

Sistemi informativi statistici

Sistemi informativi statistici

“Un dato statistico, per esprimere le sue potenzialità, dovrebbe essere parte, talora sostanziale, di un sistema informativo che contenga informazioni statistiche e non, formali ed informali, seguite dal momento della loro produzione, o raccolta, a quello della elaborazione e della diffusione, finalizzate al perseguimento di obiettivi di governo, gestione e controllo” (Chiandotto, 1985)

Sistemi informativi statistici

Nella pianificazione della produzione di dati statistici i due aspetti più rilevanti da un punto di vista strettamente tecnico-statistico sono quelli dell'**integrazione verticale** e dell'**integrazione orizzontale**.

Secondo il principio dell'integrazione verticale nella pianificazione della produzione dei dati deve essere garantita la coerenza delle scelte operate e delle definizioni adottate all'interno del singolo processo produttivo nel corso di tutte le sue fasi.

Il rispetto del principio dell'integrazione orizzontale assicura la possibilità di utilizzare informazioni provenienti da processi produttivi diversi che siano relativi ad una stessa area informativa.

L'osservanza di tali principi consente, quindi, di mettere in relazione le informazioni provenienti da fonti (raccolta di informazioni ottenute attraverso specifiche procedure di osservazione e misura) diverse.

L'insieme di processi produttivi affini e delle relazioni che possono essere instaurati fra loro forma un **sistema informativo statistico (SIS)**.

Sistemi informativi statistici

In pratica l'obiettivo di sviluppare un **SIS** può essere perseguito soltanto se si disegna un processo di produzione dei dati secondo un approccio integrato.

In altri termini il gruppo di progettazione non avrà come obiettivo solo la completezza e la coerenza interna della singola indagine, bensì anche quella del **SIS** all'interno del quale l'indagine va ad inserirsi pensando, quindi, in un ottica di **SIS** e non più di singola indagine.

All'interno di una azienda il disegno globale nel quale si innestano tutte le attività di produzione dei dati dovrebbe prevedere che per ciascun area di interesse possa essere costruito un **SIS** alimentato da processi produttivi affini.

Pertanto, ciascun **SIS** si configurerebbe come un sotto-sistema nell'ambito di una rete informativa più vasta che costituisce il sistema informativo complessivo.

Sistemi informativi statistici

Ovviamente in tale organizzazione il livello di coerenza ed organicità interno ad un sotto-sistema sarà più elevato che non a livello di sistema nel suo complesso.

Tutta la costruzione di un **SIS** poggia sul concetto di integrazione e cioè sulla possibilità di istituire confronti o legami tra dati provenienti da fonti diverse.

Condizione necessaria per l'istituzione di tali confronti o legami è che i dati vengano descritti in modo formale e non ambiguo. È cioè necessario esplicitare chiaramente quali sono le entità da osservare, (le unità di analisi, le unità di rilevazione), le dimensioni rilevanti per la comprensione del fenomeno (attributi delle entità), altre entità con le quali istituire relazioni.

Le metodologie di **progettazione concettuale** si pongono come obiettivo quello di agevolare la descrizione formale e non ambigua dei dati sia in ingresso che in uscita del processo produttivo in modo tale che possa essere interpretata in modo univoco anche da chi non appartiene al gruppo di progettazione.

Sistemi informativi statistici

Tale descrizione si riferisce esclusivamente agli aspetti formali della realtà (da qui l'appellativo di concettuale) e pertanto non solo è **indipendente** dai mezzi con i quali operativamente si producono i dati ma è anche **stabile** nel tempo.

In conclusione, il risultato dell'applicazione della metodologia di progettazione concettuale ad una specifica realtà di interesse è la **definizione di uno schema concettuale che contiene la descrizione formale e non ambigua dei dati disaggregati che dovranno essere prodotti (microdati) nonché dei dati aggregati (macrodati) che saranno il risultato dei processi di aggregazione ed elaborazione condotti su dati disaggregati.**

I risultati della fase di individuazione delle esigenze conoscitive e dell'analisi statistica stanno alla base della definizione del modello concettuale scelto per la rappresentazione della realtà.

