IMAGINING EMERGENT METADATA, REALIZING THE EMERGENT WEB

Jason Bengtson, MLIS, AHIP

University of New Mexico

Health Sciences Library and Informatics Center

Jbengtson@salud.unm.edu

KEYWORDS: informatics, information science, metadata, information technology, computer science, complexity science, emergence, emergent behavior, search and discovery, semantic web, web 3.0, artificial intelligence, self organization

Author Bio: Jason Bengtson is the Emerging Technologies/R&D Librarian for the University of New Mexico's Health Sciences Library and Informatics Center in Albuquerque, NM. He is a graduate of the University of Iowa School of Library and Information Science in Iowa City, Iowa.

Abstract: While metadata is a key ingredient of machine-semantic technologies, it also suffers from drawbacks. As it is currently formed, metadata lacks dynamic responsiveness and requires top down system modeling. The author proposes a schema and process of emergent metadata which will, if successful, allow metadata to respond to environmental conditions dynamically and exhibit self-organizational features.

INTRODUCTION

Fundamentally this paper posits one question: can we use concepts of self-organization and emergence from the Complexity Sciences to improve the utility of metadata? The model still used for a great deal of information handling, that of people directly controlling a limited amount of information, is thoroughly obsolete. There is simply too much information in proportion to those charged with organizing and maintaining it. The Complexity Sciences offer us tools to theoretically allow our information to do some of this work for us. In fact, these tools are so powerful that we may potentially see systems emerge from our information that could teach us new ways to manage it.

Modern information systems rely on the use of records as “digital proxies” for an actual discrete information unit. These records can be indexed in many ways in order to facilitate search. The records themselves consist of various kinds of “metadata”, or “information about information”. Metadata is a staple of information science. It is the tool by which information professionals bring order to the chaos of our channels of information so that information can be used meaningfully and found at will. But modern metadata has many limitations. Like all products generated by human beings, metadata is subject to a certain level of human error. Tags that could be useful in searching a topic can easily be left out, making some relevant documents hard to locate in a literature search. The very act of assigning tags is fraught with problems. Most things can be classified in multiple ways. How does a cataloger capture the essence of a thing correctly? Metadata tags also suffer from being “dead information”. That is to say, once they have been created and assigned they generally sit, unchanged, until someone explicitly edits them. This makes metadata a relatively unresponsive medium for use in our highly dynamic information environments.

Metadata needs to become smarter. As information proliferates, it is not enough to have static metadata. We need metadata that responds dynamically to changes in searching trends to make itself more useful. As we continually work toward a more semantic web, we need a more semantic metadata. We need a new dimension that we can add to this formerly flat format. I believe that this new dimension we are searching for is emergence.

WHAT IS EMERGENCE?

Emergent behavior is observed in many sciences. In emergence, individual components, by following a few simple rules, generate a new level of complexity within a system which cannot be predicted based on a reduction of that system to the components themselves (Munoz & de Castro, 2009, p.277; Claus, Køppe, & Stjernfelt, 1997, p. 83). Such systems often display effects which give the illusion of resulting from some external direction, or from the dictates of any overarching intelligence. In this way, relatively simple independent agents can exhibit behaviors that seem to belie their limited capabilities. The new complexity generated from an emergent system also exerts a backward influence (bidirectionality) over the agents that make up the system (Munoz & de Castro, 2009, p.277). Because emergent systems are not strictly reducible, emergent outcomes in a system can only be predicted in a rough fashion through the act of simulating that system (Bedau, 2008, p. 453-454). In addition, emergent systems are robust enough that failures of specific, individual components of the system should not cause the entire system to fail (Munoz & de Castro, 2009, p.277).

Self-organizing behavior, by contrast, results in an increase in order, must be robust, autonomous and dynamic, but, unlike emergence which "cannot be reduced to the behaviors and properties of the component parts of the system" (Munoz & de Castro, 2009, p.276), self-organization is reducible (Munoz & de Castro, 2009, p.276). The methodology proposed by this paper will attempt to make use of both self-organization (which we will see in 'level 2 tags') and emergence (evidenced in 'level 3 tags'), but the emphasis will be on attempts to harness the more sophisticated, emergent behavior.

Emergent systems generate increasing complexity that often defies conventional wisdom about the necessity of “top down” organizational and informational models. It is perhaps instructive to try to conceptualize the difference by describing this principle in terms of competing economic models. Centrally planned economies can be thought of as top down in organization, with production and supply being dictated by a central planning source. Free market economies employ a “bottom-up” approach that can be thought of as displaying emergent levels of complexity based on the irreducible (hence the difficulty in making economic predictions) interactions of many independent agents without having a central planning authority dictate how the economy must operate. The complexity that emerges (in the form of the job market or other aspects of the economy) exerts a backwards control over the individual agents whose interactions are responsible for the behavior. This analogy isn’t perfect of course, just as neither organizational model perfectly meets the needs of the agents that comprise it. What may be required is a blend of both types of organizational model for systems to operate to a maximal level of effectiveness. This will probably prove to be true in information systems as well, as traditional ontologies and organizational schemas are combined with emergent processes to produce an organizational sophistication that neither could achieve alone.

