# ANNARITA: Il database Object Relational dell'Anagrafe Nazionale delle Ricerche

**Alex Manzo**
*CILEA, Roma*

*Abstract*
Nell'ottica di un potenziamento e arricchimento dei dati dell'Anagrafe Nazionale delle Ricerche gestita dal CILEA per conto del Ministero dell'Università e della Ricerca, è stato pianificato un lavoro di rifacimento evolutivo del database contenente le informazioni dei soggetti iscritti allo schedario che svolgono ricerca finanziata con fondi pubblici. Questo database, denominato Annarita (ANagrafe NAzionale delle Ricerche – ITAlia), ha lo scopo di permettere una migliore gestione dei dati, attraverso l'utilizzo della tecnologia *object relational*, che sfrutta tutti i vantaggi del mondo degli oggetti ma permette una compatibilità con il modello relazionale puro.

As part of a strengthening and enrichment of the data the "Anagrafe Nazionale delle Ricerche" CILEA managed by the MUR, has scheduled a work to evolve the database from pure relational kind to object relational. This database contains a register with information about subjects involved in publicly funded research. This database called "Annarita" (ANagrafe NAzionale delle Ricerche – ITAlia) aims to enable better management of the data because it is based about evolutionary object relational technology that exploits the advantages of the world of objects but allows compatibility with the pure relational model.

*Keywords*: Object Relational, Oracle, UML.

## Introduzione

L'Anagrafe Nazionale delle Ricerche (ANR), che rappresenta uno strumento per conoscere lo stato della ricerca in Italia, gestisce uno schedario, costantemente aggiornato, degli enti finanziabili e dei finanziamenti concessi da parte degli enti erogatori per lo svolgimento di progetti di ricerca. La legge obbliga i soggetti proponenti a effettuare l'iscrizione all'ANR attraverso una scheda, con la quale si raccolgono alcune informazioni del soggetto, principalmente di natura anagrafica, quali: denominazione, indirizzo, etc. Attualmente i dati acquisiti vengono memorizzati in un database relazionale.

Nell'ottica di un potenziamento e arricchimento continuo dell'Anagrafe, è stato migliorato il corredo informativo relativo ai soggetti facenti ricerca, anche in relazione alla loro diversa natura giuridica. Esistono infatti diverse tipologie di enti che svolgono ricerca e, poiché ognuna di queste mostra aspetti strutturali diversi per competenze e ruoli, ne consegue una diversa gestione delle informazioni. Basti pensare, per esempio, alla differenza tra un'impresa privata e un'università.

Questa vasta tipizzazione ha inoltre determinato la ricerca di ulteriori fonti da cui recuperare informazioni per conseguire un arricchimento dei soggetti finalizzato non solo ad avere maggiori notizie dell'attore stesso, ma anche relative al suo "contorno". Per questo sono di particolare interesse, per esempio, i dati dei "gruppi" ossia le entità che esercitano il controllo sulle imprese iscritte all'Anagrafe. In Figura 1 è presente un esempio di gruppo. In questo modo si rendono disponibili dei controlli amministrativi e di monitoraggio, che permettono al Ministero dell'Università e della Ricerca (MUR) di conoscere lo stato della ricerca in Italia. L'inserimento di nuove informazioni nell'esistente database relazionale dell'Anagrafe ha costituito la premessa per il rifacimento evolutivo della parte di database relativa alle informazioni sui soggetti iscritti allo schedario.

Si è così sviluppata una soluzione tecnologica innovativa, che consente di rinnovare il database dei soggetti, utilizzando le caratteristiche della programmazione ad oggetti disponibili nel DBMS Oracle, che implementa lo strato della persistenza nel sistema informatico dell'Anagrafe.

