovo fornendo *clientId* e *clientSecret*.

Gestione dei Token. Nel protocollo OAuth il Client usa gli access token per avere accesso alla risorsa protetta. Gli access token non sono visibili al Client e contengono tutte le informazioni relative all'autorizzazione. Nella versione esaminata la proprietà life-time di un access token assume un ruolo fondamentale, una volta scaduta per accedere alla risorsa sarà necessario richiedere un nuovo token. Il Resource Server gestisce il life-time e, allo scadere del quale, notifica l'errore al Client che dovrà rinnovare la richiesta per ottenere un nuovo access token. In determinati casi l'Authorization Server, rilascia al Client sia un access token che un Refresh Token; quest'ultimo sarà inviato dal Client al Server al fine di ottenere come risposta un access token valido accompagnato da un nuovo Refresh Token. I Refresh Token vengono rilasciati solo alle applicazioni web e alle applicazioni client-native.

Attualmente, la piattaforma fa uso del protocollo OAuth 2.0 (Clients Credential Grant).

2.3 Meccanismi di Validation

La validazione è un processo con il quale si verifica l'attendibilità di un dato.

Nella creazione di un social networking per i trasporti tale concetto assume un'importanza fondamentale, in quanto ogni segnalazione fornita da un singolo user ha bisogno di essere approvata prima di essere messa a disposizione della community.

Questo anche al fine di non fornire informazioni false che potrebbero procurare, in alcuni casi, falsi allarmi o diminuire la credibilità del sistema.

Per questo motivo, ogni volta che una segnalazione viene pubblicata, questa subisce il processo di validazione seguente.

Per ogni valore di reputation associato ad uno user corrisponde un numero di user necessari alla validazione della segnalazione.

Più nel dettaglio, l'algoritmo ideato propone di validare la segnalazione qualora si raggiunga, tramite il meccanismo di Event Confirm, un certo numero di user che confermano tale segnalazione, calcolato tramite la seguente formula:

$$NumeroUser = \frac{K \cdot (100 - rep)}{10}$$

Dove:

- K è una costante che serve a "tarare" il numero di user necessari in caso di uguale reputation fra due o più user
- rep è la reputation dello user che ha effettuato la segnalazione ed ha un valore compreso fra 0 e 100

Si è considerato anche che il numero di user sopra citato possa variare in base alla reputation di chi sta confermando la segnalazione in accordo alle condizioni seguenti:

$$NumeroUserNew = \min(NumeroVotanti(like); NumeroVotanti - 1)$$

Dove:

- *NumeroVotanti* è il numero di utenti necessari alla validazione
- *NumeroVotanti(likers)* è il numero di utenti necessari alla validazione per lo user che conferma la segnalazione.

Allo stesso modo l'algoritmo propone un meccanismo uguale a quello descritto nel caso in cui la segnalazione voglia essere smentita, cioè considerata non vera. In questo caso il numero di utenti necessari per "smentire" la segnalazione sarà calcolato allo stesso modo.

2.4 Meccanismi di Reputation

Il concetto di reputation, per un determinato user, attiene alla credibilità che un determinato soggetto ha all'interno di un social network.

E' una caratteristica dei soggetti che deriva quindi dalle azioni del gruppo in cui sono inseriti, e in particolare dalle azioni di trasmissione di valutazione sociale.

Ogni user ha, associato al proprio username, un valore di reputation che cresce e decresce in base alla validazione o meno della segnalazione effettuata.

Questo valore può variare da 0 a 100 e sono stati previsti cinque diversi cluster per classificare gli user:

- 81 - 100: *Top*
- 61 - 80: *Advanced*
- 41 - 60: *Intermediate*
- 21 - 40: *Amateur*
- 0 - 20: *Beginner*

2.4.1 Reputation increase

Ad ogni segnalazione validata ogni user che ha contribuito alla validazione, compreso quello che ha pubblicato quest'ultima, riceve un incremento al valore di reputation.

Tale incremento presenta un andamento descritto dalla formula:

$$I = (Bel(dato)) \cdot \left(1 - \frac{Rep}{100}\right) \cdot (C)$$

Dove:

- *Bel* ("believe" della segnalazione) è inteso come rapporto tra la somma delle reputation degli utenti che partecipano alla validazione e il numero di questi. Tale valore può variare fra 0 e 1
- *Rep* ("reputation") è legata ad ogni user. Tale valore può variare tra 0 e 100
- *C* è il numero di segnalazioni consecutive validate. Tale valore può variare tra 0 e ∞ .

Il parametro *C* utilizzato nella precedente formula è utile per "premiare" uno user che è particolarmente attento e preciso nel fornire segnalazioni al resto della community, facendo in modo quindi che la sua crescita sia più veloce rispetto ad uno user meno preciso.

2.4.2 Reputation decrease

Ad ogni segnalazione che non viene validata e quindi non è considerata attendibile, lo user che effettua la segnalazione inesatta ed i relativi eventuali "liker" ricevono un decremento al valore di reputation. Tale decremento ha un andamento descritto dalla seguente formula:

$$D = (Bel(dato)) \cdot \frac{Rep}{100} \cdot (C)^2$$

Dove:

- *Bel* ("believe" della segnalazione) è intesa come rapporto tra la somma delle reputation degli utenti che partecipano alla validazione e il numero di questi. Tale valore può variare fra 0 e 1
- *Rep* ("reputation") è legata ad ogni user. Tale valore può variare tra 0 e 100
- *C* è il numero di segnalazioni consecutive non validate. Tale valore può variare tra 0 e ∞ .

Anche in questa formula, prevista per il decremento, il parametro *C* è fondamentale perché, al contrario di quanto detto in precedenza, lo user che non è preciso nella segnalazione viene penalizzato e quindi scoraggiato a fornire informazioni errate.

2.5 L'aggregatore Tweet

La piattaforma "ITS 2.0" applica il paradigma del web 2.0 al settore ITS che ha l'obiettivo di sviluppare soluzioni di trasporto innovative basate sul concetto di "mobilità collaborativa".

L'applicazione realizzata per questo scenario è un aggregatore dei tweet di viabilità, che è stato successivamente integrato con la piattaforma prototipale "ITS 2.0".

L'attività di sviluppo di tale aggregatore ha permesso di dimostrare l'utilizzo di Twitter come piattaforma agile per la pubblicazione di Open Data, in uno scenario end-to-end che vede delle "applicazioni OTT" comunicare automaticamente tramite informazioni strutturate diffuse sul social network.

2.5.1 Gli open data

Gli Open Data sono tipologie di dati liberamente accessibili e riproducibili, non sottoposti a restrizioni commerciali (come copyright o brevetti) o di altro tipo (come particolari vincoli di riservatezza).

I dati aperti sono dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti da chiunque, soggetti eventualmente alla necessità di citarne la fonte e di condividerli con lo stesso tipo di licenza con cui sono stati originariamente rilasciati.

Negli ultimi tempi gli Open Data sono diventati piuttosto "popolari" (come in passato il Cloud Computing, o più recentemente i Big Data) in seguito a una serie di iniziative globali, europee e nazionali volte a rendere progressivamente accessibili le banche dati governative.

Gli Open Data sono infatti associati principalmente al dominio della Pubblica Amministrazione, nell'ottica di rendere il funzionamento della cosa pubblica trasparente e monitorabile da parte dei cittadini, tramite la rendicontazione dei costi e la documentazione dei servizi erogati.

La Open Knowledge Foundation propone una definizione formale degli Open Data, sottolineando il tema primario della libertà di accesso, utilizzo e riproduzione dei dati. Anche la legge italiana definisce gli Open Data all'interno del Codice dell'Amministrazione Digitale, enfatizzando gli aspetti tecnologici legati ai formati di pubblicazione (aperti ovvero non proprietari) e alla facilità di manipolazione automatica (l'interoperabilità dei dati).

Gli Open Data hanno beneficiato dell'evoluzione di Internet. Il web 2.0 e le tecnologie ad esso correlate (web service, feed RSS, linguaggi di definizione di ontologie, etc.) favoriscono i processi di pubblicazione e utilizzo delle risorse digitali.

Solitamente gli Open Data vengono resi accessibili tramite siti web e portali. Soprattutto nell'ambito della Pubblica Amministrazione, la modalità di pubblicazione è di tipo "best-effort": i dati vengono offerti in una molteplicità di formati (PDF, Word, Excel, CSV, etc.), e risultano spesso privi di un apparato descrittivo sufficiente a decifrare le informazioni.

L'attuale strategia privilegia la quantità alla qualità, confidando sostanzialmente in una forma di crowdsourcing, che scarica completamente sugli utenti finali la responsabilità di integrare i dati, immaginare degli use-case di una qualche utilità e implementare delle applicazioni.

2.5.2 Gli account Twitter dedicati all'infomobilità

Twitter è uno dei social network più popolari: attualmente conta 284 milioni di utenti attivi su base mensile in tutto il mondo con 500 milioni di tweet pubblicati ogni giorno.

Gli eventi di viabilità pubblicati in tempo reale su Twitter dagli account esaminati (riferibili a gestori di sistemi di trasporto ed enti e uffici preposti alla sicurezza stradale) sono un esempio di come si possano utilizzare i social network per diffondere in tempo reale informazioni etichettabili come Open Data.

I tweet diventano immediatamente consultabili da parte degli utenti del web, tanto da suggerire l'utilizzo di Twitter come piattaforma agile per la pubblicazione di Open Data.

Si devono però rilevare alcuni limiti:

- I dati sono molto spesso incompleti, poiché chi pubblica le informazioni solitamente non gestisce il limite di 140 caratteri imposti da Twitter, e i messaggi risultano troncati
- la validità degli eventi di viabilità spesso non è chiara agli utenti, poiché nella maggior parte dei casi la conclusione di un evento (coda, incidente, pioggia, etc.) non viene segnalata;
- il formato degli eventi pubblicati è piuttosto grezzo. Potrebbe essere reso più fruibile tramite l'integrazione con un database geografico, l'arricchimento delle informazioni originali, la correlazione con altri eventi;

- l'individuazione, l'aggregazione e l'integrazione delle diverse fonti è a carico degli utenti.

Nella tabella di Figura 2.7 si evidenziano le differenze significative nel formato dei diversi tweet dei principali enti ufficiali di infomobilità.

Fonte	Account automatizzato	Tweet consistenti	Inclusione link	Comunicazione ripristino	Feed RSS	Tweet corredato da immagine	Eventi di viabilità e previsioni di traffico
ANAS	✓	✓					
CCISS	✓		✓	✓	✓	✓	✓
Viabilità Italia	✓		✓	✓	✓	✓	✓
Autovie Venete	✓		✓	✓	✓	✓	✓
Viaggiare in Trentino	✓	✓	✓	✓	✓	✓	✓
Servei Català de Trànsit (PER LA CATALOGNA)	✓	✓	✓	✓	✓	✓	✓
Luceverde		✓	✓	✓	✓	✓	✓
Polizia municipale di Torino		✓	✓	✓	✓	✓	✓
Verona Mobilità		✓					

Fig. 2.7: Differenze nel formato dei tweet

Ogni fonte Twitter di eventi di viabilità adotta un proprio formato di pubblicazione delle informazioni.

Tale formato è strutturato se la pubblicazione è effettuata automaticamente, non strutturato se la pubblicazione è manuale.

In alcuni casi i tweet possono essere inconsistenti, includere link, essere corredati da immagine e possono riguardare non solo eventi di viabilità ma anche fornire previsioni sul traffico. Inoltre solo alcuni enti comunicano il termine di un evento e sono disponibili anche tramite feed RSS.

Il patrimonio informativo pubblicato su Twitter può essere valorizzato da un aggregatore che colleziona le diverse fonti di eventi di viabilità.

In questo scenario il social network svolge il ruolo puramente infrastrutturale di piattaforma social di publish/subscribe che alimenta l'aggregatore dei tweet.

Le informazioni estratte, opportunamente integrate e convalidate, permettono di sviluppare delle applicazioni verticali destinate all'utente finale (Figura 2.8).

Per ricondurre a un unico formato le informazioni pubblicate dalle diverse fonti, l'aggregatore dei tweet di viabilità utilizza un proprio



Fig. 2.8: Applicazioni verticali dell'aggregazione dei tweet

modello dati. I tweet pubblicati in modalità automatica, al di là delle differenze linguistiche e di formattazione, fanno sostanzialmente riferimento al medesimo modello dati.

Nella tabella seguente sono riportati alcuni esempi:

ANAS	#SS7 Traffico in congestione da Piazza dei Re di Roma a Colli Albani. Velocità rilevata:15km/h
CCISS	A50 Tangenziale Ovest Di Milano code per 1,5 km causa incidente tra Svincolo Settimo Milanese-Ss11: S.Siro-Fiera (Km. 6,8) e Svincolo Ss494 Vigevanese Nuova: Lorenteggio (Km. 13,9) in direzione Allacciamento A1 Milano-Napoli
Autovie Venete	#traffico #A4 - Corsia di sorpasso chiusa tra DUINO e MONFALCONE EST dir. VENEZIA causa Lavori

Fig. 2.9: Esempi tweet di infomobilità

- nei tre casi, un evento di viabilità (traffico in congestione, code, corsia di sorpasso chiusa) ha luogo su una particolare strada o autostrada (SS7, A50 Tangenziale Ovest Di Milano, A4);
- l'evento viene circoscritto a un tratto stradale o autostradale, delimitato da due località (che corrispondono ad uscite, svincoli o intersezioni con altre strade e autostrade) ed eventualmente da una direzione;
- le due località che circoscrivono l'evento possono anche essere corredate da un'indicazione chilometrica che misura la distanza delle località dall'inizio della strada, o dalle coordinate geografiche;

- gli eventi possono essere associati a una determinata causa;
- alcuni eventi includono dei parametri aggiuntivi (la lunghezza di una coda o la velocità rilevata in un determinato tratto stradale).

Il modello logico dei tweet di viabilità utilizzato per definire la struttura dati della piattaforma di aggregazione è mostrato nella tabella di Figura 2.10:

Index	SS7	A50	A4
Road	EndPoint1	Tangenziale Ovest di Milano	
	EndPoint2		
LocationA	Description	Piazza dei Re di Roma	Svincolo Settimo Milanese.Ss11: S.siro-Fiera
	Km	6.8	
	Lat		
	Long		
LocationB	Description	Colli Albani	Svincolo Ss494 Vigavenese Nuova: Lorenteggio
	Km	13.9	
	Lat		
	Long		
Event	Description	Traffico in congestione	Code per 1,5 km
	Cause		Incidente
	Direction		AllacciamentoA1 Milano-Napoli
	Speed	15km/h	

Fig. 2.10: Struttura modello dati

Le località che permettono di delimitare i tratti stradali e autostradali su cui occorrono gli eventi di viabilità vengono censite nel database TMC (Traffic Message Channel), che descrive la topologia di una rete stradale e autostradale tramite un insieme di punti ben definiti (che corrispondono alle intersezioni tra le strade della rete).

Il TMC è unico per ogni paese dell'Unione Europea: tutti i centri di informazione e di controllo del traffico e tutti i servizi di infomobilità devono fare riferimento al medesimo database, che in Italia è gestito dal Ministero delle Infrastrutture e dei Trasporti, che provvede periodicamente agli aggiornamenti e a distribuire agli utenti la nuova versione del database.

Viabilità Italia ha predisposto un Glossario degli eventi di viabilità, per definire univocamente le diverse situazioni di traffico e uniformare il linguaggio utilizzato nel rendere pubbliche le informazioni.

Il glossario può essere utilizzato come norma di riferimento per riconciliare le differenze terminologiche tra le varie fonti Twitter.

Twitter è una piattaforma di microblogging concepita per “loggere” tutti i contenuti pubblicati, la cui successione va a formare una sorta di archivio (salvo cancellazioni esplicite da parte degli utenti).

Diversamente dai siti web che mostrano esclusivamente le condizioni di traffico in tempo reale (lo stato operativo della rete stradale), su Twitter gli eventi di viabilità già pubblicati non vengono mai cancellati dall'archivio, e possono riferirsi anche a situazioni ormai concluse.

E' l'utente a dover determinare, sulla base di ulteriori elementi di contesto, se il messaggio è ancora valido o meno.

