



Tesi di Laurea in
Informatica

Progettazione e implementazione
di una base di dati relazionale
per il percorso gravidanza e parto
nell'Ospedale di Udine

Candidato

Andrea Salvador

Relatore

Prof. Angelo Montanari

Correlatori

Prof. Lorenza Driul

Dott. Andrea Brunello

Dott. Nicola Saccomanno

CONTATTI DELL'ISTITUTO

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

Università degli Studi di Udine

Via delle Scienze, 206

33100 Udine — Italia

+39 0432 558400

<https://www.dmif.uniud.it/>

Indice

1	Introduzione	1
1.1	Obiettivo del progetto	1
1.2	Realizzazione del progetto	2
2	Analisi del dominio e dei requisiti	3
2.1	Stato di fatto	3
2.2	Tecnologie e collezioni di dati in uso	4
2.2.1	Tabella per la visita del primo trimestre	4
2.2.2	Tabella per la visita del secondo trimestre	6
2.2.3	Tabella per il parto	7
2.2.4	Rilevazione del cardiocotografo	8
2.2.5	Gynbase	9
2.2.6	Tabella per gravidanze a rischio	10
2.3	Problemi riscontrati con le soluzioni attuali	10
2.4	Requisiti funzionali del nuovo sistema	11
2.5	Scelta di una base di dati di tipo relazionale	12
2.5.1	Modello relazionale	12
2.5.2	Confronto con altri modelli	12
2.5.3	Applicazione al caso di studio	13
3	Progettazione concettuale	15
3.1	Area concettuale della gravidanza	15
3.2	Area concettuale delle visite	18
3.3	Area concettuale del parto	20
3.4	Area concettuale del neonato	20
4	Progettazione logica	25
4.1	Ristrutturazione dello schema Entità-Relazione	25
4.1.1	Ristrutturazione dell'area concettuale della gravidanza	25
4.2	Schema logico relazionale	27
5	Progettazione fisica	29
5.1	Definizione dei domini degli attributi	29
5.2	Definizione delle tabelle	29
5.3	Definizione dei vincoli	29
5.4	Definizione dei trigger	29
6	Funzionalità della base di dati	31
6.1	Esempi di query	31

1

Introduzione

Lo sviluppo di una base di dati centralizzata per la Clinica Ostetrica dell'Ospedale di Udine si inserisce in un progetto di studio più ampio, mirato alla raccolta e all'analisi dei dati relativi a gravidanze e parti, al fine di migliorare l'assistenza fornita dal personale medico e la conseguente esperienza delle pazienti. Nell'ambito medico in generale la raccolta dati investe un ruolo cruciale: attraverso l'analisi e l'interpretazione dei risultati statistici è possibile in particolare:

- migliorare la cura del paziente individuando modelli, tendenze e fattori di rischio, consentendo così ai medici di predisporre strumenti di prevenzione adatti [4, 6];
- promuovere la ricerca scientifica stimolando nuovi studi a partire dai risultati, potenzialmente inediti, estrapolati dalla mole di dati trattata dal personale ospedaliero;
- ottimizzare la gestione ospedaliera grazie alle informazioni su efficienza operativa, uso delle risorse e conseguenti risultati.

Più nello specifico, la Clinica Ostetrica presso l'Ospedale "Santa Maria della Misericordia" di Udine è interessata a sostituire i diversi strumenti software attualmente in uso con un sistema centralizzato, personalizzato secondo i requisiti dei medici e che migliori l'accesso ai dati sia per la sintesi dei dati della singola paziente sia per analisi statistiche di più ampia scala.

1.1 Obiettivo del progetto

Come verrà esposto nella Sezione 2.1, gli usuali sistemi di refertazione adottati a livello aziendale risultano inadatti per la visualizzazione sintetica e le analisi sui dati. I referti sono costituiti di solo testo, oltre a mancare di una struttura logica; ciò ha portato alla scelta di registrare i dati in documenti separati, principalmente in forma di tabelle compilate a mano a seguito delle visite.

Questo sistema parallelo, costruito progressivamente secondo le esigenze specifiche dei suoi utenti, è in uso da circa x anni. I medici hanno evidenziato delle problematiche nel suo utilizzo, relative principalmente alla velocità nell'accesso alle informazioni e all'insorgenza di ambiguità legate alla compilazione esclusivamente manuale e da parte di persone diverse.

Al fine di fornire uno strumento informatico che sia più funzionale, accessibile e coerente proponiamo lo sviluppo e l'implementazione di una base di dati di tipo relazionale che sia modellata sulle esigenze specifiche del contesto in cui verrà utilizzata.

1.2 Realizzazione del progetto

Il processo comprende una prima fase di analisi dei requisiti, volta a individuare le caratteristiche richieste dagli utenti finali e dagli *stakeholder* in generale, seguita da una fase di modellazione a livello concettuale che determina le caratteristiche del dominio di interesse. A partire dal modello concettuale del dominio costruiamo uno schema logico relazionale che permette la definizione di operazioni di base sui dati, come interrogazioni e inserimenti. L'ultima fase consiste nell'implementazione con software specifici che permettono l'interazione attraverso il linguaggio SQL. Mostreremo quindi un insieme di analisi di dati ottenibili grazie a interrogazioni sulla base di dati, per evidenziare in particolar modo le azioni rese possibili dal nuovo sistema che non erano possibili o praticabili con gli strumenti precedenti.

Analisi del dominio e dei requisiti

Il fulcro della progettazione di una base di dati è l'analisi dei requisiti, in quanto essi stabiliscono quali informazioni dovranno essere memorizzate e quali vincoli vengono posti su di esse [5]. Questa fase deve essere condotta in modo accurato ed esaustivo perché eventuali errori commessi al momento della definizione dei requisiti richiedono una quantità molto più elevata di tempo e risorse per essere corretti rispetto a quelli introdotti successivamente, ad esempio nell'implementazione. L'acquisizione dei requisiti deve superare il divario di conoscenza del dominio che si ha tra l'utente finale, in questo caso un membro del reparto di Ginecologia e Ostetricia che si interfacerà con il sistema, e il progettista che lo deve realizzare; ciò si attua con interviste dirette ai committenti e con uno studio accurato dei sistemi già in uso, rilevando le esperienze e le eventuali difficoltà avute.

2.1 Stato di fatto

Il personale medico utilizza sistemi informatici specifici, integrati e uniformati a livello di azienda sanitaria o di sistema sanitario regionale, per compilare e conservare referti di visite ed esami vari; ciò permette il funzionamento dei fascicoli sanitari digitalizzati attualmente in uso. I software in questione sono progettati per rispettare normative legali e standard di qualità e privacy, consentendo di accedere ai singoli referti, ma rendono più difficoltosa l'estrazione delle singole informazioni su pazienti o esami. Per i medici della Clinica che ha richiesto la realizzazione di un nuovo sistema informatico, la necessità è di poter inserire o visualizzare i dati in modo immediato, senza che sia necessario leggere l'intero referto, consentendo inoltre di processare i risultati a fini statistici anche con metodi automatizzati.

I diversi momenti nel corso della gravidanza in cui vengono registrati dati relativi alla paziente (come parametri fisiologici o esami svolti) si sintetizzano nei seguenti.

1. Una visita nel primo trimestre di gravidanza, svolta tra la xesima e la xesima settimana, volta principalmente all'esame chiamato *translucenza nucale* e a un accertamento delle condizioni iniziali della gravidanza.
2. Una visita nel secondo trimestre, detta *morfologica*, che punta ad analizzare lo sviluppo del feto, svolta tra la xesima e la xesima settimana.
3. Una o più visite ecografiche, dette anche *biometriche*, svolte nel corso del terzo trimestre.

4. Il travaglio e il parto, con informazioni sull'intero processo inclusi dati sui neonati.

È importante considerare che il personale medico registra dati soltanto per le visite svolte all'interno della Clinica, e accade frequentemente che una o più visite vengano effettuate in altre strutture o che anche il parto possa avvenire presso altri punti nascita.

Oltre agli esiti di queste visite si vorrebbero conservare informazioni su eventuali patologie delle pazienti, potendo contare sulla presenza di un ambulatorio specifico per le gravidanze *a rischio* che monitora anche le terapie seguite.

2.2 Tecnologie e collezioni di dati in uso

Come presentato nella Sezione 2.1 i medici lamentano la mancanza di un software, personalizzato e parallelo ai sistemi di refertazione “ufficiali” attualmente in uso, che favorisca la lettura dei dati ai fini di consultazione e analisi statistica sull'insieme di pazienti. In assenza di uno strumento centralizzato adatto, nel corso degli ultimi anni (principalmente a partire dal 2023) il personale della Clinica ha registrato i dati di interesse attraverso altri strumenti, affiancati quindi ai referti, variabili sia nel livello di strutturazione dei dati sia nelle modalità di interazione.

