
Servizio DAF. Datawarehouse

Seconda giornata di formazione agli strumenti statistici

3 Dicembre 2012

Gianni Aristelli

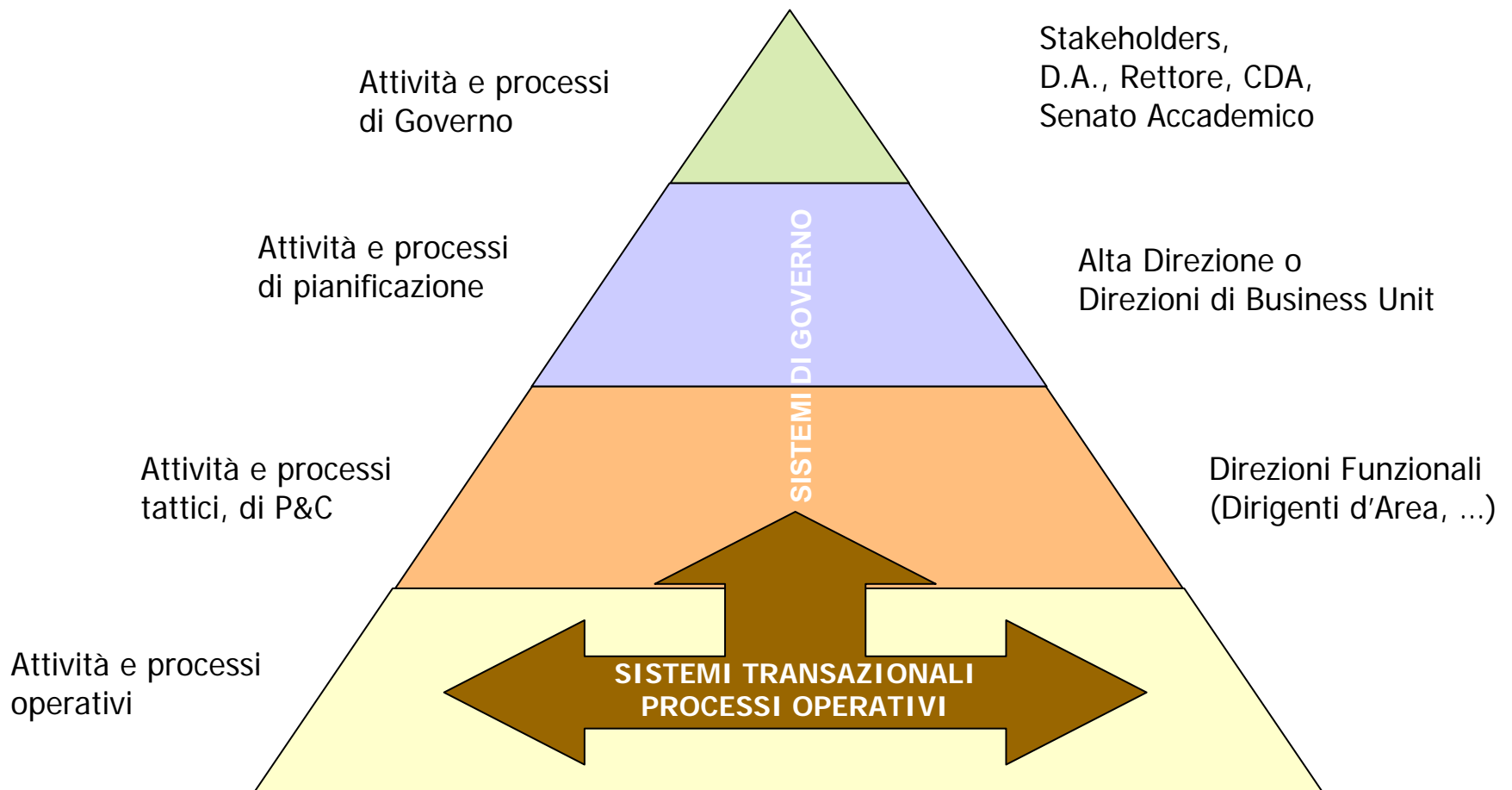
*Ufficio Servizi Statistici
Area Programmazione, Controllo e Sviluppo Organizzativo
Università degli Studi di Firenze*

Contenuti della giornata

- **Contesto**
 - Supporto alle decisioni
 - Processi dinamici
 - **Presentazione del servizio DAF**
 - Servizi di documentazione statistica
 - Servizi di datawarehouse
 - Documentazione
 - **Datawarehouse**
 - Cenni teorici e soluzione locale
 - Prima esplorazione del sistema, operatività e menù
 - Gli oggetti
 - Datamart e loro contenuto
 - **Valutazione del corso e dell'apprendimento**
-

