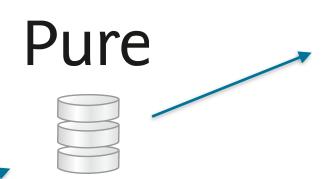Elsevier Research Intelligence

# Fueling VIVO with rich meta-data from a RIS system

*Thomas Vestdam (PhD)*

Head of Product Technology

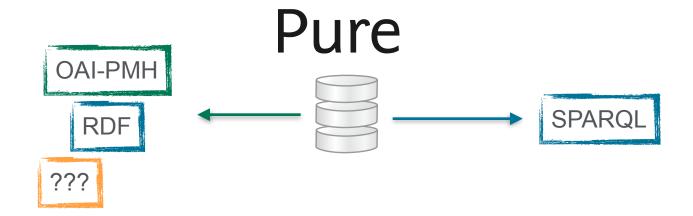Elsevier (Aalborg, Denmark)

T.Vestdam@elsevier.com

ELSEVIER

Empowering Knowledge

# The Desire

Pure

VIVO Network

- Global Network
- SPARQL

- Lots of data (fine-grained)
- Co-author network (local)
- Researcher Profiles
- …

- HR
- Finance
- Repositories
- Wos, Scopus, Pubmed, …
- …

# The Problem - harvesting vs querying

Pure

OAI-PMH

RDF

???

SPARQL

# Option 1: A SPARQL parser and query engine

Sounds like a great idea, and would be a great software engineering project …but…

- Interpreter (SPARQL Syntax)
- Build an intermediate representation
  - …based on the VIVO ontology
- Optimization
  - …rewrite to a Pure "friendly" intermediate representation
  - and, what about reasoning?
- Generate SQL queries
  - …based on the Pure "friendly" intermediate representation
- Generate a RDF/VIVO result

# Option 1: A SPARQL parser and query engine

Sounds like a great idea, and would be a great software engineering project …but…

- This is actually quite complicated
- …costly to build
- …costly to maintain
- …and, consequently error prone

# Option 2: VIVO "index"

Straight forward and "proven" solution - it works.

- Jena TDB + Fuseki

Relevant caveats

- Yet another "server" to administrate
  - And, pay for…
- Rebuild vs event driven updates

Mapping

- Fairly straight forward
- The VIVO 1.6 ontology is pretty mature
- Developers challenge was to get their head wrapped around the ontology

# Option 3: Your own aggregator

The DIY approach. Fast as h…
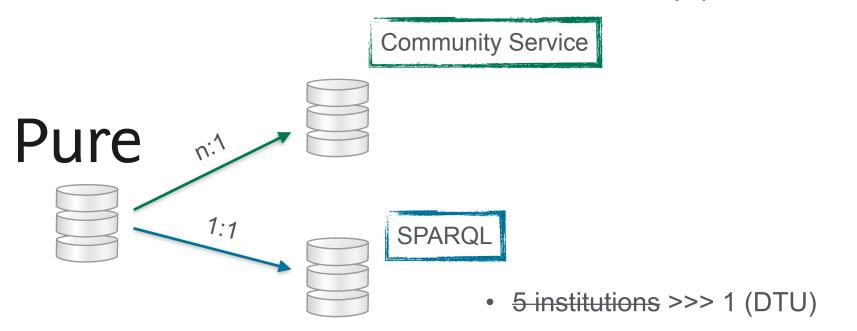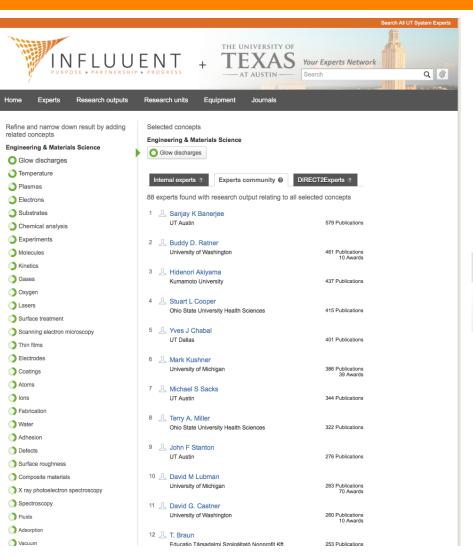- MogoDB + ElasticSearch

Relevant caveats
- Closed
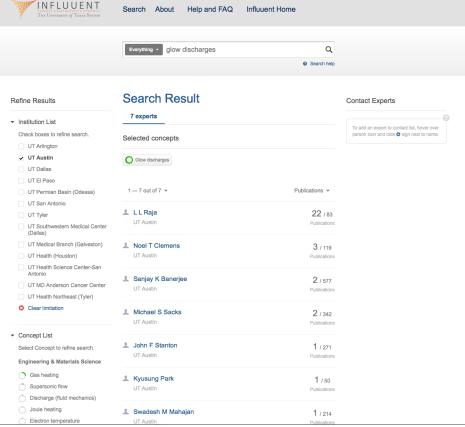- "Proprietary"
- … primarily because we are lazy developers

# What we did…

- 60 institutions
- 90.000 researchers
- 3.200.000 publications
- 115.000 grants
- 300 equipment

Community Service

Pure    n:1

SPARQL

1:1

- ~~5 institutions~~ >>> 1 (DTU)

# Towards a networking enabling standard

Programming Languages, 25%

Development Processes, 25%

Big Data, 25%

Software Architecture, 25%

*within Computer Science*

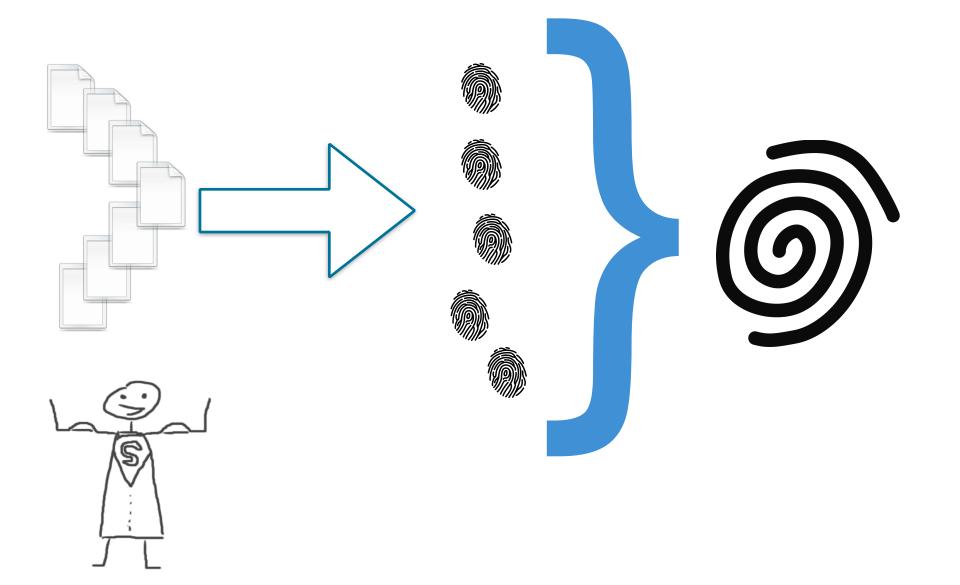Death Metal, 50%

Black Metal, 20%

Doom Metal, 20%

Heavy Metal, 10%

*within Music Genres*

… at this point in time

# Towards a networking enabling standard

# Towards a networking enabling standard

<Thesaurus>

<Concept, Weight>

.
.
.
.
.
.
.
.
.

n