DiOnEpIt

Dizionario Ontologico delle Espressioni Poliremiatiche dell'Italiano

Ph.D student Michele Stingo

DIPARTIMENTO DI SCIENZE POLITICHE E DELLA COMUNICAZIONE

UNIVERSITÁ DEGLI STUDI DI SALERNO - NETWORK CONTACTS SRL

DiOnEPiT

Obiettivi realizzativi

- Creazione di una risorsa lessicografica digitalizzata per le MWE
- Raccolta ed integrazione di diverse fonti informative
- Paradigma Open Data

DiOnEPiT

Espressioni Polirematiche

- Creazione di una risorsa lessicografica digitalizzata per le MWE
- Raccolta ed integrazione di diverse fonti informative
- Paradigma Open Data

"gruppo di parole che ha un significato unitario, non desumibile da quello delle parole che lo compongono, sia nell'uso corrente sia nei linguaggi tecnico-specialistici"
(De Mauro, 2000)

Ferro da stiro Conflitto d'interesse

DiOnEPiT

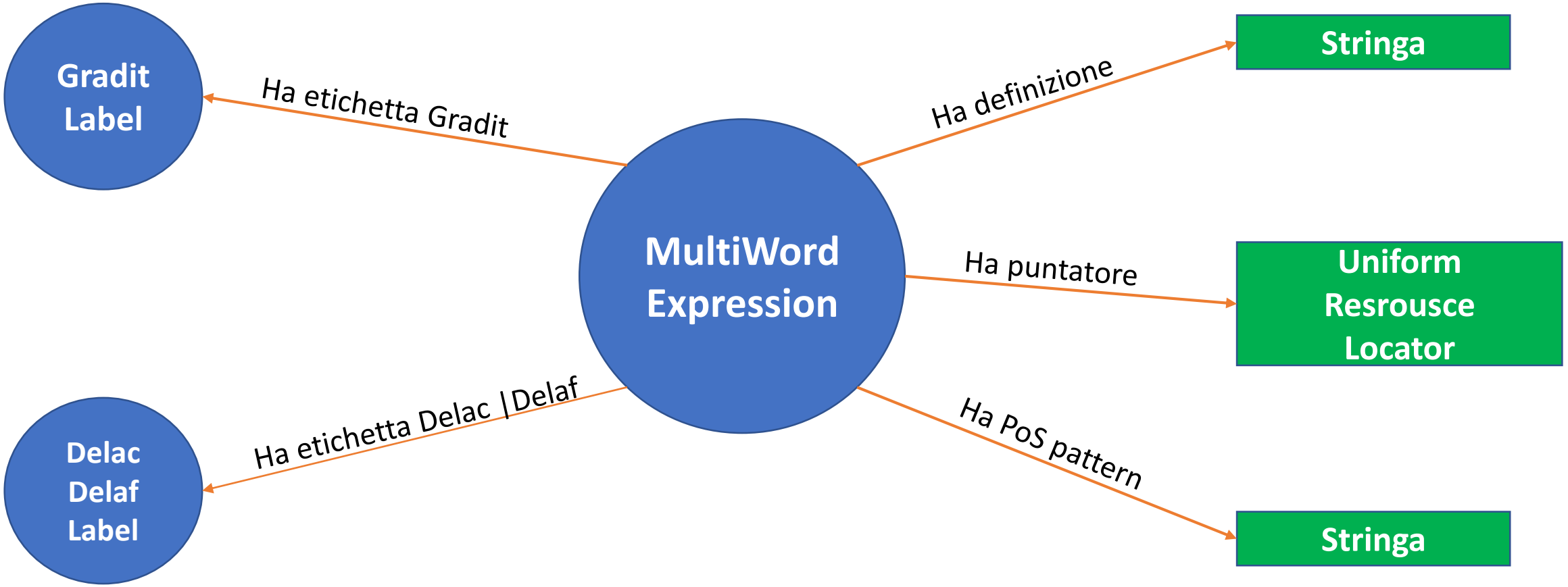
Dizionario Ontologico

- Creazione di una risorsa lessicografica digitalizzata per le MWE
- Raccolta ed integrazione di diverse fonti informative
- Paradigma Open Data

Un'ontologia è una descrizione formale ed esplicita dei concetti (classi) appartenenti ad un dominio della conoscenza, delle caratteristiche (proprietà) attribuibili ad ogni concetto e delle relative restrizioni. Quando un'ontologia è popolata dai dati, essa assume anche la funzione di Knowledge Base.

