Microformats

Presentation by,
Vikrant Nandakumar

Agenda

Definition

Why Microformats?

What are Microformats?

What microformats are not?

Principles, Purpose, Process

Microformat Specifications

Tools and demos

Definition

"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"

—microformats.org

Semantic Web vs. semantic web

	Semantic Web	semantic web
Philosophy	Build a common data format for expressing the meaning of data. Use ontologies to help machines to understand web content.	Humans first, machines second. Encode existing Web content with special tags.
Language	RDF, RDFS, OWL	Microformats
Format	Well-formed RDF documents	Valid XHTML documents
Semantic	Defined by the underlying ontology model	Loosely defined. No formal semantic model.

Slide adopted from Harry Chen's Geospatial Semantic Web Presentation

Why Microformats?

◆ Currently the information is either

- Human readable
  - Books
  - Rendered HTML

- Machine readable
  - Punchcards
  - RDF
  - XML
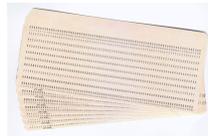
```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/
22-rdf-syntax-ns#"
xmlns:terms="http://purl.org/dc/terms
/"> <rdf:Description
rdf:about="urn:x-states:New%20York">
<terms:alternative>NY</terms:alternat
ive> </rdf:Description> </rdf:RDF>
```

```
BEGIN:VCARD VERSION:2.1
N:Gump;Forrest FN:Forrest Gump
ORG:Bubba Gump Shrimp Co.
TITLE:Shrimp Man
TEL;WORK;VOICE:(111) 555-1212
TEL;HOME;VOICE:(404) 555-1212
ADR;WORK:;;100 Waters
Edge;Baytown;LA;30314;United
States of America
LABEL;WORK;ENCODING=QUOTED-
PRINTABLE:100 Waters
Edge=0D=0ABaytown, LA
30314=0D=0AUnited States of
America ADR;HOME:;;42 Plantation
St.;Baytown;LA;30314;United States
of America
LABEL;HOME;ENCODING=QUOTED-
PRINTABLE:42 Plantation
St.=0D=0ABaytown, LA
30314=0D=0AUnited States of
America
EMAIL;PREF;INTERNET:forrestgump@ex
ample.com REV:20080424T195243Z
END:VCARD
```

◆ Microformats serve this purpose by making it possible to represent information, human readable as well as machine readable.
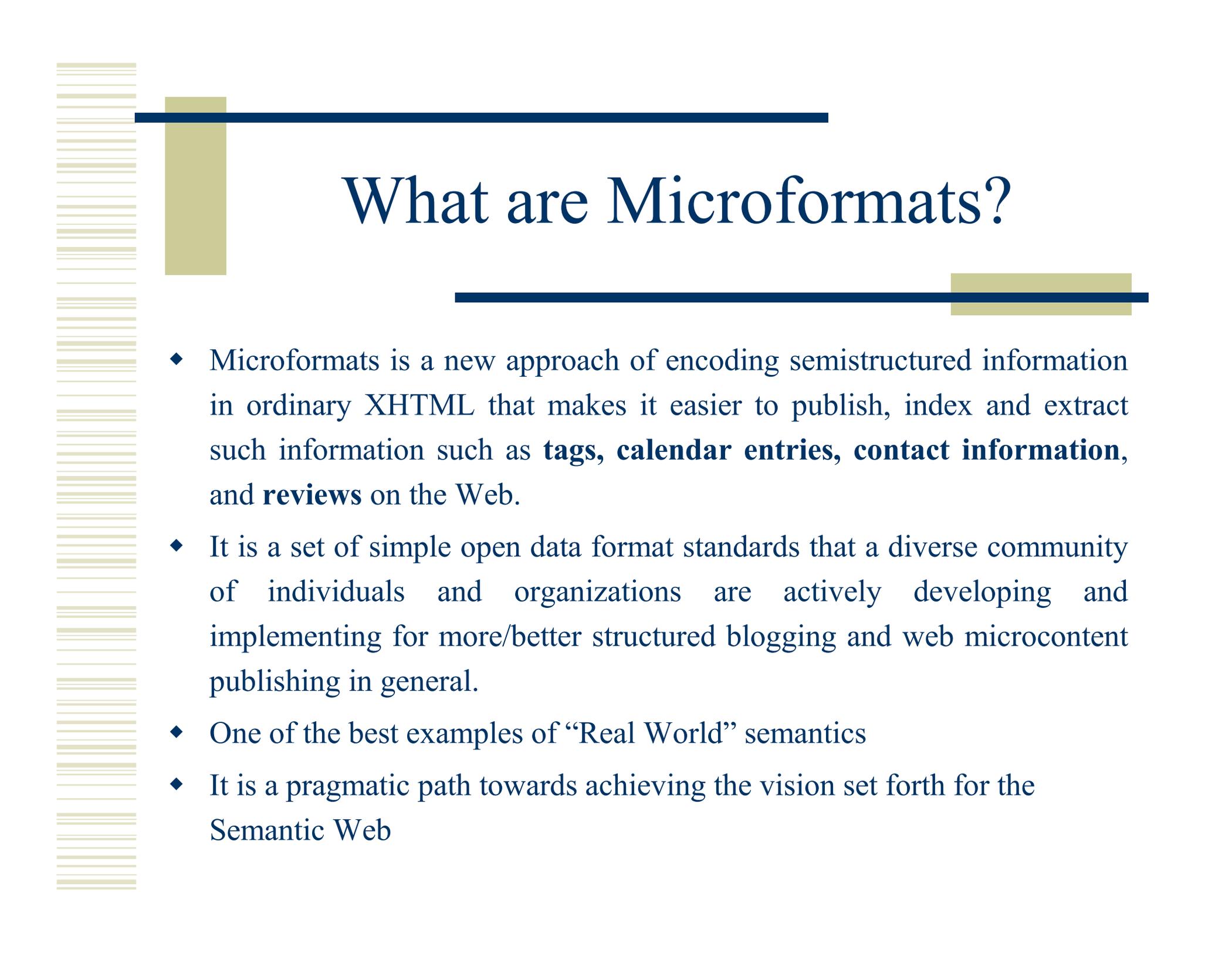
What are Microformats?

