

De la <meta/>date la microformate

Dr. Sabin-Corneliu Buraga

Facultatea de Informatică

Universitatea “A.I.Cuza” Iași

<http://www.infoiasi.ro/~busaco/>



cuprins

- Preliminarii
- Reprezentarea <meta />datelor
- Vocabulare de <meta />date
- Microformate
- Ce urmează?!...
- Resurse



preliminarii

- **World-Wide Web**
- Resursele sunt identificate prin **adresa** lor
– **identificator uniform de resurse (URI)**
- Existența unui **protocol** (e.g., **HTTP**)
pentru accesul la conținutul resurselor
- Resursele (văzute ca **documente**)
includ **marcaje** (adnotări)
- Marcajele conțin
la rândul lor URI-uri

pagini Web



Detalii: www.w3.org/TR/webarch/



preliminarii

- Ce putem afirma despre:
`<p>CV-ul meu</p>`
- Întrebări privitoare la structură:
 - `<a>` este parte din `<p>`? Ori altceva?
 - `cv.html` este o resursă (obiect)? O clasă de obiecte? Un atribut? O relație? Sau altceva?
- Întrebări referitoare la vocabular:
 - `<a>` este similar cu `<a>` din XHTML?
 - Autorul marcajului este cel care l-a și publicat? Dacă da, când? Sub ce termeni privind *copyright*-ul?



preliminarii

- **Întrebări:**
 - Cum putem preciza informații (proprietăți, caracteristici,...) despre resursele Web?
 - Cum se pot reprezenta informațiile care pot fi *identificate* pe Web, chiar dacă nu pot fi *accesate* via Web?



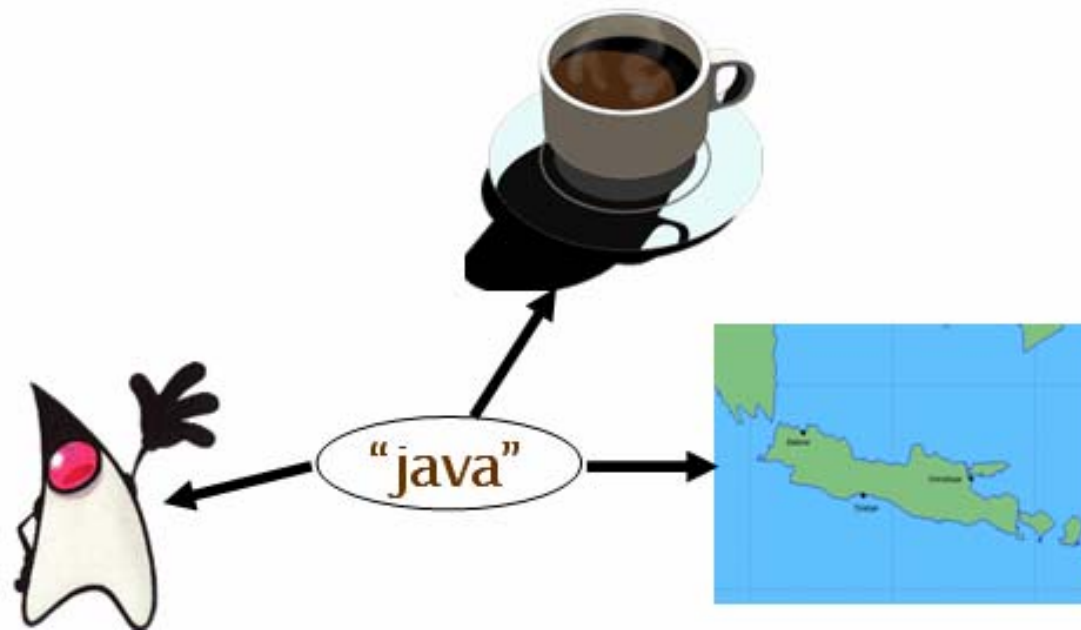
preliminariii

- **Necesități:**
 - O manieră de a atașa date privitoare la date \Rightarrow creare de **vocabulare de termeni** descriind “ceva” (proprietăți, domenii, lumi,...)
 - Un mecanism de a specifica relații între resurse \Rightarrow structuri de organizare a datelor
- \Rightarrow **Cunoștințe despre “lucruri”**
(*knowledge about Things*)



preliminariii

- Modelarea datelor trebuie să facă implicitul explicit



“java” \equiv limbaj, insulă ori sortiment de cafea?



