•

Politecnico di Torino Dipartimento di Ingegneria Meccanica e Aereospazioale (DIMEAS)

Laurea Magistrale in Ingegneria Biomedica

Tesi di Laurea Magistrale

Progettazione del data warehouse clinico oncologico per l'ospedale Humanitas - Gradenigo



Autore: Donato Caso (Matricola: 244416) Relatore: prof. Gabriella Balestra Referente aziendale: Ing. Susanna Pavanelli Referente aziendale: Ing. Tralli Augusta

A.A. 2019/2020

 $Ai\ miei\ genitori,\ per\ la\ loro\ perseveranza.$ $A\ mio\ fratello,\ fedele\ amico.$ $A\ mia\ nonna,\ per\ il\ suo\ amore.$

Sommario

eccellenza.

La ricerca clinica sta vivendo un periodo di forte innovazione e integrazione sui dati clinici informatizzati, i quali possono essere usati come supporto alle decisioni per medici, ricercatori o altri professionisti del settore sanitario. I dati clinici oggi, risiedono in più sorgenti differenti, con formati diversi, per questo motivo si richiede uno strumento informativo chiamato data warehouse clinico(DW).

Uno dei ruoli dei sistemi informativi dell'ospedale Humanitas – Gradenigo, è quello di fornire tutti gli strumenti necessari per consentire l'innovazione sui sistemi da utilizzare, ma anche per garantire il giusto supporto informatizzato alla ricerca clinica, per la prevenzione di eventuali patologie, e alla possibilità di selezione tra i diversi percorsi terapeutici. Durante il periodo di svolgimento tesi presso l'ospedale, ho percepito l'importanza dei dati clinici e fornito il mio supporto per la progettazione di un data warehouse clinico orientato sulla branca oncologica. Nel 2019, in Piemonte, la sopravvivenza netta a 5 anni dalla diagnosi su tutti i tumori è al 53%, per questo motivo uno strumento valido e che include tutti i dati psico – fisici

e percorsi terapeutici di ciascun paziente rappresenta l'innovazione tecnologica per

L'obiettivo della tesi verte nella progettazione del data warehouse oncologico, affronta tutte le fasi: dall'analisi e ricognizione delle fonti dati utilizzate dall'ospedale, fino alla fase di progettazione logica in cui si costruisce lo schema snowflake. L'analisi dei requisiti è necessaria per la comprensione del contesto, degli attori coinvolti, ma anche per lo studio dei processi ospedalieri, prima fonte necessaria per popolare il data warehouse. La progettazione concettuale consente, invece, di individuare il fatto alla base del modello multidimensionale del DW, ovvero il paziente, e di stabilire le connessioni con le diverse dimensioni ad esso legate. In questa fase si raggiunge il dettaglio dei dati archiviati, grazie alla visualizzazione di tutti gli attributi per ciascuna dimensione.

Tra gli obiettivi futuri ci sarà la necessaria progettazione dell'alimentazione per il DW e la valutazione degli strumenti da utilizzare. Passo fondamentale, sebbene rappresenti il 20% della progettazione del DW. Le fasi come l'analisi dei requisiti e la progettazione concettuale arricchiscono la documentazione dello strumento, in previsione di estendere l'utilizzo anche per la ricerca clinica su altre patologie.

Indice

1	Inti	oduzio	one	10
	1.1	Il fran	nework i2b2	12
	1.2	Il mod	dello i2b2	12
2	Ma	teriali	e metodi	15
	2.1	Le Ba	si di Dati	15
	2.2	Il mod	dello relazionale	16
		2.2.1	Vincoli di integrità	18
		2.2.2	Modello Entità - Relazione	18
		2.2.3	Costrutto entità e attributi dell'entità	19
		2.2.4	Associazioni	19
		2.2.5	Reverse Engineering	20
	2.3	Data	warehouse	21
		2.3.1	Definizione	21
		2.3.2	OLTP E OLAP	21
		2.3.3	Modello multidimensionale	22
		2.3.4	L'architettura per il Data Warehouse	22
		2.3.5	Gli strumenti ETL	24
		2.3.6	Data Mart	24
	2.4	Fasi d	i Progettazione di un Data Warehouse	24
		2.4.1	La progettazione del data warehouse	25
		2.4.2	Analisi e riconciliazione delle sorgenti	26
		2.4.3	Analisi dei requisiti	26
		2.4.4	Metodologia Tropos	26
		2.4.5	Progettazione concettuale	27
		2.4.6	Progettazione logica	27
3	Il P	ercors	o CAS	28

INDICE

	3.1	I dati clinici nel percorso CAS	28
	3.2	Synopsis, Workflow e Swim-Lane Activity	29
		3.2.1 Synopsis	29
		3.2.2 Workflow	30
		3.2.3 SWim-Lane Activity	31
4	DW	Ospedale Humanitas – Gradenigo	33
	4.1	l dominio applicativo dell'ospedale	33
	4.2	Pianificazione del DW	34
	4.3	Analisi delle fonti dati	35
	4.4	Ricognizione	35
		4.4.1 SGP	38
		4.4.2 Schema relazionale CUP99	38
		4.4.3 AREAS	41
		4.4.4 Schema relazionale IRIDE	43
		4.4.5 SOWEB	48
		4.4.6 Schema relazionale SOWDATI	48
		4.4.7 wHealth	51
	4.5	Entity - Relationship	51
	4.6	Analisi dei requisiti	55
		4.6.1 Diagramma degli attori	55
	4.7	Progettazione concettuale	57
		4.7.1 Dimensional Fact Model	63
5	Rist	ultati e conclusioni	65
	5.1	Progettazione logica del DW clinico	65
	5.2	Conclusioni	68
\mathbf{A}	Mod	dellare un processo	6 9
	A.1	Synopsis	69
	A.2	Workflow	70
В	Sch	ermate applicativi	7 2
	B.1	SGP	72
	B.2	AREAS	76
	B.3	SOWEB	80

INDICE	INDICE

Bibliografia

Elenco delle figure

1.1	Trend di utilizzo del data Warehouse per scopi di ricerca clinica	11
1.2	Schema a stella del framework i2b2	13
1.3	Demo i2b2, esempio query: pazienti europei, di età maggiore di 65 anni con diagnosi di neoplasia	14
2.1	Schema dei livelli di una base dati.	16
2.2	Definizione schema della relazione $Appuntamenti.$	17
2.3	La tabella appuntamenti è in relazione alla tabella medici mediante la FK di Codice Medico Refertante. La PK presentano il campo corrispondente colorato di verde.	17
2.4	Costrutto entità <i>Paziente</i> con i relativi attributi e identificatore	19
2.5	Associazione <i>Prenotazione</i> connessa all'entità paziente e richiesta appuntamento	19
2.6	Costruzione dello schema concettuale	20
2.7	Differenza di utilizzo OLTP vs OLAP.	22
2.8	Modello multidimensionale: l'evento paziente ha dimensioni: Data ricovero, Ricovero e Diagnosi	22
2.9	Infrastruttura del Data Warehouse suddivisa nei vari livelli	23
2.10	Costrutti Tropos per diagramma degli attori	27
3.1	Synospis diagram del percorso CAS	30
3.2	Workflow del percorso CAS	31
3.3	Swim-Lane Activity del percorso CAS	32
4.1	Analisi del dominio applicativo dell'ospedale	33
4.2	Architettura del data warehouse in progetto	34
4.3	Fonti dati delle sedi Humanitas Torino	35
4.4	Schermata di selezione del Database	36
4.5	Schermata Database CUP99 e descrizione campi utili	37
4.6	Applicativo SGP per gestione CUP	38

4.7	Formalismo utilizzato per chiave primaria e chiave esterna	38
4.8	Schema relazionale CUP99	39
4.9	Tabella di Anagrafica dei pazienti	39
4.10	Tabella delle Richieste	40
4.11	Tabella Appuntamenti	40
4.12	Tabella Prestazioni	40
4.13	Tabella Catalogo Ministeriale Prestazioni	40
4.14	Tabella Prestazioni Regionali	41
4.15	Tabella codifiche diagnosi	41
4.16	Tabella di Link tra equipe e personale medico	41
4.17	Tabella di descrizione degli stati appuntamento	41
4.18	Tabella con codice medici e unità	41
4.19	Tabella di Link tra equipe medica unità	41
4.20	Tabella Link tra medico e codice referto	42
4.21	Tabella Link tra appuntamenti e referti	42
4.22	Tabella unità/agenda	42
4.23	Applicativo Web per gestione ADT	42
4.24	Schema relazionale IRIDE	43
4.25	Tabella dei Ricoveri	44
4.26	Tabella regime di ricovero	44
4.27	Tabella di descrizione reparti	44
4.28	Tabella Diagnosi	45
4.29	Tabella richieste prericoveri	45
4.30	Tabella Interventi/Procedure	45
4.31	Tabella Pronto Soccorso	46
4.32	Tabella Anagrafica Paziente	46
4.33	Tabella Descrizione Diagnosi	46
4.34	Tabella Descrizione Interventi	47
4.35	Tabella Modalità Dimissione	47
4.36	Tabella Trasferimento ospitanti ricovero	47
	Tabella di descrizione Stati	47
4.38	Tabella degli Appuntamenti	47
4.39	Tabella Lista di Attesa	48
4.40	Tabella schema relazionale SOWDATI	48
4 41	Tabella gestione atto operatorio	40

4.42	Tabella Diagnosi associate all'atto operatorio	49
4.43	Tabella Descrizione atto operatorio	49
4.44	Tabella Interventi/Procedure Associate all'atto operatorio	49
4.45	Tabella Equipe Medica	50
4.46	Tabella Anagrafica Sala Operatoria	50
4.47	Vista Medici.	50
4.48	Vista Tecnici	50
4.49	Vista Infermieri	50
4.50	Logo Applicazione web w Health per refertazione ambulatoriale. $\ \ . \ \ .$	51
4.51	Schema Entity - Relationship CUP99	52
4.52	Schema Entity - Relationship IRIDE	53
4.53	Schema Entity - Relationship SOWDATI	54
4.54	Diagramma degli attori del sistema ospedaliero Humanitas	56
4.55	Albero degli attributi	58
4.56	Modulo di richiesta CAS	59
4.57	Modulo allergie/intolleranze	59
4.59	Modulo Accettazione	60
4.58	Modulo anagrafica paziente	60
4.60	Modulo Sceening nutrizionale	60
4.61	Modulo Valutazione del dolore	61
4.62	Modulo Valutazione geriatrica	61
4.63	Modulo Protezione famiglie fragili	62
4.64	Modulo Karnofsky performance status	62
4.65	Modulo Sintesi valutazione psicologica	63
4.66	Dimensional fact model del DW oncologico	64
5.1	Progettazione logica: schema snowflake	66
A.1	Synopsis diagram	70
A.2	Formalismo workflow	71
D 1		79
B.1	Schermata di Ricerca Unità per la visualizzazione del Piano di Lavoro.	
B.2	Schermata piano di Lavoro selezionato per Unità	
B.3	Schermata piano di Lavoro selezionato per Unità	
B.4	Menù di AREAS	
B.5	Schermata anagrafica paziente	(1

ELENCO DELLE FIGURE

B.6	Schermata scheda di ricovero	78
B.7	Schermata scheda di Pre-ricovero	79
B.8	Schermata Ricerca paziente	80
B.9	Schermata informazioni generali atto operatorio	81
B.10	Schermata Descrizione atto operatorio	82
B.11	Schermata Dettagli atto operatorio	82

Capitolo 1

Introduzione

I dati clinici sono sempre stati il cuore del miglioramento in medicina. Con l'avvento della sanità digitale, l'enorme volume di informazioni ha creato una nuova sfida: trasformare terabyte di righe di dati in informazioni significative per migliorare la qualità, la ricerca clinica e i costi in sanità. Questa trasformazione richiede una soluzione tecnologica idonea ed efficiente, per questo è necessario progettare il data warehouse, un sistema informativo essenziale per elaborare informazioni cliniche dalle sorgenti a disposizione.

Organizzare, catalogare, strutturare e storicizzare le informazioni consente di trarre enormi benefici, tra i quali condurre ricerche e analizzare gli andamenti generali di uno specifico settore aziendale. I report generati e la realizzazione di grafici garantiscono, ai team di lavoro presenti nell'organizzazione ospedaliera, di orientare l'attenzione ad aree che presentano maggiori opportunità di miglioramento. La possibilità di fare affidamento ad un'unica risorsa, capace di eliminare eventuali discrepanze tra i dati che possono occorrere se si lavora sui dati operazionali(OLTP), ci garantisce una maggiore accuratezza delle informazioni, e questo è fondamentale soprattutto per le attività di ricerca clinica, in cui un'analisi statistica fornisce un grande impulso per servizi che possono andare dalla prevenzione di eventuali patologie alla possibilità di intervento per epidemie o allarmi conclamati.

Attualmente, grandi aziende e ospedali dipendono dai dati. Il sistema informativo ospedaliero è costituito da diversi domini in relazione al processo e alle operazioni che si intendono svolgere: per esempio, prenotare una visita medica, gestire un ricovero, refertare una visita, compilare un verbale del registro operatorio e così via. Spesso i domini elencati sono legati tra loro mediante relazioni progettate per la gestione del processo ospedaliero, e non eventualmente per l'estrazione di informazioni di rilevanza clinica. Tuttavia, la manipolazione dei dati clinici è una sfida ardua per i ricercatori: selezionare un campione di pazienti rappresenta un aspetto critico, iterativo e dispendioso in termini di tempo, soprattutto quando la quantità di dati messi a disposizione non include un'integrazione orientata sul paziente [1].

La valutazione degli archivi clinici consente di studiare le tendenze e le informazioni nascoste all'interno dei dati che potrebbero incrementare significativamente la

Trend utilizzo DWH nel 2018

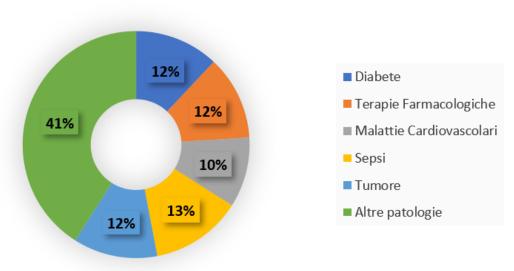


Figura 1.1: Trend di utilizzo del data Warehouse per scopi di ricerca clinica.

comprensione del progredire di una malattia e i trattamenti in uso. Sforzi passati in quest'area sono limitate primariamente a studi epidemiologici, ma di recente sulla base delle ricerche effettuate e riportate in Fig.1 si nota come il sistema data warehouse abbia esteso i suoi confini, dalle più rare alle più diffuse patologie.

In oncologia, c'è stata una recente spinta verso l'istituzione data warehouse di ricerca per supportare la medicina di precisione integrando i dati psico-fisici dei pazienti con risultati, dati di imaging e analisi molecolari e genomiche. [2]

I repository ospedalieri raccolgono una quantità enorme di dati su differenti sorgenti, contenenti informazioni relative a tutte le interazioni tra paziente e il sistema ospedaliero. In generale le informazioni estratte dal datawarehouse [3] risalgono a periodi antecedenti a quello in corso, e c'è un'attenzione particolare nel preservare l'identità del paziente mediante l'utilizzo di una codifica. I dati comunemente utili sono il sesso, l'età, il peso, il codice diagnosi del paziente, la data di dimissione del ricovero e se il paziente è ancora vivo oppure deceduto. Per l'esempio di malattie oncologiche è necessario ricostruire anche la storia clinica del paziente, considerando quindi le informazioni di eventuali procedure terapeutiche, atti operatori e riabilitazione.

L'estrazione delle informazioni necessarie per lo studio clinico deve avvalersi di un architettura del datawarehouse costruita ad *hoc* per un sistema ospedaliero. Un'architettura a tre livelli è una soluzione ideale, in quanto prevede la progettazione dei data mart, presenti nell'ultimo livello, cioè un'aggregazione di dati rilevanti per una specifica area applicativa.

1.1 Il framework i2b2

i2b2(Informatics for intergrating Biology and the Bedside) è un centro nazionale per l'informatica biomedica finanziato dal NIH, con sede presso Partners HealthCare. i2b2 ha sviluppato un framework per legare i dati di ricerca clinica e l'ampia banca di dati proveniente da ricerche scientifiche in modo da migliorare la comprensione delle basi genetiche di malattie complesse. La struttura è costituita da 2 grandi parti: la prima, back-end infrastructure, relativa alla sicurezza, accesso e gestione del repository. La seconda è un'applicazione per query e mining per permettere all'utente di effettuare domande specifiche sui dati.

i2b2 è costituito da un datawarehouse che permette di integrare informazioni sui pazienti da più risorse. Queste, possono includere cartelle cliniche elettroniche, risultati di laboratorio, dati genetici o di ricerca, così come risorse pubbliche come il registro di nascita. Le informazioni sono aggregate, 'pulite' e rese anonime. Una volta pronte sono messe a disposizione dell'utente, il quale sarà in grado di utilizzarlo per delle richieste. i2b2 è utilizzato maggiormente come piattaforma per registri di ricerca, difatti, la maggior parte delle query sono essenzialmente forme di identificazione di categorie di persone: quanti pazienti sono in terapia X o quanti aderiscono alla linea Y. Grazie alla costruzione di registri di ricerca, gli utenti possono aggiungere i propri dati collaborando con altre istituzione esterne. Le informazioni caricate nel datawarehouse i2b2 sono di tipo demografico(età, sesso, etnia, ecc.), diagnosi(ICD9), allergie, procedure, medicazioni, risultati di laboratorio.

1.2 Il modello i2b2

In [3] il data warehouse descritto in dettaglio, istituito da CCHMC(Cincinnati Children's Hospital Medical Center) è in costante evoluzione. I domini che il data warehouse i2b2 ricopre sono molteplici:

- Dati allergologici
- Dati su campioni biologici
- Dati demografici
- Dati di laboratorio
- Dati farmacologici
- Dati terapeutici
- Referti medici

Questa lista è destinata a crescere nel tempo.

Il framework i2b2 è semplice, ma molto efficiente. Consiste in *fatti* e *dimensioni*. Il fatto è l'informazione che viene richiesta, mentre le dimensioni sono i gruppi

gerarchici in grado di descrivere i fatti. I Database i2b2 utilizzano uno schema a stella che consiste in un fatto centrale circondato da più dimensioni.

