

In the service of the Federation



Jerven Bolleman
Lead Developer
Swiss-Prot Group



Swiss Institute of
Bioinformatics

SERVICE in SPARQL 1.1

- Subquery on remote database

```
SELECT *  
WHERE {  
  ?thing <relation> ?otherThing.  
  SERVICE <http://example.org/sparql> {  
    ?otherThing <otherRelation> ?coolThing .  
  }  
}
```

Federated queries

- Theoretically terrible
- Need to do it anyway
- Run time of day(s) ok

Life Sciences → lots of RDF

Swiss + Lipids

PubChem

InterMine

BioModels

ICD-II
International
Statistical
Classification
of Diseases and
Related Health
Problems
10th Revision
2013
World Health
Organization

UniCarb-DB

DDBJ
DNA Data Bank of Japan

Reactome



WORLDWIDE
PDB
PROTEIN DATA BANK

DisGeNET

UniProt

nextprot

ü

PRO
Protein Ontology

MeSH

BETA
WIKIPATHWAYS
Pathways for the People

BioSamples

Allie

Rhea



ChEMBL



e!Ensembl

Expression
Atlas

Searching UniProt using Japanese

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>

PREFIX skos:<http://www.w3.org/2004/02/skos/core#>

PREFIX faldo:<http://biohackathon.org/resource/faldo#>

PREFIX up:<http://purl.uniprot.org/core/>

PREFIX allie:<http://purl.org/allie/ontology/201108#>

SELECT ?protein ?sfr1

WHERE {

SERVICE <http://data.allie.dbcls.jp/sparql/>{

 ?pc allie:hasShortFormRepresentationOf/rdfs:label ?sfrl ;

 allie:hasLongFormRepresentationOf ?lfr .

 ?lfr rdfs:label

 "(脳や癌細胞の代謝活性を測定するPET診断薬) フッ素18-フルオロデオキシグルコース"@ja .

}

 ?protein up:encodedBy/(skos:prefLabel|skos:altLabel) ?sfr1

}

SELECT

?protein ?transcript ?exon ?order

{

?protein rdfs:seeAlso

?transcript .

?transcript up:database

database:Ensembl .

SERVICE ensembl:sparql

{

?transcript obo:S0_translates_to ?peptide .

?peptide a ens_terms:protein .

?transcript obo:S0_has_part ?exon ;

sio:SI0_000974 ?orderedPart .

?orderedPart sio:SI0_000628 ?exon ;

sio:SI0_000300 ?order .

