



International Conference on Biological Ontology & BioCreative
August 1-4, 2016, Corvallis, Oregon, USA

**CH2M HILL Alumni Conference Center
Oregon State University, Corvallis, OR, USA**

August 1-4, 2016

Food, Nutrition, Health and Environment for the 9 billion

ABSTRACT BOOK

www.icbo-conference.org



**@ICBOconf
#ICBO2016**



This work is licensed under a
[Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

Thank you to our conference sponsors



Committees

ICBO BioCreative 2016 Organizing Committee

Conference Chair: Pankaj Jaiswal, Oregon State University (OSU), USA

Program Chair: Robert Hoehndorf, KAUST, Saudi Arabia

Sponsoring and Publicity:

Sivaram Arabandi, ONTOPRO, USA

Dave Clements, Johns Hopkins, Baltimore, USA

Posters and Demos: Paul Schofield, University of Cambridge, UK

Conference Logistics: Oregon State University Conference Services (Donna Williams, Jill Soth, Carly Weber, Jennifer Stotts)

ICBO Program committee	BioCreative Organizing Committee
<ul style="list-style-type: none"> • Anika Oellrich, KCL, UK • Austin Meier, Planteome, OSU, USA • Barry Smith, NCBO and University at Buffalo, USA • Cecilia Arighi, BioCreative & University of Delaware, USA • Chris Mungall, LBNL, USA • Elizabeth Arnaud, Bioversity, France • Eugene Zhang, OSU, USA • Filipe Santana da Silva, Universidade Federal de Pernambuco, Brazil • Georgios Gkoutos, University of Birmingham, UK • Helen Parkinson, EBI, UK • Laurel Cooper, Planteome, OSU, USA • Marie Angélique Laporte, Bioversity, France • Mark Jensen, University at Buffalo, USA • Mark Schildhauer, NCEAS, USA • Mark Wilkinson, UPM, Spain • Mary Dolan, Jackson Laboratories, USA • Matthew Brush, Monarch Initiative, OHSU, USA • Nicole Vasilevsky, Monarch Initiative, OHSU, USA • Paul Schofield, University of Cambridge, UK • Pier Luigi Buttigieg, ENVO, Max-Planck-Institute, Germany • Prashanti Manda, UNC Chapel Hill, USA • Stefan Schulz, Medizinische Universität Graz, Austria • William (Bill) Hogan, University of Florida, USA • Yongqun (Oliver) He, University of Michigan, USA 	<ul style="list-style-type: none"> • Cecilia Arighi, University of Delaware, USA • Alfonso Valencia, Spanish National Cancer Centre, CNIO, Spain • Cathy Wu, University of Delaware and Georgetown University, USA • Donal Comeau, National Center for Biotechnology Information (NCBI), NIH, USA • Fabio Rinaldi, Institute of Computational Linguistics, University of Zurich, Switzerland • Kevin Cohen, University of Colorado, USA • Lynette Hirschman, MITRE Corporation, USA • Martin Krallinger, Spanish National Cancer Centre, CNIO, Spain • Rezarta Islamaj Dogan, National Center for Biotechnology Information (NCBI), NIH, USA • Sun Kim, National Center for Biotechnology Information (NCBI), NIH, USA • Zhiyong Lu, National Center for Biotechnology Information (NCBI), NIH, USA

Table of Contents

BP01:	10
Iagnet: A centrality and INO-based web system for analyzing and visualizing literature-mined networks	10
BP02:	10
Disease Named Entity Recognition Using NCBI Corpus	10
BP03:	11
Label Embedding Approach for Transfer Learning	11
BIT101-D204:	11
Large-scale Semantic Indexing with Biomedical Ontologies	11
BIT102:	11
One tagger, many uses: Illustrating the power of ontologies in dictionary-based named entity recognition	11
BIT103:	12
Scalable Text Mining Assisted Curation of PTM Proteoforms in the Protein Ontology	12
BIT104:	12
Cardiovascular Health and Physical Activity: A Model for Health Promotion and Decision Support Ontologies	12
BIT105-D106-BP04:	13
A Web Application for Extracting Key Domain Information for Scientific Publications using Ontology	13
BIT106:	13
Use of text mining for Experimental Factor Ontology coverage expansion in the scope of target validation	13
D101:	14
Plant Image Segmentation and Annotation with Ontologies in BisQue.....	14
D102:	14
SPARQL2OWL: towards bridging the semantic gap between RDF and OWL	14
D103-W12-05-IP36:	14
The Phenoscape Knowledgebase: tools and APIs for computing across phenotypes from evolutionary diversity and model organisms	14
D104:	15
Updates to the AberOWL ontology repository	15
D106-BP04:	15
Enhancing Information Accessibility of Publications with Text Mining and Ontology	15
D201:	16
Ontobull and BFOConvert: Web-based programs to support automatic ontology conversion	16
D202:	16
Reusing the NCBO BioPortal technology for agronomy to build AgroPortal	16
D203:	17
Humane OWL: RDF and OWL for Humans.....	17
BIT101-D204:	17
Large-scale Semantic Indexing with Biomedical Ontologies	17
D205:	17
Easy Extraction of Terms and Definitions with OWL2TL	17
IP02:	18
Adding evidence type representation to DIDEO	18

IP03:	18
Multi-species Ontologies of the Craniofacial Musculoskeletal System	18
IP04:	18
EGO: a biomedical ontology for integrative epigenome representation and analysis.....	18
IP05:	19
An Ontological Representation for the Transtheoretical Theory	19
IP06:	19
Building a molecular glyco-phenotype ontology to decipher undiagnosed diseases	19
IP07:	20
The Cell Line Ontology integration and analysis of the knowledge of LINCS cell lines	20
IP08:	20
Gold-Standard Ontology-Based Annotation of Concepts in Biomedical Text in the CRAFT Corpus: Updates and Extensions	20
IP09:	21
How to Summarize Big Knowledge Subjects	21
IP10:	21
Analysis of SNOMED ‘bleeding’ concepts & terms.....	21
IP11:	21
Uc_Sense: An Ontology for Scientifically-based Unambiguous Characterization of Sensory Experiences.....	21
IP12:	22
NAPRALERT, from an historical information silo to a linked resource able to address the new challenges in Natural Products Chemistry and Pharmacognosy.	22
IP13:	22
uc_Eating: Ontology for unambiguous characterization of eating and food habits	22
IP14:	22
Towards designing an ontology encompassing the environment-agriculture-food-diet- health knowledge spectrum for food system sustainability and resilience.	22
IP15:	23
uc_Milk: An ontology for scientifically-based unambiguous characterization of mammalian milk, their composition and the biological processes giving rise to their creation.....	23
IP16:	23
Definition Coverage in the OBO Foundry Ontologies: The Big Picture.....	23
IP17:	24
Comparison of ontology mapping techniques to map plant trait ontologies	24
IP18:	24
The ImmPort Antibody Ontology.....	24
IP19:	24
Opportunities and challenges presented by Wikidata in the context of biocuration	24
IP20:	25
Growth of the Zebrafish Anatomy Ontology: Expanded to support adult morphology and dynamic changes in the early embryo	25
IP21:	25
FoodON: A Global Farm-to-Fork Food Ontology	25
IP22:	26
SourceData: Making Data discoverable.....	26
IP23:	26
Towards an Ontology of Schizophrenia	26

IP24:	27
An OBI ontology Datum Proof Sheet	27
IP25:	27
The Zebrafish Experimental Conditions Ontology Systemizing Experimental Descriptions in ZFIN	27
IP26:	27
Performance Evaluation Clinical Task Ontology(PECTO)	27
IP27:	28
Dealing with elements of medical encounters: an approach based on the ontological realism	28
IP28:	29
uc_FIDO: unambiguous characterization of food interactions with drugs ontology	29
IP29:	29
Global Agricultural Concept Scheme: A Hub for Agricultural Vocabularies	29
IP30:	29
Supporting database annotations and beyond with the Evidence & Conclusion Ontology (ECO)	29
IP31:	30
Planteome Gene Annotation Enrichment Analysis	30
IP32:	30
Enhancing SciENCv through semantic research profile integration with the VIVO-ISF ontology	30
IP34:	31
Plant Reactome: A Resource for Comparative Plant Pathway Analysis	31
IT201:	32
Environmental semantics for sustainable development in an interconnected biosphere ..	32
IT202:	32
Sustainable Development Goals Interface Ontology.....	32
IT203:	32
Defining and sustaining populations and communities	32
IT204:	33
A sustainable approach to knowledge representation in the domain of sustainability: bridging SKOS and OWL.....	33
IT205:	33
Data-driven Agricultural Research for Development: A Need for Data Harmonization Via Semantics.....	33
IT206:	34
The UNEP Ontologies and the OBO Foundry.....	34
IT401:	34
An analysis of differences in biological pathway resources	34
IT402:	34
Enhancing the Human Phenotype Ontology for Use by the Layperson	34
IT403:	35
OntONeo: The Obstetric and Neonatal Ontology.....	35
IT404:	35
Ten simple rules for biomedical ontology development.....	35
IT405:	36
Building Concordant Ontologies for Drug Discovery	36

IT406-IP35:	36
The Planteome Project	36
IT407:	37
Annotating germplasm to Planteome reference ontologies.....	37
IT501:	37
A Descriptive Delta for Identifying Changes in SNOMED CT	37
IT502:	38
Visualizing the “Big Picture” of Change in NCIt’s Biological Processes.....	38
IT503:	38
Malaria study data integration and information retrieval based on OBO Foundry ontologies	38
IT504:	39
OOSTT: a Resource for Analyzing the Organizational Structures of Trauma Centers and Trauma Systems.....	39
IT505:	39
Towards a Standard Ontology Metadata Model	39
IT506:	40
An Ontological Framework for Representing Topological Information in Human Anatomy 40	
IT507:	40
Natural Language Definitions for the Leukemia Knowledge Domain.....	40
IT601:	40
Identifying Missing Hierarchical Relations in SNOMED CT from Logical Definitions Based on the Lexical Features of Concept Names	40
IT602:	41
A Semantic Web Representation of Entire Populations.....	41
IT603:	41
Improving the Semantics of Drug Prescriptions with a Realist Ontology	41
IT604:	42
Qualitative causal analyses of biosimulation models.....	42
IT605:	42
SEPIO: A Semantic Model for the Integration and Analysis of Scientific Evidence	42
IT606:	43
Measuring the importance of annotation granularity to the detection of semantic similarity between phenotype profiles	43
IT701:	43
A Quality-Assurance Study of ChEBI	43
IT702:	44
To MIREOT or not to MIREOT? A case study of the impact of using MIREOT in the Experimental Factor Ontology (EFO)	44
IT703:	44
Semantic Digitization of Experimental Data in Biological Sciences	44
IT704:	45
Representation of parts within the Foundational Model of Anatomy ontology.....	45
IT705:	45
A Realist Representation of Social Identity Data	45
W14-03:	45
The Plant Phenology Ontology for Phenological Data Integration.....	45
W14-02:	46

The Biological Collections Ontology for linking traditional and contemporary biodiversity data 46

BT101:**46**
 Cycles of Scientific Investigation in Discourse - Machine Reading Methods for the Primary Research Contributions of a Paper 46

BT102:**47**
 Collaborative Workspaces for Pathway Curation 47

BT103:**47**
 Crowdsourcing Protein Family Database Curation 47

BT104:**47**
 Opportunities and challenges presented by Wikidata in the context of biocuration 47

BT201:**48**
 Text mining to enable routine personalized cancer therapy 48

BT202:**48**
 Social Media Mining for Pharmacovigilance 48

BT203:**48**
 MutD – A PubMed Scale Resource for Protein Mutation-Disease Relations through Bio-Medical Literature Mining 48

BT204:**49**
 CancerMine: Knowledge base construction for personalised cancer treatment 49

BT205:**49**
 Text Mining for Drug Development: Gathering Insights to Support Decision Making 49

BT301:**50**
 NLP for the Institute: Developing and Deploying an NLP Capability to Accelerate Cancer Research 50

BT302:**50**
 Annotations for biomedical research and healthcare -- Bridging the gap 50

BT303:**51**
 PubAnnotation: a public shared platform for scientific literature annotation. 51

BT304:**51**
 BioCconvert: A Conversion Tool between BioC and PubAnnotation 51

W04-01.....**51**
 Ontology of the Organigram 51

W04-02.....**52**
 Relations between Institutional Roles and Deontic Roles in Biomedical Organizations 52

W04-03.....**52**
 An ontological study of healthcare corporations and their social entities 52

W04-04.....**53**
 Who's paying and who's graying? The organizations and roles associated with insurance policies, funding agencies, and national census data 53

W04-05.....**53**
 Dealing with social and legal entities in the obstetric and neonatal domain 53

W01-01.....**53**
 Big Data Visual Analysis 53

W01-02.....**54**
 Topology-Driven Data Visualization of Large-Scale Tensor Fields 54

W01-03.....**54**
 Computer Vision for Next Generation Phenomics and Tree of Life 54

W01-05	54
Uncertainty analysis and visualization in Large-Scale Vector Fields.....	54
W01-04	55
Telling a genome's story graphically.....	55
W11-01	55
Why We Need A Food Ontology.....	55
W11-02	55
International Efforts in Creating Food Vocabularies	55
W11-03	55
Sustainable food systems and food in ecosystems	55
W11-04	56
Food composition and links to human health	56
W11-05:	56
Food safety and infectious disease.....	56
W11-06	57
FoodON Use cases: Caution! Food Allergies Ahead	57
W11-07	57
IC3-Foods: An infrastructure for the next generation internet of food systems, food, and health.....	57

BP01:

Ignet: A centrality and INO-based web system for analyzing and visualizing literature-mined networks

Arzucan Ozgur, Junguk Hur, Zuoshuang Xiang, Edison Ong, Dragomir Radev and Yongqun He

Abstract: Ignet (Integrative Gene Network) is a web-based system for dynamically updating and analyzing gene interaction networks mined using all Pub- Med abstracts. Four centrality metrics, namely degree, eigenvector, betweenness, and closeness are used to determine the importance of genes in the networks. Different gene interaction types between genes are classified using the Interaction Network Ontology (INO) that classifies interaction types in an ontological hierarchy along with individual keywords listed for each interaction type. An interactive user interface is designed to explore the interaction network as well as the centrality and ontology based network analysis. Availability: <http://ignet.hegroup.org>.

