



REGIONE AUTONOMA
FRIULI VENEZIA GIULIA

Codifica automatica RT-FVG

XVIII Corso di aggiornamento per operatori dei Registri
Tumori

dal 3 al 5 ottobre 2018 - Monopoli (BA)

1

- Il **RT FVG** si colloca ora **all'interno di un più articolato progetto di sviluppo di una piattaforma tecnologica che consente il trattamento**, entro i limiti previsti dalle normative, dei **dati** provenienti dagli applicativi **gestionali** in uso in Regione FVG e **gestiti da Insiel** (società in House della Regione)
- In quest'ottica il RT-FVG può essere visto anch'esso come **uno degli strumenti** all'interno di una cassetta degli attrezzi **per la valutazione delle politiche sanitarie in Regione FVG**

RT-FVG #Collocazione

2

- **Stretto coordinamento fra il servizio centrale regionale e le strutture collocate nelle aziende** che non agiscono solo come operatori periferici ma anche, quando accreditate, **come strutture centrali regionali** (*ad esempio il RT FVG è incardinato nell'IRCCS CRO di Aviano ma opera come struttura regionale, conseguentemente il suo direttore D. Serraino agisce come direttore regionale del RT-FVG*).
- Accreditamento di una **rete regionale professionale** (60 utenti), che si estende dal centro e si dirama nelle strutture epidemiologiche del sistema, che **ha accesso a tutta la base dati regionale e collabora al suo sviluppo** (*si costituisce così una comunità professionale attorno al data base regionale*).

Modello partecipativo

- Il «**Servizio epidemiologico e dei flussi informativi**», in qualità di soggetto interessato e portatore di interesse, si occupa:
 - della politica di **sviluppo** e **presidio** della piattaforma,
 - della **ottimizzazione** ed **economicità dell'intera filiera** di produzione del dato.

Gestore piattaforma

- Il RT-FVG si **sviluppa** intorno ad un **algoritmo** utilizzato inizialmente dalla **Regione Veneto** e successivamente **adattato** e arricchito da **Insiel nel 1998**.
- Insiel ha **scelto** la tecnologia **SAS** per:
 - la **realizzazione** del sistema;
 - lo sviluppo dell'**algoritmo** che genera l'incidenza;
 - L'**interfaccia** risoluzione manuale dei casi
 - Le **attività di analisi** a fini statistici e di programmazione per l'intero sistema sanitario regionale.

RT-FVG #Soluzione 1998

L'attuale RT-FVG è un ibrido, una parte della codifica è basata su tecniche:

1. automatiche [Algoritmi espliciti]
2. manuali «su dati classificati» [Machine learning]
3. manuali «su testi» [Contextual analysis]

In **futuro** anche le componenti manuali *potrebbero* essere ulteriormente automatizzate.

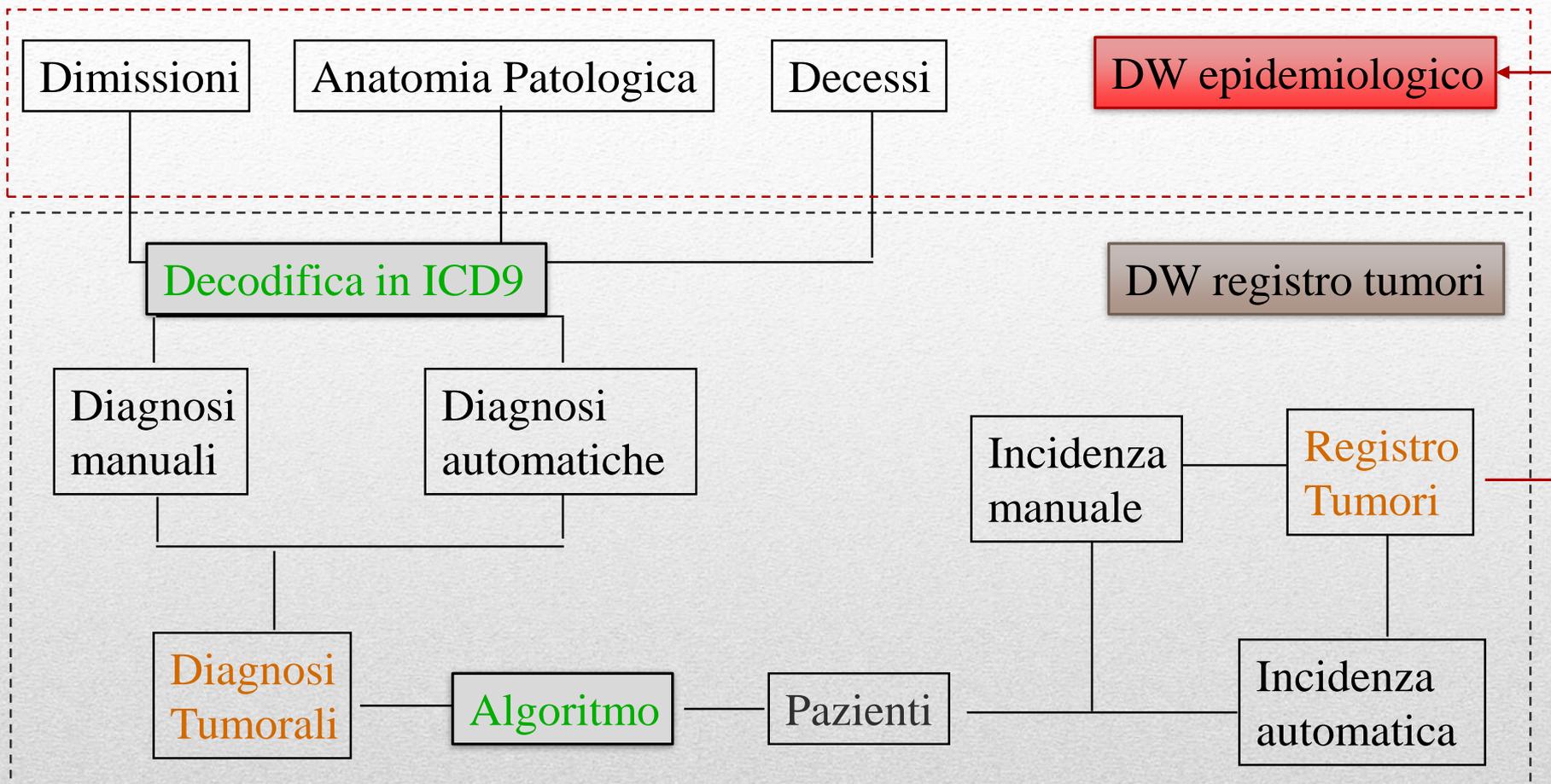
Codifica automatica ?

In realtà esistono già in Regione registri a codifica automatica, basati solo su algoritmi espliciti, per i quali viene calcolata l'incidenza e la prevalenza.