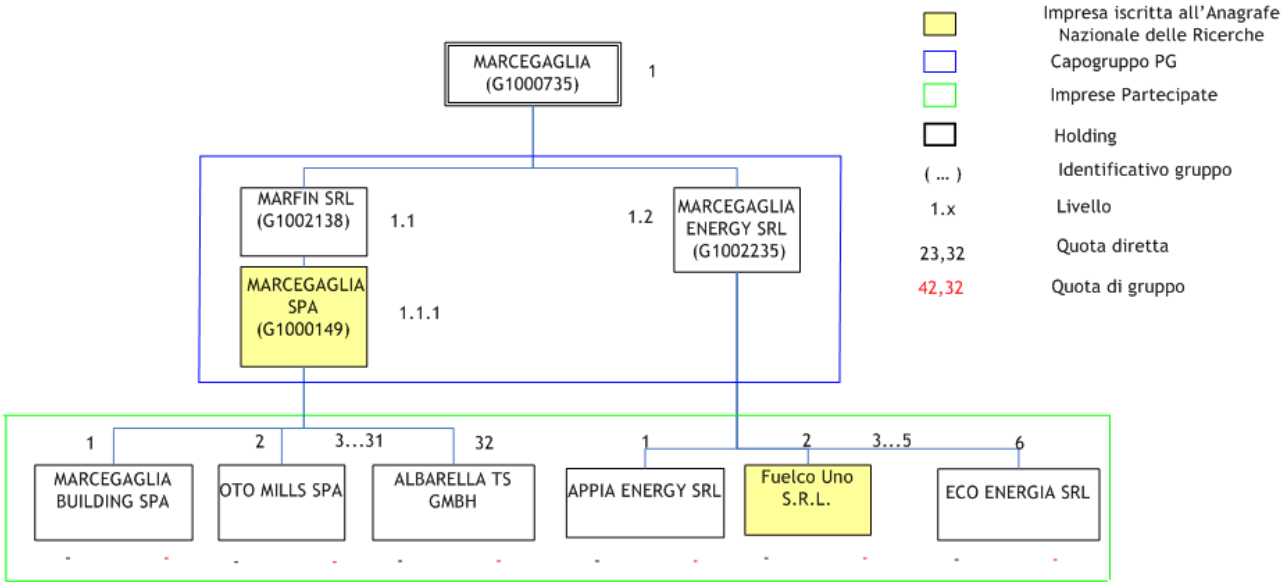


*Fig. 1 – Esempio Gruppo*

Più precisamente, la soluzione tecnologica che si è sviluppata prevede l'uso dell'approccio evoluzionistico *object relational,* che consente di avere un legame diretto tra struttura del database e applicazione software. La tecnologia Oracle a oggetti, rappresenta infatti, uno strato di astrazione sopra la tecnologia relazionale. Dietro lo "strato" dell'oggetto i dati sono sempre memorizzati in colonne e tabelle, ma ora possono essere elaborati in termini di entità del mondo reale, tali da renderli espressivi in accordo al mondo della programmazione a oggetti. Nell'interrogazione di una base di dati, invece di pensare in termini di colonne e tabelle, si può operare in termini di oggetti e entità. Ulteriore vantaggio dell'approccio object relational è il mantenimento della compatibilità con i precedenti database relazionali, utile nel caso dell'ANR, per le relazioni esistenti con altri servizi informatici utilizzati dal Ministero.

Con questa innovazione si possono così gestire in maniera più efficace tutte le complessità strutturali e di movimentazione dei soggetti. Inoltre si può avere una visione più ordinata dei soggetti stessi, che tiene conto della loro evoluzione nel tempo dopo l'iscrizione all'Anagrafe, anche per rendere più facile il loro riconoscimento futuro. Il sistema progettato è in grado di mostrare i dati presenti ANR nel modo più completo e organizzato possibile. Si è così realizzato un database object relational denominato Annarita (ANagrafe NAzionale delle Ricerche – ITAlia) per la persistenza dei soggetti iscritti nello Schedario.

## Il modello logico di Annarita

L'Anagrafe Nazionale delle Ricerche contiene il registro dei soggetti pubblici e privati che svolgono ricerca in Italia finanziata, tutta o in parte, con fondi a carico del bilancio dello stato e di enti pubblici. Come detto, la varietà di questi soggetti dipende dalla propria natura e struttura giuridica e per questo non è banale avere un *repository* comune in grado di contenere tutte le informazione anagrafiche di interesse. In seguito all'evoluzione dell'ANR e al suo continuo arricchimento di dati provenienti da più fonti esterne, si è proposto, con ancora più evidenza, il problema di avere un posto, il più possibile completo, dal quale recuperare tutte le informazioni di un soggetto iscritto utili alla sua gestione nell'ambito dell'Anagrafe.
Per tutti questi motivi si è pensato di realizzare il nuovo database Annarita con la tecnologia object relational. In questo modo si sono raggiunti gli obiettivi preposti grazie alle potenzialità degli oggetti, ma allo stesso tempo si è lasciata la persistenza dei dati compatibile con tutti gli altri sistemi preesistenti, basati sulla logica del database relazionale puro.