Le informazioni raccolte dalle visite vengono registrate principalmente in semplici fogli di calcolo tabellari, accessibili dai diversi calcolatori del reparto e compilati manualmente: elencheremo i diversi campi di queste tabelle con il relativo significato e i possibili valori che possono assumere. A questi si aggiunge un software più complesso e strutturato, ma giudicato non adatto alle esigenze del personale medico.

2.2.1 Tabella per la visita del primo trimestre

Nella visita del primo trimestre vengono annotati aspetti di base ma fondamentali della gravidanza, insieme a primi test genetici e di rischio per determinate condizioni patologiche. In questa visita viene svolto l'esame della translucenza nucale, volto a rilevare malformazioni nel feto già negli stadi iniziali della gravidanza [8]. Questa tabella, così come le successive, contiene un record per ogni gravidanza.

Nella tabella la notazione usata adotta frequentemente abbreviazioni e i campi indicati sono stati predisposti direttamente dal personale medico che poi li deve compilare, adattandosi quindi alle convenzioni a cui gli operatori sono abituati. La compilazione manuale dà ovviamente la possibilità di inserire informazioni sbagliate in tabella, siano esse dati verosimili ma non corrispondenti alla realtà oppure dati corretti ma espressi in modo inconsistente.

- **Data NT:** data della visita. La sigla NT sta per “translucenza nucale” (*nuchal translucency*).
- **Cognome, Nome, DdN:** dati personali identificativi della paziente (cognome, nome e data di nascita).
- **Età all'esame:** età della paziente, espressa come numero intero.
- **Gemellare:** indica se la gravidanza è gemellare, espresso con valori 0/1. In questo attributo e nei seguenti, dove non espresso diversamente, 0 indica un valore falso, assente o negativo e 1 indica un valore vero, presente o positivo.

- **EG:** età gestazionale, indicata come settimane più giorni (ad esempio 12+5).
- **Rischio T21, Rischio T18, Rischio T13:** rischio valutato per le trisomie 21, 18 e 13. I fattori di rischio sono indicati usualmente come classi di rischio, che vanno da alto (A), medio (M oppure I, intermedio), basso (B). In alcuni casi il rischio viene indicato in modo più preciso come rapporto 1:X, con X valore intero, e un livello basso corrisponde a un rapporto 1:1000 o inferiore, un livello medio si trova tra 1:250 e 1:1000, un livello alto è 1:250 o superiore. La scelta di indicare il rischio con una lettera, con l'intera parola o come rapporto dipende dall'operatore che compila la tabella.
- **Morfologia fetale:** indica se sono presenti anomalie nella morfologia fetale, con valori 0/1.
- **Anomalie segnalate:** se il campo precedente ha valore 1, indica le anomalie morfologiche riscontrate come descrizione testuale.
- **NT:** esito dell'esame della translucenza nucale, indica lo spessore della plica nucale in millimetri espresso come numero reale. Il risultato può essere accompagnato da una descrizione testuale.
- **Esito genetica:** risultati di test genetici, eseguiti opzionalmente, successivamente alla visita. I risultati sono molto variabili e indicati come descrizione testuale.
- **Premorfo indicata?:** indica se viene consigliata un'ecografia premorfologica, con valori 0/1 ed eventualmente la motivazione nel caso 1.
- **NIPT:** riporta se è stato effettuato il test prenatale non-invasivo (*Noninvasive Prenatal Testing*).
- **esito NIPT:** esito dell'esame NIPT, riportato come descrizione testuale.
- **BS/BSOB:** indice morfologico misurato sulle dimensioni del tronco encefalico (BS) e della distanza dall'osso occipitale (BSOB), indicato come valore reale.
- **CRL:** indice morfologico, misurato sulla lunghezza del feto in millimetri.
- **PAPP-A, freeBHCG, PLGF:** esami svolti sul sangue della paziente per la rilevazione di proteine e materiale genetico, danno valori reali.
- **UTPI:** indice di pulsatilità dell'arteria uterina, espresso come valore reale.
- **PAM:** pressione arteriosa media materna, misurata come valore intero in mmHg.
- **Esito <34, Esito <37:** esiti di esami, espressi come rapporto di rischio.
- **Prescrizione ASA:** indica se c'è stata prescrizione di cardioaspirina, con valori 0/1.
- **Spontanea, PMA, IUI, FIVET, ICSI, Ovodonazione:** informazioni su concepimento ed eventuale tipo di procreazione medicalmente assistita, tutti con valori 0/1. I campi **Spontanea** e **PMA** sono mutuamente esclusivi; gli altri campi, da **IUI** a **Ovodonazione**, sono compilati solo nel caso di procreazione medicalmente assistita.
- **Peso, Altezza, BMI:** parametri fisici della paziente alla visita.

- **Fumo:** indica se la paziente ha fumato nel periodo precedente alla visita, con valori 0/1.
- **Diabete pregestazionale:** indica se la paziente soffre di diabete pregestazionale (precedente o non dovuto alla gravidanza).
- **Malattie autoimmuni:** indica, se ne soffre, le malattie autoimmuni della paziente, elencate in forma testuale.
- **Pregressa PE:** indica se ha sofferto di preeclampsia (PE) in gravidanze precedenti.
- **PE, IUGR, Eseguita ASA, EG al parto:** attributi predisposti per la compilazione alla conclusione del percorso di gravidanza, dato che si riferiscono a dati registrati alla fine della gravidanza o al momento del parto. In realtà questi campi normalmente non vengono compilati e le informazioni si possono reperire nella tabella che contiene i dati del parto.
- **Note:** annotazioni aggiuntive.

2.2.2 Tabella per la visita del secondo trimestre

La visita del secondo trimestre è incentrata sull'esecuzione di un'ecografia morfologica, ovvero volta a esaminare lo sviluppo del corpo del feto. Si registrano anche eventuali esami svolti nel periodo precedente alla visita.

- **Data Morfo:** data di svolgimento della visita.
- **Cognome, Nome, DdN:** dati personali identificativi della paziente.
- **EG, Età all'esame:** età gestazionale ed età della paziente al momento della visita.
- **Premorfo:** indica se è stata svolta un'ecografia premorfologia e l'eventuale esito in forma testuale.
- **Gemellare:** indica se la gravidanza è gemellare, con valori 0/1.
- **Morfologia fetale:** indica se sono presenti anomalie morfologiche nel feto, con valori 0/1.
- **Anomalie segnalate:** se il valore del campo precedente è 1, elenca le anomalie riscontrate in forma testuale.
- **Decorso:** indica, in forma testuale, se ci sono stati eventi degni di nota nel periodo precedente alla visita.
- **Esito genetica:** risultati di test genetici, non sempre presenti, descritti in forma testuale.
- **Note:** annotazioni aggiuntive
- **NIPT:** indica se è stato effettuato il test NIPT, con valori 0/1.
- **Esito NIPT:** esito dell'esame NIPT. Come valori si trovano 0 (o equivalentemente: B, BR, basso, basso rischio), insufficiente (materiale mancante per dare un esito chiaro), oppure una descrizione testuale del risultato.

2.2.3 Tabella per il parto

Durante il ricovero ospedaliero per il periodo di travaglio e parto si registrano dati relativi sia al processo di espulsione (comprese tempistiche o complicazioni) sia parametri fisiologici della partoriente e dei nati. I valori delle colonne da APGAR a Sesso neonato sono riferite ai neonati: se il parto è gemellare sono presenti due (o eventualmente più) valori, separati da uno spazio, per ciascuna di queste colonne, mantenendo ovviamente lo stesso ordine.

- **Data:** data del parto.
- **PZ:** cognome e nome della paziente.
- **S.G.:** età (stato) gestazionale, espresso sempre come settimane+giorni.
- **Parità:** stringa di cifre che sintetizzano le gravidanze della paziente che hanno preceduto quella corrente. Si considerano: il numero di figli nati a termine, il numero di figli nati pretermine, il numero di aborti (spontanei così come per interruzione volontaria di gravidanza), il numero totale di figli nati vivi; la stringa risultante è costituita di questi valori, nell'ordine riportato (ad esempio 1021, 0000), e quindi è costituita generalmente di 4 caratteri, ma occasionalmente può averne di più se la paziente ha avuto un numero notevole di gravidanze.
- **Travaglio:** indica il metodo di induzione del travaglio. Può avere i seguenti valori: indotto, spontaneo, pilotato, senza travaglio. L'ultimo caso corrisponde a parti cesarei programmati.
- **Motivo induzione:** indica il motivo dell'induzione del travaglio come descrizione testuale.
- **Metodo induzione:** indica il metodo di induzione del travaglio. Solitamente corrisponde a una lista di abbreviazioni che rappresentano farmaci o metodi meccanici di induzione.
- **Parto:** indica il tipo di parto, con i seguenti valori possibili: spontaneo, cesareo, operativo.
- **Motivo parto operativo:** se il tipo di parto è operativo, riporta i motivi dell'attuazione, come descrizione testuale.
- **Se TC:** se il tipo di parto è cesareo, riporta se è stato programmato come tale, con i seguenti valori possibili: programmato, urgente.
- **Motivo TC:** se il tipo di parto è cesareo, riporta i motivi dell'attuazione, come descrizione testuale.
- **Episiotomia:** indica se è stata svolta l'episiotomia, con valori possibili Sì o No.
- **Motivo episiotomia:** se è stata eseguita l'episiotomia, ne riporta il motivo come descrizione testuale.
- **Lacerazioni:** indica il grado di lacerazioni. I valori possibili sono i seguenti: 0 (equivalenti: No, vuoto), 1, 2, 3, 4, AL (altro)
- **Tracheloraffia:** indica la presenza di lacerazioni nel collo dell'utero, con valori possibili Sì o No.
- **Perineo integro:** indica se il perineo è rimasto integro, con valori Sì o No.