- Microformats is a new approach of encoding semistructured information in ordinary XHTML that makes it easier to publish, index and extract such information such as **tags, calendar entries, contact information**, and **reviews** on the Web.

- It is a set of simple open data format standards that a diverse community of individuals and organizations are actively developing and implementing for more/better structured blogging and web microcontent publishing in general.

- One of the best examples of "Real World" semantics

- It is a pragmatic path towards achieving the vision set forth for the Semantic Web

What Microformats are not?

- A new language
- An attempt to get everyone to change their behavior and rewrite their tools
- A whole new approach that throws away what already works today
- A panacea for all taxonomies, ontologies, and other such abstractions
- Do not address implicit knowledge representation, ontological analysis, or logical inference.
- Controlled by any individual or organization
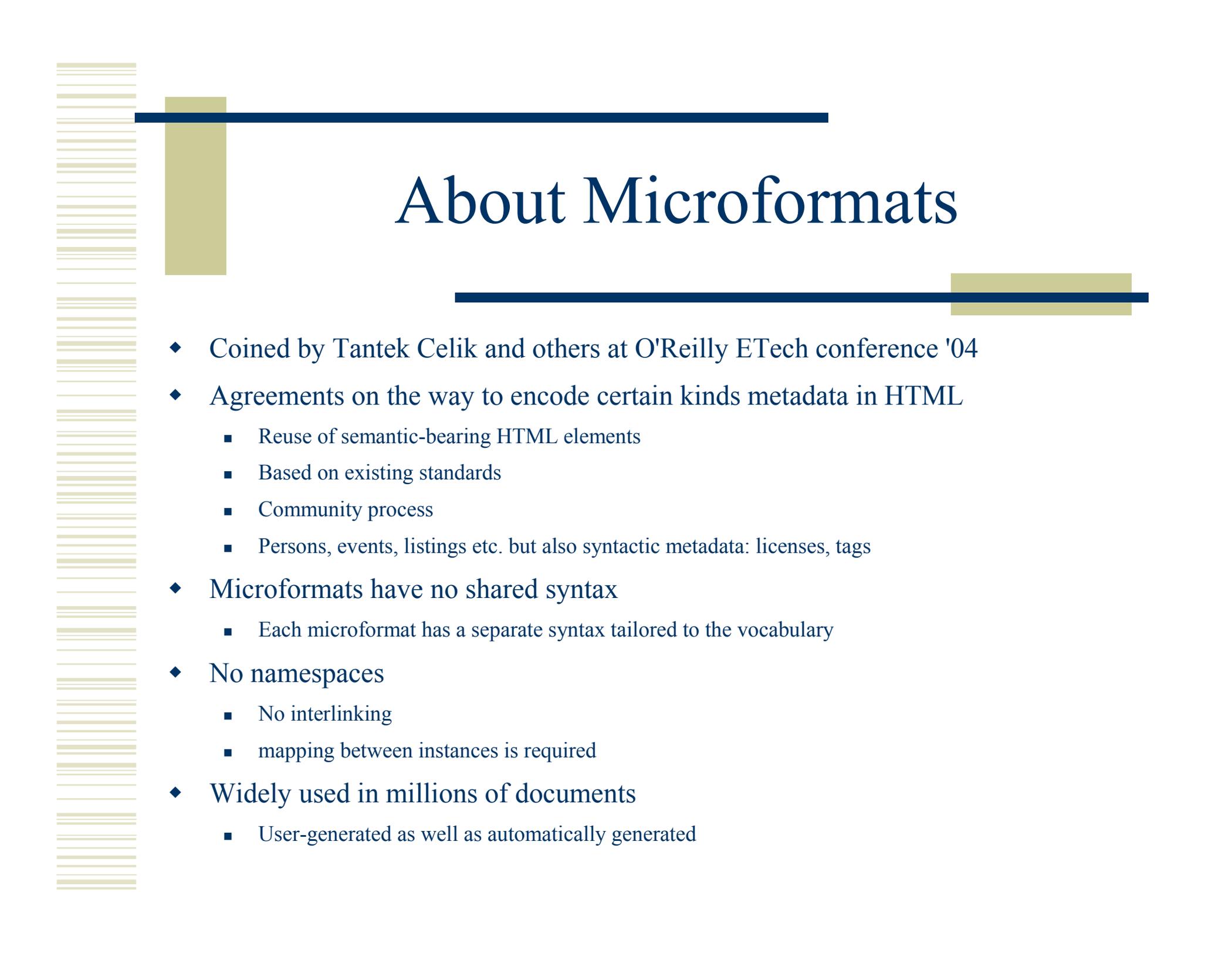
About Microformats

- Coined by Tantek Celik and others at O'Reilly ETech conference '04
- Agreements on the way to encode certain kinds metadata in HTML
  - Reuse of semantic-bearing HTML elements
  - Based on existing standards
  - Community process
  - Persons, events, listings etc. but also syntactic metadata: licenses, tags
- Microformats have no shared syntax
  - Each microformat has a separate syntax tailored to the vocabulary
- No namespaces
  - No interlinking
  - mapping between instances is required
- Widely used in millions of documents