Uno dei benefici dell'utilizzo di questo modello è rappresentato dalla facilità nell'aggiungere e integrare i dati da risorse multiple senza riprogettare il sistema o cambiare l'architettura sottostante. Tutte le nuove osservazioni sono semplicemente aggiunte alla tabella dei fatti. È possibile navigare attraverso i fatti grazie all'utilizzo dei metadati, ciò permette all'utente di creare delle categorie gerarchiche di concetti differenti nel database.

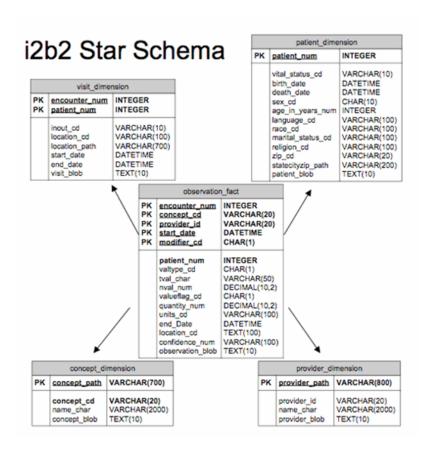


Figura 1.2: Schema a stella del framework i2b2.

In 1.3 è presente la schermata di una demo resa disponibile sul sito ufficiale di i2b2. A sinistra della finestra è presente l'elenco dei fatti clinici identificati nel datawarehouse. Ciascuno di essi contiene le dimensioni di interesse. Con un semplice drag&drop è possibile effettuare la selezione dei fatti, delle dimensioni e successivamente fare eseguire la query. In automatico questo strumento applicativo permette la visualizzazione dei report mediante grafici e dettagli sulla selezione. Come da esempio, questa ricerca consente di comprendere meglio quali fatti possono interessare per degli studi clinici, ispezionando una coorte di pazienti che presenta determinate caratteristiche.

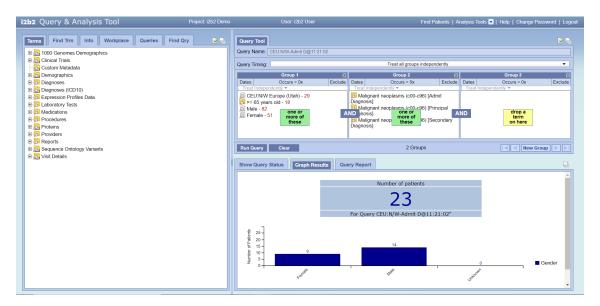


Figura 1.3: Demo i2b2, esempio query: pazienti europei, di età maggiore di 65 anni con diagnosi di neoplasia.

Capitolo 2

Materiali e metodi

2.1 Le Basi di Dati

I dati vengono raccolti dai processi, memorizzati, e possono essere reperiti in un secondo momento per estrarre informazioni. Spesso nei processi aziendali, i dati vengono raccolti per migliorare la conoscenza del processo, aumentare la produttività e i profitti, se gestiti in modo efficiente. Il database rappresenta una collezione di dati, organizzati in modo da essere fruibili dagli utenti a diversi livelli. Al dato grezzo, viene assegnata una codifica e una semantica, quest'ultima è utile per interpretare il dato e organizzarlo. I dati grezzi devono essere interpretati e correlati tra loro e inoltre, deve essere possibile ricercarli in modo organico e mirato. I dati devono essere mantenuti nel tempo, e non è conveniente memorizzarli in maniera non strutturata.

La gestione di grandi moli di dati strutturati si avvale di uno strumento software chiamato DBMS(Database Management system) [4]. Questo strumento consente:

- la creazione di nuovi database specificando la struttura logica dei dati;
- la possibilità di interrogare il database eseguendo delle query;
- di modificare dati;
- garantire sicurezza ed efficiente accesso ai dati;
- garantire l'integrità del dato evitando accessi simultanei da parte di più utenti nello stesso momento;

Tra gli obiettivi del DBMS vi è anche quello di raggiungere l'indipendenza fisica e logica dei dati. La soluzione porta alla realizzazione di un'architettura a 3 livelli, di seguito descritti e rappresentati in figura 2.1.

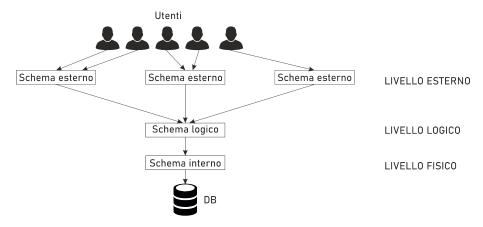


Figura 2.1: Schema dei livelli di una base dati.

- Schema esterno: è costituito dalla descrizione di una porzione del database mediante lo schema logico e consente all'utente di lavorare sul database. Su questo livello si raggiunge l'indipendenza sugli applicativi che fanno riferimento al database;
- Schema logico: struttura tabellare con le varie relazioni e con il loro schema, indipendente dalle strutture fisiche e dai programmi che vi accedono;
- Schema interno: descrive la strutture fisiche di memorizzazione.

Le modalità con cui è possibile utilizzare i dati sono diverse: si possono modificare i dati già esistenti, aggiornarli oppure si può interrogare la base dati, cioè reperire delle informazioni senza alterarne il contenuto. Le interrogazioni sono effettuate grazie ad un linguaggio SQL oppure mediante l'utilizzo di interfacce grafiche, in modo che anche un utente meno esperto può interrogare le informazione di interesse. Occorre standardizzare il modo in cui si va ad organizzare e gestire i dati. E' necessario di conseguenza definire un modello in modo da uniformare il processo. Con il tempo, i modelli si sono evoluti e la scelta di quale modello utilizzare dipende dal contesto che si vuole gestire. Ne esistono diversi: gerarchico o reticolare, ad oggetti, XML e relazionale. Uno dei modelli più utilizzato è quello relazionale. Deve il suo successo ad una proprietà chiamata indipendenza dei dati, in quanto, è indipendente dalla struttura fisica in cui vengono memorizzati i dati.

2.2 Il modello relazionale

La relazione è rappresentata da un'operazione matematica che ha come risultato un sottoinsieme del prodotto cartesiano di due o più insiemi. Il modello relazionale è una rappresentazione logica del database costituita da un insieme di concetti denominati *entità* che sono peculiari all'interno del contesto, i quali sono legati tra loro mediante delle *relazioni*. L'insieme delle proprietà comuni è detto schema della relazione.

Il concetto base di questo modello è lo schema della relazione cioè una tabella bidimensionale con una struttura fissata ed un nome definito, in cui le righe sono le istanze, cioè i valori puntuali delle entità(n-uple) e le colonne sono gli attributi. Nella progettazione si parte dai macro-concetti per poi passare a livello di memorizzazione in cui vengono definite come sono descritte le istanze.

Nonostante ci siano due concetti diversi è possibile correlare le informazioni tra schemi di relazioni distinte. Il riferimento non è creato sugli indirizzi fisici dei concetti, ma si hanno delle correlazioni tra le istanze dei due concetti attraverso i valori.

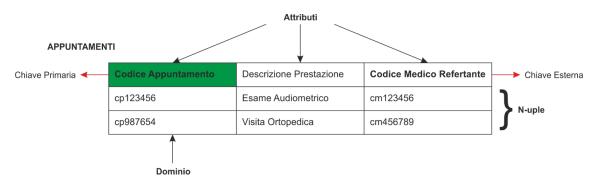


Figura 2.2: Definizione schema della relazione *Appuntamenti*.

In figura 2.3 sono mostrati gli schemi relazionali di due tabelle differenti. La relazione tra essi è creata grazie ad una correlazione tra i valori di *Codice Medico Refertante*, il cui attributo è la chiave esterna dello schema relazionale Appuntamenti, e il *Codice Medico*, il cui attributo è la chiave primaria dello schema relazionale Medici.

APPUNTAMENTI	Codice Appuntamento	Descrizione Prestazione	Codice Medico Refertante
	cp123456	Esame Audiometrico	cm123456
	cp987654	Visita Ortopedica	cm456789
MEDICI	Codice Medico	Nome	Cognome
	cm123456	Mario	Rossi
	cm456789	Giuseppe	Bianchi

Figura 2.3: La tabella appuntamenti è in relazione alla tabella medici mediante la FK di Codice Medico Refertante. La PK presentano il campo corrispondente colorato di verde.

Questo modello permette di memorizzare le informazioni relative ai diversi concetti in relazioni distinte e legate tra loro attraverso dei valori chiave. E' ovviamente necessario una consistenza tra i valori che costituiscono la relazione. Questa consistenza è garantita dai *vincoli*. Una delle problematiche che occorrono spesso è che i valori possono essere mancanti. In tal caso è necessario trattarli come "casi speciali" utilizzando degli standard specifici per ogni database. Ad esempio sarà possibile trattare tutti i valori mancanti come NULL.

2.2.1 Vincoli di integrità

Un vincolo è una proprietà che deve essere soddisfatta per tutte le istanze della base dati per mantenere la correttezza dei legami all'interno di essa. Possono essere di vario tipo:

- *Intra-relazionale:* vengono definiti dei vincoli, specifici solo su attributi della stessa relazione;
- Inter-relazionale: gli attributi di una tabella identificata come slave possono assumere soltanto dei valori specificati nella tabella identificata come master.

La tipologia di vincoli Intra-relazionali comprende:

- o Chiave primaria: deve essere possibile identificare univocamente ciascuna nupla;
- Chiave univoca: similare alla chiave primaria, ma in questo caso gli attributi possono assumere un valore nullo;
- Vincoli di tupla: esprime condizioni sul valore assunto da singole tuple, in modo indipendente dalle altre tuple della relazione. All'interno di una determinata relazione, se c'è una dipendenza tra valori, questa deve essere mantenuta;
- o *Vincoli di dominio:* esprime condizioni sul valore assunto da un singolo attributo di una tupla. Questo vincolo può essere definito a livello di database ma anche gestiti a livello applicativo.

La tipologia di vincoli **Inter-relazionali** comprende:

• Vincolo d'integrità referenziale: lega tabelle differenti. Gli attributi di una tabella slave possono assumere soltanto i valori specificati in un'altra tabella master. A questo scopo i database utilizzano determinati costrutti per rispettare questo vincolo, come ad esempio la chiave esterna.

E' fondamentale definire quali siano i vincoli affinché il DBMS possa segnalare la violazione dello stesso.

2.2.2 Modello Entità - Relazione

P.P. Chen nel 1976 ideò uno strumento di analisi che viene usato per la progettazione inziale delle basi di dati; è un modello concettuale che permette di descrivere, attraverso lo schema Entity - Relationship il mondo di interesse e consente di rispettare gli obiettivi definiti nella fase di analisi dei requisiti. I costrutti principali del modello vengono descritti nei paragrafi seguenti. La comprensione di questo modello è fondamentale in quanto permetterà di ricostruire lo schema concettuale delle sorgenti in analisi, mediante il processo di reverse engineering.

2.2.3 Costrutto entità e attributi dell'entità

L'entità in figura 2.4, rappresenta l'elemento di base su cui si costruisce il modello concettuale. Esso è definito dal nome dell'entità racchiuso all'interno di un rettangolo. Ciascuna entità è caratterizzata dalle proprietà, chiamate attributi. Gli attributi circondano il rettangolo dell'identità. L'entità è identificata univocamente mediante un identificatore che può comprendere uno o più attributi. La codifica diagrammatica usata per gli identificatori è mediante un pallino nero.

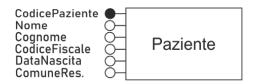


Figura 2.4: Costrutto entità *Paziente* con i relativi attributi e identificatore.

2.2.4 Associazioni

Il secondo elemento da introdurre, non meno importante, è l'associazione, il costrutto che mette in relazione due o più entità. La rappresentazione grafica del costrutto è un rombo che lega tutte le entità ad esso connesse. Anche un'associazione, così come un'entità non può avere attributi uguali.

E' necessario indicare anche il numero massimo e il numero minimo di elementi delle altre entità coinvolte dall'associazione, i valori sono detti *cardinalità*. Il numero generico maggiore di zero è indicato con n. Se la cardinalità minima è zero, l'associazione è detta parziale. Se la cardinalità minima è uguale ad 1 allora l'associazione è detta totale.

Per completezza è possibile definire anche un nome per l'associazione. Nell'esempio di figura 2.5. l'entità paziente è associata all'entità Richiesta Appuntamento. Questa associazione è necessaria dal momento che un paziente vuole prenotare una visita ambulatoriale. La cardinalità (1,N) rispetto all'entità paziente dice che da 1 ad N pazienti possono richiedere un appuntamento, al pari della richiesta appuntamento. Nello studio affrontato, in realtà è preferibile associare una sola richiesta appuntamento per ciascun paziente purché le prestazioni ad esso associate siano della stessa branca per al massimo otto prestazioni. Per includere tutte le casistiche è opportuno avere come cardinalità (1,N).



Figura 2.5: Associazione *Prenotazione* connessa all'entità paziente e richiesta appuntamento.

2.2.5 Reverse Engineering

La procedura di reverese engineering consente di ricostruire lo schema concettuale a partire dall'analisi delle componenti della base dati, raccogliendo la documentazione disponibile e la definizione di viste e tabelle. Questa procedura prevede due fasi:

- Estrazione: prevede una parte iniziale di analisi delle tabelle e delle viste, al fine di recuperarne lo schema logico descritto dal modello entità relazione. In questo specifico caso di analisi, per ogni database si analizzano le tabelle contenenti informazioni di maggiore rilevanza clinica. Per ricreare le associazioni è necessario revisionare lo script SQL, reso disponibile dalla piattaforma d'interfaccia TOAD. In questo modo sarà possibile individuare le chiavi primarie ed esterne. Un ulteriore strumento utile per l'estrazione delle informazioni è fornito direttamente dagli applicativi utilizzati, nel caso specifico SGP e AREAS.
- Concettualizzazione: la descrizione funzionale della base dati è rappresentato dallo schema entità relazione, il quale favorisce una migliore comprensione di ciascuna entità presente e delle associazioni tra esse.

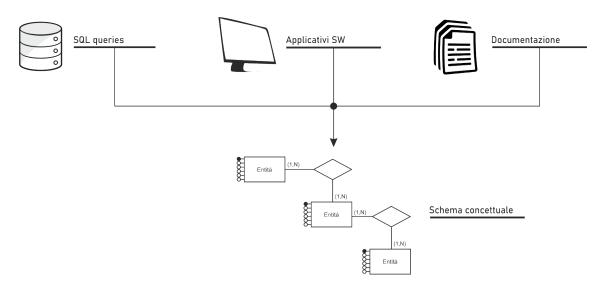


Figura 2.6: Costruzione dello schema concettuale.

2.3 Data warehouse

2.3.1 Definizione

"Data warehouse(DW): è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:

- È orientata ai soggetti di interesse;
- È integrata e consistente;
- È rappresentativa dell'evoluzione temporale e non volatile."

Questo enunciato [5] tratta da Inmon(1996) esplica non solo la definizione di DW ma anche gli obiettivi che si vogliono raggiungere avvalendosi di questa soluzione tecnologica. La collezione di dati presenti all'interno del DW non include nuove informazioni, ma solo dati collezionati negli anni. In realtà le aziende che non utilizzano il DW effettuano le interrogazioni su sistemi OLTP (On Line Transaction Processing), i quali scrivono e mostrano i record in tempo reale in base alle varie esigenze operative, perciò le query avvengono solo sulle basi dati di interesse.

2.3.2 OLTP E OLAP

I sistemi OLTP (On Line Transaction Processing), o transazionali, consentono la gestione dei record in tempo reale. L'enfasi principale di questo sistema è velocizzare i processi di interrogazione, aggiornamento e modifica dei dati, evitandone quindi la ridondanza. Il sistema OLTP effettua giornalmente delle transazioni, ovvero delle operazioni in sequenza, le quali devono risultare tutte eseguite correttamente, oppure tutte fallite in blocco. Questo tipo di proprietà è denominata atomicità. Più utenti hanno accesso ai dati di tipo transazionale e possono effettuare operazioni di rapida esecuzione.

Per i sistemi di tipo OLAP (On Line Analytical Processing) è solitamente la direzione che ha accesso ai dati, ma stavolta non si tratta di record recenti ma coprono un periodo temporale molto vasto. Questa visione più ampia pone le basi per il livello multidimensionale dei dati, fondamenta del concetto di data warehouse. Infatti, l'utente ha un accesso di sola lettura ai dati, e potrà interfacciarsi ad esso attraverso applicativi di elaborazione dati e reportistica.

Lo schema in figura 2.7 mostra l'interazione tra i due sistemi. Dal punto di vista architetturale essi vanno considerati separatamente, il collegamento fra essi è dettato da un processo chiamato ETL e che consente di effettuare la pulizia, trasformazione e caricamento dei dati all'interno del data warehouse.

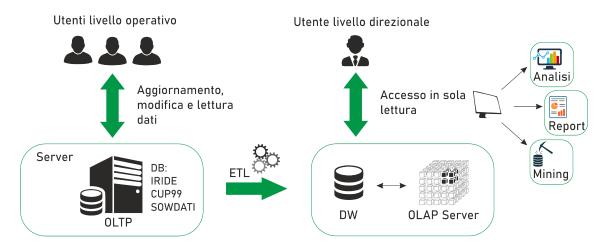


Figura 2.7: Differenza di utilizzo OLTP vs OLAP.

2.3.3 Modello multidimensionale

I dati sono ben strutturati in un cubo multidimensionale. Il modello multidimensionale è costituito dai *fatti* del mondo aziendale. La descrizione quantitativa di ciascun evento, ovvero il verificarsi di una ricorrenza per un determinato fatto, fa uso delle *misure*. La numerosità degli eventi, che possono occorrere in un'azienda, comporta la necessità di raggrupparli in uno spazio n-dimensionale i cui assi sono detti *dimensioni*.

Nell'esempio riportato in figura 2.8 l'evento potrebbe raggruppare un numero di pazienti ricoverati in una data ben precisa a cui è stata diagnosticata la stessa patologia.