Il contesto di impiego

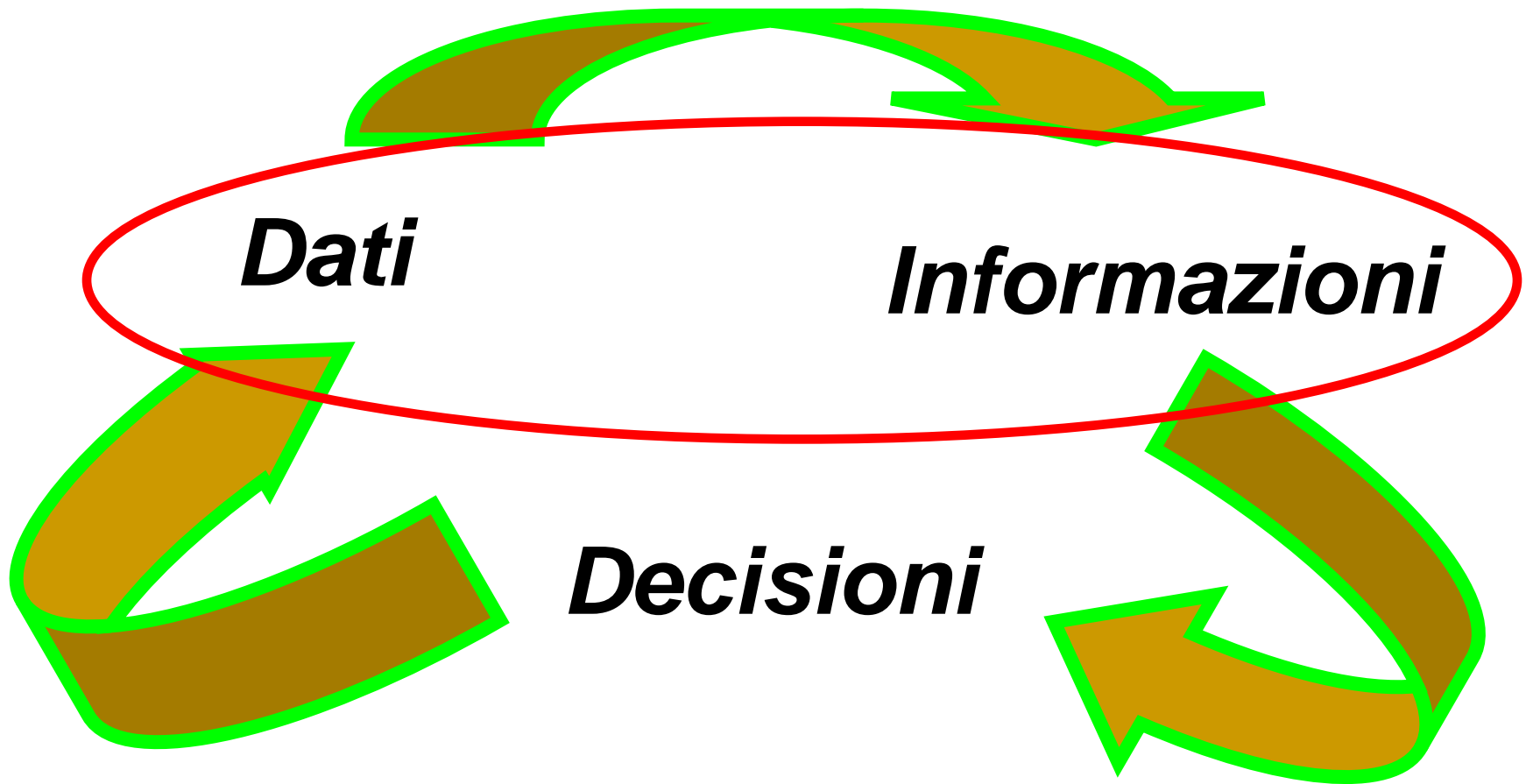
La piramide dei sistemi



Il ciclo del processo decisionale interno



Il flusso



Dinamica dei processi

A differenza del precedente corso DAF

Questo corso rispetto al precedente corso “DAF – Fonti statistiche” tenuto lo scorso Ottobre ha:

- principale somiglianza: utilizza dati di fonte amministrativa
- principale differenza: si concentra sui **dati dinamici**.

Si tratta di **dati dipendenti dal momento di osservazione**, che non hanno necessità di rappresentare situazioni ufficiali, descrittive di un fenomeno concluso e quindi invarianti nel tempo.

Si basa sui dati disponibili poco prima nei sistemi informatici gestionali centralizzati dell'Ateneo e come tale è vulnerabile alla tempestività delle registrazioni.

L'impiego dei dati è pertanto più delicato e richiede la contestuale conoscenza dei meccanismi di raccolta al momento di estrazione.

Rilevanza dell'evento generatore

- A prevalente natura amministrativa
 - Programmazione didattica
 - Gestione delle aule
 - ...
 - A prevalente richiesta dell'utente
 - Esami
 - Pagamento di fatture
 - Congedi e aspettative
 - ...
-

Le tante dimensioni del tempo ...

Fondo e flusso

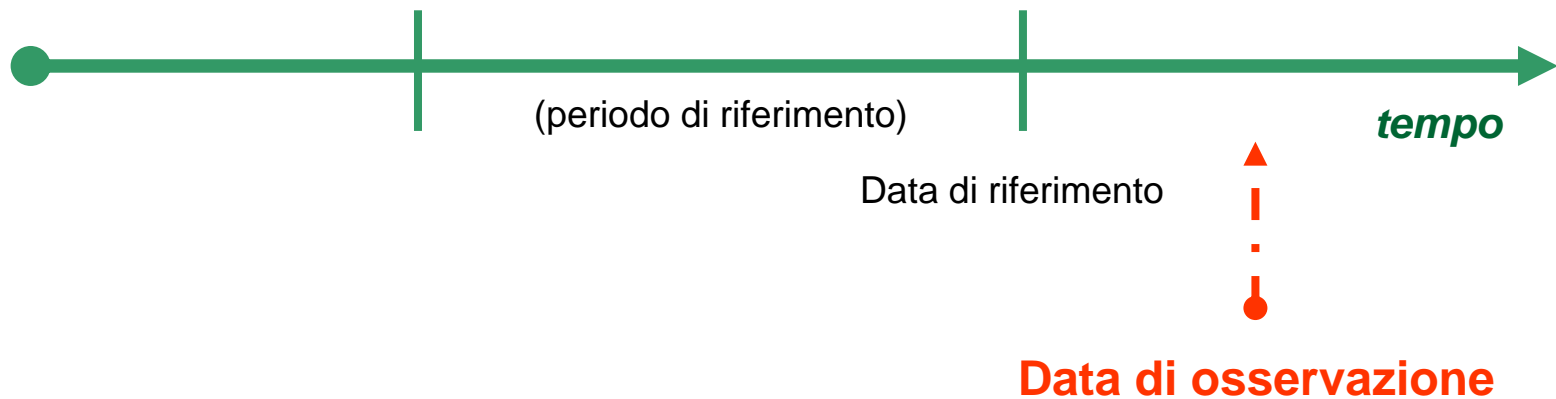
- Esami nell'anno solare
- Uscite per beni inventariabili
- ...

Flusso



- Personale in servizio
- Immatricolazioni all'A.A. ad una data
- ...

Data (periodo) di riferimento e osservazione



Il tempo nelle valutazioni

- **Ex-post**

Opera una volta che il fenomeno in oggetto si ritenga esaurito (accaduto)

- **In-itinere**

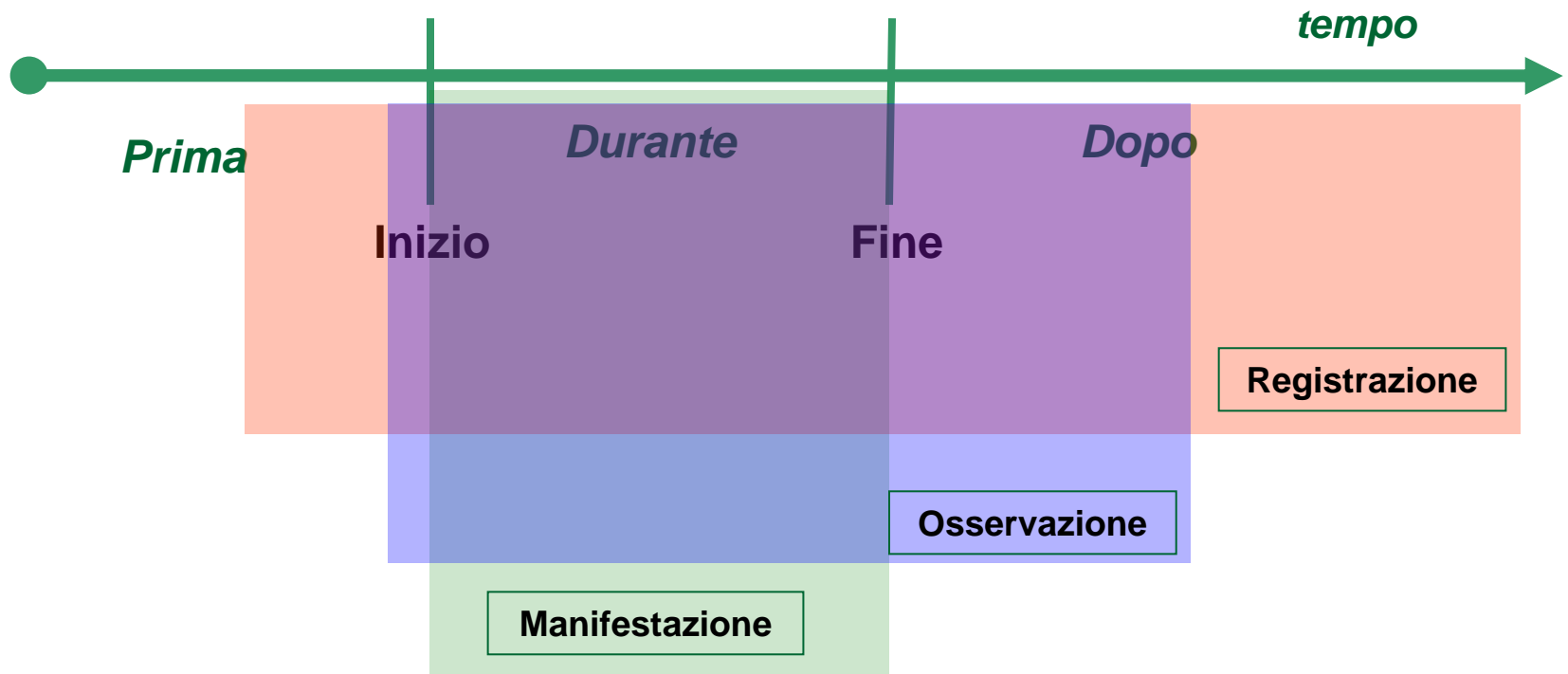
Opera tra il momento di inizio e di conclusione del fenomeno (monitoraggio)

