Heritage Connector Technical Introduction

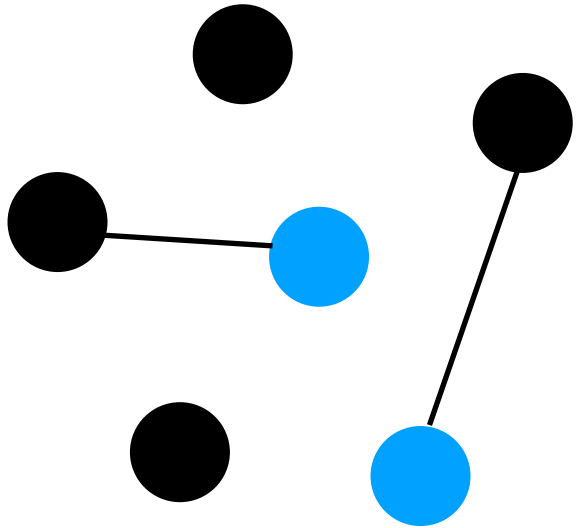
Jamie Unwin & Kalyan Dutia

22nd March 2021

“TRANSFORMING TEXT INTO DATA TO EXTRACT MEANING AND MAKE CONNECTIONS”

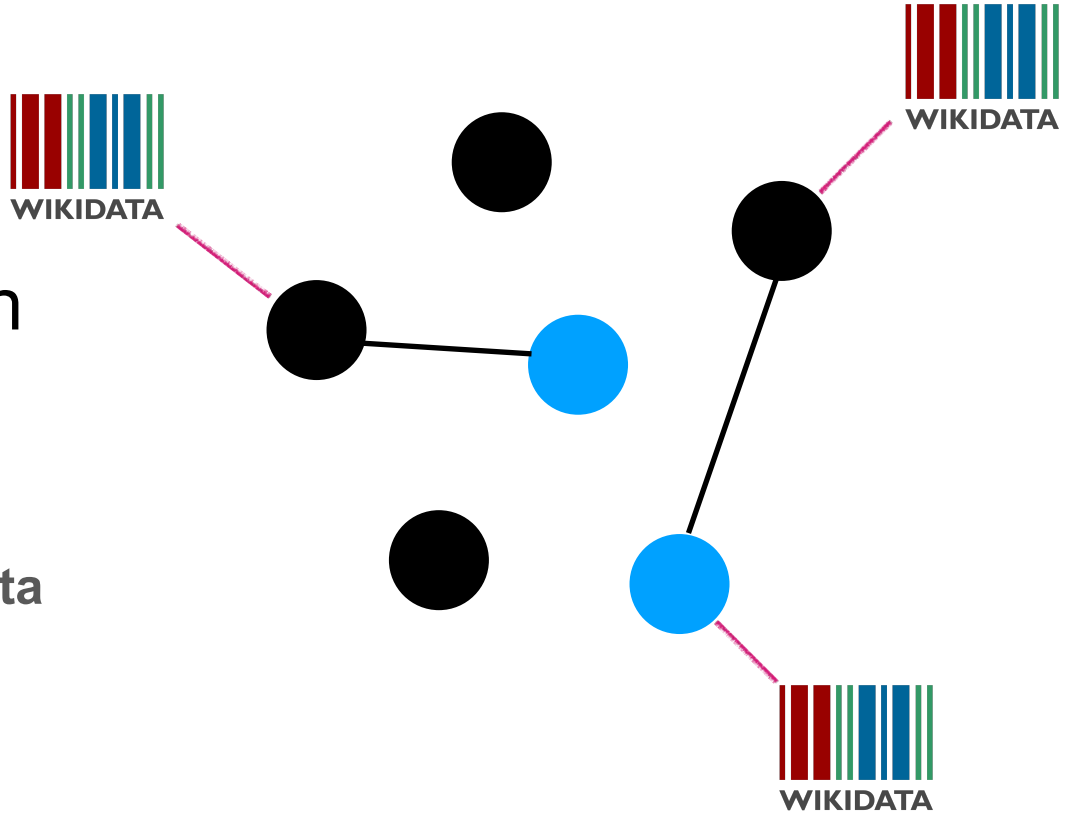
This is our collection now..

Small islands of thin data



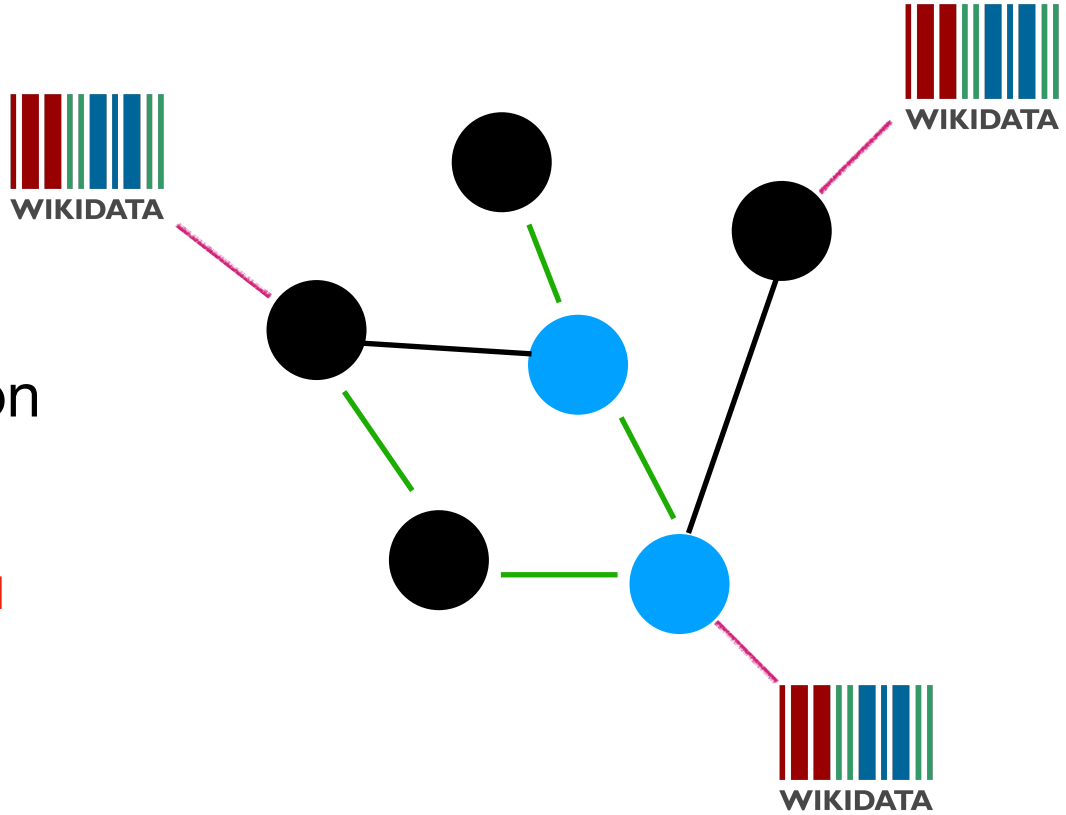
This is our collection connected to Wikidata..