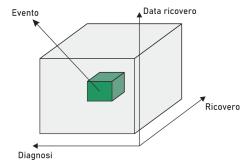


Figura 2.8: Modello multidimensionale: l'evento paziente ha dimensioni: Data ricovero, Ricovero e Diagnosi.

2.3.4 L'architettura per il Data Warehouse

L'infrastruttura del Data Warehouse [2.9] è basata su livelli al fine di creare una separazione tra i dati delle sorgenti e i dati di sintesi, garantendo in questo modo l'accessibilità alle informazioni anche quando le sorgenti dati non sono disponibili. Di seguito sono elencati e descritti brevemente i diversi livelli:

• Livello delle sorgenti: le fonti dati sono archiviate generalmente in database di produzione, spesso utilizzando anche diverse tecnologie o risorse software.

Tutti i dati contenuti nelle diverse sorgenti vengono raccolti nell'area di staging dove saranno sottoposti a controlli e validazioni;

- Livello dell'alimentazione: gli strumenti ETL(Extraction, Transformation and Loading) sono gli attori principali di questo livello, utili per estrarre i dati dalle diverse sorgenti, pulirli, validarli e caricarli nel DW;
- Livello del data warehouse: i flussi dei dati vengono convogliati in un unico contenitore che può essere direttamente consultato, ma che lascia spazio anche alla possibilità di costruire dei blocchi dipendenti dal DW primario, chiamati data mart. I data mart sono più facilmente consultabili in quanto contengono aggregazioni di dati ben strutturati e contenenti le informazioni rilevanti per una particolare categoria o area del business;
- Livello di analisi: l'elaborazione dei dati integrati è facilitata da strumenti di analisi e reportistica.

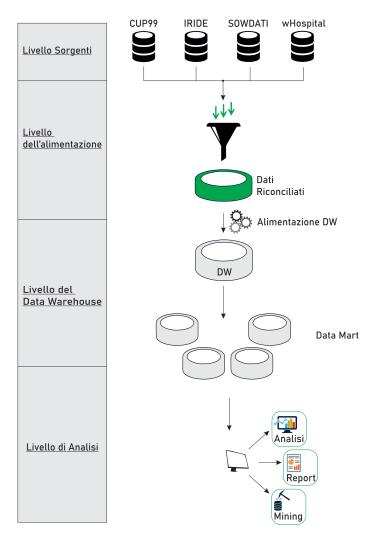


Figura 2.9: Infrastruttura del Data Warehouse suddivisa nei vari livelli.

2.3.5 Gli strumenti ETL

Le sorgenti sono classificate in differenti categorie e i dati sono salvati in formati differenti per questo motivo è necessario convertirli in un unico formato. Gli strumenti ETL presentano diverse funzioni e risorse utili per ottenere un dato che sia strutturato, pulito e validato.

- *Estrazione*: i dati sono presi da diversi database e mediante l'utilizzo di filtri è possibile selezionare i dati clinici di interesse;
- *Pulitura*: i dati estratti spesso sono 'sporchi' perciò è necessario eliminare eventuale dati duplicati, dati incompleti o inconsistenti;
- Trasformazione: i dati vengono convertiti in un unico formato;
- Caricamento: i dati vengono accolti nel DW. Questo processo può avvenire in due modi:
 - Refresh: i dati vengono riscritti completamente, sostituendoli con i nuovi;
 - Update: i dati vengono accodati o aggiornati nel DW senza sostituzione dei precedenti.

2.3.6 Data Mart

Il data mart è un'aggregazione specifica di dati, orientato ad una specifica richiesta aziendale che riguarda un singolo dipartimento. È possibile isolare l'uso, la manipolazione e lo sviluppo dei propri dati a specifiche figure aziendali incrementando il livello di sicurezza dei dati.

Il data mart consente di rendere le informazioni facilmente accessibili grazie alla progettazione strutturata dei dati presenti in esso, migliorando i tempi di risposta dell'utente finale, e consentendo a quest'ultimi di avere accesso al tipo specifico di dati che devono visualizzare più spesso.

E comune utilizzare più di un data mart per soddisfare le esigenze di ogni singola unità di business.

2.4 Fasi di Progettazione di un Data Warehouse

Quando si progetta un Data Warehouse bisogna considerare i diversi rischi che possono occorrere e che possono costituire un motivo di fallimento dell'intera progettazione. I rischi possono essere legati: alla gestione del progetto, in questo caso l'azienda potrebbe essere diffidente nel voler condividere i propri dati con le parti coinvolte al progetto; oppure legati alle tecnologie, ai dati e alla progettazione, in quanto la qualità dei dati influenza la corretta interpretazione degli stessi. Infine, non per minore importanza vi sono anche i rischi legati all'organizzazione, la direzione deve rendere consapevole l'utente finale che le informazioni estratte possano essere utili per la crescita aziendale.

La buona riuscita del progetto può essere ostacolata per i motivi precedentemente

elencati, a questo proposito è necessario affrontare la decisione metodologica per il processo di progettazione del Data Warehouse.

Esistono due approcci differenti:

- Top Down: è il metodo che richiede un maggiore investimento in termini di tempo e costi, ma permette di avere una visione integrale dell'intero progetto. Consente di integrare e riconciliare tutte le sorgenti dell'azienda;
- Bottom Up: è il metodo più utilizzato, in questo caso il Data Warehouse si ottiene costruendo e assemblando più data mart ciascuno dei quali fa riferimento ad una singola unità di business.

Al fine di realizzare l'obiettivo proposto, si segue l'approccio Top - down per l'azienda ospedaliera Humanitas – Gradenigo. La continua evoluzione degli applicativi ospedalieri e i cambiamenti delle sorgenti, in cui i dati operazionali vengono archiviati, influenza la selezione delle sorgenti da considerare. Nel corso dell'anno 2019 e verso la metà del 2020, si è portato a termine l'integrazione di un nuovo sistema di refertazione ambulatoriale, che abbandona l'archiviazione delle informazioni cliniche sul vecchio database(CUP99) per sostituirlo con un nuovo database studiato al meglio per l'organizzazione delle informazioni cliniche.

Infatti, l'introduzione di un nuovo sistema di refertazione ha permesso di strutturare i dati in maniera più organizzata all'interno del database. Questo permetterà, in seguito, di mappare più facilmente le sorgenti di interesse e di legarle tra loro per tutte le diverse aree cliniche specialistiche.

2.4.1 La progettazione del data warehouse

Il processo di realizzazione del data warehouse è descritto dalle seguenti fasi:

- 1. Analisi e riconciliazione delle sorgenti: questa fase prevede la ricognizione e mappatura delle sorgenti disponibili (ricongnizione), l'integrazione e omogeneizzazione delle stesse per eliminare le inconsistenze. In uscita da questa fase si deve ottenere uno schema riconciliato;
- 2. Analisi dei requisiti: il progettista deve coinvolgere tutte le figure professionali dell'azienda per raccogliere e stilare gli obiettivi da raggiungere. In seguito, si descrivono i metodi per l'analisi dei requisiti;
- 3. Progettazione concettuale: in questa fase si caratterizza nel dettaglio il data warehouse ed è il passo fondamentale per rappresentare concettualmente il modello multidimensionale. Si costruisce lo schema di fatto per ciascun fatto di interesse con dimensioni e misure annesse;
- 4. *Progettazione logica*: consente di creare lo schema logico utilizzando forme di ottimizzazione delle strutture dati;
- 5. Progettazione dell'alimentazione: in questa fase vengono definite le modalità di caricamento dei dati provenienti dalle sorgenti dati nel data warehouse.

Delle fasi elencate solo le prime quattro rientrano nell'obiettivo di questa tesi, a questo proposito, di seguito sono descritte le modalità e le scelte progettuali considerate.

2.4.2 Analisi e riconciliazione delle sorgenti

La fase di analisi delle sorgenti è fondamentale per acquisire una dettagliata conoscenza sulle diverse sorgenti su cui si basa l'intero sistema ospedaliero. L'analisi delle fonti dati è un processo complesso e richiede la collaborazione di tutte le figure esperte che utilizzano l'applicativo o che si occupano della manutenzione e pulizia dei database. Inoltre, è necessario tenere conto della possibilità di avere sorgenti eterogenee, in quanto utilizzano diverse tecnologie per archiviare i dati, ad esempio database relazionali, fogli excel o flat file. A tal proposito, alla fase di ricognizione e normalizzazione, che coincide con l'acquisire approfonditamente la conoscenza sugli schemi locali dei database, deve susseguirsi la fase di riconciliazione degli schemi logici precedentemente mappati, per ottenere uno schema riconciliato globale che presenta l'integrazione completa dei dati tra le diverse sorgenti.

2.4.3 Analisi dei requisiti

Per ottenere le informazioni necessarie all'utente finale per l'analisi, è importante determinare fatti, misure e dimensioni. Un metodo per semplificare la progettazione, evitando in seguito di apporre opportune integrazioni, è raccogliere e strutturare le informazioni ricavate dagli utenti finali del data warehouse. L'approccio seguito è basato sulla metodologia Tropos.

La metodologia ottimale per questa fase si ottiene eseguendo un'analisi dei requisiti basata sugli obiettivi. E' bene considerare due prospettive di analisi: modellazione dell'organizzazione centrata sugli skateholder, e la schematizzazione del processo ospedaliero di interesse .

2.4.4 Metodologia Tropos

Tropos [6] è una metodologia di sviluppo software orientata agli oggetti e basata su due concetti principali: agenti e obiettivi. I passi da seguire sono:

- 1. Analisi degli obiettivi: si deve disegnare una diagramma degli attori, successivamente gli obiettivi assegnati per ciascun attore;
- 2. Analisi dei fatti: vengono identificati i fatti;
- 3. Analisi dimensionale: a ciascun fatto si associano le dimensioni;
- 4. Analisi delle misure: a ciascun fatto si associano le misure.

Per costruire il diagramma degli attori, ci si avvale dei costrutti illustrati in figura 2.10.

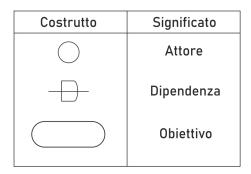


Figura 2.10: Costrutti Tropos per diagramma degli attori.

2.4.5 Progettazione concettuale

I modelli Entità – Relazione non possono costituire la base di fondamento del data warehouse, per questo motivo è necessario costruire i Dimensional Fact Model. Per raggiungere il risultato finale si seguono i seguenti passi:

- 1. costruzione dell'albero degli attributi;
- 2. potatura e innesto dell'albero degli attributi;
- 3. definizione delle dimensioni;
- 4. definizione delle misure;
- 5. creazione dello schema di fatto.

2.4.6 Progettazione logica

A partire dal modello concettuale, la fase logica permette di definire un modello implementabile dal DBMS. Il prodotto che si ottiene al termine di questa fase è denominato schema logico, in cui si mette in evidenza la multidimensionalità del data warehouse. Il modello multidimensionale che si utilizza è il MOLAP (Multidimensional On – Line Analytical processing).

Lo schema logico scelto è lo schema snowflake [7], che rappresenta un'estensione dello schema a stella,in con l'ggiunta di ramificazioni con altre dimensioni. I benefici di questo schema sono molteplici: azzeramento della ridondanza dei dati, minore occupazione di spazio per la conservazione dei dati e una maggiore velocità di accesso ai dati, contro la necessità di eseguire query più complesse per la ricerca.

Capitolo 3

Il Percorso CAS

Il Centro Accoglienza e Servizi(CAS) è la struttura di riferimento per pazienti affetti da malattia tumorale. Il CAS nell'ambito della Rete Oncologica del Piemonte e Valle d'aosta, svolge un ruolo di assistenza, coordinamento delle attività dei diversi professionisti, orientamento e supporto. Il CAS svolge molteplici ruoli:

- Accoglie il paziente, informandolo in merito ai servizi erogati, alle modalità di accesso, alle prenotazioni;
- Fornisce un progrmma di supporto psico-oncologico e psico-sociale;
- Svolge un ruolo amministrativo durante il percorso diagnostico-terapeutico;
- Coordina il GIC, ovvero il Gruppo Interdisciplinare di Cure;
- Prenota le prestazioni diagnostiche;
- Garantisce una comunicazione costante tra i diversi attori coinvolti;

3.1 I dati clinici nel percorso CAS

Il percorso CAS inzia nel momento in cui il medico di medicina generale, uno specialista ambulatoriale o i medici del pronto soccorso identificano un sospetto tumorale sul paziente. In tal caso, il soggetto abilitato all'attivazione del percorso CAS compila il primo modulo di richiesta. Nel caso in cui uno specialista richiede l'attivazione del CAS, vengono indicati, da subito, anche tutti gli esami diagnostici che il paziente dovrebbe eseguire.

In parallelo agli esami diagnostici, vi è la visita infermieristica CAS, stabilita entro pochi giorni dal momento della richiesta, la quale consente di avere un quadro delle condizioni psico-fisiche del paziente. Grazie alle informazioni raccolte, l'infermiere del CAS:

- valuta il livello di ansia e preoccupazione manifestato dal paziente;
- valuta il problema di salute, il percorso terapuetico e le abitudini di vita della persona assistita;

- valuta le condizioni cliniche del paziente;
- valuta il livello di autonomia nel svolgere le mansioni basilari;
- raccoglie l'anamnesi assistenziale;
- valuta il dolore e ne illustra su appositi moduli il sito del dolore;
- valuta la fragilità del paziente;
- effettua valutazione geriatrica
- effettua screening nutrizionale

Tutti i dati psico-fisici del paziente vengono raccolti grazie ad un sistema informatizzato ed in uso presso l'ospedale Humanitas - Gradenigo. In aggiunta, al termine degli esami di stadiazione e delle visite infermieristiche, avviene la discussione del GIC in cui si stabilisce il percorso di cura del paziente. Anche in quest'utlimo caso, i dati clinici del referto verranno salvati grazie agli appositi moduli informatizzati.

3.2 Synopsis, Workflow e Swim-Lane Activity

La comprensione del processo clinco è fondamentale per avere un quadro dettagliato dei dati che vengono inseriti all'interno dei moduli informatizzati, soprattutto perchè si deve definire, successivamente, l'integrazione tra il sistema wHealth utilizzato per la refertazione ambulatoriale e percorsi CAS con gli applicativi che riguardano le altre aree ospedaliere.

3.2.1 Synopsis

Il synospsis diagram rappresenta il contesto del processo, vengono definiti:

- Attori
- Evento che determina l'inizio del processo
- Dati di input
- Dati di output
- I risultati del processo

In 3.1 si illustra l'analisi del contesto *Percorso CAS*.

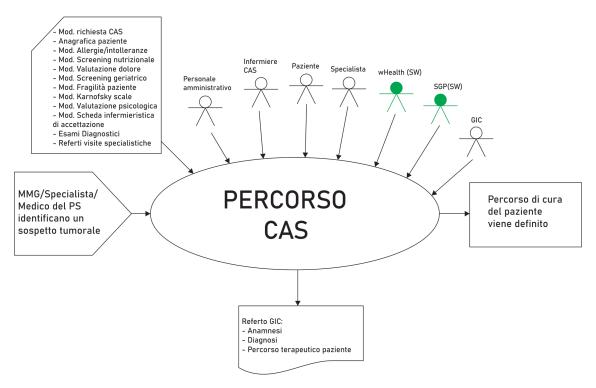


Figura 3.1: Synospis diagram del percorso CAS.

3.2.2 Workflow

Il workflow è utilizzato per descrivere le azioni compiute durante il processo. In 3.2 vengono illustrate le attività del percorso CAS. I software utilizzati, se presenti nella specifica attività, sono segnalati da un apposito riquadro. Il formalismo indicante il tempo in cui si svolge l'attività, è generico poichè i tempi di attesa dipendono dalle agende ospedaliere.

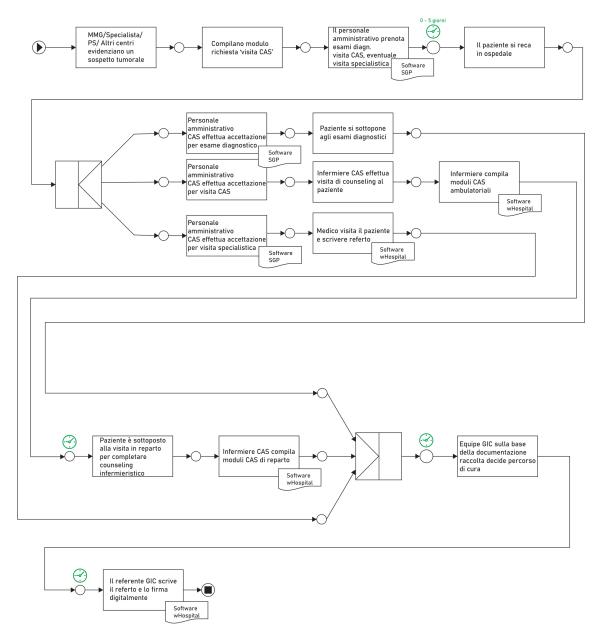


Figura 3.2: Workflow del percorso CAS.

3.2.3 SWim-Lane Activity

Lo Swim-Lane activity è utilizzato per descrivere le azioni compiute durante il processo, dai diversi attori, in maniera molto dettagliata. In 3.3 vengono illustrate le attività nel dettagio del percorso CAS. I software utilizzati, sono considerati attori facenti parte del processo.

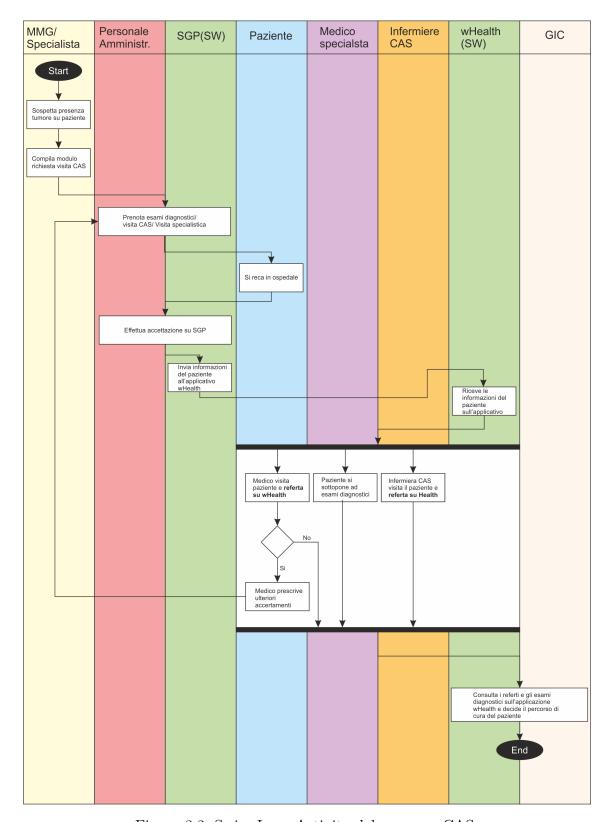


Figura 3.3: Swim-Lane Activity del percorso CAS.

