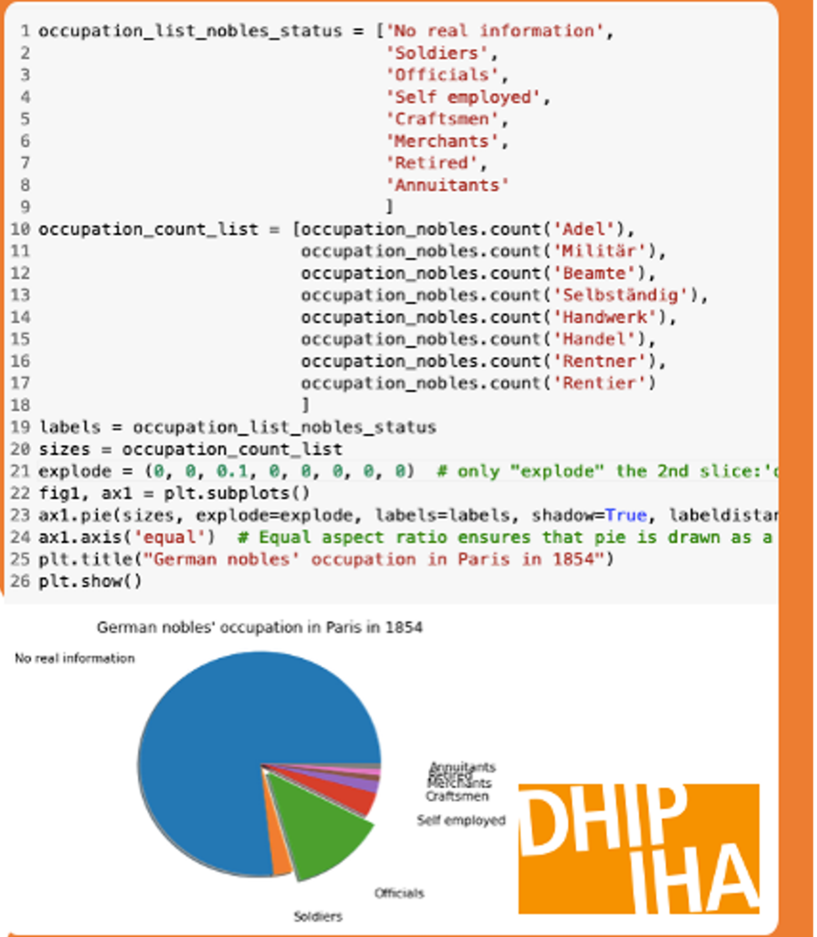
DIGITAL HUMANITIES SEMANTIC NOTEBOOKS



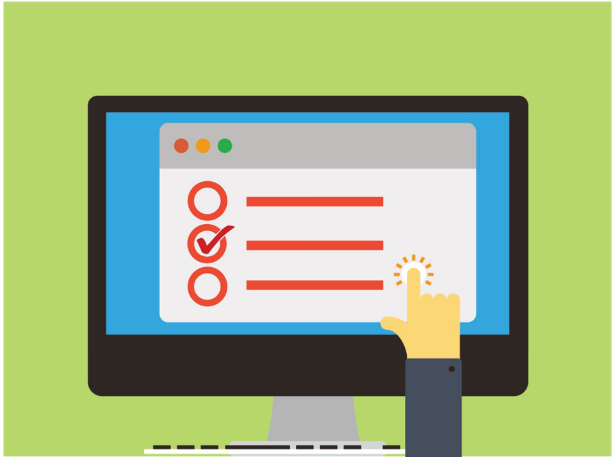
```
occupation_list_nobles_status = ['No real information',
                                 'Soldiers',
                                 'Officials',
                                 'Self employed',
                                 'Craftsmen',
                                 'Merchants',
                                 'Retired',
                                 'Annuitants'
                                 ]
occupation_count_list = [occupation_nobles.count('Adel'),
                         occupation_nobles.count('Militär'),
                         occupation_nobles.count('Beamte'),
                         occupation_nobles.count('Selbständig'),
                         occupation_nobles.count('Handwerk'),
                         occupation_nobles.count('Handel'),
                         occupation_nobles.count('Rentner'),
                         occupation_nobles.count('Rentier')
                         ]
labels = occupation_list_nobles_status
sizes = occupation_count_list
explode = (0, 0, 0.1, 0, 0, 0, 0, 0)  # only "explode" the 2nd slice:'
fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels=labels, shadow=True, labeldistar
ax1.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a
plt.title("German nobles' occupation in Paris in 1854")
plt.show()
```

gkembellec@dhi-paris.fr

- Semantic publishing?
- Executable notebook?

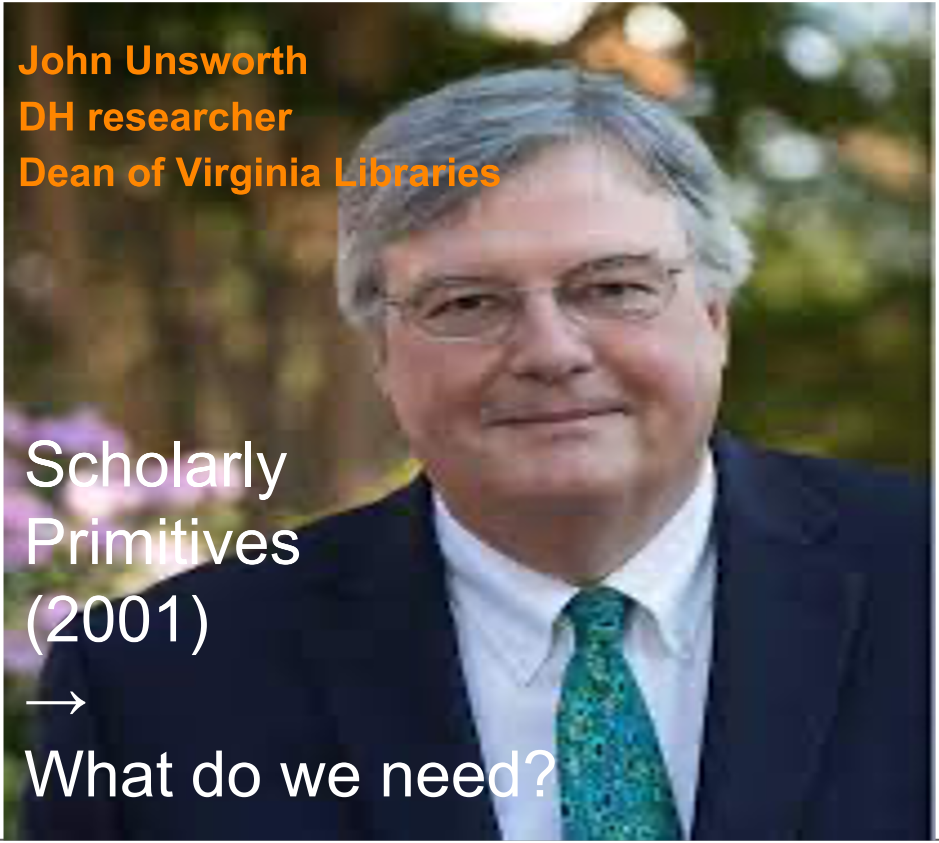
(*aka* Jupyter Notebook)

https://pingo.coactum.de/337082



Please
fill in the form

The challenges of science mediation

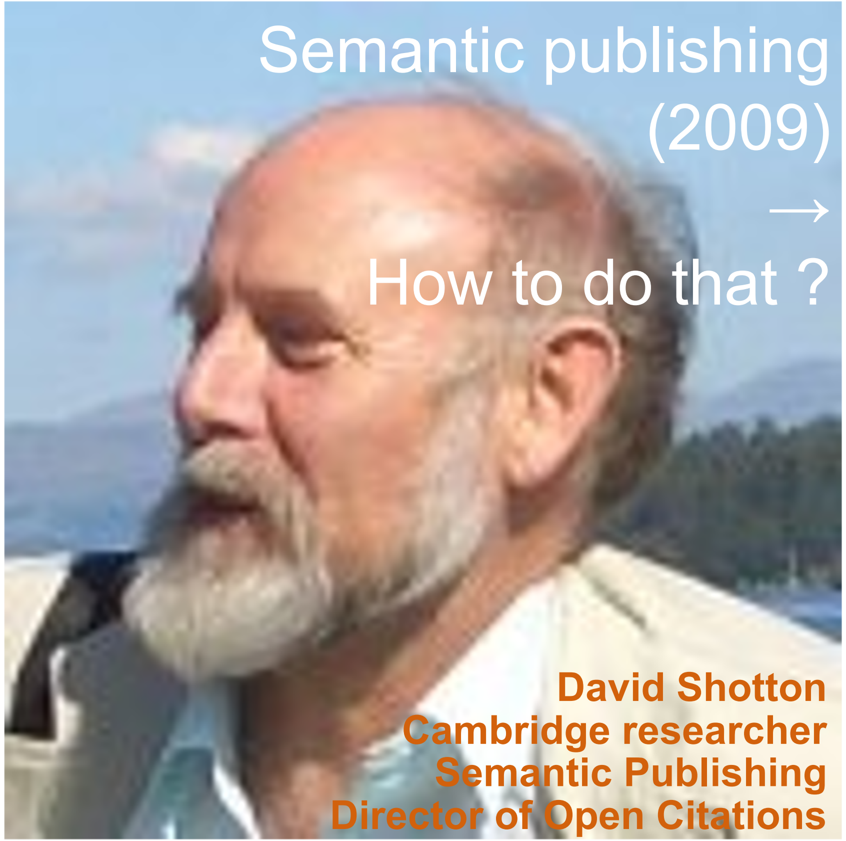


John Unsworth
DH researcher
Dean of Virginia Libraries

Scholarly Primitives (2001)

→

What do we need?



Semantic publishing (2009)

→

How to do that ?

David Shotton
Cambridge researcher
Semantic Publishing
Director of Open Citations

Scholars’ needs and ways to make linked reproducible science

What do we need as scholars ?

