

Una proposta per un catalogo degli algoritmi di riconoscimento di caso di malattia cronica da database amministrativi

Rosa Gini



Roma, 17 ottobre 2011

Indice

Introduzione

Algoritmi

- Patologie

- Fonti di dato: una criticità

- Algoritmi e TheMatrix

- Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

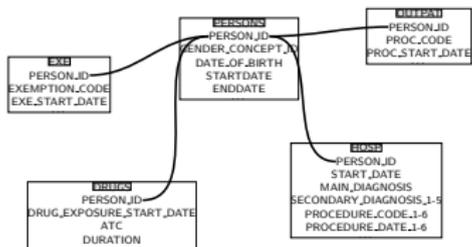
Il contesto: studi osservazionali da database amministrativi

- I database amministrativi sono entrati in modo esplosivo nella ricerca epidemiologica

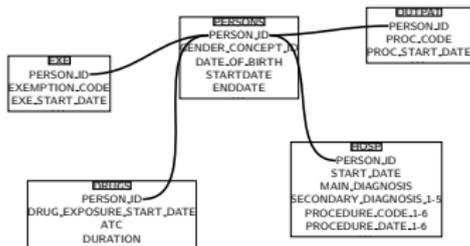
Il contesto: studi osservazionali da database amministrativi

- I database amministrativi sono entrati in modo esplosivo nella ricerca epidemiologica
- Gli studi osservazionali basati su questa fonte di dato meritano pertanto lo sviluppo di strumenti metodologici *ad hoc*

Raccolta dati di studi osservazionali da database amministrativi



Raccolta dati di studi osservazionali da database amministrativi

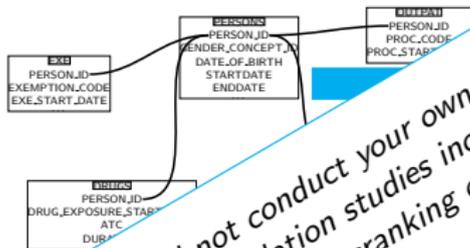


ANONYMOUS_ID
GENDER_CONCEPT_ID
BIRTHYEAR
EXPOSURE
COVARIATE₁
...
COVARIATE_N
RESPONSE

Raccolta dati di studi osservazionali da database amministrativi



Raccolta dati di studi osservazionali da database amministrativi



If you did not conduct your own validation study, be prepared to cite others who did. Validation studies increase cost and take time, but they may be the difference between cranking out analyses and sound epidemiologic research.

Hernán M. With great data comes great responsibility. *Epidemiology* 2011 May.

PERSON_ID
PROCEDURE_ID
EXPOSURE
COVARIATE₁
...
COVARIATE_N
RESPONSE

Proposta

- La letteratura sulla validazione di algoritmi per la definizione di caso da database amministrativi è ricca e in fase di espansione

Proposta

- La letteratura sulla validazione di algoritmi per la definizione di caso da database amministrativi è ricca e in fase di espansione
- Per rendere agilmente disponibile questa letteratura alla comunità epidemiologica italiana possiamo sviluppare uno strumento informativo che li cataloghi

Proposta

- La letteratura sulla validazione di algoritmi per la definizione di caso da database amministrativi è ricca e in fase di espansione
- Per rendere agilmente disponibile questa letteratura alla comunità epidemiologica italiana possiamo sviluppare uno strumento informativo che li cataloghi
- **Due obiettivi:**

Proposta

- La letteratura sulla validazione di algoritmi per la definizione di caso da database amministrativi è ricca e in fase di espansione
- Per rendere agilmente disponibile questa letteratura alla comunità epidemiologica italiana possiamo sviluppare uno strumento informativo che li cataloghi
- Due obiettivi:
 - **Pratico:** accedere a un database di algoritmi nel caso si intenda avviare uno studio, per scegliere quello con le caratteristiche più adeguate

Proposta

- La letteratura sulla validazione di algoritmi per la definizione di caso da database amministrativi è ricca e in fase di espansione
- Per rendere agilmente disponibile questa letteratura alla comunità epidemiologica italiana possiamo sviluppare uno strumento informativo che li cataloghi
- Due obiettivi:
 - Pratico: accedere a un database di algoritmi nel caso si intenda avviare uno studio, per scegliere quello con le caratteristiche più adeguate
 - Valutativo: stimare possibili bias di selezione/misclassificazione presenti in studi che utilizzano algoritmi simili