Information Science has, for most of its history, been largely concerned with top down modes of information organization. As full text keyword searching has proliferated, however, so have tools like social tags, and data mining, introducing a bottom-up element to the information recovery tools used by the sometimes skeptical Information Sciences. Additionally, some interesting work has been done at Johns Hopkins University in using emergent agents to distribute metadata actively to clients seeking information about resources (Cost et al, 2007, p.1). As the need for reliable information cataloging and recovery increases, “traditional” tools like metadata must continue to evolve using such methodologies. Despite the growing sophistication of digital technologies, too many of our information cataloging technologies remain relatively primitive.

WHAT EMERGENCE COULD MEAN TO METADATA

By applying the concept of emergence to metadata, it may be possible to develop a more useful and dynamic metadata methodology than any that has been used in the past. Self-organizing metadata systems could be created that, through a process of emergence, or through a combination of emergent principles and top down ontological organization of data relationships (acting in some sense as our basic relational rule set from which more complex relationships can be developed) could potentially revolutionize searching and data organization.

Such “emergent metadata” would be flexible, adapting to changes in semantic relationships and ontologies. As such a technology developed, rather than relying only on a particular, mandated ontology, this emergent metadata could employ applications that would seek out or even construct new ontologies based upon information found on the web. Ontologies and taxonomies could also be implicitly created through the means used to create and evaluate the metadata created through such processes. The idea is to create a system capable of novel connections and complexity building that is not reliant upon human direction. Such emergent systems could be capable of spontaneously building information nodes and frameworks in ways that human beings might never conceive of, but which could prove to be the seeds of future research in Information Science and Technology.

In order to inject metadata systems with the potential to exhibit emergent properties, however, a basic barrier to the process must be confronted; the current “dead” format of metadata. Data files do not interact with each other. They are simply data. In order to apply emergent principles, we must find ways to make these files interact. Two basic approaches to this problem are considered by this author: the use of browsing tools to facilitate "cross-pollination" between metadata units while keeping the data itself in file form, and the recreation of metadata tags as applications rather than data files. Both approaches have their strengths and drawbacks.

CROSS-POLLINATION: THE SIMPLE APPROACH

If metadata is kept in data file format (or embedded as data within another file), “comparison tools” working in concert with browsers and search tools can be used to allow the metadata in one record to "cross-pollinate" other metadata records and interact with data from other sources. In order to control the veracity of the metadata, it would be important to remove it from the control of local authors. One of the reasons that metadata sees little usage by current web search engines is the potential for abuse in order to manipulate search rankings, as was commonplace with text in websites on the web before more sophisticated search algorithms replaced simple keyword searches (Official Google Webmaster Central Blog, 2012). I proposed, in an earlier paper, that web metadata production be automated, with the generated metadata stored in off-site indexes (maintained by commercial search providers or, preferably, non-profit or governmental organizations) (Bengtson, 2010). Even if that proposed solution for generating semantic metadata on the web is deemed impractical, and web authors simply generated metadata using html meta or xml tags, it would be still be advisable for the indexes created by emergent activity to be kept off-site or in another type of format inaccessible to web authors. Fundamentally, these indexes should be machine writable only. This process would be even easier to employ within a database environment, where indexes are not normally user-accessible.

To create an emergent effect, browsers and search tools could be used in coordination with server-side “comparison tools”, functioning much like bees, carrying data from one record to the next to allow those pages to be “cross-pollinated”. The comparison tools would enable interactions between the metadata units, probably guided by ontological and taxonomic references of some sort as will be elucidated in this paper’s practical example (One Basic Model for the Web). It would be advisable to establish “levels” of metadata as outlined in appendix A. First level metadata would be automatically generated or created by a web author, and would be potentially human-editable. Other levels of metadata would be designed to be machine-writable only. As long as clear levels of metadata complexity are established, leaving the original, “first level” metadata unmodified, such a method would do nothing to degrade the original metadata assigned to a record.

There are, of course, associated problems with this methodology; the interoperability of different metadata schemas and ontologies being the primary one. Interoperability of metadata generally is a serious problem within our discipline, and as such it must be confronted as we discuss metadata interaction (Park & Tosaka, 2010; Park & Childress, 2009). A great deal of work has already been done to facilitate interoperability (Haslhofer & Klas, 2010; Lee & Jacob, 2011; Roel, 2005), and as will be elucidated further in this paper, the method outlined here does provide an intrinsic mechanism for effectively siloing incompatible metadata ontologies.