  - User-generated as well as automatically generated

Principles of Microformats

- Solve a specific problem
- Start as simple as possible
  - solve simpler problems first, then make evolutionary improvements
- Design for humans first, machines second
  - should be presentable and parsable
  - visibility and human friendliness - visible data is much better for humans than invisible metadata
  - adapt to current behaviors and usage patterns, e.g. (X)HTML, blogging
  - ease of authoring is important

Principles of Microformats

- Reuse building blocks from widely adopted standards:
    - semantic, meaningful (X)HTML i.e POSH (Plain Old Semantic HTML).
    - existing microformats (hcard, hcalendar, ..)
    - well established schemas from interoperable RFCs
- Modularity / embeddability
    - design to be reused and embedded inside existing formats and microformats
- Enable and encourage decentralized and distributed development, content, services

Goals of the principles

The objectives, goals and the effects of the principles are stated below:

- Data Integrity
  - Visible data = more accurate data
  - Not repeating yourself (DRY – Don't Repeat Yourself)
  - Multi-language integrity
- Lower barriers for publishers - more publisher-centric in design, rather than parser-centric
  - human-centric in design
- User centered data reuse
  - mark up data semantically, which enables *general* re-use

Purpose of Microformats

The purpose of Microformats is to:

- ◆ Make Web documents smarter - HTML, XHTML, BLOGS, RSS, ATOM, XML

- ◆ Enrich the semantics of Web documents.

