

Università degli Studi di Trieste

---

Corso di Laurea Magistrale in  
INGEGNERIA CLINICA

# TIPI DI DATO NELLA CARTELLA CLINICA ELETTRONICA: I DOCUMENTI TESTUALI

Corso di Informatica Medica

Docente Sara Renata Francesca MARCEGLIA



Dipartimento di Ingegneria e Architettura



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE

# GESTIONE DELL'INCERTEZZA NEI DATI MA ANCHE NEI DOCUMENTI



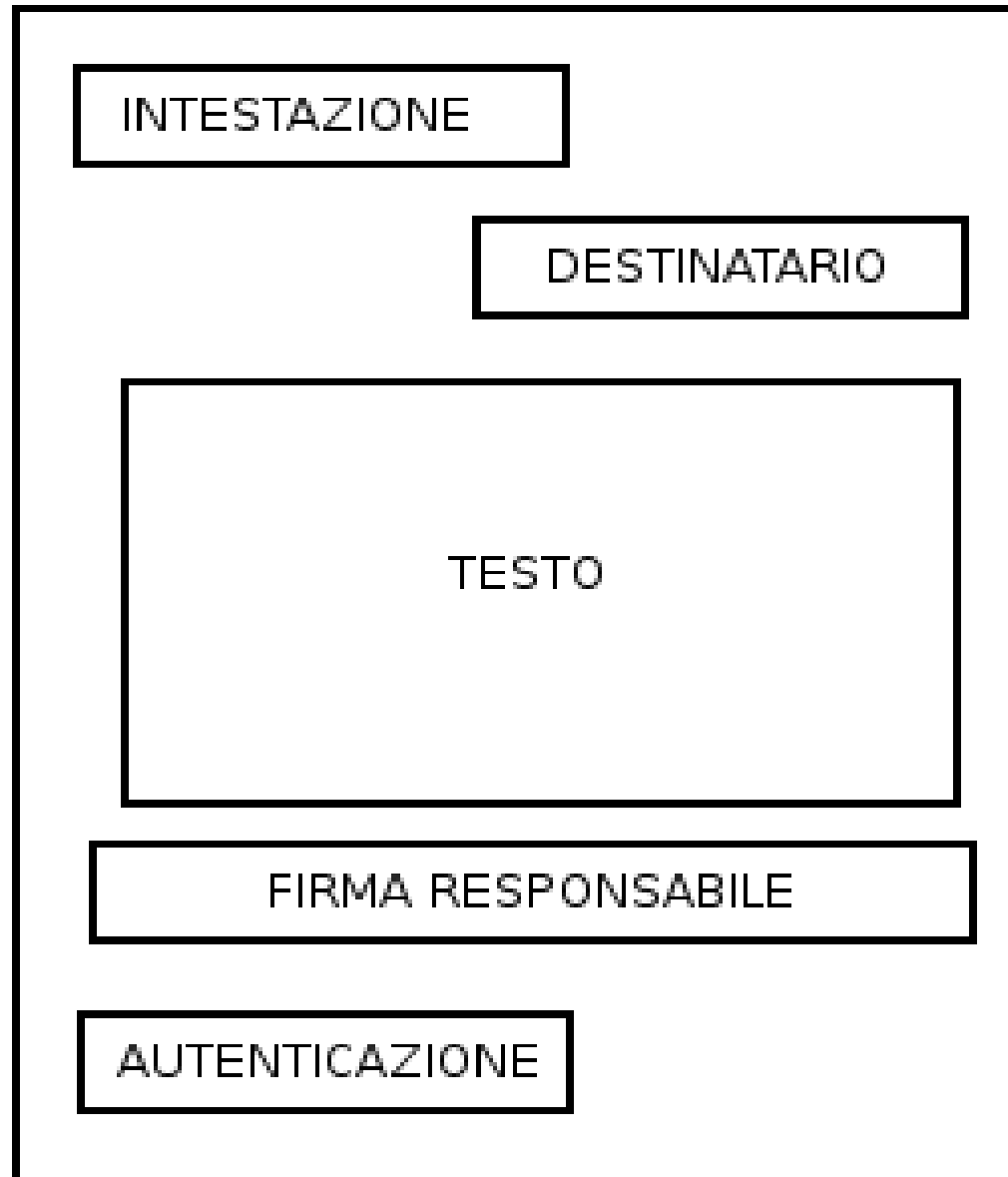
## DATI BIOMEDICI

- Valore del dato → contiene l'informazione grezza
- Come è stato registrato (sistema di misura o di rilevazione)
- In che condizione è stato registrato (condizione del paziente, condizione ambientale, etc)
- Chi ha la responsabilità della raccolta
- Chi ne è il proprietario (paziente)
- Perché è stato registrato
- Quali sono i parametri di interpretazione attuali (se esistono)

## DOCUMENTI BIOMEDICI

- ...?

# IL MODELLO DI DOCUMENTO BIOMEDICO



# GESTIONE DELL'INCERTEZZA NEI DATI MA ANCHE NEI DOCUMENTI



## DATI BIOMEDICI

- Valore del dato → contiene l'informazione grezza
- Come è stato registrato (sistema di misura o di rilevazione)
- In che condizione è stato registrato (condizione del paziente, condizione ambientale, etc)
- Chi ha la responsabilità della raccolta
- Chi ne è il proprietario (paziente)
- Perché è stato registrato
- Quali sono i parametri di interpretazione attuali (se esistono)