- **Secondamento:** indica la modalità di espulsione della placenta nel processo del parto. I valori possibili sono i seguenti: attivo, strumentale, manuale, scivolamento.
- **Perdite:** indica la perdita ematica sofferta durante il parto, quantificata in mL.
- **Robson:** classificazione del parto con valori della scala di Robson.
- **Analgesia:** indica se è stata somministrato un analgesico, con valori Sì o No.
- **Tipo analgesia:** indica il tipo di analgesia somministrata. I valori possibili sono: spinale, epidurale (equivalente: peridurale), spinale e peridurale, calinox.
- **APGAR:** valore di “vitalità” del neonato su una scala di valori interi da 0 a 10. Viene misurato al 1° minuto dalla nascita, al 5° minuto e occasionalmente anche al 10° minuto (ad esempio, 7/7/8).
- **TIN:** indica se il neonato è stato sottoposto a terapia intensiva neonatale, con valori possibili Sì o No.
- **pH, BE:** risultati delle analisi eseguite sul sangue del cordone ombelicale, espressi come valori reali.
- **Sesso neonato:** indica il sesso attribuito al neonato, con valori possibili M o F.

2.2.4 Rilevazione del cardiocotografo

Durante il travaglio viene usato il cardiocotografo, uno strumento che rileva il battito cardiaco del feto (o anche di più feto nel caso di parti gemellari) e l'entità delle contrazioni uterine [2]. Questa apparecchiatura registra e salva i dati, in modo da poterli esportare in forma grafica o tabellare. In particolare consideriamo il file risultato dall'esportazione in forma di foglio di calcolo, in cui vengono indicati alcuni dati identificativi della paziente.

- **Nome, Cognome:** informazioni personali della paziente.
- **Num. progressivo:** identificativo progressivo, indicato come numero intero, attribuito dalla macchina alla registrazione.
- **Data e ora inizio tracciato:** timestamp (in formato leggibile) del momento che la macchina assume come “0 secondi”.

I diversi sensori dello strumento sono sincronizzati e rilevano i rispettivi valori 4 volte al secondo. La tabella contiene quindi una riga per ogni rilevazione, ciascuna delle quali ha i seguenti campi.

- **Secondi, Decimi:** indicazione del momento di rilevazione, rispetto al tempo di inizio tracciato indicato all'inizio della tabella. Entrambi i valori sono interi, con **Decimi** che indica i centesimi [*sic*] di secondo (0, 25, 50, 75).
- **FCF, FCF (2), FCF (3):** indicano la frequenza cardiaca fetale, come valore intero, rilevata da ciascuno dei tre sensori di cui è dotato lo strumento. Anche se la gravidanza è singola, non è necessariamente il sensore 1 a rilevare il battito del feto. Un valore assente, sia per sensori non utilizzati sia per eventuali errori nella rilevazione, è indicato nella tabella con -1. Nel caso di parti gemellari è comunque possibile stabilire a quale neonato corrisponde ciascun tracciato.

- **TOCO:** misura la contrazione uterina, indicata come valore intero.
- **FCM:** misura la frequenza cardiaca materna, indicata sempre come valore intero.

2.2.5 Gynbase

Tra gli strumenti utilizzati dal personale medico del reparto figura il software *Gynbase*, che si occupa di gestire i dati relativi alla gravidanza e al parto. *Gynbase* è un sistema proprietario che opera su una base di dati relazionale accessibile attraverso un'interfaccia grafica dedicata. Tale programma è predisposto per contenere categorie di dati che non sono direttamente di interesse per la Clinica, per cui non sempre viene compilato quando i dati sono già presenti nelle tabelle esposte precedentemente (Sezioni 2.2.1, 2.2.2, 2.2.3). Il contenuto della base di dati, se esportato come tabella in formato CSV, arriva a comprendere circa 150 colonne; analizzeremo quindi gli attributi riportati come di interesse per il personale medico e che non si sovrappongono ad altri già esposti.

- **ID Paziente:** identificativo numerico progressivo attribuito dal sistema alla paziente.
- **ID Gravidanza:** identificativo numerico progressivo attribuito dal sistema alla gravidanza. Questo numero è unico per l'intera base di dati.
- **Cognome, Nome, Nata:** dati personali della paziente.
- **Altezza cm:** altezza della paziente, come valore intero in centimetri.
- **Kg pre gravidanza:** peso della paziente, riportato da lei oppure misurato alla visita del primo trimestre.
- **BMI pre gravidanza:** indice di massa corporea calcolato a partire da peso e altezza all'inizio della gravidanza.
- **Età al concepimento:** indica l'età alla data stimata di concepimento, riportata come numero intero.
- **UM:** riporta la data dell'ultima mestruazione.
- **EPP:** riporta la data prevista del parto.
- **PIH:** indica¹ se la paziente ha sofferto di ipertensione gestazionale (*pregnancy-induced hypertension*).
- **GDM:** indica se la paziente ha sofferto di diabete gestazionale (*gestational diabetes mellitus*).
- **Tireopatia:** indica se la paziente ha sofferto di tiropatia.
- **Data Espulsione:** indica data e ora² del parto.

¹I valori attesi della colonna PIH, così come delle seguenti GDM, Tireopatia e di altre colonne, sarebbero valori booleani del tipo Sì/No o 0/1. In realtà nel file CSV generato dal sistema viene riportato l'identificativo della paziente per significare "vero"/"sì" mentre invece è lasciato vuoto per significare "falso"/"no". Il risultato è che i dati, in questo formato, non sono facilmente leggibili dall'utente.

²Data e ora vengono indicate in formato DD/MM/YYYY HH:MM, oppure con valore "date missing" se assente.

- **Data Fase Attiva:** indica data e ora dell’inizio della fase attiva.
- **Data Membrane:** indica data e ora della rottura delle membrane.
- **Data Completa:** indica data e ora del raggiungimento della dilatazione completa.
- **Data Fase Attiva 2:** indica data e ora dell’inizio della fase di contrazioni.
- **Data Secondamento:** indica data e ora di espulsione della placenta.
- **Complicazioni:** descrizione testuale di eventuali complicazioni avvenute nel processo di travaglio e parto.
- **Indicazioni operativo:** dettagli sullo svolgimento del parto, se operativo.
- **Ind ?:** indicazione del motivo o dei motivi dell’induzione del travaglio. Sono presenti diverse colonne, ciascuna rappresentativa di una motivazione (al posto di ?): **ipertensione preesistente**, **ipertensione gestazionale**, **proteinuria isolata**, **iugr**, **gdm**, **preeclampsia**, **macrosomia**, **protratta**, **pma**, **oligoamnios**, **poliidramnios**, **altro**, **pprom**, **prom**, **colestasi**, **mef**, **prodromi prolungati**. Per indicare che l’induzione è stata svolta per uno o più determinati motivi la colonna riporta la lettera Y, altrimenti è lasciata vuota.
- **Minuti induzione-inizio travaglio:** tempo trascorso tra l’induzione del travaglio e l’inizio del travaglio, indicato come valore intero.
- **Minuti induzione-parto:** tempo trascorso tra l’induzione del travaglio e l’espulsione, indicato come valore intero.
- **Bishop:** indicatore relativo alla preparazione del collo dell’utero all’induzione del travaglio, riportato come valore intero su una scala da 0 a 16.

2.2.6 Tabella per gravidanze a rischio

Un insieme più ristretto di pazienti viene seguito nel corso della gravidanza da un ambulatorio apposito, detto appunto “gravidanze a rischio”, in presenza di determinate patologie.

- **Nome, Cognome, DdN:** dati identificativi della paziente.
- **Patologie:** diverse colonne che riportano patologie, una per colonna, sofferte dalla paziente e che sono trattate o sorvegliate da questo ambulatorio.
- **Terapie seguite:** descrizione testuale delle terapie seguite. Può non essere indicata una terapia per tutte le patologie.

2.3 Problemi riscontrati con le soluzioni attuali

L’insieme di tabelle e fogli di calcolo utilizzato attualmente dai medici del reparto è essenzialmente creato su misura, ovvero sono pensati per contenere le informazioni di loro interesse perché essi stessi

ne hanno definito lo schema. Nonostante ciò, questi strumenti sono molto più semplici di una base di dati ben strutturata e questo ne evidenzia gli aspetti negativi.