- **Ex-ante**

Opera prima o mentre il fenomeno inizia (proiezioni, previsioni, ...) in dipendenza dalla disponibilità di dati

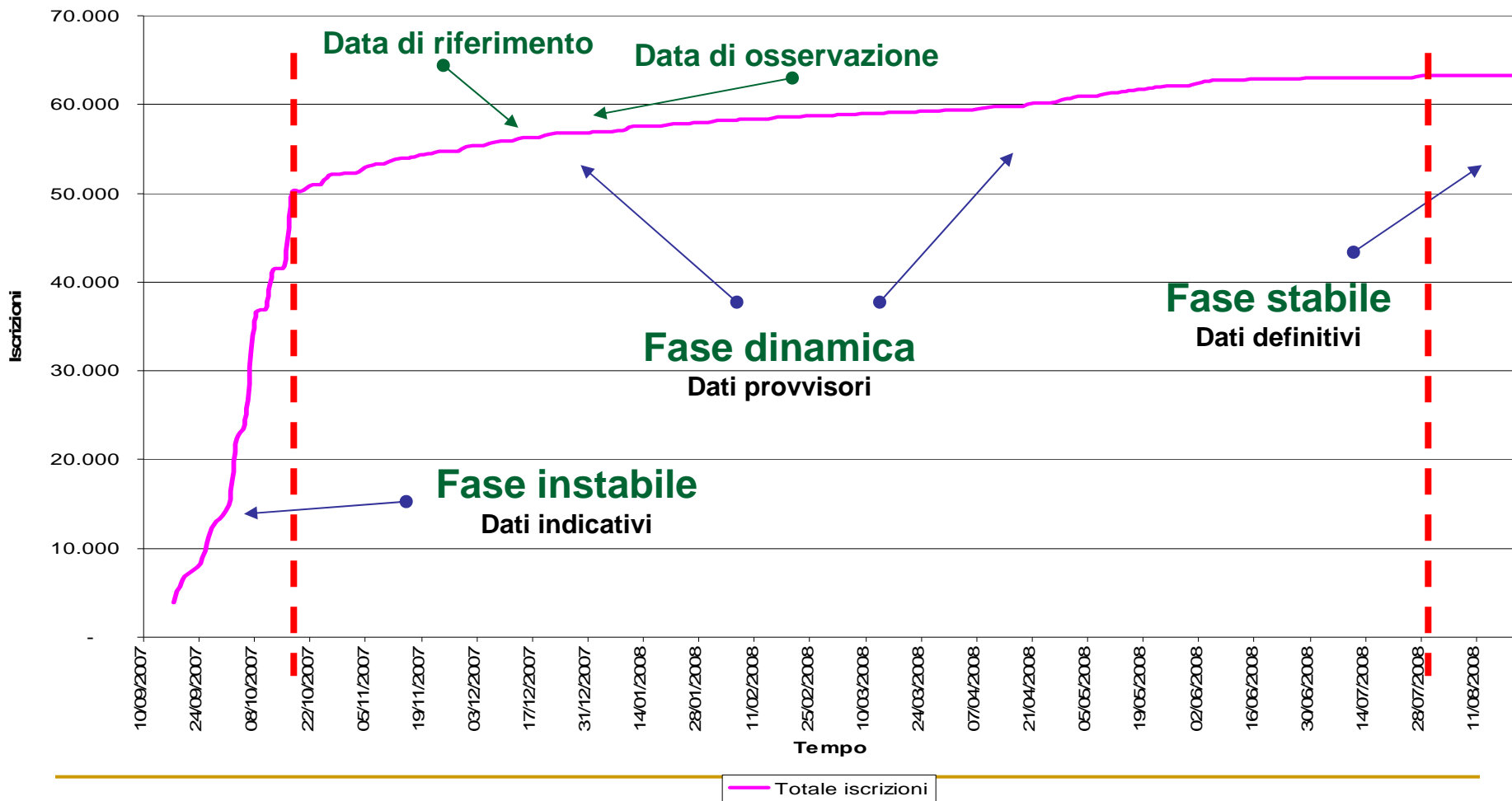


Registrazione, manifestazione e osservazione degli eventi



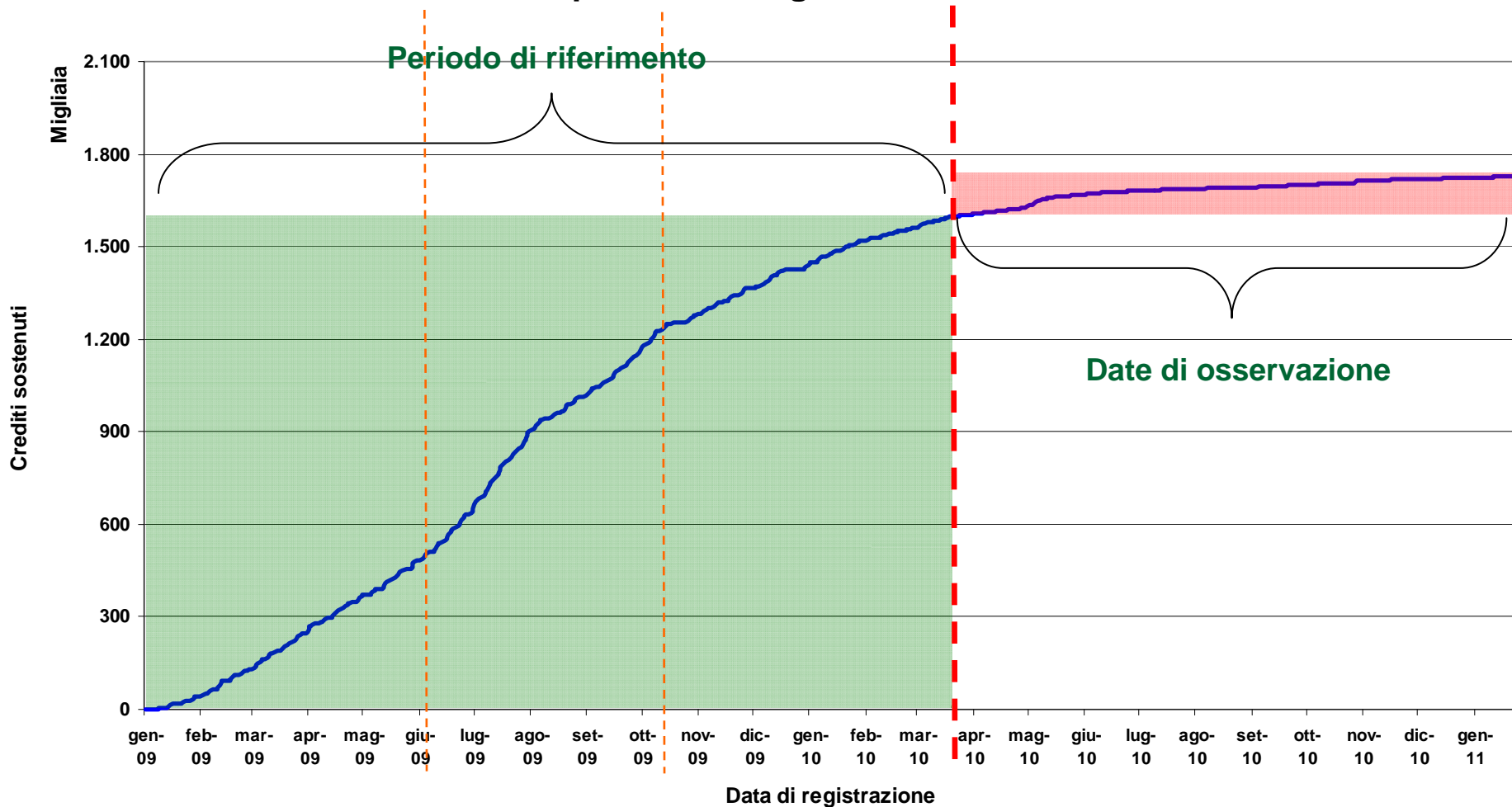
Il tempo nel processo delle iscrizioni

Avanzamento delle registrazioni di iscrizione. A.A. 2007/08



Un fenomeno a ritardo sistematico di registrazione

Cumulata dei crediti sostenuti nell'anno solare 2009
per data di registrazione

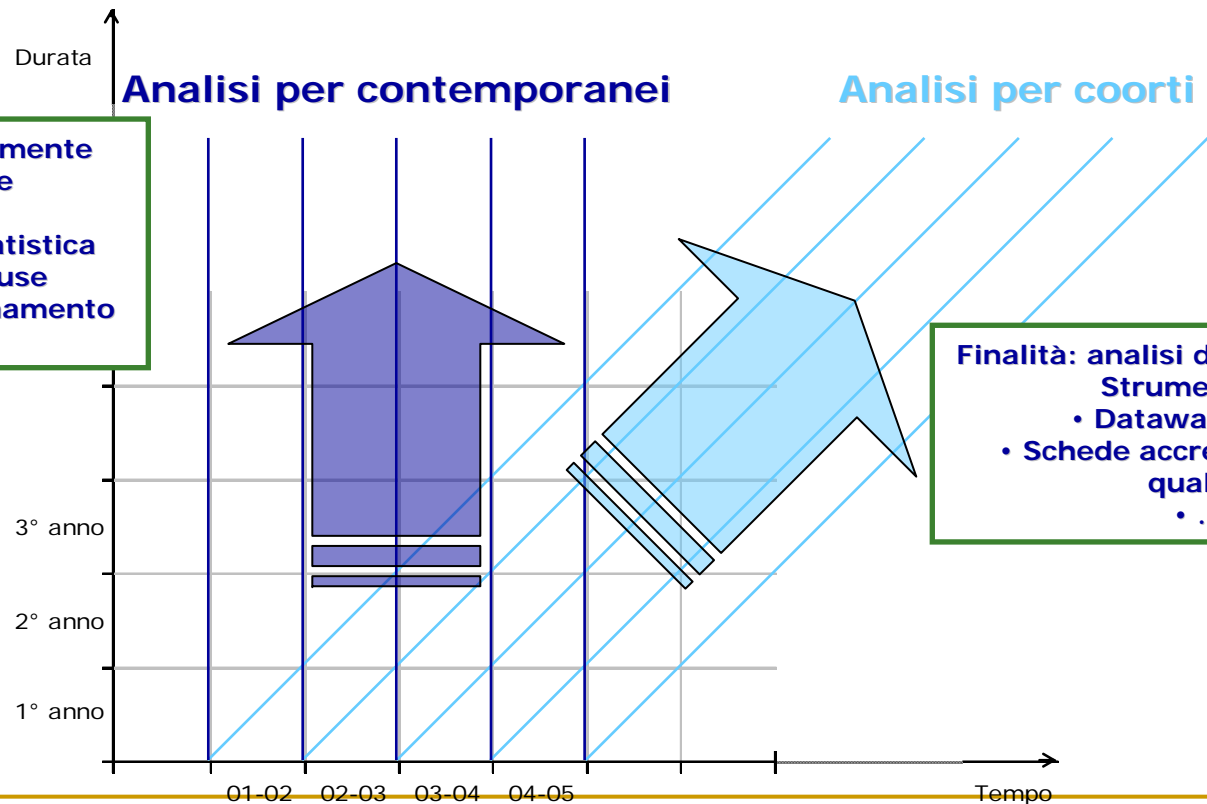


La variabilità dei tempi di registrazione degli esami

I tempi di registrazione sono estremamente variabili nelle diverse Segreterie.

Tempo di registrazione A.A. 2008/09	10% di tutti i verbali <i>Primo decile</i>	90% di tutti i verbali <i>Ultimo decile</i>
Situazione più favorevole	2 giorni	58 giorni
Situazione meno favorevole	20 giorni	338 giorni

Dati longitudinali e trasversali



Finalità: prevalentemente amministrative
Strumenti:

- Bollettino di Statistica
- Datawarehouse
- esami per insegnamento
- ...

Finalità: analisi della didattica
Strumenti:

- Datawarehouse
- Schede accreditamento / qualità
- ...

Il servizio DAF

L'origine

Il servizio DAF (Datawarehouse dell'Ateneo Fiorentino) è un servizio nato nell'Agosto 2005 in occasione dell'attivazione dei servizi di datawarehouse, per svolgere le relative funzioni di portale:

<http://www.daf.unifi.it>

Col tempo ...

Col passare del tempo il servizio si è evoluto, diventando a tutti gli effetti un sistema di distribuzione e accesso all'informazione ufficiale e non, prodotta dall'Ufficio Servizi Statistici nell'ambito della propria attività.

Che cos'è

Oggi il servizio è costituito da un sito web che distribuisce a tutti gli utenti le informazioni documentali tipiche di un portale ed ai soli utenti accreditati l'accesso ai contenuti.

In particolare gli utenti accreditati accedono a:

- l'informazione statistica ufficiale
 - il servizio di datawarehouse
 - l'ulteriore documentazione tematica.
-

Accesso al servizio

Sono utenti di diritto del servizio ed accedono senza particolari formalità o di iniziativa dell'Ufficio:

- Organi di governo centrali: Rettore, Direttore Amministrativo, Prorettori, Dirigenti
- Organi di governo delle strutture: Presidi, Direttori di UADR, Presidenti di corso di studi
- Responsabili di UADR/Ufficio.

L'accesso per tutti gli altri utenti deve essere richiesto da un utente di diritto.

Utenti e profilo

Attualmente gli utenti accreditati sono circa 400 (<http://www.daf.unifi.it/CMpro-v-p-145.html>).

Gli utenti accedono secondo due-tre profili di accesso. Esistono ulteriori accessi personalizzati solo per specifiche esigenze.

