Cours Web Semantique et Ontologies
M2P Génie Informatique UFR IM2AG – UGA

Philippe GENOUD – Danielle ZIEBELIN - LIG-Steamer
Prenom.Nom@imag.fr

TP Web Des Données
Données Liées (Linked Data)

Ce TP est une adaptation de :

What Is This Thing Called Linked Data ? – Ph. GENOUD, M. ATENCIA, J. DAVID
ACM-DocEng2015 Tutorial – 9 Septembre 2015 – Lausanne

Hands-on session

- Part I: From an Excel data file to linked open data
  - you will learn how to
    - create "cool" URIs
    - describe things with RDF
    - make links to other datasets

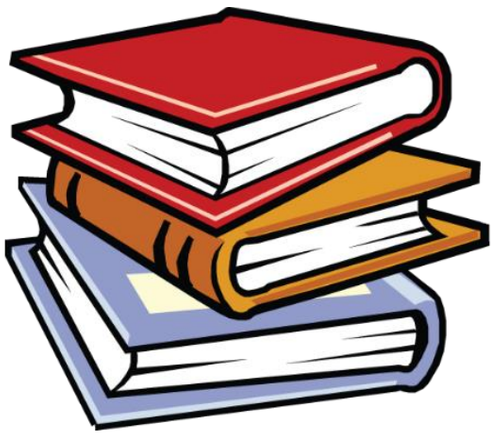
- Part II: Querying linked data with SPARQL
  - you will learn how to make queries with SPARQL
  - we will use Apache Jena - ARQ command line applications

Hands-on session

Introduction
Case study: Artemis Bookstore

- Artemis owns a bookstore. She has heard of the Linked Data technologies, but she is not fully convinced of the benefits of using these technologies.
- Artemis has provided you with a sample of her data in an Excel file: artemisBookstoreData.xlsx
- Let's convince Artemis to join Linked Open Data!

Artemis Bookstore



Hands-on session (I)

- Part I: From an Excel data file to linked open data
  - you will learn how to
    - create "cool" URIs
    - describe things with RDF
    - make links to other datasets
- Part II: Querying linked data with SPARQL
  - you will learn how to make queries with SPARQL
  - we will use Apache Jena - ARQ command line applications

Hands-on session (I)

Artemis' bookstore data

- Let's have a look at `artemisBookstoreData.xlsx`

	A	B	C	D	E	F
1	**LastName**	**FirstName**	**BookTitle**	**PagesNb**	**Isbn**	**PublisherName**
2	Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
3	Abbey	Edward	Fire on the Mountain (Abbey novel	211	0-8263-0457-5	Dial Press
4	Abbey	Edward	Hayduke Lives	352	ISBN 0-316-00411-1 (first edition,	Little, Brown and Company
5	Abbey	Edward	Jonathan Troy	374	1-131-40684-2	Dodd, Mead and Company
6	Abbey	Edward	The Monkey Wrench Gang	352	ISBN 0-397-01084-2 (hardback ed	Lippincott Williams & Wilkins
7	Abbey	Edward	The Brave Cowboy	277	0-8263-0448-6	Dodd, Mead and Company
8	Abbey	Edward	The Fool's Progress	485	0-8050-0921-3	Henry Holt and Company
9	Acito	Marc	How I Paid for College: A Novel of S	288	ISBN 0-7679-1841-X (first edition,	Broadway Books
10	Acker	Kathy	Blood and Guts in High School	165	ISBN 0-8021-3193-X (Paperback e	Grove Press
11	Adler	Warren	The War of the Roses (novel)	263	0-446-51220-6	Hachette Book Group
12	[illegible]	[illegible]	[illegible]	288	[illegible]	[illegible]

Each line correspond to a different book.

For each book we have :

- the author first and last name
- the book title
- the number of pages
- the ISBN
- the publisher's name

Hands-on session (I)

- Now, let's create RDF data!
- But before, let's recall the Linked Data principles:

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

Hands-on session (I)

- Now, let's create RDF data!
- But before, let's recall the Linked Data principles:
  1. Use URIs as names for things.
  2. Use HTTP URIs, so that people can look up those names
  3. When someon … n using the stan
  4. Include links to other URIs, so that they can discover more

What are the things (Resources) we want to describe ?

	A	B	C	D	E	F
1	LastName	FirstName	BookTitle	PagesNb	Isbn	PublisherName
2	Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
3	Abbey	Edward	Fire on the Mountain (Abbey novel	211	0-8263-0457-5	Dial Press
4	Abbey	Edward	Hayduke Lives	352	ISBN 0-316-00411-1 (first edition,	Little, Brown and Company
5	Abbey	Edward	Jonathan Troy	374	1-131-40684-2	Dodd, Mead and Company
6	Abbey	Edward	The Monkey Wrench Gang	352	ISBN 0-397-01084-2 (hardback ed	Lippincott Williams & Wilkins
7	Abbey	Edward	The Brave Cowboy	277	0-8263-0448-6	Dodd, Mead and Company
8	Abbey	Edward	The Fool's Progress	435	[illegible]	Henry Holt and Company
9	Acito	Marc	How I Paid for College: A Novel of S	288	ISBN 0-7679-1841-X (first edition,	Broadway Books
10	Acker	Kathy	Blood and Guts in High School	165	ISBN 0-8021-3193-X (Paperback e	Grove Press
11	Adler	Warren	The War of the Roses (novel)	264	0-446-51220-6	Hachette Book Group
12	Aitk	Will	The Hunte (novel)	288	[illegible]	Gay Men's Press

- Propose a RDF graph (without taking care of URIs) for the two first lines of the Spreadsheet

Hands-on session (I)

Creating RDF data

What are the things (Resources) we want to describe ?

	LastName	FirstName	BookTitle	PagesNb	Isbn	PublisherName
2	Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
109	Auster	Paul	Leviathan (Auster novel)	275	0-14-017813-9	Viking Press
110	Auster	Paul	Moon Palace	320	0-670-82509-3	Viking Press

	LastName	FirstName	BookTitle	PagesNb	Isbn	PublisherName
2	Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
3	Abbey	Edward	Fire on the Mountain (Abbey novel	211	0-8263-0457-5	Dial Press

- Book centric approach

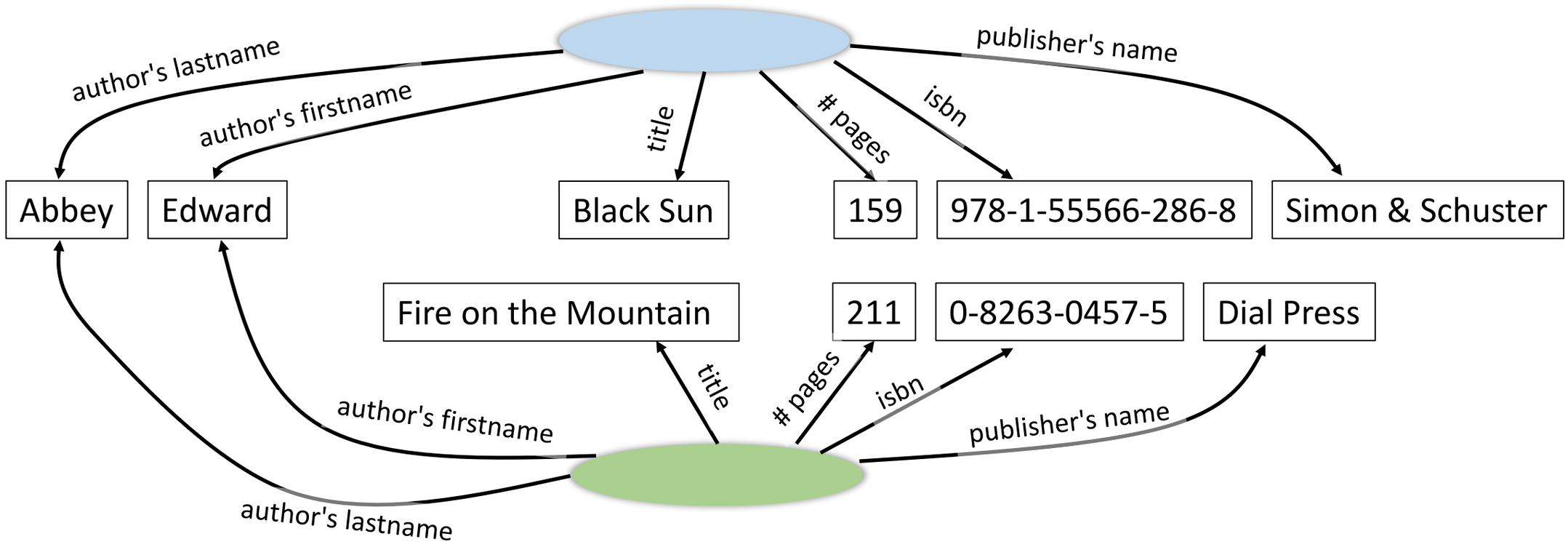
Hands-on session (I)

Creating RDF data

What are the things (Resources) we want to describe ?