Schema della proposta di catalogo

- Due database tra loro agganciati (multi-a-molti)

Schema della proposta di catalogo

- Due database tra loro agganciati (multi-a-molti)
 - uno di articoli che riferiscono studi di validazione

Schema della proposta di catalogo

- Due database tra loro agganciati (multi-a-molti)
 - uno di articoli che riferiscono studi di validazione
 - uno di algoritmi

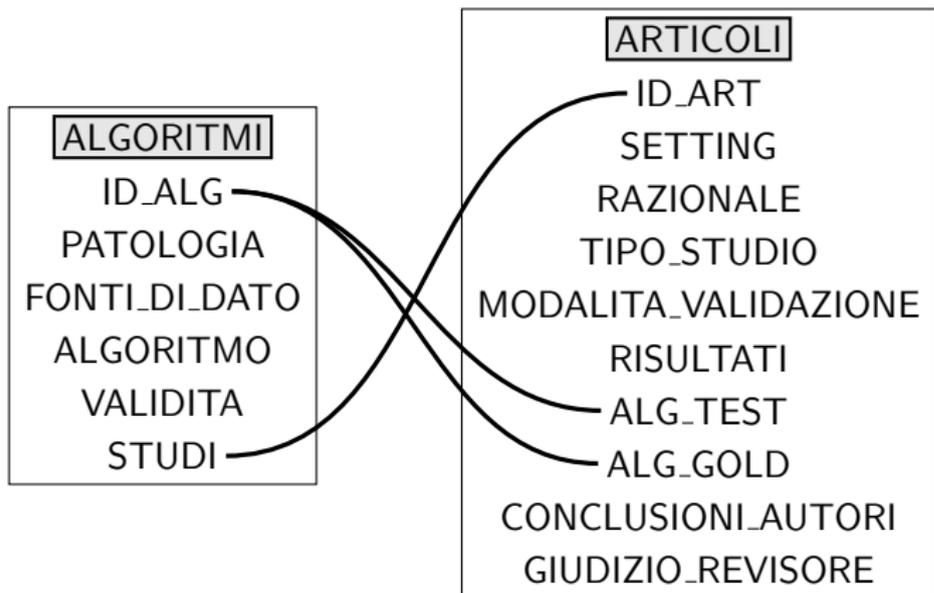
Schema della proposta di catalogo

- Due database tra loro agganciati (multi-a-molti)
 - uno di articoli che riferiscono studi di validazione
 - uno di algoritmi
- Se in un articolo si ritrovano uno o più algoritmi descritti esplicitamente, nel record relativo a quell'articolo sono contenuti gli identificativi degli algoritmi

Schema della proposta di catalogo

- Due database tra loro agganciati (multi-a-molti)
 - uno di articoli che riferiscono studi di validazione
 - uno di algoritmi
- Se in un articolo si ritrovano uno o più algoritmi descritti esplicitamente, nel record relativo a quell'articolo sono contenuti gli identificativi degli algoritmi
- Viceversa nel record di un algoritmo sono riferiti i record degli (eventuali) studi che ne esplorano la validità

Schema della proposta di catalogo



Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Almeno due campi

- Una descrizione grossolana (acronimo)

Almeno due campi

- Una descrizione grossolana (acronimo)
- Una descrizione più dettagliata

Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Criticità

Non tutti i database amministrativi sono uguali!!