Le sezioni di interesse

- Generale
 - Dati ufficiali e dati on-line
 - Bollettino di Statistica
 - Almalaurea
 - Il profilo dei laureati
 - Gli sbocchi occupazionali
 - Accesso ai dati personali
 - Analisi tematiche
 - Documentazione
 - Specifico
 - Programmazione risorse
 - Nuovi dipartimenti
 - Ulteriori elaborazioni
-

Esplorazione servizio DAF

...

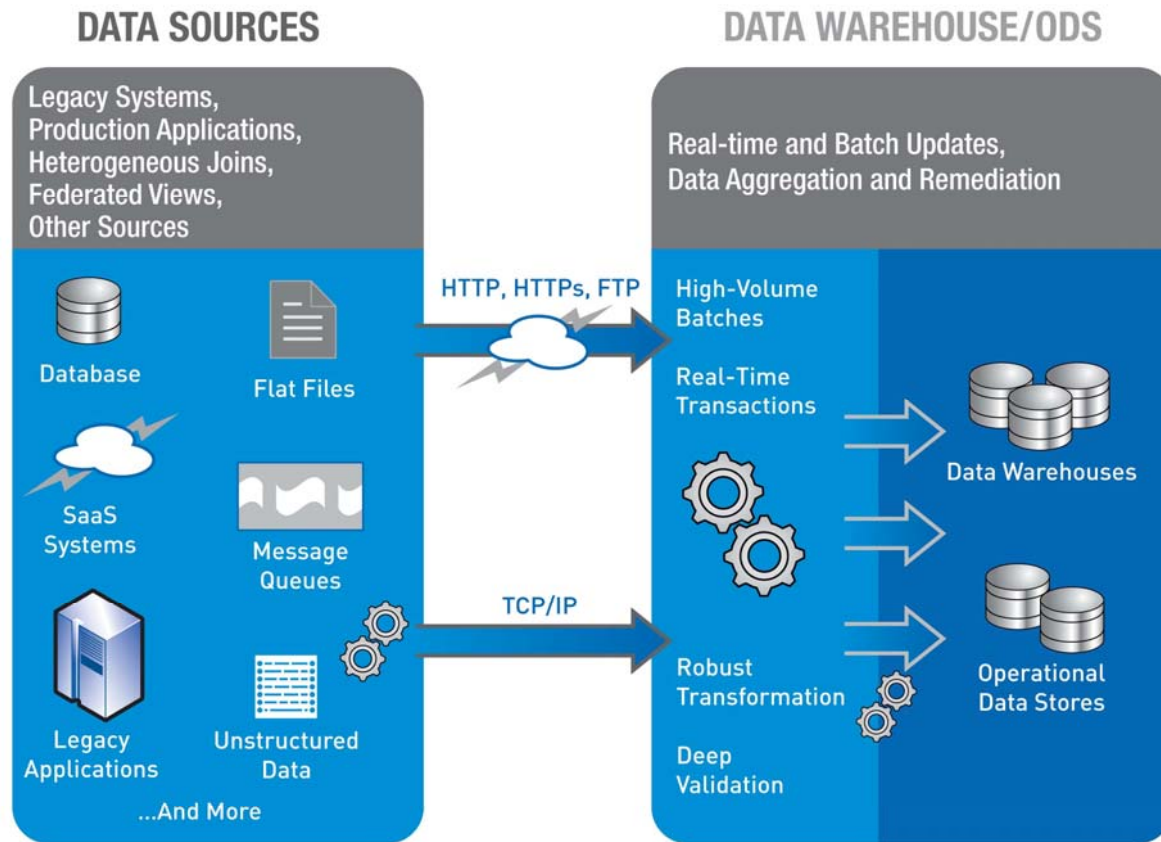
II Datawarehouse

Definizione

Un **Datawarehouse** (o **DW**, o **DWH**) (magazzino di dati) è un archivio informatico progettato per facilitare l'analisi dei dati che contiene le informazioni rilevanti di un'organizzazione.

Sono componenti essenziali anche gli strumenti per localizzare i dati, estrarli, trasformarli e caricarli, come pure gli strumenti per gestire il dizionario dei dati.

Schema semplificato lato ETL



Pervasive integration provides the robust transformation and deep validation that adds value to data sources and speeds data usability for data consumers.

 = Pervasive Integration Engine

L'integrazione e la non-volatilità

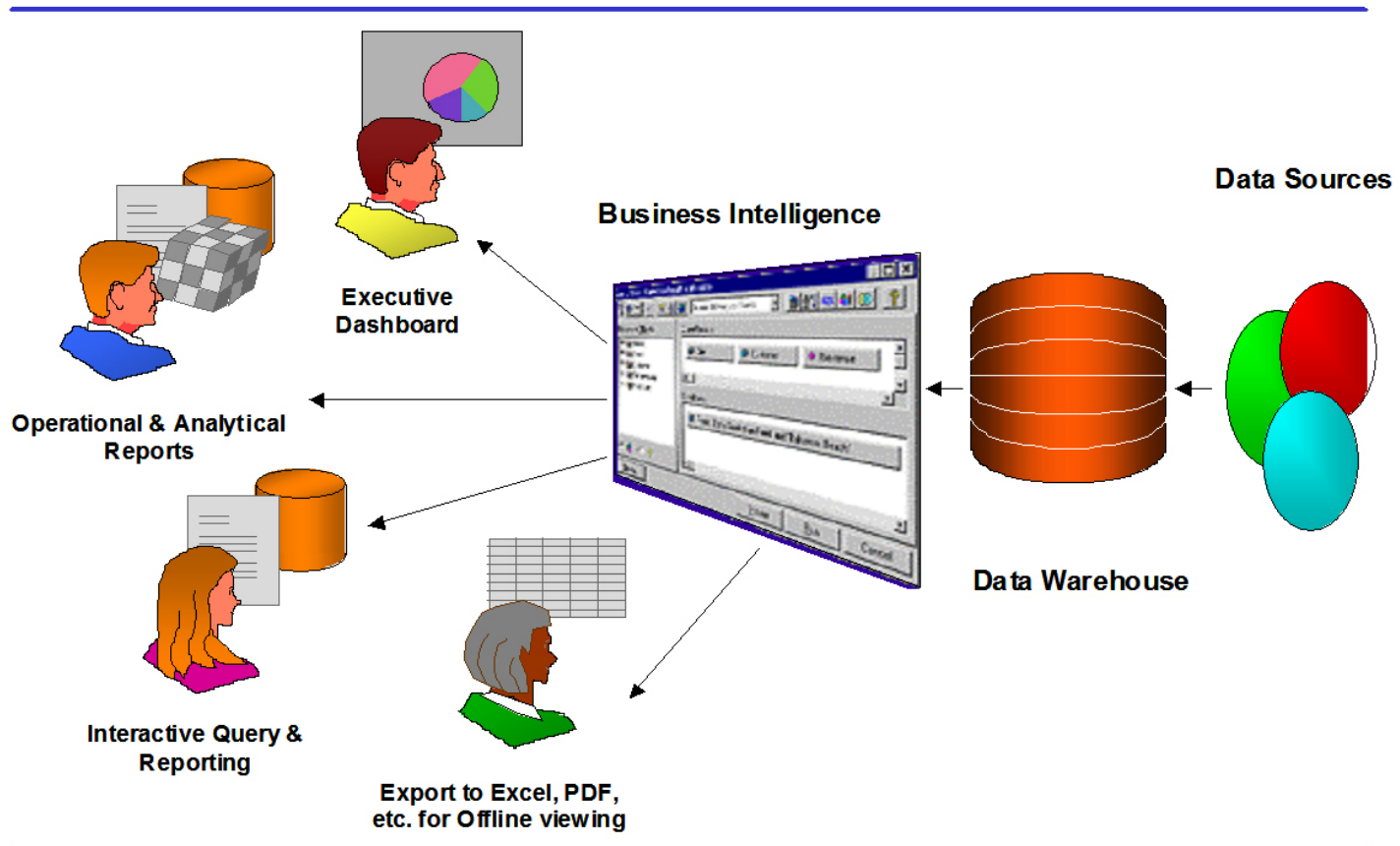
L'integrazione

- Nel datawarehouse confluiscono dati provenienti dai sistemi transazionali e da ulteriori fonti esterne.

