



UNIVERSITÀ DEGLI STUDI DI PARMA

Facoltà di Ingegneria

Corso di Laurea in Ingegneria Informatica

**SISTEMI INFORMATIVI
PER LA RICERCA:
CRIS**

Relatore:

Chiar.mo Prof. GIANNI CONTE

Correlatore:

Chiar.mo Prof. STEFANO CAGNONI

Laureando:

LUCA ZONI

Anno Accademico 2005-2006

*Alla mia famiglia
per aver creduto in me*

Ringraziamenti

Sono molte le persone che devo ringraziare sia per questo lavoro che per questi anni universitari, è anche grazie alla loro presenza e ai loro consigli che sono riuscito a raggiungere questo traguardo importante.

Il mio primo pensiero, ovviamente, va alla mia famiglia, senza i quali non sarei mai potuto arrivare a questo punto, in particolare ringrazio mio papà Luigi e mia mamma Angela, che in tutti questi anni mi hanno sempre incoraggiato e sostenuto, soprattutto nei momenti difficili. Ringrazio poi i miei nonni Ida, Gianni e Maria, che hanno sempre avuto grande fiducia in me, mio nonno Amos, a cui sarebbe piaciuto tantissimo vedermi laureato, mio cugino Marco e i miei zii Vincenza e Giorgio per l'appoggio che hanno saputo darmi.

Un ringraziamento particolare va a Tania, che in questo difficile percorso mi è sempre stata vicina, mi ha spronato ad andare avanti e mi ha sopportato nei tanti momenti in cui, a causa degli esami o del troppo studio, ero davvero insopportabile, senza di lei non so se sarei riuscito a farcela.

Vorrei poi ringraziare il prof. Gianni Conte e il prof. Stefano Cagnoni, ovvero il relatore e il correlatore della mia tesi, per i consigli e le dritte che hanno saputo darmi, in modo da svolgere questo lavoro senza particolari problemi.

Ringrazio poi tutti gli amici dell'università, non potrei non iniziare da Tullo e Boggio, che con me formavano “i tri siochetti”, grazie per avere reso più sopportabili le lezioni e le lunghe giornate di studio, che poi si trasformavano quasi sempre in interminabili partite con l'Xbox o con il computer. Ringrazio poi il “tavolo di studio”, ovvero Giava, Cirio, Fava, Baro, Giampi e Canto, per i pomeriggi passati insieme giocando a Risiko, a Uno o a Sette e mezzo. Ringrazio infine gli altri “compagni di viaggio” ad Ingegneria: Gnusso, Sbatti, Corlo, Lollo, Lore, Leo, Giusy, Furlo, Genno, Mario, Silvia e Matte.

Infine, ma non in ordine di importanza, desidero ringraziare tutti gli amici della compagnia di Fontanelle perché quando volevo rilassarmi un po' e non pensare allo studio potevo sempre contare su di loro per passare bei momenti di svago e divertimento, grazie Dani, Care, Dalle, Teo, Gabrio, Berga, Levanza, Bocchi, Gianchi, Nico, Matti, Pier, Mary, Volpi, Moli, Salva, Serena, Mango, Gemma, Isa, Sara, Vise, Silvia, Diego, Titta, Michele, Zanardi, Jardel, Plinto, Ughina, Lando, Luis, Talos, Ivan, Patto, Fra, Pelo, Pippo, Ori, Batta.

Grazie a tutti!!

Indice

Prefazione	v
1 Introduzione a CERIF	1
1.1 CERIF 2000	1
1.1.1 Aumento della copertura delle informazioni di ricerca	2
1.1.2 Flessibilità del modello di dati	2
1.1.3 Schemi di indicizzazione per soggetto e tecnologie di indicizzazione automatica.....	3
1.1.4 Approccio plurilingue (reperimento di informazioni a lingue incrociate)	4
1.1.5 Utilizzo dei metadata per integrare il modello di dati	5
1.2 Indicizzazione per soggetto	6
1.2.1 Ortelius Thesaurus	7
1.3 Gruppi di utenti interessati	8
2 Analisi documenti	11
2.1 “CERIF: Past, Present and Future: an Overview”	11
2.2 “CERIF: Information Retrieval of Research Information in a Distributed Heterogeneous Environment”	13
2.3 “Comparative Study of Metadata for Scientific Information: The place of CERIF in CRISs and Scientific Repositories”	16
2.4 “CERIF Interoperability and Compatibility”	18
3 Modello di dati di CERIF	19
3.1 Full CRIS data model	20
3.1.1 Livello 1 - Entità principali	21
3.1.2 Livello 2 - Entità secondarie.....	22
3.1.3 Livello 3 - Entità di supporto plurilingue	27
3.1.4 Livello 4 - Entità di lookup	29
3.1.5 Livello 5 - Entità di link (o entità intermedie).....	32
3.2 CERIF exchange model.....	36
3.3 Metadata data model.....	38
4 Esempi di sistemi basati su CERIF	41
4.1 EUREKA	41
4.2 HunCRIS	45
4.3 The Projekt database.....	48
4.4 WISARD	51
Conclusioni	56
Glossario	57
Bibliografia	58

Indice delle figure

Figura 3.1: relazione fra i modelli di dati di CERIF ed i CRIS esistenti.....	19
Figura 3.2: entità principali.	21
Figura 3.3: entità secondarie relative a Person.	23
Figura 3.4: entità secondarie relative a Project.....	24
Figura 3.5: entità secondarie relative a OrgUnit.....	26
Figura 3.6: entità di supporto plurilingue associate direttamente alle entità principali..	27
Figura 3.7: entità di link per le relazioni tra le entità principali.	33
Figura 3.8: entità intermedie associate a Person e relative ai dati personali.	34
Figura 3.9: entità intermedie associate a Project e relative ai risultati dei progetti.....	35
Figura 3.10: divisione delle entità secondarie nei vari modelli di scambio di dati.	37
Figura 3.11: metadata data model.	39
Figura 4.1: ricerca avanzata dei progetti in EUREKA.	42
Figura 4.2: esempio di risultati di una ricerca in EUREKA.....	44
Figura 4.3: esempio della pagina principale di un progetto in HunCRIS.....	47
Figura 4.4: esempio della pagina di un'unità organizzativa in HunCRIS.	48
Figura 4.5: form per la ricerca in Project Database.....	49
Figura 4.6: esempio della pagina principale di un progetto in Project Database.	50
Figura 4.7: sistema di ricerca dei progetti in WISARD.	52
Figura 4.8: esempio di una parte della pagina principale di un progetto in WISARD...	53

Indice delle tabelle

Tabella 3.1: codici per identificare i tipi di unità organizzativa.....	30
Tabella 3.2: esempi di codici ISO 4217 per identificare le monete europee.....	30
Tabella 3.3: esempi di codici ISO 3166 per identificare gli stati europei.	30
Tabella 3.4: esempi di codici ISO 639 per identificare le lingue europee.....	31

Prefazione

La politica di innovazione dell'Unione Europea punta a migliorare e a rafforzare la creazione e lo sfruttamento dei progetti di Ricerca e Sviluppo (R&S). Per raggiungere questo obiettivo, un requisito essenziale è l'accesso alle informazioni sulle attuali attività di ricerca in tutta Europa. In questo modo nuove idee di R&S possono emergere grazie ad un approccio pan-Europeo per la condivisione e lo scambio delle informazioni. C'è così il bisogno di un mezzo per diffondere informazioni di R&S rilevanti, in modo da incoraggiare innovazione e nuove ricerche. Questo mezzo, a livello di ogni singola nazione, è rappresentato da un CRIS (Current Research Information System), ovvero ciascun canale informativo dedicato ad offrire accesso alle informazioni di R&S; un CRIS è formato da un modello di dati che descrive i vari progetti in via di sviluppo o già realizzati e da un tool (o un set di tool) per maneggiare questi dati. Un CRIS ha due principali obiettivi: migliorare la comunicazione fra scienziati riguardo a progetti in fase di realizzazione e offrire una base di informazioni ai direttori del programma nazionale di ricerca e sviluppo.

Come è già stato detto sopra, c'è la necessità, di avere una "interfaccia comune" a livello europeo per gestire questi dati, che riesca ad integrare e comparare le informazioni derivanti dalle diverse fonti di dati; queste informazioni devono essere presentate in un modo uniforme, almeno a livello riassuntivo. Questa interfaccia comune non solo deve indirizzare il contenuto (quello che deve essere scambiato), ma anche la configurazione di tali informazioni (come dovrebbe essere presentato). Per raggiungere questi obiettivi i mezzi migliori a cui ci si può affidare sono il WWW (World Wide Web) e la tecnologia dei database. A questo proposito è stato proposto un formato comune per la gestione delle informazioni di ricerca per tutti gli stati Europei, ovvero CERIF (Common European Research Information Format).

In questa tesi si è studiato a fondo il modello di dati di CERIF, proponendo anche vari esempi reali di utilizzo di esso. Lo scopo di questo lavoro è la creazione, in futuro, di un CRIS compatibile con CERIF che contenga i vari progetti di ricerca realizzati all'interno dell'Università degli Studi di Parma.

In dettaglio il contenuto dei vari capitoli è il seguente:

- Nel capitolo 1 si parla della prima versione di CERIF e della revisione che è stata fatta per migliorarla, vengono illustrate poi le varie caratteristiche di CERIF 2000 e, infine, viene spiegato a quali gruppi di utenti potrebbe interessare.
- Nel capitolo 2 vengono analizzati quattro documenti riguardanti CERIF, questi spiegano meglio alcuni aspetti del modello di dati.
- Nel capitolo 3 vengono spiegate in dettaglio tutte le entità e gli attributi del modello di dati di CERIF.
- Nel capitolo 4 vengono presentati quattro esempi reali di sistemi che utilizzano il modello di CERIF o un modello compatibile con esso.

1 Introduzione a CERIF

Il primo CERIF fu pubblicato nel 1991 con lo scopo di facilitare gli scambi di informazioni tra i vari CRIS, questo era composto da un modello dati per descrivere progetti di ricerca che includeva:

- un elenco di dati essenziali per descrivere il progetto;
- un elenco di dati opzionali, utili per descrivere ulteriori informazioni riguardo il progetto;
- raccomandazioni riguardanti la classificazione delle attività di ricerca.

Questo modello di dati era formato da un'entità principale, ovvero il progetto, con vari attributi: il titolo, il riassunto, le persone e le organizzazioni che vi hanno collaborato, le date di inizio e di fine, i codici per la classificazione e le informazioni sui finanziamenti ricevuti. Erano inoltre presenti attributi opzionali, come i risultati ottenuti, le pubblicazioni e le attrezzature necessarie.

È da notare però che in questo primo CERIF veniva data molta importanza solamente al progetto, mentre le informazioni riguardanti le persone e le unità organizzative che vi avevano preso parte erano soltanto attributi di questa entità principale. Questa prima configurazione di CERIF è stata poi revisionata per le ragioni seguenti:

- a) il CERIF originale copriva con informazioni dettagliate solamente i progetti di ricerca, mentre gli utenti dei CRIS volevano estenderlo alla gestione dei dati riguardanti le persone, le organizzazioni e le altre entità che collaborano alla realizzazione dei progetti;
- b) lo “schema di classificazione del soggetto della ricerca” proposto in CERIF 1991 non è stato più aggiornato dal 1988, quindi è stato necessario estenderlo per coprire tutte le nuove aree di dati e migliorare la copertura di quelle esistenti;
- c) le nuove tecnologie, in particolare l'uso molto esteso di Internet e World Wide Web, hanno cambiato la natura delle attività di base dei CRIS e hanno aperto nuove strade per servire i loro utenti.

1.1 CERIF 2000

La revisione del vecchio CERIF, effettuata dal “CERIF Revision Working Group” (un gruppo di esperti degli Stati Membri dell'Unione Europea e degli Stati Associati) sotto la coordinazione della Commissione Europea, ha portato alla creazione di “CERIF 2000”, il quale ha i seguenti obiettivi:

- aiutare gli sviluppatori di nuovi sistemi di informazioni di ricerca, offrendo loro un set di componenti appropriati presi da un modello dati di informazioni di ricerca completo;

- offrire linee guida per lo scambio di dati riguardo le informazioni su R&S tra i vari CRIS e facilitare l'accesso alle varie fonti in tutta l'Unione Europea;
- fornire un esempio di un sistema per la gestione delle informazioni di ricerca per i nuovi arrivati e per le organizzazioni che vogliono espandere i propri sistemi di informazione.

Per raggiungere questi obiettivi, il “CERIF Revision Working Group”, ha apportato alcune migliorie al progetto CERIF originale che riguardano in particolare:

- 1) un aumento della copertura delle informazioni di ricerca;
- 2) un approccio flessibile per il modello di dati di CERIF 2000, questo deve essere un'ottima interfaccia per i nuovi CRIS, ma deve adattarsi perfettamente anche ai CRIS già esistenti;
- 3) indicizzazione per soggetto (catalogazione alfabetica dei soggetti che include anche la classificazione) automatica e tradizionale;
- 4) un approccio plurilingue;
- 5) utilizzo dei metadata per integrare il modello di dati.

È stata poi realizzata un'ulteriore revisione nel 2002, ma questa ha portato a modifiche lievi, che riguardano solamente alcune entità del modello di dati; sono infatti stati aggiunti alcuni attributi, altri sono stati cancellati e altri ancora solo modificati.

1.1.1 Aumento della copertura delle informazioni di ricerca

CERIF 2000 estende il raggio d'azione dei “progetti” del CERIF originale per coprire i seguenti tipi di informazioni di ricerca supplementari:

- organizzazioni, con informazioni sulle attrezzature di cui dispongono e sulle attività di cui si occupano;
- persone, con le relative competenze;
- prodotti, brevetti, pubblicazioni e gli altri “risultati” dei progetti di ricerca;
- attrezzature e mezzi necessari per lo sviluppo dei progetti.

1.1.2 Flessibilità del modello di dati

Il modello di dati di CERIF 2000 deve essere molto flessibile, in quanto non solo deve fornire un'ottima base di dati per i CRIS futuri, ma deve anche adattarsi perfettamente ai CRIS già esistenti, ovvero deve soddisfare le seguenti esigenze:

- a) definire un modello dati di CRIS completo che coprirà le strutture dei database della maggioranza dei CRIS esistenti;
- b) definire un set di modelli di dati che possano offrire esempi per lo scambio di informazioni, questo perché non è necessario implementare il modello dati completo di CRIS nel caso in cui ne sia richiesto solamente un particolare sottoinsieme;

- c) definire un modello di metadata per offrire una veduta uniforme sulle fonti di informazioni, che sono eterogenee.

1.1.3 Schemi di indicizzazione per soggetto e tecnologie di indicizzazione automatica

Con “tecniche di indicizzazione automatica” si intendono programmi software che “automaticamente” forniscono punti di accesso ai soggetti; l’indicizzazione per soggetto tradizionale, invece, usa un vocabolario controllato (schemi di classificazione e/o thesauri).

L’indicizzazione dei record usando un set predefinito di termini, come per esempio quelli trovati in un vocabolario controllato, è di supporto a livello del singolo database, ma è anche essenziale nel caso di un accesso a set di dati multipli provenienti da fonti diverse in un ambiente plurilingue. Nel caso in cui si utilizzi un vocabolario controllato, però, deve essere assicurato un aggiornamento regolare, è quindi richiesto l’intervento umano, e ciò spesso induce ad un’indicizzazione per soggetto retrospettiva sui dati esistenti.

L’indicizzazione automatica sembra quindi preferibile, ma questo metodo non è molto indicato nel caso in cui siano immessi dati plurilingue o set di dati multipli provenienti da fonti multiple (si noti che la gestione della terminologia plurilingue è molto importante visto che la piattaforma lavora in un ambiente plurilingue).

Un’alternativa futura per il problema della ricerca dei dati nel database può essere rappresentata dall’utilizzo di tecnologie emergenti, come *data mining* (estrazione di dati) o *knowledge-based systems* (sistemi basati sulla conoscenza), il CERIF Working Group ha deciso di seguire l’evoluzione di queste tecniche per un possibile utilizzo futuro.

knowledge-based systems: la tecnologia KBS si basa su agenti intelligenti che interagiscono con ontologie di dominio. Un’ontologia di dominio è come un thesaurus, ma include relazioni semantiche tra termini che non sono presenti nei thesauri tradizionali. I thesauri tradizionali si limitano fondamentalmente a sinonimi e termini gerarchicamente più alti e più bassi, le ontologie, invece, possono contenere anche altre relazioni, ad esempio relazioni causali e temporali. Tramite l’utilizzo di queste tecnologie, una ricerca può essere ampliata o ristretta automaticamente per recuperare record attinenti secondo una soglia definita dall’utente. Gli agenti intelligenti possono inoltre aiutare nell’interfaccia con l’utente, sia per la ricerca che per l’inserimento/aggiornamento dei dati. In questo modo, con l’utilizzo degli agenti intelligenti, si possono impiegare dinamiche di ontologie di dominio, superando perciò alcuni dei problemi che si presentavano con la manutenzione dei sistemi e dei thesauri. Ci sono anche tecniche per la produzione automatica di thesauri e ontologie, ma è comunque richiesto un intervento manuale ed un controllo (una sorta di funzione “redattore”), durante queste operazioni si può però ricorrere all’assistenza degli agenti per minimizzare i compiti di inserimento e amministrazione dei dati.

Data mining: questa tecnica consiste nell'analisi dei dati e nell'uso di tecniche e software per trovare modelli e regolarità in un set di dati, in pratica il computer sviluppa i modelli tramite un'identificazione delle regole fondamentali e delle caratteristiche dei dati. Il data mining rappresenterebbe quindi un attrezzo utile per sviluppare un'indicizzazione a generazione automatica, basata però su una prima fase di raggruppamento di termini e notazioni da vocabolari controllati in raccolte di dati compatibili con CERIF. La tecnologia data mining, per trovare i dati richiesti, non effettua una scansione completa della base di dati, in quanto questa operazione impiegherebbe una grande quantità di tempo e potrebbe anche produrre risultati non esatti; l'operazione svolta è invece una combinazione di tecniche di estrazione di dati con una comprensione completa dei problemi di business. Alla fine i risultati vengono presentati in un modo facilmente comprensibile per l'utente.

1.1.4 Approccio plurilingue (reperimento di informazioni a lingue incrociate)

La fornitura di un supporto plurilingue nelle consultazioni è molto importante nei CRIS europei, infatti esiste una forte necessità di trovare modi efficaci per trattare il reperimento di informazioni a lingue incrociate, ovvero l'abilità di pubblicare una consultazione in una lingua e ricevere informazioni in un'altra.

Esistono due approcci principali per accedere al contenuto dei sistemi di informazioni: il primo approccio permette all'utente di formulare le consultazioni ricercando i termini da un vocabolario controllato (vocabolario predefinito, thesaurus o sistema di classificazione), il secondo approccio è noto come ricerca libera, e permette all'utente di formulare consultazioni liberamente, in altre parole gli è permesso ricercare i termini da lui ritenuti più appropriati, successivamente viene effettuata una scansione sull'intero corpo di tutti i documenti per cercare i termini richiesti.

Anche se l'uso di un vocabolario controllato aiuta l'utente a formulare consultazioni più precise, questo approccio può provocare restrizioni in alcune espressioni quando il vocabolario controllato che si è selezionato non è abbastanza completo o preciso. Per questa ragione, molti sistemi di ricerca offrono la possibilità di effettuare una ricerca libera, spesso abbinandola all'uso di un vocabolario controllato. Nell'approccio che utilizza il vocabolario controllato, la consultazione a lingue incrociate può essere sostenuta dal mantenimento di tabelle nelle quali sono contenuti tutti i termini, insieme alle loro traduzioni.

I sistemi di ricerca libera a lingue incrociate necessitano invece di una traduzione dei termini inseriti dall'utente in tutte le lingue di interesse. L'utente introduce i termini di ricerca in una determinata lingua, poi viene consultato un dizionario plurilingue per tutte le altre lingue disponibili in modo da tradurre i termini introdotti nei termini equivalenti.

In CERIF 2000, il modello dati ha una struttura che permette le ricerche a lingue incrociate all'interno dei vari CRIS.

1.1.5 Utilizzo dei metadata per integrare il modello di dati

Comunemente i metadata vengono definiti come “dati sui dati”, ovvero informazioni strutturate e scandite in campi, relative ad un documento di cui permettono una più efficiente organizzazione e recupero. Questo concetto non è nuovo e viene usato ogni volta in cui si devono organizzare, recuperare o scambiare degli oggetti contenenti informazioni. In pratica, un sistema di metadata è un sistema di catalogazione manuale di librerie, e serve per organizzare e migliorare la ricerca dei testi. L’immagazzinamento delle informazioni di catalogazione può essere utile, ad esempio, per descrivere i beni disponibili in un negozio, le parti presenti in un magazzino, i pezzi contenuti in un museo o le vacanze offerte da una società di viaggi.

Ci sono tre categorie principali di metadata:

- a) metadata di schema (controllano la struttura del database);
- b) metadata di navigazione (offrono accesso ai dati);
- c) metadata associativi (offrono informazioni supplementari e utili riguardo i dati), a loro volta sono suddivisi in:
 - c-1) descrittivi (descrivono i dati);
 - c-2) restrittivi (controllano l’accesso ai dati);
 - c-3) condizionali (condizioni sotto le quali possono essere usati o si può avere accesso ai dati).

I metadata permettono di effettuare consultazioni assistite e interpretazioni intelligenti dei relativi risultati, offrono inoltre un controllo della qualità dei dati in ingresso, permettendo ai sistemi di scambiarsi informazioni o di partecipare a consultazioni globali su sistemi distribuiti ed eterogenei.

I “Data Dictionaries”, conformi allo standard IRDS (Information Resource Dictionary System), sono stati il primo tentativo di implementazione di un standard di metadata in una comunità di database; adesso, invece, ce ne sono molti in uso, e sono usati comunemente per lo scambio di dati nelle comunità scientifiche.

L’accesso universale fornito da WWW ha poi accentuato il bisogno di utilizzare i metadata nel web, infatti la ricchezza di fonti di informazioni presenti sul web, unita alla loro semplicità di accesso, ha dato agli utenti un assaggio di quello che può essere realizzato. Bisogna tener conto però che questa implementazione si è rivelata in alcuni casi frustrante, nel caso in cui gli utenti si trovano di fronte a risultati di ricerche generalizzate, e quindi viene richiesto loro un ulteriore lavoro per arrivare a trovare i dati a cui sono interessati. A questo proposito, il consorzio W3C ha sviluppato uno standard chiamato RDF (Resource Description Framework) che utilizza XML (Extensible Mark-up Language). Questi sono standard generalizzati che adattano la loro estensione per coprire particolari domini.

