Representing and reasoning on data semantics: survey and challenges

Marie-Christine Rousset

Data semantics

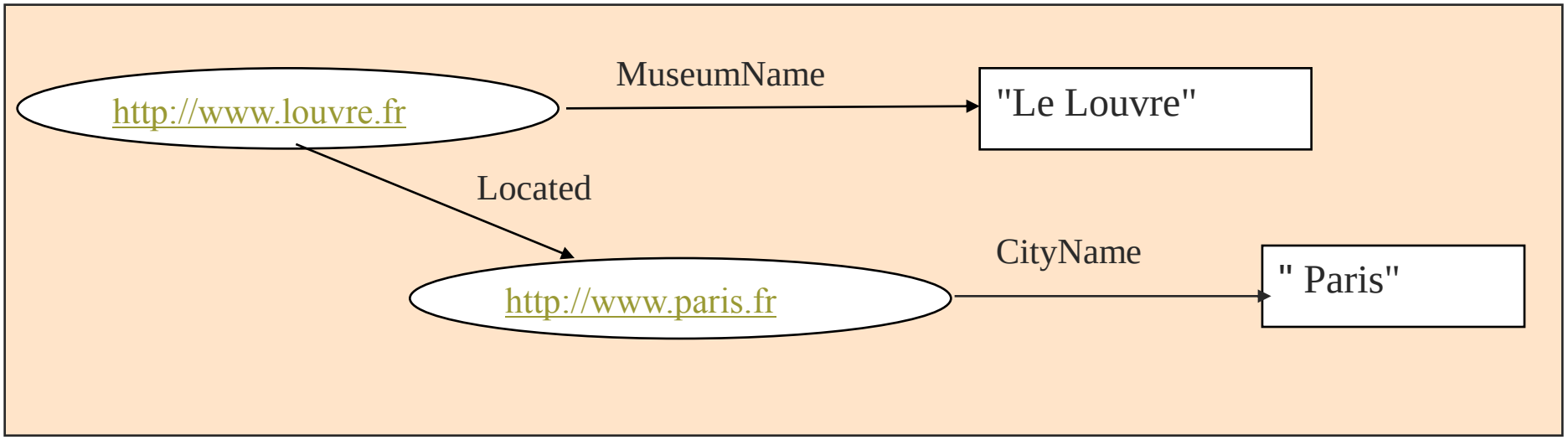
- Constraints on the vocabulary used to describe metadata
- Expressed as logical axioms on which reasoning is possible
  - For query reformulation
  - For data consistency checking
  - For data reconciliation
  - For query containment
  - For resource discovery
  - …