- ◆ Enrich the semantics of the whole Web.

  - ■ In a bottom-up, grassroots fashion, one Web document at a time.

- ◆ Create standard semantic labels.

  - ■ via a clearly defined community process.

- ◆ Leverage existing standards (don't reinvent).

Microformat Specifications

◆ XHTML and HTML standards allow for semantics to be embedded and encoded within the attributes of markup tags.

◆ Microformats take advantage of these standards by indicating the presence of metadata using the following attributes:

- class
- rel
- rev

Microformat Specifications

- ◆ rel-license
  - ■ A format for indicating content licenses in hyperlinks
- ◆ rel-nofollow
  - ■ A format for indicating that the associated hyperlink should not receive additional weight or ranking by user agents which perform link analysis upon web pages (e.g. search engines)
- ◆ rel-tag
  - ■ A format for indicating that the destination of that hyperlink is an author-designated "tag" for the current page
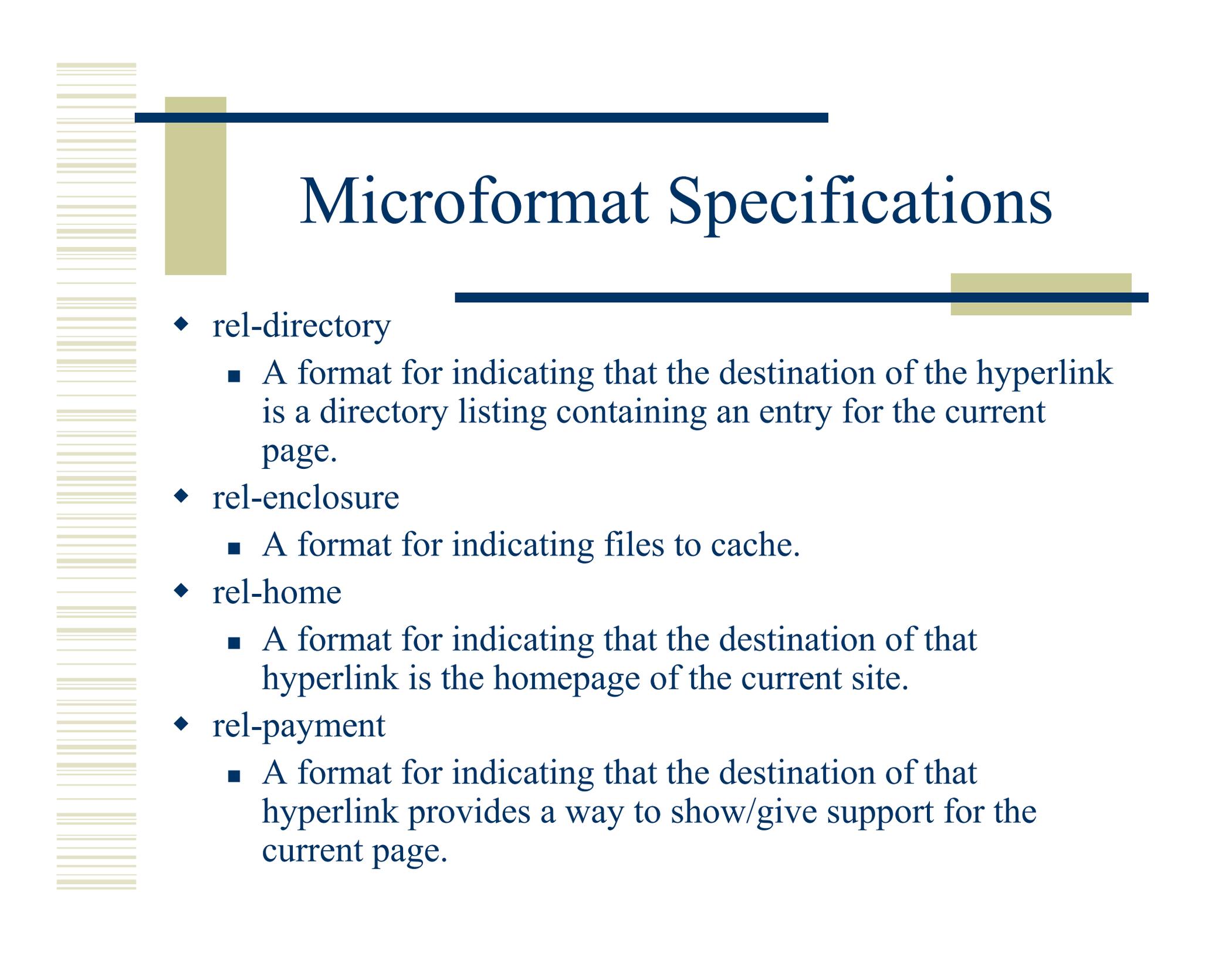
Microformat Specifications