Small islands of connected data



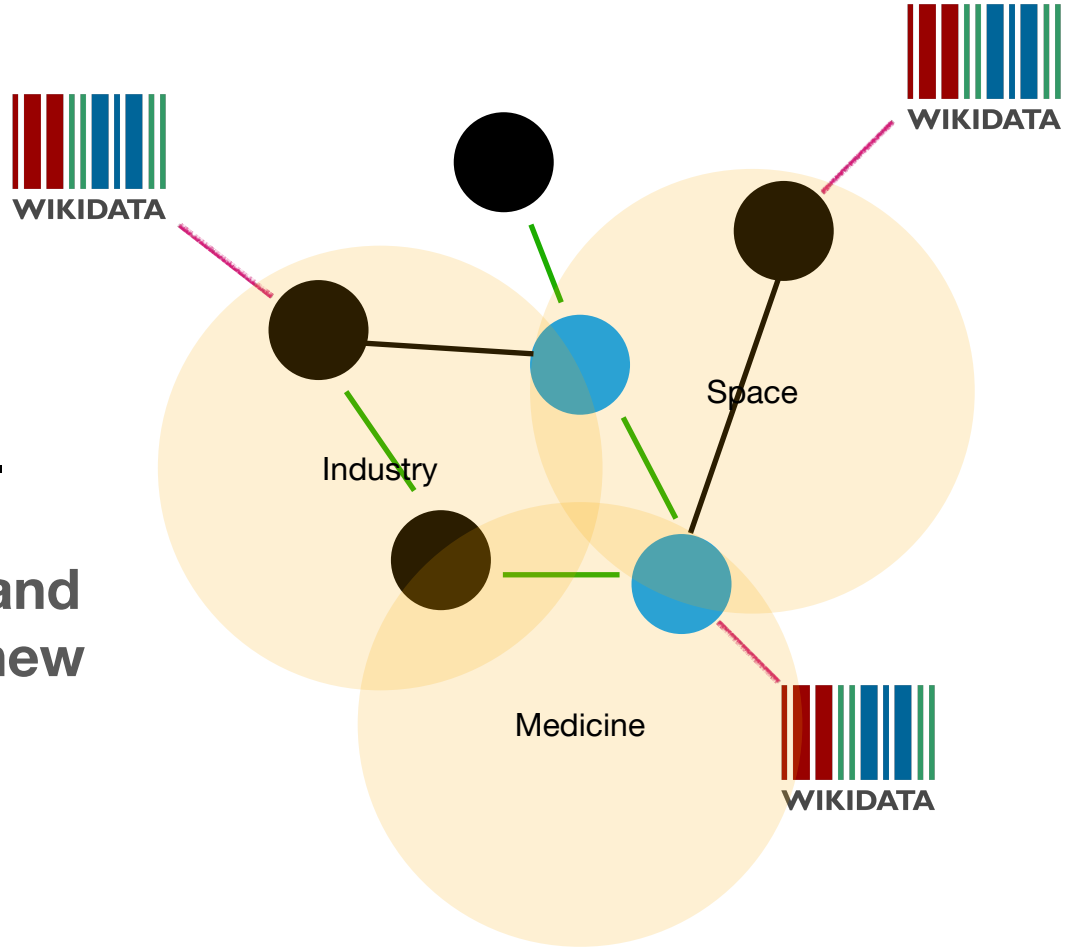
This is our collection interlinked via information extraction techniques..

Small islands of connected and interlinked data



This is our collection with new groupings..

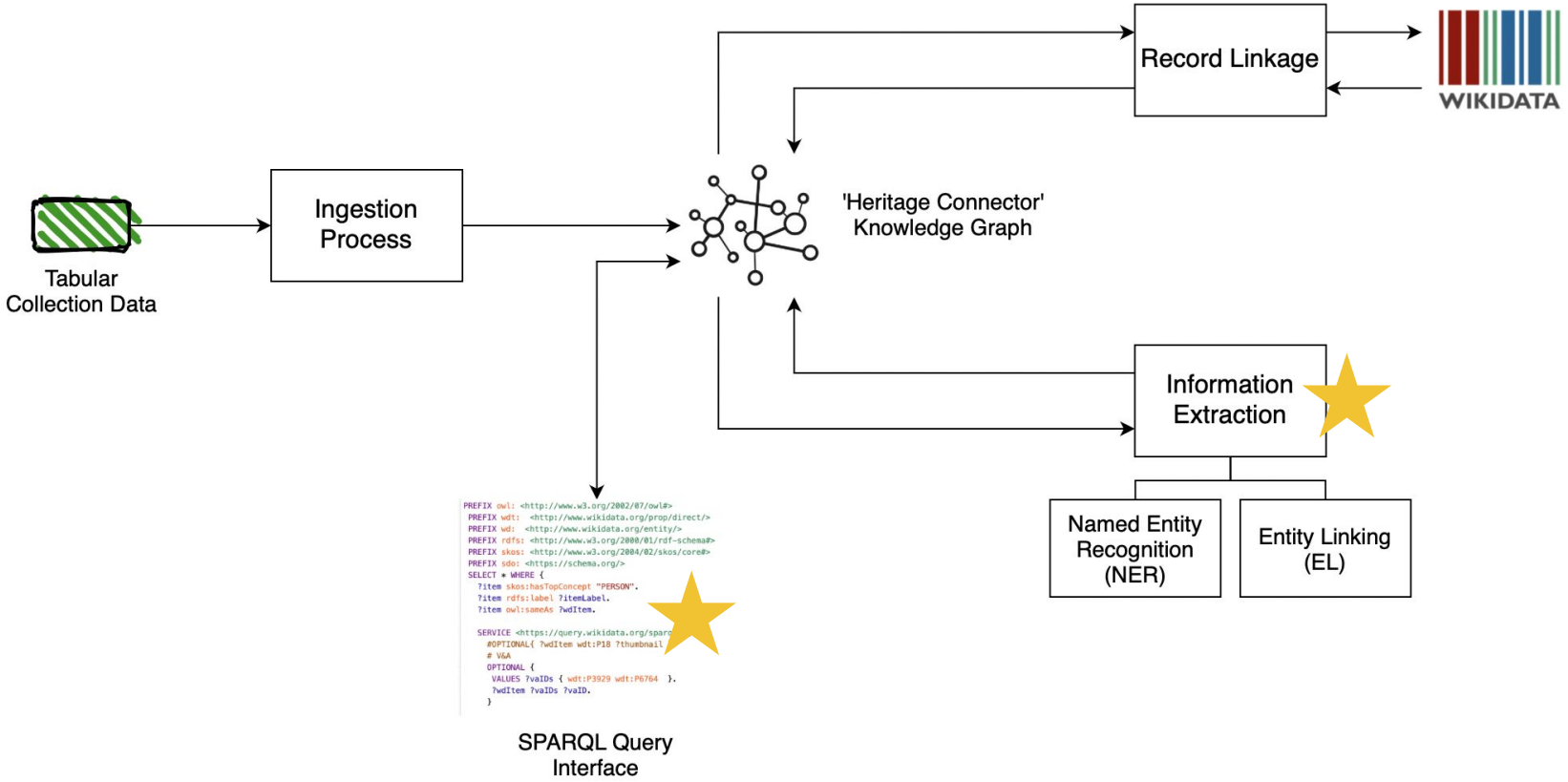
Small islands of connected and interlinked data exposing new groupings



Why are we doing this?