Semantic Web approach

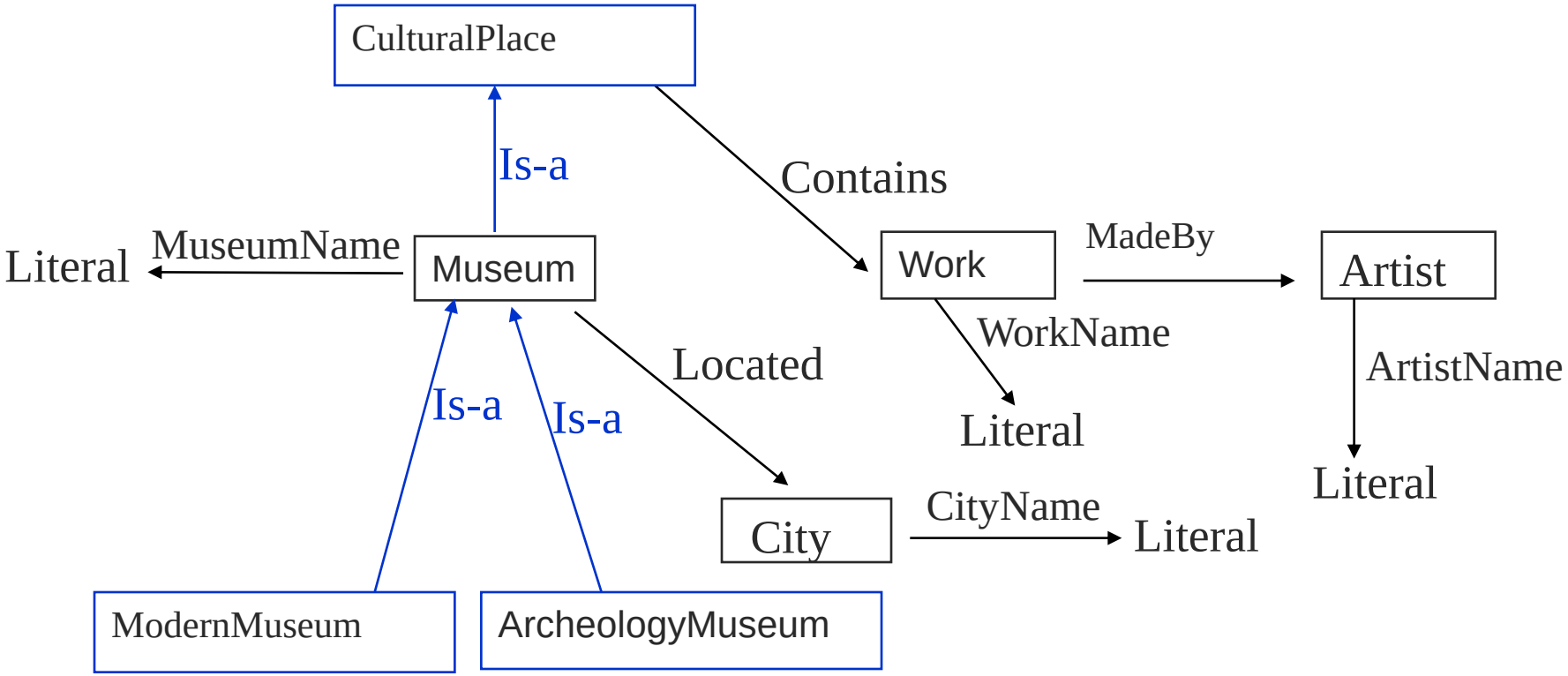
- RDF triples for describing metadata on resources identified by URIs
  - web pages
  - web services
  - fragments of XML documents
- RDFS or OWL ontologies for giving semantics to RDF triples
  - RDFS: simple logical axioms (inclusion statements)
  - OWL: (very) complex logical axioms
  - DL-Lite family: a right trade-off
- Mappings between existing ontologies for information integration

RDF: illustration

- Triple : <resource, property, value>
- Relational : property(resource, value)
- Graphical :



RDFS : illustration



RDFS : XML / logic

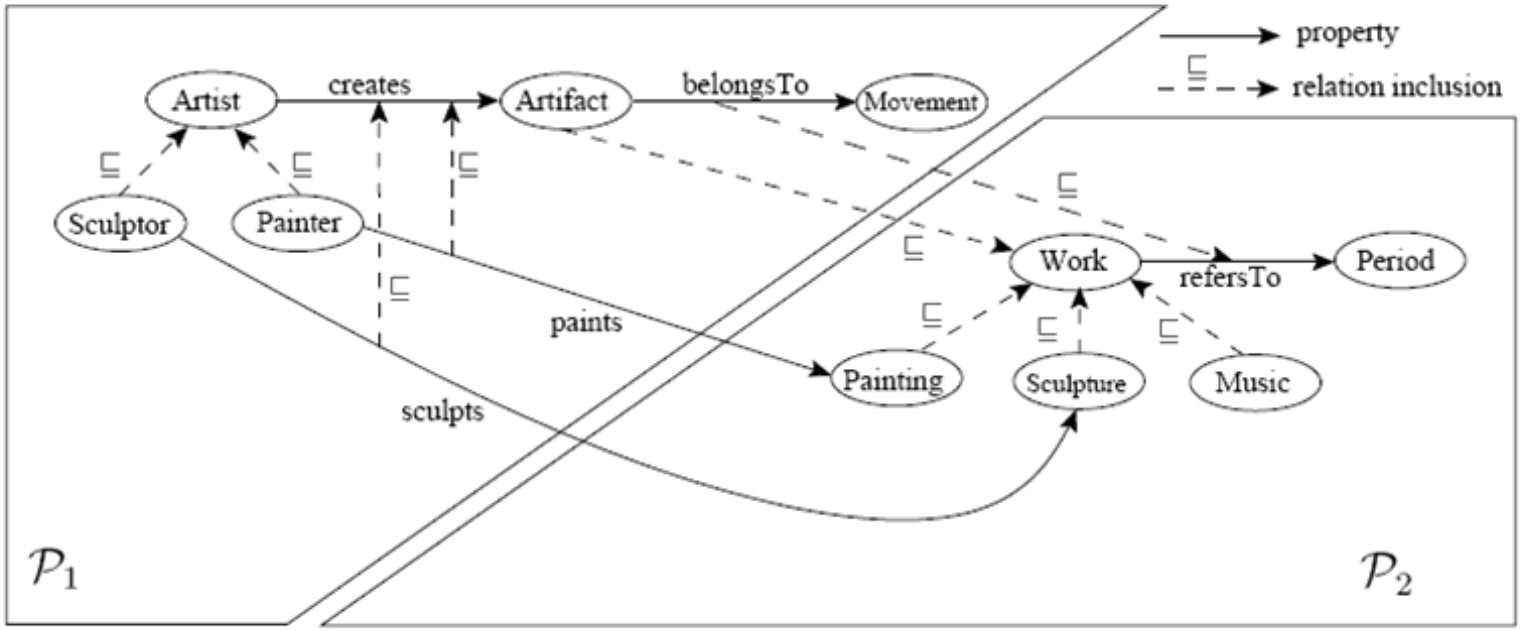
- XML serialization: RDF/XML
- Formal semantics: fragment of FOL