Employing an emergence strategy may also cause users to see an increase in low-relevance returns that outweigh any improvement in the inclusiveness of the returns. The cross-pollination approach would also require the addition of fields to current database records if this approach is employed at the reference database level. In addition, if the cross-pollination strategy is employed, this strategy will, by necessity, probably be very dependent on user navigation.

CONTROLLING METADATA PROLIFERATION

One obvious concern inherent to this system is the problem of metadata proliferation. With continual interactions between records acting as independent agents, the potential for an uncontrolled proliferation of ultimately useless metadata tags is strong. As such, a system such as this one would need to have an “evaluation cycle” built in to the interactions that limited the retained metadata tags based upon set criteria. They are many possible ways to approach this problem, but I propose two types of evaluative filter that I believe would be of particular utility: Survival of the Fittest and Strength of Weak Ties.

Survival of the Fittest: This filter is relatively self-explanatory, speaking to the heart of an evolutionary approach to information networks. Simply put, the Survival of the Fittest (SOTF) filter would retain a set number of tags with the highest frequency of emergent interactions. For more information on how this might be carried out, consult appendix A.

Strength of Weak Ties[1]: This filter requires a bit more explanation. Essentially, it operates in reverse of SOTF, retaining a set number of tags based on the rarity with which they had been generated by comparison processes. This filter is important to counteract the problem so often seen with web based adaptive algorithms, in which

[1] From the "Strength of Weak Ties" theory of Mark Granovetter (Granovetter, 1973).

users are exposed to what they are accustomed to seeing. One potential strength of emergent metadata processes is the possibility of generating novel and useful connections (as is often seen in research serendipity). By retaining Strength of Weak Ties (SOWT) tags, emergent metadata processes could potentially be far more useful than if they relied solely on commonly generated information pathways.

In addition to filters, compatibility and meaningfulness of interaction must be considered. Metadata units undergoing an interaction should have a certain established threshold of compatibility to ensure that cross-pollination should occur. This is fundamental, not only to our deployment of this metadata schema, but to the concept of emergence itself. An essential feature of emergent systems is "coherence", in which individual agents of a system are compatible, allowing the overall system to maintain a sense of identity (Munoz & de Castro, 2009, p.277). The threshold used for testing compatibility will no doubt vary by system and subject, and will need to be a flexible setting. Compatibility can be established by comparing number and position of shared metadata tags. I propose, as outlined in Appendix A, a cross-pollination method employing three cycles. The first, the handshake cycle, would employ the comparison of tags.

The handshake cycle would also evaluate time on resource by the user. If a user only remains on a particular resource with their browser for a few seconds before navigating away, the interaction is probably not meaningful enough to instigate cross-pollination. Instead, a threshold of time (again, probably a flexible one) should be used to evaluate whether the navigational interaction has been meaningful enough to initiate the cross pollination of metadata tags.

METADATA AS SOFTWARE: FROM DATABASE TO BESTIARY

An alternative to the use of cross-pollination would be the recreation of metadata as an application that could interact directly with other metadata applications. The original metadata would exist within such an application in a read-only state, with new levels of metadata generated through record interaction added as new, writable data within that application. Records could then literally interact with each other in a guided or unguided fashion according to simple rules in order to generate new levels of complexity. This direct interaction approach is far more radical than the use of cross-pollination and, while it opens the door to intriguing possibilities, it is also much more difficult to carry off at our current level of information processing technology. Take an academic database as an example. These are, literally, databases. The only active code we normally see associated with the database is the search application. In order to allow these records to interact we would have to either link the database entries to proxy metadata applications through a digital identifier of some kind, or have the entire digital record function as an application. The metadata applications, in order to interact, would have to exist outside of an established database structure on a kind of “bestiary” server, where they could be allowed to actively run in each other’s presence. This would require a great deal of storage and processing power in addition to that already being used for the database itself.

It would be somewhat easier to try this approach on the web. While web pages themselves could not actively interact, webcrawling applications could function as proxies, linked to their associated metadata record via a DOI or similar instrument, regularly writing changes in the metadata back into their associated index or parent file.

Given the enormous processing power and bandwidth needed, however, in the short term, a cross-pollination approach is almost certainly more realistic.

INTEROPERABILITY

A central issue to both the future of the semantic web and the future of metadata is interoperability (Magee, 2010). Metadata currently exists in a dizzying array of schemas and proprietary formats. It is worth asking how applying emergence to metadata would impact its interoperability, especially on a larger scale.

As the initial conception for carrying out this approach is described later in this paper, the reader should take note of the fact that, potentially, *any metadata ontology could be employed at the base level*. The "level 2" and "level 3" tags that I describe could, and should, function on a read/write layer separate from the read only access apportioned to the "level 1" (original metadata) tags. The "level 2" tags are directly derived from the "level 1" tags, whereas the "level 3" tags could be "tacked on" to virtually any base level metadata variety. This potentially makes such an approach borderline to fully "ontology agnostic", at least at this stage of conception.