Il grafico di Figura 2 presenta il diagramma delle principali entità individuate. Come indicato nella legenda, la differente colorazione segnala la diversa provenienza delle informazioni

previste in questo database; questo aspetto verrà approfondito nel paragrafo successivo. Nel modello si può notare la presenza dell'entità *Soggetto,* che rappresenta un qualsiasi soggetto di ricerca facente parte della banca dati dell'Anagrafe. I soggetti possono essere di due tipologie: coloro che realizzano in modo attivo i progetti di ricerca ossia le *Imprese*, le *Università* e gli *Enti di Ricerca* e coloro che hanno funzionalità di supervisione e controllo ossia *Ministero* e *Regioni*. Nel database sono inoltre memorizzate tutte le strutture che compongono le imprese e le università quali ad esempio le unità locali e le facoltà. È prevista inoltre la memorizzazione dei dati dei brevetti conseguiti da parte di un soggetto a seguito di una attività di ricerca finanziata. Se una impresa controlla altre imprese, ai fini della determinazione della sua dimensione si deve fare riferimento non alla singola impresa, ma all'insieme delle imprese collegate: al gruppo. Sono così memorizzate le informazioni del Gruppo di appartenenza.

Si sottolinea infine che questo diagramma è il risultato di diverse versioni che sono andate di pari passo con lo studio delle varie entità definite.

## Provenance dei dati

Nell'ambito del potenziamento dell'Anagrafe Nazionale delle Ricerche, si sono inseriti dati e informazioni che prima di questa innovazione del database non erano contemplati e previsti. Questi dati sono stati aggiunti a quelli precedentemente acquisiti dalla scheda di iscrizione che era, dei soggetti, l'unica sorgente dell'ANR. Per ogni soggetto, come visto nel diagramma del database, sono presenti diverse tipologie e quantità di informazione in relazione alla propria natura giuridica. È per questo importante fare un quadro sulle fonti dalle quali provengono questi dati. È fondamentale focalizzare e tenere traccia della fonte da cui proviene il dato del soggetto sia per capire l'informazione che si ha, sia per ricostruire e correlare l'intero corredo informativo.

Si definisce *provenance* dei dati la provenienza dell'informazione; di questa si possono sottolineare due aspetti: *source provenance* e *temporal provenance.* Per la prima l'idea è di fornire una provenienza della fonte per ciascuna tipologia di informazione (in questo caso dei singoli oggetti del modello), per la seconda l'idea è di fornire una provenienza temporale per indicarne la data dei dati presenti nel database.

L'esistenza di informazioni riguardanti uno stesso soggetto nell'ambito di diversi sistemi software e con diversi dettagli è cosa ricorrente. Avere notizie su un soggetto dell'ANR da tante fonti informative è sicuramente un vantaggio in quanto si possono avere più dati sui quali effettuare operazioni di controllo e di gestione di un soggetto, durante lo sviluppo di un progetto di ricerca da esso realizzato. Soprattutto nel caso delle aziende private, avere maggiori informazioni anagrafiche consente una loro più facile individuazione nel vasto mondo delle imprese. È importante anche approfondire i dati relativi alla natura del soggetto per avere maggiori relazioni sulle specifiche attività di cui si occupa e quindi sulla correlazione col progetto di ricerca svolto. Avere più informazioni possibili consente, inoltre, di effettuare, da parte del ministero, un più efficace monitoraggio delle spese dei fondi di finanziamento per la ricerca. Per contro, tante fonti dati portano anche degli svantaggi; uno su tutti è la difficoltà nella gestione. Importare i dati da più sorgenti significa occuparsi, attraverso più applicazioni informatiche, dell'acquisizione dei dati.

