

Consiglio Nazionale delle Ricerche



ISTITUTO DI RICERCHE
SULLA POPOLAZIONE
E LE POLITICHE SOCIALI

*Maria Girolama Caruso e
Loredana Cerbara*

Rapporto tecnico

Il sito Web dell'Istituto di
Ricerche sulla Popolazione
e le Politiche Sociali: una
vetrina virtuale in costante
evoluzione

IRPPS-CNR
Working Paper 08/2005

Working Paper

La multimedialità e le potenzialità del World Wide Web costituiscono il veicolo del terzo millennio della comunicazione scientifica. I siti web degli istituti di ricerca devono essere considerati come vetrina principale dei prodotti e dei risultati della ricerca e pertanto devono essere oggetto di cura ed attenzione costante.

Abstract

La progettazione, lo sviluppo e la manutenzione del sito Web dell'IRPPS rientrano nelle competenze del servizio "Elaborazioni informatiche e banche dati" nella sede di Roma potendo la sede di Salerno gestire una parte del sito autonomamente.

Questo lavoro è finalizzato ad illustrare le varie fasi di sviluppo del sito Web dell'IRPPS che si sono susseguite nel corso degli anni. L'obiettivo è quello di tracciare un resoconto sintetico di tutte le scelte effettuate rispetto: alle informazioni inserite nel sito, ai linguaggi di programmazione utilizzati, alle modalità di comunicazione tramite il web, alle soluzioni pratiche adottate a fronte di problemi tecnici e di esigenze che si sono verificate nel tempo in Istituto.

Nel corso degli anni sono state realizzate tre versioni: la prima molto semplice sia rispetto al linguaggio di programmazione (solo pagine html) che alla grafica utilizzata; la seconda strutturata a frame sempre in html e l'ultima progettata come un "portale", a nostro avviso molto più efficiente perché rende immediatamente disponibile l'informazione di base di tutto il sito, sulle ricerche e attività dell'Istituto. In quest'ultima versione oltre alle pagine programmate in html, sono state progettate anche pagine dinamiche pensate per la costruzione della mailing list dell'Istituto nel linguaggio Php e il MySQL utilizzato come database.

Inoltre, abbiamo cercato di applicare le linee guida di indirizzo per la produzione dell'informazione che gli organi preposti del CNR hanno consigliato di adottare per tutti i siti web del dominio "cnr.it".

Una particolare attenzione sarà rivolta ad alcune tecniche statistiche come ad esempio quelle che vanno sotto il nome di *data web maining*. Queste tecniche sono generalmente applicate ai file di log (file residenti sul server che ospita il sito e che registra quotidianamente i dati delle visite alle pagine web) e servono per conoscere il profilo degli utenti e capire come migliorare il sito in termini di facilità di accesso. Ciò, può avere ricadute inaspettate, come quella di poter monitorare e veicolare un canale di comunicazione scientifica dalle caratteristiche eccezionali, come è quello offerto da Internet.

Sommario

Introduzione	5
Le fasi della progettazione del sito	6
Progettazione del primo sito web (1997-98)	8
Progettazione del secondo sito web (2002)	10
Progettazione del terzo sito web (2004)	14
La statistica per il web	19
Conclusioni	20
Bibliografia	21

Introduzione

Il fenomeno che sta caratterizzando con maggiore forza il mondo della comunicazione è sicuramente rappresentato dallo straordinario sviluppo di Internet. E' noto che gli Istituti di ricerca sono grandi utilizzatori della rete. Poter comunicare, scambiarsi informazioni e dati con altri studiosi sparsi in tutto il mondo è una necessità basilare delle attività di ricerca e Internet, in questo senso, offre delle opportunità eccezionali.



Nel momento in cui si pensa di costruire un sito, in particolare di un Istituto di ricerche nel campo socio-demografico, si deve avere ben chiaro quali siano le funzioni essenziali che questo sito dovrebbe assolvere. Quindi per prima cosa è necessario chiedersi quali sono le informazioni e i servizi che si intende offrire al pubblico.

Nella fase di progettazione è importante prima di tutto stabilire come obiettivi prioritari quelli della evidenziazione e della promozione della propria immagine attraverso la valorizzazione delle proprie potenzialità, mettendo nella giusta luce gli obiettivi, i risultati acquisiti ed il loro valore in termini conoscitivi (Relazione, CNR, 2001).

Sicuramente, tra le informazioni essenziali rientrano: la struttura dell'Istituto (organigramma), le attività di ricerca, servizi interni e servizi esterni (come ad esempio: convegni e seminari, pubblicazioni, riviste on-line, comunicati stampa, rassegne stampa, mailing list e catalogo delle biblioteca).

Il passo successivo, consiste nell'identificare i potenziali lettori per poter considerare, già in sede di progettazione del sito, i loro eventuali bisogni. In questa categoria di utenti rientrano sicuramente oltre l'utente comune, in particolare: gli studenti, gli studiosi, i ricercatori, le amministrazioni pubbliche, ecc.

La costruzione del sito prevede una serie di attenzioni e di regole da seguire secondo il principio della "progettazione universale" (Burzagli, Graziani, 1999) in cui la progettazione deve avvenire secondo una regola fondamentale: tutti gli individui devono avere la stessa possibilità di accedere ad Internet. A tale proposito esistono delle direttive sulle accessibilità dei siti web prodotte da alcune organizzazioni come ad esempio il Consorzio W3C¹ e il WAI (Web Accessibility Initiatives)².

