

Verso un nuovo paradigma di integrazione fra dati di indagine, registri e dati amministrativi

Fabio Rapiti , Ciro Baldi

Seminario: “Innovazioni metodologiche e di processo in una rilevazione multi-source su imprese e istituzioni: la Struttura delle retribuzioni e del costo del lavoro 2012”

Istat – Aula Magna, 17 febbraio 2015

Contenuti

1. Due paradigmi più uno
2. Il diverso ruolo dei Registri nelle varie fasi di una rilevazione nei tre paradigmi
3. Come cambia l'approccio alle indagini campionarie
4. Futura RCL2016, SES2014 e GPG
5. Il sistema dei Registri nelle statistiche economiche già esiste in nuce ed è alla base della RCL
6. Considerazioni conclusive

Di quali paradigmi e modelli parliamo? E come si colloca l'esperienza della RCL?

Sample Survey based statistics

Tradizionale

Register assisted survey statistics

RCL (in posizione intermedia fra i due)

Register based (multi-source) statistics

Obiettivo principale della modernizzazione dell'Istituto

Transizione lenta in corso da anni nelle statistiche economiche anche in Italia

Al modello «sample survey» lentamente a partire dalla seconda metà degli anni '90 si è affiancato anche in Italia il nuovo modello



Nel tempo sono stati realizzati tanti progetti (Registri). Nella figura ne vengono citati solo alcuni come esempi e relativi al dominio di interesse

Consapevolezza della complessità della relazione dati amministrativi-indagini già a partire dalla fine degli anni '90

“In practice, in business statistics in some cases there is no alternative but to use administrative data...it must be done “cum grano salis”.

What we should be talking of here is integration and complementarily of survey sources rather than substituting data obtained by means of the traditional tools with data derived from administrative sources.

... the administrative source yields the mass of data, i.e. the universe of the population in question, while direct survey on small samples of the population provides a basis for appropriate processing of administrative data guaranteeing quality in terms of accuracy and comparability.

In some cases, therefore, the complementarily and integration we are dealing with appear markedly skewed, statistical survey serving largely as a qualitative support to the administrative source”.

Rapiti, Baldi (1999), “ Wages and employment official statistics using INPS data: a preliminary proposal and some methodological and quality problems” presentato alla Conferenza "Understanding the labour market using social security and administrative data" gennaio 1999 Ca' Foscari University of Venice, Economics Department pubblicato nei Contributi Istat n.16/1999.

Ruolo del Registro nelle fasi del GSBPM:

Sample survey based statistics

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design Collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination product	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Ruolo del Registro nelle fasi del GSBPM:

Register assisted survey statistics

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design Collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination product	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Ruolo del Registro nelle fasi del GSBPM:

Register based statistics

Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design Collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination product	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

Come cambia l'approccio alle indagini campionarie in un sistema di Registri (1)

