

# Intelligenza Artificiale per l'Estrazione e la Gestione della Conoscenza da Documenti Testuali Biomedicali

Mario Ciampi, Angelo Esposito, Francesco Gargiulo, Mario Sicuranza, Stefano Silvestri e Giuseppe De Pietro

*Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale della Ricerche - ICAR-CNR*

*Via Pietro Castellino 111 - 80131, Napoli, Italia*

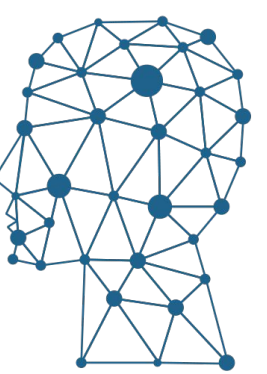
{mario.ciampi, angelo.esposito, francesco.gargiulo, mario.sicuranza, stefano.silvestri, giuseppe.depietro}@icar.cnr.it



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*



**Ital-IA**  
ITALIA INTELLIGENZA ARTIFICIALE  
cni National Lab AIIS



# Idea Progettuale

- Definizione di metodologie IA innovative necessarie per sfruttare il potenziale insito nelle grandi moli di dati testuali sanitari
  - Cartelle cliniche, referti, pubblicazioni scientifiche, annotazioni personali, etc.
- Estrazione di informazioni e costruzione di conoscenza da documenti narrativi semi-strutturati o non strutturati, contenenti parti narrative in lingua italiana
- Realizzazione di servizi e strumenti avanzati a supporto del personale sanitario, degli organismi amministrativi e dei cittadini



# Tecniche IA Utilizzate

- Modelli *Word Embeddings* (WEs) addestrati su testo di dominio biomedico
- *Character Embeddings*
- *Clustering*
- Reti neurali Deep Learning (DL):
  - di tipo *Convolutional Neural Network (CNN)*
  - di tipo *Bidirectional Long Short Term Memory (Bi-LSTM)*
- *Conditional Random Fields (CRF)*
- *Active Learning*
- *Distant Supervision*

