Informed Consent Ontology (ICO):

A BFO-based Representation of Informed Consent

Yu Lin\*, Marcelline R. Harris, Frank J. Manion, Elizabeth Eisenhauer, Bin Zhao, Wei Shi,   
Alla Karnovsky, Yongqun He\*

University of Michigan Medical School, Ann Arbor, MI 48109, USA

*Abstract* — As part of the consenting process in human subjects research, potential participants receive information about the purpose of the study, potential risks and benefits of participation, their rights, and the procedures to be undergone as part of the study. If the invidual decides to particapte, an informed consent document is signed and preserved as a record of voluntary participation in the research study. Following the OBO Foundry principles, the Informed Consent Ontology (ICO) was developed as a BFO based community-oriented ontology to logically represent the terms and term relations related to the informed consent process and content. Currently, ICO contains xxx terms including xxx ICO-specific terms and xxx terms imported from existing ontologies. Possible ICO use cases are introduced. The ICO ontology is available on the website: <http://ico-ontology.googlecode.com/>.

Keywords— Informed consent; ontology; ICO; OBO Foundry, Basic Formal Ontology (BFO), OBI ontology

# Introduction

The document that people sign when they agree to participate in research (informed consent) is in fact complex and reflects voluntary decision making of participants in research, and presumes that these participants are fully informed about the research and its risks (refs.). Different research studies such as interventional, observational, and more recently, genetic studies based on biobanked materials present different risks and benefits. The informed consent document contains essential information indicating that an individual has a clear understanding of the facts, implications, and future consequences of the research activity. The informed consent documents also include specifications on the permitted uses of and restrictions on the use of materials and personal information. Consent forms often must offer human subjects different options for participation (tiered consent). Assuring that computer-based systems “know” the permitted uses and restrictions is critical for conducting ethically responsible research.

An ontology is a set of human- and computer-interpretable terms and term relations in a specific knowledge domain. The Open Biological and Biomedical Ontologies (OBO) Foundry has established a set of principles, including openness, collaboration, and uses of common relations, for biomedical ontology development [[1](#_ENREF_1)]. This set of principles has been widely accepted by the ontology development community. Most OBO Foundry ontologies use the Basic Formal Ontology (BFO) [[2](#_ENREF_2)] as an upper level ontology that is grounded in a realism-based approach to models [[3](#_ENREF_3)]. As an upper level ontology, BFO assists in the organization and integration of other domains. BFO has been used as an upper level ontology for over 100 biological and biomedical ontologies. The Ontology for Biomedical Investigations (OBI) is a BFO based ontology co-developed and supported by some 20 different disciplinary communities. OBI provides a set of logically defined terms covering a broad range of experimental conditions, biomedical processes, and data analysis methods [[4](#_ENREF_4)].

In this study, we present an effort to develop a community-oriented Informed Consent Ontology (ICO) by aligning it with the BFO and OBI. Because OBI contains terms related to informed consent that are fully modeled in the context of BFO, we identified and imported these OBI terms to ICO as a starting point for the structure of the informed consent ontology. More terms and hierarachies of ICO were later generated based on existing informed consent templates and use cases. The rationale, design pattern, and use cases of ICO are presented in this manuscript.

In what follows, italic terms in the text represent specific ontological terms.

# Methods

## Ontology editing and representation

ICO is represented as in the W3C standard Web Ontology Language (OWL2) (<http://www.w3.org/TR/owl-guide/>). The Protégé-OWL editor 4.2 was used for ICO ontology editing.

Terms from existing ontologies (*e.g.*, OBI) were imported to ICO using the tools Ontodog [[5](#_ENREF_5)] and OntoFox [[6](#_ENREF_6)]. New ICO specific terms were generated using new ICO IDs with the prefix of “ICO\_” followed by seven auto-incremental digital numbers.

ICO was developed using a combination of top-down and bottom-up approaches as described below.

