

Modellazione di infrastrutture energetiche tramite ontologie per la pianificazione delle reti e la gestione degli asset

Enea Bionda, Francesca Soldan, Gabriele Paludetto, Alberto Maldarella, Roberto Naldi, Giulia Ronchetti, Ilaria Losa

2023

Progetto Digitalizzazione del sistema energetico integrato

Piano Triennale di Realizzazione 2022-2024 della Ricerca di Sistema Elettrico Nazionale



Accordo di programma 2022-2024 con il Ministero dell'Ambiente e della Sicurezza Energetica per le attività di ricerca e sviluppo di interesse generale per il sistema elettrico nazionale.

Progetto Digitalizzazione del sistema energetico integrato

Work Package WP 1 - Tecniche semantiche, di intelligenza artificiale e big data a supporto dei processi del sistema energetico integrato

Linea di Attività LA 1.02 - Piattaforme a supporto della transizione energetica - Ontologie per la pianificazione delle reti e la gestione degli asset

Codice identificativo RT-2.04-1.02-1

Titolo: Modellazione di infrastrutture energetiche tramite ontologie per la pianificazione delle reti e la gestione degli asset

Autori: Enea Bionda, Francesca Soldan, Gabriele Paludetto, Alberto Maldarella, Roberto Naldi, Giulia Ronchetti, Ilaria Losa

Verificatori: Giuseppe Mauri, Elisabetta Garofalo

Approvatori: Carlo Ismaele Tornelli, Claudio Luciano Bossi

Tipologia di documento: RAPPORTO

Data di emissione: 31/12/2023

Note: prima emissione

© Copyright 2023 by Ricerca sul Sistema Energetico-RSE S.p.A.

Contributo/i liberamente utilizzabile/i a condizione che venga chiaramente e visibilmente citata la società titolare.

Per la tutela dell'ambiente, prima di stampare questo documento pensa bene se è veramente necessario.



INDICE

INDICE	3
SOMMARIO.....	5
1 - INTRODUZIONE	6
2 - MODELLAZIONE DELLE INFRASTRUTTURE ENERGETICHE	8
2.1 Realizzazione di modelli CIM di reti reali e aggiornamenti all'ultima versione dello standard.....	8
2.1.1 <i>Descrizione delle sorgenti dati.....</i>	8
2.1.2 <i>Trasformazione dei dati di rete in un modello CIM.....</i>	9
2.1.3 <i>Aggiornamento IEC CIM 17 della MESP.....</i>	11
2.2 Sinergie CIM-BIM: esempio pratico per la gestione degli asset.....	13
2.2.1 <i>Modello IFC per un trasformatore di potenza AT/MT.....</i>	14
2.2.2 <i>Creazione di un trasformatore AT/MT in IEC CIM a partire dal modello IFC.....</i>	14
2.3 Definizione di un'ontologia per la modellazione delle reti gas	16
2.3.1 <i>Dati di input.....</i>	18
2.3.2 <i>Definizione dell'ontologia.....</i>	19
2.3.3 <i>Generazione del modello semantico della rete gas conforme all'ontologia definita</i>	22
2.4 Ontologie per architetture ICT	23
2.4.1 <i>IEC CIM per la descrizione di architetture ICT.....</i>	24
2.4.2 <i>Stato dell'arte di ontologie specifiche per il settore ICT.....</i>	27
3 - DIGITALIZZAZIONE DEI PROCESSI PER LA PIANIFICAZIONE MULTIENERGETICA DEL TERRITORIO ..29	
3.1 Progettazione dell'architettura informatica.....	29
3.1.1 <i>Individuazione dei requisiti.....</i>	30
3.1.2 <i>Architettura software per la pianificazione multienergetica del territorio.....</i>	31
3.1.3 <i>Trasformazione dell'osservatorio tecnologico in formato semantico.....</i>	34
3.1.4 <i>Generazione delle istanze.....</i>	36
3.2 Studio di metodologie per l'estrazione di impianti fotovoltaici da immagini satellitari	37
3.2.1 <i>Stato dell'arte sull'utilizzo di immagini da satellite per l'individuazione di impianti fotovoltaici.....</i>	38
3.2.2 <i>Proposta metodologica</i>	42
3.2.3 <i>Selezione del caso di studio.....</i>	45
4 - MESP: NUOVI SERVIZI A SUPPORTO DELLA PIANIFICAZIONE DI RETI ELETTRICHE.....	46
4.1.1 <i>Scalabilità e replicabilità.....</i>	46
4.1.2 <i>Processo di elaborazione per la SRA.....</i>	48
4.1.3 <i>Installazione nella piattaforma MESP.....</i>	53
5 - STIMA DEL POWER FLOW SU RETI ELETTRICHE TRAMITE TECNICHE DI GRAPH ANALYSIS E GRAPH MACHINE LEARNING.....	58
5.1 Selezione del <i>framework</i> tecnico di sviluppo.....	58
5.2 Traduzione dei dati di input in formato graphML	59
5.2.1 <i>Trasformazione CIM2GraphML.....</i>	59
5.2.2 <i>Trasformazione a TopologicalNode come entità di prima classe</i>	60
5.2.3 <i>Trasformazione a TopologicalNode come attributo.....</i>	61
5.3 Descrizione del compito di <i>machine learning</i>	62
5.3.1 <i>Metriche di successo e qualità.....</i>	62
5.4 Modelli basati su <i>Graph Convolutional Networks</i>	63
5.5 Modelli ensemble basati su alberi decisionali	66
5.6 Sintesi dei risultati sulla stima delle tensioni ai nodi	69
5.7 Preparazione dei dati per la stima di variabili di stato di reti gas	69



6 - PROCESSI DI MANUTENZIONE DELLE RETI.....	71
6.1 Analisi dei guasti sulle reti di distribuzione	71
6.1.1 Database degli ordini di lavoro su guasto.....	72
6.1.2 Fault Tree relativi agli ODL su guasto.....	72
6.1.3 Dati acquisiti dalle cabine secondarie in ottica manutenzione predittiva.....	73
7 - CONCLUSIONI	76
8 - BIBLIOGRAFIA.....	79
9 - ACRONIMI	85



SOMMARIO

La **digitalizzazione dei processi** di gestione ed esercizio è un fattore fondamentale per la realizzazione di un sistema energetico integrato flessibile, efficiente e resiliente. La digitalizzazione necessita di standard che favoriscano l'**interoperabilità**, sia tra differenti vettori energetici, sia tra differenti attori. Per queste ragioni il rapporto affronta tematiche di digitalizzazione dei processi, con attenzione all'**interoperabilità** a livello informativo. L'interoperabilità è facilitata da **ontologie** in grado di descrivere il dominio considerato. Il rapporto descrive un esempio concreto di utilizzo dello **standard IEC CIM**, con aggiornamento alla versione 17. Per abilitare lo sviluppo di **digital twin** del sistema energetico, viene introdotta la sinergia fra gli standard **BIM-CIM**. Considerando il caso d'uso di un trasformatore in cabina primaria vengono dimostrati i vantaggi di un connubio tra la progettazione in ambito civile ed il *planning&operation* elettrico per la gestione dell'asset durante tutto il suo ciclo di vita. Nell'integrazione dei differenti vettori è stata definita un'**ontologia per la rete gas**. Con la digitalizzazione del sistema energetico assume un ruolo rilevante il *layer* ICT, che rappresenta un punto critico per la resilienza del sistema stesso, sia a livello di *cybersecurity*, sia a livello di interruzione dei servizi ICT stessi. Per tale ragione è stato esaminato lo stato dell'arte specifiche per comprendere come modellare l'**architettura ICT tipica di un DSO**. La digitalizzazione ha un ruolo importante anche nel migliorare i processi di pianificazione e gestione delle singole aziende/*utility*. Per questo motivo è stata avviata una revisione di processi di **pianificazione multienergetica del territorio**, utilizzando soluzioni innovative in grado di semplificare la gestione dei **dati provenienti da fonti eterogenee** e di innovare la **mappatura di impianti fotovoltaici a terra** partendo da **immagini satellitari**. Per ampliare le potenzialità della piattaforma **MESP** con, sono stati aggiunti algoritmi per la pianificazione delle reti e servizi per eseguire **analisi di scalabilità e replicabilità** di soluzioni di flessibilità delle reti elettriche. Sono stati sviluppati due diversi **modelli di machine learning basati su grafi** per la **stima delle tensioni** ai nodi di una rete elettrica utilizzando tecniche *data driven*: le **Graph Convolutional Network** e **LightGBM**. Nell'ambito della digitalizzazione dei processi di manutenzione delle reti, sono state studiate alcune procedure relative ai **processi di analisi di guasto e manutenzione** con attenzione agli *asset* delle cabine secondarie.

Keywords: IEC CIM, BIM, Multi-energy, Pianificazione territoriale, Graph Machine Learning, Asset Management



1 - INTRODUZIONE

La ricerca documentata in questo rapporto fa parte della seconda annualità di un percorso triennale per sviluppare piattaforme e strumenti informatici che supportino la pianificazione e gestione delle infrastrutture delle reti energetiche, considerando anche il contesto territoriale. La creazione di un sistema energetico integrato richiede l'uso di informazioni da diversi vettori energetici e sistemi informativi, rendendo necessaria la connessione e la creazione di un sistema interoperabile basato su modelli condivisi e algoritmi innovativi.

La **digitalizzazione del settore elettro-energetico** sta rapidamente cambiando la nostra modalità di interazione con l'energia e sta portando un importante contributo verso la decarbonizzazione del sistema. Infatti, le tecnologie digitali favoriscono l'ottimizzazione del sistema, una migliore efficienza, la riduzione dei costi e dei guasti. In questo contesto, la Commissione Europea, nell'**Action Plan per la digitalizzazione del sistema energetico** [1], afferma che la digitalizzazione deve essere considerata come parte integrante della transizione energetica. Inoltre, nella comunicazione Europea "**An EU strategy for Energy System Integration**", la digitalizzazione è indicata come possibile modalità per favorire l'integrazione di diversi vettori energetici [2].

Nel capitolo 2 -, nell'ambito più generale della digitalizzazione di sistemi *multi-energy*, il rapporto si focalizza sulla gestione di dati e informazioni relative a diverse fonti energetiche tramite l'utilizzo di **ontologie**. L'evoluzione verso sistemi *multi-energy* richiede scambi informativi complessi tra diversi sistemi e attori, e risulta sempre più importante fare uso di modelli e profili di scambio standardizzati e condivisi. Le ontologie, con le loro specifiche formali dei concetti relativi ad un dominio, offrono una comprensione comune per una specifica area di interesse e quindi risultano cruciali per risolvere il problema dell'**interoperabilità** tra sistemi in domini complessi. L'evoluzione di tali modelli, descritti utilizzando tecnologie del web semantico e delle ontologie, e la loro standardizzazione, comporta un continuo aggiornamento sia a livello della descrizione delle reti energetiche rappresentate sia degli strumenti di visualizzazione e analisi delle reti. Nel settore elettro-energetico, IEC CIM rappresenta lo standard informativo per le entità del mondo elettrico. Per supportare le versioni più recenti dello standard, esaminate in un lavoro precedente [3], si è proceduto ad aggiornare la piattaforma **Multi Energy Semantic Platform** (MESP) implementando gli aggiornamenti degli strumenti *software* sviluppati e adeguando le descrizioni di reti presenti nella piattaforma. Inoltre, sfruttando dati di reti esistenti forniti da IREN sono state generate nuove descrizioni di reti nella versione 17 di IEC CIM che serviranno come base informativa per condurre analisi in ottica di manutenzione predittiva.

Per informazioni non nativamente legate al mondo elettro-energetico è possibile utilizzare modelli specifici di settore, ed estendere l'ontologia IEC CIM. Ad esempio, per il settore delle costruzioni, la **metodologia del BIM** e lo **standard IFC** sono ampiamente diffusi per la progettazione, manutenzione e dismissione degli edifici e dei relativi impianti gestendo il loro intero ciclo di vita. Nella precedente annualità di ricerca [4] è stato descritto lo stato dell'arte dell'utilizzo del BIM nel settore elettro-energetico e la possibilità di estendere le classi di IEC CIM per la descrizione della rete elettrica con quelle di IFC, per una gestione ottimale della pianificazione, manutenzione e dismissione degli asset di rete. La continuazione dell'attività, svolta in collaborazione con Enel Grids, ha riguardato un primo esempio di trasformazione delle classi che descrivono un trasformatore in cabina primaria da IFC a IEC CIM.

Un altro esempio di utilizzo delle ontologie e delle tecnologie del web semantico ha riguardato la creazione di un'ontologia derivante da IEC CIM per la descrizione della **rete gas**. I vantaggi di questa modalità, ovvero di modellizzare reti energetiche utilizzando una struttura dati simile, sono principalmente legati alla possibile gestione dei dati del nuovo vettore energetico in modo analogo a quelli della rete elettrica e alla gestione combinata e ottimale delle reti relative a più vettori energetici. In particolare, in collaborazione con Unareti, durante l'attività è stata

*Con **interoperabilità semantica** si intende l'abilità di due o più sistemi o componenti di scambiare dati e informazioni; essa viene raggiunta quando i sistemi interagenti attribuiscono lo stesso significato ai dati scambiati, il che assicura la consistenza dei dati attraverso diversi sistemi [125]. Questa interoperabilità può essere raggiunta solo tramite l'utilizzo di standard o accordi sul formato e sul significato dei dati. Attualmente diversi settori si stanno attivando con lo sviluppo di ontologie, in modo da fornire soluzioni semantiche ai problemi di interoperabilità della conoscenza e tra i dati raccolti. Anche nel settore energetico, le ontologie sono state utilizzate per supportare scambi di informazioni tra domini multipli e sono alla base di soluzioni di intelligenza artificiale e di Digital Twin.*

considerata la trasformazione delle informazioni disponibili di porzioni di reti gas esistenti in una rappresentazione secondo l'ontologia definita.

Inoltre, per studiare problemi che considerano aspetti topologici della rete assieme ad aspetti legati all'**architettura ICT** per il monitoraggio e controllo della rete, risulta utile integrare queste informazioni tramite le tecniche del web semantico. In particolare, le attività hanno valutato la possibilità di utilizzare le classi di IEC CIM per la descrizione di architetture ICT distribuite. È stato anche effettuato uno studio dello stato dell'arte sulle ontologie per la descrizione del settore ICT, che saranno da preferire nel caso in cui le classi di IEC CIM risultano troppo generali per la risoluzione di problemi specifici oppure per affrontare problemi legati alla cybersicurezza di rete. Nell'ambito della **digitalizzazione dei processi per la pianificazione multienergetica del territorio**, analizzati in precedenza [4], la gestione di flussi dati provenienti da sorgenti dati eterogenee e l'interoperabilità dei modelli costituisce una sfida per l'esecuzione delle analisi multi-obiettivo. In questo contesto, nei paragrafi 3.1 e 3.2 sono state identificate le soluzioni più flessibili e adeguate alla gestione di grandi quantità di dati eterogenei per il recepimento, l'elaborazione, l'organizzazione dei dati e l'interoperabilità dei modelli. L'analisi effettuata ha permesso la **progettazione di un'architettura informatica** dei processi ETL (*Extract/Transform/Load*) in grado di facilitare la gestione e la tracciabilità dei flussi dati. Come inizio dell'implementazione di tale architettura è stata effettuata la conversione del *database* relazionale esistente dei parametri tecnologici, utilizzati per alimentare il modello di pianificazione energetica in un **database semantico**. Si è inoltre sviluppata una metodologia per l'aggiornamento periodico e automatico della **mappatura di impianti fotovoltaici a terra** a partire da **immagini satellitari**, come elemento fondamentale per la valutazione del potenziale territoriale disponibile per nuove installazioni fotovoltaiche.

L'evoluzione delle reti energetiche richiede analisi di pianificazione che possano identificare nuove potenzialità ed eventuali criticità delle soluzioni proposte. In questo contesto, le tecniche di analisi di **scalabilità e replicabilità** (SRA) di soluzioni di flessibilità, descritte nel capitolo 4 -, possono essere di ausilio ad un operatore che voglia valutare scenari di espansione della propria rete. In questa annualità sono state integrate nella piattaforma MESP alcune nuove funzionalità necessarie per condurre questo tipo di analisi.

L'impiego di ontologie in grado di rappresentare la topologia delle reti energetiche abilita l'utilizzo di tecniche di elaborazione e di intelligenza artificiale specifiche per le strutture a grafo. Nel capitolo 5 - si descrive l'attività di sviluppo di un insieme di modelli di *machine learning* per la stima delle tensioni ai nodi di una rete elettrica, normalmente ottenute mediante un calcolo di *power flow*, senza l'ausilio di modelli analitici. In particolare, a partire da una rappresentazione a grafo delle reti elettriche, sono stati addestrati modelli di due differenti tipologie: modelli basati su **Graph Convolutional Networks** e modelli basati su **ensemble di alberi decisionali**. Le differenti tecniche sono state valutate e confrontate in termini di accuratezza e tempo di esecuzione. Lo scopo finale dell'attività triennale, in collaborazione con il Politecnico di Milano, è la valutazione delle prestazioni raggiungibili con tali tecniche per la stima dello stato di reti energetiche in casi non facilmente risolvibili tramite tecniche analitiche, ad esempio nel caso di reti estese e multivettore.

Infine, prendendo in considerazione i **processi di manutenzione delle reti elettriche**, nel capitolo 6 viene descritta la prosecuzione delle attività di analisi della gestione dei dati relativi ai guasti avviata in [4], con lo scopo di migliorare l'affidabilità e la resilienza delle reti di distribuzione.

L'attività ha riguardato uno studio sulle procedure attualmente seguite da Unareti in merito agli ordini di lavoro per gestire eventi di guasto ed interventi di manutenzione focalizzandosi, in particolare, sul caso degli **asset** delle **cabine secondarie**. Questa scelta è motivata da molteplici fattori: in primo luogo, l'entità e l'impatto che un guasto in cabina può avere sugli utenti se non gestito tempestivamente, in secondo luogo, la presenza di dati acquisiti nelle cabine secondarie telecontrollate, che ancora non vengono pienamente sfruttati ma che possono essere molto utili per lo sviluppo di algoritmi di **manutenzione predittiva**.

2 - MODELLAZIONE DELLE INFRASTRUTTURE ENERGETICHE

2.1 Realizzazione di modelli CIM di reti reali e aggiornamenti all'ultima versione dello standard

Nell'annualità precedente è stata presentata una panoramica del modello IEC CIM, la sua evoluzione nel tempo, gli attori che lo mantengono e le interazioni con gli enti di standardizzazione [4]. L'adozione dello standard IEC CIM da parte di un numero sempre più grande di *utility* ha causato una serie di aggiornamenti dello standard che riflettano i cambiamenti in atto nel settore energetico. Attualmente la versione più recente adottata nello standard IEC è la versione 17. Nelle attività di RSE, la versione di riferimento finora utilizzata era la versione 15; pertanto, i *software* sviluppati e le descrizioni di rete generate non erano conformi alla versione stabile più recente del modello IEC CIM. Perciò è stato eseguito l'aggiornamento alla versione 17 sia della descrizione delle reti utilizzate, sia dei *software* di visualizzazione e di analisi delle reti della piattaforma MESP (*Multi-Energy Semantic Platform*). L'attività è descritta nel paragrafo 2.1.3. Grazie alla collaborazione attiva con operatori di rete sono state inoltre sviluppate descrizioni CIM di nuove reti di distribuzione. Nei paragrafi 2.1.1 e 2.1.2 viene descritta l'attività di generazione del modello IEC CIM di reti di distribuzione esistenti come base informativa per l'adozione di tecniche semantiche per la gestione degli *asset*. Questa attività ha permesso di validare l'integrazione del modello IEC CIM 17 nella piattaforma MESP. In particolare, viene illustrata la procedura adottata per la trasformazione dei dati eterogenei di descrizione delle reti in un modello standard in formato IEC CIM 17.

Si ricorda che tale *standard* può essere utilizzato in differenti contesti come l'interoperabilità tra sistemi, lo scambio dei dati standardizzato e la compatibilità con la rappresentazione semantica della rete (*knowledge graph*). La rappresentazione dei dati in formato IEC CIM e la sua gestione attraverso l'utilizzo dei *knowledge graph* abilita l'utilizzo di algoritmi di intelligenza artificiale per la gestione delle reti, con particolare attenzione alla manutenzione predittiva: si possono incrociare le rappresentazioni IEC CIM topologiche e/o topografiche ottenute con dati ambientali e di esercizio per lo studio delle situazioni critiche, in concomitanza con l'attività di monitoraggio guasti e disservizi e di diagnostica delle reti.

Nei paragrafi successivi vengono spiegati in maniera dettagliata: il contesto applicativo, l'algoritmo di trasformazione dei dati, le criticità, le soluzioni adottate e gli aggiornamenti applicati alla piattaforma MESP per supportare i dati nel nuovo *standard*.

2.1.1 Descrizione delle sorgenti dati

Il caso d'uso considerato è relativo a reti in media tensione, ma l'algoritmo è facilmente adattabile e scalabile anche a reti in alta e in bassa tensione.

Per le proprie reti IREN utilizza diversi sistemi di gestione dei dati: GIS (*Geographic Information System*), SCADA (*Supervisory Control And Data Acquisition*) e sistemi gestionali (SAP). I dati GIS, in formato GeoJson¹, descrivono le parti che compongono le reti dal punto di vista geografico (Figura 2.1), raggruppando le differenti tipologie di componenti in diversi *layer*. Oltre ai dati geografici, sono presenti dettagli dei componenti, come ad esempio, la tipologia di cavo per i conduttori o la tensione delle sbarre in cabina. Sono presenti, inoltre, altri *layer* che non rappresentano componenti fisici ma che descrivono la connessione tra componenti presenti in cabina permettendo di ricostruire uno schema di collegamento interno.

¹ Formato utilizzato per archiviare una collezione di geometrie spaziali.



Figura 2.1 – Visualizzazione dei dati GIS in formato GeoJson che descrivono la rete di distribuzione esaminata. Tra i layer disponibili sono rappresentati in blu i segmenti di cavo mentre in rosso le cabine secondarie.

Nei file provenienti dallo SCADA, in formato XML², troviamo sia informazioni di diagramma, sia informazioni sullo stato dei componenti di rete, ad esempio lo stato degli interruttori; sono, inoltre, incluse nella descrizione alcune grandezze elettriche, come la reattanza, l'impedenza e la suscettanza dei cavi o tensioni nominali in ingresso e in uscita dei trasformatori, che sono utili per analisi del *power flow*.

Per ultimo, nei file del sistema gestionale SAP troviamo dati necessari per coordinare operazioni di manutenzione sugli asset della rete. Tra le varie informazioni si trovano descrizioni di impianti o di loro parti, di linee, di macchinari oppure posizioni di componenti soggetti a manutenzione.

Da una prima analisi effettuata è emersa una sostanziale separazione tra le tre sorgenti dati descritte: in particolare, gli identificativi relativi ad uno stesso componente sono diversi nelle tre sorgenti dati, a riprova dell'esistenza di cosiddetti *silos* informativi. Questo costituisce una complicazione se si vuole descrivere una rete nel formato standard IEC CIM in quanto richiede di recuperare le informazioni relative ad uno stesso asset da sorgenti dati differenti. Per poter effettuare questa operazione è necessario poter collegare le informazioni delle diverse fonti. Nel caso in esame, nella base dati SAP sono presenti le associazioni tra gli identificativi usati, per esempio, nei file GeoJson e nei file XML, ma solo in maniera parziale.

Nella realizzazione del modello CIM per le reti rese disponibili si è deciso di procedere per gradi, partendo dai file GeoJson, in quanto consentono effettuare una prima trasformazione dei dati che descrivono gli aspetti geografici ed eventualmente topologici delle reti.

La validazione finale del modello CIM sarà il calcolo del *power flow* della rete trasformata e il confronto con i risultati forniti dall'operatore. Nella sezione successiva viene descritto come i dati forniti sono stati utilizzati per la trasformazione della rete in IEC CIM 17 e i passaggi di elaborazione e validazione, concentrandoci, nell'ordine, sull'aspetto geografico e topologico.

2.1.2 Trasformazione dei dati di rete in un modello CIM

Come indicato nella sezione precedente, la sorgente dati con le informazioni di base per la descrizione della rete è costituita dai layer del file GeoJson in quanto contengono informazioni sui

² Extensible Markup Language, è il formato file standard per le ontologie.

componenti di rete e la loro collocazione geografica. Questi dati consentono di effettuare un primo passo di modellazione CIM e di validare la correttezza della trasformazione. Grazie a questi *layer* è stato semplice effettuare il primo *mapping* fra i componenti delle reti descritte nel file GeoJson e le corrispondenti classi IEC CIM. La Tabella 2.1 mostra l'associazione alle classi IEC CIM per ogni tipologia di componente definita nel file.

Tabella 2.1 - Mapping eseguito dei layer GeoJson nelle classi IEC CIM relative.

Elementi GeoJson	Classe IEC CIM
CONDUTTORI	ACLineSegment
SBARRE	BusbarSection
CABINE MT	Substation
CIRCUITI	Line
TRASFORMATORI	PowerTransformer
RACCORDI	Junction
NODI	Breaker
SEZIONATORI	Disconnecter

A ciascun componente di rete sono stati poi assegnati uno o più punti geografici (classi IEC CIM *PositionPoint*, *Location*) a seconda del tipo di dato presente nel *layer* GeoJson. Per validare la corretta trasformazione delle coordinate è stata prodotta nella piattaforma MESP una visualizzazione geografica della rete trasformata ed è stata verificata a campione la collocazione dei punti nella mappa geografica rispetto a quella originale visualizzata utilizzando l'applicazione QGIS. Per quanto riguarda l'aspetto topologico, le sorgenti dati non permettono una facile identificazione esplicita dei punti di connessione tra le parti. Tuttavia, le coordinate geografiche presenti nel GeoJson sono, in generale, molto precise e quindi adatte a dedurre le connessioni topologiche tra i diversi componenti della rete (Figura 2.2). Sfruttando le sovrapposizioni delle coordinate, in linea di principio è possibile determinare i punti di connessione associando le rispettive classi IEC CIM *ConnectivityNode* e *Terminal*.

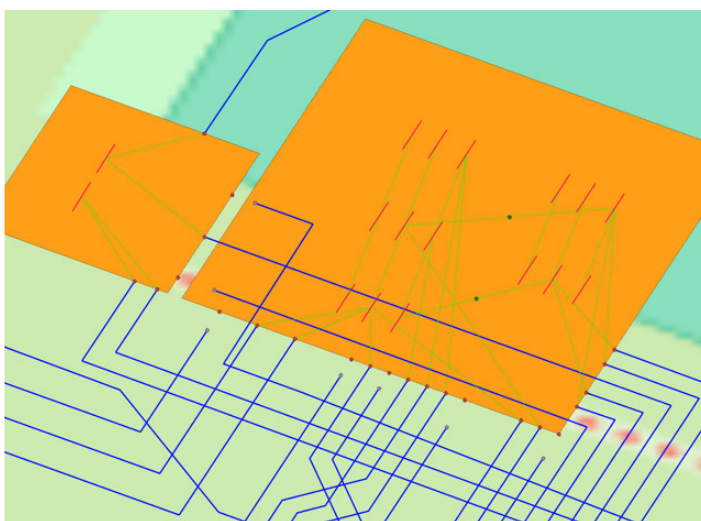


Figura 2.2 - L'immagine rappresenta una cabina in media tensione contenente un trasformatore media-bassa con tutti gli interruttori (punti rossi) e i cavi (blu) collegati alle sbarre (segmenti rossi). Sono presenti casi di sovrapposizione geografica tra gli interruttori. In verde, invece, è rappresentato un *layer* GIS che identifica le connessioni tra interruttori, sbarre e trasformatori.

I *ConnectivityNode* sono fondamentali per andare a ricavare i *TopologicalNode*, che raggruppano *ConnectivityNode* collegati tra loro tramite qualsiasi tipo di interruttore chiuso; l'algoritmo crea quindi i *ConnectivityNode*, sfruttando le coordinate geografiche dei componenti ed usandole come identificatore ID.

L'attività di modellazione si è quindi concentrata sullo sviluppo di algoritmi in grado di risolvere in modo automatico il problema di connettività tra i vari componenti. Una volta trovate le corrispondenze tra *layer* e classi IEC CIM, l'algoritmo crea un file XML che rappresenta la rete nel formato IEC CIM 17. Nella Tabella 2.2 vengono riassunti i risultati della trasformazione automatica indicando i componenti per i quali è stato possibile trovare le connessioni.

Tabella 2.2 - La tabella mostra le percentuali di componenti nel GeoJson per cui è stata trovata almeno una connessione con almeno un'altra componente utilizzando l'algoritmo descritto.

Componenti	Quantità	Numero elementi connessi	%
Conduttori	811	807	99,88
Raccordi (giunzioni)	685	531	77,51
Nodi Linea (interruttori)	153	146	95,42
Nodi Origine (interruttori)	4	4	100,00
Sezionatori	14	14	100,00
Sbarre MV	59	49	83,05
Trasformatori MT-MT	1	1	100,00
Trasformatori MT-BT	49	38	77,55

L'approccio seguito, basato sulle connessioni geografiche, permette di trovare buona parte delle connessioni tra componenti. Per verificare la correttezza della topologia di rete ricavata in modo automatico dai dati disponibili è stata eseguita una analisi topologica del modello di rete CIM ottenuto utilizzando il *software* di modellazione delle reti *Cimphony*. Nella porzione di rete considerata sono state identificate 47 isole topologiche a fronte di quattro feeder, indice del fatto che seppur le connessioni sono state trovate non sono ancora corrette. Non avendo ancora a disposizione una rappresentazione topologica fornita dal DSO, la risoluzione delle criticità riscontrate sarà oggetto di approfondimento nella prosecuzione delle attività nell'ambito della collaborazione con l'operatore di rete.

In conclusione, a partire dai dati utilizzati dai sistemi di un operatore di rete è stato possibile realizzare la trasformazione automatica dei dati in un modello CIM validato per quanto riguarda la rappresentazione geografica delle reti; tuttavia, non sono state fornite informazioni dirette sulla topologia delle reti e non è stato ancora possibile dedurla completamente in modo automatico dagli altri dati disponibili. Nella prosecuzione delle attività si punterà a validare il modello di rete CIM anche sotto l'aspetto topologico che abiliterà, tra le altre cose, il calcolo del *power flow*.

2.1.3 Aggiornamento IEC CIM 17 della MESP

Come segnalato nella precedente annualità [4], un passo importante per l'aggiornamento dei prodotti IEC CIM sviluppati da RSE è l'adeguamento per l'utilizzo del IEC CIM 17 nella piattaforma MESP. In questa piattaforma le reti elettriche vengono descritte e gestite come *knowledge graph*, in cui le entità sono descritte secondo una ontologia *standard* IEC CIM o ricavata dallo *standard*. Per aggiornare la piattaforma alla versione 17 di IEC CIM è stato necessario effettuare alcune modifiche a differenti microservizi presenti nella piattaforma.

Il passaggio alla versione IEC CIM 17 comporta l'adozione di una nuova versione dell'ontologia, che utilizza un *namespace* differente. Questo cambiamento impatta principalmente il

microservizio *Network Model Manager* (Figura 2.3) che ha come funzione l'interrogazione del *knowledge graph* della piattaforma MESP. La nuova versione dell'ontologia CIM, descritta in formato OWL (*Web Ontology Language*), è stata inserita nel *Network Model Manager* in modo da permettere al *reasoner* di fare inferenza sui dati memorizzati nel *knowledge graph* anche utilizzando dati di rete descritti secondo l'ultima versione dello standard.

Il cambiamento del modello impatta sulla struttura delle *query* SPARQL che interrogano il *knowledge graph*, sia per la definizione dei prefissi che per il corpo della *query*, poiché si utilizza un nuovo *namespace* per identificare le nuove classi.

Per quanto riguarda le classi che, nel cambiamento di versione, non hanno subito modifiche in termini di attributi e relazioni, come nel caso di classi ormai stabili quali Terminal e ConnectivityNode, l'unica variazione è nel *namespace* che identifica la classe. Pertanto, il cambio di *namespace* non influisce sul corpo delle *query* ma solo sulla parte dichiarativa dei prefissi. Nell'aggiornamento del *Network Model Manager*, per le *query* che utilizzano classi e relazioni rimaste invariate, è stata quindi operata una parametrizzazione del *namespace* che permette di far riferimento alla versione desiderata. In questo modo, è possibile gestire interrogazioni compatibili sia per le reti descritte con la versione 17, sia con versioni precedenti.

In molti casi, viste le informazioni trattate nelle reti a nostra disposizione, la parametrizzazione del *namespace* permette di gestire la maggior parte delle *query* utilizzate nella piattaforma MESP. Nel caso di cambi strutturali nel modello semantico al variare della versione, come aggiunta o deprecazione di classi, nel *Network Model Manager* è stato invece aggiornato anche il corpo delle *query*.

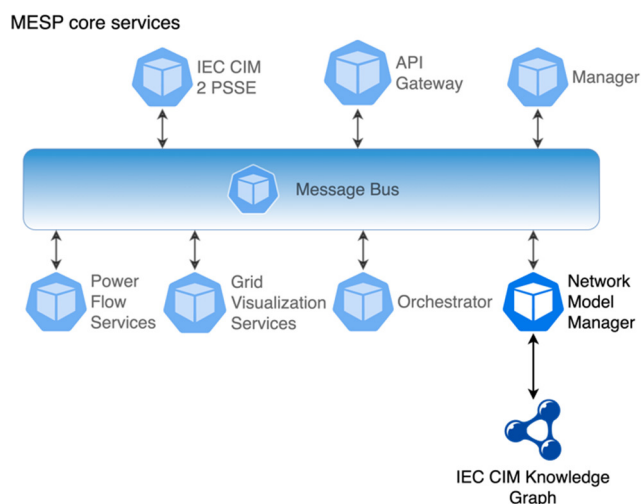


Figura 2.3 – Il microservizio Network Model Manager gestisce nella MESP le interrogazioni al *knowledge graph* richieste dagli altri microservizi tramite il Message Bus.

L'aggiornamento del *Network Model Manager* permette quindi di interrogare dati presenti nel *knowledge-graph* IEC CIM nella versione 17. Come esempio, la Figura 2.4 mostra la rappresentazione geografica ottenuta con la piattaforma MESP della rete reale descritta nel paragrafo precedente, a valle della sua trasformazione in un modello CIM in versione IEC CIM 17. Nella prosecuzione delle attività si approfondirà la modellazione di rete per supportare i casi d'uso legati alla manutenzione preventiva, valutando la possibilità di integrare informazioni provenienti anche da altre sorgenti dati, finora non utilizzate.

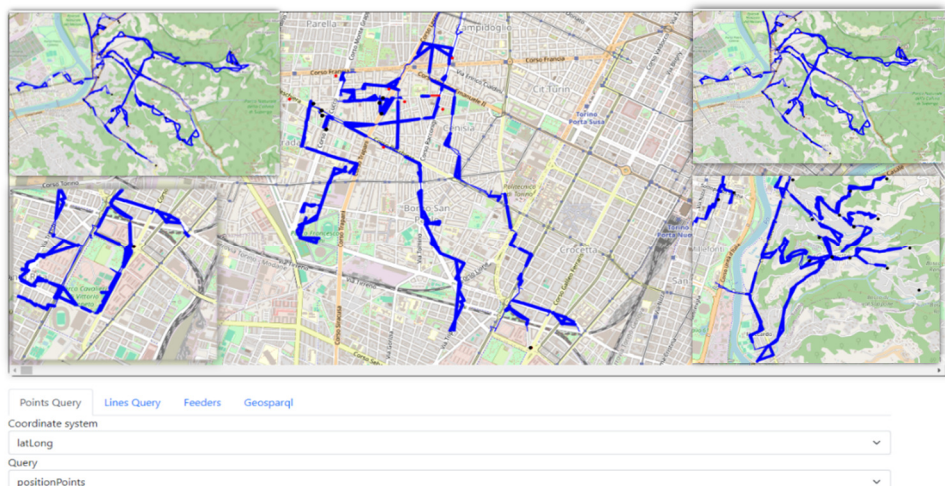


Figura 2.4 - Esempi di schermate che rappresentano porzioni della rete di Torino.

2.2 Sinergie CIM-BIM: esempio pratico per la gestione degli asset

Nelle precedenti attività [4] sono stati introdotti la **metodologia BIM** (*Building Information Modeling*) e lo **standard IFC** (*Industry Foundation Classes*) per la gestione degli asset durante il loro intero ciclo di vita. In particolare, è stata studiata la possibilità di integrare questi strumenti di tipo generale con i modelli dati specifici del settore elettro-energetico, come IEC CIM. La metodologia BIM e lo *standard IFC* sono utilizzati nel settore delle costruzioni per la gestione degli asset, ma non sono appositamente concepiti per descrivere componenti elettrici. D'altra parte, IEC CIM include solo un numero limitato di classi specifiche utili per la pianificazione, la manutenzione e la gestione degli asset di rete dalla loro installazione fino alla loro dismissione. I due modelli nascono infatti con uno scopo ben distinto: **IFC** viene tipicamente utilizzato per affrontare la **progettazione di infrastrutture** ma non è usato per scopi operativi, mentre **IEC CIM** è usato in **fase di esercizio** ma non è generalmente usato per la fase di progettazione. Combinare i due modelli rappresenta quindi la soluzione ottimale per garantire una gestione completa degli asset nel loro intero ciclo di vita.

In questa annualità del progetto, l'attività ha riguardato lo sviluppo, in collaborazione con Enel Grids, di un esempio pratico di creazione di istanze IEC CIM a partire da un modello IFC. Come caso studio è stato considerato quello di un **trasformatore AT/MT in cabina primaria**.

In primo luogo, è stato studiato il modello IFC di *input*, in modo da individuare le classi di interesse anche dal punto di vista operativo. Successivamente è stata implementata una trasformazione da tali classi IFC in istanze IEC CIM, che vanno ad alimentare il *knowledge graph* della rete. Il collegamento delle classi IFC e IEC CIM avviene tramite l'assegnazione di un codice identificativo univoco che rende possibile mantenere **allineati** i due mondi della **progettazione** (BIM) e del **planning&operation** (IEC CIM). Ciascun componente di rete è infatti identificato, secondo il modello IFC, da uno specifico codice univoco; nella fase di trasformazione nelle classi del modello IEC CIM il codice univoco viene mantenuto e utilizzato come ID della relativa classe, in modo da mantenere il collegamento tra i due mondi IFC e IEC CIM.

L'idea è di permettere che, per scopi di progettazione e gestione degli asset nel loro ciclo di vita, sia possibile continuare ad utilizzare strumenti specifici del mondo BIM. Le informazioni inserite in fase di pianificazione dal costruttore permettono di alimentare direttamente il *knowledge graph* secondo lo standard IEC CIM, che viene utilizzato a fini operativi sulla rete. Il *knowledge graph* così generato viene continuamente sincronizzato e aggiornato con le nuove informazioni, provenienti sia dagli applicativi BIM sia dai sistemi di *planning/operation*: questo rende disponibile un'unica



fonte di dati, completa da tutti i punti di vista, che permette quindi di gestire gli asset in modo ottimale nel corso del loro ciclo di vita.

2.2.1 Modello IFC per un trasformatore di potenza AT/MT

Come caso studio, è stato considerato il modello IFC che descrive i componenti di una cabina primaria Enel Grids. Le classi utilizzate sono quelle del modello sviluppato da Enel Grids per la descrizione di componenti elettrici in IFC, presentate nel rapporto dello scorso anno [4]. Il livello di dettaglio geometrico scelto per la rappresentazione dei componenti (*Level of Geometry* - LoG) è pari a 300. Con questo valore la geometria dei diversi elementi è descritta, ma non sono definiti nello specifico tutti i sottocomponenti geometrici. In Figura 2.5 viene mostrato un esempio del significato di diversi livelli di dettaglio geometrico (200-300-350) nel caso di un trasformatore.

