

Un approccio mixed-mode sequenziale all'acquisizione dati

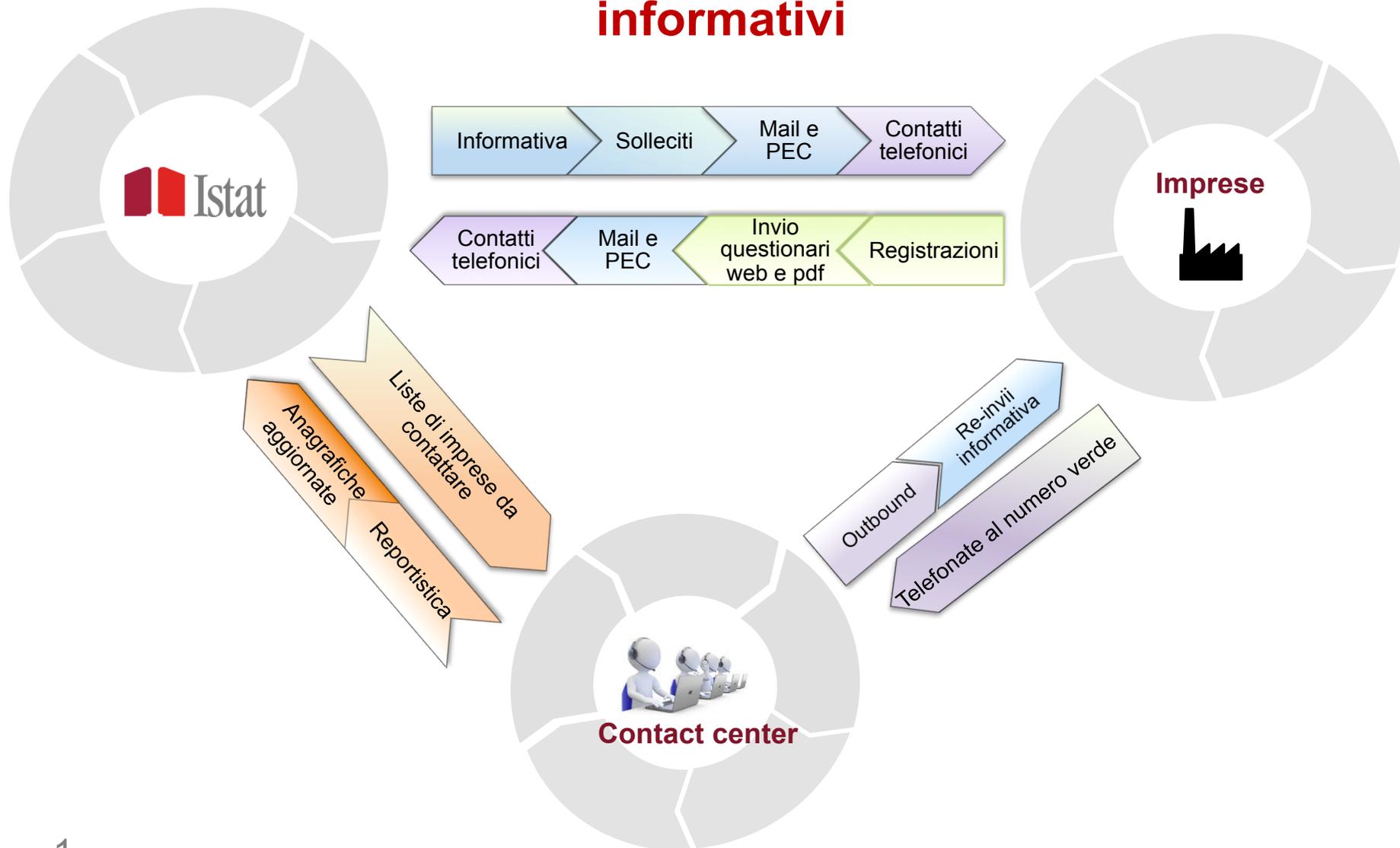
Stefano De Santis

17/02/2015

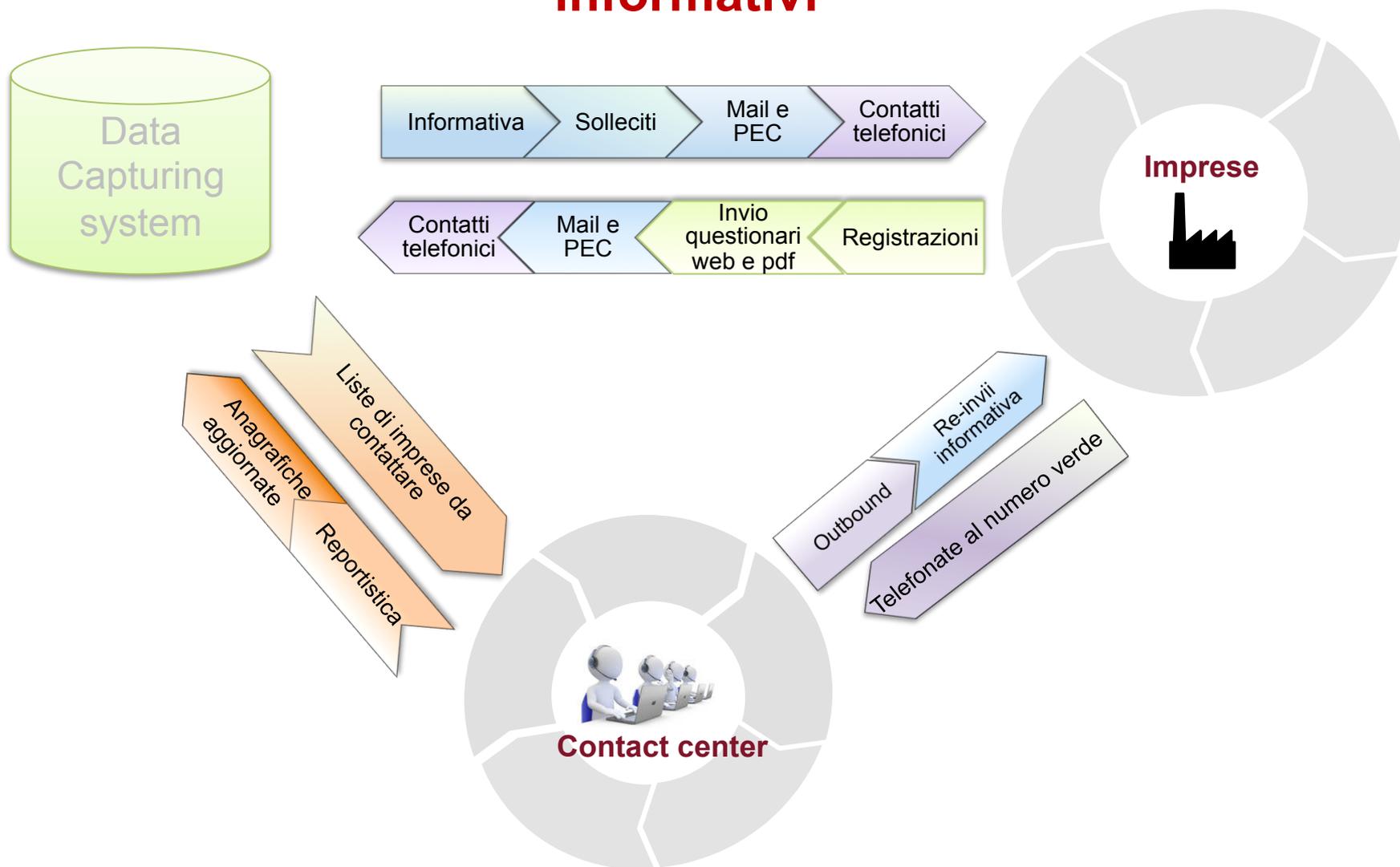
Indice

1. Il processo di indagine
2. Il questionario on line
3. Tracciato record
4. Il questionario off line
5. L'organizzazione dei dati
6. Conclusioni

