

Semantic Injection of Institutional Websites with New Microformats for Higher Education

Stefan Eicker¹, J. Peter M. Schuler², Peter M. Schuler³

¹University of Duisburg-Essen, Faculty of Business Administration and Economics, Research Group for Business Informatics and Software Engineering, Universitätsstr. 9, 45141 Essen, stefan.eicker@icb.uni-due.de.

²ibid., j.peter.m.schuler@icb.uni-due.de. ³ibid., peter.schuler@icb.uni-due.de.

Keywords

Semantic Web, Microformats, Course Catalogue, Higher Education, Bologna, MLO

1. EXECUTIVE SUMMARY

This paper discusses the use of existing and new microformats for semantic injection of institutional websites in order to make their valuable information more accessible to students - directly via client tools or indirectly via information portals.

1.1. Background

The Bologna process increased the demand for information on degree programmes. The Faculty of Business Administration and Economics at University of Duisburg-Essen uses two higher-education-specific content management solutions to satisfy this (and other) information demands by web-based course catalogues.

Thanks to pan-European standardized degrees students may face a wide market for degree programmes. In the best case institutional websites present detailed information on degree programmes and related information in an intuitive way. Of course additional effort is necessary to compare different universities, their academics and degrees due to a lack of consistency and homogeneity.

Microformats as a semantic technology enable the markup of existing websites and thus create semantically rich information. Since the two systems at the Faculty of Business Administration and economics are template based, inserting microformats seems an easily applicable solution - suitable microformats assumed.

1.2. Focus

After discussing the problem statement and introducing the concept of microformats the paper firstly presents the analysis of existing microformats for the given scenario. Secondly Metadata for Learning Opportunities (MLO) is reviewed as a possible blueprint for the semantic structure of course information to design new microformats for higher education. Thirdly scenarios for the future usage of these microformats in form of new application concepts for access of course related information are presented: web browser add-ons understanding the semantic, client tools retrieving the embedded semantic without the indirection via a web browser (e.g. a course catalogue reader for smart phones) and information portals retrieving semantically enhanced information of institutional websites from different universities for comparing their degree programmes.

1.3. Conclusion

This paper closes with a reflection of the approach, the current state of the analysis, possible alternatives and open work items.

2. THE PROBLEM STATEMENT

The Bologna process increased the demand for information on degree programmes and courses. This demand can be satisfied by institutional websites rich in content e.g. by using web-based course catalogues. Thanks to pan-European standardized degrees students may face a wide market for degree programmes - e.g. either to select their degree programme or to choose a university for spending a semester abroad. In the best case institutional websites present detailed information on study courses and related information in an intuitive way. Of course additional effort is necessary to compare different universities, their academics and degrees due to a lack of consistency and homogeneity.

2.1. The Scenario at University of Duisburg-Essen

The Faculty of Business Administration and Economics at University of Duisburg-Essen uses two content management solutions to improve student information delivery: chairT3 and bolognaT3. Both systems are based on the open source content management system TYPO3 and harmonize the presentation of information on staff, lectures, degree programmes, etc. on the faculty's website.

The valuable information managed in bolognaT3 and chairT3 at University of Duisburg-Essen comprises the elements listed in Table 1. For an introduction to the web-based course catalogue system bolognaT3 see (Eicker et al. 2008).

Table 1: Information managed by chairT3 and bolognaT3

chairT3	bolognaT3
Person	Course Catalogue
Lecture	Module Group
Thesis	Module
Publication	Course
Talk	Lecturer

With these two flexible web-based solutions at hand, the question arose how the accessibility of this valuable information may be increased. The two systems present the information to the web user who visits the pages - i.e. to the user who knows that the information he is looking for is published on the institutional website. When regarding potential consumers of the faculty in the current scenario one has to distinguish between different student groups:

1. Current students

The first group might know the faculty's information portal but might be interested in an easier and personalized way to only get the specific information they need or to get quicker to the place where it is presented. The current situation provides one single point of access to the information for every student - independent of her degree programme or current progress (e.g. <http://www.studium.wiwi.uni-due.de/>).

2. Prospective students

The second group might not even know where to get the right information on degree programmes and details on the courses but might access central information portals that try to prepare an overview of different higher education providers. These portals often apply editorial approaches so that no automatic data exchange is realized. The portals often have national focus (e.g. <http://www.studieren.de/>) - some try a pan-European approach. Searching for degree programmes on the web often results in pdf documents displaying structured information without the ability to access or compare the information in a structured way.

3. Exchange students

The third group is a mixture of the former ones. Students have to choose courses of degree programmes from abroad but at the same time ensure the suitability for their own degree programme. Besides general higher education information portals special portals try to support exchange students when choosing suitable courses. These often are provided by the home institutions - with limited scope - or try to harmonize the academic offers only for specific domains (e.g. in case of information science <http://www.is-link.org>).