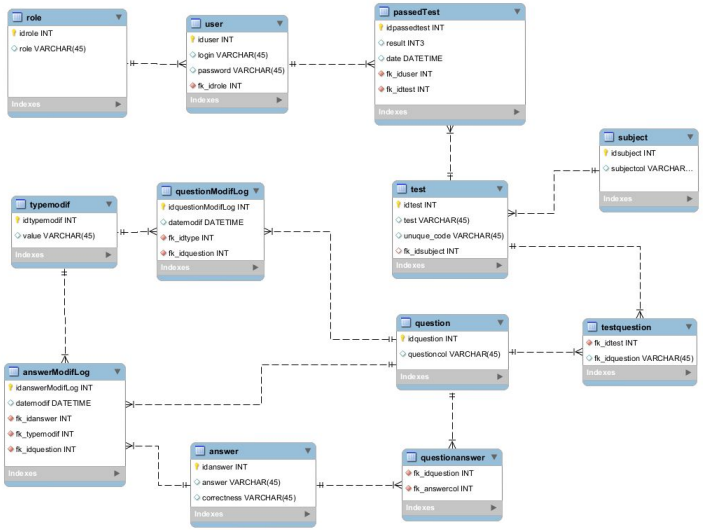
- Human resources are limited, especially expert ones
- Introduce new topics and themes not inherent in our current data
- Enrich specifics types of records/objects with additional data
- Stop our record pages becoming ‘dead ends’ to users.
- Along with all the other obvious benefits of LoD :-)

How are we doing this?


Record Linkage
WIKIDATA
Tabular
Collection Data
Ingestion
Process
'Heritage Connector'
Knowledge Graph
Information
Extraction
Named Entity
Recognition
(NER)
Entity Linking
(EL)
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX sdo: <https://schema.org/>
SELECT * WHERE {
?item skos:hasTopConcept "PERSON".
?item rdfs:label ?itemLabel.
?item owl:sameAs ?wdItem.
SERVICE <https://query.wikidata.org/sparq
#OPTIONAL{ ?wdItem wdt:P18 ?thumbnail
# V&A
OPTIONAL {
VALUES ?vaIDs { wdt:P3929 wdt:P6764 }.
?wdItem ?vaIDs ?vaID.
}
SPARQL Query
Interface

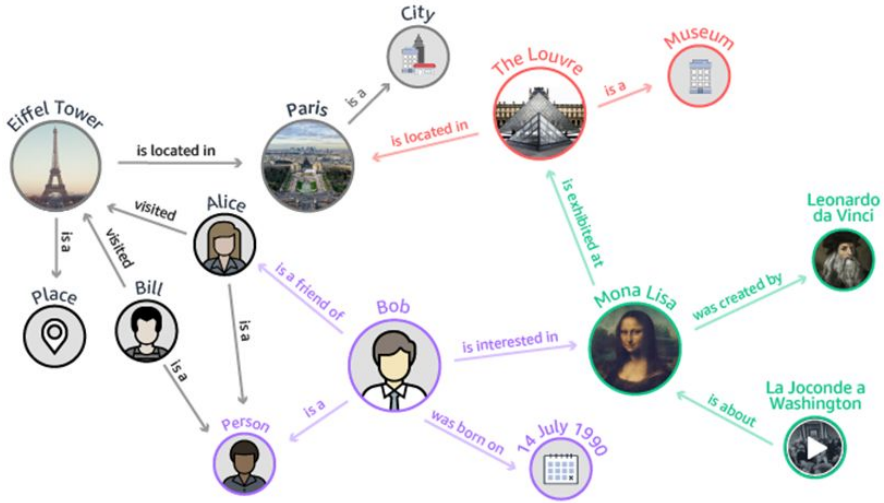
A quick primer on knowledge graphs and linked data

Traditional Relational Database



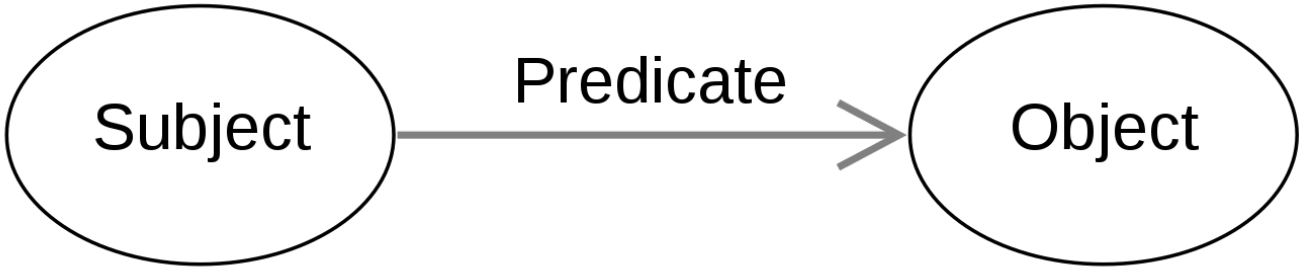
VS

Linked Data Knowledge Graph



What is Linked data?

