ContoMap

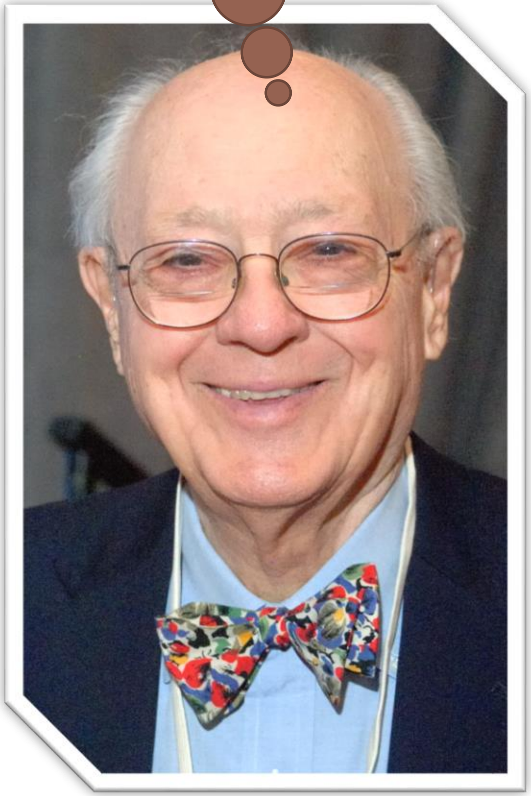
Suggesting Mappings from Relational Databases to Ontologies

Jackalopes Group:

Hamza Harkous, Julio Cesar dos Reis, Liang Beet,

Miguel Ceriani, Riccardo Porrini

Hey I'm **Charles Bachman**, and I invented relational databases!



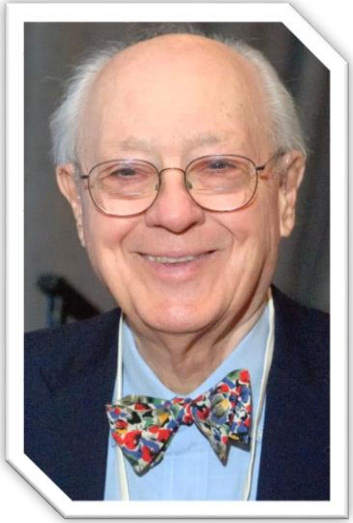
And I'm **Tim Berners-Lee**, and I invented the semantic web.

And we met in Kalamaki

The Semantic Web is a web of data, in some ways like a global **database**.

Via semantic web, we can improve heterogeneous database integration and ontology-based data access.

Aha! And via suggesting ontologies from databases, we can generate huge SW data.



PROBLEMS

Difficulty in selecting appropriate ontologies/vocabularies for reuse.

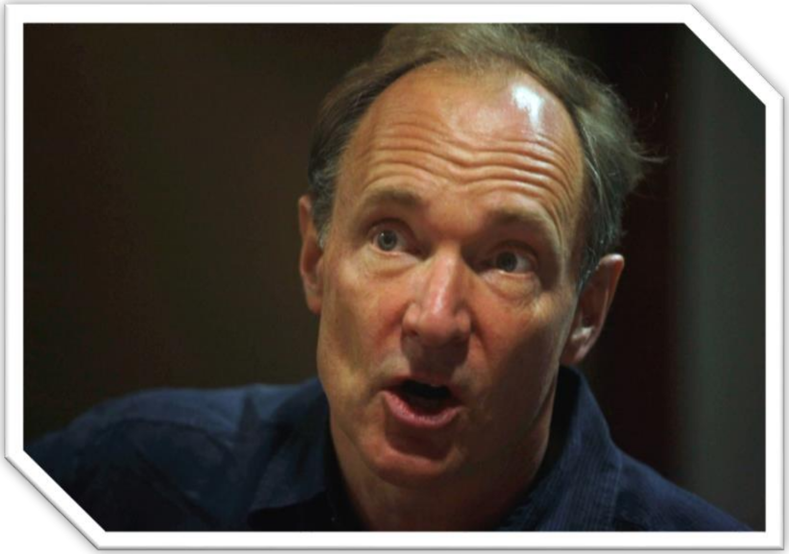
Need for (semi) automatically suggesting classes and properties suited for a given schema.