	LastName	FirstName	BookTitle	PagesNb	Isbn	PublisherName
2	Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
3	Abbey	Edward	Fire on the Mountain (Abbey novel	211	0-8263-0457-5	Dial Press
4	Abbey	Edward	[illegible]	352	[illegible]	[illegible]

- Book centric approach – One kind of data: books

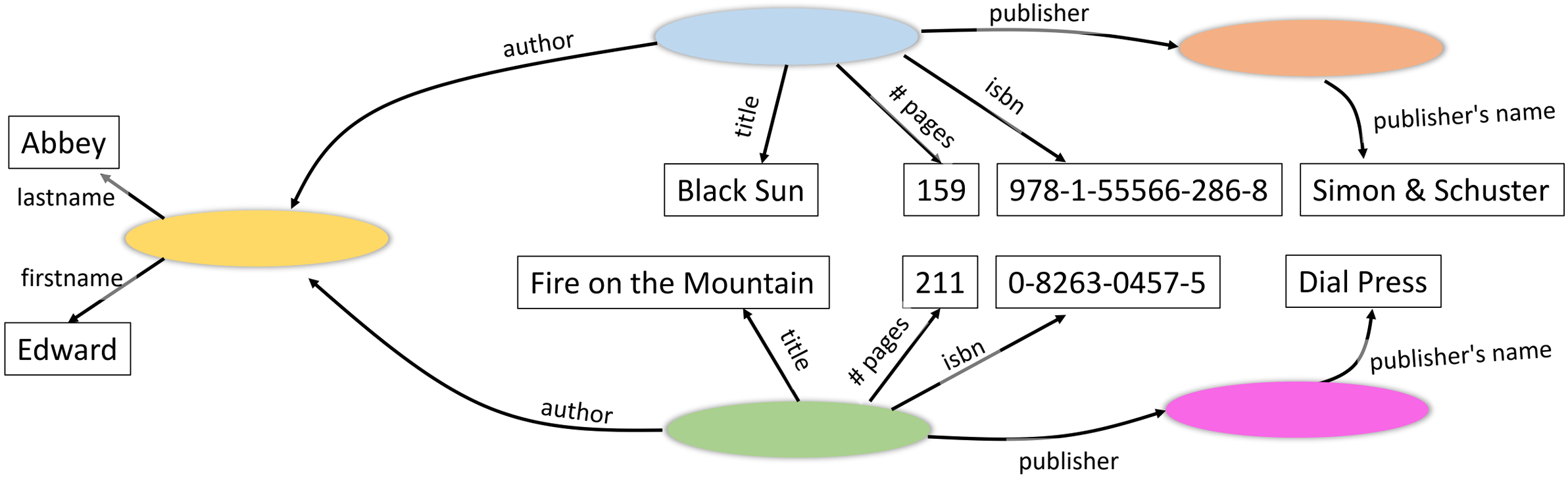


Hands-on session (I)

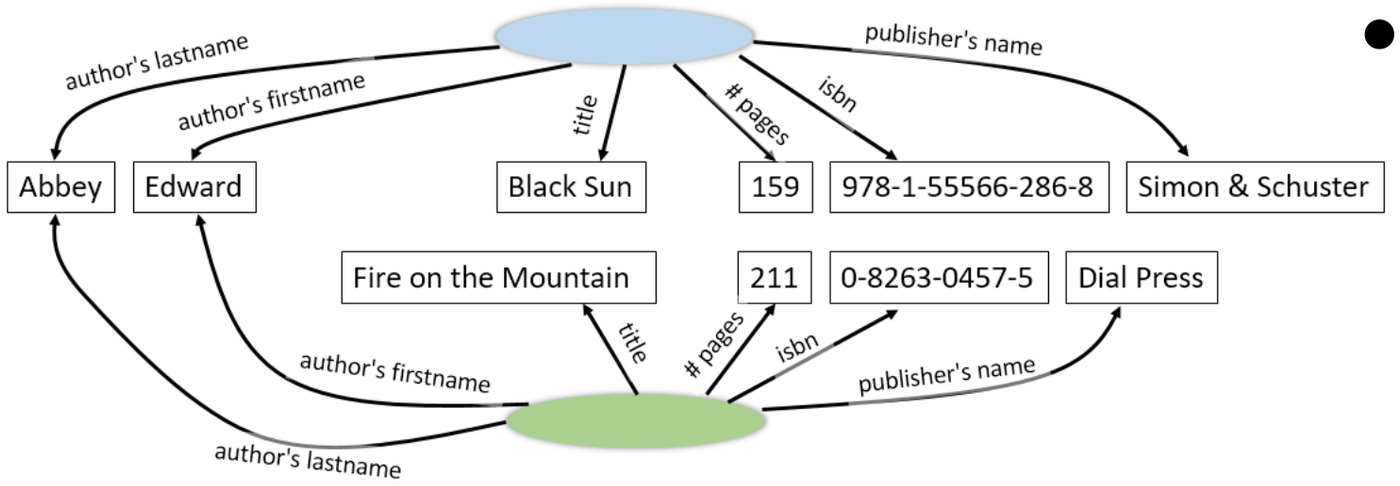
What are the things (Resources) we want to describe ?

	LastName	FirstName	BookTitle	PagesNb	Isbn	PublisherName
2	Abbey	Edward	Black Sun (Edward Abbey novel)	159	978-1-55566-286-8	Simon & Schuster
3	Abbey	Edward	Fire on the Mountain (Abbey novel	211	0-8263-0457-5	Dial Press
4	Abbey	Edward	Hayduke Lives	352	ISBN 0-316-00411-1 (first edition, hardb	Little, Brown and Company

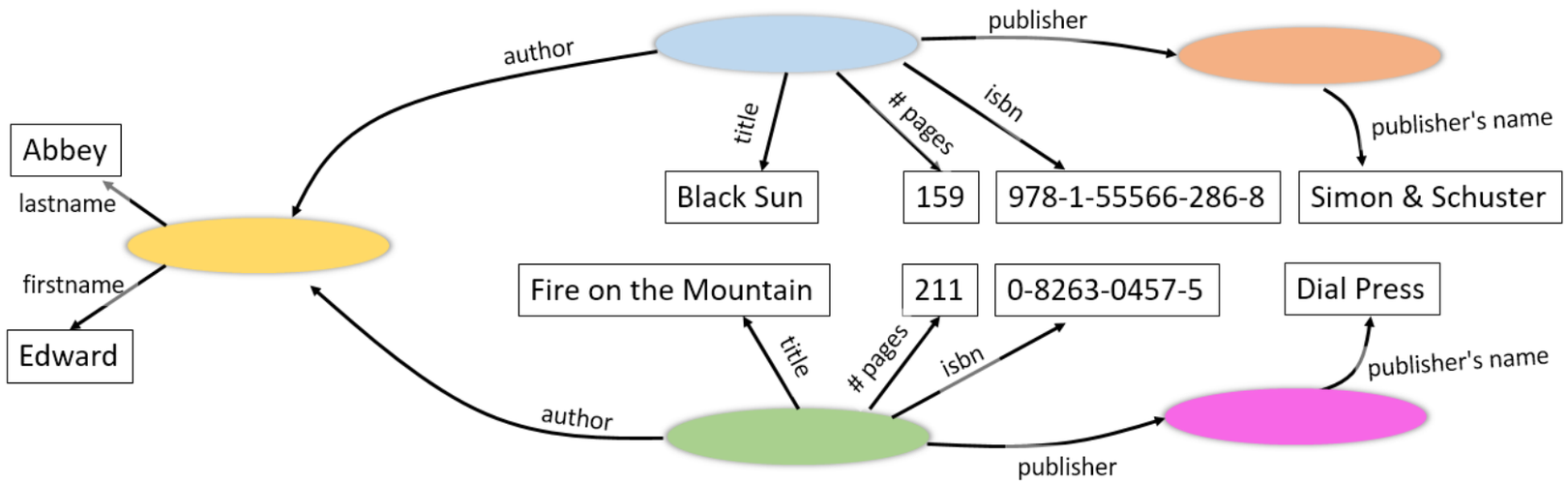
- Three kinds of data: books, authors, publishers



Hands-on session (I)



- simplicity ?