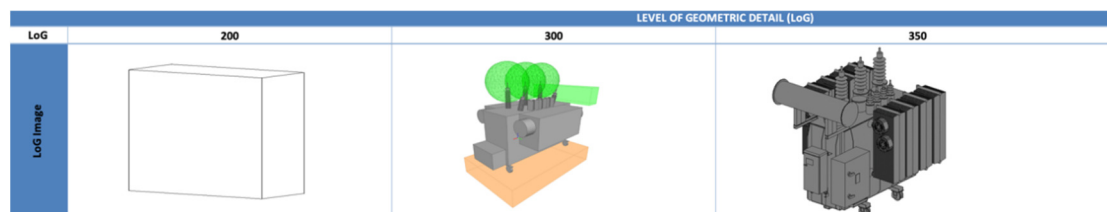


Figura 2.5 - Trasformatore descritto in IFC secondo diversi livelli di dettaglio geometrico (LoG).

Per la visualizzazione delle classi IFC è stato utilizzato un *software* gratuito *open source* **Blender** [5] per la gestione di oggetti 3D. Il vantaggio di questo strumento consiste nella possibilità di creare, visualizzare e modificare oggetti IFC tramite l'Add-on **BlenderBIM**.

Durante l'attività ci si è soffermati sulla descrizione della componente *IfcTransformer*, di cui è mostrato un dettaglio 3D nella seguente Figura 2.6, insieme agli attributi e proprietà IFC che descrivono il trasformatore.

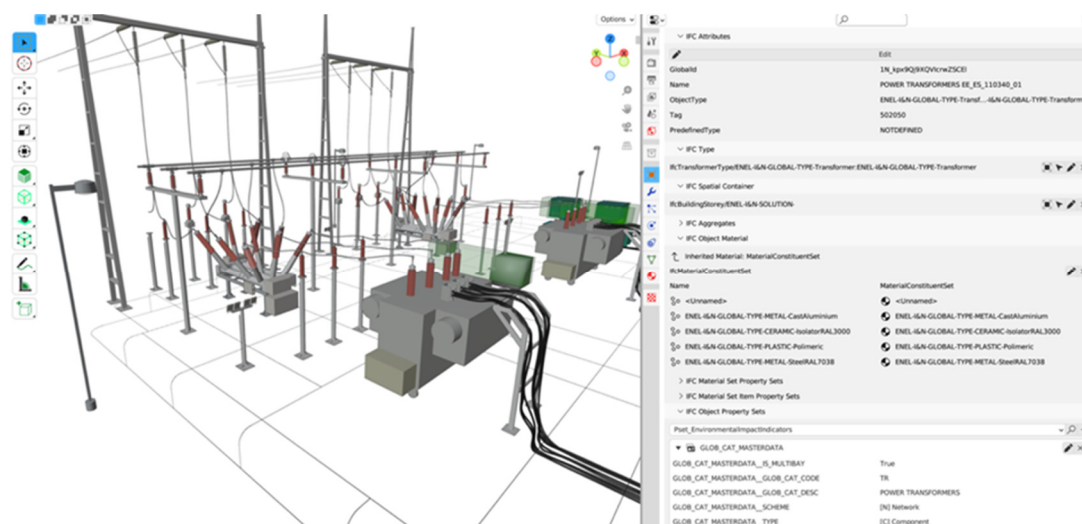


Figura 2.6 - Selezione di un elemento *IfcTransformer* tramite il software Blender e l'Add-on BlenderBIM.

2.2.2 Creazione di un trasformatore AT/MT in IEC CIM a partire dal modello IFC

L'obiettivo di questa prima attività, già introdotto al paragrafo 2.2, è avere a disposizione un unico *knowledge graph* che viene continuamente aggiornato e sincronizzato con le informazioni

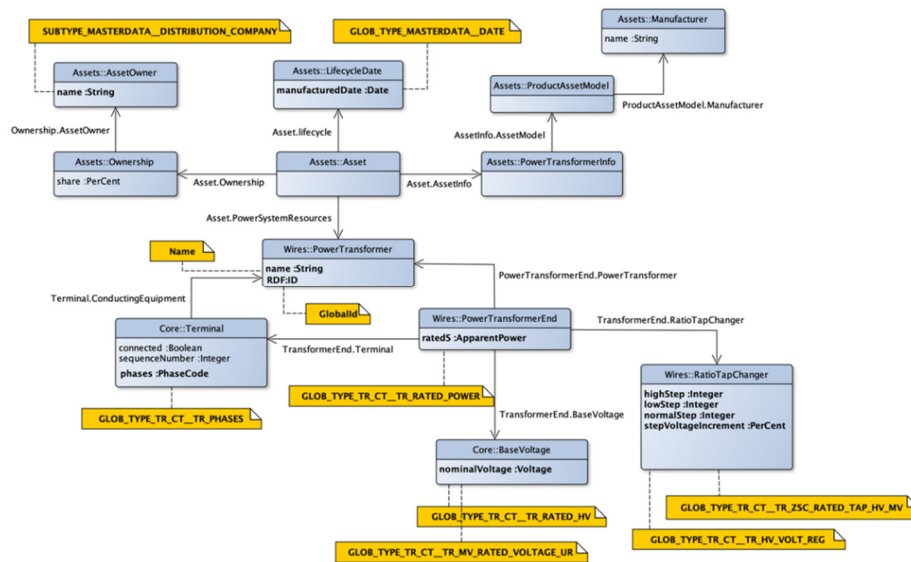
provenienti da entrambi i settori della progettazione e del *planning&operation* delle infrastrutture di rete elettrica.

La prima fase di progettazione di nuove infrastrutture elettriche tramite la metodologia BIM prevede la creazione di una loro descrizione mediante classi IFC. Successivamente, per passare alla gestione operativa delle infrastrutture, è necessario disporre dei componenti di rete descritti in formato IEC CIM. Tali descrizioni possono essere direttamente ottenute a partire dalle informazioni disponibili nello standard IFC; tuttavia, gli attributi e le proprietà IFC sono molto dettagliati e, in alcuni casi, significativi solo nel mondo BIM. Per questo motivo sono mappate da IFC a IEC CIM soltanto le informazioni necessarie per l'ambito operativo e viene mantenuto l'**allineamento con la rappresentazione BIM** tramite l'adozione di un **identificativo univoco** per lo stesso componente nei due standard di modellazione. Questo garantisce di avere a disposizione un'unica fonte di informazioni, continuamente aggiornata, che quindi permette di gestire gli asset di rete nel loro intero ciclo di vita.

Per la creazione automatica di istanze di classi IEC CIM a partire dal modello IFC del trasformatore è stata sviluppata un'applicazione in ambiente Python, facendo uso delle seguenti librerie:

- **IfcOpenShell**, per la lettura di file in formato IFC e la navigazione di classi e attributi;
- **rdflib**, per la creazione di istanze IEC CIM di *output*.

In particolare, sono state selezionate le classi IFC di tipo "ifcTransformer" e sono state mappate nelle classi dello standard CIM mostrate in Figura 2.7. Essa rappresenta il **diagramma UML** (Unified Modeling Language) delle classi CIM utilizzate nella fase di trasformazione, assieme alle loro relazioni e attributi; i commenti in giallo indicano le proprietà IFC che vengono mappate in attributi CIM.



IEC 61970 descrive le componenti da un punto di vista elettrico e specifica le loro relazioni. Nel caso d'uso specifico, a questo standard appartengono le classi per la rappresentazione funzionale del trasformatore: *PowerTransformer*, *PowerTransformerEnd*, *Terminal*, *BaseVoltage* e *RatioTapChanger*.

IEC 61968 estende il modello per la descrizione del componente dal punto di vista dell'asset. A questo standard appartengono gli elementi: *Asset*, *PowerTransformerInfo*, *ProductAssetModel*, *Manufacturer*, *Ownership*, *AssetOwner* e *LifecycleDate*.

Figura 2.7 - Diagramma delle classi utilizzate per rappresentare un trasformatore AT/MT in IFC secondo lo standard IEC CIM. In giallo sono mostrate gli attributi IFC, mentre in azzurro le classi IEC CIM.

L'*output* della trasformazione è un file contenente le classi che descrivono il trasformatore secondo la serializzazione in formato RDF XML. Dai risultati ottenuti si evidenzia che il modello IEC CIM permette la rappresentazione completa dei componenti di una rete elettrica, sia dal punto di vista funzionale, sia dal punto di vista della gestione degli asset.

Avendo a disposizione la piattaforma MESP, basata sullo standard CIM, come ricaduta dell'attività di armonizzazione BIM-CIM può risultare interessante integrare nuove funzionalità utili alla gestione del ciclo di vita degli asset, quali visualizzazioni 3D di componenti di rete o accesso ad informazioni di gestione delle infrastrutture di rete.

Si evidenzia che, nello svolgimento delle attività, è stata individuata e testata la libreria **Javascript Open BIM Components**. Questa libreria è una raccolta di oggetti 3D BIM, basata su Three.js, per la loro visualizzazione tridimensionale su pagine web. Essa potrebbe essere utilizzata nel contesto della piattaforma MESP per aggiungere funzionalità tipiche dei *software* BIM. Sono state individuate e descritte di seguito alcune funzionalità che potrebbero essere rese disponibili nella piattaforma:

- la possibilità di evidenziare nella navigazione delle classi IEC CIM quali elementi abbiano un collegamento verso istanze IFC;
- visualizzazione automatica 3D del componente di rete selezionandolo con il mouse sullo schema della rete, mediante l'apertura di una pagina basata su Open BIM Components;
- visualizzazione delle classi IFC che descrivono un oggetto selezionandolo con il mouse fra diversi elementi visualizzati.

In conclusione, l'attività svolta ha approfondito la possibilità di instaurare una **sinergia tra BIM e IEC CIM**, considerando il caso applicativo di un **trasformatore in cabina primaria**. È stata implementata una trasformazione tra le classi IFC che descrivono il trasformatore per scopi di progettazione/gestione dell'*asset* e le relative classi IEC CIM. Avere a disposizione tutte queste informazioni in un unico *knowledge graph*, mantenuto continuamente aggiornato e sincronizzato con lo stato dei componenti installati nelle reti, favorisce la gestione degli *asset* in modo ottimale, nel loro intero ciclo di vita. L'**utilizzo di un identificativo univoco** per gli stessi componenti di rete nei due standard permette di mantenere allineati i modelli BIM e IEC CIM: per scopi di **gestione del ciclo di vita** le *utility* possono continuare a utilizzare gli strumenti specifici dell'**ambiente BIM**, mentre per **scopi operativi** si sfrutta il modello di **rete IEC CIM**.

2.3 Definizione di un'ontologia per la modellazione delle reti gas

Nelle precedenti annualità è stata avviata un'attività di modellazione delle infrastrutture relative a reti energetiche di diverso tipo oltre alle reti elettriche [6, 4]. Per soddisfare i requisiti di interoperabilità tra i differenti vettori energetici per la creazione di un sistema energetico integrato è indispensabile partire dall'interoperabilità a livello informativo: i sistemi devono parlare la stessa "lingua" e capirsi. Le ontologie semplificano tale compito e, per questa ragione, dopo avere verificato la mancanza di un'ontologia di riferimento per le reti gas, si è proposto di adottare per queste reti un modello derivato dallo standard per il settore elettrico IEC CIM.

L'attività svolta nel periodo di riferimento ha riguardato la **definizione di una ontologia per il settore gas**, confrontandosi con i gestori di rete. A tal fine, le ontologie IEC CIM per il settore elettrico sono state considerate come riferimento principale, in quanto costituiscono un valido esempio di modello completo e *standard* per la descrizione di una rete energetica gestita da una *utility*. Il primo vantaggio di una descrizione derivata da CIM sarebbe la possibilità per gli operatori di rete di gestire i dati della rete gas in modo analogo a quelli della rete elettrica; ad esempio, questo rende possibile effettuare interrogazioni simili dei dati e sfruttare le funzioni tipiche dei sistemi DMS anche per la rete gas, come il calcolo di pressioni o portate (equivalente al *power flow* elettrico). Un ulteriore vantaggio è legato alla possibilità di effettuare analisi combinate di reti relative a più vettori energetici, con la risoluzione di problemi di ottimizzazione di sistemi *multi-energy*, facilitate dalla presenza di una struttura comune per il modello dati e dalla connessione tra diversi database semantici. Un esempio potrebbe essere il calcolo simultaneo di *power flow* e *gas flow* considerando la presenza di un turbospansore come punto di interscambio tra le due reti gas ed elettrica (vedi [4]). Questo dispositivo sfrutta la riduzione della pressione del gas dal livello a cui viene ricevuto a quello richiesto dagli utenti per produrre energia elettrica e permette

quindi di studiare un caso completo e rappresentativo della possibilità di interazione tra differenti reti energetiche.

Esistono differenti metodi per estendere il modello IEC CIM per la sua applicazione al vettore energetico gas. Uno di questi è descritto nel documento *Common Information Model Primer* [7] in cui si considera lo scenario in cui una *utility* vuole controllare e monitorare contemporaneamente entrambe le reti, elettrica e gas. La modalità descritta consiste nel **definire nuove classi esterne, nel contesto di un'unica ontologia IEC CIM**, collegate con relazioni di ereditarietà alle classi del modello originario (Figura 2.8). Ad esempio, si definisce una sottoclasse *ExtGasTransferEquipment* della classe *Equipment*, che diventa parte di un'estensione del modello IEC CIM stesso. In questo modo, la nuova classe eredita automaticamente tutti gli attributi della classe padre, ma può essere specializzata con nuovi attributi.

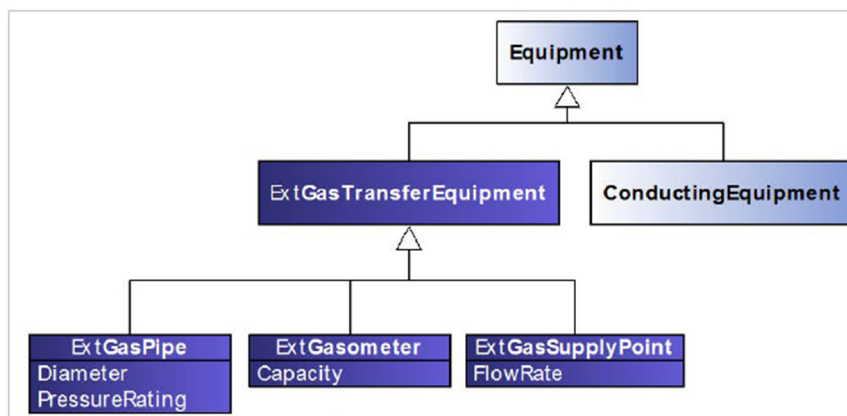


Figura 2.8 – Esempio di estensione dell'ontologia IEC CIM per l'integrazione del vettore gas [7].

Un altro esempio di estensione del modello IEC CIM per la descrizione di un sistema energetico integrato, comprendente rete elettrica, rete gas e rete termica, è proposto da Wang in [8]. L'articolo descrive in modo approfondito le classi definite per modellizzare la rete gas in analogia con IEC CIM e risulta particolarmente interessante per la descrizione dei punti di interscambio tra le reti (ad es. i compressori motorizzati, le unità P2G e P2H e i sistemi di stoccaggio *multi-energy*). Anche in questo caso la definizione di nuove classi per la rete gas avviene tramite estensione del modello IEC CIM e con l'aggiunta di classi esterne (Figura 2.9).

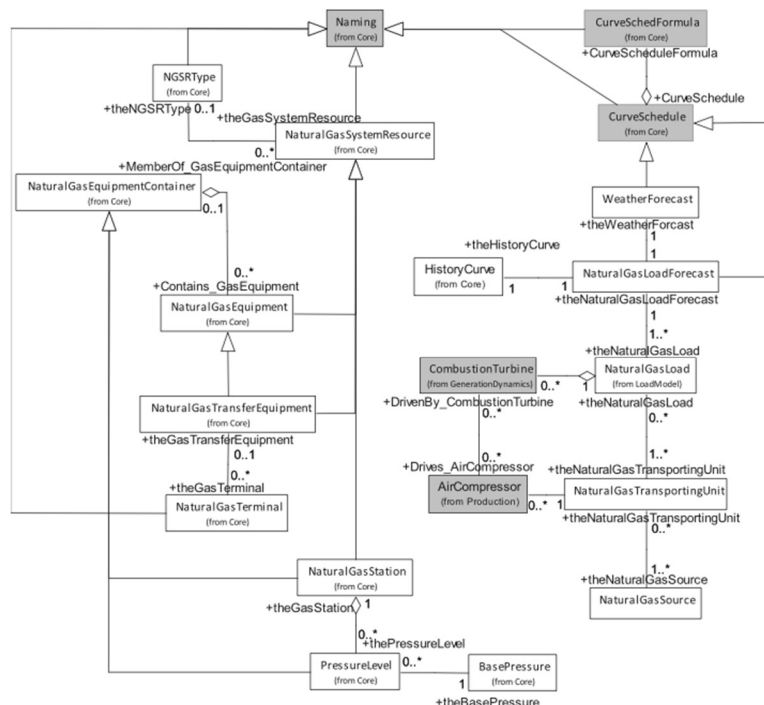


Figura 2.9 - Estensione di IEC CIM per la descrizione della rete gas [8].

A differenza di quanto fatto nei lavori citati, la metodologia adottata qui descritta si basa sulla **definizione di una nuova ontologia**, derivante da IEC CIM ma che non estende l'ontologia IEC CIM stessa. In questo modo, i due settori elettrico e gas vengono mantenuti separati, ma si ripropone, mutuando da IEC CIM, la stessa ereditarietà tra le classi, le relazioni di contenimento e di collegamento tra elementi. Il collegamento tra i due domini elettrico e gas avviene quindi tramite l'utilizzo di *query* federate su modelli di rete separati, ma caratterizzati dalla stessa struttura. Questa modalità è stata già selezionata e applicata, in precedenti attività di ricerca, per la modellizzazione della rete del teleriscaldamento con un'ontologia derivante da IEC CIM [6].

2.3.1 Dati di input

I dati di *input* per lo sviluppo di un modello semantico di reti gas e il test dell'ontologia sono stati *shapefile* descrittivi degli elementi di due porzioni di rete gas di Milano e Brescia.

Ciascuno *shapefile* contiene le informazioni relative a diversi componenti di rete, brevemente elencati e definiti in seguito:

- **anodi galvanici**: proteggono le condotte dalle correnti vaganti;
- **impianti**: si riferiscono a impianti di regolazione della pressione di diverse tipologie;
 - **REMI**: impianti di prima regolazione della pressione e misura del gas naturale. La regolazione avviene da alta a media pressione;
 - **IRI**: impianti di riduzione intermedi della pressione del gas naturale in reti di distribuzione in media pressione. La regolazione avviene da media pressione a media pressione più bassa;
 - **GRF**: gruppi di riduzione finale della pressione del gas naturale. La regolazione avviene da media a bassa pressione;
- **nodi elettrici**: comprende soltanto la tipologia "giunto isolante" e si riferisce al componente della rete che isola la rete dalla protezione catodica;
- **nodi idraulici**: si riferiscono a diverse tipologie di interruzioni meccaniche della rete;

Lo *shapefile* è un popolare formato vettoriale per sistemi informativi geografici (GIS), sviluppato da ESRI per accrescere l'interoperabilità fra i sistemi ESRI e GIS. Con *shapefile* si indica un insieme di file con estensione .shp, .dbf, .shx e altri, i quali hanno in comune il prefisso del nome. Lo *shapefile* può descrivere spazialmente punti, polilinee o poligoni.

- **interruzioni:** componenti usati per connettere due tubature a parità di diametro;
- **tappi:** componenti utilizzati per sigillare le tubature a fine tratta;
- **TePres e TeRete:** giunzioni a T (rispettivamente su rete e presa) usate per collegare 3 o 4 tubature;
- **riduzioni:** componenti usati per connettere due tubature di diverso diametro;
- **punti misura potenziale:** sono i punti (piastrine) di collegamento dei cavi elettrici per la misura del potenziale elettrico della condotta, specificati dalla normativa;
- **punti riconsegna / punti riconsegna utente:** sono i punti di confine tra l'impianto di distribuzione e l'impianto del cliente finale, dove l'impresa di distribuzione consegna il gas naturale per la fornitura;
- **piping:** sono le tubazioni all'interno degli impianti;
- **punto consegna rete nazionale:** è il collegamento della rete di distribuzione alla rete gas nazionale;
- **rete:** comprende le informazioni relative alle tubature della rete;
- **valvole:** sono i componenti in grado di permettere o bloccare il flusso attraverso le tubature.

La visualizzazione geografica della rete gas del quartiere di Lambrate, ottenuta con il *software* QGIS, è mostrata in Figura 2.10.

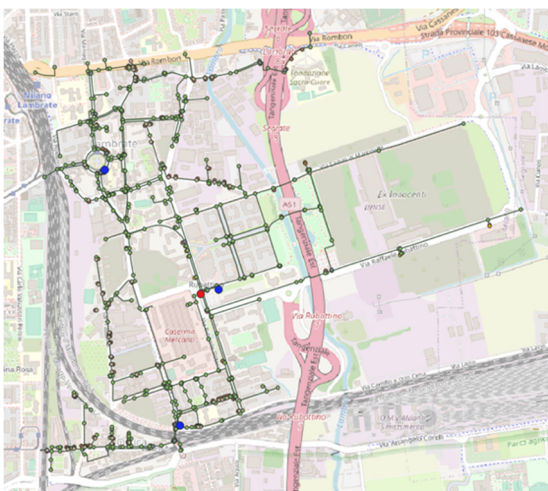


Figura 2.10 - Rappresentazione geografica in QGIS di una porzione della rete gas di Milano. In blu sono mostrati gli impianti GRF, mentre il rosso l'impianto IRI.

2.3.2 Definizione dell'ontologia

Per la **definizione di un'ontologia** in grado di modellare gli elementi precedentemente introdotti e compresi nelle reti gas sono stati considerati due principali riferimenti:

- **l'ontologia IEC CIM** per la descrizione della rete elettrica e dei suoi componenti (IEC CIM 61970 [9], IEC CIM 61968 [10], IEC CIM 62325 [11]). IEC CIM è infatti lo standard di riferimento per la descrizione di reti gestite da *utility* e permette di ottenere una modellizzazione completa di tutti gli elementi della rete elettrica. Ottenere un'ontologia così estesa e completa anche per la rete gas esula dallo scopo di questo lavoro, che invece si è focalizzato sulla definizione di un'ontologia adeguata al caso d'uso in esame. Tuttavia, si è cercato di individuare le analogie tra gli elementi della rete elettrica e quelli della rete gas, in modo da definire un'ontologia di tipo generale per le reti gas in grado di mantenere le stesse relazioni gerarchiche tra oggetti e che utilizza le stesse strutture

CIM per stabilire la connettività tra elementi (*connectivity node* e terminali). Questa soluzione garantisce la possibilità di effettuare, anche per la rete gas, interrogazioni dei dati simili a quelle che vengono già effettuate per la rete elettrica, favorendo le *utility* che gestiscono reti di entrambi i tipi;

- il **modello Gas Utility Network Configuration**, sviluppato da ESRI per la descrizione della rete gas [12]. Questo modello rappresenta l'ontologia individuata più completa per la rete gas, ed è stato preso come riferimento per una corretta denominazione delle classi e dei loro sottotipi, con i relativi attributi.

La definizione delle classi dell'ontologia per la rete gas, con le loro relazioni ed attributi è stata effettuata tramite il *software Protégé* [13], da cui può essere effettuata l'esportazione dell'ontologia direttamente in formato RDF/XML. Di seguito si riassume brevemente la procedura utilizzata per la definizione dell'ontologia per la rete gas.

In analogia con IEC CIM è stata definita una classe **IdentifiedObject**, per fornire un'identificazione comune per tutte la classi che necessitano di denominazione (Figura 2.11).



Figura 2.11 – Diagramma UML per la classe IdentifiedObject dell'ontologia per la rete gas.

La classe figlia **GasSystemResource** è stata definita in analogia con *PowerSystemResource* di IEC CIM e comprende le due classi (Figura 2.12):

- **GasEquipment**, con la rispettiva classe figlia **GasTransferEquipment**. Di questa classe fanno parte elementi di conduzione del gas (tubature), elementi di congiunzione (giunzioni di diversa tipologia) e punti di connessione di utenze (punti di immissione o di prelievo di gas). Le diverse classi contenute sono illustrate in Figura 2.13;
- **ConnectivityNodeContainer**, con la rispettiva classe figlia **GasEquipmentContainer**. Di questa classe fanno parte elementi di contenimento e raggruppamento di altre classi, ovvero *GasSubstation*, *Pipe* e *PressureLevel*.

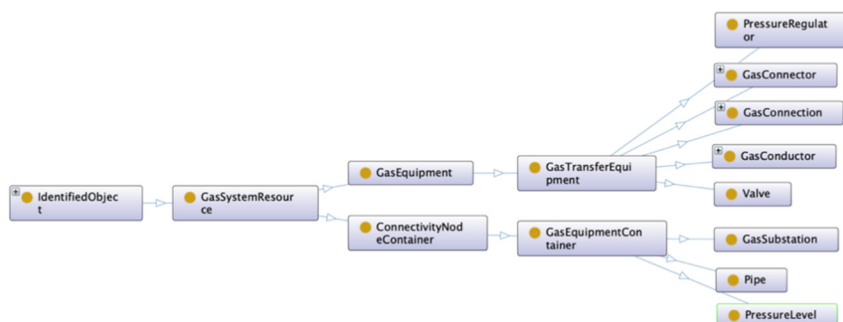


Figura 2.12 - Diagramma UML per la classe GasSystemResource dell'ontologia per la rete gas.

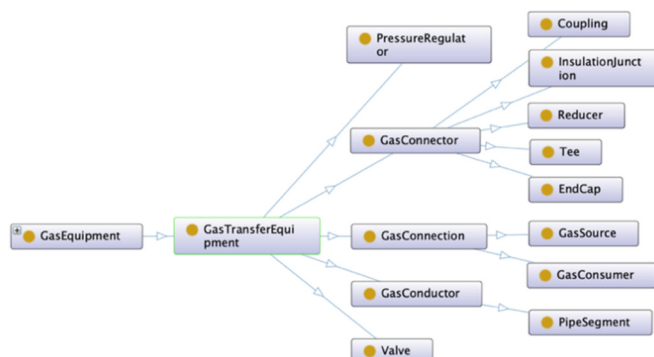


Figura 2.13 - Diagramma UML per la classe *GasTransferEquipment* dell'ontologia per la rete gas.

La modalità di descrizione della connettività tra elementi, con l'utilizzo di terminali e *connectivity nodes*, è stata mutuata da IEC CIM, come mostrato in Figura 2.14.

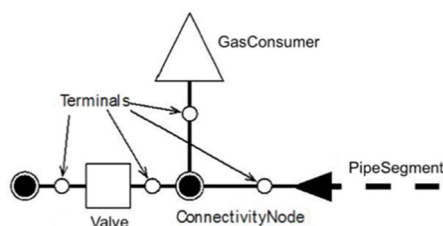


Figura 2.14 - Modalità mutuata da IEC CIM per descrivere la connessione di elementi della rete gas tramite terminali e *connectivity nodes*.

Infine, si riportano in Tabella 2.3 le classi definite, con il corrispondente elemento dell'ontologia IEC CIM per la rete elettrica, del modello ESRI e del modello GIS utilizzato dal gestore di rete.

Tabella 2.3 – Classi dell'ontologia definita per la rete gas e corrispondenze con gli elementi del modello IEC CIM, il modello GIS di Unareti e il modello GIS di ESRI.

Ontologia gas	IEC CIM	Unareti GIS	ESRI GIS
IdentifiedObject	IdentifiedObject	Non definito	Non definito
BasePressure	BaseVoltage	Non definito	Non definito
GasSystemResource	PowerSystemResource	Non definito	Non definito
RegulatorEnd	TransformerEnd	Non definito	Non definito
PressureRegulatorEnd	PowerTransformerEnd	Non definito	Non definito
GasEquipment	Equipment	Non definito	Non definito
GasEquipmentContainer	EquipmentContainer	Non definito	Non definito
GasSubstation	Substation	Non definito	Non definito
Pipe	Line	Non definito	Non definito
PressureLevel	VoltageLevel	Non definito	Non definito
GasTransferEquipment	ConductingEquipment	Non definito	Non definito
GasConnector	Connector	Non definito	Junction
GasConductor	Conductor	Non definito	Non definito
GasConnection	EnergyConnection	Non definito	Non definito
Valve	Switch	Valvola	ControllableValve
PressureRegulator	PowerTransformer	Impianto	Regulator
PipeSegment	ACLineSegment	Rete, Piping	Pipe
InsulationJunction	Junction	NodoElettrico (giunto isolante)	InsulationJunction
Coupling	Junction	NodoIdraulico (interruzione)	Coupling

Ontologia gas	IEC CIM	Unareti GIS	ESRI GIS
EndCap	Junction	Nodoldraulico (tappo)	EndCap
Reducer	Junction	Nodoldraulico (riduzione)	Reducer
Tee	Junction	Nodoldraulico (TePres, TeRete)	Tee
GasConsumer	EnergyConsumer	PuntoRiconsegna, PuntoRiconsegnaUT	simile a RuralTap
GasSource	EnergySource	ConsegnaRN	simile a Wellhead
AssetInfo, WireInfo, PerLengthLineParameter	AssetInfo, PipeInfo, PerLengthLineParameter	Non definiti	Non definiti
ConnectivityNode, Terminal, ConnectivityNodeContainer	ConnectivityNode, Terminal, ConnectivityNodeContainer	Non definiti	Non definiti
CoordinateSystem, Location, PositionPoint, GeographicalRegion, SubGeographicalRegion	CoordinateSystem, Location, PositionPoint, GeographicalRegion, SubGeographicalRegion	Definiti tramite elementi GIS	Definiti tramite elementi GIS

2.3.3 Generazione del modello semantico della rete gas conforme all'ontologia definita

Una volta definita l'ontologia per la rete gas, le classi di *input* in formato *shapefile* sono state convertite in una serializzazione RDF/XML della rete secondo l'ontologia di destinazione, utilizzando un codice Python. Per la gestione delle ontologie e la trasformazione delle classi sono state usate le seguenti librerie:

- **owlready2** per l'importazione dell'ontologia della rete gas e il controllo di conformità delle classi di output al modello di output;
- **rdflib** per lavorare con RDF in ambiente Python, in particolare per creare triple RDF, effettuare query su porzioni di grafo ed esportare la rete di output in formato RDF/XML.

Un esempio di risultato della trasformazione è mostrato in Figura 2.15.

```
<gas:Terminal rdf:ID="_5a9c5192-40c2-11ee-8082-acde48001122">
  <gas:Terminal.GasTransferEquipment rdf:resource="#_5a9c40c6-40c2-11ee-8082-acde48001122"/>
  <gas:Terminal.ConnectivityNode rdf:resource="#_5a9c4166-40c2-11ee-8082-acde48001122"/>
  <gas:IdentifiedObject.name>term-junction-GSDN00010000048293</gas:IdentifiedObject.name>
  <gas:Terminal.connected>true</gas:Terminal.connected>
  <gas:Terminal.sequenceNumber rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">1</gas:Terminal.sequenceNumber>
</gas:Terminal>
<gas:GasSubstation rdf:ID="_5a44e722-40c2-11ee-8082-acde48001122">
  <gas:IdentifiedObject.name>substation-GSDN00010001001970</gas:IdentifiedObject.name>
  <gas:GasSubstation.Region rdf:resource="#_59a02b4c-40c2-11ee-8082-acde48001122"/>
  <gas:GasSystemResource.Location rdf:resource="#_5a44ee5c-40c2-11ee-8082-acde48001122"/>
</gas:GasSubstation>
<gas:PositionPoint rdf:ID="_5adcf72e-40c2-11ee-8082-acde48001122">
  <gas:PositionPoint.Location rdf:resource="#_5adce9fa-40c2-11ee-8082-acde48001122"/>
  <gas:PositionPoint.xPosition rdf:datatype="http://www.w3.org/2001/XMLSchema#double">45.48208142695796</gas:PositionPoint.xPosition>
  <gas:PositionPoint.yPosition rdf:datatype="http://www.w3.org/2001/XMLSchema#double">9.241467176079189</gas:PositionPoint.yPosition>
</gas:PositionPoint>
<gas:PressureLevel rdf:ID="_5aa05026-40c2-11ee-8082-acde48001122">
  <gas:IdentifiedObject.name>junction-JUNC244-10Pa</gas:IdentifiedObject.name>
  <gas:PressureLevel.BasePressure rdf:resource="#_59a0be68-40c2-11ee-8082-acde48001122"/>
</gas:PressureLevel>
```

Figura 2.15 - Estratto della serializzazione in formato RDF/XML della rete gas di Lambrate secondo l'ontologia sviluppata nella presente attività.

Le porzioni di rete gas descritte secondo il formato mutuato da IEC CIM sono state importate sulla piattaforma MESP (*Multi-Energy Semantic Platform*) ed è stato possibile analizzare le reti in modo analogo a quanto fatto per le reti elettriche e di teleriscaldamento (Figura 2.16 e Figura 2.17).

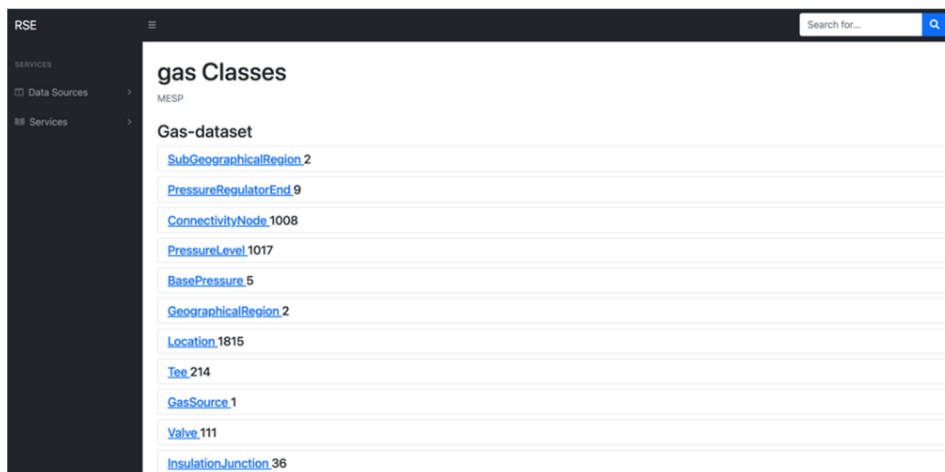


Figura 2.16 - Navigazione delle classi sulla MESP per una porzione della rete gas di Brescia.

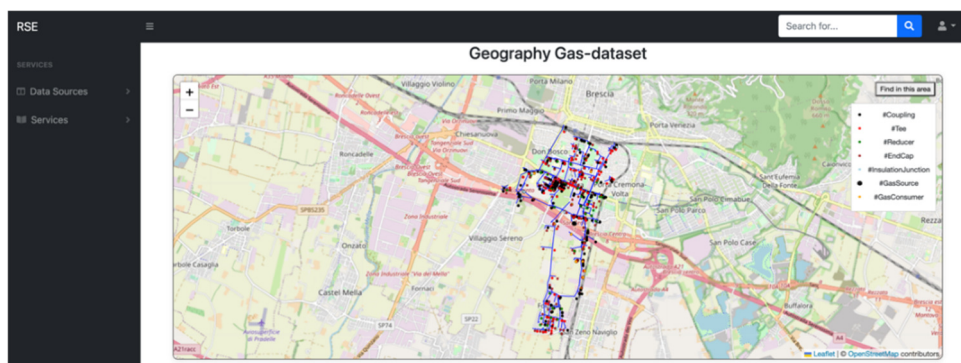


Figura 2.17 - Visualizzazione geografica sulla MESP di una porzione della rete gas di Brescia.

In conclusione, in questa attività è stata definita una nuova ontologia per la **descrizione della rete gas**, basata su IEC CIM per quanto riguarda l'ereditarietà tra le classi, le relazioni di contenimento e di collegamento tra elementi, ma che prende spunto dal modello **Gas Utility Network Configuration**, sviluppato da ESRI, per la denominazione delle classi specifiche del settore gas. Come esempio applicativo, sono state considerate porzioni di reti gas di Milano e Brescia. Le reti di *input*, convertite in una serializzazione RDF/XML secondo l'ontologia definita, sono state caricate sulla piattaforma MESP in una sezione dedicata al settore gas. Da qui sono state rese disponibili le funzionalità di **navigazione delle classi** e di **visualizzazione geografica**, presenti anche per gli altri vettori energetici. Nella prosecuzione delle attività, si procederà con l'integrazione sulla piattaforma MESP delle restanti funzionalità disponibili per gli altri vettori energetici, in particolare analisi e visualizzazioni topologiche e risoluzione di problemi di *gas-flow*.

2.4 Ontologie per architetture ICT

Negli ultimi anni, il concetto di *smart grid* ha rivoluzionato le reti di distribuzione e ha portato all'integrazione di tecniche informative e di comunicazione con la rete fisica. Lo sviluppo di **architetture ICT interagenti con reti per l'energia** risulta ancora più necessario per il passaggio a sistemi *multi-energy*, che richiedono un coordinamento efficace tra risorse energetiche distribuite anche di differenti vettori energetici, per applicazioni di gestione, controllo e protezione dei sistemi. Nasce quindi la necessità di avere una **visione complessiva e unificata delle infrastrutture energetiche, compresa l'architettura ICT al loro servizio**.

Una possibile soluzione per abilitare la gestione unitaria di questi dati consiste nell'utilizzo di tecnologie del web semantico che forniscono linguaggi, come **OWL** (*Web Ontology Language*) e **RDF** (*Resource Description Framework*), in grado di facilitare l'integrazione di informazioni provenienti da fonti eterogenee. Inoltre, l'utilizzo delle ontologie porta anche il vantaggio di poter sfruttare il **reasoning semantico** per estrarre dipendenze non esplicite tra i dati. Un esempio significativo è proposto da Syed *et al.* (2016) [14], che descrivono la possibilità di inferire la presenza di un attacco informatico tramite il *reasoning* semantico su un'infrastruttura ICT.

L'approccio basato sulle ontologie è analogo a quello seguito per descrivere le reti relative a diversi vettori energetici in modo analogo alla rete elettrica [6, 4] (tramite lo sviluppo di standard derivati da IEC CIM), in modo da poter gestire, in modo unificato in un unico *knowledge graph*, i dati di reti multi-energy.

In questo capitolo viene affrontato il tema della definizione di un'ontologia per le architetture ICT utile ad analisi congiunte con le reti dell'energia e per studiare le mutue influenze in termini di affidabilità e resilienza dei servizi energetici.

Vengono inizialmente identificate le classi di IEC CIM utilizzabili per la descrizione di architetture ICT e verificata la loro applicabilità ad architetture di interesse per i DSO. Si fa riferimento ad uno specifico esempio di architettura ICT distribuita per la gestione di una *smart grid* elettrica, ma che può essere generalizzato anche a reti *multi-energy*. Poter utilizzare un'unica ontologia per la descrizione della topologia di una rete per l'energia e delle sue componenti ICT permette di ottenere una visione olistica sulla rete stessa. Se è però necessario risolvere problemi specifici, per cui le classi di IEC CIM non risultano più sufficienti (ad es. problemi legati alla cybersicurezza oppure descrizioni più dettagliate dell'infrastruttura), è possibile fare ricorso a ontologie specifiche per il dominio ICT, di cui si riporta uno stato dell'arte.