La non-volatilità

- I dati del DWH non sono modificabili, gli accessi sono di sola lettura.
-

Schema semplificato lato impiego



La centralità delle informazioni

Il DWH si focalizza sulle informazioni

- Il motore è costituito da un'architettura ottimizzata di data-storage e navigazione
 - I dati vengono archiviati in modo da essere facilmente letti o elaborati dagli utenti
 - L'utente accede alle informazioni di cui ha bisogno, organizzate secondo la logica delle dimensioni di analisi.
-

La flessibilità

- E' pensato per la risoluzione di problemi non strutturati
 - A differenza dei tradizionali strumenti di reporting che presentano risposte preconfezionate, gli strumenti OLAP permettono la navigazione tra le informazioni, utilizzando le gerarchie e le relazioni tra le informazioni stesse come 'sentieri'
-

L'accesso e l'accessibilità

L'accessibilità

- I dati del DWH possono essere raggiunti tramite sistemi client-server proprietari e non, prevedono la profilatura degli utenti e l'impiego di browser per la consultazione.
-

La soluzione

Perché il partner CINECA

- dispone di un Dipartimento Soluzioni e Servizi per l'Amministrazione Universitaria che si occupa appunto del supporto di pianificazione e controllo, con notevoli competenze acquisite in quasi due decenni con numerosi Atenei in materia di datawarehouse e cruscotti
 - ha la quasi totale gestione dei nostri sistemi verticali
 - offre i servizi anche ad altri Atenei e supporta il MIUR per analoghe esigenze
 - utilizza il prodotto Microstrategy, dimostratosi alla prova dei fatti affidabile e completo.
 - svolge continue sperimentazioni su piattaforma open-source (Pentaho)
-

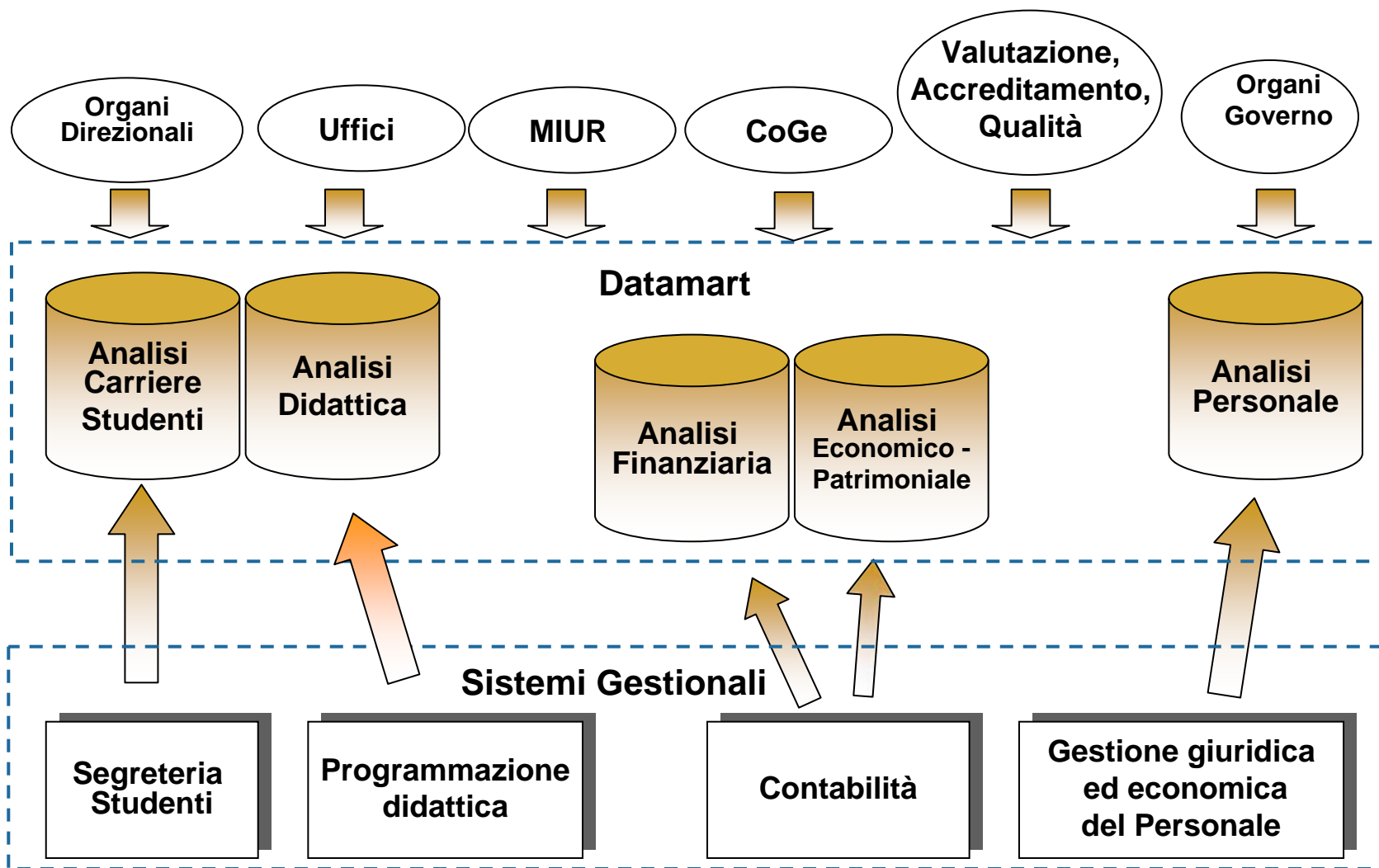
Aspetti tecnologici

- Integrazione dei dati provenienti dai diversi gestionali
 - Accesso on-line all'ambiente di analisi e reportistica
 - Possibilità di personalizzazione della reportistica
 - Esplorazione del dato a diversi livelli di aggregazione
 - Possibilità di analisi trasversale sui dati provenienti dai diversi ambienti gestionali
 - Disponibilità di dati storici o storicizzati
-

Gestione e consultazione del servizio

- L'amministrazione generale è svolta da CINECA
 - L'amministrazione locale, i contatti con CINECA e con gli utenti vengono tenuti dall'Ufficio Servizi Statistici
 - Gli utenti accedono in consultazione su un sistema di reportistica parametrica su due profili
 - Sono inibite le funzioni di modifica e creazione di nuovi oggetti
 - Sono previste occasioni di formazione per l'utenza tecnico-amministrativa. Sono allo studio occasioni di formazione per il personale docente
-