Infine, da non dimenticare che si sta progettando per il World Wide Web, ossia per la rete mondiale, quindi è necessario fornire la versione del sito in inglese (lingua di uso comune nella comunicazione internazionale), per lo meno delle pagine principali.

La prima realizzazione del sito Web dell'Istituto di Ricerche sulla Popolazione, in seguito confluito nell'Istituto di Ricerche sulla Popolazione e le Politiche Sociali, risale al 1997-98, la seconda versione al 2002 in concomitanza con la nascita dell'IRPPS e la terza al 2004.

Dalla nascita ad oggi il sito Web dell'IRPPS ha viaggiato sempre in parallelo con le attività di ricerca svolte nell'Istituto (è stato infatti considerato come lo specchio della ricerca), senza perdere mai di vista l'obiettivo principale del sito: costruire "una vetrina" sulle attività di ricerca dell'Istituto, pubblicamente esposte in una piazza virtuale dove ognuno mostra ciò che produce, innalzando il livello di competitività e della qualità della ricerca. Di conseguenza, l'aggiornamento e la manutenzione sono considerate punti indispensabili e da prevedere fin dalle prime fasi della progettazione.

Le fasi della progettazione del sito

Per la progettazione del sito web dell'Istituto sono stati seguiti i seguenti passi progettuali:

¹ Il W3C (World Wide Web Consortium) è nato nell'ottobre del 1994 e sviluppa tecnologie che garantiscono l'interoperabilità (specifiche, guidelines, software e applicazioni) per guidare il World Wide Web fino al massimo del suo potenziale agendo da forum di informazioni, comunicazioni e attività comuni. L'Ufficio W3C Italiano (W3C-IT) è il punto di contatto nazionale per le attività W3C in Italia. Il sito italiano è all'indirizzo www.w3c.it mentre il sito internazionale si trova all'indirizzo www.w3.org.

² Per accessibilità web si indica la capacità di un sito web di essere acceduto efficacemente (alla sua interfaccia e al suo contenuto) da utenti diversi in differenti contesti. Rendere un sito web accessibile significa permettere l'accesso all'informazione contenuta nel sito anche a persone con inabilità fisiche di diverso tipo e a chi dispone di strumenti hardware e software limitati.

- *Definizione del sito* – In questa fase si definiscono gli scopi, gli obiettivi relativi al web, la portata del contenuto del sito, il supporto tecnico richiesto (i browser³ e i sistemi operativi⁴ che devono supportare il sito; il supporto server web che può essere interno o fornito da un provider⁵ di servizi Internet), il budget (il costo dell'hardware, del software, dell'addestramento, degli stipendi e il costo della manutenzione) e si individua la persona che all'interno del gruppo di lavoro assumerà le funzioni di webmaster (supporto tecnico, manutenzione data base e aggiornamento).
- *Raccolta delle informazioni* – Il team di lavoro dovrà fare un inventario di tutte le informazioni esistenti e descriverne il contenuto. Una volta stabilita la struttura del sito è necessario creare dei piccoli prototipi delle singole parti per verificare in concreto il funzionamento delle procedure operative all'interno della struttura progettata.
- *Costruzione del sito* – Dopo aver stabilito il modello e la struttura del sito si passa alla realizzazione di tutte le pagine e alla creazione di un eventuale motore di ricerca o di un eventuale data base.
- *Test beta* – Deve essere fatto da lettori esterni al gruppo di lavoro, disponibili a fornire critiche costruttive, individuare eventuali bug, errori tipografici e di scrittura.
- *Marketing del sito* – E' necessario a questo punto pubblicizzare l'URL⁶ del sito attraverso mezzi di comunicazione di massa: radio, televisione, stampa, ecc; attraverso la registrazione sui alcuni motori di ricerca⁷, tipo Yahoo, Arianna e Virgilio; e infine, riportare su carta intestata il proprio URL.

³ Un browser è un programma che consente la navigazione nella rete internet, più precisamente nel World Wide Web. La funzione primaria di un browser è quella di interpretare il codice HTML (acronimo per HyperText Mark-Up Language) che è un linguaggio usato per descrivere i documenti ipertestuali disponibili su Internet. Non è un linguaggio di programmazione, ma un linguaggio di markup, ossia descrive l'aspetto che deve avere il testo. È stato sviluppato da Tim Berners-Lee al CERN di Ginevra. Il browser visualizza l'HTML (e più recentemente l'XHTML) in forma di ipertesto. L'ipertesto è un insieme di testi o pagine leggibili con l'ausilio di un'interfaccia elettronica, in maniera non sequenziale, tramite particolari parole che si chiamano hyperlink (rimandi), che costituiscono un rete raggiata o variamente incrociata di informazioni organizzate secondo criteri paritetici o gerarchici.

⁴ In informatica, un sistema operativo (abbreviato in SO) è il programma responsabile del diretto controllo e gestione dell'hardware che costituisce un computer e delle operazioni di base. Si occupa anche di controllare gli accessi degli utenti e dei processi che vengono eseguiti.

⁵ Letteralmente un provider è un fornitore, un ente che fornisce a terzi accessi ad Internet, gratuitamente o a pagamento.

⁶ URL è l'acronimo per Uniform Resource Locator. Indirizzo elettronico in Internet che identifica univocamente una risorsa. Il primo elemento di un indirizzo URL è solitamente <http://www.nome-computer-host/nome-della-risorsa-host>.

⁷ I motori di ricerca sono dei siti che poggiano su grandissimi archivi dove vengono raccolti dati e informazioni che consentono di compiere ricerche di indirizzi, immagini o argomenti sul Web.

