

L NUOVO SISTEMA DI DATA WAREHOUSE PER IL POLICLINICO UMBERTO I DI ROMA

L'utilizzo crescente dell'automazione nei processi di cura aumenta la quantità di dati sanitari raccolti dai sistemi informatici. Questa elevata disponibilità di informazioni permette analisi e studi sia a carattere scientifico che gestionale, ma per gestire l'ingente mole di dati sono necessari sistemi di elaborazione avanzati che, aggregando coerentemente le diverse fonti, permettano anche una fruibilità adeguata alle esigenze di analisi dei dati.

Offerta

Il CASPUR ha sviluppato un sistema di *data warehouse* (DWH) dedicato alla sanità e basato su meccanismi di *record linkage*, che permettono di determinare quali coppie di record, provenienti da basi di dati differenti, rappresentino la stessa entità. Per l'elevata quantità di informazioni contenute nella base di dati del DWH e per facilitarne la reperibilità, il CASPUR, inoltre, ha implementato un portale di analisi dei dati, basato su tecnologia MicroStrategy, che permette agli utenti di consultare cruscotti aziendali e produrre report, non solo statici ma anche dinamici, contenenti sia informazioni riassuntive che rappresentazioni grafiche.

Collaborazioni

Nel 2011 è proseguita con successo l'attività di collaborazione con il Policlinico Umberto I di Roma. In continuità con gli anni precedenti, il personale del Consorzio è stato coinvolto nelle attività di gestione tecnica dell'infrastruttura ICT e ha mantenuto costante il supporto alla UOC Sistemi Informativi dell'Ospedale (SIO) per le attività di progettazione e di innovazione tecnologica.

Risultati

Presso il Policlinico Umberto I, una volta definito stabilmente l'assetto tecnologico del SIO, è stato possibile passare dal trattamento dei dati delle singole applicazioni a un sistema unico, aggregato, utilizzabile sia per scopi gestionali che scientifici.

Il parco applicativo in produzione conta oltre trenta applicativi: poco meno di venti di ambito sanitario e il resto di ambito amministrativo o gestionale.

Il percorso del paziente nel corso del processo di cura è completamente tracciato grazie alla presenza di un completo strato di ADT (Ammissione, Dimissione e Trasferimento) e di un'interfaccia con il RECUP regionale per le prestazioni ambulatoriali.

Da un punto di vista tecnico l'integrazione tra i diversi applicativi, salvo poche eccezioni, avviene mediante un'architettura a stella, mostrata in Figura 1, basata su scambio di messaggi HL7, standard internazionale per la comunicazione tra applicativi in ambito sanitario.

Per stimare le dimensioni dei dati prodotti si può considerare che, nel 2011, sono state registrate circa 5.000.000 prestazioni di laboratorio, 220.000 prestazioni di radiologia e 35.000 di anatomia patologica.

I dati di produzione sono dunque distribuiti, in modo eterogeneo, in numerose basi di dati, su differenti tecnologie DBMS. Per questo motivo è necessario un sistema che funga da collettore di tutte le informazioni così da renderle facilmente accessibili ed utilizzabili come strumento per il supporto alle decisioni.

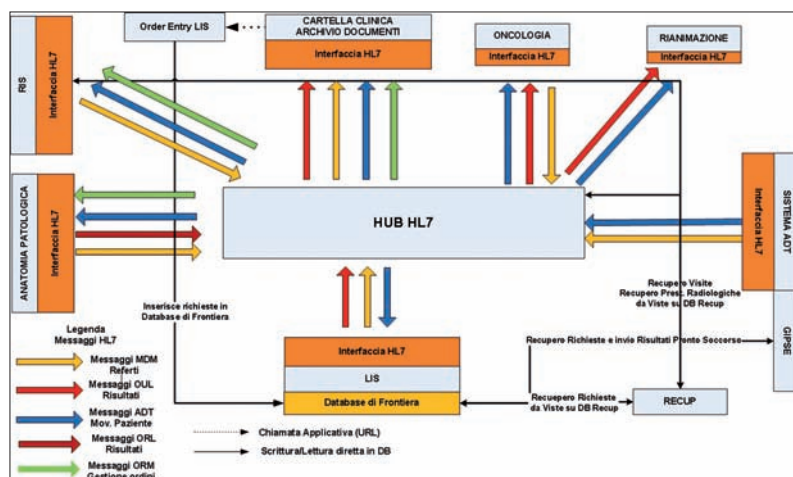


Fig. 1 Schema dell'integrazione applicativa basata su infrastruttura HL7.

L'architettura del data warehouse, mostrata in Figura 2, è basata su tre livelli in cui il primo consiste nelle basi di dati transazionali (OLTP¹) e nelle banche dati esterne utilizzate per arricchire il contenuto informativo presente all'interno del data warehouse. I dati presenti in questo livello vengono copiati all'interno dell'area di staging, dove vengono sottoposti ad operazioni di trasformazione, pulizia e infine caricamento nel data warehouse. In questa fase sono applicate le tecniche di record linkage per la riconciliazione dei dati provenienti dalle diverse fonti. I record vengono analizzati in due passi, in ognuno dei quali viene deciso lo stato di linkage di ogni coppia di record per determinare se si riferiscono alla stessa entità anagrafica. Il primo passo effettua un confronto deterministico basato esclusivamente su un identificatore anagrafico univoco, se presente. Il passo successivo usa invece un approccio non deterministico e utilizza una metrica di similarità tra stringhe (distanza di Jaro-Winkler) in cui maggiore è il valore della metrica tra le stringhe e maggiore sarà la loro similarità.

L'ultimo livello dell'architettura è rappresentato da un database dimensionale in cui sono state definite sia le metriche (tabelle dei fatti) sia le dimensioni di analisi (tabelle di dimensioni). Su questo database si aggancia infine il portale di analisi, come mostrato in Figura 2.

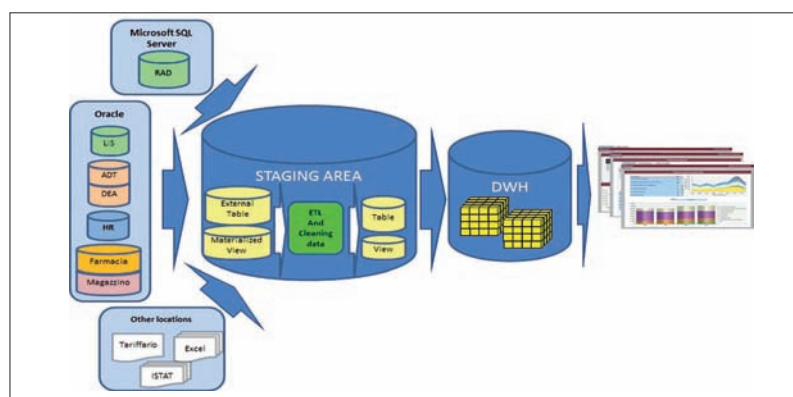


Fig. 2 Architettura del sistema di data warehouse.

Note

¹ OLTP è l'acronimo di On-line Transaction Processing ed è un insieme di tecniche per l'analisi dati che non sfruttano banche dati esterne, ma in cui le analisi vengono effettuate direttamente sui dati di esercizio.

Bibliografia essenziale

- [1] Kimball R., et al. (2009). *Kimball's data warehouse toolkit classics*. Wiley Publishing.
- [2] Verbicaro, R., Pecoraro, F., Mercurio, G. (2011). *HL7 principi di modellazione*. EDISEF.
- [3] Winkler, W.E. (2006). Overview of Record Linkage and Current Research Directions. *Research Report Series*.