}

}



Community

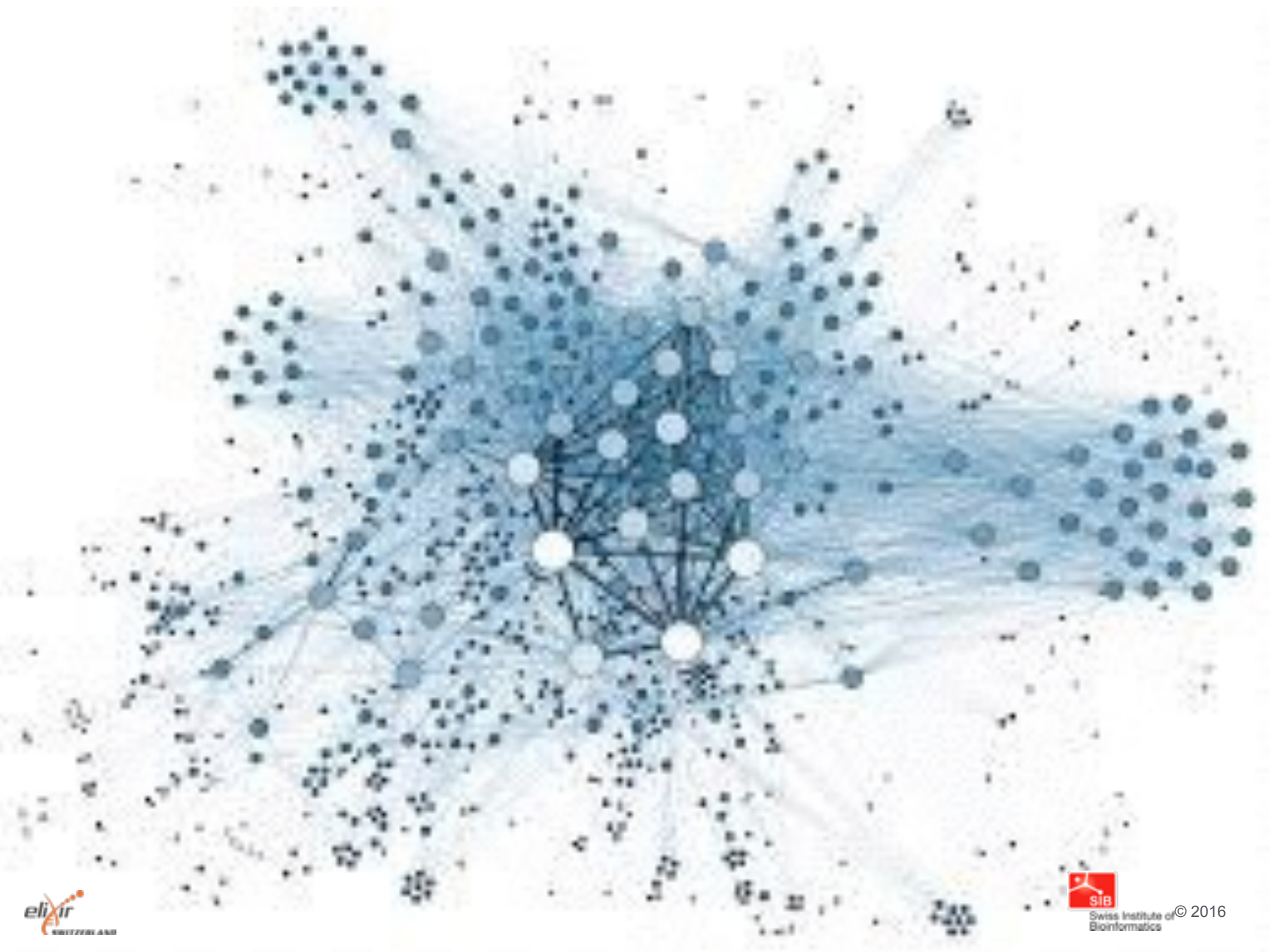


Infrastructure

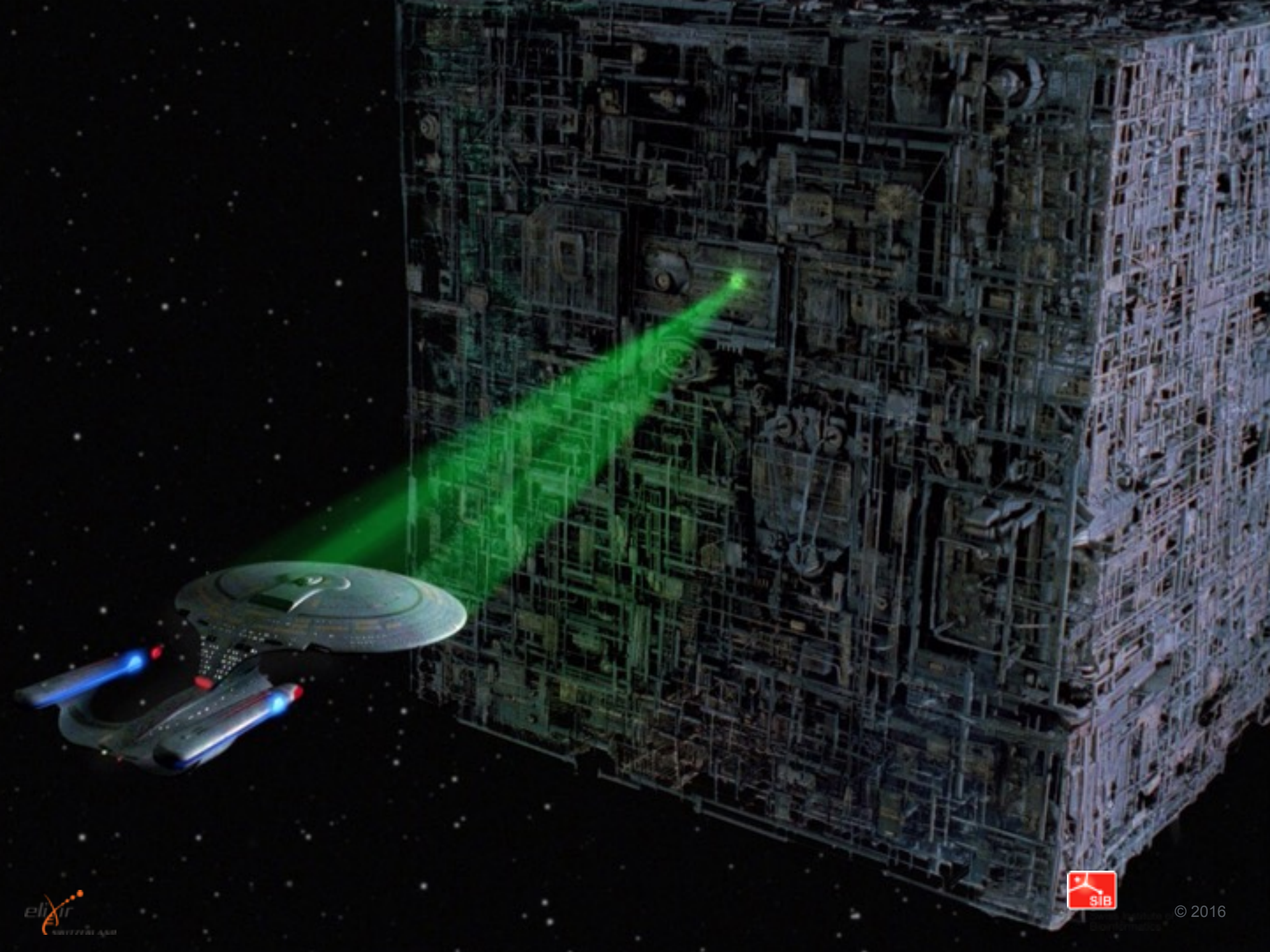


Connected





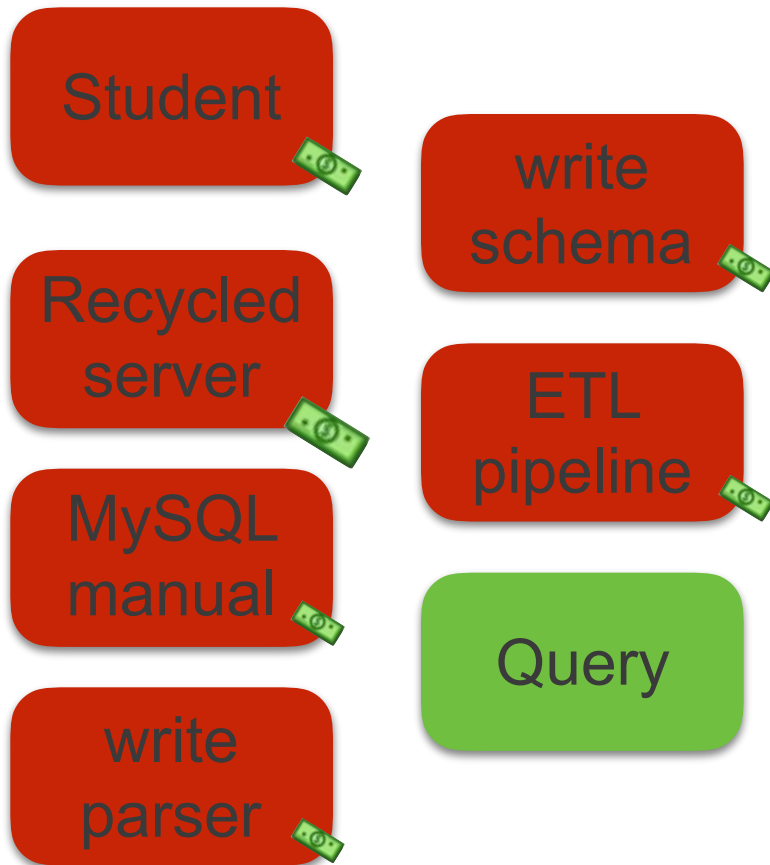




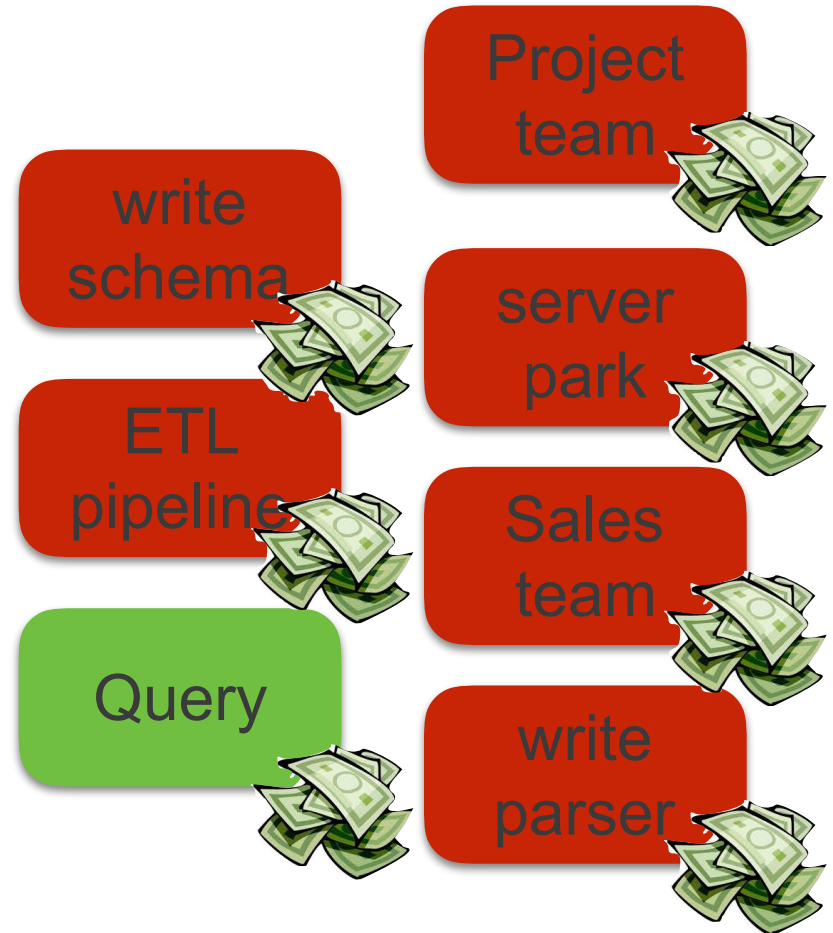
Why not ETL + Dataware house



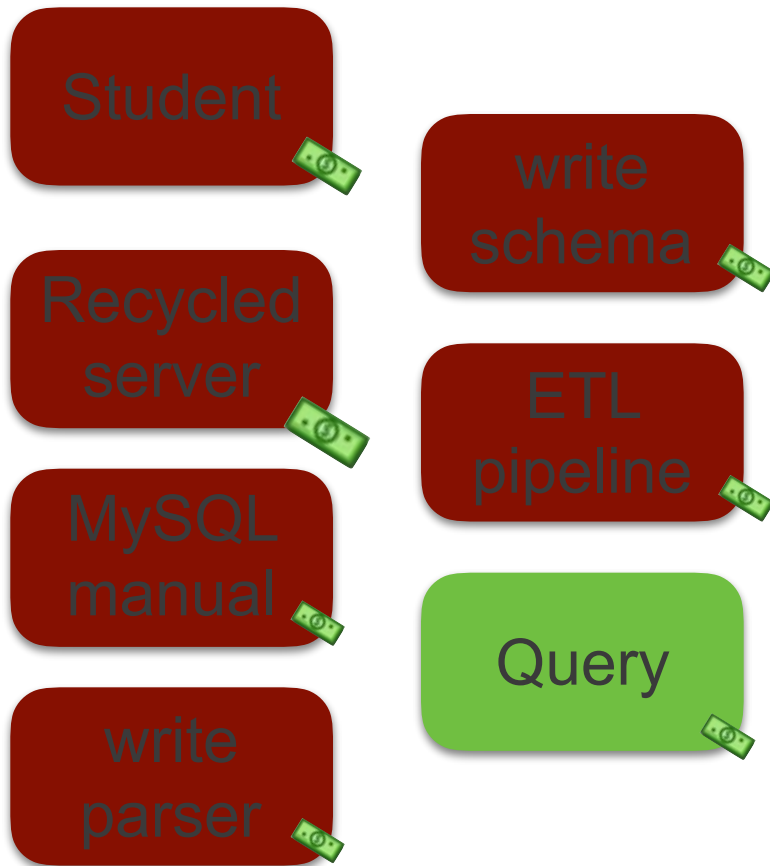
Academica



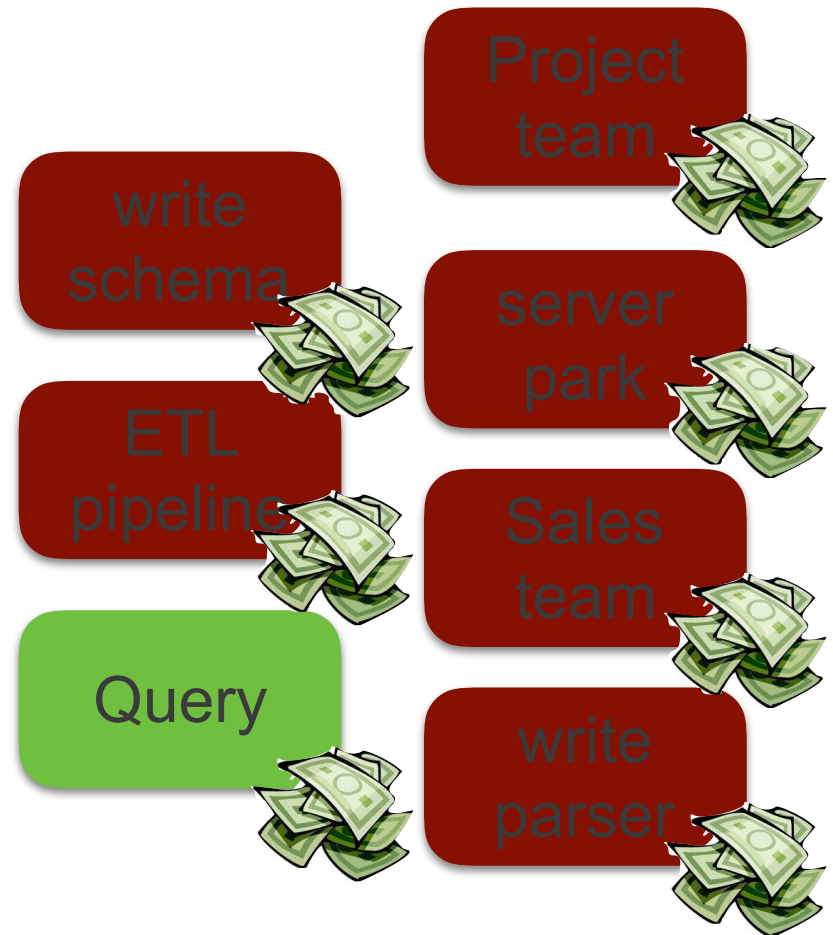
Pharma



Academica



Pharma



Why not ETL + Dataware house



Issues

- Naive implementations
- Passing on trust
- No statistics
- Cacheing

Naive implementation

```
SELECT ?gene WHERE {uniprotkb:P01000 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01001 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01002 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01003 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01004 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01005 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01006 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01007 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01008 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01009 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01010 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01011 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01012 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01013 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01014 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01015 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01016 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01017 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01018 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01019 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01020 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01021 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01022 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01023 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01024 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01025 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01026 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01027 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01028 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01029 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01030 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01031 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01032 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01033 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01034 up:encodedBy ?gene}
SELECT ?gene WHERE {uniprotkb:P01035 up:encodedBy ?gene}
```