Nel dominio delle informazioni di ricerca, il CERIF originale (1991) è stato usato come un modello di scambio di dati, ma si sarebbe potuto considerare ugualmente come se fosse stato un modello di metadata. L’applicazione di tali modelli ha condotto alla realizzazione di vari progetti pilota, tra cui il progetto svedese SAFARI e quello della Commissione Europea ERGO. Anche se questi progetti sono ancora in corso, stanno iniziando ad emergere alcune direttive su come i modelli di metadata dovrebbero essere raffinati ed estesi:

1. i modelli di metadata dovrebbero essere definiti ufficialmente con un definizione precisa di sintassi e semantica. Il modo relativamente “informale” in cui furono

definiti alcuni dei primi lavori a riguardo induce alla considerazione secondo cui l'implementazione è aperta a differenti interpretazioni, con una conseguente mancanza di omogeneità, e una ridotta possibilità di interoperazione tra i vari lavori.

2. in sistemi come CERIF, il modello di dati e il modello di metadata non dovrebbero essere considerati come mutuamente esclusivi. Nei CRIS invece non è consigliabile utilizzare i metadata, perché la loro struttura è molto complessa, con semantiche e modelli di dati precisi, utilizzando perciò i modelli di metadata generalizzati (sviluppati per entità meno strutturate, come le pagina web) si diminuirebbe di molto la ricchezza dei contenuti del CRIS originale.

1.2 Indicizzazione per soggetto

Prima si è parlato di indicizzazione per soggetto, ma di cosa si tratta di preciso?

Spesso di un documento è interessante conoscere il soggetto e i concetti che vengono rappresentati all'interno di esso. Il processo di indicizzazione per soggetto (subject indexing), consiste in un'analisi ragionata del contenuto del documento al fine di identificare i concetti in esso presenti, ed assegnare loro una qualche etichetta testuale, per permettere successivamente un rapido ritrovamento dell'informazione per un uso futuro.

Come già detto, l'indicizzazione tradizionale fa uso di un vocabolario controllato, ovvero un elenco fissato di termini. L'uso di un vocabolario controllato dà grandi vantaggi nel reperimento delle informazioni su Internet in quanto abilita l'ampliamento e il restringimento delle ricerche, semplifica l'accesso plurilingue alle raccolte di informazioni e permette la suddivisione e la manipolazione dei grandi database. Esistono due diversi tipi di vocabolario controllato: classificazione e thesaurus.

Con classificazione si intende un linguaggio informativo che usa termini artificiali (codici o notazioni). La maggior parte delle classificazioni ha una chiara struttura gerarchica con relazioni gerarchiche tra le notazioni; questa struttura è formata da classi che rappresentano una raccolta di "soggetti" con una caratteristica comune. Per alcune di esse esiste poi una sottoclasse, ovvero una parte della classe determinata da una caratteristica supplementare. In un sistema di codificazione le classi e le sottoclassi sono rappresentate da codici composti da simboli (lettere e/o numeri); ognuno di questi codici è unico e indipendente dal significato (la descrizione della classe associata), si deve invece riferire alla struttura dello schema di classificazione.

Un thesaurus invece è strumento per il controllo della terminologia mono o plurilingue e viene usato per la traduzione dei documenti dalla lingua originale in un'altra, in pratica è un vocabolario controllato dinamico che contiene termini che coprono specifici domini di conoscenza. La più importante funzione di un thesaurus è fornire un aiuto nel reperimento delle informazioni.

L'utilizzo dell'indicizzazione per soggetto in CERIF 91 copriva solamente le aree di ricerca, con l'avvento di CERIF 2000 il suo utilizzo si è allargato per includere anche i risultati delle ricerche, le informazioni sulle organizzazioni e quelle sulle persone. Questo allargamento ha richiesto quindi tre tipi diversi di indicizzazione per soggetto: uno basato sui soggetti delle ricerche, uno sulle attività economiche e un altro sui tipi di prodotto (prodotto inteso come risultato di una ricerca).

Nell'indicizzazione per soggetto basata sulle attività economiche e sui tipi di prodotto si è molto aiutati dal modello di dati di CERIF, che per queste informazioni fornisce un set predefinito di valori possibili, questi valori appartengono allo standard NACE (Nomenclature statistique des Activités économiques de la Communauté Européenne). Per quella basata sui soggetti delle ricerche è invece richiesto uno strumento apposito.

1.2.1 Ortelius Thesaurus

Per trovare lo strumento adatto all'indicizzazione per soggetto in CERIF è stato fatto un confronto tra i vari schemi di classificazione e i thesauri esistenti, in questo confronto i criteri di valutazione più importanti sono stati: una struttura gerarchica, la disponibilità elettronica, la disponibilità plurilingue, l'affidabilità della procedura di aggiornamento, la copertura del dominio di ricerca ed un possibile utilizzo internazionale/europeo.

Il sistema che è stato scelto per CERIF è stato "Ortelius Thesaurus", ovvero un thesaurus plurilingue in undici lingue (le undici lingue ufficiali dell'Unione Europea) che consente l'indicizzazione e la ricerca; questo sostiene inoltre un sistema di informazioni europeo sull'istruzione superiore. La scelta è caduta su Ortelius grazie alle garanzie offerte dai propri responsabili sul suo continuo aggiornamento, il suo miglioramento degli orientamenti riguardo la terminologia di R&S ed una situazione chiara riguardo i diritti degli utenti di Ortelius nell'ambiente di CERIF. Riguardo i criteri di valutazione sopra citati, Ortelius ha dato ottimi risultati per la sua disponibilità plurilingue e ha dimostrato di fornire una copertura estesa delle aree di ricerca grazie al suo utilizzo nell'istruzione superiore. L'interfaccia utente offerta da Ortelius è ottima; infatti vengono aggiunte alla sua struttura gerarchica caratteristiche specifiche di un thesaurus, come i riferimenti ai termini, che aiutano l'utente a spostarsi tra le varie aree di ricerca.

Ortelius Thesaurus è stato sviluppato per indicizzare e recuperare informazioni in "Ortelius Database", in particolar modo quando si tratta di informazioni riguardanti i programmi di studio. Quindi, vista l'analogia tra le aree di studio e le aree di ricerca nelle università e negli altri istituti per l'istruzione superiore, lo schema di classificazione scientifica di CERIF è stato preso come punto iniziale per lo sviluppo di Ortelius Thesaurus.

Da quando questo thesaurus è in uso si sta cercando un accordo sulla cooperazione tra i vari utenti interessati per garantire un costante aggiornamento dei termini. Il CERIF Working Group ha proposto una procedura di aggiornamento che ne prevede almeno uno ogni anno, perciò ogni utente che senta la necessità di effettuare una modifica ai termini o alle loro relazioni, oppure di aggiungere termini supplementari, ha la possibilità di spedire la propria richiesta ad un segretariato del CERIF centrale. Ad intervalli di tempo regolari queste richieste di modifica vengono controllate da un piccolo gruppo di esperti e, se ritenute valide, possono essere proposte ai gestori di Ortelius. Questa procedura include anche eventuali proposte per la traduzione dei termini in altre lingue.

1.3 Gruppi di utenti interessati

Viene ora fatta un'analisi sugli utenti che potrebbero essere interessati a questo formato comune per mantenere le informazioni di R&S. Gli utenti sono stati divisi in 7 gruppi, per ognuno dei quali viene definito in che modo essi utilizzano o potrebbero utilizzare un CRIS e che vantaggi avrebbero con l'introduzione di CERIF.

- 1. Fornitori del set di dati di CRIS:** ovvero i raccoglitori e gli editori di informazioni di ricerca (ad esempio autorità locali o nazionali); essi offrono un CRIS in quanto puntano ad aumentare la consapevolezza, l'uso e l'interesse in esso come uno strumento per comunicare le informazioni di R&S.
Per loro CERIF può essere utile in quanto gli permette di scambiare informazioni con altri CRIS in tutta Europa, inoltre, i nuovi fornitori di CRIS, oppure le organizzazioni che vogliono espandere i loro sistemi, possono trovare utili orientamenti su cui basare il loro sistema di informazioni di ricerca.
- 2. Fornitori delle informazioni contenute nei CRIS:** comprendono i centri di ricerca universitari e i reparti di ricerca all'interno di società. Questo gruppo di utenti è molto importante in quanto rappresenta il primo passo nel flusso delle informazioni di ricerca. Questi utenti usano i CRIS come un attrezzo di comunicazione, ma anche come uno strumento promozionale per le loro attività e i risultati da essi ottenuti, inoltre questi fornitori potrebbero essere anche utenti finali che cercano di trarre profitto dalle informazioni pubblicate.
CERIF 2000 utilizza una struttura di dati standardizzata e vocabolari controllati standard che facilitano la ricerca delle informazioni, per cui potrebbe essere molto utile ai ricercatori che vogliono diffondere i loro progetti.
- 3. Istituzioni:** questa categoria è costituita dalle persone attive in politica, in particolare nel contesto dell'innovazione e nella R&S a livello regionale, nazionale o europeo, questo gruppo include anche fondazioni scientifiche e di ricerca.
Essi usano i CRIS per decidere le priorità, gli obiettivi, il bilancio e i programmi nel campo dell'innovazione e della R&S. Questi sistemi possono essere loro utili per evitare duplicazioni di attività di ricerca, analizzare trend di ricerca e fornire ai cittadini esempi concreti dei risultati delle ricerche svolte.
Le istituzioni effettuano comunemente analisi statistiche sulle attività di ricerca, i CRIS possono essere utili anche in questo caso in quanto immagazzinano informazioni di ricerca in un modo tale da permettere la generazione di queste analisi, anche se CERIF non tratta il bisogno statistico come una priorità.
Con la creazione di CERIF, le istituzioni dispongono di una copertura maggiore sulle informazioni di ricerca, inoltre CERIF offre la possibilità di comparare informazioni provenienti da varie fonti. Per le istituzioni, questa veduta d'insieme, è utile nel caso di indagini riguardo le attività in una certa area di ricerca, relative ad un certo settore o un'area geografica.
- 4. Comunità di R&S:** ovvero le persone che sono coinvolte attivamente nelle attività di R&S, come ricercatori o direttori di un progetto. La comunità di R&S include: università (dipartimenti di ricerca e ricercatori individuali), centri di ricerca in

istituzioni regionali, nazionali ed internazionali, consigli di ricerca e centri di ricerca all'interno di industrie.

Per queste comunità CERIF è uno strumento che può essere utilizzato per: evitare duplicazioni di ricerche, identificare esperti per uno scambio di idee o per una collaborazione, identificare informazioni sui risultati di una ricerca in fase di sviluppo o completata, localizzare attrezzature e servizi, far conoscere le proprie attività di ricerca e comunicare i risultati ottenuti alla comunità di utenti che utilizza i CRIS.

5. **Organizzazioni intermediarie:** ossia ciascuna organizzazione che offre assistenza e sostegno nel campo dell'innovazione, della ricerca o dello sviluppo tecnologico. Esempi di organizzazioni intermediarie sono: servizi consultivi sostenuti dall'UE e le agenzie da loro coordinate, centri consultivi di imprese nazionali, servizi consultivi indipendenti e semi-indipendenti per imprese (camere di commercio ed industria, corpi associativi, agenzie di sviluppo regionali o locali), organizzazioni finanziarie di imprese.

Gli intermediari utilizzano frequentemente i CRIS per assistere i loro clienti, infatti è molto utile per loro avere una veduta d'insieme sui progetti in corso o già terminati, con lo scopo di produrre rapporti di sintesi per i loro clienti. I CRIS facilitano loro il contatto giorno per giorno con le imprese e la ricerca di informazioni pratiche e attinenti (sulle ricerche in corso e quelle completate, sui risultati di ricerca sfruttabili, sui potenziali partner e sugli esperti nei vari settori).

CERIF 2000 può aiutare queste organizzazioni allargando la copertura delle informazioni di ricerca disponibili e aggiungendo nuovi tipi di informazioni, il che offre grandi benefici pratici ad esse. Inoltre, quando CERIF 2000 verrà utilizzato da più CRIS, gli intermediari riusciranno a recuperare informazioni comparabili da fonti diverse.

6. **Imprese:** queste rappresentano la fine dei processi di ricerca e di innovazione, infatti comprendono le persone e le organizzazioni che trasformeranno le nuove idee in attività finanziarie vincenti, generando con ciò crescita economica e lavoro. Tutte le imprese, non solo quelle riguardanti la R&S, fanno parte di questo gruppo.

I CRIS possono aiutare queste imprese nei seguenti modi: informandole di quello che sta accadendo nel loro settore e nel mondo di affari in generale, promuovendo e localizzando le tecnologie trasferibili, identificando le fonti di consolidamento, identificando i potenziali partner e gli esperti presenti in tutta Europa e trasformando gli sforzi di ricerca in prodotti finiti.

La maggior parte delle ditte ha un campo di interesse relativamente stretto, di conseguenza dovrebbero riuscire ad effettuare consultazioni di informazioni concentrandosi solamente sui loro interessi, perciò per loro è di vitale importanza avere un sistema di classificazione semplice da usare e che riesca a "filtrare" le sole informazioni di interesse. A questo proposito, CERIF è un'ottima soluzione in quanto facilita le ricerche di informazioni offrendo tool per l'indicizzazione per soggetto (thesauri per la ricerca per soggetto, codici comunemente usati per l'indicazione di settori industriali o possibili domande di mercato, ecc.).

Dal canto suo, CERIF deve attirare e trattenere l'interesse delle imprese offrendo loro vantaggi pratici, come:

- facilità di utilizzo, non deve cioè richiedere nessun addestramento o attitudine speciale, deve quindi condurre in modo rapido, affidabile e diretto l'impresa che lo utilizza alle informazioni di cui ha bisogno;
- grande contenuto di informazioni relative al business, dovrebbe inoltre presentare tutte le informazioni contenute in una prospettiva di business;
- mentalità servizievole, ovvero dovrebbe tentare di capire e venire incontro alle necessità degli utenti;
- informazioni aggiornate e comprensibili.

7. **I media**, ovvero le persone che lavorano nel settore delle pubblicazioni, siano esse stampate, audio-video o virtuali, locali, nazionali o internazionali.

I media sono utenti potenziali dei CRIS, che possono essere visti come la loro fonte per trovare informazioni accurate e aggiornate su ricerche in corso, su nuove scoperte o su esperti da intervistare. Per i giornalisti CERIF 2000 sarebbe molto utile in quanto contiene informazioni molto dettagliate; inoltre lo schema di classificazione e il sistema di ricerca facilitata risultano essere molto comodi.

Nell'analisi dei rapporti degli utenti con i CRIS e con CERIF, questi sono stati suddivisi in vari gruppi. Adesso invece verranno considerate le esigenze comuni a tutti i gruppi definiti in precedenza:

- informazioni comprensive ed accurate, affidabili e di qualità;
- facilità di reperire le informazioni (interfaccia amichevole);
- fornire le informazioni nella propria lingua o in inglese;
- uso di indicizzazione per soggetto per arrivare rapidamente alle informazioni richieste;
- guida in linea/chiarimenti riguardo l'uso dell'indicizzazione per soggetto;
- protezione della privacy;
- nessun problema relativo al copyright nell'uso dei documenti e orientamenti chiari riguardo alle limitazioni per l'uso di dati;
- informazioni accurate per contattare persone, organizzazioni o accedere a documenti;
- identificazione chiara dei risultati ottenuti e del tipo di collaborazione (se c'è stata);
- informazioni sui costi.

Tutte queste caratteristiche devono essere sviluppate in CERIF per farlo diventare uno strumento utile e comodo da usare per tutti gli utenti.

2 Analisi documenti

In questo capitolo vengono analizzati alcuni documenti riguardanti CERIF, questi sono stati scritti da due componenti del CERIF Task Group, ovvero Anne Asserson e Keith G Jeffery, e da uno dei massimi esperti a riguardo, ovvero Andrei Lopatenko.

Il primo documento parla del modello di dati di CERIF ed espone alcune estensioni proposte per il futuro, il secondo propone tre approcci per i sistemi di ricerca distribuiti, nel terzo documento è spiegato in dettaglio l'utilizzo dei vari tipi di metadata in CERIF e nei CRIS, mentre nell'ultimo vengono spiegate le operazioni che i CRIS devono fare per essere compatibili con il modello di CERIF.

2.1 "CERIF: Past, Present and Future: an Overview"

Il passato di CERIF è rappresentato da CERIF 1991. Uno dei maggiori problemi con questo modello di dati era il fatto che gli sviluppatori si sono focalizzati su una singola entità (i progetti), mentre erano in uso tre classi principali di sistemi: quelli che si concentravano sui progetti, quelli che si concentravano sulle persone e quelli che si concentravano sulle unità organizzative, per ognuno di questi le altre informazioni erano solamente attributi di queste entità. Questo creava vari problemi in quanto nei CRIS che si concentravano sui progetti non venivano registrate accuratamente le informazioni sui ricercatori (veniva registrato solo il leader); inoltre sorgevano problemi con le relazioni, per esempio non era possibile registrare un progetto A come sottoprogetto del progetto B. Si rese quindi necessaria una revisione del modello di dati, in particolare era necessario definire più entità (piuttosto che un'entità unica e i suoi attributi), inoltre doveva essere possibile stabilire relazioni tra queste entità che comprendevano 1:n, n:m e la ricorsione (auto-riferimento).

Per risolvere questi problemi, è stato proposto CERIF 2000, il cui modello di dati è composto da tre entità principali (progetto, persona e organizzazione) collegate attraverso relazioni di tipo n:m e con attributi di "ruolo" (es. ruolo della persona all'interno del progetto) e di "data/tempo" (date di inizio e di fine); è inoltre consentito l'auto-riferimento per ognuna di queste entità principali (es. la relazione tra due unità "progetto" o due unità "persona").

Questo modello di dati offre grande flessibilità e robustezza in quanto tutte le informazioni di un progetto sono immagazzinate in un'entità che ha relazioni con persone e unità organizzative. Questo è utile nel caso in cui una persona abbia preso parte a più progetti coprendo anche ruoli diversi. Un'altra caratteristica importante è rappresentata dal fatto che tutti gli attributi testuali hanno entità subordinate con varianti di lingua per permettere un'implementazione plurilingue pulita, flessibile ed estensibile.

Per CERIF 2000 sono stati proposti tre modelli di dati:

- a) modello di dati “CRIS completo”, che definisce entità, attributi e relazioni per un’ottima integrazione con i vari CRIS;
- b) modello di dati “CRIS esportabile”, che è composto da un sottoinsieme di (a) e favorisce uno scambio di dati tra i CRIS, capace anche di esportare/importare CERIF;
- c) modello di dati di “metadata”, ovvero un sottoinsieme di (b) che fornisce una descrizione dei contenuti di un CRIS in una forma leggibile da ciascun CRIS e capace di importare/esportare CERIF.

Nel presente la Commissione Europea ha presentato CERIF 2000, in seguito molti dei CRIS esistenti hanno interagito con esso, l’esperienza ha inoltre mostrato che lo sviluppo dei CRIS per la ricerca pubblica nelle università è comunemente ampiamente compatibile con CERIF 2000, anche se è stato sviluppato senza conoscerlo.

Andrei Lopatenko, dell’Università Tecnologica di Vienna (TUW) ha già indicato delle deficienze e ha proposto alcune variazioni al modello di CERIF 2000:

- 1) correzioni al modello di dati e agli schemi;
- 2) estensione di CERIF per un uso sociale del modello di dati;
- 3) modifica dei dizionari di CERIF (liste di termini validi) e associazione con attributi che hanno la capacità di elencare la lista di possibili valori validi.

Questi sviluppi hanno lo scopo di estendere la capacità e l’utilizzabilità di CERIF per sostenere i CRIS, con estensioni sia in profondità (dettaglio delle informazioni) che in larghezza (tipi di ricerche coperte).

Sono state proposte alcune estensioni per CERIF 2000, una caratteristica interessante è che queste estensioni richiedono veramente poche modifiche al CERIF originale, infatti i modelli di dati originali hanno già la capacità di soddisfare i cambiamenti richiesti.

Un’estensione importante è stata proposta dalla “Universitetet i Bergen” (UiB), qui si è espressa la necessità di riferire l’entità {result_publication} all’entità {person A}, la quale ha lavorato per l’entità {orgunit N}, e all’entità {person B}, che lavorava per l’entità {orgunit M}. In altre parole, si desidera riferire un particolare {result_publication} all’intersezione di {person} e {orgunit}. Questo richiede l’induzione che, se l’intervallo di dati (con ruolo adatto) nella relazione {person-result_publication} interseca l’intervallo di dati nella relazione {person-orgunit}, allora la persona stava lavorando per quella organizzazione al tempo della pubblicazione. Ovviamente, la persona avrebbe potuto lavorare per più di un’unità organizzativa, e molte persone avrebbero potuto lavorare per la stessa organizzazione e alla stessa pubblicazione.

Per ragioni di efficienza, alla UiB decisero di perfezionare questo concetto come una relazione ternaria {person-orgunit-result_publication} senza ruolo e date; questo metodo ha il vantaggio di ridurre le operazioni di congiungimento durante le fasi di una ricerca. Esiste però anche uno svantaggio, ovvero è difficile rappresentare le relazioni di ruolo e di tempo di una persona con una pubblicazione o una organizzazione, inoltre questo rende anche più difficile la gestione delle pubblicazioni multi-autore. In pratica questa estensione si è dimostrata inefficiente nella realizzazione.

Questo lavoro però fa luce su un problema di CERIF 2000; infatti l’idea originale era che le pubblicazioni fossero registrate fuori di CERIF (si doveva tenere solamente

un puntatore alla pubblicazione). È chiaro che questo è insufficiente e così è stata proposta l'aggiunta di questa caratteristica allo standard di CERIF 2000.

Un'altra estensione importante è stata proposta dal "Council for the Central Laboratory of the Research Councils" (CCLRC), questa estensione è stata pensata in quanto il CERIF originale non sosteneva abbastanza il business relativo alla R&S, i cambiamenti al modello di dati a cui si è pensato sono:

- {project}: aggiunta di ulteriori informazioni, come il piano del progetto e i costi;
- {person}: notevole aggiunta di informazioni, tra cui la valutazione della performance annuale (obiettivi di lavoro e loro conseguimento, necessità di apprendimento e sviluppo e loro conseguimento), aggiunta delle posizioni passate all'interno dell'organizzazione, del Curriculum Vitae, delle abilità e delle competenze di una persona;
- {orgunit}: aggiunta delle informazioni finanziarie (es. bilancio annuale, fatture, ordini), di informazioni sugli obiettivi e sui termini di referenza di un'unità organizzativa.