L'uso di tabelle, rispetto al contenuto testuale dei referti, permette di organizzare almeno visivamente l'insieme di informazioni che si vogliono conservare per ciascuna paziente, ma non vengono imposti vincoli di correttezza o di coerenza dei dati effettivamente inseriti. Le operazioni di accesso ai dati (a partire dai dati identificativi della paziente) sono veloci se si considera una singola visita ma se si vuole analizzare l'intero periodo della gravidanza è necessario leggere contemporaneamente più tabelle; sotto questo aspetto l'interfaccia grafica fornita dai diversi software per visualizzare fogli di calcolo non è ottimale per tali scopi.

2.4 Requisiti funzionali del nuovo sistema

La richiesta avanzata dal personale medico è di un software integrato e centralizzato che permetta la visualizzazione, in prima analisi, e la manipolazione, in una fase successiva, dei dati relativi a gravidanze e parti, come esposto precedentemente (Sezione 2.1).

Il sistema dovrà permettere di visualizzare, aggiungere, modificare ed eliminare le informazioni che attualmente sono contenute nei diversi supporti esposti nella Sezione 2.2. In sintesi, i dati riguardano dettagli personali della paziente e di sue eventuali patologie; dettagli della gravidanza, del suo decorso e delle visite svolte; informazioni su travaglio, parto e parametri fisiologici dei neonati; esiti dei vari esami eseguiti nel corso della gravidanza.

Analizzando le tabelle attualmente in uso possiamo dedurre le informazioni di interesse che dovranno sicuramente essere rese accessibili dal sistema che realizziamo. Questo è anche il momento più indicato per valutare l'aggiunta di requisiti che prima potevano non essere stati considerati o non venivano realizzati per la mancanza di supporti adeguati. In particolare, oltre a quanto esposto nella Sezione 2.2, i medici vogliono poter memorizzare informazioni sui seguenti aspetti non trattati prima.

1. Dati relativi ai neonati, in particolare peso, altezza e circonferenza cranica, misurati alla nascita.
2. Visite ecografiche biometriche. Possono essere svolte più visite biometriche nel corso della gravidanza (tendenzialmente 1 o 2) e si raccolgono dati relativi ai seguenti indicatori:
 - BPD (diametro biparietale);
 - CC (circonferenza cranica);
 - AC (circonferenza addominale);
 - FL (lunghezza del femore);
 - EFW (peso fetale stimato);
 - PIAO ?;
 - PIACM ?;
 - stato della crescita del feto (crescita regolare, FGR oppure SGA).
3. Esiti di esami vari, eseguiti anche al di fuori delle principali visite elencate nella Sezione 2.1, di una lista eventualmente espandibile su necessità del personale medico. Gli esami danno esiti in forma

perlopiù di indici numerici, comprendenti ma non limitati a: HB, TSH, AST/ALT, glicemia, curva da carico di glucosio, sierologie varie, gruppo sanguigno, test di Coombs indiretto.

4. Dettagli su problemi concomitanti, patologie e terapie seguite (anche al di fuori di quelle controllate nell'ambulatorio “gravidezze a rischio”) e altri aspetti come interventi, trasfusioni o allergie, da indicare in forma testuale o debolmente strutturata.

2.5 Scelta di una base di dati di tipo relazionale

2.5.1 Modello relazionale

Nel modello relazionale [3] i dati vengono strutturati attraverso relazioni (o equivalentemente, tabelle), ciascuna composta di attributi (colonne) e istanziata come insieme di tuple (righe o record). Uno schema relazionale corretto e ben strutturato mira a prevenire inconsistenze e dati duplicati. Gli strumenti software che operano su basi di dati relazionali (RDBMS) permettono di definire vincoli di integrità che contribuiscono al mantenimento della coerenza dei dati, all'assenza di ambiguità e all'efficienza delle interazioni con il sistema attraverso controlli automatici. La modellazione del dominio in cui operiamo viene fatta prima a livello concettuale definendo uno schema nel modello “Entità-Relazione” e successivamente traducendolo in uno schema logico nel modello relazionale.

Le basi di dati relazionali si possono interrogare con semplici richieste (*query*) in SQL, un linguaggio dichiarativo basato sulla logica del prim'ordine. Questo linguaggio è supportato da tutti i principali linguaggi di programmazione e ambienti di sviluppo, prestandosi così a integrarsi facilmente in applicazioni complesse e che permettono, ad esempio, l'interazione attraverso un'interfaccia grafica dedicata. In questo modo solo il programmatore ha la necessità di conoscere i dettagli implementativi dello schema mentre l'utente finale, anche se inesperto, può operare sui dati a livello più astratto e vicino al modello concettuale del dominio.

2.5.2 Confronto con altri modelli

Altri modelli di basi di dati possono avere altri vantaggi specifici, ma spesso perdono su alcuni dei punti forti del modello relazionale.

- Il modello gerarchico prevede una struttura ad albero, composto da nodi che possono avere un numero arbitrario di figli ma solo un genitore ciascuno. Questo pone molti limiti sullo schema da realizzare e spesso richiede di duplicare dati.
- Il modello reticolare estende quello gerarchico permettendo un numero arbitrario di genitori, formando quindi un grafo generico. La struttura così ottenuta, eccessivamente libera nella sua topologia, rende più difficoltosa l'interrogazione.
- Il modello a documento, che sta prendendo piede come principale modello di tipo *NoSQL*, prevede che i record siano memorizzati come documenti strutturati ma senza uno schema comune. Questo facilita la memorizzazione di dati con caratteristiche eterogenee ma allo stesso tempo rende più difficoltose le operazioni che legano documenti diversi, come le *join*.

2.5.3 Applicazione al caso di studio

Considerando i requisiti esposti nella Sezione 2.4, proponiamo l'utilizzo di una base di dati basata su un modello relazionale. Le informazioni trattate dal personale medico seguono schemi ben strutturati, come dimostra l'organizzazione tabellare utilizzata finora, e a tali schemi possiamo ricondurre le entità e le relazioni a livello astratto. Riconosciamo infatti entità come la paziente, la gravidanza, la visita, il parto, il neonato, ciascuna con un insieme di attributi. Tra di esse sussistono relazioni che seguono schemi costanti, come la relazione tra gravidanze e visite, tra gravidanze e malattie o tra parti e neonati.

Una base di dati relazionale permette di gestire in modo naturale diversi vincoli di integrità, soprattutto per quanto riguarda gli attributi identificativi (detti “chiavi”) sia il numero di entità coinvolte in determinate relazioni (detto *cardinalità*). Nel caso reale che stiamo trattando è possibile porre vincoli sul numero di visite di ciascun tipo effettuate nel corso della gravidanza, così come sul numero di bambini nati da un parto o di esami svolti in ciascuna visita.

Si vuole che il sistema risultante costituisca, in prima fase, un magazzino dati (*data warehouse*³) volto principalmente alla visualizzazione contemporanea e coerente dei dati proveniente dalle diverse fonti attualmente in uso e che possa sostituirle completamente in un secondo momento.

Nel Capitolo 3 si mostra lo schema Entità-Relazione costruito a partire dai requisiti e vengono illustrate le scelte operate a livello concettuale, con tutti i dettagli relativi alle entità presenti, alle relazioni tra esse e agli attributi.

³L'accezione di *data warehouse* [5], concetto distinto da *database*, evidenzia come il sistema integri dati provenienti da più fonti in un unico schema e sia orientato alla visualizzazione più che alle operazioni di modifica.

3

Progettazione concettuale

La figura 3.1 mostra lo schema della base di dati nel modello Entità Relazione in una versione semplificata per evidenziare le entità e le relazioni presenti. I quattro contorni tratteggiati identificano quattro aree del diagramma (gravidanza, visite, parto, neonato) che verranno illustrate dettagliatamente con gli attributi presenti. L'analisi dei requisiti e dei supporti esistenti (Sezione 2.2) evidenzia come elemento centrale le gravidanze: solo poche informazioni sono legate alla paziente piuttosto che alla gravidanza e sono essenzialmente quelle che la identificano personalmente.

In generale la presenza di attributi opzionali è dovuta non tanto alla mancanza del dato, ovvero che non sia presente a livello concettuale, bensì alla mancata conoscenza di un suo valore, da attribuirsi a un'errata compilazione dei record o a un dettaglio anamnestico non riportato dalla paziente. Nell'uso del valore *NULL* si ritrova questa stessa ambiguità [7]. Si intende lasciare la possibilità di avere dati mancanti per permettere un'integrazione che sia compatibile con i sistemi attualmente in uso, in modo da implementare il sistema come *data warehouse*, oltre a renderlo tollerante rispetto a eventuali omissioni di informazioni.

3.1 Area concettuale della gravidanza

La Figura 3.2 mostra l'ingrandimento dello schema Entità-Relazione per quanto riguarda le entità Paziente, Gravidanza e Malattia.