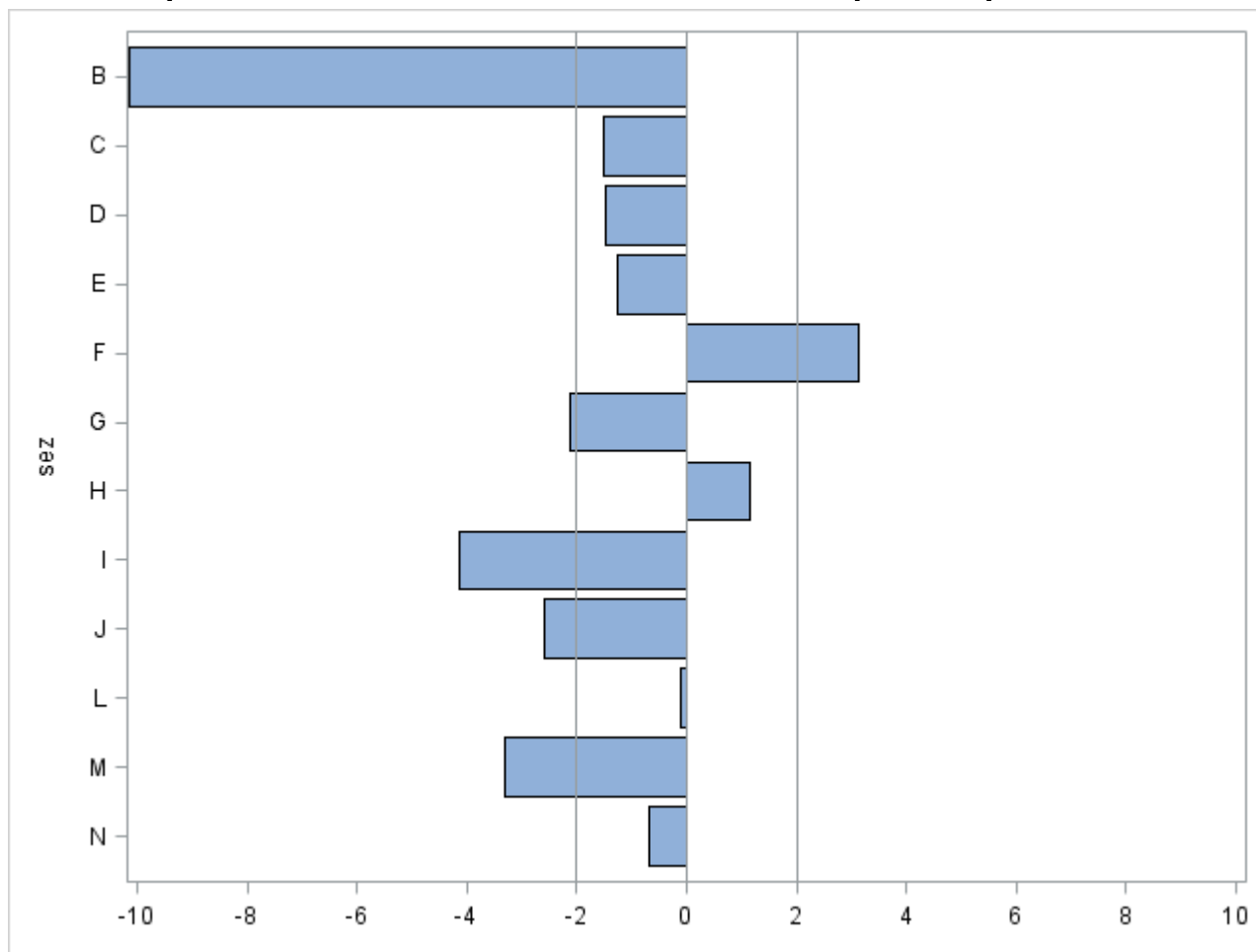
- Maggiore interdipendenza fra tutte le rilevazioni
- Maggiore coerenza-comparabilità
- Non è una linea evolutiva unica che vale per tutti le indagini campionarie (soprattutto quelle congiunturali)
 - rimangono diversi domini dove sarà necessario continuare ad effettuare indagini campionarie
 - Esempio nel dominio delle statistiche sul lavoro:
 - Posti Vacanti
 - Ore lavorate

Come cambia l'approccio alle indagini campionarie in un sistema di Registri (2)

- Progettazione dei questionari in funzione dei registri e dati amministrativi disponibili → **Register data assisted survey**
 - in modo complementare
 - da studiare/valutare 2 opzioni:
 - questionari precompilati,
 - precompilazione nascosta
- Identificazione e correzione degli errori di misura basata sui registri/dati amministrativi
 - anche «in corsa» se archivi/registri sono tempestivi
- Analisi e correzione della non-risposta e dei problemi di copertura grazie a variabili correlate da registri
 - BLS, Statistics Canada svolgono queste analisi regolarmente