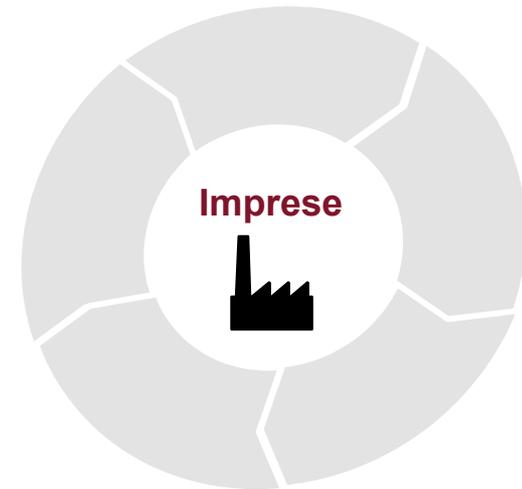
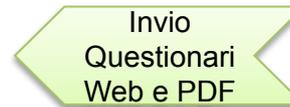
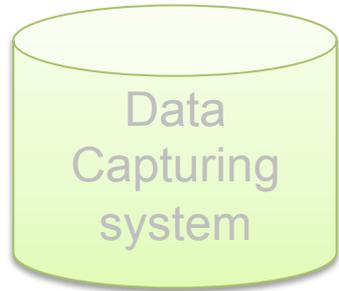
Il Processo di indagine: soggetti coinvolti e flussi informativi



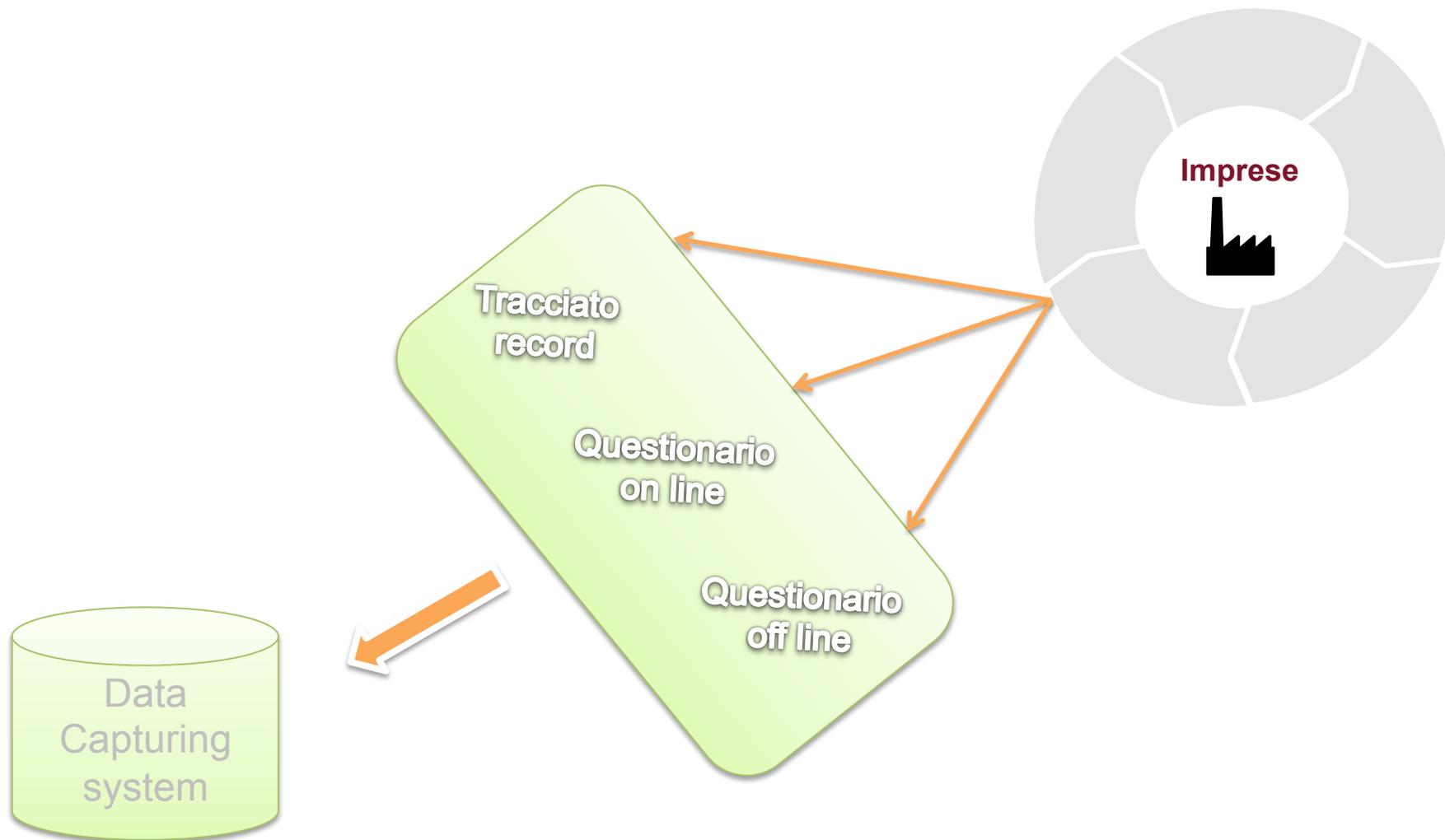
Il Processo di indagine: soggetti coinvolti e flussi informativi