Capitolo 4

DW Ospedale Humanitas – Gradenigo

L'ospedale Humanitas – Gradenigo è una società coordinata e subordinata a Humanitas S.p.A., con l'obiettivo primario di favorire un miglioramento continuo al paziente erogando cure e assistenza clinica di qualità. L'azienda ospedaliera eroga prestazioni di diagnosi e cura in regime di ricovero o in forma ambulatoriale, sia a carica del Servizio Sanitario Nazionale che in regime di libera professione.

4.1 I dominio applicativo dell'ospedale

L'analisi del dominio applicativo mette al centro del contesto del DW clinico i record del paziente. Ad esso sono associati non solo tutti i dati del ricovero, delle prestazioni effettuate e degli interventi chirurgici, ma comprende anche tutti i test del laboratorio analisi e del laboratorio di anatomia patologica. In figura 4.1 è mostrato uno schema semplificato del dominio applicativo dell'ospedale.



Figura 4.1: Analisi del dominio applicativo dell'ospedale.

È necessario specificare che il dominio applicativo non rappresenta l'elenco delle sorgenti, bensì la descrizione dello scenario ospedaliero, maggiormente approfondita nella fase di analisi dei requisiti, in cui il diagramma degli attori illustrerà i ruoli fondamentali di ciascun attore e i collegamenti fra essi.

La comprensione del dominio applicativo costituisce il primo passo fondamentale per definire meglio il contesto su cui si sta lavorando e sfruttare al meglio i dati a disposizione in modo consistente ed efficace; questa fase aiuta il progettista a raccogliere le prime informazioni sui software utilizzati dall'ospedale, quali SGP per il CUP, wHospital per la refertazione ambulatoriale, Areas per l'ADT e i vari software dipartimentali. L'associazione tra i Database e i Software utilizzati è descritta in dettaglio successivamente, nella fase di analisi delle fonti dati.

4.2 Pianificazione del DW

L'architettura scelta per il data warehouse è a tre livelli. Il livello delle sorgenti riassume le fonti dati considerate, descritte nel dettaglio nelle fasi di progettazione successive. Sono stati ricavati gli Entity – Relationship delle sorgenti CUP99, IRIDE e SOWDATI. Tuttavia, durante lo sviluppo della progettazione, il database CUP99, che archiviava tutti i dati relativi alla refertazione e visite ambulatoriali, ha subito una modifica significativa, in quanto tutti i referti vengono archiviati in un nuovo database che si interfaccia con una nuova applicazione web introdotta in quest'ultimo anno nelle strutture Humanitas – Torino. Questo progetto ha portato ad un sistema di refertazione che consente di strutturare ed organizzare meglio i dati clinici per futuri scopi di ricerca.

La Fig 4.3 mostra l'architettura di progettazione del data warehouse, la quale non esclude la possibilità di integrazione di nuove sorgenti e quindi la costruzione di nuovi data mart specifici per l'unità di interesse.

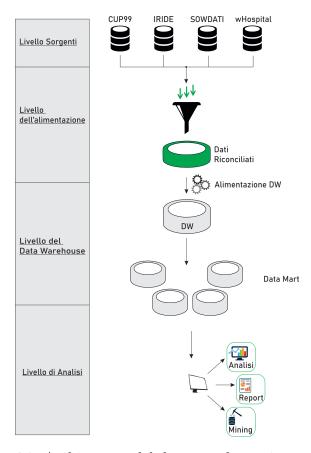


Figura 4.2: Architettura del data warehouse in progetto.

4.3 Analisi delle fonti dati

L'ospedale Humanitas – Gradenigo, Cellini, Fornaca e San Luca - utilizza tecnologie differenti a seconda del contesto sanitario, questo si evince dal numero di sorgenti da considerare per la costruzione del DW. Ciascuna di queste sorgenti si avvale di un Software specifico per la gestione dei pazienti nei vari contesti. La Figura seguente riassume l'elenco complessivo delle sorgenti.

Applicazione	Tecnologia	Sede	Contesto
SGP	Oracle	Gradenigo	Gestione paz Esterni
AREAS	Oracle	Gradenigo	Gestione paz Interni
SOWEB	Oracle	Gradenigo	Dipartimento sala operatoria
wHealth	SQL	Gradenigo, Cellini, Fornaca, San Luca	Refertazione ambulatoriale
HCS	Access	Cellini, Fornaca, San Luca	Gestione paz Esterni
HCS	M. Access	Cellini, Fornaca	Gestione paz Interni
BEAG	SQL	Cellini, Fornaca	Dipartimento sala operatoria
Openlis	Oracle	Gradenigo, Cellini, Fornaca	Laboratorio analisi
AP	Oracle	Gradenigo, Fornaca	Anatomia patologica
ENDOX	SQL	Gradenigo, Cellini, Fornaca	Endoscopia digestiva
Carestream	Oracle	Gradenigo, Cellini, Fornaca	Radiologia
MPI	SQL	Gradenigo, Cellini, Fornaca, San Luca	Anagrafica
PS-WEB	Oracle	Gradenigo	Pronto soccorso

Figura 4.3: Fonti dati delle sedi Humanitas Torino.

- CUP99, il per la gestione del CUP;
- IRIDE, il quale contiene le informazioni relative alla gestione ricoveri e pronto soccorso;
- SOWDATI, il quale contiene i dati relativi alla sala operatoria;
- wHealth, per la refertazione ambulatoriale e i moduli CAS.

Le tabelle considerate non sono indipendenti tra loro; la dipendenza è creata da entità comuni come ad esempio i dati relativi al paziente, ai medici o le prestazioni erogate. La relazione tra queste sarà esplicitata chiaramente nello schema riconciliato descritto dal diagramma Entity - Relationship globale, nella fase di riconciliazione delle fonti dati.

4.4 Ricognizione

La fase di ricognizione consente di analizzare gli schemi delle sorgenti disponibili. Il sistema informativo dei database selezionati è gestito attraverso Toad, un software applicativo per la gestione dei dati relazionali, che si interfaccia ai database ORACLE e si avvale del linguaggio SQL. L'accesso agli schemi presenti su Toad è protetto dalle credenziali di accesso. Di seguito è mostrata una schermata dell'applicativo sopracitato, figura 4.4, per la selezione del Database di interesse.

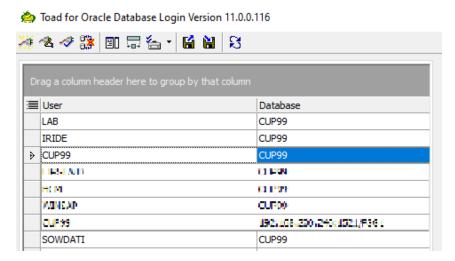


Figura 4.4: Schermata di selezione del Database.

A seguito della selezione e accesso allo schema di interesse, sarà possibile visualizzare le tabelle presenti all'interno del database relazionale. A tal proposito viene mostrata una schermata che illustra il passaggio appena descritto, figura 4.5.

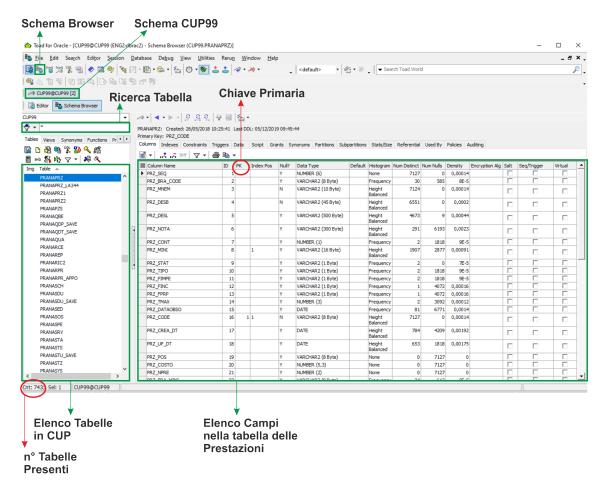


Figura 4.5: Schermata Database CUP99 e descrizione campi utili.

L'icona Schema Browser permette la visualizzazione dell'elenco completo dello schema selezionato, ad esempio CUP99. La consultazione e selezione delle molteplici tabelle presenti nel Database deve essere eseguita con la collaborazione e coordinazione di un esperto che conosce i nomi codificati delle tabelle presenti. La navigazione nel DB è facilitata dal campo di ricerca.

Una volta selezionata la tabella, l'elenco dei campi presenti all'interno di essa saranno mostrati automaticamente. I dettagli utili per ogni campo includono: ID(identificativo sequenziale del campo), PK(chiave primaria), Data Type(Tipo di dato), Num Nulls (numero di Null presenti).

Gli schemi relazionali contengono dati di tipo clinico e amministrativo, perciò il filtraggio delle informazioni necessarie per la costruzione dei data mart clinici deve essere eseguito avvalendosi dell'aiuto dei software ospedalieri utilizzati nella sede Humanitas – Gradenigo, a cui fanno riferimento i Database considerati precedentemente.

I paragrafi seguenti descriveranno l'analisi e ricognizione per ciascuna sorgente utile al nostro obiettivo.

4.4.1 SGP



Figura 4.6: Applicativo SGP per gestione CUP.

Il software SGP(Servizio Gestione Prenotazioni) è utilizzato per la gestione delle prenotazioni delle prestazioni erogate dall'ospedale, ovvero per il Centro Unico di Prenotazione(CUP). Le informazioni archiviate durante i processi di prenotazione delle prestazioni riguardano:

- Anagrafica dei pazienti;
- Anagrafica dei medici;
- Unità Diagnostiche;
- Agende(Piani di Lavoro);
- Le richieste per gli appuntamenti;
- Le prestazioni erogate;
- Gli stati degli appuntamenti;
- L'elenco delle prestazioni ministeriali e relativo tariffario;

Lo schema relazionale a cui SGP fa riferimento è CUP99. Per una migliore comprensione del processo di selezione, è presente l'appendice (A.1) in cui saranno mostrate delle schermate di SGP da cui si potranno visualizzare i campi di maggiore interesse

Il formalismo in figura 4.7 permetterà di identificare quali di questi campi siano chiavi primarie e/o chiavi esterne.

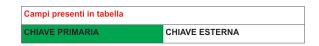


Figura 4.7: Formalismo utilizzato per chiave primaria e chiave esterna.

4.4.2 Schema relazionale CUP99

Le tabelle relazionali presenti all'interno dello schema CUP99 sono in totale 743. Dopo l'ispezione del database 15 tabelle sono state selezionate e sono riportate di seguito. Per ciascuna di esse, oltre alla codifica della stessa è riportata la corrispettiva descrizione, figura 4.8.

La prima tabella ad essere esplorata è sicuramente PRANAPAZ, figura 4.9, la quale contiene l'anagrafica dei pazienti.

Nome Relazione	Descrizione Relazione
PRANAPAZ	Campi Anagrafici Paziente
PRPRERIC	Informazioni sulle Richieste
PRPREAPP	Informazioni sugli Appuntamenti
PRANAPRZ	Descrizione Prestazioni
EG_CATALOGO_REGIONALE	Catalogo Regionale Prestazioni
PRANANTR	Catalogo Ministeriale Prestazioni
NEW_PRANAPRR	Prestazioni Ministeriali Regionali
PRANAICD	Diagnosi con codifica ICD9
PREQUPRS	Link tra Equipe e Personale Medico
PRANASTA	Descrizione e Codifica Stati
PRANAOPE	Codici dei Medici
PREQUUNI	Link tra Equipe Medica e Unità
PRRICREF	Referti
PRAPPPREF	Link tra Appuntamenti/Attività e Referti
PRANAUNI	Tabella Unità

Figura 4.8: Schema relazionale CUP99.

PRANAPAZ			
Nome Campo	Descrizione Campo	Tipo	
PAZ_COGN	Cognome Paziente	VARCHAR2(40 Byte)	
PAZ_NOME	Nome Paziente	VARCHAR2(40 Byte)	
PAZ_SESS	Sesso Paziente	VARCHAR2(1 Byte)	
PAZ_CFIS	Codice Fiscale Paziente	VARCHAR2(16 Byte)	
PAZ_DNASC	Data di Nascita Paziente	DATE	
PAZ_CCOD	Codice Paziente	VARCHAR2(16 Byte)	
PAZ_IRES	Residenza Paziente	VARCHAR2(210 Byte)	
PAZ_CAPR	Cap Residenza	VARCHAR2(5 Byte)	
PAZ_IDOM	Domicilio Paziente	VARCHAR2(210 Byte)	
PAZ_CAPD	Cap Domicilio	VARCHAR2(5 Byte)	

Figura 4.9: Tabella di Anagrafica dei pazienti.

In questo caso la chiave primaria è rappresentata dal *codice paziente*. L'anonimità sarà gestita successivamente, durante le procedure ETL per la pulizia e trasformazione dei dati. Si procede con le tabelle successive:

PRPRERIC			
Nome Campo	Descrizione Campo	Tipo	
RIC_PAZ_CODE	Codice Paziente	VARCHAR2(16 Byte)	
RIC_CODE	Codice Richiesta	VARCHAR2(12 Byte)	
RIC_TIT_CODE	Codice Tariffa	VARCHAR2(8 Byte)	
RIC_PSOCC	Eventuale Codice di PS	VARCHAR2(14 Byte)	
RIC_TIR_CODE	Tipo Richiesta	VARCHAR2(8 Byte)	
RIC_ATDATE	Data Prenotazione	DATE	
RIC_USL_CODI	Codice ASL Richiesta	VARCHAR2(8 Byte)	
RIC_USL_CODI_ASS	Codice ASL Assistito	VARCHAR2(8 Byte)	
RIC_COM_CRES	Comune Residenza assistito	VARCHAR2(8 Byte)	

Figura 4.10: Tabella delle Richieste.

PRPREAPP			
Nome Campo	Descrizione Campo	Tipo	
APP_CODE	Codice Appuntamento	VARCHAR2(12 Byte)	
APP_RIC_CODE	Codice Richiesta	VARCHAR2(12 Byte)	
APP_PRZ_CODE	Mnemonico Prestazione	VARCHAR2(8 Byte)	
APP_DATA	Data dell'appuntamento	DATE	
APP_ORAP	Ora prevista in minuti	Number(4)	
APP_EXTI	Appuntamenti per Ext e Int	VARCHAR2(1 Byte)	
APP_STAT	Stato finale dell'app.	VARCHAR2(1 Byte)	
APP_UNI_SEQ	Codice dell'unità	Number(6)	
APP_FIT_CODE	Raggruppamento Convenz.	VARCHAR(8 Byte)	
APP_STAA	Stato Iniziale App.	VARCHAR2(1 Byte)	

Figura 4.11: Tabella Appuntamenti.

PRANAPRZ			
Nome Campo	Descrizione Campo	Tipo	
PRZ_CODE	Mnemonico Prestazione	VARCHAR2(8 Byte)	
PRZ_DESB	Descrizione Breve Prestaz.	VARCHAR2(45 Byte)	
PRZ_BRA_MINI	Branca Ministeriale Nom.Naz	VARCHAR2(8 Byte)	
PRZ_DESL	Descrizione Ricerche Text	VARCHAR2(500 Byte)	
PRZ_MINI	Codice Paziente	VARCHAR2(16 Byte)	
PRZ_NTR_CODE	Codice del Catalogo Prestaz.	VARCHAR2(10 Byte)	

Figura 4.12: Tabella Prestazioni.

PRANANTR		
Nome Campo	Descrizione Campo	Tipo
NTR_CODE	Codice del Catalogo	VARCHAR2(8 Byte)
NTR_DATATARIFFA	Codice Richiesta	DATE
NTR_REG_CODI	Mnemonico Prestazione	VARCHAR2(10 Byte)

Figura 4.13: Tabella Catalogo Ministeriale Prestazioni.

PRANAPRR			
Nome Campo	Descrizione Campo	Tipo	
PRR_MINI	Codice Ministeriale Prestaz.	VARCHAR2(8 Byte)	
PRR_REG_CODI (PK & FK)	Codice Regionale Prestaz.	VARCHAR2(8 Byte)	
PRR_DESC	Descrizione Prestazione	VARCHAR2(250 Byte)	
PRR_COSTOE	Costo Prestazione	NUMBER(8,2)	
PRR_DATATARIFFA	Data Aggiornamento Tariffa	DATE	

Figura 4.14: Tabella Prestazioni Regionali.

PRANAICD			
Nome Campo	Descrizione Campo	Tipo	
ICD_CODI	Codice Diagnosi	VARCHAR2(8 Byte)	
ICD_TIPO	Tipo di Codifica	VARCHAR2(2 Byte)	
ICD_DESC	Descrizione Diagnosi	VARCHAR2(200 Byte)	

Figura 4.15: Tabella codifiche diagnosi.

PREQUPRS		
Nome Campo	Descrizione Campo	Tipo
PRS_PRS_SEQ (PK & FK)	Serial del Medico	Number(6)
PRS_EQU_SEQ (PK & FK)	Serial dell'equipe	Number(6)

Figura 4.16: Tabella di Link tra equipe e personale medico.

PRANASTA			
Nome Campo	Descrizione Campo	Tipo	
STA_CODE	Codice dello stato	VARCHAR2(1 Byte)	
STA_TIPO	Tipo dello stato	VARCHAR2(1 Byte)	
STA_DESC	Descrizione dello stato	VARCHAR2(40 Byte)	

Figura 4.17: Tabella di descrizione degli stati appuntamento.