BP02:

Disease Named Entity Recognition Using NCBI Corpus

Thomas Hahn, Hidayat Ur Rahman and Richard Segall

Abstract: Named Entity Recognition (NER) in biomedical literature is a very active research area. NER is a crucial component of biomedical text mining because it allows for information retrieval, reasoning and knowledge discovery. Much research has been carried out in this area using semantic type categories, such as *fiDNAfl*, *fiRNAfl*, *fiproteinsfl* and *figenesfl*. However, disease NER has not received its needed attention yet, specifically human disease NER. Traditional machine learning approaches lack the precision for disease NER, due to their dependence on token level features, sentence level features and the integration of features, such as orthographic, contextual and linguistic features. In this paper a method for disease NER is proposed which utilizes sentence and token level features based on Conditional Random Fields using the NCBI disease corpus. Our system utilizes rich features including orthographic, contextual, affixes, bigrams, part of speech and stem based features. Using these feature sets our approach has achieved a maximum F-score of 94% for the training set by applying 10 fold cross validation for semantic labeling of the NCBI disease corpus. For testing and development corpus the model has achieved an F-score of 88% and 85% respectively.

BP03:

Label Embedding Approach for Transfer Learning

Rasha Obeidat, Xiaoli Fern and Prasad Tadepalli

Abstract: Automatically tagging textual mentions with the concepts, types and entities that they represent are important tasks for which supervised learning has been found to be very effective. In this paper, we consider the problem of exploiting multiple sources of training data with variant ontologies. We present a new transfer learning approach based on embedding multiple label sets in a shared space, and using it to augment the training data.

BIT101-D204:

Large-scale Semantic Indexing with Biomedical Ontologies

Chih-Hsuan Wei, Robert Leaman and Zhiyong Lu

Abstract: We introduce PubTator, a web-based application that enables large-scale semantic indexing and automatic concept recognition in biomedical ontologies. Not only was PubTator formally evaluated and top-rated in BioCreative, it also has been widely adopted and used by the scientific community from around the world, supporting both research projects and real-world applications in biocuration, crowdsourcing and translational bioinformatics.

BIT102:

One tagger, many uses: Illustrating the power of ontologies in dictionary-based named entity recognition

Lars Juhl Jensen

Abstract: Automatic annotation of text is an important complement to manual annotation, because the latter is highly labour intensive. We have developed a fast dictionary-based named entity recognition (NER) system and addressed a wide variety of biomedical problems by applied it to text from many different sources. We have used this tagger both in real-time tools to support curation efforts and in pipelines for populating databases through bulk processing of entire Medline, the open-access subset of PubMed Central, NIH grant abstracts, FDA drug labels, electronic health records, and the Encyclopedia of Life. Despite the simplicity of the approach, it typically achieves 80-90% precision and 70-80% recall. Many of the underlying dictionaries were built from open biomedical ontologies, which further facilitate integration of the text-mining results with evidence from other sources.

BIT103:

Scalable Text Mining Assisted Curation of PTM Proteoforms in the Protein Ontology

Karen Ross, Darren Natale, Cecilia Arighi, Sheng-Chih Chen, Hongzhan Huang, Gang Li, Jia Ren, Michael Wang, K Vijay-Shanker and Cathy Wu

Abstract: The Protein Ontology (PRO) defines protein classes and their interrelationships from the family to the protein form (proteoform) level within and across species. One of the unique contributions of PRO is its representation of post-translationally modified (PTM) proteoforms. However, progress in adding PTM proteoform classes to PRO has been relatively slow due to the extensive manual curation effort required. Here we report an automated pipeline for creation of PTM proteoform classes that leverages two phosphorylation-focused text mining tools (RLIMS-P, which detects mentions of kinases, substrates, and phosphorylation sites, and eFIP, which detects phosphorylation-dependent protein-protein interactions (PPIs)) and our integrated PTM database, iPTMnet. By applying this pipeline, we obtained a set of ~820 substrate-site pairs that are suitable for automated PRO term generation with literature-based evidence attribution. Inclusion of these terms in PRO will increase PRO coverage of species-specific PTM proteoforms by 50%. Many of these new proteoforms also have associated kinase and/or PPI information. Finally, we show a phosphorylation network for the human and mouse peptidyl-prolyl cis-trans isomerase (PIN1/Pin1) derived from our dataset that demonstrates the biological complexity of the information we have extracted. Our approach addresses scalability in PRO curation and will be further expanded to advance PRO representation of phosphorylated proteoforms.

BIT104:

Cardiovascular Health and Physical Activity: A Model for Health Promotion and Decision Support Ontologies

Vimala Ponna, Aaron Baer and Matthew Lange

Abstract: Current cardiovascular disease decision support systems (DSS) rely primarily on ontologies that characterize and quantify disease, recommending appropriate pharmacotherapy (PT) and/or surgical interventions (SI). PubMed and Google Scholar searches reveal no specific ontologies or literature related to DSS for recommending physical activity (PA) and diet interventions (DI) for cardiovascular health and fitness (CVHF) improvement. This dearth of CVHF-PA/DI structured knowledge repositories has resulted in a scarcity of user-friendly tools for scientifically validated information retrieval about CVHF improvement. Advancement of health science depends on timely development and implementation of health (rather than disease) ontologies. We developed a time-efficient workflow for constructing/maintaining structured knowledge repositories capable of providing informational underpinnings for CVHF- PA/DI ontologies and DSS that support health promotion, including precise, personalized exercise prescription. This workflow creates conceptual lattices about effects of varied

PA on CVHF. These conceptual maps lay the foundation for accelerated creation of health-focused ontologies, which ultimately equip DSS with CVHF knowledge related PA and DI.

BIT105-D106-BP04:

A Web Application for Extracting Key Domain Information for Scientific Publications using Ontology

Weijia Xu, Amit Gupta, Pankaj Jaiswal, Crispin Taylor and Patti Lockhart

Abstract: We present demos of an ongoing project, domain informational vocabulary extraction (DIVE), which aims to enrich digital publications through entity and key informational words detection and by adding additional annotations. The system implements multiple strategies for biological entity detection, including using regular expression rules, ontologies, and a keyword dictionary. These extracted entities are then stored in a database and made accessible through an interactive web application for curation and evaluation by authors. Through the web interface, the user can make additional annotations and corrections to the current results. The updates can then be used to improve the entity detection in subsequent processed articles. Although the system is being developed in the context of annotating journal articles, it can be also be beneficial to domain curators and researchers at large.

BIT106:

Use of text mining for Experimental Factor Ontology coverage expansion in the scope of target validation

Senay Kafkas, Ian Dunham, Helen Parkinson and Johanna McEntyre

Abstract: Understanding the molecular biology and development of disease plays a key role in drug development. Integrating evidence from different experimental approaches with data available from public resources (such as gene expression level changes and reaction pathways affected by pathogenic mutations) can be a powerful approach for evaluating different aspects of target-disease associations. The application of ontologies is of fundamental importance to effective integration. The Target Validation Platform is a user-friendly interface that integrates such evidences from various resources with the aim of assisting scientists to identify and prioritise drug targets. Currently, the EFO is used as the reference ontology for diseases in the platform, importing terms from existing disease ontologies such as the Human Phenotype Ontology as required. In order to generalize the use of EFO from key target-diseases for wider use, we need to compare the target associated disease coverage in EFO with the scope of other available disease terminology resources. In this study, we address this issue by using text mining and present our initial results.

D101:

Plant Image Segmentation and Annotation with Ontologies in BisQue

Justin Preece, Justin Elser, Pankaj Jaiswal, Kris Kvilekval, Dmitry Fedorov, B.S. Manjunath, Ryan Kitchen, Xu Xu, Dmitrios Triggakis, Sinisa Todorovic and Seth Carbon

Abstract: The field of computer vision has experienced much progress in the last two decades. Image analysis of photography and video has moved out of computer science research labs and into a wide range of applications. One example of progress in image analysis concerns the segmentation of images on the basis of gray scale, color hue, texture, geometry, and other features. Such image segmentation allows for increasingly refined classification of images and their components. In a parallel development, semantic computing has pursued the creation of ontologies in hopes of capturing and defining what it is we “know” about the world, and presenting it in the form of a terminology network connected by defined relationships. This knowledge network is computable, and makes it possible to make logical inferences about facts and data annotated with ontology terms.

D102:

SPARQL2OWL: towards bridging the semantic gap between RDF and OWL

Mona Alsharani, Hussein Almashouq and Robert Hoehndorf

Abstract: Several large databases in biology are now making their information available through the Resource Description Framework (RDF). RDF can be used for large datasets and provides a graph-based semantics. The Web Ontology Language (OWL), another Semantic Web standard, provides a more formal, model- theoretic semantics. While some approaches combine RDF and OWL, for example for querying, knowledge in RDF and OWL is often expressed differently. Here, we propose a method to generate OWL ontologies from SPARQL queries using n-ary relational patterns. Combined with background knowledge from ontologies, the generated OWL ontologies can be used for expressive queries and quality control of RDF data. We implement our method in a prototype tool available at <https://github.com/bio-ontology-research-group/SPARQL2OWL>.

D103-W12-05-IP36:

The Phenoscape Knowledgebase: tools and APIs for computing across phenotypes from evolutionary diversity and model organisms

James Balhoff

Abstract: The Phenoscape Knowledgebase (KB) is an ontologydriven database that combines existing phenotype annotations from model organism databases with new phenotype annotations from the evolutionary literature. Phenoscape curators have

created phenotype annotations for more than 5,000 species and higher taxa, by defining computable phenotype concepts for more than 20,000 character states from over 160 published phylogenetic studies. These phenotype concepts are in the form of Entity–Quality (EQ) [1] compositions which incorporate terms from the Uberon anatomy ontology, the Biospatial Ontology (BSPO), and the Phenotype and Trait Ontology (PATO). Taxonomic concepts are drawn from the Vertebrate Taxonomy Ontology (VTO). This knowledge of comparative biodiversity is linked to potentially relevant developmental genetic mechanisms by importing associations of genes to phenotypic effects and gene expression locations from zebrafish (ZFIN [2]), mouse (MGI [3]), Xenopus (Xenbase [4]), and human (Human Phenotype Ontology project [5]). Thus far, the Phenoscape KB has been used to identify candidate genes for evolutionary phenotypes [6], to match profiles of ancestral evolutionary variation with gene phenotype profiles [7], and to combine data across many evolutionary studies by inferring indirectly asserted values within synthetic supermatrices [8]. Here we describe the software architecture of the Phenoscape KB, including data ingestion, integration of OWL reasoning, web service interface, and application features (Fig. 1).

D104:

Updates to the AberOWL ontology repository

Miguel Ángel Rodríguez-García, Luke Slater, Imane Boudellioua, Paul Schofield, Georgios Gkoutos and Robert Hoehndorf

Abstract: A large number of ontologies have been developed in the biological and biomedical domains, which are mostly expressed in the Web Ontology Language (OWL). These ontologies form a logical foundation for our knowledge in these domains, and they are in widespread use to annotate biomedical and biological datasets. The use of the semantics provided by ontologies requires the use of automated reasoning – inferring new knowledge by evaluating the asserted axioms. AberOWL is an ontology repository which utilises an OWL 2 EL reasoner to provide semantic access to classified ontologies. Since our original presentation of the AberOWL framework, we have developed several additional tools and features which enrich its ability to integrate and explore data, make use of the semantic and inferred content of ontologies. Here we present an overview of AberOWL and the enhancements and new features which have been developed since its conception. AberOWL is freely available at <http://aber-owl.net>.

D106-BP04:

Enhancing Information Accessibility of Publications with Text Mining and Ontology

Weijia Xu, Amit Gupta, Pankaj Jaiswal, Crispin Taylor and Patti Lockhart

Abstract: We present an ongoing effort on utilizing text mining methods and existing biological ontologies to help readers to access the information contained in the scientific

articles. Our approach includes using multiple strategies for biological entity detection and using association analysis on extracted analysis. The entity extraction processes utilizes regular expression rules, ontologies, and keyword dictionary to get a comprehensive list of biological entities. In addition to extract list of entities, we also apply natural language processing and association analysis techniques to generate inferences among entities and comparing to known relations documented in the existing ontologies.

D201:

[Ontobull and BFOConvert: Web-based programs to support automatic ontology conversion](#)

Edison Ong, Zuoshuang Xiang, Jie Zheng, Barry Smith and Yongqun He

Abstract: When a widely reused ontology appears in a new version which is not compatible with older versions, the ontologies reusing it need to be updated accordingly. Ontobull (<http://ontobull.hegroup.org>) has been developed to automatically update ontologies with new term IRI(s) and associated metadata to take account of such version changes. To use the Ontobull web interface a user is required to (i) upload one or more ontology OWL source files; (ii) input an ontology term IRI mapping; and (where needed) (iii) provide update settings for ontology headers and XML namespace IDs. Using this information, the backend Ontobull Java program automatically updates the OWL ontology files with desired term IRIs and ontology metadata. The Ontobull subprogram BFOConvert supports the conversion of an ontology that imports a previous version of BFO. A use case is provided to demonstrate the features of Ontobull and BFOConvert.