## DOCUMENTI BIOMEDICI

- Intestazione
- Destinatario
- Contenuto
- Firma
- Autenticazione

# INTESTAZIONE: CARATTERISTICHE

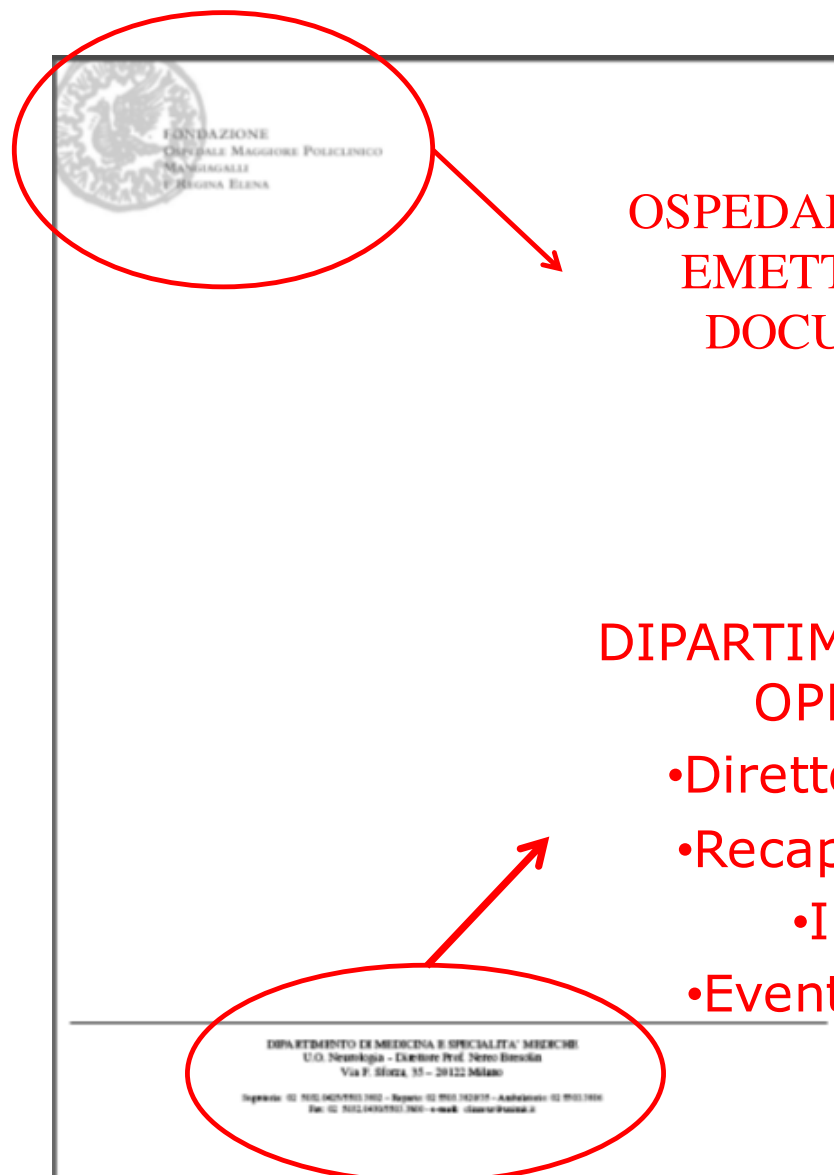


- Dichiarazione dell'identità di **chi sta emettendo il documento**
- Può essere la carta intestata dell'ospedale.
- **Esistono degli ALBI** (digitali/cartacei) che contengono tutte le possibili intestazioni vere
  - Elenco delle strutture Ospedaliere della Regione
  - Elenco delle aziende in Camera di Commercio
  - Elenco dei professionisti iscritti all'ordine professionale (es: medico)
  - Elenco dei cittadini iscritti all'anagrafe
  - Elenco degli aventi diritto all'assistenza e ai vari gradi di esenzione presso l'Azienda Sanitaria Locale

# INTESTAZIONE: ESEMPIO



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE



OSPEDALE CHE STA  
EMETTENDO IL  
DOCUMENTO

DIPARTIMENTO O UNITÀ  
OPERATIVA:

- Direttore/primario
- Recapiti telefonici
- Indirizzo
- Eventuale e-mail



# DESTINATARIO

- Dichiarazione dell'identità della **persona a cui il documento è rivolto**.
- Normalmente il **destinatario è il paziente**, anche se poi non sarà lui ad utilizzare l'informazione contenuta nel documento.
- È comunque importante capire a chi sia rivolto il documento generato (scopo del documento) →
  - ✓ Documento diagnostico
  - ✓ Referto diagnostico
  - ✓ Lettera di dimissione, ricetta...
  - ✓ Documento di protocollo sperimentale
  - ✓ Articolo scientifico

# CONTENUTO: REFERTI



Destinatario

Intestazione

Identificazione  
del referto

Conclusioni

Firma

ISTITUTO CLINICO  
HUMANITAS Istituto di Ricovero e Cura a Carattere Scientifico **ICHI**

DIPARTIMENTO DI DIAGNOSTICA PER IMMAGINI  
Direttore Dr. LUCA BALZARINI

SERVIZIO DI RADIOLOGIA DIAGNOSTICA  
Responsabile Dr. LUCA BALZARINI  
Sezione di Senologia – Capo Sezione Dr. Paolo Malerba  
Sezione Autonoma di Neuroradiologia – Capo Sezione Dr. Alberto Bizzi

SERVIZIO DI RADIOLOGIA ONCOLOGICA e INTERVENTISTICA  
Responsabile Dr. FABIO ROMANO LUTMAN  
Sezione di Radiologia Vascolare – Capo Sezione Dr. Vittorio Pedicini  
Sezione di Radiologia d'Urgenza – Capo Sezione Dr. Dario Poretti

SERVIZIO DI ECOGRAFIA  
Responsabile Dr.ssa PAOLA MAGNONI

Richiesta n° : 2130216761 Provenienza : ESTERNO

Cognome Nome : ██████████ A  
ID : 1565458  
Data di nascita : 16/03/1980  
Residenza : VIA EUROPA 18  
20097 - SAN DONATO MILANESE (MI)

Referto n. : ██████████  
842131.1  
Data referto : 11.03.2013 16.04

Prestazioni : (RMN SPALLA/BRACCIO DX SENZA CONTRASTO del 26.03.2013 TCRM Faenza  
Pierpaolo)

RMN SPALLA/BRACCIO DX SENZA CONTRASTO

Esame eseguito mediante acquisizioni sui piani ortogonali con sequenze ad eco variabile