Sistemi informativi statistici

Alla creazione dello schema concettuale vanno fatti seguire dei passi che consentono di calare lo schema concettuale ottenuto all'interno del particolare **ambiente tecnologico** individuato.

Queste fasi successive, denominate progettazione logica e progettazione fisica di un **SIS**, sono di stretta competenza degli informatici.

La fase di progettazione concettuale, che si colloca in posizione intermedia tra quella di analisi della realtà e quella di progettazione logica e fisica, coinvolge sia lo statistico sia l'informatico che **deve rendere tecnicamente realizzabili i legami che garantiscono il maggior grado possibile di integrazione orizzontale tra i processi produttivi.**

Sistemi informativi statistici

Il modello Entità-Relazione

Tra i modelli utilizzati per la rappresentazione della realtà di interesse il più diffuso è il modello entità-relazione caratterizzato da maneggevolezza ed adattabilità alle situazioni concrete.

Il modello si compone di un certo numero di elementi che devono essere isolati e messi in relazione l'uno con l'altro a partire dalla situazione reale.

Elementi rilevanti: **entità, attributi, relazioni e cardinalità.**

Sistemi informativi statistici

Il termine **entità** ha il significato più generico possibile ed indica qualunque cosa esista e sia distinguibile dalle altre. Non solo oggetti concreti ma anche concetti astratti possono essere entità. Un gruppo di entità simili forma un insieme di entità.

Esempi

Entità	Insieme di entità
Persona	Popolazione
Unità locali	Insieme di unità locali
Regione	Regioni

Sistemi informativi statistici

Gli **attributi** sono le proprietà delle entità che associano un valore estratto da un dominio di valori per quell'attributo con ogni entità di un insieme. Se un attributo o un insieme di attributi identifica univocamente ciascuna entità in un insieme di entità esso è detto **chiave** (il codice fiscale per l'entità persona).

Esempi

Entità	Attributo
Persona	Occupato
Unità locali	Attività economica
Regione	Ripartizione territoriale

“Occupato” potrebbe anche essere considerato un'entità. Stabilire se sia da considerare un'entità oppure un attributo dipende dall'importanza che si assegna allo stato di occupazione.

Se fosse infatti un'entità sarebbe possibile collegare gli occupati mediante legami logici diretti.

Nel caso in cui sia un attributo gli occupati possono essere collegati indirettamente tramite la corrispondente entità costituita dalla persona.

Sistemi informativi statistici

Una **relazione** fra insiemi di entità è una lista ordinata di insiemi di entità. Se c'è una relazione REL tra insiemi di entità $E(1)$, $E(2)$, ... $E(n)$, si presume che esista un insieme di n-ple chiamato REL. Tale insieme è detto relazione.

Esempi

Relazione binaria $E(1)$ E' MADRE DI $E(2)$

$E(1)$ insieme delle madri

$E(2)$ insieme dei figli

$[e(1);e(2)]$ singole coppie di madre e figlio identificate dalle loro chiavi

[Paola Bianchi; Mario Bianchi]

Relazione binaria $E(1)$ E' IMPIEGATO IN $E(2)$

$E(1)$ insieme degli impiegati

$E(2)$ insieme delle aziende

$[e(1);e(2)]$ singole coppie di lavoratore ed impresa identificate dalle loro chiavi

Relazione binaria $E(1)$ COSTA $E(2)$

$E(1)$ insieme delle merci

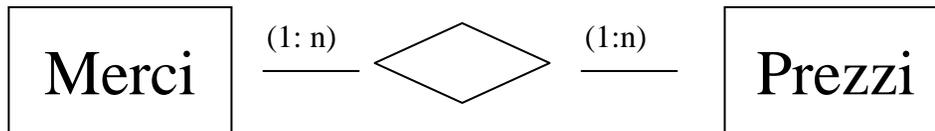
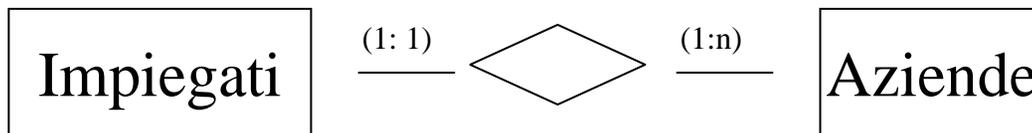
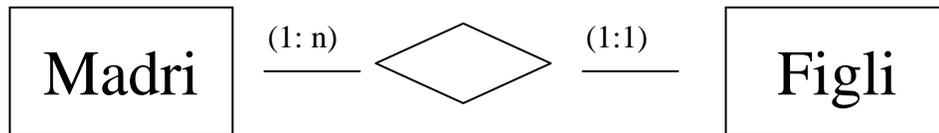
$E(2)$ insieme dei prezzi