POSSIBLE APPROACHES

Creation of a new ontology from an existing database instance

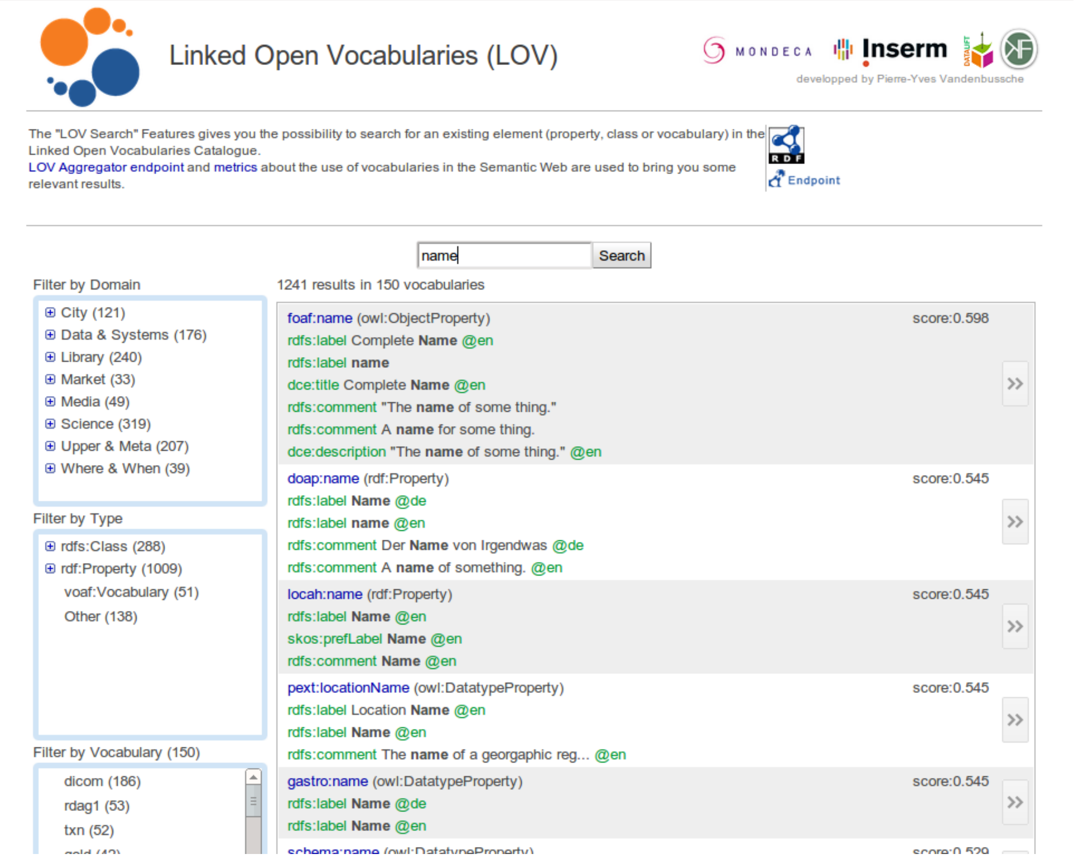
Discovery of mappings between an existing database schema and existing ontologies

We’ll go with your idea Charles!



MANUAL EXPLORATION IN LOV

- Searching for individual key words from the database table fields to retrieve vocabularies



Linked Open Vocabularies (LOV)

The "LOV Search" Features gives you the possibility to search for an existing element (property, class or vocabulary) in the Linked Open Vocabularies Catalogue.
LOV Aggregator endpoint and metrics about the use of vocabularies in the Semantic Web are used to bring you some relevant results.

name Search

Filter by Domain

- ⊞ City (121)
- ⊞ Data & Systems (176)
- ⊞ Library (240)
- ⊞ Market (33)
- ⊞ Media (49)
- ⊞ Science (319)
- ⊞ Upper & Meta (207)
- ⊞ Where & When (39)

Filter by Type

- ⊞ rdfs:Class (288)
- ⊞ rdf:Property (1009)
- voaf:Vocabulary (51)
- Other (138)

Filter by Vocabulary (150)

- dicom (186)
- rdag1 (53)
- txn (52)
- gold (42)