- *Manutenzione* – Dopo aver pubblicato il sito, inizia una nuova fase: la manutenzione. Senza questa prospettiva una pubblicazione elettronica è destinata diventare rapidamente obsoleta.

Progettazione del primo sito web (1997-98)

La prima progettazione del sito web, come è stato già detto, risale al 1997-98, le altre si sono susseguite negli anni avvenire.

Il progetto iniziale del sito dell'Istituto era suddiviso in *tre fasi*, ognuna delle quali conteneva un pacchetto di risorse implementabili separatamente in quanto dotate di una propria autonomia funzionale. La *prima fase* prevedeva la creazione di un nucleo funzionale del sito e l'implementazione delle sezioni Istituto, Demotrends, Convegni e Seminari e si componeva di una *home page*⁸ che dava l'accesso a tutte le sezioni suddette:

- La sezione *Istituto* presentava l'istituto stesso e né illustrava i componenti, le attività, l'ubicazione, etc. Ognuna di queste categorie poteva essere letta come una sottosezione a sé:
 - Istituto;
 - Staff;
 - Attività;
 - Pubblicazioni;
 - Informazioni;
 - Novità;
 - Links.
- La sezione *Demotrends*, dava l'accesso alle pagine previste per l'edizione on-line del *quadrimestrale sulla realtà demografica italiana* ed era organizzata nel seguente modo: titoli degli articoli e autore sotto forma di link che portava l'utente alla pagina contenente l'articolo richiesto. La *home page* del giornale dava l'accesso all'ultimo numero e a tutti i precedenti in formato pdf⁹.

⁸ È la schermata di ingresso a un sito Web. Scritta in linguaggio html, rappresenta il biglietto da visita grafico che una società o un singolo utente fornisce ai visitatori del proprio sito. In genere, dalla Home page alcuni link permettono di accedere ad altri documenti o informazioni presenti nel sito o fuori di esso.

⁹ Il PDF (Portable Document Format) è un linguaggio di descrizione di pagina sviluppato da Adobe Systems per rappresentare documenti in modo indipendente dall'hardware e dal software utilizzati per generarli. Un file PDF può descrivere documenti che contengono testo od immagini a qualsiasi risoluzione.

Il PDF è basato principalmente sul PostScript, un linguaggio di descrizione immagini anch'esso sviluppato dall'Adobe. Postscript permette di descrivere un'immagine come il risultato di un'esecuzione di un programma. In questo modo, ogni apparecchio capace di eseguire il programma (ossia, che abbia un interprete Postscript), sarà in grado di riprodurre tale immagine al meglio delle sue capacità. Tale caratteristica è stata quindi ereditata anche dal PDF.

- Per la sezione *Convegni e Seminari*, il numero di pagine era destinato a variare in relazione alla quantità dei convegni. I convegni già effettuati venivano comunque mantenuti come materiale di archivio.

La *seconda fase* comprendeva la realizzazione e l'implementazione della sezione Indagini e Rapporto e della sezione Mailing list. La sezione Indagini e Rapporto in linea di massima si prevedeva di organizzarla sul modello di Demotrends, non solo per fornire un metodo di navigazione unico per tutte le sezioni, ma anche perché qualsiasi altra soluzione sarebbe stata di gran lunga più costosa e lunga da realizzare. La sezione Mailing list sarebbe stata composta (ma all'epoca non fu realizzata e ci si fermò alla sola progettazione) essenzialmente da una pagina contenente una scheda da compilare a cura dell'utente per poter immettere i dati personali dello stesso e il tipo di materiale che desiderava ricevere nella propria casella postale elettronica (indagini, convegni, Demotrends, novità).

La *terza fase* concludeva virtualmente tutta l'operazione e consisteva nella pubblicazione del catalogo della biblioteca dell'Istituto.

Inoltre, il progetto prevedeva sia una versione italiana e una versione in inglese (entrambe con la stessa interfaccia grafica) da poter essere scelte dall'utente sulla home page.

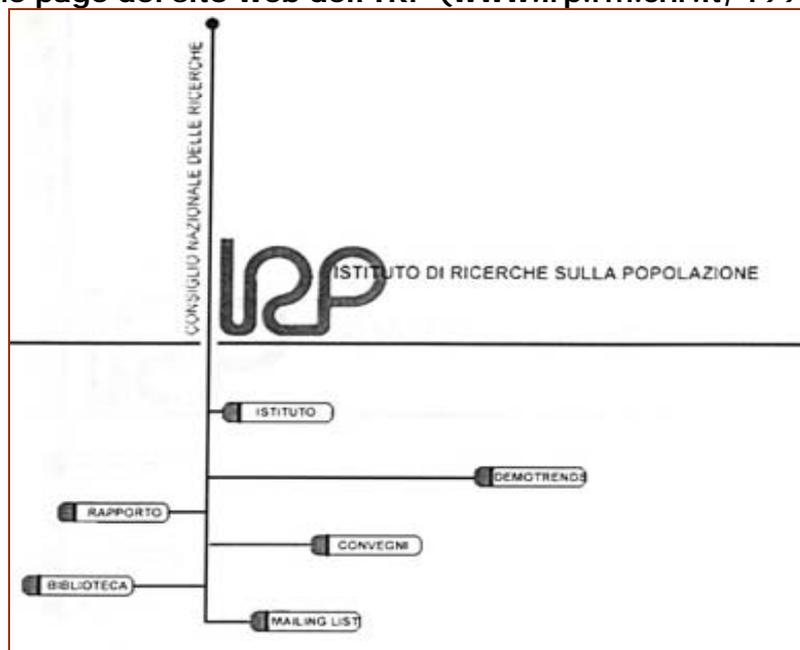
L'intero sito doveva essere programmato in HTML con l'aggiunta di alcune funzioni in Javascript¹⁰.