## Top-down ICO development

In our top-down ontology development approach, we first imported all of the BFO 2 as the ICO upper level ontology [[2](#_ENREF_2)]. To identify informed consent related terms in OBI, the software program Ontodog [[5](#_ENREF_5)] was first used. Ontodog provides an Excel template that lists all OBI terms. Each of the OBI terms was screened, and those relevant to informed consent were checked. The relevance to informed consent was measured by the direct or indirect association of an OBI term to commonly used informed consent templates in the areas of clinical research trials, including those clinical studies resulting in storage and analysis of specimens from human subjects. In this process, we have also identified many relevant terms in the Information Artefact Ontology (IAO) (Ref.), for example, *IAO:document* (IAO\_xxxxx). Informed consent forms are classified in ICO as a kind of *IAO:document*. Many other terms such as various informed consent form components are also classified as information artefact. Therefore, the inclusion of these IAO terms would help future ICO specific term generation. Similarly, many relation terms from the Relation Ontology (RO) [[8](#_ENREF_8)] were included in OBI and might also be important for ICO modeling. After filling out the Ontodog Excel template as instructed by the program, the web-based Ontodog execusion (?) step was performed to generate an OBI subset, which was then imported to ICO. The ICO subset maintains the URIs, annotations, class hierarchies, and axioms associated with the terms checked for importing in the Ontodog Excel template. Since many additional OBI terms were later found relevant to ICO, the OntoFox program [[6](#_ENREF_6)] was applied to import these individual OBI terms to ICO.

After the top level structure was generated, we added many terms under imported ontology classes. For example, we have added the ICO term ‘*informed consent form*’ (ICO\_0000001), different types of informed consent forms, ‘*part of informed consent form*’ (ICO\_000000xx), and different types of informed consent form parts. New ICO terms were also created for generating and approving the informed consent forms. Terms related to biospecimen storage and processing complying with the informed consent agreements were also generated.

Another top-down ICO development practice was the ontological modeling of informed consent workflows corresponding to different stages of informed consent studies as described in the Results section.

## Bottom-up ICO Term Expansion

After the initial top-down ICO development, the ICO development team felt overwhelmed with the possibly large number of informed consent related terms to represent at the ontology term. To better identify more terms for ICO inclusion, we adopted a bottom-up approach for further ICO term expansion. The bottom-up development approach expanded ICO based on content derived from different informed consent templates. The specific steps in our bottom-up development are described below:

### Identifying candidate terms from various resources

We manually identified and extracted a list of candidate term terms from three informed consent form templates used in the University of Michigan (UM): The template used by the Institutional Review Board of the UM Medical School (IRBMED), the Health Sciences and Behavioral Sciences Institutional Review Boards (IRB-HSBS) template, and Biolibrary informed consent. Informed consent templates from outside of UM such as the World Health Organization (WHO) have also been analysed. By using templates developed for different broad areas of research the ICO ontology developers were able to provide preferred definitions and individual candidate terms applicable to these areas.

### Categorizing candidate ICO ontology terms

After the list of candidate terms were selected, these candidate terms were mapped to preferred textual definitions from a series of existing vocabularies, including the NCI Thesaurus (Ref.), OCRe [[10](#_ENREF_10)], OBI, etc. (Question: Is this list of vocabularies complete?). These terms were then clustered into 32 categories by aligning their textual definitions as possible. For instance, category ‘authorization’ was assigned to terms as ‘*authorization for medical records release*’, ‘*authorization documentation*’ or ‘*authorization*’. [Supplementary file?] Those categories were then grouped into 14 broader topics (or modeling units), forming the subsets of possible terms to be included in ICO. Clustering and categorizing ontology terms was performed by the ontologists based on the preferred definitions. The categorized candidate ontology terms were implemented into a relational database to facilitate the review process. During this process, terms were classified according their abundance in the templates, their overall relevance to informed consent, and the major topic areas the term applies to (need example here). (**Question:** Do we want to discuss Elizabeth’s use of UMLS semantic groups here?)

### ICO specific term addition, revision, and validation

New ICO-specific terms were asserted under parent classes derived from the upper (BFO) or middle level (OBI) ontologies using the Protégé-OWL editor. If appropriate, logical axioms are defined to provide restrictions and reasoning powerful to these terms.

\* Co-Corresponding authors.