$[e(1);e(2)]$ singole coppie di merce e prezzo identificate dalle loro chiavi

Sistemi informativi statistici

La **cardinalità** rappresenta il numero di legami che un generico oggetto di una qualsiasi classe (entità o relazione) può avere con oggetti appartenenti alla classe ad essa aggregata.

Esempi



Sistemi informativi statistici

Vantaggi della progettazione concettuale

- Viene garantita un'effettiva integrazione fra statistici ed informatici evitando così equivoci ed ambiguità
- Il rigore formale che viene richiesto favorisce la comunicazione tra gli operatori e la chiarezza delle specifiche di progetto
- Lo schema concettuale costituisce la migliore documentazione possibile del progetto perché la sua struttura consente di ricostruire agevolmente l'iter della sua formulazione
- Validità stabile nel tempo, infatti lo schema concettuale è costruito in modo autonomo ed indipendente dall'ambiente tecnologico all'interno del quale verrà poi materialmente realizzato (tutela dai cambiamenti indotti dall'evolversi della tecnologia e dall'avvicendamento degli interlocutori informatici abituali)

Sistemi informativi statistici

Principali vantaggi connessi alla costruzione di un **SIS**

- Utilizzare un approccio sistemico per la produzione dei dati che ha importanti implicazioni di carattere metodologico ed organizzativo (adottare criteri classificatori omogenei nella raccolta delle varie informazioni secondo diverse chiavi; collaborare più da vicino con tutti i soggetti coinvolti nel processo di produzione dei dati; ricercare con continuità forme di miglioramento della produttività dell'azione statistica accrescendo allo stesso tempo in modo significativo anche la qualità dell'informazione statistica)
- Usufruire di un potente e flessibile strumento di supporto alle decisioni
- Costruire quadri informativi articolati in grado di rappresentare fedelmente la realtà e di esplicitare i legami esistenti fra i diversi fenomeni
- Ridurre il carico che grava sui rispondenti attraverso l'implementazione di soluzioni metodologiche ed organizzative efficienti grazie alle quali ogni singola informazione possa essere utilizzata per più finalità

Sistemi informativi statistici

Fattori legati allo sviluppo tecnologico che hanno favorito il diffondersi dei SIS

- Disponibilità di **database relazionali** che rendono molto più agevole la gestione integrata dell'informazione abbattendo i costi di progettazione e sviluppo
- Possibilità di **gestire enormi masse di dati** a costi contenuti
- Esistenza di **strumenti software che rendono la progettazione di un SIS molto flessibile** consentendo di riorientare le scelte in corso d'opera e di introdurre cambiamenti senza mettere in discussione la struttura del sistema
- Riduzione dell'orizzonte temporale degli investimenti
- Possibilità di adottare in fase di sviluppo un approccio del tipo **bottom-up** (sviluppare sottosistemi destinati ad essere interconnessi in un secondo momento)

Sistemi informativi statistici

Alcune riflessioni sugli aspetti che caratterizzano un sistema informativo (SI) come statistico

La nozione di **SI** nasce in ambito **organizzativo-gestionale** ed è ampiamente utilizzata in ambito aziendale.

Un **SI** è rappresentabile come l'insieme di **strumenti automatici** di:

- ☞ memorizzazione,
- ☞ scambio,
- ☞ acquisizione,
- ☞ elaborazione,
- ☞ interrogazione,
- ☞ diffusione

di informazioni raccolte da un'azienda per soddisfare i propri **scopi** di **esercizio**, **controllo** e **pianificazione** basato su un'architettura informatica di tipo *database*. Alcuni autori estendono il concetto includendo le risorse umane e le norme organizzative.