An RDF Triple



All knowledge can be expressed with these three pieces of information

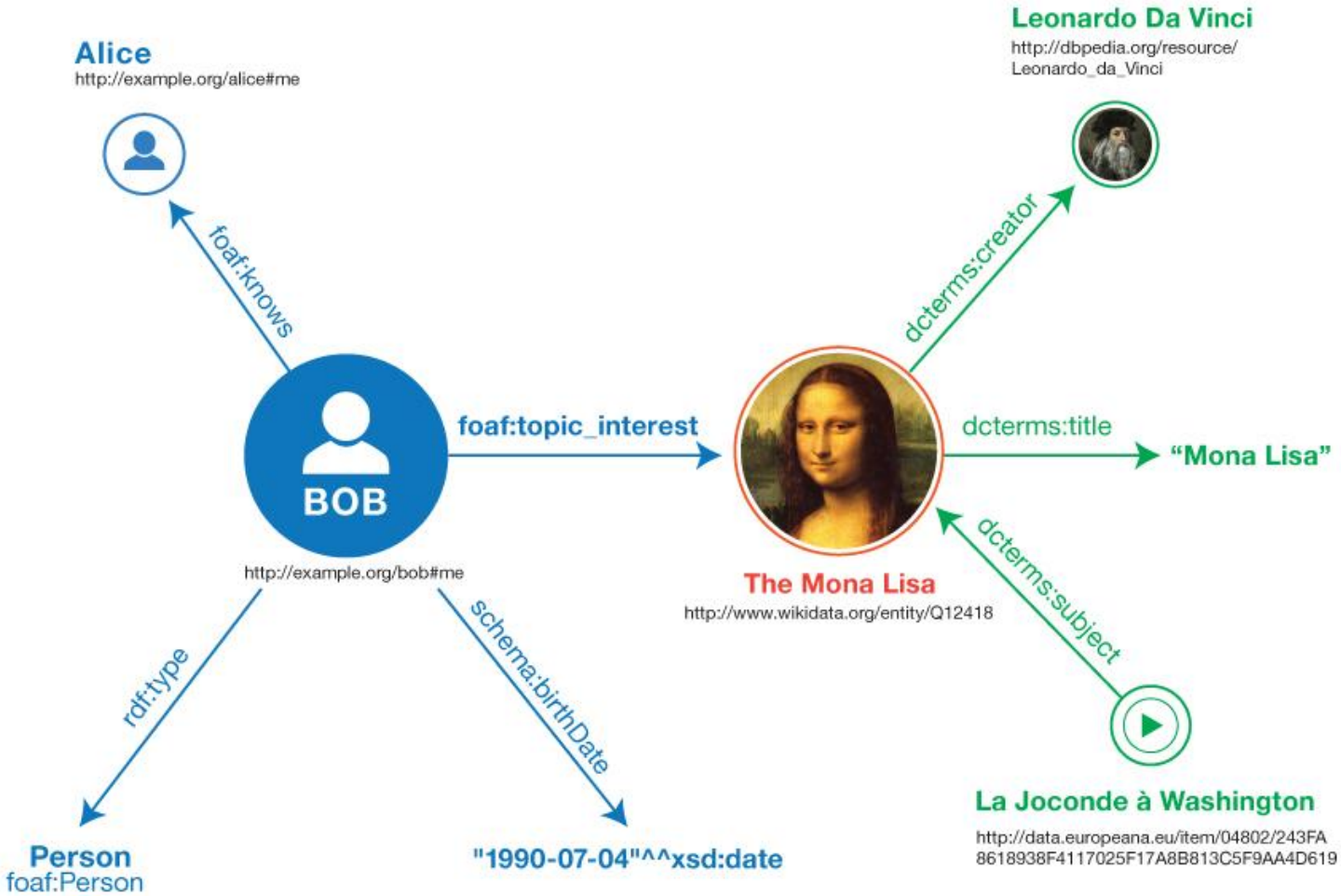
Subject -> Predicate -> Object

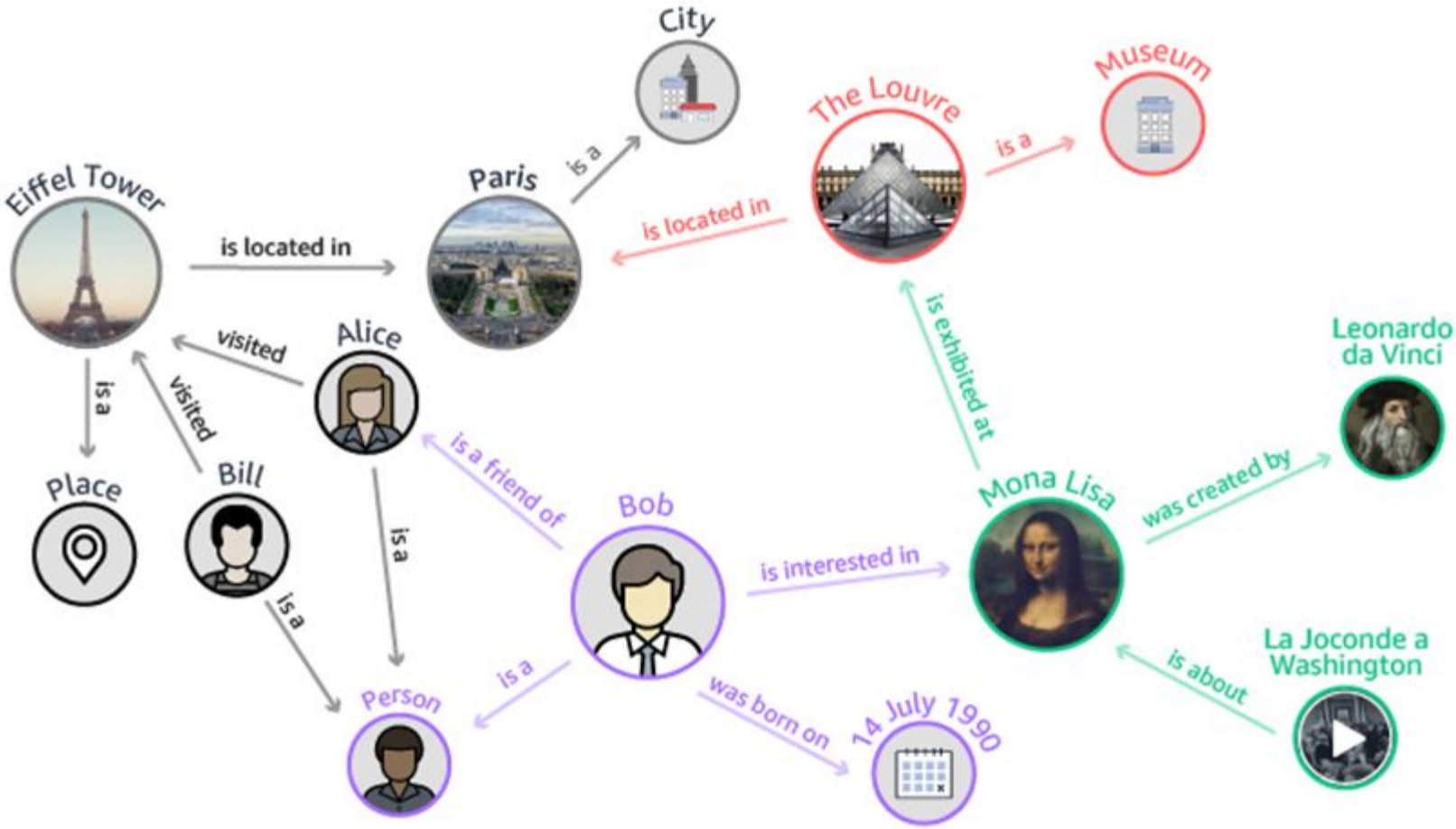
<Bob> <is a> <person>.
<Bob> <is a friend of> <Alice>.
<Bob> <is born on> <the 4th of July 1990>.
<Bob> <is interested in> <the Mona Lisa>.
<the Mona Lisa> <was created by> <Leonardo da Vinci>.
<the video 'La Joconde à Washington'> <is about> <the Mona Lisa>

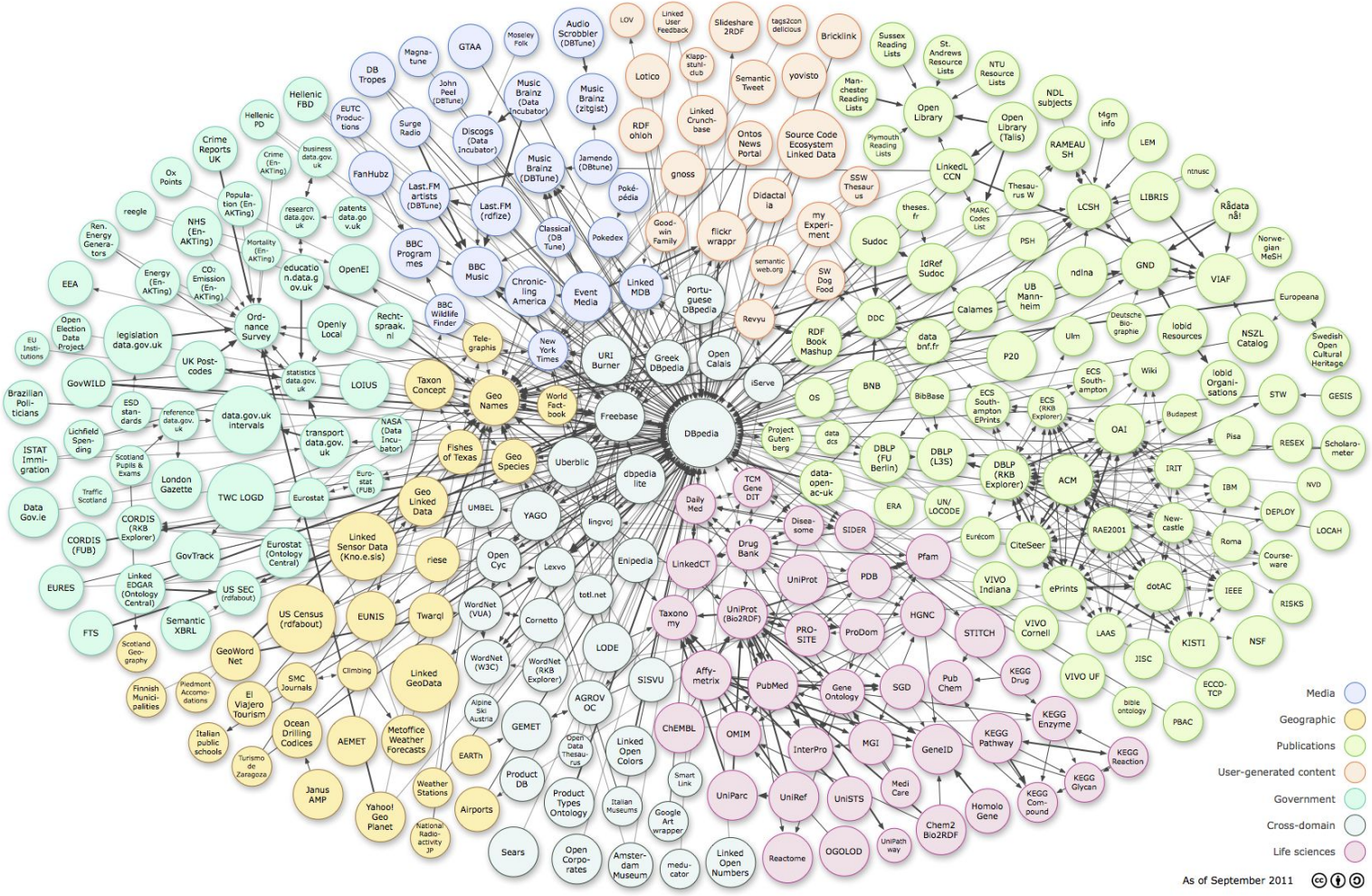
How do we know we’re all talking about the same things?