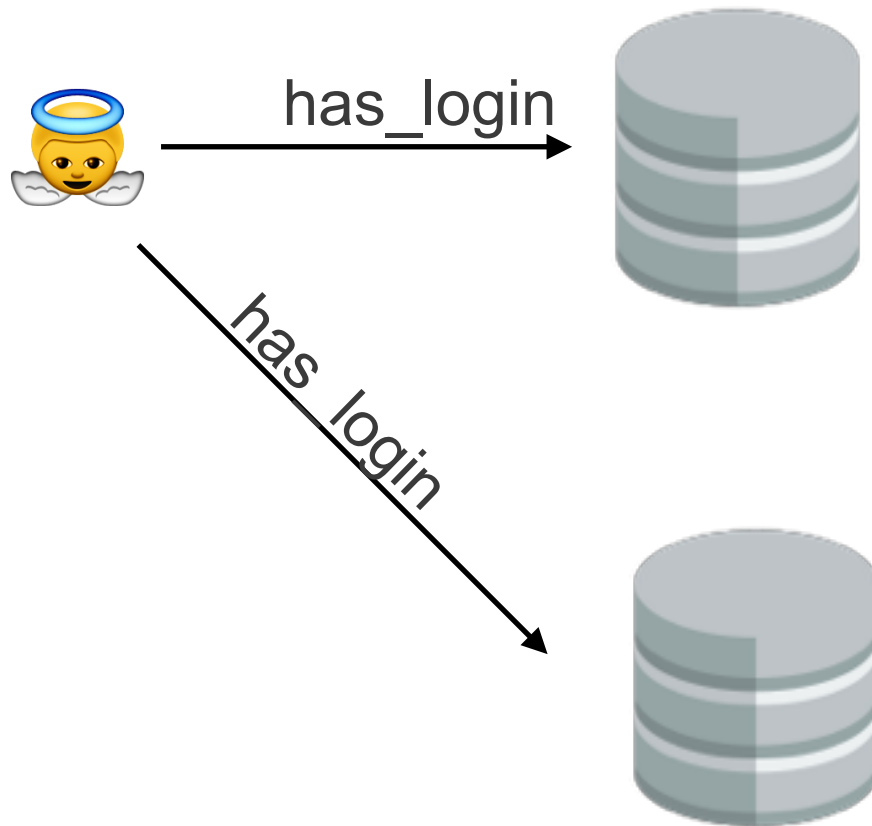
Intelligent implementation

```
SELECT ?gene WHERE {?protein up:encodedBy ?gene}
VALUES ?protein {uniprotkb:P01000 uniprotkb:P01001
uniprotkb:P01002 uniprotkb:P01003 uniprotkb:P01004 uniprotkb:P01005 uniprotkb:P01006 uniprotkb:P01007 uniprotkb:P01008
uniprotkb:P01009 uniprotkb:P01010 uniprotkb:P01011 uniprotkb:P01012 uniprotkb:P01013 uniprotkb:P01014 uniprotkb:P01015
uniprotkb:P01016 uniprotkb:P01017 uniprotkb:P01018 uniprotkb:P01019 uniprotkb:P01020 uniprotkb:P01021 uniprotkb:P01022
uniprotkb:P01023 uniprotkb:P01024 uniprotkb:P01025 uniprotkb:P01026 uniprotkb:P01027 uniprotkb:P01028 uniprotkb:P01029
uniprotkb:P01030 uniprotkb:P01031 uniprotkb:P01032 uniprotkb:P01033 uniprotkb:P01034 uniprotkb:P01035 uniprotkb:P01036
uniprotkb:P01037 uniprotkb:P01038 uniprotkb:P01039 uniprotkb:P01040 uniprotkb:P01041 uniprotkb:P01042 uniprotkb:P01043
uniprotkb:P01044 uniprotkb:P01045 uniprotkb:P01046 uniprotkb:P01047 uniprotkb:P01048 uniprotkb:P01049 uniprotkb:P01050
uniprotkb:P01051 uniprotkb:P01052 uniprotkb:P01053 uniprotkb:P01054 uniprotkb:P01055 uniprotkb:P01056 uniprotkb:P01057
uniprotkb:P01058 uniprotkb:P01059 uniprotkb:P01060 uniprotkb:P01061 uniprotkb:P01062 uniprotkb:P01063 uniprotkb:P01064
uniprotkb:P01065 uniprotkb:P01066 uniprotkb:P01067 uniprotkb:P01068 uniprotkb:P01069 uniprotkb:P01070 uniprotkb:P01071
uniprotkb:P01072 uniprotkb:P01073 uniprotkb:P01074 uniprotkb:P01075 uniprotkb:P01076 uniprotkb:P01077 uniprotkb:P01078
uniprotkb:P01079 uniprotkb:P01080 uniprotkb:P01081 uniprotkb:P01082 uniprotkb:P01083 uniprotkb:P01084 uniprotkb:P01085
uniprotkb:P01086 uniprotkb:P01087 uniprotkb:P01088 uniprotkb:P01089 uniprotkb:P01090 uniprotkb:P01091 uniprotkb:P01092
uniprotkb:P01093 uniprotkb:P01094 uniprotkb:P01095 uniprotkb:P01096 uniprotkb:P01097 uniprotkb:P01098 uniprotkb:P01099
