

TESI DI LAUREA IN
INFORMATICA

Thesis Title

CANDIDATO

Francesco Bortuzzo

RELATORE

Professor Angelo Montanari

CORRELATORE

Dottor Andrea Brunello

Dottor Nicola Saccomano

Anno Accademico 2023/2024

INDICE

1. Introduzione e obiettivo del progetto	1
1.1. Introduzione al batterio Legionella	2
1.2. Legionella in Friuli Venezia Giulia	2
1.3. Obiettivo del progetto	3
2. Glossario	5
3. Analisi critica del database relazionale	8
3.1. Analisi dei requisiti	8
3.1.1. Requisiti non strutturati	8
3.2. Schema relazionale	9
3.2.1. Notazione IDEF1X	9
3.2.2. Schema concettuale-logico	11
3.3. Analisi dello schema	11
3.3.1. Considerazioni e proposte di modifica	11
3.3.2. Diagramma E-R	13
4. Integrazione dei nuovi requisiti nella base di dati	13
4.1. Nuovi requisiti e proposte di modifica dello schema	14
4.2. Diagramma E-R	16
4.2.1. Considerazioni	16
4.3. Diagramma E-R finale	16
4.4. Schema relazionale	16
5. Progettazione fisica	16
5.1. Creazione delle tabelle	16
5.2. Implementazione dei vincoli di integrità	16
5.3. Esempi di applicazioni	16
6. Conclusioni	16
7. Bibliografia	18

INDICE

<u>1. Introduzione e obiettivo del progetto</u>	<u>1</u>
1.1. Introduzione al batterio Legionella	2
1.2. Legionella in Friuli Venezia Giulia	2
1.3. Obiettivo del progetto	3
<u>2. Glossario</u>	<u>5</u>
<u>3. Analisi critica del database relazionale</u>	<u>8</u>
3.1. Analisi dei requisiti	8
3.2. Schema relazionale	9
3.3. Analisi dello schema	11
<u>4. Integrazione dei nuovi requisiti nella base di dati</u>	<u>13</u>
4.1. Nuovi requisiti e proposte di modifica dello schema	14
4.2. Diagramma E-R	16
4.3. Diagramma E-R finale	16
4.4. Schema relazionale	16
<u>5. Progettazione fisica</u>	<u>16</u>
5.1. Creazione delle tabelle	16
5.2. Implementazione dei vincoli di integrità	16
5.3. Esempi di applicazioni	16
<u>6. Conclusioni</u>	<u>16</u>
<u>7. Bibliografia</u>	<u>18</u>

1. INTRODUZIONE E OBIETTIVO DEL PROGETTO

1.1. INTRODUZIONE AL BATTERIO LEGIONELLA

Il batterio legionella è un bacillo gram-negativo aerobio, non mobile, che prospera in ambienti acquatici e umidi, sia naturali, come acque sorgive, termali, di fiumi o laghi, sia artificiali, come tubature, serbatoi, fontane e piscine. La legionella è in grado di sopravvivere in una vasta gamma di condizioni ambientali, tra cui temperature comprese tra venti e quarantacinque gradi Celsius, pH neutro o leggermente alcalino, e presenza di nutrienti organici.

Il genere comprende sessantadue specie diverse, suddivise in settantuno sierotipi, di cui circa venti sono patogeni per l'uomo. La specie più comune è *Legionella pneumophila*, responsabile della maggior parte dei casi riportati di legionellosi¹. La malattia può essere contratta inalando aerosol contenenti il batterio, come quelli prodotti da docce, fontane, impianti di condizionamento o umidificatori.

È quindi di fondamentale importanza monitorare la diffusione di questo batterio negli ambienti umidi e acquatici; particolare attenzione deve essere rivolta alle strutture ospedaliere, termali e alberghiere, che rappresentano per loro natura ambienti a rischio di diffusione del batterio.

1.2. LEGIONELLA IN FRIULI VENEZIA GIULIA

A livello europeo, la raccolta di dati relativi alla presenza del batterio è effettuata dall'ECDC². Nel nostro Paese, invece, questa attività è svolta da diversi enti e istituzioni. Un contributo significativo proviene dall'Istituto Superiore di Sanità e dai vari organismi che costituiscono il SNPA³, di cui fa parte l'ARPA FVG⁴.