2.4.1 IEC CIM per la descrizione di architetture ICT

Per la descrizione di architetture ICT nel dominio elettro-energetico, IEC CIM 61970 comprende delle classi utilizzabili per modellare le informazioni usate dalle applicazioni SCADA (*Supervisory Control and Data Acquisition*). L'utilizzo di un'unica ontologia per la descrizione della topologia di rete e della sua infrastruttura ICT è particolarmente vantaggioso perché permette di ottenere una visione olistica della rete di distribuzione. Questo favorisce la possibilità di interrogare simultaneamente domini differenti e quindi di risolvere in modo combinato problemi di interazione fra la rete elettrica e l'architettura ICT. Esempi potrebbero essere il calcolo di indici di resilienza, tenendo in considerazione guasti della rete elettrica e/o della rete ICT, l'analisi di impatto di attacchi cyber oppure di problemi legati all'isolamento di componenti soggetti a guasti tramite il controllo dello stato di interruttori.

Di seguito sono descritte, e rappresentate in Figura 2.18, le principali classi di IEC CIM utilizzabili per la descrizione di architetture ICT di controllo e monitoraggio della rete.

- **RemoteUnit**: rappresenta una classe generale per indicare elementi come IED (*Intelligent Electronic Device*), sistemi di controllo delle cabine o centri di controllo. Una RemoteUnit è costituita da **RemotePoints**, che possono essere di tipo **RemoteControl** (nel caso in cui la RemoteUnit abbia funzione di controllo) oppure **RemoteSource** (nel caso in cui la RemoteUnit corrisponda a un punto di misura di variabili di stato telemetrate). L'unità centrale di controllo viene modellizzata tramite una RemoteUnit con attributo *remoteUnitType* pari a *ControlCenter*. La RemoteUnit viene collegata ai Terminali del Breaker a cui è associata, in modo da garantire la comunicazione con il centro di controllo. La connessione non avviene direttamente, ma tramite l'utilizzo delle classi **RemotePoint**, **RemoteSource**, **MeasurementValue** e **Measurement**;
- **CommunicationLinks**: descrivono i canali di comunicazione tra dispositivi ICT. Le proprietà dei collegamenti sono descritte tramite le classi **ComMediaAsset** (per descrivere ad esempio la tipologia di rete di comunicazione) e **ReliabilityInfo** (per associare al canale di comunicazione un *failure rate*).

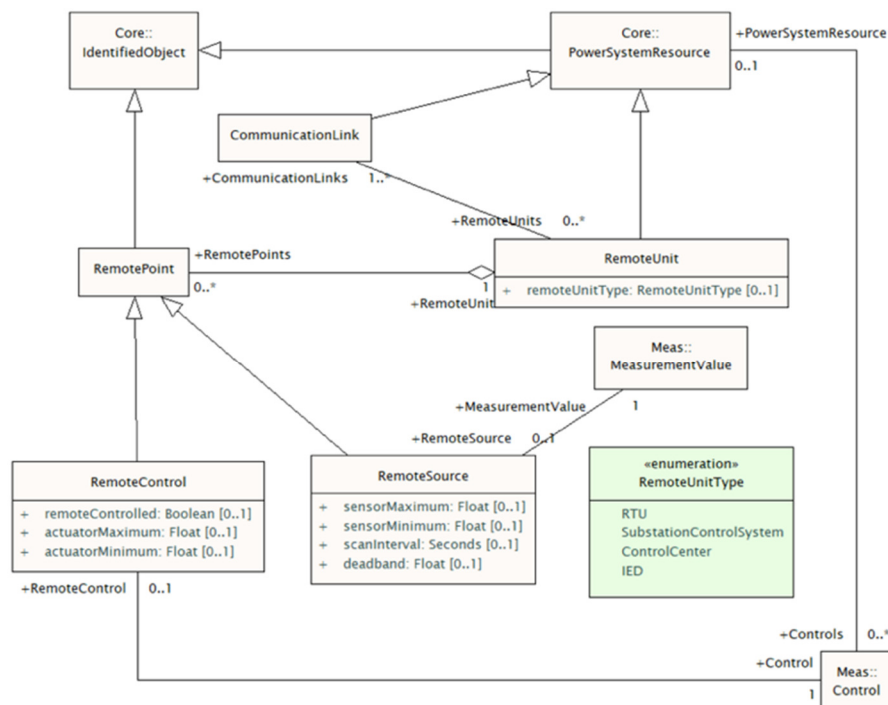


Figura 2.18 - Diagramma delle classi contenute nel pacchetto SCADA di IEC CIM 61970 [15].

In letteratura è stato trovato un solo caso di utilizzo di IEC CIM per una descrizione integrata di aspetti topologici di una rete elettrica di distribuzione e dell'infrastruttura ICT associata [16]. Gli autori hanno sviluppato un modello di "sopravvivenza" di una rete di distribuzione in caso di necessità di intervento a causa di guasti, considerando la probabilità di disponibilità della rete di comunicazione. Le principali classi IEC CIM da loro considerate per lo sviluppo del modello sono mostrate in Figura 2.19.

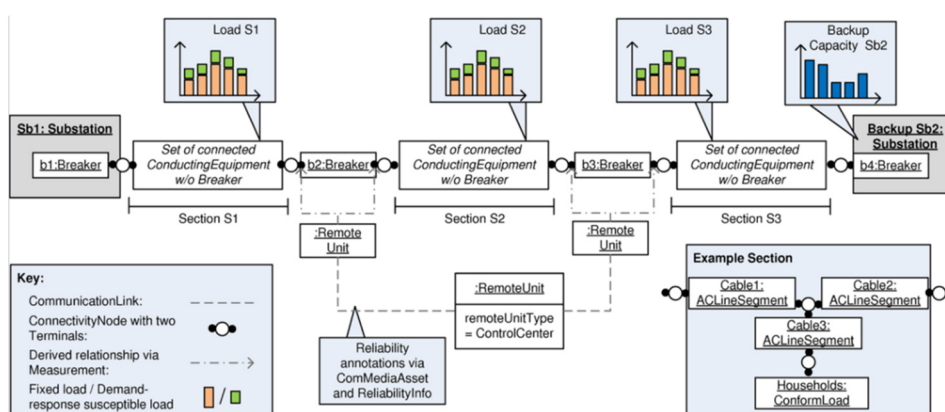


Figura 2.19 – Classi di IEC CIM considerate in [16] per lo sviluppo di un modello di "sopravvivenza" della rete in caso di guasto.

Nonostante la scarsa numerosità delle classi dedicate alle infrastrutture ICT presenti nello standard IEC CIM esse risultano sufficienti per una descrizione generale della struttura di una rete ICT dei DSO.

Si è considerato come esempio rappresentativo di architettura ICT associata alla rete elettrica quella proposta nell'ambito del progetto europeo IDE4L (IDEal grid 4 all) [17] per la gestione e l'automazione della rete di distribuzione. Il DSO Unareti è stato partner di tale progetto e ha

considerato l'architettura sviluppata come riferimento per lo **sviluppo di sistemi di automazione distribuiti**, ad esempio con logiche di automazione per la localizzazione dei guasti e il successivo ripristino del servizio [18] (Figura 2.20).

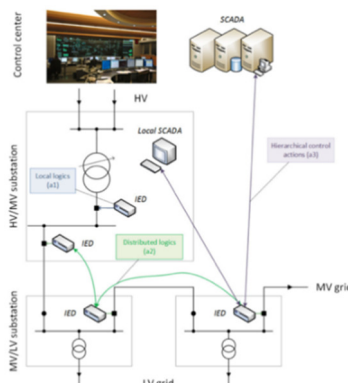


Figura 2.20 - Schema semplificato di un sistema di automazione [18].

L'infrastruttura proposta nell'ambito di IDE4L è descritta in [19] ed è in grado di supportare il monitoraggio, il controllo e la protezione di rete mediante un'architettura organizzata in livelli gerarchici e basata sui seguenti dispositivi:

- gli **Intelligent Electronic Device** (IED) sono utilizzati per il monitoraggio, controllo e protezione della rete di distribuzione a livello locale (controllo primario). In questa categoria sono presenti gli RTU (*Remote Terminal Unit*), PMU (*Phasor Measurement Unit*), Smart Meter e HEMS (*House Energy Management System*);
- le **Substation Automation Units** (SAU) sono i sistemi presenti a livello di cabina primaria e secondaria, che gestiscono le relative porzioni di rete di distribuzione. Modificano i *set point* degli IED sulla base dei risultati dell'esecuzione di algoritmi di *State Estimation* e *State Forecast* (controllo secondario nelle cabine primarie e secondarie). Comunicano sia con gli IED, sia con il *Distribution Management System* e scambiano informazioni con SAU confinanti;
- il **Distribution Management System** (DMS) è il componente presente nella cabina di controllo per la raccolta di dati dal campo. Riceve i risultati di *State Estimation* e *State Forecast* dalle SAU e ottimizza la posizione degli interruttori in modo da limitare congestioni e perdite di potenza (controllo terziario). Comunica con gli altri sistemi presenti in cabina di controllo, come l'*Automatic Meter Management* (AMM) per la raccolta dei dati dagli *smart meter* e i sistemi GIS (*Geographical Information System*), CIS (*Customer Information System*) e NIS (*Network Information System*). Può inoltre gestire l'acquisto di energia e di servizi di flessibilità da operatori di mercato.

Prendendo come esempio l'architettura distribuita proposta in IDE4L, si mostra in Tabella 2.4 l'associazione tra gli elementi dell'architettura e le relative classi presenti nello standard IEC CIM ed utilizzabili per la sua rappresentazione con un modello semantico.

Tabella 2.4 – Esempio di descrizione delle componenti principali dell’infrastruttura ICT proposta in IDE4L tramite le classi IEC CIM.

Elemento infrastruttura ICT	Classe IEC CIM
Centro di controllo (DMS)	RemoteUnit remoteUnitType = ControlCenter
Substation Automation Unit	RemoteUnit remoteUnitType = SubstationControlSystem
Intelligent Electronic Device	RemoteUnit remoteUnitType = IED
Comunicazione tra componenti	CommunicationLink

In conclusione, le classi di IEC CIM risultano sufficienti per una descrizione semplificata e generale dell’infrastruttura ICT dei DSO.

Se invece è necessario risolvere specifici casi d’uso, ad esempio orientati verso aspetti di cybersecurity o che richiedono una descrizione più dettagliata dei componenti ICT, può essere utile estendere IEC CIM tramite l’utilizzo di ontologie specifiche per il settore ICT, di cui si riporta nel successivo paragrafo un’analisi dello stato dell’arte.

2.4.2 Stato dell’arte di ontologie specifiche per il settore ICT

Un’analisi dello stato dell’arte delle ontologie specifiche per la descrizione e la gestione di architetture ICT note in letteratura ha permesso di individuare le seguenti, riassunte in Tabella 2.5.

Tabella 2.5 – Esempi di ontologie specifiche per il dominio ICT.

Nome	Applicazione	Descrizione
ThreMA [20]	Identificazione automatica delle minacce a cui sono esposti gli asset tramite utilizzo delle ontologie	Composta da: • ICT sub-ontology • Data flow sub-ontology • Thread sub-ontology Validata tramite casi studio forniti da stakeholder del settore pubblico italiano
Ontology for Cybersecurity Governance of ICT Systems [21]	Individuazione di vulnerabilità di asset IT/OT tramite dipendenze non esplicite tra i dati	Composta da: • ICT ontology • Vulnerability ontology • Attack ontology
Unified Cybersecurity Ontology (UCO) [14]	Descrizione esaustiva del dominio della cybersecurity	Nata dall’integrazione di diversi dati eterogenei, ontologie sviluppate indipendentemente e concetti dei <i>Linked Open Data</i>
High-Level Ontology Network for ICT Infrastructures [22]	Descrizione di infrastrutture ICT	Rete di ontologie composta da un’ontologia più generale e da altre 9 ontologie interconnesse, per rappresentare: organizzazioni, descrizioni di prodotti e servizi, data center, infrastruttura server, componenti di rete, software, database, componenti hardware e sicurezza di rete

Tramite l’architettura ThreMA [20] gli autori in hanno presentato una nuova modalità, basata sull’utilizzo delle ontologie, per l’identificazione automatica delle minacce a cui sono esposti gli asset (*Automatic Thread Modeling*). Questo avviene tramite la definizione di una rappresentazione standard di modelli e dati usati nelle varie fasi del processo e tramite l’identificazione di un insieme di regole di inferenza, con cui è possibile riconoscere le minacce in modo automatico.

Il meta-modello sviluppato è composto da tre ontologie separate:

- **ICT sub-ontology**: contiene le regole e il vocabolario per la modellizzazione delle infrastrutture ICT
- **Data flow sub-ontology**: per la rappresentazione dello scambio di dati tra programmi software oppure tra elementi dell’infrastruttura

- **Thread sub-ontology:** per la caratterizzazione delle minacce.

Al meta-modello si aggiunge la definizione della **Thread Modelling Logic**, espressa tramite regole di inferenza sull'ontologia. L'architettura ThreMA risulta particolarmente interessante perché è stata validata tramite casi studio forniti da *stakeholder* del settore pubblico italiano. Tuttavia, non sono stati reperiti casi di utilizzo pratico dell'ontologia, o suoi ulteriori sviluppi.

Anche gli autori in [21] sfruttano il *reasoning* semantico per individuare le vulnerabilità di *asset* IT/OT tramite dipendenze non esplicite tra i dati. Essi forniscono un modello in grado di gestire le seguenti informazioni:

- Composizione dell'infrastruttura ICT (asset, reti, funzionalità e dipendenze con entità esterne), descritta tramite l'**ICT ontology**. Essa prende come riferimento l'ontologia sviluppata dall'Agenzia per la Cybersicurezza Nazionale, per la gestione dell'asset ICT
- Potenziali vulnerabilità del sistema, descritte tramite la **vulnerability ontology**
- Informazioni su vettori di attacco e potenziali mitigazioni, descritte tramite l'**attack ontology**.

Non sono stati reperiti lavori successivi che utilizzano o espandono l'ontologia, con scopo di standardizzazione.

Come terzo esempio, l'**ontologia UCO** [14] è nata a partire da uno studio di integrazione di diversi dati eterogenei, ontologie sviluppate indipendentemente e concetti dei *Linked Open Data*, per la descrizione del dominio della cybersecurity. Essa rappresenta l'ontologia più comprensiva ed esaustiva per questo ambito e ha il vantaggio di essere ben documentata e disponibile *online*. L'ontologia, pur essendo la più completa per i concetti generali della *cybersecurity*, risulta migliorabile per la descrizione dell'infrastruttura.

Infine, gli autori in [22] propongono per la descrizione di architetture ICT una rete di ontologie composta da un'ontologia generale ("core") e da altre 9 ontologie interconnesse, in grado di rappresentare: organizzazioni, descrizioni di prodotti e servizi, *data center*, infrastruttura *server*, componenti di rete, software, *database*, componenti hardware e sicurezza di rete. L'ontologia, disponibile *online* [23], è stata sviluppata e testata dagli autori sulla base di infrastrutture ICT di Huawei.

Come emerge dal presente stato dell'arte, le ontologie specifiche per il settore ICT sono in genere orientate alla descrizione e alla gestione di qualche aspetto particolare delle architetture informatiche, ad esempio la *cybersecurity* delle infrastrutture. Nelle *smart grid*, per la rappresentazione e lo studio delle interazioni fra l'infrastruttura elettrica e quella ICT ad essa associata, lo standard CIM rimane un sicuro riferimento di base, come evidenziato al paragrafo 2.4.1. Anche l'ontologia *High-Level Ontology Network for ICT Infrastructures* [22] può essere utile per estendere la descrizione base con dettagli ICT più specifici. A fronte di architetture ICT sempre più complesse, spesso condivise da molteplici casi applicativi, le ontologie costituiscono uno strumento utile darne una rappresentazione completa e condivisa e che permette la gestione efficiente delle infrastrutture ICT.

Nella prosecuzione delle attività, le ontologie ICT individuate potranno essere applicate per la descrizione di architetture ICT utilizzate nella *Test Facility multi-energy* di RSE, di cui è presentata una panoramica nel rapporto [24]. I casi applicativi di riferimento soluzioni di monitoraggio e controllo delle reti energetiche e modelli innovativi di gestione coordinata tra diversi vettori energetici.

3 - DIGITALIZZAZIONE DEI PROCESSI PER LA PIANIFICAZIONE MULTIENERGETICA DEL TERRITORIO

Il raggiungimento degli elevati livelli di decarbonizzazione previsti dai piani nazionali ed europei richiede lo sviluppo delle risorse energetiche rinnovabili e l'integrazione ottimizzata dei diversi vettori energetici per massimizzare la generazione. L'adozione di strumenti modellistici può aiutare nella pianificazione e nella gestione sostenibile del sistema energetico, ma è fondamentale considerare la variabilità spaziale e temporale delle risorse rinnovabili, soprattutto in contesti geograficamente complessi come l'Italia.

Per sviluppare strumenti di supporto alla pianificazione energetica realmente efficaci, i modelli utilizzati devono essere interoperabili e integrati con i *database* contenenti informazioni territoriali, socioeconomiche e tecnologiche.

Il tema della pianificazione energetica viene trattato nel progetto 2.5 "Energia da fonti rinnovabili e integrazione nel territorio" con l'obiettivo di sviluppare strumenti modellistici per studiare gli effetti delle interazioni tra il sistema elettrico e altri vettori energetici nel contesto territoriale. L'attività utilizza dati aventi non solo diversa origine (es. dati istituzionali o derivanti da elaborazioni interne), ma anche risoluzioni spaziali e temporali variabili. Dunque, spesso, l'aggiornamento, l'organizzazione e l'interrogazione simultanea possono costituire passaggi critici. Risulta, pertanto, necessario identificare soluzioni tecnologiche innovative per migliorare la fruibilità, l'efficienza e la flessibilità degli strumenti sviluppati per la pianificazione energetica territoriale.

Il progetto "Digitalizzazione del sistema energetico integrato" si propone, dunque, di identificare soluzioni più flessibili e adeguate alla gestione di grandi quantità di dati eterogenei per il recepimento, l'elaborazione, l'organizzazione dei dati e l'interoperabilità dei modelli.

Tra i vari processi del sistema modellistico per la pianificazione energetica, analizzati nel dettaglio nell'ambito della precedente annualità in termini di flussi informativi e sequenza di elaborazioni necessarie [4], quest'anno si è scelto di approfondire:

1. la progettazione dell'architettura informatica per la gestione dei flussi dati e la conversione del *database* dei parametri tecnologici da cui attinge il modello di pianificazione energetica, da relazionale a semantico;
2. lo sviluppo di una metodologia per l'aggiornamento periodico e automatico della mappatura di impianti fotovoltaici a terra da immagini satellitari, come elemento fondamentale nella valutazione del potenziale territoriale disponibile per le nuove installazioni fotovoltaiche.

3.1 Progettazione dell'architettura informatica

Nella scorsa annualità sono stati presentati i processi di raccolta dati per condurre le analisi di pianificazione del territorio [4]. In particolare, ci si è focalizzati sui flussi dati che costituiscono l'input del modello CALLIOPE utilizzato per risolvere il problema della pianificazione multi-obiettivo. Parte dei dati di input vengono elaborati e preparati da processi presenti nello strumento TOTEM [25]. Una panoramica dei processi è presente nello schema MIRO³ dove sono descritte le sorgenti dati e i relativi processi di acquisizione/elaborazione. Sulla base di queste *pipeline* sono stati identificati i requisiti necessari per poter automatizzare ed efficientare tali processi. In particolare, il caso d'uso analizzato riguarda la pianificazione ottimale delle risorse multienergetiche. In questa sezione sono descritte le seguenti attività:

³ Miro è un ambiente di lavoro collaborativo visuale.

- revisione dell'architettura informatica della piattaforma per la pianificazione multienergetica del territorio sviluppata e utilizzata da RSE nell'ambito delle attività del parallelo progetto "Energia da fonti rinnovabili e integrazione nel territorio;
- implementazione della componente che descrive i parametri tecnologici con utilizzo di tecnologie semantiche.

3.1.1 Individuazione dei requisiti

Per eseguire le analisi di pianificazione il modello CALLIOPE richiede l'interrogazione di tipi di dato eterogenei (Figura 3.1) che provengono da elaborazioni talvolta articolate operate in gran parte dallo strumento TOTEM (*Territory Overview Tool for Energy Modeling*).

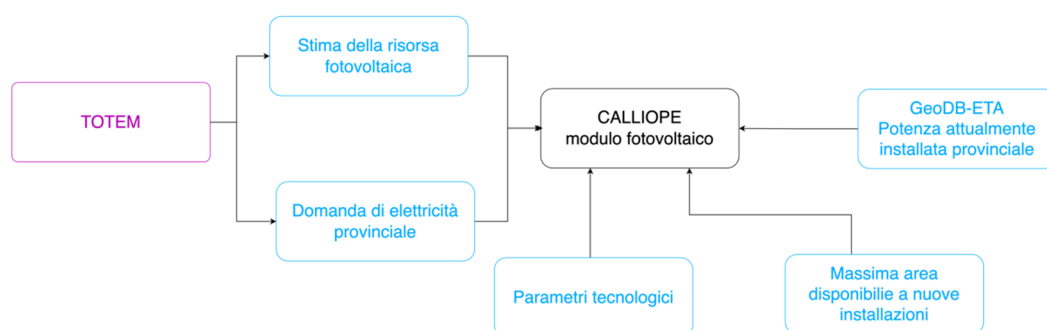


Figura 3.1 – Dati che confluiscono nel modello CALLIOPE nel caso di pianificazione per l'installazione di nuovi moduli fotovoltaici. Ogni sorgente dati viene ottenuta a partire da processi ETL a partire da sorgenti esterne (ENTSO-e, ISTAT, TERNA, etc.)

Di seguito vengono presentati i requisiti che la nuova architettura dovrà soddisfare.

3.1.1.1 Requisiti di usabilità del modello

I primi requisiti, fondamentali per l'utilizzatore della piattaforma di pianificazione sono legati alla possibilità di utilizzo del modello CALLIOPE con possibilità di effettuare analisi personalizzate.

[R-1] L'utente deve poter eseguire il modello CALLIOPE.

[R-2] L'utente deve poter eseguire analisi personalizzate su una base dati condivisa.

Per quanto riguarda la gestione di dati e metadati, sono state individuate tre macroaree tematiche dalle quali originano i requisiti: le tipologie di dati supportati, l'elaborazione dei dati e la loro osservabilità.

3.1.1.2 Requisiti di tracciabilità e schedulazione

Dal punto di vista del processamento dei dati, un meccanismo strutturato per eseguire processi di importazione, trasformazione e caricamento (*Extract, Transform, Load* - ETL) attualmente non è presente. In molti casi il download dei dati può avvenire manualmente attraverso un'interfaccia web o, nel caso la sorgente dati lo supporti, in maniera semi automatica per mezzo di *script*. La non automatizzazione dei processi di *download* rende difficile una possibile parametrizzazione dei processi di estrazione e caricamento di file – come l'interrogazione delle basi dati per provincia o area geografica - per le analisi di pianificazione. Diventa quindi importante nella nuova architettura poter disporre di "blocchi" di codice parametrizzabile che possa essere eseguito, schedato e tracciato. Laddove sia possibile automatizzare l'acquisizione dei dati, bisogna disporre di comandi per poter effettuare le estrazioni in maniera automatizzata. In molti casi le

analisi multi-scenario richiedono una iterazione sulla preparazione dei dati quindi il processo di raccolta dati ed elaborazione deve essere facilmente eseguibile.

- [R-3] L'utente deve poter importare i dati da sorgenti esterne ed eseguire operazioni di preelaborazione.
- [R-4] L'utente deve poter schedare i processi ETL.
- [R-5] L'utente deve poter visualizzare la cronologia dei processi ETL effettuati.
- [R-6] L'utente deve poter visualizzare e navigare le strutture dati coinvolte nei processi ETL.

3.1.1.3 Requisito di osservabilità dei metadati

Dal punto di vista più globale del *data management*, non c'è modo di visualizzare in maniera chiara e veloce le sorgenti dati da cui estrarre le informazioni, i relativi metadati e i processi ETL che legano le differenti sorgenti dati e la storia dei processi ETL. L'assenza di queste caratteristiche si riflette nella difficoltà di parametrizzazione dello strumento TOTEM, che si trova a dover utilizzare sorgenti dati rigide. Vista la quantità di sorgenti dati coinvolte bisogna poter avere una panoramica ad alto livello e la possibilità di entrare nel dettaglio delle differenti strutture dati coinvolte e relativi flussi di elaborazione ETL. Gli *output* di tali flussi costituiranno gli *input* degli algoritmi di pianificazione.

- [R-7] L'utente deve poter visualizzare i metadati e i processi ETL attualmente attivi nella piattaforma.

3.1.1.4 Requisito di interoperabilità e supporto dati eterogenei

Per quanto riguarda le sorgenti dati del modello, ci sono dati esterni (TERNA, ENTSO-e, ISTAT, ISPRA, etc.) che vengono importati in differenti formati (xlsx, pdf, csv, etc.). In molti casi tali dati vengono processati ed inseriti nel database GeoDB [26] che attualmente gestisce **serie temporali**, dati **GIS** e dati **strutturati** interrogabili mediante *query* SQL. I dati salvati in GeoDB possono essere ulteriormente elaborati per poi diventare l'*input* del modello Calliope. In alcuni casi l'utilizzo di database relazionali può vincolare o complicare la gestione dei dati e, in presenza di modelli dati complessi, può non rappresentare la soluzione ottimale. Un'alternativa è costituita da *datastore* senza schema (o con schema opzionale) come i *triplestore* dove, se necessario, è possibile definire uno schema utilizzando delle ontologie. In caso di dati meno strutturati, come rapporti e documenti, potrebbe essere utile invece adottare soluzioni basate sulla rappresentazione dei documenti come oggetti.

- [R-8] La base dati della piattaforma deve supportare tipologie di dati strutturati e non strutturati.

I requisiti qui elicitali sono prevalentemente funzionali e specifici per i processi di pianificazione analizzati. Ad integrazione di questi requisiti serve aggiungere requisiti non funzionali come quelli già considerati nel corso della progettazione della piattaforma MESP [27] e che possono essere riutilizzati. In questo modo l'architettura ICT a supporto della pianificazione territoriale multienergetica potrà avvantaggiarsi dell'impianto dell'architettura MESP sulla quale si innesteranno ulteriori funzionalità definite dai nuovi requisiti.

3.1.2 Architettura software per la pianificazione multienergetica del territorio

Tenendo in considerazione l'analisi dei requisiti descritta nel paragrafo precedente, per la digitalizzazione dei processi di pianificazione multienergetica si è optato per l'adozione come

riferimento dell'architettura della piattaforma MESP, con l'aggiunta dei nuovi componenti mostrati in Figura 3.2. Tale rappresentazione mostra, in verde, componenti già sviluppati, come il *layer* di visualizzazione e il *layer* di base dati interna che supporta differenti tipologie di dati. Le componenti in blu rappresentano le aggiunte necessarie per gestire i nuovi requisiti elicitati. L'area arancione rappresenta i dati provenienti da sorgenti esterne e le tipologie di formati da gestire. Nel resto della sezione sono spiegate in dettaglio le differenti componenti mostrate.

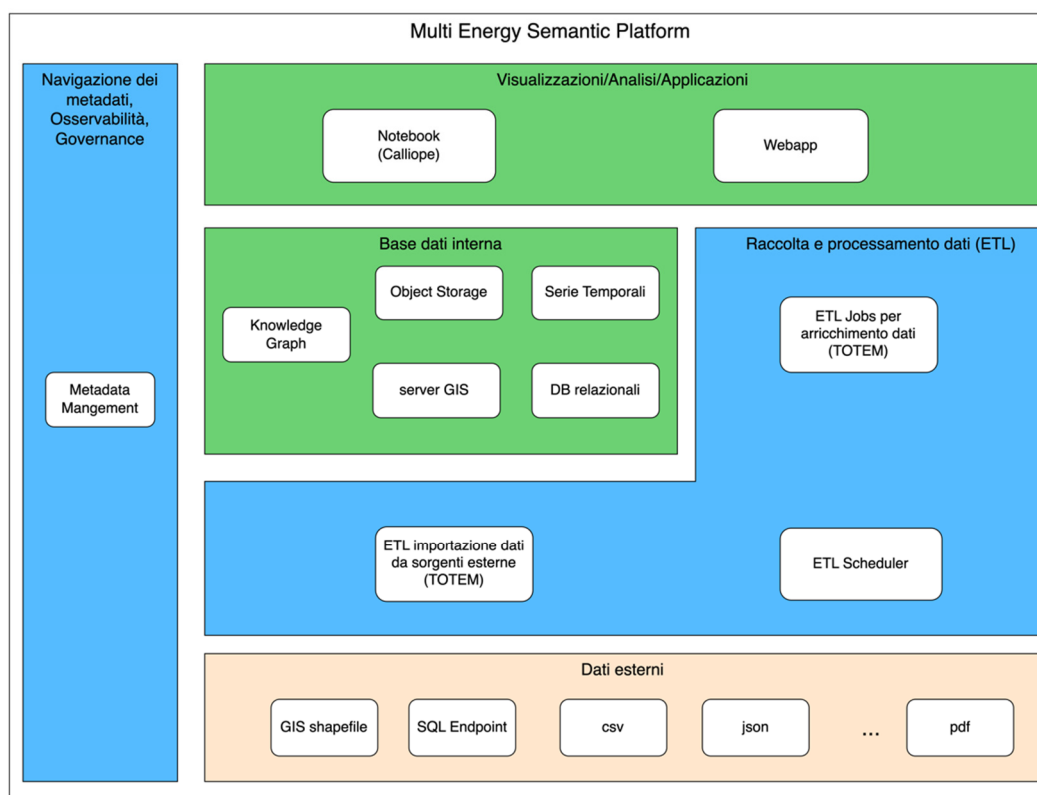


Figura 3.2 - Architettura software per la gestione dei flussi dati per la pianificazione multi-energetica del territorio. Le aree in blu relative a metadata management ed ETL rappresentano aspetti nuovi rispetto all'architettura della piattaforma MESP.

In ragione della molteplicità di tipologie di dati esterni da cui derivare gli *input* ai modelli di pianificazione presenti in Calliope, si richiede la presenza di diverse modalità di importazione che verranno gestite dall' ETL. In questo blocco sono inclusi i processi di elaborazione già implementati in TOTEM e saranno presenti differenti connettori che permetteranno di interfacciarsi direttamente con le sorgenti dati esterne, quando possibile. Come descritto in precedenza, non tutte le operazioni sono automatizzabili ma, nel caso, si provvederà a creare il blocco di codice specifico, che potrà essere riutilizzato per sorgenti dati simili. Oltre a fornire i connettori, questo blocco permetterà di caricare i dati acquisiti nell'area di memoria interna dove risiedono i dati in ingresso processati, pronti all'interrogazione da parte di TOTEM per fornire gli *input* al modello Calliope (vedi ad esempio Figura 3.1).

Talvolta sono richieste azioni di *processing* su dati già presenti nella base dati interna. Queste operazioni possono consistere nella realizzazione di viste contenenti aggregazioni sui dati importati dalle sorgenti esterne oppure incroci con dati di altre tabelle che servono da *input* ai modelli di analisi. In tal caso saranno sviluppati processi ETL, che abbiamo chiamato *arricchimento dati*, che elaboreranno dati presenti nella base dati interna e salveranno i risultati nello stesso. Per tenere traccia dell'esecuzione dei processi ETL è stata predisposta la componente ETL Scheduler

I **Notebook** sono un ambiente di programmazione interattivo che consente agli utenti di creare documenti che includono codice, output di codice, testo formattato e immagini. Sono ampiamente utilizzati per l'analisi dati e la visualizzazione.

che permette di gestire la frequenza con il quale eseguire l'importazione di nuovi dati e tenere traccia delle operazioni passate.

L'*Internal Datastore* rappresenta il luogo nel quale i dati processati dalle *pipeline* ETL vengono salvati e sono pronti all'interrogazione di strumenti come TOTEM. Questo blocco può essere considerato come l'analogo del *database* GeoDB utilizzato finora. La differenza principale risiede nelle tipologie di dato supportate. Infatti, oltre ai dati strutturati già presenti (GIS, dati interrogabili in SQL) sono previsti connettori ed interfacce per integrazioni di basi dati non relazionali come *knowledge graph* e *Object Storage*.

Per poter visualizzare, gestire, elaborare e fruire dei dati disponibili è stata prevista un'area di *Visualization/Application/Analysis*. Quest'area ospiterà le visualizzazioni dei dati per mezzo di *web app* e permetterà di eseguire interrogazioni, elaborazioni e analisi, incluse le attuali funzionalità di Calliope, con strumenti quali i *Notebook*. In particolare, i flussi dati di TOTEM e le analisi effettuate con Calliope sono fatte tipicamente utilizzando *script* in linguaggio Python. Per facilitare l'utilizzo di processi automatizzati e lo sviluppo di nuove analisi in maniera condivisa, si ritiene una buona soluzione disporre di uno strumento come i *Notebook jupyter* sui quali sviluppare *workflows* condivisi da inserire in gestori di *pipeline* ETL. Esempi utili di strumenti *open source* utilizzabili a questo scopo sono *Dagster* (Figura 3.3) o *Airflow*.

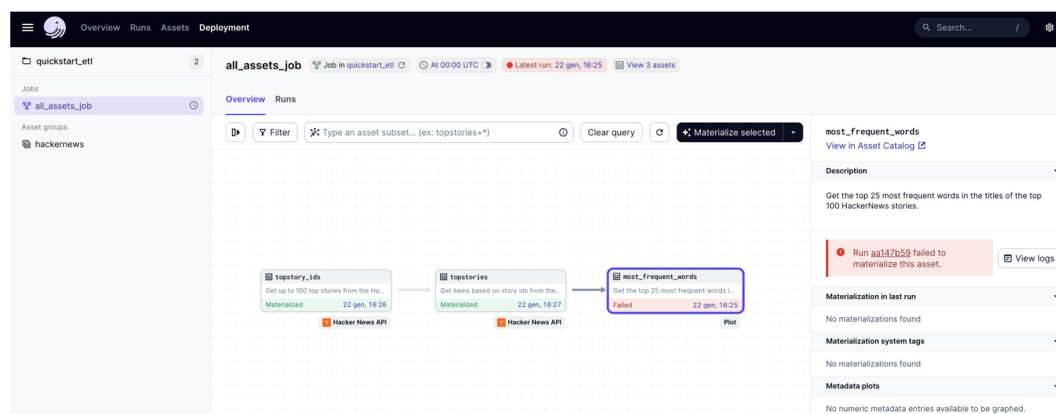


Figura 3.3 - Esempio di gestione si flussi ETL mediante piattaforme *opensource*. Un esempio di gestione di ETL è dato da Dagster.

L'eterogeneità dei dati e la quantità di processi di acquisizione necessari alla pianificazione multienergetica ha evidenziato la necessità di chiarezza relativamente ai metadati di ciascuna sorgente e le relative *pipeline* di acquisizione. Questa esigenza è comune ad altri processi che richiedono la garanzia della qualità dei dati; intatti, la derivazione e la tracciabilità del dato (*data lineage*) è fondamentale per comprendere il flusso di informazioni all'interno di un'organizzazione e garantirne la qualità. A questo scopo sono noti strumenti *open source*, come OpenMetadata, che offrono servizi di monitoraggio della *data lineage*, dando una rapida visualizzazione di come i dati vengono trasformati e utilizzati nei vari sistemi. Pertanto, nell'architettura è stato introdotto un *layer* verticale (Data discovery, observability, governance) che includerà strumenti per facilitare l'osservabilità delle sorgenti dati e dei processi dell'infrastruttura di elaborazione. Utilizzando uno strumento come questo, di cui un esempio è mostrato in Figura 3.4, è possibile monitorare l'evoluzione dei metadati di un determinato processo e i relativi task di ETL.

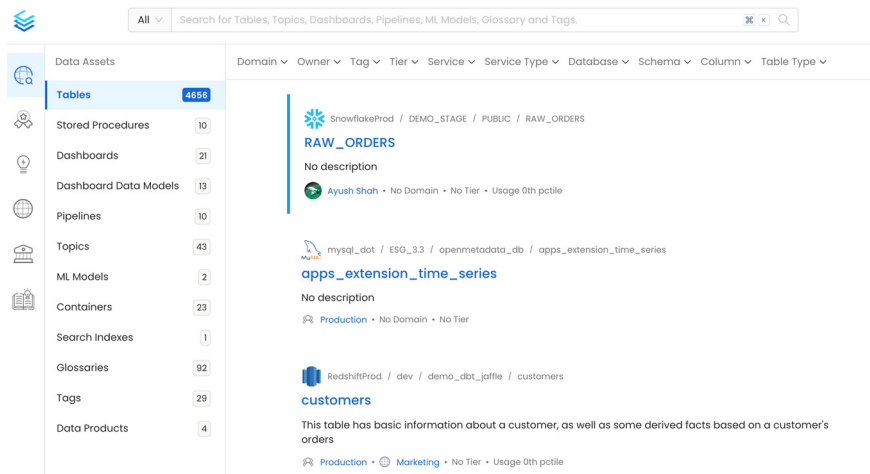


Figura 3.4 - Un esempio di visualizzazione di uno strumento di gestione dei Metadati (*OpenMetadata*). Questo strumento raggruppa gran parte degli aspetti relativi a dati e flussi dati. Tra le possibilità offerte c'è l'esplorazione di metadati di determinati database e la visualizzazione di pipeline ETL.

3.1.3 Trasformazione dell'osservatorio tecnologico in formato semantico

Con riferimento all'architettura proposta nel paragrafo precedente, è stata identificata una parte del processo di pianificazione multienergetica che potesse trarre beneficio dalla possibilità di utilizzare *knowledge graph* nella gestione dei dati. In particolare, è stato preso in considerazione l'osservatorio tecnologico dei sistemi di produzione di energia e di accumulo necessario per la pianificazione e definito un nuovo modello della base dati per i parametri che caratterizzano ciascuna tecnologica di produzione e accumulo.

Per effettuare le analisi di pianificazione (Figura 3.5) il modello Calliope utilizza come *input* i dati relativi a parametri tecnologici di sistemi di produzione di energia e di accumulo sotto forma di un *file* di configurazione [4]. Tale *file* dipende da caratteristiche specifiche della tecnologia scelta. Le differenti tecnologie (produzione FER, accumulo e conversione tra i diversi vettori energetici, etc.) possono avere proprietà molto differenti ed eterogenee. Nel corso di questo triennio, nel progetto 2.5 "Energia da fonti rinnovabili e integrazione nel territorio", è stata avviata una attività volta a realizzare un *database* consultabile che contenga i dati principali riguardanti le diverse tecnologie per la produzione di energia e per l'accumulo [28]. Sono state a tal fine individuate le fonti dati da cui attingere le informazioni, i parametri utili alla caratterizzazione delle tecnologie ed è stato valutato un primo approccio di archiviazione utilizzando un database relazionale. Alcuni aspetti critici emersi nella prima realizzazione del database sono stati l'eterogeneità dei dati, la gestione di diversi livelli di dettaglio, e la presenza di parametri differenti in funzione della tecnologia considerata per cui diventa problematico definire una struttura di memorizzazione adatta a gestire tutte le esigenze.