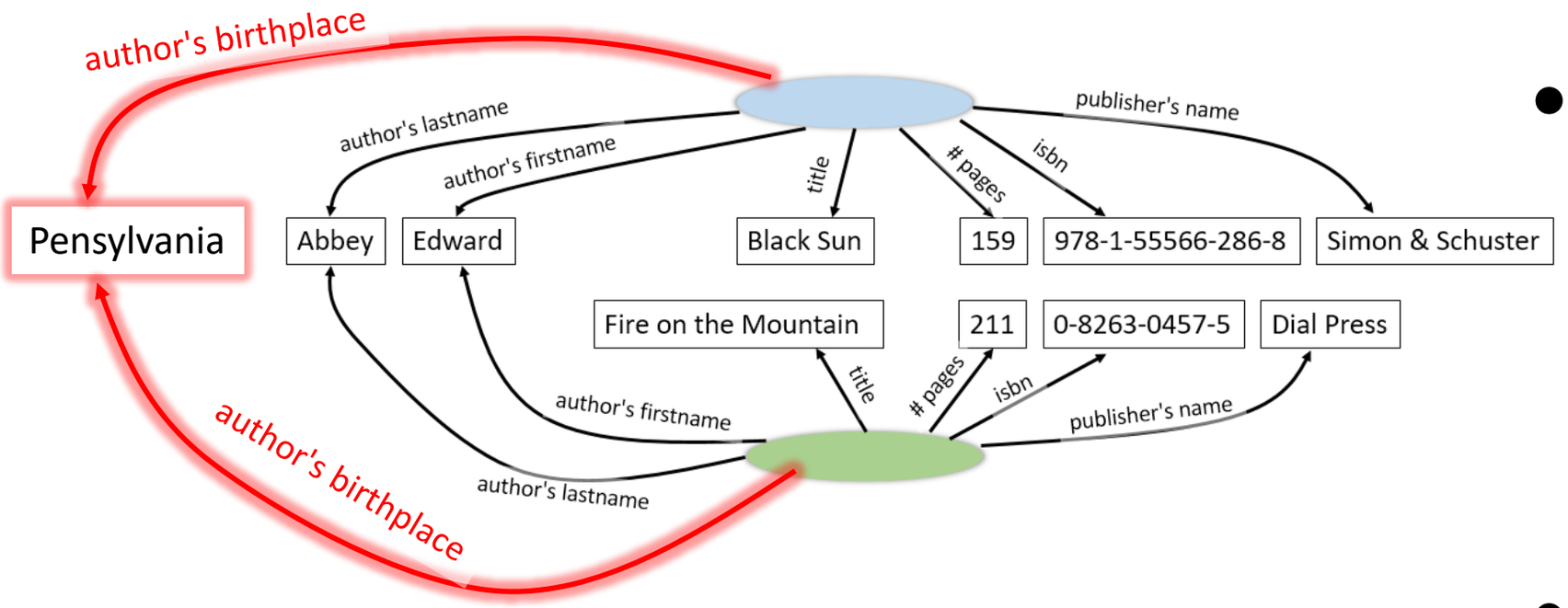


Creating RDF data

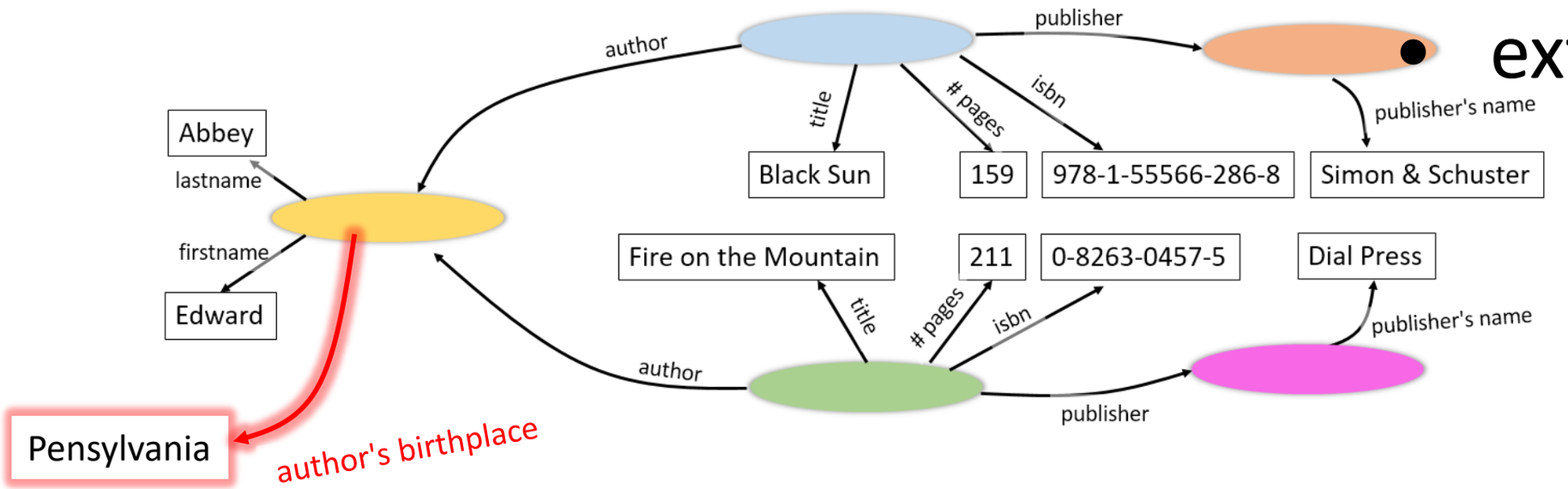
Hands-on session (I)

let's add extra information, i.e birthplace of author's.

Edward Abbey is born in Pensylvania



- simplicity ?
- flexibility
- extensibility



Hands-on session (I)

- Now, let's create RDF data!
- But before, let's recall the Linked Data principles:

1. Use URIs as names for things.

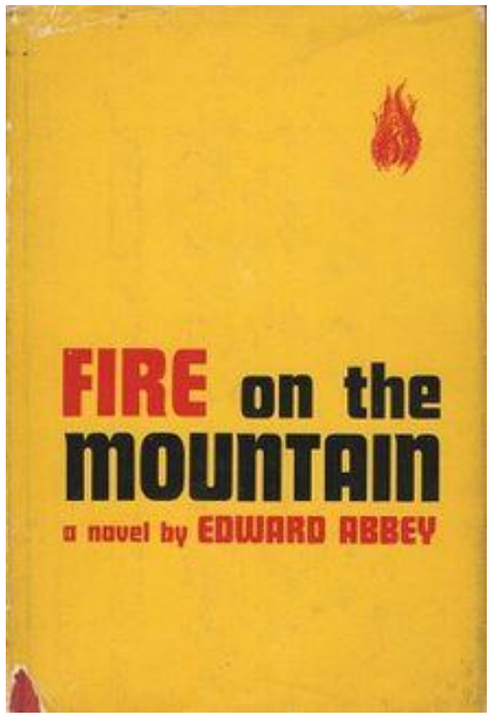
2. Use HTTP URIs, so that people can look up those names.

3. When someone looks up a URI, provide useful information using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

Hands-on session (I)

Creating RDF data
Cool URIs

- Let's consider the book "Fire on the Mountain" written by Edward Abbey. Can you find a cool URI for this resource?



REMEMBER!
web document ≠ web resource
Actually, you should find
- A URI for the real object itself.
- A URI for the related information resource that describes the real-world object and has an HTML representation.
- A URI for a related information resource that describes the real-world object and has an RDF/XML representation.

Hands-on session (I)

Creating RDF data
Cool URIs

- Let's consider the paperback book "Invisible" written by Paul Auster. Can you find a cool URI for this resource?

  - uncool URIs

    keep out of namespaces you do not control

    - `https://www.amazon.com/Fire-Mountain-Edward-Abbey/dp/0062193902`

      abstract away from implementation details

    - `http://artemisbookstore.com:8080/book.php?title=Fire+on+the+Mountain&author=Edward-Abbey&format=rdf`
    - `http://localhost:3333/Fire-on-the-Moutain-Edward-Abbey`

  - cool URIs

    this local name must be **unique** and **persistent.** i.e. ISBN or generated UUID (Universal Unique Identifier)

    - `http://artemisbookstore.com/resource/0-8263-0457-57`
    - `http://artemisbookstore.com/page/0-8263-0457-57.html`
    - `http://artemisbookstore.com/data/0-8263-0457-57.rdf`

more info at http://www.w3.org/TR/cooluris/

Hands-on session (I)

Creating RDF data
Cool URIs

- Let's consider the paperback book "Invisible" written by Paul Auster. Can you find a cool URI for this resource?
  - uncool URIs

keep out of namespaces you do not control

    - `https://www.amazon.com/Fire-Mountain-Edward-Abbey/dp/0062193902`

abstract away from implementation details

    - `http://artemisbookstore.com:8080/book.php?title=Fire+on+the+Mountain&author=Edward-Abbey&format=rdf`
    - `http://localhost:3333/Fire-on-the-Moutain-Edward-Abbey`
  - cool URIs
    - `http://id.artemisbookstore.com/book/0-8263-0457-57`
    - `http://page.artemisbookstore.com/book/0-8263-0457-57.html`
    - `http://data.artemisbookstore.com/book/0-8263-0457-57.rdf`

more info at http://www.w3.org/TR/cooluris/

Hands-on session (I)

- Now, let's create RDF data! But before
- Let's recall the Linked Data principles:

1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. **When someone looks up a URI, provide useful information using the standards (RDF, SPARQL).**
4. Include links to other URIs, so that they can discover more things.