Quesito clinico:algia

Note di artrosi acromion-claveare cui si associa riduzione in spessore della fascia adiposa sottoacromiale e impingement sottoacromiale.  
Minima disomogeneità del tendine del sovraspinato e del sottoscapolare.  
Tenosinovite del capo lungo del bicipite.  
Si segnala area di alterato segnale in adiacenza all'articolazione gleno-omeroale nelle parti molli muscoloscheletriche cui concomita modesto edema-ematoma adiacente utile valutazione clinica specialistica mirata.  
Tracce di versamento liquido a livello del recesso coracoideo  
Non espansi cistici e/o solidi di significato evolutivo a carico dei segmenti scheletrici esaminati

Dott.ssa Federica Mrakic-Sposta

COPIA CONFORME  
ALL'ORIGINALE

Sistema Sanitario Regione Lombardia

I medici del Servizio di Diagnostica per immagini sono a disposizione per chiarimenti al num. 0282246659 Pagina 1 di 1

Humanitas Mirafiori s.p.a.  
Società soggetta  
a direzione e coordinamento  
di Humanitas s.p.a.

Sede legale e operativa:  
Istituto Clinico Humanitas  
Via Manzoni 56 - 20089 Rozzano (Milano)  
Tel. 02.8224.1 - Fax 02.8224.2299

Capitale sociale € 13.840.000 (v)  
Partita IVA e Cod. Fisc. 10125410158  
Iscr. Tribunale Milano 10125410158  
(n. 310153) - 7785 - 3



# CONTENUTO: SCHEDA DI DIMISSIONE OSPEDALIERA (SDO)



- La scheda di dimissione è raccolta obbligatoriamente sia in caso di ricovero ordinario sia in caso di day hospital.
- Essa non si applica, invece, all'attività ambulatoriale né alle strutture socio-assistenziali (salvo diverse disposizioni delle Regioni).
- Le informazioni raccolte descrivono →
  - Aspetti clinici del ricovero (diagnosi e sintomi rilevanti, interventi chirurgici, procedure diagnostico-terapeutiche, impianto di protesi, modalità di dimissione)
  - Aspetti organizzativi del ricovero (ad esempio: unità operativa di ammissione e di dimissione, trasferimenti interni, soggetto che sostiene i costi del ricovero).
- Un numero ristretto di tali informazioni viene trasmesso dalle Regioni al Ministero della salute per le attività di indirizzo e monitoraggio nazionale.
- Dalla scheda di dimissione sono escluse informazioni relative ai farmaci somministrati durante il ricovero o le reazioni avverse ad essi (oggetto di altre specifici flussi informativi).

# SDO - ESEMPIO



Anno: 2003 ,Cartella : 4 BOZZA DI DIMISSIONE GIOVEDÌ 25 SETTEMBRE 2003

(20) PASSAGGIO DAL REPARTO DI AMMISSIONE A QUELLO DI COMPETENZA NOSOLOGICA	
Data	Reparto
(20) TRASFERIMENTI INTERNI	
(1) Data	Reparto
Firma del primario	Diagnosi di trattamento
(2) Data	Reparto
Firma del primario	Diagnosi di trattamento
(3) Data	Reparto
Firma del primario	Diagnosi di trattamento
(4) Data	Reparto
Firma del primario	Diagnosi di trattamento
(32) AREA FUNZIONALE DI DIMISSIONE	
AFO Medica	(33) MODALITÀ DI DIMISSIONE
(32) Data Di Dimissione	(21) Reparto di dimissione
25/09/2003	Chirurgia
(37) Peso in Kg	(35) DIAGNOSI PRINCIPALE ALLA DIMISSIONE
85	Traumatismi intracranici di altra e non specificata natura, senza menzione di ferita intracranica esposta, stato di coscienza non specificato
(39) PATOLOGIE CONCOMITANTI O COMPLICANZE DELLA MALATTIA PRINCIPALE	
1)	
2)	
3)	
4)	
5)	
6)	
7)	
(30) INTERVENTO CHIRURGICO PRINCIPALE O PARTO	
Data	Tipo di intervento/Parto
(31) ALTRI INTERVENTI O PROCEDURE	
1) Innesto di nervi cranici o periferici	04.5
2)	
3)	
4)	
5)	
Servizio da compilare solo in caso di ricovero in regime di Day Hospital	
(25) Motivo del Ricovero	(26) Nr. Giorni di presenza
FIRMA DEL PRIMARIO	FIRMA DEL MEDICO CURANTE/COMPILATORE

Intestazione → codice centro (8 cifre)

Destinatario →

- Codice comune di nascita (ISTAT)
- Codice comune di residenza (ISTAT)
- Cittadinanza (3 cifre, 100=ITA)
- CF e data di nascita
- ASL di appartenenza (codice)

Provenienza (altri reparti, altre strutture)