In case of the Faculty of Business Administration and Economics at University of Duisburg-Essen the current situation is kind of a dilemma. Through the use of chairT3 and bolognaT3 valuable structured information is available in the web but not accessible - at least not in a way the students need it. Current appendages of better information exchange are focused on special databases or editorial portals. Thus without intentionally restricting the access to the information the information delivery in fact is restricted to an institution-centric approach - which is not compliant to the intention of the Bologna process.

2.2. Applicability of Semantic Specifications

Even though standards like Metadata for Learning Opportunities (MLO) address the standardization of metadata for courses, the question still remains how these standards will be used. XML specifications are well applicable for import or export interfaces e.g. to databases. They cause the systems to create an additional interface (for the data exchange) besides the existing (user) interface - thus increasing the number of interfaces to manage.

The semantic web instead tries to inject the necessary metadata directly into the web. "Among other things, the semantic web makes information more meaningful to people by making it more understandable to machines" (Ohler 2008). Microformats as a semantic technology aim at injecting semantics into the existing interface - more precise: the web-based user interface - and have a focus on specific semantic problems.

This paper thus discusses the applicability of the concept of microformats to inject semantics into institutional websites. This way a semantic structure can be embedded directly into the website - invisible to the human user, but accessible by software understanding the embedded semantic and enabling new ways of accessing information on institutional websites. In the given scenario at the Faculty of Business Administration and Economics this approach seems - from the point of view of the technical feasibility - very attractive because chairT3 and bolognaT3 use a template based approach. Thus the implementation of microformats into the systems only implies a simple addition to the XHTML template¹ - assumed the existence of suitable microformats.

3. MICROFORMATS

The main idea of microformats is the semantic markup of content published in the World Wide Web. At first glance semantic technologies seem to be complex and time-consuming but the user benefits from the richer web of data and the easier access to it (Friedman 2009). The following quotes get the idea of microformats to the point:

»Designed for humans first and machines second, microformats are a set of simple, open data formats built upon existing and widely adopted standards.«
(Microformats.org About)

Microformats thus are a semantic technology that relies on existing - adopted - standards when creating a new set of data formats that clearly focuses on the human user of the data.

»Microformats are a way of adding simple markup to human-readable data items such as events, contact details or locations, on web pages, so that the information in them can be extracted by software and indexed, searched for, saved, cross-referenced or combined.«
(Microformats.org Introduction)

¹ The same principle applies e.g. to other XML standards like MLO, etc. but they require the systems to provide two interfaces - one for the user and one for the machines.

Microformats are simple markups added to existing web pages. The semantic is not hidden metadata but identical to the information visible to the user. Thus the information gets understandable to machines enabling better access to it and new application scenarios.

»More technically, they are items of semantic markup, using just standard "plain old semantic (X)HTML" (i.e. "POSH") with a set of common class-names and "rel" values. They are open and available, freely, for anyone to use.«
(Microformats.org Introduction)

The low technical barrier to implement microformats enables easy adoption by information providers and application developers.

The design of microformats is based on a number of key principles which contribute to the goals of data integrity, low barriers for information publishers and user-centric data re-use. In contrast to generic semantic technologies microformats aim at solving specific problems. For example RDF is a generic framework to embed semantic information. It enables institutions to create their own semantic structures. In contrast a microformat defines a standard specification that focuses on one specific problem. The microformat thus is not applicable to other contexts (except for some generic microformats) but the semantic structure is comprehensive and cross-institutional applicable. It has a smaller (semantic) scope but potentially a wider impact.

3.1. A Microformat Example: hCard

To illustrate the concept of microformats the hCard microformat will be presented. The semantic blueprint is the vCard, a specification of the data format for an electronic business card (Dawson and Howes 1998). Figure 1 illustrates the vCard.

```
BEGIN:VCARD
VERSION:3.0
N:Mustermann;Max
FN:Max Mustermann
ORG:Wikipedia
URL:http://de.wikipedia.org/
EMAIL;TYPE=INTERNET:max.mustermann@example.org
TEL;TYPE=voice,work,pref:+49 1234 56788
ADR;TYPE=intl,work,postal,parcel;;;Musterstraße 1;Musterstadt;;12345;Germany
END:VCARD
```

Figure 1: vCard Example (Source: <http://de.wikipedia.org/wiki/VCard>)

The hCard is the microformat specification of vCard, i.e. a set of classes and attributes that transfer the vCard elements into XHTML compatible tags. Figure 2 illustrates the hCard structure.

```
<div id="hcard-Max-Mustermann" class="vcard">
  <span class="fn">Max Mustermann</span>
  <div class="org">Wikipedia</div>
  <a class="email"
href="mailto:max.mustermann@example.org">max.mustermann@example.org</a>
  <div class="adr">
    <div class="street-address">Musterstraße 1</div>
    <span class="locality">Musterstadt</span>
    <span class="postal-code">12345</span>
    <span class="country-name">Germany</span>
  </div>
  <div class="tel">+49 1234 56788</div>
</div>
```

Figure 2: hCard Example

Both code snippets can be generated by tools, e.g. content management systems. The second code (hCard) can be used as XHTML code within a website and in this way integrates the information itself and its semantic structure in one place.