Scholarly Primitives, according to Unsworth (2000) :

- **Discovering** (serendipity)
- **Annotating** (Read / Write as a single process)
- **Comparing** (Theories, datasets, bibliographies)
- **Refering** (quote)
- **Sampling** (data, corpora…)
- **Illustrating** (examples, figures)
- **Representing** (disseminate in the community)

1st : semantic Web / publishing In science & culture

some meaningful concepts

Semantic Web ?

Linking and describing resources on the Web with :

- Authorities databases (ISNI, GnD, Wikidata, Dbpedia…)
- Vocabularies (schema.org, FoaF, Dublin Core...)
- A theoretical framework as a model (RDF)
- Technical implementations (TEI, HTML5 with RDFa, JSON-LD,

Sir Tim Berners Lee
Computer Scientist
Father of the Web (1991)
Father of the Semantic Web (2001)

Semantic publishing as a solution ?

Semantic publishing best practices, according to David Shotton (2009) :

- Use established **standards** wherever possible.
- Publish **raw datasets** to the Web.
- Release **article metadata**, particularly reference lists, in **machine-readable form**.

“Resource Description Framework”
acronym RDF

RDF grammar of a triple :

(Subject , Predicate , Object)

Subject : The resource (what we are talking about) ;

Predicate : the kind of resource property ;

Object : value of the property : can be a number, a text, a web URL

Document contents with RDF triples

If I write :
“Sociology of Religion has been written by Max Weber”,
it could be seen as an RDF triple this way:

Subject	: ***”Sociology of Religion”***
Predicate	: ***was written by***
Object	: “***Max Weber”***

Using the description authority “Dublin Core” we will have “creator” for predicate :

(**”Sociology of Religion”**, dc:creator , ”Max Weber”)

http://purl.org/dc/elements/1.1/creator

same

Document contents with RDF triples

As there is an authority for books (ISBN) , we can use it to be sure we are talking about the good book :

Subject	: ***isbn:***0-8070-4205-6
Predicate	: http://purl.org/dc/elements/1.1/creator
Object	: ***Max Weber***

Using the description authority "isbn" we will have 0-8070-4205-6 for subject

(***isbn:***0-8070-4205-6 ,
"http://purl.org/dc/elements/1.1/creator",
Max Weber)

Document contents with RDF triples

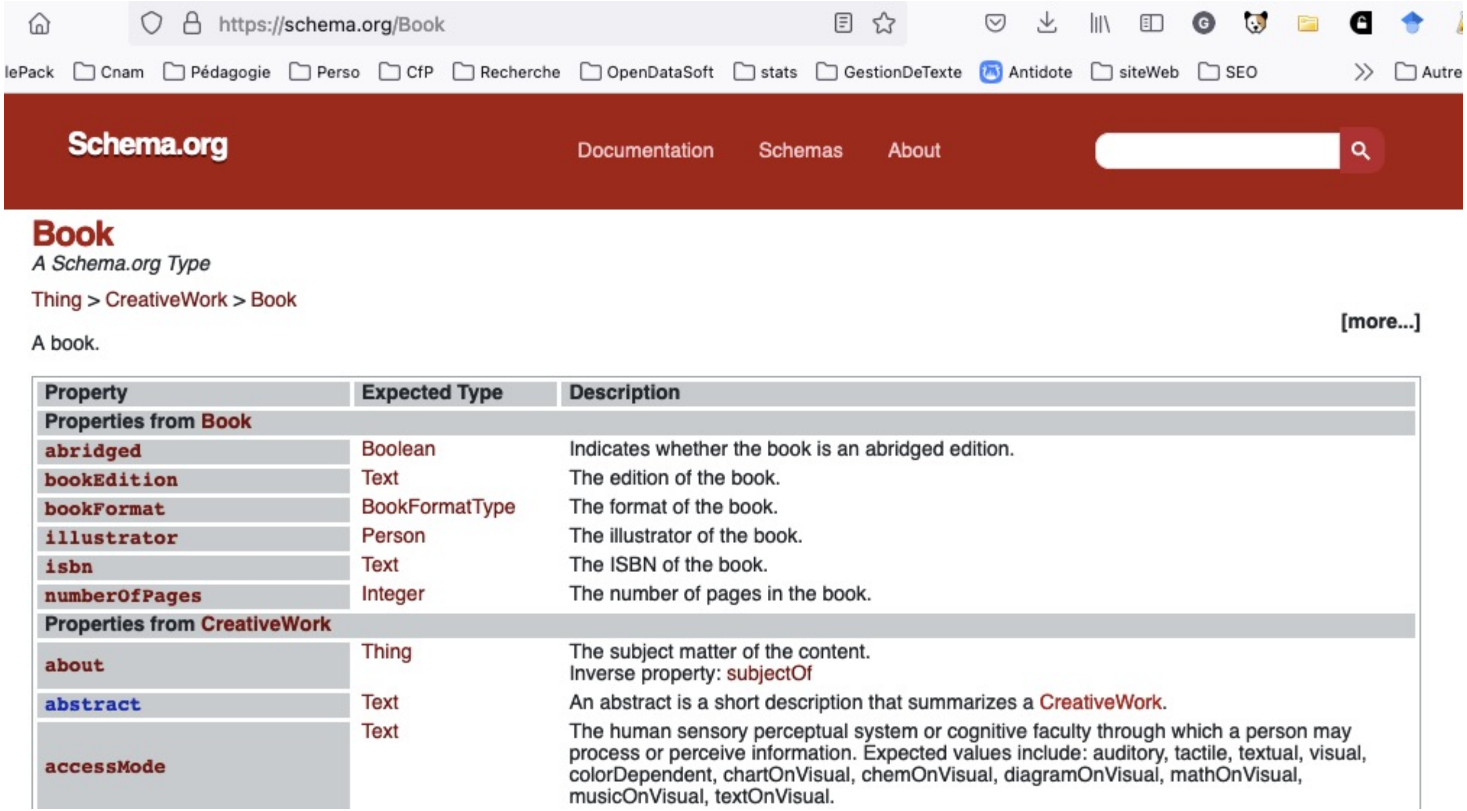
There are also several authorities for people (ISNI, PND from GND, VIAF…) , we can use it to be sure we are talking about the good guy :

Subject	: ***isbn:***0-8070-4205-6
Predicate	: http://purl.org/dc/elements/1.1/creator
Object	: http://d-nb.info/gnd/118629743

Using the description authority ***Personennamendatei*** or **PND** we will have “creator” for Object

(***isbn:***0-8070-4205-6,
http://purl.org/dc/elements/1.1/creator,
http://d-nb.info/gnd/118629743)

Document contents with RDF triples : schema.org



Book

A Schema.org Type

Thing > CreativeWork > Book

[more...]

A book.

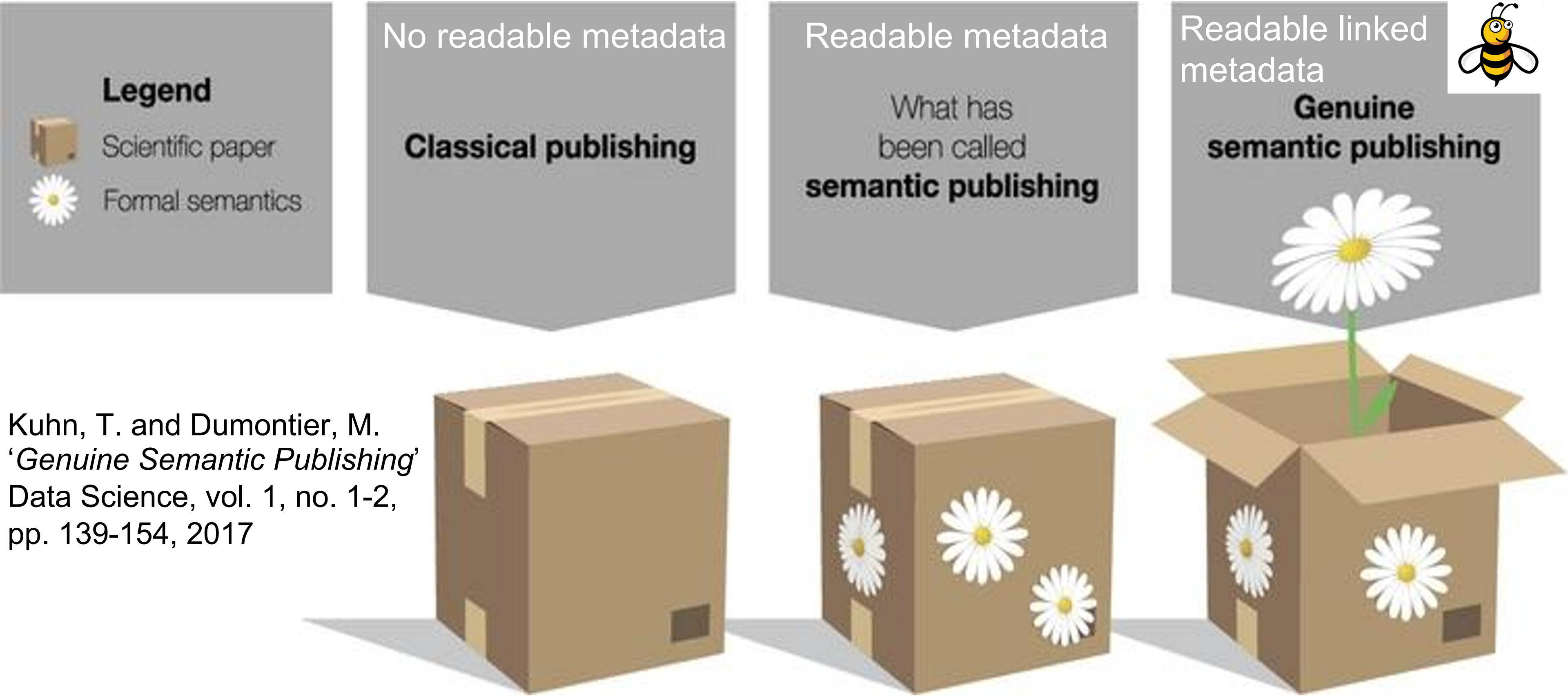
Property	Expected Type	Description
Properties from Book		
abridged	Boolean	Indicates whether the book is an abridged edition.
bookEdition	Text	The edition of the book.
bookFormat	BookFormatType	The format of the book.
illustrator	Person	The illustrator of the book.
isbn	Text	The ISBN of the book.
numberOfPages	Integer	The number of pages in the book.
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: subjectOf
abstract	Text	An abstract is a short description that summarizes a CreativeWork.
accessMode	Text	The human sensory perceptual system or cognitive faculty through which a person may process or perceive information. Expected values include: auditory, tactile, textual, visual, colorDependent, chartOnVisual, chemOnVisual, diagramOnVisual, mathOnVisual, musicOnVisual, textOnVisual.