Allora, per una serie motivi economici e tecnici, la progettazione del sito web dell'Istituto si fermò alla prima fase. La seconda e la terza fase del progetto in cui si prevedeva la realizzazione di una mailing list e di un catalogo on line della biblioteca, probabilmente, non si realizzarono non soltanto per un problema di budget, ma anche perché non fu data la giusta importanza a queste nuove forme di comunicazione, che erano ancora poco usate per questi scopi e poco conosciute.

Di fatto, le ultime due fasi del progetto si sono realizzate con la terza progettazione del sito, ed esattamente l'ultima versione presenta un collegamento sia alla mailing list che ai cataloghi on-line delle biblioteche del nostro Istituto.

¹⁰ JavaScript è un linguaggio di scripting orientato agli oggetti comunemente usato nei siti web. Fu originariamente sviluppato da Brendan Eich della Netscape Communications con il nome di Mocha e successivamente di LiveScript, ma in seguito è stato rinominato "JavaScript" ed è stato formalizzato con una sintassi più vicina a quella del linguaggio Java di Sun Microsystems.

Home page del sito web dell'IRP (www.irp.rm.cnr.it, 1997-98)



Inoltre, si ricorda che dopo la pubblicazione in rete delle pagine web, per assicurare una migliore visibilità, si proseguì con l'indicizzazione sui principali motori di ricerca con l'inserimento di *meta tag*¹¹ contenenti parole chiave, campi di testo, etc.

Progettazione del secondo sito web (2002)

La seconda progettazione del sito risale al 2002 in concomitanza con la nascita dell'IRPPS. Questo sito rispetto al precedente ha subito notevoli cambiamenti soprattutto in termini di contenuti (cambiamento della direzione, obiettivi più allargati, altre attività di ricerca, altri servizi, aumento del personale, nuova ubicazione, etc.).

In considerazione del fatto che la home page della prima versione, per la sua forma grafica, non consentiva frequenti aggiornamenti, soprattutto ponendo limiti all'introduzione di nuove voci di menu, si è pensato di passare ad un'altra versione, più snella e facilmente manipolabile.

¹¹ L'insieme delle Istruzioni in linguaggio HTML contenute all'interno delle pagine, che forniscono varie informazioni ai motori di ricerca, tra le quali il titolo della pagina, l'autore della stessa, il programma usato per la costruzione, la descrizione e le keywords.

Home page del sito web dell'IRPPS (www.irpps.cnr.it, 2002)



In questo periodo era molto in uso la tecnica dei *frame*¹², che consentivano di svincolare graficamente ma anche strutturalmente la parte della pagina riservata ai menu di navigazione da quella destinata al contenuto vero e proprio. Questa tecnica presentava anche il vantaggio di poter disporre in ogni pagina visitata dello stesso menu, assicurando quindi la navigabilità totale del sito.

Il nuovo sito è stato dunque strutturato a frame. La home page riportava un menu di navigazione sul lato sinistro con collegamenti alle sezioni: *istituto*, *ricerca*, *pubblicazioni*, *osservatori*, *indagini*, *demotrends*, *convegni e seminari*, *lavoro e formazione*, *eventi e notiziario* e *links*; al centro della home le informazioni di tipo generale: collegamento con l'Ente di appartenenza, l'indirizzo con telefono e fax, il nome del Direttore, la data di aggiornamento e un collegamento e-mail al webmaster.

La sezione *Istituto* illustrava brevemente la storia dell'Istituto, le principali attività di ricerca e riportava dei collegamenti ipertestuali alle seguenti voci: struttura dell'Istituto, Consiglio Scientifico, Comitato d'Istituto, staff e servizi.

¹² Nel linguaggio HTML indica una pagina divisa da cornici e composta da diversi documenti. Ad esempio: una parte superiore fissa con un banner pubblicitario o con la testata del sito, una parte sulla sinistra con un menu di scelta ed una parte sulla destra che visualizza il contenuto del documento scelto dal menu.

La sezione *ricerca* illustrava le varie competenze che l'Istituto aveva acquisito nel corso degli anni e dei links alle principali linee di ricerca dell'Istituto, descritte in modo sintetico (responsabile, afferenti, missione e metodi di ricerca).

La sezione *pubblicazioni* riportava un menù con collegamenti alle pubblicazioni dell'Istituto divise per collana con la possibilità di scaricare i testi dei lavori in formato .pdf o delle sintesi in html. Il menu era composto dalle seguenti voci: Comunicazioni a convegni e seminari, Rapporto sulla situazione demografica in Italia, Atti convegni e seminari, Fuori collana, Serie documenti e ristampe, Working Papers, Serie collana monografie, Demotrends, Quaderni, Cd-rom.

La sezione *osservatori* riportava un menù con collegamenti a quattro osservatori: l'osservatorio sulle intenzioni riproduttive, l'osservatorio permanente sulle migrazioni italiane, l'osservatorio sulla situazione demografica italiana e l'osservatorio europeo sull'informazione e l'educazione della popolazione; questi prevedevano un monitoraggio continuo su alcune attività di ricerca dell'Istituto.

La sezione *indagini* riportava i risultati ottenuti da tutte le indagini condotte in Istituto già dalla metà degli anni '80 su alcuni temi specifici: la famiglia, le intenzioni riproduttive, le migrazioni, i giovani, gli anziani, la conoscenza dei fatti demografici.