Subject -> Predicate -> Object

<http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me> .
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/topic_interest> <http://www.wikidata.org/entity/Q12418> .
<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/title> "Mona Lisa" .


Alice
http://example.org/alice#me
foaf:knows
BOB
http://example.org/bob#me
foaf:topic_interest
rdf:type
Person
foaf:Person
schema:birthDate
"1990-07-04"^^xsd:date
Leonardo Da Vinci
http://dbpedia.org/resource/
Leonardo_da_Vinci
dcterms:creator
dcterms:title
"Mona Lisa"
The Mona Lisa
http://www.wikidata.org/entity/Q12418
dcterms:subject
La Joconde à Washington
http://data.europeana.eu/item/04802/243FA
8618938F4117025F17A8B813C5F9AA4D619


City
Museum
The Louvre
is a
Eiffel Tower
Paris
is a
is located in
is located in
is exhibited at
Leonardo
da Vinci
visited
Alice
is a
visited
Bill
Place
is a friend of
Bob
Mona Lisa
was created by
is interested in
is a
is a
Person
was born on
14 July 1990
is about
La Joconde a
Washington


DBpedia
Freebase
Geo Names
URI Burner
Greek DBpedia
Open Calais
iServe
Linked MDB
Portu-guese DBpedia
Event Media
Chronic-ling America
BBC Music
BBC Programmes
Last.FM (rdfize)
Last.FM artists (DBTune)
Music Brainz (DBTune)
Jamendo (DBtune)
Discogs (Data Incubator)
Music Brainz (Data Incubator)
Audio Scrobbler (DBTune)
GTAA
DB Tropes
Magna-tune
Moseley Folk
John Peel (DBTune)
Surge Radio
Hellenic FBD
Hellenic PD
EUTC Produc-tions
FanHubz
Crime Reports UK
Ox Points
reegle
NHS (En-AKTing)
Popula-tion (En-AKTing)
Crime (En-AKTing)
business data.gov.uk
research data.gov.uk
patents data.go v.uk
Mortality (En-AKTing)
Energ. (En-AKTing)
CO2 Emission (En-AKTing)
Ren. Energy Genera-tors
EEA
Energy (En-AKTing)
education.data.gov.uk
OpenEI
Ord-nance Survey
Openly Local
Recht-spraak.nl
BBC Wildlife Finder
BBC Programmes
Tele-graphis
New York Times
Taxon Concept
LOIUS
statistics data.gov.uk
legislation data.gov.uk
Open Election Data Project
EU Insti-tutions
UK Post-codes
GovWILD
Brazilian Poli-ticians
ESD stan-dards
reference data.gov.uk
data.gov.uk intervals
Lichfield Spen-ding
ISTAT Immi-gration
Scotland Pupils & Exams
London Gazette
transport data.gov.uk
NASA (Data Incu-bator)
Fishes of Texas
Geo Species
Uberblic
dbpedia lite
Daily Med
TCM Gene DIT
Project Guten-berg
data dcs
data-open-ac-uk
DBLP (FU Berlin)
DBLP (L3S)
DBLP (RKB Explorer)
ACM
ERA
UN/ LOCODE
Eurécom
CiteSeer
Disea-some
SIDER
Drug Bank
LinkedCT
Enipedia
Lexvo
totl.net
Cornetto
WordNet (VUA)
WordNet (W3C)
WordNet (RKB Explorer)
LODE
AGROVOC
SISVU
Affy-metrix
Taxono my
UniProt (Bio2RDF)
UniProt
PDB
Pfam
PRO-SITE
ProDom
HGNC
STITCH
PubMed
Gene Ontology
SGD
Pub Chem
KEGG Drug
KEGG Enzyme
KEGG Pathway
KEGG Reaction
KEGG Glycan
KEGG Com-pound
Homolo Gene
GeneID
MGI
InterPro
OMIM
ChEMBL
Linked Open Colors
Open Data Thesau-rus
Smart Link
UniParc
UniRef
UniSTS
UniPath way
OGOLOD
Reactome
Linked Open Numbers
medu-cator
Google Art wrapper
Italian Museums
Amster-dam Museum
Open Corpo-rates
Sears
Product Types Ontology
Product DB
Airports
Weather Stations
National Radio-activity JP
Yahoo! Geo Planet
Janus AMP
Ocean Drilling Codices
AEMET
Metoffice Weather Forecasts
GEMET
EARTh
Alpine Ski Austria
Climbing
SMC Journals
El Viajero Tourism
Piedmont Accomo-dations
Finnish Munici-palities
Italian public schools
Turismo de Zaragoza
GeoWord Net
Linked GeoData
US Census (rdfabout)
EUNIS
Twarql
Linked Sensor Data (Kno.e.sis)
riese
Open Cyc
YAGO
UMBEL
Geo Linked Data
Eurostat
Eurostat (Ontology Central)
Euro-stat (FUB)
TWC LOGD
GovTrack
US SEC (rdfabout)
Linked EDGAR (Ontology Central)
Semantic XBRL
Scotland Geo-graphy
EURES
FTS
CORDIS (FUB)
CORDIS (RKB Explorer)
Traffic Scotland
Data Gov.ie
Recyclos
Gnoss
Ontos News Portal
Source Code Ecosystem Linked Data
Didactalia
Good-win Family
flickr wrappr
Pokedex
Classical (DB Tune)
Poké-pédia
Jamendo (DBtune)
Linked Crunch-base
RDF ohloh
Lotico
LOV
Linked User Feedback
Slideshare 2RDF
Semantic Tweet
tags2con delicious
yovisto
Bricklink
Klapp-stuhl-club
Sussex Reading Lists
St. Andrews Resource Lists
NTU Resource Lists
Man-chester Reading Lists
Plymouth Reading Lists
Open Library
Open Library (Talis)
NDL subjects
t4gm info
RAMEAU SH
LEM
LinkedL CCN
SSW Thesaur us
Thesau-rus W
MARC Codes List
theses. fr
LCSH
LIBRIS
ntnusc
Rådata nå!
PSH
Norwe-gian MeSH
Sudoc
IdRef Sudoc
GND
VIAF
Europeana
Swedish Open Cultural Heritage
my Experi-ment
semantic web.org
SW Dog Food
Revyu
RDF Book Mashup
DDC
Calames
UB Mann-heim
ndlna
Ulm
Deutsche Bio-graphie
lobid Resources
NSZL Catalog
data bnf.fr
P20
BNB
BibBase
OS
ECS South-ampton
ECS South-ampton EPrints
ECS (RKB Explorer)
Wiki
lobid Organi-sations
STW
GESIS
Budapest
OAI
Pisa
RESEX
Scholaro-meter
IRIT
IBM
NVD
DEPLOY
LOCAH
New-castle
RAE2001
Roma
Course-ware
dotAC
IEEE
RISKS
ePrints
VIVO Indiana
VIVO Cornell
LAAS
KISTI
NSF
JISC
VIVO UF
ECCO-TCP
PBAC
bible ontology
Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences
As of September 2011

“The Heritage Connector uses computational techniques to turn information held in free text into world identifiable URIs and RDF triples”

How do we use it?