D202:

[Reusing the NCBO BioPortal technology for agronomy to build AgroPortal](#)

Clement Jonquet, Anne Toulet, Elizabeth Arnaud, Sophie Aubin, Esther Dzale Yeumo, Vincent Emonet, John Graybeal, Mark A. Musen, Cyril Pommier and Pierre Larmande

Abstract: Many vocabularies and ontologies are produced to represent and annotate agronomic data. By reusing the NCBO BioPortal technology, we have already designed and implemented an advanced prototype ontology repository for the agronomy domain. We plan to turn that prototype into a real service to the community. The AgroPortal project aims at reusing the scientific outcomes and experience of the biomedical domain in the context of plant, agronomic, food, environment (perhaps animal) sciences. We offer an ontology portal which features ontology hosting, search, versioning, visualization, comment, recommendation, enables semantic annotation, as well as storing and exploiting ontology alignments. All of these within a fully semantic web compliant infrastructure. The AgroPortal specifically pays attention to respect the requirements of the agronomic community in terms of ontology formats (e.g., SKOS,

trait dictionaries) or supported features. In this paper, we present our prototype as well as preliminary outputs of four driving agronomic use cases. With the experience acquired in the biomedical domain and building atop of an already existing technology, we think that AgroPortal offers a robust and stable reference repository that will become highly valuable for the agronomic domain.

D203:

[Humane OWL: RDF and OWL for Humans](#)

James A. Overton

Abstract: Humane OWL (HOWL) is a syntax for RDF and OWL designed for manual editing. By allowing human-readable labels to be used in place of IRIs, and providing convenient syntax for OWL annotations and expressions, HOWL files can be used like source code with tools such as GitHub, then translated into any other RDF or OWL format for use with other tools.

BIT101-D204:

[Large-scale Semantic Indexing with Biomedical Ontologies](#)

Chih-Hsuan Wei, Robert Leaman and Zhiyong Lu

Abstract: We introduce PubTator, a web-based application that enables large-scale semantic indexing and automatic concept recognition in biomedical ontologies. Not only was PubTator formally evaluated and top-rated in BioCreative, it also has been widely adopted and used by the scientific community from around the world, supporting both research projects and real-world applications in biocuration, crowdsourcing and translational bioinformatics.

D205:

[Easy Extraction of Terms and Definitions with OWL2TL](#)

John Judkins, Joseph Utecht and Mathias Brochhausen

Abstract: "Facilitating good communication between semantic web specialists and domain experts is necessary to efficient ontology development. This development may be hindered by the fact that domain experts tend to be unfamiliar with tools used to create and edit OWL files. This is true in particular when changes to definitions need to be reviewed as often as multiple times a day. We developed ""OWL to Term List"" (OWL2TL) with the goal of allowing domain experts to view the terms and definitions of an OWL file organized in a list that is updated each time the OWL file is updated. The tool is available online and currently generates a list of terms, along with additional

annotation properties that are chosen by the user, in a format that allows easy copying into a spreadsheet."

IP02:

[Adding evidence type representation to DIDEO](#)

Mathias Brochhausen, Philip E. Empey, Jodi Schneider, William R. Hogan and Richard D. Boyce

Abstract: In this poster we present novel development and extension of the Drug-drug Interaction and Drug-drug Interaction Evidence Ontology (DIDEO). We demonstrate how reasoning over this extension of DIDEO can a) automatically create a multi-level hierarchy of evidence types from descriptions of the underlying scientific observations and b) automatically subsume individual evidence items under the correct evidence type. Thus DIDEO will enable evidence items added manually by curators to be automatically categorized into a drug-drug interaction framework with precision and minimal effort from curators. As with all previous DIDEO development this extension is consistent with OBO Foundry principles.

IP03:

[Multi-species Ontologies of the Craniofacial Musculoskeletal System](#)

Jose Leonardo Mejino, James Brinkley, Timothy Cox and Landon Detwiler

Abstract: We created the Ontology of Craniofacial Development and Malformation (OCDM) [1] to provide a unifying framework for organizing and integrating craniofacial data ranging from genes to clinical phenotypes from multi-species. Within this framework we focused on spatio-structural representation of anatomical entities related to craniofacial development and malformation, such as craniosynostosis and midface hypoplasia. Animal models are used to support human studies and so we built multi-species ontologies that would allow for cross-species correlation of anatomical information. For this purpose we first developed and enhanced the craniofacial component of the human musculoskeletal system in the Foundational Model of Anatomy Ontology (FMA)[2], and then imported this component, which we call the Craniofacial Human Ontology (CHO), into the OCDM. The CHO was then used as a template to create the anatomy for the mouse, the Craniofacial Mouse Ontology (CMO) as well as for the zebrafish, the Craniofacial Zebrafish Ontology (CZO).

IP04:

[EGO: a biomedical ontology for integrative epigenome representation and analysis](#)

Yongqun He, Zhaohui Qin and Jie Zheng

Abstract: Epigenomics is crucial to understand biological mechanisms beyond genome DNA. To better represent epigenomic knowledge and support data integration, we developed a prototype Epigenome Ontology (EGO). EGO top level hierarchy and design pattern are provided with a use case illustration. EGO is proposed to be used for statistically analyzing enriched epigenomic features based on given sequence data input using statistical methods.

IPO5:

[An Ontological Representation for the Transtheoretical Theory](#)

Hua Min, Robert H. Friedman and Julie Wright

Abstract: Ontologies are widely used in computer science and medicine. Ontologies may be useful in health promotion and disease prevention for intervention development. Interventionists usually use theory to guide intervention design and evaluation, but there is no standard vocabulary for health behavior theory. A formal mechanism for converting theory to a computer-based representation may provide a tool that can assist in the development of computer-based interventions. This paper demonstrates how ontology can be used to represent a health behavior theory using the Transtheoretical Model (TTM) of behavior change as an example.

IPO6:

[Building a molecular glyco-phenotype ontology to decipher undiagnosed diseases](#)

Jean-Philippe Gourdine, Thomas Metz, David Koeller, Matthew Brush and Melissa Haendel

Abstract: Hundreds of rare diseases are due to mutation on genes related to glycans synthesis, degradation or recognition. These glycan-related defects are well described in the literature but largely absent in ontologies and databases of chemical entities and phenotypes, limiting the application of computational methods and ontology-driven tools for characterization and discovery of glycan related diseases. We are curating articles and textbooks in glycobiology related to genetic diseases to inform the content and the structure of an ontology of Molecular Glyco-Phenotypes (MGPO). MGPO will be applied toward use cases including disease diagnosis and disease gene candidate prioritization, using semantic similarity and pattern matching at the glycan level with glycomics data from patient of the Undiagnosed Diseases Network.

IP07:

[The Cell Line Ontology integration and analysis of the knowledge of LINCS cell lines](#)

Edison Ong, Jiangan Xie, Zhaohui Ni, Qingping Liu, Yu Lin, Vasileios Stathias, Caty Chung, Stephan Schurer and Yongqun He

Abstract: Cell lines are crucial to study molecular signatures and pathways, and are widely used in the NIH Common Fund LINCS project. The Cell Line Ontology (CLO) is a community-based ontology representing and classifying cell lines from different resources. To better serve the LINCS research community, from the LINCS Data Portal and ChEMBL, we identified 1,097 LINCS cell lines, among which 717 cell lines were associated with 121 cancer types, and 352 cell line terms did not exist in CLO. To harmonize LINCS cell line representation and CLO, CLO design patterns were slightly updated to add new information of the LINCS cell lines including different database cross-reference IDs. A new shortcut relation was generated to directly link a cell line to the disease of the patient from whom the cell line was originated. After new LINCS cell lines and related information were added to CLO, a CLO subset/view (LINCS-CLOview) of LINCS cell lines was generated and analyzed to identify scientific insights into these LINCS cell lines. This study provides a first time use case on how CLO can be updated and applied to support cell line research from a specific research community or project initiative.

IP08:

[Gold-Standard Ontology-Based Annotation of Concepts in Biomedical Text in the CRAFT Corpus: Updates and Extensions](#)

Michael Bada, Nicole Vasilevsky, Melissa Haendel and Lawrence Hunter

Abstract: Ontologies are increasingly used for semantic integration across disparate curated biomedical resources, while gold-standard annotated corpora are needed for accurate training and evaluation of text-mining tools. Bringing together the respective power of these, we created the Colorado Richly Annotated Full-Text (CRAFT) Corpus, a collection of full-length, open-access biomedical journal articles that have been manually annotated both syntactically and semantically with select Open Biomedical Ontologies (OBOs), the first release of which includes ~100,000 annotations of concepts mentioned in the text of 67 articles and mapped to the classes of eight prominent OBOs. Here we present our continuing work on the corpus, including updated versions of these annotations with newer versions of the ontologies, new annotations made with two additional OBOs, annotations made with newly created extension classes defined in terms of existing classes of the ontologies, and new annotations of roots of prefixed and suffixed words.

IP09:

[How to Summarize Big Knowledge Subjects](#)

Ling Zheng, Yehoshua Perl, James Geller and Gai Elhanan

Abstract: One manifestation of the “Big Knowledge” challenge is providing automated tools for summarization of ontology content to facilitate user comprehension. An aggregation approach for the automatic identification and display of major subjects covered by an ontology’s content is presented. The results show that our methodology is viable in capturing the “big picture” of ontology content.

IP10:

[Analysis of SNOMED ‘bleeding’ concepts & terms](#)

Jonathan Bona, Selja Seppälä and Werner Ceusters

Abstract: We present an analysis of SNOMED CT ‘bleeding’ concepts – those concepts with descriptions that include ‘hematoma’, ‘hemorrhage’, or ‘bleeding’; or that are descended from ‘Bleeding (finding)’ in the Is-a hierarchy; or that have Hematomas or Hemorrhages as their associated morphology – to assess how consistently they are used in the ontology.

IP11:

[Uc_Sense: An Ontology for Scientifically-based Unambiguous Characterization of Sensory Experiences](#)

Aaron Baer and Matthew Lange

Abstract: Recent efforts in biological ontology go to great lengths to unambiguously categorize biological entities and phenomena of the natural world, as well as their relationships with each other. This paper illustrates the importance of unambiguously characterizing the perception of entities relative to biological apparatus required for specific modes of sensing, physiological conduction of sensed experiences, along with subjective interpretation and communication of sensed experiences. In addition to building a computable knowledge base around existing sensory science, ontological modelling of this aspect of biology will enable an increased understanding about alternate perceptions of identical stimuli. Leveraging this understanding to modulate desired behavior toward increased health and happiness outcomes is a fundamental goal of this project.

IP12:

[NAPRALERT, from an historical information silo to a linked resource able to address the new challenges in Natural Products Chemistry and Pharmacognosy.](#)

Jonathan Bisson, James McAlpine, James Graham and Guido Pauli

Abstract: NAPRALERT (<https://www.napralert.org>) is a database on natural products, including data on ethnobotany, chemistry, pharmacology, toxicology, and clinical trials from literature dating back to the 19th century. Established in 1975 by Norman R. Farnsworth, it became a web accessible resource in 2005 but soon became stagnant while literature grew exponentially. After a complete rewrite of the platform, the focus is now on connecting this resource to the rest of the existing databases and expanding its usability. The creation of a Pharmacognosy/Natural Product ontology will foster better understanding of this domain, its linking potential with other resources and the ability to automatize literature annotation and entry efficiently.

IP13:

[uc_Eating: Ontology for unambiguous characterization of eating and food habits](#)

Kimiya Taji and Matthew Lange

Abstract: The uc_Eating ontology is a standardized unambiguous characterization system for modeling human food habits and eating processes. The uc_Eating ontology along with the physiological, environmental, behavioral, and food ontologies it maps to, provide an infrastructure for annotating the relationships between food, food consumption, eating behaviors, and environments creating a foundation for computable knowledge bases around food and beverage consumption scenarios, their observation, interrogation, and manipulation at biological, behavioral, and environmental levels.

IP14:

[Towards designing an ontology encompassing the environment-agriculture-food-diet-health knowledge spectrum for food system sustainability and resilience.](#)

Ruthie Musker, Matthew Lange, Allan Hollander, Patrick Huber, Nathaniel Springer, Courtney Riggle, James Quinn and Thomas Tomich

Abstract: Feeding 9 billion people is not solely a matter of food, health, nutrition, and the environment. Promoting human health by increasing the sustainability and resilience of food systems requires integrating information from a broad range of disciplines from human nutrition/health systems and agricultural/natural systems to social, financial, physical and political systems. Ontologies serve to specify common terminologies for critical concepts and relationships within these systems, however very few ontologies have been developed with this interdisciplinary focus. Biological ontologies, whether focused on human physiology, soil quality, or nutritional value are

only part of the story when it comes to determining linkages throughout the food system that help determine human health and well-being. We seek to build an ontology of food and food systems that encompasses the relevant sustainability issues in their entirety. We have already built an ontology of sustainable sourcing of agricultural raw materials issues and indicators, but aim to expand our ontology to include attributes of resilience, and other issues along the environment-agriculture-food-diet-health knowledge spectrum. Additionally, we aim to create this ontology with the intention of quick usability for the food system decision-maker.

IP15:

[uc_Milk: An ontology for scientifically-based unambiguous characterization of mammalian milk, their composition and the biological processes giving rise to their creation](#)

Emeline Colet and Matthew Lange

Abstract: Recent efforts in biological ontology go to great lengths to unambiguously categorize biological entities and phenomena of the natural world, as well as their relationships with each other. This paper illustrates the importance of unambiguously characterizing mammalian milk because milk is a complex mixture of many chemical components and thus represents a key role in infant nourishment and development. In addition to the build of a computable knowledge base around mammalian milk, ontological modeling of this aspect of biology and chemistry enable increased understanding of mammalian milk composition and the biological structures and biochemical processes giving rise to their creation. Utilizing unambiguous vocabularies to compare human milk with other mammalian milks relative to the biological and behavioral survival challenges facing varied mammalian organisms and the phenotypic qualities each milk confers, is a fundamental goal of this project.