La sezione *convegni e seminari* conteneva lo storico dei convegni organizzati dall'Istituto a partire dal 1998.

La pagina *lavoro e formazione* riportava le eventuali opportunità di lavoro e formazione.

La pagina degli *eventi e notiziario* conteneva le informazioni più recenti o le novità.

La pagina dei *Links* riportava un elenco dei siti attinenti alle ricerche sulla popolazione con collegamenti al loro URL.

Inoltre in questa versione del sito abbiamo tentato di applicare alcune delle regole (linee guida) che gli organi proposti del CNR hanno consigliato di adottare per tutti i siti web del dominio "cnr.it". Tra queste rientrano: la doppia versione del sito (italiano e inglese) l'inserimento di meta-informazioni sia in italiano che in inglese, la presentazione dell'organo, l'organizzazione, le ricerche, le attività di formazione, attività di servizio, le manifestazioni, etc., inoltre, la home page deve contenere tutte le informazioni necessarie per far capire all'utente la tipologia del sito.

La home page del sito web dell'Istituto, anche se un po' scarna, includeva tutte le informazioni necessarie per far capire agli utenti visitatori la tipologia del sito e il suo contenuto (attraverso le voci: istituto,

pubblicazioni, ricerca, osservatori, indagini e convegni), inoltre, presentava un collegamento all'Ente di appartenenza, il nome del direttore, l'indirizzo, la data di aggiornamento, etc., purtroppo mancava la versione in inglese. Inoltre, in questa versione del sito sono state inserite le meta-informazioni necessarie per l'indicizzazione dei principali motori di ricerca.

La scelta dei meta dati utilizzati rientra, tra l'altro, in un progetto (*Progetto per la gestione della pubblicazione dei dati nel server*) esteso a cinque istituti del CNR (IDG, IRP, ISRDS, ITIM e URP) per la sperimentazione di un motore di ricerca. Per la scelta dei meta dati si è ricorso agli standard definiti dalla *Dublin Core Metadata Initiative* considerati idonei a indicizzare i materiali provenienti dai siti web del CNR. Nel "subject" (vedi oltre) sono stati inseriti "codici disciplina" (es.: EcB4 EcB6 EcD2 EcJ1 EcJ4 ErF1) utilizzati per i consuntivi e i preventivi annuali da ciascuna struttura del CNR, nonché eventuali parole chiave necessarie per descrivere l'argomento della risorsa (es.: Popolazione = Invecchiamento = Fecondità = Mortalità = Demografia della famiglia). Per il "type" (vedi oltre) sono stati inseriti "codici risultato" anch'essi utilizzati per i consuntivi e i preventivi.

Si riportano i meta tag provenienti dallo standard DCMI¹³ inseriti nell'head¹⁴ del file html della home page del sito dell'Istituto:

- <meta name="description" content="La Home page dell'IRPPS permette di consultare le seguenti sezioni: istituto, ricerca, pubblicazioni, osservatori, indagini, demotrends, convegni e seminari, links"> (spiegazione del contenuto della risorsa)
- <meta name="language" content="it"> (lingua del contenuto della risorsa)
- <meta name="publisher" content="IRPPS/CNR Istituto di Ricerche sulla Popolazione e le Politiche Sociali del Consiglio Nazionale delle Ricerche"> (entità responsabile del contenuto della risorsa)
- <meta name="subject" content="EcB4 EcB6 EcD2 EcJ1 EcJ4 ErF1 Hu4104 SoE= SoF2 SoK= SoL= = Popolazione = Invecchiamento = Fecondità = Mortalità = Demografia della famiglia = Politiche demografiche = Studi di genere = Migrazioni = Indagini d'opinione = Population = Aging = Fertility = Mortality = Family Demography = Population policy = Gender studies = Migration = Opinion survey"> (argomento scientifico disciplinare della risorsa)

¹³ DCMI è un acronimo per Dublin Core Metadata Initiative contenuto nel documento *DCMI Metadata Terms* disponibile su <http://dublincore.org/documents/2005/01/10/dcmi-terms/>.

¹⁴ Head è un tag, una delle istruzioni del linguaggio HTML. Un documento HTML si divide in due parti fondamentali: l'intestazione e il corpo del documento. L'intestazione è racchiusa tra due tag HEAD che ne indicano l'inizio e la fine e contiene una serie di informazioni necessarie al browser per una corretta interpretazione del documento, ma non visualizzate all'interno dello stesso. L'intestazione, quindi, ha un ruolo non apparente ma fondamentale.

- <meta name="type" content="19"> (natura/genere del contenuto della risorsa)
- <title>IRPPS - Istituto di ricerche sulla Popolazione e le Politiche Sociali</title>

I meta tag sono stati inseriti nella home page e nelle pagine di secondo livello.

Progettazione del terzo sito web (2004)

In previsione di una nuova progettazione del sito "la commissione web" (istituita nel frattempo in Istituto allo scopo di costituire una sorta di gruppo di redazione del sito) ha ritenuto opportuno procedere prima con un'analisi statistica con tecniche innovative di *data web mining*¹⁵ sul sito in oggetto, tecnica che consente di estrarre facilmente le informazioni sui comportamenti dei visitatori dagli archivi contenuti nel server (i file log¹⁶). Per il tipo di struttura del sito (a frame), l'analisi non ha potuto cogliere tutti gli aspetti di nostro interesse. Ciononostante, il numero di utenti che hanno usufruito dei servizi offerti è notevole, considerando il campo di interesse molto specifico, ed anche gli utenti internazionali sono presenti in misura considerevole.