PRANAOPE			
Nome Campo	Descrizione Campo	Tipo	
OPE_CODI	Codice unità	VARCHAR2(8 Byte)	
OPE_PRS_CODI	Codice del Medico	VARCHAR2(16 Byte)	

Figura 4.18: Tabella con codice medici e unità.

PREQUUNI		
Nome Campo	Descrizione Campo	Tipo
UNI_UNI_SEQ (PK & FK)	Codice Equipe	Number(6)
UNI_EQU_SEQ	Codice Unità	Number(6)

Figura 4.19: Tabella di Link tra equipe medica unità.

4.4.3 AREAS

La gestione clinica si avvale di un applicazione web denominata AREAS. Essa è utilizzata per gestire il flusso dei ricoveri. L'entità centrale su cui si regge l'intero database è ricoveri, che fa riferimento alla tabella AALRICOVERO dello

PRRICREF		
Nome Campo	Descrizione Campo	Tipo
REF_CODE	Codice Referto	VARCHAR2(16 Byte)
REF_PRS_CODI	Codice Medico	VARCHAR2(16 Byte)

Figura 4.20: Tabella Link tra medico e codice referto.

PRAPPREF		
Nome Campo	Descrizione Campo	Tipo
REF_CODE (PK & FK)	Codice Referto	VARCHAR2(16 Byte)
APP_CODE	Codice Appuntamento	VARCHAR2(12 Byte)

Figura 4.21: Tabella Link tra appuntamenti e referti.

PRANAUNI		
Nome Campo	Descrizione Campo	Tipo
UNI_SEQ	Codice unità	Number(6)
UNI_REP_SEQ	Codice raggruppamento Un.	Number(5)
UNI_MNEM	Mnemonico Prestazione	VARCHAR2(16 Byte)
UNI_DIS_CODI	Disciplina unità(branca)	VARCHAR2(8 Byte)
UNI_ATTI	Flag unità(Attiva/Non Attiva)	CHAR(1 Byte)
UNI_CODE	Identificativo Dipartimento	VARCHAR2(32 Byte)

Figura 4.22: Tabella unità/agenda.

schema relazionale IRIDE. Essa contiene tutte le informazioni relative alla tipologia del ricovero (programmato o ordinario, Day Hospital medico oncologico ed interventistico, Day Hospital chirurgico).



Figura 4.23: Applicativo Web per gestione ADT.

A questa entità si aggiunge anche quella dei pre-ricoveri, in cui il paziente prima di essere sottoposto ad un intervento chirurgico, può eseguire gli esami preliminari necessari, per questo è convocato per una giornata in pre-ricovero.

La piattaforma web AREAS figura B.4, consente inoltre un collegamento con le piattaforme per la gestione dei pazienti in Pronto Soccorso(PS-WEB) e degli interventi chirurgici(SOWEB). Quest'ultimi fanno riferimento rispettivamente agli schemi relazionali IRIDE e SOWDATI.

4.4.4 Schema relazionale IRIDE

L'ispezione delle tabelle si è rivelata più complessa rispetto al lavoro eseguito per il CUP, questo perché la gestione clinica di AREAS fa riferimento ad uno schema relazionale che conta 2052 tabelle. Dopo l'ispezione del database 15 tabelle sono state selezionate e sono riportate di seguito. Per ciascuna di esse, oltre alla codifica della stessa è riportata la corrispettiva descrizione.

Nome Relazione	Descrizione Relazione
AA_LRICOVERO	Ricoveri
AA_REGIMERICOVERO	Regime Ricovero
SA_STRUTTURA	Descrizione Reparti
AA_LDIAGNOSI	Descrizione Diagnosi
PRPRERIC	Richieste
AA_LINTERVENTO	Interventi
PS_SCHEDA	Scheda Pronto Soccorso
SA_PAZIENTE	Anagrafica Paziente
SA_DIAGNOSI	Descrizione Diagnosi
SA_INTERVENTO	Descrizione Interventi
AA_MODALITADIM	Modalità Dimissione
AA_LTRASFOSPITANTE	Trasferimento Ospitanti Ricovero
PRANASTA	Descrizione Stati
PRPREAPP	Appuntamenti
AL_LPRENOTAZIONE	Lista di Attesa

Figura 4.24: Schema relazionale IRIDE.

AA_LRICOVERO		
Nome Campo	Descrizione Campo	Tipo
RC_PAZ_CPAZ	Codice Paziente	VARCHAR2(16 Byte)
RC_CODNOSO	Codice Ricovero	VARCHAR2(16 Byte)
RC_DT_RIC	Data Ricovero	DATE
RC_DT_DIM	Data Dimissioni	DATE
RC_REP_RIC_DIM	Codice Reparto Ricovero	VARCHAR2(11 Byte)
RC_DIAGNOSI_ACCE	Diagnosi Text in entrata	VARCHAR2(4000 Byte)
RC_REGIME	Regime di Ricovero(Ord/DH)	VARCHAR2(1 Byte)
RC_PSO_PRAC	Codice Pronto Socc.	VARCHAR2(11 Byte)
RC_TIPO_RIC_ID	Codice Tipo Ricovero	VARCHAR2(11 Byte)
RC_MOTI_RIC_ID	Codice Motivo Ricovero	VARCHAR2(11 Byte)
RC_MODA_DIM_ID	Codice Modalità Dimissione	VARCHAR2(11 Byte)
RC_OPE_CERT_ID	Codice Medico Refertante	VARCHAR2(16 Byte)
RC_DT_INIZIO_PRERIC	Data Inizio Pre-Ricovero	DATE
RC_DT_FINE_PRERIC	Data Fine Pre-Ricovero	DATE
RC_PSD_ID(PK & FK)	Presidio Osp. Ricovero	VARCHAR2(11 Byte)
RC_REP_OSP_ID	Reparto Ospitante	VARCHAR2(11 Byte)
RC_STATOSK_ID	Stato scheda(DIM/REP/VAL)	VARCHAR2(11 Byte)
RC_PSO_IDSC	Scheda Pronto Socc.	VARCHAR2(14 Byte)
RC_ONERE_ID	Onere della Degenza	NUMBER(11)
RC_DRG_ID	DRG	VARCHAR2(11 Byte)
RC_IMPORTO	Costo Ricovero DRG	NUMBER(11,2)
RC_PROP_RIC_ID	Proposta Ricovero	VARCHAR2(11 Byte)

Figura 4.25: Tabella dei Ricoveri.

AA_REGIMERICOVERO		
Nome Campo Descrizione Campo Tipo		
RR_ID	ID univoca	VARCHAR2(11 Byte)
RC_COD	Codice Ricovero	VARCHAR2(11 Byte)
RR_F_CATEGORIA	Regime Ricovero	VARCHAR2(11 Byte)
RR_DESC	Desc. Regime Ricovero	VARCHAR2(256 Byte)

Figura 4.26: Tabella regime di ricovero.

SA_STRUTTURA		
Nome Campo	Descrizione Campo	Tipo
SR_ID	Codice Interno	VARCHAR2(11 Byte)
SR_CODICE	Codice Mnemonico	VARCHAR2(30 Byte)
SR_DESC	Descrizione	VARCHAR2(255 Byte)
SR_AZ	Codice Azienda	VARCHAR2(2 Byte)

Figura 4.27: Tabella di descrizione reparti.

AA_LDIAGNOSI		
Nome Campo	Descrizione Campo	Tipo
DD_CODNOSO	Codice Ricovero	VARCHAR2(16 Byte)
DD_DIA_ID	Codice Diagnosi	VARCHAR2(11 Byte)
DD_REP_ID	Codice Reparto	VARCHAR2(11 Byte)
DD_PSD_ID	Codice Presidio Ricovero	VARCHAR2(11 Byte)
DD_PROG	Progressivo Diagnosi	VARCHAR2(8 Byte)
DD_ORDINE	Diagnosi Dimissione	NUMBER(2)
DD_FLAG_PRINC	Flag Dimissione	VARCHAR2(1 Byte)

Figura 4.28: Tabella Diagnosi.

PRPRERIC		
Nome Campo	Descrizione Campo	Tipo
RIC_PAZ_CODE	Codice Paziente	VARCHAR2(16 Byte)
RIC_DRIC	Data Richiesta Ricovero	DATE
RIC_CODE	Codice Richiesta Ricovero	VARCHAR2(12 Byte)
RIC_PSOCC	Eventuale Codice PS	NUMBER(11)
RIC_PSNUMEROSCHEDA	Numero scheda PS	VARCHAR2(16 Byte)
RIC_DPRENO	Data di Prenotazione	DATE
RIC_REP_CODI	Codice Reparto Ospitante	VARCHAR2(8 Byte)
RIC_POS	Codice POS Richiedente	VARCHAR2(11 Byte)
RIC_MED	Codice Medico Prescrivente	VARCHAR2(16 Byte)
RIC_INDATE	Data di Inoltro	DATE
RIC_ATDATE	Data prevista di erogazione	DATE
RIC_STATUS	Stato della Richiesta	VARCHAR2(1 Byte)

Figura 4.29: Tabella richieste prericoveri.

AA_LINTERVENTO		
Nome Campo	Descrizione Campo	Tipo
IN_CODNOSO	Codice Ricovero	VARCHAR2(16 Byte)
IN_DATA	Data Intervento	DATE
IN_INT_ID	ID Codice Intervento	VARCHAR2(11 Byte)
IN_PSD_ID	Codice Presidio Ricovero	VARCHAR2(11 Byte)
IN_REP_ID	Codice Reparto	VARCHAR2(11 Byte)
IN_REPARTO_RIC_ID	Codice Reparto Ricovero	VARCHAR2(11 Byte)
IN_ORDINE	Ordine di Importanza	NUMBER(3)
IN_FLAG_PRINC	Flag interventi	VARCHAR2(1 Byte)
IN_DATA_FINE	Data Fine Intervento	DATE
IN_ANESTETISTA	Codice Anestetista	VARCHAR2(16 Byte)
IN_CHIRURGO1/2/3	Codice Chirurgo	VARCHAR2(16 Byte)
IN_SALA_OPE	ID Sala Operatoria	VARCHAR2(11 Byte)

Figura 4.30: Tabella Interventi/Procedure.

PS_SCHEDA		
Nome Campo	Descrizione Campo	Tipo
SC_SCHEDA	Codice Scheda PS	VARCHAR2(11 Byte)
SC_NOSOLOGICO	Codice ricovero(se presente)	VARCHAR2(16 Byte)
SC_PAZIENTE	Codice Paziente	VARCHAR2(16 Byte)
SC_DT_INIZIO	Data Inizio Contatto PS	DATE
SC_DT_FINE	Data Fine Contatto PS	DATE
SC_STRUTTURA	Codice Struttura PS	VARCHAR2(11 Byte)
SC_AMBOX	Box PS	VARCHAR2(11 Byte)
SC_MODAINV_ID	Modalità	NUMBER(11)
SC_MODACC_ID	Modalità di accesso	NUMBER(11)
SC_ESITO_ID	Esito Accettazione	NUMBER(11)
SC_URG_OUT	Colore assegnato	NUMBER(6)
SC_DATA_PRESA_IN_CAR	Data Presa in carico	DATE
SC_URG_TRG	Colore assegnato al Triage	NUMBER(6)
SC_ID	Progressivo Scheda	NUMBER(11)

Figura 4.31: Tabella Pronto Soccorso.

SA_PAZIENTE			
Nome Campo	Descrizione Campo	Tipo	
PZ_ID	Codice Paziente	VARCHAR2(16 Byte)	
PZ_COGN	Cognome Pazinete	VARCHAR2(40 Byte)	
PZ_NOME	Nome Paziente	VARCHAR2(40 Byte)	
PZ_DT_NAS	Data di Nascita	DATE	
PZ_SESSO	Sesso Paziente	VARCHAR2(1 Byte)	
PZ_COM_NAS	Comune di Nascita	VARCHAR2(11 Byte)	
PZ_CFIS	Codice Fiscale	VARCHAR2(16 Byte)	
PZ_COM_RES	Comune di Residenza	VARCHAR2(11 Byte)	
PZ_CAP_RES	Cap di Residenza	VARCHAR2(5 Byte)	
PZ_COM_DOM	Comune Domicilio	VARCHAR2(40 Byte)	
PZ_CAP_DOM	Cap Domicilio	VARCHAR2(5 Byte)	

Figura 4.32: Tabella Anagrafica Paziente.

SA_DIAGNOSI			
Nome Campo	Descrizione Campo	Tipo	
DG_ID	Codice Diagnosi	VARCHAR2(11 Byte)	
DG_CODI	Codice Regionale Diagnosi	VARCHAR2(11 Byte)	
DG_DESC	Descrizione Diagnosi	VARCHAR2(250 Byte)	
DG_CATEGORIA	Codice Categoria	NUMBER(11)	
DG_RARA	Diangosi Rara	VARCHAR2(1 Byte)	

Figura 4.33: Tabella Descrizione Diagnosi.

SA_INTERVENTO		
Nome Campo	Descrizione Campo	Tipo
IN_ID	Codice Interno Interv.	VARCHAR2(11 Byte)
IN_CODI	Codice Esterno interv.	VARCHAR2(8 Byte)
IN_DESC	Descrizione Interv.	VARCHAR2(250 Byte)

Figura 4.34: Tabella Descrizione Interventi.

AA_MODALITADIM			
Nome Campo Descrizione Campo		Tipo	
MD_ID	ID Modalità	VARCHAR2(11 Byte)	
MD_CODI	Codice Modalità Dimissione	VARCHAR2(8 Byte)	
MD_REGI	Codice Regime di Ricovero	VARCHAR2(1 Byte)	
MD_DESC	Descrizione Mod. Dimissione	VARCHAR2(256 Byte)	

Figura 4.35: Tabella Modalità Dimissione.

AA_LTRASFOSPITANTE			
Nome Campo	Descrizione Campo	Tipo	
TO_PSD_ID	Codice Presidio	VARCHAR2(11 Byte)	
TO_CODNOSO	Codice Ricovero	VARCHAR2(16 Byte)	
TO_PROG	Progressivo Ricovero	NUMBER(3)	
TO_REP_TRA_ID	Codice Reparto(Trasferito)	VARCHAR2(11 Byte)	
TO_REP_RIC_ID	Codice Reparto(Precedente)	VARCHAR2(11 Byte)	
TO_DATA_TRAS	Data Trasferimento	DATE	

Figura 4.36: Tabella Trasferimento ospitanti ricovero.

PRANASTA		
Nome Campo	Descrizione Campo	Tipo
STA_CODE	Codice dello stato	VARCHAR2(1 Byte)
STA_TIPO	Tipo dello stato	VARCHAR2(1 Byte)
STA_DESC	Descrizione dello stato	VARCHAR2(40 Byte)

Figura 4.37: Tabella di descrizione Stati.

PRPREAPP		
Nome Campo	Descrizione Campo	Tipo
APP_CODE	Codice Appuntamento	VARCHAR2(12 Byte)
APP_POS_REQ	Codice Pos Richiedente	VARCHAR2(11 Byte)
APP_POS_ERO	Codice Pos Erogante	VARCHAR2(11 Byte)
APP_RIC_CODE	Codice della Richiesta	VARCHAR2(12 Byte)
APP_PRZ_CODE	Codice della Prestazione	VARCHAR2(15 Byte)
APP_BRA_CODE	Codice della Branca	VARCHAR2(15 Byte)
APP_MED_RIC	Codice Medico Richiedente	VARCHAR2(16 Byte)
APP_DATA	Data Appuntamento	DATE
APP_STAT	Stato Appuntamento	VARCHAR2(1 Byte)
APP_MED_CODE	Codice del Medico Erogante	VARCHAR2(16 Byte)

Figura 4.38: Tabella degli Appuntamenti.

AL_LPRENOTAZIONE		
Nome Campo	Descrizione Campo	Tipo
PR_SCHEDA_ID	Codice Appuntamento	VARCHAR2(14 Byte)
PR_REP_ID	Codice Int. reparto prenotato	VARCHAR2(11 Byte)
PR_REGIME	Regime di Ricovero Previsto	VARCHAR2(1 Byte)
PR_URG_ID	Codice Int. Urgenza	NUMBER(11)
PR_OPE_PROP	Codice Int. Medico Propon.	VARCHAR2(16 Byte)
PR_DT_PRENOTAZ	Data Prenotazione Lista Att.	DATE
PR_INT_ID	Categorizzazione Prest. Int.	VARCHAR2(11 Byte)
PR_DT_PRE_INIZIO	Data Inizio Prericovero	DATE
PR_DT_PRE_FINE	Data Fine Prericovero	DATE
PR_DT_CONV	Data di Convocazione	DATE
PR_STATOSK_ID	Stato Prenotazione	VARCHAR2(11 Byte)
PR_PAZ_CPAZ	Codice Paziente	VARCHAR2(16 Byte)
PR_DIA_SINTETICA_ID	Diagnosi interna per Reparto	VARCHAR2(11 Byte)
PR_REP_EQUIPE_CHIR	Codice Equipe	VARCHAR2(11 Byte)
PR_PSD_ID	Codice Interno Presidio	VARCHAR2(11 Byte)

Figura 4.39: Tabella Lista di Attesa.

4.4.5 **SOWEB**

SOWEB è la piattaforma web che permette di gestire le procedure degli interventi chirurgici. Lo schema relazionale a cui si fa riferimento è SOWDATI.

4.4.6 Schema relazionale SOWDATI

Dopo l'ispezione del database, 9 tabelle sono state selezionate e sono riportate di seguito. Per ciascuna di esse, oltre alla codifica della stessa è riportata la corrispettiva descrizione.

Nome Relazione	Descrizione Relazione
SO01_OPERAZIONE	Gestione Atto Operatorio
SO13_DIAGNOSI_OPERAZ	Diagnosi associate all'atto operatorio
SO30_OPERAZ_DETTAGLIO	Descrizione Atto Operatorio
SO02_INTERV_OPERAZ	Interventi/Procedure Associate all'Atto Op.
SO06_COMPON_EQUIPE	Equipe medica non associate all'Atto Op.
PG63_SALA	Anagrafiche Sala Operatorie
TB40_MEDICO	VISTA: Medici
TB39_TECNICO	VISTA: Tecnici
TB38_INFERMIERE	VISTA: Infermieri

Figura 4.40: Tabella schema relazionale SOWDATI.