Document contents with RDF triples

We can here define what the book is about using the book schema from schema.org/ :

Subject	: *”Sociology of Religion”*
Predicate	: https://schema.org/Book/About
Object	: “sociology”, “religion”

Semantic publishing?


Legend
Scientific paper
Formal semantics
Classical publishing
What has been called semantic publishing
Genuine semantic publishing

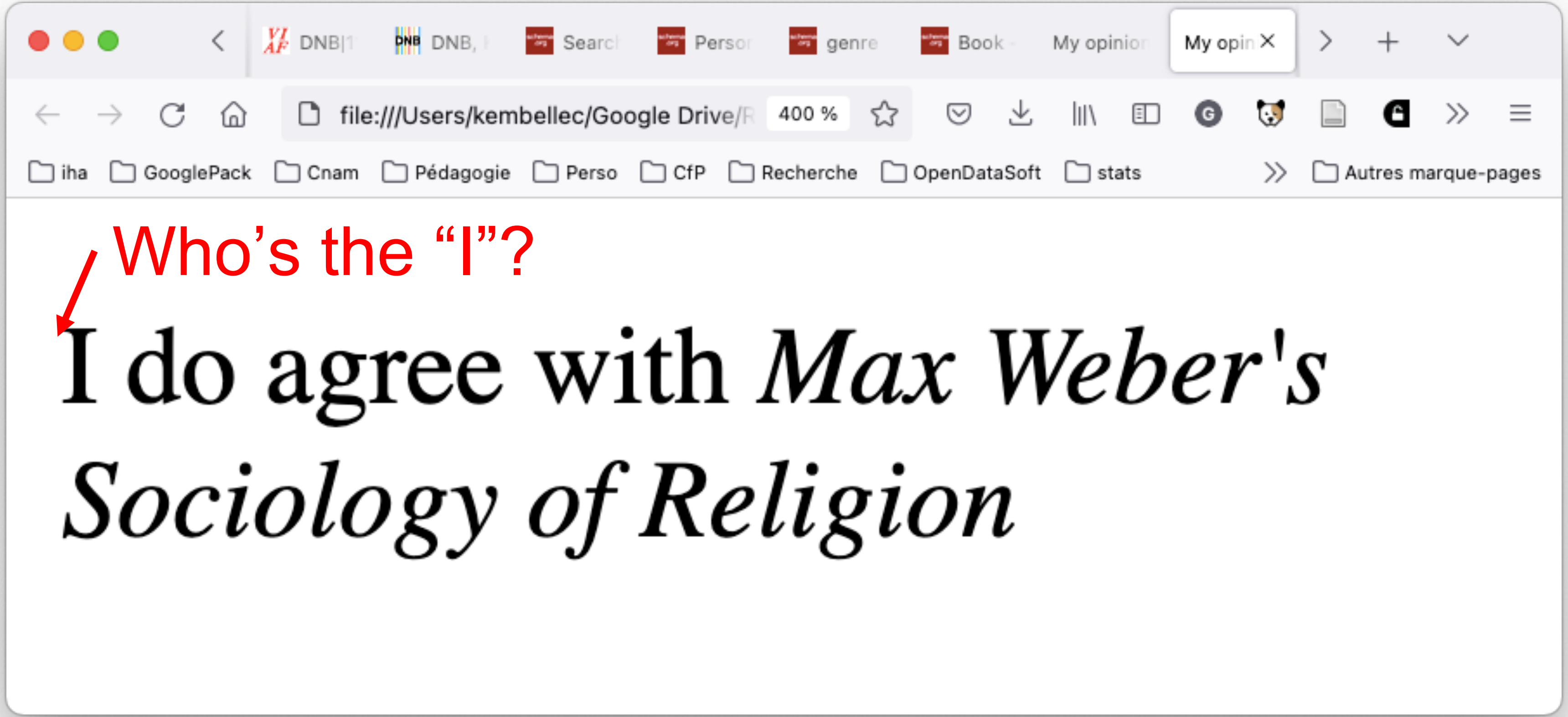
No readable metadata

Readable metadata

Readable linked metadata

Kuhn, T. and Dumontier, M.
‘*Genuine Semantic Publishing*’
Data Science, vol. 1, no. 1-2,
pp. 139-154, 2017

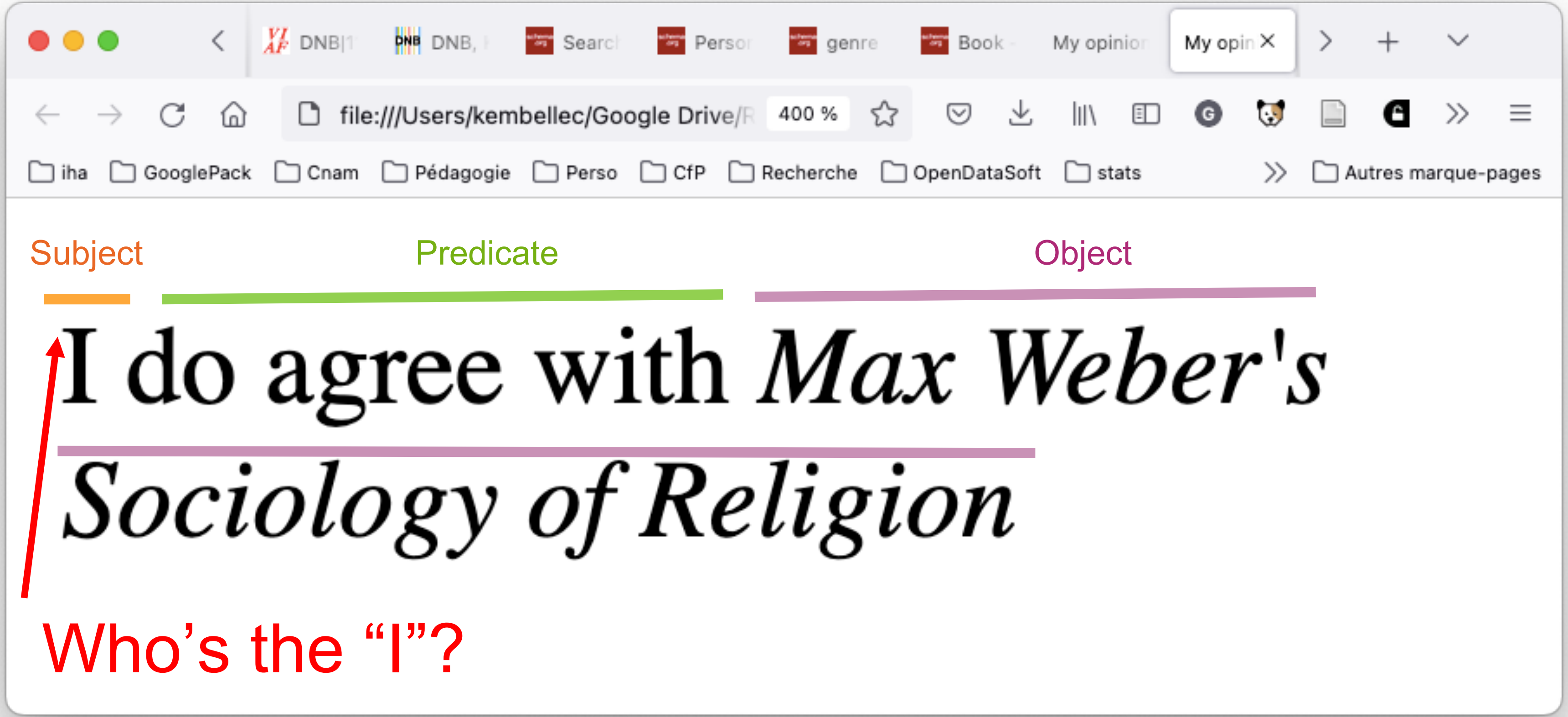
Small example : *mirodata* in a HTML page



Who’s the “I”?