I dati raccolti sono utilizzati per valutare il rischio di diffusione del batterio e adottare le misure di prevenzione e controllo indicate dal Ministero della Salute nelle *Linee guida per la prevenzione ed il controllo della legionellosi*⁵.

¹Legionellosi, o malattia del legionario, è una malattia infettiva che si presenta con sintomi simili all'influenza, come febbre, tosse, dolori muscolari e mal di testa. In alcuni casi, può evolvere in una forma polmonare, con sintomi analoghi a quelli della polmonite, e comportare complicazioni gravi, come polmonite atipica o decesso.

²Centro europeo per la prevenzione e il controllo delle malattie, istituito nel 2005.

³Sistema Nazionale per la Protezione dell'Ambiente

⁴Agenzia Regionale per la Protezione dell'Ambiente Friuli Venezia Giulia

⁵[1] «Linee guida per la prevenzione e il controllo della legionellosi». [Online]. Disponibile su: <https://www.salute.gov.it/portale/malattieInfettive/dettaglioPubblicazioniMalattieInfettive.jsp?id=2362>

In questo ambito, l'ARPA FVG ha effettuato numerose indagini sul territorio e ha pubblicato i risultati in vari report. Ad esempio, nel 2019, una collaborazione con l'Università degli Studi di Udine ha portato alla pubblicazione di un articolo⁶, relativo alla presenza di legionella nei sistemi di raccolta e distribuzione dell'acqua nella regione. Lo studio ha coperto un periodo di sedici anni, dal 2002 al 2017, durante il quale sono stati raccolti e analizzati 20.319 campioni attraverso 3.983 indagini ambientali.

I risultati riferiti alle indagini non cliniche e dunque eseguite routinariamente nell'ambito del piano regionale di sorveglianza ambientale, hanno evidenziato che la presenza di legionella è diffusa soprattutto nei cluster di impianti termali, nei quali il batterio è stato individuato nel 57,8% dei siti indagati, e ospedalieri, in cui nel 50,8% delle strutture è stata riscontrata la legionella almeno una volta, con picchi dei campioni positivi soprattutto nei mesi che segnano l'inizio del periodo autunnale.

Inoltre, si è osservato che la presenza del batterio ha registrato un notevole incremento tra la seconda metà del 2006 e l'inizio del 2009, seguito da una diminuzione fino al 2013 e da un nuovo aumento negli anni successivi. Questo andamento indica chiaramente che, per ridurre il rischio di diffusione del batterio, è essenziale implementare un piano di prevenzione adeguato, che comprenda sia la manutenzione degli impianti sia la sorveglianza ambientale.

1.3. OBIETTIVO DEL PROGETTO

Pur riconoscendo l'importanza cruciale della sorveglianza ambientale per il controllo della legionellosi, in Friuli Venezia Giulia, come in molte altre regioni, manca un sistema efficiente per la memorizzazione, la gestione e l'analisi dei dati raccolti. Tale carenza rende estremamente oneroso lavorare con la mole di informazioni raccolte nelle indagini ambientali, ostacolando così lo svolgimento di analisi e ricerche mirate.

In questo contesto, i sistemi di basi di dati giocano un ruolo fondamentale, in quanto permettono di memorizzare grandi quantità di dati e di effettuare ricerche complesse in modo rapido ed efficiente.

Il presente documento si propone di delineare gli aspetti principali per la progettazione di un database relazionale destinato alla memorizzazione dei dati relativi alla diffusione della legionella. In particolare, verrà effettuata un'analisi

⁶[2] A. Felice, M. Franchi, S. De Martin, N. Vitacolonna, L. Iacumin, e M. Civilini, «Environmental surveillance and spatio-temporal analysis of *Legionella* spp. in a region of northeastern Italy (2002–2017)», *PLOS ONE*, vol. 14, fasc. 7, p. e218687, 2019, doi: [10.1371/journal.pone.0218687](https://doi.org/10.1371/journal.pone.0218687).

critica di una soluzione esistente e saranno proposte modifiche per adattarla alle nuove esigenze emerse dai colloqui condotti in collaborazione con i ricercatori dell'ARPA FVG.