Ogni fonte presente su Twitter gestisce a proprio modo il ciclo di vita degli eventi pubblicati:

- alcune fonti notificano esplicitamente la conclusione degli eventi. In questo caso l'utente finale può ignorare una segnalazione pubblicata in precedenza;
- altre fonti non notificano la conclusione degli eventi. In questo caso l'utente finale non può sapere con certezza se una segnalazione pubblicata in precedenza è ancora valida.

L'aggregatore dei tweet gestisce esplicitamente il ciclo di vita degli eventi di viabilità:

- il modello dati specificato nella Figura 2.10 è stato esteso con uno stato, una data di inizio e una data di fine, Figura 2.11;
- ogni nuovo evento ha una data di inizio (se non è specificata nel tweet, si può far coincidere con la data di pubblicazione) e assume lo stato in corso
- se un tweet segnala la conclusione di un evento pubblicato in precedenza si assegna all'evento una data di fine e si modifica il suo stato (da in corso a concluso);

- se la conclusione dell'evento non viene notificata, si definisce una durata massima (time-to-live) per ciascuna tipologia di evento (code, incidenti, pioggia, ecc.), sulla base della quale si determina automaticamente la futura data di fine per ogni evento pubblicato, allo scadere della quale l'evento viene considerato concluso.

Status				
StartDate				
EndDate				
Road	Index	EndPoint1	EndPoint2	
LocationA	Description	km	lat	long
LocationB	Description	km	lat	long
Event	Description	Cause	Direction	Speed

Fig. 2.11: Estensione del Modello dati

2.5.3 La parsificazione, data cleansing e data validation

Twitter rende disponibile una API in streaming, che tramite l'apertura di una connessione HTTP persistente permette di accedere ai tweet pubblicati in tempo reale (<https://dev.twitter.com/streaming/overview>).

La streaming API può essere invocata specificando gli account di origine dei tweet (che nel caso in esame corrispondono alle diverse fonti degli eventi di viabilità: ANAS, CCISS, Autovie Venete, etc.).

I tweet restituiti dalla API comprendono sia il testo del messaggio che una serie di metadati (come l'autore e la data di creazione) e sono in formato JSON; essi non sono strutturati, e devono essere parsificati per estrarre le informazioni che descrivono l'evento di viabilità in base al modello dati definito in Figura 2.10.

La parsificazione dei tweet è più semplice quando la pubblicazione avviene in modalità automatica, sulla base di un template caratteristico per ciascuna fonte.

Ogni fonte richiede la propria specifica regola di parsing.

La fase di parsificazione dei tweet include anche il processo di data cleansing, che serve a individuare ed eventualmente a correggere i dati che risultano incompleti, non accurati, o corrotti. Si è visto infatti come spesso gli eventi di mobilità pubblicati su Twitter risultino troncati (e quindi incompleti) a causa del limite di 140 caratteri.

La fase di data cleansing permette anche di uniformare il dizionario adottato dalle diverse fonti per specificare le tipologie di eventi e le relative località, utilizzando come standard di riferimento il glossario degli eventi di viabilità e il database TMC delle località.

Il database TMC può anche essere utilizzato per completare le descrizioni delle località che risultano troncate, e a integrare l'evento con altre informazioni utili (come i riferimenti chilometrici e le geocoordinate associate alle località), nei casi in cui l'informazione parziale è sufficiente a determinare un match con un'entità del database.

Si vedano i seguenti due tweet:



Fig. 2.12: Tweet d'esempio

Nel primo tweet la descrizione della seconda località che identifica il tratto autostradale ("Svincolo di Casoria") è completa, mentre mancano l'indicazione chilometrica (che risulta troncata) e le geocoordinate.

Tali informazioni possono essere recuperate dal database TMC a partire dalla descrizione integrale della località.

Il secondo tweet è analogo, ma in questo caso la descrizione della località è incompleta ("Svincolo Sp113 Cernusc").

La descrizione parziale potrebbe comunque essere sufficiente a individuare univocamente la località, grazie anche alla contestuale indicazione dell'intersezione con l'autostrada "A1 Roma-Napoli".

La successiva fase di data validation convalida la consistenza e l'utilità dei dati.

Se la fase di data cleansing verifica la consistenza sintattica delle informazioni, quella di data validation è orientata a valutarne la consistenza semantica.

Le fonti ufficiali selezionate garantiscono la generale veridicità e validità dei tweet, a condizione che le informazioni pubblicate siano sufficienti a individuare un evento semanticamente consistente. L'eventuale non completezza dei dati ricavati da Twitter potrebbe infatti comprometterne l'intelligibilità.

Nel seguente esempio, l'unico dato mancante è la velocità rilevata sul tratto stradale in congestione, ma il tweet si può considerare sostanzialmente valido:



Fig. 2.13: Tweet ANAS

In altri casi la mancanza di informazioni essenziali pregiudica la validità del tweet, come per l'esempio di Figura 2.14:



Fig. 2.14: Tweet CCISS

L'evento completo, ricavato dal sito web del CCISS, è infatti il seguente:

A9 Lainate-Chiasso code causa veicolo fermo o in avaria tra Svincolo Di Como Monte Olimpino (Km. 39,4) e Svincolo Di Como Sud (Km. 33,8) in direzione Lainate dalle 15:35 del 15 gen 2015

Nel tweet risultano assenti sia il secondo estremo del tratto autostradale, sia la direzione delle code, sia la data di inizio dell'evento (che potrebbe anche essere approssimata con la data di pubblicazione). L'assenza di tali informazioni non consente di ricostruire l'evento con sufficiente precisione, e il tweet non viene considerato valido.

2.5.4 Integrazione con la piattaforma "ITS 2.0"

L'aggregatore dei tweet descritto è stato così integrato con la piattaforma "ITS 2.0", come connettore per gli eventi di viabilità pubblicati su Twitter.

L'aggregatore archivia gli eventi di viabilità reperiti su Twitter e li inoltra alla piattaforma "ITS 2.0" (Figura 2.15).

In particolare:

- si connette a Twitter tramite la streaming API
- riceve in tempo reale i tweet (e gli RSS feed) pubblicati dalle fonti selezionate
- parsifica i tweet (coerentemente con il formato di pubblicazione adottato da ciascuna fonte)
- estrae le informazioni significative, le integra (e.g. ricavando le geocoordinate delle località dal database TMC), le convalida e archivia l'evento in un formato strutturato. I tweet, oltre a segnalare nuove situazioni legate al traffico, possono anche notificare la conclusione di un evento già presente in archivio (in questo caso l'aggregatore modifica lo stato dell'evento, da attivo a concluso)
- inoltra in tempo reale gli eventi di viabilità alla piattaforma "ITS 2.0", che espone una apposita API di data ingestion

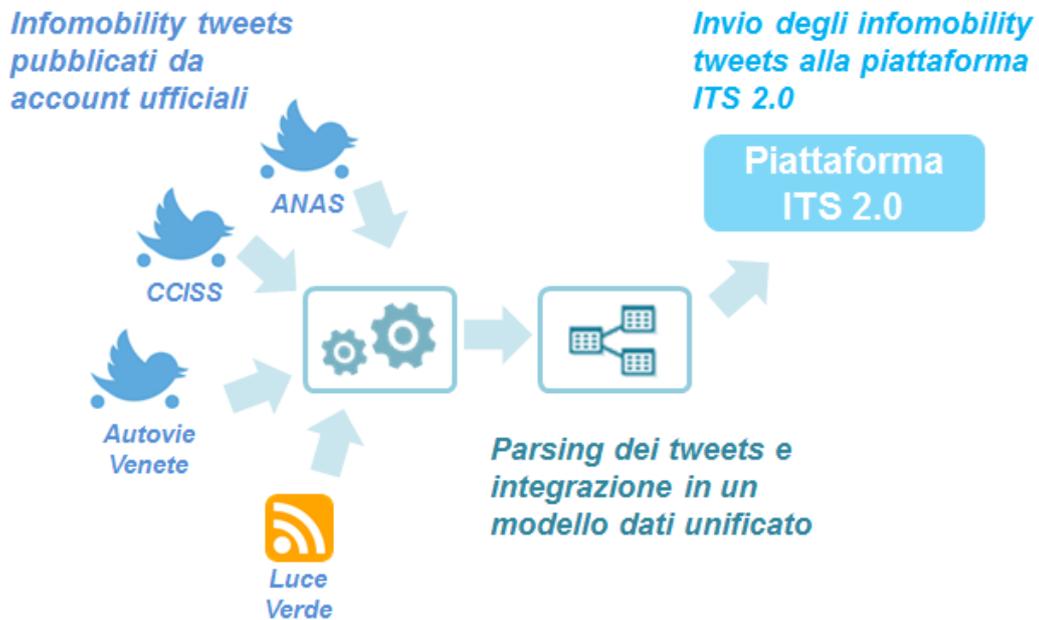


Fig. 2.15: Processo di aggregazione e inserimento in ITS 2.0

I parametri richiesti dalla API sono:

- le credenziali utilizzate per autenticare l'accesso alla API;
- le coordinate spaziali dell'evento, che possono corrispondere a seconda dei casi a un unico punto (individuato da latitudine e longitudine) oppure a una coppia di punti (utilizzati per specificare il tratto stradale in cui l'evento si verifica);
- le coordinate temporali dell'evento, ovvero la data e l'ora di inizio e di fine o se l'evento è stato pianificato in precedenza (ad esempio, la chiusura di un tratto stradale per lavori di manutenzione) la data e l'ora di fine rappresentano un istante futuro certo:
 - se l'evento non è stato pianificato ma è comunque terminato (poiché la fonte Twitter notifica la conclusione degli eventi), la data e l'ora di fine rappresentano un istante passato certo;
 - se l'evento non è stato pianificato e non è possibile conoscere il momento in cui terminerà (poiché la fonte Twitter non notifica

la conclusione degli eventi), la data e l'ora di fine rappresentano un istante futuro probabile in cui si suppone che l'evento sia terminato, determinato sulla base di una durata massima (time-to-live) assegnata a ciascuna tipologia di evento (code, incidenti, pioggia, etc.).

- la categoria dell'evento (1-traffico, 2-meteo, 3-buche, 4-lavori in corso). Le 55 tipologie di eventi elencate dal glossario adottato dall'aggregatore devono quindi essere assegnate alle 4 categorie definite dalla piattaforma "ITS 2.0";
- la descrizione dell'evento (una generica stringa di 250 caratteri), ovvero una traduzione testuale omogenea delle informazioni archiviate dall'aggregatore (che adotta un unico formato di riferimento, diversamente dalle fonti Twitter originali, eterogenee tra loro).

CAPITOLO 3

ANALISI DATI

L'analisi dati è un processo tramite il quale i dati grezzi (*Raw Data*), ovvero i dati direttamente osservati e quindi non soggetti ad elaborazione, mediante meccanismi di *Data Cleaning*, qualora necessari, e successivamente di *Data Mining* vengono trasformati in informazioni che vanno ad accrescere la nostra conoscenza (*Knowledge*).



Fig. 3.1: Processo di analisi Dati

3.1 Anomalie Database

Nel seguito verranno esposte le anomalie riscontrate nel database fornito da Telecom Italia.

Le suddette anomalie possono essere raggruppate in sei macro-categorie:

- incoerenza tra attributi e valori;
- incoerenza tra i valori di uno stesso record;
- incongruenza del formato dei valori di uno stesso attributo in record diversi;
- disomogeneità nelle nomenclature riferite ad uno stesso oggetto;
- presenza di anomalie nella georeferenziazione delle segnalazioni.

Le analisi sono state effettuate attraverso apposite query eseguite sulle tabelle *event* ed *event_point*.

3.1.1 Incoerenza tra attributi e valori

Le incoerenze tra attributi e valori riscontrate riguardano:

- gli attributi *street* e *city* e i corrispondenti valori (Figura 3.2);
- l'attributo *direction* e i corrispondenti valori (Figura 3.3);
- l'attributo *road_code* e i corrispondenti valori. In Figura 3.4 si riportano alcuni valori dell'attributo in questione (*#CassiasBis*, *#Bergamo*, *#ViaPontina*, *#ViaAnagnina*, etc.).

Tali anomalie si riscontrano in molti record riguardanti tutte le fonti Twitter.

Row	id	city	no...	street	source
14	59		SS11	Padana Superiore	www.milano.lucoverde.it
15	70		AS2	Tangenziale Nord Di Mi...	www.milano.lucoverde.it
16	71	Via Maffucci		Milano Via Adolina Patti	www.milano.lucoverde.it
17	72	Via degli Imbriani		Milano Via Adolina Patti	www.milano.lucoverde.it
18	73	Via dei Tronzi		Milano Via Termopoli	www.milano.lucoverde.it
19	74	Via Roggia Scagna		Milano Via Termopoli	www.milano.lucoverde.it
20	75	Via Cittadini		Milano Via Arisa	www.milano.lucoverde.it
21	76	Via Orsi		Milano Via Arisa	www.milano.lucoverde.it
22	77	Via Saponaro		Milano Via Costantino B...	www.milano.lucoverde.it
23	78	Via Feraboli		Milano Via Costantino B...	www.milano.lucoverde.it
24	79	Via Valera		Milano Via Benaco	www.milano.lucoverde.it
25	80	Piazza Geremia Boninelli		Milano Via Benaco	www.milano.lucoverde.it
26	91	Svincolo Strampelli	SS...	Portina	regionelazio.lucoverde.it
27	92	Usca Pomenza Centro/Castelli Romani	SS...	Portina	regionelazio.lucoverde.it
28	SS9	Via Enfila	www.milano.lucoverde.it
29	...	Svincolo Strampelli	SS...	Portina	regionelazio.lucoverde.it
30	...	Usca Pomenza Centro/Castelli Romani	SS...	Portina	regionelazio.lucoverde.it

Row	id	city	road_code	street	source
1	1	Svincolo Ardeatina e Usca Pomenza Centro/Castelli Romani	SS148	Portina	regionelazio.lucoverde.it
2	2	Portonaccio - Galleria Placida e Tangenziale Est. in direzione Tangenziale Est.	AS4	Tratto Libiano	regionelazio.lucoverde.it
3	3	Portonaccio	A1	Roma-Napoli	regionelazio.lucoverde.it
4	4	Ceprano	A1	Roma-Napoli	regionelazio.lucoverde.it
5	5	Orte	A1	Firenze-Roma	regionelazio.lucoverde.it
6	6	Magliana Sabina	A1	Firenze-Roma	regionelazio.lucoverde.it
7	7	10	SP2c	Via Laurentina	roma.lucoverde.it
8	11	Barrera Di Napoli Sud-Biara	A3	Napoli - Salerno	www.cciss.it
9	12	Grande Raccordo Anulare Di Roma	SS115	Sud Occidentale Scula	www.cciss.it
10	13	Proto Ennodio/Innesto S51 192ter Sud Occ. Scula	SS115	Sud Occidentale Scula	www.cciss.it
11	24	Svincolo Di Busto Arsizio	AS	Milano - Varese	www.cciss.it

Row	id	city	road_code	street	source
23	47	Bivio Per Foligno/Innesto S63 Flaminia e Colfiorito	SS77	Della Val Di Chienti	www.cciss.it
24	48	Ortona	A14	Pescara - Bari	www.cciss.it
25	49	Lanciano	A14	Pescara - Bari	www.cciss.it
26	50	Martova Sud	A22	Autostrada Del Brennero	www.cciss.it
27	51	Pegognaga	A22	Autostrada Del Brennero	www.cciss.it
28	60	Cergnola Est	A14	Pescara - Bari	www.cciss.it
29	61	Foggia	A14	Pescara - Bari	www.cciss.it
30	62	Alloccamento Dikamazione Per Ravenna	A14	Bologna - Ancona	www.cciss.it
31	63	Faenza	A14	Bologna - Ancona	www.cciss.it
32	66	Martova Sud	A22	Autostrada Del Brennero	www.milano.lucoverde.it
33	67	Pegognaga	A22	Autostrada Del Brennero	www.milano.lucoverde.it
34	70	...	AS2	Tangenziale Nord Di M...	www.milano.lucoverde.it
35	71	Via Maffucci		Milano Via Adolina Patti	www.milano.lucoverde.it
36	72	Via degli Imbriani		Milano Via Adolina Patti	www.milano.lucoverde.it