I do agree with *Max Weber's Sociology of Religion*

Small example : *mirodata* in a HTML page



Subject

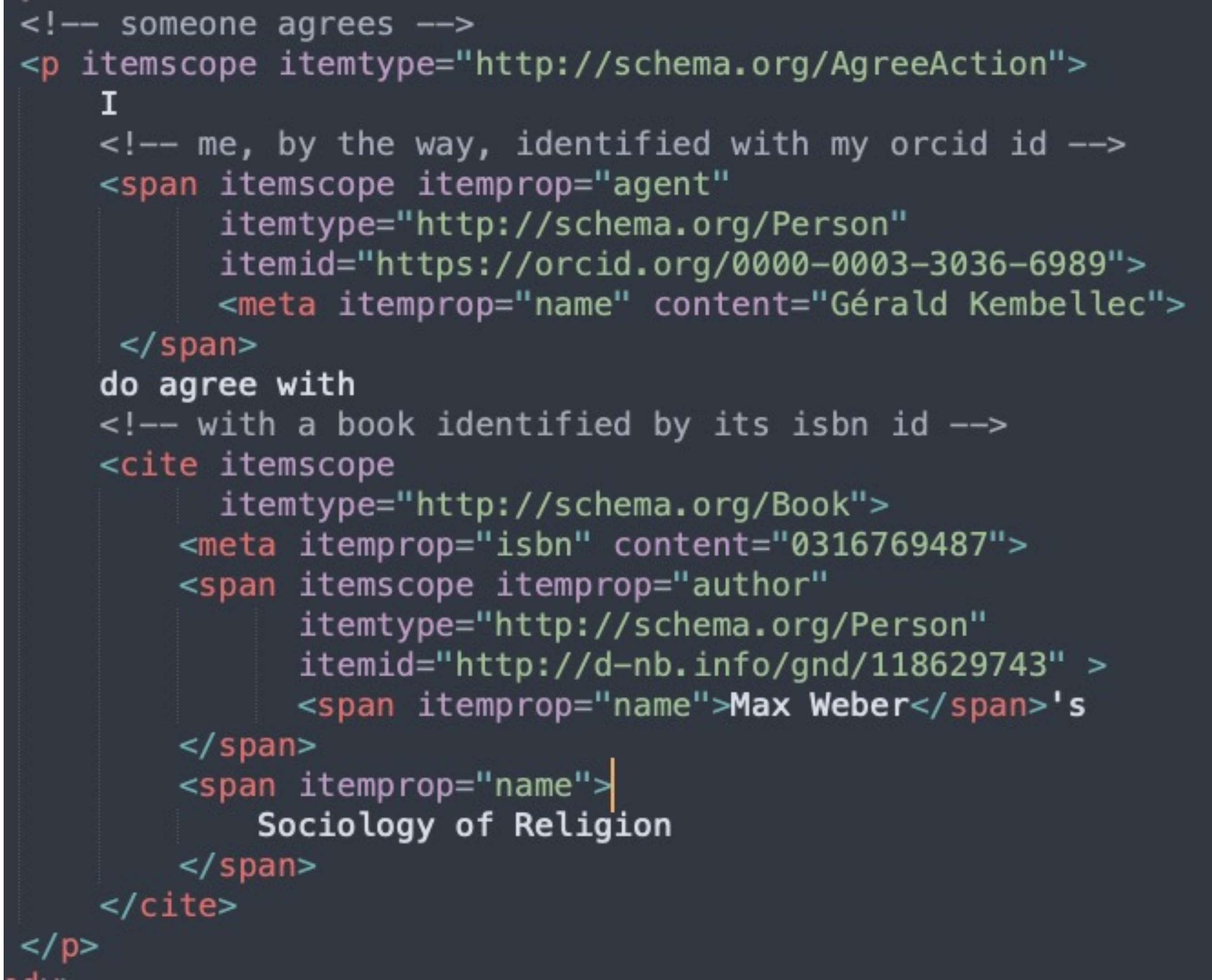
Predicate

Object

I do agree with *Max Weber's Sociology of Religion*

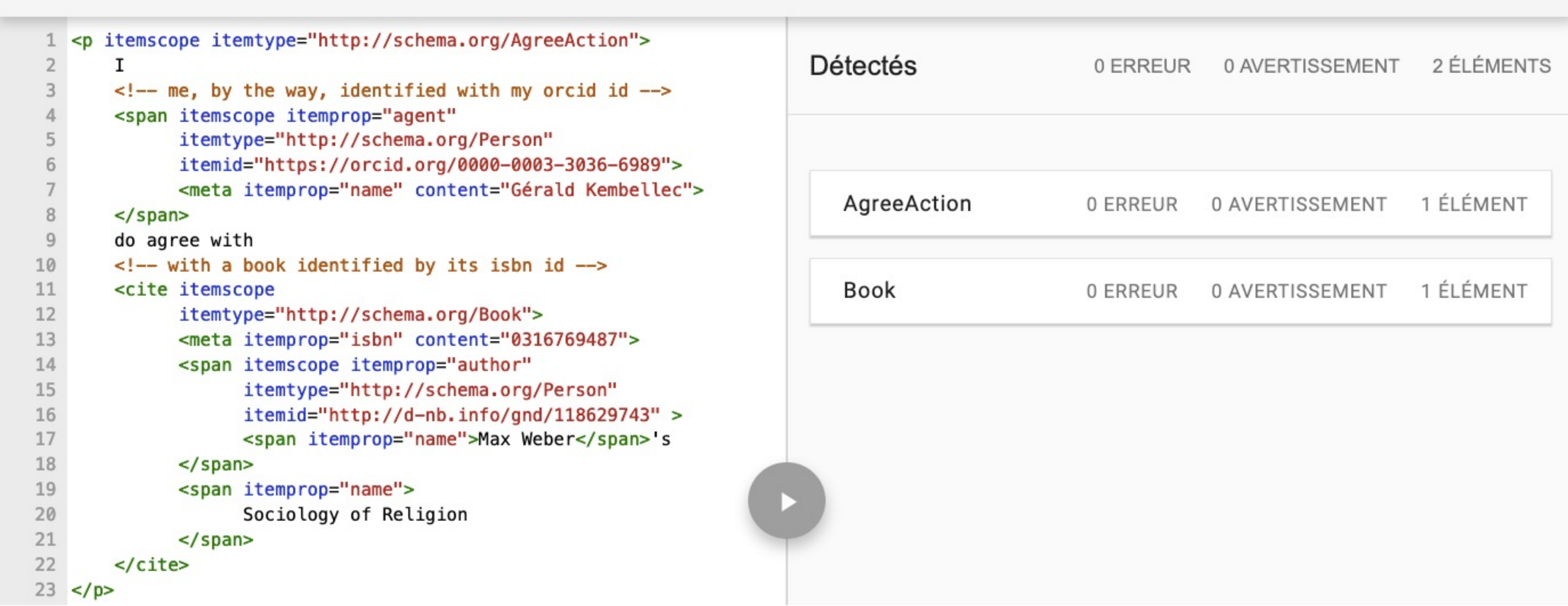
Who's the "I"?

Small example : *microdata* in a HTML page



```
<!-- someone agrees -->
<p itemscope itemtype="http://schema.org/AgreeAction">
    I
    <!-- me, by the way, identified with my orcid id -->
    <span itemscope itemprop="agent"
          itemtype="http://schema.org/Person"
          itemid="https://orcid.org/0000-0003-3036-6989">
          <meta itemprop="name" content="Gérald Kembellec">
     </span>
    do agree with
    <!-- with a book identified by its isbn id -->
    <cite itemscope
          itemtype="http://schema.org/Book">
        <meta itemprop="isbn" content="0316769487">
        <span itemscope itemprop="author"
              itemtype="http://schema.org/Person"
              itemid="http://d-nb.info/gnd/118629743" >
              <span itemprop="name">Max Weber</span>'s
        </span>
        <span itemprop="name">
            Sociology of Religion
        </span>
    </cite>
</p>
```

Small example : *microdata* in a HTML page test with https://validator.schema.org/ 1/3



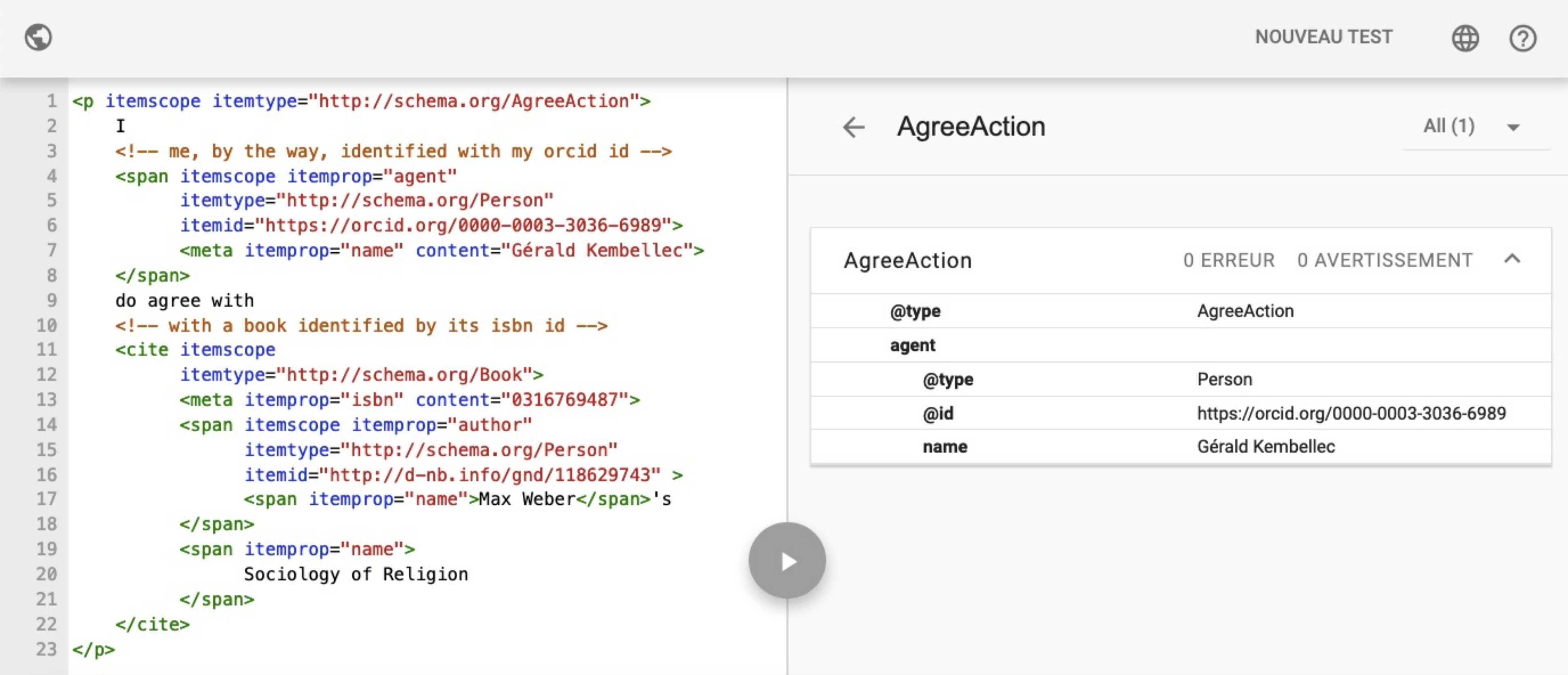
```
<p itemscope itemtype="http://schema.org/AgreeAction">
    I
    <!-- me, by the way, identified with my orcid id -->
    <span itemscope itemprop="agent"
          itemtype="http://schema.org/Person"
          itemid="https://orcid.org/0000-0003-3036-6989">
          <meta itemprop="name" content="Gérald Kembellec">
    </span>
    do agree with
    <!-- with a book identified by its isbn id -->
    <cite itemscope
          itemtype="http://schema.org/Book">
          <meta itemprop="isbn" content="0316769487">
          <span itemscope itemprop="author"
                itemtype="http://schema.org/Person"
                itemid="http://d-nb.info/gnd/118629743" >
                <span itemprop="name">Max Weber</span>'s
          </span>
          <span itemprop="name">
                Sociology of Religion
          </span>
    </cite>
</p>
```