2. GLOSSARIO

Al fine di facilitare la comprensione del documento, è redatto il seguente glossario contenente le definizioni dei termini tecnici utilizzati.

Termine	Definizione
Aerosol	Particelle sospese nell'aria, contenenti gocce d'acqua, che possono trasportare il batterio Legionella
Analisi	Esame di laboratorio effettuato su campioni di acqua prelevati durante un'indagine ambientale
Analisi Colturale	Esame di laboratorio che permette di isolare e identificare le unità formanti colonia (UFC_L) di legionella in un campione di acqua
Campione	Piccola quantità di acqua da sottoporre a esame
Categoria	Classificazione di un sito, o più specificamente di un edificio, in base alla sua destinazione d'uso, come ad esempio ospedaliero, termale o alberghiero
Componente idraulica	Componente di un sistema idraulico da cui viene prelevato un campione di acqua, come un rubinetto o un filtro di un impianto di condizionamento
Entità	In riferimento allo schema E-R, descrive una classe di oggetti con esistenza autonoma, con particolare significato nel contesto in esame. (sin. tabella)

Glossario

Termine	Definizione
FollowUp Clinico	Indagine ambientale, o indagini ambientali, condotte a seguito di uno o più casi di legionellosi. Tali indagini non si limitano al domicilio del paziente, ma possono estendersi a tutti i luoghi frequentati dal malato nei dieci giorni precedenti l'insorgenza dei sintomi. La decisione di effettuare tali indagini è lasciata al competente servizio territoriale che «deve valutare di volta in volta l'opportunità di effettuare o meno dei campionamenti ambientali, sulla base della valutazione del rischio» ⁷
Indagine Ambientale	Collezione di campioni prelevati da un sito specifico in una data specifica
PCR	Polymerase Chain Reaction, è una «tecnica di laboratorio per produrre rapidamente (amplificare) milioni o miliardi di copie di uno specifico segmento di DNA, che può poi essere studiato in modo più dettagliato. La PCR prevede l'uso di brevi frammenti di DNA sintetico chiamati primer per selezionare un segmento del genoma da amplificare, e quindi più cicli di sintesi del DNA per amplificare quel segmento» ⁸ .
PCR Qualitativa	Esame di laboratorio che fornisce un'informazione dicotomica sulla presenza di legionella in un campione
PCR Quantitativa	Esame di laboratorio rapido che rileva e quantifica il DNA o l'RNA di legionella presenti in un campione. (sin. Real-Time PCR)
Relazione	In riferimento allo schema E-R, legame che rappresenta la connessione logica, e significativa per la realtà modellata, tra due o più entità.
Richiedente	Ente o istituzione che ha richiesto un'indagine ambientale

Termine	Definizione
Sierotipo	Livello di classificazione di batteri di legionella inferiore a quello specie. Il laboratorio ARPA distingue tre sierotipi: sierotipo 1, sierotipo 2-15 e sierotipo sp (sin. sierogruppo)
Sito	Edificio presso il quale è condotta un'indagine ambientale
UFC_L	Unità formanti colonie per litro, unità di misura utilizzata per indicare la concentrazione di Legionella in un campione d'acqua destinato all'analisi colturale.
UG_L	Microgrammi per litro, unità di misura utilizzata per determinare la concentrazione di Legionella in un campione d'acqua mediante PCR quantitativa.

Glossario

⁷[1, Linee guida per la prevenzione ed il controllo della legionellosi, p. 30])

⁸[3] M. Smith, «Polymerase Chain Reaction». Consultato: 31 agosto 2024. [Online]. Disponibile su: <https://www.genome.gov/genetics-glossary/Polymerase-Chain-Reaction>

3. ANALISI CRITICA DEL DATABASE RELAZIONALE

Come accennato nel capitolo introduttivo, una delle principali sfide riscontrate nell'attuale sistema di gestione dei dati riguarda la realizzazione di soluzioni efficienti per la memorizzazione delle informazioni raccolte durante le indagini ambientali. In questa sezione si procede a un'analisi critica di un database relazionale utilizzato per archiviare i dati relativi alla diffusione della legionella. Il database oggetto di analisi è stato sviluppato dal dottor Dario Garlatti nell'ambito della sua tesi di laurea triennale in informatica, dal titolo "Base di dati e applicazione web per il monitoraggio del batterio della legionella".