Differenze tra database amministrativi citati nella letteratura internazionale

- In letteratura ci si riferisce a *administrative databases* o *claims databases*

Differenze tra database amministrativi citati nella letteratura internazionale

- In letteratura ci si riferisce a *administrative databases* o *claims databases*
- In realtà ogni organizzazione sanitaria ha la propria tipologia di dati amministrativi

Differenze tra database amministrativi citati nella letteratura internazionale

- In letteratura ci si riferisce a *administrative databases* o *claims databases*
- In realtà ogni organizzazione sanitaria ha la propria tipologia di dati amministrativi
- Per esempio in Canada, ma anche nelle assicurazioni statunitensi, vengono routinariamente registrate le visite mediche (generali o specialistiche) con tanto di codice diagnosi: questo rende quei database molto più sensibili (e meno specifici)

Differenze tra database amministrativi citati nella letteratura internazionale

- In letteratura ci si riferisce a *administrative databases* o *claims databases*
- In realtà ogni organizzazione sanitaria ha la propria tipologia di dati amministrativi
- Per esempio in Canada, ma anche nelle assicurazioni statunitensi, vengono routinariamente registrate le visite mediche (generali o specialistiche) con tanto di codice diagnosi: questo rende quei database molto più sensibili (e meno specifici)
- È necessario sviluppare una classificazione schematica che riassume le caratteristiche di ogni particolare database amministrativo, da utilizzare nei campi che si riferiscono alla fonte di dato

Bozza di classificazione delle fonti *elettroniche* di dati

- Amministrativi

Bozza di classificazione delle fonti *elettroniche* di dati

- Amministrativi
 - Raccolti *ex ante* (es. ricette farmaceutiche, ricette per visite specialistiche)

Bozza di classificazione delle fonti *elettroniche* di dati

- Amministrativi
 - Raccolti *ex ante* (es. ricette farmaceutiche, ricette per visite specialistiche)
 - Raccolti *ex post* con diagnosi (es. schede dimissioni ospedaliere; in Canada: schede visite specialistiche)

Bozza di classificazione delle fonti *elettroniche* di dati

- Amministrativi
 - Raccolti *ex ante* (es. ricette farmaceutiche, ricette per visite specialistiche)
 - Raccolti *ex post* con diagnosi (es. schede dimissioni ospedaliere; in Canada: schede visite specialistiche)
- Clinici (MMG, specialisti, ospedali, . . .)

Bozza di classificazione delle fonti *elettroniche* di dati

- Amministrativi
 - Raccolti *ex ante* (es. ricette farmaceutiche, ricette per visite specialistiche)
 - Raccolti *ex post* con diagnosi (es. schede dimissioni ospedaliere; in Canada: schede visite specialistiche)
- Clinici (MMG, specialisti, ospedali, . . .)
- Registri (morte, tumori, malformazioni, . . .)

Bozza di classificazione delle fonti *elettroniche* di dati

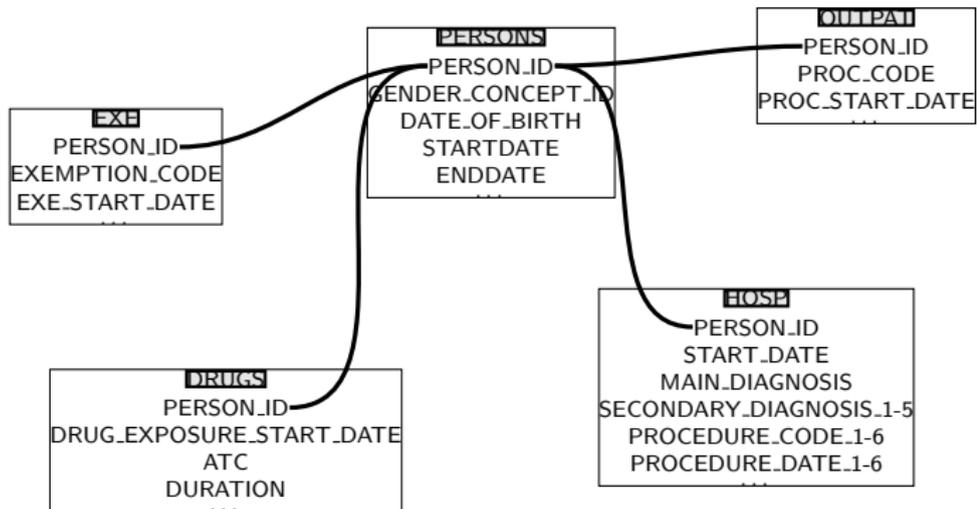
- Amministrativi
 - Raccolti *ex ante* (es. ricette farmaceutiche, ricette per visite specialistiche)
 - Raccolti *ex post* con diagnosi (es. schede dimissioni ospedaliere; in Canada: schede visite specialistiche)
- Clinici (MMG, specialisti, ospedali, . . .)
- Registri (morte, tumori, malformazioni, . . .)
- **Indagini campionarie (es. Multiscopo)**