Détectés 0 ERREUR 0 AVERTISSEMENT 2 ÉLÉMENTS

AgreeAction 0 ERREUR 0 AVERTISSEMENT 1 ÉLÉMENT

Book 0 ERREUR 0 AVERTISSEMENT 1 ÉLÉMENT

Small example : *microdata* in a HTML page test with https://validator.schema.org/ 2/3



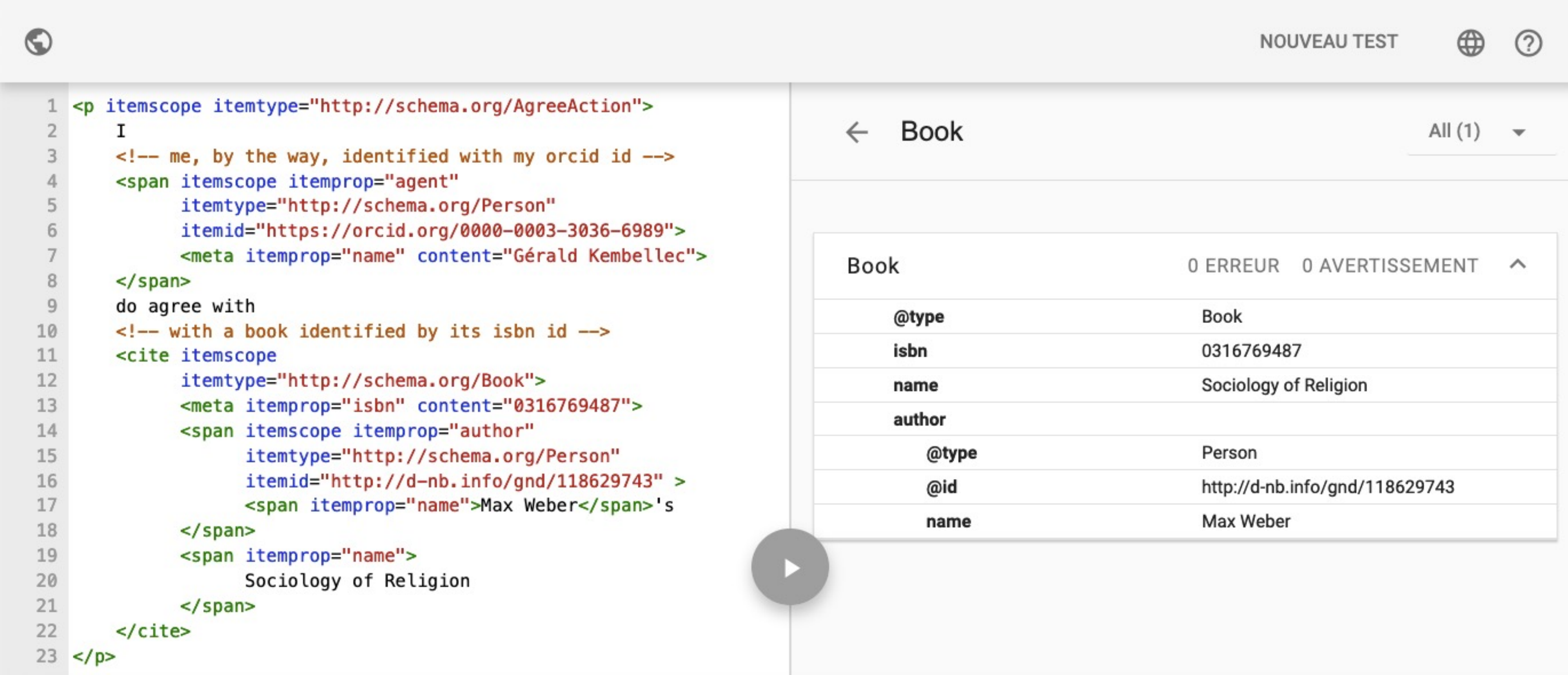
```
<p itemscope itemtype="http://schema.org/AgreeAction">
    I
    <!-- me, by the way, identified with my orcid id -->
    <span itemscope itemprop="agent"
          itemtype="http://schema.org/Person"
          itemid="https://orcid.org/0000-0003-3036-6989">
          <meta itemprop="name" content="Gérald Kembellec">
    </span>
    do agree with
    <!-- with a book identified by its isbn id -->
    <cite itemscope
          itemtype="http://schema.org/Book">
          <meta itemprop="isbn" content="0316769487">
          <span itemscope itemprop="author"
                itemtype="http://schema.org/Person"
                itemid="http://d-nb.info/gnd/118629743" >
                <span itemprop="name">Max Weber</span>'s
          </span>
          <span itemprop="name">
                Sociology of Religion
          </span>
    </cite>
</p>
```

AgreeAction

AgreeAction 0 ERREUR 0 AVERTISSEMENT

@type	AgreeAction
agent	
@type	Person
@id	https://orcid.org/0000-0003-3036-6989
name	Gérald Kembellec

Small example : *microdata* in a HTML page test with https://validator.schema.org/ 3/3



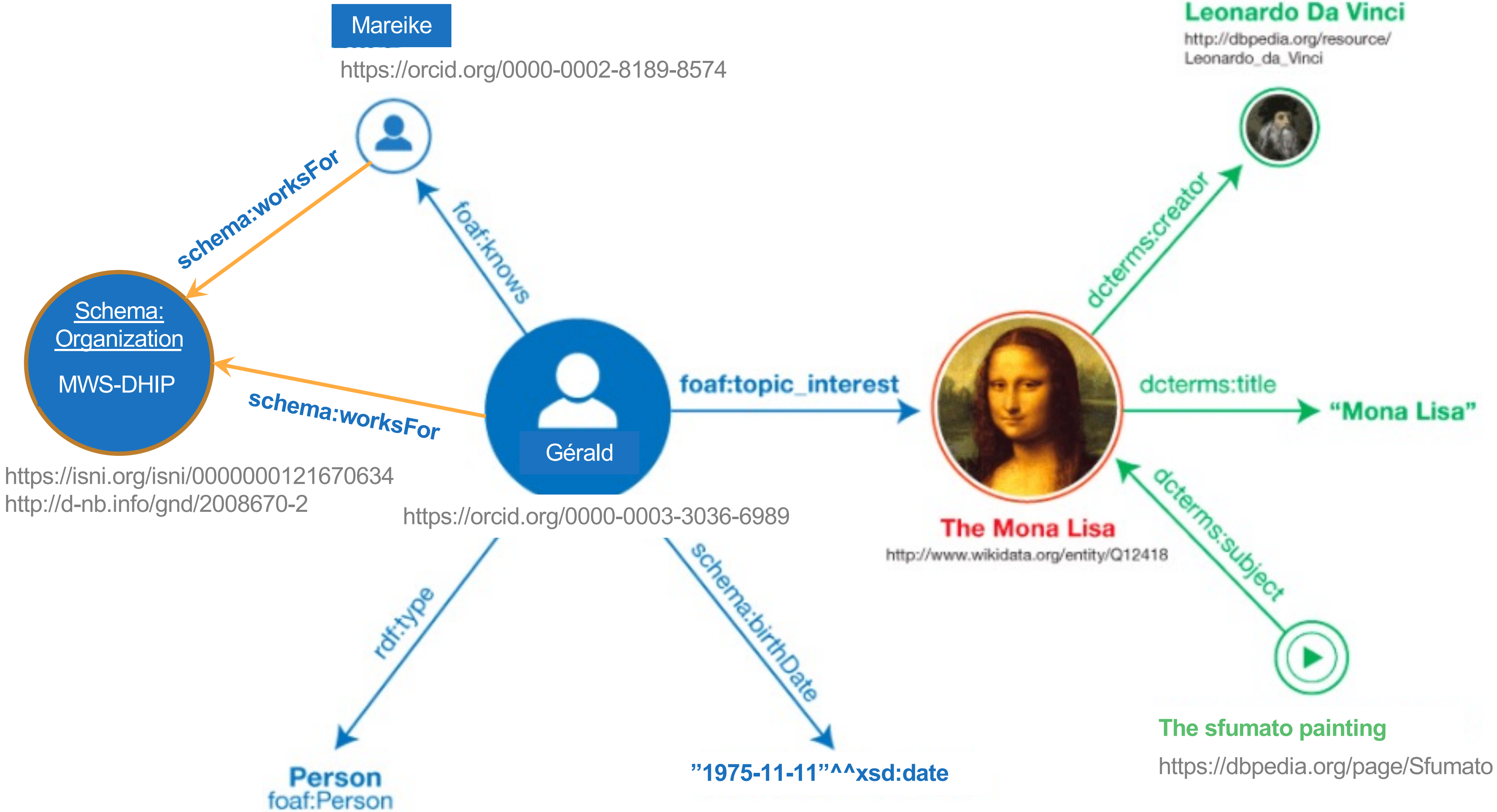
```html
<p itemscope itemtype="http://schema.org/AgreeAction">
    I
    <!-- me, by the way, identified with my orcid id -->
    <span itemscope itemprop="agent"
          itemtype="http://schema.org/Person"
          itemid="https://orcid.org/0000-0003-3036-6989">
          <meta itemprop="name" content="Gérald Kembellec">
    </span>
    do agree with
    <!-- with a book identified by its isbn id -->
    <cite itemscope
          itemtype="http://schema.org/Book">
          <meta itemprop="isbn" content="0316769487">
          <span itemscope itemprop="author"
                itemtype="http://schema.org/Person"
                itemid="http://d-nb.info/gnd/118629743" >
                <span itemprop="name">Max Weber</span>'s
          </span>
          <span itemprop="name">
                Sociology of Religion
          </span>
    </cite>
</p>
```