IP16:

[Definition Coverage in the OBO Foundry Ontologies: The Big Picture](#)

Daniel Schlegel, Selja Seppälä and Peter Elkin

Abstract: High quality ontologies have both textual and logical definitions for their terms. Definitions serve many purposes: good textual definitions allow for experts and non-experts alike to understand the content of an ontology and use it in the manner the authors intended; logical definitions are necessary for reasoners to verify that an ontology is consistent, and may make application of the ontology easier for users. Ideally, logical and textual definitions would convey the same information, and each can provide an accuracy check on the other.

IP17:

[Comparison of ontology mapping techniques to map plant trait ontologies](#)

Marie-Angélique Laporte, Léo Valette, Laurel Cooper, Chris Mungall, Austin Meier, Pankaj Jaiswal and Elizabeth Arnaud

Abstract: Crop specific ontologies for phenotype annotations in breeding have proliferated over the last 10 years. Across-crop data interoperability involves linking those ontologies together. For this purpose, the Planteome project is mapping the Crop Ontology traits (www.croponontology.org) to the reference ontology for plant traits, Trait Ontology (TO). Manual mapping is time-consuming and not sustainable in the long-run as ontologies keep on evolving and multiplying. We are thus working on developing reliable automated mapping techniques to assist curators in performing semantic integration. Our study shows the benefit of the ontology matching technique based on formal definitions and shared ontology design patterns, compared to standard automatic ontology matching algorithm, such as AML (AgreementMakerLight).

IP18:

[The ImmPort Antibody Ontology](#)

William Duncan, Travis Allen, Jonathan Bona, Olivia Helfer, Barry Smith, Alan Ruttenberg and Alexander D. Diehl

Abstract: Monoclonal antibodies are essential biomedical research and clinical reagents that are produced by companies and research laboratories. The NIAID Immunology Database and Analysis Portal (ImmPort) is a sustainable data warehouse for data generated by NIAID, DAIT and DMID funded studies designed to allow long-term archiving and re-use of immunological data [1]. A variety of immunological data in ImmPort is generated using techniques that rely upon monoclonal antibody reagents, including flow cytometry, immunofluorescence, and ELISA. In order to facilitate querying, integration, and reuse of these data, standardized terminology for describing monoclonal antibody reagents and their targets needs to be used for annotating data submitted to ImmPort.

IP19:

[Opportunities and challenges presented by Wikidata in the context of biocuration](#)

Benjamin Good, Timothy Putman, Andrew Su, Andra Waagmeester, Sebastian Burgstaller-Muehlbacher and Elvira Mitraka

Abstract: Wikidata is a world readable and writable knowledge base maintained by the Wikimedia Foundation. It offers the opportunity to collaboratively construct a fully open access knowledge graph spanning biology, medicine, and all other domains of knowledge. To meet this potential, social and technical challenges must be overcome

most of which are familiar to the biocuration community. These include community ontology building, high precision information extraction, provenance, and license management. By working together with Wikidata now, we can help shape it into a trustworthy, unencumbered central node in the Semantic Web of biomedical data.

IP20:

[Growth of the Zebrafish Anatomy Ontology: Expanded to support adult morphology and dynamic changes in the early embryo](#)

Ceri Van Slyke, Yvonne Bradford and Christian Pich

Abstract: The Zebrafish Anatomy Ontology (ZFA) is an application ontology used by ZFIN to support curation of expression and phenotype. The research community also uses the ontology to support annotation of high throughput studies. As the research focus of the zebrafish community evolves it drives changes in the ZFA. Here we provide an update on the changes made to support research carried out in adult fish and describe the changes in modeling of the neural crest in the ontology in order to bring the structure of the ontology into closer accord with the morphological changes that occur during development.

IP21:

[FoodON: A Global Farm-to-Fork Food Ontology](#)

Emma Griffiths, Damion Dooley, Pier Luigi Buttigieg, Robert Hoehndorf, Fiona Brinkman and William Hsiao

Abstract: Several resources and standards for indexing food descriptors currently exist, but their content and interrelations are not semantically and logically coherent. Simultaneously, the need to represent knowledge about food is central to many fields including biomedicine and sustainable development. FoodON is a new ontology built to interoperate with the OBO Library and to represent entities which bear a . It encompasses materials in natural ecosystems and food webs as well as human-centric categorization and handling of food. The latter will be the initial focus of the ontology, and we aim to develop semantics for food safety, food security, the agricultural and animal husbandry practices linked to food production, culinary, nutritional and chemical ingredients and processes. The scope of FoodON is ambitious and will require input from multiple domains. FoodON will import or map to material in existing ontologies and standards and will create content to cover gaps in the representation of food-related products and processes. As a robust food ontology can only be created by consensus and wide adoption, we are currently forming an international consortium to build partnerships, solicit domain expertise, and gather use cases to guide the ontologies development. The products of this work are being applied to research and clinical datasets such as those associated with the Canadian Healthy Infant Longitudinal

Development (CHILD) study which examines the causal factors of asthma and allergy development in children, and the Integrated Rapid Infectious Disease Analysis (IRIDA) platform for genomic epidemiology and foodborne outbreak investigation.

IP22:

SourceData: Making Data discoverable

Nancy George, Sara El-Gebali and Thomas Lemberger

Abstract: In molecular and cell biology, most of the data presented in published papers are not available in formats that allow for direct analysis and systematic mining. The goal of the SourceData project (<http://sourcedata.embo.org>) is to make published data easier to find, to connect papers containing related information and to promote the reuse and novel analysis of published data. The main concept underlying the project is that the structure of a dataset provides information about the design of the study in question and can be exploited in powerful data-oriented search strategies. SourceData has therefore developed tools to generate machine-readable descriptive metadata from figures in published manuscripts. Experimentally tested hypotheses are represented as directed relationships between standardized biological entities. Once processed, a comprehensive 'scientific knowledge graph' can be generated from this data (see demo video1 at <https://vimeo.com/sourcedata/kg>), making the body of data efficiently searchable. Importantly, this graph is objectively grounded in published data and not on the potentially subjective interpretation of the results.

IP23:

Towards an Ontology of Schizophrenia

Fumiaki Toyoshima

Abstract: The paper presents an Ontology of Schizophrenia (OS) that is designed to provide a formal representation of schizophrenia-related entities crucial to the treatment and study of schizophrenia. OS is developed in accordance with the OBO (Open Biomedical Ontology) Foundry principles and constructed in compliance with Basic Formal Ontology (BFO) as its upper-level ontology. It uses mid-level and domain ontologies built in conformity with BFO such as the Ontology for General Medical Science (OGMS) and the Mental Functioning Ontology (MFO). OS is developed using Protégé 4.3 and is implemented in OWL2. Having imported BFO2.0 OWL and the development version of OGMS, OS adds approximately 30 schizophrenia-related terms and attempts to provide both formal and textual definitions for each. The terms in OS are in addition annotated with ontology metadata such as labels, textual definitions, definition sources, term editors and editor notes.

IP24:

[An OBI ontology Datum Proof Sheet](#)

Damion Dooley, Emma Griffiths, Fiona Brinkman and William Hsiao

Abstract: There are numerous past and current examples of ontology-driven projects that provide auto-generated user interfaces for managing entities and relations, each presenting its own varied and complex data model. Our Datum Proof Sheet application aims to simplify the application development landscape by building community consensus about the way basic categorical, textual and numeric datum fields should be described within the OBOFoundry community of ontologies. The proof sheet shows selected datums (grouped under the context of an OBI “data representational model” item) as form inputs on an HTML page, enabling an application ontology’s contents to be presented to end users (ranging in our case from epidemiologists to software developers) for review without necessarily having a working application to showcase them in. The basic relations and cases necessary for presenting datums in a user interface are mostly satisfied by OBI’s design, but we introduce a few extra elements to bring more clarity to datum specifications, and to provide user interface term labels and definitions that may differ from those that ontologists prefer in the “backend”.

IP25:

[The Zebrafish Experimental Conditions Ontology Systemizing Experimental Descriptions in ZFIN](#)

Yvonne Bradford, Ceri Van Slyke, Sabrina Toro and Sridhar Ramachandran

Abstract: The Zebrafish Experimental Conditions Ontology (ZECO) defines the major experimental conditions used in research studies that employ the zebrafish, *Danio rerio*. We are systematically building the ontology to encompass both standard control conditions and experimental conditions and it is designed to allow better data curation and more precise information retrieval.

IP26:

[Performance Evaluation Clinical Task Ontology\(PECTO\)](#)

Jose F Florez-Arango, Santiago Patiño-Giraldo, Jack W Smith and Sriram Iyengar

Abstract: This poster presents a proposed Clinical Tasks Ontology (PECTO) designed to evaluate effects of technologies on human performance under controlled conditions, such as clinical simulations scenarios (CSS), across multiple clinical domains including prehospital care. In recent years there has been an explosion of technologies, including Information and Communications Technologies (ICTs) that are designed to assist health workers and improve their performance across a spectrum of clinical activities from pre-hospital care to post-surgical care. However, each new technology introduces its own

requirements on the health worker and has the potential to either increase or decrease the perceived workload on the health-worker. Since perceived workload can have significant effects on health worker performance [4], it is important to carefully measure work-load changes and relate these to health worker performance measures such as task errors, and procedure compliance. Clinical Simulation Scenarios are often used to perform controlled experiments in which health workers' performance on clinical conditions, simulated by various means, including Human Patient Simulators, is observed and measured with and without the technology being studied[5]. The Clinical Tasks Ontology (PECTO) was developed, among other applications, to help design such evaluation experiments. A major objective of such studies is to evaluate the performance of health workers as they perform specific clinical tasks. In this context, the PECTO presents a novel approach for task classification and analysis since previous approaches [6]–[8] do not account for sources of workload, and measurement of human performance in terms of errors and protocol compliance. An ontological approach was selected to build the classification system enabling tasks to have multiple properties that can be related to dependent variables. Previous work in task analysis and task classification include an ontological approach to plans and processes[6], some modeling of event evaluation, and a clinical task model of care plans[8], as well as comprehensive approaches to model human factors and workload.

IP27:

Dealing with elements of medical encounters: an approach based on the ontological realism

Fernanda Farinelli, Mauricio Almeida, Peter Elkin and Barry Smith

Abstract: Electronic health records (EHRs) serve as repositories of documented data collected in a health care encounter. An EHR records information about who receives, who provides the health care and about the place where the encounter happens. We also observe additional elements relating to social relations in which the healthcare consumer is involved. To provide a consensus representation of common data and to enhance interoperability between different EHR repositories we have created a solution grounded in formal ontology. Here, we present how an ontology for the obstetric and neonatal domain deals with these general elements documented in health care encounters. Our goal is to promote the interoperability of information among EHRs created in different specialties. To develop our ontology, we used two main approaches: one based on ontological realism, the other based on the principles of the OBO Foundry, including reuse of reference ontologies.

IP28:

[uc_FIDO: unambiguous characterization of food interactions with drugs ontology](#)

Constantine Spyrou and Matthew Lange

Abstract: uc_FIDO is an ontology that unambiguously characterizes food interactions with drugs in the human body. This ontology is part of a group of food ontologies describing food and the human experience at the International Center for Food Ontology Operability, Data and Semantics (IC-FOODS) at UC Davis. The first of its kind, uc_FIDO characterizes relations between food, medicine, and human health. uc_FIDO brings together several existing ontologies related to anatomy, metabolic pathways, biological processes, drug ingredients and food structures. Through these ontologies, uc_FIDO annotates relationships between food and drug bioactives, human physiological conditions, and biological reaction pathways. Relationships that link together fully characterize various food interactions with drugs and their effects. The current dearth of ontologies for characterizing foods limits advancement of informatics solutions for improving health. As ontologies of foods are developed, it becomes necessary to describe ingredients, bioactive molecules, potential toxins, and other molecules in food interacting with drugs and the human body.

IP29:

[Global Agricultural Concept Scheme: A Hub for Agricultural Vocabularies](#)

Caterina Caracciolo, Tom Baker and Elizabeth Arnaud

Abstract: Thesauri are used to tag semi-structured documents, texts, while more complex semantic structures are used to describe (annotate) scientific data. We are creating a Global Agricultural Concept Scheme (GACS) by mapping AGROVOC, CABT and NALT to three major thesauri in the area of food and agriculture, with a beta release in May 2016. We see GACS as a hub linking user-oriented thesauri with semantically more precise domain ontologies linking, in turn, to datasets about food and agriculture, in order to make that data more interoperable and reusable

IP30:

[Supporting database annotations and beyond with the Evidence & Conclusion Ontology \(ECO\)](#)

Marcus Chibucos, Suvarna Nadendla, James Munro, Elvira Mitraka, Dustin Olley, Nicole Vasilevsky, Matthew Brush and Michelle Giglio

Abstract: The Evidence & Conclusion Ontology (ECO) is a community standard for summarizing evidence in scientific research in a controlled, structured way. Annotations at the world's most frequented biological databases (e.g. model organisms, UniProt, Gene Ontology) are supported using ECO terms. ECO describes evidence derived from

experimental and computational methods, author statements curated from the literature, inferences drawn by curators, and other types of evidence. Here, we describe recent ECO developments and collaborations, most notably: (i) a new ECO website containing user documentation, up-to-date news, and visualization tools; (ii) improvements to the ontology structure; (iii) implementing logic via an ongoing collaboration with the Ontology for Biomedical Investigations (OBI); (iv) addition of numerous experimental evidence types; and (v) addition of new evidence classes describing computationally derived evidence. Due to its utility, popularity, and simplicity, ECO is now expanding into realms beyond the protein annotation community, for example the biodiversity and phenotype communities. As ECO continues to grow as a resource, we are seeking new users and new use cases, with the hope that ECO will continue to be a broadly used and easy-to-implement community standard for representing evidence in diverse biological applications. Feel free to visit two ECO-sponsored workshops at ICBO 2016 to learn more: 1. “An introduction to the Evidence and Conclusion Ontology and representing evidence in scientific research” and 2. “OBI-ECO Interactions & Evidence”.