Another consideration is the implicit siloing that would take place based upon the setting used to determine compatibility for cross-pollination between metadata units. By requiring a certain threshold of compatibility between metadata units before cross-pollination can occur, the browsing mechanism could avoid generating low quality second and third level tags. Even in the worst case scenario of this approach, in which a large number of low quality second and third level tags were continually produced, the first level metadata would not be affected.

As a result of these factors, this approach should not interfere with interoperability; in fact, many metadata units in a variety of schemas should be able to function side by side in the same network without an improper cross-pollination issue, in much the same way that multiple ontologies can be simultaneously employed in formats like RDF.

ONE BASIC MODEL FOR THE WEB

The following example describes one theoretical way that this approach could be put into practice on the World Wide Web. This example relies on relational and definitional logic spelled out in more detail in appendix A. For practical reasons, the metadata discussed in this example, as well as that described in appendix A, is “subject term” metadata (the metadata type which probably lends itself most readily to such a process). MeSH terms will be used as the subject taxonomy. That is to say that all of the "tags" that will be employed in this example are tags from the MeSH field of a hypothetical metadata record. This model utilizes a two part system, with subject term metadata (our MeSH headings) located offsite in a descriptive index referencing the website. In this theoretical, emergent metadata schema, all tags have the “level” attribute, while all tags with a “level” attribute of 2 or 3 also possess the “counter” attribute.

As the browser moves to a site and renders it, the browser, referencing the metadata stored offsite, will then query a backend comparison tool (software on the servers of the search provider) that accesses the NLM's MeSH tables to reference the relationships of terms. This tool can make meaningful comparisons between the tags found for each page in the provider's indexes.

The tool's algorithms can be tuned to preference, but in this example the comparison tool is going to generate self-organizing "second layer" metadata through two comparisons that essentially evaluate *metaphor* and *metonymy* in the extant metadata terms. First, the tool will seek out synonyms to the sites' metadata subject tags from established references (either through a medical dictionary or through a thesaurus table generated by a body such as NLM), regardless of their position within the tag hierarchy of either site (or the taxonomy in use), fulfilling a metaphor comparison. Then it will evaluate metonymy by examining the MeSH tables to discover if any of the terms tagged on the sites occupy a similar level underneath the same general term for both sites (or, to use a Computer Science description, to see if the terms form another instance of the same class). Such terms, if they are found to be compatible, could be exchanged between the index entries for the sites. While these tags would be identical in most respects to the original, "level one" metadata tags, there would be one significant difference. The "level" attribute of the tag would be set to "2", indicating that, unlike the original metadata, this metadata was to be fully editable by comparison tools. The "counter" attribute of the tag would record the number of times the tag had been generated in similar interactions.

Next, our comparison tool would generate emergent “third level” data by comparing the first and second layer MeSH terms and, based on their number and relationships, add more general MeSH terms that multiple first and second level terms fall under, or add narrower terms or subheadings that it finds underneath the first and second level terms. Our algorithms can also search semantically for additional terms that relate to those on the first two levels. These new third level terms that are generated would be written into the metadata index as well. In addition, our comparison tool could create new, emergent tags (a few prospective types are described in Appendix A) designed to be solely machine semantic, recording information about navigation, references, and other data. Depending on how these tags are used in indexes or by browsing tools, they could form the basis for novel information pathways. As with our level two tags, these tags will retain a “counter” value.

When the user navigates away from the page, the browser will transmit the urls of the former and current site to the comparison tool, continuing the cycle.

Now that the basic model has been described, it can be stepped through with a practical example. For purposes of this example a search for *neoplasms* AND *hippocampus* will be assumed. The first site visited has the tags in question, as well as tags for *hypothalamus* and *amygdala*. The site also has a tag for *hospital administration*.