Book

All (1)

Book 0 ERREUR 0 AVERTISSEMENT

@type	Book
isbn	0316769487
name	Sociology of Religion
author	
@type	Person
@id	http://d-nb.info/gnd/118629743
name	Max Weber



Mareike

https://orcid.org/0000-0002-8189-8574

Leonardo Da Vinci

http://dbpedia.org/resource/
Leonardo_da_Vinci

schema:worksFor

foaf:knows

dcterms:creator

Schema:
Organization

MWS-DHIP

foaf:topic_interest

dcterms:title

“Mona Lisa”

schema:worksFor

Gérald

https://isni.org/isni/0000000121670634
http://d-nb.info/gnd/2008670-2

https://orcid.org/0000-0003-3036-6989

The Mona Lisa

http://www.wikidata.org/entity/Q12418

dcterms:subject

rdf:type

schema:birthDate

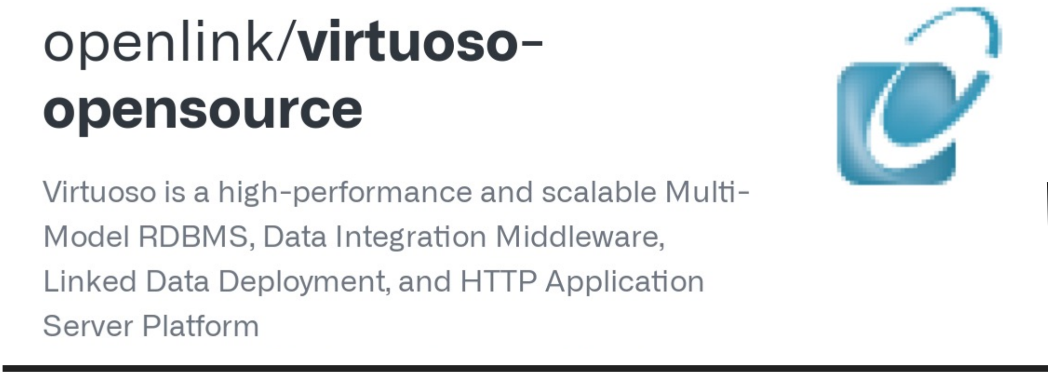
Person
foaf:Person

”1975-11-11”^^xsd:date

The sfumato painting

https://dbpedia.org/page/Sfumato

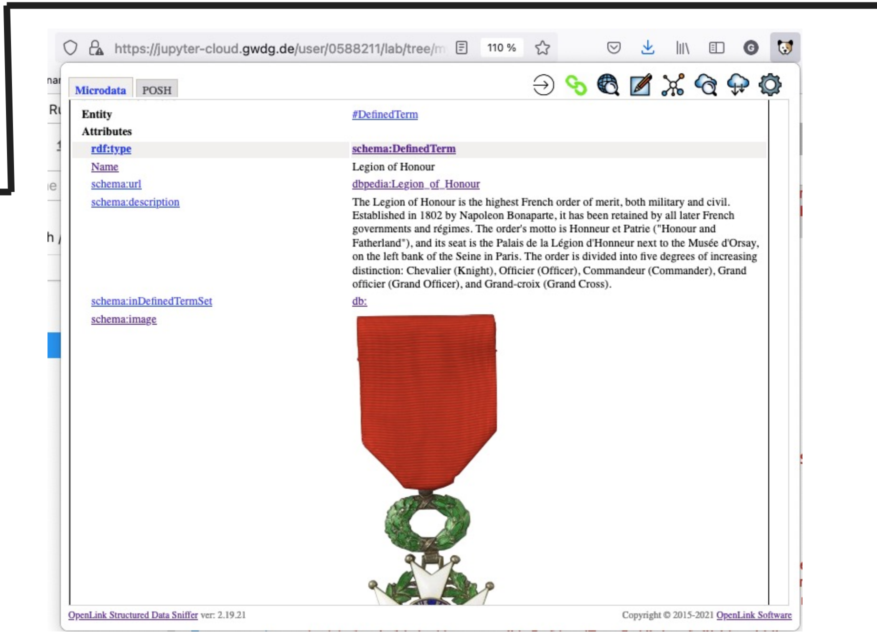
Openlink Softwares used in the demonstration



openlink/**virtuoso-opensource**

Virtuoso is a high-performance and scalable Multi-Model RDBMS, Data Integration Middleware, Linked Data Deployment, and HTTP Application Server Platform

SparQL Enpoint : Virtuoso
https://virtuoso.openlinksw.com/



Entity	#DefinedTerm
Attributes	
rdf:type	**schema:DefinedTerm**
Name	Legion of Honour
schema:url	dbpedia:Legion_of_Honour
schema:description	The Legion of Honour is the highest French order of merit, both military and civil. Established in 1802 by Napoleon Bonaparte, it has been retained by all later French governments and régimes. The order's motto is Honneur et Patrie ("Honour and Fatherland"), and its seat is the Palais de la Légion d'Honneur next to the Musée d'Orsay, on the left bank of the Seine in Paris. The order is divided into five degrees of increasing distinction: Chevalier (Knight), Officier (Officer), Commandeur (Commander), Grand officier (Grand Officer), and Grand-croix (Grand Cross).
schema:inDefinedTermSet	db:
schema:image	

OpenLink Structured Data Sniffer ver: 2.19.21

Copyright © 2015-2021 OpenLink Software

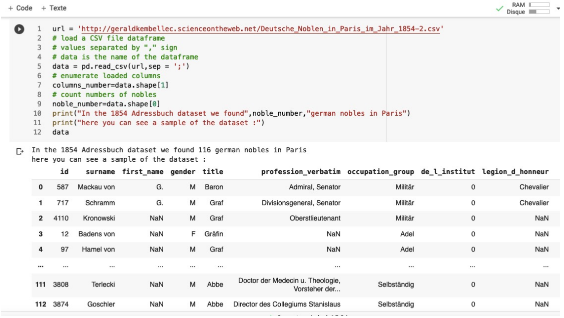
Browser's plugin
for semantic discovering tool :
Open Link structured data sniffer
https://addons.mozilla.org/fr/firefox/addon/openlink-structured-data-sniff/

2nd concept :
The Jupyter Notebook

What is a Jupyter notebook ?

This is a notebook

This is a laptop computer notebook



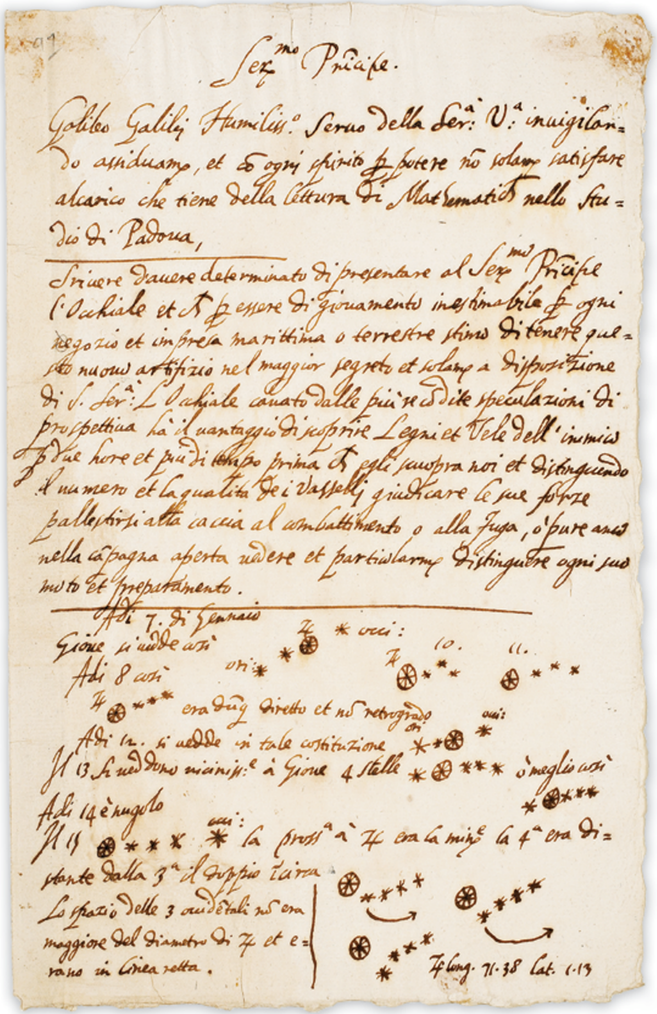
```
url = 'http://geraldkembellec.scienceontheweb.net/Deutsche_Noblen_in_Paris_im_Jahr_1854-2.csv'
# load a CSV file dataframe
# values separated by "," sign
# data is the name of the dataframe
data = pd.read_csv(url,sep = ';')
# enumerate loaded columns
columns_number=data.shape[1]
# count numbers of nobles
noble_number=data.shape[0]
print("In the 1854 Adressbuch dataset we found",noble_number,"german nobles in Paris")
print("here you can see a sample of the dataset :")
data
```

```
In the 1854 Adressbuch dataset we found 116 german nobles in Paris
here you can see a sample of the dataset :
```

	id	surname	first_name	gender	title	profession_verbatim	occupation_group	de_l_institut	legion_d_honneur
0	587	Mackau von	G.	M	Baron	Admiral, Senator	Militär	0	Chevalier
1	717	Schramm	G.	M	Graf	Divisionsgeneral, Senator	Militär	0	Chevalier
2	4110	Kronowski	NaN	M	Graf	Oberstlieutenant	Militär	0	NaN
3	12	Badens von	NaN	F	Gräfin	NaN	Adel	0	NaN
4	97	Hamel von	NaN	M	Graf	NaN	Adel	0	NaN
...	...	...	...	...	...	...	...	...	...
111	3808	Terlecki	NaN	M	Abbe	Doctor der Medecin u. Theologie, Vorsteher der...	Selbständig	0	NaN
112	3874	Goschler	NaN	M	Abbe	Director des Collegiums Stanislaus	Selbständig	0	NaN

This is a Jupyter notebook

The first "notebook" by Galileo : manuscript from 17th century : "Observations of Jupiter and four of its moons".



