IMT595 – Efthimiadis/Barzilai-Nahon
Peter Ellis
June 9, 2009
Capstone Project Summary

Project Title: Environmental Sustainability Metadata at the University of Washington
Project Members: Peter Ellis

Project Summary

The purpose of this project was originally to create a metadata schema to encapsulate metadata about the environmental sustainability efforts of the University of Washington based on my experiences with the Green Computing Special Interest Group on campus, and also inspired by a previous capstone proposal that would have coordinated with the College of the Environment to capture student-created projects focusing on sustainability initiatives across campus. The idea was simple: create a structure to capture the data spread across campus and attempt to give it a shared format that would allow a single source for all of the University of Washington's assessment work in the area. The scope of the project was defined as the definition and assessment of a workable XML schema for use by the University of Washington. Primary stakeholders included faculty, staff, and researchers at the University of Washington, followed by more general interest stakeholders such as students and the general public.

The work involved four phases: a needs assessment and research phase, a card sort, creation of the metadata schema itself, and a usability testing round of that schema.

The involvement of the Office of Environmental Stewardship and Sustainability came about as a happy accident as I began the project – referred by not one but two of the respondents to my initial needs assessment survey, I contacted them and had a sit-down meeting to describe the project. They agreed to be the steward for the schema that would be generated, and quite a bit of the project was driven by the pointers and questions that they had.

Overall, the schema's aim was to provide a first draft attempt at trying to provide a structure for a wide range of data that the University of Washington must process in order to assess its environmental performance and to centralize the management and stewardship of that data.

Lessons Learned

This capstone project let me really drill into the process of trying to understand what it means to maintain data from the standpoint of creating new methods of maintaining legacy data. Not only is it imperative to recognize the range of data that might be present in a specific environment, but it's also crucial to ensure that the structure that you create reflects the actual needs of the users rather than simply being fit to the stacks of data that are already present.

In this regard, I think the project did extraordinarily well in using more user-centered methodologies as a series of checks and balances to ensure that the best possible implementation was used. I was inherently limited by the number of participants in the project overall, which never topped 15 (which, given the University's size, is a very, *very* small subset). This reflects real-world projects fairly well – it is never easy to find the contacts you need in every project,

and it is never easy to engage those contacts that you do find. I might have been able to get more of a representative sample if I had pushed to get the correct contacts in place earlier and if I had engaged someone from the Office of Environmental Stewardship and Sustainability as the primary project sponsor. However, my selection of Terry Dievendorf, one of the co-leaders of the Green Computing Special Interest Group on campus and an applications developer for the Foster School of Business, worked out well in the long run – his technical expertise was invaluable in our discussions of the options that we had as we proceeded through the project.

It was pointed out to me a couple of times over the course of the final project presentations and in other presentations that I've given on this work that I could have benefitted from a thorough review of sustainability assessment methodologies; while the point is well taken, and I am at least minimally familiar with several of the methodologies out there by virtue of my past work, I argue that any assessment methodology must be applied to the data itself, and so therefore is out of scope. While it's true that the methodology and the data cannot be separated, the methodology really only cares about the *data* and less about the structure that the data is in. So long as the data structure is robust enough to describe the data appropriately, the methodology is irrelevant.

If anything, the failure point in the entire project was the time frame that I had to work with. Eight weeks is simply not enough time for such a large-scale undertaking; I am quite grateful that the scope of my project was well-defined, for it's quite clear to me that this is something that someone could very well end up spending decades working on before reaching anything resembling a suitable solution for 90% of the use cases. It was due to time constraints that I was unable to do a really thorough process of looking at exactly what kinds of data exist across campus. I was given hints and indicators by the various stakeholders that I had meetings with and by my own independent research, but remain convinced that I barely even scratched the surface. However, the ability to use the needs assessment tools that I garnered throughout the program and the card sort methodologies that I was introduced to via classes like IMT540 and the Information Architecture Institute was extremely useful.