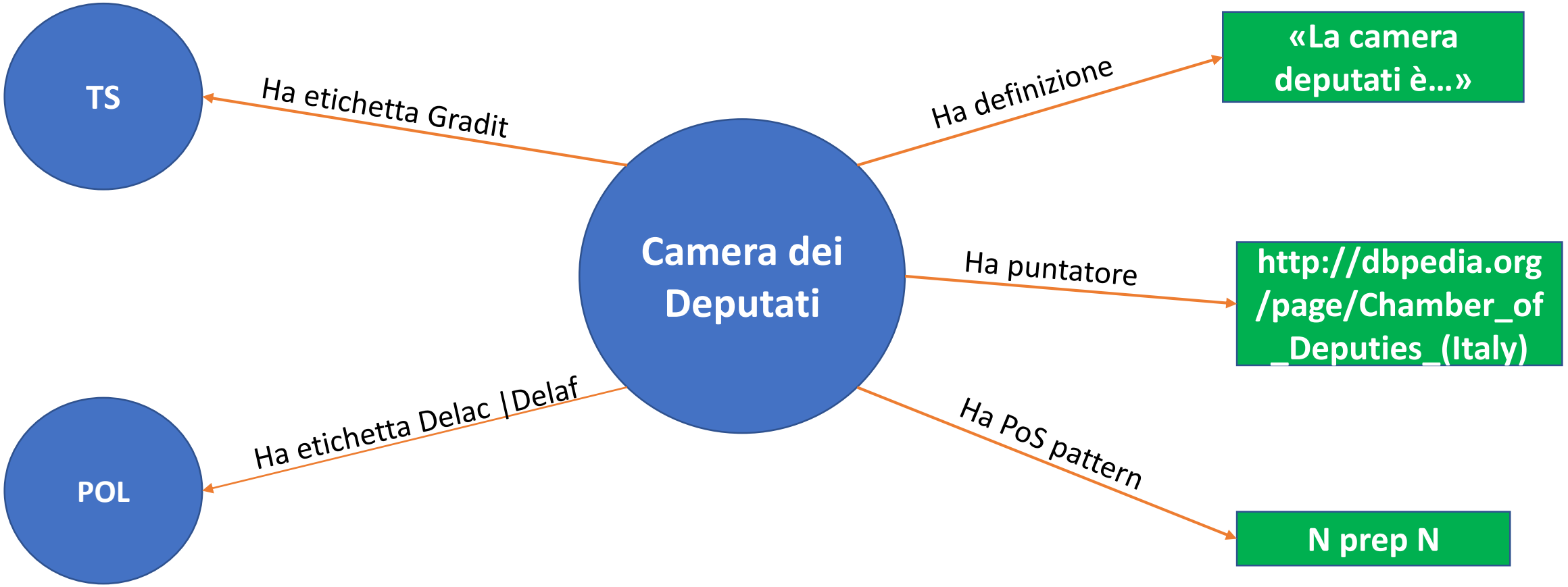
DiOnEPiT

Dizionario Ontologico: schema



DiOnEPiT

Dizionario Ontologico: schema



DiOnEPiT

PoliSdict (Trotta, Stingo *et al.*, 2018)

- Creazione di una risorsa lessicografica digitalizzata per le MWE
- Raccolta ed integrazione di diverse fonti informative
- Paradigma Open Data

Dizionario elettronico in lingua italiana composto da *espressioni poliriematiche* occorrenti nel *parlato spontaneo* estratte a partire dal PoliModalCorpus (Trotta, Elia, *et al.*, 2018)