Regime di ricovero

Onere

Diagnosi

Data di dimissione

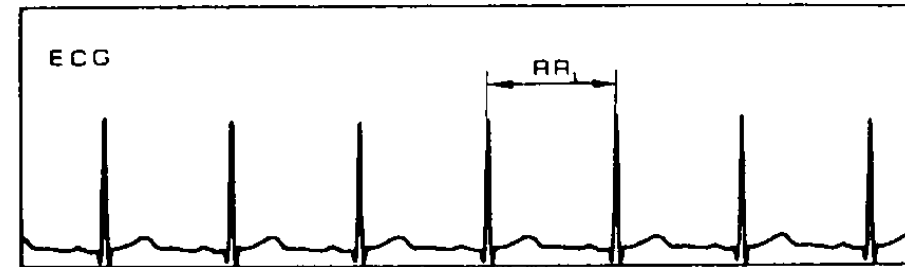
Firma

# CONTENUTO: REPERTI



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE

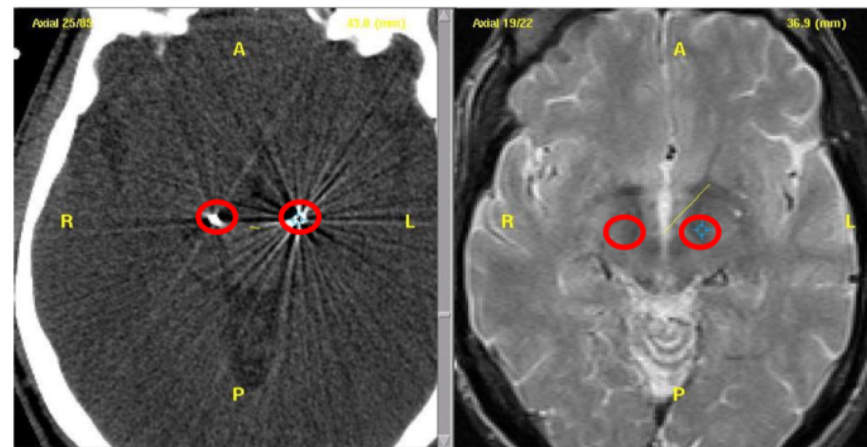
## BIOSEGNALI



## BIODATI

Crasi	
Globuli rossi	5.66 × 1000/ul
Globuli bianchi	7.76 /ul
HB	14.4 g/dl
HCT	43.0 %
MCV	76.0 fl
MCH	25.4 pg
MCHC	33.5 g/dl

## BIOIMMAGINI



## BIOFILMATI



## FIRMA DEL RESPONSABILE

- Identificazione del **soggetto che si prende la responsabilità** della generazione del documento biomedico e quindi del suo contenuto.
- Si tratta, solitamente, di un medico/tecnico autorizzati ad effettuare l'esame diagnostico.
- La conferma dell'intestazione presuppone lo stabilire dei collegamenti, che spesso non sono immediati, ad esempio tra l'intestazione della struttura ospedaliera e la firma di un responsabile di tale struttura (oppure tra la denominazione dell'azienda e chi vi è titolare di cariche sociali).



# AUTENTICAZIONE

- È l'operazione con cui qualcuno degno di fede testimonia che effettivamente la firma è stata apposta in sua presenza da chi si è identificato a lui come l'individuo in questione.
- Non è sempre presente.
- Alcuni esempi →
  - ✓ Notaio
  - ✓ Segretario comunale
  - ✓ Pubblico ufficiale riconosciuto allo scopo
  - ✓ Testimone generico
- Si può parlare di **autenticazione digitale** se esistono degli algoritmi che sono utilizzati per validare la firma del documento (firma elettronica/digitale).

# FIRMA E AUTENTICAZIONE: IL CAD



- È nato da qualche anno il CAD - “codice dell’amministrazione digitale”
- Regolamenta la validità giuridica dei documenti digitali
- I documenti devono essere firmati:
  - Firma elettronica
  - Firma elettronica autenticata
  - Firma elettronica qualificata
  - Firma digitale
- Sono validi a tutti gli effetti i documenti firmati con firma elettronica autenticata, digitale o qualificata
- Le copie (elettroniche o cartacee di documenti nati come elettronici o cartacei) sono ritenute conformi se non esplicitamente dichiarato il contrario

# TIPI DI FIRMA



- q) **firma elettronica**: l'insieme dei **dati in forma elettronica**, allegati oppure connessi tramite associazione logica ad altri dati elettronici, utilizzati come metodo di identificazione informatica;
- q-bis) **firma elettronica avanzata**: insieme di **dati in forma elettronica** allegati oppure connessi a un documento informatico che consentono l'identificazione del firmatario del documento e **garantiscono la connessione univoca al firmatario**, creati con **mezzi sui quali il firmatario può conservare un controllo esclusivo**, collegati ai dati ai quali detta firma si riferisce in modo da consentire di **rilevare** se i dati stessi siano stati **successivamente modificati**;
- r) **firma elettronica qualificata**: un particolare tipo di **firma elettronica avanzata che sia basata su un certificato qualificato** e realizzata mediante un dispositivo sicuro per la creazione della firma;
- s) **firma digitale**: un particolare tipo di **firma elettronica avanzata basata su un certificato qualificato e su un sistema di chiavi crittografiche**, una pubblica e una privata, correlate tra loro, che consente al titolare tramite la chiave privata e al destinatario tramite la chiave pubblica, rispettivamente, di rendere manifesta e di verificare la provenienza e l'integrità di un documento informatico o di un insieme di documenti informatici;

# IL DOCUMENTO TESTUALE STRUTTURATO



	SINDROME O MALATTIA		Caratteristiche Temporali	Eziologia
		Localizzazione		
01				
02				
03				
04				
05				



# VANTAGGI E SVANTAGGI

## Testo libero:

- Documento non strutturato
- Documento in linguaggio non standard
- Può essere reso strutturato a posteriori tramite tecniche di Natural Language Processing (NLP) e di estrazione dei concetti (anche standard)
- Dà maggior flessibilità nell'interpretazione (si possono utilizzare concetti non codificati, si possono inserire osservazioni non contemplate dagli attributi proposti nel testo strutturato)

## Testo strutturato e standardizzato:

- Garantisce la comprensione da parte di più professionisti
- Possibilità di effettuare ricerche
- Possibilità di riutilizzare l'informazione contenuta
- Possibilità di contestualizzare semanticamente l'informazione contenuta
- Possibilità di confrontare testi scritti da operatori diversi (potrebbero non contenere tutte le informazioni necessarie)

# DOCUMENTO STRUTTURATO E STANDARDIZZATO



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE

STRUTTURATO = documento organizzato secondo uno schema che può essere standard

STANDARDIZZATO = documento che utilizza una terminologia medica condivisa per garantire la comprensibilità dei contenuti anche da un calcolatore



Original Investigations  
JAMIA

*Model Formulation* ■

## HL7 Clinical Document Architecture, Release 2

ROBERT H. DOLIN, MD, LIORA ALSCHULER, SANDY BOYER, BSP, CALVIN BEEBE,  
FRED M. BEHLEN, PhD, PAUL V. BIRON, AMNON SHABO (SHVO), PhD