Fonti informative e attori



Cadenze di aggiornamento

Datamart	Tipo aggiornamento	Cadenza aggiornamento	Riferimento aggiornamento
Segreteria studenti	Totale il Lunedì e incrementale gli altri giorni	Lunedì, Mercoledì, Venerdì	Sera del Venerdì, Lunedì, Mercoledì precedenti
Programmazione didattica	Totale	Lunedì, Mercoledì, Venerdì	Giorno stesso
Economico-finanziaria	Totale (notare che l'esercizio chiuso è invariante nel tempo)	Lunedì, Mercoledì, Venerdì	Sera del Venerdì, Lunedì, Mercoledì precedenti
Personale	Totale	16 di ogni mese	Ultimo giorno del mese precedente

Aspetti organizzativi e funzionali

- Aspetti organizzativi

- Certificazione ed univocità del dato
- Semantica unica
- Possibilità di definire un SOLO processo organizzativo finalizzato alla produzione e diffusione dei dati

- Aspetti funzionali

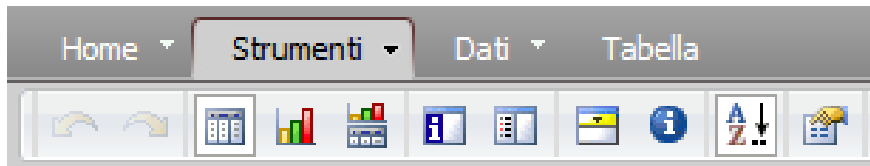
- Qualità e tempestività dell'informazione
 - Facilità di distribuzione massiva delle informazioni
-

Esplorazione del datawarehouse

Operatività col DWH

- Esecuzione di report e document
 - Drillatura (drilling)
 - Menù e funzioni disponibili
 - Esportazione di tabelle
 - Esportazione di documenti
-

Menù contestuali



Datawarehouse, gli oggetti

Oggetti

- Fatti
 - Metriche
 - Attributi
 - Filtri, prompt, ...
-

Fatti

I fatti sono gli oggetti di interesse per l'analisi, che guidano lo sviluppo degli elementi del modello (metriche, attributi, report, ...).

Esempio: iscrizioni, entrate, CFU, ...

Metriche o misure

Una misura o metrica è una proprietà numerica di un fatto e ne descrive un aspetto quantitativo di interesse per l'analisi. Esempio: numero di iscrizioni, ammontare delle entrate, CFU conseguiti ...

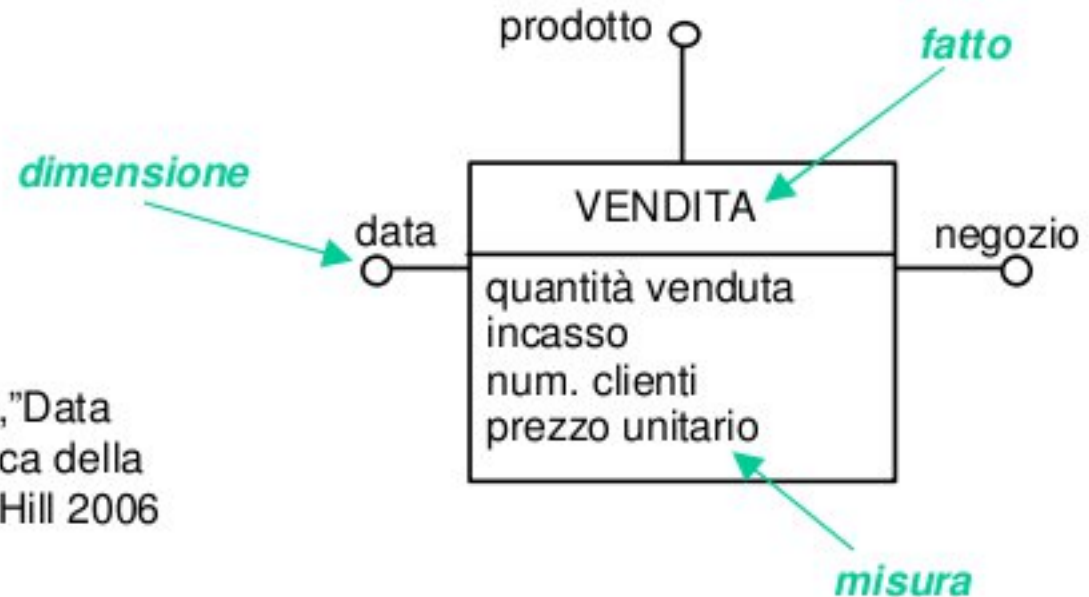
Dimensioni

Prospettive attraverso cui vengono analizzati i fatti d'interesse: una dimensione è una proprietà con dominio finito di un fatto e ne descrive una coordinata di analisi.

Le dimensioni sono caratterizzate da attributi logicamente correlati, generalmente di tipo categorico. Un fatto ha in genere più dimensioni che ne determinano la granularità minima di rappresentazione.

Esempio: tempo, dipendente ...

Fatti, misure, dimensioni



Tratto da Golfarelli, Rizzi, "Data warehouse, teoria e pratica della progettazione", McGraw Hill 2006

Attributi

Gli attributi agiscono come contenitori e riferimento delle informazioni, aggiungendo un contesto ai fatti misurati con le metriche. Sono tipicamente testuali e contengono valori discreti.

Esempio: corso di studio, UADR, Esercizio ...

Gerarchie

Relazioni logiche tra attributi che definiscono un percorso di navigazione dei dati all'interno delle dimensioni, e quindi il livello di aggregazione dei dati.

Esempio: anno, mese, giorno; Polo, UADR ...