Figure 3 illustrates a website with an embedded hCard making the semantic accessible to the browser addon Operator (<https://addons.mozilla.org/de/firefox/addon/4106>).



Figure 3: Example of chairT3.Person with embedded hCard semantic

3.2. The Microformat Standardization Process

The definition of new microformats follows a community-driven standardization process (Microformats.org Process): As an introductory step the semantic injection of the own website is suggested to gain “real world” experience with microformats.

The next step is to identify the specific problem that should be solved by the new microformats and to document the current user behavior. If these prerequisites are fulfilled new microformats might be proposed.

Concerning the semantic definition these should be oriented on existing semantic specifications. The process of standardization itself iterates several times through the phases problem statement, discussion/research and proposal/draft.

4. EXISTING & NEW MICROFORMATS FOR INSTITUTIONAL WEBSITES

Concerning the information elements managed by chairT3 and bolognaT3 some elements can be identified where existing microformats might be applied. Other areas obviously need additional elements to enable a semantic injection of the institutional website.

Table 2: Mapping of information elements to existing microformats

Information element	Existing microformats
chairT3.Person	hResume (hCard)
chairT3.Lecture	hEvent
chairT3.Thesis	-
chairT3.Publication	citation
chairT3.Talk	hEvent
bolognaT3.Course-Catalogue	-
bolognaT3.Module-Group	-
bolognaT3.Module	-
bolognaT3.Course	-
bolognaT3.Lecturer	hCard
Attachments in Lectures, Publications, Talks, etc.	rel="enclosure"
General address information, when hCard does not apply (e.g. rooms/institution)	adr

Table 2 shows that there are areas where existing microformats can be used to add semantics to institutional websites. E.g. the hCard microformat can be embedded into the personal website of employees.

5. MLO AS BLUEPRINT FOR NEW MICROFORMATS

The more academic information with focus on the degree programmes is not yet supported by specific microformats - except for citation. Even if a lecture might well be described as an hEvent the specific information elements in the higher education sector definitely miss semantic containers in form of microformats that are able to classify the hEvent lecture as a lecture.

MLO aims at the harmonization of several course description metadata specifications in the European learning environment. Thus MLO lends itself to be used as the semantic blueprint for microformats in the higher education area. MLO covers a small set of entities:

- Learning Opportunity Provider
- Learning Opportunity Specification
- Learning Opportunity Instance

The MLO specification has been reviewed to check the possibility to use the MLO elements as a blueprint for new microformats. Table 3 illustrates the result of the analysis.

Table 3: Proposal of new microformats (μ f) for HE-specific information elements

Information element	Existing μ f	Proposed μ f
chairT3.Person	hResume (hCard)	-
chairT3.Lecture	hEvent	-
chairT3.Thesis	-	-
chairT3.Publication	citation	-
chairT3.Talk	hEvent	-
bolognaT3.Course-Catalogue	-	degree-programme
bolognaT3.Module-Group	-	- / part of degree-programme (?)
bolognaT3.Module	-	course-unit (?)
bolognaT3.Course	-	course-unit
bolognaT3.Lecturer	hCard	course-unit.contributer
Attachments in Lectures, Publications, Talks, etc.	rel="enclosure"	-
General address information, when hCard does not apply (e.g. rooms/institution)	adr	-

The comparison of information elements, existing microformats and possible new microformats shows the need for higher-education-specific semantics. A semantic blueprint is definitely necessary to start the definition of new microformats. MLO provides the benefit of a European approach to harmonize course description information. These elements fit well to the elements of bolognaT3 even if some changes might be necessary. A detailed analysis of other specifications including the data fields of each element is required.

Information elements of chairT3 are not covered by MLO but instead by existing microformats. This means that they are easier to adapt and the benefit for the user might be achieved more quickly. The injection of semantics, e.g. by implementing hCard, hEvent, etc. into the template of chairT3 enables the use of the semantics without defining new microformats.

6. APPLICATION SCENARIOS

This section sketches three application scenarios enabled by semantic injection of institutional websites: Easier access to information embedded in semantic websites, specific client application - e.g. for smart phones and information portals understanding institutional websites. All three

scenarios make use of the semantic injection by microformats. Through unification of the website and the semantic interface there is no need for an additional interface.