reprezentarea metadatelor

- **Metadata** = date despre date
- Reprezentarea metadatelor:
 - **Resursă** – specificată de URI-uri
 - **Proprietate** – aspect, caracteristică, atribut sau relație care descrie o resursă
 - **Declarație** – triplu {subiect, predicat, obiect}
- **RDF** (*Resource Description Framework*)
 - **www.w3.org/RDF**



reprezentarea metadatelor

- Autorul resursei desemnate de URI-ul *http://www.infoiasi.ro/~web/* este Sabin Buraga

```
<rdf:RDF>
```

```
<rdf:Description rdf:about="http://www.infoiasi.ro/~web/">
```

```
<s:Autor>Sabin Buraga</s:Autor>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```

- Alternative sintactice (exemple):
 autor ("http://www.infoiasi.ro/~web/", "Sabin Buraga")
 descriere ("http://www.infoiasi.ro/~web/", "<Web />")
 angajat ("#csb", "http://www.uaic.ro/")



reprezentarea metadatelor

- Metadatele pot fi asociate unor resurse deja definite:

```
<rdf:Description rdf:ID="csb">
```

```
  <s:Rol>prezentator la &lt;Web /&gt;</s:Rol>
```

```
</rdf:Description>
```

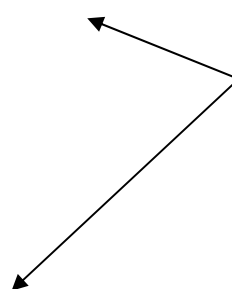
```
...
```

```
<rdf:Description rdf:about="#csb">
```

```
  <s:Adresa>busaco@infoiasi.ro</s:Adresa>
```

```
</rdf:Description>
```

Vocabular
de termeni



vocabulare de metadate

- **DCMI (*Dublin Core Metadata Initiative*)**
proprietăți pentru descrierea resurselor: <dc>Title>, <dc>Description>, <dc>Date>, <dc>Type>, <dc:Subject>, ...
 - www.dublincore.org/
- **RSS (*Really Simple Syndication*) 2.0**
mediatizarea (*syndication*) siturilor Web:
 - Noutăți & anunțuri (eventual, pentru *blog*-uri)
 - *Listing*-uri de documente (resurse noi, documentații adăugate, fișiere de *download*)
 - Calendare (evenimente, *deadline*-uri, ...)
 - <http://blogs.law.harvard.edu/tech/rss>
- **Atom**
 - www.atomenabled.org



vocabulare de metadate

- **Aplicații RSS/Atom** \Rightarrow *Social Web*
 - Acces: RSS Reader (C#, .NET), Firefox, Flock, Opera
 - Reflectă dinamica *blog*-urilor, forumurilor etc.
 - *Social Web tools*: Basecamp, CiteULike (Tcl, Common Lisp, Perl), Connotea (Perl), del.icio.us (Perl), Flickr.com, Google (Python), Frassle (Perl), Furl (Java), Spurl.net (PHP), Technorati.com, unalog (Python), Yahoo!
 - Agregatori – colectează & sumarizează
 - “Semne de carte” (*bookmarking*)
 - Imagini (*image sharing & aggregation*)
 - Notițe (*notes*)



vocabulare de metadate

- FOAF (*Friend Of A Friend*)
relații între persoane
 - Exprimă informații privitoare la o persoană și la relațiile acesteia cu alte persoane
 - Suport pentru e-comunități (*social networking*)
 - Relația *knows* (nu se specifică relații mai rafinate: *e.g.*, *friendOf*, *collaboratorOf*)
 - “Sabin o cunoaște pe Adria”:
 @prefix foaf: <<http://xmlns.com/foaf/0.1>>
 <#Sabin> <foaf:knows> <#Adria> .
 - Specializări: <http://purl.org/vocab/relationship>
 - <http://rdfweb.org/>



vocabulare de metadata

- **DOAP (Description Of A Project)**
descrierea unui proiect (software)
 - <http://usefulinc.com/ns/doap>
- **XMP (Extensible Metadata Platform)**
metadata “scufundate” în documentele produse de aplicațiile Adobe
 - SDK disponibil gratuit (ECMAScript)
 - <http://www.adobe.com/xmp>



vocabulare de metadate

- Specificarea relațiilor dintre trei persoane:

```
<foaf:Person rdf:ID="adria">  
  <foaf:name>Lenuta Alboaic</foaf:name>  
  <foaf:homepage rdf:resource="http://www.infoiasi.ro/~adria"/>  
</foaf:Person>  
<foaf:Person rdf:ID="cristi">  
  <foaf:name>Cristian Olaru</foaf:name>  
</foaf:Person>  
<foaf:Person rdf:ID="sabin">  
  <foaf:name>Sabin Buraga</foaf:name>  
  <foaf:knows rdf:resource="#cristi" />  
</foaf:Person>
```



vocabulare de metadate

- Specificarea unei imagini + metadatele aferente:

```
<foaf:Image rdf:about="http://www.sit.ro/bucuresti.jpg">  
  <dc:date rdf:datatype="xsd:date">2005-10-03</dc:date>  
  <dc:title>Intilnirea de la Bucuresti</dc:title>  
  <foaf:depicts rdf:resource="#cristi" />  
  <foaf:depicts rdf:resource="#sabin" />  
  <foaf:maker rdf:resource="#adria" /> <!-- 'fotograful' -->  
</foaf:Image>
```



microformate

- Idee: vocabularele de metadata să fie specificate direct în cadrul unor formate de redare a datelor
- Reutilizarea vocabulelor XML existente, disponibile liber și standardizate, și nu inventarea altelor noi
- **Microformate** ≡ seturi de vocabulare pentru realizarea de adnotări semantice direct în XHTML și alte limbaje de prezentare a datelor
- **www.microformats.org**



microformate

- Exemple:
 - *hCard* – informații de contact despre persoane, companii, organizații etc.
 - *hCalendar* – evenimente și orare
 - *hReview* – date despre comentariile emise de vizitatorii unui sit (*e.g.*, *blog*)
 - *XOXO (Extensible Open XHTML Outlines)* – bazat pe diverse module XHTML pentru a facilita procesarea conținutului de diverse entități software și, totodată, redarea acestuia în cadrul *browser*-ului



microformate

- Marcaje XHTML (e.g., `<div>`, ``) pentru specif. datelor & structurii lor, “clase” CSS pentru prezentare & asocierea de descrieri ale metadatelor

hCard Creator

[hCard](#)-o-matic

name

company

street

city **state** **zip**

phone


url

photo url

code