3.1. ANALISI DEI REQUISITI

Prima di procedere con lo studio del database, è necessario definire i requisiti del sistema informativo. Questi requisiti sono di natura qualitativa e descrivono le caratteristiche che il sistema deve possedere per soddisfare le esigenze degli utenti e degli stakeholder. I requisiti alla base della progettazione della soluzione in analisi riguardano l'intera fase di acquisizione dei dati relativi alle indagini ambientali portate a termine dai ricercatori dell'ARPA FVG, per il monitoraggio della legionella in regione.

3.1.1. REQUISITI NON STRUTTURATI

Il sistema deve consentire la registrazione delle indagini ambientali riguardanti la presenza di Legionella nei sistemi di adduzione e conservazione dell'acqua.

Un'indagine ambientale è definita dal tipo, dalla data e dal sito presso cui viene eseguita, ed è associata al richiedente nel caso di indagine di follow-up.

Un sito è identificato dall'indirizzo e dalla categoria a cui appartiene.

L'indagine comprende il prelievo di campioni che saranno analizzati per rilevare la presenza del batterio legionella. Ciascun campione prelevato è collegato a una specifica indagine, caratterizzato dal punto di prelievo all'interno del sito, e identificato da un codice univoco.

Tutti i campioni prelevati devono essere sottoposti a diverse analisi per accertare la presenza o l'assenza di legionella:

1. PCR qualitativa: una tecnica di reazione a catena della polimerasi che consente di rilevare la presenza del DNA di *Legionella* nei campioni prelevati.
2. PCR quantitativa: permette di misurare la concentrazione di legionella nei campioni prelevati (espressa in µg/l).
3. Analisi colturale: consente di isolare e identificare le unità formanti colonia (UFC_L) e, nel caso in cui il campione risulti positivo al batterio, di determinare il sierogruppo.

Si segnala che la PCR non costituisce un metodo diagnostico definitivo per la legionellosi, ma piuttosto un test di screening che necessita di conferma attraverso la coltura. Infatti, «poiché, così come specificato nella norma ISO “*Water quality- Detection and quantification of Legionella spp and/or Legionella pneumophila by concentration and genic amplification by quantitative polymerase chain reaction (qPCR)*” (ISO/TS 12869, 2012), la qPCR non dà informazione riguardo lo stato delle cellule, la quantificazione dovrà sempre essere determinata mediante esame colturale»⁹. I metodi analitici impiegati per la rilevazione del batterio, descritti nell'allegato 4 delle Linee Guida per la prevenzione e il controllo della legionellosi, variano a seconda della matrice da analizzare (acqua, biofilm, aria); tuttavia, i risultati ottenuti sono espressi in maniera uniforme, indipendentemente dalla tipologia di analisi condotta.

3.2. SCHEMA RELAZIONALE

Di seguito viene presentato lo schema concettuale-logico del database sviluppato dal dottor Garlatti. Tale schema è stato modellato utilizzando il linguaggio IDEF1X¹⁰. Questo linguaggio appartiene alla famiglia dei linguaggi di modellazione IDEF¹¹. Per una corretta comprensione dello schema, è essenziale definire i concetti di entità e relazione, che rappresentano i fondamenti della modellazione dei dati.

3.2.1. NOTAZIONE IDEF1X

Nella notazione IDEF1X, le entità sono rappresentate attraverso tabelle contenenti attributi che ne descrivono le proprietà, e ciascuna entità è identificata da una chiave primaria, costituita da un singolo attributo o da una combinazione di attributi in grado di identificare univocamente ogni riga della tabella. Un'en-

⁹[1, Linee guida per la prevenzione ed il controllo della legionellosi, p. 21])

¹⁰Integration DEFinition for information modeling

¹¹<https://www.idef.com/>

tità può essere classificata come indipendente se può essere identificata senza necessità di relazioni con altre entità, mentre si considera dipendente quando il suo significato emerge solo in relazione a un'altra tabella associata.

