Mashup, Microformats, RDFa and GRDDL
Mashup

- A mashup application combines contents or services coming from several, possibly heterogeneous, applications.
- In the case of a website, a mashup aggregates content coming from different web sites in order to design a new one.
Mashup

- Differences between mashups:
  - where the integration happens (server side vs client side),
  - what is being integrated,
  - how the data is retrieved (API vs screen scraping)
Web scraping (harvesting)

- When a site does not provide an API or the API does not offer a needed feature
- A script/program extracts data from a web site.
Intro sur RDF

- RDF: Resource Description framework
- W3C Recommandation
- A simple data model:
  - Subject, predicate, object
  - Ex: Ernest loves cookies
  - Graph representation:
    ![Graph representation](Ernest loves cookies)
RDF suite

- Several syntaxes: XML/RDF, N3, N-triples, graphic-based.
- N3 Notation:

```xml
@prefix pref: <http://exemple.com/vocabulaire#>.
<#ernest> <pref:loves><#cookies>.
```
- `pref` refers to a vocabulary.
Vocabulary

- A vocabulary provides meaning (sémantics) to triples
- Machines can interpret RDF graphs using these vocabularies.
FOAF

- FOAF stands for Friend Of A Friend and is a vocabulary used in social networks.
- In N3:

```n3
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
<#olivier> <foaf:knows><#marie>.
```
Semantic Web applications
Motivation

- Most Web pages are being consumed by human beings.
- Data intended for machines are available but they are generally stored in complementary documents, in a different format and with a limited correspondence with the documents aimed at humans.
Motivation

- How can we integrate data in Web pages that can be used for both rendering and processing.
- Several solutions are emerging.
- Two important goals are to enable mashups and to extract metadata from web pages.
RDFa: RDF in html attributes

- Describe data through (X)HTML format such that it is used for rendering and processing
- Transfer Transférer des données data from an application to another application via a Web format.
RDFa: RDF in html attributes

- RDFa uses URI
- CURIE: compact URI
- Ex: foaf:name
- where foaf refers to the http://xmlns.com/foaf/0.1 URI and name is foaf predicate
RDFa example 1

```xml
<body xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <span about="http://www.univ-mlv.fr/~ocure/
    property="foaf:name">
    Olivier Curé
  </span>
</body>
```

- Équivalent à

```xml
@prefix foaf: <http://xmlns.com/foaf/0.1/>.
<#olivier> <foaf:knows><#marie>.
```
RDFa example 2

```html
<body xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <span about="http://www.univ-mlv.fr/~ocure/" typeof="foaf:Person" property="foaf:name">
    Olivier Curé
  </span>
</body>
```

To specify the type of the subject:

- `http://www.univ-mlv.fr/~ocure/`
- `Olivier Curé`
- `Person`
RDFa example 3

<body xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <span about="http://www.univ-mlv.fr/~ocure/
    instanceof="foaf:Person"  property="foaf:name">
    Olivier Curé
  </span>

  <span about="http://www.exemple.com/marie/
    instanceof="foaf:Person"  property="foaf:name">Marie
  </span>

  <span about="http://www.univ-mlv.fr/~ocure/
    rel="foaf:knows"  resource="http://www.exemple.com/marie/">olivier knows marie</span>
</body>
XHTML attributes

- @rel: a whitespace separated list of CURIEs, used for expressing relationships between two resources ('predicates' in RDF terminology);
- @rev: a whitespace separated list of CURIEs, used for expressing reverse relationships between two resources (also 'predicates');
- @content: a string, for supplying machine-readable content for a literal (a 'plain literal object', in RDF terminology);
XHTML attributes

- `@href`: a URI for expressing the partner resource of a relationship (a 'resource object', in RDF terminology);
- `@src`: a URI for expressing the partner resource of a relationship when the resource is embedded (also a 'resource object').
- `@about`: a URI or SafeCURIE, used for stating what the data is about (a 'subject' in RDF terminology);
XHTML attributes

- @property : a whitespace separated list of CURIEs, used for expressing relationships between a subject and some literal text (also a 'predicate');
- @resource: a URIorSafeCURIE for expressing the partner resource of a relationship that is not intended to be 'clickable' (also an 'object');
XHTML attributes

- @datatype: a CURIE representing a datatype, to express the datatype of a literal;
- With
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  span property="dbp:dateOfBirth" datatype="xsd:date">1879-03-14</span>
XHTML attributes

- @typeof: a whitespace separated list of CURIEs that indicate the RDF type(s) to associate with a subject.

- Example:

```html
<div typeof="foaf:Person">
  <span property="foaf:name">Albert Einstein</span>
  <span property="foaf:givenname">Albert</span>
</div>
```
Popularité de RDFa?

- Yahoo! Searchmonkey (a framework to enable site editors to publish detailed information about their sites) uses RDFa.
- Google, Drupal (CMS open source), Slideshare, MySpace, etc. qlo start to use RDFa.
Microformats

- A solution to add structured information in web pages
- It is possible to create RDF data from microformats using XSLT
- But each microformat requires its own XSLT!
Steps

• Declare the schemas you are going to use:

```html
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:cal="http://www.w3.org/2002/12/cal/icaltzd#"
xmlns:xs="http://www.w3.org/2001/XMLSchema#"/>
```
Steps

- Use attributes to mark, type and add (meta)data
  - `<div id="hcard-Olivier-Cure" class="vcard">
    <a class="url fn" href="http://www.univ-mlv.fr/~ocure/"><font size=+4>&nbsp;Olivier Curé</font></a>
  </div>`

- Let agents extract RDF from your web pages.
Usage

- It is possible to exploit data associated to events:
  - By integrating them in calendar related applications
  - Downloading informations in iCal format
RDFa vs Microformats

- Microformats specify both a syntax for embedding structured data into HTML documents and a vocabulary of specific terms for each microformat.
- RDFa specifies only a syntax and relies on independent specification of terms (often called vocabularies or taxonomies) by others.
GRDDL

- Gleaning Resource Descriptions from Dialects of Languages
- Proposes a standard method to map microformat data, or another structured format, to RDF.