Constructor	FOL axiomatization
Class inclusion	$\forall X(C_1(X) \Rightarrow C_2(X))$
Property inclusion	$\forall X \forall Y(P_1(X,Y) \Rightarrow P_2(X,Y))$
Domain typing of a property	$\forall X \forall Y(P(X,Y) \Rightarrow C(X))$
Range typing of property	$\forall X \forall Y(P(X,Y) \Rightarrow C(Y))$

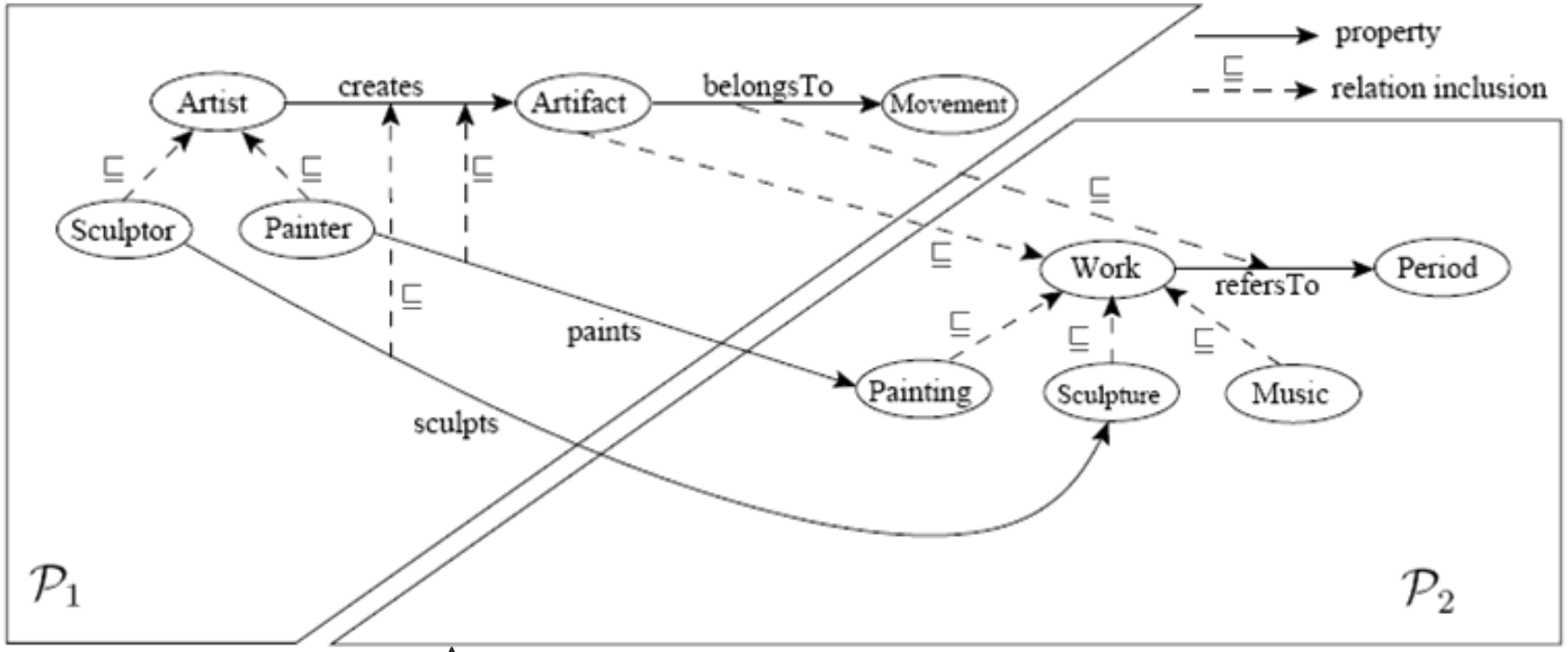
  - on which we can plug a standard relational query language