**Abstract** Clinical Document Architecture, Release One (CDA R1), became an American National Standards Institute (ANSI)-approved HL7 Standard in November 2000, representing the first specification derived from the Health Level 7 (HL7) Reference Information Model (RIM). CDA, Release Two (CDA R2), became an ANSI-approved HL7 Standard in May 2005 and is the subject of this article, where the focus is primarily on how the standard has evolved since CDA R1, particularly in the area of semantic representation of clinical events. CDA is a document markup standard that specifies the structure and semantics of a clinical document (such as a discharge summary or progress note) for the purpose of exchange. A CDA document is a defined and complete information object that can include text, images, sounds, and other multimedia content. It can be transferred within a message and can exist independently, outside the transferring message. CDA documents are encoded in Extensible Markup Language (XML), and they derive their machine processable meaning from the RIM, coupled with terminology. The CDA R2 model is richly expressive, enabling the formal representation of clinical statements (such as observations, medication administrations, and adverse events) such that they can be interpreted and acted upon by a computer. On the other hand, CDA R2 offers a low bar for adoption, providing a mechanism for simply wrapping a non-XML document with the CDA header or for creating a document with a structured header and sections containing only narrative content. The intent is to facilitate widespread adoption, while providing a mechanism for incremental semantic interoperability.

■ J Am Med Inform Assoc. 2006;13:30-39. DOI 10.1197/jamia.M1888.



# DA DOCUMENTO NON STRUTTURATO...

EXAMPLE 1 – Medical report at enrollment visit

**Piacenza, Nov 12, 2015**

**Case History.** Mr. John Doe complains having difficulty in hearing in both ears since 10 years ago when he had a car accident resulting in head injury. He suffers from diabetes type 2, hypertension and arteriosclerosis. He reports tinnitus in both ears since last 2 years. He was diagnosed with Meniere. No family history of hearing loss.

Mr. Doe denies ear pain, ear fullness and balance problems.

## **Evaluation.**

Otoscopic inspection: normal bilaterally

Immittance: type A bilaterally

Pure tone audiometry: bilateral sensorineural hearing loss at 2000-8000 Hz, moderate at right ear, profound at left ear.

Speech audiometry: max intelligibility 100% at 60 dB right ear and 0% at 90 dB at left.

Word discrimination: 80%

Word recognition: 85%

Sound field speech audiometry in noise: max intelligibility 35%, S/N 5 dB.

CT scan: normal

MRI: normal

## **Psychometric Results**

Impact of tinnitus: THI score 76, grade 4

## **Conclusions**

Based on history and current exams, Mr. Doe has a bilateral sensorineural hearing loss of profound degree for the left ear with Meniere disease and tinnitus. It is recommended to have a cochlear implant at the left ear.

# ... A DOCUMENTO STRUTTURATO



Section	Sub-section	Coded concept	Auditory disorders axis (and sub-class)	Risk factor group
Medical history evidences	Presenting complaint(s)	[C1384666] Hearing impairment	Hearing disorder ()	-
	Significant diseases/illnesses/disorders/signs having still present relevance	[C0011860] Diabetes Mellitus, Non-Insulin-dependent	-	General medical conditions
		[C0020538] Hypertensive disease	-	General medical conditions
		[C0003850] Arteriosclerosis	-	Vascular disease
		[C0040264] Tinnitus	Hearing disorder ()	-
		[C0025281] Meniere Disease	Ear disorders (peripheral vertigo)	-
	Traumatic injuries	[C0744612] Head injury trauma	-	General medical conditions
Family history of HL	None	-	-	
Examinations – audiometric	Audiometric thresholds for tones	Date: 2015/11/12. Diagnostic procedure: [C2022139] Pure tone threshold audiogram. Type of HL: [C0018784] Sensorineural Hearing Loss (disorder). Laterality: [C0238767] Bilateral. HL degree right ear: [C0205081] Moderate (severity modifier). HL degree left ear: [C0439808] Profound. Range of frequencies (Hz): 2000/8000.	Hearing disorder (hearing loss)	-

- È possibile effettuare analisi su testi non strutturati?
- È possibile estrarre informazione anche se il testo non è codificato/standardizzato?



## TECNICHE DI NATURAL LANGUAGE PROCESSING

Esempio: <http://linguistic-annotation-tool.italianlp.it>



## I PASSI PRINCIPALI DEL NLP

### SENTENCE TOKENIZER

- Identificazione dei confini delle frasi (token)
- Può considerare parole (congiunzioni), spazi, o segni di punteggiatura

### PART OF SPEECH (POS) TAGGER

- Definizione della categoria grammaticale (nome, verbo, etc)

### PARSER

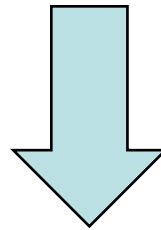
- Identificazione delle relazioni grammaticali tra le parti identificate



# SENTENCE TOKENIZER



**OCULAR COMPLICATIONS OF MYASTHENIA GRAVIS.**



**OCULAR**

**COMPLICATIONS**

**OF**

**MYASTHENIA**

**GRAVIS**

**.**

# PART OF SPEECH TAGGER



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE

NOME

NOME  
COMPOSITO

OCULAR COMPLICATIONS OF MYASTHENIA GRAVIS .