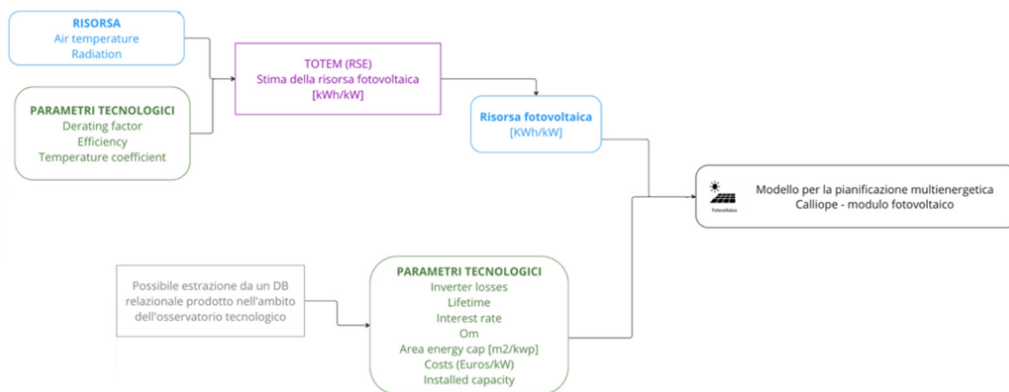


Figura 3.5 - Porzione del processo nel quale i parametri tecnologici vengono utilizzati.

A fronte di queste criticità nell'attività di ricerca è stata investigata la possibilità di rappresentare le informazioni relative ai parametri tecnologici di sistemi di produzione di energia e di accumulo utilizzando un modello semantico. È stata quindi definita un'ontologia per descrivere questo tipo di informazioni ed in seguito sono state generate le prime istanze del modello.

3.1.3.1 Definizione dell'ontologia

Viene qui descritto il processo che ha portato alla definizione del modello semantico adottato per la descrizione dei dati dell'osservatorio tecnologico per i sistemi di produzione di energia e di accumulo. Per la definizione del modello è stato adottato un metodo sistematico di sviluppo denominato Ontology101 [29] che, seppur con qualche limitazione, è il metodo più indicato per lo sviluppo agile di ontologie.

I passi eseguiti sono stati quindi i seguenti:

1. identificazione dello scopo dell'ontologia;
2. identificazione dei termini importanti;
3. gerarchizzazione delle classi;
4. identificazione delle proprietà tra classi;
5. definizione dei vincoli sulle proprietà;
6. generazione delle istanze.

Per mezzo di interazioni con gli esperti di dominio i passi indicati sono stati iterati fino a quando non è stato raggiunto un insieme stabile di classi idonee in termini di intuitività, estendibilità e manutenibilità. Essendo un processo iterativo, il modello ontologico potrà quindi essere migliorato nel corso del tempo in base alle necessità che nasceranno dall'utilizzo dei dati nel modello Calliope.

L'ontologia risultante è incentrata prevalentemente su tre classi principali:

- **Technology;**
- **EnergySource;**
- **TechnologyParameter.**

In ottica di un possibile riutilizzo da parte di altri soggetti e/o in progetti europei l'ontologia è stata sviluppata facendo uso di termini in lingua inglese.

Sfruttando l'ereditarietà fra le classi sono state definite le classi principali e le relative relazioni, mostrate in Figura 3.6.

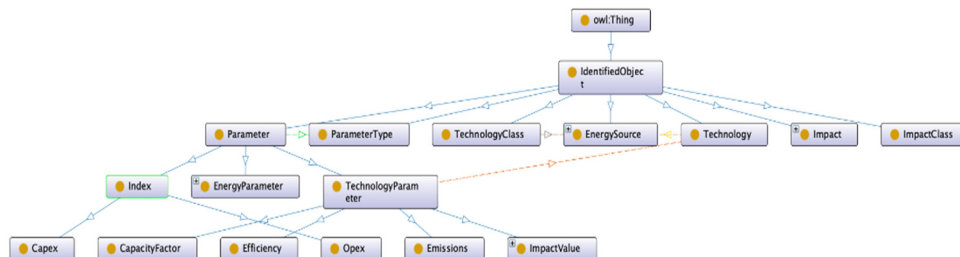


Figura 3.6 - Visualizzazione con il tool Protege delle classi e delle relazioni principali definite nel modello.

Insieme agli esperti di dominio sono state formulate alcune domande in linguaggio naturale (*competency question*) per identificare l'obiettivo a cui deve rispondere la base dati:

- Quali sono le tecnologie appartenenti ad una determinata classe?
- Data una tecnologia qual è il suo costo?
- Data una tecnologia quali sono i parametri principali?
- Quali sono gli impatti di una specifica tecnologia?

Queste domande sono state quindi riscritte nella forma di interrogazioni ad un *knowledge graph* utilizzando i termini definiti nell'ontologia e il linguaggio di interrogazione dei grafi RDF SPARQL. Tali *query* sono state utilizzate per validare la completezza dell'ontologia nel descrivere il dominio.

3.1.4 Generazione delle istanze

Una volta definito il modello ontologico, si è iniziato a popolare il *knowledge graph* corrispondente. A partire da un campione delle tecnologie archiviate nel *database* relazionale, le informazioni associate descrittive della tecnologia sono state trasformate in istanze delle differenti classi mediante uno script Python. L'*output* della trasformazione (Figura 3.7) viene serializzato in formato XML/RDF e può essere quindi caricato nel *datastore* semantico assieme all'ontologia, per poter essere successivamente interrogato mediante *query* in linguaggio SPARQL.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:base="file:///Users/base/pianificazione_territorio/output/osservatorio.xml#"
  xmlns:land="http://land-planning.rse-web.it/v1/#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <rdf:Description rdf:about="fb69e695-b24f-4113-8b9b-7932a19468ce">
    <rdf:type rdf:resource="http://land-planning.rse-web.it/v1/#Capex"/>
    <land:Parameter.referenceYear rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2023.0</land:Parameter.referenceYear>
    <land:Parameter.maxValue rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1680.0</land:Parameter.maxValue>
    <land:Parameter.minValue rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1200.0</land:Parameter.minValue>
    <land:maxSize rdf:datatype="http://www.w3.org/2001/XMLSchema#double">850.0</land:maxSize>
    <land:minSize rdf:datatype="http://www.w3.org/2001/XMLSchema#double">850.0</land:minSize>
    <land:hasTechnology rdf:resource="733362b4-4415-46bb-9989-91f7472ad94c"/>
  </rdf:Description>
  <rdf:Description rdf:about="26031f3a-867d-413b-80f4-b51c62e2ecdd">
    <rdf:type rdf:resource="http://land-planning.rse-web.it/v1/#Opex"/>
    <land:Parameter.referenceYear rdf:datatype="http://www.w3.org/2001/XMLSchema#double">2019.0</land:Parameter.referenceYear>
    <land:Parameter.value rdf:datatype="http://www.w3.org/2001/XMLSchema#double">0.0</land:Parameter.value>
    <land:maxSize rdf:datatype="http://www.w3.org/2001/XMLSchema#double">20.0</land:maxSize>
    <land:minSize rdf:datatype="http://www.w3.org/2001/XMLSchema#double">1.0</land:minSize>
    <land:hasTechnology rdf:resource="ce8c7f1d-b024-4c9b-9685-40a919fa445a"/>
  </rdf:Description>
  <rdf:Description rdf:about="7064c3a0-d798-4eba-a3eb-727915bde67">
    <rdf:type rdf:resource="http://land-planning.rse-web.it/v1/#EnergySource"/>
    <land:IdentifiedObject.name>Energia geotermica</land:IdentifiedObject.name>
    <land:isRenewable rdf:datatype="http://www.w3.org/2001/XMLSchema#boolean">true</land:isRenewable>
    <land:sourceType>Primary</land:sourceType>
  </rdf:Description>
</rdf:RDF>
```

Figura 3.7 - Rappresentazione xml/RDF delle istanze dei parametri tecnologici.

Accedendo al *datastore* in cui i dati sono memorizzati come istanze delle classi dell'ontologia è possibile quindi di validare il nuovo archivio utilizzando le *competency question* descritte

precedentemente. In Figura 3.8 è mostrato un esempio di *query* in linguaggio SPARQL che fa uso delle classi definite nell'ontologia.

/parametri-tecnologici

query add data edit info

SPARQL Query

To try out some SPARQL queries against the selected dataset, enter your query here.

Example Queries

Selection of triples Selection of classes

Prefixes

rdf rdfs owl xsd

SPARQL Endpoint

Content Type (SELECT)

Content Type (GRAPH)

/parametri-tecnologici/query