Pochi algoritmi sono davvero applicabili

Applicabilità

Tra i campi relativi alla fonte di dato va indicato se l'algoritmo è applicabile ai dati amministrativi italiani

Dati amministrativi italiani (IAD)



Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Dimensioni dell'algoritmo

- codici di selezione

Dimensioni dell'algoritmo

- codici di selezione
- anni di followup

Dimensioni dell'algoritmo

- codici di selezione
- anni di followup
- rappresentazione logica (AND e OR)

Dimensioni dell'algoritmo

- codici di selezione
- anni di followup
- rappresentazione logica (AND e OR)
- eventuali raffinamenti (ad es: eliminare record riferiti a soggetti che hanno avuto negli ultimi 3 anni almeno un ricovero per...)

Dimensioni dell'algoritmo

- codici di selezione
- anni di followup
- rappresentazione logica (AND e OR)
- eventuali raffinamenti (ad es: eliminare record riferiti a soggetti che hanno avuto negli ultimi 3 anni almeno un ricovero per...)
- se l'algoritmo è applicabile ai dati amministrativi italiani:
script di TheMatrix

Cos'è TheMatrix

- TheMatrix è un software in Java che è in corso di sviluppo nel progetto MATRICE dell'Agenas

Cos'è TheMatrix

- TheMatrix è un software in Java che è in corso di sviluppo nel progetto MATRICE dell'Agenas
- Si tratta di un software da installare presso sistemi informativi regionali o aziendali, che

Cos'è TheMatrix

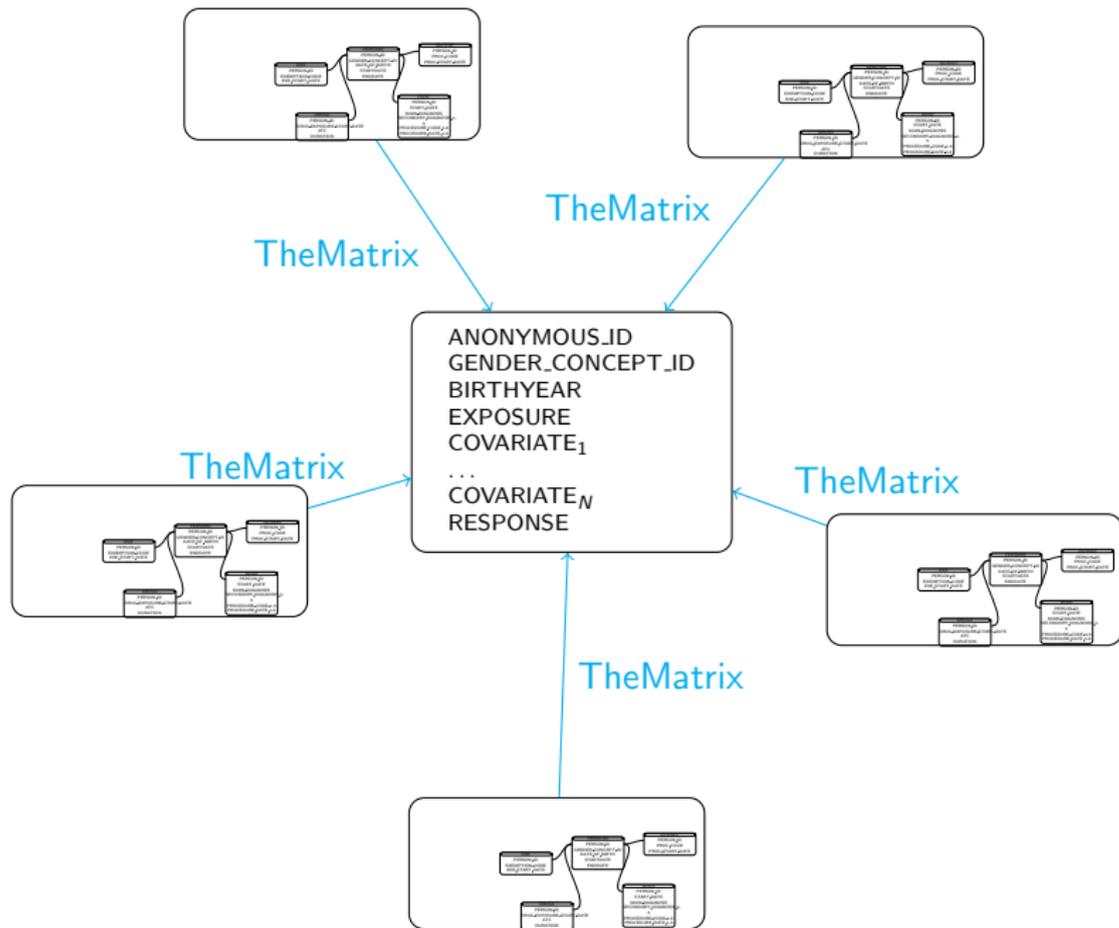
- TheMatrix è un software in Java che è in corso di sviluppo nel progetto MATRICE dell'Agenas
- Si tratta di un software da installare presso sistemi informativi regionali o aziendali, che
 - all'atto dell'installazione mappa i dati amministrativi in un tracciato record standard