Nel futuro CERIF ha ancora molto da offrire ai responsabili dei vari CRIS, fornendo un modello di dati per l'importazione/esportazione di dati da un CRIS verso altri sistemi. CERIF offre un modello affidabile e formalizzato, e sembra soddisfare completamente i requisiti dei CRIS; inoltre può essere esteso per altri scopi, incluso un modello dati relativo al business. CERIF inoltre può essere usato per lo scambio di dati tra CRIS diversi: il CRIS A fornisce alcuni dati a CERIF, il quale li converte secondo il suo modello dati e li trasmette al CRIS B, questo li riceve nel formato di dati di CERIF e li immagazzina secondo il proprio formato; in questo modo sono pronti per la consultazione da parte degli utenti di CRIS B. Un altro utilizzo di CERIF è quello di accedere ai dati di un CRIS (costituisce un portale per accedere a tutti i CRIS). Il portale permette consultazioni in nella lingua desiderata e una traduzione di quella consultazione secondo il formato del CRIS.

2.2 "CERIF: Information Retrieval of Research Information in a Distributed Heterogeneous Environment"

Questo testo espone i requisiti per l'integrazione dei dati riguardanti la ricerca e descrive i risultati già realizzati in questo ambito, descrive poi quali approcci sono in corso di sviluppo da parte del CERIF Task Group, spiega inoltre le richieste degli utenti riguardo la qualità e la completezza dei dati e, infine, suggerisce soluzioni per consultazioni distribuite di dati.

Le necessità tipiche degli utenti dei CRIS non sono limitate ad un solo sistema di informazioni, ma sistemi diversi per quanto riguarda l'area di appartenenza delle informazioni o il paese da cui provengono. Questa ricerca di informazioni richiede la conoscenza del sistema dove queste sono immagazzinate e molti sforzi per visitare tutti i sistemi; inoltre potrebbero sorgere dei problemi nel caso in cui i sistemi non fossero compatibili tra loro, ovvero le informazioni provenienti da uno di essi non possono

essere riutilizzate in un altro. Sono stati sviluppati molti progetti per creare sistemi distribuiti per la ricerca di informazioni, tra questi sono da segnalare tre importanti approcci definiti dal CERIF Task Group.

a) Approccio Database Distribuito (CERIF-DD)

Lo scopo di questo approccio è quello di offrire una ricerca molto semplice da sviluppare che sia però efficiente su database multipli. L'applicazione migliore di questo approccio è l'unificazione dei database di ricerca che appartengono a rami diversi di un'organizzazione. Ogni database deve esportare un set di viste conformi ai requisiti di CERIF. Esiste poi, per ogni rete, un server centrale che registra tutti i database e fornisce all'utente finale i risultati di una ricerca complessiva su tutti questi. Quando un database esporta le proprie viste, questo deve essere registrato al server centrale della rete, con la descrizione della fonte dei dati, la qualità, l'aggiornamento, il settore di ricerca e le informazioni di accesso; dopo questa registrazione, il database sarà incluso nella ricerca complessiva tramite il server centrale. La sorgente dati dovrebbe essere registrata al server centrale attraverso l'interfaccia di SOAP (Simple Object Access Protocol), alla quale basta inviare solo la descrizione secondo lo standard RDF della sorgente di dati.

Attualmente in CERIF-DD sono possibili ricerche per le entità usate più frequentemente, come persone, progetti, unità organizzative, pubblicazioni e collegamenti.

- Vantaggi: l'operazione di ricerca troverà dati adeguati anche se descritti da vocabolari diversi, è possibile accedere direttamente ai dati ed utilizzare i tool di JDBC per la loro l'analisi e immagazzinamento.
- Svantaggi: possibili problemi di sicurezza derivati dall'accesso diretto al database (deve essere protetto), grandi richieste per la configurazione manuale.

b) Approccio Web Semantico (CERIF-SW)

Con il termine web semantico si intende la trasformazione del World Wide Web in un ambiente dove è possibile pubblicare non più solo documenti (pagine HTML, file, immagini, e così via) ma anche informazioni e dati relativi ai documenti stessi in un formato adatto all'interrogazione, all'interpretazione e, più in generale, all'elaborazione automatica. Così facendo sarà possibile un'interpretazione più significativa del contenuto dei documenti stessi, l'effettuazione di ricerche estremamente precise, e numerose altre operazioni innovative.

L'approccio a web semantico ha lo scopo di risolvere i problemi relativi alla ricerca di informazioni distribuite nei casi in cui: la struttura e la semantica sono diverse in fonti di dati diverse, sono richieste operazioni di ricerca di informazioni sofisticate, lo schema delle fonti di dati nuove o già esistenti può essere cambiato, devono essere usati dati di web semantici già pubblicati, non può essere fornito l'accesso diretto al database o quando è importante la compatibilità con altre architetture basate sul web semantico.

Una delle caratteristiche principali della soluzione a web semantico è la facile integrazione con fonti di dati molto eterogenee (biblioteche digitali, database). Quando si vogliono integrare nella rete sistemi con la struttura dei dati o il loro significato diversi da quelli di CERIF, l'ontologia di questi sistemi deve essere descritta nei termini di CERIF (se questo è possibile). Questo permette l'integrazione anche con fonti di dati che non utilizzano CERIF, ma che sono compatibili con esso (es. DublinCore, MathNet, ecc). Per fare questo però ogni fonte di dati deve pubblicare la sua ontologia sul web

ed esprimere i dati nel formato RDF, l'ontologia deve adattarsi al formato DAML (DARPA Agent Markup Language) e, per rendere disponibili i propri dati nelle ricerche tramite CERIF, deve convertire i termini locali in termini adottati da CERIF. In pratica, per rendere i dati accessibili nelle consultazioni di CERIF, questi devono essere trasformati in RDF (questo è possibile usando il toolkit fornito da CERIF); i dati RDF possono poi essere pubblicati sul web per essere raggiunti direttamente attraverso HTTP.

Un'altra caratteristica importante di questa soluzione è la possibilità di usare un sistema di reperimento delle informazioni intelligente e ontologicamente guidato, ovvero le stesse informazioni possono essere rappresentate nel sistema di informazioni in modi diversi, descritte da punti di vista differenti dovuti alla diversa provenienza dei dati, ad una politica di restrizioni o anche alla posizione dell'autore/compiler. Si noti che le vedute delle stesse informazioni possono essere anche molto diverse per categorie diverse di utenti.

- Vantaggi: reperimento delle informazioni intelligente e ontologicamente guidato, compatibilità con un set di progetti e attività emergenti, nessun problema di sicurezza in accessi diretti al database.
- Svantaggi: adesso molto inefficiente in quanto le operazioni di ricerca sui database richiedono il trasferimento di molti dati (non può fornire un accesso veloce), l'uso effettivo della tecnologia richiede grande conoscenza del web semantico e delle ontologie, infine l'immagazzinamento, l'analisi e le altre operazioni sui database non possono usare i dati dal web semantico direttamente.

c) Approccio Web Service (CERIF-WS)

Lo scopo di questo approccio è quello di risolvere i problemi relativi alla ricerca distribuita di informazioni, ovvero punta a: realizzare fonti di informazioni di ricerca compatibili con gli standard emergenti, offrire servizi web per la ricerca dei dati, impedire un accesso diretto al database (per motivi di sicurezza), utilizzare il formato XML/SGML (compatibile con i sistemi già sviluppati), offrire un protocollo di livello trasporto efficiente per CERIF-SW.

CERIF-WS consiste di uno schema XML basato sull'ontologia di CERIF e di un'implementazione SOAP di ogni fonte di dati, la quale offre funzioni di ricerca e recupero dati, questo approccio inoltre fornisce una ricerca facilitata che permette all'utente l'accesso ai dati di ricerca provenienti da qualsiasi sistema. I CRIS possono partecipare facilmente alla rete di CERIF (anche quelli che hanno uno schema di dati diverso) pubblicando, attraverso una "copertina", i loro dati in XML, questo è favorito dal fatto che CERIF-WS è stato costruito in modo da facilitare agli sviluppatori la creazione o la modifica della codifica XML dei loro dati nella copertina.

- Vantaggi: l'implementazione del web service per i CRIS è molto facile, i tool per sostenerlo sono descritti in molte lingue, la piattaforma è già disponibile, la soluzione dei web service per CERIF può utilizzare il web semantico per consultazioni sofisticate o per l'integrazione di fonti di dati eterogenee.
- Svantaggi: adesso questo tipo di soluzione non è molto efficiente rispetto alla ricerca di dati distribuita, servono risorse per insegnare l'utilizzo dei web service, i web service non sono ancora molto maturi, alcuni tool non sono efficienti o hanno molti bug, l'attuale implementazione di CERIF con i web service permette di usare solamente consultazioni molto semplici, è difficile

perfezionare consultazioni sofisticate che costringono a controllare le relazioni tra oggetti.

La fine di questo documento parla della qualità dei dati, i parametri principali per definirla sono la completezza, l'attualità e la correttezza. La completezza di dati di un CRIS corrisponde alla presenza di informazioni riguardanti tutte le entità che sono soggette a quel determinato CRIS, con attualità invece si intende la presenza di informazioni il più aggiornate possibili sul soggetto a cui si riferiscono, la correttezza infine indica se le informazioni sono corrette o meno.

2.3 “Comparative Study of Metadata for Scientific Information: The place of CERIF in CRISs and Scientific Repositories”

I metadata sono la chiave per realizzare l'interoperazione tra i vari CRIS e tra i CRIS e i depositi scientifici e bibliografici. I metadata sono divisi in tre categorie principali:

1. metadata di schema: sono una descrizione intensionale di un esempio estensionale. Tipicamente uno schema è composto da un database {nome, dimensione, autorizzazioni di sicurezza} e dagli attributi {nome, tipo, costrizioni}. Lo schema (intensione) ha una relazione logica formale con i dati (estensione), questo è importante per assicurare la qualità di essi. In pratica, i metadata di schema costringono i dati da essi descritti a mantenere la loro integrità.
2. metadata di navigazione: offrono informazioni su come arrivare ad una fonte di informazioni, i meccanismi includono: nome del file, nome del database + algoritmo di navigazione, nome del database + predicato (query), URL (Uniform Resource Locator), URL + predicato (query) o varie combinazioni di questi. I metadata di navigazione non hanno una relazione logica formale con i dati, ma gli attributi di questi metadata possono descrivere il nome della raccolta di dati (es. nome del file).
3. metadata associativi: offrono informazioni supplementari per fornire assistenza nelle applicazioni. L'assistenza può migliorare le performance, l'accuratezza o la precisione del sistema oppure aiutare l'utente finale attraverso un'interfaccia d'appoggio. Di solito non hanno una relazione logica formale con dati, anche se ci possono essere relazioni di associazione sistematiche. A loro volta si dividono in descrittivi (record di catalogo), restrittivi (stima di contenuto, sicurezza o privacy) e d'appoggio (dizionari, thesauri, hyperglossari).

Per stimare i diversi formati di metadata, viene fatta una valutazione su tre dimensioni: tipo (metadata di schema, navigazione o associativi), destinazione di uso, compatibilità con CERIF. La compatibilità con CERIF può essere di due tipi: 1) se il formato di metadata è compatibile e può essere sviluppata una soluzione interoperabile; 2) se il formato di metadata descrive entità diverse da quelle di CERIF, ma possono essere comunque usati per offrire servizi supplementari per i CRIS.

I **metadata di schema** nei CRIS possono essere usati per perfezionare le soluzioni di accesso ad informazioni distribuite, per integrare i dati, per aiutare l'utente a trovare CRIS attinenti ai suoi requisiti e per descrivere le risorse del web.

I **metadata di navigazione** nei CRIS sono utilizzati per descrivere l'ubicazione dei dati nei sistemi distribuiti e per fornire un accesso persistente ad essi. Questo meccanismo offre l'estensione primaria dei CRIS convenzionali per includere dati provenienti da depositi scientifici e bibliografici. CERIF usa gli URI per offrire metadata di navigazione che si riferiscono a fonti di informazioni più particolareggiate.

I **metadata associativi descrittivi** offrono in CERIF un contesto di informazioni formale per la considerazione di dati provenienti da depositi scientifici o bibliografici, riferendo i dati particolareggiati del deposito ad un modello canonico che descrive le informazioni di ricerca. A loro volta sono suddivisi in metadata per depositi scientifici e metadata per informazioni di ricerca.

I **metadata per depositi scientifici** descrivono dati scientifici o raccolte di dati scientifici e rappresentano oggetti scientifici o risultati di esperimenti scientifici, permettono inoltre lo scambio di questi dati tra le applicazioni, l'immagazzinamento in database e la pubblicazione tramite WWW. Nei CRIS possono essere usati per: estendere le ricerche degli utenti ad informazioni scientifiche e tecniche particolareggiate e fornire una descrizione precisa e formale dei soggetti delle informazioni di ricerca in modo da migliorare l'efficacia della ricerca stessa.

I **metadata per informazioni di ricerca** descrivono le fonti di informazioni (come depositi bibliografici) con entità simili a quelle descritte da CERIF (es. biblioteche digitali). La traduzione dei metadata di CERIF in altri schemi di metadata permette il riutilizzo dei dati di CERIF da parte di altri sistemi; questo può essere utile per attirare nuovi utenti e per fornire un'offerta di servizi supplementari. Prendendo invece le informazioni di ricerca espresse in altri formati di metadata, e traducendole nel formato di CERIF, è possibile caricare nuovi dati in CERIF ed effettuare uno scambio di dati tra i CRIS attraverso i metadata di CERIF.

I **metadata associativi restrittivi** restringono l'accesso degli utenti (o dei sistemi di software) alle fonti di dati dei CRIS, ne limitano l'utilizzo o forniscono servizi di sicurezza. CERIF si occupa di questi metadata mettendo costrizioni nelle relazioni che rappresentano il rapporto, per esempio, tra un autore e una pubblicazione o un utente e una pubblicazione. Questo offre molta più flessibilità rispetto ad un immagazzinamento degli attributi di associazione restrittivi con la pubblicazione o con la persona (sia essa autore o utente) siccome è facilmente mutevole ed estensibile.

I **metadata associativi d'appoggio** in CRIS sono usati per creare valori di dominio per i metadata descrittivi e offrono altri servizi di appoggio, includono inoltre dizionari, thesauri, hyperglossari, e ontologie di dominio. CERIF offre una struttura tale da consentire l'utilizzo di quasi tutti i metadata associativi d'appoggio. Nelle ricerche di informazioni viene utilizzato un sistema a termini controllati, questo migliora la precisione e consente il successo della ricerca anche nel caso di risultati in lingua straniera.

Nei CRIS viene utilizzato un altro tipo di metadata, ossia i **metadata per la valutazione della qualità**, questi servono per descrivere la qualità delle informazioni in modo da aiutare gli utenti a filtrare le risorse e scartare quelle a bassa qualità. La descrizione della qualità delle informazioni è creata dal fornitore delle informazioni o da un esperto.

Uno dei requisiti chiave per CERIF è la compatibilità con altri formati di metadata, per soddisfarlo si è basato su standard e formati molto usati (RDBMS, XML, DAML + OIL, RDF). Inoltre, il modello di CERIF2002, estende quello attuale e contiene definizioni di entità, di proprietà e di costrizioni comprensibili dalle macchine, questo è utile per rendere più automatica la compatibilità di CERIF con altri formati di metadata. Questo permette anche, automaticamente, di integrare i CRIS basati su CERIF con gli altri non compatibili con esso.

2.4 “CERIF Interoperability and Compatibility”

Lo scopo di questo documento è quello di offrire orientamenti per la compatibilità con CERIF ai CRIS nuovi e a quelli già esistenti. Ciascun CRIS che desidera realizzare questa interoperazione deve utilizzare il modello di dati completo proposto dal CERIF Task Group.

Il modello di dati completo di CERIF è formato dall'insieme normalizzato e regolarizzato delle entità e degli attributi scelti in modo da soddisfare i requisiti dei CRIS usati più comunemente, risulterà quindi minore dell'insieme completo delle entità e degli attributi di tutti i CRIS. Questo è anche chiamato “CERIF Recommended Model” in quanto è il modello di dati ideale per un CRIS.

Il modello di metadata di CERIF, chiamato anche “CERIF Core Model”, è definito per essere usato come una rappresentazione succinta dei contenuti di un CRIS, una specie di catalogo (non rappresenta un modello per lo scambio di dati). Questo modello è in parte un sottoinsieme del modello di dati completo di CERIF ed in parte un'astrazione. Il sottoinsieme è formato dalle entità primarie {Project}, {Person} e {OrgUnit} con le rispettive relazioni, dall'entità secondaria {Contact} con le sue relazioni e dalle entità relative alla lingua per gli attributi delle tre entità principali. L'astrazione invece è formata da tre entità: {ProjectAdditional}, {PersonAdditional} e {OrgUnitAdditional}, ognuna di queste consiste di attributi booleani con possibili valori [Y / N] che indicano se il CRIS descritto dal metadata possiede le entità e gli attributi che corrispondono all'attributo descritto in queste entità supplementari.

Se un CRIS desidera interoperare con CERIF, deve documentare e rendere disponibili il suo modello di dati completo, la parte del suo modello di dati disponibile per l'esportazione e il suo modello dati di metadata, deve inoltre accettare consultazioni ed esportare le informazioni che le soddisfanno, il tutto usando il modello di dati completo di CERIF, questo incoraggia i CRIS a modificare il loro modello di dati per renderlo uguale a quello. Per interoperare, un CRIS dovrà inoltre esporre un modello di metadata di CERIF con all'interno elementi derivati dalle sue informazioni immagazzinate.

Per realizzare la compatibilità con CERIF, un CRIS non è però obbligato ad immagazzinare le informazioni secondo il database di CERIF, può anche immagazzinarle secondo un modello diverso, deve però poi essere in grado di effettuare una conversione dal/al modello di CERIF per scambiare informazioni con gli altri CRIS e per fornire informazioni in risposta alle consultazioni.

3 Modello di dati di CERIF

In questo capitolo vengono descritte le entità e gli attributi che compongono il modello di dati proposto da CERIF, per i documenti ufficiali (aggiornati nel settembre 2002) si rimanda alla pagina web del CERIF Task Group (vedi Bibliografia) nella sezione “Downloadable files”.

Gli obiettivi principali del modello di dati di CERIF 2000 sono:

1. fornire un modello di dati flessibile, in modo da permettere alla maggioranza dei CRIS esistenti di adattare ad esso le loro strutture di database;
2. fornire una struttura di base per lo scambio dei dati.

Questi obiettivi non sono però compatibili al 100% con tutti i sistemi di informazioni, in termini di un solo modello dati. Per esempio, un modello dati che copre tutte le strutture dei database dei CRIS esistenti non sarebbe un modello molto pratico per lo scambio di dati, così come ogni CRIS non è interessato a tutte le parti di database di tutti gli altri CRIS. L’approccio che si è utilizzato punta a venire incontro ad entrambi gli obiettivi, ovvero:

- a) definire un modello di dati di CRIS completo che copra le strutture dei database della maggioranza dei CRIS esistenti (Full CRIS data model);
- b) definire un set di modelli di dati che potrebbero offrire esempi per lo scambio di dati (Exchange data model), questo è utile perché non rende necessario implementare il modello dati di CRIS completo se ne è richiesto solamente un particolare sottoinsieme;
- c) definire un modello di metadata per offrire una veduta uniforme sulle fonti di informazioni (Metadata data model).

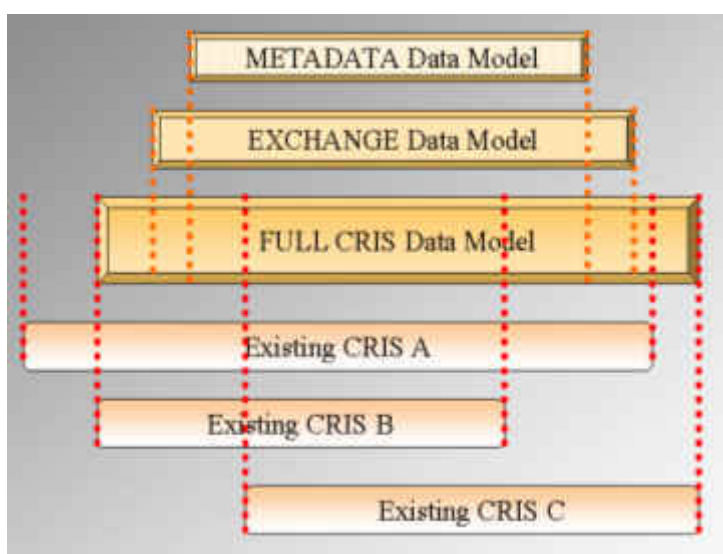


Figura 3.1: relazione fra i modelli di dati di CERIF ed i CRIS esistenti.

Dalla figura 3.1 si nota che il Metadata data model e il Exchange data model sono sottoinsiemi del Full CRIS data model, in particolare il Metadata data model è un sottoinsieme sia del Full CRIS data model che del Exchange data model. Si nota inoltre che in questo esempio il Full CRIS data model è formato dall'unione del CRIS B e del CRIS C, mentre una parte del CRIS A è fuori da questo modello, non è un problema perché, come già detto, un CRIS non è interessato a tutte le parti di database di tutti gli altri CRIS.

3.1 Full CRIS data model

Il Full CRIS data model è formato da molte entità suddivise in 5 livelli di astrazione a seconda dell'importanza delle informazioni contenute, questo per semplificare l'inserimento e la ricerca delle informazioni stesse, questi livelli sono:

- livello 1 (base tables): è formato dalle tre entità principali: Project, Person e OrgUnit (unità organizzative), queste entità contengono le informazioni di maggiore importanza in quanto da esse derivano poi tutte le informazioni secondarie, per questo motivo l'inserimento di questi dati è obbligatorio;
- livello 2 (secondary base): include entità secondarie, come i risultati e la classificazione di un progetto, i contatti e gli eventi in cui si parla di un progetto o a cui prendono parte le persone, l'inserimento delle informazioni relative a queste entità è caldamente consigliato;
- livello 3 (language-field base): include le varie traduzioni per molte delle entità comprese nel livello 2, l'inserimento di queste informazioni è utile nel caso in cui si vogliono rendere disponibili i propri dati durante ricerche in altre lingue;
- livello 4 (lookup tables): include entità che forniscono supporto riguardo gli attributi con un set predefinito di possibili valori, per esempio l'attributo "Titolo Accademico" (che può contenere solo i valori Dott., Ing, Prof., ecc.), queste entità sono utili per facilitare l'inserimento di alcuni valori ed evitare errori.
- livello 5 (link tables): include le entità intermedie necessarie a rappresentare le relazioni "molti-a-molti".