1241 results in 150 vocabularies

foaf:name (owl:ObjectProperty) score:0.598
rdfs:label Complete **Name** @en
rdfs:label **name**
dce:title Complete **Name** @en
rdfs:comment "The **name** of some thing."
rdfs:comment A **name** for some thing.
dce:description "The **name** of some thing." @en

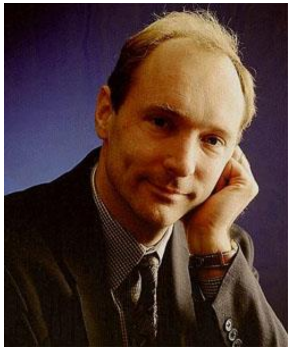
doap:name (rdf:Property) score:0.545
rdfs:label **Name** @de
rdfs:label **name** @en
rdfs:comment Der **Name** von Irgendwas @de
rdfs:comment A **name** of something. @en

locah:name (rdf:Property) score:0.545
rdfs:label **Name** @en
skos:prefLabel **Name** @en
rdfs:comment **Name** @en

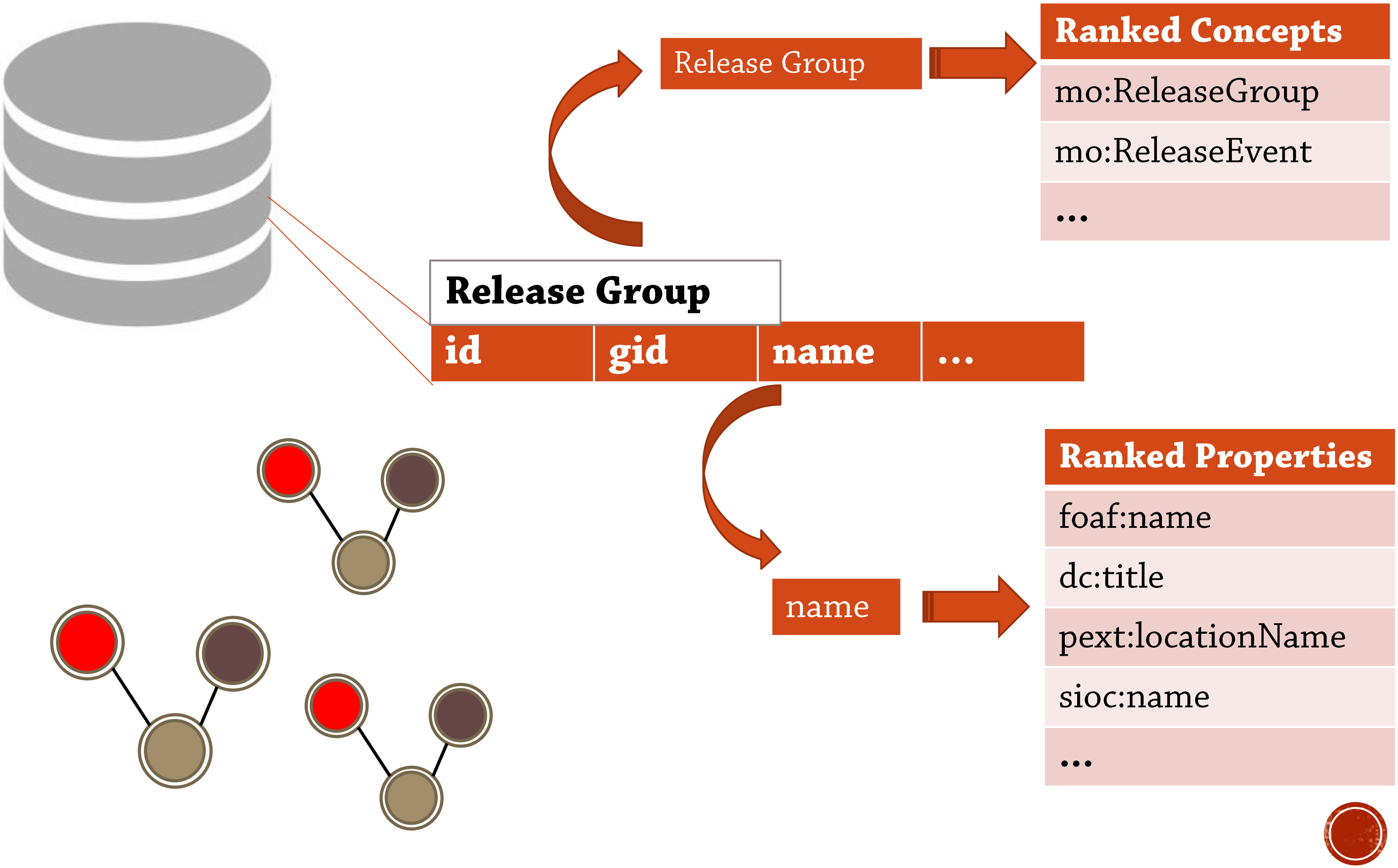
pext:locationName (owl:DatatypeProperty) score:0.545
rdfs:label Location **Name** @en
rdfs:label **Name** @en
rdfs:comment The **name** of a georgaphic reg... @en