Q(X): Museum(X) ∧ Contains(X,Y)∧ MadeBy(Y,Z) ∧ ArtistName(Z, »Picasso »)

Mappings


property
relation inclusion
Artist
creates
Artifact
belongsTo
Movement
Sculptor
Painter
paints
sculpts
Work
refersTo
Period
Painting
Sculpture
Music
$\mathcal{P}_1$
$\mathcal{P}_2$

Querying by reformulation



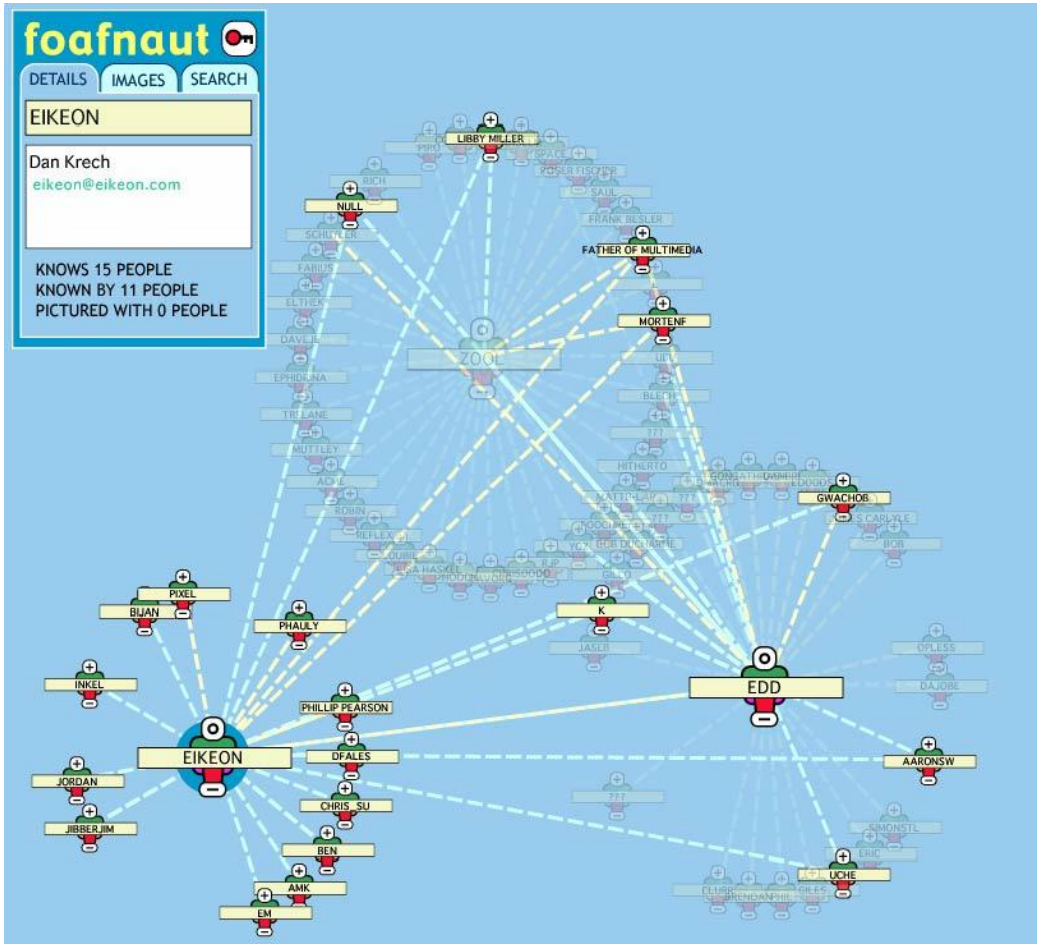
$$Q_2(X,Y) \equiv \mathcal{P}_2{:}Painting(X) \wedge \mathcal{P}_2{:}refersTo(X,Y)$$