Il Processo di indagine: gli strumenti di acquisizione



Il Processo di indagine: gli strumenti di acquisizione



Il questionario on line: punti di forza e svantaggi

GINO++ (Gathering Information Online)

Il questionario è:

- soggetto a regole (controlli di coerenza, calcolo automatico, vincoli di obbligatorietà, salti);
- corredato da aiuti alla compilazione (suggerimenti, box note, *tooltip*, *link* a fonti esterne)

Il questionario on line: punti di forza e svantaggi

GINO++ (Gathering Information Online)

E' articolato in Variabili (domande) che:

- assumono varie forme (*radio button, checkbox* ecc.)
- consentono l'immissione libera o obbligata dell'informazione (lista di *item* o completamento automatico ecc.)

Il questionario on line: punti di forza e svantaggi

GINO++ (Gathering Information Online)

Le Variabili sono raccolte in Quesiti e Sezioni al fine di consentire:

- la compilazione sequenziale o a salti
- la visualizzazione del questionario su pagine web distinte o su un'unica pagina
- il salvataggio e/o il controllo delle regole al cambio di sezione o a richiesta

Il questionario on line: punti di forza e svantaggi



VANTAGGI

Maggiore interattività con il rispondente

Disponibilità immediata e qualità dei dati raccolti

Possibile uso di questionari di “secondo livello ” per la correzione interattiva

Presenza di un database organizzato (Oracle) – Online Transaction Processing (OLTP) system

SVANTAGGI

Rischi sulla risposta (cultura informatica/esigenze differenziate degli utenti)

