

CAPITOLO SECONDO

Individuazione dell'ambiente e delle aree di intervento

2.1. PREMESSA

La natura delle informazioni oggetto di questo studio di fattibilità è distribuita tra più ambienti informatici ed è quindi opportuno:

- evitare duplicazioni di dati presso i vari sistemi informativi ed uffici, operando con un ambiente fortemente integrato in cui si instaurano collegamenti logici tra i diversi sistemi senza la necessità di duplicazioni che rendono problematico l'allineamento delle informazioni;
- organizzare le operazioni sui dati in modo che possano essere eseguite da tutti i soggetti interessati, senza che ogni ambiente debba realizzare in proprio applicazioni già realizzate da altri e con la garanzia che il risultato della stessa operazione sia sempre identico qualunque sia il soggetto e l'ambiente da cui viene invocata.

Queste considerazioni impongono scelte tecniche che consentano la condivisione di dati e applicazioni tra ambienti informatici diversi e, in generale, un coordinamento dei processi di automazione volto a garantire la compatibilità, a razionalizzare l'ambiente evitando sprechi di risorse, ad abilitare la cooperazione applicativa.

In altri termini è necessaria una visione progettuale che abbracci i problemi nella loro globalità, non limitata alle esigenze di un singolo soggetto e ai vincoli di un unico ambiente informatico

Un approccio di questo tipo impone anche una standardizzazione dei meccanismi di colloquio tra ambienti informatici. L'infrastruttura tecnologica per tale standardizzazione può essere individuata nella Rete unitaria della pubblica amministrazione (RUPA). Essendo i sistemi informativi del Consiglio di Stato (CdS) e dei Tribunali amministrativi regionali (TAR) omogenei può essere conveniente prevedere una *intranet* con un'unica porta applicativa sulla RUPA.

Le azioni sul piano tecnico rappresentano una condizione necessaria per la collaborazione applicativa, perché mettono a disposizione gli strumenti, ma è comunque necessaria un'azione di indirizzo e di coordinamento dell'iniziativa per garantirne il successo e per ridurre i tempi di adozione delle procedure automatizzate. È infatti necessario iniziare a modificare comportamenti e modi di operare consolidati, per utilizzare al meglio le nuove funzionalità che i sistemi automatizzati mettono a disposizione. È importante ricordare che l'automazione non deve essere

intesa soltanto come una modalità diversa, più efficiente e sicura, di svolgere operazioni attualmente eseguite in modo manuale, ma soprattutto come una tecnologia che rende possibili nuovi modi di operare realizzando una progressiva evoluzione che, in termini aziendali, è conosciuta con il termine *business process re-engineering* cioè una riprogettazione del modo di perseguire le finalità istituzionali.

Questa evoluzione non si esaurisce con gli aspetti tecnici informatici, ma coinvolge aspetti organizzativi e normativi che esulano dagli obiettivi di questo studio e che devono essere valutati e adottati con gradualità, facendo evolvere di pari passo organizzazione e informatizzazione.

2.2. AMBIENTE TECNOLOGICO

Nell'ambito dell'intero sistema esistono diversi tipi di documenti: quelli prodotti direttamente dal CdS e dai TAR (provvedimenti), quelli prodotti dagli avvocati (ricorsi, memorie, istanze, documenti allegati, ecc.), e quelli ai quali i magistrati accedono per la loro attività (normativa, giurisprudenza). Inoltre deve essere tenuta traccia di ogni transizione ed evento, che si verifichi all'interno del processo amministrativo, mediante memorizzazione di opportuni dati: funzionalità realizzata dalle attuali procedure di automazione dei registri.

Gli interventi per una gestione corretta ed efficace dei vari tipi di documenti e/o dati sono differenziati in base alla loro origine e tipologia, e condizionati da vincoli organizzativi. Indipendentemente dalle specifiche tipologie di dati, va comunque tenuto presente che il successo di un'azione di informatizzazione è sempre fortemente condizionato da una accurata progettazione del sistema informatico di gestione dei dati e da una opportuna opera di strutturazione del dato.

All'interno del CdS e dei TAR è necessario prevedere sistemi informatici basati su due grandi categorie tecnologiche:

- sistemi RDBS (*Relational Data Base System*), per la rappresentazione di dati strutturati, quali sono i registri;
- sistemi IRS (*Information Retrieval System*), per la gestione ed il reperimento di documenti testuali.

La categoria dei RDBS, che comprende Oracle v. 8, rappresenta oggi lo standard più diffuso per la realizzazione di basi di dati strutturati. Rispetto ad altri modelli esso offre il vantaggio di consentire la realizzazione di basi di dati che rispecchiano il modello concettuale della realtà rappresentata, modello che risulta da un insieme di entità e di loro associazioni.