Row	id	city	road_code	street	source
142	...	Poggio Imperiale	A14	Pescara - Bari	www.cciss.it
143	...	Termoli-Molise	A14	Pescara - Bari	www.cciss.it
144	...	Carnia	A23	Palmanova - Tarvisio	www.cciss.it
145	...	Portofino	A23	Palmanova - Tarvisio	www.cciss.it
146	...	Martova Sud	A22	Autostrada Del Brennero	www.milano.lucoverde.it
147	...	Pegognaga	A22	Autostrada Del Brennero	www.milano.lucoverde.it
148	...	Svincolo 27: Via C. Colombo e Svincolo 24: Via Ardeatina in direzio...	A40-GRA	Grande Raccordo Anulare Di R...	regionelazio.lucoverde.it
149	...	Martova Sud	A22	Autostrada Del Brennero	www.cciss.it
150	...	Pegognaga	A22	Autostrada Del Brennero	www.cciss.it
151	...	Pescara N. Citta' S. Angelo	A14	Pescara - Bari	www.cciss.it
152	...	Grande Raccordo Anulare Di Roma - Raccordo Salerno-Avellino	A3	Pescara - Bari	www.cciss.it
153	...	Svincolo Mercato S. Severino e Salerno - Raccordo Salerno-Avellino	A30	Caserta - Salerno	www.cciss.it
154	...	Sarzanza/Innesto S61 Aurelia e S. Stefano Di Magra	SS62	Della Cisa	www.cciss.it
155	...	Palazzo	A4	Milano - Brescia	www.cciss.it

Row	id	city	road_code	street	source
109	...	Pegognaga	A22	Autostrada Del Brennero	www.cciss.it
110	...	Caserta Nord	A1	Roma - Napoli	www.cciss.it
111	...	Capua	A1	Roma - Napoli	www.cciss.it
112	...	Brodano-Portici	A3	Napoli - Salerno	www.cciss.it
113	...	Torre Del Greco	A3	Napoli - Salerno	www.cciss.it
114	...	Svincolo Di Brennero	A22	Autostrada Del Brennero	www.cciss.it
115	...	Chiusa-Valgröden	A22	Autostrada Del Brennero	www.cciss.it
116	...	Casale Svinco: A22 Torino-Bardonecchia	SS395	Di Bardonecchia	www.cciss.it
117	...	Genova Est	A12	Genova - Rosignano	www.cciss.it
118	...	Alloccamento A7 Milano-Genova	A12	Genova - Rosignano	www.cciss.it
119	...	Capua	A1	Roma - Napoli	www.cciss.it
120	...	Capua	A1	Roma - Napoli	www.cciss.it
121	...	Bergamo	A4	Milano - Brescia	www.cciss.it
122	...	Grande Raccordo Anulare S. Severino e Salerno - Raccordo Salerno-Avellino	A3	Caserta - Salerno	www.cciss.it
123	...	Martova Sud	A22	Autostrada Del Brennero	www.milano.lucoverde.it

Fig. 3.2: Incongruenze tra gli attributi city e street e i corrispondenti valori

	varying(100)	street character varying(100)	km double precision	description character varying(500)	order_in_event integer	event_id bigint	direction character varying(500)
293		Milano - Brescia	203.4		1	167	
294	Veneto Nord	Mestre - Belluno	58.9	lavori	0	168	Venezia
295	Veneto Sud	Mestre - Belluno	52.3		1	168	
296	periale	Pescara - Bari	507	lavori	0	169	Pescara
297	molise	Pescara - Bari	477		1	169	
298	Mercato S. Severino e Salerno	Caserta - Salerno	55.3		0	170	Salerno dalle 13:15 del 08 mag 2015
299		Palmanova - Tarvisio	59.6	lavori	0	171	
300		Palmanova - Tarvisio	92.4		1	171	
301	ud	Autostrada Del Brennero		incidente	0	172	Modena
302	a	Autostrada Del Brennero			1	172	
303		Roma-Napoli	669.6		0	173	Roma si tratta di vettura finita fuori
304	o	Roma-Napoli	658.3		1	173	
305	27: Via C. Colombo e Svincolo	Grande Raccordo Anulare Di Roma			0	174	
306	ud	Autostrada Del Brennero	265	incidente	0	175	Modena
307	a	Autostrada Del Brennero	276.7		1	175	
308	l. Citta' S. Angelo	Pescara - Bari	364.2	lavori	0	176	Bari
309	mento A25 Torano-Pescara	Pescara - Bari	378.4		1	176	
310	erlengo	Milano - Bologna	37.9	perdita accidentale	0	177	Milano
311		Milano - Bologna	22.7		1	177	
312	Mercato S. Severino e Salerno	Caserta - Salerno	55.3		0	178	Salerno dalle 13:15 del 08 mag 2015
313	Innesto S61 Aurelia e S. Stefano	Della Cisa	6	causa lavori	0	179	
314		Milano - Brescia	195.2		0	180	Brescia dalle 13:42 del 08 mag 2015
315		Milano - Brescia	203.4		1	180	
316	to	Ancona - Pescara	351.8	lavori	0	181	Ancona
317	egli Abruzzi	Ancona - Pescara	344		1	181	
318		Autostrada Del Brennero	302.2	traffico intenso	0	182	Modena
319	mento Al Milano-Napoli	Autostrada Del Brennero	313.1		1	182	
320	l. Cantone	Milano - Genova	100.7	lavori	0	183	Genova
321	lvia	Milano - Genova	106.5		1	183	
322	24: Via Ardeatina	Grande Raccordo Anulare Di Roma	48		0	185	carreggiata interna incidente avvenuto
323	vest	Brescia - Padova	329	causa lavori	0	186	Padova
324	est	Brescia - Padova	336.6		1	186	

Fig. 3.3: Incongruenze tra l'attributo direction e i corrispondenti valori

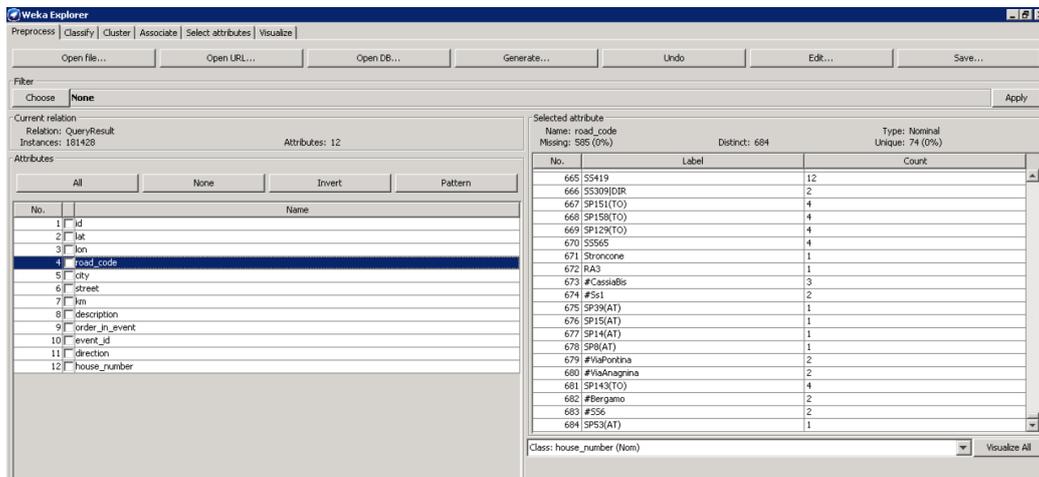


Fig. 3.4: Incongruenza tra l'attributo *road_code* e i corrispondenti valori

3.1.2 Incoerenza tra i valori di uno stesso record

Molti record presentano incongruenze logiche tra valori di diversi attributi. In Figura 3.5 si riporta un esempio.

In particolare, si mostra che l'evento con $id = 90117$ appartiene alla categoria dei parcheggi ($category_id = 4$), si riferisce ad un incidente ($description = causaincidente$) ed è localizzato da due coppie di coordinate diverse ($lat = -3.2296647$ e $lon = -52.2623839$; $lat = 44.42609$ e $lon = 12.15309$).

Si tratta di esempio di anomalia in quanto:

- la categoria parcheggio prevede una localizzazione univoca;
- il primo punto ($lat = -3.2296647$ e $lon = -52.2623839$) fa riferimento ad una località in non italiana (Brasile);
- il valore dell'attributo *city* non è quello di una città.

Row	description	direction	id	lat	lon	category_id	road_code	street	city	source	ts_create	order_in_event
1			90117	-3.2296647	-52.2623839	4	A14	Diramazione Per Ravenna	Svincolo...	CCISS_Ministero	2015-06-19 00:00:00.0	1
2	causa incidente		90117	44.42609	12.15309	4	A14	Diramazione Per Ravenna	Svincolo S.S.16 Adriatica	CCISS_Ministero	2015-06-19 00:00:00.0	0

Fig. 3.5: Incoerenze tra i valori di uno stesso record

3.1.3 Incongruenza del formato dei valori di uno stesso attributo in record diversi

Si utilizzano diversi formati per i valori di uno stesso attributo in record diversi. In Figura 3.6 si riporta un esempio in cui il valore di *road_code* è espresso in due differenti modi (#A4 e A4) pur riferendosi ad uno stesso oggetto (A4).

Row	id	lat	lon	city	road_code	street	source
1	13	45.56535	10.07003	Ospitaletto	A4	Milano - Brescia	www.cciss.it
2	12	45.57686	10.00818	Rovato	A4	Milano - Brescia	www.cciss.it
3	14	45.56699	9.34192	Allacciamento Tangenziale Est Di Milano	A4	Milano - Brescia	www.cciss.it
4	15	45.56984	9.36119	Agrate	A4	Milano - Brescia	www.cciss.it

Row	id	lat	lon	city	road_code	street	source	category_id
1	125370	45.5614828	9.332778000000001	Barriera di Milano Est	#A4	tratto urbano # traffico	LuceverdeMilano	1
2	125637	45.498527	9.1961173	Sesto San Giovanni-Viale Zara	#A4	Tratto Urbano	LuceverdeMilano	1
3	125638	45.4958889	9.139014999999999	Viale Certosa	#A4	Tratto Urbano	LuceverdeMilano	1
4	125656	45.4138233	10.8818161	#Sommacampagna	#A4	#TorinoVenezia	LuceverdeMilano	1
5	125657	31.7346128	-106.4022221	#Verona Sud	#A4	#TorinoVenezia	LuceverdeMilano	1
6	125740	31.7346128	-106.4022221	Verona Sud	#A4	#TorinoVenezia	LuceverdeMilano	3

Fig. 3.6: Formato differente per l'oggetto A4

3.1.4 Disomogeneità nelle nomenclature per uno stesso oggetto

Sono stati riscontrati casi in cui uno stesso oggetto viene indicato con nomenclature diverse. Ciò inficia sulla qualità dell'analisi dei dati in quanto risulterebbe complessa l'individuazione dell'uguaglianza semantica.

In Figura 3.7 si riporta un esempio in cui si mostra che le fonti Twitter *Luceverde Roma* e *Luceverde Regione Lazio* a volte vengono indicate con *roma.luceverde.it* e *regionelazio.luceverde.it* ed altre volte con *LuceverdeRoma* e *LuceverdeLazio*

Row	description	lat	lon	street	city	source
1	Km	44.6865136	7.827983400000001	- #T...	#TangenzialeOvest	LuceverdeMilano
2	code	45.67453140000001	9.7137887	- Dell...	Grassobbio Centro/Zona Industriale/Seriato	LuceverdeMilano
3		45.6768089	9.685160699999999	- Dell...	Orio al Serio	LuceverdeMilano
4		45.703688850727	8.7224840502167	- Dir...	Sesto Calende-Vergiate	LuceverdeMilano
5	#MilanoLaghi	44.017046	10.9245059	- Dir...	#A26 dei Trafori	LuceverdeMilano
6	code	45.9711136	9.193689299999999	- Re...	Lenno	LuceverdeMilano
7		45.9408234	9.128215899999999	- Re...	Argegno	LuceverdeMilano
8		45.4958889	9.139014999999999	- #Mil...	V.Le Certosa	LuceverdeMilano
9		45.519606499999999	9.1129301	- #Mil...	Fiera Milano	LuceverdeMilano
10	Olimpico.	41.9377375	12.5082287		Via #Salaria	LuceverdeLazio
11		41.8908925	12.5212715		#Castrense	LuceverdeRoma
12	Olimpico.	41.9377375	12.5082287		Via #Salaria	LuceverdeRoma
13	Giovanni.	41.8914125	12.5210689		#Prenestina	LuceverdeRoma
14		41.8908925	12.5212715		#Castrense	LuceverdeLazio
15	Giovanni.	41.8914125	12.5210689		#Prenestina	LuceverdeLazio

Row	description	lat	lon	category_id	street	city	source	ts_create	order_in_event	expire_when
26		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-06-30 00:00:00.0	0	2015-06-30 01:00:00
27		4...	1...	4	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-01 00:00:00.0	0	2015-07-08 00:00:00
28		4...	1...	4	Via del Foro Italico	regionelazio.luceverde.it	regionelazio.luceverde.it	2015-07-01 00:00:00.0	0	2015-07-08 00:00:00
29		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-01 00:00:00.0	0	2015-07-01 01:00:00
30		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-01 00:00:00.0	0	2015-07-01 01:00:00
31		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-01 00:00:00.0	0	2015-07-01 01:00:00
32		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-02 00:00:00.0	0	2015-07-02 01:00:00
33		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-02 00:00:00.0	0	2015-07-02 01:00:00
34		4...	1...	1	Via del Foro Italico	roma.luceverde.it	roma.luceverde.it	2015-07-02 00:00:00.0	0	2015-07-02 01:00:00

Fig. 3.7: Nomenclature diverse per Luceverde Roma e Luceverde Regione Lazio

3.1.5 Anomalie nella georeferenziazione delle segnalazioni

È stata riscontrata la presenza di punti anomali di localizzazione delle segnalazioni ovvero di valori delle coppie (*lat,lon*) che non si collocano all'interno del territorio italiano (Figura 3.8)

Si tratta di anomalie in quanto:

- i tweet si riferiscono a dati italiani;
- le restanti informazioni testuali dello stesso record fanno riferimento a località italiane.

Mediante una query che rileva le coppie (*lat,lon*) esterne al territorio italiano (individuato da un'area rettangolare avente come lati i valori estremi di latitudine e longitudine dello stesso) è stato possibile stabilire la numerosità di tali anomalie.

Si tratta di 284 osservazioni (sui 181.428 record totali della tabella *event_point*).

Seppur all'interno dell'area suddetta, in Figura 3.9 si evince la presenza di un'ulteriore localizzazione anomala (indicata dalla freccia) che fa

riferimento ad una segnalazione collocata indicativamente nel Mare Adriatico ma che presenta, tuttavia, il valore Bologna-Firenze sotto la voce *street*.