La complessità sta inoltre nella provenienza temporale accennata prima. I dati dei soggetti possono essere abbastanza dinamici nel tempo e per questo è importante avere il dato aggiornato. Questo deve essere rapportato alla specificità del database in cui i dati sono contenuti: per esempio è più importante, per la gestione dell'ANR, avere aggiornato il dato dell'indirizzo della sede di un soggetto, rispetto all'informazione dell'acronimo. Un altro aspetto interessante da considerare è la duplicazione di una stessa informazione. È infatti possibile avere lo stesso campo da due fonti diversi; in questo caso si devono gestire le possibili differenze del valore in esso contenuto e capire quale delle due è più affidabile. Effettuare confronti di questo tipo permette di raggiungere l'obiettivo di avere un maggiore interscambio di dati a tutto vantaggio della correttezza del contenuto informativo dell'Anagrafe, con l'obiettivo di avere i dati del soggetto più aggiornati possibile. Ad esempio, nel diagramma di Figura 2, l'informazione dell'indirizzo della sede legale di un'impresa è presente sia nel *Soggetto* che nell'*Impresa*; i dati del primo sono fonte Anagrafe, quelli del secondo sono fonte Registro Imprese (InfoCamere). Per tutti questi motivi si è presa in considerazione, in fase di progettazione e realizzazione della struttura del database, la provenienza delle informazioni.