Hands-on session (I)

Creating RDF data
Describing things with RDF

- Which kind of information should we include in the description of the book "Invisible" written by Paul Auster?
  - triples that describe the resource with literals (e.g. title)
  - triples that describe the resource by linking *to* other resources, or *outgoing links* (e.g. author)
  - triples that describe the resource by linking *from* other resources, or *incoming links* (e.g. has written)
  - triples describing related resources (e.g. the name of the resource's creator)
  - triples describing the description itself (e.g. licensing terms)
  - triples about the broader dataset of which this description is a part

Hands-on session (I)

Creating RDF data
Describing things with RDF

- Which kind of information should we include in the description of the book "Invisible" written by Paul Auster?
  - triples that describe the resource with literals (e.g. title)
  - triples that describe the resource by linking *to* other resources, or *outgoing links* (e.g. author)
  - triples that describe the resource by linking *from* other resources, or *incoming links* (e.g. has written)
  - triples describing related resources (e.g. the name of the resource's creator)
  - triples describing the description itself (e.g. licensing terms)
  - triples about the broader dataset of which this description is a part

Hands-on session (I)

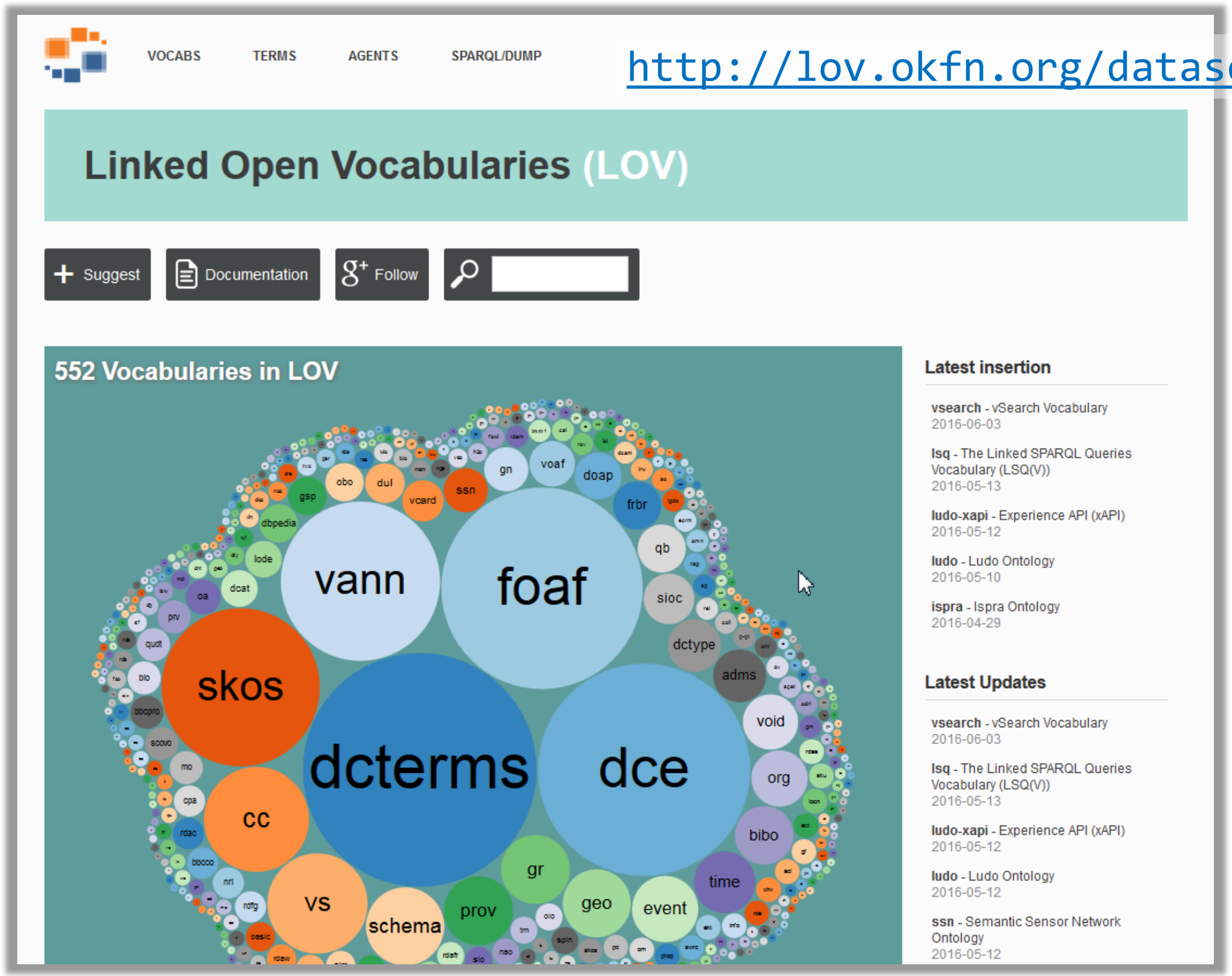
Creating RDF data
Describing things with RDF

- Which kind of information should we include in the description of the book "Invisible" written by Paul Auster?
  - triples that describe the resource with literals (e.g. title)
  - triples that describe the resource by linking *to* other resources, or *o*
  - triples that des
    resources, or *in*
  - triples describi
    resource's crea
  - triples describi
  - triples about th
    a part

and, for writing this description, you can use:
- your own defined terms
- terms from existing vocabularies
  - Dublin Core (e.g. dcterms:title)
  - FOAF (e.g. foaf:name)
  - RDFS (rdfs:comment and rdfs:label for annotations),
  - OWL (owl:sameAs for links)
  - ...

Hands-on session (I)

Reusing Existing Terms

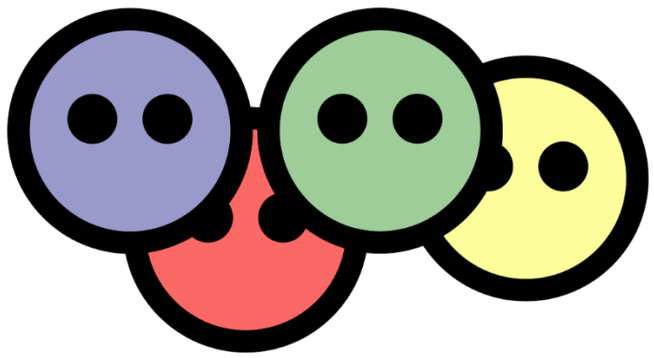


Hands-on session (I)

Reusing Existing Terms
FOAF vocabulary

- The FOAF ("Friend of a Friend") vocabulary can be used for describing persons, their activities and their relations to other people and objects.
- FOAF = RDF + Social Web
- http://xmlns.com/foaf/spec/
- namespace:

  foaf: <http://xmlns.com/foaf/0.1>



Hands-on session (I)

Reusing Existing Terms
FOAF vocabulary

- An example:

```
@base <http://exmo.inrialpes.fr/about> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<#jdavid>
  a foaf:Person ;
  foaf:name "Jérôme David"@fr ;
  foaf:nick "JD" ;
  foaf:mbox <mailto:jerome.david@i
  foaf:homepage "http://exmo.inria
  foaf:depiction <http://exmo.inrialpes.fr/jdavid_img_small.jpg> ;
  foaf:knows <http://exmo.inrialpes.fr/about#matencia> .
             <h
             [
```

jdavid is a person, with the name of "Jérôme David" and the nickname of "JD", who has an email address of "jerome.david@inria.fr". His homepage is... He is depicted in the image... He knows... He also knows a person whose name is "Ewan David" and who is 5 years old.