La progettazione di basi di dati relazionali è quindi basata sulla definizione di uno schema concettuale (*class diagram*) ovvero una descrizione ad alto livello indi-

pendente dal particolare *software* adottato per la gestione del *database*, del contenuto informativo delle basi di dati. Un aspetto rilevante è costituito dal fatto che lo schema concettuale è da ritenersi invariante per un periodo medio-lungo, in quanto una sua modifica è prevedibile solo a fronte di mutamenti della realtà circostante o del tipo di attività svolta dall'organizzazione, e quindi la progettazione dello schema concettuale può riuscire utile anche nei casi in cui alla fine l'implementazione del sistema informativo non faccia ricorso alla tecnologia *database*.

Nello studio di fattibilità è stato prodotto uno schema concettuale per l'attività giurisdizionale del TAR e del CdS, che deve essere considerato un esempio, necessariamente non completo, del modo di operare che deve essere usato nello sviluppo dei nuovi sistemi.

La categoria dei sistemi di *information retrieval* (IRS) è invece specifica per la gestione ed il reperimento di documenti testuali. Tradizionalmente, l'efficacia di un sistema di reperimento dell'informazione si misura in base al valore di due parametri: la precisione (cioè la percentuale di documenti ritrovati e pertinenti rispetto alla totalità di documenti ritrovati), e il richiamo (cioè la percentuale di documenti ritrovati e pertinenti rispetto al numero di documenti pertinenti esistenti nell'intera collezione di documenti). Va sottolineato che l'efficacia di un sistema viene misurata tenendo conto di entrambi i parametri. È accertato che l'utilizzo di *thesauri* e dizionari per la formulazione ed il raffinamento delle richieste (*query*) contribuisce in maniera sostanziale al miglioramento di entrambi i parametri.

È ormai consuetudine realizzare interfacce di tipo ipertestuale verso basi di dati o sistemi di *information retrieval*, ma va tenuto presente che uno dei problemi connessi con questo tipo di approccio è il sovraccarico cognitivo. Una parte dell'informazione recuperata non interessa l'utente e determina il fastidioso effetto «rumore» nella ricerca. È quindi opportuno intervenire sull'insieme dei documenti reperiti mediante una adeguata operazione di «filtraggio», che deve avvenire sia a livello di selezione dei documenti reperiti dalla *query* che a livello di *link* attivi (*link* dinamici con tipo), e deve rispondere agli interessi specifici dell'utente, rappresentati mediante un opportuno profilo. Questo significa che il risultato di una *query* potrà modificarsi, al cambiare del profilo utente, sia come numero di documenti reperiti, sia come tipi di informazioni visualizzate o *link* attivi.

In estrema sintesi, comunque, si ritiene opportuno ricordare che i problemi affrontati si riconducono spesso a due tipologie: le modalità di strutturazione del documento e il metodo di indicizzazione.

Riguardo al primo aspetto, va identificata in modo consistente l'unità documentaria. Per quanto concerne il secondo aspetto, tutti si sono posti il problema di migliorare l'efficacia del sistema, misurata in base ai parametri di precisione e richiamo, uti-

lizzando opportuni strumenti di normalizzazione e organizzazione dei concetti da ricercare, affrontando il non lieve compito di indicizzare in maniera opportuna il documento. A fronte di questa unità di intenti, tuttavia, va segnalato che le modalità adottate per l'organizzazione delle conoscenze è diversa nei vari casi, e quindi si pone con significativa rilevanza il problema di interagire con basi di dati documentarie in cui i documenti sono classificati in base a schemi diversi. La difficoltà di interoperabilità è dovuta, evidentemente, non a semplici varianti linguistiche, ma a una reale differenza nella organizzazione della gerarchia dei concetti.

Il problema è stato affrontato e discusso anche in altri ambiti disciplinari, e si è constatata più di una volta la non riconciliabilità di schemi di classificazione differenti, soprattutto quando l'albero di classificazione raggiunge una profondità significativa, e quindi diventa molto sottile la differenza tra i vari termini. Inoltre, una classificazione troppo specifica richiede un livello di competenza elevato sia da parte dell'indicizzatore che da parte dell'utente, e rende assolutamente necessario disporre di adeguati meccanismi per la consultazione del *thesaurus*.