```
<div class="vcard">
  
  <a class="url fn" href="http://www.infoiasi.ro/~busaco/">Sabin
  Buraga</a>
  <div class="org">UAIC</div>
  <div class="adr">
    <div class="street-address">Berthelot, 16</div>
    <span class="locality">Iasi</span>,
    <span class="region">IS</span>
    <span class="postal-code">400475</span>
  </div>
```

preview



Tails

- Fairmont Hotel
 Fairmont Hotel
 San Jose, CA
- XTech
 2006-05-16 - 2006-05-19
 Amsterdam, Netherlands
[Website](#)
- WWW 2006
 2006-05-23 - 2006-05-26
 Edinburgh, Scotland
[Website](#)
- Where 2.0 conference:
 Geo Microformats BOF II:
 Developing a "directions"
 Microformat
 20060613T1930-0700 -
 20060613T2030-0700
 Paseo Room, Fairmont
 Hotel, San Jose, CA
[Website](#)
- @media
 2006-06-15 - 2006-06-18
 London, UK
[Website](#)

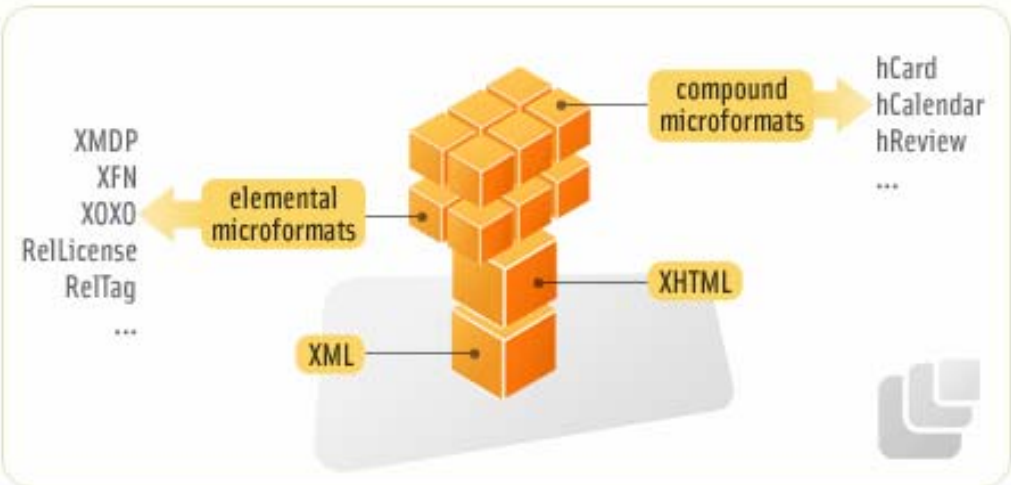


- blog
- wiki
- code**
- discuss
- about

About microformats

- about microformats**
- thanks
- people
- colophon

Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards. Instead of throwing away what works today, microformats intend to solve simpler problems first by adapting to current behaviors and usage patterns (e.g. XHTML, blogging).



microformats are:

- a way of thinking about data

Overview of

- People and [hCard](#)
- Calendars and [hCalendar](#)
- Opinions, [VoteLinks](#)
- Social Net [XFN](#)
- Licenses: [rel-license](#)
- Tags, Key [rel-tag](#)
- Lists and [XOXO](#)
- more... See [micro](#)

Upcoming ev

- [Subscribe...](#) ([M](#) [wiki](#))

microformate

- Triplele RDF și vocabularele de metadate (*e.g.*, DCMI, FOAF) pot fi exprimate via microformate
 - **RDF/A** – “scufundarea” RDF în XHTML
www.w3.org/TR/xhtml-rdfa-primer/
 - **XHTML 2.0** – suport direct pentru metadate
www.w3.org/TR/xhtml2



microformate

- **Studiu de caz:**

Pentru documentul Web de la adresa

<http://www.infoiasi.ro/~web/>

dorim să specificăm următoarele:

- Are titlul „Situl Summer <Web />”
- A fost vizitat de „ceva” care se numește Dragoș
- Acel „ceva” are situl personal la adresa
<http://www.adt.ro/>



microformate

- Folosind RDF, vom putea scrie:

```

<rdf:Description rdf:about="http://www.infoiasi.ro/~web/">
  <viz:vizitatDe>
    <rdf:Description>
      <foaf:homepage rdf:resource="http://www.adt.ro/" />
      <foaf:name>Dragoş</foaf:name>
    </rdf:Description>
  </viz:vizitatDe>
  <dc:title>Situl Summer &lt;Web /&gt;</dc:title>
</rdf:Description>

```



microformate

- În XHTML 2.0, vom specifica:
<html xmlns="http://www.w3.org/2002/06/xhtml12/">
<meta about="http://www.infoiasi.ro/~web/">
 <meta property="viz:vizitatDe">
 <meta property="foaf:name">Dragoş</meta>
 <link rel="foaf:homepage" href="http://www.adt.ro/" />
 </meta>
 <meta property="dc:title">
 Situl Summer <Web /></meta>
</meta>
</html>



microformate

- Alt exemplu XHTML 2.0:

```
<html xmlns="http://www.w3.org/2002/06/xhtml2/" xml:lang="ro">
  <head>
    <title>Fotografii personale</title>
    <meta property="dc:creator">Sabin Buraga</meta>
  </head>
  <body>
    <section>
      <h>Fotografii personale</h>
      <separator />
      <p src="foto" type="image/png, image/*">Fotografia mea</p>
      <meta about="#csb" property="title">Situl personal</meta>
      Alte fotografii sunt disponibile <p id="csb"
        href="http://www.infoiasi.ro/~busaco/">aici.</p>
    </section>
  </body>
</html>
```



microformate

- **Proiectare (inginerie):**
 - Alegerea unei probleme *specifice, simple*, de publicare a conținutului resurselor Web și definirea ei
 - Studierea & documentarea *manierei actuale de publicare pe Web* a conținutului dorit
 - Studierea *formatelor existente* conforme cu problematica în cauză
 - *Specificarea* unei scheme implicite sau *reutilizarea* unui vocabular existent
 - *Rafinarea* formatului în cadrul comunității



microformate

- **Principii de proiectare:**
 - Specifice unei probleme specifice
 - Cât mai simple posibil
 - Îmbunătățiri pe parcurs (evoluție)
 - Focalizate asupra utilizării de către oameni, apoi de către mașină
 - Caracter prezentațional & facil de procesat
 - Adaptare la comportamentul actual al *browser*-elor
 - Reutilizare/rafinare a standardelor adoptate
 - (X)HTML semantic, scheme XML existente etc.
 - Modularitate / capacitate de includere (“scufundare”) în alte formate
 - Dezvoltare, conținut, servicii descentralizate
 - Încurajarea explicită a “spiritului Web”



ce urmează?!...

- Cum interogăm metadatele?
Există un limbaj de interogare specializat?
- Declarațiile RDF (RDF/A) se situează la nivel de instanță (indivizi)
- Lipsește un mecanism de declarare a unui vocabular