Cos'è TheMatrix

- TheMatrix è un software in Java che è in corso di sviluppo nel progetto MATRICE dell'Agenas
- Si tratta di un software da installare presso sistemi informativi regionali o aziendali, che
 - all'atto dell'installazione mappa i dati amministrativi in un tracciato record standard
 - può lanciare degli script che eseguono estrazioni e data management anche sofisticati producendo dataset per studi epidemiologici

Cos'è TheMatrix

- TheMatrix è un software in Java che è in corso di sviluppo nel progetto MATRICE dell'Agenas
- Si tratta di un software da installare presso sistemi informativi regionali o aziendali, che
 - all'atto dell'installazione mappa i dati amministrativi in un tracciato record standard
 - può lanciare degli script che eseguono estrazioni e data management anche sofisticati producendo dataset per studi epidemiologici
- Quindi in particolare un algoritmo applicabile ai dati amministrativi italiani può essere rappresentato da uno script di TheMatrix: chi abbia installato il software TheMatrix e voglia applicare quell'algoritmo può scaricarlo e lanciarlo



Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Se disponibili

- informazioni sul potere predittivo positivo

Se disponibili

- informazioni sul potere predittivo positivo
- informazioni sulla sensibilità

Se disponibili

- informazioni sul potere predittivo positivo
- informazioni sulla sensibilità
- altro

Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)
- Riferimento alle schede degli algoritmi utilizzati per il test

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)
- Riferimento alle schede degli algoritmi utilizzati per il test
- Descrizione del gold standard: dati raccolti ad hoc oppure uso secondario di altra sorgente di dati (eventuale riferimento alle schede degli algoritmi corrispondenti)

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)
- Riferimento alle schede degli algoritmi utilizzati per il test
- Descrizione del gold standard: dati raccolti ad hoc oppure uso secondario di altra sorgente di dati (eventuale riferimento alle schede degli algoritmi corrispondenti)
- Modalità di validazione seguita (per dati individuali o aggregati)

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)
- Riferimento alle schede degli algoritmi utilizzati per il test
- Descrizione del gold standard: dati raccolti ad hoc oppure uso secondario di altra sorgente di dati (eventuale riferimento alle schede degli algoritmi corrispondenti)
- Modalità di validazione seguita (per dati individuali o aggregati)
- Risultati

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)
- Riferimento alle schede degli algoritmi utilizzati per il test
- Descrizione del gold standard: dati raccolti ad hoc oppure uso secondario di altra sorgente di dati (eventuale riferimento alle schede degli algoritmi corrispondenti)
- Modalità di validazione seguita (per dati individuali o aggregati)
- Risultati
- Conclusioni degli autori

Dimensioni da rilevare (adattato da progetto CCM)

- Razionale
- Setting
- Popolazione
- Tipo di studio (se applicabile: coorte storica, coorte prospettica, caso-controllo, casistica, trasversale)
- Riferimento alle schede degli algoritmi utilizzati per il test
- Descrizione del gold standard: dati raccolti ad hoc oppure uso secondario di altra sorgente di dati (eventuale riferimento alle schede degli algoritmi corrispondenti)
- Modalità di validazione seguita (per dati individuali o aggregati)
- Risultati
- Conclusioni degli autori
- Giudizio del revisore

Indice

Introduzione

Algoritmi

- Patologie

- Fonti di dato: una criticità

- Algoritmi e TheMatrix

- Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Collocazione

Richiesta

Epicentro?