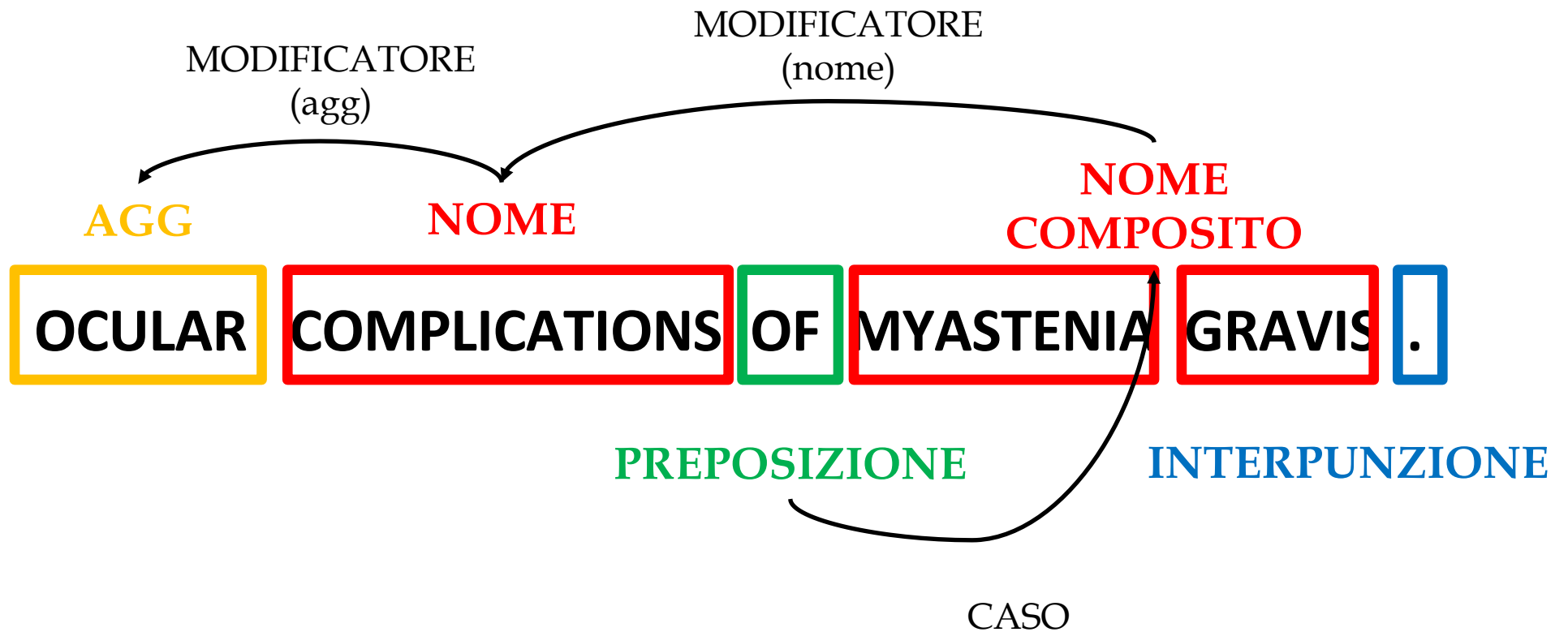
AGGETTIVO

PREPOSIZIONE

INTERPUNZIONE

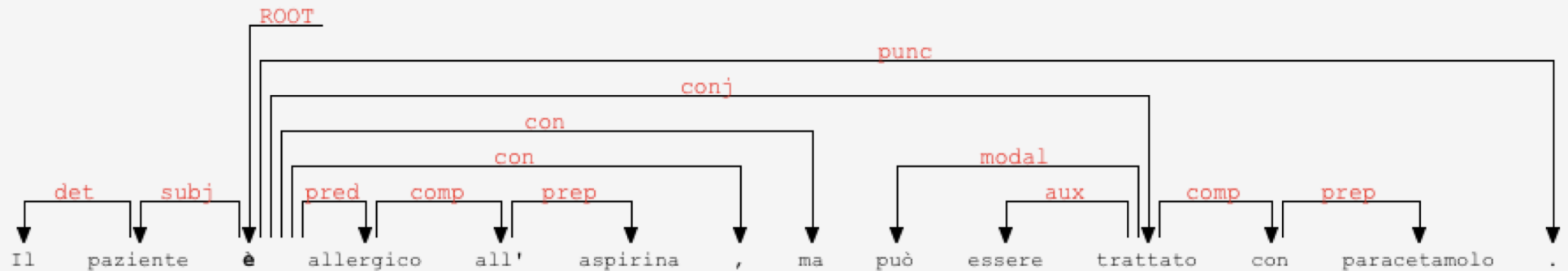


# PARSER



# ESEMPIO

Il paziente è allergico all'aspirina, ma può essere trattato con paracetamolo.



# DAL NLP ALL'INFORMATION EXTRACTION



## Named Entity Recognition

- Categorizzazione dei sostantivi (persona, sostanza, entità geografica, ...)
- Es: Adam Smith lavora alla NASA → Adam Smith = persona, NASA = azienda

## Semantic Role Recognition

- Riconoscimento del ruolo semantico
- Es: Adam Smith sta alzando la mano → Adam = agent; la mano=parte del corpo; alzare = azione

## Entity Relation Recognition

- Identificazione delle relazioni tra le entità
- Es: Adam Smith lavora alla NASA → Adam Smith = persona, lavora per = relazione, NASA = azienda

## Timex and Time Line Recognition

- Riconoscimento delle espressioni temporali
- Es: oggi, domani, settimana scorsa, 11 maggio, etc

## Word Sense Disambiguation

- Discriminazione degli omografi
- Es: Il bambino mangia la pesca vs Il bambino pesca la trota

## Negation Detection Identification

- Comprensione della negazione
- Es: il paziente non presenta segni di infarto → infarto = assente