IP31:

Planteome Gene Annotation Enrichment Analysis

Botong Qu, Jaden Diefenbaugh, Eugene Zhang, Justin Elser, Pankaj Jaiswal, Seth Carbon and Christopher Mungall

Abstract: Annotation enrichment analysis of a gene list helps biologists to identify the potential biological functions associated with it. With the extensions of plant ontology categories, the discovery of significant ontology terms associated with a gene list becomes more and more informative. We introduce a tool to help biologists to find out these terms based on the expanding ontology database of the Planteome project. In addition, we propose some new visualization schemes to help users construct a meaningful interpretation of the results guided by the ontology tree.

IP32:

Enhancing SciENCv through semantic research profile integration with the VIVO-ISF ontology

Marijane White, Matthew Brush, Shahim Essaid, Robin Champieux, Adrienne Zell, Melissa Haendel, Colin Grove, Syeda Momina Tabish and David Eichmann

Abstract: SciENCv [1], the US Federal Science Experts Network Curriculum Vitae, is an online system for simplifying the creation of researcher biographical sketches or biosketches, which are required when applying for federal funding. SciENCv profiles are curated and controlled by researchers themselves - they own the data, they control what data are public, and they edit and maintain the information

contained within the profiles, which includes expertise, employment history, educational background, and professional accomplishments. The system leverages data from myNCBI and eRA Commons, and includes links to ORCID [2] researcher identifiers. The structure of SciENCv biosketches is defined by an XML Schema Definition and profiles can be downloaded in XML format. The system aims to eliminate the need for researchers to repeatedly enter biosketch information and reduce the administrative burden associated with federal grant submission and reporting requirements, as well as creating a repository of researcher profile data where researchers can describe their scientific contributions in their own language.

IP34

Plant Reactome: A Resource for Comparative Plant Pathway Analysis

Sushma Naithani, Justin Preece, Parul Gupta, Peter D'Eustachio, Justin Elser, Antonio Mundao, Joel Weiser, Sheldon McKay, Lincoln, Doreen Ware, Pankaj Jaiswal

Abstract: The Plant Reactome database (<http://plantreactome.gramene.org/>) hosts metabolic, genetic and signaling pathways for several model and crop plant species. The Reactome data model organizes gene products, small molecules and macromolecular interactions into reactions and pathways in the context of their subcellular location to build a systemslevel framework of a plant cell. The Plant Reactome features *Oryza sativa* (rice) as a reference species, built by importing the RiceCyc metabolic network and curating new metabolic, signaling and genetic pathways. The Plant Reactome database now contains 241 rice reference pathways and orthology-based pathway projections for 58 plant species. Plant Reactome allows users to i) compare pathways across various plant species; ii) query and visualize curated baseline and differential expression data available in the EMBL-EBI's Expression Atlas in the context of pathways in the Plant Reactome; and iii) analyze genome-scale expression data and conduct pathway enrichment analysis to enable researchers to identify pathways affected by the stresses or treatments studied in their data sets. Plant Reactome links out to numerous external reference resources, including the gene pages of Gramene, Phytozome, SoyBase, Legume Information System, PeanutBase, Uniprot, as well as ChEBI for small molecules, PubMed for literature supported evidences, and GO for molecular function and biological processes. Users can access/download our data in various formats from our website and via APIs. The presentation will discuss tools for pathway enrichment analysis and homologue pathway comparison, development of the Plant Reactome portal, curation of reference rice pathways, and phylogeny-based analyses of projected pathway annotations. The project is supported by the Gramene database award (NSF IOS-1127112) and the Human Reactome award (NIH: P41 HG003751, ENFIN LSHG-CT-2005-518254, Ontario Research Fund, and EBI Industry Programme).

IT201:

Environmental semantics for sustainable development in an interconnected biosphere

Pier Luigi Buttigieg, Mark Jensen, Ramona Walls and Christopher Mungall

Abstract: Development unavoidably impacts the ecosystems constituting often producing complex outcomes across a range of spatial and temporal scales. The interconnectivity of global ecosystems and their varied responses to disturbance necessitate great caution when using or encountering terms such as sustainable and sustainability. The Environment Ontology (ENVO; www.environmentontology.org) is coordinating with the Sustainable Development Goals Interface Ontology (SDGIO) to improve semantic representation of environments in the context of global development. In this talk, we will present progress towards this goal, emphasizing the potential of ecosystem semantics in bridging and seeding new domain ontologies.

IT202:

Sustainable Development Goals Interface Ontology

Mark Jensen

Abstract: Ontologies are built in a variety of circumstances, for different applications. Some are quite small and built with minimal consideration (or need) for extensibility. Others, such as many of the large-scale efforts in the OBO Foundry, have a team of developers, domain specialists and collaborators actively engaged in interactive development [1-2]. The size of the ontology, or the complexity of the domain, aren't necessarily indicators of the development methodology and to what extent efforts are made to integrate the ontology with existing ontologies or other semantic resources. The use of ontologies by large organizations and government agencies is growing. For example, the Common Core ontologies by the US Army, the Joint Doctrine ontology, United States Geological Survey [3-4]. Within the context of integrating ontologies with sizable agencies, problems arise over harmonizing current vocabularies, agency doctrine and standards, while still attempting to build a useful tool that still adheres to best practice in developing modular and extensible realism-based ontologies.

IT203:

Defining and sustaining populations and communities

Ramona Walls

Abstract: The UN Sustainable Development Goals (SDGs) cover subjects such as ending poverty, achieving gender equality, ensuring access to water, energy, and food, and protecting ecosystems. Every SDG includes wording that refers directly or indirectly to a population or community or organisms, humans or otherwise. Therefore, the Population and Community Ontology (PCO) plays a crucial role in defining the language of the SDGs

and their targets and indicators. This talk will describe the PCO and its applicability to sustainability studies.

IT204:

A sustainable approach to knowledge representation in the domain of sustainability: bridging SKOS and OWL

Chris Mungall, Mark Jensen, Marie-Angélique Laporte and Pier Buttigieg

Abstract: n/a

IT205:

Data-driven Agricultural Research for Development: A Need for Data Harmonization Via Semantics

Medha Devare, Céline Aubert, Marie-Angélique Laporte, Léo Valette, Elizabeth Arnaud and Pier Luigi Buttigieg

Abstract: Addressing global challenges to agricultural productivity and profitability increasingly requires access to data from a variety of disciplines, and the ability to easily combine and analyze related data sets. Innovation in agricultural research for development must therefore be mediated by reliable and consistently annotated information resources across disciplinary domains. Leveraging semantics ensures this consistency and ease of reuse, and the global CGIAR Consortium that includes 15 agricultural research for development Centers is attempting to harness this promise through efforts such as its Open Access, Open Data Initiative. CGIAR™s Crop Ontology project plays a key role in this, and will soon be enhanced by an Agronomy Ontology (AgrO). AgrO is being built to represent traits identified by agronomists and the simulation model variables of the International Consortium for Agricultural Systems Applications (ICASA). Further, it will coordinate its semantics with existing ontologies such as the Environment Ontology (ENVO), Unit Ontology (UO), and Phenotype And Trait Ontology (PATO). Once stable, it is anticipated to address one of the domains temporarily represented in the Sustainable Development Goals Interface Ontology (SDGIO), pertaining to multiple SDGs such as the elimination of hunger and poverty. AgrO will complement existing crop, livestock, and fish ontologies to enable harmonized approaches to data collection, facilitating data sharing and reuse. Further, AgrO will power an Agronomy Management System and fieldbook, similar to the Crop Ontology-based Integrated Breeding Platform (IBP) and fieldbook. There is substantial interest from agronomists and modelers in such a fieldbook to standardize agronomic data collection, and the ontology itself as a means of facilitating hitherto missing linkages with breeding and other data, and enabling wider sharing and reuse of agronomic research data.

IT206:

The UNEP Ontologies and the OBO Foundry

Barry Smith and Mark Jensen

Abstract: It is now generally accepted that wherever we need to address multiple different kinds of data deriving from multiple different kinds of sources, a strategy is required to ensure interoperability across the various systems involved. The United Nations Environmental Program (UNEP) is developing an ontology for use in their knowledge management platform—the Sustainable Development Goals Interface Ontology (SDGIO). The UNEP approach to achieving interoperability is modeled on the OBO (Open Biomedical Ontologies) Foundry, an initiative of ontology developers working in a range of life science domains who have been working since 2004 to develop a suite of ontologies which would work together consistently in such a way as to advance the integration of data [1]. To serve such integration the OBO Foundry members formulated and tested an evolving set of principles for ontology development which are now being used by ontology developers also in other areas, including manufacturing, geology, transport and security.

IT401:

An analysis of differences in biological pathway resources

Lucy Wang, John Gennari and Neil Abernethy

Abstract: Integrating content from multiple biological pathway resources is necessary to fully exploit pathway knowledge for the benefit of biology and medicine. Differences in content, representation, coverage, and more occur between databases, and are challenges to resource merging. We introduce a typology of representational differences between pathway resources and give examples using several databases: BioCyc, KEGG, PANTHER pathways, and Reactome. We also detect and quantify annotation mismatches between HumanCyc and Reactome. The typology of mismatches can be used to guide entity and relationship alignment between these databases, helping us identify and understand deficiencies in our knowledge, and allowing the research community to derive greater benefit from the existing pathway data.

IT402:

Enhancing the Human Phenotype Ontology for Use by the Layperson

Nicole Vasilevsky, Mark Engelstad, Erin Foster, Chris Mungall, Peter Robinson, Sebastian Köhler and Melissa Haendel

Abstract: In rare or undiagnosed diseases, physicians rely upon genotype and phenotype information in order to compare abnormalities to other known cases and to

inform diagnoses. Patients are often the best sources of information about their symptoms and phenotypes. The Human Phenotype Ontology (HPO) contains over 12,000 terms describing abnormal human phenotypes. However, the labels and synonyms in the HPO primarily use medical terminology, which can be difficult for patients and their families to understand. In order to make the HPO more accessible to non-medical experts, we systematically added new synonyms using non-expert terminology (i.e., layperson terms) to the existing HPO classes or tagged existing synonyms as layperson. As a result, the HPO contains over 6,000 classes with layperson synonyms.

IT403:

[OntONeo: The Obstetric and Neonatal Ontology](#)

Fernanda Farinelli, Mauricio Almeida, Peter Elkin and Barry Smith

Abstract: This paper presents the Obstetric and Neonatal Ontology (OntONeo). This ontology has been created to provide a consensus representation of salient electronic health record (EHR) data and to serve interoperability of the associated data and information systems. More generally, it will serve interoperability of clinical and translational data, for example deriving from genomics disciplines and from clinical trials. Interoperability of EHR data is important to ensuring continuity of care during the prenatal and postnatal periods for both mother and child. As a strategy to advance such interoperability we use an approach based on ontological realism and on the ontology development principles of the Open Biomedical Ontologies Foundry, including reuse of reference ontologies wherever possible. We describe the structure and coverage domain of OntONeo and the process of creating and maintaining the ontology.

IT404:

[Ten simple rules for biomedical ontology development](#)

Melanie Courtot, James Malone and Chris Mungall

Abstract: Biomedical ontology development is often a time and resource consuming endeavor. To maximize efficiency of the process, we present a set of 10 simple rules covering basic technical requirements such as scoping and versioning, while considering additional elements such as licensing and community engagement. When applied, the rules will help avoid common pitfalls and jump-start ontology building.

IT405:

Building Concordant Ontologies for Drug Discovery

Hande Küçük-McGinty, Saurabh Metha, Yu Lin, Nooshin Nabizadeh, Vasileios Stathias, Dusica Vidovic, Amar Koleti, Christopher Mader, Jianbin Duan, Ubbo Visser and Stephan Schürer

Abstract: In this study we demonstrate how we interconnect three different ontologies, the BioAssay Ontology (BAO), LINCS Information Framework ontology (LIFEo), and the Drug Target Ontology (DTO). The three ontologies are built and maintained for three different projects: BAO for the BioAssay Ontology Project, LIFEo for the Library of Integrated Network-Based Cellular Signatures (LINCS) project, and DTO for the Illuminating the Druggable Genome (IDG) project. DTO is a new ontology that aims to formally describe drug target knowledge relevant to drug discovery. LIFEo is an application ontology to describe information in the LIFE software system. BAO is a highly accessed NCBO ontology; it has been extended formally to describe several LINCS assays. The three ontologies use the same principle architecture that allows for re-use and easy integration of ontology modules and instance data. Using the formal definitions in DTO, LIFEo, and BAO and data from various resources one can quickly identify disease-relevant and tissue-specific genes, proteins, and prospective small molecules. We show a simple use case example demonstrating knowledge-based linking of life science data with the potential to empower drug discovery.