Le relazioni di connessione, o associazioni, sono rappresentate mediante linee che collegano due entità, segnalando l'esistenza di un legame tra di esse. In particolare, si distinguono due tipi di relazioni: le associazioni identificative, in cui l'entità figlia è identificata in relazione all'entità genitore e la cui chiave primaria include quella del genitore, rappresentate da una linea continua; e le associazioni non identificative, in cui l'entità figlia è comunque identificata in relazione all'entità genitore, ma la chiave primaria della figlia non include quella del genitore, rappresentate da una linea tratteggiata. La cardinalità di queste associazioni è indicata da lettere: "p" denota una relazione uno a uno o molti, "z" indica una relazione uno a zero o uno, e "n" specifica una relazione uno a esattamente "n".

Le relazioni di categorizzazione, invece, sono rappresentate da linee che collegano un'entità genitore a una o più entità figlie, sottolineando che queste ultime ereditano le proprietà dell'entità genitore, pur mantenendo attributi distintivi. Le entità di categoria¹² sono mutuamente esclusive e si distinguono grazie a un attributo discriminatore, il cui valore è univoco per ciascuna entità di categoria. Esistono due tipologie di categorizzazione: le categorizzazioni complete, in cui ogni entità genitore deve essere associata a una figlia, rappresentate da un pallino vuoto e due linee; e le categorizzazioni incomplete, in cui un'entità genitore può non essere associata a nessuna entità figlia, rappresentate da un pallino pieno e una linea.