Property: foaf:knows

knows - A person known by this person (indicating some level of reciprocated interaction between the parties).
Status: stable
Domain: having this property implies being a Person
Range: every value of this property is a Person

The knows property relates a Person to another Person that he or she knows.

from http://xmlns.com/foaf/spec/#term_knows

Hands-on session (I)

Reusing Existing Terms
The Dublin Core Schema

- The Dublin Core Schema is a vocabulary of terms that can be used to describe web resources (videos, images, web pages) and physical resources (books, publications, CDs).
- Dublin Core Metadata Initiative (DCMI)
- two namespaces:
  - Dublin Core Metadata Set version 1.1

    `dc: <http://purl.org/elements/1.1/>`

    domains and ranges are not specified (e.g. `dc:creator` may be used with both literal and non-literal values)
  - DCMI Metadata Terms

    `dcterms: <http://purl.org/dc/terms/>`

    domains and ranges are specified (e.g. the range of `dcterms:creator` is the class `dcterms:Agent`)

    `dcterms:creator` is a subproperty of `dc:creator`

Hands-on session (I)

Reusing Existing Terms
The Dublin Core Schema

- An example:

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix ex: <http://www.example.org/> .

ex:myPaper a dcterms:BibliographicResource ;
           dcterms:title "What is this thing called Linked Data?" ;
           dcterms:creator ex:
           dcterms:subject "Li
           dcterm
           dcterm
           dcterm                                              ne " .

ex:matencia a foa
            foaf:
            foaf:

...
```

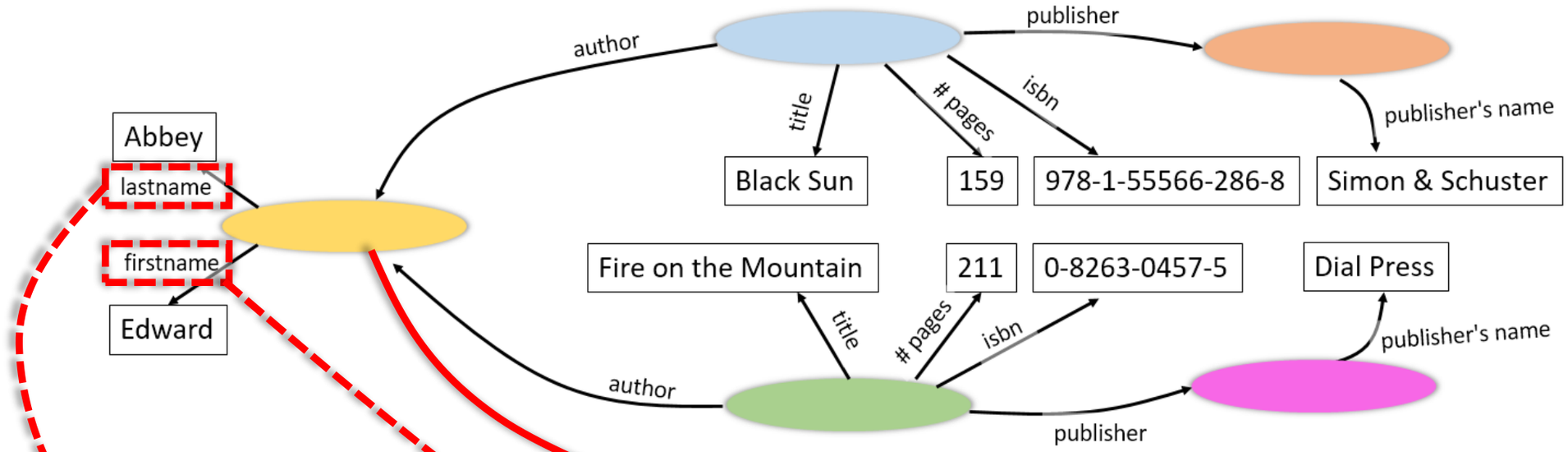
from http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#terms-creator

Term Name: creator	
URI:	http://purl.org/dc/terms/creator
Label:	Creator
Definition:	An entity primarily responsible for making the resource.
Comment:	Examples of a Creator include a person, an organization, or a service.
Type of Term:	Property
Refines:	http://purl.org/dc/elements/1.1/creator
Refines:	http://purl.org/dc/terms/contributor
Has Range:	http://purl.org/dc/terms/Agent
Version:	http://dublincore.org/usage/terms/history/#creatorT-002
EquivalentProperty:	http://xmlns.com/foaf/0.1/maker

Hands-on session (I)

Creating RDF data
Reusing existing terms

- Which properties/classes could we use for describing Artemis' bookstore data?



foaf vocabulary

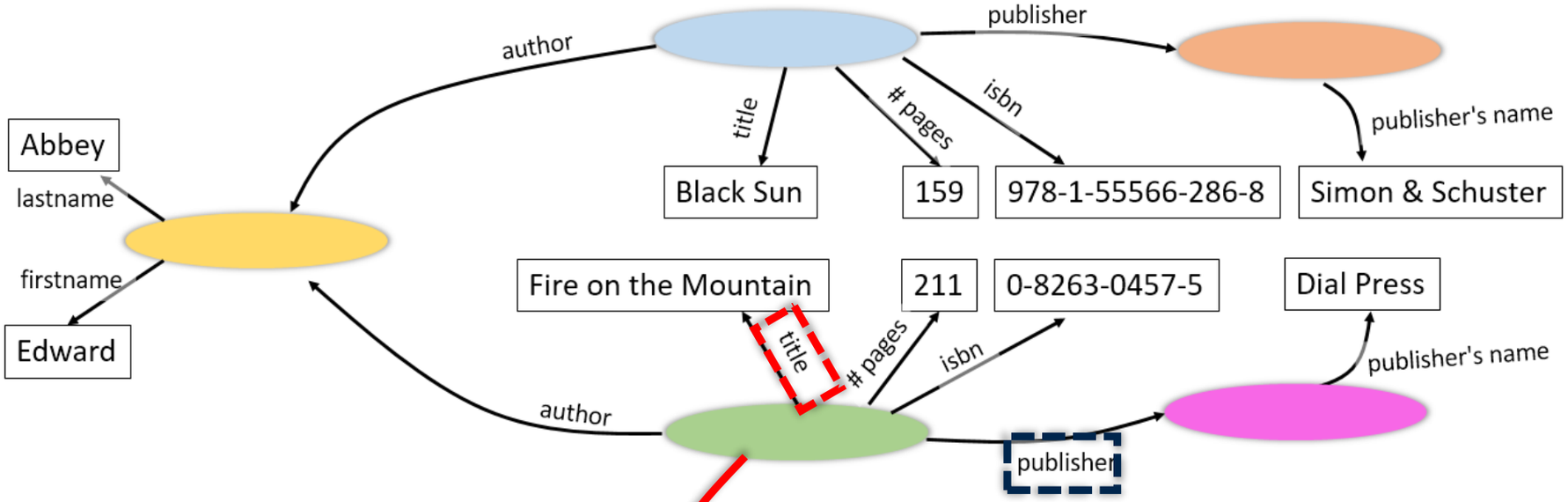
Classes: | Agent | Document | Group | Image | LabelProperty | OnlineAccount | OnlineChatAccount | OnlineEcommerceAccount | OnlineGamingAccount | Organization | Person | PersonalProfileDocument | Project |

Properties: | account | accountName | accountServiceHomepage | age | aimChatID | based_near | birthday | currentProject | depiction | depicts | dnaChecksum | familyName | family_name | firstName | focus | fundedBy | geekcode | gender | givenName | givenname | holdsAccount | homepage | icqChatID | img | interest | isPrimaryTopicOf | jabberID | knows | lastName | logo | made | maker | mbox | mbox_sha1sum | member | membershipClass | msnChatID | myersBriggs | name | nick | openid | page | pastProject | phone | plan | primaryTopic | publications | schoolHomepage | sha1 | skypeID | status | surname | theme | thumbnail | tipjar | title | topic | topic_interest | weblog | workInfoHomepage | workplaceHomepage | yahooChatID |

Hands-on session (I)

Creating RDF data
Reusing existing terms

- Which properties/classes could we use for describing Artemis' bookstore data?



rdf:type

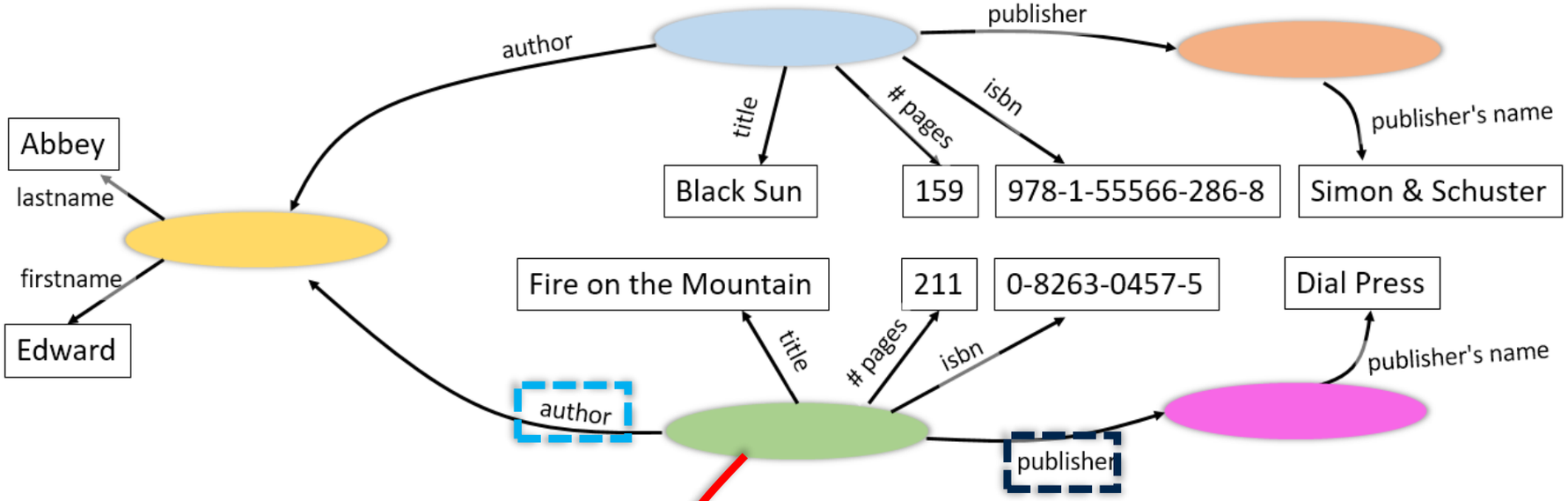
dublin core vocabulary

Properties in the */terms/* namespace	abstract , accessRights , accrualMethod , accrualPeriodicity , accrualPolicy , alternative , audience , available , bibliographicCitation , conformsTo , contributor , coverage , created , creator , date , dateAccepted , dateCopyrighted , dateSubmitted , description , educationLevel , extent , format , hasFormat , hasPart , hasVersion , identifier , instructionalMethod , isFormatOf , isPartOf , isReferencedBy , isReplacedBy , isRequiredBy , issued , isVersionOf , language , license , mediator , medium , modified , provenance , publisher , references , relation , replaces , requires , rights , rightsHolder , source , spatial , subject , tableOfContents , temporal , title , type , valid
Classes	Agent , AgentClass , BibliographicResource , FileFormat , Frequency , Jurisdiction , LicenseDocument , LinguisticSystem , Location , LocationPeriodOrJurisdiction , MediaType , MediaTypeOrExtent , MethodOfAccrual , MethodOfInstruction , PeriodOfTime , PhysicalMedium , PhysicalResource , Policy , ProvenanceStatement , RightsStatement , SizeOrDuration , Standard

Hands-on session (I)

Creating RDF data
Reusing existing terms

- Which properties/classes could we use for describing Artemis' bookstore data?



rdf:type

dublin core vocabulary

Dublin Core Metadata Initiative

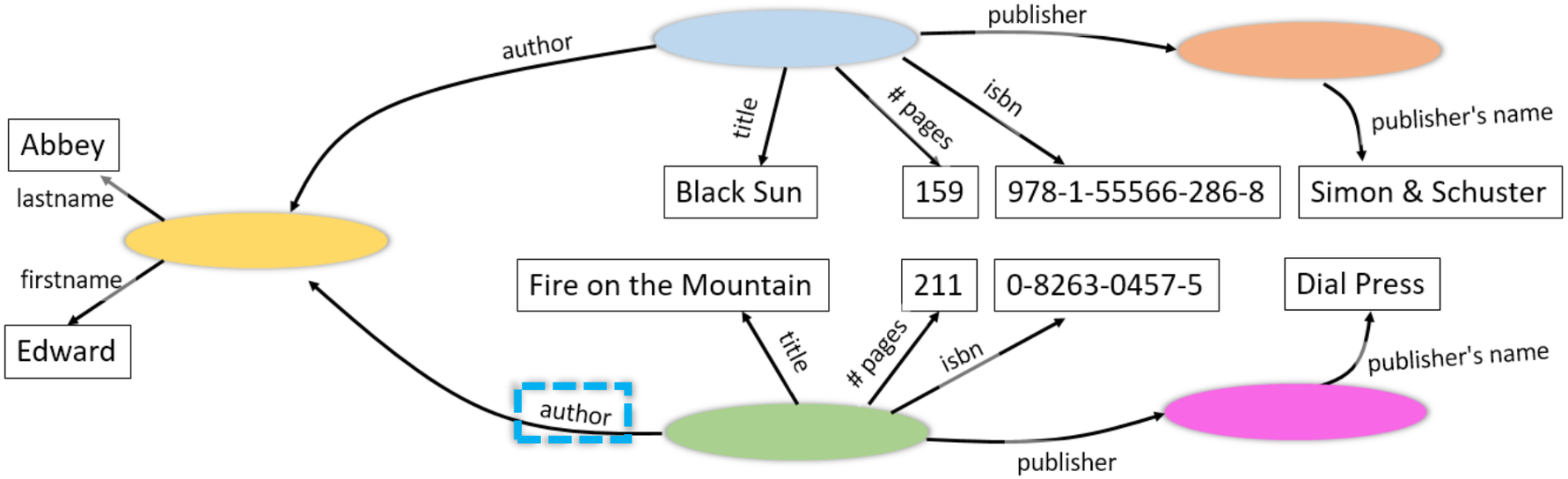
well... not so optimal

Properties in the */terms/* namespace	abstract , accessRights , accrualMethod , accrualPeriodicity , accrualPolicy , alternative , audience , available , bibliographicCitation , conformsTo , contributor , coverage , created , creator , date , dateAccepted , dateCopyrighted , dateSubmitted , description , educationLevel , extent , format , hasFormat , hasPart , hasVersion , identifier , instructionalMethod , isFormatOf , isPartOf , isReferencedBy , isReplacedBy , isRequiredBy , issued , isVersionOf , language , license , mediator , medium , modified , provenance , publisher , references , relation , replaces , requires , rights , rightsHolder , source , spatial , subject , tableOfContents , temporal , title , type , valid
Classes	Agent , AgentClass , BibliographicResource , FileFormat , Frequency , Jurisdiction , LicenseDocument , LinguisticSystem , Location , LocationPeriodOrJurisdiction , MediaType , MediaTypeOrExtent , MethodOfAccrual , MethodOfInstruction , PeriodOfTime , PhysicalMedium , PhysicalResource , Policy , ProvenanceStatement , RightsStatement , SizeOrDuration , Standard

Hands-on session (I)

Creating RDF data
User-defined vocabularies

- Which properties/classes could we use for describing Artemis' bookstore data?



your own vocabulary/ontology **use hash URIs for small datasets (e.g. ontologies)**

http://artemisBookstore.com/ontology#author *rdfs:subPropertyOf dcterms:creator*

http://artemisBookstore.com/ontology#isbn

http://artemisBookstore.com/ontology#pages

http://artemisBookstore.com/ontology#datePublication

http://artemisBookstore.com/ontology#dateAcquired

http://artemisBookstore.com/ontology#Novel

RFDS/OWL
to describe vocabularies (classes and properties)

rdfs:subPropertyOf dcterms:date

rdfs:subPropertyOf dcterms:date

rdfs:subClassOf
dcterms:BibliographicResource

Hands-on session (I)

- You should end up with something like this:

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix abo: <http://artemisBookstore.com/ontology#> .
@prefix dcterms: <http://purl.org/dc/terms/> .


<http://artemisBookstore.com/id/author/1> a foaf:Person ;
        foaf:firstName "Paul" ;
        foaf:lastName "Auster" .

<http://artemisBookstore.com/id/book/571142200> a dcterms:BibliographicResource ;
        dcterms:title "Moon palace" ;
        abo:author <http://artemisBookstore.com/id/author/1> ;
        abo:bookType "Paperback" ;
        abo:isbn "571142200" ;
        abo:language "English" ;
        abo:pages "320"^^<http://www.w3.org/2001/XMLSchema#int> ;
        abo:price "12.5"^^<http://www.w3.org/2001/XMLSchema#double> ;
        abo:datePublication "2004-02-01" ;
        abo:dateAcquired "2013-02-01 15:00:00" ;
        dcterms:publisher "Faber & Faber" .

<http://artemisBookstore.com/id/author/2> a foaf:Person ;
        foaf:firstName "Siri" ;
        foaf:lastName "Hustvedt" .
```

