



**11<sup>th</sup> Annual Data Harmony Users Group (DHUG) Meeting 2015**  
**Beyond Subject Metadata; or, So you have a Taxonomy! ... now what?**  
**Monday, February 16, 2014 - Training**

**8:00 AM – 9:00 AM                    REGISTRATION AND BREAKFAST**

**9:00 AM – 5:00 PM                    Building Taxonomies: Fundamentals, Construction, and Application**  
*Access Innovations, Inc. staff*

To open the first day of training, we will present a thorough introduction to basic taxonomy concepts. This introduction will lay the groundwork for later sessions, establish basic terminology, and ensure we all understand our shared controlled vocabulary describing taxonomies and thesauri.

We will explain how to develop a taxonomy from scratch and how to modify an existing vocabulary. Join us as we demonstrate how to structure a controlled vocabulary based on a hierarchical sorting of terms while detailing the differences between each taxonomic classification. The session will cover top-down and bottom-up approaches to thesaurus construction. Additionally, we will cover best practices in thesaurus management in order to ensure fully functional and descriptive vocabularies.

We will also cover approaches to integrating your taxonomy into your production platform to maximize your investment.

We will take morning, lunch and afternoon breaks as appropriate.

**5:30 PM –7:00 PM                    NETWORKING RECEPTION**

Join us in the Lobo Room at the Hampton Inn. The Lobo Room is on the first floor, just past the elevators, on the right (south) side of the hall. Visit with the Data Harmony and Access Innovations staff and other Data Harmony software users at this an informal reception. Beer and wine, soft drinks, and light hors d’oeuvres will be available. Use this time to find a buddy with whom you can explore some of our wonderful local eateries!



**11<sup>th</sup> Annual Data Harmony Users Group (DHUG) Meeting 2015**  
**Beyond Subject Metadata; or, So you have a Taxonomy! ... now what?**  
**Tuesday, February 17, 2014**

Note about locations: Tuesday and Wednesday – all sessions, and the morning and afternoon breaks, are at the NMSU Mega Classroom in the building southwest of the Access Innovations building; breakfast and lunch are served at Access Innovations’ offices.

**8:00 AM – 9:00 AM                    REGISTRATION AND BREAKFAST**

**9:00 AM – 12:00 PM                WELCOME AND GREETING, FEATURES UPDATE**  
*Marjorie Hlava, Access Innovations, Inc.*

What is new and what has changed in the last year for the Data Harmony software? Each year we discuss new features in the software and related services, as well as new directions for the coming year.

**12:00 PM – 1:00 PM                LUNCH**

**1:00 PM – 1:45 PM                Case Study: The JTHES as Part of the Intelligence Layer for the Sustainability Collection Prototype**  
*Sharon Garewal and Ron Snyder, [JSTOR](#)*

The JSTOR Sustainability Collection, which will launch in 2015, is composed of journals, reports, and working papers selected in consultation with scholars, policy researchers, and subject librarians. The collection features journal titles from academic publishers, scholarly societies, and industry groups, as well as a substantial library of indexed reports and working papers from leading research institutes and university centers. It addresses the emerging interdisciplinary discussion about how the environment and human activities and economic gains can be made durable over the long term. Along with this broad set of content, the collection will feature specialized functionality to support research in this emerging field, including a semantic indexing feature that helps researchers locate related terms and concepts that may have varying names across disciplines.

Ron Snyder, ITHAKA Labs Director of Research and Development, and Sharon Garewal, Senior Metadata Librarian, will discuss how the JSTOR Thesaurus (JTHES) was applied as part of the intelligence layer for the Sustainability collection prototype. This includes adding a facet for sustainability within the JTHES to tag terms as part of the collection, working with SME's across disciplines, and applying the curated terms into a live data portal.

Sharon Garewal is a Senior Metadata Librarian for JSTOR, a digital library of academic journals, books, and primary sources that helps people discover and use a wide range of content, and that preserves this content for future generations. She is responsible for the maintenance of the JSTOR Thesaurus which includes acting as an organizational resource for education and training on the thesaurus as well as conducting content audits to measure thesaurus accuracy. In this capacity, Sharon has previously presented JSTOR case studies at DHUG and via webinars. She is considered to be the Martha Stewart of the JSTOR Thesaurus (minus the jail time). Sharon has a Masters of Library Science degree from Kent State University and began her glittering JSTOR career in 1999. In her spare time .... Wait, she has an infant son and therefore has no spare time.