- rel-directory
  - A format for indicating that the destination of the hyperlink is a directory listing containing an entry for the current page.
- rel-enclosure
  - A format for indicating files to cache.
- rel-home
  - A format for indicating that the destination of that hyperlink is the homepage of the current site.
- rel-payment
  - A format for indicating that the destination of that hyperlink provides a way to show/give support for the current page.

Microformat Specifications

- **VoteLinks**
  - A format expressing three new values for the rev attribute of the <a> (hyperlink) tag in HTML:
    - vote-for
    - vote-abstain
    - vote-against
- **XFN (XHTML Friends Network)**
  - A simple way to represent human relationships using hyperlinks. Eg: Brad
- **XMDP (XHTML Meta-Data Profiles)**
  - A format for defining HTML meta data profiles.
- **XOXO (Extensible Open XHTML Outlines)**
  - A simple, open outline format

Microformat specifications

- **hCard** – format for representing people, companies, organizations and places, using a 1:1 representation of vCard properties and values in semantic HTML or XHTML

- **hCalendar** – distributed calendaring and events format, based on the iCalendar standard, suitable for embedding in HTML or XHTML, Atom, RSS, and arbitrary XML

- **hReview** – Enables and encourages the sharing, distribution, syndication and aggregation of reviews

- **hResume** – is used for publishing resumes and CVs. It is based on a set of fields common to numerous resumes published on the web. Possible field names are chosen and reused from **preexisting microformats**
    - Contact info - use hCard; use <address> + hCard.
    - Experience - One or more hcalendar events with the class name 'experience', with an embedded hCard indicating the job title, name of company, address of company etc.
    - Education - One or more hcalendar events with the class name 'education', with an embedded hCard indicating the name of school, address of school etc.

Easy creation of Microformat specifications

- ◆ hCard – for contact information
  - ● http://microformats.org/code/hcard/creator
- ◆ hCalendar - for events
  - ● http://microformats.org/code/hcalendar/creator
- ◆ hReview - for reviews
  - ● http://microformats.org/code/hreview/creator
- ◆ hResume - for resumes or CVs
  - ● http://hresume.weblogswork.com/hresumecreator/

Uses of Microformats

- Using microformats within HTML code provides additional formatting and semantic data that can be used by applications

- Applications could be the ones that collect data about on-line resources, such as web crawlers, or desktop applications such as e-mail clients or scheduling software

- Can be used to facilitate "mash ups" such as exporting all of the geographical locations on a web page into Google Maps, to visualize them spatially

Tools, plug-ins

- ◆ "Operator" extension for Mozilla Firefox
  - ◆ Demo   View Phone
- ◆ "Oomph" plug-in for MS Internet Explorer
  - ◆ Demo
- ◆ Yahoo Query Language can be used to extract microformats from web pages
  - ◆ Demo
- ◆ Share your microformats!
  - ■ Pingerati receives updates of pages with microformats from numerous sources, and sends those pings to services that support microformats
    - ◆ How?