gastro:name (owl:DatatypeProperty) score:0.545
rdfs:label **Name** @de
rdfs:label **Name** @en

schema:name (owl:DatatypeProperty) score:0.529



OBJECTIVE



Release Group

Ranked Concepts
mo:ReleaseGroup
mo:ReleaseEvent
…

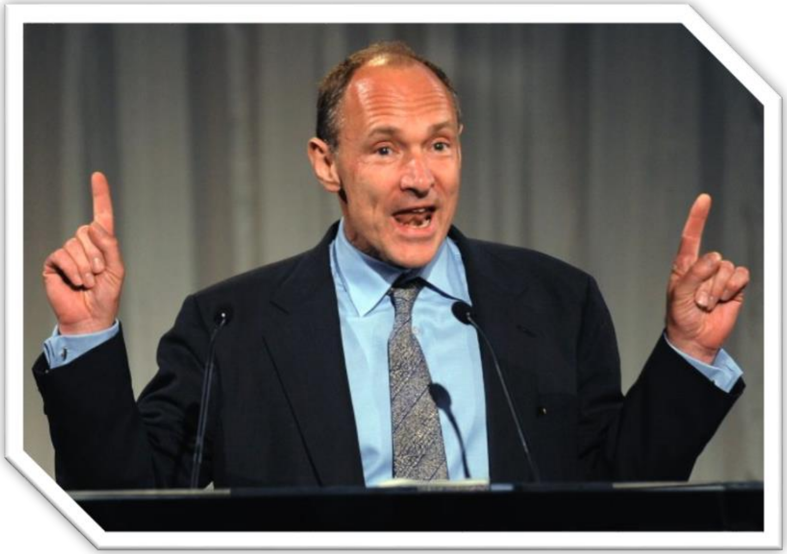
Release Group

id	gid	name	…

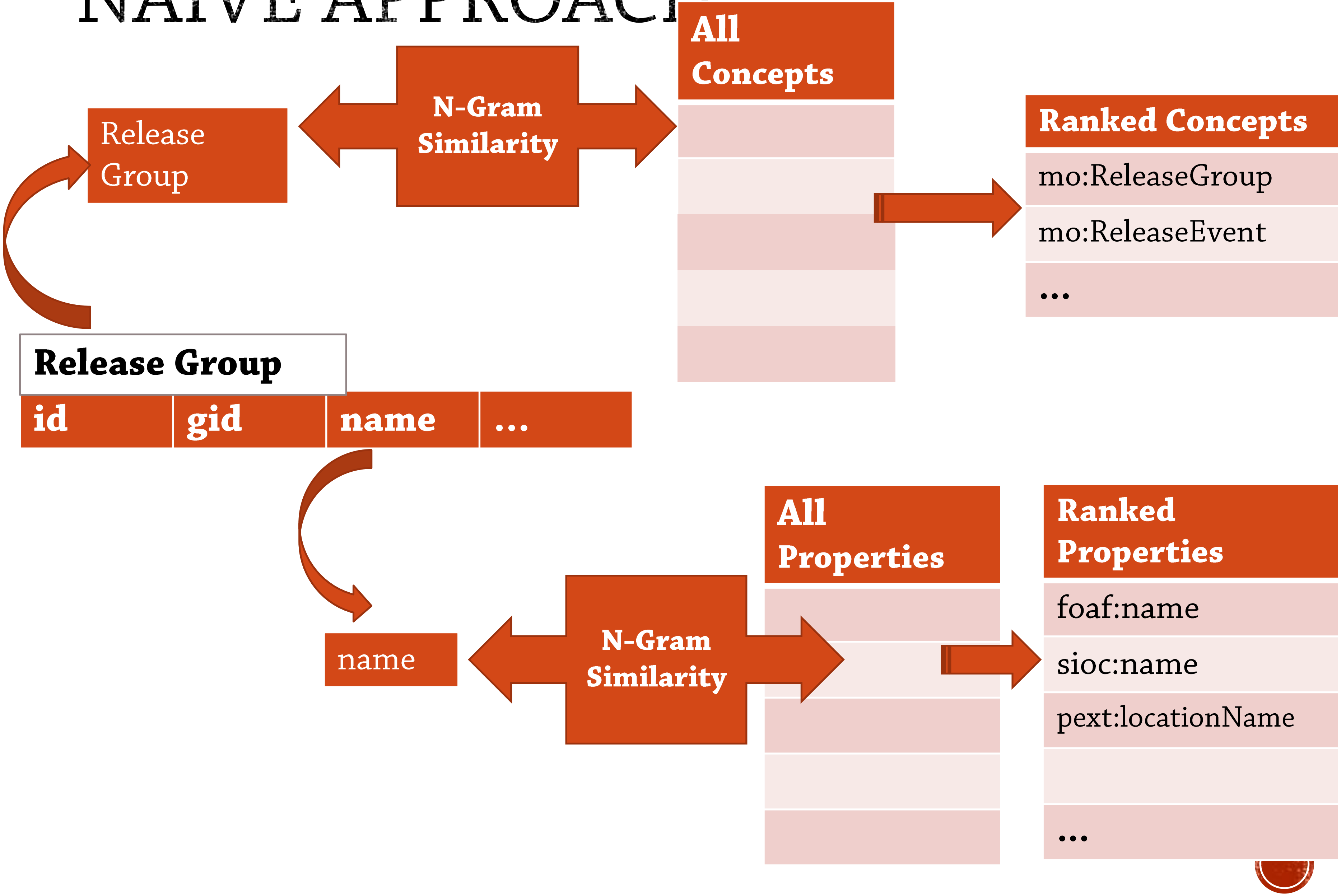
name

Ranked Properties
foaf:name
dc:title
pext:locationName
sioc:name
…

So here are our ideas!