Paziente

Si introduce un identificativo per la paziente che risulta essere effettivamente univoco, ovvero il codice fiscale, assumendo che tutte le pazienti trattate abbiano un codice fiscale assegnato. Ciò semplifica la rappresentazione dell'identificativo personale, che nello schema logico verrà inserito in quasi tutte le tabelle, rispetto alla tripla composta da nome, cognome e data di nascita, e previene i possibili, seppur rarissimi, casi di corrispondenza di questi dati per persone diverse.

Gravidanza

A questa entità fanno riferimento alcuni attributi che vengono registrati durante la visita del primo trimestre, oltre all'esito della gravidanza e alla data del primo ingresso. Decidere se assegnare tali

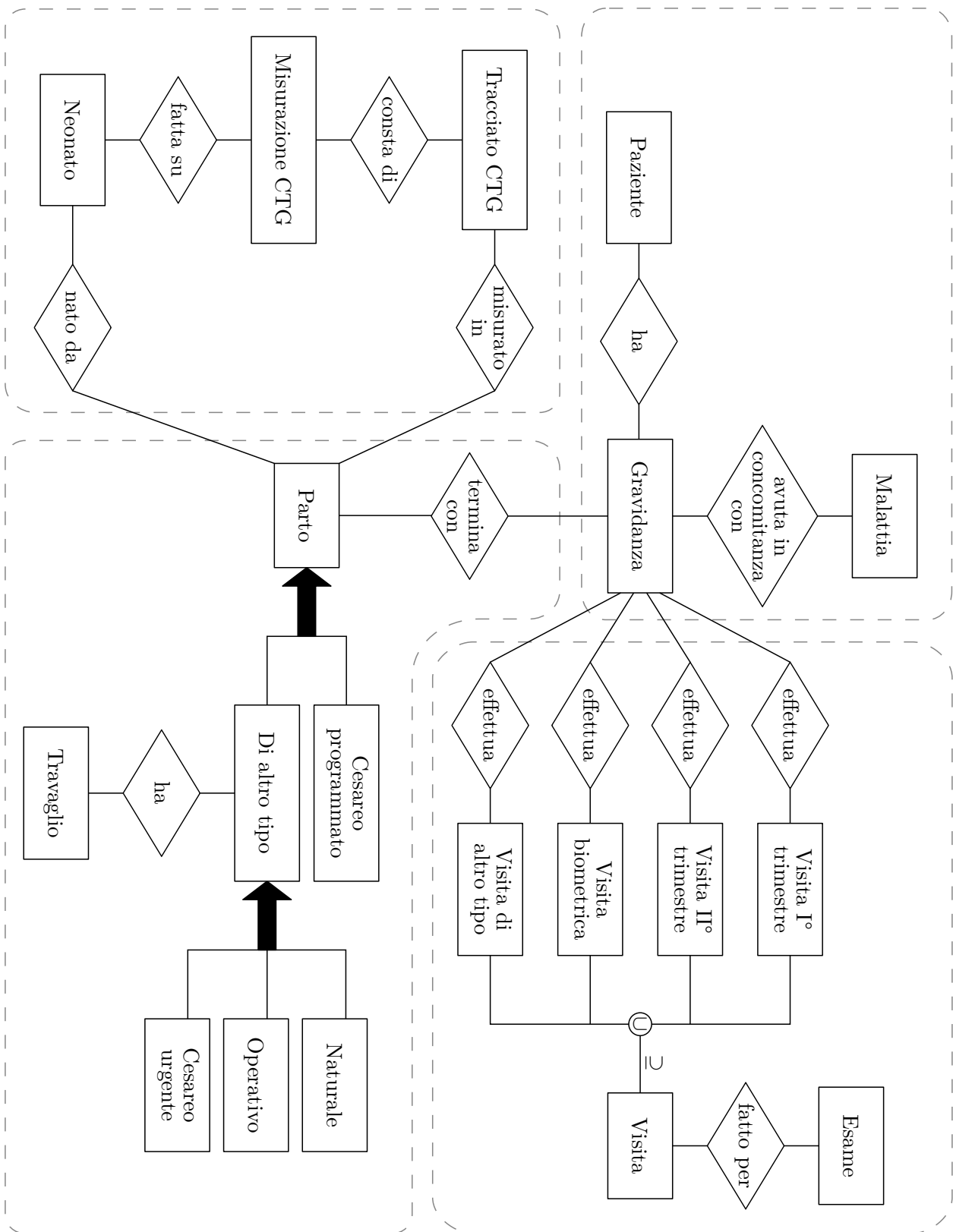


Figura 3.1: Schema Entità-Relazione. Visione d'insieme semplificata, contenente soltanto entità e relazioni del modello. Le aree tratteggiate delimitano i diversi ingrandimenti nei quali vengono specificati gli attributi delle entità e relazioni coinvolte.

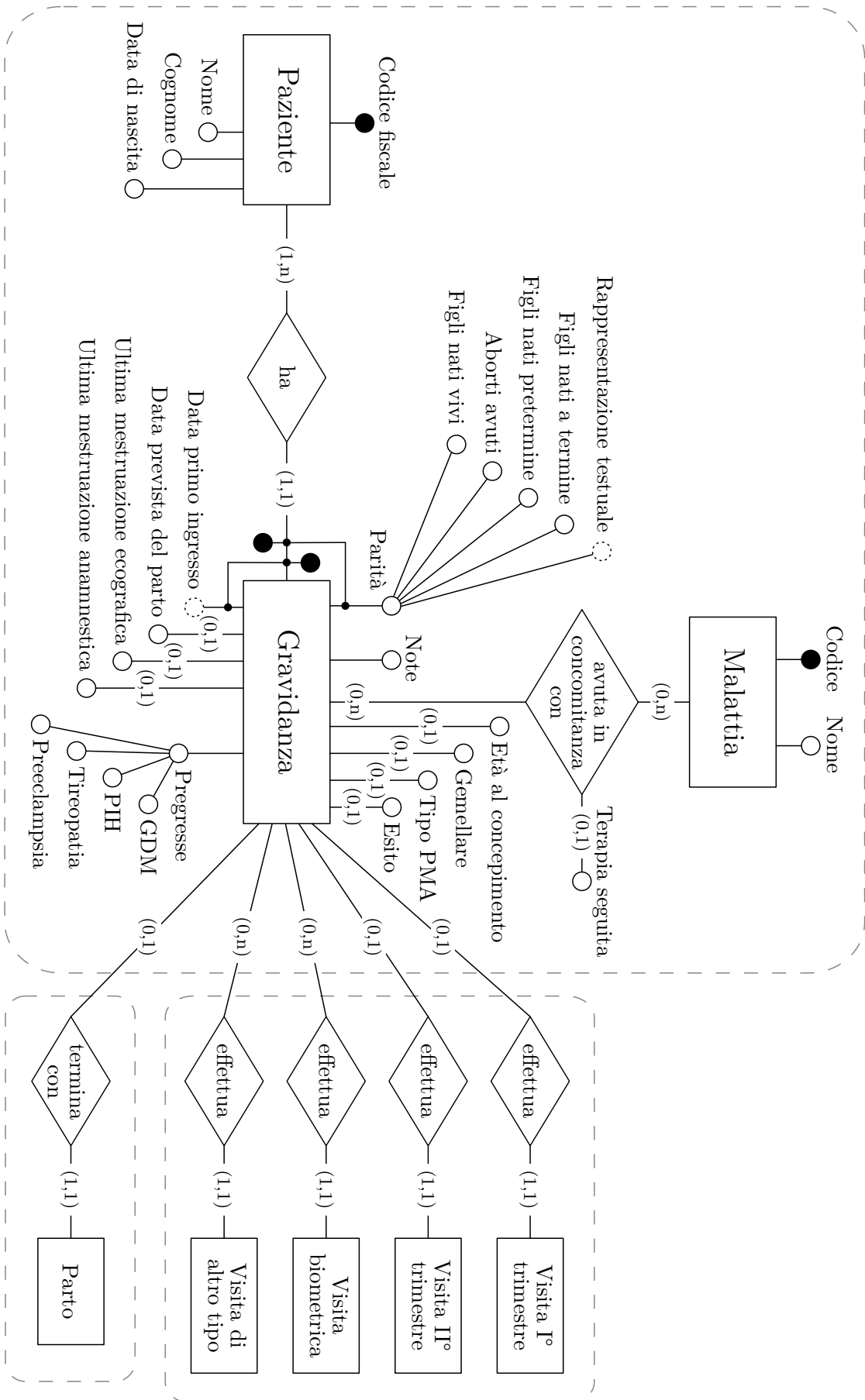


Figura 3.2: Schema Entità Relazione. Ingrandimento dell'area gravidanza.

attributi a “Gravidanza” oppure a “Visita I° trimestre” può essere talvolta arbitrario: preferiamo quindi assegnare alla prima quelle informazioni che si possono considerare come proprie della gravidanza in sé (come la presenza di feti gemellari o il tipo di PMA), mentre alla seconda le informazioni sulla paziente che si riferiscono al giorno o al periodo della visita (come ad esempio peso o altezza).