Comparison of Microformat vs RDFa

Microformat	RDFa
No namespaces (e.g., URI, dirs, prefixes)	Namespaces
Limited vocabulary	Any vocabulary (e.g., FOAF)
Common problems and solutions	Open
Not extensible	Extensible
Visible data favoured	Meta data is okay
Community decides definition	Anyone can create their own definitions
Fairly easy to implement	Moderately easy
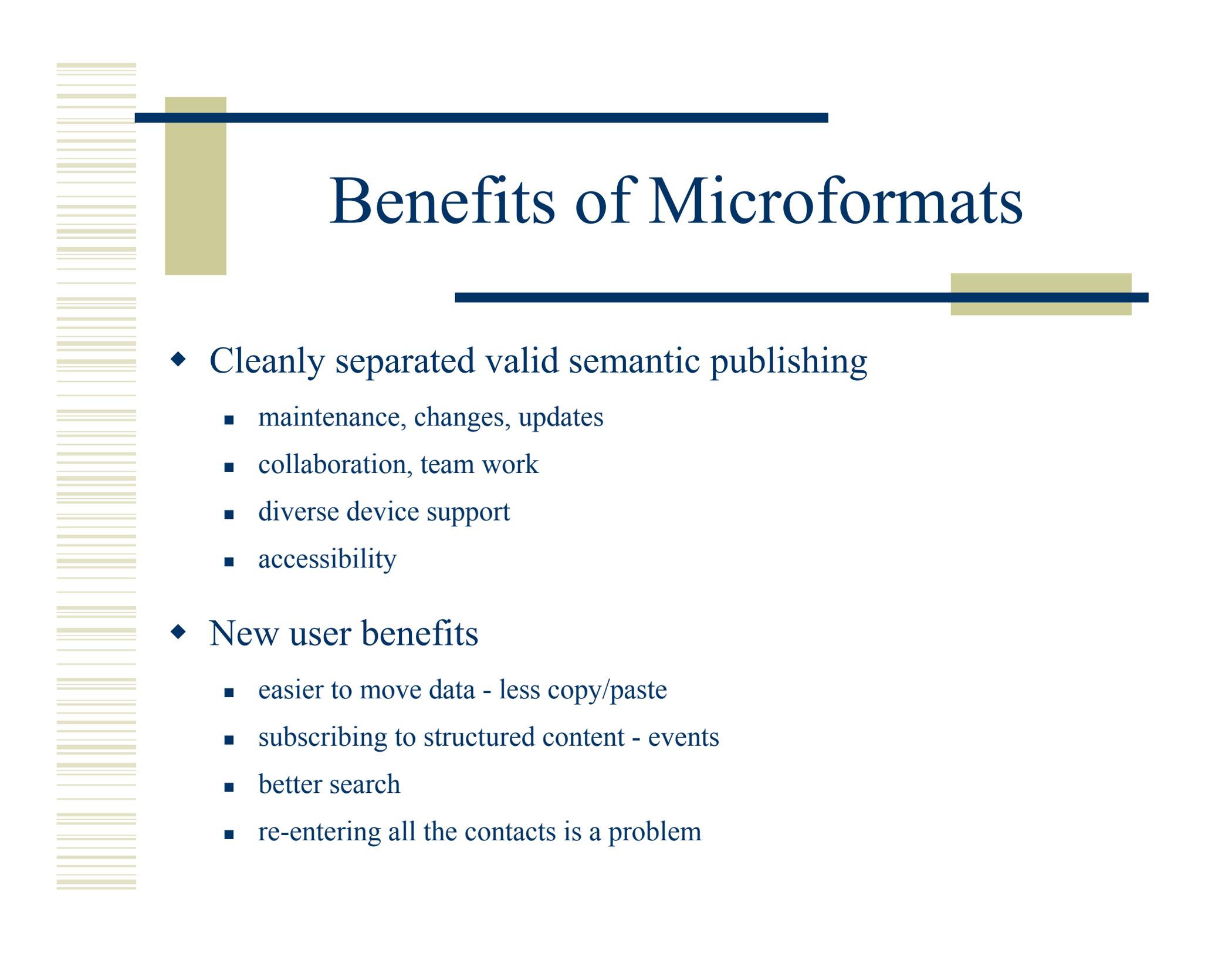
Benefits of Microformats

- Cleanly separated valid semantic publishing
  - maintenance, changes, updates
  - collaboration, team work
  - diverse device support
  - accessibility