```
<rdf:Description rdf:about="http://www.infoiasi.ro/~web/">
  <s:Organizator>Sabin-Corneliu Buraga</s:Organizator>
</rdf:Description>
```

- În loc de *organizator* putem folosi *fondator*?
- Resursa mai are și alți organizatori?
- Ce alte proprietăți specifice sunt asociate?



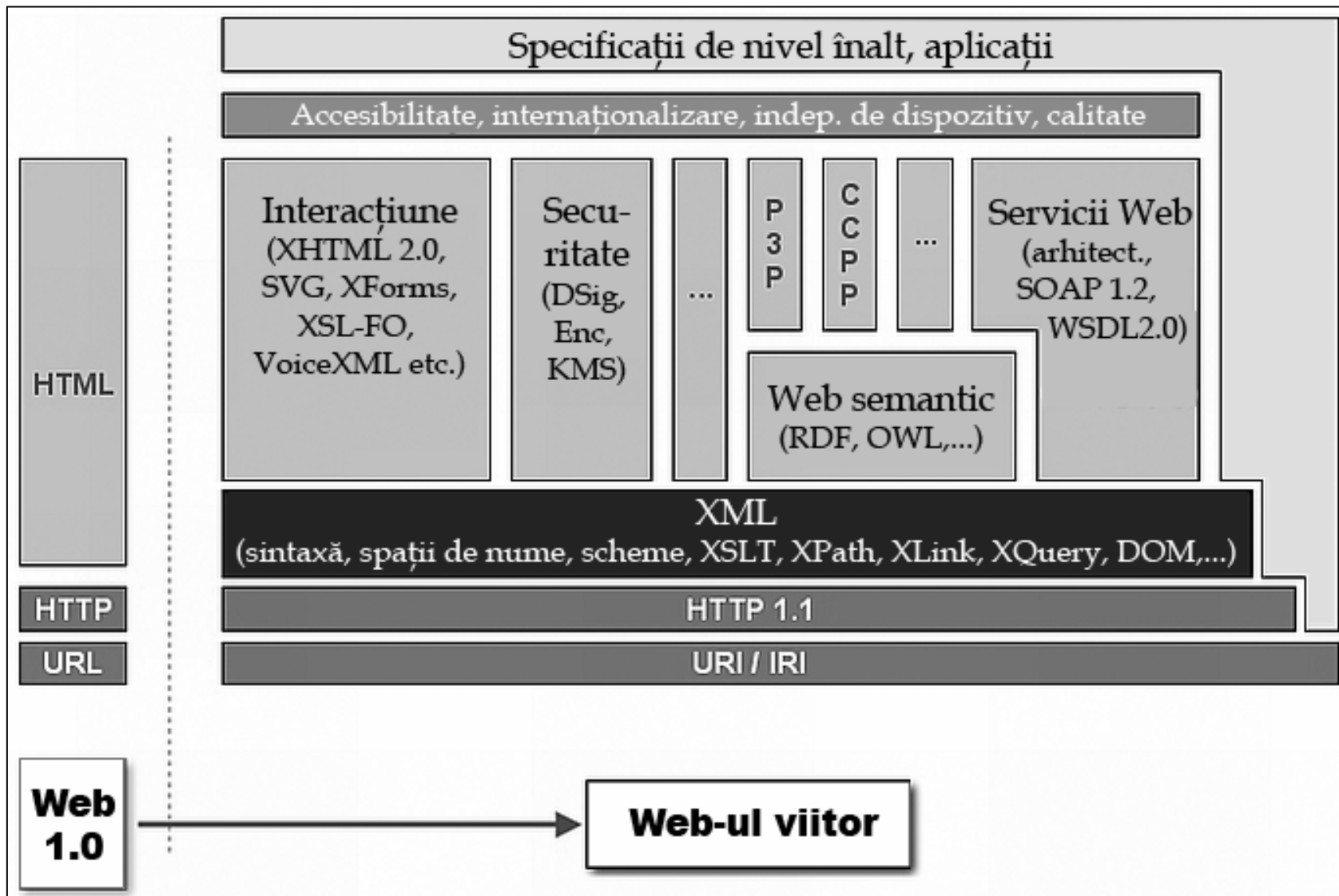
ce urmează?!...

- Limbaj de interogare pentru RDF: **SPARQL**
www.w3.org/TR/rdf-sparql-query/
- Modalități de modelare a cunoștințelor:
 - Taxonomii: **RDFS** (*RDF Schema*)
 - Tezaure: **SKOS** (*Simple Knowledge Organizational System*)
 - Ontologii: **OWL** (*Web Ontology Language*)
 - Reguli de raționament (automat): **RuleML**, **SWRL** (*Semantic Web Rule Language*)
- De urmărit evoluția spațiului Web
la **www.w3.org**



ce urmează?!...

De la Web 1.0 la Web-ul viitor



resurse

- S. Buraga, *Tehnologii XML*, Polirom, 2006
- S. Buraga, *Semantic Web*, Matrix Rom, 2004
- S. Buraga (coord.), *Tendințe actuale în proiectarea și dezvoltarea aplicațiilor Web*, Matrix Rom, 2006
- S. Buraga (coord.), *Situri Web la cheie*, Polirom, 2004
- Detalii la www.infoiasi.ro/~busaco/books.html



“exercițiu” surpriză (premiu!)

- Modelarea resurselor sitului Web privitor la participarea la **Summer <Web />** a unui grup de prieteni (studenți)
 - Descrieri (textuale) succinte ale prezentărilor și ale activităților desfășurate – în mai multe limbi
 - Colecții de fotografii făcute de participanții la **Summer <Web />** în diferite momente
 - Legături spre resurse de interes, corespunzătoare temelor prezentate, spre situl locației evenimentului și către siturile organizatorilor & persoanelor participante
 - Notificări ulterioare ale disponibilității de fotografii, de știri privitoare la premiile acordate și la siturile edițiilor viitoare etc.



rezumat

- Preliminarii
- Reprezentarea <meta />datelor
- Vocabulare de <meta />date
- Microformate
- Ce urmează?!...
- Resurse



De la <meta/>date la microformate

Mulțumiri pentru atenție!

Întrebări?