1. $R_1^2(X,Y) \equiv \mathcal{P}_2{:}Painting(X) \wedge \mathcal{P}_2{:}refersTo(X,Y)$
2. $R_2^2(X,Y) \equiv \mathcal{P}_2{:}Painting(X) \wedge \mathcal{P}_1{:}belongsTo(X,Y)$
3. $R_3^2(X,Y) \equiv \exists Z\ \mathcal{P}_1{:}paints(Z,X) \wedge \mathcal{P}_2{:}refersTo(X,Y)$
4. $R_4^2(X,Y) \equiv \exists Z\ \mathcal{P}_1{:}paints(Z,X) \wedge \mathcal{P}_1{:}belongsTo(X,Y)$

Real applications: FOAF

The *Friend of a Friend* (FOAF) project is about creating a Web of machine-readable homepages describing people, the links between them and the things they create and do.

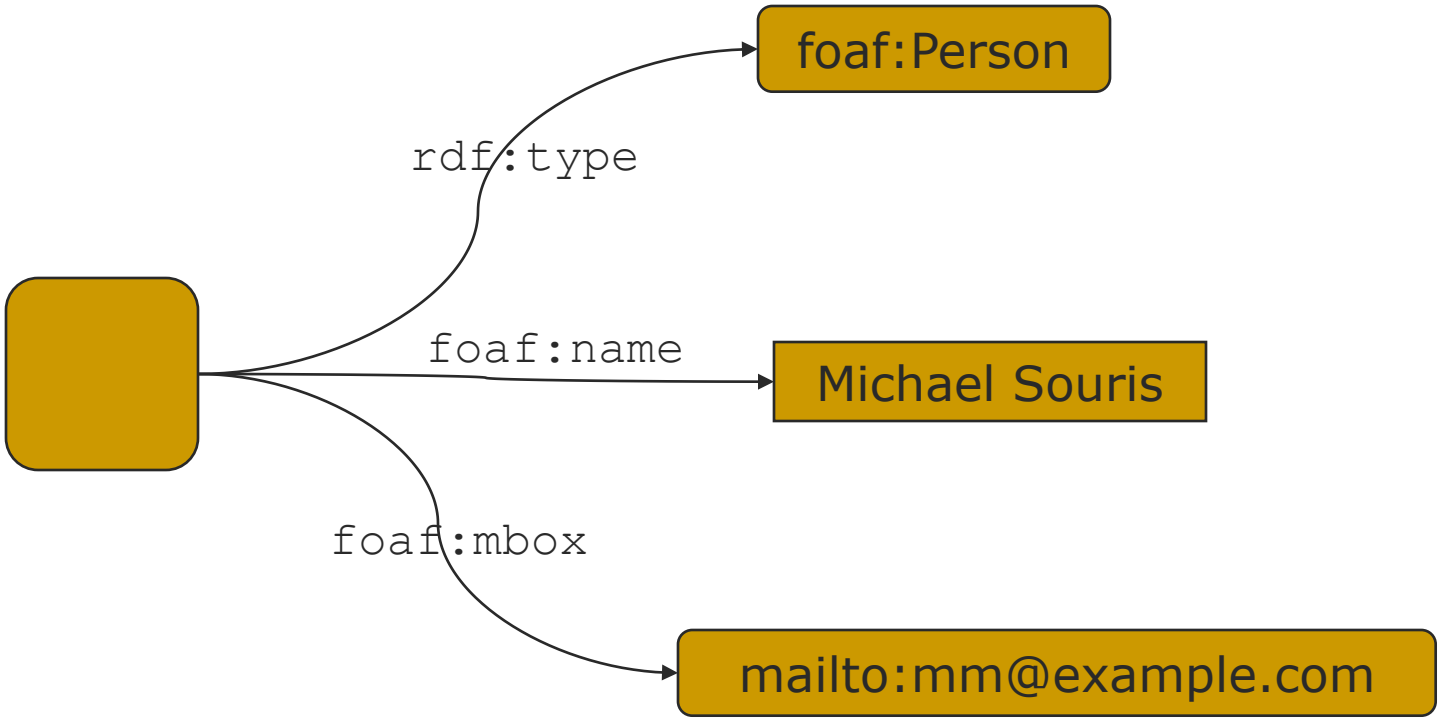
Distributed RDF/XML records describing people, who they know, projects they work on…