- New user benefits
  - easier to move data - less copy/paste
  - subscribing to structured content - events
  - better search
  - re-entering all the contacts is a problem
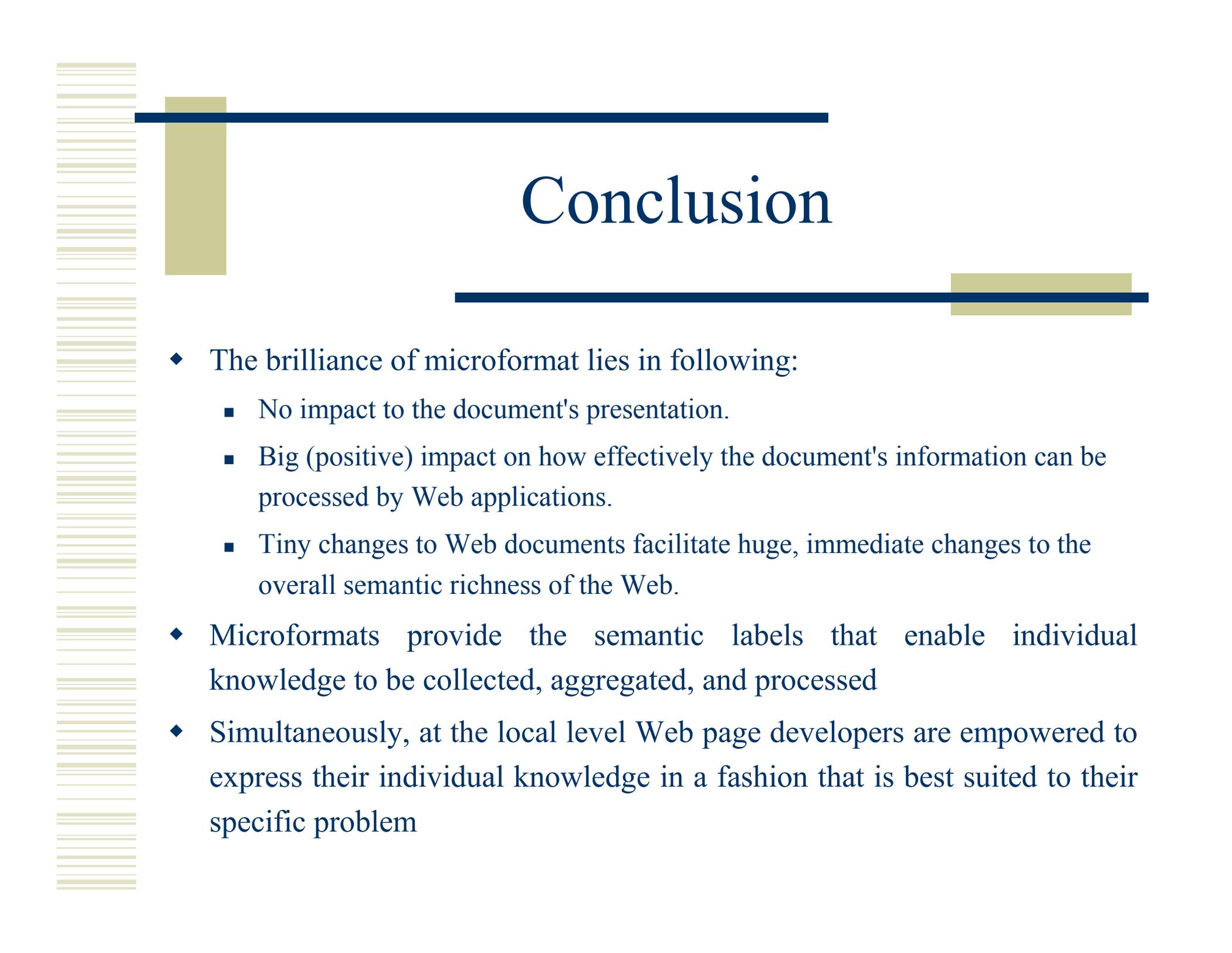
Conclusion

- The brilliance of microformat lies in following:
  - No impact to the document's presentation.
  - Big (positive) impact on how effectively the document's information can be processed by Web applications.
  - Tiny changes to Web documents facilitate huge, immediate changes to the overall semantic richness of the Web.
- Microformats provide the semantic labels that enable individual knowledge to be collected, aggregated, and processed
- Simultaneously, at the local level Web page developers are empowered to express their individual knowledge in a fashion that is best suited to their specific problem

References

- http://microformats.org/
- http://www.deitel.com/ResourceCenters/Web20/Microformats/MicroformatsSpecifications/tabid/966/Default.aspx
- http://en.wikipedia.org/wiki/Microformat
- Rohit Khare, Tantek Çelik *"Microformats: a Pragmatic Path to the Semantic Web"*

Thank you