¹²entità che costituisce un sottotipo di un'altra

3.2.2. SCHEMA CONCETTUALE-LOGICO

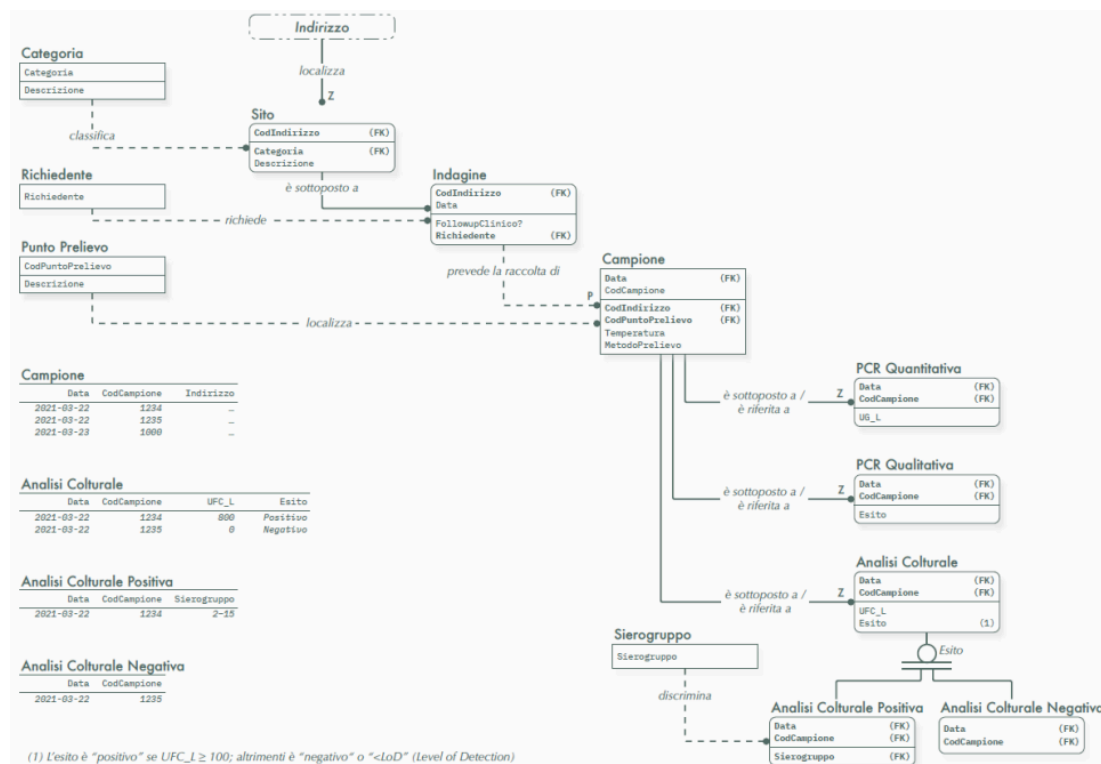


Figura 1: Diagramma ER

3.3. ANALISI DELLO SCHEMA

Lo schema illustrato è stato concepito per rispondere ai requisiti di memorizzazione dei dati relativi alla diffusione della legionella. Tuttavia, durante una prima fase di analisi del database, sono stati individuati alcuni difetti che richiedono un'accurata valutazione e una potenziale revisione dello schema.

3.3.1. CONSIDERAZIONI E PROPOSTE DI MODIFICA

Alcune entità, come *indirizzo* e *categoria*, sono state inizialmente progettate come entità autonome, ma potrebbe essere più efficace trattarle come attributi dell'entità *sito*. Questo approccio non solo semplificherebbe lo schema, ma migliorerebbe anche la sua chiarezza strutturale. In particolare, l'attributo descrizione dell'entità *categoria* è superfluo, poiché il nome della categoria dovrebbe bastare a identificarla in modo univoco. Inoltre, l'aggiunta di un attributo "nome" all'entità *sito* potrebbe facilitare la consultazione dei dati, specialmente per quanto riguarda gli ospedali, che sono generalmente riconosciuti dalla combinazione di nome e città, piuttosto che unicamente dall'indirizzo. In aggiunta, si propone di arricchire l'entità *sito* con nuovi attributi che ne descrivano le

caratteristiche principali nel contesto specifico. Questi attributi includerebbero dettagli sull'impiantistica del sito, come la tipologia di caldaia, il materiale delle tubature, l'uso del cloro, e altre informazioni di carattere generale, come l'anno dell'ultima ristrutturazione.

Un ulteriore elemento di riflessione riguarda l'associazione dell'entità *richiedente* alle *indagini ambientali*. Superando quanto indicato nei requisiti, si propone che l'entità *richiedente* possa essere messa in relazione con indagini che non siano unicamente di follow-up. Inoltre, si suggerisce l'introduzione di una nuova entità denominata *follow-up clinico*, associata a una o più indagini ambientali. Questa modifica si dimostra particolarmente efficace nella gestione dei dati relativi ai pazienti affetti da legionellosi e nella valutazione del rischio di diffusione del batterio. Infatti, «per avere un quadro globale della situazione, è fondamentale disporre, per ciascun paziente affetto da legionellosi, di informazioni precise su una eventuale esposizione a rischio nei dieci giorni precedenti l'insorgenza dei sintomi»¹³. La possibilità di associare un paziente a una o più indagini ambientali risulterebbe, dunque, vantaggiosa.

Si osserva che l'entità *follow-up clinico* potrebbe essere ulteriormente arricchita con attributi volti a descrivere il paziente e la sua esposizione al rischio, quali la data di insorgenza dei sintomi, il luogo di residenza, il luogo di lavoro e le attività svolte nei dieci giorni precedenti l'insorgenza dei sintomi. Questi dettagli, tuttavia, non sono modellati nello schema attuale né saranno inclusi nello schema finale, poiché non sono stati considerati nei requisiti né approfonditi con i ricercatori. Tuttavia, potrebbero rivelarsi utili per una valutazione più accurata del rischio di diffusione del batterio.

Per quanto concerne l'entità *campione*, è opportuno valutare l'introduzione di un attributo "volume" per specificare la quantità d'acqua prelevata per l'analisi. Sebbene non strettamente necessario, tale attributo trova pertinenza nel definire parametri di riferimento relativi al prelievo dei campioni, come il volume minimo d'acqua richiesto per eseguire tutte le analisi previste. Inoltre, poiché è possibile prelevare campioni di diversa matrice ambientale, come acqua, biofilm o aria, si propone di introdurre un attributo "matrice" che consenta di specificare il tipo di campione analizzato.