Questi dati allora sono serviti da stimolo a migliorare le pagine on line sia dal punto di vista dei servizi offerti, sia dal punto di vista della navigabilità e facilità di accesso agli stessi.

È stata quindi predisposta una nuova versione del sito dalla struttura completamente diversa e molto più ricca di contenuti sulla ricerca effettuata dall'Istituto. La home page si presenta molto semplificata nella struttura, ma, a nostro avviso, molto più efficiente perché rende immediatamente disponibile l'informazione di base di tutto il sito, proprio come se si trattasse di un 'portale' sulle ricerche e le attività dell'Istituto.

¹⁵ Il Data Mining è un processo analitico e automatico per l'esplorazione di grandi quantità di dati, finalizzato alla ricerca di relazioni sistematiche tra variabili e alla validazione dei risultati ottenuti applicando le regolarità individuate a nuovi sottoinsiemi di dati (M.J.A. Berry, G. Linoff, 1997). L'estensione dei concetti di Data Mining al web ha generato un nuovo filone di ricerca detto *Data Web Mining*.

¹⁶ In generale nella lingua inglese il termine log indica il diario di bordo di una nave. Per estensione lo si applica anche ad altri veicoli e ad altre macchine tra cui il computer. Nel caso specifico del server il file di **log** è creato automaticamente dai server Web che registra permanentemente le operazioni compiute all'interno di un sito Web. Attraverso l'analisi dei file log i programmi di statistiche risalgono a dati aggregati e grafici, anche di lungo periodo.

Home page del sito web dell'IRPPS (www.irpps.cnr.it, 2004)

**ISTITUTO DI RICERCHE SULLA POPOLAZIONE
E LE POLITICHE SOCIALI**

Consiglio Nazionale delle Ricerche

istituto | sezioni di ricerca | osservatori | indagini | convegni e seminari | eventi e notiziario | servizi | prodotti

istituto

chi siamo
struttura

sezioni di ricerca

Dinamiche di popolazione e governance
Mobilità e territorio
Analisi dei comportamenti sociali e demografici e valutazione delle politiche
Sistemi di Welfare e politiche sociali
Risorse umane e società della conoscenza

osservatori

Osservatorio sulle intenzioni riproduttive
Osservatorio sulle migrazioni italiane
Osservatorio sulla situazione demografica italiana
Osservatorio Europeo sull'Informazione e l'Educazione della Popolazione

indagini

convegni e seminari

prodotti

pubblicazioni banche dati software

servizi

links lavoro e formazione

mailing list

biblioteca di Roma
biblioteca di Salerno

demotrends

credits | risoluzione 1024 x 768 | copyright | come raggiungereci

Google Cerca nel sito

La struttura del sito si presenta nel seguente modo:

- La sezione *Istituto* presenta l'istituto stesso e né illustra la struttura, lo staff e i servizi interni all'Istituto. Dalla pagina dello staff si accede ad alcune pagine personali contenenti un breve curriculum e una breve descrizione sui principali interessi di ricerca, le linee di ricerca a cui ogni ricercatore o tecnologo partecipa e una lista di pubblicazioni.
- La sezione *ricerca* presenta le attività delle cinque sezioni di ricerca: *Dinamiche di popolazione e governance*, *Mobilità e territorio*, *Analisi dei comportamenti sociali e demografici e valutazione delle politiche*, *Risorse umane e società della conoscenza* e *Sistemi di Welfare e Politiche Sociali*. Ogni singola sezione di ricerca è composta da più pagine. Una home della sezione stessa, in cui si riportano gli obiettivi

prioritari e il collegamento all'e-mail del responsabile. La pagina *ricercatori* contiene la lista degli afferenti alla sezione con i links ad alcune pagine personali. Le pagine *linee di ricerca* (sia quelle in corso che quelle concluse di cui si è voluto lasciare traccia) riportano una breve descrizione dei progetti di ricerca, i partecipanti al progetto e alcune pubblicazioni di riferimento. Nel corso degli anni per alcuni progetti di ricerca sono state realizzate delle sottosezioni all'interno del sito stesso per poter illustrare più dettagliatamente i contenuti dei progetti. Tra queste ricordiamo:

- Studi sulla popolazione dell'Etiopia realizzato nel 2001;
- Le indagini "I giovani e la popolazione, le famiglie e gli immigrati" realizzato dal 2002;
- Il progetto MobiDis realizzato nel 2004.
-

Home page della sezione di ricerca "Mobilità e territorio"

The screenshot shows the website interface for the 'Mobilità e territorio' section. At the top, there is a navigation bar with links: 'Istituto', 'sezioni di ricerca', 'osservatori', 'indagini', 'convegni e seminari', 'eventi e notiziario', 'servizi', and 'prodotti'. The main header reads 'ISTITUTO DI RICERCHE SULLA POPOLAZIONE E LE POLITICHE SOCIALI' and 'Consiglio Nazionale delle Ricerche'. The left sidebar lists various research sections: 'Dinamiche di popolazione e governance', 'Mobilità e territorio' (highlighted), 'Analisi dei comportamenti sociali e demografici e valutazione delle politiche', 'Sistemi di Welfare e politiche sociali', and 'Risorse umane e società della conoscenza'. Under 'Mobilità e territorio', there are links for 'ricercatori', 'linee di ricerca', and 'indagini', as well as 'Osservatorio sulle migrazioni italiane' with links to 'pubblicazioni' and 'migrazioni in cifre'. The main content area is titled 'Mobilità e territorio' and lists the responsible researcher as 'Corrado Bonifazi'. The text describes the section's focus on mobility and territory, mentioning its role in providing knowledge, interpretation, and intervention tools. It also lists current research topics like international migration, Italian emigration, immigration, and urbanization dynamics.