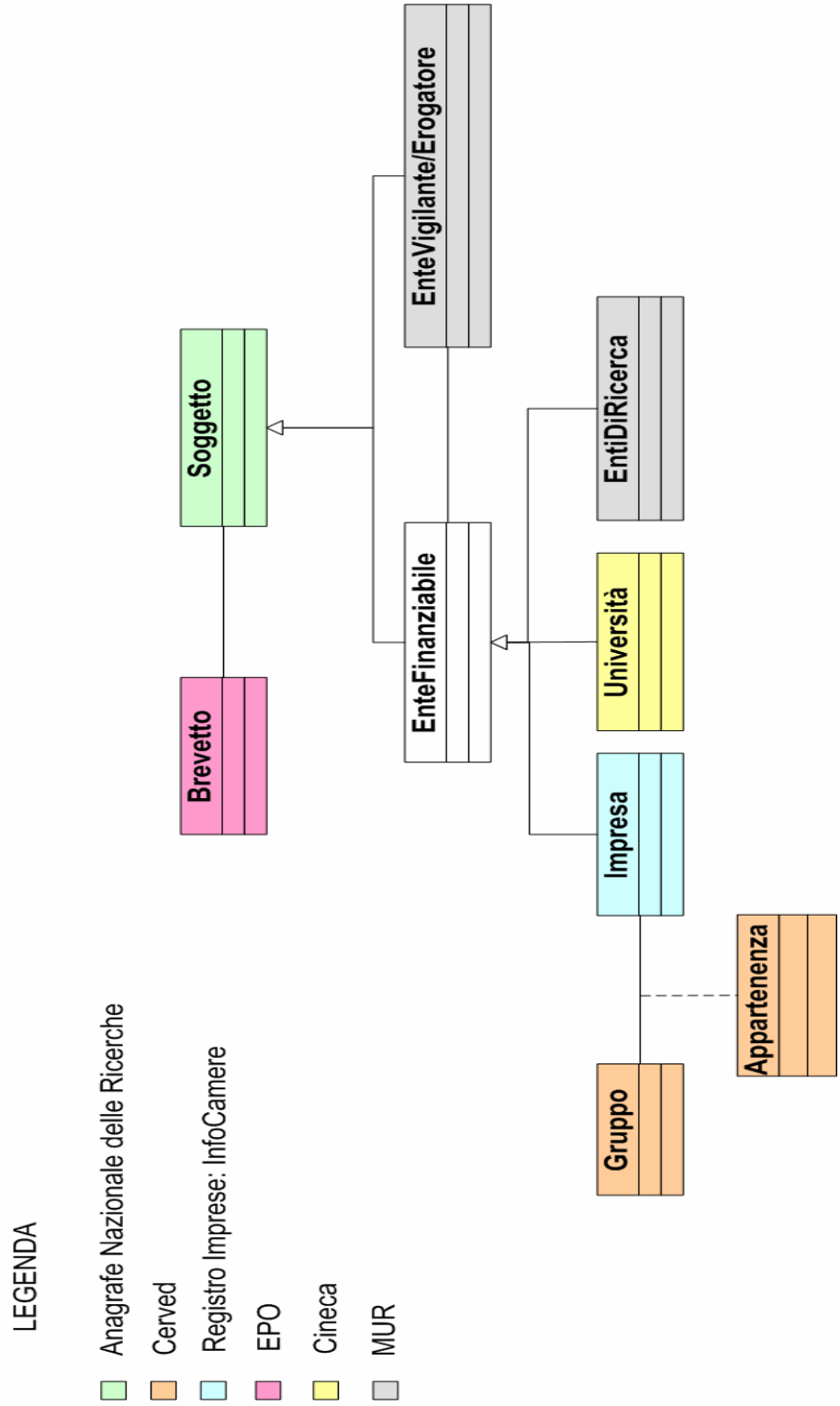


*Fig. 2 – Modello logico di Annarita*

## Annarita: vantaggi

I benefici fondamentali ottenuti dalla creazione di Annarita con l'approccio object relational sono da imputarsi alle caratteristiche della programmazione ad oggetti: su tutti la possibilità di definire la funzionalità sui dati direttamente nel database. La difficoltà di una gestione dinamica dei cambiamenti di un soggetto quale un'impresa iscritta all'Anagrafe era, fino alla creazione di Annarita, demandata al software. Si trovavano, inoltre, notevoli difficoltà di gestione della persistenza delle informazioni di un cambiamento giuridico, per esempio una fusione tra due imprese iscritte. Ora tutto questo è direttamente gestito nel database, attraverso apposite funzioni che effettuano le opportune modifiche dei dati derivanti da una fusione.

Un altro notevole beneficio, che si riscontra in Annarita, rispetto al precedente database relazionale, è la possibilità di salvare in modo più caratterizzante e il più possibile aderente alla realtà la struttura di un ente iscritto all'Anagrafe. Anche in questo caso si riporta l'esempio dell'entità Gruppo: memorizzare in tabelle relazionali le strutture di un gruppo sopra definite significa effettuare un notevole lavoro sul codice dell'applicazione che vuole interpretarli per generare grafici come quelli di Figura 1. Utilizzare questa tecnologia permette di superare queste difficoltà.

## Sintesi - Conclusioni

Il CILEA ha pianificato un lavoro di rifacimento evolutivo del database contenente le informazione dei soggetti iscritti allo schedario che svolgono ricerca finanziata con fondi pubblici. Questo database è stato denominato Annarita (ANagrafe NAzionale delle Ricerche – ITAlia).

Il lavoro si è concretizzato nella realizzazione di un database object relational per l'Anagrafe Nazionale delle Ricerche che consente una nuova memorizzazione dei dati e gestione dello schedario dei soggetti che svolgono ricerca in Italia.

Il lavoro è stato caratterizzato dallo studio del dominio dell'ANR, ovvero degli attori della ricerca in Italia, e della tecnologia usata. L'attività è proseguita con la definizione, attraverso varie versioni, del modello logico del database. Da questo è stato generato Annarita. Successivamente si sono esaminate le diversi fonti dati che arricchiscono questo database in modo da strutturare le procedure di caricamento del database.

Si è così raggiunto l'obiettivo di realizzare un database dell'Anagrafe più ricco di informazioni e che consente, attraverso le caratteristiche del mondo degli oggetti proprie all'approccio object relational, una migliore gestione dello schedario.

Un'interessante prospettiva di valutazione per eventuali sviluppi futuri è rappresentata dalla possibilità di ampliare questo database con l'inserimento dei dati dei progetti di ricerca svolti dai soggetti iscritti nello schedario. Il progetto prevede la realizzazione di un altro database object relational definito sul modello logico derivato dalla struttura delle informazioni riguardanti i progetti. In questo modo l'integrazione fra soggetti e progetti può diventare diretta e consentire quindi un monitoraggio della ricerca italiana sempre più completo e preciso, grazie ai vantaggi derivanti dal mondo della programmazione ad oggetti.

## Bibliografia

[1] Object Relational Features 10g Realise2 Application Developer's Guide – June 2005 ORACLE.

[2] Anagrafe Nazionale delle Ricerche URL: http://www.anagrafenazionalericerche.it

[3] Database Object Relational URL: http://en.wikipedia.org/wiki/Object-relational_database

[4] *Basi di dati – Architetture e linee di evoluzione* Paolo Atzeni, Stefano Ceri, Piero Fraternali, Stefano Paraboschi, Riccardo Torlone McGraw-Hill.

[5] ML for Database Design Eric J. Naiburg, Robert A. Maksimchuk Addison-Wesley Professional.