Popolamento

- Inserire revisione bibliografica effettuata nel progetto CCM

Popolamento

- Inserire revisione bibliografica effettuata nel progetto CCM
- Inserire le revisioni bibliografiche in corso nel progetto
MATRICE

Aggiornamento

- Nomina di un piccolo comitato editoriale (1-2 persone)

Aggiornamento

- Nomina di un piccolo comitato editoriale (1-2 persone)
- I ricercatori che lo desiderano sottopongono le schede articolo o le schede algoritmo al comitato

Aggiornamento

- Nomina di un piccolo comitato editoriale (1-2 persone)
- I ricercatori che lo desiderano sottopongono le schede articolo o le schede algoritmo al comitato
- **Revisione della forma**

Aggiornamento

- Nomina di un piccolo comitato editoriale (1-2 persone)
- I ricercatori che lo desiderano sottopongono le schede articolo o le schede algoritmo al comitato
- Revisione della forma
- Pubblicazione a nome del ricercatore (che ha quindi responsabilità del contenuto)

Indice

Introduzione

Algoritmi

Patologie

Fonti di dato: una criticità

Algoritmi e TheMatrix

Stime di validità

Studi di validazione

Organizzazione del catalogo

Sintesi

Concetti

- Gli algoritmi in letteratura internazionale sono raramente applicabili ai dati italiani

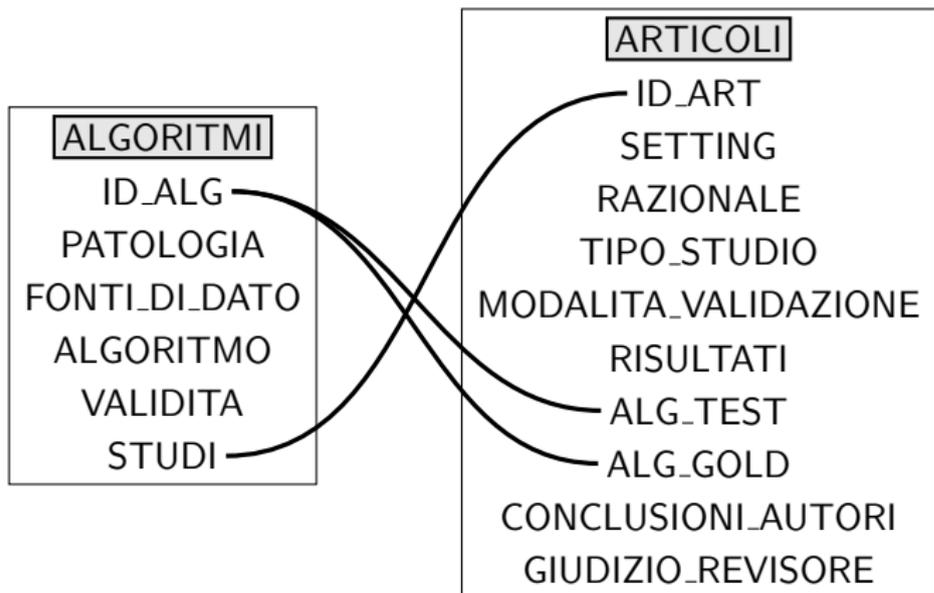
Concetti

- Gli algoritmi in letteratura internazionale sono raramente applicabili ai dati italiani
- È ciononostante utile costruire un catalogo di tali algoritmi e degli articoli che ne studiano la validità

Concetti

- Gli algoritmi in letteratura internazionale sono raramente applicabili ai dati italiani
- È ciononostante utile costruire un catalogo di tali algoritmi e degli articoli che ne studiano la validità
- **Nello stesso catalogo può essere sintetizzata la conoscenza relativa agli algoritmi applicabili ai dati amministrativi italiani**