DiOnEPiT

PoliSdict (Trotta, Stingo *et al.*, 2018)

*"The extracted MWEs were **manually verified using the GRADIT** (De Mauro, 2000). This operation has allowed us to identify 356 MWEs compared to 882 identified by DELAC- DELACF and to attribute to each compound expression the respective frequency label documented by the GRADIT"*

DiOnEPiT

Patrimonio terminologico Network Contacts

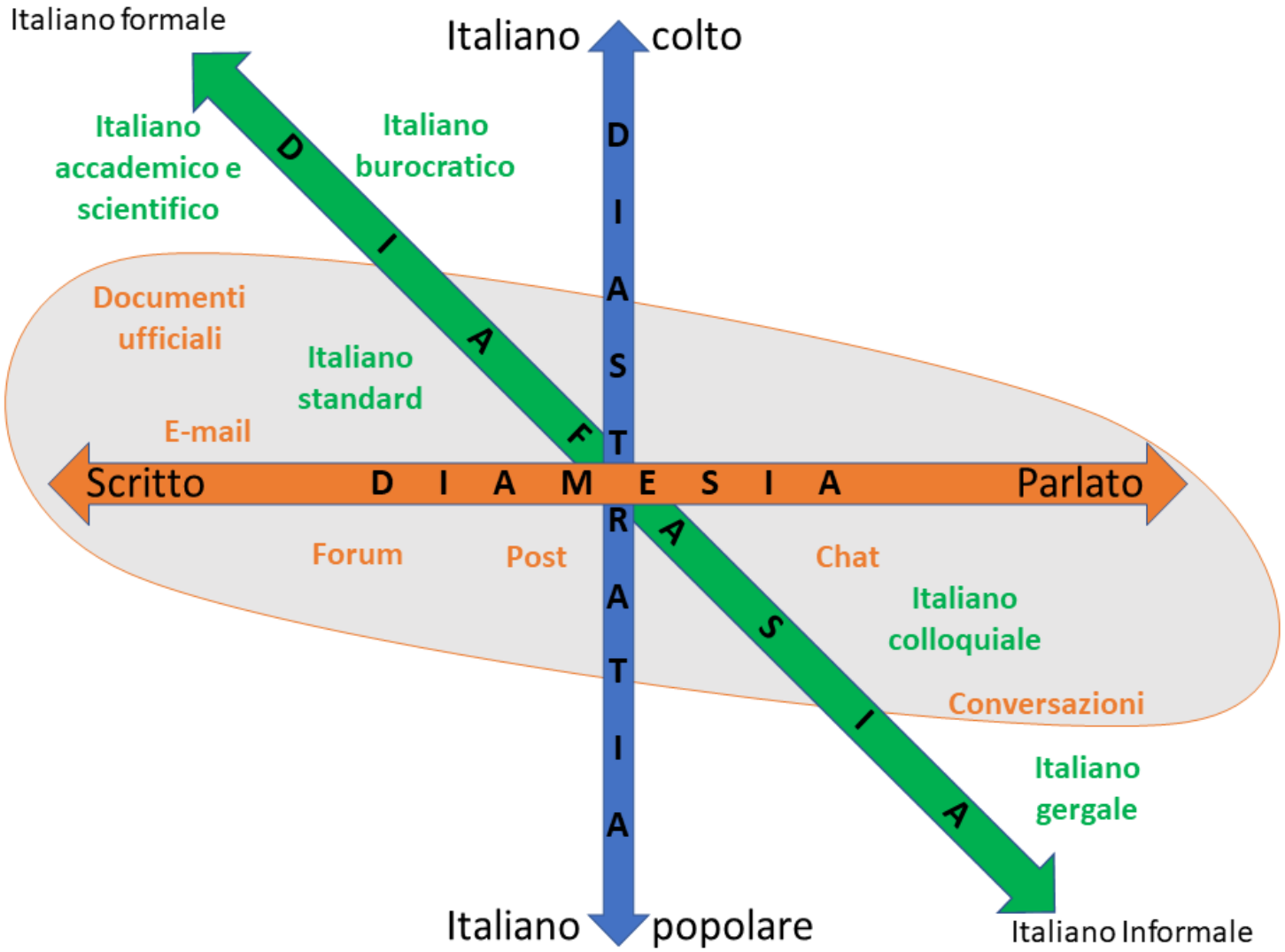
- Creazione di una risorsa lessicografica digitalizzata per le MWE
- Raccolta ed integrazione di diverse fonti informative
- Paradigma Open Data