Infine, si propone di riorganizzare la disposizione delle entità indagine ambientale e campione all'interno dello schema. In particolare, per come definita nella sezione 2, un'indagine ambientale non è altro che una collezione di campioni prelevati in un sito specifico in una data determinata. Pertanto, risulta più coerente associare solo l'entità *campione* alle informazioni spaziali contenute nelle tabelle *punto di prelievo* e *sito*. Si noti che tale modifica comporta l'intro-

¹³[1, Linee guida per la prevenzione ed il controllo della legionellosi, p. 30])

In questo contesto, appare vantaggioso apportare una lieve modifica alla struttura delle entità *sito* e *punto di prelievo* nel modo seguente: si consiglia di aggiungere l'attributo *coordinate* all'entità *sito*, associandolo a una coppia di coordinate, ad esempio riferite al centro geografico o all'ingresso principale dell'edificio, che costituirebbero una chiave per l'entità. Inoltre, l'entità *punto di prelievo* potrebbe essere trasformata in un'entità debole rispetto al *sito*, implicando il vincolo imposto dall'associazione di un *punto di prelievo* a un *sito*, secondo il quale un *punto di prelievo* deve essere situato all'interno del perimetro del *sito* di cui fa parte. Al *punto di prelievo* potrebbero essere associati attributi che ne descrivano la posizione all'interno del *sito*, come il piano, la stanza o il tipo di componente idraulico, da cui è stato prelevato il campione.

3.3.2. DIAGRAMMA E-R

13

4. INTEGRAZIONE DEI NUOVI REQUISITI NELLA BASE DI DATI

Come accennato in precedenza, la progettazione concettuale della base di dati deve essere adeguata alle nuove esigenze emerse a seguito dei colloqui con i ricercatori dell'ARPA FVG. In questa sezione si procede con l'integrazione dei nuovi requisiti nella base di dati, partendo dallo schema concettuale-logico proposto alla conclusione del capitolo precedente.

4.1. NUOVI REQUISITI E PROPOSTE DI MODIFICA DELLO SCHEMA

Le nuove informazioni sono finalizzate a rendere la base di dati più completa e funzionale. In particolare, è stata considerata l'opportunità di introdurre ulteriori entità e attributi, allo scopo di memorizzare dati aggiuntivi relativi ai campioni raccolti nel corso delle indagini ambientali e ai siti coinvolti. Di seguito sono elencati i requisiti non strutturati che hanno guidato l'integrazione dei nuovi elementi.

Si ritiene opportuno mantenere le informazioni relative agli aspetti meteorologici e climatici dei siti in cui vengono condotte le indagini ambientali, poiché tali dati possono essere utili per valutare l'impatto delle condizioni ambientali sulla diffusione del batterio e per individuare eventuali correlazioni tra la presenza di *Legionella* e i fattori climatici. Tali informazioni sono raccolte presso le stazioni meteorologiche presenti sul territorio e comprendono dati relativi a temperatura, umidità e pressione atmosferica. Nella base di dati si propone di introdurre un'entità denominata *stazione meteorologica*, identificata dalla posizione geografica, che può essere rappresentata attraverso l'indirizzo e/o le coordinate e che conserva i dati meteorologici raccolti. Questa entità è associata all'entità *sito* nel seguente modo: ogni sito è in relazione con la stazione meteorologica più vicina, la quale fornisce i dati relativi alle condizioni climatiche del luogo.

Una seconda considerazione riguarda l'opportunità di ampliare il campo di azione delle analisi condotte sui campioni prelevati durante le indagini ambientali. In particolare, si propone di introdurre un nuovo tipo di analisi, denomi-

nata *analisi del pH*, volta a misurare il livello di acidità o alcalinità dell'acqua campionata. Questo parametro è di fondamentale importanza per valutare la qualità dell'acqua e la presenza di Legionella, poiché il batterio prospera in acque con pH neutro o leggermente alcalino.