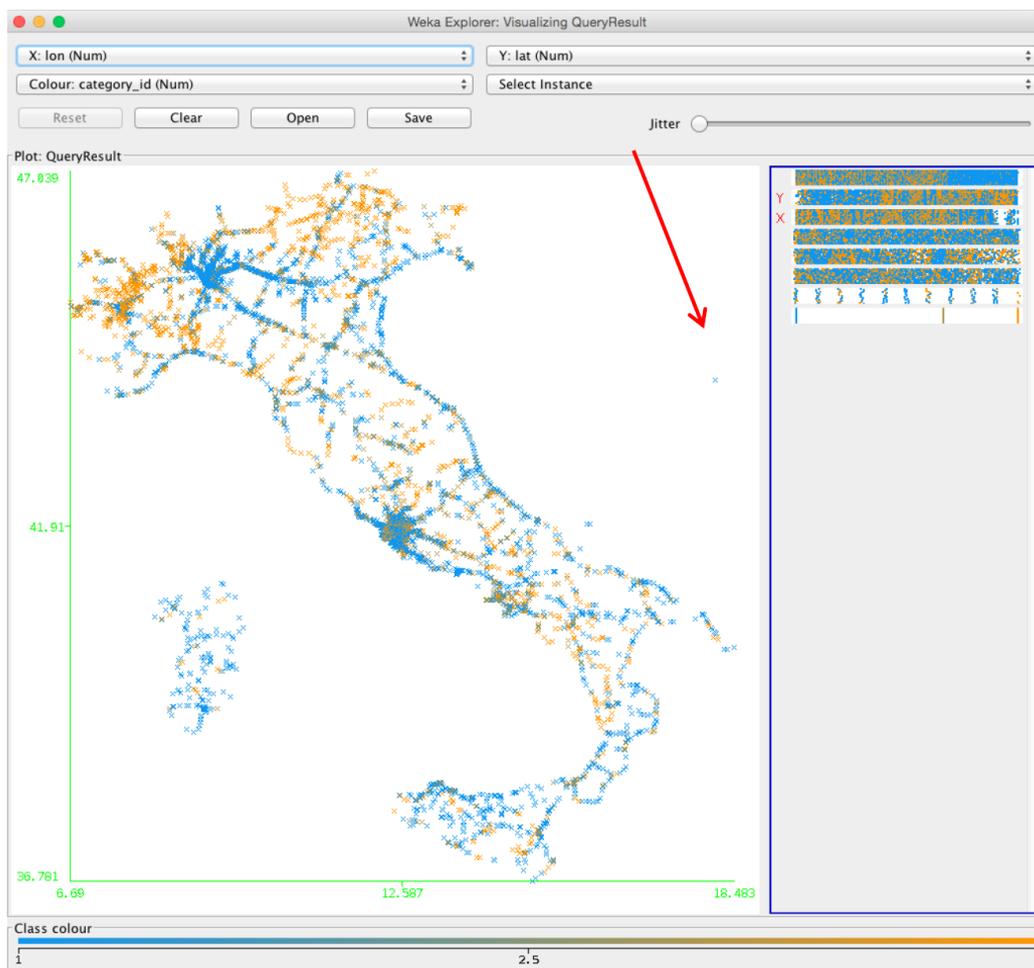


Fig. 3.8: Rappresentazione sul piano cartesiano dei valori di latitudine e longitudine delle segnalazioni interne al perimetro italiano

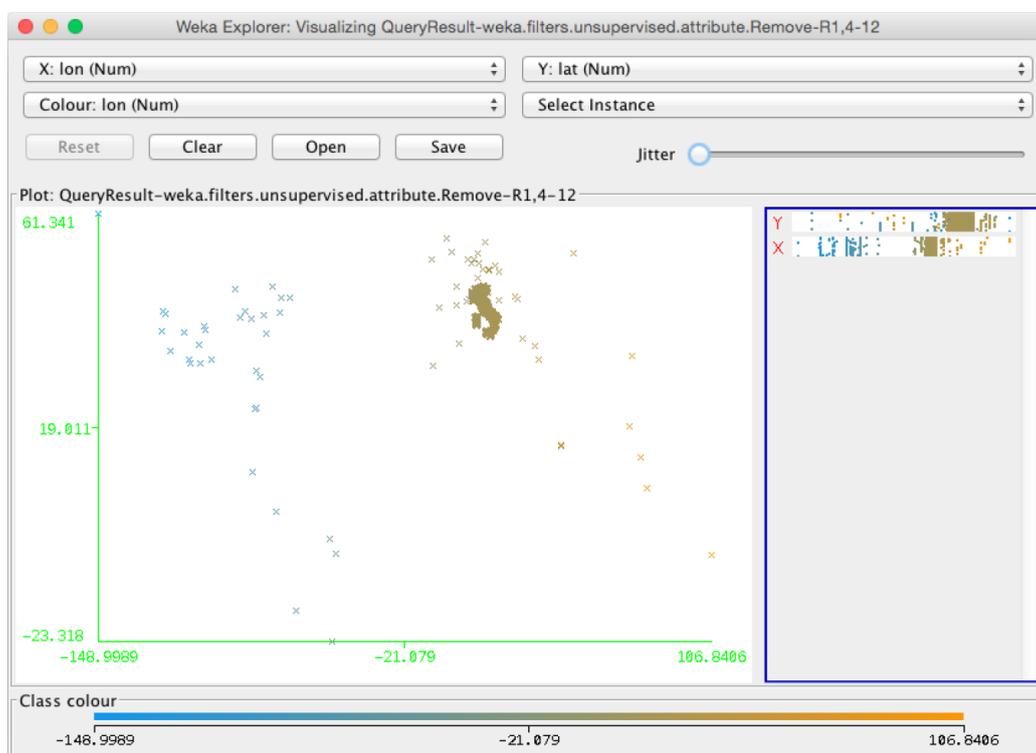


Fig. 3.9: Rappresentazione sul piano cartesiano dei valori di latitudine e longitudine delle segnalazioni

3.2 Data Cleansing

Il *Data Cleansing*, o *Data Cleaning*, si occupa di individuare e rimuovere gli errori e le incongruenze dei dati al fine di migliorare la qualità degli stessi. I problemi di qualità dei dati sono presenti in singole raccolte di dati, come file e database, ad esempio, a causa di errori di ortografia durante l'immissione di dati, informazioni mancanti o dati non validi. Quando più fonti di dati devono essere integrate, per esempio, nei Data Warehouse, nei sistemi di database federati o sistemi informativi web-based a livello mondiale, si rende necessario un processo di Data Cleansing significativo.

Questo perché le fonti spesso contengono dei dati ridondanti in diverse rappresentazioni.

Al fine di fornire l'accesso a dati precisi e coerenti diventa necessario il consolidamento di differenti rappresentazioni dei dati e l'eliminazione di informazioni duplicate.

3.2.1 L'approccio al Data Cleansing

In generale, il Data Cleansing implica diverse fasi:

- **Analisi dei dati:** Al fine di individuare quali tipi di errori e incoerenze sono da rimuovere, è richiesta una dettagliata analisi dei dati. In aggiunta ad una ispezione manuale dei dati, programmi di analisi dovrebbero essere utilizzati per acquisire i metadati relativi alle proprietà dei dati e individuare i problemi della qualità dei dati.
- **Definizione della trasformazione del flusso di lavoro e della mappatura di regole:** a seconda del numero delle fonti di dati, del loro grado di eterogeneità e della dirtyness dei dati, possono essere eseguiti un gran numero di trasformazioni dei dati e di fasi di pulizia. A volte, una traduzione dello schema viene utilizzato per mappare le fonti ad un modello comune di dati (e.g. rappresentazione relazionale). I primi step del Data Cleansing sono in grado di correggere i problemi di istanze di sorgenti singole e preparare i dati

per l'integrazione. I successivi step hanno a che fare con l'integrazione di schemi/dati e con i problemi di pulizia di istanze di sorgenti multiple (e.g. i duplicati). Per consentire la generazione automatica del codice di trasformazione, la procedura di Data Cleansing deve essere specificata da una query dichiarativa e da un linguaggio di mapping.

- **Verifica:** la correttezza e l'efficacia della trasformazione di un flusso di lavoro dovrebbe essere testata e valutata (e.g. su un campione o su una copia dei dati di origine) per migliorarne la definizione. Più iterazioni delle fasi di analisi, di progettazione e di verifica possono essere necessarie, ad esempio, dal momento che alcuni errori emergono solo dopo l'applicazione di alcune trasformazioni.
- **Trasformazione:** esecuzione degli step della trasformazione eseguendo il flusso di lavoro o durante la risposta delle query sui dati di sorgenti multiple.
- **Riflusso di dati puliti:** Dopo che gli errori vengono rimossi, i dati puliti dovrebbero sostituire i dati "sporchi" delle fonti originali, al fine di dare alle applicazioni legacy dati migliori cercando di evitare di rifare il lavoro di pulizia per future estrazioni di dati.

Il processo di trasformazione richiede ovviamente una grande quantità di metadati, quali schemi, caratteristiche dei dati a livello di istanza, mappature di trasformazione, definizioni del flusso di lavoro, etc. etc. Per coerenza, flessibilità e facilità di riuso, questi metadati deve essere mantenuti in un repository DBMS-based. Per supportare la qualità dei dati, devono essere registrate informazioni dettagliate sul processo di trasformazione, sia nel repository sia nelle istanze trasformate, in particolare quelle relative alla completezza e all'aggiornamento dei dati di origine e quelle relative alle informazioni di derivazione circa l'origine degli oggetti trasformati e le modifiche a loro applicate.

3.2.2 L'algoritmo implementato

L'algoritmo di Data Cleansing implementato rende omogenei e corretti i dati di infomobilità relativi ad eventi avvenuti sulla rete stradale urbana.

A tal fine, sono state utilizzate le *Google Geocoding API*, in particolare le funzionalità *Geocoding* e *Reverse Geocoding*:

- Il **Geocoding** consente di convertire l'indirizzo di una via, una piazza o del centro città, nelle corrispondenti coordinate geografiche di latitudine e longitudine;
- Il **Reverse Geocoding** consente di ricavare l'indirizzo testuale dalle coordinate geografiche di latitudine e longitudine. Fornendo in input le coordinate geografiche di latitudine e longitudine restituisce in risposta i seguenti valori:
 - *street_number*
 - *route* (Strada)
 - *locality* (Città)
 - *administrative_area_level_3* (Comune)
 - *administrative_area_level_2* (Provincia)
 - *administrative_area_level_1* (Regione)
 - *country* (Stato)
 - *postal_code* (CAP)

Come già precisato, l'algoritmo applica eventuali correzioni a record riferenti ad eventi che non avvengono su strade extraurbane. L'identificazione di questi ultimi avviene verificando la presenza del luogo di accadimento dell'evento nel database TMC e, in caso di esito negativo, verificando che il valore di *street_number* restituito dalla Reverse Geocoding Google API in seguito all'inserimento delle coordinate di latitudine e longitudine è pari a "*null*". Tutti i record processati vengono immagazzinati in una nuova tabella.

Di seguito si riportano i passi che costituiscono l'algoritmo e che si ripetono per ogni record della tabella *event_point*:

1. Si verifica la presenza dei valori di *road_code* e *city* del record nel database TMC. Se sono presenti allora il record viene copiato nella nuova tabella.

Altrimenti:

- (a) Prende in ingresso i valori di *lat* e *lon* del record ed interroga la Reverse Geocoding Google API;
- (b) Se il valore di "*street_number*" è pari a "*null*" allora il record viene copiato nella nuova tabella.

Altrimenti:

- i. Si verifica se i valori di *locality* (o alternativamente *administrative_area_level_3* nel caso in cui *locality* è pari a "*null*") e *route* sono presenti nei campi di *street*, *city* o *road_code*
- ii. Se sono presenti, al fine di ottenere omogeneità nei dati, i valori di *route* e *locality* (*administrative_area_level_3*) si sovrascrivono, rispettivamente, nei campi *street* e *city*. Il record corretto si copia nella nuova tabella.

Altrimenti:

- Si forniscono alla Geocoding Google API i valori di *street*, *city* e *road_code*. I valori delle coordinate date in risposta si sovrascrivono sui campi *lat* e *lon*, al fine di ottenere dati corretti. L'omogeneità, invece, è garantita sovrascrivendo, come in precedenza, *route* e *locality* (*administrative_area_level_3*) nei campi *street* e *city*. Il record corretto viene copiato nella nuova tabella.

Nel seguito si riporta un esempio di correzione di un record.
Si consideri il seguente record Figura 3.10

id	lat	lon	road_code	city	street
484383	41,90278	12,49637	#Roma	via di Malafede	#incidente con possibili

Fig. 3.10: Record da correggere

1. La coppia (#Roma, via di Malafede) non è presente in TMC.
 - (a) I valori (41.9027835, 12.4963655) vengono forniti al Reverse Geocoding Google API
 - (b) In risposta si ottiene la risposta mostrata in Figura 3.11
 - (c) Il valore di *street_number* non è pari a "null".
 - i. I valori di *route* e *locality*, rispettivamente pari a "Via Giuseppe Romita" e "Roma" non sono presenti entrambi in *street*, *city* o *road_code*. Si passa al *ii* passo dell'algoritmo;
 - ii. Si fornisce alla Geocoding della Google API i valori (#Roma; via di Malafede; #incidente con possibili). In risposta si ottiene quanto mostrato in Figura 3.12.
 - iii. I valori (*lat*, *lon*, *administrative_area_level_3*, *route*) rispettivamente pari a (41.7687147, 12.4030473, Roma, Via di Malafede) vengono sovrascritte nei campi *lat*, *lon*, *city*, *street* del record.

```
{
  "results" : [
    {
      "address_components" : [
        {
          "long_name" : "2",
          "short_name" : "2",
          "types" : [ "street_number" ]
        },
        {
          "long_name" : "Via Giuseppe Romita",
          "short_name" : "Via Giuseppe Romita",
          "types" : [ "route" ]
        },
        {
          "long_name" : "Roma",
          "short_name" : "Roma",
          "types" : [ "locality", "political" ]
        },
        {
          "long_name" : "Roma",
          "short_name" : "Roma",
          "types" : [ "administrative_area_level_3", "political" ]
        },
        {
          "long_name" : "Città Metropolitana di Roma",
          "short_name" : "RM",
          "types" : [ "administrative_area_level_2", "political" ]
        },
        {
          "long_name" : "Lazio",
          "short_name" : "Lazio",
          "types" : [ "administrative_area_level_1", "political" ]
        },
        {
          "long_name" : "Italia",
          "short_name" : "IT",
          "types" : [ "country", "political" ]
        },
        {
          "long_name" : "00185",
          "short_name" : "00185",
          "types" : [ "postal_code" ]
        }
      ]
    }
  ],
}
```

Fig. 3.11: Risposta Reverse Geocoding Google API

```

{
  "results" : [
    {
      "address_components" : [
        {
          "long_name" : "Via di Malafede",
          "short_name" : "Via di Malafede",
          "types" : [ "route" ]
        },
        {
          "long_name" : "Municipio Roma X",
          "short_name" : "Municipio Roma X",
          "types" : [ "sublocality_level_1", "sublocality", "political" ]
        },
        {
          "long_name" : "Roma",
          "short_name" : "Roma",
          "types" : [ "administrative_area_level_3", "political" ]
        },
        {
          "long_name" : "Città Metropolitana di Roma",
          "short_name" : "RM",
          "types" : [ "administrative_area_level_2", "political" ]
        },
        {
          "long_name" : "Lazio",
          "short_name" : "Lazio",
          "types" : [ "administrative_area_level_1", "political" ]
        },
        {
          "long_name" : "Italia",
          "short_name" : "IT",
          "types" : [ "country", "political" ]
        },
        {
          "long_name" : "00125",
          "short_name" : "00125",
          "types" : [ "postal_code" ]
        }
      ],
      "formatted_address" : "Via di Malafede, 00125 Roma RM, Italia",
      "geometry" : {
        "bounds" : {
          "northeast" : {
            "lat" : 41.7929804,
            "lng" : 12.4196144
          },
          "southwest" : {
            "lat" : 41.7447021,
            "lng" : 12.3892431
          }
        },
        "location" : {
          "lat" : 41.7687147,
          "lng" : 12.4030473
        },
        "location_type" : "GEOMETRIC_CENTER",
        "viewport" : {
          "northeast" : {
            "lat" : 41.7929804,
            "lng" : 12.4196144
          },
          "southwest" : {
            "lat" : 41.7447021,
            "lng" : 12.3892431
          }
        }
      },
      "partial_match" : true,
      "place_id" : "ChIJhzYZuhrzJRMruw0JVOD22YU",
      "types" : [ "route" ]
    }
  ],
  "status" : "OK"
}

```

Fig. 3.12: Risposta Geocoding Google API

Il record corretto, che verrà copiato nella nuova tabella, è il seguente:

id	lat	lon	road_code	city	street
484383	41,76871	14,40305	Roma	Roma	via di Malafede

Fig. 3.13: Record corretto

3.3 Data Mining

Si definisce *Data Mining* il processo che impiega una o più tecniche di apprendimento computerizzate per analizzare automaticamente ed estrarre le conoscenze dai dati contenuti in un database. Scopo di una sessione di Data Mining è identificare trend e modelli nei dati. La conoscenza estratta da una sessione di Data Mining è data da un modello o da una generalizzazione dei dati.

Esistono numerose tecniche di Data Mining; tuttavia tutti i metodi di Data Mining applicano l'apprendimento basato sull'induzione.

L'apprendimento basato sull'induzione è il processo che genera definizioni di concetti generali dall'osservazione di esempi specifici che necessitano di essere appresi.

In generale, il Data Mining può essere definito come un processo costituito da quattro passi.