# ESTRAZIONE DI INFORMAZIONI TRAMITE COGNITIVE COMPUTING



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE

Testo libero



NLP e parti del discorso



Modello di conoscenza del dominio



Annotazione del testo



Addestramento del Sistema

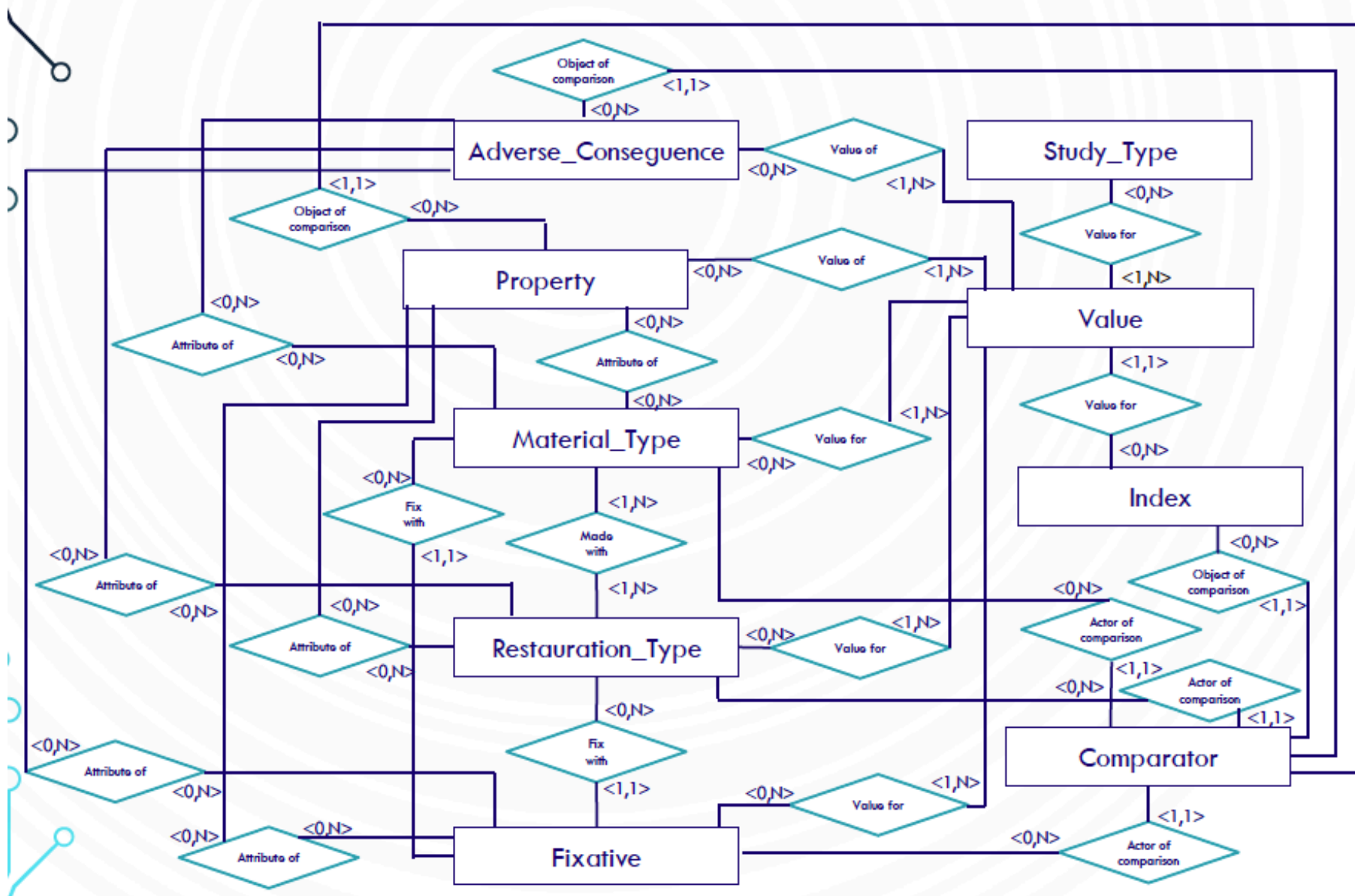


Applicazione

# ESEMPIO



## MODELLO DI CONOSCENZA



# ESEMPIO



## Annotazione delle Entità

IBM Watson Knowledge Studio

serena.narbone@gmail.com

Back to Workspaces | Back to Task | Open document list

View Details | Attribute View | Completed

### Fracture resistance of different implant.txt

1 Fracture resistance of different implant abutments supporting all-ceramic single crowns after aging.

2 Stimmelmayer M, Heiß P, Erdelt K, Schweiger J, Beuer F.

3 OBJECTIVE: To test the mechanical properties of three different restorative materials for implant abutments supporting all-ceramic single crowns.

4 MATERIALS AND METHODS: Thirty implants with butt-joint connections were distributed into three test groups: Group A with 10 one-piece zirconia abutments, Group U with 10 titanium abutments, and Group T with 10 titanium-zirconia hybrid abutments.

5 Monolithic zirconia single crowns were cemented and artificially aged.

6 The crowns were loaded at a 30-degree angle in a universal testing machine until fracture or bending.

7 Additionally, after removal of the restorations, the implant-abutment interface of the fixtures was inspected using a scanning electron microscope (SEM).

8 RESULTS: In Group A, the abutments failed on average at 336.78 N, in Group U at 1000.12 N, and in Group T at 1296.55 N. The mean values between Groups T and U ( $P = 0.009$ ), and between Group A and Groups T and U ( $P < 0.001$ ) were significantly different.

9 The abutments in Group A failed early due to fractures of the internal parts and parts close to the implant neck.

10 In Groups T and U, failures occurred due to bending of the implant neck.

11 CONCLUSION: This experimental study proves that hybrid and titanium abutments have similar mechanical properties.

Entity	Mention	
Type	Subtype	Role
A	Adverse_Consequence	
C	Comparator	
F	Fixative	
U	Follow_up_period	
I	Index	
M	Material_Type	
P	Property	
R	Restauracion_Type	
S	Study_Type	
V	Value	

# ESEMPIO



## Annotazione delle Relazioni

IBM Watson Knowledge Studio

Completed

Relation Type

- Actor\_of\_comparison
- Attribute\_of
- Fix\_with
- Made\_with
- Object\_of\_comparison
- Value\_for
- Value\_of

13

Comparator

There were no significant differences among groups.

13

Object\_of\_comparison

Value\_of

Property

Value

Value

Value

Comparator

The percentages of satisfactory tooth-composite internal adaptation varied from 81 to 98 % in occlusal dentin, from 63 to 90 % in axial dentin, and from 71 to 84 % in cervical dentin without any statistical difference.