Studio terminologico sul dominio TELCO (telecomunicazioni) per la creazione di risorse linguistiche ausiliarie a sistemi di Information Retrievial e Question Answering

- ❑ All Inclusive Summer Edition
- ❑ Scatto alla risposta
- ❑ Area Clienti

DiOnEPiT

Patrimonio terminologico Network Contacts

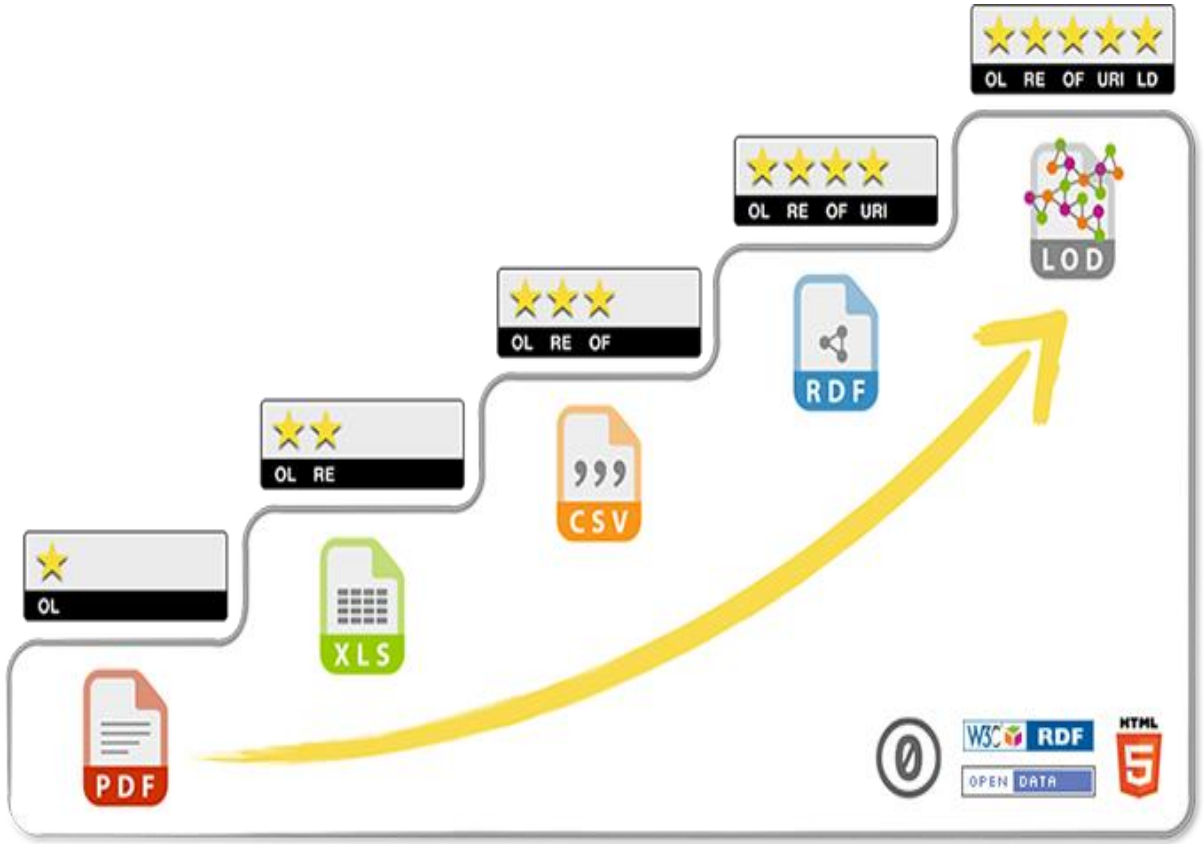


MWE estratte a partire da un corpus originale, costituito a supporto di soluzioni NLP richiamate da servizi informatici di customer care in area TELCO (telecomunicazioni)