Hands-on session (I)

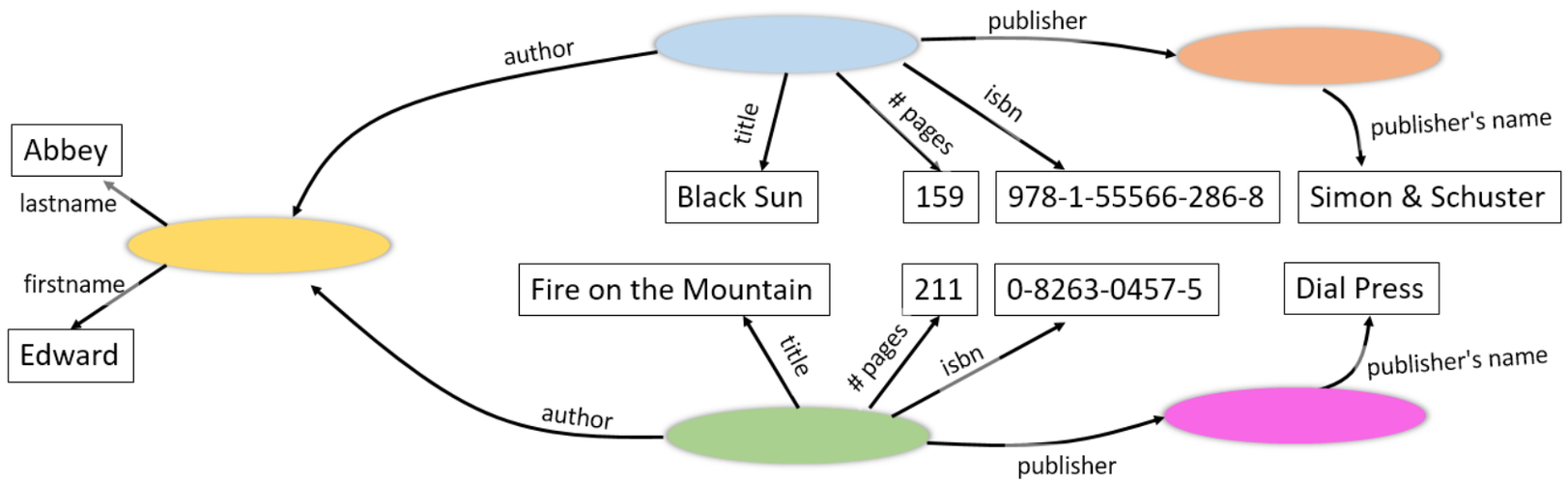
Creating links to external datasets

- Now, let's create RDF data! But before
- Let's recall the Linked Data principles:

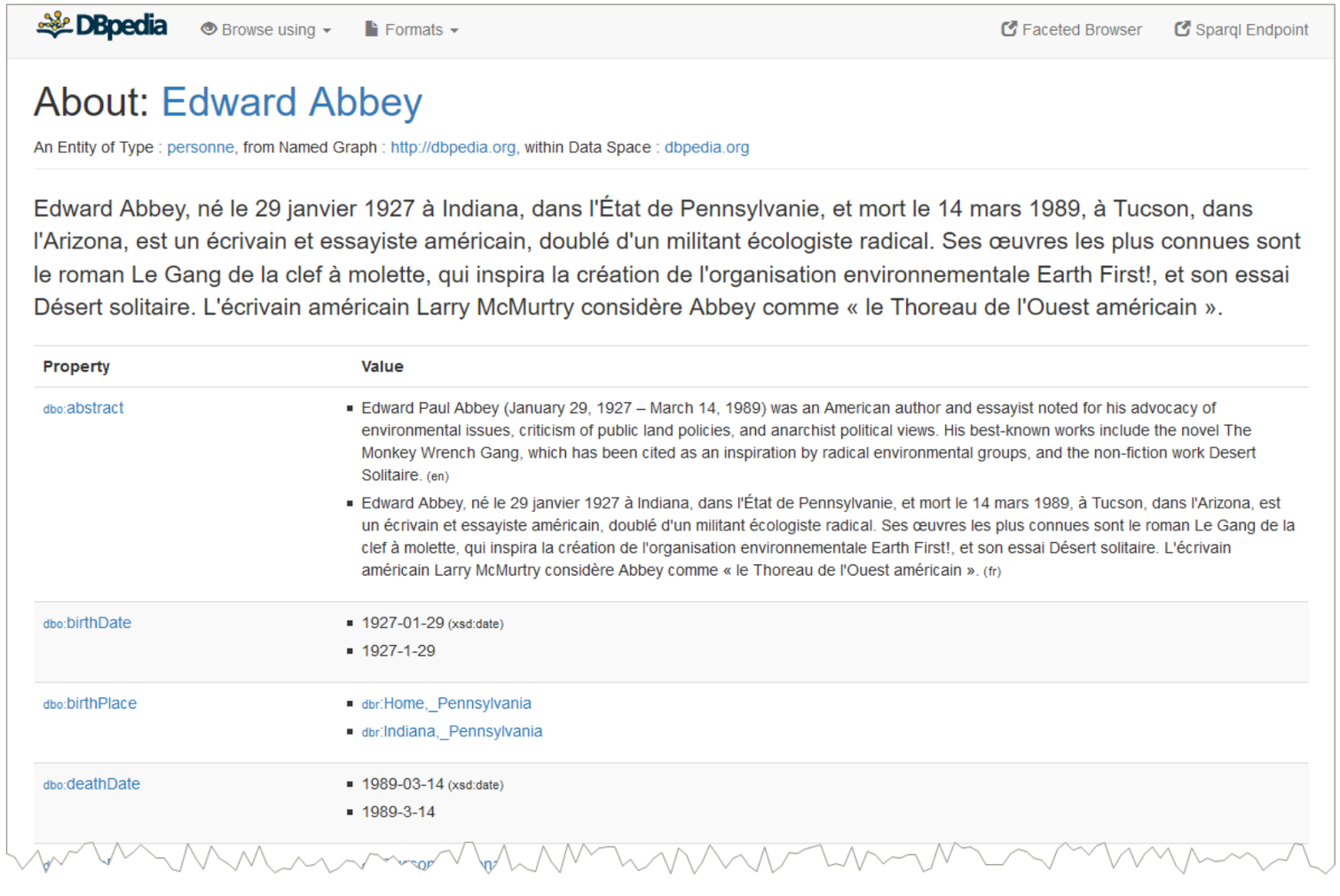
1. Use URIs as names for things.
2. Use HTTP URIs, so that people can look up those names.
3. When someone looks up a URI, provide useful information using the standards (RDF, SPARQL).
4. **Include links to other URIs, so that they can discover more things.**

Creating RDF data

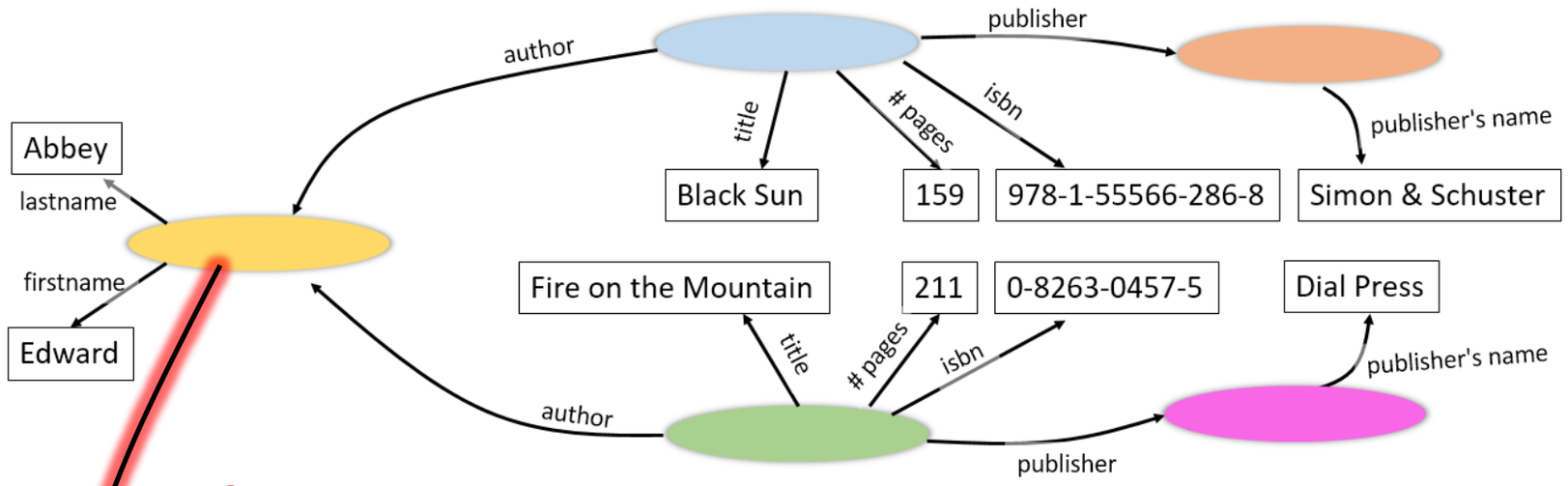
Hands-on session (I)