13

Actor\_of\_comparison

Made\_with

Attribute\_of

Material\_Type

Material

Rest...

Object\_of\_comparison

Property

Comparator

CONCLUSIONS: The results of the present study support the use of flowable or restorative composites as a liner underneath ceramic, CAD/CAM inlays, producing marginal and internal adaptation which is not different from restorations placed directly on dentin.



# ESEMPIO

## Knowledge Studio - SV

Language of documents  
English

### Number of documents per set



70 Training Set  
18 Test Set  
22 Blind Set



### Training Set

[View Ground Truth](#)

### Test Set

[View Ground Truth](#)

[View Decoding Results](#)

### Last trained on:

Jul 3, 2019 12:13:46 PM

### Last evaluated on:

Jul 3, 2019 12:16:32 PM

[Train and evaluate](#)

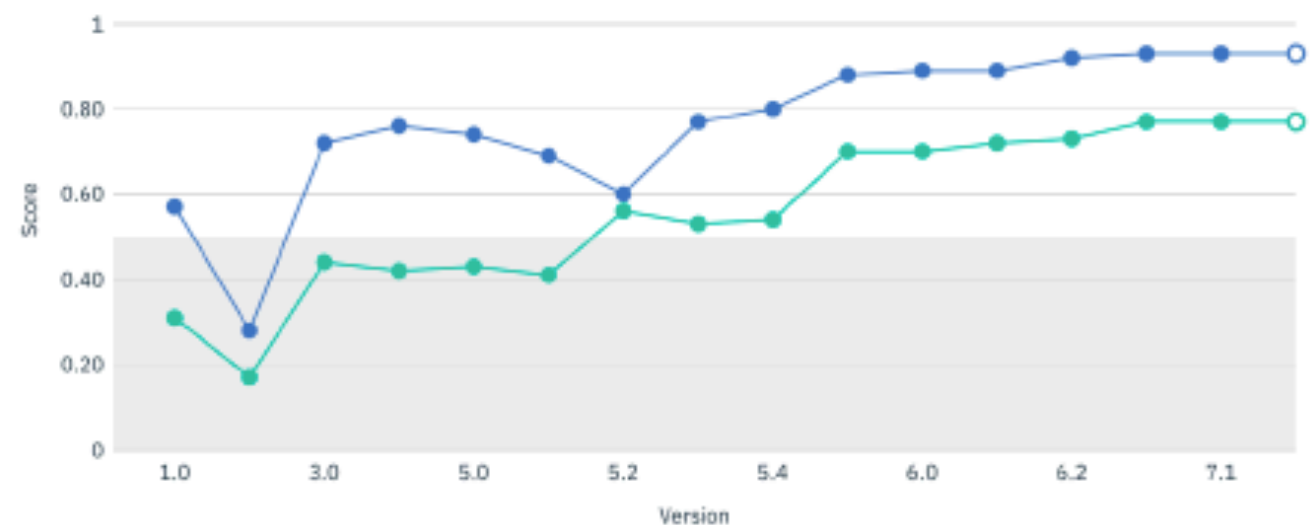
Low performance? [Click here to train.](#)

**PRECISION:** frazione di annotazioni corrette (rispetto all'annotatore umano)

**RECALL:** frazione di annotazioni correttamente create relative ad una certa etichetta (concetto o relazione) rispetto a quelle che sarebbero dovute essere create in base all'annotazione umana

## Document set evaluation ?

### Model over time



[View Log](#)

### Mention

**0.93** Precision: 0.95  
Recall: 0.92

### Relation

**0.77** Precision: 0.76  
Recall: 0.78

### Coreference

-- Precision: --  
Recall: --

Low performing range



# IL VALORE DEL TESTO LIBERO



UNIVERSITÀ  
DEGLI STUDI DI TRIESTE

*Journal of the American Medical Informatics Association*, 26(11), 2019, 1189–1194

doi: 10.1093/jamia/ocz119

Advance Access Publication Date: 12 August 2019

Research and Applications



OXFORD

Research and Applications

## Real world evidence in cardiovascular medicine: ensuring data validity in electronic health record-based studies

Tina Hernandez-Boussard,<sup>1,2,3</sup> Keri L Monda,<sup>4,5</sup> Blai Coll Crespo,<sup>4</sup> and Dan Riskin<sup>1,3,6</sup>

### ABSTRACT

**Objective:** With growing availability of digital health data and technology, health-related studies are increasingly augmented or implemented using real world data (RWD). Recent federal initiatives promote the use of RWD to make clinical assertions that influence regulatory decision-making. Our objective was to determine whether traditional real world evidence (RWE) techniques in cardiovascular medicine achieve accuracy sufficient for credible clinical assertions, also known as “regulatory-grade” RWE.

**Design:** Retrospective observational study using electronic health records (EHR), 2010–2016.

**Methods:** A predefined set of clinical concepts was extracted from EHR structured (EHR-S) and unstructured (EHR-U) data using traditional query techniques and artificial intelligence (AI) technologies, respectively. Performance was evaluated against manually annotated cohorts using standard metrics. Accuracy was compared to pre-defined criteria for regulatory-grade. Differences in accuracy were compared using Chi-square test.

**Results:** The dataset included 10 840 clinical notes. Individual concept occurrence ranged from 194 for coronary artery bypass graft to 4502 for diabetes mellitus. In EHR-S, average recall and precision were 51.7% and 98.3%, respectively and 95.5% and 95.3% in EHR-U, respectively. For each clinical concept, EHR-S accuracy was below regulatory-grade, while EHR-U met or exceeded criteria, with the exception of medications.

**Conclusions:** Identifying an appropriate RWE approach is dependent on cohorts studied and accuracy required. In this study, recall varied greatly between EHR-S and EHR-U. Overall, EHR-S did not meet regulatory grade criteria, while EHR-U did. These results suggest that recall should be routinely measured in EHR-based studies intended for regulatory use. Furthermore, advanced data and technologies may be required to achieve regulatory grade results.