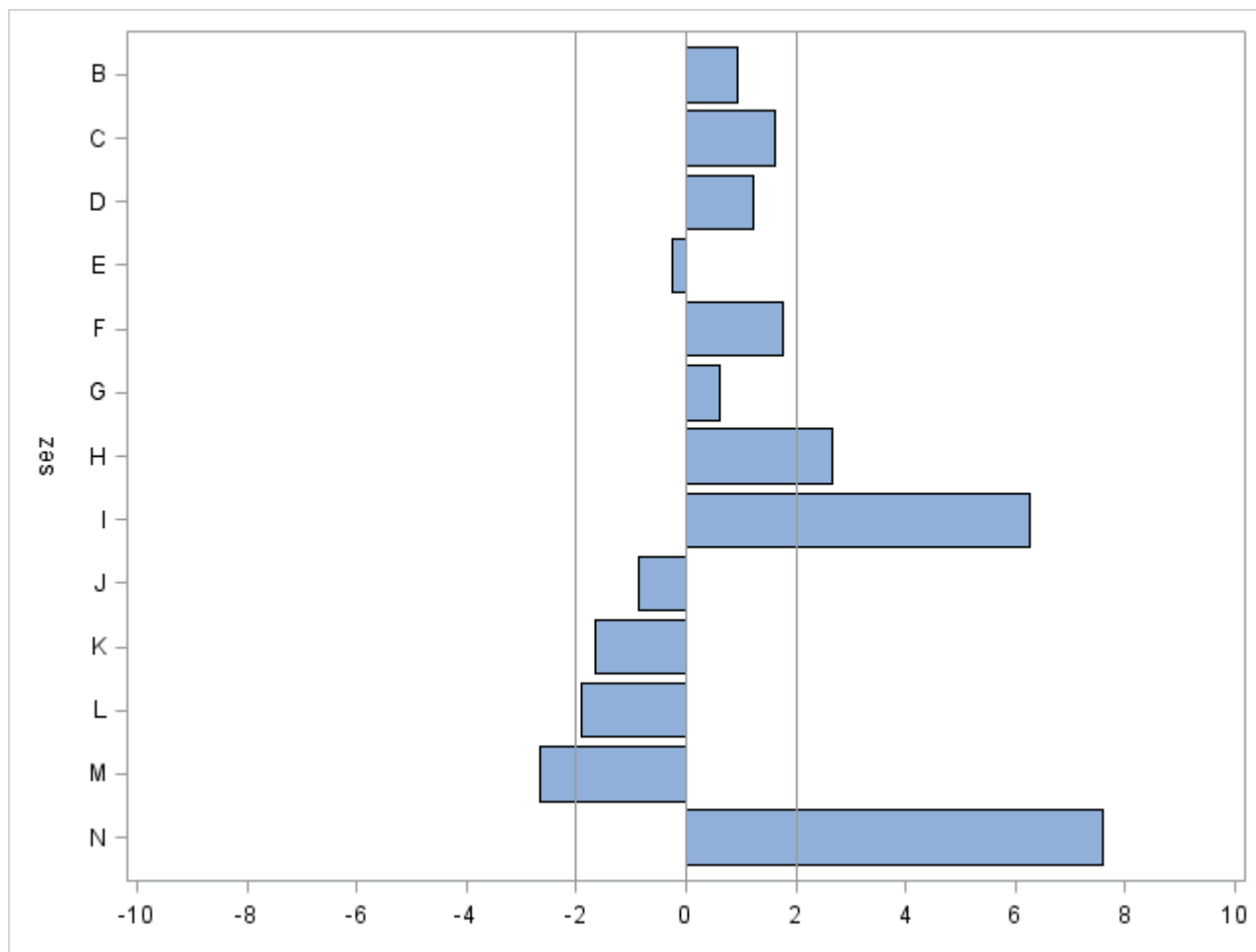
Comparabilità: validazione SBS VS LCS

Differenza percentuale nelle retribuzioni per dipendente SBS-LCS

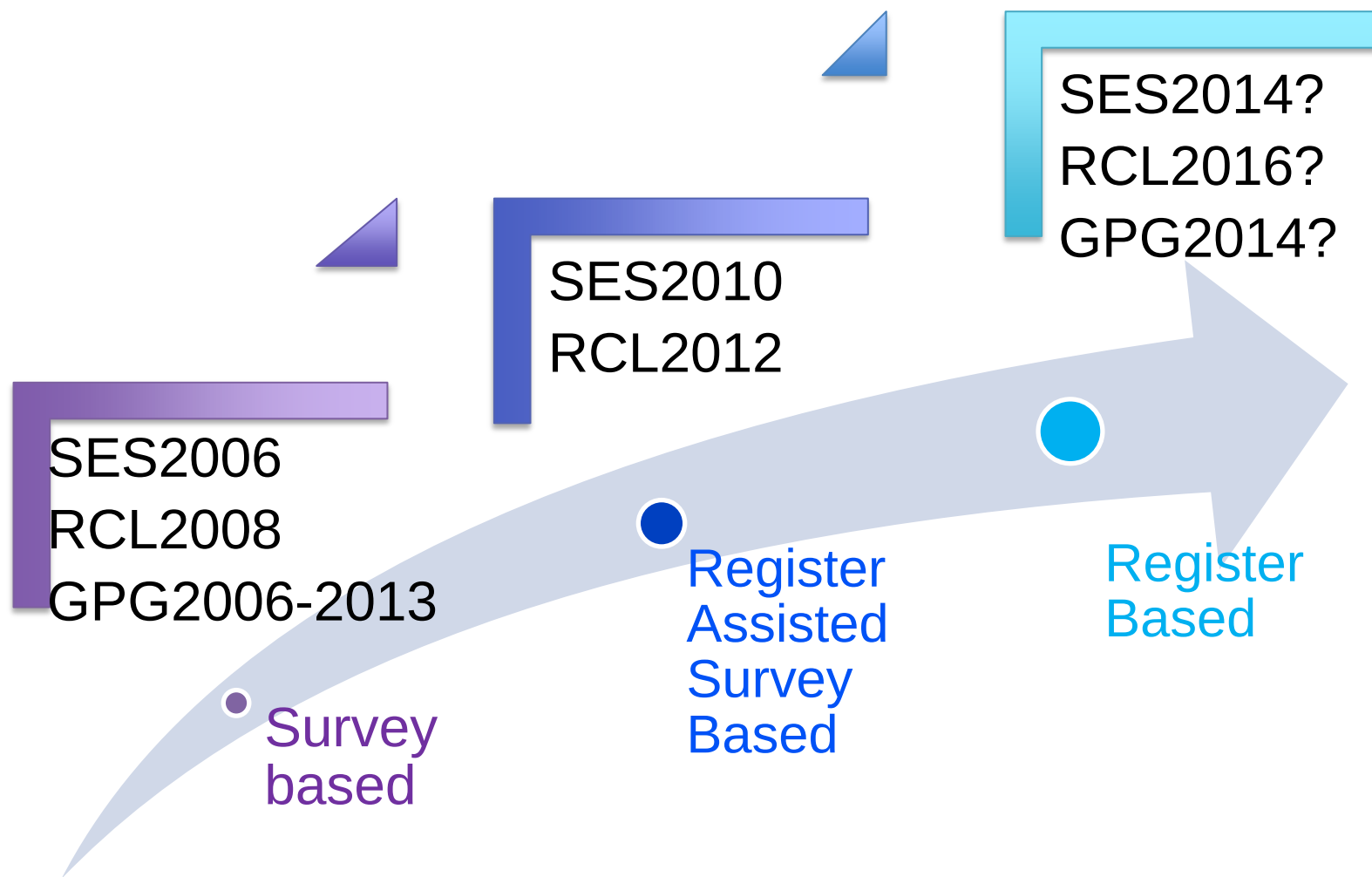


Comparabilità: validazione VELAGI VS LCS

Differenza percentuale nelle ore lavorate per dipendente VELAGI-LCS



Lo sviluppo delle rilevazioni del settore privato



RCL 2016: possibili evoluzioni

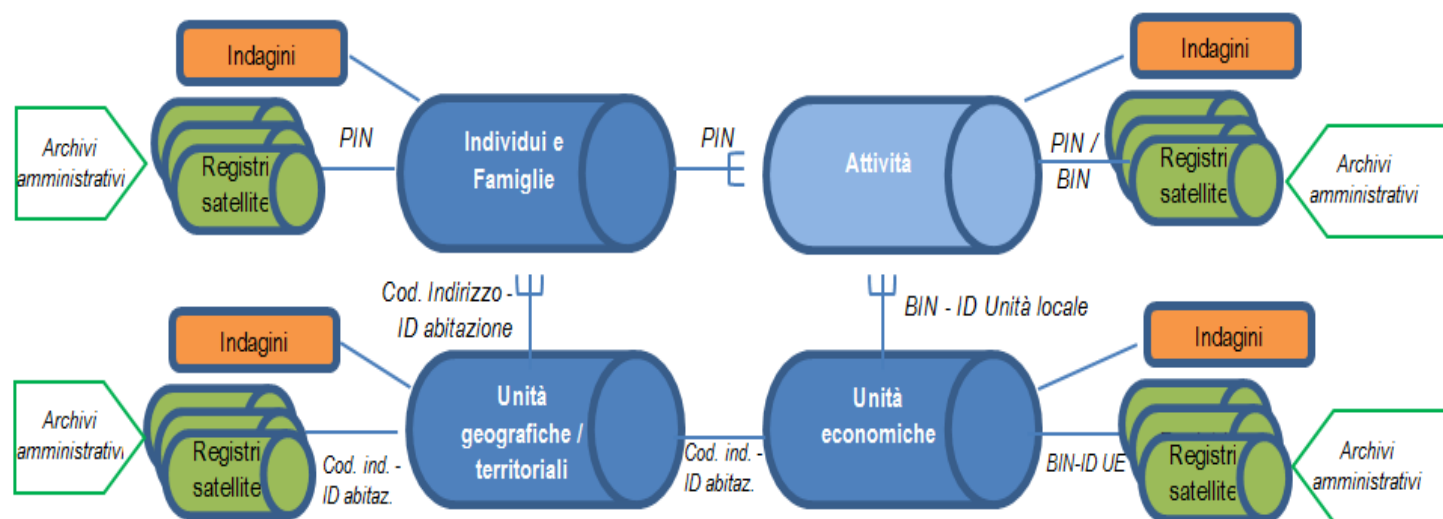
- Dipende anche da come si evolverà il sistema dei registri
 - Maggiore utilizzo dell'Uniemens (INPS)
- Si utilizzeranno altre variabili del Registro (contributi sociali, ecc.)
- Quasi sicuramente sarà necessario fare ancora una indagine diretta
- Si potrebbe ridurre la numerosità del campione
- Questionario precompilato o precompilato nascosto con dati amministrativi all'interno di alcune variabili

SES2014 e GPG annuale

- La SES precedente aveva un campione teorico di 20.000 imprese (8.000 rispondenti con 215 mila lavoratori)
- Studi e valutazioni ancora in corso (GdL)
- Opzione A: **Register-based survey**; rilevazione basata solo su combinazione di dati preesistenti
 - Registri +
 - Rilevazioni dirette sulle famiglie: FL, EUSILC, ecc
- Opzione B: **Register assisted survey**; ancora una indagine diretta ma con pochissime variabili e un piccolo campione
- GPG (Gender Pay Gap) stima annuale basata solo su combinazione di dati preesistenti **Register-based estimation**