```

1 PREFIX lp: <http://land-planning.rse-web.it/v1/#>
2 PREFIX base: <file:///Users/mac/PycharmProjects/pianificazione_territorio/output/>
3 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
4 SELECT ?techId ?techName ?lifetime ?name ?className
5 WHERE {
6   ?techId a lp:Technology;
7           lp:hasTechnologyClass ?techClass;
8           lp:lifetime ?lifetime;
9           lp:identifiedObject.name ?techName.
10  ?techClass lp:identifiedObject.name ?className.
11 }
12 ORDER BY ?subject
13 LIMIT 10

```

Table Response 10 results in 0.137 seconds

techId	techName	lifetime	name	className
<file:///Users/base/pianificazione_territorio/output/67b9c481-18e6-4b71-8114-6c008dcd31c>	Batteria	*25**<http://www.w3.org/2001/XMLSchema#integer>		Accumulo
<file:///Users/base/pianificazione_territorio/output/d5d8f102-d406-4e49-9e33-20f8dccb5aa>	Serbatoio idrogeno	*25**<http://www.w3.org/2001/XMLSchema#integer>		Accumulo
<file:///Users/base/pianificazione_territorio/output/9cb9f882-cecf-47e7-b1dc-065fc8f5479d>	Caldia a pellet	*25**<http://www.w3.org/2001/XMLSchema#integer>		Caldia a combustibile
<file:///Users/base/pianificazione_territorio/output/bbdfb874-57d0-41f8-9756-45188a637c53>	Caldia a GPL	*25**<http://www.w3.org/2001/XMLSchema#integer>		Caldia a combustibile
<file:///Users/base/pianificazione_territorio/output/8c67bef-5757-4165-8309-5d38d970a37f>	Dummy tech	*20**<http://www.w3.org/2001/XMLSchema#integer>		Dummy class
<file:///Users/base/pianificazione_territorio/output/1ac21cf-9033-4dff-b6a3-9d0ada5736bd>	PEM	*10**<http://www.w3.org/2001/XMLSchema#integer>		Elettrolizzatori
<file:///Users/base/pianificazione_territorio/output/6189c84e-d450-46cc-8352-f1325e99b78c>	Elettrolizzatore alcalino	*25**<http://www.w3.org/2001/XMLSchema#integer>		Elettrolizzatori

Figura 3.8 - Esempio di interrogazione con il linguaggio SPARQL delle istanze generate in formato XML/RDF. In particolare, questa query risponde alla domanda "Quali sono le tecnologie appartenenti ad una determinata classe?".

L'utilizzo di una ontologia nel contesto descritto ha il vantaggio di poter descrivere in maniera naturale un modello dati complesso, che conta differenti proprietà associate ai particolari tipi di tecnologie di generazione ed accumulo, senza richiedere di alterare tutta la struttura della base dati. L'utilizzo con l'ontologia di funzionalità di *reasoning*, come l'ereditarietà e l'inversione, permette inoltre di inferire nuove informazioni da quelle archiviate senza doverle inserire esplicitamente nel *knowledge graph*. L'utilizzo dell'ontologia e del modello semantico favorisce inoltre la condivisione di conoscenza. Più in generale il formato semantico risulta utile come metodo per scambiare informazioni in modalità *machine readable*, abilitando in questo modo l'interazione tra agenti e tra algoritmi e favorendo lo sviluppo di soluzioni basate sull'intelligenza artificiale.

Riassumendo, il primo passo per la trasformazione dei dati dell'osservatorio sulle tecnologie di generazione e accumulo in un *knowledge graph* è stato completato ed è reso disponibile ai processi di pianificazione un servizio interrogabile mediante il linguaggio SPARQL. Compilando le *query* opportune sarà possibile usare l'*output* delle stesse come *input* delle trasformazioni di TOTEM per alimentare il modello CALLIOPE. Il processo di aggiornamento del *knowledge graph* sarà realizzato come una componente ETL da richiamare qualora vengano introdotte nuove tecnologie o sia necessario un aggiornamento dell'ontologia esistente sulla base di nuovi utilizzi.

3.2 Studio di metodologie per l'estrazione di impianti fotovoltaici da immagini satellitari

Nell'attività di caratterizzazione del territorio necessaria per la pianificazione energetica, un ruolo fondamentale è ricoperto dalla valutazione del potenziale territoriale disponibile per le nuove installazioni fotovoltaiche. A questa valutazione concorrono naturalmente anche gli impianti già installati, in particolar modo a terra.

Sebbene la localizzazione e la potenza degli impianti sia nota ai gestori, spesso mancano dati liberamente accessibili che permettano di analizzare la distribuzione spaziale delle installazioni attuali e la loro evoluzione nel tempo. Nel contesto italiano, per alcune regioni, una mappatura delle aree su cui insistono impianti fotovoltaici è ricavabile attraverso la mappa regionale di uso del suolo, quando la classificazione è dettagliata fino al quarto livello⁴, ma questo livello di dettaglio non è disponibile per tutto il territorio nazionale; inoltre, la frequenza di aggiornamento delle mappe è variabile e soprattutto non è disponibile il dato di potenza associato agli impianti stessi.

Una mappatura dettagliata dell'occupazione degli impianti fotovoltaici (FV) a terra con potenza superiore a 100 kW è stata svolta durante le precedenti annualità di RdS, grazie ad una collaborazione con GSE [30], con riferimento a dati del 2019; tuttavia, considerato l'attuale contesto fortemente dinamico del fotovoltaico, risulta già obsoleta e necessita ulteriori aggiornamenti. Si è quindi pensato di sviluppare una metodologia per l'aggiornamento periodico e automatico della mappatura degli impianti fotovoltaici a terra in termini di posizione e perimetrazione utilizzando, oltre al *database* esistente di riferimento (mappatura svolta da RSE [30]), anche le informazioni estratte dalle immagini satellitari a media risoluzione spaziale, come le immagini Sentinel-2 messe a disposizione attraverso il programma europeo di osservazione della Terra Copernicus [31].

Infatti, le immagini satellitari sono state utilizzate con successo per l'identificazione e la mappatura di elementi sul territorio anche a scopi di pianificazione energetica. In questo triennio di RdS, si punterà all'impiego di moderne tecniche di analisi delle immagini satellitari, che prevedano l'utilizzo dell'intelligenza artificiale, con l'obiettivo di giungere ad una rappresentazione del territorio più accurata e applicabile a diverse scale di analisi. Limitatamente al caso del monitoraggio di impianti FV a terra, l'integrazione di immagini satellitari e algoritmi di intelligenza artificiale permette il riconoscimento di oggetti in maniera automatica, più efficace e immediata rispetto ad un'identificazione manuale e replicabile nel tempo, garantendo un frequente aggiornamento della mappatura stessa. Inoltre, tali procedure possono essere arricchite da modelli di stima della potenza, in modo da associare alla mappatura degli impianti anche l'informazione della potenza.

Nei paragrafi che seguono viene presentata un'analisi di letteratura riguardo il tema della mappatura di impianti fotovoltaici a partire da immagini tele-rilevate e viene proposta una metodologia applicabile per tale scopo che sarà testata in uno specifico caso studio selezionato nel contesto italiano.

3.2.1 Stato dell'arte sull'utilizzo di immagini da satellite per l'individuazione di impianti fotovoltaici

Negli ultimi anni, ai fini del raggiungimento degli obiettivi di decarbonizzazione, si è assistito ad un sensibile incremento delle installazioni di impianti FV, considerata come fonte di energia rinnovabile fondamentale per la transizione energetica. Nonostante il crescente interesse verso questa tecnologia, in Italia, così come nel resto del mondo, non sono disponibili dati sulla distribuzione degli impianti FV nel territorio, e, quando disponibili, spesso non sono aggiornati, quindi poco utili in un contesto in continua evoluzione [32].

I più recenti tentativi di mappatura di impianti FER (fonti energetiche rinnovabili) a scala nazionale o continentale propongono l'utilizzo di dati derivanti da iniziative di *crowdsourcing* (es. *OpenStreetMap*), *database* esistenti, opportunamente armonizzati, o campagne di misura. Ad esempio, il *database* degli impianti di energia rinnovabile esistenti, in costruzione, o proposti nel

⁴ La classificazione dell'uso del suolo è suddivisa in livelli di dettaglio, fino ad un massimo di 5 livelli, dal più generico (livello 1) al più specifico (livello 5), [126]. Gli impianti FV sono una classe prevista al quarto livello di dettaglio, dei livelli superiori: Territori modellati artificialmente (livello 1), Insediamenti produttivi, commerciali, dei servizi pubblici e privati, delle reti e delle aree infrastrutturali (livello 2); Reti ed aree infrastrutturali stradali e ferroviarie e spazi accessori, aree per grandi impianti di smistamento merci, reti ed aree per la distribuzione idrica e la produzione e il trasporto dell'energia (livello 3).



continente africano (RePP Africa) [33] è stato generato attraverso la raccolta delle coordinate e dei valori di potenza degli impianti idroelettrici, eolici e solari sul territorio africano, estratti da cataloghi e *database* ufficiali e sottoposti a procedure di omogeneizzazione per garantire congruenza tra le diverse fonti. Opportune procedure di elaborazione e validazione sono state applicate anche ai dati disponibili in *OpenStreetMap* (OSM) per generare il primo *dataset* di impianti eolici e fotovoltaici presenti in tutto il mondo [34]. Analogamente, per il Regno Unito, più di 260 mila installazioni FV, inclusi piccoli impianti residenziali, sono state geolocalizzate a partire da estensive campagne di raccolta dati utilizzando come piattaforma OSM [35]. Tali iniziative di *crowdsourcing*, seppure di grande utilità soprattutto per scopi umanitari, si basano però su comunità di persone che volontariamente contribuiscono al mantenimento e all'aggiornamento delle informazioni mappate, per cui l'estensione e la precisione della mappatura non sono uniformi, ma dipendono strettamente dagli sforzi di ciascuna comunità. Inoltre, questi dati spesso contengono errori e necessitano di essere verificati e validati, e includono solo informazioni di tipo geografico, quindi, nel caso specifico della mappatura degli impianti FER, è possibile ricavare il valore di potenza associato all'impianto solo applicando specifici modelli estimativi, che a loro volta possono includere errori. Il telerilevamento (o *Remote Sensing*, RS) è una tecnica di rilievo non invasiva, che permette di studiare e monitorare ciò che è presente sulla superficie terrestre tramite l'osservazione da lontano ed è stato efficacemente utilizzato in passato per risolvere diverse esigenze legate allo sviluppo del sistema FV [36]. In letteratura, si possono trovare diversi esempi in cui dati telerilevati sono stati applicati per stimare il potenziale FV in un territorio [37, 38, 39], per rilevare e monitorare guasti nel sistema FV [40, 41] e per mappare impianti FV a terra o su tetto [42, 43, 44]. In particolare, grazie all'estensione spaziale dei dati, alla frequenza di aggiornamento, alla possibilità di usufruire di molteplici tipologie di acquisizione, molte delle quali gratuitamente, e alla grande disponibilità di algoritmi e tecniche di processamento dati, il telerilevamento da satellite o aereo può essere applicato con successo per l'identificazione di installazioni FV, offrendosi come alternativa ai già citati metodi di raccolta dati da *database* esistenti, rilievi in campo o campagne di *crowdsourcing* [45]. A partire dai primi studi sul tema [46, 47], si possono identificare due principali metodi utili alla mappatura di impianti FV: 1) i cosiddetti approcci fisicamente basati, a partire da immagini iperspettrali e 2) l'applicazione di algoritmi di *Machine Learning* (ML) ad immagini multispettrali. Gli approcci fisicamente basati sfruttano le caratteristiche di riflettanza dei pannelli FV nelle diverse bande dello spettro elettromagnetico, per identificare l'impianto rispetto al *background* circostante. La composizione a strati dei moduli FV genera una firma spettrale caratterizzata da bassi valori di riflettanza nelle lunghezze d'onda del visibile (tra 400 nm e 700 nm), una rapida crescita della riflettanza tra 900 nm e 1150 nm e due forti buche di assorbimento intorno a 1730 nm e 2200 nm [47]. Questi dettagli però possono essere rilevati esclusivamente attraverso il telerilevamento iperspettrale, cioè l'acquisizione di immagini satellitari o aeree composte da numerose (> 100) bande spettrali per coprire un vasto range di lunghezze d'onda dello spettro elettromagnetico tra 400 nm e 2500 nm. In [48], gli autori hanno stimato l'area occupata da impianti FV tramite immagini acquisite dal sensore DESIS (*DLR Earth Sensing Imaging Spectrometer*), montato sulla Stazione Spaziale Internazionale (ISS) e in grado di acquisire fino a 235 bande nell'intervallo di lunghezza d'onda compreso tra 400 nm e 1000 nm, con risoluzione spettrale e spaziale rispettivamente pari a 2,35 nm e 30 m. In un recente studio [45], immagini iperspettrali aeree del sensore AVIRIS-NG (*Airborne Visible InfraRed Imaging Spectrometer-Next Generation*) e satellitari della missione PRISMA (PRecursore IperSpettrale della Missione Applicativa), con risoluzione spaziale rispettivamente uguale a 5,3 m e 30 m, sono invece state utilizzate per rilevare installazioni FV di potenza superiore a 10 kW in Germania, con accuratezze variabili dal 65% all'88%. Sebbene questi studi abbiano dimostrato l'efficacia dell'approccio fisicamente basato per il riconoscimento di impianti FV, il metodo necessita di conoscere a priori le caratteristiche spettrali dei materiali che compongono i moduli solari e l'ambiente in cui sono posizionati. La varietà di tipologie degli impianti FV esistenti e l'interesse nella ricerca verso nuovi materiali sempre più performanti richiedono, quindi, aggiornamenti costanti delle firme spettrali tipiche dei pannelli solari, per garantirne un corretto riconoscimento;

inoltre, con questo metodo, si originano errori di mappatura dovuti alla confusione tra gli impianti e le superfici con proprietà spettrali simili, come pellicole utilizzate in agricoltura, coperture in polietilene ed erba sintetica utilizzata nei campi sportivi. La diffusione di questo tipo di approccio, infine, è limitata dalla disponibilità e facilità di accesso a dati derivanti da missioni iperspettrali, come PRISMA, missione italiana che garantisce l'acquisizione di 240 bande spettrali nelle lunghezze d'onda del visibile (VIS), vicino infrarosso (NIR) e corto infrarosso (SWIR), lanciata solo nel 2019 ma che si concluderà già a marzo 2024 [49]. Il secondo metodo per identificare impianti FV prevede l'utilizzo di algoritmi di ML accoppiati con immagini tele-rilevate multispettrali. Le immagini multispettrali si differenziano dalle immagini iperspettrali per minor numero di bande spettrali (circa una decina) e risoluzione spettrale (tra 10 nm e 200 nm), ma garantiscono acquisizioni con maggior tempo di rivisitazione (anche giornaliero) e maggior risoluzione spaziale (fino a pochi centimetri in alcuni casi). Queste caratteristiche, in aggiunta al numero sempre crescente di sensori disponibili e alla facilità di accesso ai dati, hanno permesso la diffusione dell'uso delle immagini multispettrali in diversi campi di applicazione, soprattutto legati all'osservazione in continuo dei processi in atto sulla superficie e atmosfera terrestre e al monitoraggio delle risorse naturali [50, 51, 52, 53]. Inoltre, i recenti sviluppi di tecniche di analisi di immagini, tra cui algoritmi di ML e *deep learning* (DL), consentono di estrarre in modo rapido ed efficiente informazioni dalle immagini telerilevate [54]. Rispetto al tema dell'identificazione di impianti FV, gli studi disponibili in letteratura si differenziano principalmente in funzione della risoluzione spettrale e spaziale delle immagini tele-rilevate oggetto dell'analisi e della tecnica di processamento dei dati, soprattutto classificazione con algoritmi ML o estrazione di oggetti tramite modelli di DL. Riguardo la risoluzione spettrale e spaziale delle immagini, si possono distinguere studi basati sull'utilizzo di immagini multibanda a media risoluzione spaziale (tra 10 m e 30 m) [55, 56, 57, 58, 59, 60, 61, 62, 63, 64] e studi che sfruttano immagini a colori naturali (RGB) ad altissima risoluzione spaziale (< 1 m) [54, 65, 66, 42, 67]. Le immagini multibanda comprendono acquisizioni in diverse bande dello spettro elettromagnetico, principalmente nelle lunghezze d'onda del VIS e NIR, e offrono il vantaggio di poter sfruttare anche le caratteristiche spettrali degli oggetti per individuare i moduli FV in un'area di studio. Ad esempio, le proprietà spettrali degli impianti FV hanno permesso di identificare le aree occupate da grossi impianti di potenza superiore a 1 MW nella regione Kanto in Giappone, caratterizzata da una forte varietà di paesaggi, tra cui aree urbane, montuose e grandi foreste [64]. Per migliorare il riconoscimento degli impianti FV, gli autori suggeriscono di combinare le diverse bande spettrali attraverso semplici formule algebriche per ottenere indici spettrali [58, 62]. Tali indici infatti permettono di enfatizzare le proprietà spettrali degli oggetti e facilitarne l'identificazione. I principali indici spettrali utilizzati in letteratura per l'estrazione di moduli FV da immagini telerilevate sono elencati in Tabella 3.1. Si evidenzia una prevalenza verso l'uso di indici di vegetazione (i.e. BSI, EVI, NDBI, NDVI), indici per valutare il contenuto di acqua presente in una superficie (i.e. LSWI, NDWI, mNDWI, NDSI) e indici per la caratterizzazione di un ambiente costruito (i.e. BUI, NDBI). Gli impianti FV a terra, infatti, generalmente restituiscono bassi valori per gli indici di vegetazione e di acqua e alti valori per gli indici del costruito, a differenza del *background* in cui sono inseriti (tipicamente campi agricoli o terreni con vegetazione rada) da cui quindi possono essere distinti.

Tabella 3.1 – Principali indici spettrali utilizzati in letteratura per il riconoscimento di impianti FV a terra.

Indice	Descrizione	Formula	Riferimento
BSI	Bare Soil Index	$\frac{(Red + SWIR2) - (NIR + Blue)}{(Red + SWIR2) + (NIR + Blue)}$	[68]
BUI	Built-Up Index	$NDBI - NDVI$	[69]
EVI	Enhanced Vegetation Index	$G \times \frac{NIR - Red}{NIR + c_1 \times Red - c_2 \times Blue + L}$	[70]
LSWI	Land Surface Water Index	$\frac{NIR - SWIR2}{Green + SWIR2}$	[71]

Indice	Descrizione	Formula	Riferimento
NDBI	Normalized Difference Built-up Index	$\frac{SWIR1 - NIR}{SWIR1 + NIR}$	[72]
NDTI	Normalized Difference Tillage Index	$\frac{SWIR1 - SWIR2}{SWIR1 + SWIR2}$	[73]
NDVI	Normalized Difference Vegetation Index	$\frac{NIR - Red}{NIR + Red}$	[74]
NDWI	Normalized Difference Water Index	$\frac{Green - NIR}{Green + NIR}$	[75]
mNDWI	Modified Normalized Difference Water Index	$\frac{Green - SWIR1}{Green + SWIR1}$	[76]
NDSI	Normalized Difference Snow Index	$\frac{Green - SWIR2}{Green + SWIR2}$	[77]
PAROLE CHIAVE: - Blue, Green, Red, NIR, SWIR ₁ , SWIR ₂ : bande spettrali - G, C ₁ , C ₂ , L: coefficienti definiti in funzione delle immagini satellitari			

Inoltre, alcuni studi dimostrano di poter raggiungere accuratezze ancora maggiori per l'estrazione di impianti FV, accoppiando bande e indici spettrali con altri tipi di dati, quali le variabili di *textures* [55, 57], informazioni geometriche [78, 79], o immagini satellitari radar, comprese le acquisizioni del satellite Sentinel-1 [56, 59, 60]. Negli studi analizzati, le immagini multibanda sono acquisite dalle costellazioni Landsat e Sentinel-2, caratterizzate da media risoluzione spaziale e per questo applicate soprattutto all'identificazione di impianti FV a terra, in un'area di studio estesa, generalmente corrispondente ad una nazione. Inoltre, queste costellazioni mettono a disposizione serie temporali di immagini di media (< 10 anni) o lunga (> 10 anni) durata, tali da poter analizzare l'evoluzione del fenomeno nel tempo e verificare come il territorio sia variato con la penetrazione crescente delle installazioni FV. Ad esempio, [61] dimostra come serie temporali di immagini Landsat siano state impiegate per caratterizzare lo sviluppo degli impianti FV a terra nel nord-ovest della Cina tra il 2007 e il 2019, evidenziando che la conversione verso sistemi FV abbia interessato soprattutto territori desertici o sabbiosi e aree contraddistinte da vegetazione erbosa. Le immagini RGB ad alta risoluzione spaziale invece vengono utilizzate principalmente per il riconoscimento di moduli FV su tetto, in quanto le dimensioni ridotte dei pannelli posizionati su tetto necessitano di acquisizioni di dettaglio, con risoluzione metrica o centimetrica, ma risultano efficaci anche per la mappatura di impianti a terra [54, 66, 80]. Gli studi basati su analisi di immagini RGB non sfruttano le proprietà spettrali delle installazioni FV, bensì le loro caratteristiche geometriche per identificare la forma dei pannelli ed estrarli dal *background* circostante. In [65], la posizione esatta dei moduli FV presenti su tetto viene individuata a partire da immagini RGB con risoluzione spaziale pari a 12 cm e viene fornita anche una stima della potenza generata dagli stessi. L'elevata risoluzione spaziale di questo tipo di immagini però preclude il loro utilizzo su larga scala, come dimostrato dagli studi qui considerati, che riportano analisi su un'area di interesse limitata, generalmente corrispondente ad una città [54, 42, 65]. Per quanta riguarda le tecniche di processamento delle immagini, dall'analisi di letteratura si evidenzia una preferenza verso l'uso di algoritmi di classificazione con ML e di segmentazione con DL. Gli algoritmi di classificazione ML, tra cui il più diffuso risulta l'algoritmo *Random Forest* (RF)

[81], sono tecniche di analisi dati che permettono di attribuire una specifica classe ad un *pixel* sull'immagine, combinando informazioni ricavate autonomamente dalle immagini stesse. Nel caso del riconoscimento degli impianti FV, algoritmi di ML sono stati efficacemente applicati ad immagini satellitari multibanda o RGB, a media o alta risoluzione, per individuare moduli FV con accuratezze fino a 98% [55, 60, 61, 62, 79]. I metodi DL, invece, sono considerati un sottogruppo degli algoritmi ML, basati sull'uso di reti neurali artificiali, tra cui le più usate ai fini dell'estrazione di impianti FV risultano essere le reti neurali convoluzionali (CNN), implementate per la segmentazione semantica [63, 82, 83]. Tuttavia, gli autori evidenziano tra i principali limiti di questi metodi, la necessità di disporre di un elevatissimo numero di dati in *input* per la calibrazione del modello, la difficoltà di generalizzare ed estendere i modelli, spesso troppo legati alle caratteristiche specifiche delle immagini usate in fase di calibrazione, e valori di accuratezza dei modelli in generale alti ma derivati da un forte sbilanciamento tra i *pixel* rappresentativi degli impianti FV e del *background*, tipicamente in numero sensibilmente maggiore [54]. Per risolvere i limiti di questi metodi e aumentare l'efficacia nel riconoscimento degli impianti FV, gli autori propongono anche l'uso combinato di algoritmi di ML e DL: i primi utilizzati per mappare grossolanamente aree occupate dalle installazioni FV e, successivamente, i secondi impiegati per proporre una mappa di dettaglio ed estrarre l'esatta posizione e forma dei moduli FV [78].

Tra gli studi considerati in questa analisi, i più rilevanti ai fini di proporre una mappatura degli impianti FV a terra a scala nazionale, e mantenerla aggiornata, sfruttano immagini satellitari a media risoluzione (e.g. 10 m) della costellazione Sentinel-2 combinate con raffinate reti neurali [80, 58, 63]. In [63], viene descritta la metodologia per individuare la posizione degli impianti FV a terra in Brasile: la mappa con la migliore accuratezza (92%) si raggiunge a partire da immagini multibanda (i.e., RGB + NIR) da Sentinel-2, impiegate come *input* di un modello di segmentazione semantica con architettura U-Net [84]. Lo stesso modello è stato applicato in [80] ad immagini satellitari a media risoluzione per generare un *dataset* degli impianti FV a scala globale, con accuratezza, anche in questo studio, vicina al 90% e che include, oltre alla posizione dei moduli, anche la data di installazione. Analogamente, in [58], acquisizioni multispettrali della costellazione Sentinel-2 vengono impiegate per addestrare un modello di segmentazione semantica e individuare la localizzazione e la data di installazione di impianti fotovoltaici a terra distribuiti in un territorio vasto ed eterogeneo quale l'India.

3.2.2 Proposta metodologica

Gli studi riportati al paragrafo precedente dimostrano l'efficacia dell'utilizzo di immagini telerilevate per la mappatura di impianti FV a terra, ancor più se sfruttate all'interno di una rete neurale finalizzata alla segmentazione di oggetti. Questi risultati hanno portato, in questa annualità RdS, alla definizione di una metodologia per l'aggiornamento automatico della mappatura di impianti FV a terra, a partire da immagini satellitari a media risoluzione acquisite dalla costellazione Sentinel-2 combinate con un modello di segmentazione semantica. La metodologia individuata è rappresentata nel *flowchart* in Figura 3.9 e si compone di tre fasi principali, quali: 1) la preparazione del *dataset*; 2) la generazione del modello; 3) l'applicazione del modello stesso a partire da un nuovo set di acquisizioni.

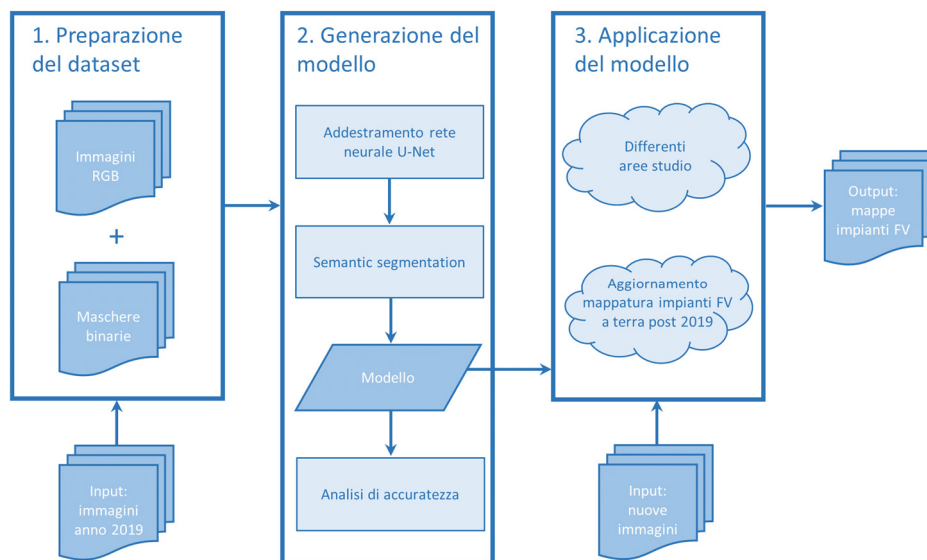


Figura 3.9 – Flowchart di sintesi delle diverse fasi della metodologia.

La prima fase della metodologia consiste nella preparazione del *dataset* utile all'addestramento della rete neurale. Il *dataset* è composto in realtà da due *set* di immagini: da una parte immagini RGB (*Red, Green, Blue*) che rappresentano l'intera area di studio, dall'altra maschere binarie (cioè immagini con *pixel* che possono assumere solo valori pari a 1 o 0) che rappresentano la posizione e la forma degli impianti FV visibili nell'area di studio. A scopo esemplificativo, si riportano in Figura 3.10 una immagine RGB e la rispettiva maschera binaria. Nella maschera, le aree bianche identificano la presenza di un impianto FV a terra (valori dei *pixel* uguali a 1), mentre le aree nere corrispondono al *background* con valori dei *pixel* uguali a 0.

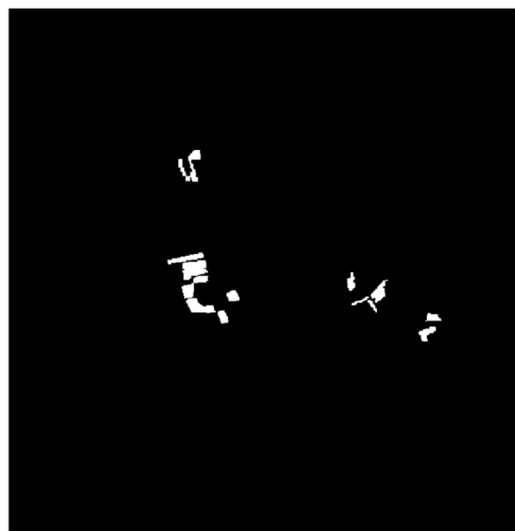


Figura 3.10 – Esempio di immagine RGB (a sinistra) e della rispettiva maschera binaria (a destra). Nella maschera, le porzioni bianche identificano la presenza di un impianto FV.

In questo studio, come immagini RGB sono state scelte immagini telerilevate dalla costellazione di satelliti Sentinel-2, acquisite durante l'anno 2019, con risoluzione spaziale equivalente a 10 m. Allo scopo di rendere minime le procedure di preelaborazione dei dati satellitari, sono state

utilizzate immagini orto-rettificate, geometricamente e atmosfericamente corrette, per cui ciascun *pixel* rappresenta il valore di riflettanza della superficie (Level 2A) [85].

Per la generazione del *dataset*, sono comunque necessarie alcune operazioni di preelaborazione per eliminare dalle immagini disturbi causati dalla presenza di nuvole, ombre, acqua, *pixel* con valori a saturazione, che potrebbero causare errori nelle successive fasi automatiche di classificazione. Tra queste, fondamentale è la rimozione dalle immagini dei *pixel* corrispondenti a piccole nuvole e/o ombre, svolta mediante la mappa di classificazione (*Scene Classification Layer*, SCL) fornita tra i metadati di ciascuna acquisizione, e *pixel* corrispondenti ad aree marine, per evitare che le increspature del mare interferiscano con le elaborazioni.

Le maschere binarie, invece, sono derivate a partire dalla mappatura degli impianti FV a terra svolta da RSE nel 2019 [30]. Ad ogni immagine RGB corrisponde una e una sola maschera, che rappresenta la perimetrazione di eventuali impianti FV presenti nell'immagine RGB stessa. Per favorire l'addestramento della rete neurale [54, 42], sia le immagini RGB che le rispettive maschere degli impianti sono opportunamente suddivise in sotto-blocchi, ciascuno di dimensioni pari a 549x549 *pixel* (equivalente ad una superficie di circa 30 km²).

La seconda fase della metodologia consiste nella generazione del modello di segmentazione semantica degli impianti FV all'interno delle immagini satellitari. Il *dataset*, creato durante la prima fase e costituito da immagini satellitari RGB e dalle rispettive maschere binarie, viene utilizzato per addestrare la rete neurale a riconoscere gli impianti nelle immagini mediante la tecnica della *semantic segmentation*. La *semantic segmentation* è un'attività di *computer vision* che assegna un'etichetta di classe ai *pixel* di un'immagine utilizzando un algoritmo di *deep learning* (DL). La rete neurale scelta per l'addestramento del modello è U-Net [84]; tale architettura è stata implementata con una struttura del tipo *encoder-decoder* composta da:

- un percorso iniziale di *downsampling* (*encoder*) utilizzato dalla rete neurale per ottenere tutte le informazioni derivanti dal contesto e dalle caratteristiche degli oggetti da segmentare all'interno dell'immagine;
- un percorso finale di *upsampling* (*decoder*) che utilizza le informazioni acquisite durante la prima fase per apprendere la posizione precisa degli oggetti e fornire come *output* la loro segmentazione.

In questo studio, la segmentazione semantica viene impiegata per il riconoscimento degli impianti FV all'interno delle immagini satellitari; il modello è dunque addestrato a distinguere gli impianti FV dal *background*.

La fase di *training* del modello viene eseguita su una macchina virtuale equipaggiata di GPU, al fine di ottimizzare i processi e ridurre le tempistiche. Una volta terminata la fase di addestramento, l'efficacia del modello deve essere misurata su un *set* di immagini di *test* mediante l'utilizzo di due metriche tra loro complementari: l'*Intersection over Union*⁵ (IoU) e la *Dice loss*⁶. Entrambe le metriche sono molto usate in letteratura e, di fatto, misurano, in modo diverso, la stessa quantità, ovvero, la sovrapposizione tra la maschera predetta e quella reale, presente fra le maschere del *test set*. Una volta valutata l'accuratezza del modello nell'identificare e segmentare gli impianti FV nelle immagini satellitari, sarà possibile utilizzare il modello per l'inferenza.

Infine, nella terza fase della metodologia, il modello così ottenuto viene applicato per generare nuove mappe di impianti FV a terra. Il modello, addestrato a partire da immagini acquisite durante l'anno 2019 e rappresentativo di una specifica area di studio, è utilizzato per aggiornare in maniera automatica la perimetrazione di nuovi impianti FV a terra installati in anni successivi al 2019,

⁵ La *Intersection over Union* (IoU) è una metrica che serve per quantificare il grado di sovrapposizione tra la maschera originale (quella che segmenta gli impianti) e quella predetta dal modello. Il suo intervallo di valori varia tra 0 e 1: se è vicino a 1, il modello è robusto; viceversa, se è intorno a 0 il modello è pessimo.

⁶ La *Dice loss* è una funzione che deriva dal coefficiente di Dice, molto simile alla IoU ma pari al doppio dell'area di sovrapposizione (fra le maschere reale e predetta) divisa per il numero totale di pixel in entrambe le immagini. La *Dice loss* è definita come la differenza fra uno ed il coefficiente di Dice. L'andamento della *Dice loss* è inverso rispetto a quello della metrica IoU.

oppure per effettuare la mappatura in altre aree di interesse. In queste applicazioni, il modello sfrutta, come dati di ingresso, nuove acquisizioni della costellazione Sentinel-2, con le stesse caratteristiche radiometriche e geometriche (ad esempio: immagini RGB con risoluzione spaziale 10 m), per cui sono necessarie le stesse operazioni di preelaborazione previste per le immagini utilizzate in fase di generazione del *dataset* di addestramento della rete.

3.2.3 Selezione del caso di studio

Secondo i rapporti statistici forniti dal GSE, negli ultimi anni la maggiore penetrazione dei pannelli fotovoltaici installati a terra è stata osservata nelle regioni meridionali e in particolare in Puglia e Basilicata, dove nel 2021 si registra un'incidenza di impianti a terra relativamente molto elevata (rispettivamente, 72% e 64% del totale degli impianti FV della regione) [86]. Nelle regioni settentrionali, al contrario, si osserva una scarsa penetrazione della capacità degli impianti a terra, al di sotto del 10% in Liguria, Lombardia, Valle d'Aosta e nelle province di Trento e Bolzano.

La Puglia è, inoltre, la regione con la maggiore potenza installata e con la dimensione media degli impianti a terra più elevata [86], aspetto peraltro confermato dal rapporto sul consumo di suolo del 2022 redatto da ISPRA [87], in cui tra le varie forme di consumo di suolo viene indagata anche quella relativa all'occupazione degli impianti fotovoltaici a terra.

A partire da queste considerazioni si è deciso, nella prosecuzione delle attività, di testare la metodologia proposta in questo rapporto su un'area corrispondente alla provincia di Lecce, poiché caratterizzata dalla maggiore concentrazione di potenza fotovoltaica nel 2021 e, tra le province del Sud, quella in cui si concentra la quota maggiore di nuovi impianti [86].

In corrispondenza di tale area sono state scaricate e pre-processate 19 immagini a 10 m di risoluzione spaziale, acquisite dal satellite Sentinel-2 durante il corso del 2019 e aventi una copertura nuvolosa inferiore al 10%. L'intera area di studio è contenuta in un unico fotogramma di acquisizione di 110 x 110 km², corrispondente al *tile* T33TYE. In Figura 3.11, sono rappresentate la provincia di Lecce, oggetto dello studio, e la dimensione delle acquisizioni satellitari.

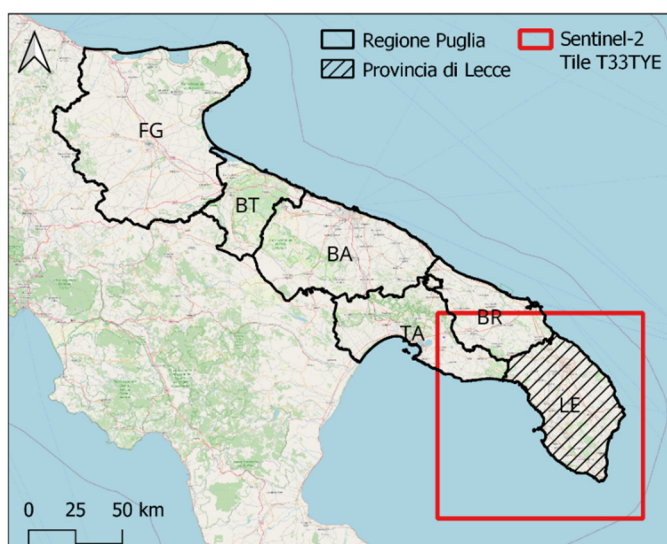


Figura 3.11 – La provincia di Lecce, selezionata come area di studio, e il *tile* T33TYE (in rosso), corrispondente alle acquisizioni della costellazione Sentinel-2. *Coordinate Reference System (CRS):* WGS84/UTM zone 33 N. *Basemap:* ©OpenStreetMap contributors.

4 - MESP: NUOVI SERVIZI A SUPPORTO DELLA PIANIFICAZIONE DI RETI ELETTRICHE

L'evoluzione delle reti energetiche richiede analisi di pianificazione che possano tenere conto delle potenzialità offerte dall'utilizzo di nuove soluzioni di controllo ed identificare loro eventuali criticità. Nel contesto del progetto europeo Platone [88] sono stati proposti metodi per l'analisi di scalabilità e replicabilità (SRA) di soluzioni di flessibilità per le reti elettriche che possono essere di ausilio ad un operatore che voglia monitorare o valutare scenari di evoluzione della propria rete.

La struttura modulare e interoperabile della piattaforma MESP si è mostrata idonea ad accogliere nuove funzioni e algoritmi per l'analisi delle reti, sviluppati da partner del progetto Platone, finalizzati all'analisi di scalabilità e replicabilità delle soluzioni di controllo proposte nei dimostratori del progetto. Nell'attività RdS queste nuove funzionalità sono state integrate nella piattaforma MESP per realizzare ed automatizzare il processo completo di elaborazione necessario alle valutazioni SRA per un insieme di reti di distribuzione di diverse dimensioni e con differenti caratteristiche.

Nella fase di progettazione e implementazione della catena elaborativa, sono stati attentamente considerati alcuni vincoli fondamentali. La soluzione è pensata per garantire scalabilità ed efficienza, adattandosi dinamicamente alle dimensioni e alla complessità delle reti elettriche da esaminare. Inoltre, è stato fatto uso di standard interoperabili, dove possibile, per favorire l'interconnessione e l'integrazione con altri sistemi e tecnologie. Questa attenzione ai vincoli architetturali contribuisce a garantire la flessibilità e la compatibilità della *Multi Energy Semantic Platform*.

In questo capitolo, a valle di un'introduzione generale relativa alle analisi di scalabilità e replicabilità, viene descritta l'integrazione delle nuove funzionalità come microservizi della piattaforma MESP e lo sviluppo nella piattaforma dell'intero processo di analisi SRA per le reti elettriche di distribuzione.

4.1.1 Scalabilità e replicabilità

La **scalabilità** è definita come la capacità di un sistema, rete o processo di aumentare la sua dimensione/ambito/intervallo al fine di soddisfare adeguatamente una crescita della domanda. La **replicabilità** è definita come la capacità di un sistema, rete o processo di essere duplicato in un'altra area geografica o con differenti condizioni al contorno [89].

Tali analisi sono elementi fondamentali per preparare il *deployment* su larga scala di nuove soluzioni di gestione e controllo delle reti energetiche, evitandone il fallimento commerciale e identificando preventivamente barriere tecniche, economiche e regolatorie che potrebbero limitarne lo sviluppo. Pertanto, sono necessarie simulazioni e analisi qualitative che tengano in considerazione i seguenti elementi:

- analisi del contesto tecnico ed economico che caratterizza lo scenario in cui queste soluzioni verranno sviluppate;
- quadro regolatorio;
- accettazione delle nuove soluzioni da parte dei differenti stakeholders del settore energetico.

L'analisi della **scalabilità** mira a rispondere alla domanda "cosa ci si può aspettare se il caso d'uso venisse implementato su una scala più ampia con le stesse condizioni al contorno?". L'implementazione di un caso d'uso su una scala più ampia potrebbe significare l'attuazione di un grado maggiore di intelligenza, una maggiore area di azione, il coinvolgimento di un maggior numero di consumatori, la penetrazione di volumi più elevati di risorse distribuite, etc. A questo proposito, la scalabilità può essere classificata secondo le due dimensioni principali:

- **scalabilità in densità:** questa analisi include la valutazione degli effetti dell'aumento della diffusione di una determinata soluzione nella stessa area del caso pilota di riferimento: ad esempio, un grado maggiore di generazione distribuita nella rete, un grado maggiore di flessibilità dei consumatori, un grado maggiore di automazione della rete, ecc.;
- **scalabilità in dimensione:** questa analisi include la valutazione degli effetti dell'implementazione di una determinata soluzione su una scala più ampia coinvolgendo aree più estese all'interno di una regione o di un paese.

Riferendosi alle reti elettriche, gli aspetti tecnici includono parametri come: dimensione della rete, numero e dimensioni dei consumatori, domanda di picco, numero, dimensioni e posizione delle risorse distribuite, ecc. Aspetti tecnici aggiuntivi da considerare potrebbero includere effetti di saturazione della *hosting capacity* della rete, sovraccarico di linee e trasformatori, fattori di simultaneità, ecc. Gli aspetti economici comprendono principalmente i segnali economici ricevuti da diversi soggetti coinvolti nel caso applicativo (incentivi, tariffe, ecc). Altri aspetti legati agli *stakeholder* riguardano la potenziale accettazione della soluzione da parte dei principali *stakeholder*.

L'analisi della **replicabilità** mira a rispondere alla domanda "cosa ci si può aspettare se il caso d'uso venisse implementato in un altro luogo o posizione, dove si possono trovare condizioni di contesto diverse?" Per analizzare la replicabilità, è necessario considerare diversi scenari e valutare la sensibilità ai principali parametri che costituiscono le condizioni di confine del caso pilota di riferimento. L'analisi della replicabilità ha due dimensioni principali:

- **replicabilità intra-nazionale:** affronta l'analisi della replicabilità della stessa soluzione nello stesso paese che ospita la dimostrazione, ma in situazioni in cui le condizioni tecniche al contorno differiscono (livelli di tensione, tipologia di rete urbana o rurale ecc.) pur mantenendo le medesime condizioni economiche e regolatorie dell'operatore. Queste analisi studiano la variazione dei KPI di progetto al variare del grado di diffusione della soluzione studiata (ad esempio la percentuale di utenti che partecipano a schemi di flessibilità di rete);
- **replicabilità internazionale:** affronta l'analisi della replicazione della stessa soluzione quando tutti i tipi di condizioni di contesto possono differire da quelle del sito di dimostrazione a causa di schemi di regolamentazione e incentivi diversi, situazioni economiche diverse, strategie diverse da parte di decisori politici e aziende di distribuzione, tipi di reti diversi, differenti caratteristiche degli stakeholders, ecc.