Per effettuare una sessione di Data Mining è necessario:

1. Assemblare l'insieme dei dati da analizzare;
2. Presentare i dati a un programma di Data Mining;
3. Interpretare i risultati. L'interpretazione dei dati richiede l'esame dell'output degli strumenti di Data Mining per determinare se ciò che abbiamo scoperto sia utile e interessante. Se i risultati non sono soddisfacenti, è possibile ripetere i passi del processo di Data Mining utilizzando nuovi attributi e/o casi. In alternativa, si può decidere al warehouse e ripetere il processo di estrazione.

4. Applicare i risultati a nuovi problemi. Si riporta una lista di esempi che mostra come il Data Mining possa essere applicato con successo per risolvere problemi reali:
 - (a) **Scoperta di frodi.** Il Data Mining viene utilizzato per individuare chiamate fraudolente, segnalare possibili transazioni con carte di credito false, individuare attività di riciclaggio di denaro, frodi nelle reti di telefonia mobile.
 - (b) **Cura della salute.** Il Data Mining viene utilizzato per aiutare a scoprire cure farmaceutiche meno costose ma ugualmente efficaci, per certe categorie di pazienti.
 - (c) **Economia e finanza.** Il Data Mining viene utilizzato per determinare i rischi assicurativi, gestire i portafogli finanziari e identificare le persone e le aziende che hanno un rischio di credito alto o basso, identificare i clienti che utilizzano determinati prodotti e quali tipi di prodotti sono preferiti al fine di ottenere il giusto mix di prodotti e servizi che soddisfano le esigenze dei clienti.
 - (d) Applicazioni scientifiche. L'analisi statistica clusterizzata è stata utilizzata per scoprire una terza classe di esplosioni di raggi gamma.
 - (e) **Sport e giochi.** Alcuni allenatori utilizzano Advanced Scout (uno strumento di Data Mining) per creare le migliori marcature e identificare gli schemi di gioco più adatti.

Nonostante il campo di applicazione delle tecniche di Data Mining sia in continuo cambiamento, alcune strategie basilari sono rimaste pressoché invariate. Mentre una strategia di Data Mining definisce le linee guida dell'approccio alla risoluzione dei problemi, una tecnica di Data Mining applica una strategia ad un insieme di dati.

3.3.1 Tecniche di Data Mining

Le tecniche di Data Mining sono definite da un algoritmo e da una struttura di conoscenze.

Le caratteristiche comuni che distinguono le varie tecniche di Data Mining sono i tipi di apprendimento, supervisionato o non supervisionato (aspetto approfondito nel seguito), e la natura dell'output, categorico o numerico. Tecniche di Data Mining includono, ad esempio, gli alberi decisionali, i generatori di regole produttive, le reti neurali, le regole associative e le tecniche di clusterizzazione.

Le tecniche di clusterizzazione utilizzano alcuni parametri di similitudine per classificare i casi in classi disgiunte.

I metodi di clusterizzazione sono frequentemente utilizzati per determinare l'insieme ideale di attributi di input per costruire i modelli di apprendimento supervisionato.

Un albero decisionale è una struttura ad albero i cui nodi intermedi rappresentano test su uno o più attributi, mentre i nodi terminali riflettono i risultati delle decisioni.

Un albero di regressione è un albero decisionale i cui nodi foglia hanno valori numerici anziché categorici.

Una regola produttiva è una regola della forma: IF condizioni antecedenti, THEN condizioni conseguenti

Un albero decisionale può essere convertito in una regola produttiva tuttavia non è necessario avere una struttura ad albero iniziale per generare le regole produttive. Le regole associative sono regole produttive il cui termine conseguente può contenere svariate condizioni e relazioni tra gli attributi.

Le reti neurali sono un insieme di nodi interconnessi ideati per riprodurre le funzionalità del cervello umano.

3.3.2 Strategie di Data Mining

Le strategie di Data Mining possono essere classificate in generale come strategie supervisionate e strategie non supervisionate.

L'apprendimento supervisionato costruisce modelli tramite l'utilizzo di attributi di input per predire i valori degli attributi di output.

Molti algoritmi di Data Mining supervisionato ammettono un solo attributo di output.

Altri strumenti di apprendimento supervisionato permettono di specificare uno o più attributi di output. Gli attributi di output sono detti anche variabili dipendenti, in quanto i loro risultati dipendono dai valori di uno o più attributo di input. Gli attributi di input sono detti anche variabili indipendenti.

Quando l'apprendimento non è supervisionato, non esiste un attributo di output; di conseguenza tutti gli attributi utilizzati per costruire i modelli sono variabili indipendenti.

Le strategie di apprendimento supervisionato possono essere ulteriormente classificate in funzione del fatto che gli attributi di output siano discreti o categorici, sia in base al fatto che i modelli siano progettati per determinare una condizione corrente o per predire una situazione futura. Un'ulteriore strategia di Data Mining che si discosta dalle due grandi categorie sopra individuate è la market basket analysis.

Scopo della market basket analysis è ricercare relazioni interessanti tra i prodotti delle vendite al dettaglio al fine di ideare campagne promozionali, la disposizione dei prodotti negli scaffali, nei cataloghi e di sviluppare strategie di marketing.

3.3.3 Apprendimento supervisionato

Si analizzano di seguito i seguenti algoritmi di Data Mining supervisionato:

- C4.5
- CART
- Algoritmo genetico
- Classificatore bayesiano

Algoritmo C4.5

L'algoritmo di Data Mining C4.5 permette di creare sia alberi decisionali con split binario sia alberi che eseguono split multipli sulle features discrete (può essere un modello di previsione).

CART

È stato il primo algoritmo a generare alberi di regressione. Si tratta del più diffuso algoritmo per la costruzione di alberi di classificazione e di regressione. L'algoritmo CART permette di generare alberi binari, ovvero alberi in cui ad ogni nodo corrispondono due soli rami.

L'algoritmo permette di:

- Individuare il miglior classificatore di N individui appartenenti a J gruppi che siano internamente omogenei ed esternamente eterogenei (variabile di risposta in classi)
- Predire una variabile dipendente (continua) attraverso M variabili esplicative (variabile di risposta continua)

Vantaggi:

- Si tratta di un algoritmo veloce
- Tratta tutti i tipi di variabili
- Invariante alle trasformazioni monotone
- Interpretabili
- Capace di gestire dati mancanti e le cui variabili sono correlate

Svantaggi:

- Algoritmo di tipo sequenziale (le scelte effettuate ad un passo influenzano quelle successive) tutti i campi divisori vengono scelti sequenzialmente e ogni campo divisore è di fatto dipendente dai

precedenti. Ciò implica che tutti i futuri campi divisori sono dipendenti dal nodo radice dell'albero, a tal punto che una modifica del campo divisore del nodo radice potrebbe portare alla costruzione di un albero completamente differente

- I risultati sono instabili
- Il guadagno in termine di accuratezza è modesto rispetto alla regressione logistica
- Quando l'algoritmo sceglie, in un determinato nodo, un opportuno campo divisore non è assolutamente detto che quello sia (in assoluto) il miglior campo divisore per quel nodo; sicuramente è il miglior campo divisore per quel nodo in quel preciso istante ed alla luce delle informazioni che l'algoritmo possiede in quell'istante; il problema è che, alla luce di nuove informazioni, l'algoritmo non andrà a modificare i campi divisori dei nodi precedenti.

L'algoritmo parte dalla totalità dei record appartenenti al training set e comincia con una divisione binaria in classi. Il criterio con cui viene effettuata questa divisione si basa sull'information gain e consiste nell'individuare l'attributo che mi da il massimo guadagno di informazione, al fine di stabilire l'appartenenza di un record ad una classe.

Confronto tra C4.5 e CART CART effettua sempre suddivisioni binarie dei dati senza tener conto della natura numerica o categorica delle variabili; CART utilizza i dati test per effettuare lo sfoltimento e poi generalizzare l'albero binario ottenuto, mentre c4.5 usa soltanto i dati di training per creare le strutture ad albero finale. Entrambi operano sia con variabili categoriche che numeriche al contrario di CHAID (solo categoriali) che è un altro algoritmo di creazione di alberi decisionali.

Algoritmo genetico

Gli algoritmi genetici applicano un approccio evolutivo all'apprendimento induttivo, sono stati sviluppati da John Holland nel

1986 e sono basati sul principio di selezione di Darwin. Gli algoritmi genetici possono essere sviluppati per il Data Mining supervisionato e non supervisionato.

Vantaggi:

- Si scelgono, rispetto agli altri tipi di ottimizzazione, quando ci sono molti parametri che descrivono un fenomeno e tra di essi non c'è una evidente correlazione.
- La ricerca stocastica dell'ottimo globale viene fatta in parallelo e questo li rende più efficienti e meno suscettibili agli ottimi locali. Essendo in parallelo, infatti, quello che avviene in una certa porzione della popolazione non dipende da quello che avviene in altre parti della popolazione. In pratica effettuano una ricerca Montecarlo molto efficiente con in più una progressione evolutiva.

Svantaggi:

- non ci sono garanzie che una data soluzione sia il risultato di un'ottimizzazione locale anziché di un'ottimizzazione globale
- l'impiego di una funzione di fitness può richiedere elevati tempi di elaborazione
- spiegano i loro risultati nell'ambito della comprensione della funzione di fitness
- trasformare i dati in un formato adatto agli algoritmi genetici può diventare un'impresa ardua

Algoritmo genetico e apprendimento supervisionato L'algoritmo genetico relativo all'apprendimento supervisionato permette di costruire un modello di classificazione o previsione.

Si presenta una algoritmo genetico di base:

1. si inizializza una popolazione P di n elementi, detti cromosomi, come soluzione potenziale

2. finche non è soddisfatta una specifica condizione di conclusione (numero di iterazioni o gli elementi della popolazione devono tutti soddisfare un determinato criterio minimo):
 - (a) si utilizza una funzione di fitness e dei dati di training per valutare ogni elemento della soluzione corrente. Se un elemento supera i criteri di fitness rimane nella popolazione
 - (b) la popolazione adesso contiene m elementi ($m \leq n$). Si utilizzano gli operatori genetici per creare n nuovi elementi da aggiungere.

Gli operatori genetici più largamente utilizzati sono il crossover, la mutazione e la selezione. Terminato l'algoritmo, si ha a disposizione un modello per classificare nuove osservazioni o per predire risultati futuri. Per usare il modello, si può mettere a confronto una nuova osservazione sconosciuta con gli elementi della popolazione finale. Una semplice tecnica consiste nel dare alle osservazioni sconosciute la stessa classificazione dell'elemento della popolazione cui sono più simili. Una seconda possibilità consiste nel registrare gli m elementi della popolazione più simili all'osservazione presentata per la classificazione, poi l'algoritmo sceglie casualmente uno degli m elementi e fornisce all'osservazione sconosciuta la classificazione dell'elemento scelto a caso.

Classificatore Bayesiano

È una tecnica di classificazione supervisionata. Il modello assume che le variabili siano indipendenti tra di loro e di uguale importanza. La struttura è basata sul teorema di Bayes. Permette data una nuova osservazione di stabilire quale sarà il valore di un attributo non noto.

Apprendimento non supervisionato

L'apprendimento non supervisionato crea cluster che possono essere **gerarchici o non gerarchici**.

Il clustering gerarchico determina una gerarchia, ovvero una struttura di cluster in cui l'insieme di unità viene bipartita ricorsivamente in corrispondenza di diversi livelli di aggregazione.

Il clustering non gerarchico determina una partizione, l'insieme delle unità viene partizionato in k cluster omogenei disgiunti.

Clustering non gerarchico

Algoritmo delle k-medie Si tratta di una tecnica di clusterizzazione non supervisionata.

L'algoritmo k-means riceve in input un parametro k e partiziona un insieme di n oggetti in k cluster in modo tale che la similarità intra-cluster risultante sia alta mentre la similarità inter-cluster sia bassa. La similarità dei cluster è misurata rispetto al valore medio degli oggetti in un cluster; tale valore può essere visto come il centro di gravità del cluster.

L'algoritmo k-means procede nel seguente modo. Innanzitutto, riceve in input il numero di cluster da generare, k , successivamente seleziona randomicamente k degli oggetti, ciascuno dei quali rappresenta inizialmente la media o il centro di un cluster. Ciascuno degli oggetti rimanenti viene associato al cluster più simile basandosi sulla distanza tra l'oggetto e la media del cluster.

Esso, quindi, calcola la nuova media per ciascun cluster. Tale processo viene iterato fino a quando non si raggiunge la convergenza di una determinata funzione criterio.

Vantaggi:

- creazione di cluster stabili (non coincidente necessariamente con quella ottimale)

Svantaggi:

- configurazione finale dei cluster dipendente dalla scelta casuale dei k punti iniziali scelti dall'algoritmo
- l'algoritmo opera solo con valori reali
- scelta del numero di cluster
- l'algoritmo funziona meglio quando i cluster presenti nei dati hanno la stessa dimensione. Se la soluzione migliore è rappresentata da cluster di dimensioni diverse l'algoritmo non sempre è in grado di trovare la soluzione ideale
- noi dobbiamo interpretare i risultati ottenuti. Si possono usare strumenti di datamining supervisionato per capire meglio la natura dei cluster generati
- non c'è modo di sapere quali variabili siano significative per determinare i cluster

Una strategia interessante, che spesso restituisce buoni risultati, consiste nell'applicare prima un algoritmo di agglomerazione gerarchica per determinare il numero di cluster e per trovare una classificazione iniziale, e nell'utilizzare successivamente la rilocazione iterativa per migliorare la classificazione.

Un'altra variante del k-means è il metodo k-modes che estende il paradigma k-means per clusterizzare i dati categorici, rimpiazzando le medie dei cluster con le mode, adottando nuove misure di dissimilarità per

trattare oggetti categorici e utilizzando metodi basati sulla frequenza per aggiornare le mode.

K-Medoids L'algoritmo k-means è sensibile agli outlier, dal momento che un oggetto con un valore estremamente grande può distorcere sensibilmente la distribuzione dei dati.

Per diminuire la sensitività, invece di prendere come punto di riferimento il valore medio degli oggetti in un cluster, può essere utilizzato il medoid, ovvero l'oggetto localizzato più centralmente in un cluster.

In questo modo è possibile comunque eseguire il metodo di partizionamento cercando di minimizzare la somma delle dissimilarità tra ciascun oggetto e il suo punto di riferimento corrispondente.

La strategia sostituisce iterativamente medoidi con non-medoidi e termina quando la qualità del clustering risultante non può più essere migliorata. Questa qualità viene stimata utilizzando una funzione di costo che misura la dissimilarità media tra un oggetto e il medoide del suo cluster.

Il metodo k-medoids è più robusto di k-means in presenza di rumore e di outlier perché un medoide è meno influenzato dagli outlier o da altri valori estremi rispetto ad una media. Tuttavia, l'elaborazione di k-medoids è più costosa rispetto al metodo k-means.

Entrambi i metodi richiedono all'utente di specificare k , il numero di cluster.

Algoritmo genetico e clusterizzazione non supervisionata L'algoritmo genetico relativo alla clusterizzazione non supervisionata clusterizza i dati dopo aver ricevuto in input il numero di cluster da generare, k .

Supponiamo di avere una popolazione di P elementi e che a ciascuno di essi siano associati n valori, l'algoritmo genera m possibili soluzioni.

Ogni soluzione contiene k punti n -dimensionali. Tramite la funzione di fitness e gli operatori genetici al termine si individueranno m soluzioni; la soluzione da scegliere è quella che presenta il valore di fitness migliore tra le stesse.

Clustering gerarchico

Nel seguito si spiega nel dettaglio un algoritmo di clustering gerarchico: algoritmo agglomerativo.

Clusterizzazione agglomerativa Si tratta di una tecnica di clusterizzazione non supervisionata.

Comincia assumendo che ogni record crea un cluster a sé stante e successivamente unisce i cluster più simili; la procedura termina quando tutte le unità risultano aggregate in un unico cluster.

La clusterizzazione da preferire si determina mediante diverse tecniche. Possono essere applicate sia misure euristiche che statistiche.

La clusterizzazione agglomerativa viene comunemente utilizzata come preludio alle altre tecniche di clusterizzazione. Per esempio si può applicare tale tecnica per individuare un valore iniziale per k delle k -means ponendolo pari alla media dei cluster individuati dall'algoritmo.