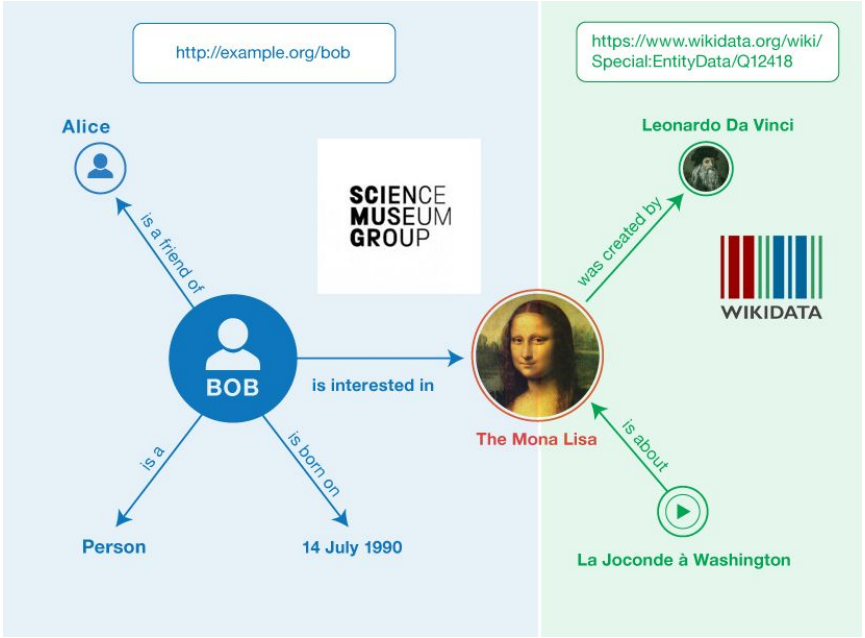
SPARQL = SQL for Linked Data

```
SELECT ?item ?itemLabel
WHERE
{
  ?item wdt:P31 wd:Q146.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

Federated queries are one of the more powerful features of SPARQL

```
SELECT ?item ?itemLabel ?wdItem ?wdItemlabel WHERE {
 # Change PERSON below to OBJECT or ORGANISATION
 ?item skos:hasTopConcept 'PERSON'.
  ?item owl:sameAs ?wdItem.
  ?item rdfs:label ?itemLabel.