SO01_OPERAZIONE			
Nome Campo	Descrizione Campo	Tipo	
CS01_NOSOLOGICO	Codice Ricovero	NUMBER	
CS01_PROGR_ANAG	Codice Paziente	NUMBER	
CS01_PROGR_OPERAZ	Progr. Atto Operatorio	NUMBER	
CS01_C_REPARTO	Reparto Ricovero Deg	VARCHAR2(11 Byte)	
CS01_DT_INIZIO_OPERAZ	Data ora Inizio Ope.	DATE	
CS01_DT_FINE_OPERAZ	Data ora Fine Ope.	DATE	
CS01_F_ESAME_ISTOLOGICO	Flag Esame Istologico	VARCHAR2(1 Byte)	
CS01_F_ESAME_ISTOLOGICO_ESTEMP	Flag E. I. Estempor.	VARCHAR2(1 Byte)	
CS01_F_SCOPIA	Flag Scopia	VARCHAR2(1 Byte)	
CS01_C_SALA	Codice Sala Ope.	VARCHAR2(8 Byte)	
CS01_F_OPERAZ_CHIUSA	Stato Atto Operatorio	VARCHAR2(1 Byte)	
CS01_F_OPERAZ_ANNULLATA	Stato Atto Operatorio	VARCHAR2(1 Byte)	
CS01_C_MEDICO	Cod. Primo Operatore	VARCHAR2(30 Byte)	
CS01_C_TIPOLOGIA_OPERAZIONE	Tipologia Operazione	VARCHAR2(2 Byte)	
CS01_C_REPARTO_CHIR	Cod. Reparto Chirurg.	VARCHAR2(11 Byte)	
CS01_REGISTRO_REP	Registro Ope.Chirurg.	NUMBER	
CS01_TECNICA_CHIRURGICA	Tecnica Chirurgica	VARCHAR2(2 Byte)	
CS01_TECNICA_ANEST	Tecnica Anestesiolog.	VARCHAR2(2 Byte)	
CS01_DURATA_SCOPIA	Durata Scopia	NUMBER(4)	
CS01_F_GRAFIA	Grafia	NUMBER(4)	
CS01_DURATA_GRAFIA	Durata Grafia	NUMBER(4)	
CS01_DURATA_ESP_RAD_IONI	Dur. Esposizione Rad.	NUMBER(4)	
CS01_F_ASSIST_ANEST	Assistente Anestetista	VARCHAR2(1 Byte)	

Figura 4.41: Tabella gestione atto operatorio.

SO13_DIAGNOSI_OPERAZ		
Nome Campo	Descrizione Campo	Tipo
CS13_PROGR_OPERAZ(PK & FK)	Progressivo Atto Op.	NUMBER
CS13_N_DIAGNOSI	Progressivo Diagnosi	NUMBER
CS13_C_DIAGNOSI	ID Diagnosi	VARCHAR2(11 Byte)
CS13_F_ANNUL_DIAGNOSI	Flag Annullamento Dia	VARCHAR2(1 Byte)

Figura 4.42: Tabella Diagnosi associate all'atto operatorio.

SO30_OPERAZ_DETTAGLIO		
Nome Campo	Descrizione Campo	Tipo
CS30_PROGR_OPERAZ(PK & FK)	Progressivo Atto Op.	NUMBER(2)
CS30_NUM_EQUIPE	Progr. Equipe Op.	NUMBER
CS30_DESC_ATTO_OPERAT	Descrizione Text Op.	VARCHAR2(4000 Byte)
CS30_DESC_ATTO_OPERAT_2	Estensione Desc. Op.	VARCHAR2(4000 Byte)

Figura 4.43: Tabella Descrizione atto operatorio.

SO02_INTERV_OPERAZ			
Nome Campo	Descrizione Campo	Tipo	
CS02_PROG_OPERAZ(PK & FK)	Progressivo Atto Op.	NUMBER(2)	
CS02_N_INTERVENTO	Progr. Interventi	NUMBER	
CS02_C_INTERVENTO	Codici Interventi Effett.	VARCHAR2(11 Byte)	
CS02_F_ANNULL_INTERV	Flag Annullamento Int.	VARCHAR2(1 Byte)	
CS02_NUM_EQUIPE	Prog. Equipe Interv	NUMBER(2)	

Figura 4.44: Tabella Interventi/Procedure Associate all'atto operatorio.

SO06_COMPON_EQUIPE			
Nome Campo	Descrizione Campo	Tipo	
CS06_C_EQUIPE	Codice compon. i-equ.	VARCHAR2(30 Byte)	
CS06_C_TIPO_PERSONALE	Tipologia Personale	VARCHAR2(3 Byte)	
CS06_NUM_EQUIPE	Equipe di Appartenenz	NUMBER(2)	
CS06_PROGR_OPERAZ	Progr. Atto Operatorio.	NUMBER	
CS06_C_RUOLO_SO	Ruolo in Sala Op.	VARCHAR2(8 Byte)	

Figura 4.45: Tabella Equipe Medica.

PG63_SALA			
Nome Campo	Descrizione Campo	Tipo VARCHAR2(11 Byte)	
CG63_ID_SALA	ID Sala Operatoria		
CG63_D_SALA	Descrizione Sala Op.	VARCHAR2(100 Byte)	
CG63_C_SALA	Codice Sala Op.	VARCHAR2(8 Byte)	

Figura 4.46: Tabella Anagrafica Sala Operatoria.

TB40_MEDICO (vista)			
Nome Campo	Descrizione Campo	Tipo	
C40_C_MEDICO	Codice Medico	VARCHAR2(44 Byte)	
C40_C_NOME_MEDICO	Nome Medico	VARCHAR2(127 Byte)	

Figura 4.47: Vista Medici.

TB39_TECNICO (vista)			
Nome Campo	Descrizione Campo	Tipo	
C_USER	Codice Tecnico	VARCHAR2(44 Byte)	
C39_NOME_TECNICO	Nome Tecnico	VARCHAR2(127 Byte)	

Figura 4.48: Vista Tecnici.

TB38_INFERMIERE (vista)			
Nome Campo	Descrizione Campo	Tipo	
C_USER	Codice Infermiere	VARCHAR2(44 Byte)	
C39_NOME_INFERMIERE	Nome Infermiere	VARCHAR2(127 Byte)	

Figura 4.49: Vista Infermieri.

4.4.7 wHealth



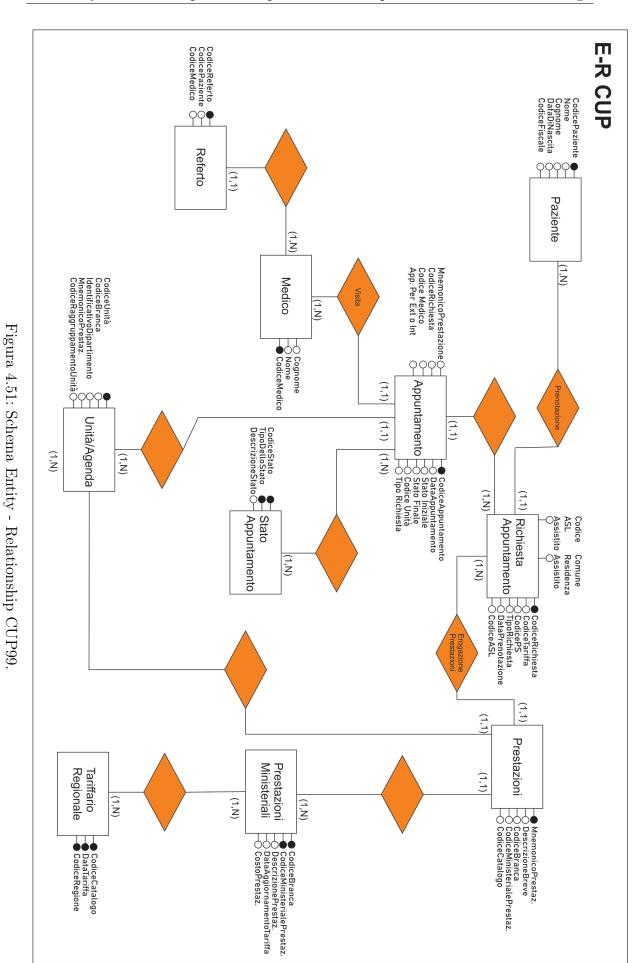
Figura 4.50: Logo Applicazione web wHealth per refertazione ambulatoriale.

Verso la metà del 2019 è stato introdotto un nuovo sistema di refertazione ambulatoriale, il quale ha fornito un importante svolta per quanto riguarda la completezza dei dati clinici del paziente. La possibilità di creare dei moduli di refertazione informatizzati, pensati per ogni area clinica specialistica ha consentito una strutturazione ed organizzazione del dato clinico, subito utilizzabile per la piattaforma di data warehouse. L'obiettivo della tesi è rappresanto dalla progettazione del DW in ambito oncologico, ma le potenzialità estese di questo applicativo consentiranno, in futuro, l'estensione del data warehouse per altri studi clinici quali cardiologici, ortopedici, neurologici ecc.

L'integrazione di wHealth con i restanti applicativi dell'ospedale si ha grazie alla comunicazione bidirezionale con il software SGP, grazie al codice paziente, e tutte le tabelle di frontiera relative alle visite ambulatoriali. Tutti i moduli informatizzati legati al processo CAS sono utilizzabili tramite wHealth.

4.5 Entity - Relationship

La realizzazione degli schemi E- R è stato il punto di partenza per il raggiungimento dell'integrazione tra le diverse basi dati presenti nel sistema ospedaliero. Il modello di ciascun database mappato è di tipo relazionale, ad eccezione di CUP99, che, per motivi associati alla storicità di questo database, rappresenta una versione ibrida tra un modello relazionale e un modello di tipo gerarchico .



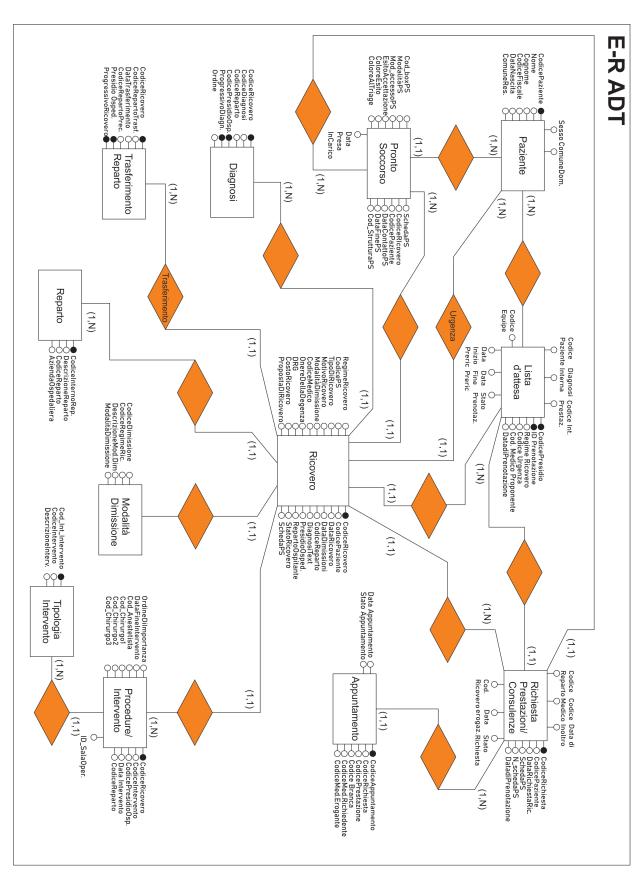


Figura 4.52: Schema Entity - Relationship IRIDE.

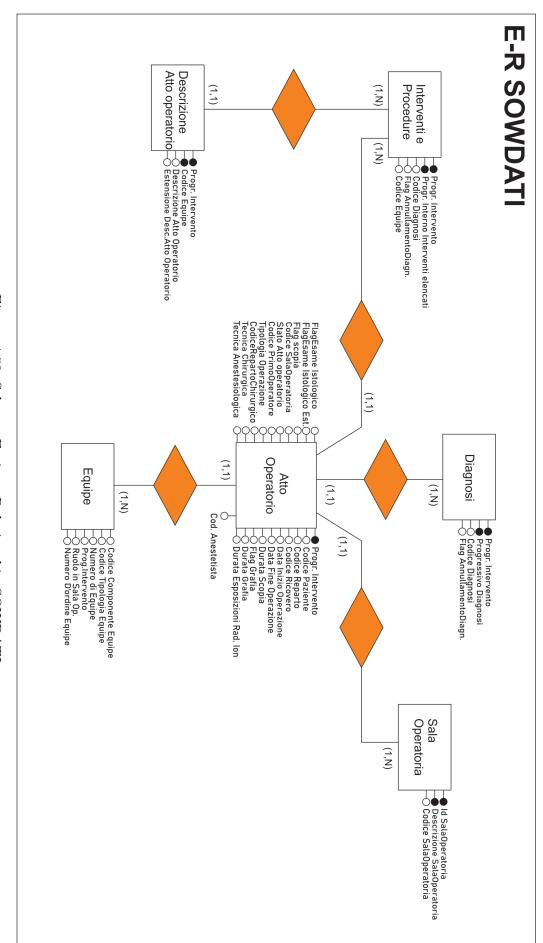


Figura 4.53: Schema Entity - Relationship SOWDATI

4.6 Analisi dei requisiti

Il raggiungimento degli obiettivi preposti è influenzato dalle informazioni che il progettista raccoglie durante la fase di analisi dei requisiti, più sono i dettagli raccolti e maggiore sarà la granularità del data warehouse che si vuole progettare. Questa proprietà deve essere ben manipolata, in quanto avere un livello di dettaglio sui dati ottimale consente di estrarre più informazioni, ma inficerebbe sulle prestazioni. La fase di analisi dei requisiti segue la metodologia Tropos descritta nel 2.4.4.

4.6.1 Diagramma degli attori

La panoramica degli attori coinvolti in un sistema ospedaliero è ottenuta grazie al diagramma degli attori e racchiude le attività principali che vengono svolte in ospedale. Gli attori principali che fanno parte del sistema ospedaliero sono: OSPEDALE, CUP, PAZIENTE, UFFICIO RICOVERI, MEDICI, AMBULATORI, REPARTI, PRONTO SOCCORSO, LABORATORI, TECNICI, SALE OPERATORIE e DIREZIONE OSPEDALIERA.

Il diagramma degli attori è riportato in figura 4.57.

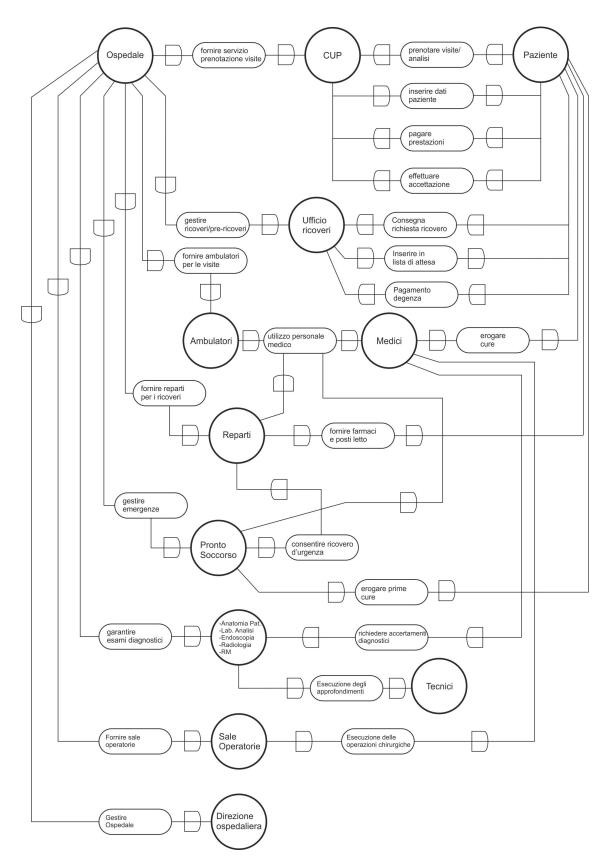


Figura 4.54: Diagramma degli attori del sistema ospedaliero Humanitas.

4.7 Progettazione concettuale

La costruzione dell'albero degli attributi costituisce il primo passo per la definizione del fatto di interesse. La progettazione concettuale rappresenta la fase di transizione tra il data warehouse e i database, in quanto consente di elidere e sintetizzare le informazioni, considerando esclusivamente quelle di maggiore interesse clinico, selezionando gli attributi rilevanti e modificando le relazioni tra essi. Si cominciano a delineare anche le dimensioni, ovvero i filtri necessari per raggruppare e analizzare i dati raccolti nel data warehouse.

Nello schema riportato in 4.55 sono individuate le dimensioni, e i rispettivi attributi, raggruppati in gerarchie. Il *Codice Paziente* permette l'integrazione con tutte le basi dati considerate, e rappresenterà nel *Dimenosional Fact Model* il *fatto*, a cui saranno legate le diverse *dimensioni*. Si anticipa, inoltre, che il collegamento tra alcune dimensioni non è diretto al Codice Paziente, e questo sarà il motivo per cui nella progettazione logica si ottiene un'estensione dello schema a stella.

L'area indicata in giallo, rappresenta la fonte dati di wHealth con i dati clinici dei pazienti. Il codice del modulo CAS è univoco per ogni compilazione associata ad un determinato paziente. La compilazione dei moduli CAS infermieristici consentono di caratterizzare l'aspetto psico-fisico del paziente. I moduli sono elencati nel capitolo 3.

La fonte dati wHealth include anche le informazioni del percorso di cura deciso dal GIC. I moduli informatizzati usati per la refertazione contengono i campi di anamnesi, diagnosi e la terapia personalizzata per il paziente affetto da patologie tumorali. Per comodità di rappresentazione, l'albero degli attributi, raffigurato in figura 4.55 non contiene gli attributi legati ai moduli di riferitmento. Di seguito vengono illustrati nel dettaglio i campi contenuti in ciascun modulo CAS.