NAÏVE APPROACH



Release Group

N-Gram Similarity

All Concepts

Ranked Concepts

mo:ReleaseGroup

mo:ReleaseEvent

…

Release Group

id	gid	name	…

name

N-Gram Similarity

All Properties

Ranked Properties

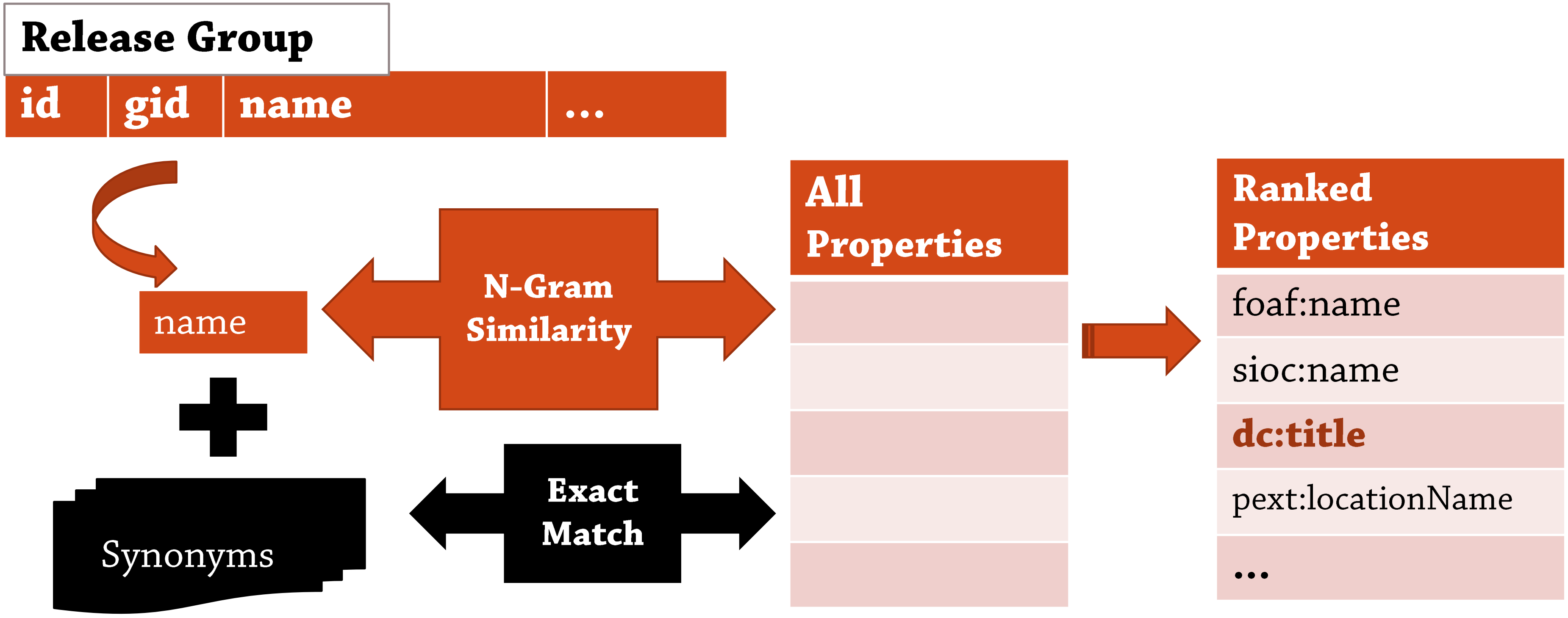
foaf:name

sioc:name

pext:locationName

…

SYNONYMS-ENHANCED APPROACH



Release Group

id	gid	name	…

name

Synonyms

N-Gram Similarity

Exact Match

All Properties

Ranked Properties

foaf:name

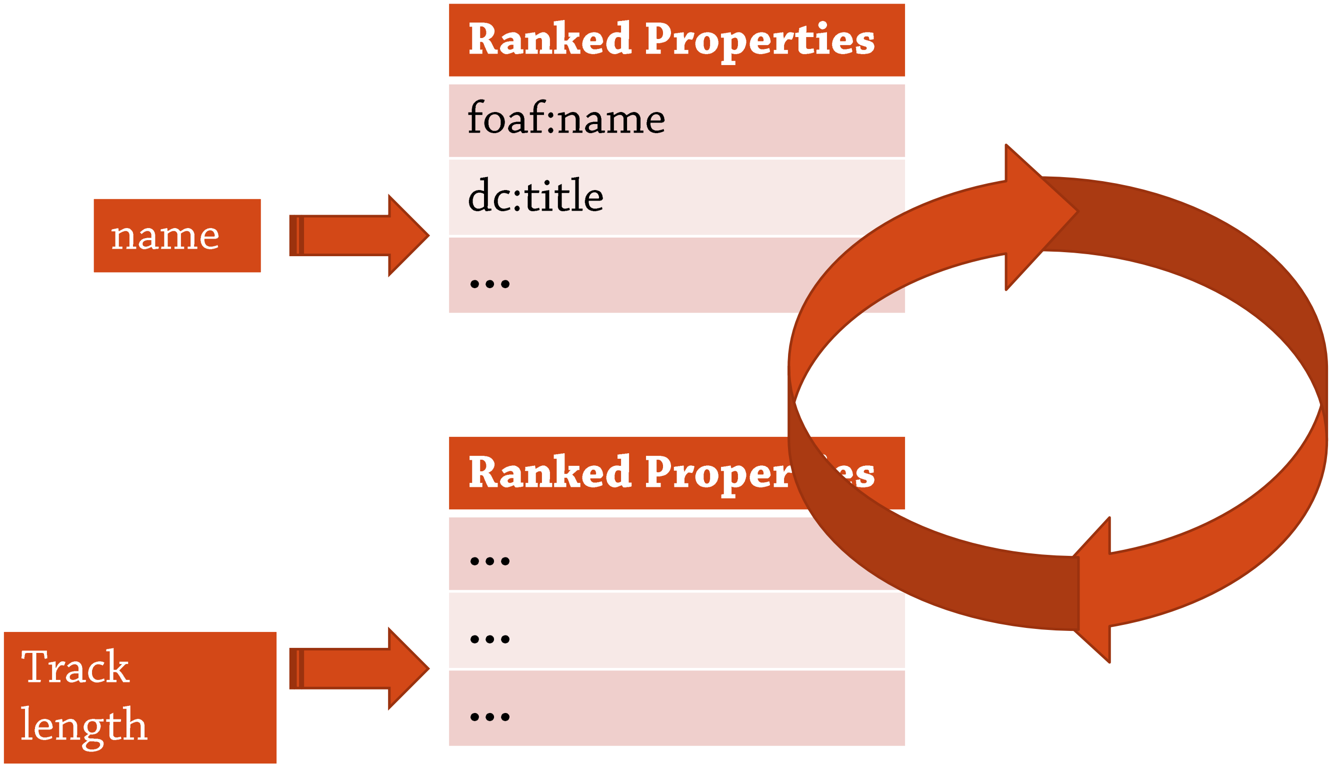
sioc:name

dc:title

pext:locationName