Un modo di operare consistente con le tendenze che vanno emergendo nell'area disciplinare dell'*information retrieval* e del reperimento dell'informazione distribuita in rete, e che potrebbe portare al superamento dei vari problemi, è quello che prevede di:

- strutturare i documenti in base a uno schema che abbia una valenza semantica, e non meramente sintattica;
- organizzare semanticamente i concetti da utilizzare per indicizzare i documenti;
- implementare opportuni meccanismi di restituzione delle informazioni che attivino collegamenti ipertestuali tra i documenti e ne permettano una visualizzazione personalizzata in base a un profilo utente che ne caratterizzi gli interessi.

2.3. PRIMA INDIVIDUAZIONE DELLE AREE DI INTERVENTO

Per una prima individuazione delle aree di intervento è necessario analizzare, seppur in maniera qualitativa, il flusso dei dati partendo dalle richieste provenienti dall'esterno verso i TAR od il CdS. Per quanto riguarda la situazione attuale del flusso dei dati sia presso i TAR che presso il CdS si può vedere la Figura 2.1, che prende in considerazione l'attività giurisdizionale.

La Figura 2.1 indica il flusso qualitativo dei dati, senza prendere in considerazione l'iter del ricorso. Come si può dedurre, l'automazione degli uffici riguarda essenzialmente la segreteria ed in particolare la tenuta dei registri. La mole di docu-

menti cartacei è molto elevata: il ricorso, le memorie, le note sono contenute nel fascicolo cartaceo, che accompagna tutto il procedimento.

La redazione dei provvedimenti è eseguita mediante videoscrittura, quale sostituto della macchina da scrivere, utilizzando il prodotto Microsoft Word. Sebbene redigere i provvedimenti in forma elettronica permetta una più facile correzione ed archiviazione degli stessi, il formato Word non è direttamente trasferibile su *World Wide Web* (*Web*) od in banche dati, essendo il formato proprietario della Microsoft.

Presso il CdS, una volta completato l'iter processuale, il provvedimento è archiviato, per la stragrande maggioranza quale immagine, mediante un sistema di archiviazione dei provvedimenti. Il sistema attualmente in uso non ha la potenza di ricerca di un *information retrieval* e non è, attualmente, consultabile dall'esterno. Questo limita di fatto la possibilità di reperimento dei precedenti.

Il CdS e i TAR hanno aggiornato completamente la loro strumentazione informatica sostituendo, quasi completamente, il loro parco macchine ed acquisendo nuovo *software* di gestione degli archivi. È, inoltre, intenzione del CdS realizzare le procedure opportune per la creazione di «fascicoli virtuali». Questa scelta si coniuga con l'intenzione del CdS di acquisire i ricorsi in forma elettronica sia per mezzo di dischetti che per via telematica. Questa impostazione impone quindi di analizzare l'intero procedimento prevedendo l'integrazione dei vari sistemi informativi. Infatti l'esperienza maturata in altri settori ha dimostrato che automatizzare solo alcune fasi del procedimento non è garanzia di una maggiore efficienza. È infatti facile verificare come il flusso dei dati, attualmente non automatizzato, che si svolge tra magistrati e segreteria sia molto elevato.

Nella situazione che si può ipotizzare di raggiungere, i dati giungono dall'esterno per via telematica. Essi possono così essere immessi direttamente nel sistema informativo. Questo può portare, col tempo, alla eliminazione della maggior parte della documentazione cartacea e nel caso di acquisizione di documenti opportunamente strutturati può ridurre od eliminare l'impegno legato alla immissione dei dati.

L'adozione di una porta *Web* permette una naturale condivisione delle informazioni contenute nel sistema informativo, compreso il contenuto dei fascicoli virtuali, con i *personal computer* dei magistrati. Deve essere comunque previsto l'utilizzo della firma digitale per l'accesso alle informazioni. L'iter dei ricorsi o dei pareri potrebbe essere assistito da un sistema di *workflow* documentale.

I magistrati potranno accedere, mediante i loro *personal computer*, ad un ambiente di lavoro integrato che permette loro di visualizzare il contenuto dei fascicoli virtuali di loro competenza, fare ricerche sugli archivi del sistema informativo, ricerche di pre-

cedenti giurisprudenziali o di atti normativi, memorizzare appunti o note ed infine redigere il provvedimento in modo altamente standardizzato ed agevolato.