Attributi e gerarchie

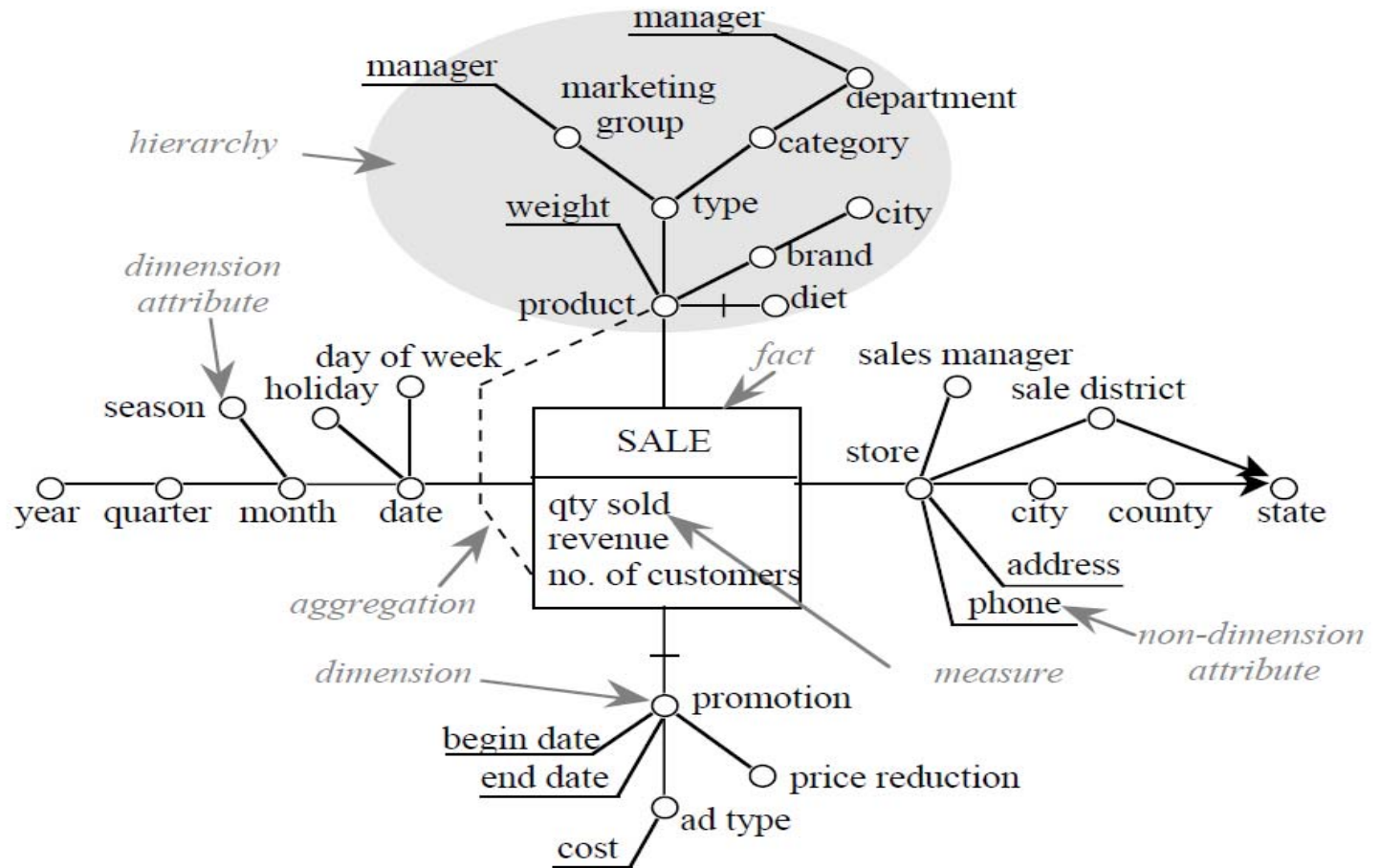


Drilling

Si riferisce alle operazioni e trasformazioni di dati multidimensionali

- Down/Up: è l'operazione di “esplosione/implosione” del dato nelle sue determinanti che sfrutta le gerarchie esistenti
 - Verso altre direzioni: è l'operazione mediante la quale si naviga liberamente dalle gerarchie
-

Dimensional Fact Model



Report o grafico

Un'opportuna composizione di oggetti presentata in forma tabellare prende il nome di REPORT, presentata in forma di raffigurazione di GRAFICO.

Possono essere generati da:

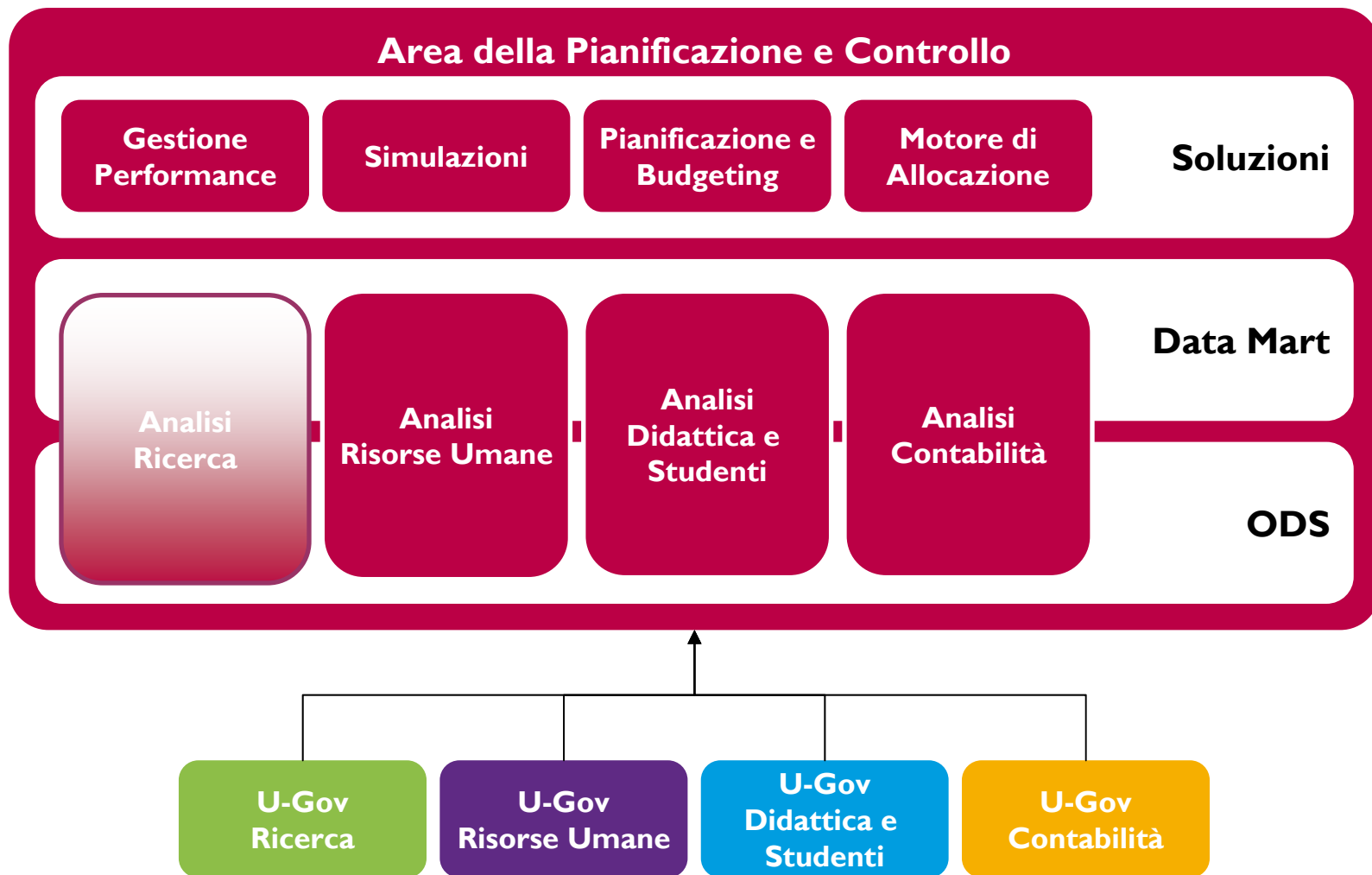
- attributi di classificazione
 - metriche per eseguire calcoli sui fatti
 - filtri e prompt per selezionare i dati da impiegare
 - custom group per raggruppare in classi valori continui
 - ...
-

Document o dashboard

Un document è una struttura capace di contenere una pluralità di supporti informativi: tabelle, grafici, testo, immagini. Il dashboard presenta le stesse informazioni del document ma in forma di cruscotto (attualmente non disponibile).

Le prospettive

La soluzione prospettiva UGov



Prospettive di breve e medio periodo di ampliamento del DWH

- Segreteria studenti
 - Abbandoni
 - Altri ingressi (trasferimenti, passaggi, ...)
 - Esami di stato
 - Programmazione didattica
 - Analisi dei regolamenti
 - Analisi su TAF e ambiti
 - Personale
 - Analisi economica
 - Contabilità finanziaria
 - Analisi per funzioni obiettivo
-

Prospettive di impiego rapido

- Supporto ai processi di accreditamento / qualità dei corsi (AVA / Scheda SUA?)
 - Supporto ai processi di miglioramento
 - Supporto ai processi di valutazione
 - Supporto ai processi decisionali di allocazione di risorse di Dipartimento o Corso di Studi
-

Grazie per l'attenzione