  SERVICE <https://query.wikidata.org/sparql> {
    ?wdItem rdfs:label ?wdItemlabel.
    FILTER(LANG(?wdItemlabel) = "en").
  }
} LIMIT 100
```

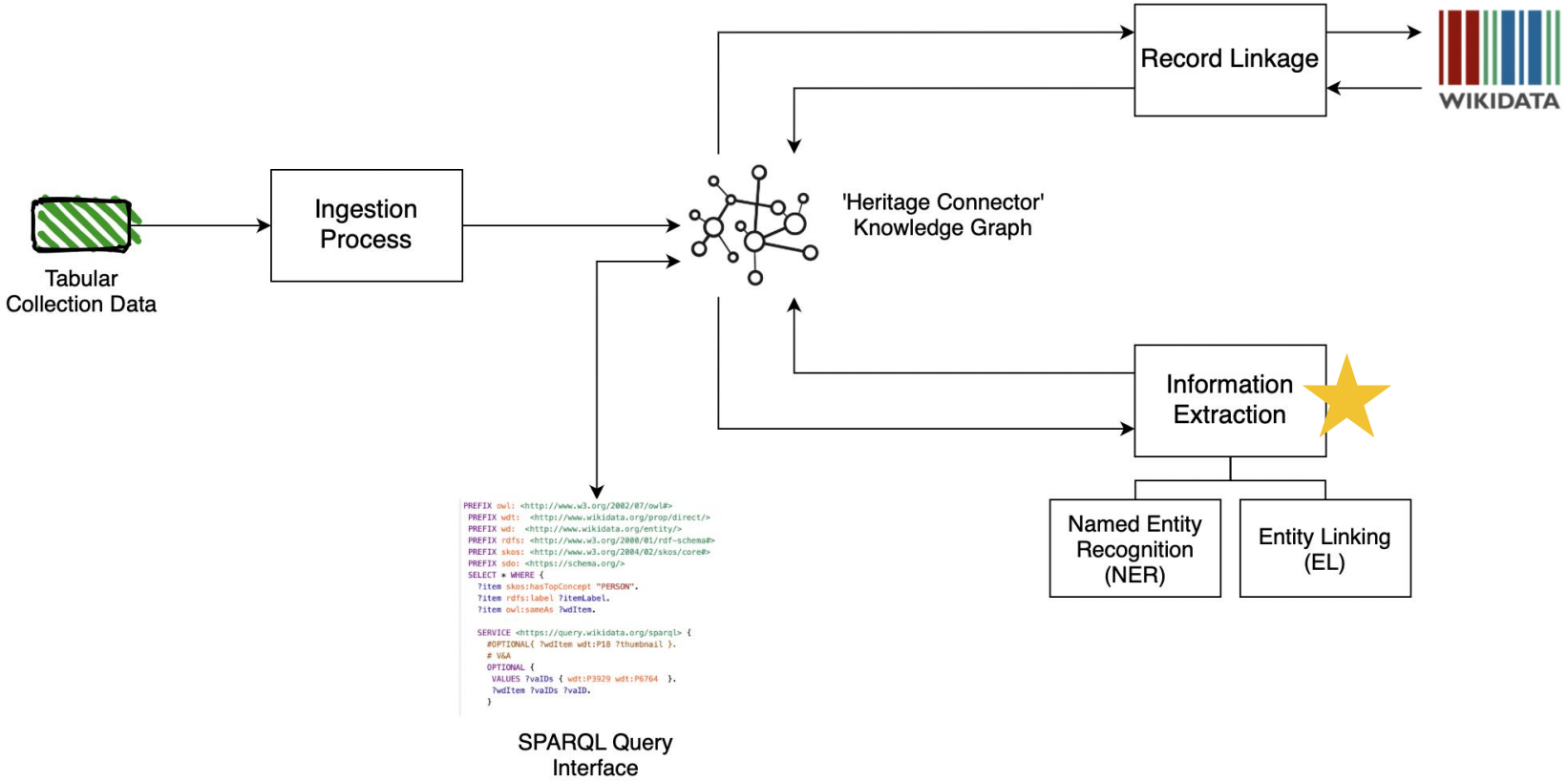


Let’s try some SPARQL queries..

SPARQL DEMO

Making new connections:

Named Entity Recognition and Entity Linking


Record Linkage
WIKIDATA
Tabular
Collection Data
Ingestion
Process
'Heritage Connector'
Knowledge Graph
Information
Extraction
Named Entity
Recognition
(NER)
Entity Linking
(EL)
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX sdo: <https://schema.org/>
SELECT * WHERE {
?item skos:hasTopConcept "PERSON".
?item rdfs:label ?itemLabel.
?item owl:sameAs ?wdItem.
SERVICE <https://query.wikidata.org/sparql> {
#OPTIONAL{ ?wdItem wdt:P18 ?thumbnail }.
# V&A
OPTIONAL {
VALUES ?vaIDs { wdt:P3929 wdt:P6764 }.
?wdItem ?vaIDs ?vaID.
}
SPARQL Query
Interface

ai4lam | 2021-02-16 ai4lam Community | Sokol space suit | Science Mus

https://collection.sciencemuseumgroup.org.uk/objects/co8538105/sokol-space-suit-space-suit

Search objects, people, categories, object numbers etc.

Sokol space suit

MADE: 1991 in Russia

British astronaut, Helen Sharman's Sokol spacesuit made by Zvezda. Sharman wore this rescue suit during the space flight on board the SOYUZ-TM-12 and MIR spacecraft in May 1991. Space suit model number KV-2 No. 167.

Sokol-KV-2 rescue suit worn by Helen Sharman during the Juno mission to the Mir space station, 1991

СПАСАТЕЛЬНЫЙ СКАФАНДР

Helen Sharman was the first British person in space. Sharman wore this suit for two hours on the ground to check its fit. Lying back, she tried to read but her arms ached from holding the book for so long. Despite the suit's cooling systems she sweated 2 litres during the mission launch. Once she could remove the suit, she dried it thoroughly to ensure it would not go mouldy.

The Sokol suit was developed after three unsuited cosmonauts asphyxiated on the Soyuz 11 mission in 1971 when their descent module depressurised during the return to Earth. Every cosmonaut now wears one during launch and return from space. It will keep the wearer alive for a number of hours in the event of a cabin depressurisation. Each suit is tailor made to the individual cosmonaut and comprises an inner, airtight 'bladder' of rubberised plastic and an outer layer of nylon canvas. There are connecting rings on the lower abdomen for air (cooling) and oxygen supplies and a centrally positioned pressure adjustment valve control on the chest; the pressure gauge is on the left wrist. The helmet and boots are integral with the rest of the suit; the gloves are attached with anodized aluminium bayonet fixings. Today's Sokol design is little changed from the original.

Source: Zvezda

ON DISPLAY

Science Museum: Exploring Space Gallery

If you are visiting to see this object, please contact us in advance to make sure that it will be on display.

RELATED PEOPLE

Helen Sharman

RELATED ARTICLES

National Science and Media Museum

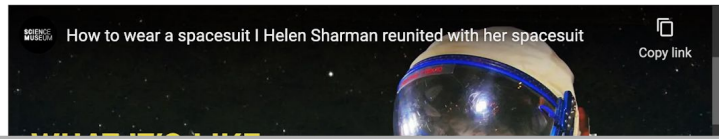
Bring the National Science and Media Museum collection home in Animal Crossing

Science Museum

Highlights on display

Science Museum announces National Lottery ticket sales trial as Helen Sharman spacesuit goes back on display

UK tour of Tim Peake's spacecraft attracts 1.3 million visitors as Science Museum marks Apollo anniversaries with Summer of Space



LOOK CLOSER

Helen Sharman on her Sokol space suit

Applying NER to Catalogue Descriptions

British NORP astronaut, Helen Sharman's PERSON Sokol OBJECT spacesuit made by Zvezda ORG . Sharman PERSON wore this rescue suit during the space flight on board the SOYUZ-TM-12 and MIR spacecraft in May 1991 DATE . Space suit model number KV-2 No. 167 CARDINAL .

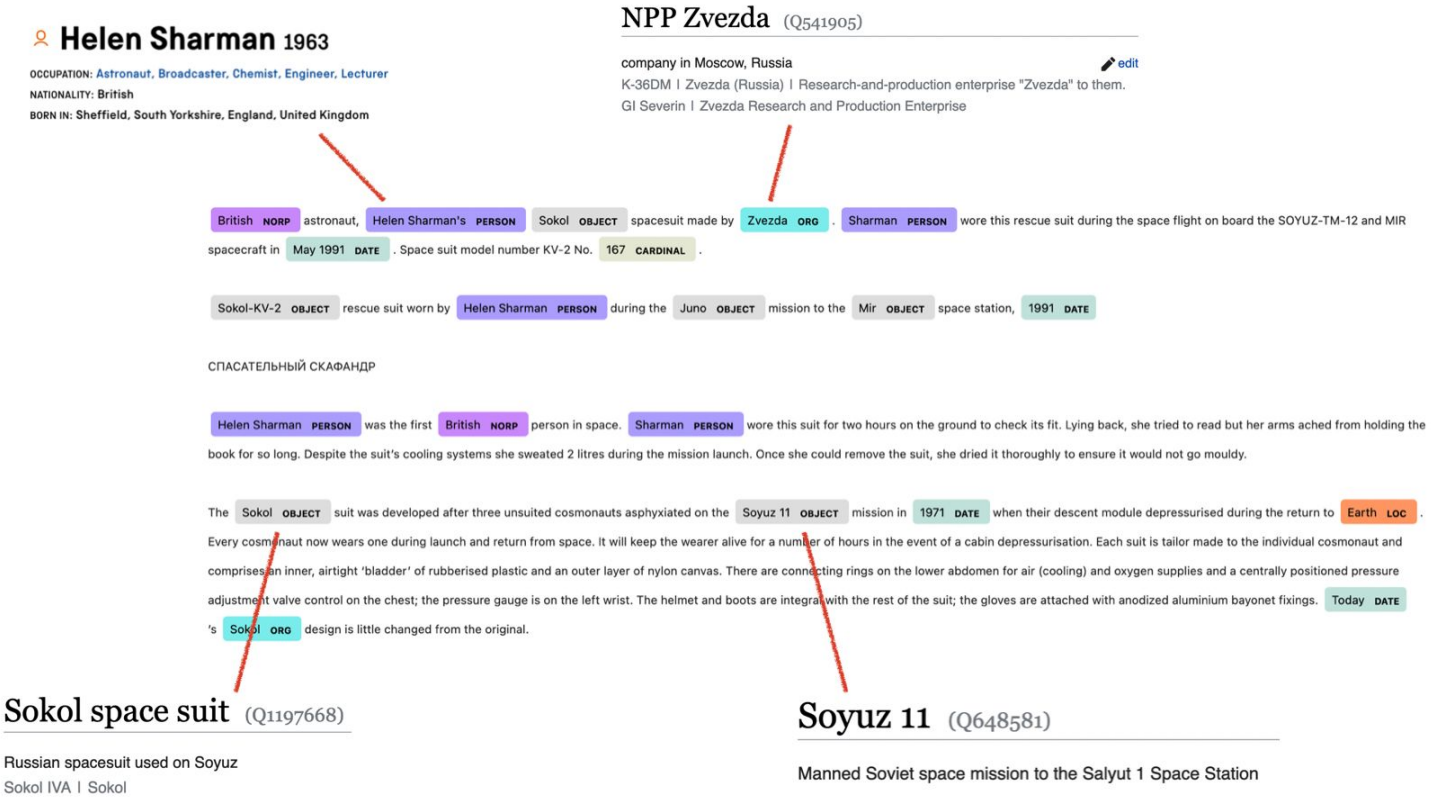
Sokol-KV-2 OBJECT rescue suit worn by Helen Sharman PERSON during the Juno OBJECT mission to the Mir OBJECT space station, 1991 DATE

СПАСАТЕЛЬНЫЙ СКАФАНДР

Helen Sharman PERSON was the first British NORP person in space. Sharman PERSON wore this suit for two hours on the ground to check its fit. Lying back, she tried to read but her arms ached from holding the book for so long. Despite the suit's cooling systems she sweated 2 litres during the mission launch. Once she could remove the suit, she dried it thoroughly to ensure it would not go mouldy.

The Sokol OBJECT suit was developed after three unsuited cosmonauts asphyxiated on the Soyuz 11 OBJECT mission in 1971 DATE when their descent module depressurised during the return to Earth LOC . Every cosmonaut now wears one during launch and return from space. It will keep the wearer alive for a number of hours in the event of a cabin depressurisation. Each suit is tailor made to the individual cosmonaut and comprises an inner, airtight 'bladder' of rubberised plastic and an outer layer of nylon canvas. There are connecting rings on the lower abdomen for air (cooling) and oxygen supplies and a centrally positioned pressure adjustment valve control on the chest; the pressure gauge is on the left wrist. The helmet and boots are integral with the rest of the suit; the gloves are attached with anodized aluminium bayonet fixings. Today DATE 's Sokol ORG design is little changed from the original.

Predicting knowledge graph connections with an Entity Linker



Named Entity Recognition and Entity Linking in 3 steps

Named Entity Recognition and Entity Linking in 3 steps

1. Mention Detection (NER) spaCy

A cyanotype of Lastrea Filix-Mas, from 'Cyanotypes of British and Foreign Ferns', made by Anna Atkins PERSON (1799-1871) in 1853.

Named Entity Recognition and Entity Linking in 3 steps

1. Mention Detection (NER) spaCy

A cyanotype of Lastrea Filix-Mas, from 'Cyanotypes of British and Foreign Ferns', made by Anna Atkins PERSON (1799-1871) in 1853.

2. Candidate Generation

A cyanotype of Lastrea Filix-Mas, from 'Cyanotypes of British and Foreign Ferns', made by Anna Atkins PERSON (1799-1871) in 1853.

`Anna Atkins (person)` →