Sempre in relazione alle analisi condotte sui campioni, durante i colloqui è emersa la volontà di memorizzare le informazioni genomiche relative al batterio. In particolare, si intende raccogliere dati sulla presenza, o assenza, di specifici geni e individuare i fattori genetici che influenzano la diffusione del batterio. A tale scopo, è opportuno eseguire un'analisi genomica sui campioni prelevati per identificare la sequenza del DNA di Legionella. Questa informazione è memorizzata in un'entità *analisi genomica*, che rappresenta una specializzazione dell'entità *analisi*, e contiene l'intera sequenza del DNA di legionella, espressa mediante le quattro lettere che indicano le basi azotate (A, T, C, G).

A ciascun genoma sequenziato si intende associare i geni noti di Legionella, presenti nei database di riferimento di BLAST¹⁴corrispondenti. Tali geni sono memorizzati in un'entità *gene*, identificata univocamente mediante una chiave corrispondente al proprio protein ID¹⁵ e caratterizzata dal nome del gene, se presente nel database utilizzato per l'analisi. A questa entità, che ha lo scopo di conservare informazioni stabili e ben definite sui geni noti di legionella, si propone di associare un'entità *gene del genoma*, che rappresenta i geni individuati per ogni genoma sequenziato. Si tiene traccia, tramite i parametri riportati, del fattore di similarità tra i geni noti e quelli individuati tramite l'analisi di BLAST. Questo approccio ha lo scopo di consentire, in futuro, a seguito del progresso delle tecniche di riconoscimento genetico e dell'espansione dei dati disponibili, una rivalutazione dei geni identificati al fine di determinare se emergano geni con maggiore somiglianza rispetto a quelli attualmente presenti nel genoma analizzato.

Infine, si intende registrare per ciascun *gene del genoma* la sua posizione relativa rispetto agli altri geni all'interno del profilo genetico sequenziato. Questa informazione è essenziale per valutare la prossimità tra i geni e per identificare eventuali pattern di distribuzione specifici all'interno del genoma di Legionella. In termini pratici, si propone di introdurre una relazione auto-referenziale che coinvolga l'entità *gene del genoma*, al fine di stabilire un legame tra i geni identificati e la loro posizione, relativa¹⁶ all'interno del genoma sequenziato. La cardinalità di tale relazione sarà definita come 0,2 a 0,2, indicando che ogni gene può essere associato a zero, uno, o al massimo due geni vicini. Questa con-

¹⁴Basic Local Alignment Search Tool <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

¹⁵Identificativo univoco associato a ciascuna proteina mappata nei database di riferimento di BLAST

¹⁶definita in relazione alla prossimità ad altri geni all'interno del genoma sequenziato

figurazione tiene conto della limitata conoscenza attuale sui geni di Legionella, che potrebbe comportare l'assenza di associazioni per alcune aree del genoma.

4.2. DIAGRAMMA E-R

4.2.1. CONSIDERAZIONI

4.3. DIAGRAMMA E-R FINALE

4.4. SCHEMA RELAZIONALE

5. PROGETTAZIONE FISICA

5.1. CREAZIONE DELLE TABELLE

5.2. IMPLEMENTAZIONE DEI VINCOLI DI INTEGRITÀ

5.3. ESEMPI DI APPLICAZIONI

6. CONCLUSIONI

7. BIBLIOGRAFIA

- [1] «Linee guida per la prevenzione e il controllo della legionellosi». [Online]. Disponibile su: <https://www.salute.gov.it/portale/malattieInfettive/dettaglioPubblicazioniMalattieInfettive.jsp?id=2362>
- [2] A. Felice, M. Franchi, S. De Martin, N. Vitacolonna, L. Iacumin, e M. Civitini, «Environmental surveillance and spatio-temporal analysis of *Legionella* spp. in a region of northeastern Italy (2002–2017)», *PLOS ONE*, vol. 14, fasc. 7, p. e218687, 2019, doi: [10.1371/journal.pone.0218687](https://doi.org/10.1371/journal.pone.0218687).
- [3] M. Smith, «Polymerase Chain Reaction». Consultato: 31 agosto 2024. [Online]. Disponibile su: <https://www.genome.gov/genetics-glossary/Polymerase-Chain-Reaction>