Besides the sketched application scenarios search engines are able to make use of the semantic rich content of the institutional websites.

6.1. Browser Addons

As seen in Figure 3 browser addons are available that retrieve information framed by microformats and pass them to other applications for further processing. This way the user easily accesses the data placed in the website currently visited. E.g. a website with several hCards must not be scrolled but the relevant contact data might directly be submitted to the address book application.

As with hCard the same might apply to higher education specific information. E.g. the student accesses the faculty's website to fill the schedule for the upcoming semester. While reading the information on an interesting course, the browser plugin indicates hEvent data embedded in the website enabling the student to transfer the lecture data directly into the calendar application.

Specific data like publications marked up by citation microformat might be used to pass the data to a word processing application to reference the work of researchers in a paper or thesis.

6.2. SmartPhone Applications

The second application type replaces the browser and thus accesses the World Wide Web with a specific intention. Students that want to access specific data with relevance to the current semester of their degree program might not need to visit the institutional website. An application like a course catalogue reader, e.g. for smart phones, might parse the semantic information of the website and prepare it in a way suitable for mobile phones. Students would be enabled to access information on their courses with a fingertip and choose their courses for the upcoming semester, identify lecturers or access course materials.

In order to enable the application to differentiate between relevant information (that is the semantic injected data) and irrelevant information (e.g. XHTML elements that specify the design and layout of the website) microformats are used. The application does not need to access a separate XML interface or download the data - it just can understand the semantic enhanced data of the websites.

6.3. Portals understanding information

The limitations of current information portals (editorial effort, no semantic, scope, proprietary interface, actuality) can be solved by the use of microformats for higher education. Instead of editorial work or interface implementation an information portal might retrieve the semantically markedup information directly from the institutional website. The use of microformats helps the portal to understand the information presented in the website and thus enables it to make it comparable and searchable. Institutions could advertise their courses automatically by just maintaining their own institutional websites.

7. CONCLUSION

The proposed approach is a smart data integration solution based on microformats as a technology of the semantic web. The sketched application concepts are only three possible improvements of the way students currently access information. Even if some details of the microformats design process have to be validated the main task will remain to clarify the semantic background. MLO seems a suitable candidate because it aims at the harmonization of the course metadata in the European learning space.

The current project only addresses the data provided by the systems chairT3 and bolognaT3. There might evolve additional areas where information is provided by higher education institutions. Once the universities have semantic injected websites additional information might increase the value for students: Semantically annotated reviews of courses provided by a huge community of higher education consumers might increase the value of information portals aggregating details on several institutions.

Institutional websites injected with microformats make a step forward to Ohler's vision on the future of semantically enriched institutional websites:

"At some point, institutions will describe courses and degrees semantically, probably just to help their own internal functioning, but with the secondary effect of making many of the components of education at least somewhat comparable across institutions. It is a short leap from that point to students being able to identify comparable coursework and experiences from several educational providers and, in the process, even meet the graduation requirements of yet another. Smart schools will get ahead of this and figure out just what the inevitable institutional inter-connectedness will mean for them." (Ohler 2008)

8. REFERENCES

- Dawson, F., Howes, T. (1998). The Internet Society. RFC 2426. *vCard MIME Directory Profile*. Retrieved May 24, 2009, from: <http://www.ietf.org/rfc/rfc2426.txt>
- Eicker, S., Schuler, J.P.M., Schuler, P.M. (2008). *Management and Distribution of Course Catalogues with BolognaT3*. EUNIS 2008 - Visions for IT in Higher Education. Århus. Denmark. Retrieved May 24, 2009, from: <http://eunis.dk/papers/p116.pdf>
- Friedman, V. (2009). *Praxisbuch Web 2.0*. 2nd Edition. Bonn: Galileo Press.
- Microformats.org (n.d.). *About microformats*. Retrieved May 24, 2009, from: <http://microformats.org/about/>
- Microformats.org (n.d.). *Introduction to microformats*. Retrieved May 24, 2009, from: <http://microformats.org/wiki/introduction>
- Microformats.org (n.d.). *Process*. Retrieved May 24, 2009, from: <http://microformats.org/wiki/process>
- MLO (n.d.). *Metadata for Learning Opportunities (MLO) - Advertising*. Retrieved May 24, 2009, from: <http://mjukis.blogg.skolverket.se/files/2008/10/mlo-ad-v5.pdf>
- Ohler, J. (2008). The semantic Web in Education: What happens when the read-write web gets smart enough to help us organize and evaluate the information it provides? *Educause Quarterly*, 4 2008, pp. 7-9. Retrieved May 24, 2009, from: <http://net.educause.edu/ir/library/pdf/EQM0840.pdf>