La Regione ha definito con apposita legge l'**istituzione dei registri di patologia** che già da 1998 fanno uso sistematico degli applicativi gestionali in uso nelle aziende del SSR. Attualmente sono in attività in Regione FVG i seguenti registri:

- Tumori
- Eventi acuti cardiovascolari
- **Ictus e tia**
- **Diabete**
- **Trattamenti sostitutivi renali (dialisi e trapianti)**
- Resistenze batteriche
- Malattie professionali
- Infortuni sul lavoro
- Incidenti stradali

I registri di patologia



RT-FVG #workflow 1998

**RT - FVG
Incidence 2006 - 2007**

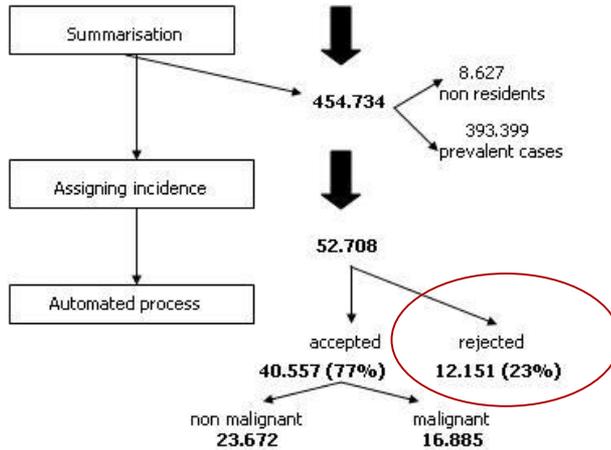
Hospital Discharges (1985-2007)	5.766.653
Pathology records (1983-2007)	4.147.751
Death certificates (1989-2007)	282.092
Cancer register (1995-2005)	112.501

Total 10.308.997

Selecting codes 140-239

Hospital discharges	887.148
Pathology records	618.250
Death certificates	87.090
Cancer register	112.487

Total 1.704.975



1998

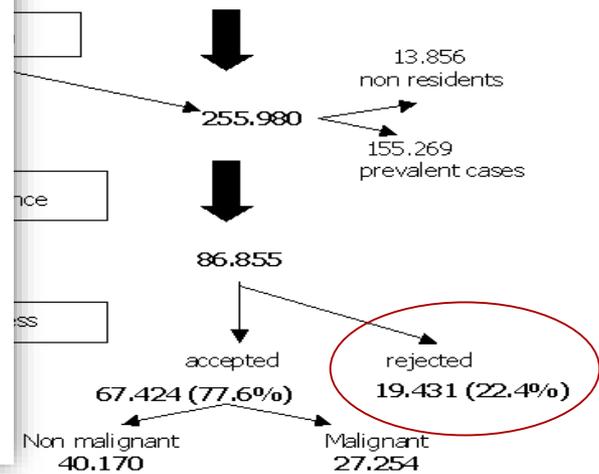
Hospital Discharges (1985-2007)	3.701.943
Pathology records (1983-2007)	1.920.775
Death certificates (1989-2007)	151.349
Cancer register (1995-2005)	112.501

5.774.067

Selecting codes 140-239

Hospital discharges	495.663
Pathology records	273.308
Death certificates	45.419
Cancer register	112.487

814.390



La maggior parte del processo è automatizzato

Negli ultimi anni il registro ha rallentato ...

- SW monolitico non rinnovato da 20 anni?
- Processo manuale frammentato in origine (SW locale)
- Dispersione su altri progetti?
 - Repository epidemiologico
 - Altri registri di patologia
 - Programmi di screening
-

a fine 2018 si conclude il 2013

RT-FVG #Limiti attuali

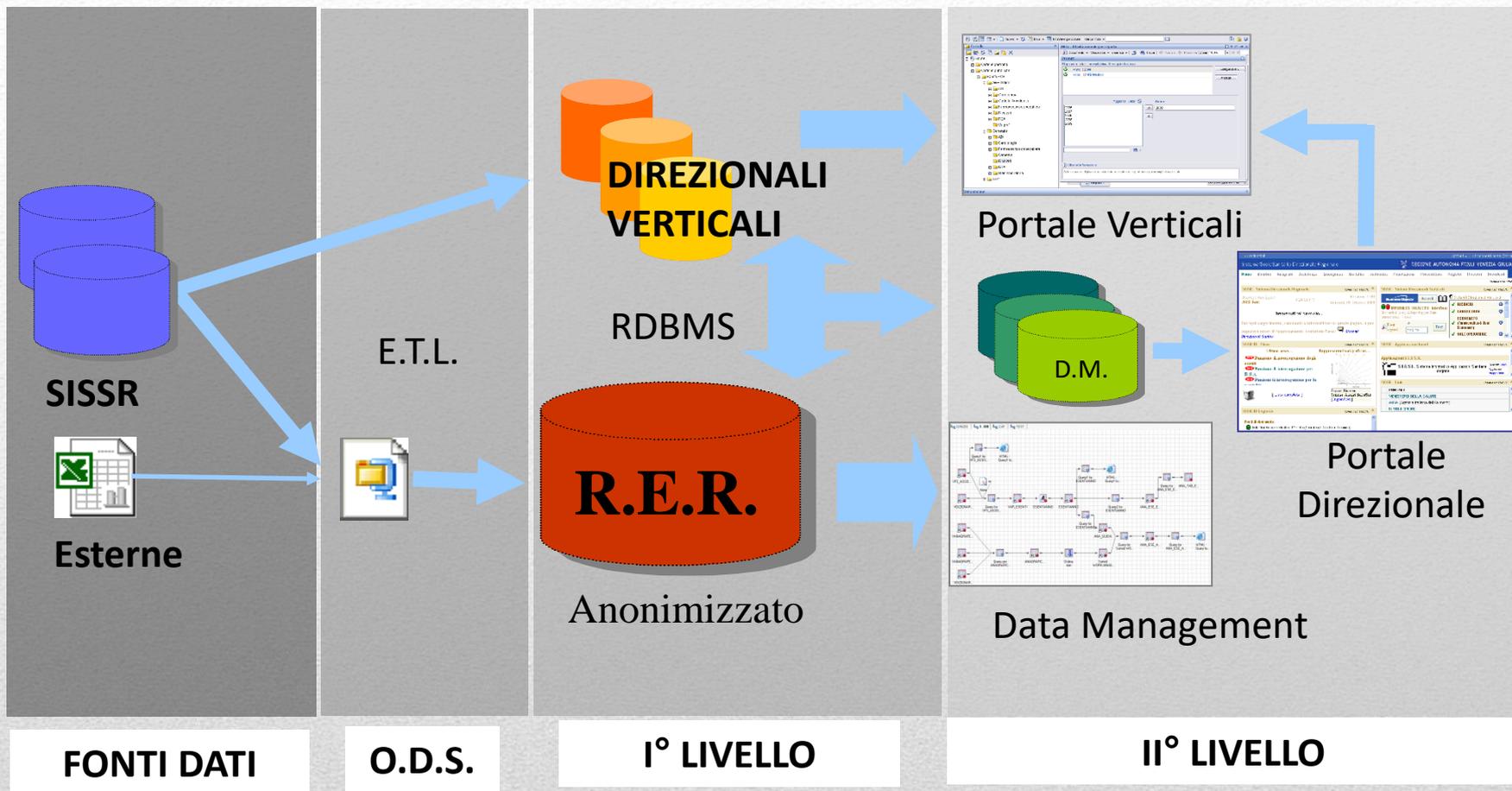
10

Risultato, il problema è del

La versione SW 2018 (in rilascio) comprende alcune novità rispetto il 1998 a livello di:

- Fonti dati strutturate
- Fonti dati non strutturate
- Revisione del processo metodologico
 - Intervento sul RER
 - Intervento sul RT
- Revisione della soluzione tecnologica
 - Data base
 - Interfaccia

RT-FVG #Novità V.2018



Piattaforma direzionale

Indice

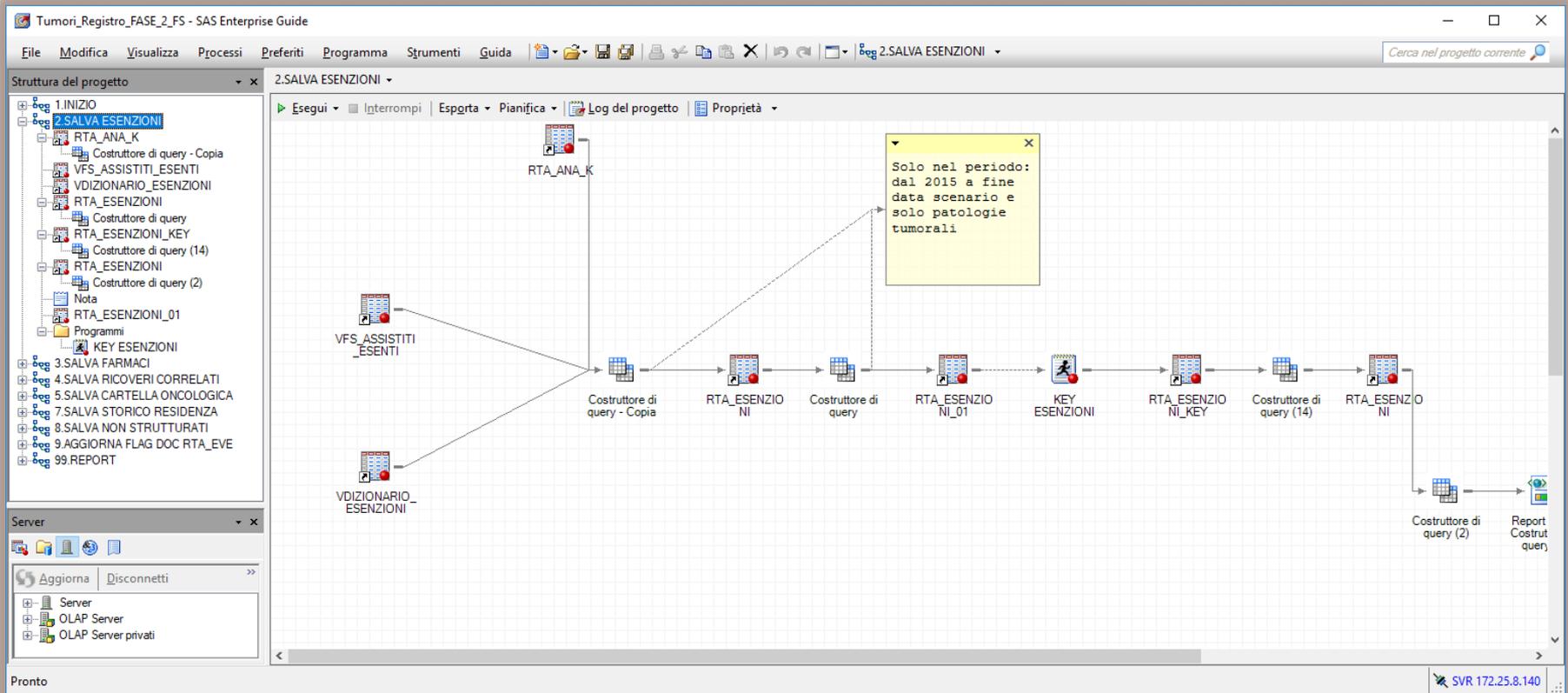
1. Relazioni tra le tabelle	3
1.1. ANAGRAFE	4
1.2. ANATOMIA PATOLOGICA DIAGNOSI	5
1.3. ANATOMIA PATOLOGICA PRESTAZIONI	6
1.4. ASSISTENZA DOMICILIARE	7
1.5. ASSISTENZA DOMICILIARE MEDICO (ADM)	8
1.6. CARTELLA ONCOLOGICA – GESTIONE CHEMIOTERAPIA	9
1.7. CARTELLA ONCOLOGICA – GESTIONE CHEMIOTERAPIA (PROTOCOLLI COMPOSTI)	9
1.8. CARTELLA ONCOLOGICA – CARTELLA	10
1.9. FARMACEUTICA CONVENZIONATA	11
1.10. PRESTAZIONI AMBULATORIALI & CUP	12
1.11. PRONTO SOCCORSO	13
1.12. RICOVERI OSPEDALIERI	14
1.13. RESIDENZE SANITARIE ASSISTENZIALI (RSA)	15
1.14. VACCINAZIONI	16
1.15. POPOLAZIONE	17
1.16. NASCITA	18
1.17. MORTALITÀ	19
1.18. CARDIOLOGIA	20
1.19. SALE OPERATORIE	21
1.20. FARMACEUTICA OSPEDALIERA	22
1.21. FARMACEUTICA DIRETTA	23
1.22. LABORATORIO ANALISI	24
1.23. CARTELLA SOCIALE	25

**Nel 1998 esisteva solo un prototipo di DWH con:
anagrafe, DCO, SDO, AP, ad hoc per RT**

+ 5 Registri di patologia e 3 programmi di screening

Manuale utente - R.E.R.