- La sezione *osservatori* introduce i quattro osservatori in via di realizzazione: l'Osservatorio sulle intenzioni riproduttive, l'Osservatorio sulle migrazioni italiane, l'Osservatorio sulla situazione demografica italiana e l'Osservatorio Europeo sull'Informazione e l'Educazione della Popolazione (concluso). L'osservatorio sulla situazione demografica italiana ha messo a disposizione un breve rapporto in formato .pdf sulla situazione demografica italiana, proponendosi di redigere e tenere aggiornato un Atlante Demografico, che metta a disposizione degli studiosi, dei politici, degli amministratori e del pubblico comunque interessato materiale multimediale (dati, testi, mappe, foto, ecc.) concernente la popolazione e debitamente organizzato per territorio e trend storici.

L'osservatorio sulle migrazioni italiane punta a seguire con continuità le diverse dinamiche migratorie che interessano il nostro Paese. Esso è composto da più pagine: le pagine *pubblicazioni* suddivise secondo l'attività editoriale e di destinazione del prodotto, le pagine dei *links* a siti che si occupano di migrazioni, ed infine sono in costruzione le pagine per la voce "*migrazioni in cifre*" in cui si metterà a disposizione degli utenti visitatori materiale multimediale (dati, testi, mappe, ect.).

- La sezione *indagini* è composta da più pagine. La home page riporta una breve presentazione delle indagini svolte in Istituto fin dalla metà degli anni '80 e inoltre, offre la possibilità di scaricare un CD-ROM sugli Atteggiamenti degli italiani su temi di popolazione - *Italians' attitudes towards population issues*, prodotto dall'IRP nel 2001 e di scaricare dei filmati realizzati per alcune indagini. Le altre pagine, considerate di approfondimento riportano le informazioni metodologiche, il questionario allegato e le pubblicazioni di riferimento.

Home page della sezione di indagini

Fin dalla metà degli anni '80, l'Istituto studia gli atteggiamenti e le opinioni degli italiani verso le recenti tendenze demografiche attraverso lo strumento delle indagini campionarie. I temi trattati sono molteplici: la famiglia, le intenzioni riproduttive, le migrazioni, i giovani, gli anziani, la conoscenza dei fatti demografici.

CD-ROM - Surveys - Atteggiamenti degli italiani su temi di popolazione - *Italians' attitudes towards population issues*, IRP, 2001.

surveys.zip (empezza 20MB) si consiglia di scaricarlo solo se si dispone di una connessione veloce.

Per alcune indagini sono stati realizzati i seguenti filmati:

Per ascoltare i file audio è necessario aver installato nel proprio computer Real Player.

Indagine sulla permanenza dei giovani in famiglia
Survey on "stay-at-home" children
Anziani: lavoro, famiglia, tempo libero e tecnologie
The Elderly: work, family, free time and the new technologies
Delphy study

- La sezione *convegni e seminari* è composta da una pagina e contiene lo storico dei convegni organizzati dall'Istituto a partire dal 1998 ad oggi.
- La sezione *eventi e notiziario* è composta anche essa da una pagina, inoltre, per dare maggiore rilievo all'evento è stato inserito sulla home page un piccolo banner a scopo pubblicitario.

- La sezione *prodotti* è composta da più pagine: la pagina dei *software* in cui sono descritti i software prodotti dall'Istituto; la pagina *banche dati* (in costruzione) e infine le pagine *pubblicazioni*. In particolare, per questa sottosezione abbiamo voluto conservare un l'elenco anche se parziale delle pubblicazioni degli Istituti (IRP, IRIDISS, ISRDS) che dal 2002 ad oggi sono confluiti nell'Istituto di Ricerche sulla Popolazione e le Politiche Sociali.
- La sezione *servizi* riporta una pagina con *links* di siti attinenti e una pagina sulle opportunità di lavoro e formazione offerte dall'Istituto.
- La sezione *mailing list* è composta da una pagina contenente una scheda da compilare a cura dell'utente per poter immettere i dati personali dello stesso e il tipo di materiale che si desidera ricevere nella propria casella postale (il *notiziario*, per la ricezione via e-mail di tutte le novità: call for papers, pubblicazioni, seminari, eventi dell'IRPPS; e la *newsletter*, per la sola ricezione della copia elettronica del quadrimestrale demotrends). Questa parte del sito è stata sviluppata in Php (linguaggio usato per la programmazione delle pagine dinamiche) con l'utilizzo del database MySql per l'archiviazione dei dati. Tutte le pagine sono protette da eventuali accessi non autorizzati per garantire il rispetto delle norme sulla privacy degli utenti che aderiscono al servizio di mailing list.
- La sezione *biblioteca* offre l'accesso on line a due cataloghi distinti: uno realizzato presso la biblioteca della sede di Roma e l'altro presso la biblioteca della sede di Salerno. Il catalogo on line della sede di Roma è ospitato dal server della Biblioteca dell'Istituto di Scienze e Tecnologia dell'Informazione – ISTI – del CNR di Pisa (<http://pc-biblio6.area.pi.cnr.it/>). Il catalogo on line della sede di Salerno è stato sviluppato con File Maker e consente la consultazione solo via Intranet.