Questa suddivisione in livelli facilita l'inserimento delle informazioni in quanto lascia molta libertà riguardo alla quantità, al tipo e al livello di dettaglio dei dati che si vogliono inserire; infatti oltre alle entità principali obbligatorie, le altre sono tutte facoltative, è ovvio però che per fornire una buona spiegazione del lavoro svolto sarebbe meglio completare tutte le entità presenti. È da notare inoltre che, per ogni livello di astrazione, esistono entità diverse relative alle persone, ai progetti e alle unità organizzative; questo lascia un'ampia scelta, a chi vuole pubblicare le informazioni, sulla o sulle entità a cui dare più importanza, ovvero si potrebbero descrivere dettagliatamente i progetti (es. arrivare alle entità di livello 4) e solo accennare alle persone e alle unità organizzative che vi hanno preso parte (es. completare solo le entità di livello 1).

Il modello di dati forma una "maschera" dalla quale le implementazioni dei vari CRIS possono scegliere le entità e gli attributi da mostrare a seconda dei loro interessi,

in pratica ogni particolare CRIS è un sottoinsieme del modello di dati completo. Alcuni database specializzati in R&S possono avere anche alcune entità e/o attributi supplementari, ma una grande percentuale di modelli di dati dovrebbero intersecarsi con il Full CRIS data model. Procedo adesso ad un'analisi dettagliata delle entità che formano i vari livelli del modello di dati.

3.1.1 Livello 1 - Entità principali

Come già detto le entità di maggiore importanza sono Project, Person e OrgUnit, tutte le altre sono associate ad una di queste entità principali. Le relazioni tra di loro sono di tipo “multi-a-molti”, questo vuole dire che ad un progetto possono aver collaborato molte persone e molte unità organizzative; lo stesso vale per le persone e per le unità organizzative, le quali possono aver partecipato a più progetti. Queste relazioni “multi-a-molti” saranno realizzate tramite entità intermedie che comprendono anche attributi di ruolo e di intervallo di tempo (se ne parlerà più specificatamente nel livello 5).

Ci possono essere inoltre relazioni tra due elementi persona, tra due elementi progetto e tra due elementi unità organizzativa, sempre con l'aggiunta degli attributi ruolo e intervallo di tempo. Questo vuole dire che nel modello di dati può essere evidenziata, per esempio, la cooperazione, per un certo periodo di tempo, tra la persona A, con il ruolo di “project manager”, e la persona B, con il ruolo di “programmatore”. Un altro esempio di utilizzo di queste relazioni potrebbe essere rappresentato dalla relazione tra un progetto X e i suoi sottoprogetti Y e Z, o tra il progetto X e il progetto associato W. Nel caso delle unità organizzative, possono essere messe in evidenza le relazioni tra i vari reparti di esse oppure tra un'unità e le sue sotto-unità.

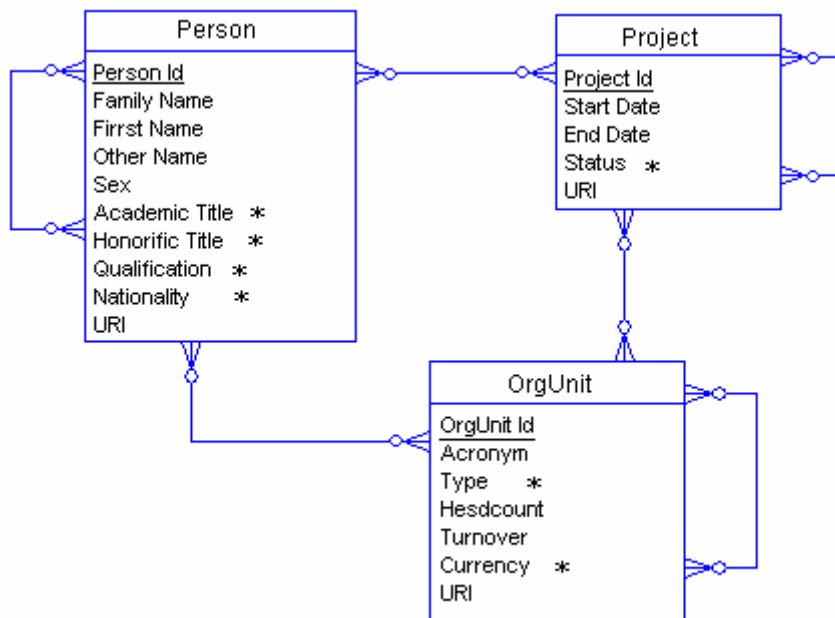


Figura 3.2: entità principali.

Nella figura 3.2 sono rappresentate le entità principali, gli attributi sottolineati rappresentano le chiavi primarie delle relative entità e sono quindi obbligatori, mentre gli attributi con l'asterisco (*) hanno un set predefinito di possibili valori, quindi saranno sviluppate tramite le tabelle del livello 4. Come già detto le relazioni tra le entità principali sono di tipo "multi-a-molti", nel disegno il "molti" è rappresentato dalla ramificazione del collegamento.

Vado adesso ad analizzare in dettaglio gli attributi delle tre entità principali:

- a) Person: contiene attributi riguardanti i dati anagrafici (nome, cognome, sesso, nazionalità), le qualifiche (titolo accademico, eventuale titolo onorario e appunto qualifiche) e l'home page di un individuo (URI, Uniform Resource Identifier), l'unico attributo obbligatorio è il cognome (Family Name). Per descrivere le qualifiche (Qualification) di una persona si utilizza lo standard ISCED (International Standard Classification of Education) dell'UNESCO;
- b) OrgUnit: contiene l'acronimo (Acronym), il tipo di unità organizzativa (Type), il numero di dipendenti (Headcount), il budget annuale (Turnover), un codice "ISO 4217" di 4 lettere che indica la valuta in cui è indicato il budget (Currency) e l'home page (URI), l'unico attributo obbligatorio è il tipo.
- c) Project: contiene le date di inizio e di fine (se il progetto è stato finito), lo stato del progetto (completato, in corso, in fase di stallo,...) e l'home page (URI), sono obbligatori solamente la data di inizio e lo stato.

Analizzando le entità principali si nota una sostanziale differenza tra Person, OrgUnit e Project; infatti nelle prime due, gli attributi disponibili forniscono una buona spiegazione di una persona o di un'unità organizzativa, mentre gli attributi di Project non spiegano niente riguardo il progetto. È quindi ovvio che, nel descrivere un progetto, non ci si può fermare al primo livello, ma si deve procedere a spiegazioni più dettagliate. Nella descrizione di una persona o di un'unità organizzativa, invece, ci si potrebbe anche fermare al primo livello in quanto si forniscono già un buon numero di informazioni, è però consigliabile andare avanti con gli altri livelli per fornire le informazioni con un maggiore dettaglio.

3.1.2 Livello 2 - Entità secondarie

Il livello 2 del modello di dati include quelle che sono definite come entità secondarie; queste sono le entità associate direttamente a quelle principali, ed hanno quindi una grande importanza, in particolar modo per quanto riguarda l'entità Project dal momento che, come detto prima, in questa viene spiegato veramente poco del progetto.

Alcune entità appartenenti a questo livello sono associate a due delle entità principali, ed altre a tutte e tre, per realizzare i collegamenti tra queste e le entità principali si utilizzano tabelle intermedie che appartengono al livello 5. Nelle entità di questo livello non vengono approfondite le relazioni "multi-a-molti" (se ne parlerà più specificatamente nella parte relativa alle entità di livello 5); infatti anche nei disegni non vengono indicate le relative entità intermedie. Anche in questo caso nei disegni ho indicato con l'asterisco gli attributi con un set predefinito di possibili valori, questi saranno contenuti nelle tabelle del livello 4.

a) Inizio parlando delle entità associate a Person:

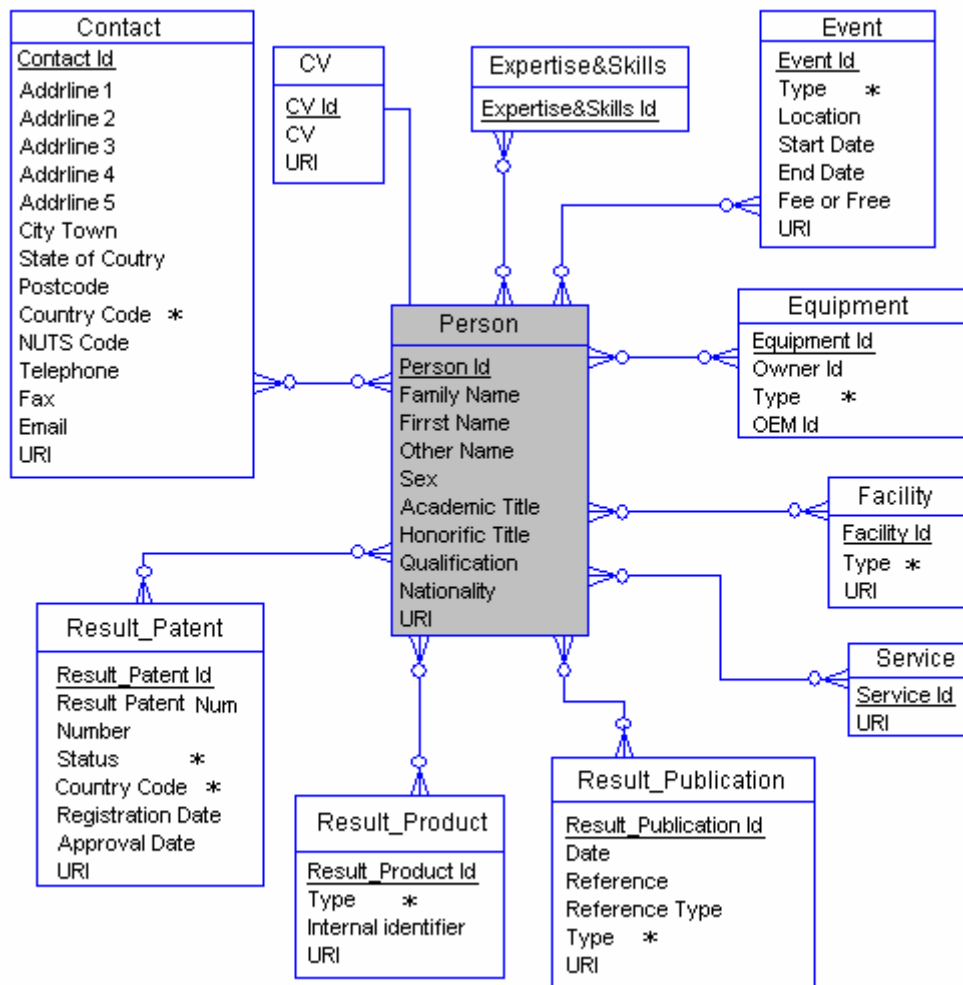


Figura 3.3: entità secondarie relative a Person.

Contact contiene l'indirizzo dettagliato di una persona che comprende la città e lo stato di residenza, il codice postale, un codice di 4 lettere per indicare la nazione (Country Code) e un codice NUTS (codice indicativo di regione), sono presenti inoltre i numeri di telefono e di fax, l'indirizzo e-mail e la home page. Gli attributi obbligatori sono la città, il codice postale e il codice della nazione. Questi dati sono inseriti in un'entità associata e non in quella principale per gestire indirizzi multipli di un individuo (es. ufficio e laboratorio) o un indirizzo a cui lavorano più persone, per questo motivo la relazione con Person è del tipo "multi-a-molti". Questa entità è di grande importanza perché permette agli utenti di CERIF di che hanno svolto un progetto di essere contattati.

Event contiene i congressi e i meeting a cui hanno partecipato o parteciperanno le persone indicate, sono specificati il tipo e il luogo dell'evento, le date di inizio e di fine, se questo è gratuito o a pagamento (Fee or free) e l'eventuale home page. Gli attributi obbligatori sono il tipo di evento e l'home page. Questa entità è associata anche a

Project e a OrgUnit perché una persona, che potrebbe appartenere ad una certa unità organizzativa, durante un evento potrebbe parlare di un progetto da lui realizzato.

CV contiene il curriculum vitae di una certa persona, questa è una cosa unica e personale, quindi la relazione con Person sarà del tipo “uno-a-uno”.

Expertise&Skills contiene tutte le competenze e le abilità di una persona, sarà quindi relazionata a Person come “multi-a-molti” perché una persona può avere più competenze e più persone possono avere la stessa.

Mentre queste ultime due entità sono associate solamente a Person, le entità Result Patent, Result Product e Result Publication sono riferite in particolar modo all’entità Project (quindi ne parlerò dopo), ma una persona potrebbe essere comunque collegata ad esse (es. autore di una pubblicazione).

Service è associata più direttamente a OrgUnit, ma può essere riferito anche all’entità Person per indicare il ruolo di essa (es. manager, tecnico).

General Facility e Particular Equipment sono entità simili e comprendono le installazioni generali e le attrezzature particolari richieste per lo sviluppo di un progetto, queste sono associate più fortemente con Project (es. mezzi usati per il progetto) oppure con OrgUnit (es. l’istituzione che possiede la risorsa), anche in questo caso però l’entità Person può avere un ruolo con queste due entità (es. responsabile dell’utilizzo, tecnico della manutenzione).

b) Passo adesso ad analizzare le entità secondarie associate a Project:

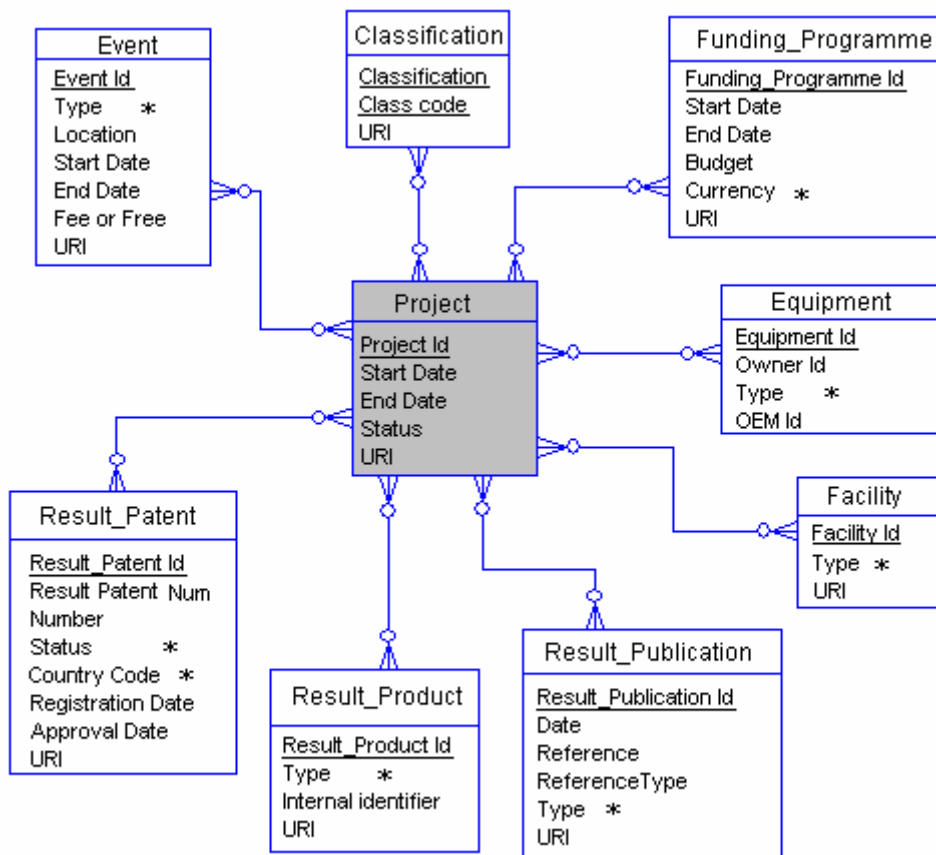


Figura 3.4: entità secondarie relative a Project.

Event, come già detto, è la stessa entità associata a Person.

Result Publication contiene le pubblicazioni relative al progetto (Reference) con l'anno di pubblicazione (Date), il tipo (Reference_Type, es. Vancouver, APA, BibteX), dove è stata pubblicata (Type, es. giornale, sito web) e l'URI, tutti questi attributi, tranne l'ultimo, sono obbligatori.

Result Patent contiene gli eventuali brevetti registrati per il progetto con specificati il numero del brevetto, lo stato (es. applicato, pubblicato, accordato), il codice dello stato in cui è stato registrato, le date di registrazione e di approvazione e l'URI, sono obbligatori solamente il numero di brevetto e il codice dello stato. Per definire lo stato del brevetto viene utilizzato lo standard IPC (International Patent Classification).

Result Product riguarda gli eventuali prodotti realizzati tramite lo sviluppo di un progetto, di questi è presente il tipo (es. in vendita, disponibile per lo sviluppo), un codice identificativo (unico campo obbligatorio) e l'URI.

Funding Programme riguarda gli eventuali finanziamenti ricevuti per lo sviluppo del progetto, questa entità comprende le date di inizio e di fine dei suddetti finanziamenti, il loro ammontare (Budget), il codice che ne indica la valuta (Currency) e l'URI. Questa entità può inoltre auto-riferirsi, ovvero ci può essere una relazione tra due elementi di essa, per esempio nel caso di un programma di finanziamenti che è a sua volta un sotto-programma di un altro, per realizzare questo tipo di relazione si utilizza una tabella intermedia (livello 5).

General Facility spiega le installazioni generali richieste per il progetto, questa entità contiene solamente il tipo di installazione e l'URI, infatti tutti gli altri attributi utili sono contenuti nella tabella intermedia tra questa e Project.

Particular Equipment spiega le attrezzature particolari che servono per lo sviluppo del progetto, di queste è contenuto il tipo di attrezzatura (Type, es. laser, reattore,...), un codice che ne identifica il proprietario (Owner Id, campo obbligatorio), e un altro che identifica il fabbricante (OEM Id ovvero Original Equipment Manufacturer Id), anche in questo caso c'è una tabella intermedia che spiega meglio l'utilizzo di queste attrezzature. Gli aggettivi particolare e generale sono usati con attrezzatura e installazione per distinguere tra un progetto che utilizza un particolare pezzo di un'attrezzatura con un altro che ha un accesso generale ad un'installazione, e che ha quindi a disposizione tutte le attrezzature relative.

Classification viene utilizzata per l'indicizzazione per soggetto, serve per identificare lo schema di classificazione o il thesaurus utilizzato per l'inserimento dei dati e facilitarne così la successiva ricerca, per comodità può essere a sua volta associata ad un'entità Classification_Scheme che contiene tutti gli schemi di classificazione possibili. Questa entità Classification, anche se nella figura 3.4 non è indicato, è associata anche a Result_Publication, a Result_Patent, a Result_Product e a Service, anche in questo caso serve per semplificare la ricerca.

c) Analisi infine le entità secondarie associate a OrgUnit:

Event è la stessa entità associata a Person e a Project.

Contact è uguale a quella associata a Person, con gli stessi attributi, ma non è la stessa perché in questo caso sono contenute le informazioni per contattare le varie unità organizzative.

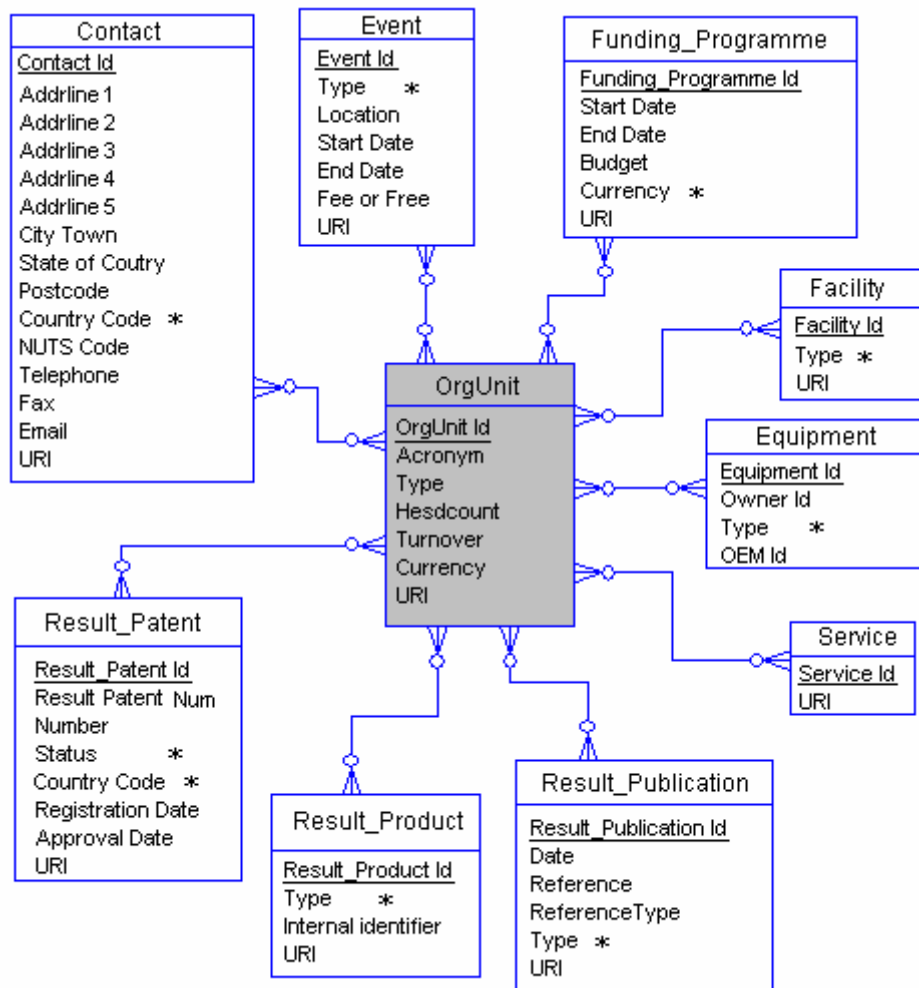


Figura 3.5: entità secondarie relative a OrgUnit.

Le entità Result Publication, Result Patent e Result Product sono associate in particolar modo a Project, ma un'unità organizzativa potrebbe essere comunque collegata ad esse, per esempio potrebbe essere l'intestatario del brevetto (l'unità organizzativa o un suo dipendente) o la responsabile della produzione del relativo prodotto.

Funding Programme è la stessa entità associata a Project ed è utilizzata nel caso un'unità organizzativa finanzia un progetto.

General Facility e Particular Equipment contengono le installazioni e le attrezzature fornite dall'unità organizzativa per lo sviluppo di un progetto, è la stessa entità associata a Person e a Project.

Service spiega i servizi offerti o ricevuti da un'unità organizzativa, questa entità contiene solamente l'URI, infatti tutti gli altri attributi utili sono contenuti nella tabella intermedia tra questa e OrgUnit.

3.1.3 Livello 3 - Entità di supporto plurilingue

Questo livello contiene entità che aggiungono nuove informazioni a quelle già contenute nelle entità principali e in quelle secondarie; queste nuove informazioni aggiunte possono essere scritte in un'unica lingua oppure tradotte in molte lingue, questo per favorire ricerche plurilingue. Questa traduzione delle informazioni non è obbligatoria, ma diventa molto importante nel caso in cui si voglia aumentare la visibilità delle proprie ricerche anche al di fuori del proprio paese, gli sviluppatori di CERIF hanno prestato molta attenzione a questo approccio plurilingue perché sta alla base della condivisione delle informazioni tra i vari paesi europei.