13



RT-FVG #Fonti strutturate

14

Nota: fonti censite in RER

Nella piattaforma informativa epidemiologica:

- Anatomia patologica (referti dal 1982)
- Dimissione ospedaliera (lettere dal 2008)
- Cartella oncologica (testi dal 2005)
- Prestazioni cliniche (referti dal 2000)

RT-FVG #FONTI NON strutturate anonimizzate

15

Nota: fonti da censire in RER

Estrazione dati

Tumori_Registro_FASE_1
Tumori_Registro_FASE_2
Tumori_Registro_FASE_2_TESTI
Tumori_Registro_FASE_2_FS

Algoritmi automatici

Tumori_Registro_FASE_3A
Tumori_Registro_FASE_3B
Tumori_Registro_FASE_4
Tumori_Registro_FASE_5

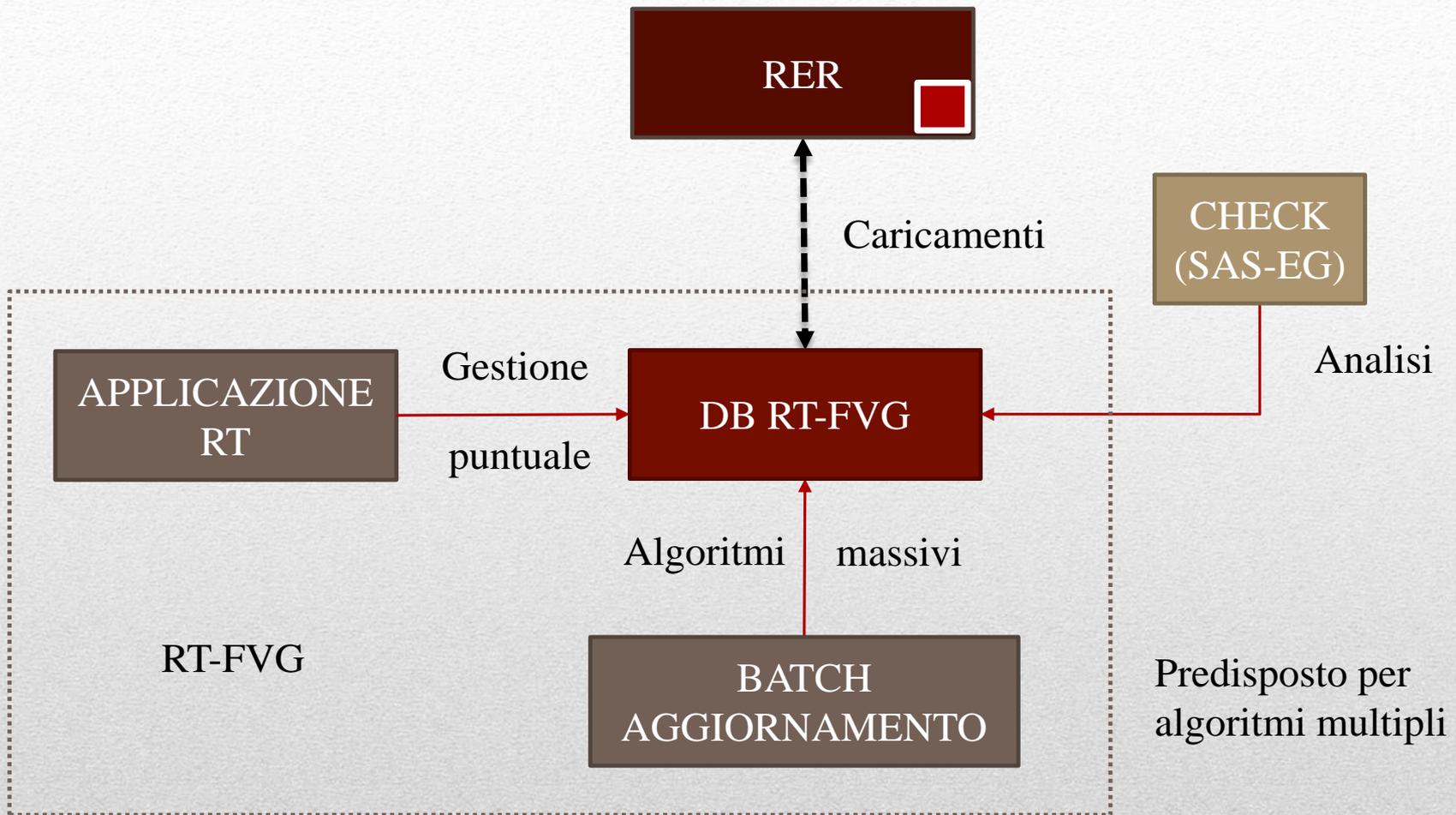
Gruppi di trattamento

Tumori_Registro_FASE_6

Aggiornamento DB

Tumori_Registro_FASE_9_DB

RT-FVG #Fasi



RT-FVG #Multi Processo

17

Nota: La fase degli algoritmi è distinta

PROCESSO	SOGGETTI	EVENTI	DIAGNOSI
00.RC_CERTIFICATI DECESSO - EVENTI INIZIALI	129.600	129.601	
00.RC_DIMISSIONI OSPEDALIERE - EVENTI INIZIALI	450.280	1.248.083	
00.RC_REFERTI_ANATOMIA - EVENTI INIZIALI	526.139	1.056.367	
10.RC_TUTTE LE FONTI - BASE DATI INIZIALE	658.408	2.434.051	
11.RC_TUTTE LE FONTI - BASE DATI INIZIALE+REGISTRO	658.408	2.588.205	3.165.064
12.RC_TUTTE LE FONTI - ESCLUSI DIAGNOSI > DATA FINE INCIDENZA	576.156	2.225.212	2.726.548

es. 2010-2013

STATO	CLASSIFICAZIONE SOGGETTI	SOGGETTI
0.DA TRATTARE	CAVALIERE - DA TRATTARE	11.839
0.DA TRATTARE	SCENARIO ATTUALE - DA TRATTARE	48.053
3.AUTOMATICO	CAVALIERE - SOSPETTO PREVALENTE PER TUTTE DIAGNOSI	9.153
3.AUTOMATICO	SCENARI PRECEDENTI - VECCHIO CASO	114.434
5.NON CASO	CAVALIERE - NON RESIDENTE NELLO SCENARIO ATTUALE	61
5.NON CASO	CAVALIERE - SOLO DIAGNOSI NON MALIGNI NELLO SCENARIO ATTUALE	3.829
5.NON CASO	PREVALENTE INIZIALE ANTE REGISTRO	115.156
5.NON CASO	SCENARI PRECEDENTI - IN SOSPESO	5
5.NON CASO	SCENARI PRECEDENTI - NON CASO	24.471
5.NON CASO	SCENARI PRECEDENTI - NON DEFINITO	279
5.NON CASO	SCENARI PRECEDENTI - NON MALIGNO	127.937
5.NON CASO	SCENARI PRECEDENTI - NON RESIDENTE	44.552
5.NON CASO	SCENARI PRECEDENTI - SOSPETTO PREVALENTE ANTE REGISTRO PER TUTTE LE DIAGNOSI	4.760
5.NON CASO	SCENARIO ATTUALE - NON MALIGNO	56.329
5.NON CASO	SCENARIO ATTUALE - NON RESIDENTE	15.298
		576.156