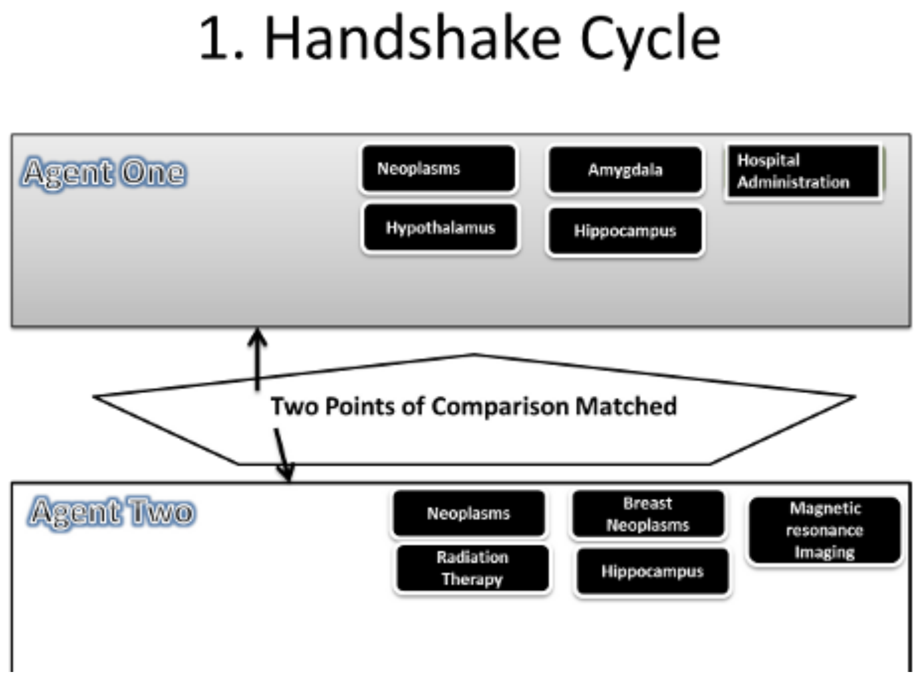

figure1: Handshake Cycle

The browser moves on to a second site. After a certain threshold of *time on agent* (thirty seconds, perhaps) is met, the urls of the two sites are transmitted to the comparison tool, which enters the *Handshake Cycle* of the interaction. It compares the tags of both sites from their metadata indexes and determines that the shared tags within the MeSH framework that it detects constitute sufficient compatibility to initiate cross-pollination. After this determination is made, the comparison tool enters the *Interaction Cycle*.

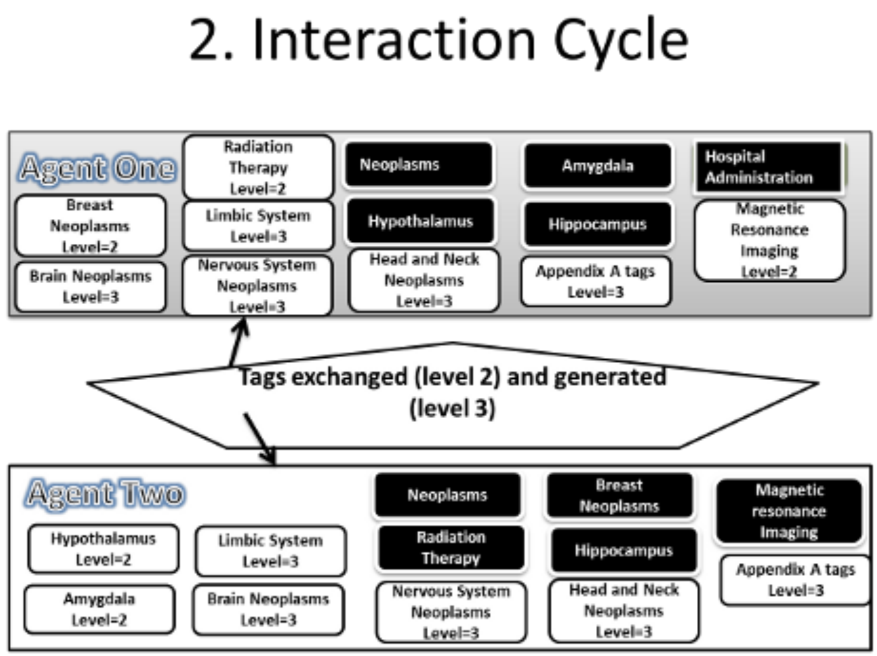

figure2: Interaction Cycle

The second site, beyond the tags being explicitly searched for, also possesses subject tags for *breast neoplasms*, *radiation therapy*, and *magnetic resonance imaging*. The comparison tool examines the tags stored in the metadata index. It finds no relationship to justify retaining *hospital administration* from the first site as a second level tag in the new site, so this MeSH term is not pollinated. *Hypothalamus* and *amygdala*, however, are found to be narrower terms beneath *limbic system* next to *hippocampus* in the MeSH hierarchy, so those two terms are added as tags to the index of the second site with a level value of 2. A next stage of analysis by our comparison tool causes it to pull in *limbic system* as a term, since we now have three of its narrower terms in the index entry. Further semantic analysis and comparison to the MeSH tables shows that *limbic system* falls beneath *brain*. The combination of *neoplasms* and *brain* has a rough semantic equivalence with the MeSH terms *brain neoplasms*, *head and* *neck neoplasms*, and *nervous system neoplasms*, so those terms are generated as subject tags with a level value of 3. Additional level 3 tags are generated based on navigational and other data, to be added to the indexes of both sites. Thanks to this process of cross pollination and the resultant emergent creation of higher orders of complexity within the system, additional meaningful tags have been applied to both sites, making them potentially easier to find.