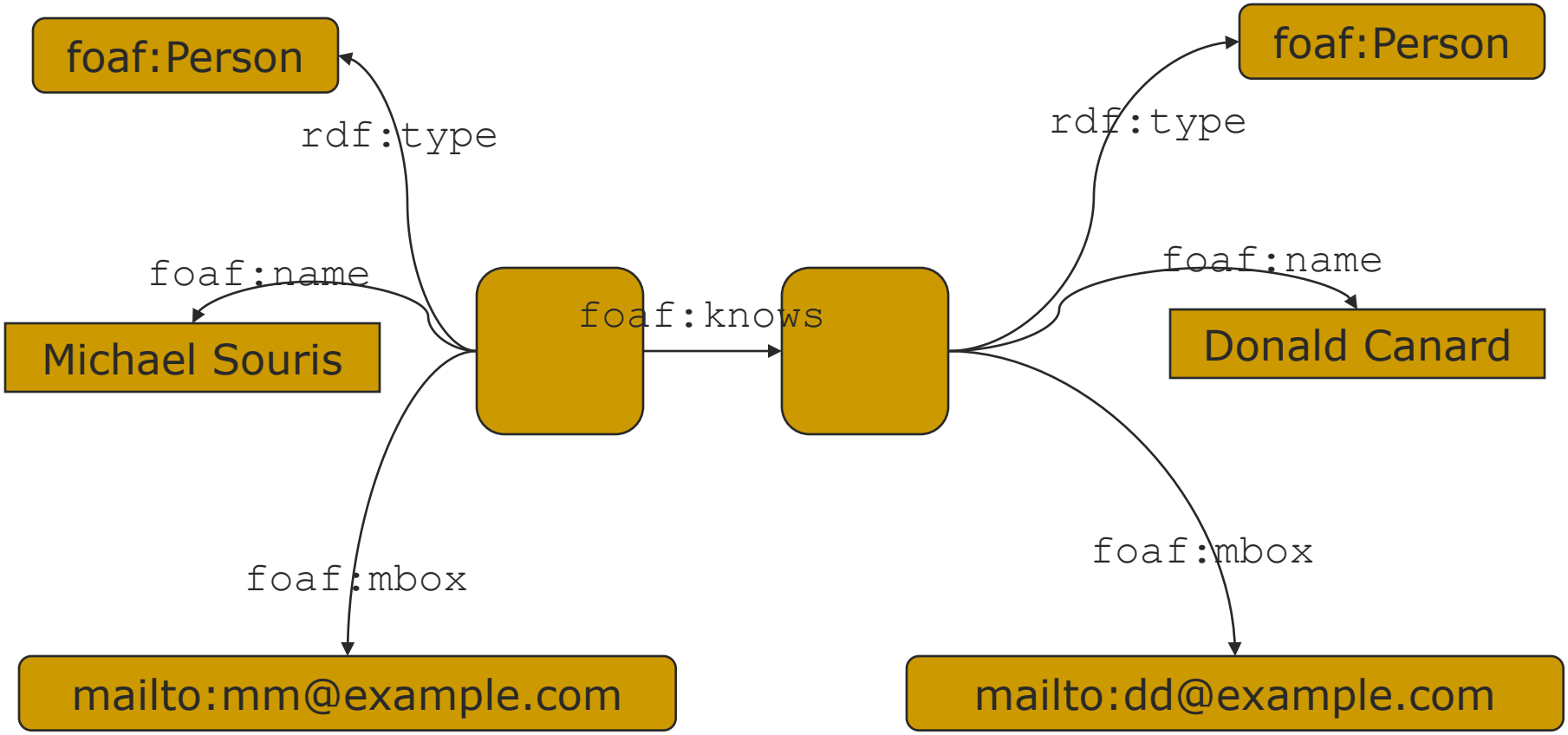
FOAF - motivations

- Augment e-mail filtering by prioritizing mails from trusted colleagues
- Locate people with interests similar to yours
- ‘Find an expert’ in knowledge communities
- Photo co-depiction

A Simple FOAF Model



A More Complex FOAF Model

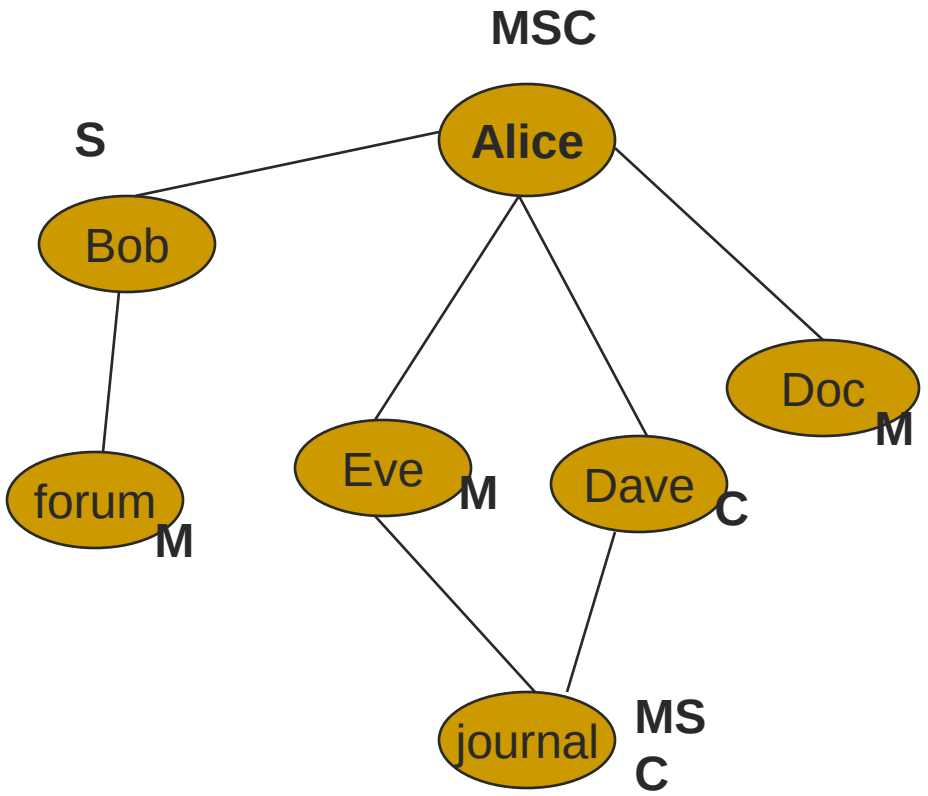


Challenges raised by a P2P setting (WebDam)

- At each peer, representing and reasoning on its **partial knowledge** about other peers data
- Query answering
  - by gathering evidence on possibly incomplete or inconsistent answers
  - and by comparing them for ranking or aggregating them
    - depending on their lineage and the trust of the querying peer towards the peers involved in the answers
- Requires a logic making explicit the notion of belief and knowledge
  - modal operators
- Scalability issues of decentralized reasoning

Illustrative scenario

- Alice trusts Doc for Medecine
- Alice trusts Bob and believes that he knows about Sport
- Alice does not trust Eve
- Alice trusts Dave and believes that he knows about Cinema



Illustrative scenario (ctd)

- **For a query related to Medecine**
  - Reasoning for finding the reliable sources
    - Forum.
    - Journal.
    - Doc
  - Querying them
  - Comparing, aggregating or ranking the answers
  - Evaluating them and updating the knowledge about trust and belief towards the peers involved in the answers