RT-FVG # Estrazione dati

18

Primo cambiamento: classificazione dei dati extra risoluzione

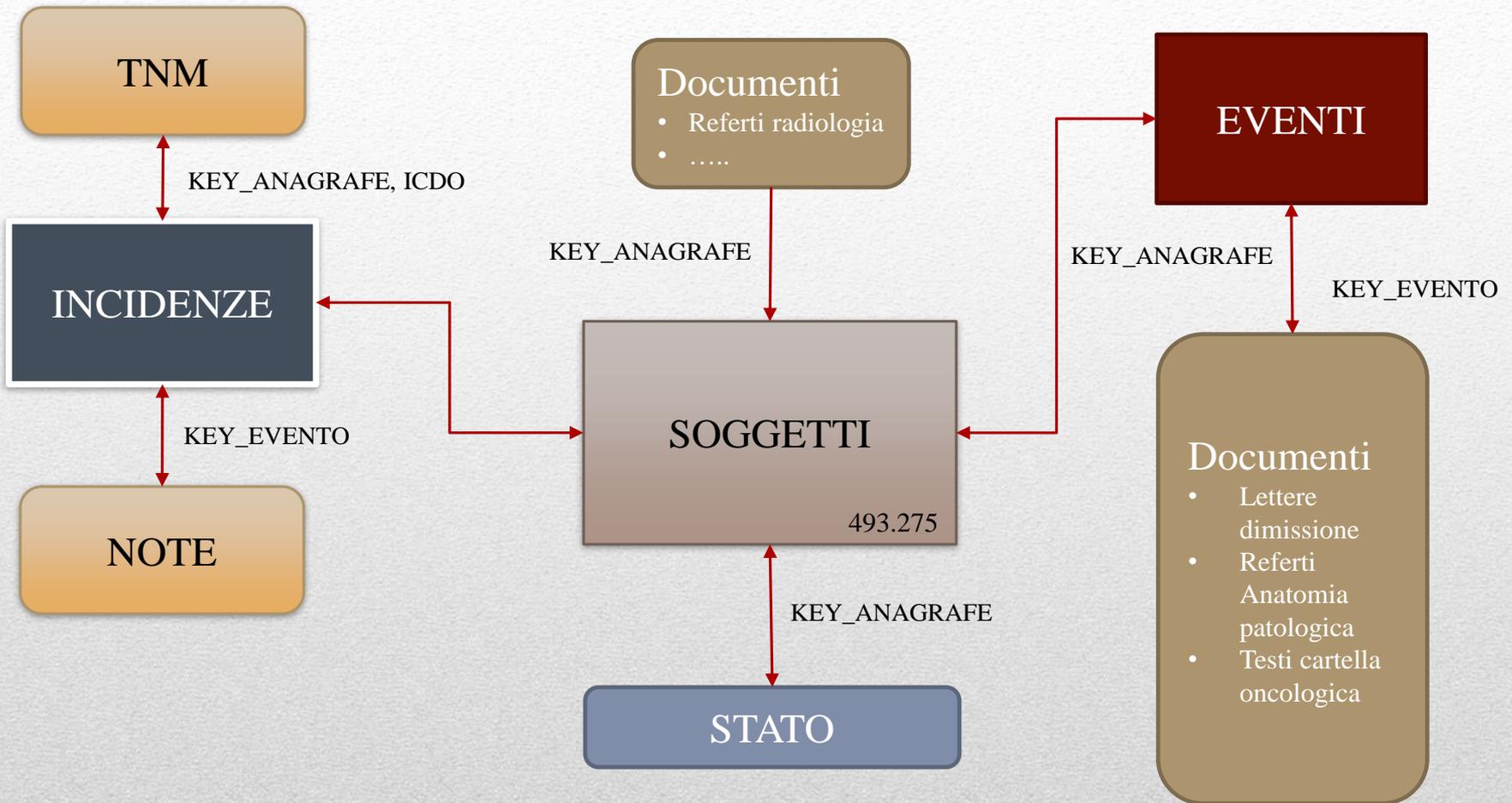
CLASSIFICAZIONE SOGGETTO	STATO	SOGGETTI
CAVALIERE - DA TRATTARE	0.DA TRATTARE	10.268
CAVALIERE - DA TRATTARE	2.PREVALENTE	1.193
CAVALIERE - DA TRATTARE	3.AUTOMATICO	251
CAVALIERE - DA TRATTARE	4.IN SOSPEO	1
CAVALIERE - DA TRATTARE	7.NON RESIDENTE	126
SCENARIO ATTUALE - DA TRATTARE	0.DA TRATTARE	20.773
SCENARIO ATTUALE - DA TRATTARE	2.PREVALENTE	1.918
SCENARIO ATTUALE - DA TRATTARE	3.AUTOMATICO	18.232
SCENARIO ATTUALE - DA TRATTARE	4.IN SOSPEO	2.713
SCENARIO ATTUALE - DA TRATTARE	6.NON DEFINITO	3.826
SCENARIO ATTUALE - DA TRATTARE	7.NON RESIDENTE	591
	TOTALE	59.892
	MANUALI (56,4%)	33.755