3.3.4 L'algoritmo implementato

Al fine di far collassare eventi istituzionali coincidenti, ovvero più occorrenze di uno stesso evento, è stato implementato il seguente algoritmo.

Si ipotizza, innanzitutto, che occorrenze in *event_point* con uguali valori di *city*, *street*, *road_code* e *category_id* ed *order_in_event* = 0 sono potenziali occorrenze di uno stesso evento, diventano reali occorrenze di uno stesso evento se la distanza tra i valori dell'attributo *ts_create* corrispondenti, individuati mediante un operazione di join tra le tabelle *event* ed *event_point* sugli attributi rispettivamente *id* ed *event_id*, è inferiore al lifetime della corrispondente categoria di appartenenza dell'evento.

Il processo di identificazione dei gruppi (cluster) di occorrenze potenziali secondo il criterio di similarità dell'ora di pubblicazione dell'evento, è avvenuto secondo l'approccio non supervisionato.

L'approccio all'*unsupervised clustering* seguito è il clustering gerarchico.

Con il clustering gerarchico viene creata una gerarchia di partizioni caratterizzata da un numero crescente o decrescente di gruppi, visualizzabile mediante un diagramma ad albero (dendrogramma).

I principali algoritmi del *hierarchical clustering* sono due, si distinguono in base all'approccio seguito:

- approccio **bottom-up**: si creano tanti clusters quanti sono gli oggetti (singleton), e per passaggi successivi si uniscono gli oggetti tra loro più vicini, fino ad ottenere un gruppo contenente tutti gli oggetti (hierarchical agglomerative clustering)
- approccio **top-down**: da un unico cluster contenente tutti gli oggetti, si splitta l'insieme fino ad ottenere clusters contenenti ciascuno un solo oggetto (hierarchical divisive clustering).

Il metodo usato nella presente analisi è quello bottom-up. Lo strumento utilizzato per applicare la tecnica di Data Mining scelta è il *software R*.

R è un ambiente di sviluppo open source per l'analisi statistica.

La funzione utilizzata è la funzione *hclust*, del pacchetto *stats*, la quale riceve in input la matrice delle distanze (ovvero l'output della funzione *dist*) costruita sull'attributo sul quale si vuole clusterizzare.

Nella presente analisi si tratta dell'attributo *date*, nel seguito tale aspetto verrà trattato nel dettaglio.

I valori di *date* si sono ottenuti trasformando i valori di *ts_create* in secondi al fine di permettere il calcolo della matrice delle distanze. Esempio: il valore 2015-05-08 19 : 50 : 00 sotto l'attributo *ts_create* è stato trasformato nel valore 1431107400 sotto l'attributo *date*.

La funzione *dist* riceve in input il vettore delle date delle occorrenze potenziali e restituisce una matrice in cui l'elemento *i-j* esimo rappresenta la differenza semplice tra l'elemento *i* e l'elemento *j* del vettore *date*.

La funzione *dist* mette a disposizione diverse metodologie di calcolo delle distanze ("euclidean", "maximum", "manhattan", "canberra", "binary" o "minkowski"), ai fini della nostra analisi si è ritenuto opportuno, tuttavia, calcolare la differenza semplice in valore assoluto.

```

> query<-paste0("select date, id, identifier,
ts_create,category_id,street,city,road_code, source from joined_table where
(source='regionelazio.luceverde.it' or source='roma.luceverde.it' or
source='www.milano.luceverde.it' or source='5T' or source='www.cciss.it' or
source='CCISS_Ministero' or source='StradeANAS' or source='LuceverdeMilano' or
source='LuceverdeLazio' or source='LuceverdeRoma') and identifier in (select
identifier from joined_table limit 1 offset ", i ,") ")
> a<-dbGetQuery(conn,query)
> b<-a$date
b
1431082740 1435742580 1435746420 1435758660 1435759080 1435858860 1435862220
c<-matrix(b,nrow=length(b),ncol=1)
distance<-dist(c,"eucl")
distance
      1      2      3      4      5      6
2 4659840
3 4663680    3840
4 4675920   16080   12240
5 4676340   16500   12660    420
6 4776120  116280  112440  100200   99780

```

Fig. 3.14: Esempio di calcolo della matrice delle distanze

Si riportano i comandi implementati e un esempio su quanto detto:

In altri termini, dati i valori di *date* in posizione 1 e 5, rispettivamente pari a 1431082740 e 1435759080, il valore della quinta riga e prima colonna della matrice *distance* è pari a $4676340 = |431082740 - 1435759080|$.

La funzione *hclust* prende in ingresso la matrice delle distanze e, inizialmente, tratta ogni unità come un cluster. Successivamente, mediante un algoritmo iterativo, unisce, ad ogni step, i due clusters più simili fino ad ottenere un unico cluster contenente tutte le unità.

La visualizzazione del dendrogramma permette di individuare graficamente l'insieme delle partizioni annidate che sono state calcolate.

È possibile altresì visualizzare quali oggetti vengono inglobati e a quale livello (altezza) di fusione *h* ovvero il valore della distanza massima tra 2 elementi consecutivi di un cluster.

Il valore di *h*, nel presente caso, è espresso in secondi, poiché l'algoritmo di clusterizzazione si applica sull'attributo *date*, come precedentemente precisato.

Esso dipende dal valore di *category_id* degli eventi potenziali.

In particolare, è posto pari al valore del rispettivo lifetime, dunque:

- $h = 3600$ se $category_id = 1$
- $h = 43200$ se $category_id = 2$ o 3
- $h = 604800$ se $category_id = 4$
- $h = 86200$ se $category_id = 5$

Fissato il valore per h , l'algoritmo restituisce per ogni stringa il cluster di appartenenza; stringhe appartenenti ad uno stesso cluster sono avranno così uguali valori per $street$, $road_code$, $city$, $category_id$ e saranno in media distanziati da h secondi, dunque afferiscono ad uno stesso fatto.

Il popolamento della tabella *aggregated_event* è avvenuto selezionando per ogni cluster l'evento con ts_create minore, ovvero la stringa che presenta data di pubblicazione più remota al fine di evidenziare il momento di inizio dell'evento.

Si ottiene dunque una tabella contenente una sola occorrenza per uno stesso evento.

Si riporta la procedura di individuazione di occorrenze reali di uno stesso evento:

1. Selezione eventi potenziali.

Si selezionano gli eventi potenziali tra quelli pubblicati dalle fonti ufficiali, ovvero, le stringhe con uguale valore di identifier (si fissa $i = 25$ ovvero il valore di *identifier* in posizione 25) (Figura 3.15).

I valori di *identifier* si ottengono concatenando i valori di *road_code*, *street*, *city* e *category_id*.

2. Verifica della condizione di applicabilità dell'algoritmo.

Si verifica se il numero di elementi selezionati è maggiore di uno. Si rende necessaria tale verifica in quanto altrimenti la funzione *hclust* non restituisce i risultati voluti. Poiché il numero di eventi potenziali è pari a 6, la condizione di applicabilità è verificata.

```

i=25
query<-paste0("select date, id, identifier,
ts_create,category_id,street,city,road_code, source from joined_table where
(source='regionelazio.luceverde.it' or source='roma.luceverde.it' or
source='www.milano.luceverde.it' or source='5T' or source='www.cciss.it' or
source='CCISS_Ministero' or source='stradeANAS' or source='LuceverdeMilano' or
source='LuceverdeLazio' or source='LuceverdeRoma') and identifier in (select
identifier from joined_table limit 1 offset ", i ,") ")
a<-dbGetQuery(conn,query)
b<-a$date
a

```

	date	id	identifier	ts_create	category_id	street	city	road_code
			A14Forli'					
1431078840		26	Bologna Ancona4	2015-05-08 11:54:00+02	4	Bologna Ancona	Forli'	A14
			A14Forli'					
1431114480		872	Bologna Ancona4	2015-05-08 21:48:00+02	4	Bologna Ancona	Forli'	A14
			A14Forli'					
1431328140		2637	Bologna Ancona4	2015-05-11 09:09:00+02	4	Bologna Ancona	Forli'	A14
			A14Forli'					
1435590600		52911	Bologna Ancona4	2015-06-29 17:10:00+02	4	Bologna Ancona	Forli'	A14
			A14Forli'					
1435659300		53837	Bologna Ancona4	2015-06-30 12:15:00+02	4	Bologna Ancona	Forli'	A14
			A14Forli'					
1435836600		55906	A14Forli' Bologna Ancona4	2015-07-02 13:30:00+02	4	Bologna Ancona	Forli'	A14

Fig. 3.15: Selezione eventi potenziali

3. Applicazione funzione *hclust*.

Si applica l'algoritmo gerarchico che permette di individuare gli eventi vicini, secondo la misura h , dal punto di vista temporale.

(a) Si individuano in maniera iterativa tutti i cluster, (Figura 3.16):

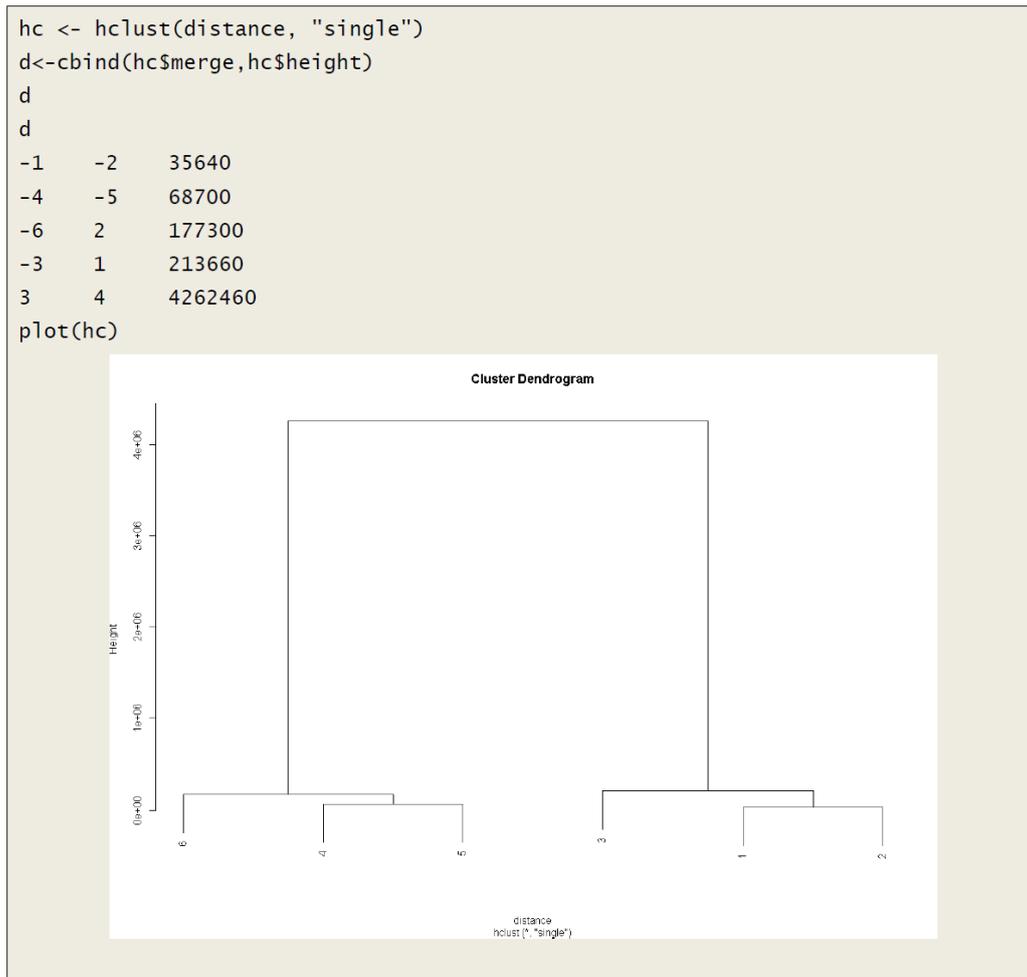


Fig. 3.16: Cluster individuati in maniera iterativa

Dal dendogramma, "Cluster Dendogram", o equivalentemente, dalla risposta della funzione *hclust*, emerge che:

- i. La prima iterazione ha comportato l'inserimento in uno stesso cluster (cluster 1) degli elementi in posizione 1 e 2;
- ii. La seconda l'inserimento degli elementi in posizione 4 e 5 nel cluster 2
- iii. La terza l'inserimento dell'elemento in posizione 6 nel cluster 2 e quindi la creazione del cluster 4
- iv. La quarta l'inclusione dell'elemento in posizione 4 nel cluster 1 (cluster 5)
- v. E infine l'unione dei cluster 3 e 4.

Tutti gli elementi, nell'ultima iterazione, fanno parte dello stesso cluster.

(b) Individuazione del valore di h .

Poiché $category_id = 4$, $h = 604800$ (Figura 3.17).

```
query_peso<-paste0("select lifetime from event_category where
id=",a$category_id[1],")
peso_mill_sec<-dbGetQuery(conn,query_peso)
peso<-peso_mill_sec/1000
peso
lifetime
604800
```

Fig. 3.17: Individuazione del valore di h

(c) Individuazione dei cluster i cui elementi pubblicati consecutivamente distano al più di h (Figura 3.18)

```
cluster<- cutree(hc,h=peso)
cluster
1 1 1 2 2 2
```

Fig. 3.18: Individuazione dei cluster i cui elementi distano in media di h

Dunque, gli elementi in posizione 1, 2 e 3 appartengono al cluster 1 e i rimanenti al cluster 2.

I cluster individuati sono mostrati nella Figura 3.19



Fig. 3.19: Cluster selezionati

La Figura 3.20 mostra l'accurata della clusterizzazione degli eventi. Infatti, eventi appartenenti ad uno stesso cluster, essendo *category_id* = 4, sono pubblicati nella stessa settimana.

```
x<-cbind(cluster,a$id_e,a$street,a$city,a$road_code,
a$category_id,a$date/1000000,a$ts_create)
x
```

cluster	id	street	city	road_code	category_id	date	ts_create
1	26	Bologna Ancona	Forlì'	A14	4	1,43E+09	2015-05-08 11:54:00
1	872	Bologna Ancona	Forlì'	A14	4	1,43E+09	2015-05-08 21:48:00
1	2637	Bologna Ancona	Forlì'	A14	4	1,43E+09	2015-05-11 09:09:00
2	52911	Bologna Ancona	Forlì'	A14	4	1,44E+09	2015-06-29 17:10:00
2	53837	Bologna Ancona	Forlì'	A14	4	1,44E+09	2015-06-30 12:15:00
2	55906	Bologna Ancona	Forlì'	A14	4	1,44E+09	2015-07-02 13:30:00

Fig. 3.20: Clusterizzazione degli eventi

In generale:

1. Crea la tabella *aggregated_event*, nella quale saranno inseriti gli eventi aggregati con attributi *cluster*, *event_id*, *street*, *city*, *road_code*, *category_id*, *date*, *ts_create*, *source*
2. Crea la tabella *joined_table* frutto del join delle tabelle *event* ed *event_point* sull'attributo *id* e *event_id* rispettivamente con gli attributi *ts_create*, *id*, *category_id*, *street*, *city*, *road_code*, *source*, *date* (presenta i valori di *ts_create* trasformati in secondi) e *identifier* (concatena i valori di *category_id*, *street*, *city*, *road_code*)
3. Seleziona le occorrenze potenziali di uno stesso evento, ovvero le stringhe che presentano stessi valori per *street*, *city*, *road_code* e *category_id* e che sono pubblicati da enti ufficiali
4. Se il numero di stringhe selezionato è uguale a 1, la stringa selezionata viene copiata in *aggregated_event* e il ciclo termina; altrimenti:

- (a) Ogni unità viene considerata come un cluster, successivamente, in maniera iterativa , i clusters più simili, secondo la misura temporale, si uniscono fino ad ottenere un unico cluster contenente tutte le unità.
 - (b) Si individua il valore della distanza massima tra elementi pubblicati consecutivamente che deve caratterizzare la clusterizzazione finale
 - (c) Si individua la clusterizzazione che rispetta il requisito presedente.
 - (d) Per ogni cluster, si seleziona la stringa più remota e la inserisce nella tabella *aggregated_event* specificando il valore per source pari ad "Ente ufficiale"
5. Seleziona le occorrenze potenziali di uno stesso evento, ovvero le stringhe che presentano stessi valori per *street*, *city*, *road_code* e *category_id* e che sono pubblicati da soggetti appartenenti alla community
6. Si ritorna al passo 4 e termina al passo 4d.