Identifichiamo due chiavi candidate, entrambe parziali in quanto l’entità è debole verso Paziente:

- la data del primo ingresso, attributo derivato e corrispondente alla data della prima visita effettuata, se ce n’è stata una, altrimenti alla data del parto;
- la parità registrata all’inizio della gravidanza.

L’attributo “Parità” è composto. Nei sistemi attualmente in uso è scritto solitamente in un unico campo, usando quindi la rappresentazione testuale che è calcolata come concatenazione del numero di figli nati a termine, del numero di figli nati pretermine, del numero di aborti (comprendente sia aborti spontanei sia interruzioni volontarie di gravidanza) e del numero totale di figli nati vivi, in questo ordine.

Sempre sull’attributo “Parità” si pone il seguente vincolo.

Se una paziente ha più di una gravidanza, per ogni coppia di gravidanze successive i valori degli attributi (non derivati) sono non decrescenti e almeno uno di essi è strettamente crescente.

Le relazioni che l’entità Gravidanza intrattiene con le diverse visite e con Parto hanno tutte partecipazione parziale, per cui viene imposto il seguente vincolo.

Una gravidanza viene registrata solo se viene effettuata almeno una visita oppure se termina con un parto.

Malattia

L’entità “Malattia” rappresenta le patologie che possono presentarsi in concomitanza con la gravidanza. Attualmente vengono registrate in modo sistematico solo per le pazienti seguite nell’ambulatorio “gravidanze a rischio” (vedi Sezione 2.2.6) ma è applicabile alla gravidanza di qualsiasi paziente. È importante considerare che le malattie possono variare tra le diverse gravidanze della stessa paziente. Le informazioni riguardo alle eventuali terapie seguite sono memorizzate nell’attributo della relazione che lega gravidanza e malattie.

3.2 Area concettuale delle visite

La Figura 3.3 mostra l’ingrandimento dello schema Entità-Relazione per quanto riguarda le entità relative alle visite e all’entità Esame.

Visite

Dividiamo le possibili visite in quattro entità: Visita I° trimestre, Visita II° trimestre, Visita biometrica, Visita di altro tipo. Gli attributi di queste entità fanno riferimento principalmente alla data e alle informazioni rilevate dalla paziente come età, epoca gestazionale o anomalie fetali. Gli esiti degli esami

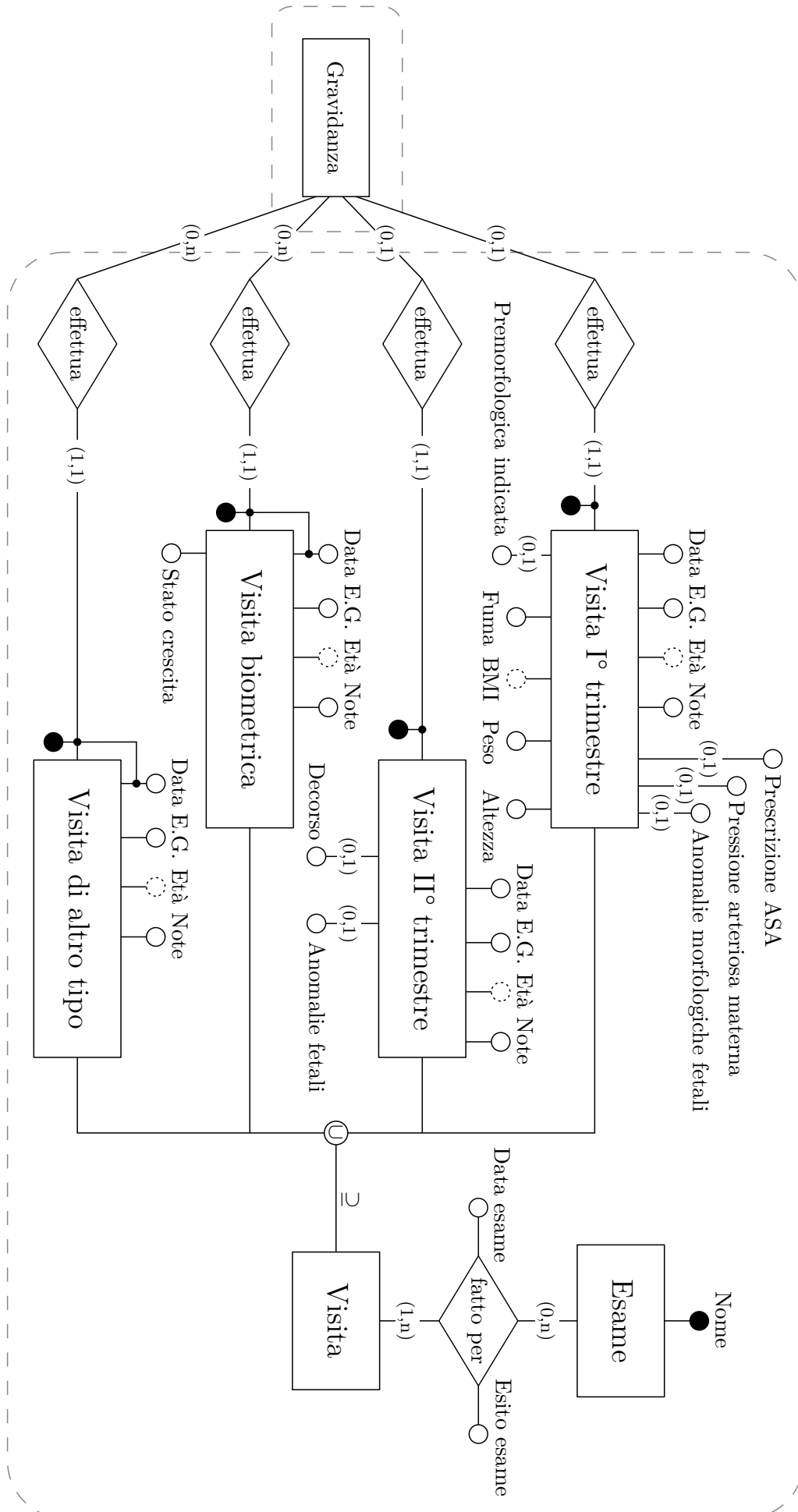


Figura 3.3: Schema Entità Relazione. Ingrandimento dell'area visite.

svolti per una determinata visita non sono modellati come attributi delle visite ma come attributi della relazione che lega Visita e Esame, in modo che gli esami da registrare possano essere modificati nel tempo senza agire sullo schema della base di dati.

I diversi attributi “Età”, presenti in tutte le entità, sono calcolati a partire dalla data di nascita della paziente. L’entità “Visita di altro tipo” raccoglie le istanze di visite che altrimenti non rientrerebbero in nessuna delle tre categorie specifiche ma per le quali si vogliono comunque registrare esiti di esami svolti.

La scelta di organizzare l’entità “Visita” come unione (o categoria) piuttosto che come una generalizzazione è dovuta alla cardinalità delle relazioni che i quattro tipi di visita intrattengono con l’entità Gravidanza. Tutte le visite sono entità deboli rispetto a Gravidanza, ma le visite biometriche o di altro tipo possono essere svolte più volte nel corso di una gravidanza e quindi richiedono un attributo che sia chiave parziale, che invece sarebbe scorretto mettere per le visite del primo e del secondo trimestre le quali sono effettuate una sola volta ciascuna. Se fossero state modellate come specializzazioni di un’unica entità Visita si sarebbero creati numerosi cicli e vincoli aggiuntivi da imporre.

Esame

L’entità Esame rappresenta l’esame in generale, mentre l’esito di uno specifico esame svolto si ritrova nella relazione tra Esame e Visita. Il tipo di dato dell’esito dipende dall’esame, ma l’aspetto implementativo viene trattato a livello di schema logico.

3.3 Area concettuale del parto

La Figura 3.4 mostra l’ingrandimento dello schema Entità-Relazione per quanto riguarda le entità Parto, con le sue specializzazioni, e Travaglio.

Parto

Le diverse specializzazioni dell’entità Parto sono totali e disgiunte. Solo alcune informazioni, come data o informazioni su analgesia e perdite ematiche, sono applicabili ad ogni tipo di parto: gli altri attributi si trovano nelle entità che specializzano Parto, in alcuni casi con sovrapposizioni. Ad esempio, i dati sul travaglio e sulle tempistiche del parto sono applicabili a tutti i tipi di parto eccetto i parti cesarei programmati, mentre le lacerazioni e il secondamento sono dati presenti nei parti naturali e operativi. La data del parto è coerente con le tempistiche indicate se il parto non è cesareo urgente.

Travaglio

Le informazioni sul travaglio sono applicabili a tutti e soli i parti che non sono cesarei programmati. Se il travaglio è indotto sono presenti anche i dati relativi al metodo e alle tempistiche dell’induzione.