DiOnEPiT

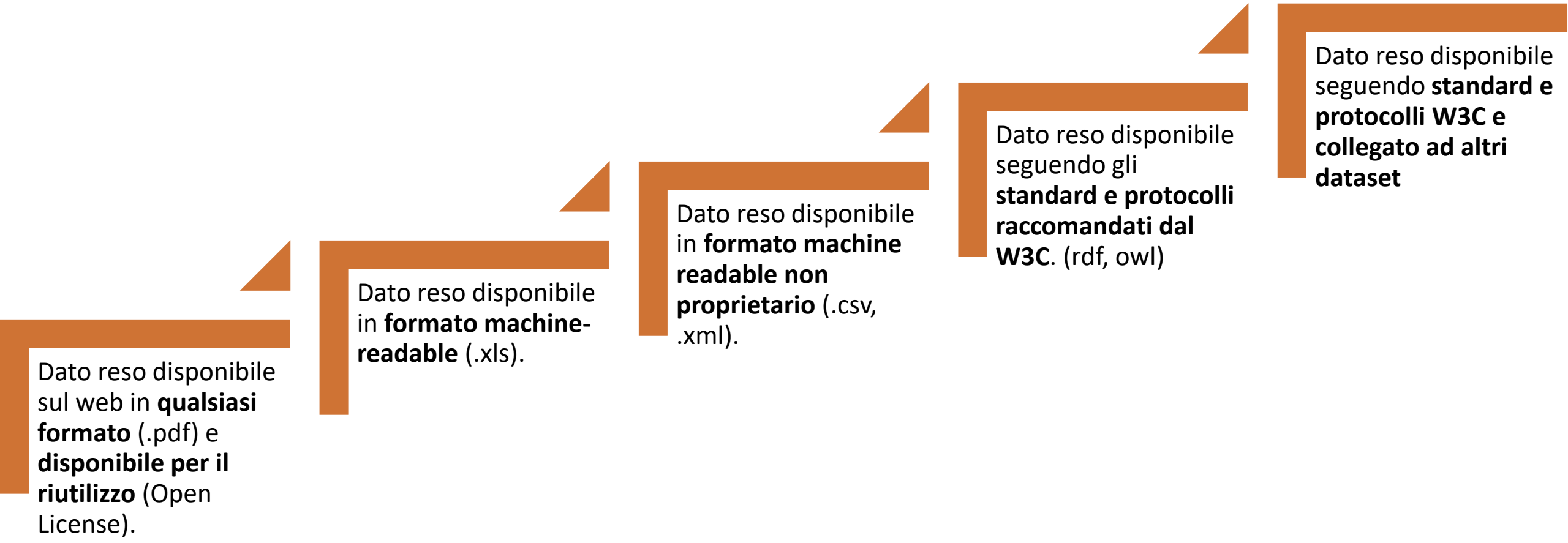
Scala d'apertura dei Web Data

- Creazione di una risorsa lessicografica digitalizzata per le MWE
- Raccolta ed integrazione di diverse fonti informative
- Paradigma Open Data