Uno schema illustrativo della metodologia SRA è riportato in Figura 4.1.

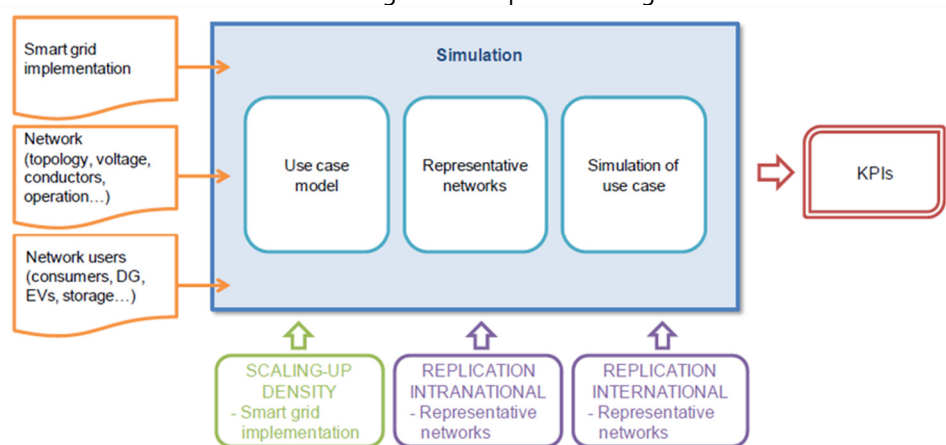


Figura 4.1 - Procedura per l'analisi di scalabilità e replicabilità di un progetto pilota [90].

Per i casi d'uso del progetto Platone selezionati per la valutazione SRA i partner hanno sviluppato strumenti e algoritmi *ad hoc* per poter simulare le strategie avanzate di funzionamento delle reti di distribuzione che consentono agli operatori di sistema di sfruttare i servizi di flessibilità forniti dalle risorse distribuite per risolvere congestioni locali. Questo insieme di strumenti e algoritmi è stato utilizzato per replicare i risultati delle strategie di controllo implementate nelle dimostrazioni del progetto.

Nel contesto della SRA, sono state considerate le due strategie di controllo di una sottorete ritenute più significative per gli operatori: quella in cui viene mantenuto un valore di scambio di potenza desiderato con la rete prevalente e quello in cui questo scambio viene azzerato.

Il caso d'uso "**scambio di potenza desiderato**" offre agli operatori di rete una strategia di controllo che consente ad esempio di regolare nel tempo la quantità di energia importata (o esportata) dalla (verso la) rete principale. Il vincolo sulla potenza di scambio è compensato dalla fornitura di servizi di flessibilità da parte delle fonti di flessibilità locali. Questo caso d'uso potrebbe rappresentare anche la situazione in cui il DSO, al fine di prevenire potenziali congestioni di rete che possono verificarsi durante i giorni di picco, chiede ai gestori delle fonti distribuite di flessibilità di modificare le loro curve di produzione e consumo.

Il caso d'uso "**scambio di potenza zero**" mira a simulare una strategia di controllo che consente al DSO di annullare lo scambio di potenza tra reti MT e BT o MT e AT in determinati intervalli di tempo. In questo caso, per soddisfare la domanda di energia, che in funzionamento normale sarebbe stata fornita agli utenti dalla rete principale, la rete può fare affidamento solo sulla flessibilità delle fonti locali, ovvero può solo consumare l'energia prodotta o immagazzinata nelle batterie locali e può fare affidamento sulla flessibilità dei carichi.

4.1.2 Processo di elaborazione per la SRA

La Figura 4.2 illustra la *pipeline* sviluppata da RSE per eseguire le valutazioni SRA. Gli *input* necessari per eseguire lo schema di elaborazione sono i seguenti:

- profili di carico e generazione (combinazioni di profili di generazione e carico);
- file di topologia della rete: modelli di rete IEC CIM;
- limiti per le capacità delle linee, deviazioni di tensione, limiti del trasformatore di collegamento della sottorete alla rete prevalente [91].

Referendosi alla Figura 4.2, data una rete elettrica (1) con i relativi profili di generazione e carico in cabina primaria (2), vengono **generati N-scenari** che simulano una evoluzione futura della rete, modulando generazione e carico per differenti istanti temporali (3)-(6). Ogni istante temporale di ciascuno scenario viene controllato eseguendo un **Power Flow** (PF) per verificare la presenza di eventuali congestioni di rete (8). In assenza di congestioni la configurazione non ha bisogno di ulteriori analisi, se invece si rilevano violazioni dei limiti di esercizio viene eseguito un calcolo di **Optimal Power Flow** (OPF) (10)-(12) per risolvere, se possibile, la congestione.

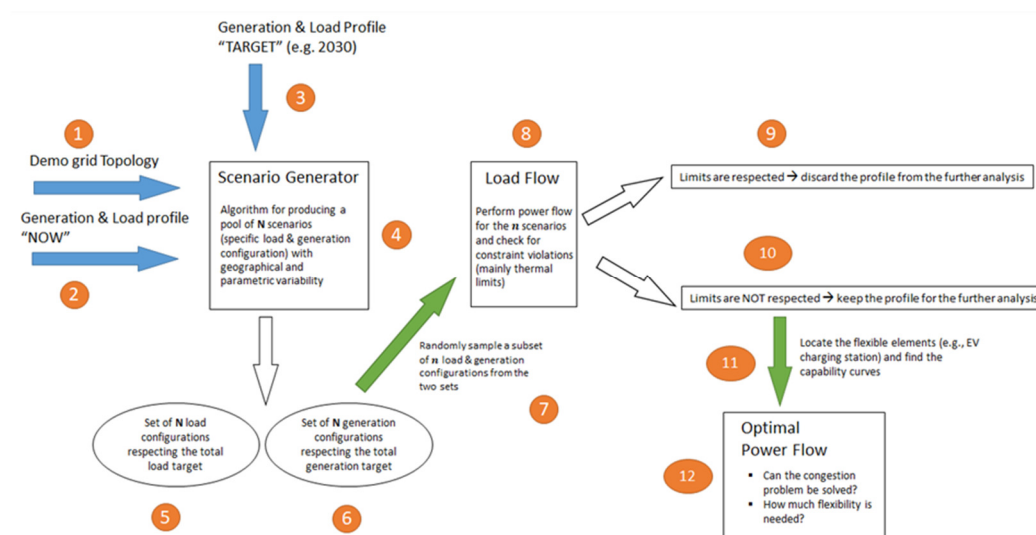


Figura 4.2 - Descrizione generale della pipeline sviluppata per la SRA.

Nelle seguenti sezioni vengono introdotte più in dettaglio le componenti della *pipeline* e la loro integrazione nella piattaforma MESP.

4.1.2.1 Generazione di scenari

L'obiettivo del generatore di scenario è produrre una famiglia di N scenari casuali a partire dai dati dei profili di generazione e carico aggregati. La casualità è intesa sia geograficamente (ad esempio, diversi valori di potenza in nodi diversi) sia parametricamente (diversi valori di potenza in un nodo specifico).

Si prenda per esempio una rete composta da 3 nodi per la quale siano noti due profili giornalieri (Figura 4.3):

- profilo "attuale" che rappresenti un profilo di carico e generazione della rete "ad oggi" (es. 2023),
- profilo "target" che rispecchi l'evoluzione della domanda e della generazione ad un certo periodo futuro (es. 2030). Tale valore è stimato tra un limite inferiore (*targetMIN*) e un limite superiore (*targetMAX*).



Figura 4.3 - Profilo totale attuale (*as is*, in blu) con i profili totali target generati dall'algoritmo (arancione, grigio).

Assumiamo inoltre che, in un dato momento t , i valori totali di carico dei profili "attuale" e "target" siano ad esempio quelli definiti nella Tabella 4.1.

Tabella 4.1 - Carico totale per gli scenari "attuale" e "target".

	As is	Target MIN	Target MAX
$P_{Ltot}(t)$ [MW] al tempo t	200	500	600

La Formula 3.1 illustra un esempio dell'algoritmo sviluppato per calcolare le curve di carico e generazione negli scenari *target*. In particolare, il carico totale "così com'è" P_{Ltot} è dato dalla somma di tutti i carichi presenti. Si riporta un esempio con 3 carichi A, B, C:

$$P_{Ltot}(t) = P_L^A + P_L^B + P_L^C = 50 + 110 + 40 \text{ MW} \quad 3.1$$

Nell'esempio illustrato, la percentuale di aumento di P_{Ltot} dai profili "attuali" ai profili "target" è compresa tra il 150% (200 → 500 MW) e il 200% (200 → 600 MW). Di conseguenza, la variabilità di ogni carico nodale è inclusa nei seguenti intervalli di valori:

$$\begin{aligned} P_L^A &= [75, 150] \text{ MW} \\ P_L^B &= [165, 330] \text{ MW} \\ P_L^C &= [60, 120] \text{ MW} \end{aligned}$$

Si è ipotizzato che la variabilità di questi parametri sia descritta da una distribuzione di probabilità uniforme. Le probabilità associate ad ogni valore di carico vengono utilizzate per eseguire una simulazione Monte Carlo allo scopo di calcolare i corrispondenti valori casuali di P_{Ltot} , come illustrato nella Figura 4.4.

	P_L^A	P_L^B	P_L^C	P_{Ltot}
1	145.3634	220.8928	100.4420	466.6983
2	75.5968	253.4841	105.4548	434.5357
3	81.1613	206.0780	96.9483	384.1876
4	129.7379	245.4418	79.1338	454.3135
5	113.8326	288.3107	108.9068	511.0501
6	85.5956	212.1338	87.4511	385.1805
7	84.1655	218.8771	91.8977	394.9402
8	122.8320	260.6600	70.4207	453.9127
9	100.2135	239.1611	89.5491	428.9238
10	78.5503	309.9603	115.9471	504.4577
11	132.0064	226.7369	62.7801	421.5235
12	79.8207	228.0481	116.1845	424.0532
13	79.4794	273.8412	107.3646	460.6851
14	87.2233	234.2408	103.6155	425.0796
15	79.0018	247.8604	68.0813	394.9434
16	102.3170	317.7481	67.0103	487.0754
17	81.6019	311.3529	93.6508	486.6057
18	137.4236	223.6857	118.1698	479.2791
19	139.8409	271.6768	95.2255	506.7432
20	124.2397	276.9576	94.0511	495.2484
21	124.0177	260.1801	66.1808	450.3786

Figura 4.4 - Applicazione del metodo illustrato per la creazione di un set di scenari *target* per il carico totale P_{Ltot} . Tra le configurazioni generate sono evidenziati in rosso i valori di potenza dei nodi per cui il totale è compreso tra i valori di *target* minimo e massimo.

Ogni insieme di combinazioni di valori di carico nodale rappresenta uno scenario di carico. L'insieme degli scenari di carico che rispettano il vincolo per il carico totale della sottorete, che deve essere compreso tra i valori target minimo e massimo stabiliti, rappresenta la famiglia di scenari di carico utili per l'analisi SRA.

In maniera simile si procede per calcolare gli scenari dei profili di generazione. Dopo aver generato l'insieme degli scenari di carico e generazione, può essere eseguito un campionamento casuale all'interno di questi due insiemi ed estrarre da questi un numero N , selezionato dall'utente, di scenari di carico casuali da utilizzare per l'analisi SRA. Per creare un profilo temporale di analisi, ossia un insieme di scenari corrispondenti a diversi istanti, il processo finora descritto viene ripetuto per tutti gli istanti di tempo desiderati (ad esempio, 24 volte per un profilo orario giornaliero), considerando le condizioni "attuale" e "target".

Il programma "*scenario generator*", sviluppato in collaborazione con i partner del progetto Platone, crea automaticamente diversi scenari che descrivono l'evoluzione prevista delle reti elettriche prese in considerazione, prendendo come *input* i profili attuali misurati presso le sottostazioni secondarie o primarie ed un insieme di informazioni che descrivono l'evoluzione prevista globale delle reti. Queste informazioni sono ottenute esaminando i piani di sviluppo delle reti e mediante interviste con gli operatori di sistema. I parametri di *input* che descrivono l'evoluzione prevista della rete, come ad esempio la percentuale di aumento dei nodi di generazione o l'incremento complessivo dei carichi, sono riassunti in Tabella 4.2.

Tabella 4.2 - Tabella dati di input dello scenario generator.

Variabile	Unità	Descrizione
n_nodes	#	numero di nodi sottesi alla stazione esaminata
Cos		power factor
perc_increase_load	[%]	crescita attesa del carico rispetto al baseline scenario
uncertain_load	[%]	errore associato alla crescita stimata del carico
perc_increase_gen	[%]	crescita attesa della generazione rispetto al baseline scenario
uncertain_gen	[%]	errore associato al tasso di crescita atteso
perc_nodes_gen	[%]	% di nodi connessi ad un generatore nello scenario target
gen_types	labels	tipi di generatori connessi alla rete nello scenario target [default: PV; PV e accumuli]
gen_percs	[%]	percentuale di ciascuna tipologia di generatore (la somma deve essere pari a 100%)
load_types	labels	tipi di carichi connessi alla rete nello scenario target (default: residenziale, EV, fissi, accumulo)
load_percs	[%]	percentuale di ciascuna tipologia di carico (la somma deve essere pari a 100%)
min_contracted_power	[kW]	potenza contrattuale minima considerata nello scenario target
med_contracted_power	[kW]	potenza contrattuale media considerata nello scenario target
max_contracted_power	[kW]	potenza contrattuale massima considerata nello scenario target
perc_min	[%]	% di carico aventi potenza contrattuale minima

4.1.2.2 Power flow e rilevazione delle congestioni

I profili calcolati in 4.1.2.1 per i generatori e i carichi della rete in esame vengono quindi passati alla funzione di calcolo dei flussi di potenza nella rete. I dati necessari per eseguire l'elaborazione sono i seguenti:

- profili di carico e generazione ;
- file di descrizione della topologia della rete;
- limiti delle capacità delle linee, deviazioni di tensione; limiti del trasformatore (parametri tecnici fissati dalla normativa nazionale).

Dopo aver generato i dati di configurazione per ogni caso di analisi, identificato dallo scenario considerato e dall'istante temporale, viene quindi eseguito il *power flow*. Dall'esame dei risultati vengono rilevate le situazioni in cui si verificano congestioni di rete, per le quali risulta utile sfruttare il controllo della flessibilità delle risorse.

Per i casi in cui si è verificata una violazione dei limiti di esercizio della rete viene eseguito il calcolo di *Optimal Power Flow* (OPF) per identificare i nuovi punti di lavoro di carichi e generatori che possono essere assegnati per consentire al sistema di evitare congestioni sfruttando la flessibilità locale.

Le informazioni trasferite all'OPF a valle del calcolo di *power flow* sono le seguenti:

- un file con tutte le informazioni di base necessarie per eseguire l'OPF. Tra questi parametri possiamo elencare, ad esempio, il numero di nodi flessibili nella rete, i limiti di flessibilità;
- un file denominato "warm start" che viene utilizzato dall'OPF per inizializzare l'algoritmo iterativo;
- un file Excel che descrive le caratteristiche della rete, inclusi i profili di generatori e carichi che hanno causato la congestione rilevata dal *power flow*.

4.1.2.3 Optimal Power Flow

Il progetto Platone ha reso disponibile l'algoritmo di *Optimal Power Flow* (OPF) che è stato integrato nella piattaforma MESP ed utilizzato nel processo complessivo di analisi SRA.

Nel problema classico di *Optimal Power Flow* lo scopo è quello di determinare la suddivisione ottima della domanda totale di generazione in un sistema tra i generatori convenzionali presenti, in base ai loro costi operativi. Questa finalità determina la funzione obiettivo del problema di ottimizzazione, soggetta ai seguenti vincoli:

- equilibrio tra la generazione totale e la domanda totale nel sistema;
- flussi massimi di potenza nelle linee;
- limiti operativi della tensione ai nodi del sistema;
- limiti di capacità delle unità di generazione di potenza del sistema;
- limiti di capacità dei flussi attraverso le linee del sistema.

Per ottenere un'esecuzione più rapida dell'OPF, sono stati proposti algoritmi completamente distribuiti. Questo approccio garantisce anche la scalabilità e la modularità dell'algoritmo di OPF, poiché l'algoritmo può essere più facilmente applicato a sistemi di qualsiasi dimensione e con qualsiasi insieme di unità di generazione. Questo significa che l'algoritmo OPF distribuito può essere facilmente modificato per integrare nel sistema unità che non sono disacciate in base ai loro costi operativi, ad esempio carichi flessibili. Inoltre, l'algoritmo può essere facilmente modificato per simulare scenari specifici di funzionamento del sistema, ad esempio per considerare la generazione della potenza desiderata solo da parte di determinate unità o iniezioni di potenza da specifici nodi. Questi adattamenti possono essere eseguiti facilmente grazie alla formulazione nodale del problema e per la natura distribuita dell'algoritmo OPF, senza richiedere modifiche alle grandi strutture dati che descrivono l'intero sistema.

Il problema OPF classico, nella sua formulazione nodale, è stato quindi modificato per includere nel sistema i carichi flessibili e simulare il funzionamento di rete desiderato per i due casi d'uso considerati, ovvero uno scambio di potenza prefissato oppure nullo con la rete esterna.

Per considerare la presenza di carichi flessibili nelle reti, vengono introdotte nella formulazione del problema OPF le corrispondenti variabili di potenza. Queste variabili sono incluse nella funzione obiettivo, con i relativi fattori di costo, così come nei vincoli del bilancio energetico. Vengono inoltre introdotti ulteriori vincoli per descrivere i limiti di azione ammessi per queste variabili nella risoluzione del problema. Il limite inferiore di potenza dei carichi flessibili è determinato in termini percentuali rispetto al limite superiore. Va notato che i fattori di costo per l'utilizzo della flessibilità dei carichi, utilizzati nella funzione obiettivo del problema OPF, non devono necessariamente avere un'interpretazione monetaria; possono essere utilizzati anche come fattori di priorità o penalità, per forzare o evitare l'utilizzo dei carichi flessibili.

Per simulare il caso d'uso della rete in isola virtuale (scambio di potenza nullo), i parametri corrispondenti all'unità di generazione che rappresenta la rete esterna vengono impostati con il limite inferiore di questa variabile di potenza a zero ed assegnando al relativo fattore di costo un valore elevato rispetto a tutti gli altri. In questo modo, questa unità di generazione di potenza è considerata "costosa" nel problema OPF e l'algoritmo le assegna il suo valore minimo, che è zero. Per simulare un caso d'uso di iniezione di potenza desiderata nella rete esterna, la formulazione del problema OPF viene modificata assegnando un valore negativo di potenza all'unità di generazione che rappresenta la rete esterna. Il vincolo del bilancio energetico al nodo della rete esterna viene modificato di conseguenza. In particolare, il limite inferiore di potenza per la rete esterna viene impostato uguale al valore di iniezione desiderato assegnando a questa unità di generazione, nella funzione obiettivo, un fattore di costo elevato rispetto agli altri. Pertanto, questa variabile di potenza viene così forzata ad assumere il suo valore minimo, uguale alla quantità di iniezione di potenza desiderata dalla sottorete nella rete esterna.

4.1.3 Installazione nella piattaforma MESP

Seguendo la logica architetturale della piattaforma MESP, i servizi necessari per eseguire le analisi SRA sono stati implementati come *container*. Tale scelta garantisce una maggiore portabilità del *software* e offre la possibilità di orchestrazione dei servizi per mezzo di strumenti come *kubernetes*. Il seguente diagramma di sequenza (Figura 4.5) illustra l'ordine delle operazioni e gli scambi informativi tra le componenti SRA per ottenere una simulazione completa nella MESP.

Nel diagramma di sequenza sono presenti le componenti per eseguire la SRA:

- Scenarios Generator;
- Power Flow & congestion check;
- OPF.

Nel diagramma sono anche rappresentate due componenti della piattaforma MESP (*Notebook* e *Data Exchange*) non specifiche della SRA ma necessarie per coordinare la *pipeline* di elaborazione. Ciascun microservizio all'interno della piattaforma MESP si interfaccia con gli altri utilizzando il *message broker* RabbitMQ e scambiando messaggi conformi allo standard CIM IEC 61968-100. Per evitare scambi di messaggi troppo voluminosi è stata introdotta un'area di scambio (Redis) dove vengono salvati i *file* necessari alle componenti software per eseguire i loro compiti. In questo modo gli *output* delle singole funzioni che compongono la SRA vengono salvate nell'area di scambio e possono essere recuperate dalla successiva funzione come *input*.

Di seguito sono specificate le tipologie di messaggi scambiati tra i differenti microservizi.

La componente di **generazione degli scenari** richiede i seguenti file di *input*:

- file di configurazione Json che descrive le caratteristiche della rete e la previsione;
- il carico totale attuale;
- la generazione totale attuale.

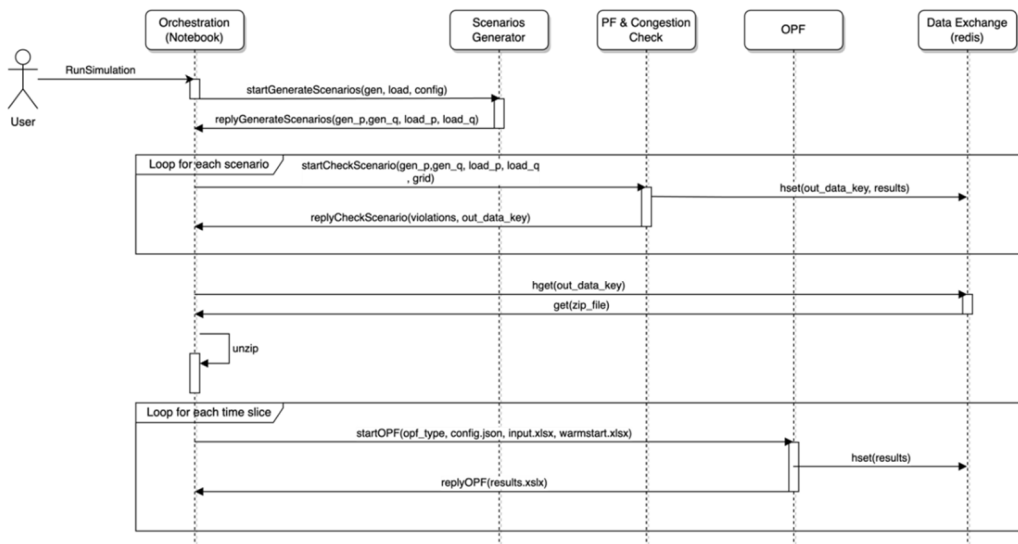


Figura 4-5 - Diagramma di sequenza che descrive gli scambi informativi tra i microservizi della SRA nella piattaforma MESP.

In Figura 4.6 viene mostrata la sintassi dei messaggi scambiati in *input* e *output* del microservizio di generazione di scenari.

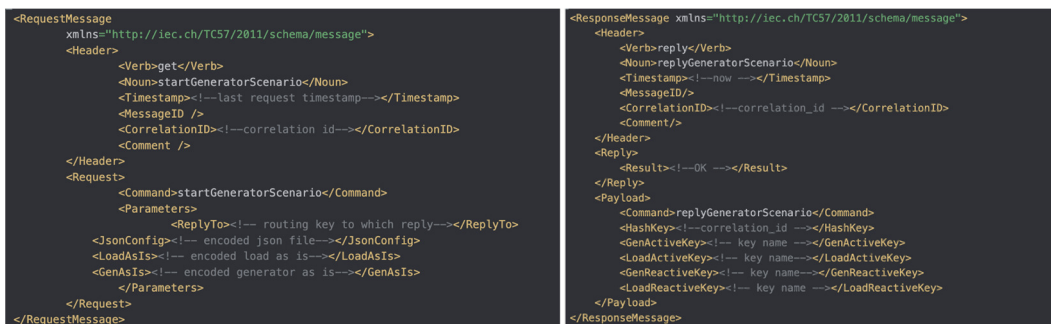


Figura 4.6 – Messaggi XML di *input* e di *output* per il microservizio di generazione scenari.

L'*output* è costituito da quattro *file* csv, contenenti rispettivamente i risultati di potenza attiva e reattiva di generazione e carico per ciascuno scenario. Questi sono i valori che dovranno essere usati come *input*, assieme alla topologia della rete, per il calcolo del *power flow*.

Il microservizio che effettua il *power flow* prende in carico una collezione di scenari ed esegue per ogni scenario e per ogni istante temporale il **power flow**. Viene eseguito anche il **controllo automatico di violazioni** dei limiti su tensioni e correnti nella rete. Tale servizio si occupa di fornire in risposta l'*output* del *power flow* e i file di *input* necessari per eseguire il successivo OPF per i casi in cui si sono verificate congestioni di rete. È stato adottato un risolutore di *power flow* basato su Matpower e Octave, garantendo un approccio *open source* nell'implementazione. Questo consente l'accesso e la condivisione del codice sorgente, promuovendo la trasparenza e la collaborazione nella comunità. Il risolutore, basato su tali tecnologie, offre in modo affidabile le funzionalità fondamentali necessarie per l'esecuzione di un *power flow*.

I dati di input necessari sono:

- parametri di configurazione dell'algoritmo;
- *file* di potenza attiva e reattiva per ciascun nodo di rete e ciascuno scenario.

Una volta eseguito il controllo delle violazioni viene pubblicato un messaggio che include:

- *file* delle violazioni;
- *file* zip con i risultati di *power flow* e i dati necessari per eseguire l'OPF.

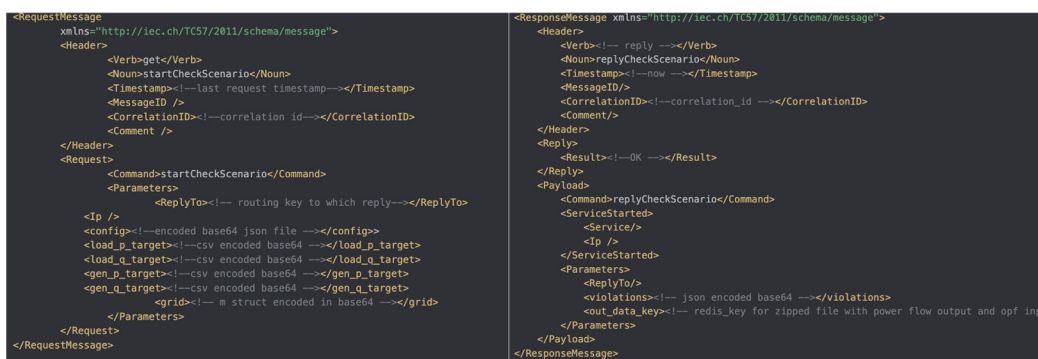


Figura 4.7 - Messaggi XML di input e di output per il microservizio di *power flow* e controllo delle congestioni.

Il microservizio che effettua l'OPF per la risoluzione delle congestioni riceve invece i seguenti *file* di *input*:

- il tipo di OPF da eseguire ("scambio di potenza assegnato" o "scambio di potenza zero");
- il *file* di *input*;
- il *file* di *warmstart* per ottimizzare le condizioni iniziali dell'OPF;
- i file con configurazioni aggiuntive per l'esecuzione dell'OPF.

Una volta eseguito l'OPF viene pubblicato il messaggio con il *file* dei risultati.



Figura 4.8 . Messaggi XML di *input* e di *output* per il microservizio di *Optimal Power Flow*.

Nell'esecuzione delle analisi SRA del progetto Platone utilizzando gli strumenti di elaborazione predisposti nella piattaforma MESP, l'algoritmo di OPF ha presentato criticità dovute all'oneroso carico computazionale che richiederebbe buone capacità di calcolo parallelo distribuito. Infatti, tale algoritmo si appoggia su librerie mpi che permettono di distribuire il calcolo su cluster HPC (*High Performance Computing*). Nel cloud privato di RSE in cui è installata la piattaforma MESP non è disponibile un *cluster* con queste caratteristiche ma è possibile lavorare su singolo nodo per il singolo OPF. In caso di più richieste di OPF contemporanee la scalabilità è ottenuta distribuendo i differenti *task* su più macchine virtuali. La soluzione implementata è comunque compatibile con una eventuale futura integrazione di *cluster* HPC.

La distribuzione del calcolo su differenti macchine virtuali e la conseguente riduzione dei tempi di simulazione ha permesso di ricalibrare più velocemente i dati di *input* in modo da trovare le

configurazioni migliori per garantire la convergenza dell'OPF. Infatti, l'esperienza nell'attività SRA di Platone ha mostrato come la convergenza dell'OPF richieda un *tuning* molto delicato dei parametri di configurazione in ingresso, evidenziando aspetti di miglioramento dell'algoritmo per superare queste criticità.

Le nuove componenti software sviluppate sono state installate nella piattaforma MESP secondo lo schema presentato in Figura 4.9. Utilizzando il cloud privato *Openstack* di RSE è stata istanziata una macchina virtuale con i servizi della piattaforma MESP. Sfruttando le potenzialità del *cloud computing* sono state create alcune macchine virtuali come *worker* sulle quali distribuire il calcolo. I servizi più onerosi dal punto di vista del calcolo (*power flow* e *optimal power flow*) sono stati distribuiti e avviati all'interno dei *worker*. La parte destra della figura mostra invece la connessione tra i differenti microservizi istanziati e rispecchia il *deployment* della piattaforma MESP applicata all'analisi SRA del progetto Platone.

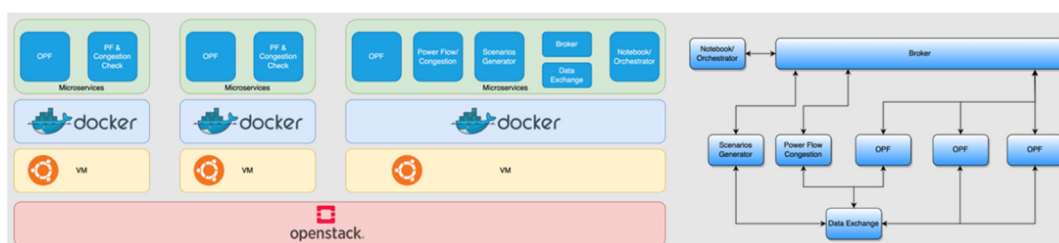


Figura 4.9 - Implementazione dell'architettura software della MEPS per il caso d'uso SRA. A sinistra l'installazione sul *cloud* Openstack, a destra i container con le relative interazioni.

L'utente utilizza il *notebook* (Figura 4.10) per coordinare l'esecuzione dei *task* di ciascun microservizio necessario per l'analisi. In sequenza viene richiesta la generazione di scenari in base ai parametri passati nel file di configurazione e al profilo di generazione e carico aggregato *attuale* (percentuale di crescita dei nodi di generazione sulla rete, aumento del carico, etc.). Una volta creati gli scenari viene passata la rete con i relativi profili di generazione e carico al servizio che esegue il *power flow* e rilevate le eventuali violazioni ai limiti di corrente e tensione. Se vengono rilevate violazioni la risposta del microservizio conterrà l'*output* del *power flow* e l'*input* per l'algoritmo di risoluzione dell'*optimal power flow*, tutto in un file compresso. A questo punto, per ogni istante temporale per cui si è verificata una congestione di rete verrà eseguito l'*optimal power flow* e i risultati delle simulazioni salvati sul *notebook* Jupyter per ulteriori analisi.

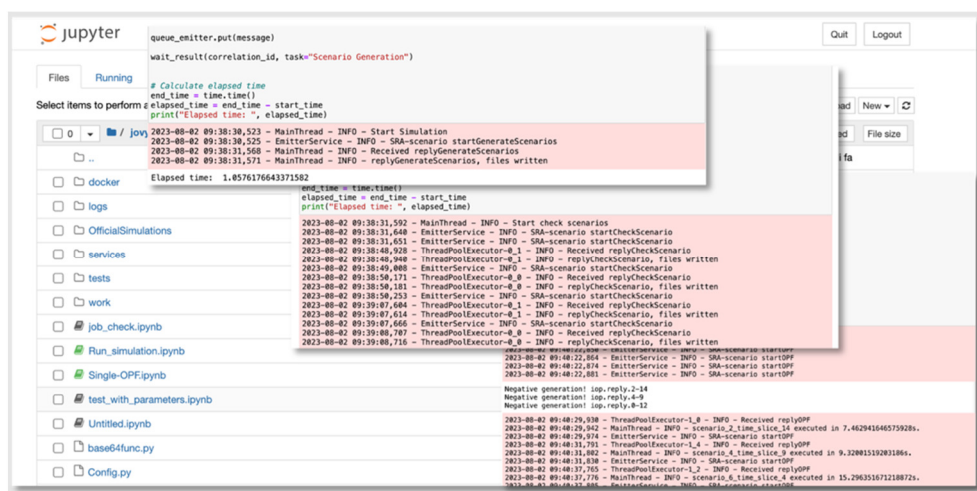


Figura 4.10 - Jupyter Notebook utilizzato per implementare la *pipeline* di esecuzione di simulazioni distribuite con la piattaforma MESP.



L'attività svolta ha permesso quindi di integrare nella piattaforma MESP la metodologia SRA sviluppata nel contesto del progetto Platone. Tale realizzazione è stata utilizzata per eseguire le analisi richieste nel contesto del progetto europeo ed ha arricchito la piattaforma di nuovi servizi, disponibili agli utilizzatori della piattaforma MESP che abbiano necessità di utilizzare singole funzioni o vogliano eseguire analisi SRA per valutare scenari di evoluzione delle proprie reti.



5 - STIMA DEL POWER FLOW SU RETI ELETTRICHE TRAMITE TECNICHE DI GRAPH ANALYSIS E GRAPH MACHINE LEARNING

In questa attività, svolta in collaborazione con il Politecnico di Milano [92], è stata approfondita la possibilità di utilizzo di tecniche basate sui grafi (*Graph Analysis* e *Graph Machine Learning*) per la **stima dei valori di tensione nei nodi di reti di distribuzione elettriche**. Lo scopo finale dell'attività triennale è la valutazione delle prestazioni raggiungibili con tali tecniche per la stima dello stato di reti energetiche in casi non facilmente risolvibili tramite tecniche basate su modelli analitici, ad esempio nel caso di reti estese e multivettore. Un approccio di questo tipo ha infatti il vantaggio di sfruttare i moderni e performanti database basati sui grafi per la gestione dei dati di rete e, allo stesso tempo, di utilizzare le tecniche di calcolo specifiche del mondo dei grafi per la risoluzione di algoritmi di analisi delle reti. Durante la scorsa annualità [4] sono stati individuati e validati i dati che saranno usati come *input*: i file IEC CIM relativi a 20 isole topologiche della rete di distribuzione MT di Milano, i profili normalizzati di potenza associati ai nodi topologici e i relativi valori di tensione, ottenuti in seguito al calcolo di *power flow*. Inoltre, è stato effettuato uno studio dello stato dell'arte per comprendere come le tecniche di *Graph Machine Learning* siano state applicate nel settore elettro-energetico.

La prosecuzione delle attività ha riguardato l'effettivo sviluppo di un insieme di modelli di *machine learning* per la stima dei valori di tensione $V_i(t)$ per ciascun nodo topologico delle reti senza l'ausilio di tecniche analitiche, avendo in ingresso il *Knowledge Graph* (KG) rappresentante ciascuna rete in uno specifico istante temporale t . La scelta di focalizzarsi sulla stima delle tensioni ai nodi riguarda la possibilità di valutare in anticipo eventuali scostamenti della tensione rispetto ai limiti normativi per abilitare possibili interventi correttivi finalizzati alla regolazione della tensione. Allo stesso modo, una volta affrontata questa riduzione del problema, si potrà considerare l'applicazione di tecniche di *Graph Machine Learning* similari per la stima di altre grandezze di stato della rete, come le correnti sui rami, completando il compito di stima del *power flow*.

5.1 Selezione del *framework* tecnico di sviluppo

Nella scorsa annualità di Ricerca di Sistema, sono state esaminate le caratteristiche dei *framework* disponibili per lo sviluppo di modelli di *Graph Machine Learning*. Il confronto ha permesso di individuare le tre piattaforme più promettenti: **Deep Graph Library** (DGL) [93], **PyTorch Geometric** (PyG) [94] e **StellarGraph** [95].

Nella prosecuzione delle attività, tali piattaforme sono state valutate sperimentalmente rispetto a un **compito benchmark di classificazione**, considerando come *input* i seguenti *dataset* noti nello stato dell'arte: CORA [96], CiteSeer [97] e PubMed [98]. I risultati del confronto hanno mostrato che StellarGraph, nonostante sia più agevole da utilizzare rispetto agli altri *framework* considerati, non riesca mai a raggiungere i migliori valori di accuratezza. Inoltre, presenta tempi di inferenza sensibilmente maggiori rispetto a quanto ottenuto con gli altri ambienti. Queste motivazioni hanno portato a una prima esclusione di StellarGraph per l'applicazione al presente progetto.

Come seconda comparazione, il compito di stima dei valori di tensione nei nodi di reti elettriche è stato parzialmente ricondotto a un compito di classificazione: questo è stato ottenuto suddividendo i valori di tensione calcolati per ciascun nodo in un numero limitato di bande, in modo da ottenere un dominio discreto e limitato. In questo modo, il compito *benchmark* di classificazione è stato valutato anche considerando i *dataset* di *input* che saranno utilizzati per l'addestramento dei modelli di *machine learning* per la stima della tensione ai nodi. Dai risultati ottenuti, si è notato che il *framework* PyTorch Geometric presenta una maggior aggressività e stabilità nel processo d'addestramento, ma fallisce nello scavalcare zone piatte presenti nello spazio di ottimizzazione dei parametri del modello. In modo contrario, *Deep Graph Library* è in

La **classificazione di nodi** all'interno di un grafo è un compito che costituisce la base teorica e applicativa per compiti più complessi come quelli di completamento del grafo. Il compito di classificazione di nodi si pone l'obiettivo di **assegnare a ciascun nodo di un grafo un'etichetta**, ossia determinare, data una caratteristica categorica (a dominio discreto e limitato), quale valore di tale caratteristica sia più appropriato assegnare in base agli attributi di tale nodo, dei nodi vicini e della topologia del vicinato. In letteratura, la classificazione di nodi è ampiamente studiata e profilata grazie all'impiego di **dataset noti** per le loro caratteristiche in grado di metterne alla prova vari aspetti. Esempi eminenti di tali dataset sono le **reti di citazioni** (citation networks), ossia grafi rappresentanti delle pubblicazioni scientifiche (nodi del grafo), suddivise in differenti categorie (ad esempio, per area disciplinare d'afferenza), in cui ogni arco mette in relazione due pubblicazioni in cui una cita l'altra come fonte.



grado di scavalcare tali zone e di individuare una migliore ottimizzazione dei parametri dei modelli, il che permette di guadagnare un ulteriore 10% di accuratezza. Per le motivazioni indicate, **Deep Graph Library** è il *framework* tecnologico scelto per il proseguo della presente ricerca.

5.2 Traduzione dei dati di input in formato graphML

I modelli di *Graph Machine Learning* (applicati in questa attività per la stima dei valori di tensione) agiscono principalmente in modo **convolutivo**. Con ciò si intende che, per ogni punto modellato, le **caratteristiche del dataset** da considerare nella modellazione di tale punto **provengono dal vicinato del punto stesso**, dato un criterio di definizione di tale vicinato.

I dati di *input* per l'attività si riferiscono a reti elettriche di distribuzione in media tensione in ambito urbano, descritte secondo il *Common Information Model* (CIM), standardizzato principalmente all'interno degli standard IEC 61970 e IEC 61968, e serializzate in formato RDF/XML.

Secondo il *Resource Description Framework* (RDF), ogni entità rilevante per il problema è rappresentata da un nodo nel grafo, indipendentemente dalla sua funzione: le entità primarie su cui il problema opera, cioè gli elementi della rete elettrica, sono nodi del grafo, ma anche le loro reciproche relazioni e gli attributi sono rappresentate da nodi. Questo richiede, per poter raccogliere sufficiente informazione nel modellare un'entità o un suo attributo, di estendere il vicinato preso in considerazione ad un numero di nodi che diventa ingestibile in termini pratici, portando i tempi di addestramento e inferenza del modello a crescere fuori dai limiti utili a qualsiasi applicazione. È quindi necessario derivare, da un *Knowledge Graph* di partenza, una rappresentazione più densa, che tratti come nodi del grafo risultante solo le entità oggetto di modellazione, trasformando le relazioni in archi del grafo risultante e integrando i nodi di tipo attributo direttamente come attributi dei nodi e degli archi così determinati.

Durante il precedente triennio di Ricerca di Sistema (2019-2021) è stata implementata una trasformazione **CIM2GraphML** [99], in grado di convertire una rete IEC CIM in graphML, formato utile per l'applicazione di tecniche di *graph analysis* e *graph machine learning*. Tuttavia, non è generalmente determinabile un'unica trasformazione che operi in questi termini, ma risulta necessario, per ogni scenario applicativo, determinare le trasformazioni di interesse per facilitare il compito di apprendimento. Nei seguenti paragrafi sono descritte le tre diverse trasformazioni delle reti in input in formato graphML, implementate per l'attività in oggetto. Esse si differenziano per la modalità di trattamento delle entità CIM di tipo *TopologicalNode*, che rappresentano i punti per cui il modello dovrà determinare i valori di tensione.

5.2.1 Trasformazione CIM2GraphML

Un requisito fondamentale per lo sviluppo di un modello in grado di stimare i valori delle tensioni nei nodi di una rete elettrica consiste nell'avere un **grafo risultante con nodi che rappresentino i nodi topologici della rete elettrica in modo esplicito**. Questo non è garantito per nessuna delle trasformazioni già sviluppate in precedenza [99]. Tuttavia, la forte modularità della procedura con cui la famiglia di trasformazioni è stata progettata garantisce di poter sviluppare una sequenza di azioni di trasformazione che permette di ottenere un grafo con le seguenti caratteristiche:

- i nodi corrispondono alle entità di tipo **TopologicalNode**, **EnergyConsumer**, **BusbarSection**, **Terminal**, **PowerTransformerEnd**, **SynchronousMachine**, **Junction**, **EnergySource**. Gli attributi di tali nodi sono tutti gli attributi provenienti dal *Knowledge Graph*, dai profili normalizzati di potenza associati ai nodi topologici e dai valori di tensione calcolati tramite il calcolo di *power flow*;
- gli archi corrispondono alle entità di tipo **ACLineSegment**, **Breaker**, **LoadBreakSwitch**, **PowerTransformer**, **Disconnecter**, con la potenziale aggiunta di connettori ideali, introdotti per mantenere totalmente reversibile la trasformazione.

In caso di modellazione su spazi euclidei discreti (ad esempio, immagini rappresentate in pixel o volumi rappresentati in voxel), il **vicinato di un punto** (un pixel o un voxel) è immediatamente determinato da tutti gli altri punti (pixel o voxel) che presentano una distanza (distanza Manhattan) minore o pari ad un valore di soglia (solitamente, da 1 a 3), il quale sarà uno degli iper-parametri del modello stesso.

Nel caso di spazi non euclidei, come appunto i grafi, i punti sono dati dai nodi del grafo e la metrica di distanza non è data dallo spazio stesso, ma è definita anch'essa come iper-parametro del modello.

Il requisito con cui le trasformazioni della famiglia CIM2GraphML sono state progettate è la **reversibilità**. Con questo si intende che nessuna informazione è creata o distrutta come parte della trasformazione da *Knowledge Graph* (a meno di filtri relativi ad informazione non d'interesse per il problema trattato) e che è sempre possibile eseguire una trasformazione inversa che ricostruisca esattamente il *Knowledge Graph* di partenza.



La trasformazione CIM2GraphML è in grado di ridurre il cammino minimo fra entità rilevanti per la modellazione, ma non migliora la **vicinanza di differenti TopologicalNode corrispondenti a nodi della rete elettrica direttamente connessi**. Infatti, essi non risultano a una distanza di un arco, ma anche a distanze superiori (due o più archi di distanza). Un ulteriore problema consiste nell'**inserimento di artefatti**, ovvero archi di tipologia "Connettore Ideale" non presenti nell'ontologia CIM, ma necessari per mantenere invariato il contenuto informativo a valle della trasformazione stessa.

Vista la necessità di maggior compattezza del risultato rispetto alle entità di tipo TopologicalNode, le quali sono l'obiettivo della predizione, e l'introduzione di un numero non trascurabile di artefatti, i quali potrebbero deviare l'attenzione del processo di addestramento, la trasformazione CIM2GraphML non presenta le caratteristiche necessarie al raggiungimento dell'obiettivo dell'attività. Per questo motivo, sono state derivate dalla trasformazione fondamentale CIM2GraphML le due trasformazioni "trasformazione a TopologicalNode come entità di prima classe" e "trasformazione a TopologicalNode come attributo".

5.2.2 Trasformazione a TopologicalNode come entità di prima classe

Questa trasformazione è pensata per **favorire la modellizzazione dell'interazione globale dei nodi topologici a livello di rete**. In primo luogo, tutte le entità di tipo TopologicalNode vengono vincolate ad essere rappresentate come nodi del grafo risultante. Inoltre, si impone che ogni coppia di TopologicalNode direttamente connessi debba essere a 1 arco di distanza (concetto di "prima classe"). Un ultimo requisito fondamentale della trasformazione è di non introdurre alcun artefatto, in modo da ridurre il rischio che essi possano influenzare negativamente il processo di addestramento di modelli convolutivi.

In seguito ai passaggi di trasformazione, si ottiene un grafo caratterizzato da **due tipi di vicinato a distanza 1** (Figura 5.1):

1. il vicinato a distanza 1 di un nodo di tipo **ConductingEquipment** per un TopologicalNode che include sempre solo il nodo di tipo **TopologicalNode** a cui si riferisce;
2. il vicinato a distanza 1 di un nodo di tipo **TopologicalNode** che include sempre
 - a. l'intero insieme di **ConductingEquipment** che vi si riferiscono;
 - b. l'insieme di tutti gli altri nodi di tipo **TopologicalNode** che sono direttamente collegati allo stesso.

Inoltre, ogni vicinato di distanza superiore ad 1 deve necessariamente essere raggiunto tramite l'inclusione nel cammino di nodi di tipo TopologicalNode. Una peculiarità della trasformazione appena presentata consiste nella modellizzazione delle entità di tipo BusbarSection come archi multipli insistenti, da un capo, sul TopologicalNode con cui si relazionano e, dall'altro capo, su ciascuno dei nodi contenuti nel nodo topologico stesso (ad esempio le entità di tipo EnergyConsumer ed EnergySource).

Le caratteristiche topologiche della rete in graphML ottenuta con la "trasformazione a TopologicalNode come entità di prima classe" sono estremamente affini alla risoluzione del problema del power flow con applicazione di tecniche convolutive. Infatti, i nodi di tipo TopologicalNode, per i quali sarà necessario ottenere la regressione del valore di tensione, potranno, innanzitutto, ricevere le informazioni da tutti i nodi che possono essere considerati ConductingEquipment a loro direttamente riferiti, ottenendo ciascuno una completa caratterizzazione locale da una convoluzione sul proprio vicinato a distanza 1. Successivamente, estendendo la convoluzione su vicinati più ampi, l'informazione aggregata su ciascun nodo di tipo TopologicalNode potrà influenzare i nodi vicini dello stesso tipo, portando il modello a scoprire come far convergere la regressione della tensione in base ad un'informazione globale.

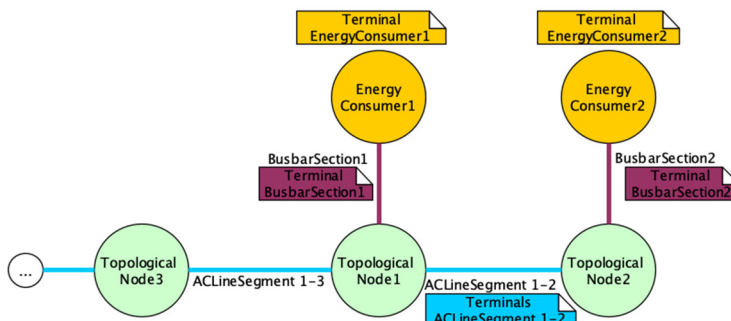


Figura 5.1 – Grafo ottenuto come output della “trasformazione a TopologicalNode come entità di prima classe”. Le etichette associate a nodi e archi indicano i rispettivi attributi.

5.2.3 Trasformazione a TopologicalNode come attributo

Questa trasformazione è pensata per **focalizzare la modellizzazione su ogni singolo nodo topologico** e favorire la **modellizzazione dell’interazione a livello locale delle entità che costituiscono il nodo topologico stesso**. Una potenziale debolezza della trasformazione presentata nella sezione precedente deriva infatti dall’ambiguità residua di vicinato a distanza 1 di un TopologicalNode. Si presentano infatti due casi in cui l’essere incluso nel vicinato di un altro nodo indica, alternativamente, il fatto che il nodo in esame (ad esempio per la presenza di EnergyConsumer ed EnergySource) influenza i valori di potenza, attiva e reattiva, che si costituiscono nell’altro nodo, oppure che il nodo in esame influenza come tali grandezze di potenza, attiva e reattiva, vengono scambiate verso il resto della topologia della rete elettrica. Questo causa la necessità da parte dei modelli convolutivi di una notevole quantità di dati di addestramento per poter essere in grado di discriminare in modo efficace tra differenti topologie i contributi all’interno del vicinato.