IT406-IP35:

The Planteome Project

Laurel Cooper, Austin Meier, Justin Elser, Justin Preece, Xu Xu, Ryan Kitchen, Botong Qu, Eugene Zhang, Sinisa Todorovic, Pankaj Jaiswal, Marie-Angélique Laporte, Elizabeth Arnaud, Seth Carbon, Chris Mungall, Barry Smith, Georgios Gkoutos and John Doonan

Abstract: The Planteome project is a centralized online plant informatics portal which provides semantic integration of widely diverse datasets with the goal of plant improvement. Traditional plant breeding methods for crop improvement may be combined with next-generation analysis methods and automated scoring of traits and phenotypes to develop improved varieties. The Planteome project (www.planteome.org) develops and hosts a suite of reference ontologies for plants associated with a growing corpus of genomics data. Data annotations linking phenotypes and germplasm to genomics resources are achieved by data transformation and mapping species-specific controlled vocabularies to the reference ontologies. Analysis and annotation tools are being developed to facilitate studies of plant traits, phenotypes, diseases, gene function and expression and genetic diversity data across a wide range of plant species. The project database and the online resources provide researchers tools to search and browse and access remotely via APIs for semantic

integration in annotation tools and data repositories providing resources for plant biology, breeding, genomics and genetics.

IT407:

Annotating germplasm to Planteome reference ontologies

Austin Meier, Laurel Cooper, Justin Elser, Pankaj Jaiswal and Marie-Angélique Laporte

Abstract: An expected use case of plant phenotype ontologies will be the identification of germplasm containing particular traits of interest. If phenotype data from experiments is annotated using ontologies, it makes sense to include annotations to that germplasm source. A lack of standardized data formatting reduces the utility of these data. Standardizing germplasm data, including links to germplasm databases, or distribution locations improves collaboration, and benefits both researchers and the scientific community as a whole. All plant traits contained in the Planteome reference ontologies are searchable, and interconnected through relationships in the ontology. All data annotated to these reference ontologies will be displayed, shareable, and computable through the Planteome website (www.planteome.org) and APIs. This manuscript will discuss the advantages of standardizing germplasm trait annotation, and the semi-automated process developed to achieve such standardization.

IT501:

A Descriptive Delta for Identifying Changes in SNOMED CT

Christopher Ochs, Yehoshua Perl, Gai Elhanan and James Case

Abstract: SNOMED CT is a large and complex medical terminology. Thousands of editing operations are applied to its content for each new release. Understanding what changed in a release is important for the end user and SNOMED CT editors. Each SNOMED CT release comes with release notes that provide a brief description of the changes that occurred and a set of delta files that identify individual changes in the content. The release notes are brief and changes to thousands of concepts may be described in a few sentences, whereas the delta files contain tens of thousands of individual changes. To better identify how SNOMED CT content changes between releases we introduce a methodology of creating a descriptive delta that captures the editing operations that were applied to SNOMED CT content in a given release in a more comprehensible form. We use this methodology to analyze editing operations that were part of a recent remodeling effort of the Congenital disease and Infectious disease subhierarchies in the large Clinical finding hierarchy.

IT502:

Visualizing the “Big Picture” of Change in NCIt’s Biological Processes

Yehoshua Perl, Christopher Ochs, Sherri de Coronado and Nicole Thomas

Abstract: The National Cancer Institute thesaurus (NCIt) is a large and complex ontology. NCIt is frequently updated; a new release is made available approximately every month. Tracking structural changes in NCIt is important for the editors of its content. In this paper we describe a methodology and tool using diff partial-area taxonomies to visually summarize structural changes between two NCIt releases. Diff partial-area taxonomies provide a comprehensible view of the overall impact of the changes. This methodology is illustrated using the Biological Process hierarchy. Specifically, we illustrate how diff partial-area taxonomies reflect change that occurred due to major restructuring of this hierarchy between September 2004 and December 2004. During this time the hierarchy nearly doubled in size and a large portion of the classes were extensively modified. Several kinds of change patterns are identified and discussed.

IT503:

Malaria study data integration and information retrieval based on OBO Foundry ontologies

Jie Zheng, Jashon Cade, Brian Brunk, David Roos, Chris Stoeckert, San James, Emmanuel Arinaitwe, Bryan Greenhouse, Grant Dorsey, Steven Sullivan, Jane Carlton, Gabriel Carrasco-Escobar, Dionicia Gamboa, Paula Maguina-Mercedes and Joseph Vinetz

Abstract: The International Centers of Excellence in Malaria Research (ICEMR) projects involve studies to understand the epidemiology and transmission patterns of malaria in different geographic regions. Two major challenges of integrating data across these projects are: (1) standardization of highly heterogeneous epidemiologic data collected by various ICEMR projects; (2) provision of user-friendly search strategies to identify and retrieve information of interest from the very complex ICEMR data. We pursued an ontology-based strategy to address these challenges. We utilized and contributed to the Open Biological and Biomedical Ontologies to generate a consistent semantic representation of three different ICEMR data dictionaries that included ontology term mappings to data fields and allowed values. This semantic representation of ICEMR data served to guide data loading into a relational database and presentation of the data on web pages in the form of search filters that reveal relationships specified in the ontology and the structure of the underlying data. This effort resulted in the ability to use a common logic for storing and display of data on study participants, their clinical visits, and epidemiological information on their living conditions (dwelling) and geographic location. Users of the Plasmodium Genomics Resource, PlasmoDB, accessing the ICEMR data will be able to search for participants based on environmental factors such as type of dwelling, location or mosquito biting rate, characteristics such as age at enrollment,

relevant genotypes or gender and visit data such as laboratory findings, diagnoses, malaria medications, symptoms, and other factors.

IT504:

[OOSTT: a Resource for Analyzing the Organizational Structures of Trauma Centers and Trauma Systems](#)

Joseph Utecht, John Judkins, Mathias Brochhausen, Terra Colvin Jr., J. Neil Otte, Nicholas Rogers, Robert Rose, Maria Alvi, Amanda Hicks, Jane Ball, Stephen M. Bowman, Robert T. Maxson, Rosemary Nabaweesi, Rohit Pradhan, Nels D. Sanddal, M. Eduard Tudoreanu and Robert Winchell

Abstract: Organizational structures of healthcare organizations has increasingly become a focus of medical research. In the CAFÉ project we aim to provide a web-service enabling ontology-driven comparison of the organizational characteristics of trauma centers and trauma systems. Trauma remains one of the biggest challenges to healthcare systems worldwide. Research has demonstrated that coordinated efforts like trauma systems and trauma centers are key components of addressing this challenge. Evaluation and comparison of these organizations is essential. However, this research challenge is frequently compounded by the lack of a shared terminology and the lack of effective information technology solutions for assessing and comparing these organizations. In this paper we present the Ontology of Organizational Structures of Trauma systems and Trauma centers (OOSTT) that provides the ontological foundation to CAFÉ's web-based questionnaire infrastructure. We present the usage of the ontology in relation to the questionnaire and provide the methods that were used to create the ontology.

IT505:

[Towards a Standard Ontology Metadata Model](#)

Hua Min, Stuart Turner, Sherri de Coronado, Brian Davis, Trish Whetzel, Robert R. Freimuth, Harold R. Solbrig, Richard Kiefer, Michael Riben, Grace A. Stafford, Lawrence Wright and Riki Ohira

Abstract: Bio-ontologies are becoming increasingly important in semantic alignment for data integration, information exchange, and semantic interoperability. Due to the large number of emerging bio-ontologies, it is challenging for ontology for their applications. Therefore, it is important to have a consistent terminology metadata model and a resource for discovering appropriate ontologies or other resource for use in annotating data. This paper aims to seek a common, shareable, and comprehensive method to create, disseminate, and consume metadata about terminology resources.

IT506:

[An Ontological Framework for Representing Topological Information in Human Anatomy](#)

Takeshi Imai, Emiko Shinohara, Masayuki Kajino, Ryota Sakurai, Kazuhiko Ohe, Kouji Kozaki and Riichiro Mizoguchi

Abstract: Medical ontologies have been a focus of constant attention in recent years as one of the fundamental techniques and knowledge bases for clinical decision support applications. In this paper, we discuss the description framework of our anatomy ontology with a focus on representing topological information, which is required for anatomical reasoning in clinical decision support applications. Our framework has major advantages over preceding studies with respect to: (1) representations of branching sequence; (2) combined representation of relevant knowledge with the use of “general structural component”; and (3) cooperation with the disease and abnormality ontologies.

IT507:

[Natural Language Definitions for the Leukemia Knowledge Domain](#)

Amanda Damasceno De Souza and Maurício Barcellos Almeida

Abstract: The creation of natural definitions is a phase of any methodology to build formal ontologies. In order to reach formal definitions, one should first create natural language definitions according to sound principles. We gather a set of principles available in literature and organize them in a list of stages that one can use to create good definitions in natural language. In order to test the set of principles, we conducted a case study in which we create definitions in the domain of cancer, more specifically, definitions for acute myeloid leukemia. After creating and validating the definition of this specific kind of leukemia, we offer remarks about the experiment.

IT601:

[Identifying Missing Hierarchical Relations in SNOMED CT from Logical Definitions Based on the Lexical Features of Concept Names](#)

Olivier Bodenreider

Abstract: Objectives. To identify missing hierarchical relations in SNOMED CT from logical definitions based on the lexical features of concept names. Methods. We first create logical definitions from the lexical features of concept names, which we represent in OWL EL. We infer hierarchical (subClassOf) relations among these concepts using the ELK reasoner. Finally, we compare the hierarchy obtained from lexical features to the original SNOMED CT hierarchy. We review the differences manually for evaluation purposes. Results. Applied to 15,833 disorder and procedure concepts, our approach identified 559 potentially missing hierarchical relations, of which 78% were

deemed valid. Conclusions. This lexical approach to quality assurance is easy to implement, efficient and scalable.

IT602:

A Semantic Web Representation of Entire Populations

Daniel Welch, Amanda Hicks, Josh Hanna and William Hogan

Abstract: Accurately representing demographic realities is a critical component in creating useful, agent-based epidemiological models of infectious disease. Synthetic ecosystems are generated from Census data microsamples in a statistically-sound manner to maintain population-level demographic characteristics. These highly detailed representations of populations are the basis of many advanced simulations of infectious disease epidemics. Creating a standard, machine-readable representation of synthetic ecosystem data would enable easier use and integration with epidemic simulator software. Here we describe an ontology-based representation in Resource Description Framework (RDF) and Web Ontology Language (OWL) of version 1.0 of the 2010 U.S. Synthetic Population database by RTI International. Our representation draws upon applicable classes from several reference ontologies, including the Ontology of Medically Related Social Entities (OMRSE). After failing to find suitable ontological representations of several key data elements in the Synthetic Population dataset, we created new classes in OMRSE for representing employment status, employee roles, workplaces, residences, households, and age measurements. We loaded a test RDF dataset (structured according to ontologies in OWL) of synthetic individuals into a commercial triple store (Stardog) and validated the representation with SPARQL queries.

IT603:

Improving the Semantics of Drug Prescriptions with a Realist Ontology

Jean-Francois Ethier, Ryeyan Taseen, Luc Lavoie and Adrien Barton

Abstract: Electronic prescriptions are supported as a means to reduce adverse drug events, but the ambiguities and overspecificities of prescription semantics along with their lack of standardization reduce adoption, limit interoperability and are potential sources of error. Ontologies in the OBO Foundry, founded on realist methodology, have been successful in fostering the logical, scientifically accurate data standards that the domain of drug prescriptions is currently in need of. This paper illustrates some problems regarding the structuration of current electronic prescriptions, and demonstrates how the Prescription of Drugs Ontology (PDRO) addresses these issues with improved semantics founded on OBO and realist principles. PDRO reuses classes and object properties from IAO, OBI, OGMS, OMRSE and DRON, introducing new entities within its scope and proposing entities within those of its imported domains that may be useful to other health care and information artifact-related ontologies in the OBO

Foundry. PDRO aims at improving the semantics of drug prescriptions and prospectively enabling the interoperability of prescription data.

IT604:

Qualitative causal analyses of biosimulation models

Maxwell Neal, John Gennari and Daniel Cook

Abstract: We describe an approach for performing qualitative, systems-level causal analyses on biosimulation models that leverages semantics-based modeling formats, formal ontology, and automated inference. The approach allows users to quickly investigate how a qualitative perturbation to an element within a model's network (an increment or decrement) propagates throughout the modeled system. To support such analyses, we must interpret and annotate the semantics of the models, including both the physical properties modeled and the dependencies that relate them. We build from prior work understanding the semantics of biological properties, but here, we focus on the semantics for dependencies, which provide the critical knowledge necessary for causal analysis of biosimulation models. We describe augmentations to the Ontology of Physics for Biology, via OWL axioms and SWRL rules, and demonstrate that a reasoner can then infer how an annotated model's physical properties influence each other in a qualitative sense. Our goal is to provide researchers with a tool that helps bring the systems-level network dynamics of biosimulation models into perspective, thus facilitating model development, testing, and application.

IT605:

SEPIO: A Semantic Model for the Integration and Analysis of Scientific Evidence

Matthew Brush, Kent Shefchek and Melissa Haendel

Abstract: The Scientific Evidence and Provenance Information Ontology (SEPIO) was developed to support the description of evidence and provenance information for scientific claims. The core model represents the relationships between claims, their lines of evidence, and the data items that comprise this evidence, as well as the methods, tools, and agents involved in the creation of these artifacts. SEPIO was initially developed to support the data integration and analysis efforts of the Monarch Initiative, where it provides a unified and computable representation of evidence and provenance metadata for genotype-phenotype associations aggregated across diverse model organism and clinical genetics databases. However, additional requirements were collected from diverse community partners in an effort to provide a shared community standard, with a core model that is domain independent and extensible to represent any type of claim and its associated evidence. In this report we describe the structure and principles behind the SEPIO model, and review its applications in support of data integration, curation, knowledge discovery, and manual and computational evaluation

of scientific claims. The SEPIO ontology can be found at <http://github.com/monarch-initiative/SEPIO-ontology/blob/master/src/ontology/sepio.owl>.