- Dbpedia has data about Abbey Edwards



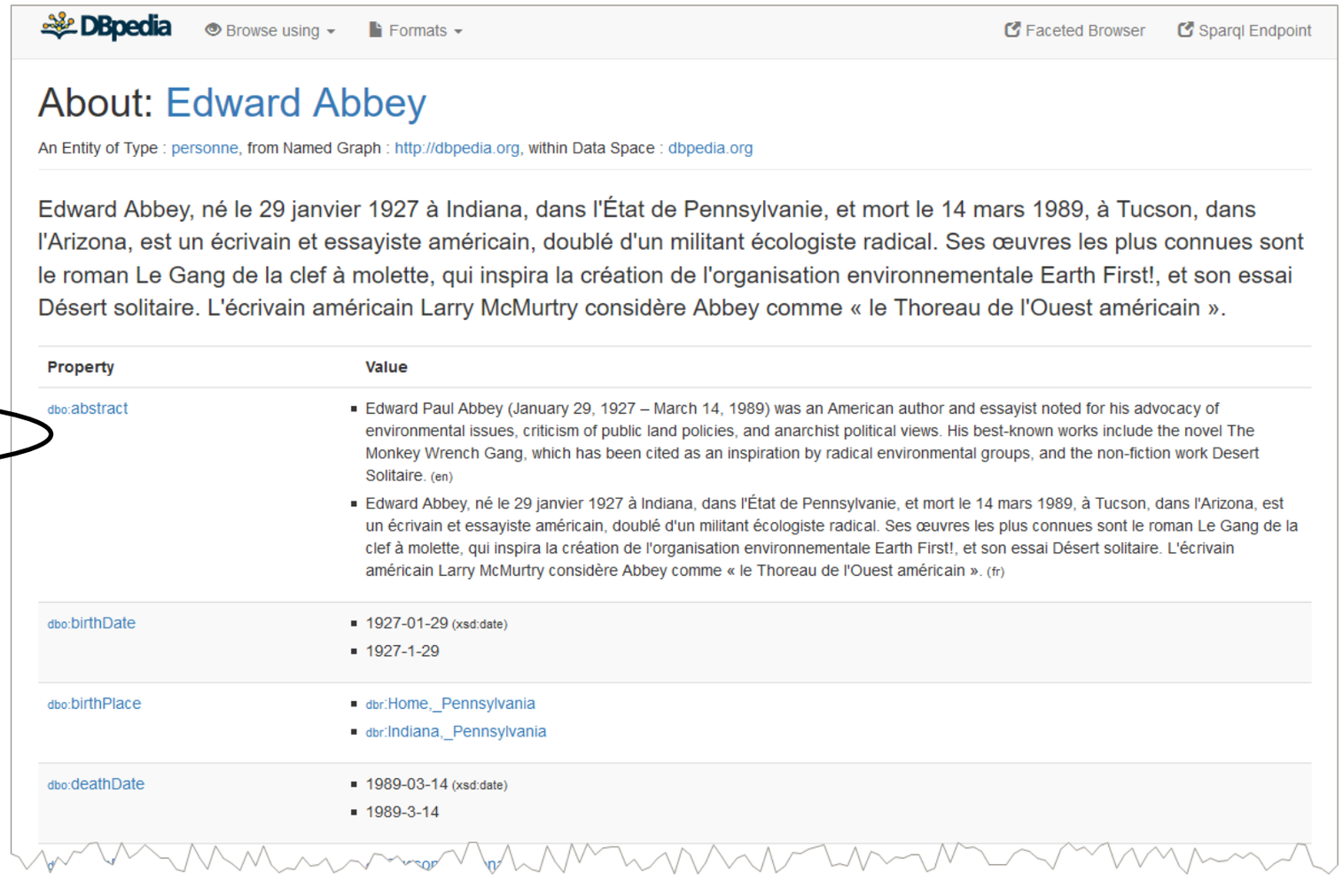
Hands-on session (I)



owl:sameAs

- Dbpedia has data about Abbey Edwards

http://dbpedia.org/page/Edward_Abbey



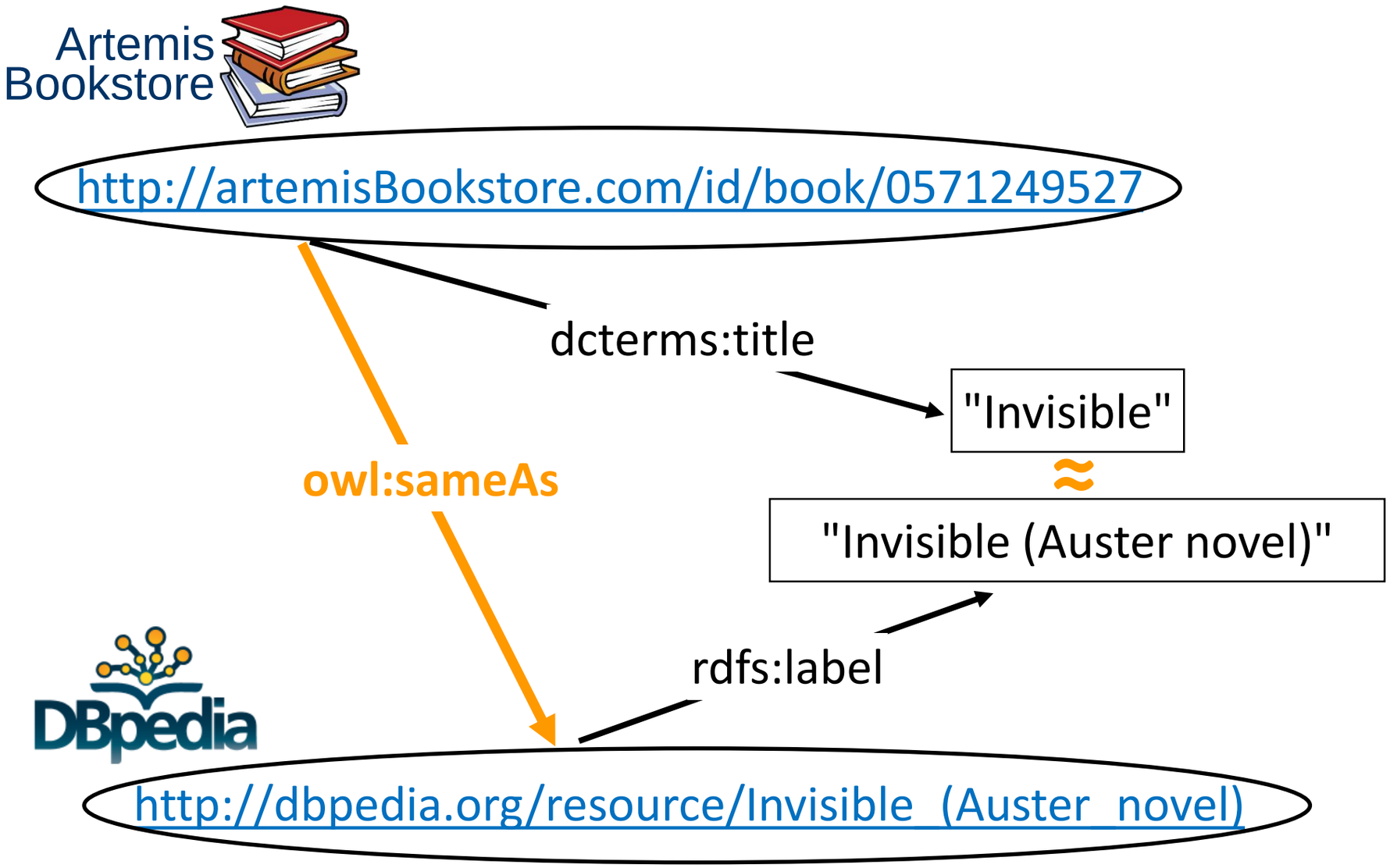
Hands-on session (I)

- Data interlinking
  - Not an easy task
    - a large quantity of data
    - semantic heterogeneity
    - noisy data
  - Performed manually or (semi-)automatically
  - Key-based approaches vs similarity-based approaches
  - Data interlinking tools:
    - Silk - Link Discovery Framework
    - LiMES - Link Discovery Framework for Metric Spaces
    - Link Keys
    - OpenRefine + Reconciliation services

Hands-on session (I)

OpenRefine
Reconciliation services

- Reconciliation services based on SPARQL endpoints



Hands-on session (I)

Creating links to external datasets

- You should end up with something like this

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix abo: <http://artemisBookstore.com/ontology#> .
@prefix dcterms: <http://purl.org/dc/terms/> .


<http://artemisBookstore.com/id/author/1> a foaf:Person ;
        foaf:firstName "Paul" ;
        foaf:lastName "Auster" ;
        owl:sameAs <http://dbpedia.org/resource/Auster,_Paul> .

<http://artemisBookstore.com/id/book/571142200> a dcterms:BibliographicResource ;
        dcterms:title "Moon palace" ;
        abo:author <http://artemisBookstore.com/id/author/1> ;
        abo:bookType "Paperback" ;
        abo:isbn "571142200" ;
        abo:language "English" ;
        abo:pages "320"^^<http://www.w3.org/2001/XMLSchema#int> ;
        abo:price "12.5"^^<http://www.w3.org/2001/XMLSchema#double> ;
        abo:datePublication "2004-02-01" ;
        abo:dateAcquired "2013-02-01 15:00:00" ;
        dcterms:publisher "Faber & Faber" ;
        owl:sameAs <http://dbpedia.org/resource/Moon_Palace> .

<http://artemisBookstore.com/id/author/2> a foaf:Person ;
        foaf:firstName "Siri" ;
        foaf:lastName "Hustvedt" ;
        owl:sameAs <http://dbpedia.org/resource/Siri_Hustvedt> .
```

Work to do

- write a Java Program that reads the artemisBookstoreData.csv CSV files and generates a RDF file in turtle using strings.
- validate your rdf file using `riot` Jena command line tool
- *optional: modify your program to create a Jena core API Model when parsing the csv file and to output it in various RDF serialization formats (Trutle, RDF/XML…)*
- using Jena core API write a Java program that reads your RDF file and given an author's name finds and displays all the titles of all the books he wrote.