Sistemi informativi statistici

Il processo di costruzione di un **SI** è normalmente caratterizzato dalle seguenti fasi:

1. definizione dei contenuti informativi;
2. studio di fattibilità;
3. progettazione concettuale dei dati;
4. analisi delle funzioni;
5. progettazione logica del sistema;
6. progettazione fisica;
7. implementazione e test;
8. manutenzione.

Sistemi informativi statistici

Il processo di costruzione di un **SIS** non si differenzia in modo sostanziale da quello di un **SI** infatti si è definito il **SIS** come un'ampia base informativa, generalmente costituita da diverse fonti d'informazione (eventualmente gestite da differenti organismi) strutturate e rese disponibili in funzione del loro utilizzo per lo studio statistico di particolari fenomeni.

In altre parole, un **SIS** attraverso l'utilizzo della tecnologia informatica dei *database* **permette di rappresentare dati provenienti da un gran numero di fonti con modalità di presentazione omogenee garantendo, tra l'altro, una elevata rapidità di accesso alle informazioni.**

Ciò che differenzia un **SI** da un **SIS** è la finalità statistica (in qualche modo un **SIS** può anche essere visto come un concetto che generalizza quello di indagine o quello di fonte di informazione) e pertanto nella specifica dei requisiti di un **SIS** bisogna tenere conto di aspetti quali: la dinamicità dell'offerta informativa di tipo statistico; la necessità di offrire all'utente un archivio di metadati.

Sistemi informativi statistici

I requisiti di un SIS

- **Estensibilità** così da potersi arricchire di informazioni aggiuntive
- Offerta di funzioni *on-line* e *user-friendly* per l'effettuazione da parte di un utente anche non esperto di **elaborazioni più o meno complesse**
- Possibilità di **gestione diretta di documentazione** relativa al contenuto informativo del sistema sotto forma di metadati
- Disponibilità di funzioni che consentono di **accedere direttamente alle fonti e ai dati** di interesse specificandone la definizione
- Integrità ed affidabilità
- Semplicità d'uso

Sistemi informativi statistici

Il ruolo della metainformazione

Un utilizzatore di un **SIS** è in grado di effettuare correttamente il confronto, sia in termini di contenuto che di metodologia, tra i dati di interesse soltanto se è disponibile a supporto del **SIS** un sistema di documentazione concepito e realizzato come un sistema integrato di metainformazioni.

Già da alcuni anni, in ambito internazionale, sono state attivate linee di ricerca tendenti ad individuare le **classi di metainformazione richieste dagli utenti** dell'informazione statistica e a proporre **sistemi informatizzati di organizzazione, gestione e distribuzione di metainformazione**.

In sostanza gli studi si sono concentrati nell'individuare e descrivere le classi di metainformazioni standard, connesse da relazioni anche complesse che sono **associate** ai concetti chiave con i quali si descrivono i **processi produttivi** ed i **contenuti delle indagini**.

Sistemi informativi statistici

E' possibile individuare, sulla base dei principali risultati della ricerca in tema di definizione di nozione di metadato, le seguenti tipologie:

1. CONTENUTI
2. LIVELLO DI ASTRAZIONE
3. SCOPO

OSS.: Differenti classi di **utenti**, così come diverse **attività del processo di produzione dei dati**, richiedono la loro peculiare classe di metadati

Sistemi informativi statistici

CONTENUTI

Metadati che descrivono il:

- a) contenuto informativo dell'indagine (fenomeni osservati, unità di rilevazione, unità di analisi, variabili e classificazioni corredate di definizioni)
- b) contenuto informativo dell'indagine vista come processo di produzione (fasi, operazioni [rilevazione, correzione dei dati, registrazione, revisione], azioni di controllo)

Sistemi informativi statistici

LIVELLO DI ASTRAZIONE

Metadati che descrivono:

- le operazioni ed il contesto organizzativo - informano sull'attività di produzione a livello di **singola procedura** con riferimento al contesto organizzativo (modalità operative di esecuzione [fasi, operazioni svolte, agenti delle operazioni, enti che coordinano agenti o operazioni], modalità di controllo dei processi produttivi [azioni di controllo per prevenire, correggere, stimare gli errori, agenti delle azioni di controllo], depositi di dati creati o utilizzati dall'indagine [questionario, file dei dati grezzi, file dei dati puliti])
- gli aspetti più propriamente statistici – informano su cosa è stato osservato e come è stato osservato (fenomeni, unità di rilevazione, unità di analisi, variabili e classificazioni corredate di definizioni, disegno di indagine)

Sistemi informativi statistici

SCOPO

I metadati si classificano come:

- **locali** – descrivono ogni fonte
- **globali** – descrivono in modo omogeneo le caratteristiche ed il contenuto complessivo di un insieme di indagini. Ovviamente è possibile definire i metadati globali solo se i metadati locali sono standardizzati, confrontati ed integrati

Sistemi informativi statistici

In sintesi:

I **SIS** sono sistemi multi-fonte e multi-utente pertanto:

- ☞ necessitano di **metadati globali**
- ☞ i progettisti dei SIS devono preoccuparsi di rendere disponibili non solo i dati ma anche i metadati globali sviluppando **sistemi specializzati nella definizione e gestione** di tali tipi di metadati che non sono di facile implementazione poiché la costruzione di un metadato globale richiede un delicato e complesso **lavoro di standardizzazione ed integrazione**.

Sistemi informativi statistici

Esempio di una possibile strategia per lo sviluppo di sistemi di gestione di metadati

Sviluppare **due** sistemi, nell'ottica di un'integrazione successiva, per:

1. gestire tutte le classi di metadati che descrivono l'**indagine come processo**
2. gestire le classi di metadati che riguardano i **contenuti informativi delle fonti** e fornire gli strumenti per confrontare ed integrare le diverse definizioni dei **contenuti informativi**

Sistemi informativi statistici

Un esempio di SIS: il Sistema Informativo Universitario (SIU)

Obiettivo: *costituzione di un Sistema Informativo Universitario orientato alla valutazione.*

Piano operativo:

1. Individuazione del fabbisogno informativo
2. Reperimento dei dati non disponibili
3. Sviluppo degli indicatori derivabili

Scelte di fondo:

1. Rendere il SIU utile ai diversi livelli decisionali (Istat, MIUR, CRUI, CNVSU, Atenei, Regioni, Province, Comuni)
2. Mettere a disposizione sia i dati elementari che un set minimo di indicatori descrittivi

Sistemi informativi statistici

Il database SIU - contenuti:

- Studenti
- Personale
- Finanze
- Ricerca
- Edilizia
- Contesto

Fonti:

- Atenei
- CINECA
- indagini sugli sbocchi occupazionali dei laureati
- spese sostenute dalle famiglie per la formazione
- Miur (FFO, diritto allo studio)
- Istat (statistiche demografiche, forze lavoro, bilanci universitari)

Il sistema di supporto all'attività di ricerca, accesso ed elaborazione dell'informazione di interesse offre diverse funzioni di manipolazione dei dati, che consentono in tempo reale l'estrazione della propria informazione di interesse, e l'accesso alle funzionalità offerte dai sistemi OLAP/Datawarehousing

Sistemi informativi statistici

Osservazioni

Nella progettazione del SIU si è previsto fin dalla fase iniziale di tener conto degli indicatori che avrebbero potuto essere utili per il monitoraggio del settore. Tale scelta ha comportato che i progettisti non perdessero mai di vista l'integrazione delle diverse fonti enfatizzando la **caratteristica del SIS come sistema per il raccordo delle conoscenze su un fenomeno**, piuttosto che come un bacino in cui convogliare informazioni diverse, seppure riconducibili ad uno stesso fenomeno.

Il SIU appare come un sistema concepito come base di conoscenza ed effettivamente modellato sulle esigenze dell'utenza.

L'esperienza di realizzazione del SIU ha consentito di rendere esplicita la distinzione dei due momenti che caratterizzano la progettazione di un **SIS** quella in cui si delinea un sistema coerente ed integrato di fonti e quella in cui si progettano i sistemi di supporto alla gestione ed all'uso dei dati che concretamente lo realizzano.