Poco consono alle esigenze di particolari classi di utenti

Problemi di concorrenza nell'accesso alle risorse informatiche

2

Tracciato come raccordo dei sistemi

ID SEZIONE	NOME SEZIONE	ID QUESITO	TESTO QUESITO	ID VARIABILE	NOME VARIABILE	LABEL OUTPUT	NOME VIEW	FORMATO VARIABILE	LUNG VAR	ID CLASS	MODALITA
1	Dati indagine	1	Nome indagine e tematica	89	Nome Indagine	Nome Indagine	VAR_89	S	100	0	
2	Dati rispondente	1	Periodo di riferimento	92	Anno riferimento indagine	Anno di riferimento:	VAR_92	S	100	0	
2	Dati rispondente	2	Dati Rispondente	66	Cod anagrafica rispondente	CODICE UTENTE:	VAR_66	S	100	0	
2	Dati rispondente	2	Dati Rispondente	91	Nome rispondente	RAGIONE SOCIALE:	VAR_91	S	100	0	
2	Dati rispondente	2	Dati Rispondente	4	Codice Fiscale	CODICE FISCALE:	VAR_J1	S	100	0	
3	INFORMAZIONI GENERALI	14	1.1 Indicare le ripartizioni territoriali in cui l'impresa è localizzata	158	Ripartizione Nord Ovest - Filtro	NORD OVEST	FILTRO_NO	N	1	8	1="si" 2="no"
3	INFORMAZIONI GENERALI	14	1.1 Indicare le ripartizioni territoriali in cui l'impresa è localizzata	152	Ripartizione Nord Ovest - Unità locali		UL_NO	N	5	0	
3	INFORMAZIONI GENERALI	14	1.1 Indicare le ripartizioni territoriali in cui l'impresa è localizzata	157	Ripartizione Nord Ovest - Totale Dipendenti		TOT_DIP_NO	N	9	0	
3	INFORMAZIONI GENERALI	14	1.1 Indicare le ripartizioni territoriali in cui l'impresa è localizzata	165	Ripartizione Nord Est - Filtro	NORD EST	FILTRO_NE	N	1	8	1="si" 2="no"
3	INFORMAZIONI GENERALI	14	1.1 Indicare le ripartizioni territoriali in cui l'impresa è localizzata	159	Ripartizione Nord Est - Unità Locali		UL_NE	N	5	0	
					[...]						
5	'OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	221	Ore ordinarie Dipendenti a tempo pieno	Dipendenza tempo pieno	DIP_FT_ORE_1	N	9	0	
5	OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	222	Ore ordinarie Dipendenti a tempo parziale	Dipendenza tempo parziale	DIP_PT_ORE_1	N	9	0	
5	OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	225	Ore ordinarie Totale Dipendenti	Totale Dipendenti	TOT_ORE_LAV_1	N	9	0	
5	OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	223	Ore ordinarie Apprendisti	DI CUI: Apprendisti	APP_ORE_1	N	9	0	
5	OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	221	Ore di straordinario retribuite Dipendenti a tempo pieno	Dipendenza tempo pieno	DIP_FT_ORE_2	N	9	0	
5	OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	222	Ore di straordinario retribuite Dipendenti a tempo parziale	Dipendenti a tempo parziale	DIP_PT_ORE_2	N	9	0	
5	OCCUPAZIONE, ORE, RETRIBUZIONI	13	2.2.2 Indicare il numero delle ore effettivamente lavorate	225	Ore di straordinario retribuite Totale Dipendenti	Totale Dipendenti	TOT_ORE_LAV_2	N	9	0	
					[...]						

e come base per costruire un ambiente di data warehouse...

ID RISPONDENTE RILEV	...	ID VARIABLE	...	VALORE	...	ID CLASS
1		111		3		
1		112		7		
1		100		via Anagni 24		
2		100		via Bellegra 35		
2		111		7		
2		112		13		

[...]



CODICE_UNITA	INDIRIZZO	DIPFT	DIPPT	DIP	RETRIMP
1	via Bellegra 35	7	3	10	20.000
2	via Anagni 24	13	7	20	40.000

[...]

...definizione di un database organizzato (Oracle) di analisi -
On-Line Analytical Processing (OLAP) system

Il tracciato record: punti di forza e svantaggi



VANTAGGI

- Aumento del tasso di risposta → idoneo per soggetti con elevata cultura informatica
- Possibilità di dialogo con i sistemi informativi complessi
- Inserimento massivo di grande mole di dati (riduce il carico sui sistemi)
- Possibilità di usare il tracciato come sintesi dei dati introdotti

SVANTAGGI

Possibile *bias* da mode selection e bassa interattività

Necessità di successive correzioni interattive / iterazione dell'operazione di upload

Necessità di raccordo con un database di produzione

4 Disponibilità differita dei dati e costi di gestione

Il questionario off line: punti di forza e svantaggi



VANTAGGI

Aumento del tasso di risposta → idoneo per soggetti con bassa cultura informatica