Se, per esempio, un ricercatore italiano pubblicasse i risultati della propria ricerca solo in italiano, ci sarebbero poche possibilità che qualcuno legga questi risultati al di fuori dell'Italia; invece, traducendoli anche solamente in inglese, aumenterebbe di molto il numero di persone che potrebbe accedere in quanto l'inglese è compreso quasi in tutto il mondo.

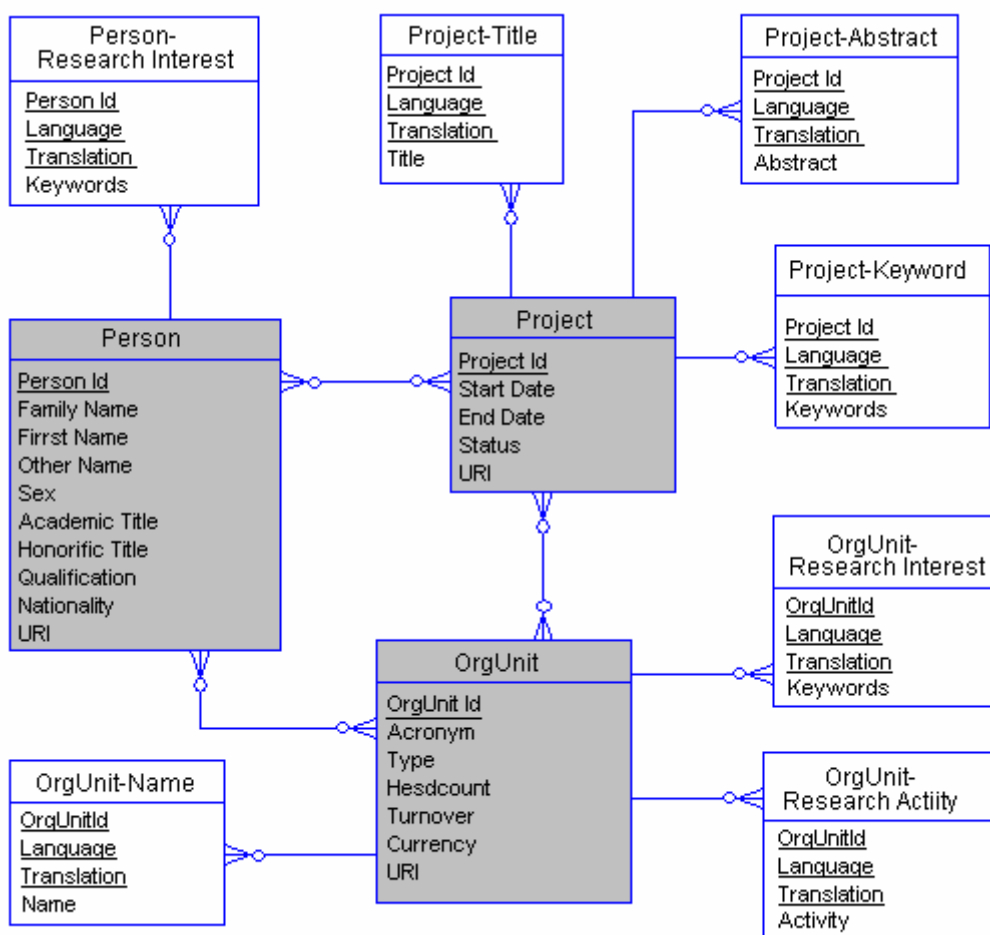


Figura 3.6: entità di supporto plurilingue associate direttamente alle entità principali.

Queste entità del livello 3 sono tutte simili tra loro ed estremamente semplici, in ognuna di esse è contenuto il riferimento all'entità associata, la parola o il testo, la

lingua in cui è scritto e un campo “Translation” che può contenere i valori “o” (se il testo è l’originale), “h” (se il testo è stato tradotto da una persona) o “m” (se è stato tradotto da un particolare software, e quindi la traduzione potrebbe essere poco precisa).

La figura 3.6 non rappresenta tutte le entità appartenenti a questo livello ma solamente quelle che forniscono traduzione ad informazioni relative alle sole entità principali; non sono state rappresentate tutte in quanto, come si vede e come ho già detto, queste entità sono tutte quasi uguali tra loro, gli unici attributi che cambiano sono l’identificativo dell’entità associata e il testo. Le relazioni tra queste entità e quelle a cui sono riferite sono di tipo “uno-a-molti” in quanto sono possibili più traduzioni dello stesso testo; nel caso però il testo sia presente in una sola lingua la relazione diventa di tipo “uno-a-uno”. Inoltre, come si può vedere, la chiave primaria di queste tabelle è rappresentata dalla combinazione di tre attributi, ovvero l’identificativo dell’entità associata, il tipo e la lingua del testo. Per facilitare la comprensione, le entità di questo livello sono state suddivise in due sotto-categorie: quelle associate alle entità principali e quelle associate alle entità secondarie.

a) Inizio l’analisi dettagliata delle entità di questo livello parlando di quelle relative alle 3 entità principali. All’entità Person vengono aggiunte, a questo livello, poche informazioni (solamente gli interessi di ricerca della persona), per questo motivo nell’inserire i dati di un individuo ci si può benissimo fermare al livello precedente.

Per quanto riguarda OrgUnit vengono aggiunte tre nuove informazioni: il nome, la cui traduzione in varie lingue non è particolarmente importante e può essere tralasciata, gli interessi di ricerca, analoga a quella relativa a Person, e le attività di ricerca. Questa ultima informazione è di grande importanza per capire di cosa si occupa una determinata unità organizzativa, effettuando quindi la traduzione di questa informazione si può far conoscere la suddetta in tutta Europa.

Relativamente a Project, a questo livello vengono fornite informazioni fondamentali per capire cosa riguarda il progetto, è quindi molto conveniente esprimerle in varie lingue. In particolare è consigliabile fornire varie traduzioni delle parole chiave (keywords) riguardanti il progetto, perché alcuni sistemi di ricerca, volendo fornire una soluzione rapida, controllano solamente queste senza effettuare la scansione dell’intero testo. A questo proposito, inserendo molte traduzioni delle parole chiave di un progetto, è possibile che questo venga recuperato durante ricerche effettuate in varie lingue. Un’altra informazione importante aggiunta a questo livello è il riassunto, ovvero una breve sintesi del progetto che spiega come è stato sviluppato e a cosa potrebbe servire. Anche questa informazione è di grande importanza in quanto, attraverso essa, si vuole attirare l’attenzione di un eventuale lettore e fare in modo che legga poi tutto il materiale relativo al progetto; traducendolo inoltre in varie lingue (o almeno in inglese) si aumenta il numero di persone che possono leggerlo. Infine viene aggiunto il titolo del progetto; anche in questo caso è consigliabile inserirlo in varie lingue (o almeno in inglese) dal momento che è la prima cosa che viene letta da una persona e, se è scritto in una lingua che questa non capisce è difficile che poi provi a leggere tutte le informazioni relative ad esso.

b) Passo adesso a parlare delle entità di livello 3 associate alle entità secondarie, si noti che quasi tutte le entità secondarie hanno alcune informazioni aggiuntive in varie lingue, questo a testimonianza dell’importanza che gli sviluppatori di CERIF hanno dato alla gestione plurilingue delle informazioni.

In particolare per l'entità `Result_Publication` è possibile aggiungere il titolo e per l'entità `Result_Patent` il titolo e il riassunto, come già detto sopra è importante esprimere queste due informazioni in varie lingue in quanto sono le prime cose che si leggono.

Per l'entità `Result_Product` vengono aggiunti il nome, che si può tranquillamente indicare in una sola lingua, e la descrizione, di quest'ultima informazione è consigliabile fornire varie traduzioni per far capire a più persone i prodotti che si è riusciti a realizzare tramite il progetto realizzato.

Nell'entità `Classification` viene aggiunto l'attributo `Description`, questo è molto utile per far capire che schema di classificazione si è utilizzato.

Infine, alle entità `General_Facility`, `Particular_Equipment`, `Expertise&Skills`, `Service`, `Funding_Programme` ed `Event`, vengono aggiunti gli stessi due attributi: il nome e la descrizione, sarebbe molto conveniente tradurre entrambi in varie lingue in quanto questi attributi servono a spiegare meglio le informazioni contenute nella relativa entità.

3.1.4 Livello 4 - Entità di lookup

Il livello 4 comprende le cosiddette "entità di lookup", ovvero le entità di supporto nel caso di attributi con un set predefinito di possibili valori (nelle figure precedenti questi attributi erano indicati da un asterisco). Dentro queste entità infatti vengono immagazzinati tutti i possibili valori che può assumere l'attributo, non devono quindi essere compilate dall'utente ma dai gestori del CRIS. L'utente deve solamente scegliere il valore dell'attributo tra quelli possibili; i valori inseriti dagli utenti saranno poi immagazzinati in una tabella intermedia che conterrà l'identificativo dell'entità e quello del valore scelto (il contenuto di questa verrà spiegato meglio nelle entità di livello 5). In questo modo viene controllato l'inserimento del valore di un attributo di un'entità, viene quindi assicurato un inserimento corretto ed evitati eventuali errori in quanto l'utente non può inserire valori diversi da quelli proposti.

Le entità di questo livello sono quasi tutte uguali, ognuna infatti possiede un attributo che contiene i vari valori predefiniti (`nome_attributo_full`) e un altro che ne contiene un identificativo (`nome_attributo`); le analizzo ora in dettaglio suddividendole in quattro sotto-categorie: le entità di lookup associate agli attributi delle entità principali, quelle associate agli attributi delle entità secondarie, quelle associate all'attributo di ruolo nelle link table e le due nuove entità aggiunte.

a) Inizio elencando le entità di lookup associate agli attributi delle entità principali, facendo anche alcuni esempi di possibili valori:

`Project-Status` (completato, in corso, in fase di stallo), `Person-Honorific Title` (Sir, Lady), `Person-Academic Title` (Dr, Prof), `Person-Qualification` (DSc, DPhil, PhD, MSc, Mphil, BSc, RNDr), `Person-Nationality` (nome completo, non codice), `OrgUnit-Type` (vedi tabella 3.1) e `OrgUnit-Currency` (alcuni esempi nella tabella 3.2), per quest'ultima entità viene utilizzato lo standard internazionale "ISO 4217", stabilito dall'Organizzazione Internazionale per le Standardizzazioni, il quale fornisce codici di tre lettere per definire i nomi delle valute (le prime due lettere del codice corrispondono al codice nazionale definito dallo standard "ISO 3166-1 alpha-2" per eliminare il

problema causato da nomi come Dollaro, Franco e Sterlina utilizzati in dozzine di nazioni diverse, ma con valori differenti).

BES	settore di business, include anche i consulenti individuali
HES	settore di alta educazione
RPR	centri di ricerca privati/commerciali
RPN	centri di ricerca privati senza scopo di lucro
RPU	centri di ricerca pubblici
JRC	centri di ricerca comuni
PUS	settori pubblici non di ricerca
PNP	settori privati senza scopo di lucro non di ricerca
INO	organizzazioni internazionali
OTH	altri

Tabella 3.1: codici per identificare i tipi di unità organizzativa.

BGN	Lev bulgaro
CHF	Franco svizzero
CSD	Dinaro serbo
CZK	Corona ceca
DKK	Corona danese
EUR	Euro
GBP	Sterlina inglese
HRK	Kuna croata
HUF	Fiorino ungherese
ISK	Corona islandese
NOK	Corona norvegese
PLN	Zloty polacco
RUB	Rublo russo
SEK	Corona svedese
TRL	Lira turca

Tabella 3.2: esempi di codici ISO 4217 per identificare le monete europee.

BE	Belgio
DK	Danimarca
FR	Francia
DE	Germania
GR	Grecia
IE	Irlanda
IT	Italia
NL	Olanda
PT	Portogallo
ES	Spagna
GB	Gran Bretagna

Tabella 3.3: esempi di codici ISO 3166 per identificare gli stati europei.

b) Passo adesso alle entità di lookup associate agli attributi delle entità secondarie: Result_Publicatio-Type (giornale, sito web), Result_Product_Type (in vendita, disponibile per lo sviluppo), Result_Patent-Status (applicato, pubblicato, accordato), Result_Patent-Country Code (vedi esempi nella tabella 3.3), Facility-Type, Equipment-Type, Contact-Country Code (vedi tabella 3.3), Contact-NUTS Code (codice che identifica le varie regioni di uno stato), Event-Type (conferenza, evento culturale, esibizione, fiera commerciale), Funding_Programme-Currency (vedi tabella 3.3).

c) Passo adesso alla spiegazione delle entità di lookup relative alle tabelle intermedie; si noti che adesso questa spiegazione potrebbe risultare poco comprensibile in quanto queste ultime non sono ancora state spiegate, ma sarà tutto più chiaro dopo aver letto la parte relativa ad esse. Le tabelle intermedie servono per rappresentare le relazioni di tipo “molti-a-molti”, quindi ce ne sarà una per ogni relazioni di questo tipo esistente tra le tre entità principali o tra una di queste e le sue entità associate (ovvero tutte tranne CV per l’entità Person). In queste entità intermedie c’è un attributo che indica il ruolo, e proprio a questo viene associata una tabella di lookup che contiene tutti i vari ruoli possibili tra le due entità.

d) Oltre ad appoggiare attributi di entità già esistenti, vengono anche aggiunte due nuove entità: Language, associata a Person, e Multimedia_Type, associata a CV.

Language contiene tutte le lingue europee e serve per indicare le lingue conosciute e parlate da una persona, gli attributi di questa entità sono: un codice di due lettere che identifica una lingua (Language), il nome completo della lingua (Language full name) e la sua traduzione in inglese (Language english name). Per identificare le diverse lingue viene utilizzato lo standard “ISO 639”, questo identifica ogni lingua con un codice di due lettere, nella tabella 3.4 sono riportati quelli per le lingue europee più importanti.

Multimedia_Type serve invece a descrivere il formato multimediale in cui è scritto il curriculum vitae di una persona, anche in questo caso la relazione tra questa entità e CV è di tipo “molti-a-molti”. Questa entità possiede solamente due attributi, ovvero il tipo di formato (Multimedia type full) ed un identificativo per esso (Multimedia type).

DA	Danese
FR	Francese
EL	Greco
EN	Inglese
IT	Italiano
NL	Olandese
PT	Portoghese
RU	Russo
ES	Spagnolo
DE	Tedesco

Tabella 3.4: esempi di codici ISO 639 per identificare le lingue europee.

3.1.5 Livello 5 - Entità di link (o entità intermedie)

Quelle appartenenti a questo ultimo livello sono le cosiddette “link table”, ovvero le entità necessarie alla realizzazione delle relazioni di tipo “multi-a-molti”. Grazie a queste, infatti, è possibile trasformare una relazione “multi-a-molti” in una coppia di relazioni “uno-a-molti” che collegano le due entità ad una tabella intermedia (l’entità di link) che contiene gli identificativi di un elemento di ognuna delle due entità. In queste tabelle gli unici attributi obbligatori sono i due identificativi, gli altri servono a spiegare meglio la relazione tra le due entità.

In quasi tutte le entità di questo livello è presente un attributo “Role” che potrebbe indicare la cooperazione tra due persone, la partecipazione di una persona o di un’unità organizzativa nella realizzazione di un progetto, ecc. Come detto, questo attributo possiede un set predefinito di possibili valori diversi, a seconda del tipo di entità che sono relazionate.

Come si è visto studiando le entità dei livelli superiori, quasi tutte le relazioni del modello di dati di CERIF sono del tipo “multi-a-molti”, quindi le entità appartenenti a questo livello sono molto numerose; per questo motivo, per semplificarne la spiegazione, divido questo livello in quattro sotto-categorie: le entità di link per le relazioni tra le sole entità principali, quelle tra le entità principali e le relative entità secondarie riguardanti i dati personali e quelle riguardanti i progetti svolti, infine quelle tra sole entità secondarie.

a) Le entità intermedie più importanti sono quelle che servono per realizzare le relazioni tra le tre sole entità principali e gli auto-riferimenti di esse, queste sono anche chiamate entità di link primarie.

Come si vede dalla figura 3.7, gli attributi di queste tabelle sono gli identificativi degli elementi delle due entità a cui sono associate, oppure di due elementi della stessa tabella (nel caso di auto-riferimenti), l’attributo di ruolo, che indica l’associazione tra i due elementi, e due attributi “Start date” e “End date” che indicano la durata della collaborazione. Come già detto, il set di valori dell’attributo Role cambia a seconda della relazione che rappresenta, in particolare alcuni esempi di possibili valori sono:

- Person-Project: responsabile del progetto, ricercatore, contabile;
- Person-OrgUnit: capo della divisione, leader del gruppo;
- Project-OrgUnit: gruppo, divisione, reparto, laboratorio;
- Person-Person: collega, manager;
- Project-Project: sotto-progetto, progetto parallelo;
- OrgUnit-OrUnit: sotto-gruppo, gruppo associato.

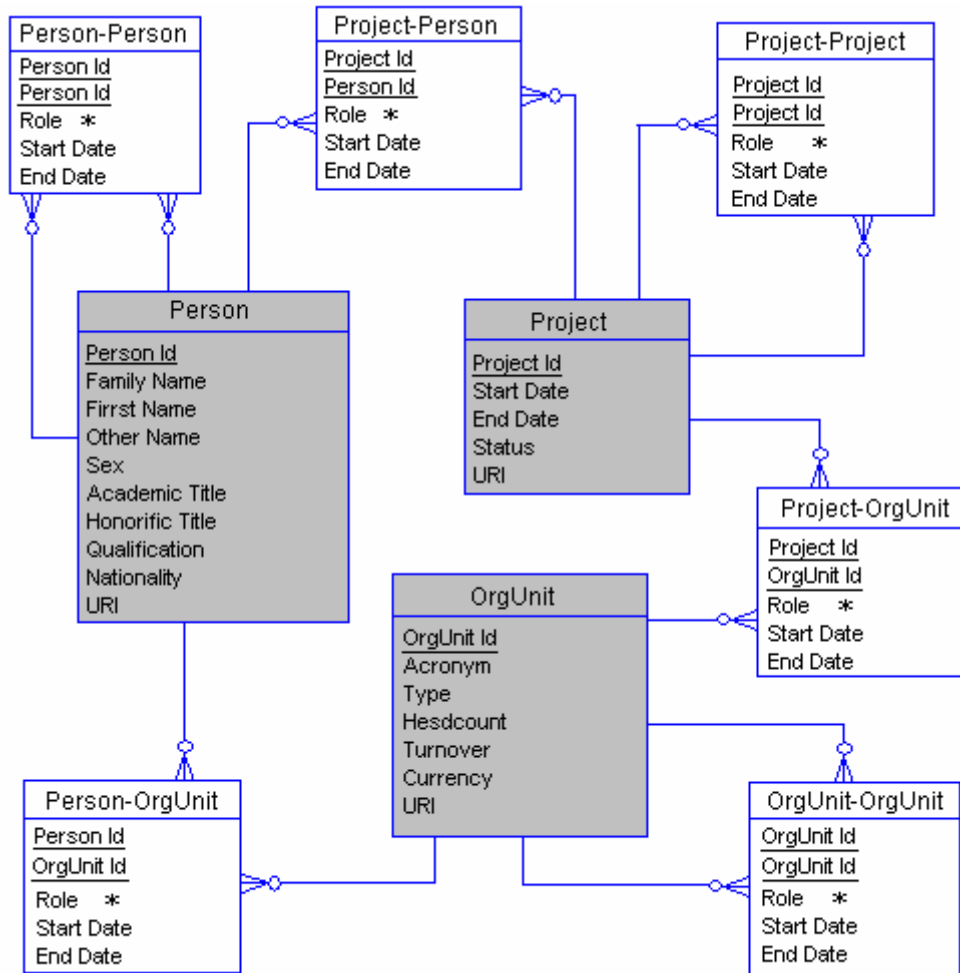


Figura 3.7: entità di link per le relazioni tra le entità principali.

b) Passo adesso a descrivere le tabelle di link che realizzano le relazioni tra le entità principali e le relative entità associate riguardanti i soli dati personali, queste sono diverse per ognuna delle entità principali. Nella figura 3.8 sono mostrate quelle relative a Person come esempio, non vengono riportate quelle relative a Project e ad OrgUnit in quanto sono associate entrambe ad una sola tabella intermedia. Si noti che nella figura 3.8 non sono riportate le entità secondarie a cui si riferiscono le tabelle intermedie per renderla più comprensibile ed evitare di fare confusione, è comunque semplice capire di quale entità secondaria si tratta dal nome della tabella intermedia (es. Person-Contact è associata a Contact). Queste entità intermedie sono ben diverse da quelle descritte in precedenza, solamente Person-Contact ha gli stessi attributi delle altre, dove Role può assumere i valori centro di distribuzione, ufficio, laboratorio, ecc.

Person-Expertise&Skills contiene la descrizione delle prestazioni lavorative di una persona, oltre ai soliti attributi, dove Role può assumere i valori consulente, tirocinante, esperto, ecc., questa entità contiene altre informazioni importanti, come la disponibilità di una persona, le sue condizioni e il prezzo delle prestazioni. Non è presente un attributo che indica la valuta in cui è espresso il prezzo delle prestazioni in quanto è obbligatorio inserirlo in Euro.