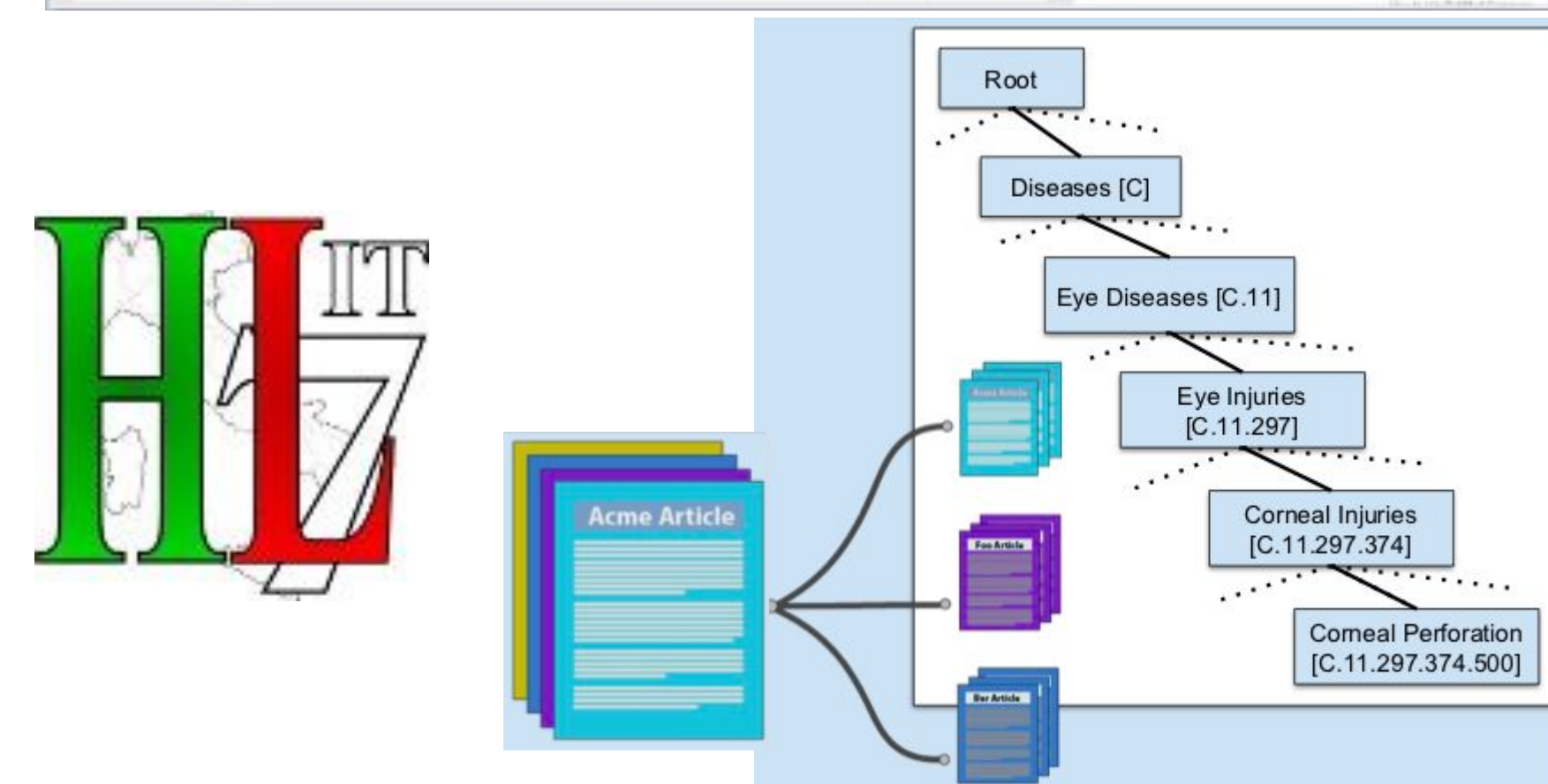