L'algoritmo va richiamato ogni x secondi dove $x = lifetime$ che dipende dalla categoria dell'evento. Dovrà agire sulla tabella contenente i nuovi eventi e , inoltre, sull'evento più recente già clusterizzato Y . Nella tabella *aggregated_event* gli eventi appartenenti allo stesso cluster di Y , saranno aggiunti al cluster già esistente.

CAPITOLO 4

APPLICAZIONE ANDROID

Al fine di rendere fruibili le informazioni raccolte dalla piattaforma "ITS 2.0" per mezzo dell'aggregatore Twitter e dell'API di Data Ingestion, si è resa necessaria la realizzazione di una applicazione Android.

I requisiti della nuova app possono essere annoverati all'interno delle seguenti voci:

- **Visualizzazione eventi:** attraverso l'app deve essere possibile il reperimento di informazioni di infomobilità real time posizionati sulla mappa;
- **Invio segnalazioni:** l'app deve consentire l'invio di segnalazioni raccolte real time dagli utenti;
- **Validazione degli eventi:** gli eventi pubblicati da canali non ufficiali necessitano di essere validati dagli utenti della community ai fini della valutazione della loro veridicità;
- **Navigazione:** l'app deve permettere di impostare un itinerario (date location di partenza e destinazione) e fornire indicazioni circa il percorso da seguire;

- **Individuazione eventi lungo il percorso individuato:** l'app deve permettere di visualizzare i singoli eventi di infomobilità rilevati lungo il percorso selezionato tramite la funzionalità di navigatore.

Si è ritenuto che nell'integrazione di tali funzionalità si dovesse perseguire l'obiettivo di una User Experience adeguata all'utilizzo in modalità "alla guida": si vuole intendere con ciò che l'app non deve favorire la distrazione del guidatore ma facilitarlo attraverso un'interfaccia adeguata al contesto d'uso.

Al fine di rispondere a tale esigenza si è considerata utile l'implementazione di un'interfaccia d'uso basata su *Comandi Vocali*: l'utilizzatore dell'app, inserito nel contesto urbano, avrà così modo di interagire con la piattaforma senza distrarsi dalla guida ma avendo al tempo stesso la possibilità di inviare segnalazioni e di contribuire al reperimento e alla condivisione di eventi di infomobilità.

L'applicazione dovrà tuttavia anche fornire la possibilità di inviare segnalazioni in modalità touch tradizionale al fine di condividere segnalazioni in tempo reale anche in un contesto d'uso diverso da quello di guida; tale funzionalità dovrà persistere nel perseguire l'obiettivo di user experience facilitata per consentire all'utente di partecipare al reperimento di informazioni con velocità e senza dispendio di tempo.

La funzionalità di navigatore che si è deciso di implementare assume notevole importanza soprattutto nell'individuazione di un itinerario sul quale effettuare ricerche di eventi in corso tali da richiedere una modifica del percorso.

4.1 Casi d'uso

L'applicazione, come si è detto, ha come target quello degli automobilisti; si rivolge infatti a coloro che intendono percorrere strade urbane/extra-urbane con il proprio mezzo di trasporto e che hanno necessità di conoscere gli eventi di infomobilità real-time al fine di adottare azioni correttive circa il proprio itinerario. Tuttavia, è opportuno pensare

che alcune tipologie di segnalazioni si rendono inviabili in maniera asincrona rispetto alla loro osservazione: basti pensare, ad esempio, al rilevamento di un cantiere stradale lungo l'itinerario percorso e alla necessità che ne potrebbe conseguire di segnalazione dell'evento una volta raggiunta la destinazione (perché l'app, ad esempio, non era attiva al momento dell'osservazione). Lo studio dell'interfaccia dell'applicazione è dunque passato per la caratterizzazione delle due modalità d'uso che caratterizzano il suo utilizzo descritte nella seguente figura:

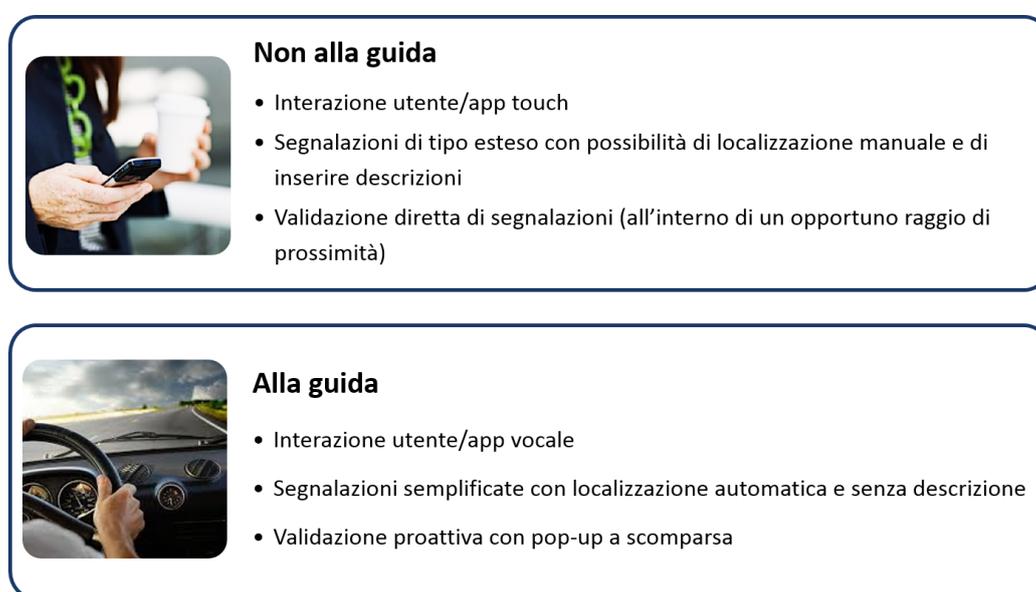


Fig. 4.1: Modalità d'uso previste per l'applicazione

Tale caratterizzazione ha guidato la definizione della UI nella sua globalità, sia per ciò che concerne l'interazione funzionale con le azioni previste dall'applicazione sia per quanto riguarda l'ergonomia attraverso cui tali funzionalità si rendono accessibili.

L'approccio seguito ha previsto una prima fase di studio delle caratteristiche di cui le app competitor analizzate sono state ritenute deficitarie. In particolare, è stato ritenuto che l'eccessivo utilizzo di spazio grafico delle schermate potesse risultare un elemento di caratterizzazione negativa per un'app rivolta ad un uso automobilistico: sebbene sia valida l'intenzione di implementare funzionalità a controllo vocale, non è

possibile prescindere da una benché minima interazione touch/visiva con l'applicazione e si rende dunque necessario che la reperibilità visiva dell'elemento grafico di interesse sullo schermo sia adeguata al contesto d'uso.

Al fine di porre le basi per lo sviluppo effettivo dell'applicazione, è stato ritenuto opportuno procedere con uno studio delle interfacce grafiche connesse alle funzionalità previste per mezzo della realizzazione di mockup; tali riproduzioni hanno successivamente ispirato lo sviluppo effettivo delle interfacce grafiche dell'app.

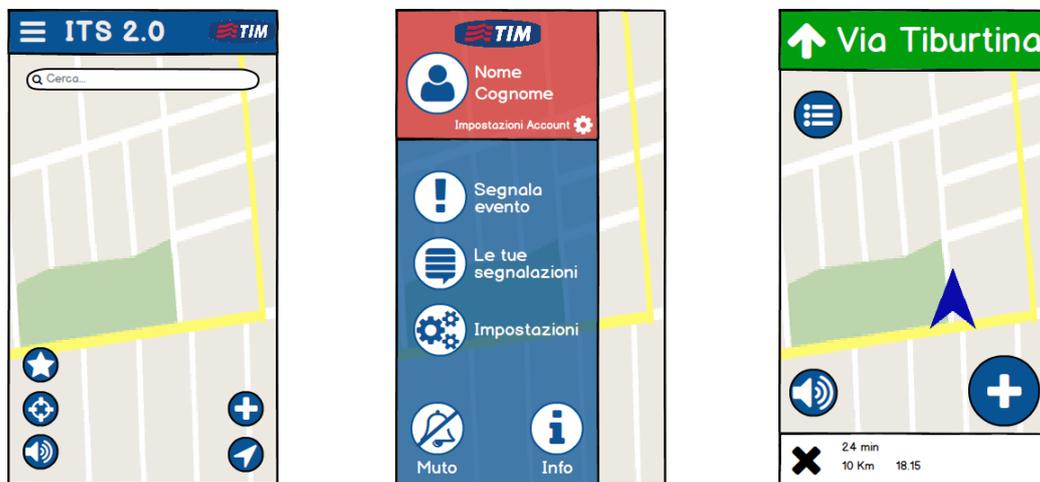


Fig. 4.2: Esempi di Mockup di interfacce grafiche dell'applicazione: App Home, Navigation Drawer, Modalità Navigazione

4.2 Nome e Logo Applicazione

In virtù dell'accezione pubblica del prodotto da sviluppare, è stato ritenuto opportuno individuare un Nome ed un Logo per l'applicazione mobile della piattaforma "ITS 2.0".

Prendendo spunto dal filone dei servizi TIM per l'entertainment (e.g. TIMmusic, TIMvision, etc.), si è pensato di adottare la medesima struttura grammaticale andando a sottendere semanticamente l'ambito in cui tale progetto si colloca.

La scelta è ricaduta sulla soluzione **TIMobility**, che, da una parte vuole enfatizzare il brand "TIM" e il concetto di mobilità, dall'altro vuole essere una esortazione alla mobilitazione, "Ti mobiliti".

Per quanto concerne il logo, si è pensato anche in tale contesto di prendere ispirazione dal logo aziendale volendo al tempo stesso esprimere il concetto di mobilità.



Fig. 4.3: Logo App TIMobility

4.3 Descrizione funzionalità

In questo paragrafo verranno descritte le funzionalità sviluppate per l'applicazione, ora identificata con il nome TIMobility. Tali funzionalità saranno esposte attraverso la presentazione degli screenshot delle relative schermate a cui verranno allegate le opportune descrizioni delle soluzioni implementative adottate.

L'applicazione è stata sviluppata per mezzo dell'IDE di sviluppo Android Studio in linguaggio di programmazione Java.

4.3.1 Login tramite credenziali fornite da Identity Manager

L'accesso alle informazioni della piattaforma è possibile tramite autenticazione con credenziali fornite dal processo di registrazione dell'Identity Manager sviluppato da Telecom Italia. Si mostra di seguito la relativa schermata di login.

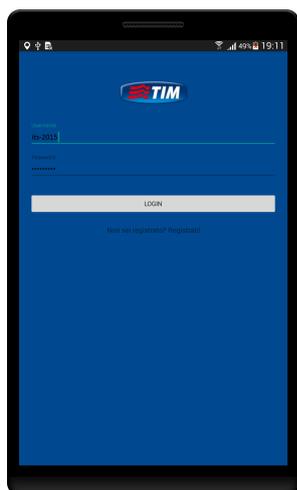


Fig. 4.4: Schermata di Login

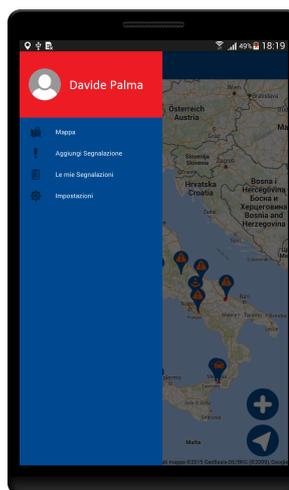


Fig. 4.5: Menu principale

4.3.2 Visualizzazione eventi di infomobilità

Per mezzo dell'app si devono poter visualizzare gli eventi di infomobilità che la piattaforma "ITS 2.0" raccoglie per mezzo delle fonti istituzionali presenti su twitter, canali RSS e, a valle della sperimentazione, dei contributi singoli degli utenti che saranno solo allora possibili.

Tali informazioni possono essere recuperate dal server residente presso il TILab di Torino per mezzo di un'apposita API di Data Ingestion fornita da Telecom Italia e accessibile mediante chiamate http con informazioni in input somministrate tramite strutturazione di request in formato JSON.

Una volta eseguito l'accesso, si potrà effettuare la localizzazione della posizione occupata e, dunque, degli eventi di infomobilità in corso.

La Figura 4.6 e la Figura 4.7 mostrano rispettivamente i marker posizionati in corrispondenza degli eventi di traffico reperiti dalla banca

dati "ITS 2.0" e differenziati per categoria di evento e l'infowindow visualizzata al click di uno specifico marker e riportante le informazioni del relativo evento; tali informazioni appartengono ad appositi campi del response JSON (es. localizzazione geografica, strada di riferimento, km di riferimento).

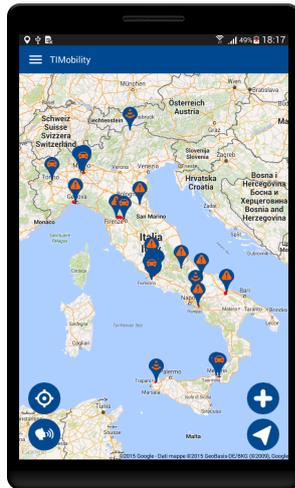


Fig. 4.6: Eventi di infomobilità reperiti real time dal database

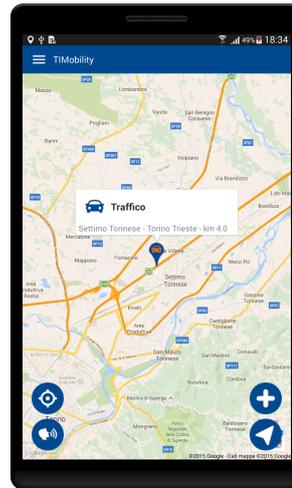


Fig. 4.7: Infowindow di un marker relativo ad un evento

Le categorie e sottocategorie degli eventi gestiti dalla piattaforma sono le seguenti:

ICONA	DESCRIZIONE
	Traffico (coda, lieve, altro)
	Pericolo (incidente, lavori, ostacoli, altro)
	Meteo (pioggia, neve, nebbia, altro)
	Chiusura (strada, galleria, corsia d'emergenza, altro)
	Cantieri (lavori in corso, cantiere, pulizia strade, altro)

Fig. 4.8: Categorie degli eventi

4.3.3 Invio segnalazioni

L'app dovrà permettere di inviare segnalazioni di eventi di infomobilità rilevati real time (crowd sourcing). Sulla base degli casi d'uso individuati sono state implementate 2 differenti modalità di segnalazione:

- **Modalità di segnalazione touch:** modalità attraverso la quale possono essere inviate segnalazioni per mezzo dell'inserimento di informazioni tramite tap/selezione di voci.

L'impostazione grafica adottata è tale per cui con soli n.4 tap dell'interfaccia si è in grado di inviare una segnalazione esplicitiva circa l'evento rilevato (ulteriori campi testuali possono completare le informazioni già inserite).

Accedendo al menu laterale dell'app (Figura 4.5) e selezionando la voce "Segnala evento", si accede alla schermata mostrata dalla

Figura 4.9; ciascuna delle voci di categoria possiede una sotto-classificazione in sottocategorie (alla selezione delle opzioni viene composto un package informativo che verrà recepito dal server, decodificato e memorizzato opportunamente.

- **Modalità di segnalazione vocale:** modalità attraverso la quale possono essere inviate segnalazioni per mezzo dell'inserimento di informazioni tramite comandi vocali. Premendo il tasto contraddistinto dal segno "+" (Figura 4.6), si attiva il pop-up menu che consente di selezionare, rispondendo alle richieste della voce guida, la tipologia di evento che si vuole comunicare e la sottocategoria descrittiva; il sistema rileverà la posizione occupata dal dispositivo e trasmetterà alla piattaforma tramite request JSON l'informazione composta da categoria-sottocategoria-localizzazione.