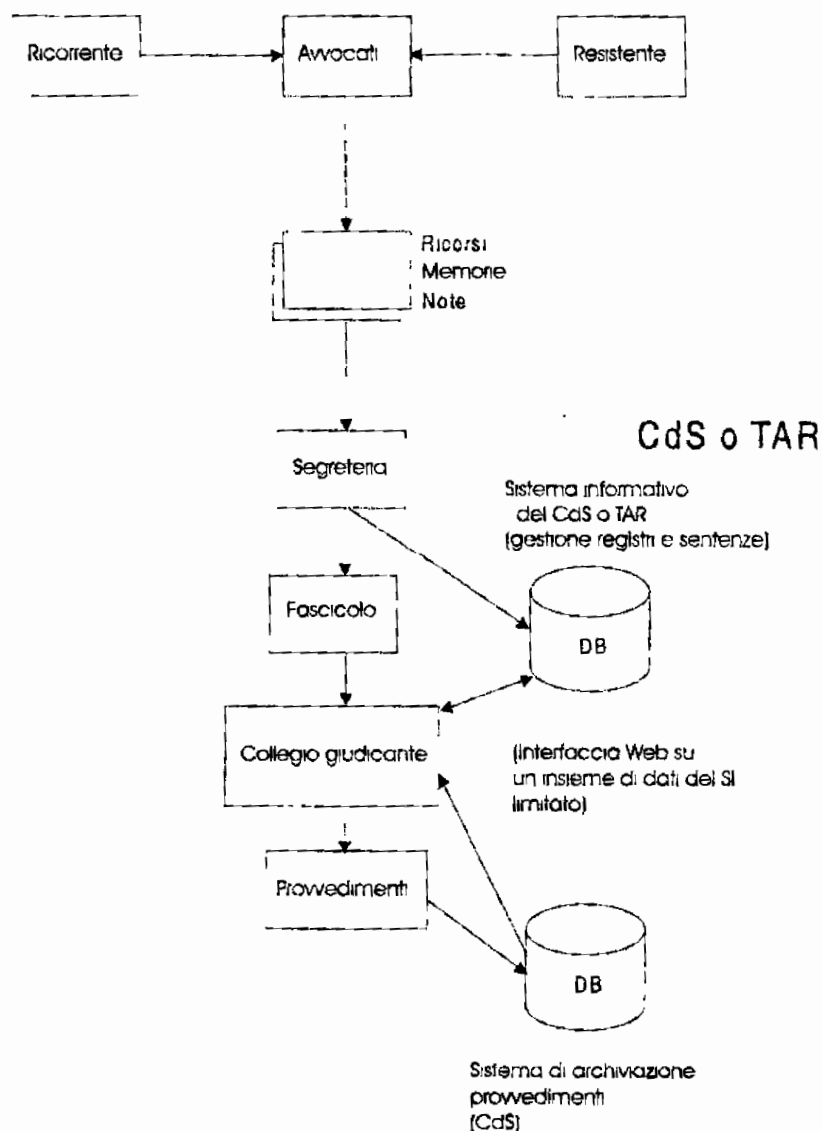


Figura 2.1 - Flusso dei dati nella situazione attuale

Una volta completato l'iter processuale, il provvedimento dovrà essere trasferito sul sistema di pubblicazione *on line* e reso accessibile ai magistrati ed all'esterno mediante un sistema di *information retrieval* interrogabile via interfaccia *Web*.

Per ottenere una facile pubblicazione dei provvedimenti è opportuno che questi siano redatti in un formato di interscambio tra il sistema di aiuto alla redazione ed il sistema di *information retrieval* che contenga informazioni sulla struttura del provvedimento ovvero identifichi gli elementi informativi in esso contenuti. Questo è possibile utilizzando il linguaggio XML (*eXtensible Markup Language*) che consente di incapsulare nel testo tale patrimonio informativo supplementare attraverso l'uso di appositi marcatori. Il risultato è un unico oggetto che fonde insieme il semplice testo con gli identificatori di ciascuna parte significativa del testo stesso. L'importanza di questo linguaggio consiste nello svincolare l'organizzazione dei dati dall'ambiente elaborativo/applicativo nel quale verranno poi utilizzati per soddisfare esigenze di varia natura.

È utile osservare come la redazione del ricorso in XML possa permettere un automatico inserimento delle informazioni necessarie nelle procedure di gestione dei registri, essendo le informazioni necessarie facilmente individuabili in automatico.

2.4. IL LINGUAGGIO XML

L'XML è un dialetto semplificato del linguaggio SGML. L'SGML (*Standard Generalized Markup Language*) è una tecnologia riconosciuta dall'*International Standardization Organization* (Organizzazione internazionale per la standardizzazione, comunemente nota con il nome ISO) per la marcatura dei testi. I linguaggi di marcatura descrivono i meccanismi di strutturazione e di rappresentazione del testo utilizzando delle convenzioni standardizzate e derivano concettualmente dai sistemi usati nelle tipografie per «marcare» le porzioni di testo indicandone le caratteristiche. Essi sono particolarmente adatti all'identificazione delle partizioni testuali, dei riferimenti e delle informazioni rilevanti in un testo.