One of the decisions that I was forced to make early on was whether or not this metadata effort would be implemented in XML or in a database format. My justification for choosing XML has remained the same throughout the entire project: it's fast to implement and it allows for a much greater flexibility in implementing changes, and, if the XML files are structured properly or if a source management system is sued, is much more amenable to version control. However, XML could potentially break down under the rather substantial amount of data that the University of Washington could end up loading into the schema. Databases are far more scalable, but take a lot more care to implement. However, the data modeling concepts are essentially the same for both. My choice of XML may have resulted in generating further work down the road if the Office of Environmental Stewardship and Sustainability decides to continue work on this project.

Another potential mistake I made was to not work harder in securing the project's future. While recommendations and the deliverables for the project were sent to the Office of Environmental Stewardship and Sustainability, no real followup plan was put in place to secure the likelihood that the Office could continue the work without significant guidance from an information management professional. While next steps were provided with the documentation and

somewhat fleshed out, the Office lacks any real manpower to be able to follow through with those (intentionally) vague next steps.

I point at two sources for this issue: first is a failure to establish a sturdy rapport with the Office staff. While I did have an initial meeting with them and I involved their staff in the card sort and usability testing exercises, I am not convinced that I laid a sound foundation to allow them to comfortably engage the Information School in further dialogs and collaboration on the topic. The second is the nature of the scope of the project: the scope only called for the creation of the schema and some brief testing to ensure that the model was done correctly. While this is not an excuse to hide behind, it (plus the fact that the Office of Environmental Stewardship and Sustainability was something I didn't know about at the time of planning) helps to explain why the project timeline did not incorporate any strong attempts at establishing further project stewardship.

Impact
Within the University of Washington, the impacts of this project are significant, even though they have not been widely communicated:

1. *This schema allows for centralization of disparate data sources primarily maintained by Facilities Services, but can also capture other data that may not currently be collected.* This is a huge benefit, simply because it allows for interested parties to go to one place in order to get all the information they need. It also positions the Office of Environmental Stewardship and Sustainability as a central data repository.
2. *The project lays the foundation for future collaboration on the subjects of environmental assessment and environmental sustainability.* Though there are other projects on campus that either have or are attempting to document the University of Washington's environmental sustainability efforts, this structure will hopefully serve as a catalyst to begin a conversation about what information should be tracked by what people. Since the Office of Environmental Stewardship and Sustainability owns the schema, it also establishes a management point of contact for that information.
3. *The research and the schema attempt to surface data and information that are currently not easily findable.* This speaks to one of the core ideas of the MSIM program – engaging information in such a way that it can be used and manipulated rather than ignored.

Personally, I was quite surprised at forgetting one of the crucial tools of metadata and information architecture: card sorts. I had not even considered using a card sort until a meeting with Mike Crandall, where he urged me to consider it as a tool that might help to flesh out existing preconceptions and ideas about how to properly structure the data. It reminded me that it is quite easy to get stuck into a rut when approaching a particular project – even though it may appear that you have considered all the options, there may be a tool in the MSIM toolbox that might challenge your own perceptions of the issues.

In general, the program prepared me quite well for this particular project. Though I did not use some of the concepts presented in classes like IMT530 – ontologies, preexisting metadata schemas, faceted classification *a la* Ranganathan's colon classification – they certainly could

have been applied to such a project. The project, though, failed at really stretching my boundaries; that is not to imply a lack of a challenge. Rather, I feel as if it did not teach me anything particularly new about information management through the process of doing the work; it reminded me of some things that I had forgotten over the last two years, but there was nothing fresh. It is not imperative that *every* project teach *something* new about information management. In this particular case, it was enough to be engaged in something that bridged two of my interests: environmental sustainability and information management. However, considering the relative freshness of the material (since I have not been applying the lessons learned for as long as a seasoned professional), I had hoped for more.