Screenshot invio segnalazione - modalità touch



Fig. 4.9: Selezione categoria Traffico



Fig. 4.10: Selezione categoria Pericolo



Fig. 4.11: Selezione categoria Meteo



Fig. 4.12: Selezione categoria Chiusura



Fig. 4.13: Selezione categoria Cantieri



Fig. 4.14: Caso d'uso: Chiusura-Strada-Loc. manuale

Screenshot invio segnalazione - modalità vocale

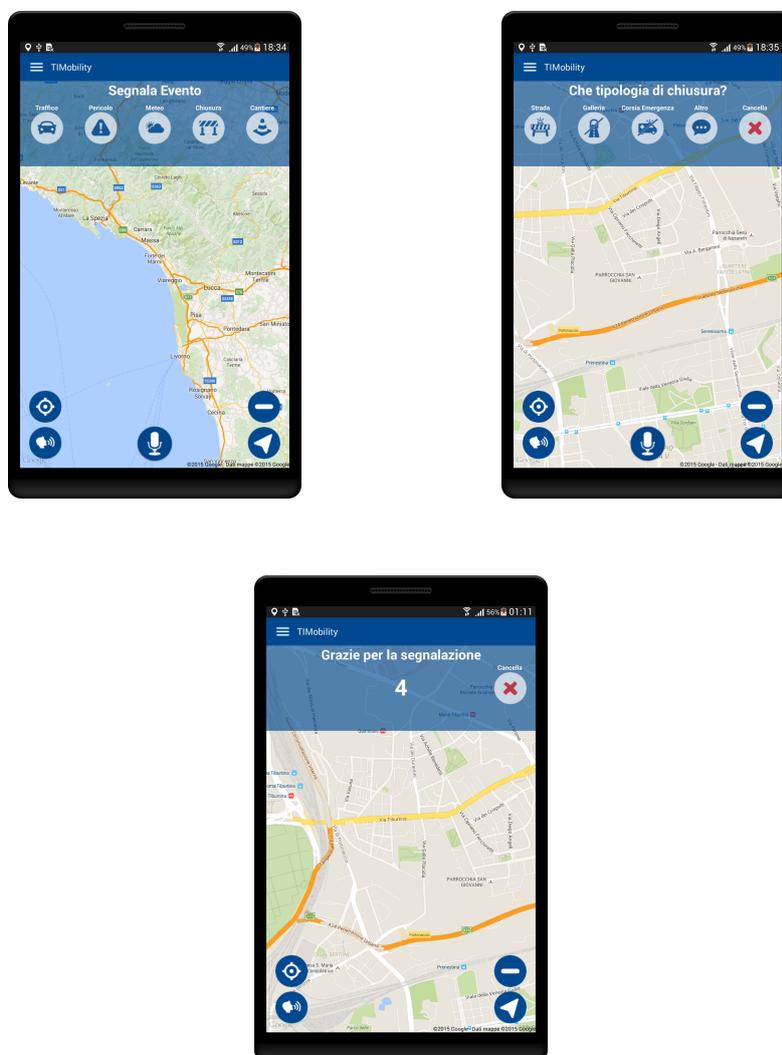


Fig. 4.15: Invio della segnalazione tramite comandi vocali: alla pressione del tasto di aggiunta, viene mostrato il pop-up di scelta della categoria e a seguire, dopo la pressione di una delle voci previste, il tab di sotto-categoria associato alla categoria selezionata inizialmente

4.3.4 Validazione degli eventi

La community di utenti può partecipare al processo di validazione delle segnalazioni inviate da canali non istituzionali ai fini della valutazione della veridicità degli eventi associati.

È chiaro infatti che non vi è alcun motivo di valutare se un evento pubblicato da uno dei canali ufficiali coinvolti dalla piattaforma è veritiero (lo si considera tale a prescindere) mentre è utile verificare che lo siano quelli trasmessi dai singoli partecipanti al processo di raccolta.

Tale procedimento si rende possibile attraverso l'operazione di validazione: alla selezione di un marker, l'infowindow ad esso relativa presenterà gli indicatori di conferma e smentita (Figura 4.16) e al suo tap apparirà un dialog con la richiesta di conferma della veridicità dell'evento.

La risposta selezionata costituirà a sua volta una comunicazione al server della banca dati che provvederà ad assegnare l'opportuno valore di validation all'evento e reputation al suo owner.

4.3.5 Navigatore e Lista Eventi lungo il percorso

TIMobility è in grado di fungere da navigatore. Infatti, inserendo una location di partenza ed una di destinazione, è possibile reperire tramite l'API Directions di Google l'itinerario del percorso comprensivo dei suggerimenti da seguire per lo spostamento e le indicazioni circa i tratti stradali da percorrere.

Il percorso viene mostrato attraverso una polilinea di colore blu che ricalca i tratti stradali da percorrere; la posizione del guidatore viene indicata attraverso un marker specifico che si aggiorna in corrispondenza dello spostamento rilevato.

I suggerimenti stradali vengono visualizzati coerentemente con la posizione occupata nella parte alta dello schermo all'interno di un apposito frame, il contenuto testuale viene allo stesso tempo letto dalla funzionalità di sintesi vocale integrata con il sistema Android.

All'attivazione della navigazione viene inoltre reso possibile il reperimento di eventi di infomobilità localizzati lungo il percorso

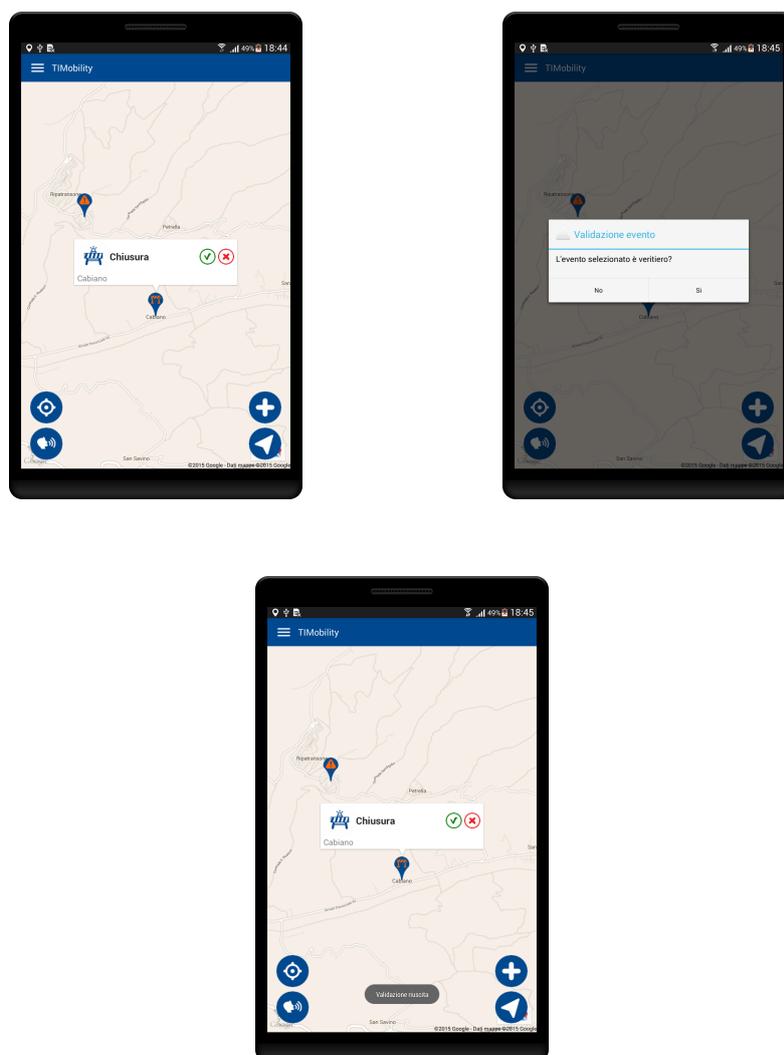


Fig. 4.16: Infowindow di un evento segnalato da un utente (con i caratteristici indicatori di validazioni), dialog attivato al tap dell'infowindow, indicazione di avvenuta validazione

individuato; alla pressione di uno specifico pulsante, viene visualizzato un dialog con la lista di tali eventi, il tutto al fine di ottenere delle informazioni utili per l'adozione di azioni correttive nel corso dello spostamento. L'app è inoltre in grado di fornire alert all'avvicinarsi di un evento con pop-up a comparsa/scomparsa volti ad avvisare il guidatore circa l'accadimento all'interno del proprio raggio di prossimità.

Il pop-up permette all'utente di validare/negare l'evento indicato per mezzo di appositi pulsanti in sovrapposizione o di nascondere il frame qualora non si voglia interagire con la piattaforma.

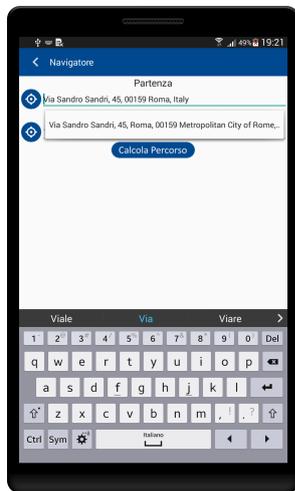


Fig. 4.17: Inserimento Partenza e Arrivo con funzionalità di rilevamento della posizione

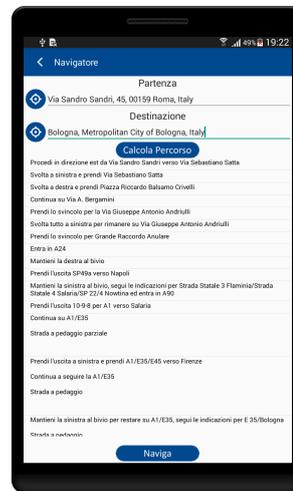


Fig. 4.18: Indicazioni fornite in riferimento all'itinerario richiesto

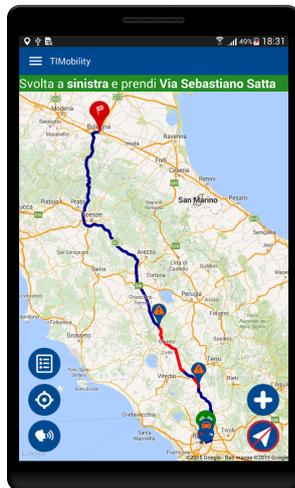


Fig. 4.19: Visualizzazione del percorso calcolato

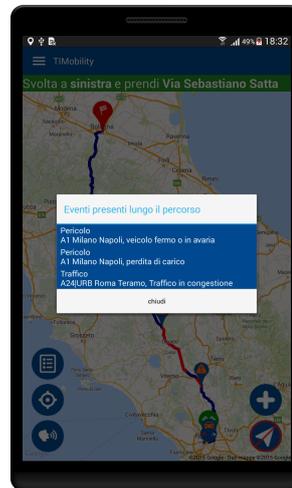


Fig. 4.20: Lista eventi attivi lungo il percorso

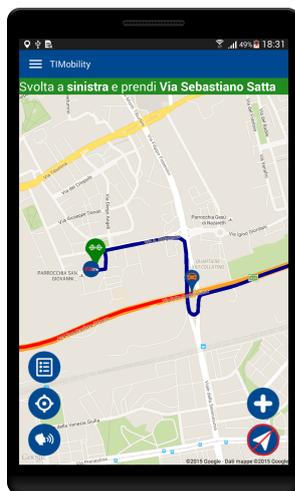


Fig. 4.21: Punto di partenza con suggerimento visualizzato

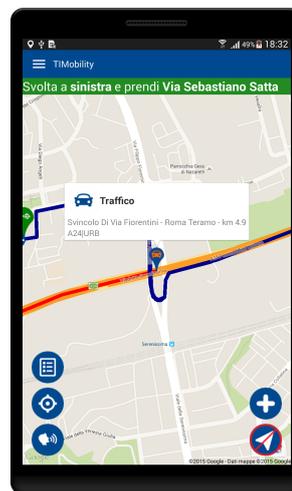


Fig. 4.22: Dettaglio infowindow di evento lungo il percorso

CAPITOLO 5

CONCLUSIONI

Il lavoro di tesi presentato pone il focus sull'evoluzione della piattaforma di Telecom Italia "ITS 2.0".

Innanzitutto si è studiato il contesto nel quale si muove il progetto di tesi, ovvero ITS (Intelligent Transport System): sistemi di trasporto intelligenti in grado di apportare benefici alla società in termini di efficienza, ecosostenibilità e sicurezza.

La piattaforma viene alimentata da informazioni circa la mobilità provenienti dal social network Twitter. In particolare analizza i tweet di enti ufficiali quali CCISS, ANAS, Luceverde, etc., li parsifica e li inserisce nel proprio database.

Dall'analisi si evince che la parsificazione non è sempre ottimale, provocando numerose anomalie sistematiche nel database della piattaforma.

Da qui è nata la prima attività svolta che mira alla correzione delle anomalie tramite meccanismi di Data Cleansing.

La fase di Data Cleansing ha avuto due attori principali: il database TMC e la Geocoding/Reverse Geocoding API di Google. Infatti tutti i record del database della piattaforma che trovavano corrispondenza con

il database TMC sono stati assunti validi e non da correggere, i restanti invece sono stati messi sotto esame dalle API di Google e quindi corretti.

Una volta resi corretti ed omogenei i dati presenti nel database, si è potuto procedere alla fase di Data Mining volta a estrapolare informazioni non note a priori.

Si è proceduto con il raggruppare tutti gli eventi di infomobilità avvenuti nello stesso luogo e caratterizzati dalla stessa categoria, dopodichè si sono aggregati tutti quegli eventi vicini temporalmente (distanti tra loro al più del lifetime specifico della categoria).

Adottando il suddetto algoritmo su una cospicua porzione di record, a valle del Data Cleansing, per l'esattezza 15808 record, si è giunti ad avere 8902 record aggregati, rilevando il 44% di record superflui che possono essere eliminati.

Infine, per rendere fruibili i dati di infomobilità processati dalla piattaforma "ITS 2.0", si è progettata e sviluppata un'applicazione Android che potesse fungere da navigatore e che permettesse agli utenti di visualizzare gli eventi, sia provenienti da fonti ufficiali che dalla community, di inviare/validare le segnalazioni e di individuare gli eventi lungo un percorso scelto. Tutto questo riponendo sempre una particolare attenzione alla User Interface e alla semplicità d'uso.

A valle del progetto di tesi, i benefici attesi da parte di Telecom Italia si possono riassumere nel seguente scenario:

- Telecom Italia offrirà un servizio agli utenti accurato che si distingue dai competitor per l'affidabilità delle informazioni;
- Gli utenti potranno usufruire del servizio tramite l'applicazione sul proprio smartphone;
- L'utilizzo dell'applicazione contribuirà ad arricchire la piattaforma tramite funzioni social;
- Il database arricchito, oltre ad offrire un servizio migliorato e competitivo, costituirà un **asset** per l'azienda.

BIBLIOGRAFIA

- [1] Donand Shoup, *Cruising for Parking*, ACCESS Magazine UCLA, 2007
- [2] Direttiva 2010/40/UE del Parlamento europeo e del Consiglio, del 7 luglio 2010, sul quadro generale per la diffusione dei sistemi di trasporto intelligenti nel settore del trasporto stradale e nelle interfacce con altri modi di trasporto,
Gazzetta ufficiale n. L207 del 06/08/2010, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32010L0040:IT:HTML>
- [3] Marco Annoni, *Intelligent e Smart Transport System*, Notiziario Tecnico Telecom Italia, 03/2013
- [4] Marco Annoni, Fabrizio Gatti, Luigi Grossi, *Muoversi in città: la sfida diventa smart?*, Notiziario Tecnico Telecom Italia, 02/2012
- [5] Marco Annoni, Umberto Trincherò, Eric Locci, Luca Rota, *Twitter e gli Open Data - Un aggregatore di eventi di viabilità*, 2015
- [6] *DATEX II V2.0 USER GUIDE*, European Commission - Directorate General for Transport and Energy, 2015
- [7] Enzo Contini, Anna Murru, Davide Frola, Rieks Joosten, Tobias Schlauch, *Development of MOBiNET Identity Manager*, 2014

- [8] Fabrizio Gatti, Anna Murru, Davide Frola, *WP4.5: CSE enablers documentation*, 2015
- [9] Roiger Richard, Geatz Michael, *Introduzione al Data Mining*, 2003
- [10] Erhard Rahm, Hong Hai Do, *Data Cleaning: Problems and Current Approaches*, University of Leipzig, Germany
- [11] *Documentazione ITS-2015*, CONSEL, 2015