uniprotkb:P01100 uniprotkb:P01101 uniprotkb:P01102 uniprotkb:P01103 uniprotkb:P01104 uniprotkb:P01105 uniprotkb:P01106
uniprotkb:P01107 uniprotkb:P01108 uniprotkb:P01109 uniprotkb:P01110 uniprotkb:P01111 uniprotkb:P01112 uniprotkb:P01113
uniprotkb:P01114 uniprotkb:P01115 uniprotkb:P01116 uniprotkb:P01117 uniprotkb:P01118 uniprotkb:P01119 uniprotkb:P01120
uniprotkb:P01121 uniprotkb:P01122 uniprotkb:P01123 uniprotkb:P01124 uniprotkb:P01125 uniprotkb:P01126 uniprotkb:P01127
uniprotkb:P01128 uniprotkb:P01129 uniprotkb:P01130 uniprotkb:P01131 uniprotkb:P01132 uniprotkb:P01133 uniprotkb:P01134
uniprotkb:P01135 uniprotkb:P01136 uniprotkb:P01137 uniprotkb:P01138 uniprotkb:P01139 uniprotkb:P01140 uniprotkb:P01141
uniprotkb:P01142 uniprotkb:P01143 uniprotkb:P01144 uniprotkb:P01145 uniprotkb:P01146 uniprotkb:P01147 uniprotkb:P01148
uniprotkb:P01149 uniprotkb:P01150 uniprotkb:P01151 uniprotkb:P01152 uniprotkb:P01153 uniprotkb:P01154 uniprotkb:P01155
uniprotkb:P01156 uniprotkb:P01157 uniprotkb:P01158 uniprotkb:P01159 uniprotkb:P01160 uniprotkb:P01161 uniprotkb:P01162
uniprotkb:P01163 uniprotkb:P01164 uniprotkb:P01165 uniprotkb:P01166 uniprotkb:P01167 uniprotkb:P01168 uniprotkb:P01169
uniprotkb:P01170 uniprotkb:P01171 uniprotkb:P01172 uniprotkb:P01173 uniprotkb:P01174 uniprotkb:P01175 uniprotkb:P01176
uniprotkb:P01177 uniprotkb:P01178 uniprotkb:P01179 uniprotkb:P01180 uniprotkb:P01181 uniprotkb:P01182 uniprotkb:P01183
uniprotkb:P01184 uniprotkb:P01185 uniprotkb:P01186 uniprotkb:P01187 uniprotkb:P01188 uniprotkb:P01189 uniprotkb:P01190
uniprotkb:P01191 uniprotkb:P01192 uniprotkb:P01193 uniprotkb:P01194 uniprotkb:P01195 uniprotkb:P01196 uniprotkb:P01197
uniprotkb:P01198 uniprotkb:P01199 uniprotkb:P01200 )
```