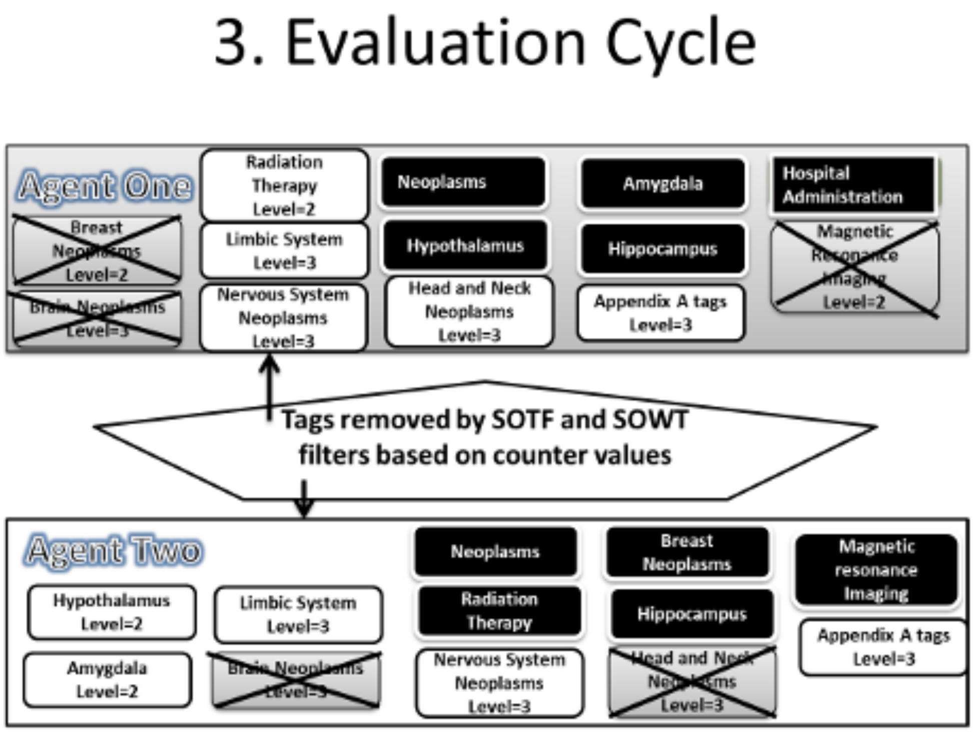

figure3: Evaluation Cycle

In the last cycle of the exchange, the *Evaluation Cycle*, the level 2 and 3 tags of both sites are evaluated for retention by the *Survival of the Fittest* and *Strength of Weak Ties* filters. Tags that fall below the SOTF threshold setting but above the SOWT threshold setting are discarded. Note that only the level 2 and 3 tags are eligible for deletion; no level 1 tags may be altered.

CONCLUSION

Part of the potential in this proposal is the idea that everything from structured metadata tags to user assigned tag clouds may interact in unpredictable but potentially useful ways in much the same fashion as is seen in nature if the capability for emergent interaction is added to them. Such cross-platform metadata interactions would require a more complex method of semantic analysis than that described here, but would build on the framework described in this paper. Rather than simply accessing MeSH tables, comparison tools could access large, server-based applications with sophisticated abilities to examine a broad range of ontologies and taxonomies. This could potentially make emergent metadata schemas a powerful feature for discovery tools to employ. However, while a discovery context is used as the primary example in this paper, it is important to note that emergent metadata potentially has implications that go far beyond taxonomies. These adaptive systems could be employed to actively seek out cross-bridging solutions, or form novel connections based on any number of criteria, improving the interoperability of networked resources.

In a sense, this approach is a form of "crowdsourcing", using navigational data crowdsourced from information consumers, and metadata crowdsourced from the information itself. As with all crowdsourced systems, there would be very little need or opportunity for oversight of tag validity. This reduces the overhead of such a system, and increases the likelihood of novel connections between information sources, but it does generate the specter of invalid tagging. Of course, such tagging occurs even in systems with significant oversight. It will be the task of further experimentation and usage in the wild to help determine the optimal settings and qualities for such a methodology to minimize the percentage of inaccurate tags and maximize performance.

From a Cost/Benefit ratio perspective, this approach is particularly promising. While storage and bandwidth needs for networks would increase (an inevitable fact, anyway) this approach would allow the production of level 2 and 3 metadata tags to occur through an automated process that requires little to function beyond the initial work of designing the system and a minimal level of oversight to tweak variable settings (such as those for compatibility and the evaluation cycle). In a worst case scenario, if problems arise with the emergent levels of metadata, search and other interoperability tools could simply be instructed to ignore the second and third level tags associated with an information source until the issues with the tags were resolved.