Schema della proposta di catalogo

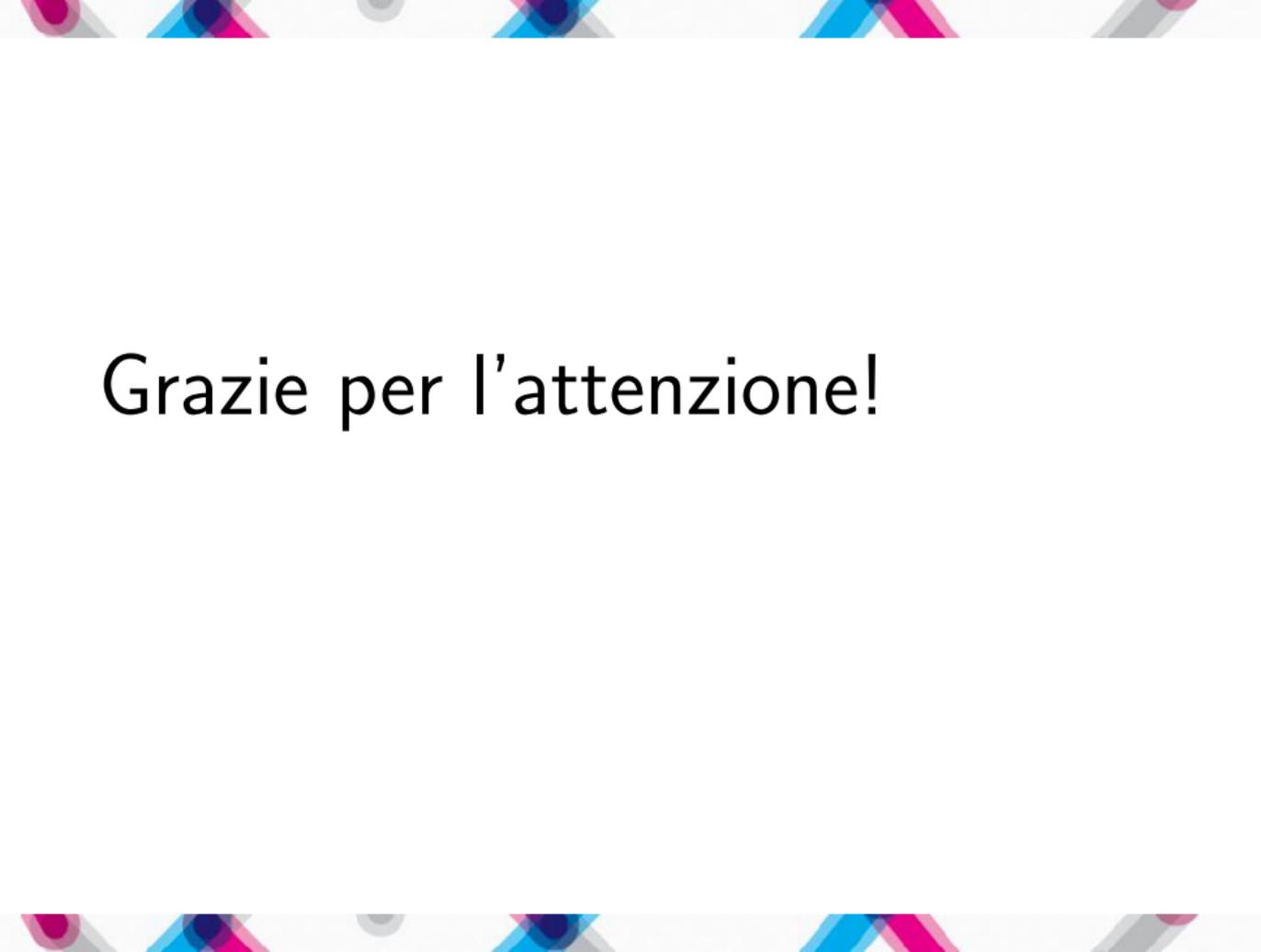




Ringraziamenti

Grazie a Cristina Canova e a Lorenzo Simonato





Grazie per l'attenzione!