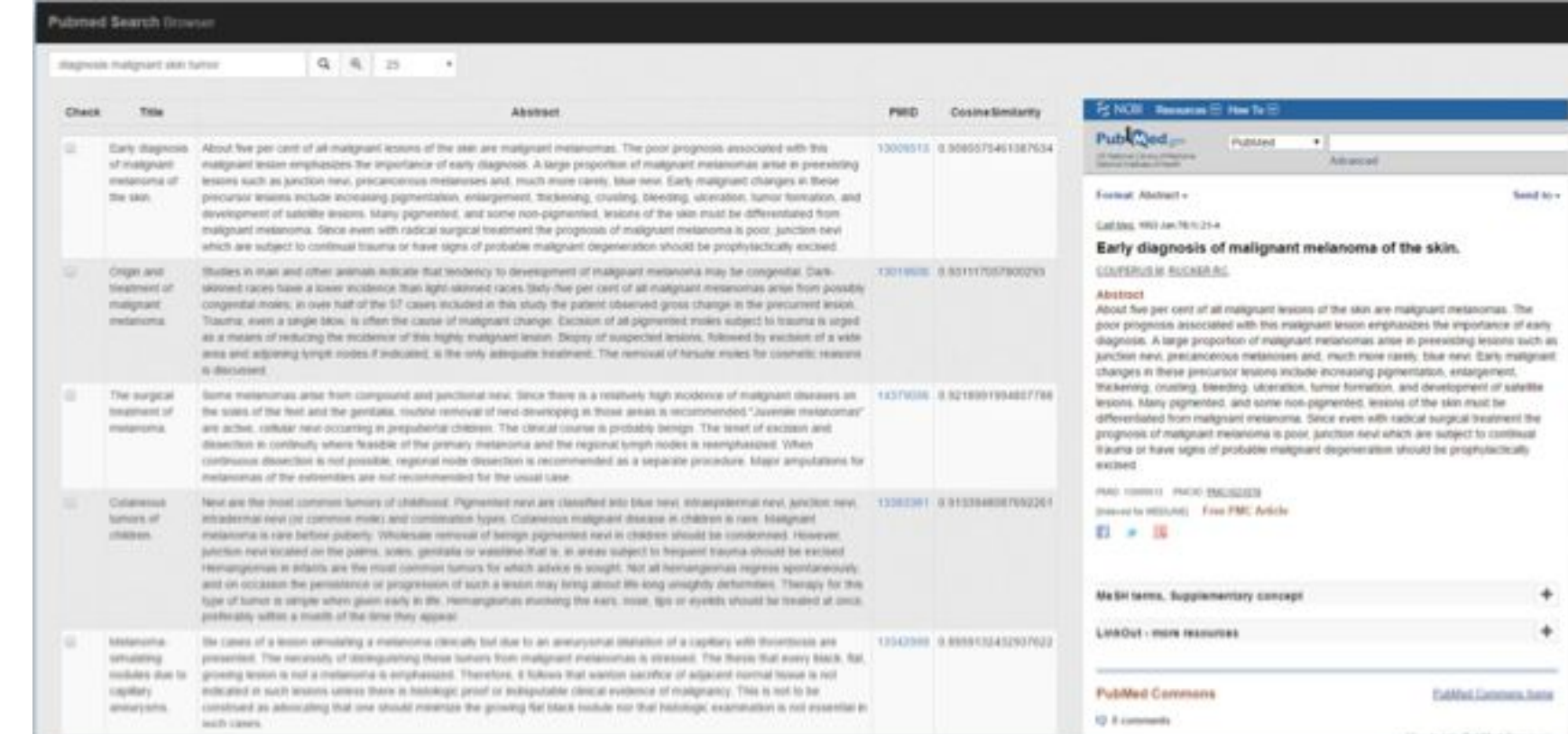


Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni



# Risultati

- Definizione e realizzazione di un motore di ricerca semantico per articoli scientifici
- Sistemi per il supporto alla validazione automatica delle regole di conformità in linguaggio naturale nelle Implementation Guide CDA R2 di HL7 Italia
- Classificazione di testo biomedico di tipo *eXtreme Multilabel Text Classification*
- Annotazione automatica real-time di termini medici mediante Named Entity Recognition (NER) specializzato per la lingua italiana



Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni



Ital-IA  
ITALIA INTELLIGENZA ARTIFICIALE  
ini National Lab AIIS



# Impatto Atteso

- Indicizzazione automatica di documenti biomedicali non strutturati (ad es. in sistemi di Fascicolo Sanitario Elettronico)
- Collegamento tra informazioni estratte da documenti e letteratura scientifica
- Conversione automatica delle informazioni testuali in formati strutturati standard (HL7 CDA R2 e FHIR)
- Annotazione di corpora biomedicali in lingua italiana
- Sistemi di analisi e di supporto delle decisioni in ambito clinico



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*



# Criticità

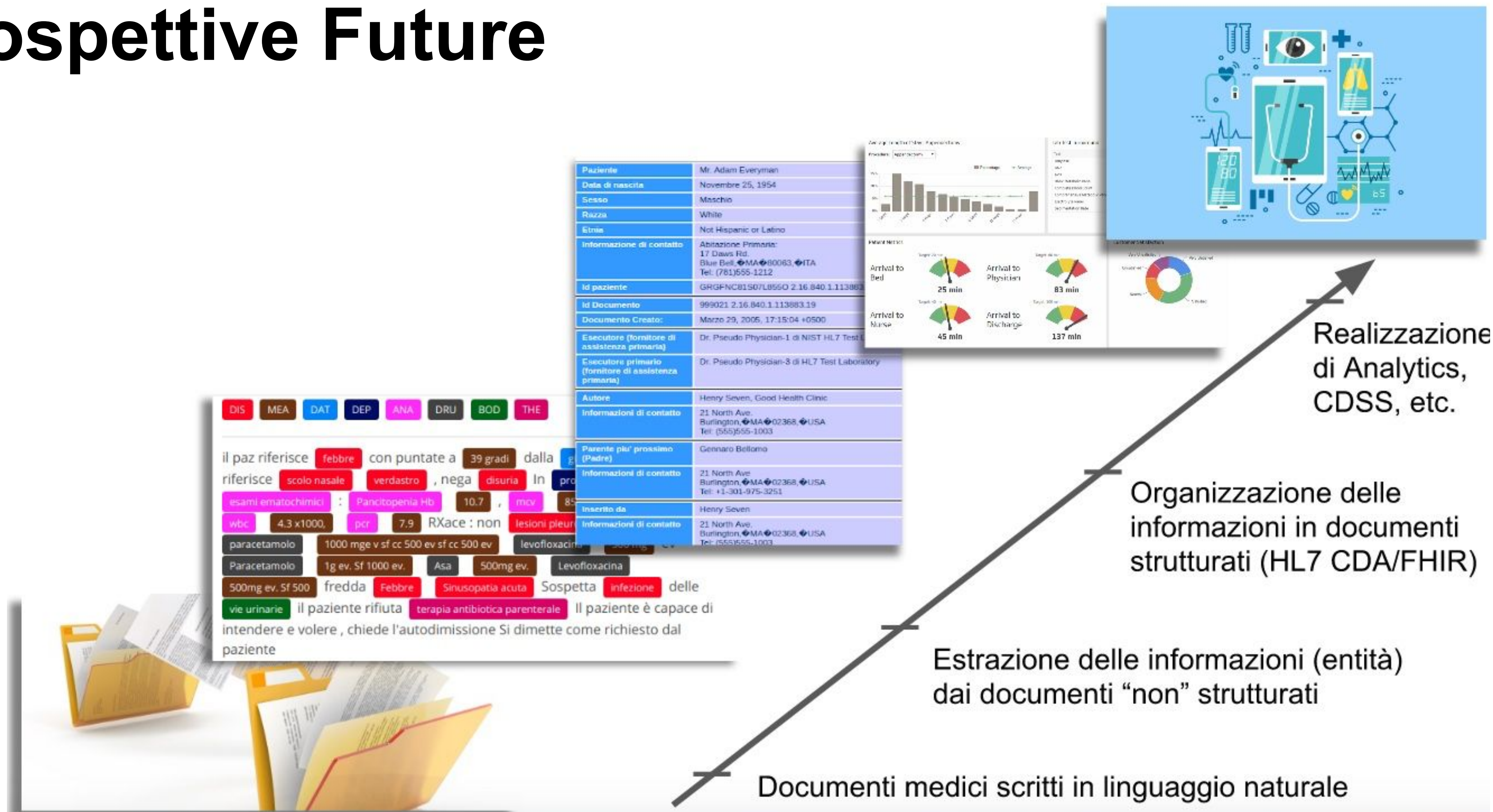
- Difficoltà nel trattamento automatico del linguaggio naturale e di testi di dominio medico in italiano
  - Lingua complessa e variabile, polisemica, locuzioni e acronimi specifici del dominio, termini tecnici, etc.
- Scarsa disponibilità di strumenti specifici per l'italiano e/o per il dominio biomedicale
- Necessità di definire metodi performanti per l'estrazione di relazioni tra entità
- Grosse moli di dati da processare
- Elevata complessità computazionale
- Scarsità di dati medici/clinici annotati in italiano



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*



# Prospettive Future



# Ricadute Progettuali (1/2)

- **FSE6** - Realizzazione di servizi e strumenti a favore delle Pubbliche Amministrazioni per l'attuazione del Fascicolo Sanitario Elettronico
- **FSE7** - Ottimizzazione del workflow e dei processi del Fascicolo Sanitario Elettronico
  - Convenzioni operative tra Agenzia per l'Italia Digitale (AgID) e CNR
  - Definizione e progettazione dell'architettura del framework nazionale per l'interoperabilità tra i sistemi regionali di FSE, in conformità al D.L. 179/2012 e ss.mm.ii.
- **BDA4PHR** - Big Data Analytics for Personal Health Record
  - Definizione e realizzazione di una piattaforma integrata basata su servizi di Big Data Analytics (BDA) associati a meccanismi di raccolta, estrazione e gestione di dati sanitari su piattaforma Cloud e analisi attraverso sistemi di IA



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*





# Ricadute Progettuali (2/2)

- **RARE.PLAT.NET** - Innovazioni diagnostiche e terapeutiche per tumori neuroendocrini, endocrini e per il glioblastoma attraverso una piattaforma tecnologica integrata di competenze cliniche, genomiche, ICT, farmacologiche e farmaceutiche
  - Estrazione, integrazione, analisi e studio di correlazioni tra i dati sanitari, clinici, genomici e farmacologici attraverso tecniche di IA e BDA



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*



# Grazie per l'Attenzione



*Consiglio Nazionale delle Ricerche  
Istituto di Calcolo e Reti ad Alte Prestazioni*



**Ital-IA**  
ITALIA INTELLIGENZA ARTIFICIALE  
**ini** National Lab **AIIS**