IT606:

Measuring the importance of annotation granularity to the detection of semantic similarity between phenotype profiles

Prashanti Manda, James P. Balhoff and Todd J. Vision

Abstract: In phenotype annotations curated from the biological and medical literature, considerable human effort must be invested to select ontological classes that capture the expressivity of the original natural language descriptions, and finer annotation granularity can also entail higher computational costs for particular reasoning tasks. Do coarse annotations suffice for certain applications? Here, we measure how annotation granularity affects the statistical behavior of semantic similarity metrics. We use a randomized dataset of phenotype profiles drawn from 57,051 taxon-phenotype annotations in the Phenoscape Knowledgebase. We compared query profiles having variable proportions of matching phenotypes to subject database profiles using both pairwise and groupwise Jaccard (edge-based) and Resnik (node-based) semantic similarity metrics, and compared statistical performance for three different levels of annotation granularity: entities alone, entities plus attributes, and entities plus qualities (with implicit attributes). All four metrics examined showed more extreme values than expected by chance when approximately half the annotations matched between the query and subject profiles, with a more sudden decline for pairwise statistics and a more gradual one for the groupwise statistics. Annotation granularity had a negligible effect on the position of the threshold at which matches could be discriminated from noise. These results suggest that coarse annotations of phenotypes, at the level of entities with or without attributes, may be sufficient to identify phenotype profiles with statistically significant semantic similarity.

IT701:

A Quality-Assurance Study of ChEBI

Hasan Yumak, Ling Chen, Michael Halper, Ling Zheng, Yehoshua Perl and Gai Elhanan

Abstract: Ontologies are important components of many health-information systems. The Chemical Entities of Biological Interest (ChEBI) ontology has become a standard reference for chemicals appearing in biological contexts. As such, assuring the quality of its content is imperative. In fact, ChEBI has a dedicated Web page at which errors and inconsistencies in its concepts can be reported. A study of the correctness of a random sample of ChEBI concepts is carried out. The results show that quite a large number of ChEBI concepts suffer from some kind of problematic modeling. For example, we found

that 15.5% of the sample concepts exhibited severe errors of commission, including incorrect hierarchical (is a) and lateral relationships. Errors of omission were also prevalent. The overall results of our quality-assurance (QA) study are presented. Suggestions for enhancing the QA processes in place for ChEBI are discussed.

IT702:

[To MIREOT or not to MIREOT? A case study of the impact of using MIREOT in the Experimental Factor Ontology \(EFO\)](#)

Luke Slater, Georgios Gkoutos, Paul Schofield and Robert Hoehndorf

Abstract: MIREOT is a mechanism for the selective re-use of individual ontology classes in other ontologies. Designed to minimise effort and to support orthogonality, it is now in widespread use. The consequences for ontology integrity and automated reasoning of using the MIREOT mechanism have so far not been fully assessed. In this paper, we perform an analysis of the Experimental Factor Ontology (EFO), an ontology which uses the MIREOT process to gather classes from a large range of other ontologies. Our study examines the effect of combining EFO with the ontologies it references by actually importing them into the EFO. We then evaluate the consistency and status of the combined ontologies. Through our investigation, we reveal that EFO in combination with all its referenced ontologies is logically inconsistent. Furthermore, when EFO is individually combined with many of the ontologies it references, we find a large number of unsatisfiable classes. These results demonstrate a potential problem within a major ontological ecosystem, and reveals possible disadvantages to the use of the MIREOT system for developing ontologies.

IT703:

[Semantic Digitization of Experimental Data in Biological Sciences](#)

Saurabh Raghuvanshi

Abstract: A major bulk of published experimental data, referred to as “Gold Standard” data, is available in a format that cannot be easily accessed by computers unless effectively curated. Most curation techniques bank on mining the text for information. Here we propose and demonstrate the efficacy of curating the experimental data itself. The data models facilitate digitization of the every aspect of the information associated with the experimental data. The models utilize several universally accepted ontologies as well as in-house developed alphanumeric notations for digitizing different aspect of the data. The data models have sufficient flexibility to address the extensive variability in experimental data. They have a very generic nature and can be used to curate and digitize experimental data from any organism. The digitized data is easily stored in a relational database management system and can thus be rapidly searched and integrated. These models have been successfully used to digitize data from over 20,000

experiments spanning over 500 research articles on rice biology. The entire dataset is available as a database entitled Manually Curated Database of Rice Proteins at www.genomeindia.org/biocuration.

IT704:

[Representation of parts within the Foundational Model of Anatomy ontology](#)

Melissa Clarkson

Abstract: As biomedical ontologies grow in size and complexity it is crucial to develop methods for detecting inconsistencies within ontologies. The Foundational Model of Anatomy (FMA) ontology represents knowledge of human anatomy, with structural organization provided by class and part relationships. Using a manual audit, I identify types of inconsistencies arising from class and regional part relationships for regions of the body and the parts of organs. Inconsistencies arise from both explicitly declared relationships and relationships that are implied by the lexical constructs of class names. The purpose of this work is to propose methods of structural organization and lexical consistency that will make the FMA more compatible with computational auditing and increase its usability.

IT705:

[A Realist Representation of Social Identity Data](#)

Amanda Hicks

Abstract: Social identities merit special treatment in realist ontologies. Their ontological status is unsettled, so we should model them in a manner that is agnostic with respect to their ontological status. Nevertheless, there is a clear criterion for determining whether a specific person has a particular identity, namely, whether that person asserts that they do. This social act forms the basis for a realist representation, not of social identities themselves, but of data about social identities. We report the representation of social identities in the Ontology of Medically Related Social Entities and show that it supports data integration and retrieval.

W14-03:

[The Plant Phenology Ontology for Phenological Data Integration](#)

Brian J. Stucky, John Deck, Ellen Denny, Robert P. Guralnick, Ramona L. Walls and Jennifer Yost

Abstract: Plant phenology the timing of life-cycle events, such as flowering or leafing-out has cascading effects on multiple levels of biological organization, from individuals to ecosystems. Despite the importance of understanding phenology for managing

biodiversity and ecosystem services, we are not currently able to address continent-scale phenological responses to anticipated climatic changes. This is not because we lack relevant data. Rather, the problem is that the disparate organizations producing large-scale phenology data are using non-standardized terminologies and metrics during data collection and data processing. Here, we preview the Plant Phenology Ontology, which will provide the standardized vocabulary necessary for annotation of phenological data. We are aggregating, annotating, and analyzing the most significant phenological data sets in the USA and Europe for broad temporal, geographic, and taxonomic analyses of how phenology is changing in relation to climate change.

W14-02:

The Biological Collections Ontology for linking traditional and contemporary biodiversity data

Ramona Walls and Rob Guralnick

Abstract: Biodiversity data comes from many sources, ranging from museum specimens to field surveys to genomic sequences. Domain specific standards provide vocabularies for many types of these data, but they do not fully support integrating data across methods, scales, and domains. The Biological Collections Ontology (BCO) was designed to bridge the terminology gap between traditional museum-based specimen collections and more contemporary environmental sampling methods, such as metagenomic sequencing, by providing a logically defined set of terms for biodiversity that map to standards such as the Darwin Core and Minimum Information for any Sequence. The BCO is expanding to encompass observational biodiversity data such as field surveys and taxonomic inventories. A key design principle of the BCO is to clearly distinguish the different types of processes involved in biodiversity data collection along with the inputs and outputs of those processes. The BCO has applications to plant biodiversity studies for linking herbarium specimens to sequence data, connecting trait data to specimens, and describing survey data.

BT101:

Cycles of Scientific Investigation in Discourse - Machine Reading Methods for the Primary Research Contributions of a Paper

Gully A. Burns; Anita de Waard; Pradeep Dasigi; Eduard H. Hovy

Abstract: We describe a novel approach to machine reading of the primary scientific literature. We treat a description of an experiment as a discourse, viewing a scientific corpus not merely into a collection of documents, but also an extended conversation formed by the collective set of experiments, their introductions and interpretations. This paper introduces this approach as a methodology called 'Cycles of Scientific Investigation in Discourse' (CoSID). In CoSID, we capture the central conceptual

structure of a paper as a series of nested reasoning loops, composed of passages in results sections, which describe individual research findings. We ground our work with a number of worked examples based on data from the MINTACT and Pathway Logic databases, and illustrate the idea in the context of machine-enable biocuration.

BT102:

Collaborative Workspaces for Pathway Curation

Funda Durupinar-Babur; Metin Can Siper; Ugur Dogrusoz; Istemi Bahceci; Ozgun Babur; Emek Demir

Abstract: We present a web based visual biocuration workspace, focusing on curating detailed mechanistic pathways. It was designed as a flexible platform where multiple humans, NLP and AI agents can collaborate in real-time on a common model using an event driven API. We will use this platform for exploring disruptive technologies that can scale up biocuration such as NLP, human-computer collaboration, crowd-sourcing, alternative publishing and gamification. As a first step, we are designing a pilot to include an author-curation step into the scientific publishing, where the authors of an article create formal pathway fragments representing their discovery- heavily assisted by computer agents. We envision that this “micro-curation” use-case will create an excellent opportunity to integrate multiple NLP approaches and semi-automated curation.

BT103:

Crowdsourcing Protein Family Database Curation

Matt Jeffryes; Maria Liakata; Alex Bateman

Abstract: We propose a novel method for crowdsourcing a protein family database. We discuss how we intend to identify novel groupings of proteins from user sequence similarity search, and how text mining will be applied to assist in annotation of these novel groupings, and more broadly as an enrichment of protein sequence similarity search results.

BT104:

Opportunities and challenges presented by Wikidata in the context of biocuration

Benjamin Good; Sebastian Burgstaller-Muehlbacher; Elvira Mitraka; Timothy Putman; Andrew Su ; Andra Waagmeester

Abstract: Wikidata is a world readable and writable knowledge base maintained by the Wikimedia Foundation. It offers the opportunity to collaboratively construct a fully open access knowledge graph spanning biology, medicine, and all other domains of

knowledge. To meet this potential, social and technical challenges must be overcome most of which are familiar to the biocuration community. These include community ontology building, high precision information extraction, provenance, and license management. By working together with Wikidata now, we can help shape it into a trustworthy, unencumbered central node in the Semantic Web of biomedical data.

BT201:

Text mining to enable routine personalized cancer therapy

Hua Xu

Abstract: Genomic profiling information is frequently available to oncologists, enabling targeted cancer therapy. Because clinically relevant genomic information is rapidly emerging in narrative data sources such as biomedical literature and clinical trials documents, there is a need for text mining technologies to support targeted therapies. In this talk, we will present two projects about developing text-mining tools to enable personalized cancer therapy, including 1) to identify molecular effects of drugs in biomedical literature, and 2) to create a knowledge base of cancer treatment trials with annotations about genetic alterations. We believe such tools would be valuable for physicians and patients who are seeking information about personalized cancer therapy, thus facilitating their decision making.

BT202:

Social Media Mining for Pharmacovigilance

Graciela Gonzalez

Abstract: N/A

BT203:

MutD – A PubMed Scale Resource for Protein Mutation-Disease Relations through Bio-Medical Literature Mining

Ravikumar Komandur Elayavilli; Majid Rastegar-Mojarad; Hongfang Liu

Abstract: A large amount of information about the role of gene variants and mutations in diseases is available in curated databases such as OMIM, ClinVar, and UniprotKB. However, much of this information remains 'locked' in the unstructured form in the scientific publications. Since manual curation involves significant human effort and time there is always a lag in the information between the curated databases and the literature. The recent findings published in the literature takes significant time to find its way into the curated knowledgebase. Text mining approaches can accelerate the process of assembling this knowledge from the published literature. However,

developing a text-mining system with semantic understanding capability in the biomedical domain is very challenging. In an earlier work, we described MutD, a literature mining system that extracts relationship between protein point mutation and diseases from bio-medical abstracts. In this abstract, we present access to a PubMed scale resource through a web interface that allows users to retrieve protein point mutation-disease relations extracted through biomedical literature mining.

BT204:

CancerMine: Knowledge base construction for personalised cancer treatment

Jake Lever; Martin Jones; Steven Jm Jones

Abstract: Knowledge of the relevant genomic aberrations that drive a particular cancer type is necessary to accelerate efficient interpretation of genomic data and enable large-scale endeavors in precision medicine. Currently, this field is limited by the lack of focused and scalable literature curation tools that can reliably capture the required information. Here we present a knowledge-base of genes that have been described in the literature as drivers, oncogenes or tumour suppressors with respect to a specific type of cancer. We have annotated a large body of literature which reports oncogenic aberrations using a custom designed annotation tool. We then applied VERSE, an in-house relation extraction tool, to catalogue driver mutations and illustrate the ability to build a useful resource for clinical interpretation of genomic data for personalized treatment approaches.

BT205:

Text Mining for Drug Development: Gathering Insights to Support Decision Making

Sherri Matis-Mitchell

Abstract: Drug discovery in Pharma R&D is an information driven process requiring many disparate bits of data from many different sources, both structured and unstructured. Text mining is the key methodology used to extract entities and relationships from unstructured text in the quest for the knowledge needed to bring a safe and effective drug to market and beyond. Much of the insight needed in early drug research to identify drug target to disease relationships and progress a potential drug target, comes from published literature and internal reports. Later stage drug development requires many additional sources of information including case reports, clinical trials, competitive intelligence and other diverse sources. In this publication, I will present 4 different use cases on how text mining is used to drive decision making in drug discovery and development and also how it can be used to identify patient insights from sources such as social media.