This paper was originally presented in an extremely rough form at the Macalester College Technology in Libraries Conference in March of 2012. The audience to the presentation was small, and the response to this idea was mixed. There was some dubiousness by at least one attendee, while two other attendees approached me after the presentation and expressed a great deal of enthusiasm for the potential of this approach. Based solely on this unscientific piece of feedback, I expect the reaction from Librarianship as a whole to also be rather mixed. Harnessing the power of emergent phenomena has great potential, but it is a technique that will have to prove itself.

The next step in exploring this approach will be simulation. Emergent systems have outcomes that can only be determined through simulation, and I am currently exploring options for a simulation protocol to establish the utility (or lack thereof) which may potentially exist in the application of emergent metadata agents. It would be a mistake in any technology project to attempt to achieve ambitious outcomes without having first made provision to obtain the necessary resources (Bengtson, 2011). Investigation of the utility of emergent metadata will require time and funding. I am currently in the first stages of applying for a grant to expedite the process.

If given a simple set of rules to follow and left to create their own complexity, emergent metadata agents may yield ways of referencing and looking at data that currently haven't even been conceived of. As such, these systems could function not only as a practical way to improve access and cross-purposing for data, but as a laboratory to discover new ways to process and mark-up that data. Metadata could be "released into the wild", to develop in novel ways that could potentially lead to watershed moments in human understanding of information.

References

Bedau, M. A. (2008). Is weak emergence just in the mind? *Minds & Machines, 18*(4), 443-459.

Bengtson, J. (2011). How tech projects fail: Learning lessons for the future. *Journal of Hospital Librarianship, 11*(2), 188-194

Bengtson, J. (2010). Why I can't love the homemade semantic web. *B Sides,* http://ir.uiowa.edu/bsides/20

Cost, R. S., McCubbin, C., Cole, J., Dale, M., Scheidt, D., & Mayfield, J. (2007). Active metadata framework and emergent metadata distribution with scout agents. *2nd International Workshop on Engineering Emergence in Decentralised Autonomic Systems*

Emmeche, C., Køppe, S., & Stjernfelt, F. (1997). Explaining emergence: Towards an ontology of levels. *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie, 28*(1), pp. 83-119.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380.

Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys, 42*(2), 7-7.37. doi:10.1145/1667062.1667064

Lee, S., & Jacob, E. K. (2011). An integrated approach to metadata interoperability construction of a conceptual structure between MARC and FRBR. Library Resources & Technical Services, 55(1), 17-32.

Magee, L. (2010). A framework for assessing commensurability of semantic web ontologies. *Electronic Journal of Knowledge Management, 8*(1), 91-101.

Munoz, Y. J., & de Castro, L. N. (2009). Self-organisation and emergence in artificial life: Concepts and illustrations. *Journal of Experimental & Theoretical Artificial Intelligence, 21*(4), 273-292. doi:10.1080/09528130902823649

Official Google Webmaster Central Blog: Google does not use the keywords meta tag in web ranking. (n.d.). Retrieved July 14, 2012, from http://googlewebmastercentral.blogspot.com/2009/09/google-does-not-use-keywords-meta-tag.html

Park, J., & Childress, E. (2009). Dublin core metadata semantics: An analysis of the perspectives of information professionals. *Journal of Information Science, 35*(6), 727-739.

Park, J. & Tosaka, Y. (2010). Metadata creation practices in digital repositories and collections: Schemata, selection criteria, and interoperability. *Information Technology & Libraries, 29*(3), 104-116.

Roel, E. (2005). The MOSC project: Using the OAI-PMH to bridge metadata cultural differences across museums, archives, and libraries. *Information Technology and Libraries, 24*(1), 22-24.

APPENDIX A:

BASIC RULES OF INTERACTION AND LOGICAL STRUCTURE PROPOSED BY THE AUTHOR FOR INITIAL IMPLEMENTATION/TESTING

Levels

Level 1 (base level): Metadata originally assigned to the agent. Cannot be comparison tool edited in any iteration cycle. Human semantic, may be machine semantic.

Level 2 (navigation level): Metadata generated through iterative process based solely on user navigation between agents. All tags with the condition Level [2] also possess the condition Counter [value]. Level 2 tags are not user editable and might exist only in a remote index. Human semantic, may be machine semantic.

Level 3 (emergent level): Metadata generated through iterative process based on semantic and navigational evaluation. Final product designed to be machine semantic. All tags with the condition Level [3] also possess the condition Counter [value]. Level 3 tags are not user editable and might exist only in a remote index.

General Definitions

Browsing Mechanism: The tool used to move from one Agent to another. May be a search tool, a web browser, a file browser, or some combination thereof.

Comparison Tool: The tool used to compare metadata from two Agents and expedite cross-pollination. May reference other resources. Should run separately from the Browsing Mechanism to reduce bandwidth load and Browsing Mechanism sophistication.