## ICO licensing, ontology deposit, and visualization

ICO is released under creative commons by 3.0 License and has been deposited into Ontobee (the default OBO Foundry ontology dereferencing software program) [[7](#_ENREF_7)]: <http://www.ontobee.org/browser/index.php?o=ICO>. ICO has also been deposited into NCBO BioPortal (note: provide the link) Both Ontobee and BioPortal provide interactive ontology visualization.

## ICO-specific data query and analysis:

An ICO-specific website was created out of our collaborative study: <http://iconect.hegroup.org/>. In this website, we generated two ICO web query interfaces. One is based on the SPARQL Protocol and RDF Query Language (SPARQL) at <http://iconect.hegroup.org/sparql/>. Another is based on a web form query that does not require knowledge of SPARQL.

# Results

## BFO-aligned ICO hierarchy:

ICO is designed to be aligned with the BFO and developed by following the OBO Foundry principles [[1](#_ENREF_1)]. The importing of BFO and selected OBI and IAO terms provides a basic syntaxtic and semantic framework for further ICO development. Such a strategy makes ICO seamlessly integrated with over 100 other biological and biomedical ontologies developed using the same strategy. Many of the other ontologies, such as xx and xx, are clinic-oriented. Therefore, ICO and these other ontologies can efficiently share terminological terms and relations for more profound clinical and translational research.

Fig. 1 demonstrates the top level hierarchical structure of the ICO ontology and key ICO ontology terms. The structure of BFO is divided into two major branches: continuant and occurrent. The continuant branch represents time-independent entities such as material entities, and the occurrent branch represents time-related entities such as time regions and processes. Informed consent related terms are ultimately placed under continuant and occurent. For example, identified participants are under continuant branch: *inveistgator (*[*ICO\_0000045*](http://www.ontobee.org/browser/rdf.php?o=ICO&iri=http://purl.obolibrary.org/obo/ICO_0000045)*),* *human subject* (ICO\_0000073), *informed consent form (*[*ICO\_0000001*](http://www.ontobee.org/browser/rdf.php?o=ICO&iri=http://purl.obolibrary.org/obo/ICO_0000001)*)*, and different sections of informed consent form such as the signature section of the form*. Likewise,* identified processes are placed under occurent: *study design,* *informed consent form design (ICO\_0000002)*, *approval of informed consent form(ICO\_0000004)*, *human subject enrollment (OBI\_0600004)* (Fig. 1)*.*

**Fig. 1. ICO ontology hierarchy.**

## Modeling informed consent workflows

Using the ICO framework, different informed consent workflows were ontologically modeled. Based on study scenarios, the major processes involved in informed consent were divided into three categories: (i) pre-informed consent processes, (ii) informed consent processes, and (iii) post-informed consent processes.

### Pre-informed consent processes

During the pre-informed consent processes, an investigator first designs a study involving human subjects. An informed consent form is then designed and submitted to a human research Institutional Review Board (IRB) organization for approval. After reviewing, the informed consent study and form are approved. A final approved informed consent form is generated. Human subject candiates can submit their applications though paper form or electronic form. Based on the eligibility criteria, a list of eligible candicates are selected (Fig. 2).



**Fig. 2. Obtaining informed consent form for a clinical study.**

### Informed consent processes

In the informed consent process, an investigor or research assistant would explain to a human subject candidate about the informed consent study and the details from the informed consent form. Different dimensions of the consenting are provided, for example, where the candidate agrees on the protocol of the study (*e.g.*, duration, time, and visits), the benefits and risks of their participation, how their samples and the data generated from their samples will be processed and used, and the confidentiality of data collected as part of the study.

After communications, the candidate or legal guardian of the candidate will sign the informed consent form. The signed consent form is a legal representation of the consent by the candidate to participate in the study (Fig. 3).



**Fig. 3. Informed consent process.**

### Post-informed consent processes

In the post-informed consent processes, different participants of the study, including coordinator, investigator, research team member, and human subjects, will work together in the pre-designed study. Human samples collected as part of the study are typically stored in a biobank. Data collected through therapy, clinical intervention, discussion, or observation are typically stored in a database of some sort, be it an electronic health record, clinical trial management system, or some other system In any case records from these systems, and the biobank if applicable, are typically associated with details of the informed consent status. Note that it is possible that a human subject will not complete the study due to a variety of reasons including, but not limited to, study suspension, study termination, or subject withdrawal (Fig. 4).