…

CONTEXT-BASED APPROACH - 1



name

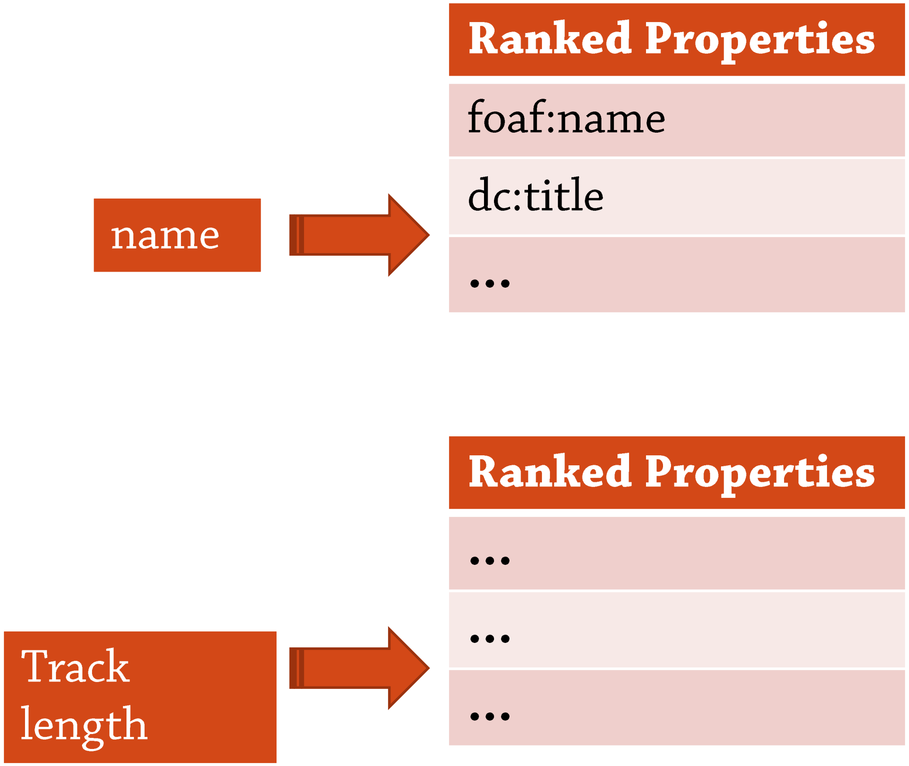
Ranked Properties
foaf:name
dc:title
…

Track length

Ranked Properties
…
…
…

Refine the property rankings, by trying to group them based on common domains.

CONTEXT-BASED APPROACH -2



name

Ranked Properties
foaf:name
dc:title
…

Track length

Ranked Properties
…
…
…

Exploit the real entities from datasets (e.g. DBpedia)

SUGGESTED EXPERIMENTS

- MusicBrainzcase study
- 44mappings used as gold standard
- Compare the different approaches (*e.g.*, naïve, synonym-enhanced).
- Calculate metrics of precision, recall and f-score.

IMPLEMENTATION

- Datasets:
  - Subset of ontologies from LOV
  - WordNet synsets and hyponims

- Technologies
  - Local graph store (Jena Fuseki)
  - Java standalone app
  - SQL -> acquiring database schema
  - SPARQL -> querying on LOV
  - Symmetrics Lib -> string similarity

FUTURE WORK

- Experimental evaluation via the MusicBrainz case study
- Refine the approach considering user's selections via interactive interfaces.

Thanks!