**1:45 PM – 2:00 PM**                      **Case Study: Using Taxonomy**  
*Charlotte McNaughton, ASCE*

ASCE is using the Data Harmony XIS as a repository for all of their metadata records with links to the full text and images. The enriched file serves as the back end of the CEKE (Civil Engineering Knowledge Environment) search system supporting the ASCE library as well as the former CEDB on the ASCE Web site. The metadata feeds back and forth to the ASCE Atypon implementation. Fully connected to the Website and the several databases of the ASCE, it offers access to the journals, proceedings, and books published by the society.

Charlotte McNaughton is currently Director of Publishing Technologies for the American Society of Civil Engineers. She has worked on ASCE's Thesaurus with Xi Van Fleet over the past 5 years, and is in the process of finding ways to integrate it with the ASCE Library, hosted by Atypon.

**2:00 PM – 2:45 PM**                      **Case Study: Development and Implementation of the Site Browser - a  
Faceted Navigation Tool for Browsing NIOSH Mining Web Site Content**  
*Audrey F. Glowacki, National Institute for Occupational Safety and Health*

In 2007, NIOSH started a project to improve customer access to information on its Mining web site. This culminated in the Site Browser, a faceted navigation tool for browsing web site content. This presentation will describe the tools and processes that were developed to reach this goal. First, a custom Mining Safety and Health Thesaurus was developed, now stored in MAIstro. Next, a custom web content management system (WCMS) was developed to let users build web pages. The WCMS also allows for automated and manual indexing of content with thesaurus terms. Finally, the Site Browser tool was written, pilot tested, and implemented. This tool operates off of a SQL Server database that underlies the WCMS. Indexing is now an integral part of web content development. Our researchers and other page developers have the first opportunity to tag their own content, but workflows provide two

approval levels before content is published. While we are still in the process of indexing our older content, anecdotal feedback indicates that our customers are finding the new tool to be very useful.

Ms. Glowacki received her BS in Computer Science from the University of Pittsburgh in 1984. Currently a Computer Scientist with the National Institute for Occupational Safety and Health (NIOSH), she has 30 years of experience as a software, web, and database developer. She has developed software in the areas of mining accident data analysis, 3D graphics-based mine equipment design, computer simulation and animation of mine fire escapes, and computerized laboratory data acquisition. Recently, she became an “accidental taxonomist” as a result of a project to improve access to NIOSH Mining web site information through content indexing

**2:45 PM – 3:15 PM**            **BREAK**

**3:15 PM – 3:45 PM**            **Case Study: The Fate Predictor Project**  
*Helen Atkins, [PLOS](#)*

Now that the PLOS data is semantically enriched, there are many things that we can learn from it. Using statistics collected about acceptance or rejection of submitted papers we can predict, with excellent success, the fate of a new submission, saving the PLOS editors thousands of hours.

Helen Atkins has been working in publishing (East and West Coasts) for more years than she cares to remember. She joined PLOS in 2013; prior to that, she was Editorial Director in the publishing division of the American Association for Cancer Research, worked at HighWire Press, and spent many years at ISI (now Thomson Reuters) in a variety of positions.

**3:45 PM – 4:15 PM**            **Case Study: Constructing a Taxonomy for Genes**  
Arthur Buchberg, AACR

**4:15 PM – 5:00 PM**            **Presentation: Implementation of Taxonomy Triples from Data Harmony Exports**  
*James Wonder, MarkLogic*

By implementing taxonomy triples exported from the Data Harmony software, users enjoy more robust semantic search on the MarkLogic platform.

**6:00 PM – 9:00 PM**      **DINNER:** Come and enjoy a sumptuous “Hummingbird Buffet” at the Indian Pueblo Cultural Center. A variety of native American and southwestern options will be available, including non-spicy selections. The Shumakolowa gift shop will be open during the cocktail hour for those wanting to purchase gifts and souvenirs. Enjoy networking with other attendees and relaxing after the first day of DHUG case studies!



**11<sup>th</sup> Annual Data Harmony Users Group (DHUG) Meeting 2015**  
**Beyond Subject Metadata; or, So you have a Taxonomy! ... now what?**  
**Monday, February 16, 2014 - Training**  
**Wednesday, February 18, 2015**

Note about locations: Tuesday and Wednesday – all sessions, and the morning and afternoon breaks, are at the NMSU Mega Classroom in the building southwest of the Access Innovations building; breakfast and lunch are served at Access Innovations’ offices.