Applicazione
di UN
algoritmo

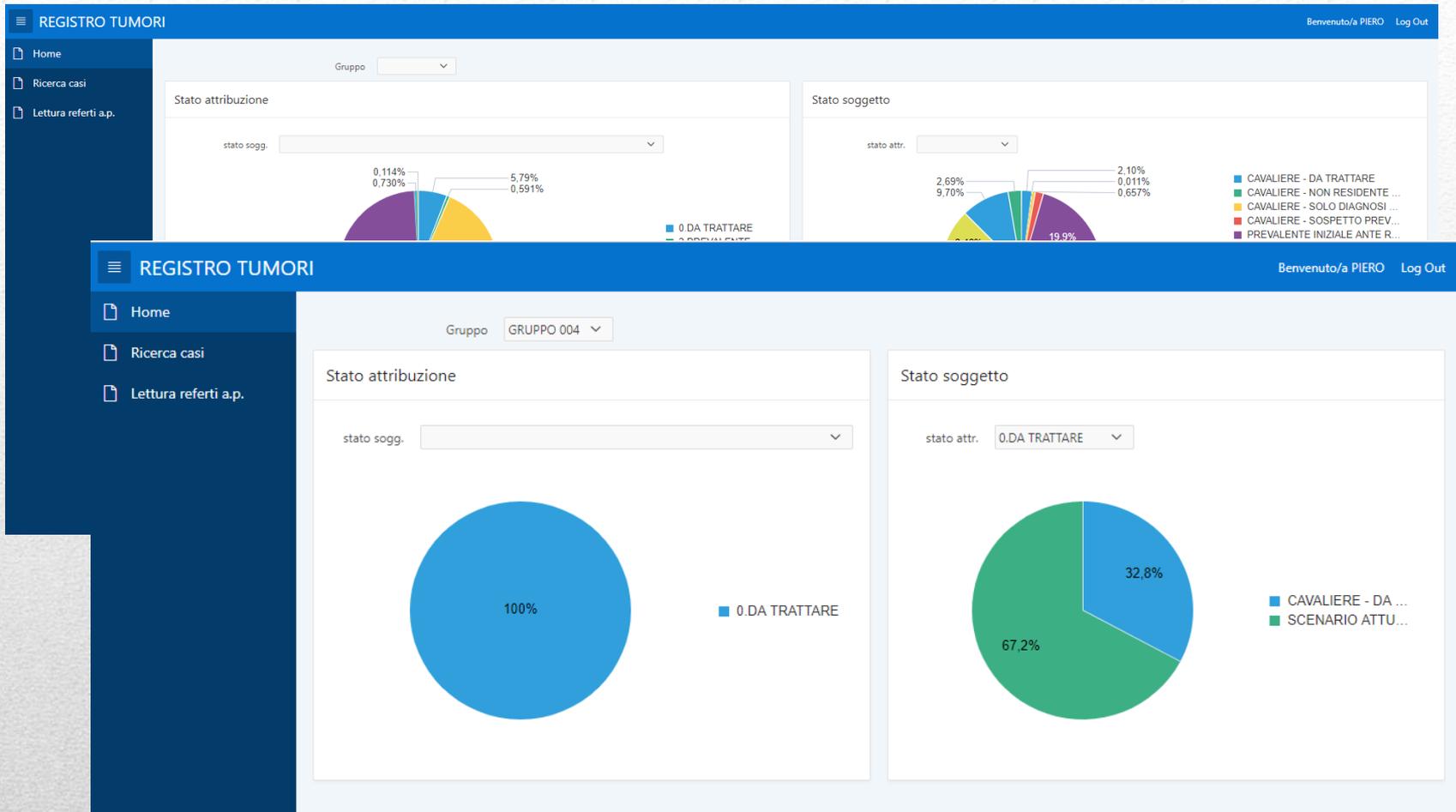
RT-FVG # es. 2010-2013

19

Secondo cambiamento: la risoluzione diventa un processo autonomo e distinto



RT-FVG # DB logico



RT-FVG #Interfaccia

21

Valutazione dello stato di avanzamento della risoluzione manuale

- Home
- Ricerca casi
- Letture referti a.p.

Assistito

Id ana.: 12509 - sc: M

Stato attribuzione

3.AUTOMATICO

Modifica stato

Esci

Elenco eventi e/o incidenze

Finestre pop-up Una

agg. /edit.	Tipo	Data	Chiave	Tipo doc. associato	Id documento	Comune di residenza evento	ICD9	Topografia ICDO	Morfologia ICDO	DIA_T	DIA_M	Codifica
+	AP	10/09/2012	2012120130108810091860201	RAP	102237	032006	2113	T1531	M82110	T59440	M82110	SNOMED-3
+	AP	10/09/2012	2012120130108810091860201	RAP	102237	032006	2113	T1532	M82110	T59460	M82110	SNOMED-3
+	AP	10/05/2012	201212008129CC10051860201	RAP	97468	032006	226	T1939	M83300	TB6100	M83300	SNOMED-3
+	CO	25/01/2012	201200232501SC18814030401	RCO	49436	032006	1533	T1533	M81403	C18.7	M81403	ICDO-3
+	AP	04/01/2012	2012120001348804011860201	RAP	87844	032006	1533	T1533	M81403	T59470	M81403	SNOMED-3
+	RO	03/01/2012	201200000211R603011101018	LDI	61539	032006	1541	T1548	M99903			ICD9_CM_2007
+	AP	10/11/2011	2011110162618810111860201	RAP	83320	032006	1533	T1533	M81403	T59470	M81403	SNOMED-3
+	AP	10/11/2011	2011110162618810111860201	RAP	83320	032006	2113	T1536	M82630	T59420	M82630	SNOMED-3
>	CI	10/11/2011	615			032006	153	T1533	M81403			

1 - 9

Chiave assistito ↑	Sesso	Data di nascita	Stato	Stato attribuzione	Stato soggetto	Num.diagnosi da anatomia patologica	Num.diagnosi da ricovero	Num.diagnosi da scheda di morte	Numero diagnosi totale	Lettere dimissione	Referti anatomia pat.	Referti cartella oncologica	Gruppo
449	F	05/12/1931	ATTIVO	4.IN SOSPESO	SCENARIO ATTUALE - DA TRATTARE	1	1	0	2	1	2	1	GRUPPO 004
831	M	24/04/1933	ATTIVO	3.AUTOMATICO	CAVALIERE - DA TRATTARE	2	1	0	2	1	1	0	GRUPPO 004
3187	M	12/05/1951	ATTIVO	3.AUTOMATICO	SCENARIO ATTUALE - DA TRATTARE	2	0	0	2	0	1	0	GRUPPO 004
5982	F	30/11/1945	ATTIVO	3.AUTOMATICO	SCENARIO ATTUALE - DA TRATTARE	2	1	0	2	0	2	1	GRUPPO 004
12509	M	28/02/1943	ATTIVO	3.AUTOMATICO	SCENARIO ATTUALE - DA TRATTARE	3	1	0	4	1	4	1	GRUPPO 004

1 - 5 di 100

RT-FVG #Ricerca casi

22

DEFINIZIONE di apprendimento automatico: replicare il processo di apprendimento mirato all'esecuzione di un compito tramite la realizzazione di algoritmi generici il cui scopo è quello di creare dei modelli per i dati trattati. Grazie a questi modelli è possibile superare le limitazioni derivanti da programmi o algoritmi espliciti (statici). Gli algoritmi di Machine Learning **migliorano le loro prestazioni in modo “adattivo”** mano a mano che gli “esempi” da cui apprendere aumentano (es. sistemi di raccomandazione in rete).

Il RT è candidabile?