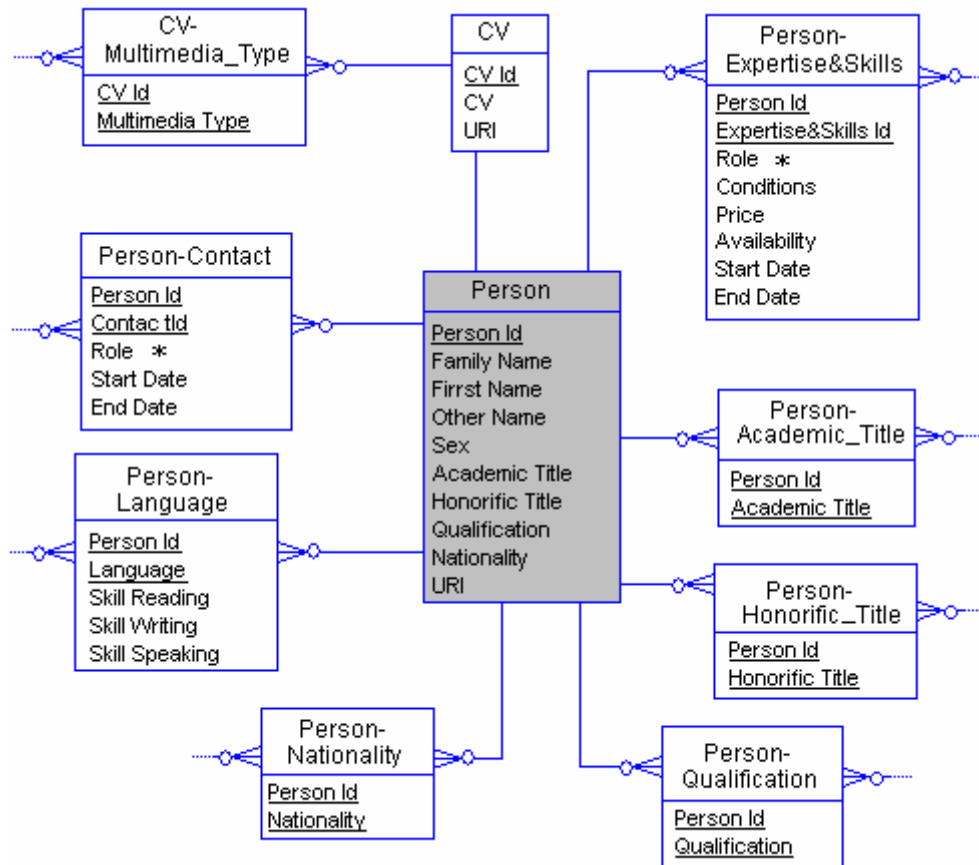


Figura 3.8: entità intermedie associate a Person e relative ai dati personali.

Person-Language serve ad indicare le conoscenze linguistiche della persona in questione, in particolare vengono descritte, per ogni lingua da essa conosciuta, le capacità di lettura, di scrittura e di comunicazione orale. Le altre tabelle intermedie associate a Person sono invece molto semplici, infatti contengono solamente gli identificativi degli elementi delle due entità associate.

Le altre due entità principali sono associate solamente ad una tabella intermedia riguardante i dati personali, infatti Project ha solamente una tabella intermedia che la relaziona a Classification, questa contiene gli identificativi delle due entità (Project Id e Classification code) e lo schema di classificazione utilizzato. OrgUnit, invece, ha solamente una tabella intermedia che la relaziona a Contact, questa è uguale a quella descritta nel caso di Person.

c) Passo adesso a descrivere le tabelle intermedie tra le entità principali e le relative entità associate riguardanti i risultati dei progetti, queste sono quasi le stesse per tutte le entità principali. Nella figura 3.9 sono mostrate quelle relative a Project come esempio, non vengono riportate le figure relative a Person e ad OrgUnit in quanto le differenze sono minime. Si noti che neanche nella figura 3.9 sono riportate le entità secondarie a cui si riferiscono le tabelle intermedie, e anche in questo caso si può capire di quale entità secondaria si tratta dal nome della tabella intermedia.

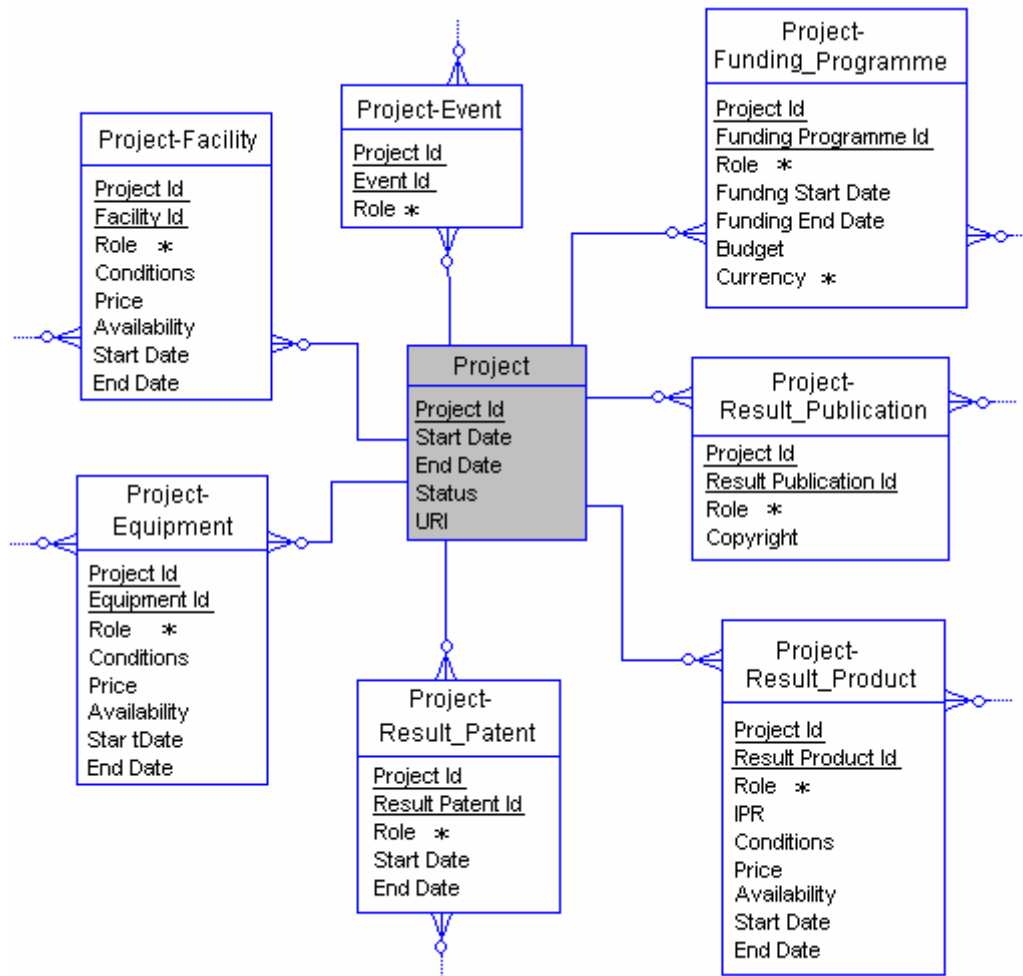


Figura 3.9: entità intermedie associate a Project e relative ai risultati dei progetti.

Le entità di link appartenenti a questa sotto-categoria sono quasi tutte diverse tra loro, vado ora ad analizzarle in dettaglio.

Project-Event è un'entità molto semplice e descrive il ruolo che ha l'evento in questione con il progetto (es. presentazione del progetto), oltre agli identificativi delle due entità è presente solamente l'attributo Role.

Project-Funding Programme descrive i programmi di finanziamento di cui gode il progetto, di questi viene spiegato il ruolo (es. finanziatore totale o parziale), le date di inizio e di fine, l'ammontare (Budget) e la valuta (Currency).

Project-Facility e Project-Equipment collegano un progetto con le installazioni generali e le attrezzature particolari di cui necessita, gli attributi di queste due entità sono uguali, ovvero il ruolo (es. utente), il prezzo da pagare per l'utilizzo (Price, valore da inserire in Euro), la disponibilità (Availability), le condizioni (Conditions) e le date di inizio e di fine dell'utilizzo.

Project-Result Patent relaziona un progetto con i suoi eventuali brevetti, questa entità è molto semplice e contiene il ruolo e le date di inizio e di fine della validità del brevetto.

Project-Result Publication descrive le pubblicazioni relative al progetto, contiene gli attributi Role (es. rapporto tecnico) e Copyright.

Project-Result Product spiega infine la relazione tra il progetto ed i prodotti realizzati grazie ad esso, di questi viene descritto il ruolo (es. sviluppo del progetto), il prezzo (in Euro), la disponibilità, le condizioni e le date di inizio e di fine della produzione.

Per quanto riguarda le entità associate a Person, queste sono quasi le stesse, le uniche differenze sono la mancanza di Person-Funding_Programme e l'aggiunta di Person-Service, questa tabella spiega i servizi svolti da una persona riguardanti un certo progetto, i suoi attributi sono il ruolo (es. contatto, manager), le condizioni, il prezzo, la disponibilità e le date di inizio e di fine della prestazione del servizio. Si deve inoltre notare che, nelle entità descritte per Project, che sono presenti anche in Person, gli attributi hanno tutti lo stesso significato, tranne Role, in cui il set di possibili valori cambia, in particolare ecco alcuni esempi:

- Person-Event: organizzatore, oratore;
- Person-Facility e Person-Equipment: operatore, direttore, responsabile marketing;
- Person-Result_Patent: inventore;
- Person-Result_Publication: autore, redattore, recensore;
- Person-Result_Product: assegnatario.

Le entità di link associate a OrgUnit, infine, sono uguali a quelle di Project, con l'aggiunta di OrgUnit-Service, che descrive i servizi offerti da un'unità organizzativa, questa entità è analoga a quella descritta in Person. Anche in questo caso i possibili valori dell'attributo Role sono diversi rispetto a quelli elencati in precedenza, ovvero:

- OrgUnit-Event: comitato di organizzazione, segretariato di conferenza;
- OrgUnit-Funding_Programme: coordinatore, direttore;
- OrgUnit-Service , OrgUnit-Facility e OrgUnit-Equipment: possessore, manager;
- OrgUnit-Result_Patent: inventore;
- OrgUnit-Result_Publication: pubblicatrice o compilatrice di un rapporto;
- OrgUnit-Result_Product: proprietaria, assegnataria, acquirente.

d) L'ultima sottocategoria comprende le tabelle intermedie che realizzano relazioni tra sole entità secondarie, queste riguardano in particolare le classificazioni e hanno tutte gli stessi attributi, ovvero l'identificativo della tabella associata, il codice e lo schema di classificazione (questo cambia a seconda della tabella), queste entità sono: Result_Patent-Classification, Result_Product-Classification, Result_Publication-Classification, Service-Classification. Esiste poi un'ultima tabella intermedia che è utilizzata per effettuare un auto-riferimento dell'entità Funding_Programme, questa serve per descrivere una possibile relazione tra due programmi di finanziamenti, contiene il ruolo (es. divisione all'interno di un'organizzazione) e le date di inizio e di fine di questa relazione.

3.2 CERIF exchange model

Il CERIF exchange model è il modello utilizzato per lo scambio di dati tra i vari CRIS o tra i fornitori di CRIS e gli utenti finali. Questo non è un modello unico, ma ci sono

molte varianti, a seconda dell'insieme di dati che si vogliono trasferire, ogni variante di questo modello comprende un sottoinsieme di entità del Full CRIS data model.

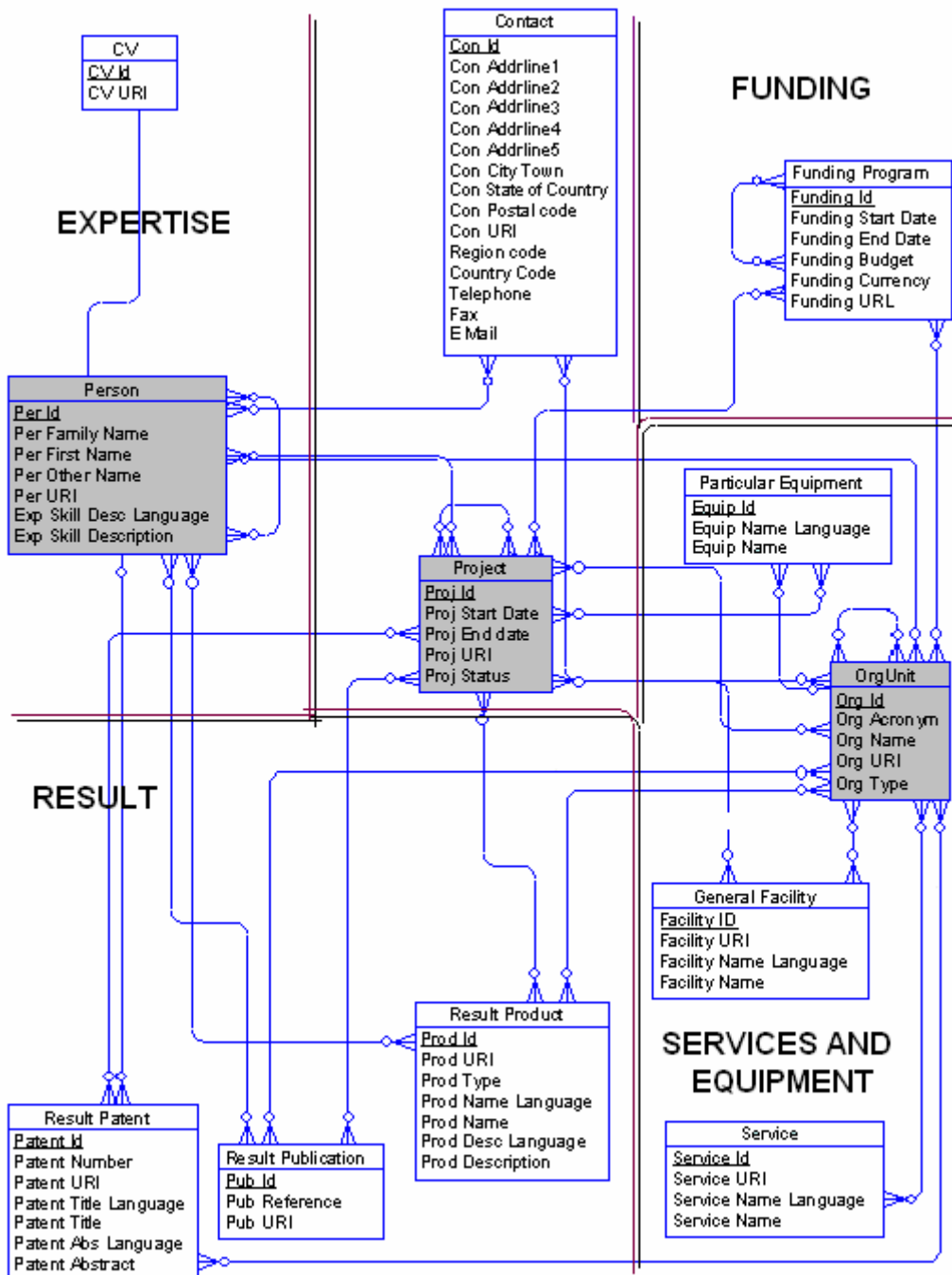


Figura 3.10: divisione delle entità secondarie nei vari modelli di scambio di dati.

Nella figura 3.10 è rappresentata la suddivisione delle entità principali e di quelle secondarie nei vari modelli di scambio di dati, questi non sono i soli esistenti ma sono quelli più comunemente utilizzati. Dalla figura 3.10 si può notare come le entità Project e Contact siano comuni a tutti i modelli di scambio, questo perché il progetto è l'entità fondamentale del modello di dati, quella da cui dipendono tutte le altre, mentre l'entità

Contact è importantissima per raggiungere le persone che hanno contribuito alla realizzazione di esso. Vado ora ad analizzare in dettaglio i modelli di scambio che sono stati proposti:

- **Expertise:** comprende l'entità Person (a cui sono stati aggiunti attributi riguardanti le abilità e le competenze della persona) e il suo Curriculum Vitae. Questo modello è utilizzato nel caso in cui si vogliono fare conoscere i realizzatori del progetto, vengono infatti tralasciate le informazioni riguardanti le unità organizzative e i risultati ottenuti.
- **Result:** questo modello è formato dai soli risultati del progetto, vengono infatti descritte le pubblicazioni, i brevetti e i prodotti realizzati. Questo modello è indicato nel caso in cui si voglia far conoscere dettagliatamente il progetto realizzato, senza soffermarsi sui realizzatori.
- **Service and Equipment:** questo sottoinsieme di entità comprende OrgUnit, Service, General_Facility e Particular_Equipment. Questo modello è utile per far conoscere le varie unità organizzative, vengono infatti riportate informazioni dettagliate riguardo esse, i servizi offerti e le attrezzature messe a disposizione.
- **Funding:** in questo modello sono presenti solamente le informazioni riguardo i programmi di finanziamento ricevuti per le realizzazioni dei progetti.

Nella figura 3.10 sono rappresentate solamente le entità principali e quelle secondarie, ma possono essere aggiunte ai modelli anche entità di altri livelli. Questo dipende dalla quantità e dal dettaglio delle informazioni che si vogliono scambiare; infatti i CRIS che vogliono scambiarsi informazioni possono scegliere il modello che più si addice alle proprie esigenze.

3.3 Metadata data model

Per gli utenti finali e per i gestori dei CRIS è importante ottenere modelli di scambio di dati adatti alle proprie esigenze, per questo motivo questi devono riuscire a selezionare dal modello completo le entità, gli attributi e le relazioni di cui sono più interessati. Per effettuare questa selezione è molto utile una rappresentazione riassuntiva di tutte le informazioni contenute nel full CRIS data model o nel exchange data model, questa rappresentazione è fornita dal modello di metadata. La definizione di questo modello di metadata richiede la ricerca dell'insieme minimo di entità, attributi e relazioni che permettono all'utente finale di trovare il più precisamente possibile le informazioni da lui richieste. In pratica questo modello è usato come un catalogo per selezionare le informazioni di interesse contenute nei vari CRIS, poi ogni CRIS esporta queste informazioni utilizzando un determinato modello di scambio di dati.

Nella figura 3.11 sono rappresentate le entità che costituiscono il modello di metadata di CERIF, questo è un sottoinsieme del modello di dati completo, con l'aggiunta delle entità Project Additional, OrgUnit Additional e Person Additional. In questa figura non sono presenti gli attributi delle entità intermedie e dell'entità Contact in quanto queste tabelle sono uguali a quelle descritte nel full CRIS data model.

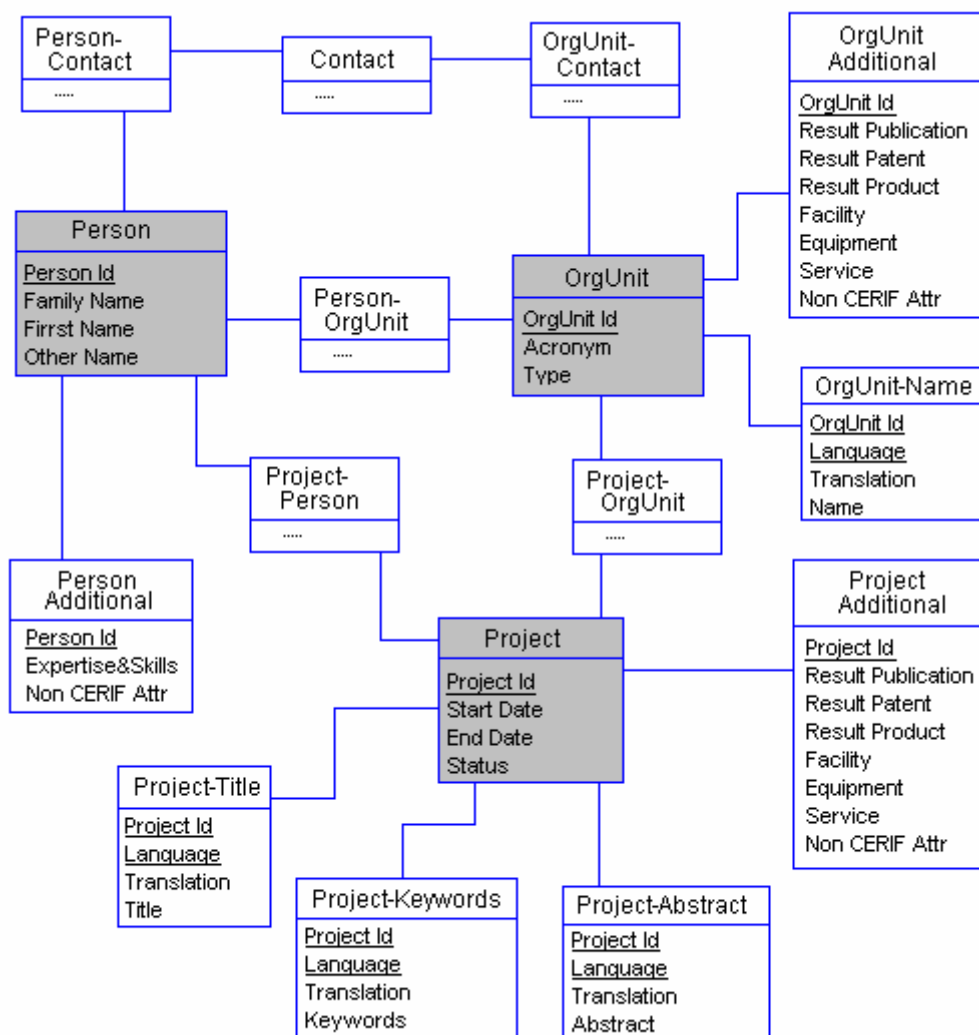


Figura 3.11: metadata data model.

Come si vede dalla figura 3.11, nelle tabelle principali sono presenti solamente alcuni attributi, ovvero quelli necessari a dare un'idea generale dell'entità descritta. Tra le entità secondarie del modello di dati completo è invece presente soltanto Contact, con tutti i relativi attributi, in quanto questa entità è molto importante per rintracciare una persona o un'unità organizzativa. Come già detto, sono poi state aggiunte alcune entità, queste contengono l'identificativo dell'entità a cui sono associate ed una serie di flag (con possibili valori [Y / N]) che indicano se nel modello completo o in quello di scambio di dati è presente o meno la relativa entità.

Per esempio si supponga che i valori di metadata per un particolare progetto presenti in Project Additional siano: Result Publication = Y, Result Patent = N, Result Product = N, Facility = N, Equipment = Y, Service = N. Questo vuole dire che, nel modello di dati completo o in quello di scambio, relativamente a questo progetto, si possono trovare una o più (non viene indicato il numero) istanze contenenti informazioni riguardo pubblicazioni sui risultati del progetto e sulle attrezzature richieste, non saranno presenti invece informazioni su brevetti, prodotti realizzati, installazioni o servizi richiesti.

Tra le entità di supporto plurilingue sono presenti solo quelle di maggiore importanza, ovvero quelle che forniscono traduzioni al nome di un'unità organizzativa e, soprattutto, al titolo, il riassunto e le parole chiave di un progetto.

Le tabelle di lookup non sono presenti in quanto questo modello rappresenta soltanto un riassunto del modello completo e non vi è la necessità di facilitare l'inserimento dei dati, in questo modo si evita l'aggiunta di numerose tabelle.

Le tabelle intermedie presenti, infine, sono le uniche necessarie data la scarsa presenza di entità secondarie, queste sono complete di tutti i loro attributi.

4 Esempi di sistemi basati su CERIF

Dopo aver spiegato in dettaglio il modello di dati utilizzato da CERIF, adesso verranno presentati alcuni esempi reali di database che contengono informazioni sulla ricerca e che utilizzano questo modello di dati:

1. EUREKA, questo è un CRIS pan-europeo per la ricerca e lo sviluppo industriale orientato al mercato che coinvolge 30 paesi europei;
2. “Hungarian Current Research Information System”, ovvero il CRIS ungherese;
3. “The Projekt database”, contiene informazioni su progetti di ricerca realizzati all’interno della Åbo Akademi University;
4. WISARD, ovvero “Web based Information Services for Agricultural Research for Development”.