L'SGML risulta tuttavia troppo complicato per essere utilizzato direttamente per la distribuzione di testi in *Internet* tramite *Web*. Il linguaggio XML è studiato per permetterne una rapida ed efficiente implementazione nei browser del *World Wide Web* di *Internet*, pur mantenendo le caratteristiche fondamentali e la potenza nel descrivere documenti dell'SGML. Il *World Wide Web Consortium* (W3C), dopo un'attività durata circa un anno, nel febbraio del 1998 ha rilasciato la versione 1.0 delle specifiche di XML. Il progetto prevede la definizione di un gruppo di specifiche, che raccolga oltre a XML anche quelle per la gestione dei *link* (XLI) e per la rappresentazione (XSL).

L'XML piuttosto che un linguaggio può essere definito un metalinguaggio, ovvero un linguaggio per la definizione di altri linguaggi o applicazioni. Infatti si oc-

cupa esclusivamente della definizione dei marcatori da usare e della loro strutturazione. XML permette di definire documenti per il *Web* separando struttura e rappresentazione (visualizzazione) dei dati. Separando la struttura e il contenuto dalla presentazione, lo stesso sorgente, scritto una sola volta, può essere visualizzato in modi diversi. Questo vuol dire che un documento scritto secondo specifiche XML può essere veicolato attraverso dispositivi diversi non necessariamente presi in considerazione all'atto della sua stesura. XML è utile ogni qual volta si debba definire e produrre un documento, a prescindere dal mezzo di trasmissione.

XML offre rispetto ad HTML, il linguaggio usato per la costruzione di pagine *Web*, alcuni notevoli vantaggi:

- l'estensibilità che permette la definizione di set personalizzati di *markup*;
- la definizione degli elementi strutturali contenuta in un computer esterno chiamato *Document Type Definition* (DTD). Il DTD contiene le regole che definiscono i tag usati nel documento XML, ovvero in pratica definisce la struttura del documento;
- la possibilità di controllare la correttezza di ogni documento mediante verifica di conformità alle regole definite nel DTD.

Un'applicazione informatica che legga un documento XML consta di due parti: il *Parser* che esegue il controllo semantico e gestisce gli errori e il *Processor* che, utilizzando un altro computer in cui è definita la formattazione dei vari tag, visualizza il documento. Il controllo da parte del *Parser* avviene su due livelli: valutando la conformità del documento al DTD di riferimento prima e, in caso di non conformità, eseguendo un altro controllo relativo però alle regole generali della sintassi XML. Proprio per via di questo doppio controllo i documenti possono essere di due tipi, *Valid* e *Well-Formed*. *Valid* sono quelli conformi al proprio specifico DTD, *Well-Formed* quelli conformi alla sintassi generale.

Compito principale nella definizione e strutturazione di un documento è dunque la definizione del DTD. Una volta definito il DTD è necessario associare al documento stesso un foglio di stile che ne descriva le regole relative alla rappresentazione. Queste non devono essere necessariamente univoche, ma devono poter variare al variare del dispositivo di output o anche in seguito all'interazione dell'utente. Il W3C sta attualmente definendo le caratteristiche di *Extensible Style Language* (XSL). Poiché le specifiche XML nulla dicono delle regole di rappresentazione, si può usare uno dei linguaggi di stile già conosciuti: il *Cascading Style Sheet* (CSS), legato al linguaggio HTML, o il *Document Style Semantics and Specification Language* (DSSSL), che è il linguaggio di stile di SGML.

Associato ad XML vi è inoltre il linguaggio XLL, dedicato allo sviluppo degli *hyperlink*. Per questo linguaggio la situazione è ancora lontana dall'essere sufficientemente definita. La prima bozza delle raccomandazioni è dell'aprile 1997 ed in seguito sono stati redatti i principi generali cui il linguaggio dovrà rispondere. Il linguaggio è diviso in due sottogruppi: *Xpointer*, che permette di indirizzare porzioni di documento, e *Xlink*, che permette di controllare eventi associati allo *hyperlink*.

Le caratteristiche del linguaggio XML, che permettono di descrivere la struttura del documento e di marcare sul testo le informazioni rilevanti, lo rendono uno strumento importante per l'IR, per i RDBS e nell'interscambio di documenti in generale.

È opportuno osservare che l'introduzione di una tecnologia che non ha ancora raggiunto la maturità, come appunto l'XML, richiede molta cautela e comunque si pone la necessità di prevedere sistemi che possano funzionare anche con formati alternativi.