BT301:

NLP for the Institute: Developing and Deploying an NLP Capability to Accelerate Cancer Research

Aaron Cohen

Abstract: "It has been well documented that a great deal of data useful for medical research is present in clinical narrative text. There is perhaps less discussion about how often what was structured data at its origin has become inaccessible except in free text form. This problem is further compounded in tertiary care institutions, like the OHSU Knight Cancer Institute, where the entire history of a referred patient's condition may only be present in the electronic health record (EHR) as free text. At the same time, future medical advances, such as in cancer research, will require much more complete patient data than has been previously available. Such advances include the discovery of new cures, expanding early detection, and realizing the promise of precision medicine. Phenotype description and outcome characterization are two areas in particular where text sources could greatly supplement our current data. The OHSU Knight Cancer Institute has begun a program to create a natural language processing (NLP) capability to extract, store, and link data from free text sources at the patient level, and make this data available to researchers in a continuous, reusable, efficient and timely manner through services delivery from the Translational Research Hub (TRH). This talk will present the challenges, progress, and future goals of our program to build NLP capabilities that can help us use free text from the EHR to first support the transformation of cancer research with the hopes of positively impacting clinical care in the future."

BT302:

Annotations for biomedical research and healthcare -- Bridging the gap

Olivier Bodenreider

Abstract: "Characterizing protein products from various model organisms with Gene Ontology terms, indexing the biomedical literature with MeSH descriptors, and coding clinical data with ICD10-CM all constitute examples of annotation tasks, i.e., the extraction and summarization of knowledge related to a biological entity, article or patient, in reference to some controlled vocabulary or ontology. However, the annotations made in biomedical research and healthcare environments tend to rely on different terminologies and ontologies, making it difficult to reconcile these annotations for translational research purposes. We will discuss how terminology integration systems, such as the Unified Medical Language System (UMLS) and BioPortal, can help bridge the gap between annotations made by biomedical researchers and physicians, and argue that more efforts are needed to foster interoperability between the resources developed by these two communities."

BT303:

PubAnnotation: a public shared platform for scientific literature annotation.

Jin-Dong Kim

Abstract: "In the last decade, the technology for biomedical literature annotation made a significant progress in terms of accuracy and speed. Now, some annotation systems claim that they have reached a production level. However, there still remain critical issues which we believe hinder further progress of the community. Among them, a relatively well known issue is "interoperability" of annotation resources. We also recognize that the community is missing a general solution for "storage infrastructure". The talk will present the PubAnnotation project which aims at addressing these two issues. In the end, a new model for "sustainable shared tasks", which is implemented on PubAnnotation, will be introduced as well."

BT304:

BioCconvert: A Conversion Tool between BioC and PubAnnotation

Donald C. Comeau; Rezarta Islamaj Doğan; Sun Kim; Chih-Hsuan Wei; W. John Wilbur; Zhiyong Lu

Abstract: BioC is a simple XML data format for text, annotations, and relations. PubAnnotation is a repository of text annotations focused on the life science literature. A conversion tool between BioC XML and the JSON import / export format of PubAnnotation has been developed, BioCconvert. As a demonstration, the Ab3P gold standard abbreviation annotations are being made available through PubAnnotation.

W04-01

Ontology of the Organigram

B. Smith

Abstract: Basic Formal Ontology (BFO) is a domain-neutral top-level ontology designed to serve as the starting point for development of domain ontologies designed to support the consistent annotation not only of scientific research data but also of data arising through clinical practice, hospital administration, and regulatory oversight. In each of these areas data are generated relating to what are called deontic entities -- obligations, duties, contracts, permissions, consents, licenses, and so forth. I will sketch how we can understand entities of these sorts within the BFO framework.

W04-02

Relations between Institutional Roles and Deontic Roles in Biomedical Organizations

Otte J.N., Brochhausen M.

Abstract: Doctors, nurses, surgeons, and other healthcare professionals are bearers of institutional roles (e.g. physician role), and it is common to think of these roles as having deontic powers as parts. This is reflected in sentences like: "Part of my job involves an obligation to oversee patient care" and "Doctors at this hospital have the following privileges". The Document Act Ontology (d-acts; <http://purl.obolibrary.org/obo/iao/d-acts.owl>) enables us to represent how deontic powers are created by document acts. Deontic powers are realizable entities that we treat as deontic roles (e.g. obligor role). This means deontic powers are roles themselves. In this talk, we turn to the question of the relationships between institutional roles and deontic roles. We propose that institutional roles can have deontic roles as parts. We illustrate how different kinds of deontic powers give rise to different parthood relations, and we conclude by arguing that such relations can better capture the nature of organizational structure than the traditional reliance on institutional roles alone.

W04-03

An ontological study of healthcare corporations and their social entities

M.B. Almeida

Abstract: Healthcare corporations have a primary goal to provide high-quality health care to those who seek its assistance. The quality and safety of care provided by a healthcare corporation depends on many factors involving both medical and business decisions. One crucial factor in healthcare corporations function is the information management mainly performed through information systems, which process information for both clinical decision making and for fulfilling internal and external legal obligations. In this lecture, we propose a sketch of an ontology-based model for healthcare corporations with the aim of facilitating the coordination of the information systems involved in either medical or management activities. In order to accomplish this, we focus on three efforts: i) to shed some light on the ontological status of corporations; ii) to clarify the relations that exist between the corporation as whole and both its members and units; iii) to explain how an corporation administers its duties and responsibilities on behalf of people who compose it.

W04-04

Who's paying and who's graying? The organizations and roles associated with insurance policies, funding agencies, and national census data

A. Hicks & W.R. Hogan

Abstract: Many social roles are necessarily related to organizations. Employee roles and primary insured roles are just a few examples. However, the relations between such roles and organizations have yet to be systematically worked out in the context of BFO-based ontologies. As an initial step toward developing such a systematic representation, we present use-case driven representations connecting roles to organizations in the context of insurance policies, scientific grants, and U.S. National Census data in OWL/RDF.

W04-05

Dealing with social and legal entities in the obstetric and neonatal domain

F. Farinelli & M.B. Almeida

Abstract: OntONeo is an ontology for the obstetric and neonatal domain, which has been created to provide a consensus representation of salient electronic health record (EHR) data and in order to serve interoperability of the associated data and information systems. Regardless of medical specialty, in general, every EHR deal with data about the entities involved in a health care appointment. We observed the existence at least three actors: health care facility, the physician, and the patient. Besides dealing with the social entities that participating in a health care appointment, we must also deal with the characteristics that define them and the events that are involved. Here, we demonstrate the utility of ontologies of social entities in the obstetric and neonatal domain. We present how OntONeo is dealing with the material entities who perform or participates in a health care encounter. The definition of these actors plus their related characteristics and events will also contribute to turning on the interoperability of information among EHR from different specialties. OntONeo is being developed with an approach based on ontological realism and the principles of OBO Foundry, including reuse of reference ontologies. Among our reusable ontologies, we can mention for instance the OMRSE, d-acts, PATO and OBI.

W01-01

Big Data Visual Analysis

Johnson C.

Abstract: We live in an era in which the creation of new data is growing exponentially such that every two days we create as much new data as we did from the beginning of mankind until the year 2003. One of the greatest scientific challenges of the

21st century is to effectively understand and make use of the vast amount of information being produced. Visual data analysis will be among our most important tools to understand such large and often complex data. In this talk, I will present state-of-the-art visualization techniques, applied to important Big Data problems in science, engineering, and medicine.

W01-02

Topology-Driven Data Visualization of Large-Scale Tensor Fields

Zhang Y.

Abstract: Spatial-temporally varying simulation data sets are growing at a scale that traditional visualization techniques are inadequate to handle. In this talk, we review topology-driven techniques which focus on extracting the key (topological) features in 3D symmetric tensor fields, which can provide a more compact visualization of the data.

W01-03

Computer Vision for Next Generation Phenomics and Tree of Life

Todorovic S.

Abstract: To build the Tree of Life, scientists collect data on all heritable features – both genotypes (e.g., DNA sequences) and phenotypes (e.g., anatomy, behavior, physiology) for all living and extinct species. The collection of phenomic data for tree-building has lagged far behind the collection of genomic data. Advances in computer vision have the potential to change this situation. In this talk, I will present a computer vision system developed in our lab for extending phenomic matrices, and in this way building the Tree of Life. Rows of a phenomic matrix represent images of specimens belonging to various species of interest, and columns represent scores of their phenomic characters. Given a phenomic matrix where only few rows are manually annotated with character scores, our vision system extends the matrix row-wise by populating missing character scores of the remaining species in the matrix. The talk will present our experimental results on scoring phenomic characters in images of bat skulls, nematocysts, and leaves, available in the Morphobank and Bisque data repositories.

W01-05

Uncertainty analysis and visualization in Large-Scale Vector Fields

Zhang E.

Abstract: Vector fields are omnipresent in science, engineering, and medicine. Due to the big-data nature of simulated vector field data, uncertainty in the data can lead to

difficulties in their physical interpretations. In this talk, we will describe how Morse-decomposition can serve as a means to quantify and control the uncertainty in the data.

W01-04

Telling a genome's story graphically

Lewis S.

Abstract: Scientific research is inherently a collaborative task; a dialog among different researchers to reach a consensus understanding of the underlying biology. Information graphics facilitate this dialog because humans are visually wired and can absorb well-executed graphics more quickly than they can absorb the written word or columns of numbers. Visual representations communicate complex ideas quickly and clearly.

W11-01

Why We Need A Food Ontology

Damion Dooley

Abstract: The need to represent knowledge about food is central to many fields such as health, food safety, nutrition, food allergy, sustainable development, trade, ecosystems etc. Academic, national, provincial and departmental databases are all silos of food terminology and data models. Several resources and standards exist for indexing food descriptors however their content and architecture are not semantically and logically coherent. Here we present a unified approach to developing a Farm-to-Fork food ontology which will facilitate data sharing and interoperability between different health, regulatory, development and research communities worldwide.

W11-02

International Efforts in Creating Food Vocabularies

Robert Hoehndorf and Matthew Lange

Abstract: A standardised system for classifying and describing food makes it easier to compare data from different sources and perform more detailed types of data analyses. As such, there have been many agency-specific, project-specific and international efforts to create food vocabularies fit for different purposes. Here we describe the different existing and ongoing efforts.

W11-03

Sustainable food systems and food in ecosystems

Pier Luigi Buttigieg

Abstract: This brief talk will outline the need for a global food ontology to flexibly represent food across human and natural ecosystems. From an anthropocentric point of view, the sustainability and resilience of the global food system - including the sustainability of ecosystems and human-made networks which support it - should be closely interlinked with entities which realise food roles and the global policy objectives to secure food supply for all. From a more "natural" point of view, a truly global food ontology should be flexible enough to link taxa (including humans) to their consumers via the simultaneous realisation of prey, detrital, and food roles. This feature would provide a semantic basis to model food webs and, in combination with compositional inventories, nutritional profiles for ecoinformatics. These anthropogenic and natural perspectives will inevitably converge as a biospheric representation of trophic patterns emerges, a process which a flexible food ontology can greatly accelerate. Vitally, these aims will require coordination across multiple established and emerging ontologies to be feasible in the long term and a number of potential synergies with the Environment Ontology, the Agronomy Ontology, and the Sustainable Development Goal Interface Ontology will be proposed.

W11-04

Food composition and links to human health

Miguel Rodriguez Garcia

Abstract: Data on the composition of foods are essential for a diversity of purposes. It provides detailed sets of information on the nutritionally important components such as proteins, carbohydrates, vitamins and minerals. In this talk, we will describe our approach to employ ontologies to identify food components in recipes described in natural language. By combining several public data sources, we link the food components to chemical compounds and their physiological and pathological effects.

W11-05:

Food safety and infectious disease

William Hsiao

Abstract: Globalization of food manufacturing, distribution and consumption require food microbiology testing programs to perform high-throughput pathogen diagnostics within a short time frame. Whole-genome sequencing (WGS) can now assemble and type bacterial genomes in near real-time with improved resolution, making WGS a viable diagnostic tool during foodborne illness outbreak investigations. However, genomic data must be combined with epidemiological, clinical, laboratory and other health care data (contextual data) to be meaningfully interpreted for actionable interventions. Canada's Integrated Rapid Infectious Disease Analysis (IRIDA) project includes the development of a Genomic Epidemiology Application Ontology (GenEpiO),

with the development of standardized food vocabulary as a priority area. Standardized food descriptors are essential for data sharing between public health agencies and health responders, accreditation and reproducibility of WGS pipelines, source attribution and risk assessment.

W11-06

FoodON Use cases: Caution! Food Allergies Ahead

Emma Griffiths

Abstract: Millions of people worldwide live with food allergies, including all those at risk for life-threatening anaphylaxis. The lack of a standardized food vocabulary impacts food source risk assessment, food hazard control, consistent food allergy policy implementation and food-allergy research. The Canadian Healthy Infant Longitudinal Study (CHILD) examines causal factors of asthma and allergy during childhood development. The development of FoodON will benefit food allergy research by standardizing food descriptors across child cohorts, enable the correlation of food antigens with biological causation of immune response, and streamline guidelines for parents.

W11-07

IC3-Foods: An infrastructure for the next generation internet of food systems, food, and health.

Matthew Lange

Abstract: IC3-FOODS, The International Conference/Consortium/Center for Food Ontology, Operability, Data and Semantics, is a new effort at UC Davis, assembling ontological and semantic infrastructure components for next generation internet of food and health. It consists of 3 specific efforts:

- i) The International Conference for FOODS assembles stakeholders desiring to integrate data and informatics systems currently residing along the Environment ↔ Ag ↔ Food ↔ Diet ↔ Health knowledge spectrum, into the:
- ii) International Consortium of FOODS, which maintains membership of representative stakeholders from academia, industry and (non-)governmental organizations to guide research priorities and development trajectories carried out by:
- iii) The International Center for FOODS, whose mission consists of hosting the IConference--FOODS, administering the I-Consortium-FOODS, and designing, assembling, and coordinating ontological and infrastructure underpinnings for them.

-----END-----