3.4 Area concettuale del neonato

La Figura 3.5 mostra l’ingrandimento dello schema Entità-Relazione per quanto riguarda le entità Neonato e quelle relative alle misurazioni del cardiocotografo.

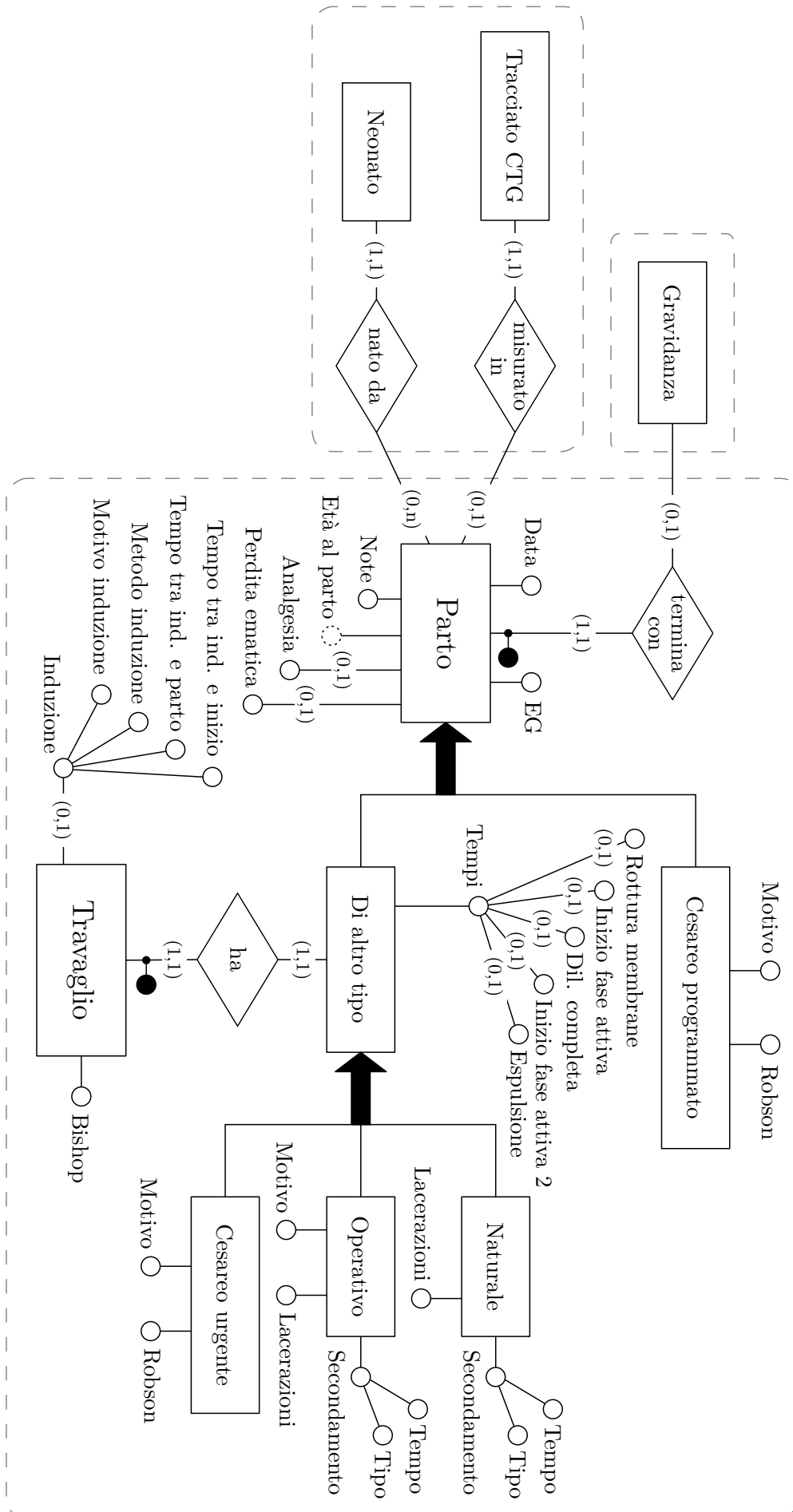


Figura 3.4: Schema Entità Relazione. Ingrandimento dell'area parto.

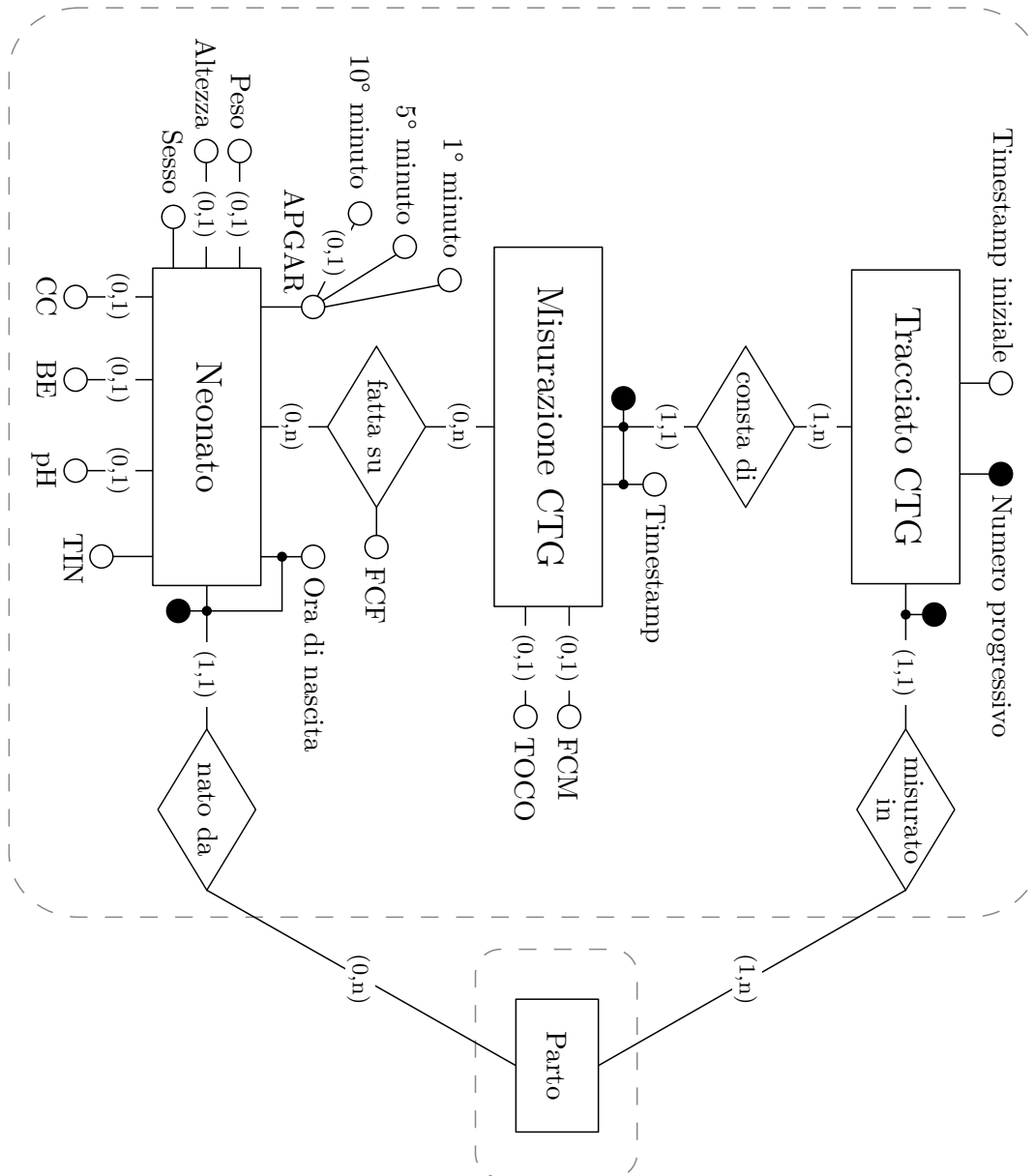


Figura 3.5: Schema Entità Relazione. Ingrandimento dell'area neonato.

Neonato

Nell'entità Neonato la chiave parziale è l'ora di nascita. Questo, ugualmente ad altri indicatori dell'ordine di nascita, è necessario per distinguere i gemelli nati in una stessa gravidanza.

Le informazioni relative al neonato sono le seguenti:

- peso, altezza e sesso attribuito;
- indice di Apgar, misurato in diversi momenti dopo la nascita;
- BE: *base excess*, eccesso di basi;
- pH: acidità del sangue;
- CC: circonferenza cranica;
- TIN: terapia intensiva neonatale.

Tracciato e misurazioni

Nello schema la sigla CTG sta per “cardiotocografo”. Un tracciato è identificabile sia con il parto in cui viene registrato sia con il numero progressivo, che viene assegnato unico dal sistema in uso nel reparto.