4.1 EUREKA

Creato come un’iniziativa intergovernativa nel 1985, EUREKA punta ora a migliorare la competitività europea attraverso il suo appoggio alle società, ai centri di ricerca e alle università che eseguono progetti pan-europei per sviluppare prodotti innovativi e servizi; offre inoltre un aiuto nella ricerca di partner finanziari, tecnici e commerciali. Questo sistema punta molto sul lato economico dei progetti, quindi nel suo database sono contenute informazioni dettagliate solamente riguardo il loro costo e le loro potenzialità di mercato; mentre le informazioni riguardanti le persone e le unità organizzative che vi hanno partecipato sono le minime indispensabili, le informazioni contenute inoltre sono solamente in lingua inglese. Attraverso la sua rete flessibile e decentralizzata, EUREKA offre agli utenti un accesso rapido ad una grande quantità di informazioni; è importante notare che per accedere al database non è necessario effettuare nessuna registrazione, questo permette a tutti gli utenti di dare un’occhiata ai progetti a cui sono interessati.

Oltre all’archivio contenente i progetti (di cui si parlerà in seguito), in EUREKA ne è presente un altro in cui sono presenti le news, gli eventi e i progetti che hanno avuto successo in passato e le pubblicazioni, queste ultime non relative ai singoli progetti ma alla rete di EUREKA e alla ricerca in generale. Tramite un motore di ricerca molto semplice (è possibile scegliere la categoria e inserire alcune parole chiave e un intervallo di tempo) ogni utente può reperire i documenti che più gli interessano; di questi viene fornita una breve descrizione e, dove presente, un link al sito o al documento ufficiale. Per quanto riguarda poi le pubblicazioni, alcune di queste sono presenti in varie lingue.

Altri due servizi molto importanti offerti da EUREKA sono i “cluster” e gli “umbrella”. I “cluster” sono iniziative industriali a lungo termine strategicamente

significative che hanno un gran numero di partecipanti e puntano a sviluppare tecnologie di grande importanza per la competitività europea. Queste inizialmente riguardavano solamente tecnologie di comunicazione e informazione, ma più recentemente si sono estese anche a energia e biotecnologia. I cluster permettono la collaborazione delle grandi società con le SME (piccole e medie imprese), gli istituti di ricerca e le università, dividendo così il rischio e i benefici dell'innovazione. Gli "ombrella" invece sono reti tematiche all'interno della struttura di EUREKA che si concentrano su una specifica area di tecnologie o un settore di affari; il loro obiettivo principale è quello di facilitare la generazione dei progetti EUREKA nella propria area di competenza. Le attività di ombrella sono coordinate e perfezionate da un working group con al suo interno alcuni rappresentanti di EUREKA ed alcuni esperti industriali.

Passo ora all'analisi del database contenete i progetti (project portfolio), per il reperimento di questi è fornito un motore di ricerca di base che permette di cercare progetti in base ad una persona che vi ha collaborato, ad un identificativo o un acronimo di EUREKA; è inoltre disponibile un sistema di ricerca avanzato molto completo e specifico.

Figura 4.1: ricerca avanzata dei progetti in EUREKA.

Come si vede dalla figura 4.1, anche in questo sistema di ricerca è possibile inserire il numero o l'acronimo EUREKA del progetto a cui si è interessati (se si conosce); inoltre si possono inserire alcune parole chiave e scegliere un paese europeo che ha collaborato al progetto, in questo elenco sono presenti tutti e 37 gli stati europei. È poi possibile scegliere l'area tecnologica di appartenenza del progetto tramite i due relativi menù a tendina: il primo contiene le categorie ed il secondo le varie sottocategorie per ognuna di esse. Questo permette una scelta molto precisa della categoria e riduce i risultati approssimativi della ricerca. Le categorie tra cui si può scegliere sono:

1. Electronics, it and telecomms technology;
2. Industrial manufacturing, material and transport;
3. Other industrial technologies;
4. Energy technology;

5. Chemistry, physical and exact sciences;
6. Biological sciences;
7. Agriculture and marine resources;
8. Agrofood technology;
9. Measurements and standards;
10. Technology for protecting man and the environment.

Il menù a tendina a destra permette poi di scegliere l'area commerciale del progetto tra le seguenti:

1. Communications;
2. Computer related;
3. Other electronics related;
4. Biotechnology/molecular biology;
5. Medical/health related;
6. Energy;
7. Consumer related;
8. Industrial product/manufacturing;
9. Transportation;
10. Construction and building products;
11. Agriculture, forestry and fishing;
12. Services.

Gli ultimi due parametri di ricerca che si possono impostare sono molto importanti. Quello a sinistra permette di scegliere lo stato dei progetti che si vogliono visionare, ovvero di selezionare solamente i progetti ancora in fase di sviluppo (running), già terminati (finished) oppure quelli che sono ancora una semplice proposta (proposal). Questo parametro è molto importante, ad esempio, per una società che cerca un progetto a cui collaborare o una proposta di progetto da sviluppare. Il pulsante di scelta a destra infine consente di cercare i soli progetti i cui realizzatori sono in cerca di partner per proseguire il loro lavoro; questo è molto utile sia per i realizzatori di un progetto, che potrebbero riuscire a trovare facilmente qualcuno che li aiuti, ma anche per una società in cerca di un progetto da finanziare.

Nella figura 4.2 è rappresentato il risultato di una ricerca in cui si sono impostati come unici parametri la categoria (Electronics, it and telecomms technology) e la sottocategoria (Multimedia) del progetto. La prima colonna è quella principale, in essa sono contenuti, in verde, il codice identificativo e l'acronimo del progetto, appena sotto invece è indicato se quello è un normale progetto oppure fa parte di un'umbrella. Sotto queste informazioni, in alcuni casi, è scritto che per quel progetto si è alla ricerca di partner per una collaborazione; è molto importante che questa informazione sia stata inserita in questa pagina di riassunto perché in questo modo un'eventuale organizzazione interessata a collaborare con qualche progetto può vedere subito chi ne ha bisogno. Nella colonna centrale sono presenti le informazioni generali sul progetto, innanzitutto è inserito lo stato del progetto, che può essere finito, in corso (annunciato o approvato) o solo una proposta, e di seguito la data in cui è entrato in quello stato. Sono poi presenti la categoria tecnologica e il settore commerciale di appartenenza, le date di inizio e di fine (data prevista se non è ancora finito), la durata ed il costo complessivo. Nella colonna a destra infine è presente una breve spiegazione del progetto; questa serve

solo a dare un'idea di quello in cui consiste, sarà poi spiegato meglio nella sezione riguardante il progetto.

E! 2964 STEREOVISOR Project	Finished 01-Jul-2005 Technology : Electronics and Telecomms Technology Start date : 08-Nov-2002 Duration : 30 months Total cost : 1.05 MEur End date : 08-May-2005	Summary The Cartographic Visualisor Is An Independent Application That Enables Cartographic Data To Be Visualised For Raster Formats, Vectorial And Alphanumeric Information.
.....		
E! 3118 EUROLEARN EUROPEAN WELDER Sub-Umbrella	Announced 18-Jun-2004 Technology : Electronics and Telecomms Technology Market : Computer Related Start date : 01-Jun-2003 Duration : 60 months Total cost : 2.05 MEur End date : 01-Jun-2008	Summary Development Of A Blended Learning System That Meets The Requirements Of Harmonized European Guidelines For The Education Of Welders. The System Will Cover The Theoretical Content As Well As Practical Welding Tasks.
.....		
E! 3701 IDA-MRI Project	Proposal 09-Oct-2006 Technology : Electronics and Telecomms Technology Market : Medical / Health Related Start date : 01-Jan-2006 Duration : 36 months Total cost : 1.45 MEur End date : 01-Jan-2009	Summary Development Of Intelligent Magnetic Resonance Acquisition And Processing Methods Exploiting Complementary And Redundant Information For Robust And Comprehensive Assessment Of Cardiac Anatomy And Function With Minimal User Interaction.
<input checked="" type="checkbox"/> looking for partners		

Figura 4.2: esempio di risultati di una ricerca in EUREKA.

Aperto la pagina relativa ad un progetto si presentano quattro sezioni: general info, project outline, participants e printable project sheet (che contiene tutte le informazioni delle altre sezioni in formato pdf), in queste sono riportate le informazioni relative al progetto in forma descrittiva.

In “general info” sono riportate nuovamente le informazioni generali relative al progetto con l’aggiunta di una spiegazione dettagliata dei costi e della durata, viene illustrato quanto si è speso (in soldi e in tempo) nella fase di definizione e in quella di implementazione. È poi presente una tabella che spiega il ruolo e la percentuale con cui hanno contribuito i vari stati europei e, infine, vengono elencate le unità organizzative che hanno collaborato e il loro ruolo.

Nella sezione “project outline” è inizialmente spiegato in dettaglio in che cosa consiste il progetto e come è stato realizzato, ovviamente se si tratta di una proposta viene spiegato soltanto cosa si intende fare. Viene indicata poi una previsione dello sviluppo tecnologico del progetto, ovvero come può essere utilizzato e ulteriormente migliorato. Infine vengono spiegate le applicazioni commerciali del progetto, ovvero chi potrebbe essere interessato all’acquisto e perché.

In “participant” sono contenute le unità organizzative e le persone che hanno partecipato al progetto, di queste viene riportato l’indirizzo, i vari modi per contattarle (telefono, fax, e-mail), l’home page e, nel caso delle organizzazioni, un responsabile;

sono poi riportati, per ogni partecipante al progetto, il suo contributo e le sue competenze.

Nel caso in cui una persona o un'organizzazione volessero inserire il loro progetto nel sistema di EUREKA ci sono due possibilità: potrebbero proporre solamente un'idea di base e aspettare eventuali collaboratori per svilupparla, oppure presentare il progetto con tutte le relative informazioni dettagliate. Nel primo caso è sufficiente compilare un modulo on-line indicando l'idea alla base del progetto, le proprie competenze, il contributo che si intende dare e quello che si sta cercando (contributo economico, tecnologico, ecc.) per lo sviluppo del progetto; ovviamente si devono anche inserire i dati anagrafici e tutte le informazioni necessarie per essere contattati. In questo caso qualunque persona può inviare un modulo, mentre nel caso della proposta del progetto serve la collaborazione di almeno due membri di EUREKA. In questo secondo caso si deve compilare un altro modulo in cui si deve spiegare dettagliatamente in che cosa consiste il progetto, le spese, la data di inizio e la durata, i collaboratori e il loro contributo e le possibili applicazioni commerciali.

4.2 HunCRIS

HunCRIS è il sistema che gestisce i progetti di R&S ungheresi, questo si basa sul modello di dati proposto da CERIF e contiene, oltre alle informazioni relative ai progetti, anche quelle riguardanti le persone e le unità organizzative. A differenza dell'esempio precedente (EUREKA), questo sistema non è incentrato sul lato economico dei progetti, ma punta solamente alla diffusione dei risultati di questi; inoltre in questo caso vengono offerti meno servizi: è possibile solamente visualizzare i progetti e aggiungerne altri. Per favorire una maggiore visibilità alle informazioni contenute anche al di fuori dell'Ungheria, il sito web del database è disponibile sia in ungherese che in inglese; questo fornisce un buon livello di multilingualità in quanto, anche se non sono disponibili altre lingue, le informazioni in inglese sono comprensibili quasi in tutto il mondo. Alcune informazioni contenute nella base di dati sono solamente in ungherese, ma le più importanti sono tradotte anche in inglese. Per aggiungere nuovi progetti al database è necessaria l'iscrizione al sistema e la compilazione di due moduli; questi sono interamente scritti in ungherese ma, visto che questo CRIS contiene i soli progetti realizzati in questa nazione, sarebbe inutile scriverli in altre lingue. Per accedere alle informazioni del database sono possibili due tipi di consultazioni: query combinata e a testo libero.

La **query a testo libero** è molto semplice, infatti permette all'utente di inserire un testo a piacere per la ricerca ed impostare il campo dove andare a cercare questo testo, è possibile cercarlo nei titoli dei progetti, nelle parole chiave, nei nomi delle unità organizzative, delle persone o delle città. È inoltre possibile impostare l'anno entro cui è stato terminato il progetto.

La **query combinata** consente invece una ricerca più specifica in quanto permette di impostare molti parametri, per evitare errori è inoltre impossibile inserire manualmente il testo da cercare, questo infatti deve essere selezionato da una lista contenete tutti i possibili valori presenti nel database. Il primo parametro che si può impostare è l'identificativo del programma di finanziamento, che permette di identificare univocamente un progetto in base al programma di finanziamento di cui gode (questo

testo viene inserito manualmente). Viene poi chiesta la città in cui è stato sviluppato il progetto, non è obbligatorio inserire proprio una città ma è possibile indicare anche una regione, per comprendere più progetti. Infatti, aprendo la lista di possibili valori, si presentano le varie regioni ungheresi, è possibile selezionare una di esse oppure passare alle città per aumentare la precisione della ricerca. È possibile poi selezionare una persona che ha partecipato al progetto, in questo caso si presenta una lista ordinata alfabeticamente di tutte le persone contenute nel database, in questa lista sono presenti solamente il nome, il cognome, l'indirizzo e-mail e il titolo accademico. Successivamente si può selezionare un'unità organizzativa, anche in questo caso si presenta una lista ordinata alfabeticamente di tutte le unità organizzative. E poi possibile selezionare la categoria di appartenenza del progetto, per effettuare questa selezione è possibile scegliere di utilizzare uno tra i due standard proposti: quello del governo ungherese (scritto in ungherese) e quello di Ortelius Thesaurus (in inglese), quest'ultimo per favorire la comprensione agli utenti stranieri. Le possibili categorie di Ortelius Thesaurus sono:

- Health sciences;
- Humanities;
- Natural sciences;
- Physical sciences;
- Social sciences;
- Technological sciences.

Per ognuna di queste categorie di base sono poi presenti varie sottocategorie e, per ognuna, altre sottocategorie, questo permette di restringere la ricerca del progetto con grande precisione. L'ultimo parametro di ricerca che è possibile impostare riguarda la parole chiave, queste sono suddivise in base alla categoria di appartenenza del progetto.

La pagina web contenente i risultati della ricerca offre meno informazioni rispetto a quella di EUREKA, in questo caso infatti consiste solamente nell'elenco dei titoli dei progetti trovati (questi sono di colore diverso a seconda se sono già stati terminati oppure sono ancora in corso) e, per ognuno, l'elenco delle informazioni contenute; in pratica è indicato se è presente solo la descrizione del progetto oppure anche le informazioni riguardanti i risultati ottenuti.

Cliccando sul titolo di un progetto si accede alla relativa pagina principale (vedi figura 4.3), questa non contiene alcuna informazione relativa ad esso, si tratta invece di un elenco delle informazioni relative a quel progetto presenti nel database. Nella parte superiore è indicato se sono presenti o meno le informazioni generali del progetto (sono presenti in qualsiasi progetto) e quelle relative ai risultati (sono presenti solo in alcuni progetti finiti). Nella parte centrale prendono posto le unità organizzative che hanno collaborato al progetto, mentre in quella inferiore le persone. I pulsanti di scelta posizionati sulla sinistra servono a selezionare l'entità che si desidera vedere, nell'esempio in figura si vede che non sono ancora presenti nel database le informazioni sui risultati del progetto (not available).

Aprendo la pagina relativa ad una persona appaiono tutte le informazioni relative ad essa, queste sono abbastanza misere, in quanto corrispondono soltanto a quelle di livello 1 del modello proposto da CERIF. Per prima cosa è indicato il ruolo di questa nello sviluppo del progetto e la durata del suo incarico, poi viene riportato il suo ruolo all'interno dell'organizzazione amministrativa. In seguito sono riportati i dati anagrafici

e, se presente, l'home page, non sono presenti informazioni per poter contattare questa persona.

Project description form	Project results form
<input type="radio"/> Open	<input type="radio"/> Not available
<hr/>	
Organisation data form	
<input type="radio"/> Department of Materials Handling and Logistics <i>University of Miskolc/Faculty of Mechanical Engineering/</i>	Open
<hr/>	
Personal data form	
<input type="radio"/> <u>Bazsó Fülöp PhD</u> <i>University of Miskolc/Faculty of Mechanical Engineering/ Department of Materials Handling and Logistics</i>	Open
<input type="radio"/> <u>Érdi Péter a kémiai tudományok doktora</u> <i>University of Miskolc/Faculty of Mechanical Engineering/ Department of Materials Handling and Logistics</i>	Open

Figura 4.3: esempio della pagina principale di un progetto in HunCRIS.

Nella figura 4.4 viene mostrato un esempio di pagina relativa ad un'unità organizzativa, come si vede è divisa in due parti: nella prima è presente solamente il nome dell'organizzazione, suddiviso in livelli (università, facoltà, dipartimento), nel secondo il ruolo assunto da essa nello sviluppo del progetto. Come si vede, a fianco di ogni nome di unità organizzativa è presente un bottone, questo serve per accedere alle informazioni dettagliate dell'entità relativa, ovvero il nome (in ungherese e inglese), l'indirizzo, i numeri di telefono e fax, l'home page e l'indirizzo e-mail.

Le informazioni relative ad un progetto sono divise in due parti: quelle generali e quelle riguardanti i risultati ottenuti. La pagina relativa alle informazioni generali è suddivisa in quattro parti: nella prima è presente soltanto il nome del progetto (in ungherese e inglese), nella seconda parte sono riportati il nome abbreviato (in ungherese e inglese) e gli identificativi del programma di finanziamenti ricevuto e del contratto relativo ad esso. La terza parte contiene solamente il riassunto del progetto, anch'esso in ungherese e in inglese, nell'ultima parte invece sono presenti le parole chiave (in entrambe le lingue), le classificazioni secondo lo standard del governo ungherese e secondo Ortelius Thesaurus (in entrambe le lingue), le date di inizio e di fine e l'home page. La pagina relativa ai risultati del progetto è molto simile alla precedente, infatti è anch'essa suddivisa in quattro parti di cui le prime due sono identiche a quelle descritte prima, in questo caso, però, nella terza parte è presente il riassunto dei risultati ottenuti

(in ungherese e in inglese), e nella quarta il numero di brevetti e pubblicazioni relative ad esso.

2/1

a) Organisation (unit) data

Name of organisation (unit): Budapest University of Technology and Economics	<input type="button" value="Org/unit"/>
Name of organisation (unit): Faculty of Civil Engineering	<input type="button" value="Org/unit"/>
Name of organisation (unit): Department of Structural Engineering	<input type="button" value="Org/unit"/>

2/2

b) Role of organisation (unit) in the project

Role of organisation (unit)

Contractor
▼

Figura 4.4: esempio della pagina di un'unità organizzativa in HunCRIS.

Le informazioni contenute in questo database non sono molto dettagliate, infatti per le persone ci si limita ai dati anagrafici (livello 1), mentre per le unità organizzative vengono aggiunte anche quelle per poter contattare la relativa entità, per quanto riguarda i progetti, vengono fornite più informazioni, come le parole chiave, il riassunto e i risultati. Si noti come questo sistema fornisca un ottimo approccio plurilingue, infatti quasi tutte le informazioni contenute sono presenti sia in ungherese che in inglese, risultano quindi comprensibili alla grande maggioranza degli utenti del mondo.

Come si vede dalle informazioni contenute, questo sistema è molto diverso da EUREKA, infatti nel primo esempio si puntava molto sulle potenzialità commerciali di un progetto, in questo invece si punta sulla diffusione dei progetti, il lato commerciale non viene neanche preso in considerazione.

4.3 The Projekt database

Questo database contiene i progetti realizzati all'interno di un'università finlandese, la Åbo Akademi University; l'amministrazione e l'aggiornamento sono gestiti dalla biblioteca della stessa università. Questo sistema di informazioni è stato realizzato nel 1998, utilizza quindi il modello di dati proposto per CERIF 1991. Nel database infatti sono descritti i vari progetti realizzati in modo dettagliato, mentre per quanto riguarda le informazioni sulle persone e sulle unità organizzative (in questo caso i vari dipartimenti dell'università), queste comprendono solamente i rispettivi nomi. Nella spiegazione dei progetti, invece, sono contenuti:

- il nome del progetto;
- il leader del progetto ed i ricercatori che vi hanno collaborato;
- la durata approssimativa;
- i finanziamenti ricevuti;
- i partner;
- i codici di classificazione;
- le parole chiave;
- un riassunto.

Le informazioni contenute in questo database sono principalmente in svedese (che è la seconda lingua in Finlandia), ma quelle più importanti, come il nome del progetto, il riassunto, e le parole chiave, sono tradotte anche in inglese.

Per inserire le informazioni riguardanti un progetto all'interno del database è necessario compilare un modulo, quest'ultimo però è disponibile solamente dal dominio "abo.fi" (dominio dell'università), questo per evitare l'intromissione di progetti realizzati al di fuori dell'università.

Searching the Projekt database

Research projects at Åbo Akademi University

Freetext:

Project: **INDEX**

Project leader: **INDEX**

Faculty: **INDEX**

Department: **INDEX**

Keywords: **INDEX**

Starting date of the project:

You can combine the search terms with **AND** Sort order: **Project title**

Figura 4.5: form per la ricerca in Project Database.

Come si vede dalla figura 4.5, la form per la ricerca dei progetti è molto semplice, questa infatti comprende un casella per una query a testo libero, che cerca il testo inserito in tutti i campi del database, una casella di testo per impostare la data di inizio del progetto che si vuole recuperare e vari campi in cui è possibile utilizzare un indice. Questo è molto utile nel caso non si conoscano, per esempio, i nomi delle facoltà o quello dei leader dei progetti, infatti grazie al pulsante "INDEX" corrispondente si

accede ad una lista completa di tutti i possibili valori di quel campo contenuti nel database ordinati alfabeticamente, con l'aggiunta del numero di volte in cui quel valore compare nel database.

Correct Component Software (CoCoS)

Date: 1.1.1996-31.12.1998
Code: 2665
Department: Åbo Akademi University / Faculty of Mathematics and Natural Sciences (MNF), Dept. of Computer Science
Address: Lemminkäinenkatu 14, FIN-20520 Åbo
 Phone +358-2-2154 673, Fax +358-2-2154 732
 E-mail Jockum.Wright@abo.fi
Project leader: FD Ralph-Johan Back, professor (1.1.1996-31.12.1998)
Researchers: FD Kaisa Sere (1.1.1996-31.12.1998)
 FD Joakim von Wright (1.1.1996-31.12.1998)
Type of research: 0 (0=Within duty, 1=Ordered research, 2=Co-operation)
 - basic research 100 %
Finnish funding organizations: Finlands Akademi, Helsingfors () FIM 1400000
Man months: 12 months in 1998; Totally: 36 months
Keywords: programmering, ohjelmointi, mjukvarokomponenter, formella metoder, preciseringskalkyl, mekanisk verifiering, inbäddade system, software components,

The project will investigate the theory [...]