Nonostante la trasformazione preveda nuovamente l’introduzione di artefatti (Figura 5.2), in questo caso essi hanno sempre una semantica univoca, ossia indicano che un ConductingEquipment è fisicamente collegato a un’entità di tipo BusbarSection. Una limitazione della rappresentazione ottenuta con la “trasformazione a TopologicalNode come attributo” è l’eliminazione dal grafo dell’obiettivo di modellazione, ossia le entità di tipo TopologicalNode. Questo significa che, sebbene questa rappresentazione possa aiutare a focalizzare la modellazione sugli aspetti locali relativi a ciascun nodo topologico, non è una rappresentazione sufficiente per modellare efficacemente l’interazione globale fra differenti nodi topologici.

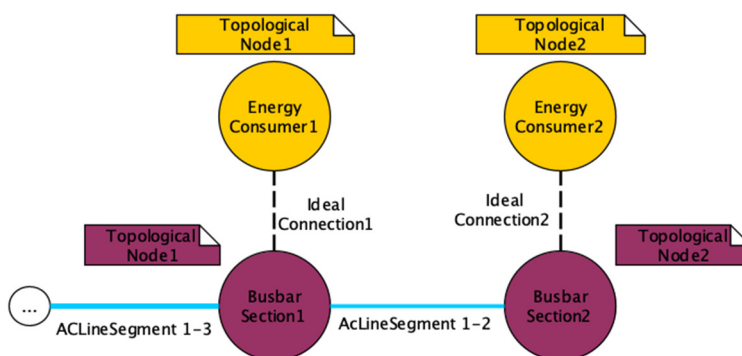


Figura 5.2 - Grafo ottenuto come *output* della “trasformazione a TopologicalNode come attributi”. I “Connettori Ideali”, indicati in grigio, sono artefatti della trasformazione e non sono entità dell’ontologia IEC CIM. Le etichette associate a nodi e archi indicano i rispettivi attributi.

5.3 Descrizione del compito di *machine learning*

L'obiettivo da raggiungere consiste nello **sviluppo di un modello di *machine learning* per la stima della tensione ai nodi di una rete di distribuzione in Media Tensione**. In questa sezione si delineano le caratteristiche e i requisiti necessari a raggiungere tale obiettivo, oltre alle metriche per valutare la qualità dei risultati.

È data una rete di distribuzione dell'energia in cui il vettore di trasporto dell'energia è di tipo elettrico, rappresentata come un grafo $G(N, E, A_N, A_E)$, in cui N e E sono rispettivamente nodi e archi del grafo che rappresentano entità della rete definite secondo l'ontologia IEC CIM, A_N sono gli attributi dei nodi del grafo e A_E sono gli attributi degli archi del grafo. Per i nodi di tipo **TopologicalNode** si definiscono:

- la potenza nominale attiva $\bar{P}$ e reattiva $\bar{Q}$ per il nodo;
- la potenza istantanea attiva $P(t)$ e reattiva $Q(t)$ per il nodo in un istante di tempo t ;
- la tensione nominale $\bar{V}$ per il nodo;
- la tensione istantanea $V(t)$ per il nodo in un istante di tempo t .

La scelta è stata quella di effettuare l'addestramento e il test separatamente su singole topologie di rete: come primo test si è provato infatti ad addestrare un modello per ciascuna rete, in grado di gestire eventualmente cambiamenti di topologia, ma non generalizzabile per reti di diversa dimensione o tipologia. Il compito di *machine learning*, di regressione di nodo, è quello di apprendere un modello in grado di inferire la tensione istantanea $V(t)$ per ogni **TopologicalNode**, a partire da tutte le caratteristiche del grafo $G(N, E, A_N, A_E)$ in cui da A_N sia esclusa proprio la tensione istantanea $V(t)$ per ogni **TopologicalNode**. Il modello deve rispettare le metriche di successo e ottimizzare le metriche di qualità illustrate nella sezione seguente.

5.3.1 Metriche di successo e qualità

Un modello di *machine learning* prodotto dalla metodologia oggetto di questa fase del progetto si considera **valido** se, la seguente **metrica di successo** (o **accuratezza minima**)

$$S(t) := \frac{\left| \left\{ n \in N : \text{tipo}(n, \text{TopologicalNode}) \wedge \left| \frac{\tilde{V}_n(t) - V_n(t)}{V_n(t)} \right| \leq 10\% \right\} \right|}{|\{n \in N : \text{tipo}(n, \text{TopologicalNode})\}|}$$

per ogni t appartenente al dataset di test, rispetta il vincolo $S(t) = 100\%$.

Nella definizione di $S(t)$, $\text{tipo}(a, b)$ risulta VERO se un nodo n ha attributo *tipo* pari a *TopologicalNode*, $V_n(t)$ è la tensione istantanea per il nodo n come definita nel dataset di test, $\tilde{V}_n(t)$ è la tensione istantanea per il nodo n come inferita dal modello in fase di test. Il vincolo imposto implica che il modello non può mai commettere errori di predizione oltre il 10% per alcun nodo topologico presente nella rete.

Stante un modello valido, ossia che rispetti il vincolo sulla metrica di successo, un modello si considera tanto migliore quanto più prossima a 100% è la **metrica di qualità** (o **accuratezza ideale/accuratezza obiettivo**), definita come:

$$A(t) := \frac{\left| \left\{ n \in N : \text{tipo}(n, \text{TopologicalNode}) \wedge \left| \frac{\tilde{V}_n(t) - V_n(t)}{V_n(t)} \right| \leq 1\% \right\} \right|}{|\{n \in N : \text{tipo}(n, \text{TopologicalNode})\}|}$$

Questo significa che il modello verrà spinto ad ottimizzare i propri parametri in modo da aumentare il più possibile il numero di nodi topologici per cui la previsione della tensione sia entro l'1% d'errore, non ottimizzando ulteriormente oltre tale soglia.

A parità di qualità del modello, si prediligeranno i modelli in grado di fornire il risultato con **tempo di inferenza** minore.



5.4 Modelli basati su *Graph Convolutional Networks*

Definiti l'obiettivo del modello di *machine learning* e le metriche per valutarne il successo e la qualità dei risultati, il primo passo è stato quello di applicare modelli che, allo stato dell'arte, risultano essere efficaci [4]. Nell'ambito dell'analisi del *power flow*, il compito di *node regression* viene risolto prevalentemente con modelli denominati *Graph Convolutional Networks* (GCN) [100, 101, 102].

Le GCN [103] sono reti composte da **layer multipli in cascata**. Ad ogni *layer*, i nodi ricevono informazioni relative allo stato dei nodi vicini, che vengono aggregate e aggiunte alla rappresentazione del nodo corrente. Tale processo è denominato "**message passing**" e può essere iterato più volte per inserire *layer* multipli e consentire così a ciascun nodo di scambiare informazioni con nodi sempre più distanti nel grafo. Questo aiuta a catturare le relazioni a lungo raggio e a migliorare l'informazione complessiva aggregata dei nodi. Durante il processo di *message passing*, l'informazione proveniente da ogni nodo viene ponderata secondo dei pesi e poi aggregata secondo una **funzione di attivazione non lineare**. Tale informazione ottenuta è anche denominata "**Graph embedding**" ed è una rappresentazione latente multidimensionale del nodo in uno spazio in cui la vicinanza tra i nodi è una misura della loro similarità. Tale rappresentazione latente viene poi pesata per inferire il valore di tensione del nodo considerato. Le GCN imparano i pesi ottimali dei processi di *message passing* e di aggregazione finale minimizzando una **funzione di perdita (loss function)**, che quantifica l'errore che la GCN sta compiendo nella stima delle tensioni dei nodi. Tale funzione di perdita può minimizzare la somma degli errori in valore assoluto (MAE - *Mean Absolute Error*) o la somma degli errori al quadrato (MSE - *Mean Squared Error*) [104]. In modelli complessi come le GCN, all'aumentare del numero di *layer*, è frequente che i pesi imparati tendano ad assumere valori molto elevati. Tale comportamento è sintomo di **overfitting** [105], fenomeno che determina la specializzazione del modello sul *set* di dati di addestramento e la sua incapacità di essere altrettanto accurato su un *set* di dati differente. Per questo motivo, all'interno delle funzioni di perdita, si inserisce un **fattore di regolarizzazione**, che penalizza valori dei pesi troppo elevati. Si parla di regolarizzazione *L1* (o *Lasso*) quando il termine di penalizzazione inserito è la somma dei pesi in valore assoluto; si parla di regolarizzazione *L2* (o *Ridge*) quando il termine di penalizzazione inserito è la somma dei pesi al quadrato; si parla di regolarizzazione *L1L2* (o *ElasticNet*) quando vengono inseriti entrambi termini di penalizzazione [106, 107]. Un ulteriore metodo di riduzione dell'*overfitting* è il **dropout** [108]. In questo caso, durante ciascun passaggio di messaggi, alcuni nodi vengono selezionati casualmente e temporaneamente "eliminati" dal processo, venendo ignorati in quella specifica iterazione. La variabilità introdotta dal *dropout* aiuta a evitare che il modello diventi troppo dipendente dai dettagli specifici del *set* di addestramento, riducendo così il rischio di *overfitting*.

5.4.1.1 Protocollo di addestramento

Con un modello GCN l'algoritmo di stima delle tensioni ai nodi parte dai dati di una rete elettrica convertita in grafo con la "**trasformazione a TopologicalNode come entità di prima classe**" che tuttavia, ha come *output* una rete contenente nodi e archi eterogenei (vedi 5.2.2). Le GCN allo stato dell'arte possono però processare soltanto grafi omogenei e, per questo motivo, sono stati presi in considerazione soltanto le entità di prima classe (nodi topologici) e ignorate le entità considerate come attributo. Gli archi che collegano i nodi topologici alle entità di attributo vengono perciò tagliati e ignorati durante l'addestramento, mentre restano gli archi che collegano tra di loro i TopologicalNode.

Ogni nodo topologico del *dataset* è caratterizzato da *features* statiche e dinamiche. In particolare, le *features* dinamiche sono i valori di potenza attiva e reattiva per ciascun istante temporale. Tra le *features* statiche di maggiore rilevanza ci sono la potenza attiva nominale *P* e la potenza reattiva nominale *Q*. Ottenuto un *dataset* idoneo all'addestramento di una GCN, è necessario dividere i

Un **embedding** è una rappresentazione latente a bassa dimensionalità di oggetti come immagini, testo, video, grafi, che rende più semplice il compito che il modello di *machine learning* deve attuare. L'oggetto di partenza viene dunque trasformato, tramite tecniche di vario tipo, in un vettore che risiede in uno spazio che è dimensionalmente inferiore rispetto a quello di partenza. I metodi per l'ottenimento di embedding si basano, dunque, sulla mappatura di tali oggetti di partenza su spazi euclidei in cui la distanza tra due istanze rappresenta una metrica accurata della loro similarità (più sono vicini due oggetti nello spazio di destinazione, più sono simili tra di loro). Nel nostro caso d'interesse, ossia i grafi, le relazioni non sono caratteristiche dello spazio ma della forma; l'obiettivo, infatti, è quello di mappare i nodi dei grafi su uno spazio dove la distanza è la metrica di similarità. Per effettuare tale compito, in passato, la creazione di embedding veniva guidata tramite l'adempimento di operazioni elementari (come traslazioni, rotazioni e rototraslazioni), imparando i pesi ottimali di tali operazioni per ottenere una rappresentazione equivalente accurata dei nodi. In seguito, tali operazioni sono state sostituite da reti neurali convoluzionali che sono capaci di imparare pattern più complessi e ottenere embedding più accurati, capaci di rappresentare le distanze tra i nodi in uno spazio euclideo in maniera più efficace rispetto a ciò che si otteneva in passato limitandosi ad operazioni fisse. Tali rappresentazioni nello spazio finale ottenuto, tuttavia, non hanno alcun significato fisico rispetto alla formalizzazione di partenza della struttura a grafo; lo spazio latente finale possiede caratteristiche completamente diverse dallo spazio di partenza e ha l'unico obiettivo di rendere significative le distanze fra i nodi.

dati in tre *set* distinti, di addestramento, validazione e *test*. Tale separazione può essere effettuata con molteplici tecniche [109], tra le quali:

- *temporal ordering*: i dati mantengono coerenza temporale, in quanto sono separati in maniera tale che tutti i dati del set di validazione siano successivi a quello di *training*, e che tutti i dati del set di test siano successivi ad entrambi gli altri due *set* di dati. In questo scenario, viene rispecchiato un caso d'uso reale, in cui i dati su cui viene addestrato il modello sono sempre temporalmente precedenti ai dati su cui il modello viene utilizzato per effettuare le predizioni;
- *random Sampling*: i dati vengono riordinati in modo casuale e in seguito separati in tre *set* distinti.

In entrambi i casi, le percentuali di dati attribuite ai *set* di addestramento, validazione e *test* sono rispettivamente 50%, 10% e 40%. Tra le due tecniche si è scelto di utilizzare il *Random Sampling*. In futuro, si prevede di provare anche il *Temporal ordering* al fine di valutare se possa condurre a prestazioni migliori.

5.4.1.2 Esplorazione degli iper-parametri e risultati sperimentali

Le prestazioni del modello sono state valutate in termini di *accuratezza minima* e *accuratezza obiettivo* (paragrafo 5.3.1). Sono stati presi in considerazione soltanto i modelli capaci di garantire un'accuratezza minima del 100%, e il confronto tra essi è stato effettuato in termini di accuratezza obiettivo. Innanzitutto, le prestazioni del modello sono state studiate al variare dei singoli iper-parametri [110], ovvero i parametri esterni al modello che non vengono appresi durante l'addestramento, ma ne definiscono la struttura e la complessità. La scelta adeguata degli iper-parametri è cruciale per ottenere prestazioni ottimali del modello. Nel caso descritto delle GCN, gli **iper-parametri** fondamentali sono i seguenti:

- **numero di *layer* convolutivi**: più è grande, più la rete diventa complessa e capace di imparare *pattern* sempre più elaborati. Tuttavia, allo stesso tempo, la rete diventa molto sensibile al *dataset* e prona all'*overfitting*; in più, i tempi di inferenza crescono [111]. Nei test effettuati, l'aumento del numero di *layer* non ha evidenziato alcun beneficio, se non sulla stabilità nel processo di apprendimento. Di conseguenza si è deciso di procedere con il numero minimo di *layer*, ossia 2;
- **tipologia di funzione di perdita**: la scelta è ricaduta su **MAE**, in quanto ha presentato risultati decisamente superiori rispetto a MSE;
- **tipologia di regolarizzazione**: sono state testate le regolarizzazioni L1, L2 e L1L2. Tuttavia, l'inserimento di tali termini di regolarizzazione non ha risultato portare alcun beneficio e di conseguenza si è scelto di **non utilizzarli**;
- **inserimento del *dropout* e del relativo tasso di spegnimento dei nodi**: il tasso di spegnimento dei nodi nel *dropout* è stato ricercato tra 0,7 e 0,9 come suggerito dallo stato dell'arte [112]. La migliore configurazione è risultata con **rate di *dropout* pari a 0,9**;
- **funzione di attivazione dei *layer* intermedi**: non è stata modificata e la scelta è stata a favore della ReLU (*Rectified Linear Unit*) [113];
- **funzione di attivazione del *layer* finale**: è stata scelta tra sigmoide, tanh e lineare. La migliore configurazione è risultata con **tanh** come funzione di attivazione.

Si riportano in Tabella 5.1 i risultati di stima ottenuti utilizzando tali iper-parametri, separatamente per ciascuna rete di distribuzione considerata.

Tabella 5.1 – Risultati delle predizioni del modello GCN addestrato per ciascuna rete elettrica. Come metrica di qualità si indica la percentuale di nodi per cui la stima di tensione è entro l'1% rispetto al valore calcolato. In verde sono evidenziate le isole topologiche per cui l'errore medio risulta inferiore a 1%.

Isola topologica	n° TopologicalNode	Percentuale di nodi con stima di tensione entro 1%	Errore medio (%)
RETE1	1146	64,9	1,6
RETE2	1416	59,6	1,7
RETE3	1690	35,8	2,8
RETE4	886	77,8	0,8
RETE5	1007	83,0	0,6
RETE6	2073	42,6	1,9
RETE7	2087	32,1	2,5
RETE8	1664	62,0	1,4
RETE9	2617	25,5	2,6
RETE10	1404	76,6	0,8
RETE11	2470	30,8	1,9
RETE12	2605	38,5	2,2
RETE13	3532	31,7	2,2
RETE14	4189	30,9	2,2
RETE15	489	99,6	0,2
RETE16	658	35,2	2,0
RETE17	928	37,3	2,1
RETE18	1202	14,1	3,2
RETE19	398	38,4	1,4
RETE20	519	79,0	0,6

Si osserva che soltanto per 4 reti l'errore percentuale medio di previsione è inferiore all'1% e raggiunge valori fino a 3,2%; tuttavia, per tutte le reti si rispetta il vincolo di tensione entro il 10% rispetto al valore calcolato. In Figura 5.3 si riporta il confronto tra tensione calcolata e stimata tramite le GCN procedendo lungo una linea MT dell'isola topologica RETE19.

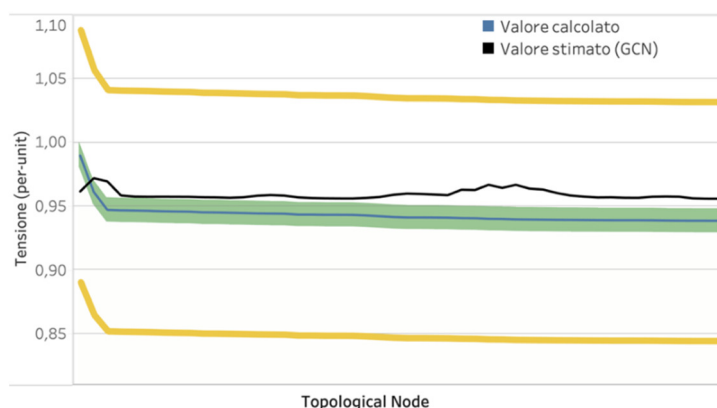


Figura 5.3 – Confronto dell'andamento della tensione calcolata e stimata (GCN) lungo una linea MT dell'isola topologica RETE19. La banda verde indica il limite 1% rispetto al valore calcolato, mentre le curve gialle indicano il limite 10% rispetto al valore calcolato.

Si nota che il vincolo di tensione stimata entro l'1% non viene rispettato, mentre lo è, per tutti i nodi, il vincolo di tensione stimata entro il 10%. Dal confronto tra i profili nel tempo della tensione calcolata e stimata per un singolo nodo topologico della RETE19 (Figura 5.4), si osserva come gli andamenti generali di crescita e decrescita della tensione vengano correttamente stimati dalla rete neurale.

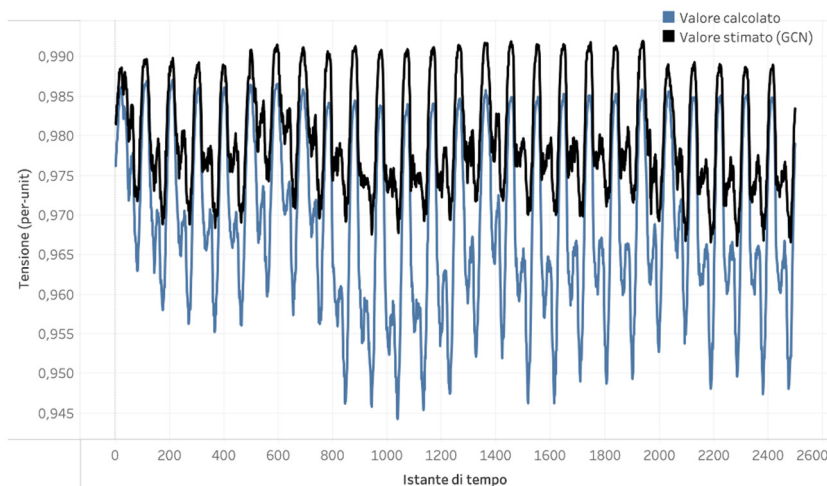


Figura 5.4 - Confronto tra la tensione calcolata e stimata (GCN) in funzione del tempo per un nodo topologico appartenente all'isola topologica RETE19.

Infine, le prestazioni computazionali delle reti neurali ottenute sono state valutate in termini di tempi di inferenza. In particolare, per l'elaborazione dei dati relativi a 1000-1200 nodi topologici e profili di 100 istanti temporali, è stato ottenuto un tempo medio di inferenza pari a circa 5 secondi utilizzando una *GPU NVIDIA Tesla T4*.

Il principale problema delle GCN implementate consiste nel fatto di poter lavorare solo con nodi omogenei (*TopologicalNode*) e di non tenere conto delle diverse tipologie di nodo della rete elettrica. L'inserimento di variabili di categoria per utilizzare tali informazioni causerebbe infatti una significativa crescita delle dimensioni dell'*input*, difficilmente gestibile da una rete neurale artificiale. Considerando l'importanza della diversa tipologia di nodi di nel calcolo del *power flow* (PV, PQ, slack), questa problematica di modellizzazione potrebbe essere la causa degli errori di predizione ottenuti.

Come tentativo, sono stati perciò implementati dei modelli di **Relational GCN** (RGCN) [114], un'estensione di GCN che permette di descrivere reti di input con nodi omogenei e archi eterogenei. L'implementazione originale è stata modificata per poter spostare l'eterogeneità dagli archi ai nodi, in modo da tenere in considerazione la presenza di nodi di diversa tipologia. Tuttavia, anche in questo caso, non sono stati ottenuti risultati soddisfacenti. Infatti, a fronte di un maggiore costo computazionale (con dilatazione sia dei tempi di addestramento sia dei tempi di inferenza), le RGCN hanno mostrato performance confrontabili a quelli delle GCN classiche.

5.5 Modelli ensemble basati su alberi decisionali

Per migliorare le stime di tensioni ai nodi, è stata provata una differente tipologia di modelli, in grado di osservare i nodi fisici specifici a basso livello, considerando la loro tipologia, quella dei vicini più stretti e tutti gli attributi statici. In particolare, l'idea è stata di testare modelli in grado di considerare informazioni estratte dal grafo di rete, ma in **formato tabulare**: i **modelli ensemble basati su alberi** [115, 116]. Lo specifico modello preso in considerazione è un **Gradient Boosting**, denominato *LightGBM* [117].

5.5.1.1 Protocollo di addestramento

In questo caso, la predizione viene eseguita a partire da una rete elettrica convertita in grafo con la "**trasformazione a TopologicalNode come attributo**". Per ciascun nodo in *input* è nota, ad uno specifico istante di tempo, la potenza attiva e reattiva istantanea, il tipo di nodo, i suoi attributi statici, il numero di vicini per tipologia di nodo e la potenza attiva e reattiva istantanea per alcuni di essi. Lo scopo è di stimare la tensione istantanea per il nodo stesso. La rappresentazione del

Un modello di **Gradient Boosting** viene addestrato combinando diversi modelli ad albero in modo sequenziale, additivo e graduale. Questo permette di produrre un modello predittivo più accurato a partire da un insieme di modelli predittivi deboli.

dataset contenente alcuni esempi degli attributi considerati è presente nella Tabella 5.2. Anche in questo caso, come per le GCN, la tecnica di separazione del *dataset* utilizzata è il *Random Sampling* e le percentuali di separazione dei *set* di addestramento, validazione e *test* sono rispettivamente 60%, 20% e 20%.

Tabella 5.2 – Esempio di rappresentazione tabulare del *dataset*. Si riportano tipo, potenza all'istante temporale considerato e potenza nominale per il nodo di riferimento e due tra i suoi vicini. Le ultime colonne riguardano il numero di nodi vicini per tipologia.

	Tipo nodo	Tipo vicino01	Tipo vicino02	p(t) nodo	p(t) vicino01	p(t) vicino02	P_nom nodo	P_nom vicino01	P_nom vicino02
1	BusSec	BusSec	BusSec	0,089	0,857	0,362	0	0,283	0
2	BusSec	BusSec	EnCons	0,494	0,247	0,494	0,287	0,253	0,287
	Q_nom nodo	Q_nom vicino01	Q_nom vicino02	N° vicini EnCons	N° vicini BusBar	N° vicini PowerTrEnd	N° vicini Junc	N° vicini SynMach	N° vicini EnSource
1	0	0,137	0	0	2	0	0	0	0
2	0,112	0,083	0,112	2	2	0	0	0	0

5.5.1.2 Esplorazione degli iper-parametri e risultati sperimentali

I metodi di ensemble sono tecniche di Machine Learning che combinano le predizioni di più modelli al fine di ridurre l'errore. Facendo sì che ogni modello veda un diverso aspetto del problema, ognuno di essi prenderà una strada leggermente diversa durante il processo di addestramento. Combinando, infine, i differenti punti di vista dei vari modelli, si ottiene una predizione più accurata. I metodi ensemble solitamente raggiungono performance elevatissime, soprattutto se rapportati ai modelli singoli. Per comprenderne il motivo, basti pensare ad un esempio relativo ad un problema di classificazione di una figura geometrica: se l'istanza è un cilindro, il modello A che ha un punto di vista laterale osserva un rettangolo, il modello B che ha un punto di vista dall'alto osserva un cerchio. Solo combinare le predizioni dei due modelli permette di ottenere un risultato accurato. Tale esempio permette di capire le potenzialità delle tecniche di ensemble. È importante sottolineare che tale esempio ha unicamente lo scopo di dare un'intuizione del motivo per cui le tecniche di ensemble siano così efficaci, ma il reale processo è molto più complesso di quanto descritto in tale esempio.

Gli iper-parametri esplorati sono alcuni di quelli messi a disposizione dal modello LightGBM: il **numero di alberi**, il **numero massimo di foglie di ogni albero**, la **funzione di perdita** da minimizzare e altri **iper-parametri tecnici relativi alla complessità del modello**. L'esplorazione degli iper-parametri è stata effettuata mediante *Optuna* [118], un *framework* di ottimizzazione automatica degli iper-parametri. Al momento un iper-parametro non esplorato riguarda il numero di vicini per cui viene estratta la tipologia e la potenza attiva e reattiva: per ciascun nodo di riferimento vengono estratte le informazioni relative a solo due vicini, ma sarebbe utile approfondire meglio le conseguenze di questa scelta.

Una volta ottimizzati gli iper-parametri, è stato addestrato il modello LightGBM per la stima dei valori di tensione ai nodi. Per poter paragonare le prestazioni del modello con quelle della GCN (che predice direttamente sul nodo topologico), è stato necessario aggregare le predizioni effettuate dal presente modello sui nodi della rete trasformata secondo la "trasformazione a TopologicalNode come attributo". Per effettuare tale operazione, sono state utilizzate alcune tecniche di *ensembling* allo stato dell'arte [119], che hanno permesso di aggregare le predizioni in maniera efficiente.

In Tabella 5.3, si osserva dai risultati per le singole reti che l'errore medio di previsione si mantiene nel range 0,1-0,6, il che mostra un netto miglioramento del modello rispetto alle GCN.

Tabella 5.3 - Risultati delle predizioni del modello LightGBM addestrato per ciascuna isola topologica. Come metrica di qualità si indica la percentuale di nodi per cui la stima di tensione è entro l'1% rispetto al valore calcolato. In verde sono evidenziate le isole topologiche per cui l'errore medio è inferiore a 1%.

Isola topologica	n° TopologicalNode	Percentuale di nodi con stima di tensione entro 1%	Errore medio (%)
RETE1	1146	93,3	0,3
RETE2	1416	93,4	0,3
RETE3	1690	89,1	0,4
RETE4	886	99,4	0,1
RETE5	1007	99,9	0,1
RETE6	2073	91,7	0,3
RETE7	2087	89,6	0,4
RETE8	1664	95,4	0,2
RETE9	2617	88,1	0,4

Isola topologica	n° TopologicalNode	Percentuale di nodi con stima di tensione entro 1%	Errore medio (%)
RETE10	1404	99,6	0,1
RETE11	2470	92,1	0,3
RETE12	2605	92,2	0,3
RETE13	3532	90,9	0,4
RETE14	4189	90,8	0,4
RETE15	489	100,0	0,1
RETE16	658	82,4	0,5
RETE17	928	89,4	0,6
RETE18	1202	83,2	0,6
RETE19	398	89,2	0,4
RETE20	519	99,8	0,2

Come per le GCN, si riportano degli esempi dell'andamento della tensione calcolata e stimata lungo una linea MT della RETE19: in particolare, si riporta l'andamento per l'istante temporale che ha riscontrato i migliori errori di previsione (Figura 5.5), per il quale la tensione stimata da LightGBM si mantiene entro l'1% rispetto al valore calcolato per quasi tutti i nodi, e un esempio per l'istante temporale che ha riscontrato i peggiori errori di previsione (Figura 5.6), per il quale la tensione stimata si mantiene comunque entro il 10% rispetto al valore calcolato.

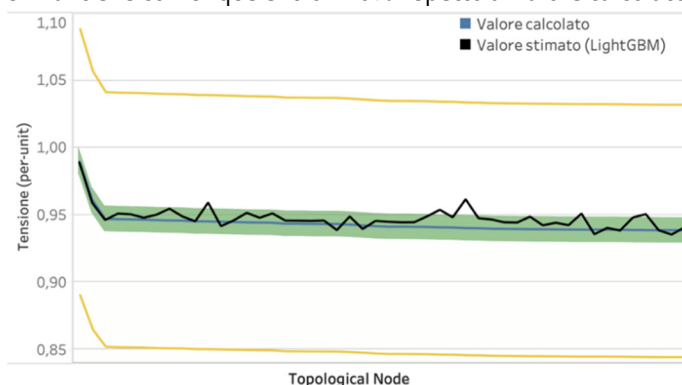


Figura 5.5 - Confronto dell'andamento della tensione calcolata e stimata (LightGBM) lungo una linea MT dell'isola topologica RETE19 per l'istante temporale per cui sono stati ottenuti i migliori errori di previsione. La banda verde indica il limite 1% rispetto al valore calcolato, mentre le curve gialle indicano il limite 10% rispetto al valore calcolato.

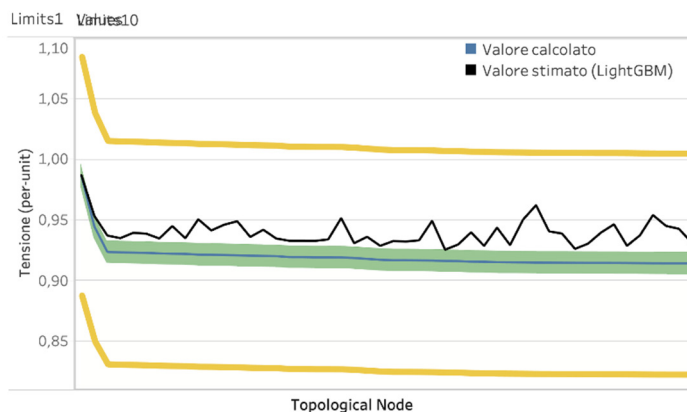


Figura 5.6 - Confronto dell'andamento della tensione calcolata e stimata (LightGBM) lungo una linea MT dell'isola topologica RETE19 per l'istante temporale per cui sono stati ottenuti i peggiori errori di previsione. La banda verde indica il limite 1% rispetto al valore calcolato, mentre le curve gialle indicano il limite 10% rispetto al valore calcolato.

Infine, la Figura 5.7 mostra un confronto tra i profili temporali delle tensioni calcolata e stimata tramite LightGBM per un singolo nodo topologico della RETE19. Anche in questo caso, gli andamenti generali di crescita e decrescita vengono correttamente stimati dalla rete neurale, con un maggiore avvicinamento ai valori massimi/minimi rispetto alle GCN.

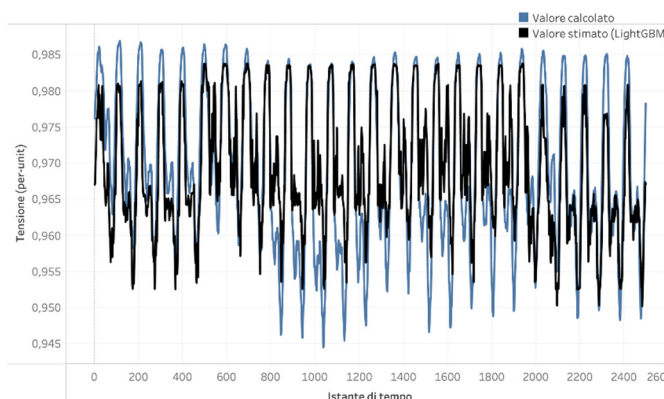


Figura 5.7 - Confronto tra la tensione calcolata e stimata (LightGBM) in funzione del tempo per un nodo topologico appartenente all'isola topologica RETE19.

5.6 Sintesi dei risultati sulla stima delle tensioni ai nodi

In conclusione, per poter **stimare le tensioni ai nodi di una rete elettrica** senza l'ausilio di tecniche numeriche sono stati sviluppati due diversi **modelli di machine learning basati su grafi**: un modello basato sulle *Graph Convolutional Network* e uno di LightGBM. Come attività preliminare, le reti elettriche descritte secondo lo standard IEC CIM sono state convertite in graphML, formato utile per l'applicazione di tecniche di *graph analysis* e *Graph Machine Learning*. Successivamente, si è passati all'effettiva implementazione di modelli per la stima dei valori di tensione separatamente per ciascuna isola topologica. Il modello basato sulle **Graph Convolutional Network** ha permesso di ottenere un errore medio di previsione rispetto al valore della tensione calcolata **nel range 0,2-2,8% per le 20 isole topologiche**. Il passaggio a LightGBM ha permesso di ottenere migliori risultati e di passare a errori percentuali medi nel range di **0,1-0,6%**. Possibili sviluppi futuri consistono nell'ibridazione dei due principali modelli ottenuti. Tale ibridazione può essere ottenuta mediante due tecniche: un classico *ensemble con media pesata* [120] delle predizioni e la *tecnica dello stacking* [121]. Nel primo caso, le predizioni generate dai due modelli per ciascun nodo vengono pesate tramite un parametro α secondo la regola:

$$\alpha \cdot P_{GCN} + (1 - \alpha) \cdot P_{LGBM}$$

Nel secondo caso, la predizione generata dalla GCN viene usata come *feature* aggiuntiva per LightGBM. Quest'ultimo ne farà uso durante il suo addestramento per comprendere quando una predizione della GCN è affidabile e quando non lo è. Potendo contare su tale informazione aggiuntiva, le performance di LightGBM, potenzialmente, cresceranno.

5.7 Preparazione dei dati per la stima di variabili di stato di reti gas

Gli algoritmi sviluppati con le GCN per la rete elettrica verranno presi come riferimento anche per lo sviluppo di un modello di **stima delle grandezze di stato di reti energetiche** per vettori differenti da quello elettrico, sempre senza far uso di tecniche analitiche. Come primo esempio verranno prese in considerazione **reti gas** di Brescia e Milano.

L'attività preparatoria svolta ha riguardato l'analisi delle reti descritte tramite l'ontologia derivante da IEC CIM per il gas, di cui al paragrafo 2.3. L'obiettivo è stato quello di analizzare la

possibilità di stabilire degli isomorfismi fra reti con differenti vettori energetici, in modo da avere sufficiente sicurezza nell'applicare metodologie sviluppate per il vettore elettrico anche ad altri vettori. Adattando la trasformazione progettata nello scorso triennio di Ricerca di Sistema [99], è stata sviluppata una trasformazione preliminare delle 2 reti di Folzano/Lamarmora (Brescia) e Lambrate (Milano) che, a partire dalla serializzazione RDF/XML ottiene una rappresentazione preliminare della rete energetica orientata all'applicazione di tecniche *Graph Machine Learning*. La trasformazione è stata adattata considerando la similitudine del ruolo ricoperto dagli elementi della rete gas descritti secondo le classi dell'ontologia derivante da IEC CIM rispetto a agli elementi della rete gas descritti secondo l'ontologia IEC CIM (si veda la Tabella 2.3). A titolo di esempio, le entità *PressureRegulator* e *PressureRegulatorEnd* per la rete gas corrispondono alle entità *PowerTransformer* e *PowerTransformerEnd* per la rete elettrica. Il risultato dell'applicazione della trasformazione preliminare della rete di Lambrate è mostrato in Figura 5.8.

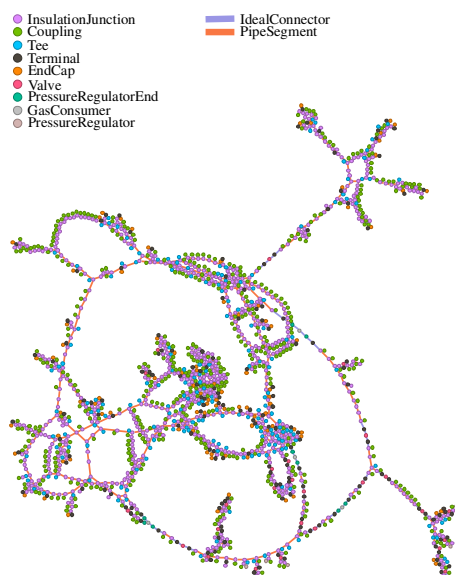


Figura 5.8 - Grafo risultante dalla trasformazione da RDF/XML a graphML per la porzione di rete gas di Lambrate.

Per lo sviluppo di un algoritmo per la stima delle variabili di stato su una rete gas, saranno considerati gli esempi di simulazione di *gas flow* forniti da Unareti e realizzati tramite il software di modellazione e simulazione Synergi [122]. Data una rete gas con una specifica topologia e a un certo livello di pressione, per ciascun nodo sono considerati come dati di input i flussi di gas prelevati in tale punto (in m³/h). Il risultato dell'output della simulazione corrisponde alle pressioni effettive per ciascun nodo e alla velocità del flusso di gas sulle condotte, assieme ad altri parametri che caratterizzano il flusso nella condotta (come il numero di Reynolds).

Queste simulazioni sono state considerate come riferimento e confrontate ai risultati ottenuti con le simulazioni di *gas flow* in ambiente python, tramite le funzionalità della libreria *pandapipes* [123]. In questo modo, a partire da un numero limitato di simulazioni, sarà possibile generare nuovi dati di input e ottenere l'output del *gas flow* sulle reti da analizzare e quindi preparare il set di dati per l'addestramento del modello di *Graph Machine Learning* per la stima delle variabili di stato sulla rete gas.

6 - PROCESSI DI MANUTENZIONE DELLE RETI

Come prosecuzione delle attività mirate alla digitalizzazione dei processi di manutenzione delle reti per l'energia, avviate nelle precedenti annualità [99], questo capitolo descrive analisi di dati relativi ai guasti di componenti della rete elettrica, con lo scopo di contribuire al miglioramento dell'affidabilità e della resilienza delle reti di distribuzione. L'attività, svolta in collaborazione con il DSO Unareti, ha riguardato uno studio sulle procedure attualmente seguite dai DSO in merito agli ordini di lavoro per gestire gli eventi di guasto e i processi manutenzione focalizzandosi, in particolare, sulle cabine secondarie. Questa scelta è motivata da molteplici fattori: prima di tutto l'entità e l'impatto che un guasto in cabina secondaria può avere sugli utenti, se non gestito tempestivamente. Inoltre, nelle cabine secondarie, specialmente quelle telecontrollate, sono presenti dati acquisiti che ancora non vengono pienamente sfruttati ma che possono essere utili per alimentare algoritmi di manutenzione predittiva. L'obiettivo triennale rivolge particolare attenzione a tecniche di *machine learning* e *big data* per individuare le posizioni dei guasti, sia dal punto di vista topologico che geografico e la correlazione con eventuali situazioni ambientali (es. ondate di calore, disastri naturali, etc.) per offrire ai DSO strumenti di gestione avanzata e di pianificazione degli interventi sulle reti.

6.1 Analisi dei guasti sulle reti di distribuzione

In generale i DSO si avvalgono di piattaforme per la gestione degli *asset* aziendali. Tali sistemi sono progettati per fornire un controllo completo sul ciclo di vita degli *asset*, dalla loro acquisizione alle periodiche manutenzioni, fino al loro smaltimento. Queste piattaforme si distinguono per la loro capacità di integrare e analizzare un'ampia e variegata gamma di dati relativi agli *asset*, consentendo di pianificare, programmare ed eseguire efficacemente le operazioni di manutenzione.

Tali sistemi permettono di registrare con precisione gli ordini di lavoro (ODL) per le manutenzioni e di tenere traccia di tutti i guasti avvenuti. Questi dati vengono raccolti in *database* dedicati, che comprendono informazioni dettagliate sugli *asset* coinvolti, sulle modalità e cause dei guasti, sulle azioni di risoluzione intraprese e sui tipi di guasto. Inoltre, i sistemi forniscono analisi approfondite per identificare modelli e tendenze ricorrenti: le analisi del *Fault Tree* (FT), che saranno dettagliate nel paragrafo 6.1.2. Parallelamente, i DSO integrano l'uso dei suddetti sistemi con *software* di analisi dati. Questa integrazione permette di estrarre, elaborare e visualizzare i dati raccolti, e di trasformarli in *insight* e report interattivi. Un esempio è riportato in Figura 6.1 dove viene indicata la frequenza delle segnalazioni ricevute per categoria di *asset*. Tale visualizzazione aiuta a identificare eventuali categorie per le quali approfondire il processo di manutenzione per renderlo più efficiente. Si noti, ad esempio, la rilevanza dei guasti per l'*asset* batteria nel sistema di telecontrollo delle cabine secondarie.

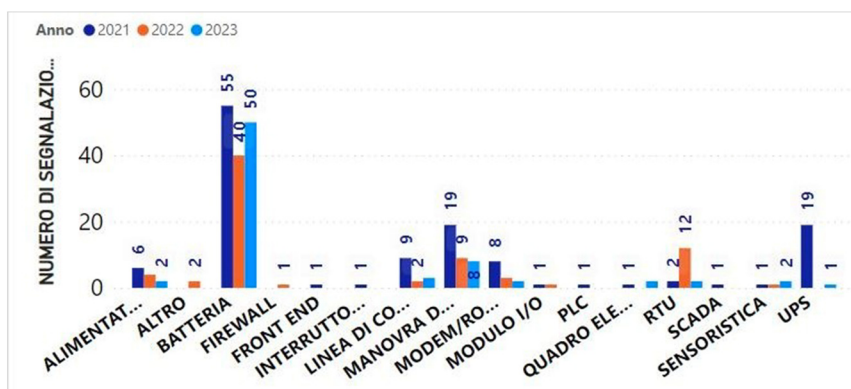


Figura 6.1 – Esempio di frequenza dei guasti per categoria di *asset* di una rete di distribuzione.

6.1.1 Database degli ordini di lavoro su guasto

In questo paragrafo si riporta, a titolo di esempio, la struttura dei dati collezionati in un ODL su guasto, così come attualmente archiviati da Unareti. La tabella contiene diversi campi che forniscono informazioni dell'intervento sugli *asset* conseguente ad eventi di guasto. In particolare, si hanno i seguenti valori:

- **ODL**: numero univoco identificativo dell'ODL;
- **fault Tree (FT)**: percorso nell'albero dei guasti che categorizza il tipo di guasto o problema riscontrato;
- **descrizione ODL**: breve descrizione del tipo di guasto o problema riscontrato;
- **stato ODL**: stato attuale dell'ODL (es. aperto, chiuso, in attesa);
- **data stato ODL**: data in cui è stato registrato l'ultimo stato dell'ODL;
- **data segnalazione**: data in cui il guasto o il problema è stato segnalato;
- **descrizione collocazione**: descrizione della collocazione dell'*asset*;
- **latitudine**: coordinata geografica latitudine dell'*asset*;
- **longitudine**: coordinata geografica longitudine dell'*asset*.

6.1.2 Fault Tree relativi agli ODL su guasto

Nel contesto della gestione degli *asset*, i *Fault Tree* (FT) rappresentano un tassello fondamentale per la valutazione dei rischi e la gestione dei guasti, mediante l'applicazione della *Fault Tree Analysis* (FTA). Un FT è un diagramma che rappresenta la combinazione logica di vari eventi di guasto, che possono portare ad un evento indesiderato o al malfunzionamento del sistema. Questi diagrammi sono strutturati in modo gerarchico, con l'evento di guasto in cima, seguito da una serie di possibili cause e sotto-cause, il tutto organizzato in nodi e collegamenti. Questo approccio analitico è utilizzato per identificare le possibili cause di guasti o malfunzionamenti in un sistema, permettendo agli ingegneri e ai responsabili della manutenzione di prevedere e prevenire potenziali problemi prima che si verifichino. Infatti, proprio grazie a questi alberi di guasto, relazionati agli ODL, è possibile individuare le principali cause di guasto delle singole cabine secondarie (CS). Per esempio, nel caso di apertura/chiusura degli interruttori dei sezionatori di manovra in CS, conseguenti alle normali riconfigurazioni post-guasto, si è notato che tali interruttori non sempre rispondevano ai comandi inviati da remoto, a causa di malfunzionamenti delle batterie di *backup* all'alimentazione principale delle unità di controllo remoto (*Remote Terminal Unit* - RTU) delle CS, collegate al sistema SCADA (*Supervisory Control and Data Acquisition*) dell'operatore.

Se ci si concentra ad analizzare il FT per la CS, mostrato in Figura 6.2, e si focalizza l'attenzione sulla batteria di *backup*, il rispettivo FT rivela un aspetto cruciale riguardo gli ODL di tale *asset*, e cioè la mancanza di dettagli relativi agli eventi di guasto. Per l'*asset* batteria di *backup* è presente soltanto una voce generica "guasto". Questa mancanza di dettaglio impedisce una comprensione precisa delle specifiche anomalie che hanno afflitto le batterie di *backup*. In termini più generali, sarebbe, dunque, necessaria una diversa definizione del FT, maggiormente esplicativa della tipologia di guasto avvenuto e più vicina a quella mostrata, a titolo di esempio, in Figura 6.3.

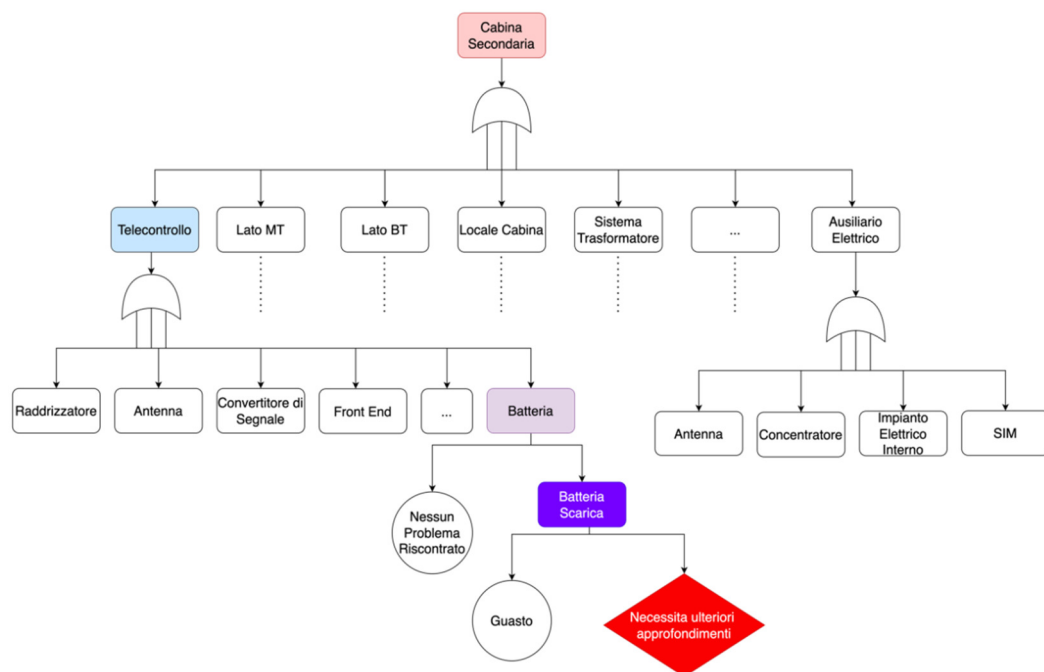


Figura 6.2 - FT delle CS con evidenziato il ramo di riferimento dell'asset batteria di backup.

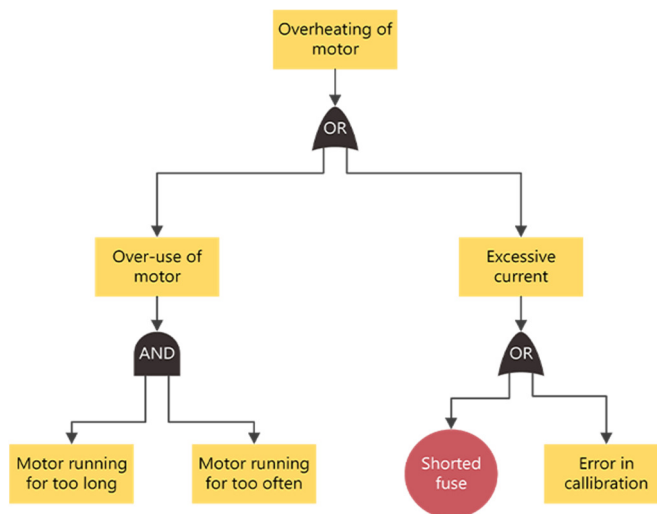


Figura 6.3 - Esempio di fault tree per l'analisi dei possibili guasti ad un motore elettrico.

6.1.3 Dati acquisiti dalle cabine secondarie in ottica manutenzione predittiva

Questa sezione riassume ed analizza i dati acquisiti dai DSO nelle CS in ottica di manutenzione predittiva. La Figura 6.4 schematizza i principali asset presenti nelle CS. All'interno di ogni CS è presente una sezione di media tensione (MT) e una di bassa tensione (BT). Sul lato MT, sono tipicamente presenti uno o più interruttori con i relativi sensori e protezioni. Sono poi presenti uno o più trasformatori MT/BT che collegano la sezione di MT a quella di BT. Quest'ultima è dotata di interruttori BT ed altri dispositivi a cui fanno capo le partenze delle linee BT che escono dalla CS per distribuire l'energia agli utenti finali. Tutti i dati acquisiti dal campo vengono concentrati in un'unità di controllo remoto (RTU) che scambia tali dati e le informazioni acquisite localmente

tramite una rete cablata o *wireless* con la CP di afferenza che può, viceversa, inviare segnali e comandi alla RTU se questa è predisposta per riceverli ed implementarli.

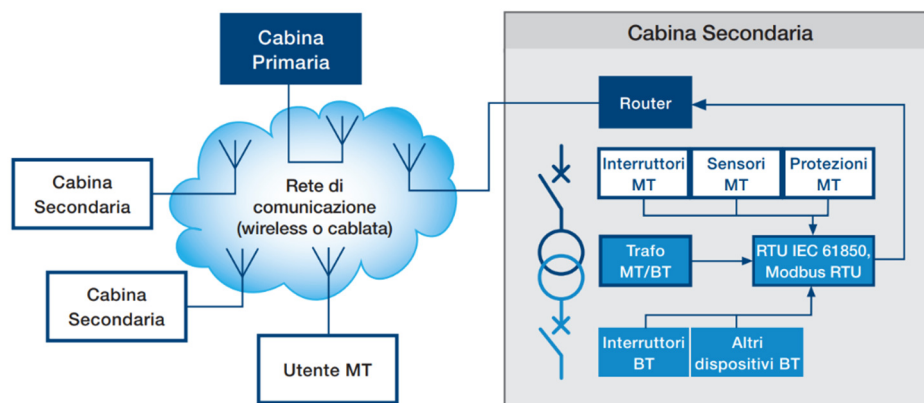


Figura 6.4 - Schema degli asset presenti in cabina secondaria [124].