Attualmente :

- È generato a partire da milioni di record di dati
- Contiene circa 500 mila soggetti esaminati con 100 mila circa incidenze in 15 anni

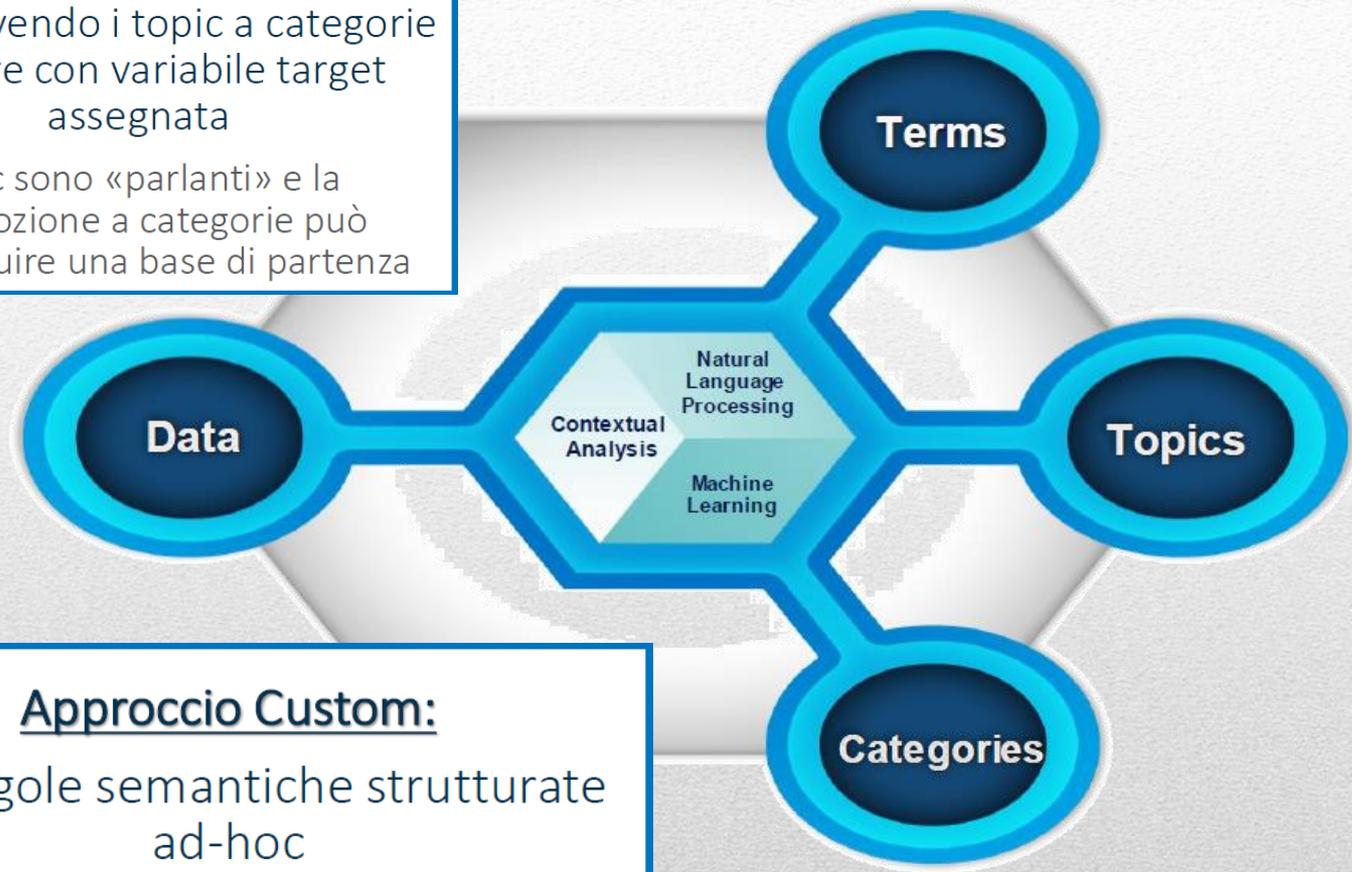
**RTV-FVG #Machine
learning**



Approccio Machine Learning:

partendo da una [start-list](#) e promuovendo i topic a categorie oppure con variabile target assegnata

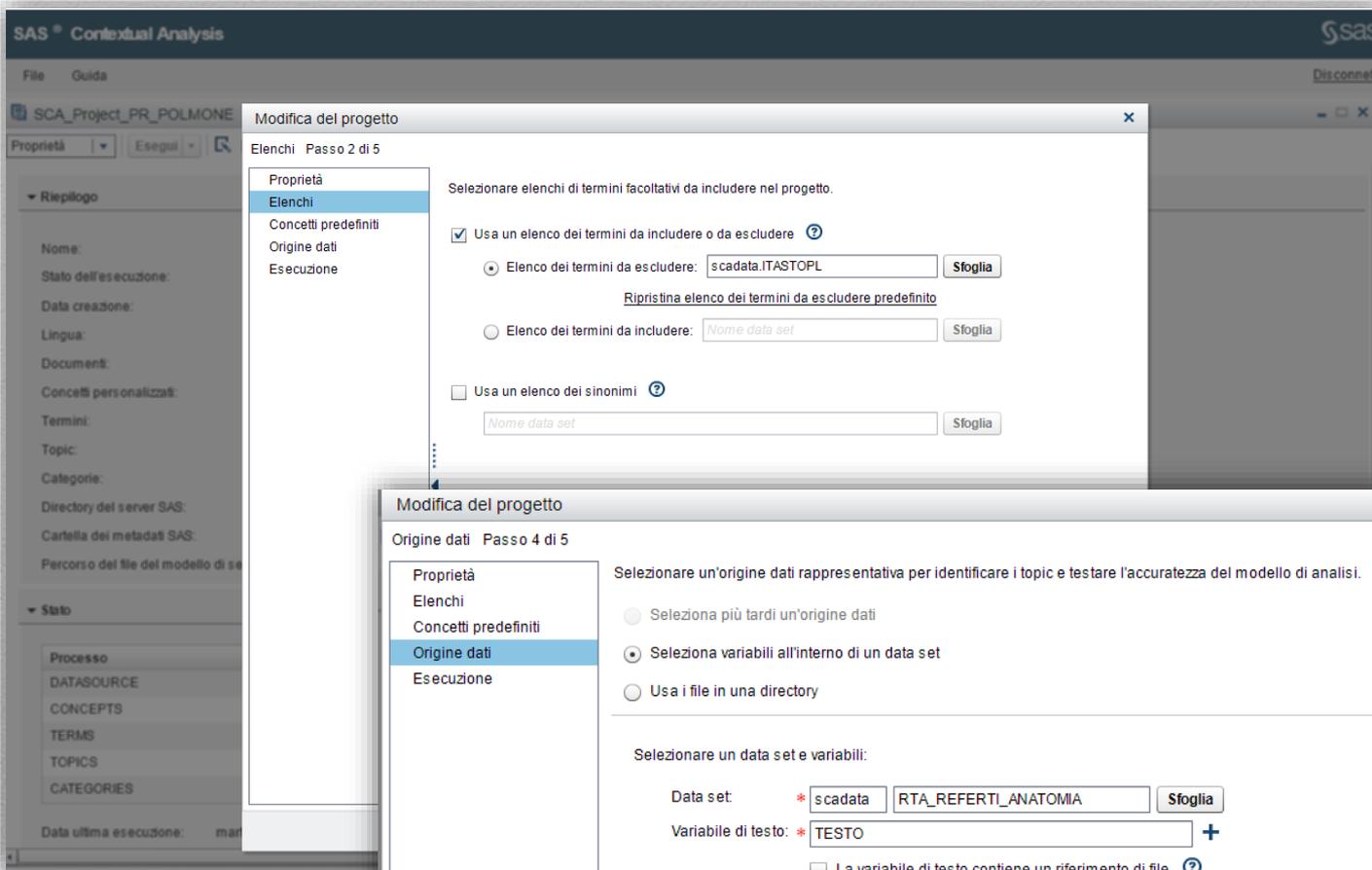
- I topic sono «parlanti» e la promozione a categorie può costituire una base di partenza