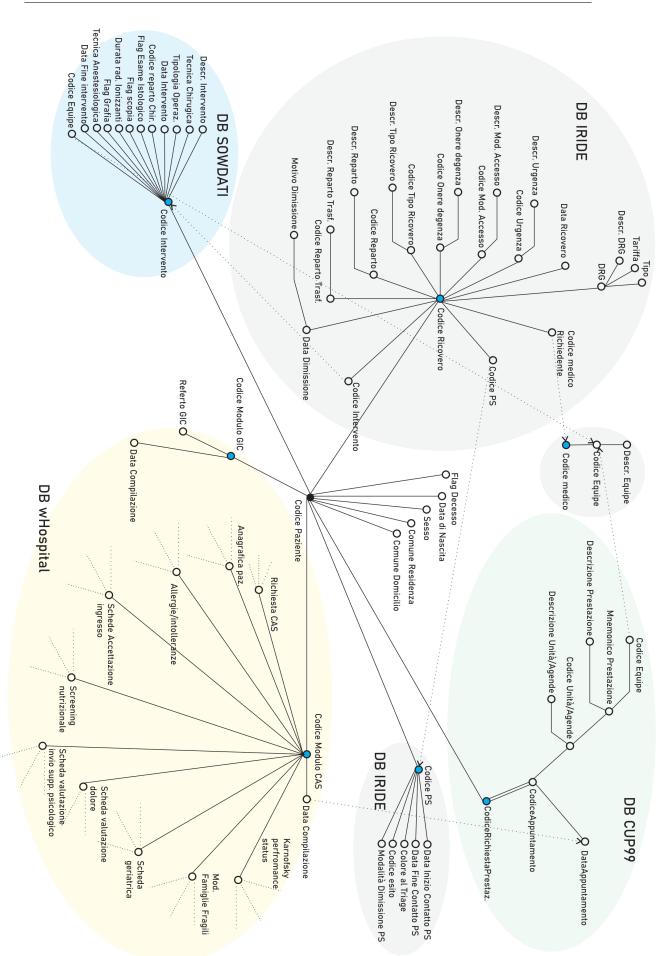


Figura 4.55: Albero degli attributi.

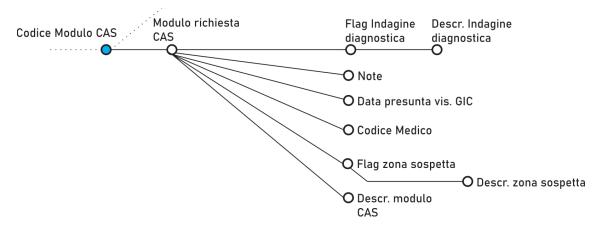


Figura 4.56: Modulo di richiesta CAS.

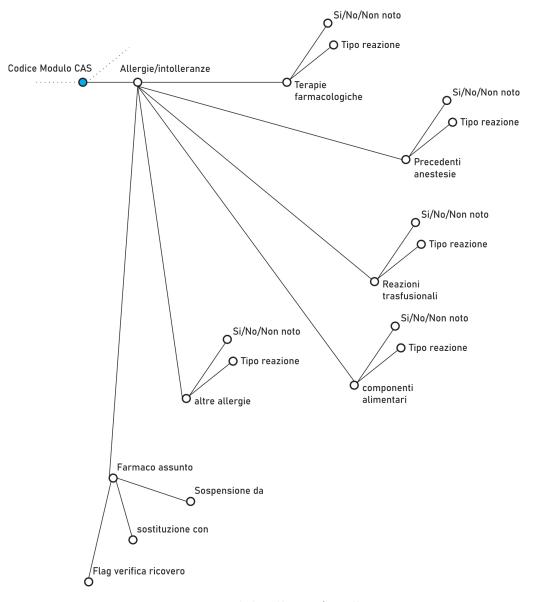


Figura 4.57: Modulo allergie/intolleranze

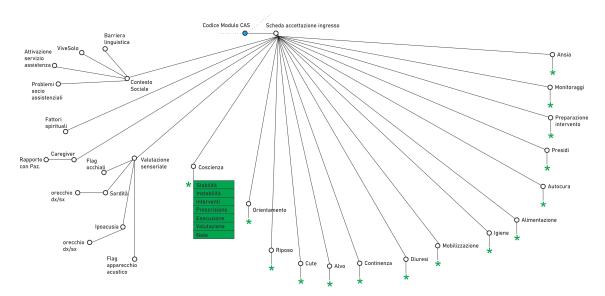


Figura 4.59: Modulo Accettazione.

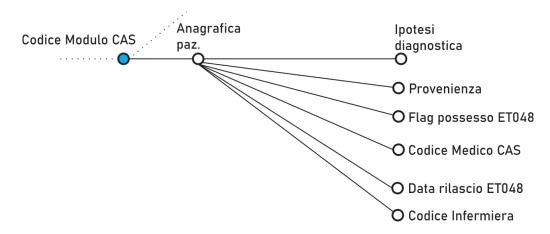


Figura 4.58: Modulo anagrafica paziente.

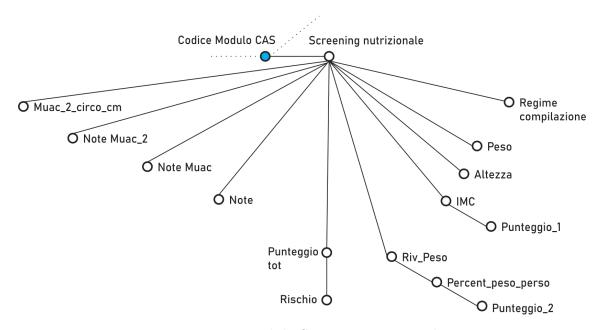


Figura 4.60: Modulo Sceening nutrizionale.

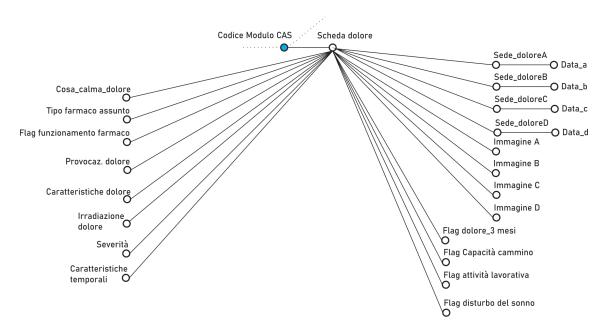


Figura 4.61: Modulo Valutazione del dolore.

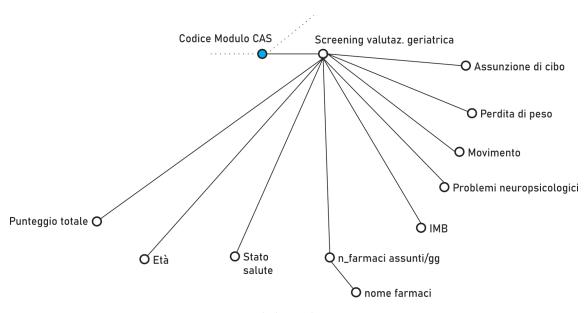


Figura 4.62: Modulo Valutazione geriatrica

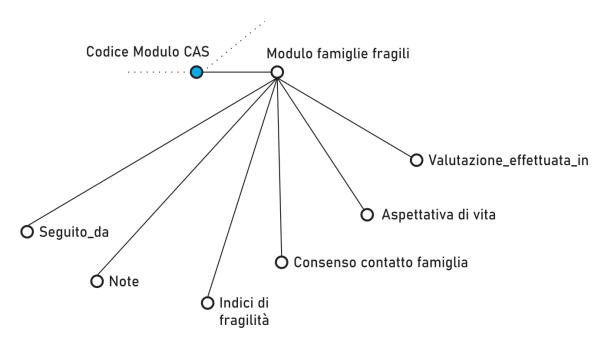


Figura 4.63: Modulo Protezione famiglie fragili

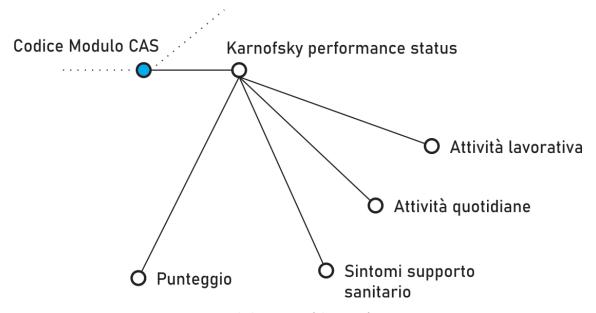


Figura 4.64: Modulo Karnofsky performance status

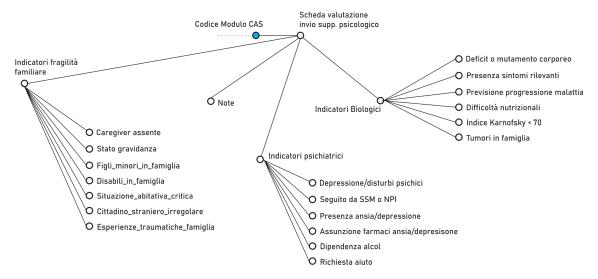


Figura 4.65: Modulo Sintesi valutazione psicologica

4.7.1 Dimensional Fact Model

Successivamente alla costruzione ed eventuale potatura dell'albero degli attributi, l'ulteriore passo per fornire una documentazione chiara ed esaustiva agli utenti finali che utilizzeranno il DW per scopi di ricerca clinica, viene costruito il Dimensional Fact Model, un diagramma che ha l'obiettivo di creare un ambiente in cui le query degli utenti possano essere formulate in modo intuitivo, e rendere possibile la comunicazione tra progettisti e utenti finali per formalizzare le specifiche dei requisiti. Il Dimensional Fact Model del DW oncologico è riportato in 4.66

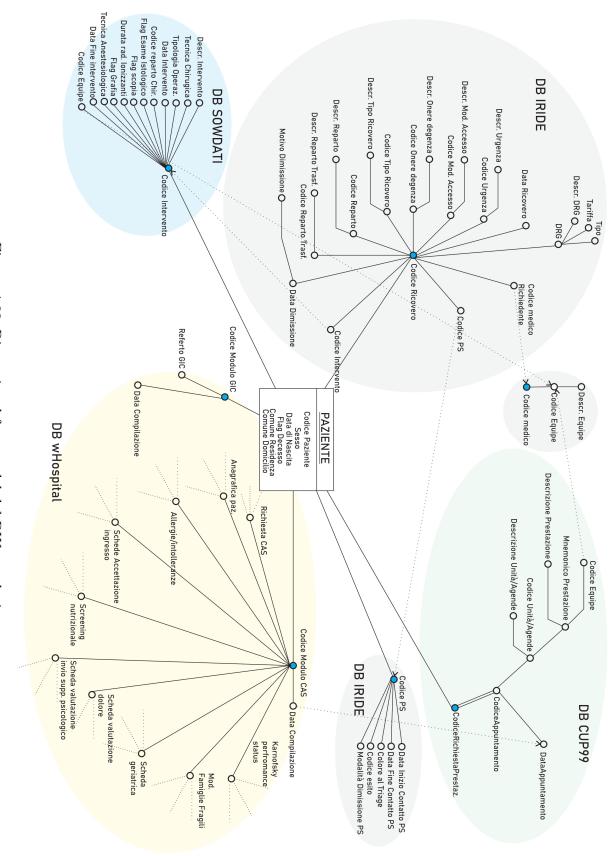


Figura 4.66: Dimensional fact model del DW oncologico

Capitolo 5

Risultati e conclusioni

In quest'ultimo capitolo conclusivo si mostrano i risultati della progettazione del DW clinico per ricerca oncologica. Il risultato ottenuto, successivo alle fasi seguite nel corso del Capitolo 4, è lo schema a fiocco di neve, chiamato anche snowflake. La progettazione logica rappresenta il passo che precede la fase di alimentazione del DW, la quale include le procedure ETL per la pulizia dei dati e il caricamento sul sistema fisico scelto.

5.1 Progettazione logica del DW clinico

Uno dei principali utilizzi del data warehouse clinico è selezionare una coorte di pazienti per studiare quanto sono comuni le malattie, le loro cause e le loro prognosi. Le coorti vengono selezionate sulla base di determinate caratteristiche condivise, ad esempio sulla base di un fattore di rischio, legata ad una possibile causa di una patologia. Inoltre, si cerca di aumentare la conoscenza sulle diagnosi e terapie per creare un'associazione statistica tra una terapia per una malattia, e un beneficio. Queste informazioni sono rese disponibili solo nel momento in cui sono state raccolte sufficienti informazioni relative ai pazienti che hanno usufruito delle cure presso l'ospedale.

La visione centrale del paziente, legata a tutte le dimensioni di interesse è la soluzione più adatta per un sistema di data warehousing. Questa scelta, garantisce all'utente finale di eseguire delle query con grande facilità, selezionando un certo gruppo di pazienti, con delle determinate caratteristiche psico – fisiche o che seguono una determinata terapia, considerando le chiavi primarie di ciascuna dimensione, presenti anche all'interno della tabella fatto *paziente*, e aggiungendo tutti gli attributi necessari per la ricerca clinica che si vuole effettuare.

Lo schema proposto in 5.1, è il risultato della progettazione logica. Si propone uno schema snowflake, un'estensione dello schema a stella, che in più presenta dei legami con altre dimensioni, costruendo in questo modo relazioni chiamati *archi*.

La scelta progettuale è motivata da molteplici vantaggi:

• se una gerarchia è comune a più dimensioni, è possibile utilizzare la medesima dimension table secondaria all'interno di più gerarchie;

- aumenta la leggibilità dello schema, evidenziando le dipendenze funzionali tra gli attributi, sebbene queste siano maggiormente esplicite nella progettazione concettuale;
- consente di ottimizzare l'aggregazione dei dati e ridurre le ridondanze di informazioni presenti nelle dimension table.

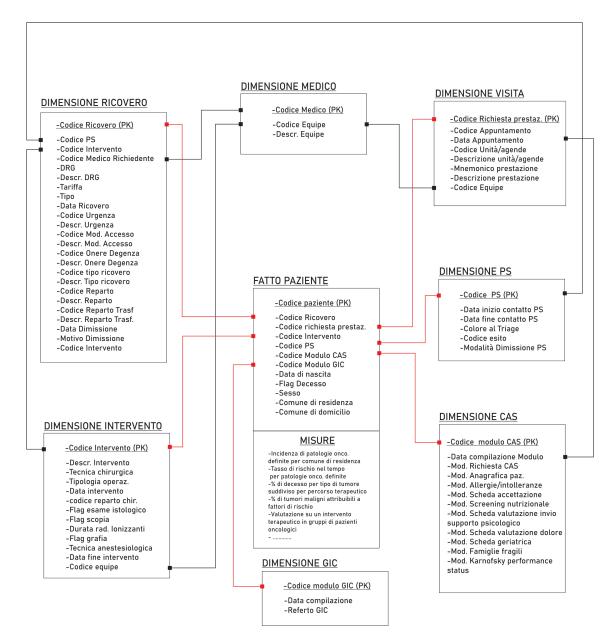


Figura 5.1: Progettazione logica: schema snowflake.

Per il fatto paziente sono state individuate le seguenti dimensioni:

- 1. Ricovero
- 2. Medico
- 3. Visita

- 4. Pronto Soccorso
- 5. CAS
- 6. GIC
- 7. Intervento

Le scelta delle misure prende ispirazione da EUPATI in [8] (European Patients' Academy for Therapeutic Innovation, EUPATI), un progetto che mette a disposizione dei protocolli per lo sviluppo degli studi clinici: vi è una suddivisione degli studi clinici in studi epidemiologici e studi trasversali. I primi, si dividono a loro volta in:

- *studi osservazionali*: verificano la frequenza delle malattie e le relazioni tra lo stato di salute e le sue variabili;
- studi sperimentali: prevedono un intervento diretto con l'applicazione di protocolli clinici o preventivi per cambiare un determinante di malattia;

Gli studi trasversali rappresentano un'indagine di intere popolazioni in un istante di tempo, per la valutazione della frequenza e distribuzione delle variabili, quali ad esempio i fattori di rischio.

Le misure scelte sono:

- 1. incidenza di patologie oncologiche definite per comune di residenza;
- 2. tasso di rischio nel tempo per patologie oncologiche definite;
- 3. % di decesso per tipo di tumore suddiviso per percorso terapeutico;
- 4. % di tumori maligni attribuibili a fattori di rischio
- 5. valutazione su un intervento terapeutico in gruppi di pazienti oncologici.

5.2 Conclusioni

Le integrazioni tra le basi dati esistenti nell'ospedale Humanitas - Gradenigo, e la costruzione di un modello che collocasse al centro dei collegamenti, il paziente, sono state le sfide principali della tesi. La possibilità di estrarre delle informazioni ben costruite e definite dalle misure rappresenta la potenzialità dello strumento per gli utenti a cui è stato pensato: ricercatori, medici e professionisti nel settore sanitario. Il DW progettato è delimitato all'area oncologica, ma non si esclude la possibilità di estendere il progetto alle diverse aree cliniche, mantenendo sempre il livello di dettaglio sui dati archiviati. Il metodo migliore per l'ampliamento del data warehouse dipende strettamente dall'architettura scelta nel progetto: l'introduzione dei data mart consente infatti, di creare delle aggregazioni specifiche di dati e che possono riguardare le diverse aree cliniche. Si potrebbe pensare di introdurre data mart sulla chirurgia, cardiologia, urologia e così via, incrementando le potenzialità dello strumento di supporto per la ricerca clinica. Sebbene la progettazione concettuale e logica siano le fondamenta dello strumento data warehouse, è necessario studiare gli strumenti ETL per la pulizia e validazione dei dati che poi andranno ad alimentare il data warehouse. La pulizia dei dati non è stata argomentata in questa tesi, ma è importante riconoscere che le informazioni cliniche estrapolate saranno integrate con tutti i dati necessari, saranno più affidabili e ancora più precise.

Il progetto vede, quindi, possibilità di sviluppo a lungo termine, ma rappresenterà lo strumento informativo più importante per garantire il giusto supporto per la prevenzione di eventuali patologie, o di selezione tra i diversi percorsi diagnostici e terapeutici.