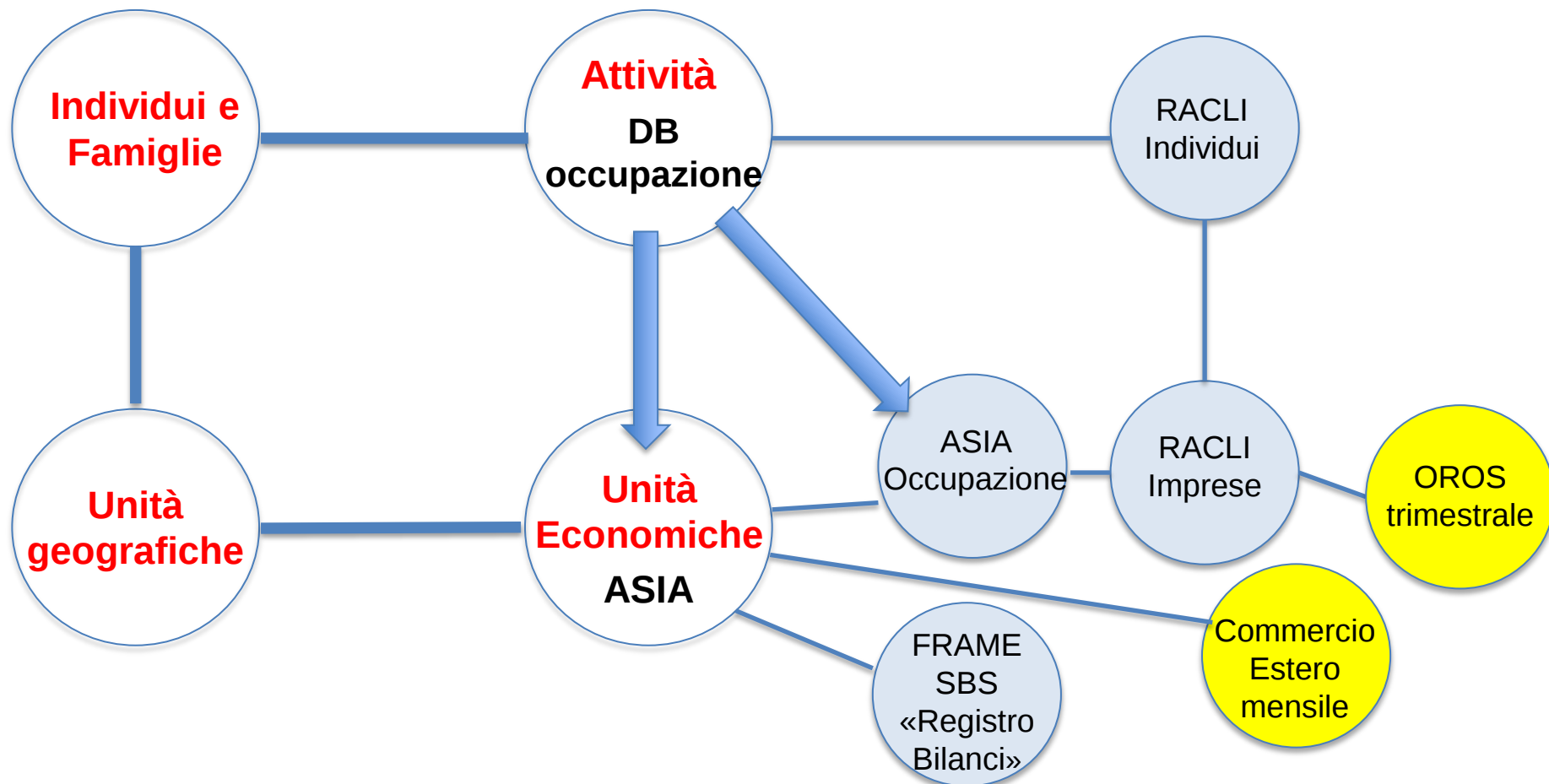
La proposta del modello svedese del sistema dei registri

- Bisogna solo vedere le cose in prospettiva
- Partiamo dal sistema dei 4 registri di base (**Unità economiche, Attività, Individui e Famiglie, Unità geografiche**) ispirati al modello svedese



Legenda: *legame uno a molti* —E; BIN=Identificativo impresa (Business ID number); PIN=Identificativo personale (Personal ID number); UE=Unità economica

Una parte del sistema dei Registri relativa alla popolazione di ASIA già esiste



Considerazioni finali

- La RCL2012 ha avuto ottimi risultati
 - uso dei registri (register-assisted survey)
 - ottimizzazione del contatto con le unità rispondenti
 - combinata una efficiente divisione del lavoro in alcune fasi (Call center e CATI) e una equilibrata collaborazione/condivisione in altre fasi (campionamento, controllo e correzione, calibrazione, ecc.)
- Nelle statistiche economiche e nel dominio «lavoro» già esistono le basi del nuovo sistema Register-Based proposto nella modernizzazione
- Nel nuovo sistema le indagini campionarie non scompaiono, si riducono nel numero e nella dimensione campionaria
- In alcuni domini le indagini campionarie possono diventare Register-based assisted survey con notevoli miglioramenti qualitativi e riduzione dell'onere sui rispondenti
- Il sistema dei Registri dovrebbe fornire strumenti per migliorare la valutazione della *non-response* e di *coverage*
- Dalla sample-survey isolata al sistema di registri c'è una fortissima interdipendenza che coinvolge tutti e accentua le responsabilità reciproche. Tre C di Statistics Norway: Coordinamento, Cooperazione, Comunicazione

Grazie
per l'attenzione

Fabio Rapiti
Fabio.rapiti@istat.it

Ciro Baldi
baldi@istat.it