Ogni misurazione del tracciato riporta i valori FCM (frequenza cardiaca materna) e TOCO (toco-gramma). I valori FCF (frequenza cardiaca fetale) possono essere multipli in ogni misurazione, quindi non sono contenute in un attributo dell'entità Misurazione TCG bensì nella relazione tra essa e Neonato. Si impone quindi il seguente vincolo.

Per ogni misurazione deve essere presente almeno uno tra: il valore FCM, il valore TOCO e un valore FCF per un neonato associato.

4

Progettazione logica

Lo schema concettuale presentato nel Capitolo 3 rappresenta in modo accurato il dominio a livello astratto ma la traduzione in uno schema logico nel modello relazionale richiede la semplificazione di alcune strutture concettuali. In seguito a questo passaggio di ristrutturazione si definisce uno schema logico, costituito da un insieme di relazioni con eventuali vincoli posti su di esse.

4.1 Ristrutturazione dello schema Entità-Relazione

Il formalismo del modello Entità-Relazione permette di utilizzare strutture piuttosto complesse a livello concettuale, come ad esempio le relazioni di specializzazione o gli attributi multivalore. Per facilitare la traduzione in uno schema logico nel modello relazionale è necessario operare una ristrutturazione o semplificazione dello schema iniziale [1], con l'obiettivo di ottenere soltanto entità senza specializzazioni o categorie, attributi semplici e relazioni binarie. Nella fase di ristrutturazione si valuta se è opportuno mantenere gli attributi ridondanti per motivi di efficienza e si scelgono gli attributi chiave delle entità che hanno più candidati possibili.

Lo schema ristrutturato, mostrato in Figura ??, rimane espresso nel modello Entità-Relazione, ma non si può più considerare puramente concettuale perché introduce modifiche dettate non dalle caratteristiche del dominio ma da aspetti implementativi e ottimizzazione delle prestazioni. Dopo la ristrutturazione, la traduzione può essere fatta in modo meccanico perché i costrutti più elementari del modello Entità-Relazione hanno dei corrispettivi diretti nel modello relazionale.

4.1.1 Ristrutturazione dell'area concettuale della gravidanza

La Figura 4.1 mostra l'ingrandimento dello schema Entità-Relazione ristrutturato per quanto riguarda le entità Paziente, Gravidanza e Malattia.

Gravidanza

Nell'entità Gravidanza semplifichiamo gli attributi composti in attributi semplici, legandoli quindi direttamente alla gravidanza. Si rimuove la “rappresentazione testuale”, presente nello schema concettuale come attributo derivato e direttamente calcolabile a partire dagli altri componenti dell'attributo compo-

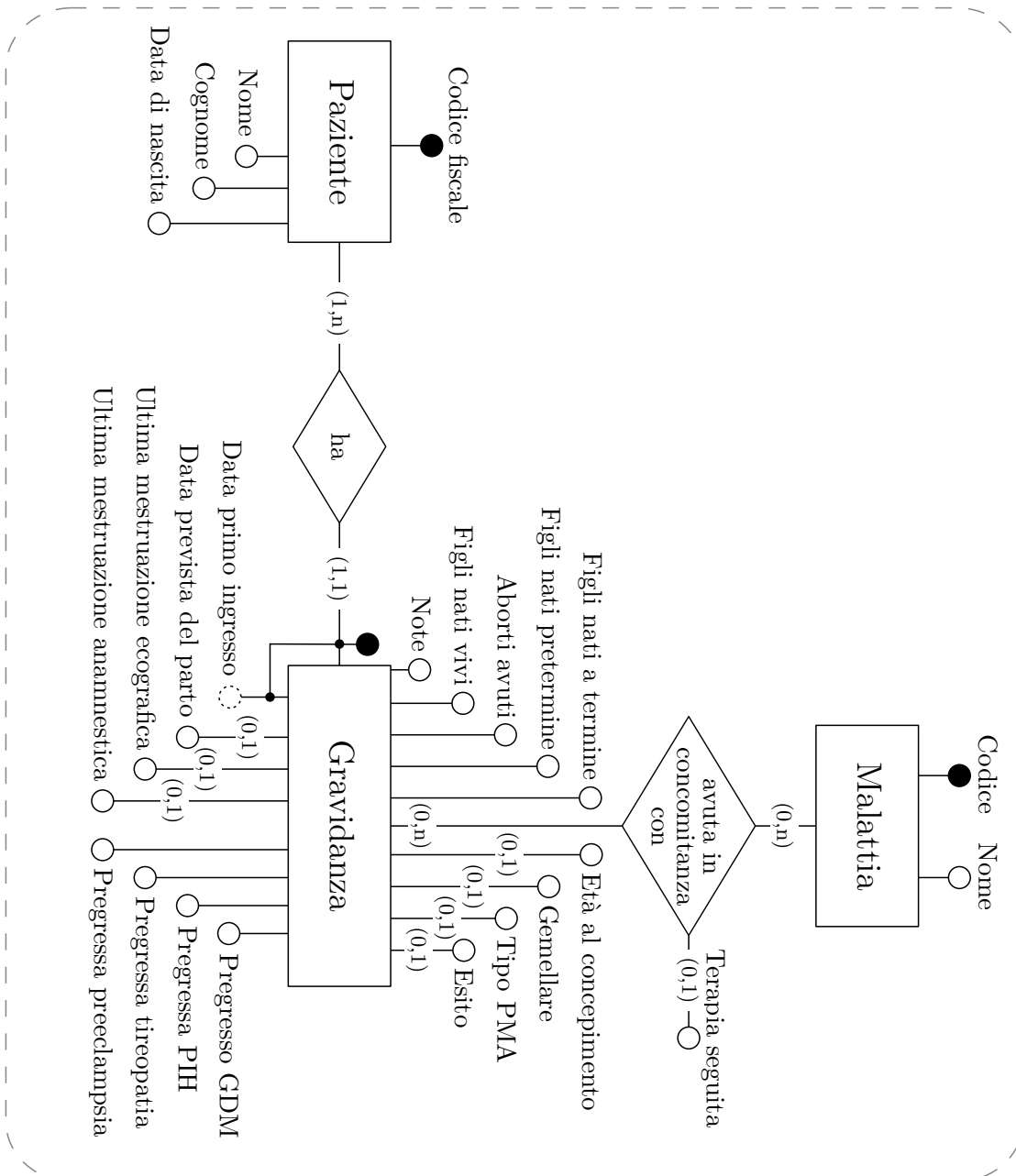


Figura 4.1: Schema Entità-Relazione ristrutturato. Ingrandimento dell'area gravidanza.

sto Parità. Viene mantenuto il vincolo posto sulla non decrescenza o crescita stretta dei componenti dell'attributo Parità.

Gravidanza è entità debole ed è identificata dalla relazione con Paziente. Tra le due chiavi parziali possibili si sceglie l'attributo Data primo ingresso perché è costituito da un solo attributo invece che da quattro diversi, risultando quindi molto più semplice da utilizzare come identificatore della gravidanza nelle diverse relazioni in cui questa entità partecipa. La quadrupla di attributi risultante dalla decomposizione di Parità, in quanto chiave candidata nel modello concettuale, mantiene le proprietà di chiave e quindi si pone un ulteriore vincolo di unicità¹.

La quadrupla di attributi che compongono la parità deve essere unica per ogni paziente.

4.2 Schema logico relazionale

¹Il vincolo di non avere valore *NULL* è già espresso dall'assenza dell'indicazione (0,1) su tutti gli attributi, quindi nessuno di essi è opzionale.

5

Progettazione fisica

5.1 Definizione dei domini degli attributi

5.2 Definizione delle tabelle

5.3 Definizione dei vincoli

5.4 Definizione dei trigger

6

Funzionalità della base di dati

6.1 Esempi di query

Bibliografia

- [1] P. Atzeni e altri. *Basi di dati*. McGraw-Hill, 2023.
- [2] D. Ayres-de-Campos. Electronic fetal monitoring or cardiotocography, 50 years later: what's in a name? *American Journal of Obstetrics & Gynecology*, 218(6):545–546, 2018.
- [3] E.F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.
- [4] L. Corey, A. Vezina, e R.B. Gala. Using technology to improve women's health care. *Ochsner Journal*, 20(4):422–425, 2020.
- [5] N. Dedić e C. Stanier. An evaluation of the challenges of multilingualism in data warehouse development In *Proceedings of the 18th International Conference on Enterprise Information Systems (ICEIS 2016)*. A cura di S. Hammoudi e altri, volume 1, pp. 196–206, 2016.
- [6] J.S. McCullough, S.T. Parente, e R. Town. Health information technology and patient outcomes: the role of information and labor coordination. *The RAND Journal of Economics*, 47(1):207–236, 2016.
- [7] A. Silberschatz, H.F. Korth, e S. Sudarshan. *Database System Concepts*. McGraw-Hill, 2011.
- [8] A.P. Souka e altri. Increased nuchal translucency with normal karyotype. *American Journal of Obstetrics & Gynecology*, 192(4):1005–1021, 2005.