Buona interattività con il rispondente

Sicurezza, semplicità ed integrità nella raccolta dati

Annulla i costi di servizi esterni di data entry

SVANTAGGI

Possibile *bias* da mode selection

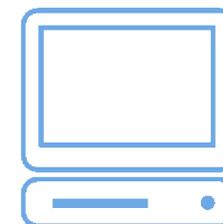
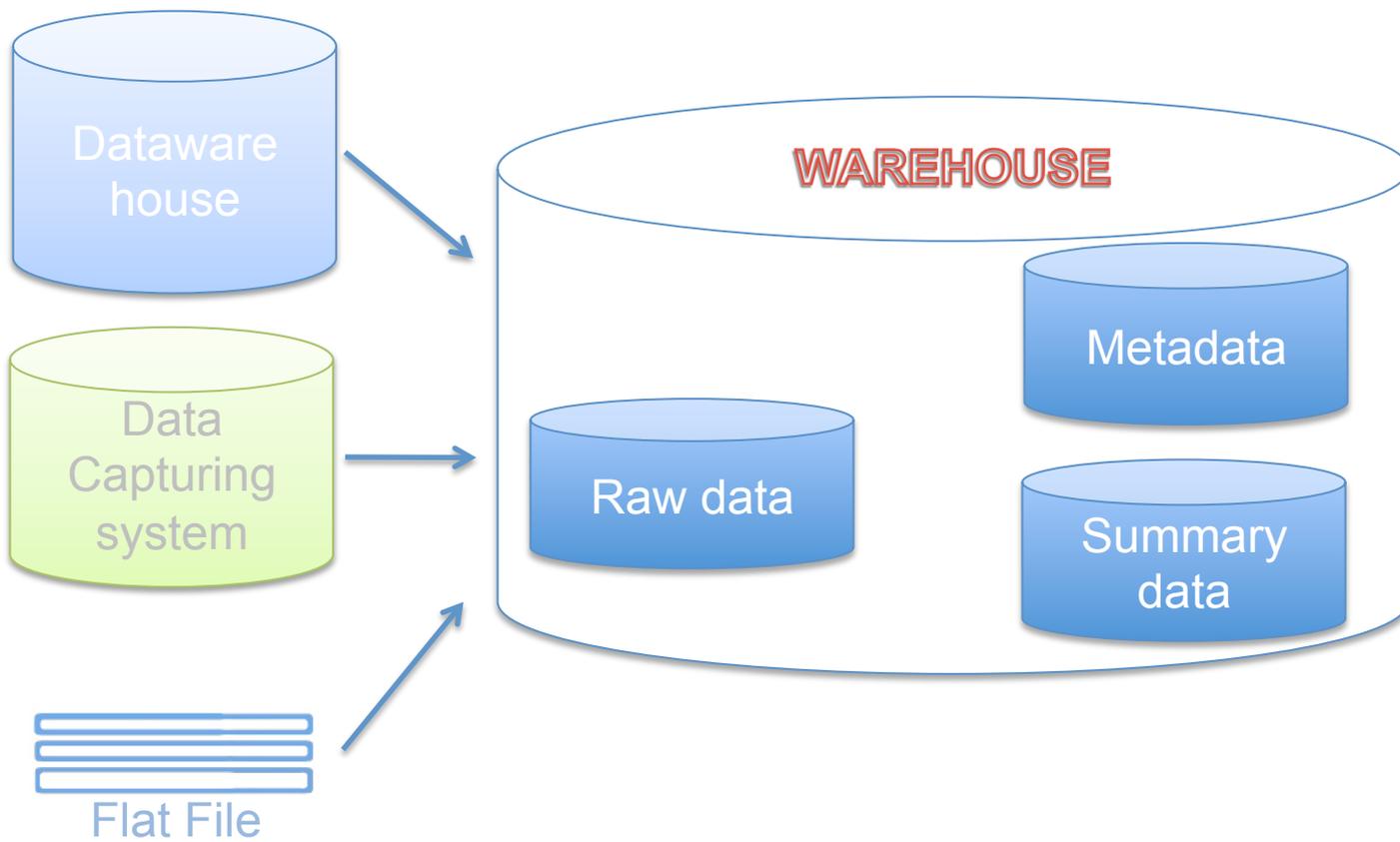
Disponibilità differita dei dati e costi di gestione

Necessità di raccordo con un database di produzione

Fonti di dati

Architettura del Data Warehouse

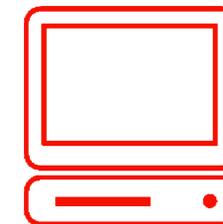
Utilizzo



Reporting



Monitoraggio



Analisi

Conclusioni

Multi-mode:

Aumento del tasso di risposta, ma possibile *bias* da mode selection

Match con le esigenze degli utenti, ma maggiore costo di gestione

Necessità di un database organizzato (OLTP /OLAP)

Database organizzato:

Monitoraggio in tempo reale (a livello micro e aggregato)

Analisi dei dati, reportistica, data mining

Consistenza dei dati a tutti i livelli di operatività del database

Grazie dell'attenzione!