La versione del sito in inglese è incompleta. Sono state realizzate solo alcune pagine: la home page, la pagina dello staff con links alle pagine personali e le pagine di due sezioni di ricerca: Mobilità e Territorio e Risorse umane e società della conoscenza.

Home page della sezione biblioteca



La statistica per il web

Abbiamo accennato nei paragrafi precedenti all'uso di tecniche statistiche applicate a dati rilevati su server web. Lo scopo di queste analisi è molteplice e, nel corso del tempo, ci ha aiutato a migliorare sia il servizio offerto con il sito, sia la comunicazione con gli utenti visitatori. Infatti, le tracce che i visitatori lasciano quando visitano un sito, possono essere analizzate per capire sia le preferenze dei visitatori stessi, sia le eventuali lacune del sito.

L'ultima progettazione del sito si è avvalsa di queste informazioni almeno in due modi: sono stati individuati i documenti che hanno suscitato il maggior interesse dei visitatori ed è stato eliminato l'uso dei frame.

Tra i documenti che sembravano aver catturato l'interesse della maggior parte dei visitatori sono state individuate due tipologie, una più tecnica dedicata ad un pubblico esperto di studi sulla popolazione, e l'altra più divulgativa per tutti gli utenti. Avendo compreso che i nostri visitatori si componevano anche di una quota (molto importante) di utenti comuni non in grado di comprendere terminologie e questioni troppo specifiche, abbiamo cercato di dare maggior rilievo alle pubblicazioni di materiale divulgativo, invitando anche i colleghi dell'Istituto a produrre altro materiale di questo tipo.

L'eliminazione dei frame ha consentito la progettazione di una struttura a *portale*¹⁷ con molte informazioni già nella prima pagina, in modo da orientare immediatamente il visitatore verso l'informazione che più lo

¹⁷ Un portale Web è un sito web che costituisce un punto di partenza, una porta di ingresso ad un gruppo consistente di risorse di Internet o di una Intranet.

interessava. Questa struttura, però è risultata piuttosto laboriosa da realizzare ed è già in fase di progettazione una nuova struttura che potrà avere gli stessi vantaggi in termini di accessibilità, ma, sfruttando le proprietà dei *fogli di stile*¹⁸ e le potenzialità dei linguaggi di programmazione sul web, sarà molto semplificata nella realizzazione e nell'aggiornamento.

Conclusioni

Con l'ultima progettazione del sito (2004) si sono realizzate tutte e tre le fasi già previste nella prima progettazione (1997-1998), dove ognuna delle quali conteneva un pacchetto di risorse implementabili separatamente in quanto dotate di una propria autonomia funzionale. Quindi, la prima fase ha riguardato l'approfondito le seguenti voci: istituto, sezioni di ricerca, osservatori, indagini, convegni e seminari, eventi e notiziario, prodotti e servizi. La seconda e la terza fase la creazione di una mailing list e la realizzazione del catalogo on line della biblioteca.

La gestione del sito web dell'istituto ha richiesto nel tempo competenze sempre più affinate e mirate ad ottimizzare il servizio offerto con le pagine pubblicate sia per gli utenti interni che per gli esterni. In tale ottica si pone la necessità di sviluppare ed applicare competenze di data mining ai dati disponibili in rete: ciò consente, oltre alla rilevazione delle aree del sito di maggior interesse, anche l'ottimizzazione del sito stesso in termini di fruibilità e facilità di accesso. Oggi è attivo un osservatorio per il monitoraggio delle visite e dell'utilizzo del sito. L'analisi statistica degli accessi ha lo scopo di verificare l'effettivo utilizzo del materiale pubblicato e di formulare ipotesi sulle aspettative e le richieste di materiale di ricerca e di divulgazione sui temi di interesse per l'Istituto.

La progettazione del sito è in continua attività e fa tesoro sia dei suggerimenti degli utilizzatori del sito, sia dei risultati dell'analisi statistica delle visite, sia, infine, dei progressi tecnologici che nel tempo stanno mettendo a disposizione degli sviluppatori del web nuove opportunità di realizzare servizi innovativi e sempre più efficienti.

¹⁸ I fogli di stile o fogli di stile a cascata - dall'inglese CSS, Cascading Style Sheets - sono un linguaggio di stile che permette di dare una formattazione ad un documento HTML o XHTML.

Bibliografia

- M. J. Berry, G. Linoff, (1997) *Data Mining Techniques*, New York, Wiley.
- L. Cerbara, M. G. Caruso, (2004) *Un anno di rete: analisi statistica del report log del sito www.irpps.cnr.it, 22/05/2003 – 16/06/2004*, WP 04, IRPPS.
- CNR, (2002) *Corso di formazione "I siti Web delle strutture CNR"*, Roma.
- Patrick J. Lynch, Sarah Horton, (2001) *Web Guida di stile – Progettazione dei siti web*, Apogeo, Milano.
- Giudici P., (2001) *Data mining – metodi statistici per le applicazioni aziendali*, McGraw-Hill, Milano.
- Silani S., Tarantino M., (2001) *Tecniche di Data Mining con SAS Enterprise Miner*, SAS Institute.
- CNR, (2001) *Sito Web del CNR*, Relazione del gruppo di lavoro, Roma.
- L. Burzagli, P. Graziani, (1999) *Accessibilità di siti Web Problematiche reali e soluzioni tecniche*, Istituto di Ricerca sulle Onde Elettromagnetiche, CNR.

www.ifac.cnr.it/smid/accesso/accesso.htm.

www.html.it.

www.wikipeida.org.