Approccio Custom:

con regole semantiche strutturate ad-hoc

RTV-FVG #SAS Contextual



Input

Termini mantenuti

Cerca

Termini e sinonimi	Numero di documenti	Concetto
▶ citologico	1079	
▶ differenziato	1045	
▶ dare	1012	
▶ cm	1007	
□ non	907	
▶ destro	901	
▶ quadro	892	
▶ neoplasia	879	
▶ piccolo	845	
▼ adenocarcinoma	841	
□ adenocarcinoma	835	
□ adenocarcinoma	1	
▶ sinistro	783	
▶ bc	727	
▶ superiore	727	

Termini mantenuti Documenti

Cerca

Termini e sinonimi	Numero di documenti	Concetto
▶ materiale	3156	
▶ microscopico	3098	
▶ pervenire	3046	
▶ diagnosi	2827	
▶ clinico	2797	
□ descrizione	2596	
▶ macroscopico	2157	
▶ carcinoma	1911	
▶ diagnostico	1792	
▶ polmonare	1726	
□ quesito	1699	PROP_MISC
□ a	1688	
▶ cellula	1612	
▶ frammento	1102	
□ diagnosi	1093	PROP_MISC
□ descrizione	1087	PROP_MISC
▶ citologico	1079	
▶ differenziato	1045	
▶ dare	1012	
▶ cm	1007	
□ non	907	
▶ destro	901	
▶ quadro	892	
▶ neoplasia	879	

Termini

SCA_Project_PR_POLMONE

Topic [dropdown] Esegui [dropdown] [refresh] [share] [print] Visualizza [dropdown]

Topic



Topic	Numero di documenti
▼ Tutti i topic (3525)	
+estemporaneo,x,+margine,+sezione,+libero	299
+vetrino,laboratorio,+papanicolaou,+tale elemento,+elemento	266
+washing,+respiratorio,+block,+cell,citologico tratto respiratorio	119
+broncoaspirato,presenza,+citologico,broncoaspirato,+piccolo cellula	210
+frammento,+bc,+mucosa,+vario frammento,+lembo	313
anatomy,+cardiopatía,+patologico,+polmone,+base	191
+polipo,+displasia,+cm,basso grado,+basso	290
versamento spontaneo,+versamento,+spontaneo,+pleurico,+striscio	86
+notizia,+negativo,+citoincluso,n,+includere	196
+frustolo,+tissutale,+inclusione,+biopsia,+bronco	310
pap,+striscio,+colorazione,+taglia,+allestire	184

Termini Documenti



Documenti (210 di 3.525)

Topic > +broncoaspirato,presenza,+citologico,broncoaspirato,+piccolo cellula

Materiale pervenuto. Broncoaspirato Diagnosi clinica/Quesito diagnostico. Sosp. ca. polmonare Descrizione microscopica. Materiale citologico broncoaspirato - Presenza di cellule tumorali maligne. Il quadro citologico è compatibile con carcinoma non piccole ...	1
Materiale pervenuto. Broncoaspirato Diagnosi clinica/Quesito diagnostico. Sosp. Ca polmonare Descrizione microscopica. Materiale citologico broncoaspirato - Presenza di cellule tumorali maligne. Il quadro citologico è compatibile con carcinoma non piccole ...	0,997
Materiale pervenuto. Broncoaspirato Diagnosi clinica/Quesito diagnostico. Sospetto ca. polmonare sn. Descrizione microscopica. Materiale citologico broncoaspirato - Presenza di cellule tumorali maligne. Il quadro citologico è compatibile con carcinoma non piccole ...	0,995
Materiale pervenuto. Broncoaspirato Diagnosi clinica/Quesito diagnostico. Sospetto Ca polmonare Descrizione microscopica. Materiale citologico broncoaspirato - Presenza di cellule tumorali maligne. Il quadro citologico è compatibile con carcinoma non piccole ...	0,986
Materiale pervenuto. Broncoaspirato Diagnosi clinica/Quesito diagnostico. Sosp. Ca polmonare. Descrizione microscopica. Materiale citologico broncoaspirato - Presenza di gruppi di cellule neoplastiche. Il quadro citologico è compatibile con carcinoma non piccol...	0,98
Materiale pervenuto. Broncoaspirato Diagnosi clinica/Quesito diagnostico. Sospetto Ca	0,974

Topics

Es. TESTO: "La biopsia ha mostrato risultati istologici coerenti con il **carcinoma a cellule basali**. Questo tipo di **carcinoma** può tipicamente essere curato con l'escissione chirurgica. "

CONCEPT	CONCEPT DEFINITION RULE
TUMORLONG	CLASSIFIER:basal cell carcinoma
TUMORSHORT	CLASSIFIER:carcinoma
TUMORSHORT	REMOVE_ITEM:(ALIGNED,"_c{TUMORSHORT}","TUMORLONG")

La regola REMOVE_ITEM impedisce al concetto TUMORSHORT di estrarre "carcinoma" (testo in rosso) quando esiste una corrispondenza per "carcinoma a cellule basali" (testo blu) nello stesso documento.

NLP - Disambiguazione

L'utilizzo di soluzioni per l'analisi dei test offre un valore aggiunto al sistema per :

- La verifica delle classificazioni (SNOMED, ICD9,..) presenti nei dati
- La semplificazione dei test da rappresentare per la risoluzione dei casi
- Il miglioramento della qualità e completezza del tracciato RT
- L'integrazione automatica del tracciato RT (es. TNM)
- La compensazione di carenze informative (es. laboratorio)
- La riduzione dei tempi di lettura (migliaia di test letti)
- Il miglioramento della qualità e completezza dei test
- L'incremento della risoluzione automatica dei casi
- per sopperire alla povertà dei dati strutturati (amministrativi) e

Analisi test

Implementazioni successive attuate a seguito del rilascio del registro tumori:

- ✓ Favorire l'accesso ai dati del RT ai fini di controllo, pianificazione e gestione del SSR
- ✓
- ✗ Sperimentare la codifica automatica della casistica, mediante utilizzo di «**reti neurali artificiali**», al fine di ridurre la numerosità delle attribuzioni da risolvere manualmente



Grazie per l'attenzione!

pierantonio.romor@insiel.it