Ser.mo Principe.

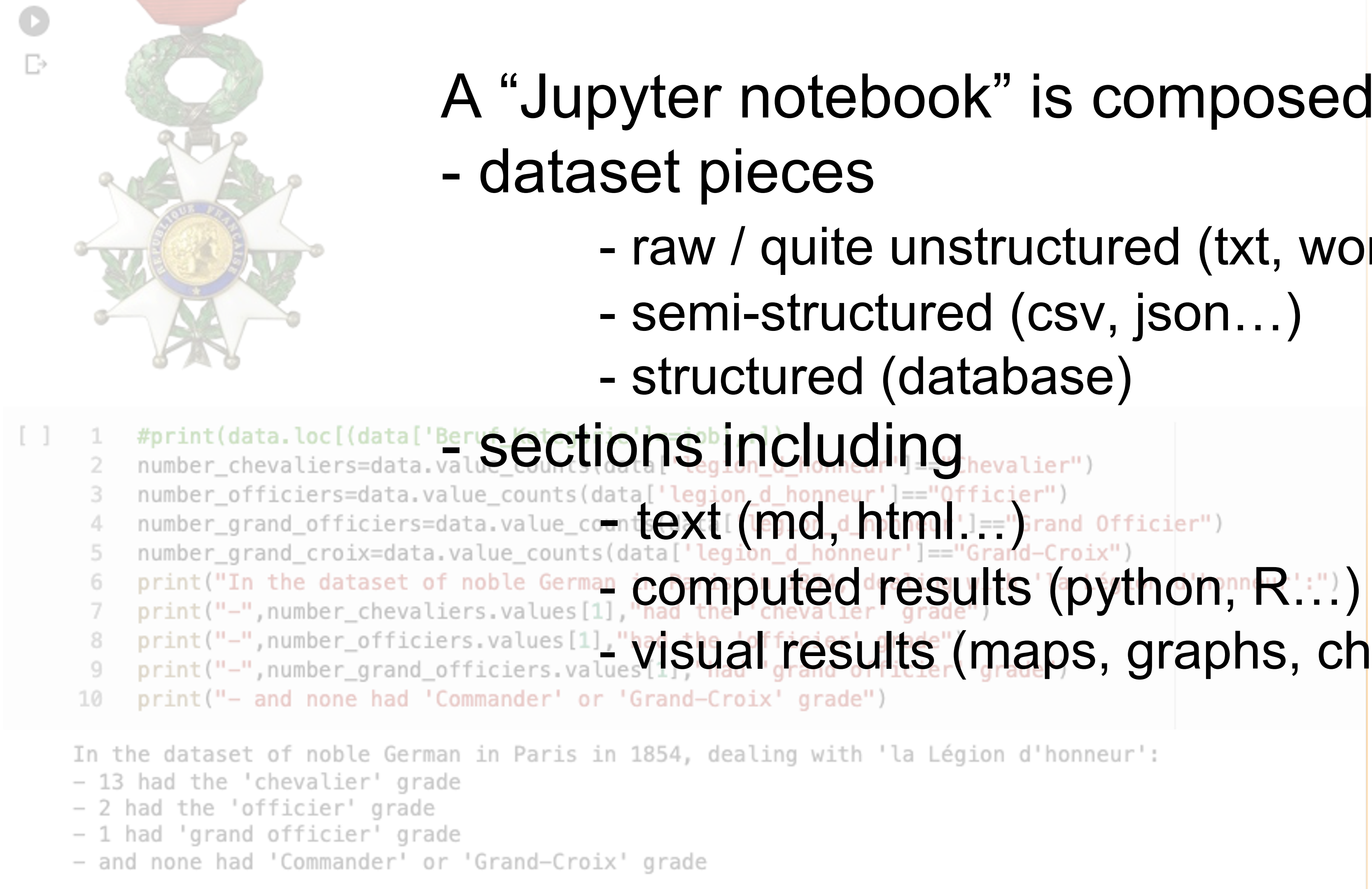
Galileo Galilei Humiliss.o Seruo della Ser.tà V.a inuigilando assiduam.te, et con ogni spirito per potere non solam.te satisfare al carico che tiene della Lettura di Matematiche nello Studio di Padova,

Sriuere dauere determinato di presentare al Ser.mo Principe l'Occhiale et quello per essere di giouamento inestimabile per ogni negozio et impresa marittima o terrestre stimo di tenere questo nuouo artifizio nel maggior segreto et solam.te a dispositione di V. Ser.tà L'Occhiale cauato dalle più recondite speculazioni di prospettiua ha il uantaggio di scoprire Legni et Vele dell'inimico per due hore et più di tempo prima che egli scuopra noi et distinguendo il numero et la qualità dei Vasselli giudicare le sue forze pallestirsi alla caccia al combattimento o alla fuga, o pure ancora nella campagna aperta uedere et particolarm.te distinguere ogni suo moto et preparamento.

Adi 7. di Gennaio
Giove si vedde così ♃ occ:
Adi 8 così ori: ♃ 10. ♃ 11.
♃ era dunq diretto et non retrogrado
Adi 12. si vedde in tale costituzione ori occ:
Il 13 si veddono vicinissime à Giove 4 stelle o meglio così
Adi 14 è nugolo
Il 15 occ: la pross.a à ♃ era la min.e la 4.a era distante dalla 3.a il doppio circa
Lo spatio delle 3 occidentali non era maggiore del diametro di ♃ et erano in linea retta.
♃ long. 71.38 lat. 1.13

A “Jupyter notebook” is composed by :

- dataset pieces
  - raw / quite unstructured (txt, word…)
  - semi-structured (csv, json…)
  - structured (database)
- sections including
  - text (md, html…)
  - computed results (python, R…)
  - visual results (maps, graphs, charts…)



```
#print(data.loc[(data['Beruf_Kategorie']=='nobel')])
number_chevaliers=data.value_counts(data['legion_d_honneur']=="Chevalier")
number_officiers=data.value_counts(data['legion_d_honneur']=="Officier")
number_grand_officiers=data.value_counts(data['legion_d_honneur']=="Grand Officier")
number_grand_croix=data.value_counts(data['legion_d_honneur']=="Grand-Croix")
print("In the dataset of noble German in Paris in 1854, dealing with 'la Légion d'honneur':")
print("-",number_chevaliers.values[1],"had the 'chevalier' grade")
print("-",number_officiers.values[1],"had the 'officier' grade")
print("-",number_grand_officiers.values[1],"had 'grand officier' grade")
print("- and none had 'Commander' or 'Grand-Croix' grade")
```

```
In the dataset of noble German in Paris in 1854, dealing with 'la Légion d'honneur':
- 13 had the 'chevalier' grade
- 2 had the 'officier' grade
- 1 had 'grand officier' grade
- and none had 'Commander' or 'Grand-Croix' grade
```

A “Jupyter notebook” can be seen as a scientific production :