Appendice A

Modellare un processo

Un processo è una serie di attività portate avanti da attori. Quest'ultimi posso essere interni o esterni all'organizzazione con determinati privilegi e non. Il processo ha un inizio segnato da più eventi, e più fini indicati come 'stati'. Dietro i processi c'è un sistema che si evolverà in stati intermedi fino ad arrivare allo stato finale. Durante il processo i dati possono essere già presenti nell'organizzazione, essere prodotti durante il processo o anche usati e conservati per studi futuri. [9] Per modellizzare un processo ci si avvale di specifici strumenti, tra cui synopsis diagram, workflow.

A.1 Synopsis

Il synospsis diagram rappresenta il contesto del processo [10], vengono definiti:

- Attori
- Evento che determina l'inizio del processo
- Dati di input
- Dati di output
- I risultati del processo

In A.1 viene mostrato il formalismo del diagramma.

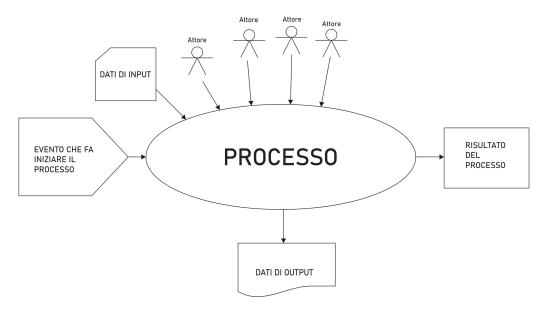


Figura A.1: Synopsis diagram

A.2 Workflow

Il workflow consente di modellare le macro-attività del processo [11]. Il livello di dettaglio che si sceglie ha una certa dipendenza dal pocesso che si vuole rappresentare. I costrutti fondamentali di questo diagramma sono due: place e transition, che rappresentano rispettivamente gli stati e le attività svolte. Le limitazioni logiche di questo modello sono:

- -non possono esserci due stati in successione, poichè il processo non può avere due stati differenti senza aver svolto un'attività nel mezzo;
- -non possono esserci due transizioni consecutive perchè a ciascuna azione il processo avrà comunque cambiato stato. Il formalismo è rappresentato in figura A.2

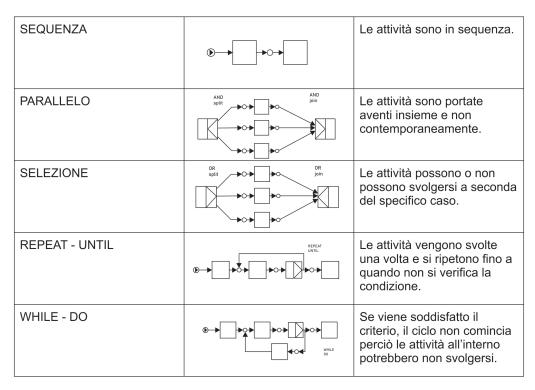


Figura A.2: Formalismo workflow.

Appendice B

Schermate applicativi

B.1 SGP

Selezionando la voce Piano di Lavoro su SGP, compare una finestra che permette la ricerca per unità/agenda. Questa suddivisione dipende dal presidio ospedaliero. A fronte di questa scelta, sarà possibile selezionare una delle unità, ovvero un raggruppamento delle prestazioni erogate dallo stesso regime. Si termina la selezione quando anche il codice mnemonico corrispondente al tipo di prestazione da erogare viene scelto. La figura seguente mostra la finestra SGP che permette la ricerca per unità.

Il Piano di Lavoro mostrato in B.2 fornisce tutte le informazioni relative agli appuntamenti del giorno corrente, per l'unità selezionata. Per ciascun paziente sono riportate le informazioni più salienti quali: Codice della richiesta, codice dell'appuntamento, codice e descrizione della prestazione, lo stato dell'appuntamento(Accettato/Eseguito, Non Presentato, Refertato, Registrato, Prenotato, Disdetta Tardiva, Da Confermare, Sospeso, Pre-appuntamento, Forzato, Arrivato, Stato No DEMA, In Refertazione). Sono presenti dei filtri che consentono di visualizzare esclusivamente i pazienti per Stato Appuntamento e Tipo Paziente.

A seguito della selezione di un appuntamento relativo ad un paziente già refertato, è possibile visualizzare i dati relativi alla refertazione cliccando su Refertazione. La schermata che si sussegue è detta "Lista Referti" B.3, la quale presenta tutti i dati relativi alla refertazione, quale il Codice di referto, Medico refertante e prescrittore, Unità, Data di refertazione, corrispondente codice richiesta e codice appuntamento.

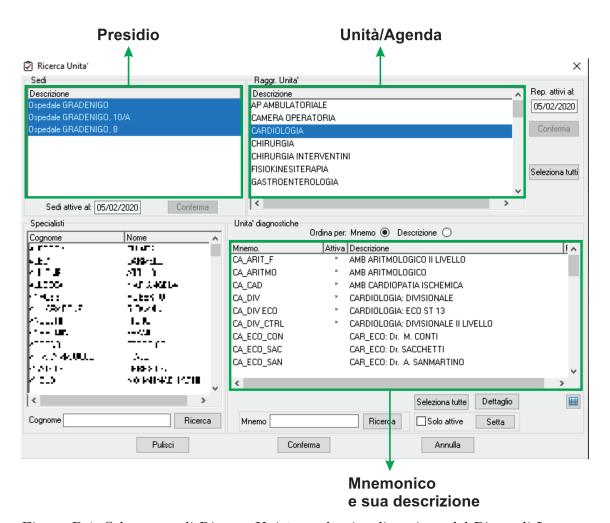


Figura B.1: Schermata di Ricerca Unità per la visualizzazione del Piano di Lavoro.

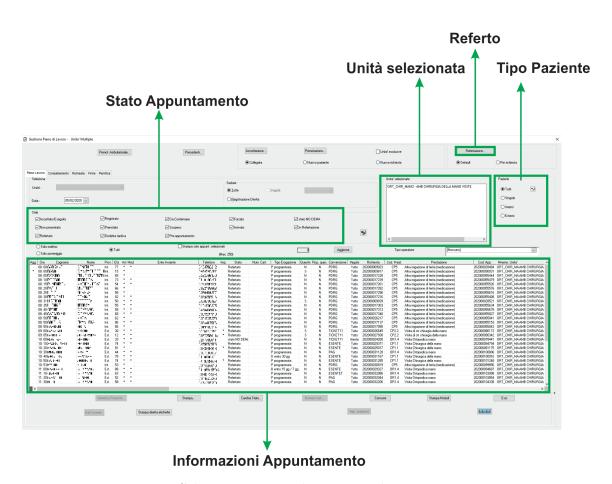


Figura B.2: Schermata piano di Lavoro selezionato per Unità.

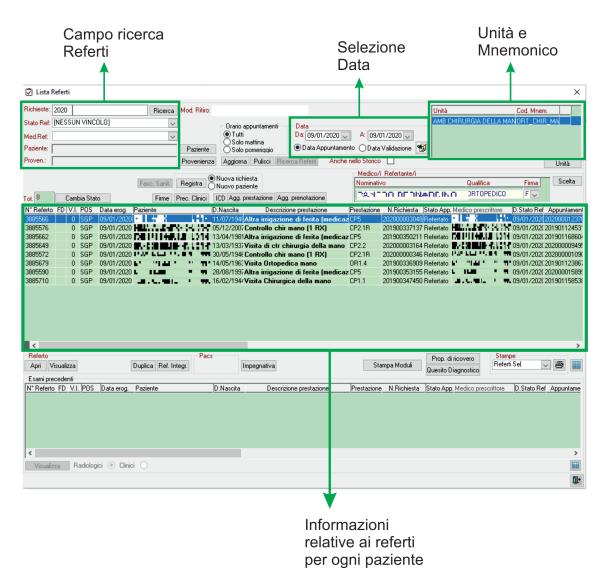


Figura B.3: Schermata piano di Lavoro selezionato per Unità.

B.2 AREAS

Si procede nella visualizzazione delle schermate principali della piattaforma per mostrare come sono stati individuati i campi delle tabelle del database che contengono le informazioni di maggiore interesse clinico. Successivamente sarà mostrata lo schema relazionale IRIDE con le tabelle selezionate. Ciascuna tabella mostrerà i campi individuati.

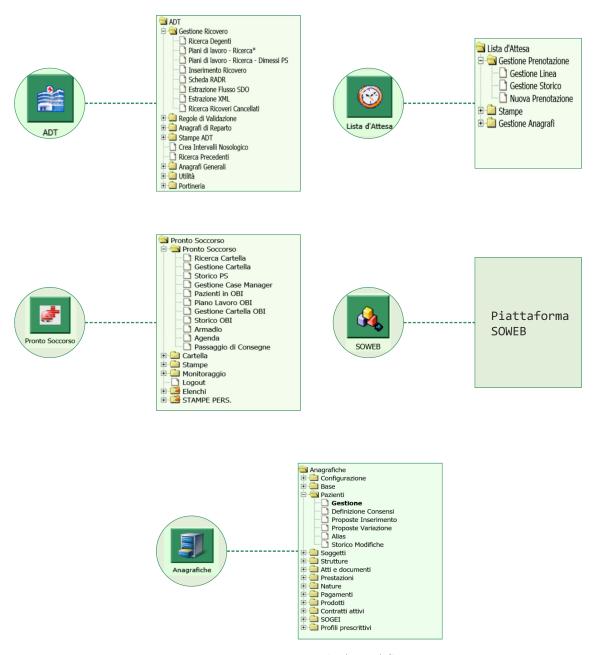


Figura B.4: Menù di AREAS.

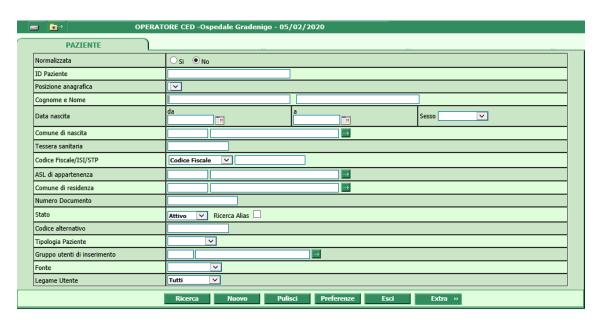


Figura B.5: Schermata anagrafica paziente.

Per accedere alle anagrafiche dei pazienti è necessario cliccare su "Anagrafiche" presente nel menù di Areas e successivamente "Gestione Pazienti" nel menù a tendina corrispondente. Nella schermata di B.5 sono mostrati tutti i campi utili per la ricerca di un paziente ricoverato; vi è la possibilità di aggiungere un nuovo paziente cliccando su "Nuovo" nella stessa schermata di ricerca.

Le informazioni relative al ricovero dei degenti sono riportate nella schermata in B.6. L'accesso alle schede di ricovero è possibile dall'icona "ADT" presente nel menù.

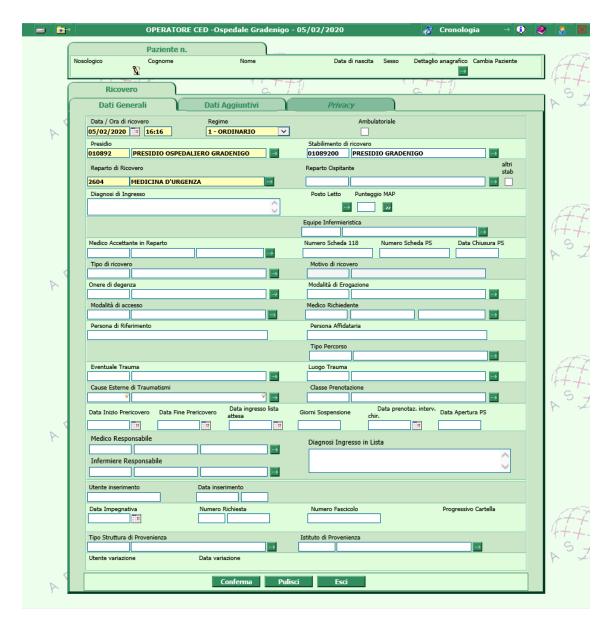


Figura B.6: Schermata scheda di ricovero.

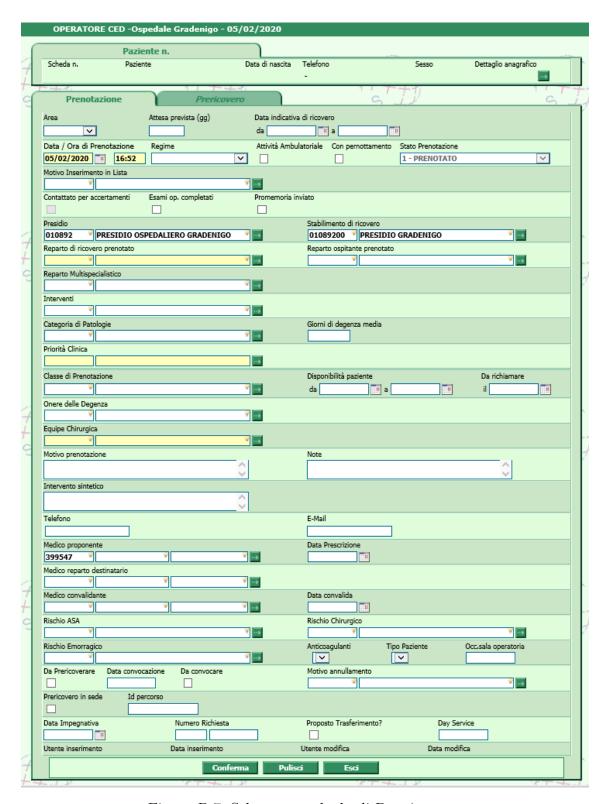


Figura B.7: Schermata scheda di Pre-ricovero.

B.3 SOWEB

Si procede alla visualizzazione delle schermate principali della piattaforma per mostrare come sono stati individuati i campi delle tabelle del database che contengono le informazioni di maggiore interesse clinico. Successivamente sarà mostrata lo schema relazionale SOWDATI con le tabelle selezionate. Ciascuna tabella mostrerà i campi individuati.

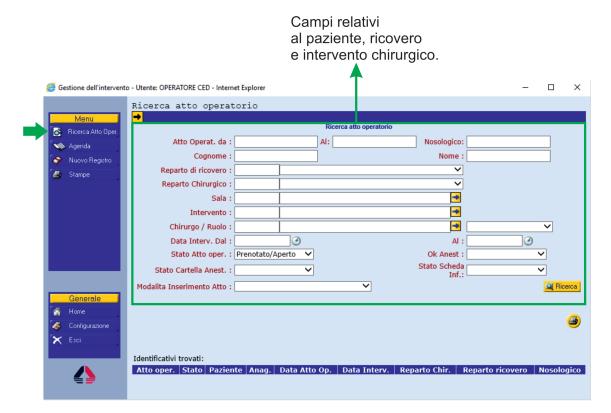


Figura B.8: Schermata Ricerca paziente.

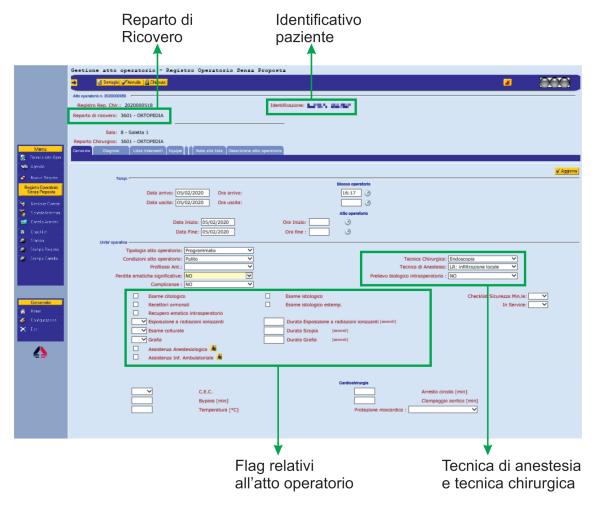


Figura B.9: Schermata informazioni generali atto operatorio.



Figura B.10: Schermata Descrizione atto operatorio.

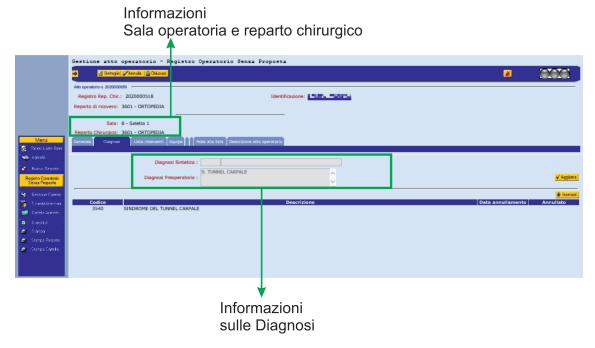


Figura B.11: Schermata Dettagli atto operatorio.

Bibliografia

- [1] Meystre Stéphane M. Mitchell Joyce A. Påvel G Deshmukh, Vikrant G. Evaluating the informatics for integrating biology and the bedside system for clinical research. *Jour*, 9, 2009.
- [2] Martin G Seneviratne, Tina Seto, Douglas W Blayney, James D Brooks, and Tina Hernandez-Boussard. Architecture and implementation of a clinical research data warehouse for prostate cancer. *eGEMs*, 6(1), 2018.
- [3] i2b2 training manual.
- [4] E. Di Buccio G.M. Di Nunzio. Basi di dati, progettazione concettuale, logica e sql. 1, 2017.
- [5] E. Di Buccio G.M. Di Nunzio. Data warehouse: teoria e pratica della progettazione. 1(Seconda edizione), 2006.
- [6] Paolo Giorgini, Manuel Kolp, John Mylopoulos, and Marco Pistore. The Tropos Methodology, pages 89–106. Springer US, Boston, MA, 2004.
- [7] Mark Levene and George Lozou. Why is the snowflake schema a good data warehouse design? *Information Systems*, 28:225–240, 05 2003.
- [8] EUPATI European Patients' Academy for Therapeutic Innovation. Sviluppo clinico e studi clinici.
- [9] Prof.ssa Gabriella Balestra. Slide corso progettazione di software medicali lezione 1. 2018/2019.
- [10] Prof.ssa Gabriella Balestra. Slide corso progettazione di software medicali lezione 2. 2018/2019.
- [11] Prof.ssa Gabriella Balestra. Slide corso progettazione di software medicali lezione 3. 2018/2019.