A seguito di una analisi delle misure acquisite nelle cabine secondarie, sono stati estratti i dati relativi ad alcune batterie di *backup* con una risoluzione temporale spinta, dell'ordine di un campione al secondo. Si è quindi scoperto che, siccome l'archiviazione dei dati è ottimizzata attraverso un approccio basato su soglie, viene memorizzato un nuovo valore delle variabili solo quando si verifica una variazione superiore al 3% rispetto all'ultimo valore misurato. Questo approccio conservativo può portare a lacune importanti nel *dataset*, specialmente se i parametri della batteria di *backup* rimangono relativamente stabili come, per esempio, i valori di tensione. Infatti, utilizzando questo approccio a soglie, se i valori di tensione rimangono stabili o fluttuano all'interno di un piccolo intervallo, molti dati non vengono registrati, portando alla presenza di numerosi *NaN*⁷ nel *dataset*. Un *dataset* di questo genere è in grado di dare evidenza solo dei cambiamenti più significativi, garantendo, allo stesso tempo, una base dati efficace e un'archiviazione efficiente per la maggior parte delle applicazioni legate alla pianificazione e gestione della manutenzione dei sistemi elettrici. Tuttavia, una tale base dati potrebbe non essere adeguata a rendere sufficientemente evidenti le variazioni delle variabili utili allo sviluppo di algoritmi di manutenzione predittiva. Queste lacune nei dati possono quindi rappresentare un ostacolo per ulteriori analisi e per l'addestramento di modelli avanzati di manutenzione predittiva. In particolare, ai fini dell'implementazione di algoritmi di manutenzione predittiva sulla batteria di *backup*, sarebbe opportuna un'acquisizione continua almeno delle variabili principali di funzionamento dell'*asset*, quali, per esempio, la tensione e la corrente in uscita e la sua temperatura. In funzione dell'algoritmo che si desidera utilizzare, la risoluzione temporale minima necessaria può essere leggermente diversa da quella attuale, ma sicuramente sarebbe più che raccomandato avere un intervallo di campionamento superiore ad 1 s. Per concludere, la Figura 6.5 riporta un esempio di andamento delle variabili della batteria di *backup* così come attualmente acquisite.

⁷ NaN (Not a Number) è un acronimo per indicare un valore non numerico sconosciuto, in genere viene usato per indicare lo spazio di un dato mancante.

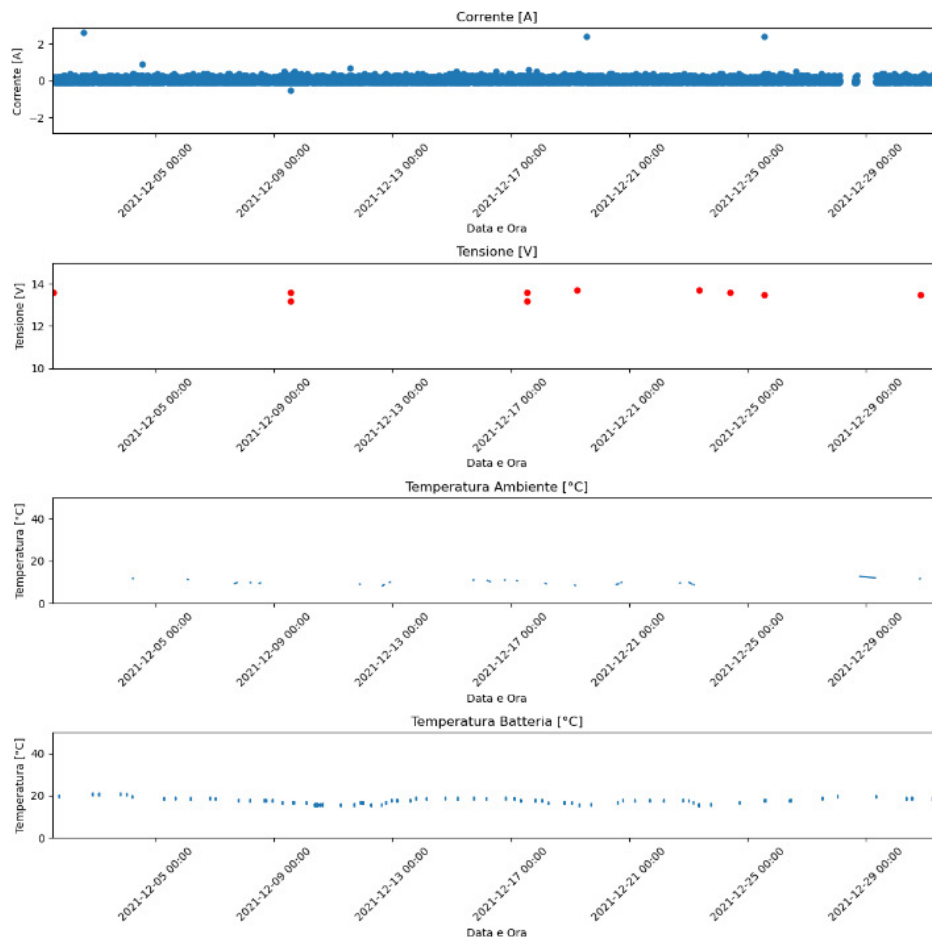


Figura 6.5 - Esempio dell'andamento delle variabili di stato della batteria di *backup* su un arco temporale mensile. Si notino i valori di tensione, particolarmente poco popolati in confronto alla variabile corrente.

Si può notare come la corrente abbia generalmente un andamento costante intorno al valore nullo con la presenza di un limitato rumore di fondo. Il *dataset* della corrente risulta molto più popolato rispetto alle altre variabili. In particolare, il grafico della tensione rivela una situazione da notare: i dati acquisiti con l'approccio a soglia sono in numero notevolmente inferiore, evidenziando una raccolta dati molto sporadica. Questo è evidente dai pochi puntini rossi presenti dal secondo andamento di Figura 6.5, che rappresentano i rari casi in cui le variazioni di tensione hanno superato la soglia impostata per la registrazione del dato nel *database*.

In conclusione, prendendo in considerazione le attuali procedure di gestione degli ordini di lavoro relativi ai processi di guasto e manutenzione, si è focalizzata l'attenzione sul caso degli *asset* delle cabine secondarie. Sono state presentate le modalità di tracciamento del ciclo di vita degli *asset*, la struttura degli ordini di lavoro e gli alberi di guasto utilizzati. Nel caso delle cabine secondarie queste informazioni possono essere utilizzate congiuntamente ai dati acquisiti dalle RTU per implementare algoritmi di manutenzione predittiva. In quest'ottica, una prima analisi dei dati provenienti dalle RTU relativi alle batterie di *backup* ha evidenziato alcuni aspetti del campionamento delle politiche di archiviazione dei dati che possono essere migliorate per il loro utilizzo in algoritmi di previsione. Queste considerazioni possono essere di aiuto al DSO per migliorare la qualità dei dati e per lo sviluppo di algoritmi di manutenzione predittiva.

7 - CONCLUSIONI

Nel presente rapporto sono state trattate tematiche relative alla digitalizzazione dei processi che interessano le reti elettro-energetiche, con particolare attenzione all'interoperabilità a livello informativo tra i sistemi delle utility anche relativi a differenti vettori energetici.

Focalizzandosi sul mondo elettrico, il modello semantico IEC CIM rappresenta lo standard di riferimento per la descrizione dei componenti di rete. Le attività descritte hanno riguardato l'utilizzo della versione più recente dello standard, la versione 17, con l'aggiornamento a questa versione della piattaforma MESP. A partire da sorgenti di dati eterogenee (GIS, SAP, SCADA), è stata convertita in formato IEC CIM la descrizione di reti di distribuzione esistenti. Per l'identificazione automatica della topologia di rete sono stati sviluppati algoritmi di prossimità geografica che hanno permesso di identificare, a seconda del tipo di connessione, tra l'80% ed il 100% delle connessioni topologiche. Le rappresentazioni IEC CIM delle reti saranno utilizzate nella prosecuzione delle attività come base informativa per sviluppare nuove funzionalità di analisi e di manutenzione predittiva delle reti di distribuzione.

Relativamente ai processi di gestione degli asset, ci si è concentrati sulla possibilità di instaurare una sinergia tra la metodologia **BIM** e lo **standard IEC CIM**, con lo scopo di creare un unico *knowledge graph* per la descrizione dei componenti di rete che consenta di poterli gestire in modo ottimale durante il loro intero ciclo di vita. A tal fine, è stata implementata, per un **trasformatore di cabina primaria**, una trasformazione tra le classi IFC e le classi IEC CIM, focalizzandosi sulle informazioni necessarie in ambito operativo per la descrizione del trasformatore. L'**utilizzo di un identificativo univoco** permette di mantenere allineati i due mondi di progettazione BIM e di *planning&operation*: per scopi di **gestione del ciclo di vita** le utility possono continuare a utilizzare tool specifici dell'**ambiente BIM**, mentre per **scopi operativi** si sfrutta il **modello IEC CIM**. Nella prosecuzione verrà integrato sulla piattaforma MESP uno strumento per analisi, modifica e visualizzazione dei modelli IFC: dalle classi IFC sarà possibile passare alla navigazione del *knowledge graph* sfruttando la corrispondenza con le classi IEC CIM. Viceversa, a partire dalla rete IEC CIM, sarà possibile passare al mondo BIM sempre tramite il codice identificativo univoco. Questa integrazione dei diversi modelli in un unico *knowledge graph* garantisce una gestione ottimale e combinata degli asset di rete, durante il loro intero ciclo di vita.

Sempre sulla base di IEC CIM è stata definita una nuova ontologia per la **descrizione di reti gas**. Tramite questa ontologia è stato possibile rappresentare parti di reti gas reali e attualmente descritte in un modello GIS proprietario tramite *shape file*. La definizione dell'ontologia per le reti gas ha riproposto la stessa ereditarietà tra le classi, le relazioni di contenimento e di collegamento tra elementi di **IEC CIM**. Il secondo riferimento utilizzato è stato il modello **Gas Utility Network Configuration**, sviluppato da ESRI; da questa ontologia sono state considerate le definizioni delle classi, dei loro sottotipi e i relativi attributi. Le reti di input, convertite in una serializzazione RDF/XML secondo l'ontologia di destinazione, sono state caricate sulla piattaforma MESP in una sezione dedicata al settore gas. Da qui sono state rese disponibili le funzionalità di **navigazione delle classi** e di **visualizzazione geografica**, presenti anche per gli altri vettori energetici. Nella prosecuzione delle attività, si procederà con l'integrazione sulla piattaforma MESP delle restanti funzionalità disponibili per gli altri vettori energetici, in particolare analisi e visualizzazioni topologiche e risoluzione di problemi di *gas-flow*.

Il **monitoraggio e controllo delle reti energetiche** richiedono un'adeguata **infrastruttura ICT** associata. Per gestire le informazioni e misure provenienti da diversi strumenti e sensori, risulta opportuno poter disporre di una visione unificata della rete energetica e dell'architettura ICT. Nel rapporto è stato illustrato come le classi di IEC CIM possano essere usate anche per descrivere componenti ICT generiche del settore elettro-energetico (ad esempio le classi *Remote Unit* e *Communication Link*). Come esempio, si è dimostrato come queste classi, purché semplici e generiche, sono sufficienti per rappresentare le componenti dell'infrastruttura ICT di un DSO. Infine, è stato presentato uno stato dell'arte relativo alle principali ontologie esistenti per il settore



ICT, che potrebbero essere utilizzate per casi d'uso più complessi. Nello sviluppo dell'attività si prenderanno come riferimento casi d'uso implementati nella Test Facility *multi-energy* di RSE per la descrizione dell'architettura ICT ad essi associata tramite le ontologie.

Nell'ambito della **pianificazione multienergetica del territorio**, sono stati elicitati i principali requisiti per la **digitalizzazione dei processi** di pianificazione ed è stata progettata un'architettura informatica in grado di efficientarli. Si è partiti con l'analisi della parte del processo di preparazione dei dati di ingresso del modello Calliope che si occupa della gestione dei dati dei parametri tecnologici delle tecnologie di generazione. La revisione ha comportato il passaggio da un modello dati relazionale ad un modello semantico basato su un'ontologia. Inoltre, è stato riportato lo studio dello stato dell'arte per lo sviluppo di una metodologia che faccia uso di tecniche di intelligenza artificiale per l'aggiornamento periodico ed automatico della **mappatura di impianti fotovoltaici a terra** a partire da **immagini satellitari**. Nel prosieguo delle attività, sarà sviluppato un componente *software* basato su tecniche di segmentazione semantica e su un modello di intelligenza artificiale che verrà addestrato al riconoscimento degli impianti da immagini satellitari presenti nella Provincia di Lecce e rilevati da Sentinel-2 nel 2019. In seguito, verrà testato su altre provincie ed utilizzato come strumento ausiliario per la valutazione del potenziale territoriale disponibile per le nuove installazioni fotovoltaiche.

Un altro contributo nell'ambito dei processi di pianificazione, in questo caso di reti elettriche, è consistito nella progettazione, sviluppo **ed integrazione** nella piattaforma MESP di servizi aggiuntivi, utili per l'analisi di **scalabilità e replicabilità** (SRA) di soluzioni di flessibilità. Tale realizzazione è stata utilizzata per eseguire le analisi richieste nel contesto del progetto europeo correlato Platone ed ha arricchito la piattaforma MESP di nuovi servizi, disponibili agli utilizzatori che vogliano valutare scenari di evoluzione delle proprie reti.

Per poter **stimare le tensioni ai nodi di una rete elettrica** senza l'ausilio di tecniche analitiche sono stati sviluppati e testati diversi **modelli di machine learning basati sui grafi**. Inizialmente sono stati **valutati diversi framework** per lo sviluppo di modelli di *Graph Machine Learning*: Deep Graph Library (DGL), PyTorch Geometric (PyG) e StellarGraph. La valutazione è stata effettuata rispetto ad un compito *benchmark* di classificazione, sia facendo uso di *dataset* noti dello stato dell'arte, sia considerando il *dataset* preparato per l'attività di stima della tensione. Il confronto delle prestazioni di modellazione e computazionali ha portato alla selezione del framework **Deep Graph Library**. In una seconda fase, a partire dalla descrizione IEC CIM delle reti di test sono state implementate diverse trasformazioni in modelli in formato **graphML**, utili per l'applicazione di tecniche di *graph analysis* e *Graph Machine Learning*. I modelli si differenziano per la differente modalità di trattare le entità di tipo TopologicalNode, che rappresentano gli elementi per cui determinare i valori di tensione. Implementando e applicando i **modelli per la stima dei valori di tensione**, separatamente per ciascuna delle 20 reti elettriche del *dataset* disponibile, il primo modello sviluppato è una **Graph Convolutional Network**, per la quale l'errore medio di previsione rispetto al valore della tensione calcolata è risultato nel range 0,2-3,2%. Applicando un modello **LightGBM**, un esempio di modello *ensemble* basato su alberi decisionali, che permette di tenere in considerazione le informazioni dei nodi vicini per ciascun nodo della rete, si è ottenuto un errore medio di previsione inferiore, compreso nel range 0,1-0,6%. Una possibilità interessante da tenere in considerazione per lo sviluppo dell'attività consiste nell'ibridare i modelli GCN e LightGBM in modo da sfruttare i punti di forza delle due modalità ed ottenere modelli ancora più performanti. Gli algoritmi sviluppati per la rete elettrica verranno presi come riferimento anche per lo sviluppo di un modello per la **stima di grandezze di stato della rete gas**, sempre senza l'ausilio di tecniche analitiche. Come primo passaggio, è stata confermata la possibilità di stabilire degli isomorfismi tra rete elettrica e rete gas, e quindi di poter quindi applicare le metodologie basate su grafi sviluppate per il vettore elettrico anche su un vettore differente.



Infine, nell'ambito dei **processi di manutenzione delle reti**, l'attività relativa all'analisi dei dati sui guasti nelle reti di distribuzione si è focalizzata sull'analisi degli ordini di lavoro e le procedure attualmente seguite per gestire eventi di guasto e di manutenzione, in particolare, per gli **asset** delle **cabine secondarie**. L'analisi dei dati storici ha evidenziato una maggior criticità nei guasti dovuti al malfunzionamento delle batterie di *backup*. L'approfondimento dell'analisi dei dati di tensione, corrente e temperatura delle batterie, provenienti dai sistemi di telecontrollo, ha permesso di formulare indicazioni di miglioramento del monitoraggio di questi componenti per abilitare lo sviluppo e l'utilizzo in algoritmi di **manutenzione predittiva**.



8 - BIBLIOGRAFIA

- [1] European Commission, «Digitalising the energy sector – EU action plan,» 2022. [Online]. Available: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13141-Digitalising-the-energy-sector-EU-action-plan_en.
- [2] European Commission, «EU strategy on energy system integration,» 2020. [Online]. Available: https://energy.ec.europa.eu/topics/energy-systems-integration/eu-strategy-energy-system-integration_en.
- [3] G. Paludetto; E. Bionda; F. Soldan, «MESP - An Interoperable Platform for Multi-Energy Systems,» 2022 AEIT International Annual Conference (AEIT), Roma, 2022.
- [4] E. Bionda, F. Soldan, G. Paludetto, M. Aiello, A. Amaranto, «Modellazione delle reti energetiche per la pianificazione e la gestione delle infrastrutture,» Ricerca di Sistema, RSE, n. 22013991, Milano, 2022.
- [5] Blender Team, «Blender,» [Online]. Available: <https://www.blender.org/>.
- [6] E. Bionda, G. Paludetto, F. Soldan, «Evoluzione della piattaforma dimostrativa Smart Grid Semantic Platform (SGSP) verso una Multi Energy Semantic Platform (MESP),» Ricerca di Sistema, RSE, n. 20010464, Milano, 2020.
- [7] EPRI, «Common Information Model Primer, Eight Edition,» Palo Alto (USA), 2022.
- [8] S. Wang; Z. Meng; S. Yuan, «IEC 61970 standard based common information model extension of electricity-gas-heat integrated energy system,» *Electrical Power and Energy Systems*, vol. 118, n. 105846, 2020.
- [9] IEC 61970, «Energy Management System application program interface (EMS-aPI) - Part 301: Common Information Model (CIM) Base,» 2013.
- [10] IEC 61968, «Application integration at electric utilities - System interfaces for distribution management - Part 11: Common Information Model (CIM),» 2013.
- [11] IEC 62325, «Framework for energy market communications».
- [12] ESRI, «Gas Utility Network Configuration Data Dictionary,» [Online]. Available: <https://solutions.arcgis.com/gas/help/gas-utility-network-configuration/DataDictionary/DataDictionary/>.
- [13] Protégé, «A free, open-source ontology editor and framework for building intelligent systems,» [Online]. Available: <https://protege.stanford.edu/>.
- [14] Z. Syed; A. Padia; T. Finin; L. Mathews; A. Joshi, «UCO: A Unified Cybersecurity Ontology,» *Association for the Advancement of Artificial Intelligence*, 2016.
- [15] International Electrotechnical Commission, «Energy Management System application program interface (EMS-API) - Part 301: Common Information Model (CIM) Base,» 2013.
- [16] A. Koziolk; L. Happe; A. Avritzer; S. Suresh, «A Common Analysis Framework for Smart Distribution Networks Applied to Survivability Analysis of Distribution Automation,» *2012 First International Workshop on Software Engineering Challenges for the Smart Grid (SE-SmartGrids)*, pp. 23-29, 2012.
- [17] Commissione Europea, «Ideal Grid for All,» [Online]. Available: <https://cordis.europa.eu/project/id/608860/it>.
- [18] D. Della Giustina; A. Alvarez de Sotomayor; A. Dedè; F. Ramos, «A Model-Based Design of Distributed Automation Systems for the Smart Grid: Implementation and Validation,» *Energies*, 2020.
- [19] A. Angioni et al., «A distributed automation architecture for distribution networks, from design to implementation,» *Sustainable Energy, Grids and Networks*, vol. 15, pp. 3-13, 2018.
- [20] F. De Rosa; N. Maunero; P. Prinetto; F. Talentino; M. Trussoni, «ThreMA: Ontology-Based Automated Threat Modeling for ICT Infrastructures,» *IEEE Access*, vol. 10, pp. 116514-116526, 2022.
- [21] F. De Rosa; N. Maunero; L. Nicoletti; P. Prinetto; M. Trussoni, «Ontology for Cybersecurity Governance of ICT Systems,» *ITASEC'22: Italian Conference on Cybersecurity, June 20-23, 2022, Rome, Italy*, 2022.
- [22] O. Corcho et al., «A High-Level Ontology Network for ICT Infrastructures,» *The Semantic Web – ISWC 2021*, vol. 12922, 2021.



- [23] «DevOps Infrastructure Ontology Network,» [Online]. Available: <https://github.com/oeg-upm/devops-infra>.
- [24] R. Terruggia; E. Albanese; R. Urban; D. Bartalesi, «Tecnologie di digitalizzazione e comunicazione innovative per l'interoperabilità tra le Test Facility multi-energy di RSE,» Mission Innovation Italy - Progetto MISSION, RSE, n. 22008327, Milano, 2023.
- [25] F. Giudici e L. Stucchi, «Modello multienergetico: software, interoperabilità e applicazione ad un caso di studio,» Ricerca di Sistema, RSE, n. 21010336, Milano, 2021.
- [26] G. Stella; M. Vergata; M. Volonterio, «Il geo DB per il sistema energetico nazionale,» Ricerca di Sistema, RSE, n. 19012819, Milano, 2019.
- [27] Enea Bionda, Gabriele Paludetto, Francesca Soldan, «Validazione dei requisiti iniziali della piattaforma MESP,» Ricerca di Sistema, RSE, n. 21008811, Milano, 2021.
- [28] Roberto Naldi, Marco Tangi, «Sviluppo di un DB per l'organizzazione dei dati delle tecnologie per i sistemi multienergetici,» Ricerca di Sistema, Rapporto RSE n.23013094, Milano, 2023.
- [29] Natalya F. Noy, Deborah L. McGuinness, «Ontology Development 101: A Guide to Creating Your First Ontology».
- [30] E. Garofalo, M. Aiello, D. Airoidi e A. Gargiulo, «Aree idonee: disponibilità potenziale a livello regionale per FV ed eolico anche alla luce di nuove soluzioni tecnologiche,» Ricerca di Sistema, RSE, n.21010392, Milano, 2021.
- [31] «Copernicus - Europe's eyes on Earth,» [Online]. Available: <https://www.copernicus.eu/it>. [Consultato il giorno 20 08 2023].
- [32] J. de Hoog, S. Maetschke, P. Ilfrich e R. R. Kolluri, «Using satellite and aerial imagery for identification of solar pv: State of the art and research opportunities,» in *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, Virtual Event Australia, 2020.
- [33] R. Peters, J. Berlekamp, K. Tockner e C. Zarfl, «RePP Africa—a georeferenced and curated database on existing and proposed wind, solar, and hydropower plants,» *Scientific Data*, vol. 10, n. 1, p. 16, 2023.
- [34] S. Dunnett, A. Sorichetta, G. Taylor e F. Eigenbrod, «Harmonised global datasets of wind and solar farm locations and power,» *Scientific Data*, vol. 7, n. 1, p. 130, 2020.
- [35] D. Stowell, J. Kelly, D. Tanner, J. Taylor, E. Jones, J. Geddes e E. Chalstrey., «A harmonised, high-coverage, open dataset of solar photovoltaic installations in the UK,» *Scientific Data*, vol. 7, n. 1, p. 394, 2020.
- [36] Q. Chen, X. Li, Z. Zhang, C. Zhou, Z. Guo, Z. Liu e H. Zhang, «Remote sensing of photovoltaic scenarios: Techniques, applications and future directions,» *Applied Energy*, vol. 333, p. 120579, 2023.
- [37] T. Sun, M. Shan, X. Rong e X. Yang, «Estimating the spatial distribution of solar photovoltaic power generation potential on different types of rural rooftops using a deep learning network applied to satellite images,» *Applied Energy*, vol. 315, p. 119025, 2022.
- [38] T. Zhong, Z. Zhang, M. Chen, K. Zhang, Z. Zhou, R. Zhu, Y. Wang, G. Lü e J. Yan, «A city-scale estimation of rooftop solar photovoltaic potential based on deep learning,» *Applied Energy*, vol. 298, p. 117132, 2021.
- [39] A. Tiwari, I. A. Meir e A. Karnieli, «Object-based image procedures for assessing the solar energy photovoltaic potential of heterogeneous rooftops using airborne LiDAR and orthophoto,» *Remote Sensing*, vol. 12, n. 2, p. 223, 2020.
- [40] U. Hijjawi, S. Lakshminarayana, T. Xu, G. P. Malfense Ferro e M. Rahman, «A review of automated solar photovoltaic defect detection systems: Approaches, challenges, and future orientations,» *Solar Energy*, vol. 266, p. 112186, 2023.
- [41] G. Roggi, A. Niccolai, F. Grimaccia e M. Lovera, «A Computer Vision Line-Tracking Algorithm for Automatic UAV Photovoltaic Plants Monitoring Applications,» *Energies*, vol. 13, p. 838, 2020.
- [42] R. Zhu, D. Guo, M. S. Wong, Z. Qian, M. Chen, B. Yang, B. Chen, H. Zhang, L. You, J. He e J. Yan, «Deep solar PV refiner: A detail-oriented deep learning network for refined segmentation of photovoltaic areas from satellite imagery,» *International Journal of Applied Earth Observation and Geoinformation*, vol. 116, p. 103134, 2023.

- [43] X. Hou, B. Wang, W. Hu, L. Yin e H. Wu, «SolarNet: a deep learning framework to map solar power plants in China from satellite imagery,» *arXiv preprint*, vol. arXiv:1912.03685, 2019.
- [44] K. Bradbury, R. Saboo, T. L. Johnson, J. M. Malof, A. Devarajan, W. Zhang, L. M. Collins e R. G. Newell, «Distributed solar photovoltaic array location and extent dataset for remote sensing object identification,» *Scientific Data*, vol. 3, n. 1, pp. 1-9, 2016.
- [45] C. Jörges, H. S. Vidal, T. Hank e H. Bach, «Detection of Solar Photovoltaic Power Plants Using Satellite and Airborne Hyperspectral Imaging,» *Remote Sensing*, vol. 15, n. 13, p. 3403, 2023.
- [46] J. M. Malof, R. Hou, L. M. Collins, K. Bradbury e R. Newell, «Automatic solar photovoltaic panel detection in satellite imagery,» in *2015 International Conference on Renewable Energy Research and Applications (ICRERA)*, Palermo, 2015.
- [47] D. W. Czirjak, «Detecting photovoltaic solar panels using hyperspectral imagery and estimating solar power production,» *Journal of Applied Remote Sensing*, vol. 11, n. 2, p. 026007, 2017.
- [48] D. Cerra, C. Ji e U. Heiden, «Solar Panels Area Estimation Using the Spaceborne Imaging Spectrometer Desis: Outperforming Multispectral Sensors,» *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, n. 2022, pp. 9-14, 2022.
- [49] European Space Agency (ESA), «The CEOS Database,» [Online]. Available: <https://database.eohandbook.com/database/missionsummary.aspx?missionID=396>. [Consultato il giorno 18 Gennaio 2024].
- [50] T. Zhang, W. Zhang, R. Yang, Y. Liu e M. Jafari, «CO₂ capture and storage monitoring based on remote sensing techniques: A review,» *Journal of Cleaner Production*, vol. 281, p. 124409, 2021.
- [51] J. Li, Y. Pei, S. Zhao, R. Xiao, X. Sang e C. Zhang, «A Review of Remote Sensing for Environmental Monitoring in China,» *Remote Sensing*, vol. 12, n. 7, p. 1130, 2020.
- [52] D. Radočaj, J. Obhodaš, M. Jurišić e M. Gašparović, «Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review,» *Land*, vol. 9, n. 11, p. 402, 2020.
- [53] A. M. Melesse, Q. Weng, P. S. Thenkabail e G. B. Senay, «Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling,» *Sensors*, vol. 7, n. 12, pp. 3209-3241, 2007.
- [54] Z. Guo, Z. Zhuang, H. Tan, Z. Liu, P. Li, Z. Lin, W.-L. Shang, H. Zhang e J. Yan, «Accurate and generalizable photovoltaic panel segmentation using deep learning for imbalanced datasets,» *Renewable Energy*, vol. 219, p. 119471, 2023.
- [55] X. Wang, X. Xiao, X. Zhang, H. Ye, J. Dong, Q. He, X. Wang, J. Liu, B. Li e J. Wu, «Characterization and mapping of photovoltaic solar power plants by Landsat imagery and random forest: A case study in Gansu Province, China,» *Journal of Cleaner Production*, vol. 417, p. 138015, 2023.
- [56] W. Jiang, B. Tian, Y. Duan, C. Chen e Y. Hu, «Rapid mapping and spatial analysis on the distribution of photovoltaic power stations with Sentinel-1&2 images in Chinese coastal provinces,» *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, p. 103280, 2023.
- [57] Z. Chen, Y. Kang, Z. Sun, F. Wu e Q. Zhang, «Extraction of Photovoltaic Plants Using Machine Learning Methods: A Case Study of the Pilot Energy City of Golmud, China.,» *Remote Sensing*, vol. 14, n. 11, p. 2697, 2022.
- [58] A. Ortiz, D. Negandhi, S. Mysorekar, S. Nagaraju, J. Kiesecker, C. Robinson, P. Bhatia, A. Khurana, J. Wang, F. Oviedo e J. Ferres, «An artificial intelligence dataset for solar energy locations in India,» *Scientific Data*, vol. 9, n. 1, p. 497, 2022.
- [59] V. Plakman, J. Rosier e J. van Vliet, «Solar park detection from publicly available satellite imagery,» *GIScience & Remote Sensing*, vol. 59, n. 1, pp. 462-481, 2022.
- [60] J. Wang, J. Liu e L. Li., «Detecting Photovoltaic Installations in Diverse Landscapes Using Open Multi-Source Remote Sensing Data,» *Remote Sensing*, vol. 14, n. 24, p. 6296, 2022.
- [61] Z. Xia, Y. Li, R. Chen, D. Sengupta, X. Guo, B. Xiong, Y. Niu, Y. Li, R. Chen, D. Sengupta, X. Guo, B. Xiong e Y. Niu, «Mapping the rapid development of photovoltaic power stations in northwestern China using remote sensing,» *Energy Reports*, vol. 8, pp. 4117-4127, 2022.



- [62] X. Zhang, S. Wang, Y. Huang, Z. Xie e M. Xu, «Mapping photovoltaic power plants in China using Landsat, random forest, and Google Earth Engine,» *Earth System Science Data Discussions*, vol. 2022, pp. 1-20, 2022.
- [63] M. Costa, O. Carvalho, A. Orlandi, I. Hirata, A. Albuquerque, F. Silva, R. Guimarães, R. Gomes e O. Júnior, «Remote sensing for monitoring photovoltaic solar plants in Brazil using deep semantic segmentation,» *Energies*, vol. 14, n. 10, p. 2960, 2021.
- [64] T. Kouyama, N. Imamoglu, M. Imai e R. Nakamura, «Verifying Rapid Increasing of Mega-Solar PV Power Plants in Japan by Applying a CNN-Based Classification Method to Satellite Images,» in *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, Waikoloa, HI, USA, 2020.
- [65] D. Jurakuziev, S. Jumaboev e M. Lee, «A framework to estimate generating capacities of PV systems using satellite imagery segmentation,» *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106186, 2023.
- [66] M. Kleebauer, C. Marz, C. Reudenbach e M. Braun, «Multi-Resolution Segmentation of Solar Photovoltaic Systems Using Deep Learning,» *Remote Sensing*, vol. 15, n. 24, p. 5687, 2023.
- [67] J. Yu, Z. Wang, A. Majumdar e R. Rajagopal, «DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States,» *Joule*, vol. 2, n. 12, pp. 2605-2617, 2018.
- [68] S. Diek, F. Fornallaz, M. E. Schaepman e R. D. Jong., «Barest pixel composite for agricultural areas using landsat time series,» *Remote Sensing*, vol. 9, n. 12, p. 1245, 2017.
- [69] C. He, P. Shi, D. Xie e Y. Zhao, «Improving the normalized difference built-up index to map urban built-up,» *Remote Sensing Letters*, vol. 1, n. 4, pp. 213-221, 2010.
- [70] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao e L. G. Ferreira, «Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices,» *Remote Sensing of Environment*, vol. 85, n. 1-2, pp. 195-219, 2002.
- [71] K. Chandrasekar, M. V. R. Sesha Sai, P. S. Roy e R. S. Dwevedi, «Land Surface Water Index (LSWI) response to rainfall and NDVI using the MODIS Vegetation Index product,» *International Journal of Remote Sensing*, vol. 31, n. 15, pp. 3987-4005, 2010.
- [72] Y. Zha, J. Gao e S. Ni, «Use of normalized difference built-up index in automatically mapping urban areas from TM imagery,» *International Journal of Remote Sensing*, vol. 24, n. 3, pp. 583-594, 2003.
- [73] A. P. van Deventer, A. D. Ward, P. H. Gowda e J. G. Lyon, «Using thematic mapper data to identify contrasting soil plains and tillage practices,» *Photogrammetric Engineering and Remote Sensing*, vol. 63, pp. 87-93, 1997.
- [74] J. W. Rouse, R. H. Haas, J. A. Schell e D. W. Deering, «Monitoring vegetation systems in the Great Plains with ERTS,» *NASA Spec. Publ.*, vol. 351, n. 1, p. 309, 1974.
- [75] S. K. McFeeters, «The use of normalized difference water index (NDWI) in the delineation of open water features,» *International Journal of Remote Sensing*, vol. 17, n. 7, pp. 1425-1432, 1996.
- [76] H. Xu, «Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery,» *International Journal of Remote Sensing*, vol. 27, n. 14, pp. 3025-3033, 2006.
- [77] D. K. Hall e G. A. Riggs, «Normalized-difference snow index (NDSI),» *Encyclopedia of snow, ice and glaciers*, 6 Maggio 2010.
- [78] Y. Wang, D. Cai, L. Chen, L. Yang, X. Ge e L. Peng, «A Downscaling Methodology for Extracting Photovoltaic Plants with Remote Sensing Data: From Feature Optimized Random Forest to Improved HRNet,» *Remote Sensing*, vol. 15, n. 20, p. 4931, 2023.
- [79] X. Zhang, M. Zeraatpisheh, M. M. Rahman, S. Wang e M. Xu, «Texture is important in improving the accuracy of mapping photovoltaic power plants: A case study of Ningxia Autonomous Region, China,» *Remote Sensing*, vol. 13, n. 19, p. 3909, 2021.
- [80] L. Kruitwagen, K. T. Story, J. Friedrich, L. Byers, S. Skillman e C. Hepburn, «A global inventory of photovoltaic solar energy generating units,» *Nature*, vol. 598, n. 7882, pp. 604-610, 2021.
- [81] L. Breiman, «Random forests,» *Machine learning*, vol. 45, pp. 5-32, 2001.
- [82] K. Ioannou e D. Myronidis, «Automatic detection of photovoltaic farms using satellite imagery and convolutional neural networks,» *Sustainability*, vol. 13, n. 9, p. 5323, 2021.



- [83] X. Hou, B. Wang, W. Hu, L. Yin e H. Wu, «SolarNet: A Deep Learning Framework to Map Solar Power Plants In China From Satellite Imagery,» *arXiv preprint*, vol. arXiv:1912, p. 03685, 2019.
- [84] O. Ronneberger, P. Fischer e T. Brox, «U-net: Convolutional networks for biomedical image segmentation,» in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference*, Munich, Germany, 2015.
- [85] European Space Agency (ESA), «Sentinel-2 MSI User Guide,» [Online]. Available: <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi>. [Consultato il giorno 29 08 2023].
- [86] A. Agrillo, V. Surace e P. Liberatore, «Rapporto statistico Solare Fotovoltaico 2021. Il solare fotovoltaico in Italia. Stato di sviluppo e trend del settore,» Gestore dei Servizi Energetici S.p.A., Roma, 2022.
- [87] ISPRA, «Consumo di suolo, dinamiche territoriali e servizi ecosistemici. Edizione 2022,» ISPRA, Roma, 2022.
- [88] «Platone (PLATform for Operation of distribution NETworks),» [Online]. Available: <https://www.platone-h2020.eu/>.
- [89] I. Losa, R. Cossent, A. Rodriguez, «Scalability and replicability analysis of smart grids demo projects: An overview of selected European approaches,» *Economics and policy of energy and the environment*, 2016.
- [90] Platone Consortium, «D7.6 Main findings and recommendations,» 31 08 2023. [Online]. Available: https://www.platone-h2020.eu/data/deliverables/864300_M48_D7.6.pdf. [Consultato il giorno 02 01 2024].
- [91] Council of European Energy Regulators (CEER), «6TH CEER benchmarking report on the quality of electricity and gas supply,» 2016. [Online]. Available: <https://www.ceer.eu/documents/104400/-/-/484ca68c-2966-2bfa-f591-of3a1eaf1f52>. [Consultato il giorno 12 01 2024].
- [92] NECSTLab del Politecnico di Milano, «Applicazione di Graph Analysis e Graph Machine Learning a reti energetiche,» RSE, n. 23012119, Milano, 2023.
- [93] DGL Team, «Deep Graph Library,» [Online]. Available: <https://www.dgl.ai>.
- [94] PyG Team, «PyG Documentation,» [Online]. Available: <https://pytorch-geometric.readthedocs.io/en/latest/>.
- [95] StellarGraph Team, «StellarGraph - Machine Learning on Graphs,» [Online]. Available: <https://github.com/stellargraph/stellargraph>.
- [96] A. Kachites McCallum; K. Nigam; J. Renni; K. Seymore, «Automating the Construction of Internet Portals with Machine Learning,» *Information Retrieval*, pp. 127-163, 2000.
- [97] C. Lee Giles; K. D. Bollacker; S. Lawrence, «CiteSeer: an automatic citation indexing system,» *DL '98: Proceedings of the third ACM conference on Digital libraries*, 1998.
- [98] P. Sen; G. Namata; M. Bilgic; L. Getoor; B. Galligher; T. Eliassi-Rad, «Collective Classification in Network Data,» *AI Magazine*, 2008.
- [99] E. Bionda, G. Paludetto, F. Soldan, A. Maldarella, «Analisi ed elaborazione di serie temporali di dati di esercizio e diagnostici del sistema elettrico,» *Ricerca di Sistema, RSE*, n. 21008828, Milano, 2021.
- [100] A. B. Jeddi e A. Shafieezadeh, «A physics-informed graph attention-based approach for power flow analysis,» in *2021 20th IEEE International conference on machine learning and applications (ICMLA)*, 2021, pp. 1634--1640.
- [101] V. Bolz, J. Ruess e A. Zell, «Power flow approximation based on graph convolutional networks,» in *2019 18th IEEE international conference on machine learning and applications*, 2019, pp. 1679--1686.
- [102] D. a. Z. K. Wang, Q. Chen, G. Luo e X. Zhang, «Probabilistic power flow solution with graph convolutional network,» in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, 2020, pp. 650--654.
- [103] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu e K. Weinberger, «Simplifying graph convolutional networks,» in *International conference on machine learning*, 2019, pp. 6861--6871.
- [104] C. J. Willmott e K. Matsuura, «Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,» *Climate research*, vol. 30, n. 1, pp. 79--82, 2005.
- [105] D. M. Hawkins, «The problem of overfitting,» *Journal of chemical information and computer sciences*, vol. 44, n. 1, pp. 1--12, 2004.



- [106] L. Melkumova e S. Y. Shatskikh, «Comparing Ridge and LASSO estimators for data analysis,» *Procedia engineering*, vol. 201, pp. 746-755, 2017.
- [107] H. Zou e T. Hastie, «Regularization and variable selection via the elastic net,» *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, n. 2, pp. 301--320, 2005.
- [108] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever e R. Salakhutdinov, «Dropout: a simple way to prevent neural networks from overfitting,» *The journal of machine learning research*, vol. 15, n. 1, pp. 1929--1958, 2014.
- [109] Z. Reitermanova, «Data splitting,» in *WDS*, 2010, pp. 31--36.
- [110] L. Yang e A. Shami, «On hyperparameter optimization of machine learning algorithms: Theory and practice,» *Neurocomputing*, vol. 415, pp. 295--316, 2020.
- [111] A. Canziani, A. Paszke e E. Culurciello, «An analysis of deep neural network models for practical applications,» *arXiv*.
- [112] Y. Rong, W. Huang, T. Xu e J. Huang, «Dropege: Towards deep graph convolutional networks on node classification,» *arXiv*, 2019.
- [113] A. F. Agarap, «Deep learning using rectified linear units (relu),» *arXiv*, 2018.
- [114] S. Vashishth, S. Sanyal, V. Nitin e P. Talukdar, «Composition-based multi-relational graph convolutional networks,» *arXiv*, 2019.
- [115] S. B. Kotsiantis, «Decision trees: a recent overview,» *Artificial Intelligence Review*, pp. 261--283, 2013.
- [116] L. Grinsztajn, E. Oyallon e G. Varoquaux, «Why do tree-based models still outperform deep learning on typical tabular data?,» *Advances in Neural Information Processing Systems*, vol. 35, pp. 507--520, 2022.
- [117] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye e T.-Y. Liu, «Lightgbm: A highly efficient gradient boosting decision tree,» *Advances in neural information processing systems*, vol. 30, 2017.
- [118] T. Akiba, S. Sano, T. Yanase, T. Ohta e M. Koyama, «Optuna: A next-generation hyperparameter optimization framework,» in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623--2631.
- [119] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince e F. Herrera, «A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches,» *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, n. 4, pp. 463--484, 2011.
- [120] T. G. Dietterich, «Ensemble learning,» *The handbook of brain theory and neural networks*, vol. 2, n. 1, pp. 110--125, 2002.
- [121] B. Pavlyshenko, «Using stacking approaches for machine learning models,» in *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, 2018, pp. 255--258.
- [122] Det Norske Veritas group, «Hydraulic Modelling and Simulation Software - Synergi Gas,» [Online]. Available: <https://www.dnv.com/services/hydraulic-modelling-and-simulation-software-synergi-gas-3894>.
- [123] pandapipes Team, [Online]. Available: <https://www.pandapipes.org/>.
- [124] ABB S.p.A., «Quaderno di Applicazioni Tecniche No. 18 Le smart grid 2. La cabina secondaria smart.,» [Online]. Available: <https://library.abb.com/d/1VCP000620>. [Consultato il giorno 15 Ottobre 2023].
- [125] ISO/IEC 19941:2017, «Information technology - Cloud computing - Interoperability and portability,» [Online]. Available: <https://www.iso.org/standard/66639.html>.
- [126] Regione Emilia-Romagna, «2020 - Coperture vettoriali uso del suolo di dettaglio - Edizione 2023,» [Online]. Available: <https://geoportale.regione.emilia-romagna.it/download/dati-e-prodotti-cartografici-preconfezionati/pianificazione-e-catasto/uso-del-suolo/2020-coperture-vettoriali-uso-del-suolo-di-dettaglio-edizione-2023>. [Consultato il giorno 15 Giugno 2023].
- [127] A. Bosisio e N. Di Giacomo, «Machine learning e big data per la manutenzione predittiva e l'analisi dell'affidabilità delle reti di distribuzione,» Università di Pavia, n. 23012650, 2023.

9 - ACRONIMI

Acronimo	Descrizione
AMM	Automatic Meter Management
AVIRIS-NG	Airborne Visible InfraRed Imaging Spectrometer–Next Generation
BIM	Building Information Modeling
BT	Bassa Tensione
CIM	Common Information Model
CIS	Customer Information System
CNN	Convolutional Neural Network
CP	Cabina Primaria
CPU	Central Processing Unit
CS	Cabina Secondaria
DER	Distributed Energy Resources
DESI	DLR Earth Sensing Imaging Spectrometer
DGL	Deep Graph Library
DL	Deep Learning
DSO	Distribution System Operator
ESRI	Environmental System Research Institute
ETL	Extract, Transform, Load
FER	Fonti Energetiche Rinnovabili
FT	Fault Tree
FTA	Fault Tree Analysis
FV	Fotovoltaico
GBM	Gradient Boosting Machine
GCN	Graph Convolutional Network
GIS	Geographic Information System
GPU	Graphics Processing Unit
GRF	Gruppo di Riduzione Finale
GSE	Gestore dei Servizi Energetici
HEMS	House Energy Management System
ICT	Information and Communication Technology
IDE4L	Ideal Grid for All
IEC	International Electrotechnical Commission
IED	Intelligent Electronic Device



Acronimo	Descrizione
IFC	Industry Foundation Classes
IoU	Intersection over Unit
IRI	Impianto di Riduzione Intermedio
ISS	Stazione Spaziale Internazionale
KG	Knowledge Graph
LoG	Level of Geometry
MAE	Mean Absolute Error
MGCC	MicroGrid Central Controller
ML	Machine Learning
MSE	Mean Squared Error
MT	Media Tensione
NIR	Near infrared
NIS	Network Information System
ODL	Ordine di lavoro
OPF	Optimal Power Flow
OSM	OpenStreetMap
OWL	Web Ontology Language
P2G	Power To Gas
P2H	Power To Hydrogen
PMU	Phasor Measurement Unit
PRISMA	PRecursore IperSpettrale della Missione Applicativa
PyG	PyTorch Geometric
RDF	Resource Description Framework
RdS	Ricerca di Sistema
REMI	REgolazione e MIsurata
ReLU	Rectified Linear Unit
RF	Random Forest
RGB	Red Green Blue
RGCN	Relational Graph Convolutional Network
RS	Remote Sensing
RTU	Remote Terminal Unit
SAP	Systeme, Anwendungen, Produkte in der Datenverarbeitung
SCADA	Supervisory Control and Data Acquisition
SCL	Scene Classification Layer



Acronimo	Descrizione
SRA	Scalabilità e Replicabilità
SWIR	Short-wave infrared
UML	Unified Modeling Language
XML	eXtensible Markup Language
VIS	Visibile