```
Anna Atkins | PERSON
1853 Cyanotype by Anna Atkins | OBJECT
Anna Katrina Zinkeisen | PERSON
Anna Eavis | PERSON
```

Named Entity Recognition and Entity Linking in 3 steps

1. Mention Detection (NER) spaCy

A cyanotype of Lastrea Filix-Mas, from ‘Cyanotypes of British and Foreign Ferns’, made by Anna Atkins PERSON (1799-1871) in 1853.

2. Candidate Generation

A cyanotype of Lastrea Filix-Mas, from ‘Cyanotypes of British and Foreign Ferns’, made by Anna Atkins PERSON (1799-1871) in 1853.

Anna Atkins (person) →

```
Anna Atkins | PERSON
1853 Cyanotype by Anna Atkins | OBJECT
Anna Katrina Zinkeisen | PERSON
Anna Eavis | PERSON
```

3. Candidate Ranking

A cyanotype of Lastrea Filix-Mas, from ‘Cyanotypes of British and Foreign Ferns’, made by Anna Atkins PERSON (1799-1871) in 1853.

Anna Atkins (person) →

```
92% Anna Atkins | PERSON
43% Anna Katrina Zinkeisen | PERSON
31% Anna Eavis | PERSON
 3% 1853 Cyanotype by Anna Atkins | OBJECT
```

Demo 2
(NER)

Try them yourself

SPARQL Query Interface

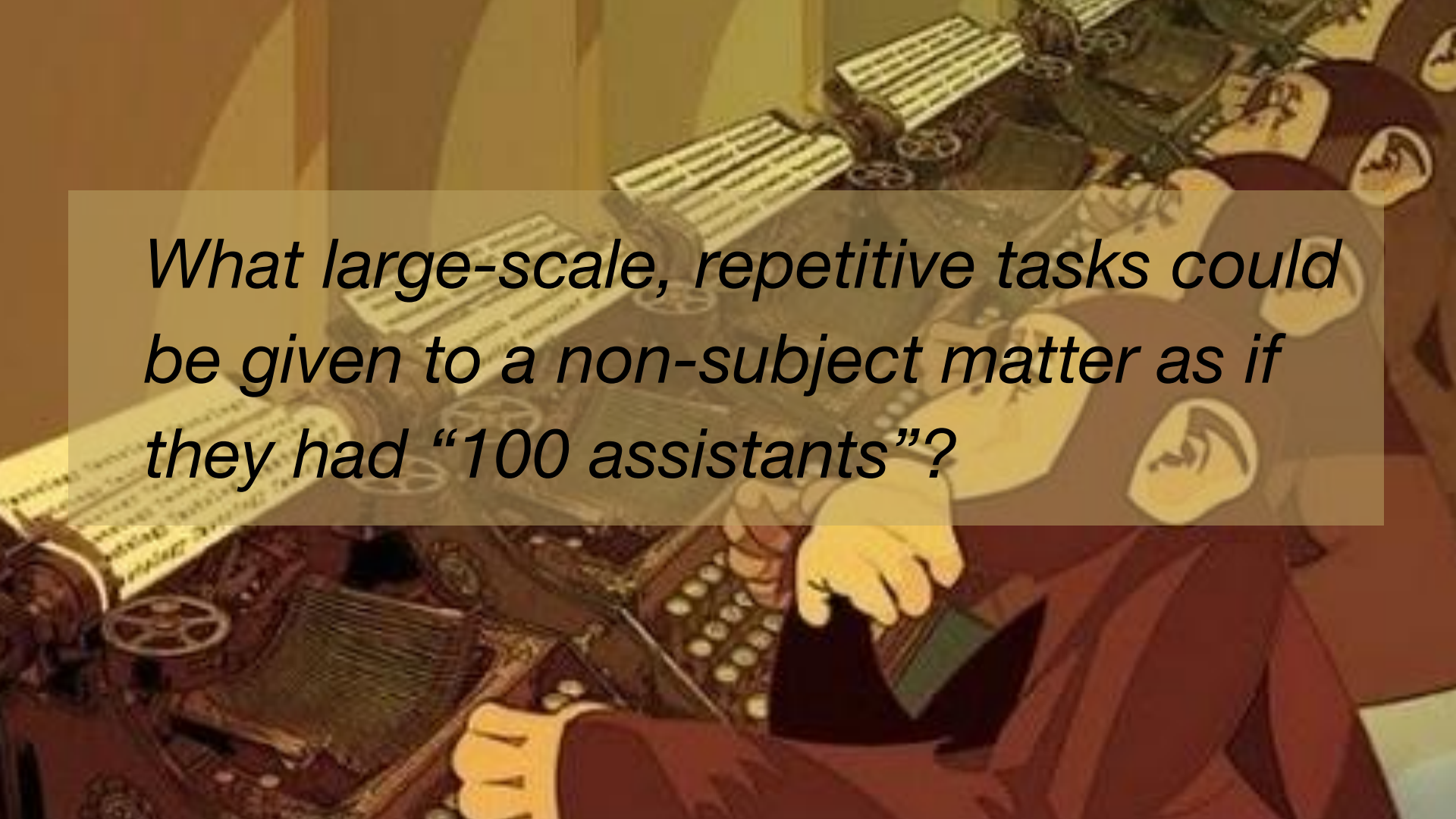
http://63.33.68.17

Named Entity Recognition and Entity Linking Interactive Demo

http://54.217.36.83:8501

Exercise 1

“100 assistants”


What large-scale, repetitive tasks could be given to a non-subject matter as if they had “100 assistants”?

Hopes & Fears Exercise

Hopes = expectations about what can be accomplished in the remainder of the project

- *I hope that…*
- *It would be great if…*

Fears = Doubts or concerns about the remainder of the project

- *I'm concerned that…*
- *This would be great, but...*

Hopes & Fears Exercise

1. Write down your hopes and fears in the Google Doc (~3 mins)
2. I'll cluster ideas into themes
3. Group discussion about hopes, fears and major themes (~10 mins)