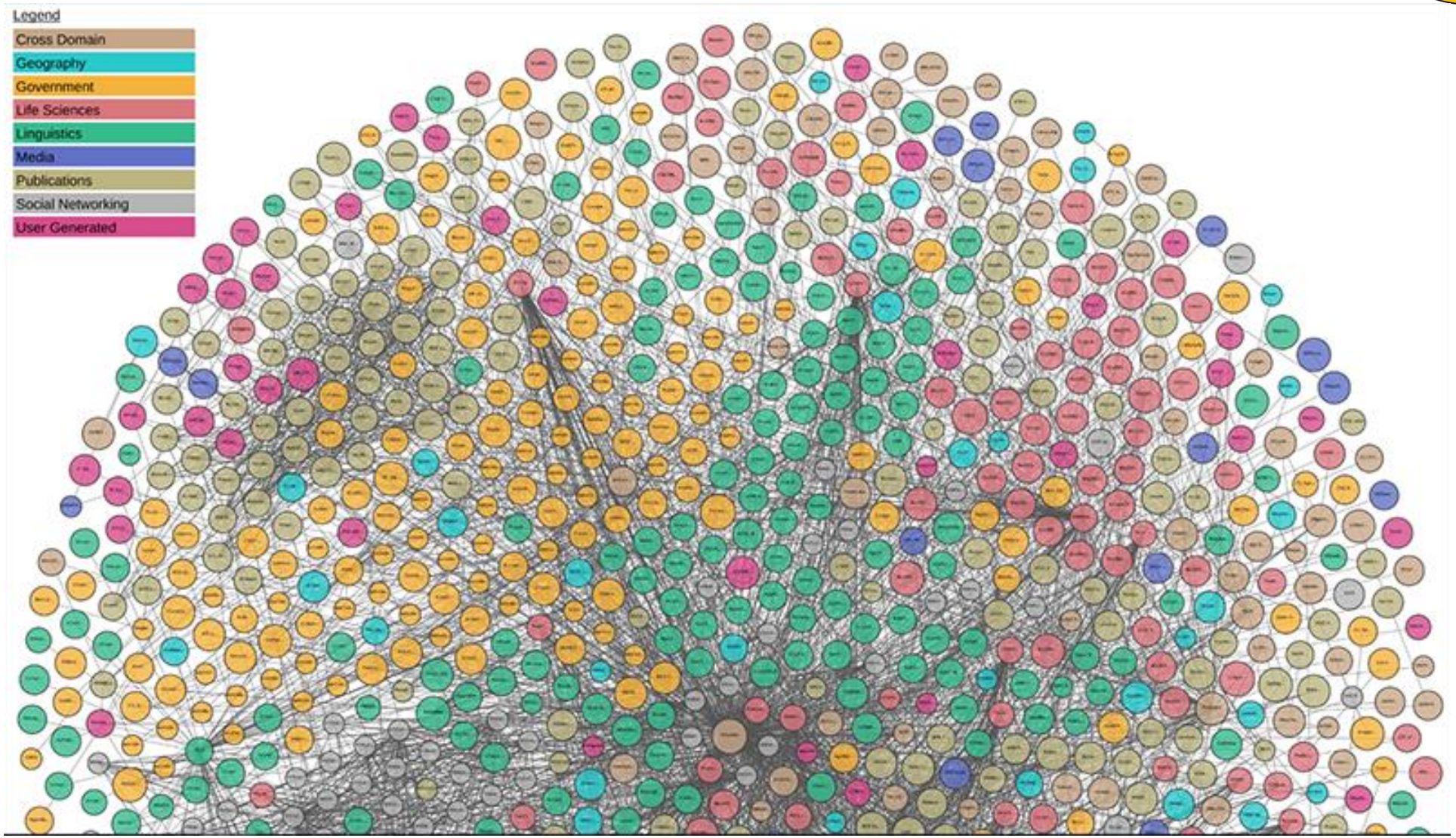
DiOnEPiT

Scala d'apertura dei Web Data



DiOnEPiT

Linked Open Data Cloud



DiOnEPiT

Prossimi step

Opportunità

- Collegamento Dataset di dominio (medico, gastronomico, etc.)
- Potenziamento modello metadescrittivo (acronimi, contesti d’uso, etc.)
- Approccio multilinguistico

Criticità

- Implementazione «time consuming»
- Manutenzione periodica

DiOnEpIt

Dizionario Ontologico delle Espressioni Polirematiche dell'Italiano

GRAZIE PER LA VOSTRA ATTENZIONE