Item: Discrete object, digital or otherwise

Agent: Indexed digital proxy (i.e. record) representing a particular Item. Item may be integrated into the Agent as a field, but usually will not be.

Field: The largest sub-division of an Agent.

Part: The largest subdivison of a Field. Represents a complete semantic branch of tags. Represents >= 1 tag(s).

Tag: A single, semantic, metadata term.

Level: A condition of any given tag reflecting its functionality and accessibility. All tags have the Level condition. Tags in the same Part may have distinctly different Levels.

Counter: A condition of any tag possessing a Level condition consisting of values 2 or 3 which reflects the number of iteration cycles which have assigned the tag to the Agent. Tags in the same Part may have distinctly different Counters.

Package: A condition of any tag possessing a Level condition consisting of value 3 which reflects the variety of information that it carries (see Package variety list).

Machine semantic: Designed to be understood by mechanistic rather than human agency.

Handshake Cycle: In the cross-pollination interaction model this represents the initial movement of a browser from one Agent to another, at which point both Agents are evaluated for cross-pollination compatibility based on sufficient numbers of shared Tags in the appropriate Fields.

Interaction Cycle: In the cross-pollination interaction model this represents the act of exchanging and generating Level 2 and 3 tags in both Agents based on a positive evaluation of compatibility in the Handshake cycle.

Evaluation Cycle: In the cross-pollination interaction model this represents the initial movement of a browser from one Agent to another, at which point the Tags contained by an Agent possessing Levels 2 and 3 are evaluated for retention in their respective indexes based on the value of the Counter condition.

Survival of the Fittest (SOTF): Iteration type that forms part of the evaluation cycle. Retains the tags in each respective level with the highest Counter values (above a user-established threshold). In the case of tying values, it can apply a random number generator so that only the user specified number of tags are retained.

Strength of Weak Ties (SOWT): Iteration type that forms part of the evaluation cycle. Retains the tags in each respective level with the lowest Counter values (below a user-established threshold). In the case of tying values, it can apply a random number generator so that only the user specified number of tags are retained. After this filter is run, the tags not marked for retention on Levels 2 and 3 are deleted.

Prospective Package Variety List

Taxonomy: Synonyms for tags from level 1 and 2 drawn from external taxonomy libraries specified by the Comparison Tool.

Folksonomy: Synonyms for tags from level 1 and 2 drawn from external folksonomy libraries specified by the Comparison Tool.

CrossLink: Links to other Agents that have been the subject of a successful exchange.

References: References from the Agent metadata and from any Agents that have been the subject of a successful exchange.

Search: The search terms present in the browsing mechanism at the time of a successful exchange.

Identity: Presents as string. Three separate tags generated. Top three Taxonomy tags+top three Folksonomy tags as determined by Counter values and random number if too many Counter values are equal.

Path: Presents as string. Three separate tags generated. Top three Reference tags+top three Crosslink tags as determined by Counter values and random number if too many Counter values are equal.

Route: Presents as string. Three separate tags generated. Top three Search tags+top three Identity tags as determined by Counter values and random number if too many Counter values are equal.

Rules of Handshake

- Fields to be evaluated should be a setting in the Comparison Tool; subject field or similar is most appropriate.
- Handshake=true if 1 of the tags of the current Agent matches 1 of the tags of the other Agent AND if the total number of tags of Agent with lowest number of tags =1 AND (time on Agent for both Agents>one minute OR user clicks through to Item from Agent).
- Handshake=true if 1/2 of the tags of the current Agent match 1/2 of the tags of the other Agent AND if the total number of tags of Agent with lowest number of tags (>1 AND <=10) AND (time on Agent for both Agents>one minute OR user clicks through to Item from Agent).
- Handshake=true if 1/3 of the tags of the current Agent match 1/3 of the tags of the other Agent AND if the total number of tags of Agent with lowest number of tags>10 AND (time on Agent for both Agents>one minute OR user clicks through to Item from Agent).

Rules of Exchange

<u>Level 2 and 3 general</u>

If Handshake=true AND tagfromotherAgent=false then add tag

If Handshake=false AND tagfromotherAgent=false then do not add tag

If Handshake=false AND tagfromotherAgent=true then do not add tag

If Handshake=true AND tagfromotherAgent=true then do not add tag

<u>If Package=Path</u>

iterates 3 times, reductive iterations (If Handshake=true then References(where Counter=highest)+Crosslink(where Counter=highest))

<u>If Package=Identity</u>

iterates 3 times, reductive iterations (If Handshake=true then Taxonomy(where Counter=highest)+Folksonomy(where Counter=highest))

<u>If Package=Route</u>

iterates 3 times, reductive iterations (If Handshake=true then Identity(where Counter=highest)+Search(where order=random))