**Fig. 4. Study execution according to informed consent agreement.**

The graphic modeling of different processes during the whole informed consent study (Fig. 2-4) provides the scope and coverage of the ICO and supports the ICO further development.

## Results from bottom-up ICO development

## As described in the Methods section, a time-intensive effort in our ICO ontology development process was spent on identifying ICO candidate terms and their categorization. As a result, we have identifed xx informed consent related terms and represented them in ICO.

## Statistics of ICO

Table 1 shows the statistics of the ICO. As of April 28, ICO contains xxx ontology terms, including xxx terms imported from xx existing ontologies, and xxx terms having ICO specific namespace (Table 1).

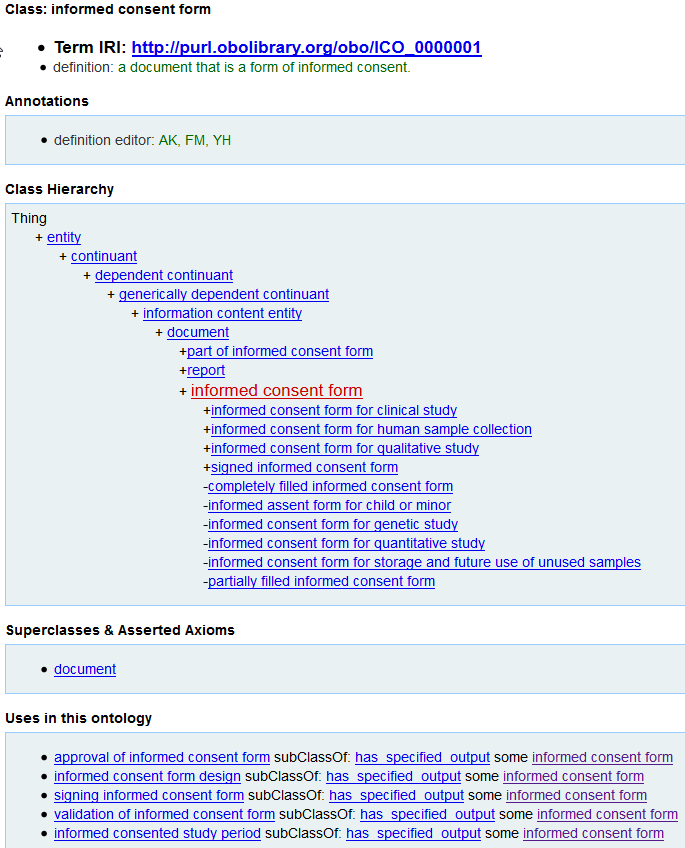
**Table 1: Statistics of ICO**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ontology | Classes | Object  Properties | Annotation Properties | *Total* |
| OBI | 135 | 8 | 2 | 145 |
| ICO | 71 | 1 | 0 | 72 |
| BFO | 22 | 38 | 0 | 60 |
| IAO | 16 | 2 | 11 | 29 |
| NCBITaxon | 7 | 0 | 0 | 7 |
| PATO | 6 | 0 | 0 | 6 |
| CHEBI | 5 | 0 | 0 | 5 |
| CARO | 2 | 0 | 0 | 2 |
| UBERON | 2 | 0 | 0 | 2 |
| OGMS | 1 | 0 | 0 | 1 |
| RO | 0 | 1 | 0 | 1 |
| GO | 1 | 0 | 0 | 1 |
| owl | 1 | 0 | 0 | 1 |
| CL | 1 | 0 | 0 | 1 |
| No Prefix | 0 | 0 | 4 | 4 |
| ***Total*** | **270** | **50** | **17** | **337** |

**Note**: this table may not include the most updated information. It will be updated before paper submission.

## ICO term display and query:

Fig. 5 provides an example of how ICO term is displayed in Ontobee [[7](#_ENREF_7)]. In this example, the ICO ‘informed consent form’ (ICO\_0000001) represents a form of informed consent. Different types of informed consent form are displayed in here.