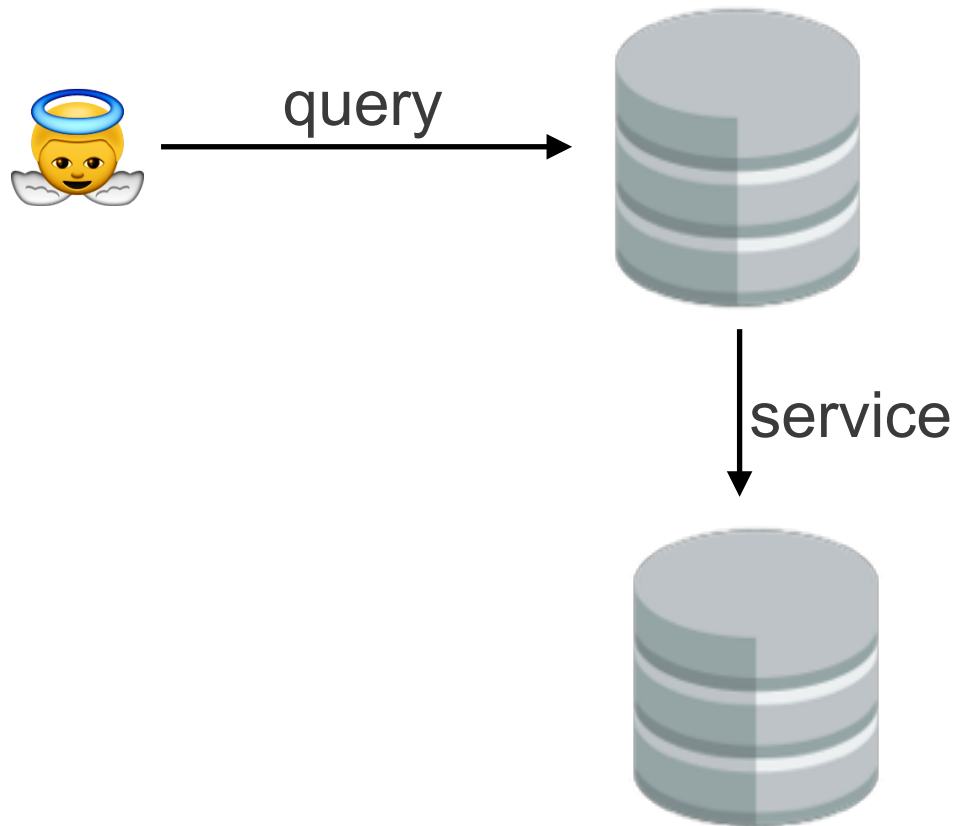
SPARQL 1.2 Protocol suggestion

- Compressed form-data?
 - ?query=SELECT ?x
 - ?compressed_query=U0VMRUNUID94IOKApG%3D%3D

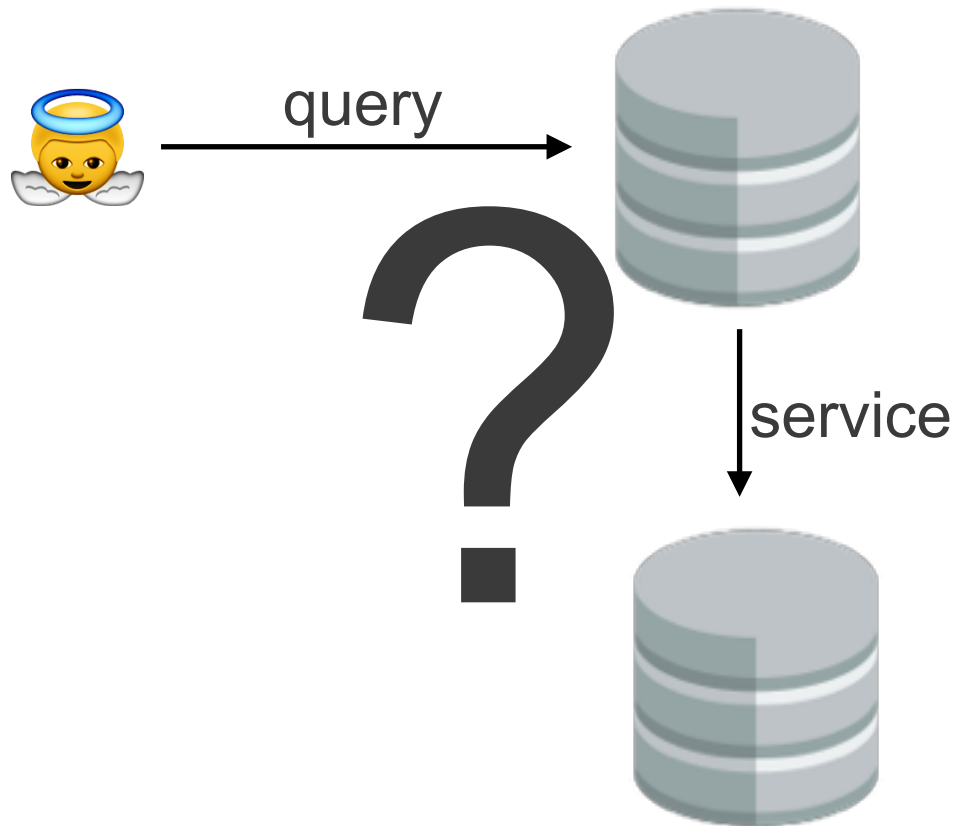
Passing on trust



Passing on trust



Passing on trust



No Statistics

- Service description + VOID

```
<has_predicate_x> a void:Property_Partion ;  
  void:property <predicate_x> ;  
  void:triple 1'000'000 .
```

```
<has_predicate_y> a void:Property_Partion ;  
  void:property <predicate_y> ;  
  void:triple 1'000'000 .
```

Cacheing

- HTTP headers
- Service description
 - update frequency

Data Mining in the field