- with its own formalism
- including

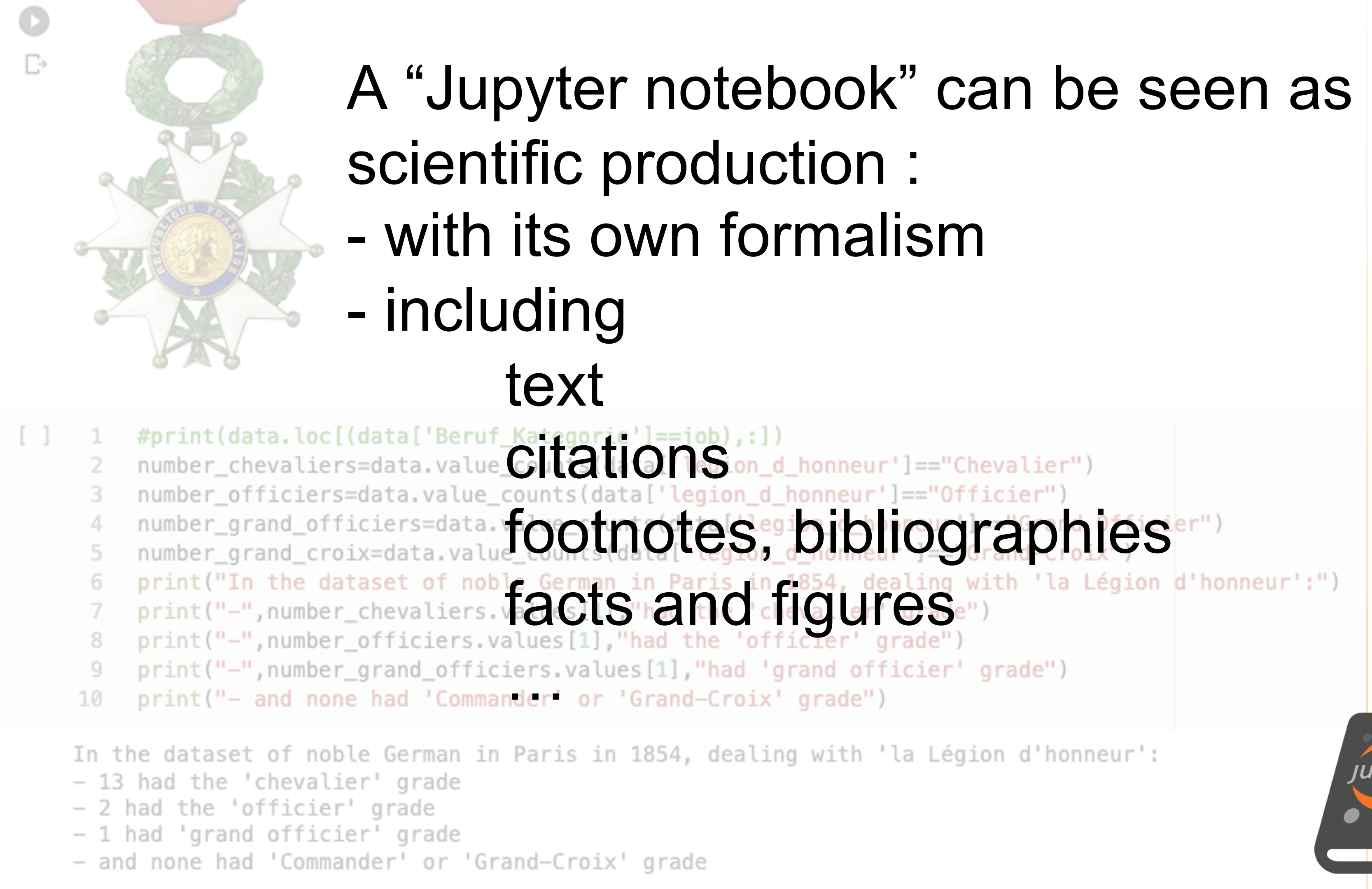
  text

  citations

  footnotes, bibliographies

  facts and figures

  …



```
#print(data.loc[(data['Beruf_Kategorie']==job),:])
number_chevaliers=data.value_counts(data['legion_d_honneur']=="Chevalier")
number_officiers=data.value_counts(data['legion_d_honneur']=="Officier")
number_grand_officiers=data.value_counts(data['legion_d_honneur']=="Grand Officier")
number_grand_croix=data.value_counts(data['legion_d_honneur']=="Grand Croix")
print("In the dataset of noble German in Paris in 1854, dealing with 'la Légion d'honneur':")
print("-",number_chevaliers.values[1],"had the 'chevalier' grade")
print("-",number_officiers.values[1],"had the 'officier' grade")
print("-",number_grand_officiers.values[1],"had 'grand officier' grade")
print("- and none had 'Commander' or 'Grand-Croix' grade")
```

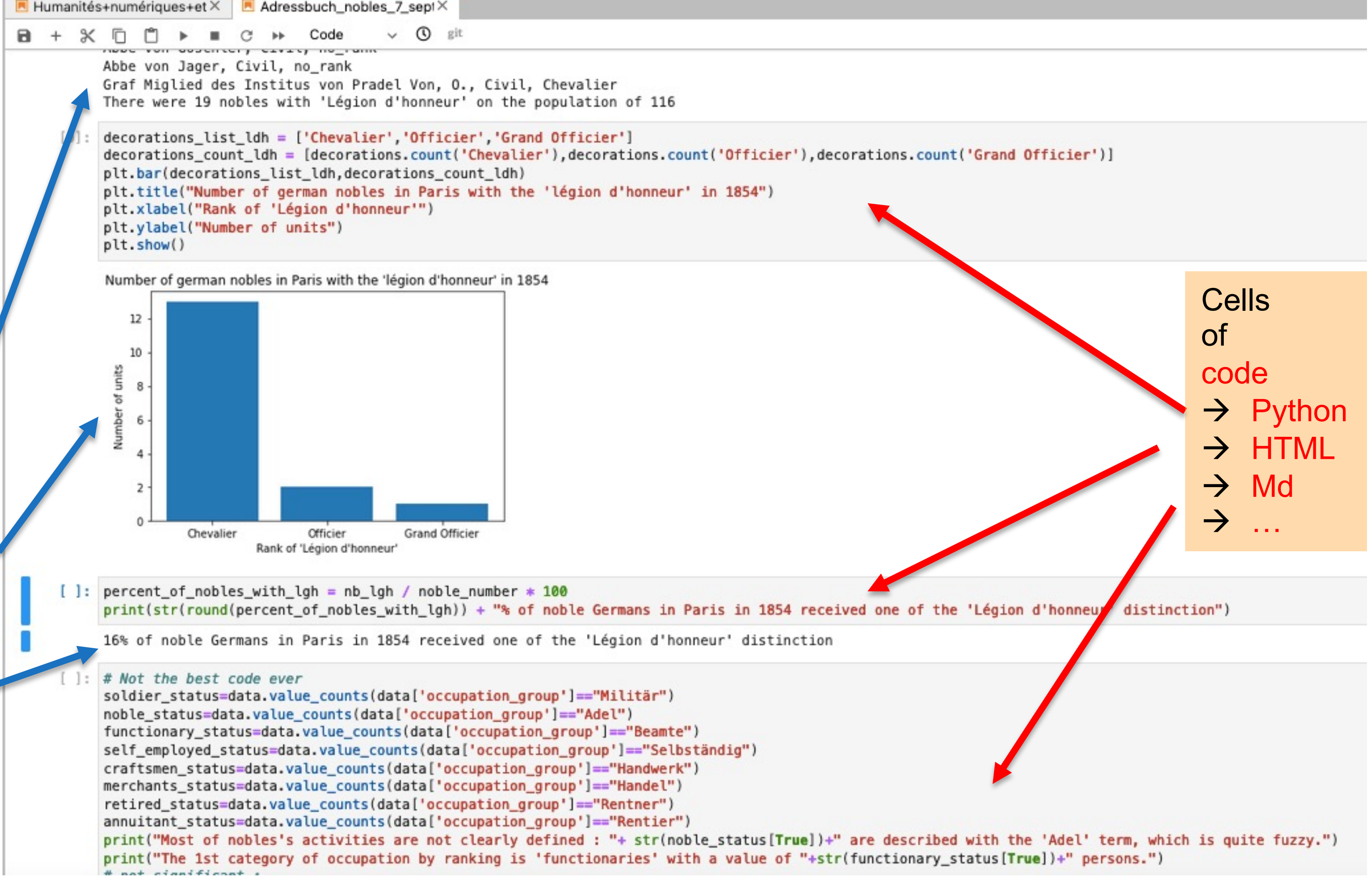
```
In the dataset of noble German in Paris in 1854, dealing with 'la Légion d'honneur':
- 13 had the 'chevalier' grade
- 2 had the 'officier' grade
- 1 had 'grand officier' grade
- and none had 'Commander' or 'Grand-Croix' grade
```

What does it look like ?

Cells of results

- → Text
- → Pics
- → Graphs
- → Charts…

Can include semantics contents ☺



```
Abbe von Jager, Civil, no_rank
Graf Miglied des Institus von Pradel Von, O., Civil, Chevalier
There were 19 nobles with 'Légion d'honneur' on the population of 116
```

```
decorations_list_ldh = ['Chevalier','Officier','Grand Officier']
decorations_count_ldh = [decorations.count('Chevalier'),decorations.count('Officier'),decorations.count('Grand Officier')]
plt.bar(decorations_list_ldh,decorations_count_ldh)
plt.title("Number of german nobles in Paris with the 'légion d'honneur' in 1854")
plt.xlabel("Rank of 'Légion d'honneur'")
plt.ylabel("Number of units")
plt.show()
```

```
[ ]: percent_of_nobles_with_lgh = nb_lgh / noble_number * 100
     print(str(round(percent_of_nobles_with_lgh)) + "% of noble Germans in Paris in 1854 received one of the 'Légion d'honneur' distinction")
```

```
16% of noble Germans in Paris in 1854 received one of the 'Légion d'honneur' distinction
```

```
[ ]: # Not the best code ever
     soldier_status=data.value_counts(data['occupation_group']=="Militär")
     noble_status=data.value_counts(data['occupation_group']=="Adel")
     functionary_status=data.value_counts(data['occupation_group']=="Beamte")
     self_employed_status=data.value_counts(data['occupation_group']=="Selbständig")
     craftsmen_status=data.value_counts(data['occupation_group']=="Handwerk")
     merchants_status=data.value_counts(data['occupation_group']=="Handel")
     retired_status=data.value_counts(data['occupation_group']=="Rentner")
     annuitant_status=data.value_counts(data['occupation_group']=="Rentier")
     print("Most of nobles's activities are not clearly defined : "+ str(noble_status[True])+" are described with the 'Adel' term, which is quite fuzzy.")
     print("The 1st category of occupation by ranking is 'functionaries' with a value of "+str(functionary_status[True])+" persons.")
```

Cells of code

- → Python
- → HTML
- → Md
- → …

Ok, what do we need to do that ?

- A notebook infrastructure
    - https://jupyter.org/try
    - https://jupyter-cloud.gwdg.de/ (DE)
    - https://colab.research.google.com/

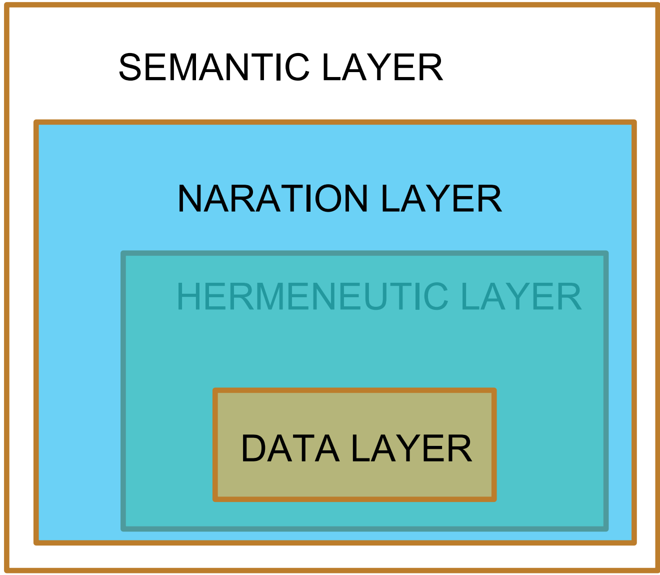
- A research dataset (txt, json, geojson, csv…)
- A programming language (Python, R, Shell…)
- A presentation language (HTML or markdown)
- A metadata exposing language (COinS, RDFa, microdata…)
- Lot of work…

Does it exists for real in the history field ?

Journal of Digital History

C²DH

DE G



SEMANTIC LAYER

NARATION LAYER

HERMENEUTIC LAYER

DATA LAYER

https://www.c2dh.uni.lu/news/digital-history-journal-transmedia-storytelling-hermeneutics-data

Let’s take a tour on Adressbuch1854’s project

… and go for demonstration…

… then we will discuss “semantic publishing” and “notebooks”