[Publication\(s\)](#)

Figura 4.6: esempio della pagina principale di un progetto in Project Database.

Procedo ora con l'analisi della pagina web che contiene le varie informazioni riguardanti un progetto (vedi figura 4.6), queste informazioni sono abbastanza dettagliate, anche se non spiegano molto il lavoro che è stato fatto. Le date di inizio e di fine del progetto sono messe come prima informazione, questo per far capire subito se il progetto è stato realizzato da poco oppure, come in questo caso, è un po' vecchio. Successivamente sono mostrati il codice che identifica univocamente ogni progetto, il dipartimento che lo ha realizzato e le varie informazioni per poterlo contattare: indirizzo, telefono, fax ed e-mail. Sono poi elencate le persone che hanno collaborato alla realizzazione del progetto, ovvero il leader e i ricercatori, per ognuno di essi vengono indicate tra parentesi le date di inizio e di fine della loro partecipazione al progetto. Queste sono le uniche informazioni che riguardano le persone riportate nel database, questo perché il sistema utilizza il modello di CERIF 1991; come già detto in precedenza questo modello lasciava poco spazio alle persone e alle unità organizzative, ed è questo uno dei motivi che hanno poi portato alla sua revisione. In seguito viene spiegato il tipo di ricerche effettuate, per indicarlo viene utilizzato un codice con tre possibili valori (ricerca dovuta, ordinata o cooperazione); è importante sottolineare che

viene poi riportata anche la legenda per una maggiore comprensibilità. La percentuale che è presente successivamente indica la quantità di ricerche di base e di ricerche applicate che il progetto ha richiesto, nell'esempio sono state svolte solamente ricerche di base. Vengono poi riportate le informazioni sui programmi di finanziamenti ricevuti, ovvero la somma ricevuta e i finanziatori, questi ultimi possono essere l'università stessa, altre università (come nell'esempio) oppure aziende. Sono poi mostrati la durata del progetto, i partner (nell'esempio non sono presenti) e le parole chiave utili per il recupero di esso durante le ricerche. Infine viene riportato il riassunto del progetto, questo è molto breve e comprende una spiegazione approssimativa che non entra nei particolari, inoltre non vengono neppure accennati i risultati ottenuti. Questi potrebbero essere esposti in alcune pubblicazioni riguardanti il progetto, tramite il link posto a fondo pagina è possibile vedere l'elenco completo.

Analizzando le informazioni riportate nel Project Database della Åbo Akademi University si nota che viene data poca importanza alle persone e alle unità organizzative, vengono invece forniti molti dati riguardanti i progetti; questi però servono solamente a dare un'idea di base riguardo a quello che è stato realizzato, in quanto manca completamente una spiegazione dettagliata. Nel database mancano inoltre le informazioni riguardanti i prodotti che potrebbero essere realizzati e gli eventuali brevetti, viene invece fornito un buon approccio plurilingue infatti, a parte i nomi delle persone e delle unità organizzative, tutte le altre informazioni sono tradotte in inglese.

4.4 WISARD

WISARD è un CRIS utilizzato per condividere le informazioni sulle organizzazioni, gli esperti, i progetti e le produzioni relative (documenti, articoli, pagine web) riguardanti l'agricoltura, la ricerca e lo sviluppo rurale. Questo è un sistema decentralizzato, quindi chiunque lo può utilizzare facilmente per pubblicare informazioni o controllare quelle già pubblicate; inoltre fornisce agli utenti un set di tool molto utili, come la classificazione e la traduzione automatica in varie lingue. WISARD utilizza un modello di dati che si interfaccia con i più importanti standard internazionali, tra cui CERIF; questo modello di dati contiene informazioni dettagliate non solo sui progetti, ma anche sulle produzioni relative ad essi, sulle persone e sulle unità organizzative che vi hanno collaborato. Il sistema di ricerca proposto è suddiviso in base all'entità che si vuole recuperare, infatti dalla home page si può scegliere tra una delle quattro ricerche possibili: progetto, persona, organizzazione o produzione.

Il sistema di ricerca di un progetto è mostrato in figura 4.7, questo sistema è molto completo e comprende tre tipi di ricerca:

1. ricerche pre-strutturate: come si vede sono possibili ricerche per classificazione, per stato in cui è stato svolto il progetto, per nazionalità dell'organizzazione, per organizzazione o per il termine "agrovoc" (Agrovoc è un thesaurus che comprende tutti i termini relativi all'agricoltura). Questo tipo di ricerca è molto semplice e veloce da utilizzare in quanto, per esempio, fornisce un elenco di tutti i progetti realizzati in un determinato stato, o che appartengono ad una determinata categoria.

2. ricerca a testo libero: ricerca semplice che cerca una o più parole inserite in tutti i campi relativi all'entità progetto.
3. ricerca avanzata: è abbastanza simile alle ricerche pre-strutturate, infatti è composta da vari menù a tendina tramite cui l'utente può selezionare tutti i progetti con una stessa caratteristica, la differenza consiste nel fatto che si possono combinare più parametri nella stessa ricerca. I parametri che si possono scegliere sono: la locazione, lo stato (in corso o terminato), oppure la classificazione in base a quattro criteri: ambito biofisico, ambito socio-economico, catena di produzione, gruppi di prodotti realizzati.

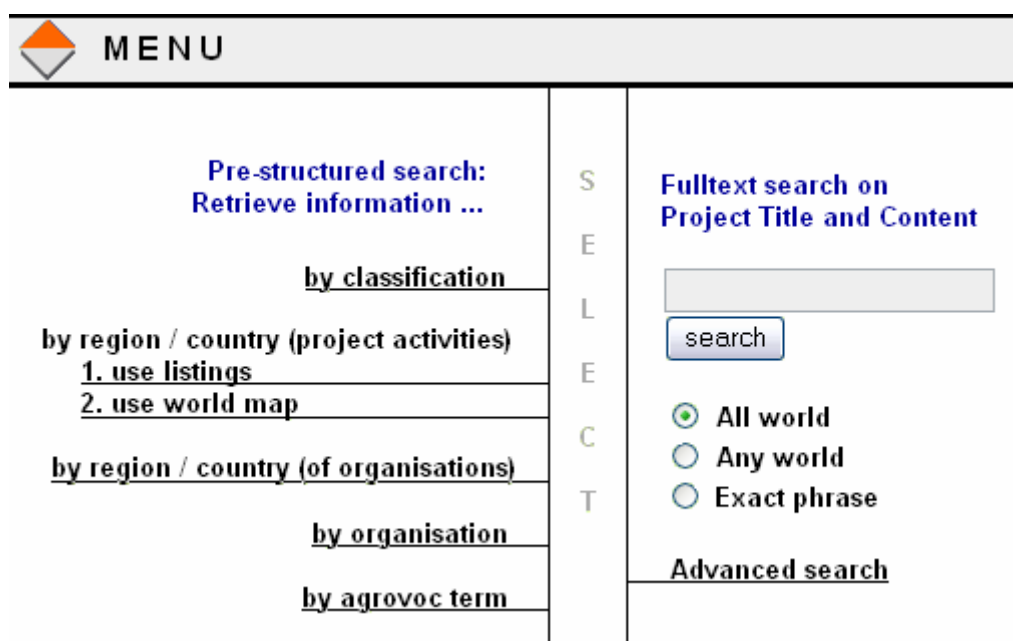


Figura 4.7: sistema di ricerca dei progetti in WISARD.

Il sistema di ricerca di una persona è analogo a quello appena descritto per i progetti, cambiano solamente i parametri di ricerca che si possono impostare. Nelle ricerche pre-strutturate è possibile scegliere il nome (si deve inserire nome e/o cognome della persona), la classificazione (esperto in geografia, educazione, ecc.), la nazione di residenza o l'organizzazione di appartenenza. Nella ricerca avanzata, invece, è possibile scegliere la nazione di residenza e la materia in cui la persona è specializzata, i criteri di classificazione sono gli stessi di prima con l'aggiunta dell'ambito geografico.

Anche il sistema di ricerca delle organizzazioni è simile a quelli appena descritti; in questo caso le ricerche pre-strutturate a disposizione sono per acronimo (scegliendo l'iniziale di questo), per profilo ARD (categoria di appartenenza, es. università, centro di ricerca, ecc.) o per mandato (ambito di lavoro, es. ricerca, pubblica amministrazione, ecc.), in questi due casi è necessario prima selezionare la nazione di appartenenza e poi la caratteristica. Nel caso delle organizzazioni non è prevista la ricerca avanzata.

Il sistema di ricerca delle produzioni è completamente differente dagli altri, infatti queste entità non sono collegate alle altre; esse sono contenute nel database AGRIS ("international information system for the agricultural sciences and technology"), ovvero una raccolta di documenti che copre tutti gli aspetti della scienza e della

tecnologia agricola. In questo sistema sono possibili due sole ricerche, oltre alla solita ricerca a testo libero, ne viene proposta una avanzata. Nella ricerca avanzata è possibile inserire alcune parole e ricercarle in un campo specifico del database, i campi su cui è possibile effettuare le ricerche sono: il riassunto, il nome dell'autore, l'organizzazione a cui appartiene l'autore, la data o l'anno di pubblicazione, il titolo della rivista (se sono stati pubblicati su di una rivista), l'editore, le parole chiave e il titolo.

Sustainable control of the horse chestnut leaf-miner	
Funding	- BBW (Swiss Federal Office for Education and Science) , Switzerland
Coordination	- CABI-SWITZ (CABI: CABI international offices: CABI Bioscience Switzerland Centre)
Location(s)	Austria Bulgaria Czech Republic France Germany Greece Italy Switzerland
Duration	from January 2001 until December 2004
Size	not known

Type of funding

European Commission Fifth Framework Programme.

Collaborating institution(s)

Germany, Italy, Switzerland, Austria, Czech Republic, France, Greece, Bulgaria

Background

The horse-chestnut leafminer, 'Cameraria ohridella', is a moth of unknown origin which has spread recently over vast areas of Europe. In some countries such as Austria or the Czech Republic, [...]

Figura 4.8: esempio di una parte della pagina principale di un progetto in WISARD.

Procedo adesso con l'analisi della pagina principale di un progetto, un esempio di una di queste è presentato nella figura 4.8. Le pagine principali dei progetti sono molto complete e contengono informazioni dettagliate; la prima parte è schematica e contiene le informazioni generali riguardanti il progetto, ovvero il nome, la o le organizzazioni che lo finanziano, l'organizzazione coordinatrice, le locazioni in cui si svolge, la durata e la dimensione del progetto (questa dipende dal numero di persone che vi lavorano e dai soldi spesi per svilupparlo), è poi possibile trovare un elenco delle organizzazioni che hanno collaborato al progetto (nell'esempio di figura non è presente). La seconda parte, invece, è maggiormente descrittiva e punta a esporre in che cosa consiste il progetto, in figura 4.8 ne è riportata solamente una parte. Le prime informazioni presenti riguardano il tipo di finanziamento ricevuto, il budget a disposizione e le istituzioni che hanno collaborato al progetto. Il campo chiamato "Background" serve poi a dare informazioni preliminari sull'argomento trattato, queste sono molto utili in quanto permettono anche gli utenti non esperti di capire perché è stato fatto il progetto. In seguito è presente il campo "Objectives", che spiega a che cosa si vuole arrivare

grazie alla realizzazione di questo progetto. Successivamente sono presenti i campi “Achievements”, che illustra i successi raggiunti grazie a questo progetto, e “(Planned) Outputs”, che spiega quali documenti o prodotti riguardanti il progetto sono stati realizzati, ovviamente questi campi sono presenti solamente nei progetti già terminati. In seguito vengono riportate le parole chiave, i materiali utilizzati, oppure il link al sito che li produce, e la classificazione. Quest’ultimo campo è suddiviso in tre parti:

1. General ARD: spiega il tipo di attività svolta (es. ricerca, laboratorio, comunicazione);
2. Thematic focus: riguarda i criteri spiegati per la ricerca avanzata dei progetti, è possibile inserire un valore per ognuno di essi;
3. AGRIS/CARIS research themes: classificazione secondo questo thesaurus.

Vengono poi riportati i nomi dei vari esperti che hanno collaborato alla realizzazione del progetto, è importante notare che questi nomi e tutti quelli delle unità organizzative riportati nelle pagine dei progetti non sono semplice testo, ma sono link all’entità corrispondente, questo favorisce una maggiore velocità di spostamento tra le varie entità collegate. Vengono infine riportati il nome della persona che ha inserito queste informazioni e le date di inserimento e di ultima modifica.

La struttura delle pagine web relative alle persone è uguale a quella delle pagine relative ai progetti, infatti anche queste sono divise in due parti. Nella prima parte sono riportate le informazioni principali di una persona, ovvero l’organizzazione a cui appartiene, il suo indirizzo, i numeri di telefono e fax e l’e-mail. La seconda parte invece contiene le informazioni secondarie, ovvero le competenze e le esperienze passate della persona, le conoscenze linguistiche, l’elenco delle pubblicazioni scritte e l’elenco dei progetti realizzati (questi ultimi sono link a pagine di progetti in WISARD). È poi presente la classificazione della persona, in questo caso il campo è diviso in quattro parti:

1. Education: riguarda le discipline studiate e le lauree possedute;
2. Employment: contiene le informazioni sull’occupazione della persona;
3. Thematic expertise: riporta le competenze della persona secondo i criteri spiegati per la ricerca avanzata (Biophysical, Socio-economic, Production chain e Target commodity group);
4. Geographic expertise: riguarda le informazioni sulle competenze geografiche della persona (climi, altitudini, umidità).

Anche la struttura delle pagine web relative alle unità organizzative è uguale alle altre; nella prima parte sono riportate le informazioni principali, ovvero il nome, l’acronimo, l’indirizzo, i numeri di telefono e fax, l’e-mail e l’home page, vi è poi un utilissimo link ad una pagina web contenente tutti i progetti a cui l’organizzazione ha collaborato. Nella seconda parte invece sono presenti informazioni secondarie; inizialmente viene mostrato se l’organizzazione possiede delle sotto-unità organizzative o se è essa stessa una sotto unità (es. una università ha come sotto-unità una facoltà, che ha come sotto-unità un dipartimento). Sono poi presenti vari campi che descrivono dettagliatamente le varie organizzazioni: “Description” serve a spiegare in che cosa consiste e cosa fa l’organizzazione, in “Mission statement” vengono illustrati gli scopi dei progetti da essa realizzati, in “Sections” invece le varie sezioni e i vari reparti in cui è suddivisa ed i loro compiti, “Focus” infine descrive l’obiettivo principale dell’unità. È poi presente la classificazione, in questo caso il campo è diviso in tre parti:

1. Legal status: riporta lo stato legale dell'organizzazione (es. governativa, privata);
2. Mandate / focus of activity: spiega di che cosa si occupa (es. ricerca, pubblica amministrazione);
3. Sector: illustra il settore a cui appartiene (es. agricoltura, conservazione della natura).

Come già detto in precedenza, le produzioni si trovano su un altro database, quindi la pagina web che le contiene è differente, innanzi tutto in questa pagina non è contenuto il documento vero e proprio, ma solamente le informazioni riguardanti esso; viene però fornita la possibilità, tramite un link, di effettuare una ricerca utilizzando Google per recuperarlo.

La struttura di questa pagina web è molto schematica, in essa sono contenute le seguenti informazioni: numero di record (identifica univocamente il documento), titolo, autori, data e luogo di pubblicazione, nome dell'editore, categoria di appartenenza (utilizzando sia la classificazione AGRIS che quella "agrovoc"), lingua, note, riassunto e giornale o rivista di pubblicazione (se è stato pubblicato su una di esse), con anche il numero e la data della pubblicazione.

Analizzando le informazioni relative alle varie entità contenute nel database si nota che, per quanto riguarda i progetti, le persone e le unità organizzative, viene utilizzato il modello di CERIF quasi nella sua completezza; il risultato è che le informazioni sono complete ed esaurienti, in quanto spiegano dettagliatamente le caratteristiche delle varie entità. Per quanto riguarda le produzioni, anche in questo caso vengono riportate molte informazioni, però l'unico campo "interessante" per capire di cosa tratta il documento è il riassunto, il quale non sempre è presente.

L'ultima considerazione su questo database riguarda il supporto plurilingue, infatti tutte le pagine e le informazioni del sito web sono in inglese, e già questo potrebbe andare bene, vi è però un ulteriore tool che è possibile utilizzare: un traduttore on-line che traduce le pagine web dei progetti in francese, tedesco, spagnolo, portoghese, italiano, cinese, giapponese e coreano ("AltaVisata Babelfish Service").

Conclusioni

In questo lavoro di tesi è stata fornita una panoramica completa riguardo CERIF, in particolare sono state spiegate in dettaglio tutte le entità e gli attributi che compongono il modello di dati con le relative relazioni. Sono stati illustrati inoltre i possibili approcci per sviluppare il sistema di ricerca delle informazioni e le possibili estensioni future al modello di dati. Sono poi state spiegate le operazioni necessarie per rendere un CRIS compatibile con CERIF; sono infine stati analizzati alcuni esempi pratici di CRIS che utilizzano il modello di CERIF.

Tutto questo lavoro di analisi potrebbe servire in futuro per la creazione di un CRIS che contenga i vari progetti di ricerca realizzati all'interno dell'Università degli Studi di Parma. Questo CRIS dovrebbe essere compatibile con il modello di CERIF in modo da permettere una maggiore visibilità ai progetti anche a livello europeo e fornire la possibilità di uno scambio di informazioni con altri sistemi compatibili con CERIF.

In particolare, per la creazione di questo CRIS si potrebbe prendere spunto dal "Project Database" realizzato alla Åbo Akademi University, questo sistema è molto semplice da implementare e fornisce un'interfaccia utente intuitiva, è inoltre perfettamente compatibile con CERIF, ma il motivo principale che lo rende l'esempio ideale da cui partire per realizzare il CRIS è la sua dimensione universitaria.

Glossario

AGRIS: sistema di informazioni internazionale per le scienze e le tecnologie agricole creato dalla FAO.

Agrovoc: thesaurus che comprende tutti i termini relativi all'agricoltura.

CERIF: Common European Research Information Format.

Classificazione: un linguaggio di informazioni che usa termini artificiali (codici o notazioni), la maggior parte delle classificazioni ha una struttura gerarchica chiara con relazioni gerarchiche tra le notazioni. Una classificazione consiste di classi che rappresentano una raccolta di "soggetti" con una caratteristica comune.

CRIS: Current Research Information Systems.

Entità di lookup: campo di dati per il quale il valore deve essere selezionato da un elenco di valori predefiniti.

HTTP: Hyper Text Transport Protocol, questo protocollo è alla base del WWW.

Indicizzazione: è un sistema di linguaggio che permette la rappresentazione dei contenuti dei documenti o delle entità di informazioni in un modo analitico (specificamente) o in un modo sintetico, questo ha lo scopo di trovare documenti pertinenti o informazioni in risposta a domande basate sui loro contenuti.

Indicizzazione per soggetto: processo che consiste in un'analisi ragionata del contenuto di un documento al fine di identificare i concetti in esso presenti, ed assegnare loro una qualche etichetta testuale, per permettere successivamente un rapido ritrovamento dell'informazione per un uso futuro.

ISO: International Standard Organization.

Metadata: informazione strutturata riguardo altre informazioni.

R&S: Ricerca & Sviluppo.

RDF: Resource Description Framework.

SOAP: Simple Object Access Protocol.

Thesaurus: vocabolario controllato dinamico, mono o plurilingue, che copre gli specifici domini della conoscenza, e viene usato per tradurre i documenti dalla lingua originale in altre lingue, la più importante funzione di un thesaurus è fornire un attrezzo nel reperimento delle informazioni.

UE: Unione Europea.

URI: Uniform Resource Identifier.

Vocabolario controllato: un elenco di termini (descrittori o notazioni) usati nell'indicizzazione per soggetto, tra i termini sono indicate anche relazioni semantiche e gerarchiche. Ci sono due tipi di vocabolario controllato: classificazione e thesaurus.

XML: eXtensible Mark-up Language.

Bibliografia

- A. Asserson, K. Jeffery, “*CERIF Course*”, 2002,
www.ub.uib.no/avdeling/fdok/cris/taskgroups/presentations.htm.
- A. Asserson, K. Jeffery, “*CERIF Interoperability and Compatibility*”, 2002,
www.ub.uib.no/avdeling/fdok/cris/taskgroups.
- A. Asserson, K. Jeffery, A. Lopatenko, “*CERIF - Information Retrieval of Research Information in a Distributed Heterogeneous Environment*”, 2002,
www.ub.uib.no/avdeling/fdok/cris/taskgroups.
- A. Asserson, K. Jeffery, A. Lopatenko, “*CERIF: Past, Present and Future: An Overview*”, 2002, www.ub.uib.no/avdeling/fdok/cris/taskgroups.
- A. Asserson, K. Jeffery, A. Lopatenko, “*Comparative Study of Metadata for Scientific Information: The place of CERIF in CRISs and Scientific Repositories*”, 2002,
www.ub.uib.no/avdeling/fdok/cris/taskgroups.
- A. Asserson, K. Jeffery, “*Full CRIS data model*”, 2002,
www.ub.uib.no/avdeling/fdok/cris/taskgroups/deliverables.htm.
- A. Asserson, K. Jeffery, “*Metadata data model*”, 2002,
www.ub.uib.no/avdeling/fdok/cris/taskgroups/deliverables.htm.
- Commissione della Comunità Europea, “*CERIF 1991 manual*”, 1991,
[ftp.cordis.lu/pub/cerif/docs/cerif1991.htm](ftp://cordis.lu/pub/cerif/docs/cerif1991.htm).
- CERIF Revision Working Group, “*CERIF 2000 guidelines*”,
[ftp.cordis.europa.eu/pub/cerif/docs/cerif2000.htm](ftp://cordis.europa.eu/pub/cerif/docs/cerif2000.htm).
- CERIF Task Group, *Home page*,
www.ub.uib.no/avdeling/fdok/cris/taskgroups/Index6.htm.
- EUREKA, *Home page*: www.eureka.be.
- HunCRIS, *Home page*: www.info.omikk.bme.hu/nkr1/HunCRIS_eng.htm.
- K. Jeffery, “*Achievements and Experience of the euroCRIS Organisation*”, 2005,
pubs.cclrc.ac.uk/bitstream/990/ExperienceAchievementsEuroCRIS.ppt.
- The Project Database, *Home page*: trip.abo.fi/aadb/projekt/infoe.htm.
- WISARD, *Home page*: www.wisard.org.