<b>8:00 AM – 9:00 AM</b>	<b>BREAKFAST</b>
<b>9:00 AM – 9:30 AM</b>	<b>Case Study: AAAS</b>
<b>9:30 AM - 10:00 AM</b>	<b>Presentation: Linked Data – Bob Kasenchak, Access Innovations, Inc.</b>
<b>10:00 AM – 10:30 AM</b>	<b>Case Study: Scholastic Press</b>
<b>10:30 AM – 11:00 AM</b>	<b>BREAK</b>
<b>11:00 AM – 11:45 AM</b>	<b>Case Study: AIP</b>
<b>11:45 AM – 12:15 PM</b>	<b>Case Study: IEEE</b>
<b>12:15 PM – 1:15 PM</b>	<b>LUNCH</b>
<b>1:15 PM – 1:45 PM</b>	<b>Presentation: Digital Pulp</b>
<b>1:45 PM – 2:15 PM</b>	<b>Case Study: Triumph Learning</b>
<b>2:15 PM – 2:45 PM</b>	<b>Presentation: Caption Colorado</b>
<b>2:45PM – 3:15 PM</b>	<b>BREAK</b>
<b>3:15 PM – 4:00 PM</b>	<b>Presentation: Integracoder &amp; Code-A-Note</b>
<b>4:00 PM – 5:00 PM</b>	<b>Wrap-Up</b> <i>Margie Hlava</i>



**11<sup>th</sup> Annual Data Harmony Users Group (DHUG) Meeting 2015**  
**Beyond Subject Metadata; or, So you have a Taxonomy! ... now what?**  
**Thursday, February 19, 2014 - Training**

**8:00 AM – 9:00 AM            REGISTRATION AND BREAKFAST**

**9:00 AM – 5:00 PM            TRAINING**

Thursday consists of a full-day hands-on training session with Data Harmony software. Using a sample taxonomy, our staff will lead you through exercises to help you feel comfortable using Data Harmony's Thesaurus Master, M.A.I. and the Admin Module.

**1.        Thesaurus Master™**

We'll start by thoroughly covering the basics of viewing and manipulating subject terms in Thesaurus Master. We'll discuss the practice and principles of building and managing a navigable hierarchy and enriching it with conceptual relationships and synonyms.

**2.        M.A.I.™**

This portion of the session will introduce users to the basics of machine-aided indexing, starting with navigating MAIstro's Rule Builder interface, then covering how to write rules, including basic syntax and stylistic conventions. It will also address the "why" of writing complex rules, and knowing when simple rules are not enough. Testing indexing with the Test MAI function and use of the statistics module will also be introduced.

**3.        Interoperability! The Admin Module, Imports/Exports, APIs, etc.**

During this session we will discuss and demonstrate features of the Data Harmony line of software that are designed for interoperability and implementation of your thesaurus projects. Topics will include import and export formats and techniques for thesaurus projects, settings in the Admin Module that affect indexing results, and using API calls to leverage the power of M.A.I. at various points in your workflow.

We will take morning, lunch and afternoon breaks as appropriate.



**11<sup>th</sup> Annual Data Harmony Users Group (DHUG) Meeting 2015**  
**Beyond Subject Metadata; or, So you have a Taxonomy! ... now what?**  
**Friday, February 20, 2015 - Training**

**8:00 AM – 9:00 AM            REGISTRATION AND BREAKFAST**

**9:00 AM – 5:00 PM            TRAINING**

**1.        XIS Projects and MAI Batch Interface**

The morning training will detail development a XIS project using a sample DTD. The tutorial will teach how to create a project, edit database tables, and utilize built-in M.A.I. functionality. The second portion of the demonstration will cover using the new MAI Batch Interface to index a random sample of full-text documents. We will cover applications for both modules and how they can improve indexing accuracy and precision.

**2.        Dive in Deep – Advanced Concepts in Rule Building**

In this session, we will delve more deeply into advanced techniques of rule building. With a little care and creativity, deceptively simple syntactic structures like nesting, parentheses and ELSE statements can turn your rules into intricate and powerful indexing tools. This session will also cover common pitfalls and debugging tips you'll be glad to know when working with more elaborate rules.

**3.        Sandbox Session – One-on-One Interactions**

Friday's afternoon session features hands-on work with Data Harmony's Thesaurus Master and M.A.I. modules. Participants will practice using these taxonomy tools to fashion a thesaurus, test content, and write rules under the guidance of our professional taxonomists.

Access Innovations staff will also be available for individual consulting on your existing thesaurus projects, so bring your vocabulary, rules, questions, and problems and we'll help you work through them.

We will take morning, lunch and afternoon breaks as appropriate.