**Fig. 5. Example of ICO display in Ontobee.**

## ICO web queries.

Most ontology-based web query is based on SPARQL. By following this trend, we have generated an SPARQL web interface: <http://iconect.hegroup.org/>. A pitfall for the SPARQL-based query is that it requires SPARQL programing, which cannot be done without sufficient programming training. Thereofre, the SPARQL-based query is not user-friendly to those who do not know how to program.

To provide the query probability for those who do not know SPARQL programming, we have also developed a web query interface using the traditional web form format: <http://iconect.hegroup.org/webQuery.php> (Fig. 6).



**Fig. 6. ICO web query without SPARQL programming requirement.**

## ICO use cases:

ICO can potentially be applied for different use cases. In our ICO development, the following use cases have been identified and examined for possible ICO applications:

### Automatic informed consent template generation

To automatically generate informed consent forms, it is required that the general information of each section of a form is predefined to comply with specific IRD regulations. It is noted that different institutes may have slightly different IRD regulations, the common features for a specific type of informed consent, such as a specific biospecimen collection and usage, may be recorded in ICO. Combined with software programs and components that are not necessarily recorded in ICO, it is possible to automatically generate electronic informed consent forms or templates for clinical uses.

It is also noted that such an ICO-based tool will not replace domain expert examination and legal approval. However, it may provide a framework that supports part of the electronic consent form generation effort.

### Informed consent validation

It is likely that electronic consenting program will be debeloped to facilitate electronic consent efforts. Such an electronic consenting program may be developed to check: (i) Is an informed consent complete? Or: Did the patient really understand it or not? (ii) Is the human subject candidate eligible? and (iii) Can the data of the human subject be released?

To achieve these goals, a computer- and human-interpretable terminology should be used. ICO is targeted to be such a controlled, logically defined terminology system. In addition, ICO may logically represent and specify exclusive and inclusive criteria for human subject eligibility. Formatted as OWL, ICO supports logical reasoning and thus might be eventually used to support informed consent validation.

### Biospecimen storage, processing, and data release

A typical type of clinical or translational research involves the extraction and usage of biospecimen from human subjects. Different biospeciment may be stored and processed differently compliant with the informed consent. In a research oriented university hosptical such as the one in the University of Michigan, hundreds of informed consent forms might have been generated. Correspondingly, different biospecimen have been stored in biobanks and processed. However, it has often become a huge challenge on how to query and possibly reuse the biospecimen stored in the biobanks.

Since we have been able to represent many informed consent forms, we are exploring how to use ICO as a possible hub to connect to different clinical studies associated with different informed consent forms. By doing so, we can possibly develop an ICO-based data management system that can integrate different biobank data from our studies. Typical use case questions we can ask include: (i) Can we retrieve all possible candidates that meet condition? (ii) Can the genetic information of a patient be released? (iii) Where an investigator can find biospecimen needed for a study under a compliant informed consent?

The ICO ontology is to be used as a central component for an ontology-based data management system for a Genomic DNA BioLibrary being developed by the Michigan Institute for Clinical and Health Research (MICHR; <https://www.michr.umich.edu/>). As a member of the national Clinical and Translational Science Award (CTSA) consortium, MICHR promotes integrated ontology-based data storage, sharing, and analysis. We also look forward to collaborating with interested users and developers around the USA and worldwide to further develop our system.

# Discussion

The informed consent is one of the fundamental pillars of research involving ethically responsible human subjects. However, currently no BFO-based informed consent research has been reported. The report of the first BFO-based ICO would strongly support the clinical informed consent biomedical research. By aligning ICO with BFO, we will be able to use seamlessly align and use all other BFO-based ontologies. This will bring us power in better representing, sharing, and analyzing the data related to informed consent.

Many terminology systems contain informed consent-related terms associated with clinical studies. For example, many systems inside the Unified Medical Language System (UMLS) [[9](#_ENREF_9)], including the National Cancer Institute (NCI) Thesaurus (ref.), HL7v3 (ref.), and SNOMED (ref.), include many useful terms and term relations associated informed consent processes. The newly developed Ontology of Clinical Research (OCRe) [[10](#_ENREF_10)] and Permission Ontology [[11](#_ENREF_11)] introduce new terms. ThePermission Ontology focuses on representation of permissions (*e.g.*, being recontacted, reusing and sharing data, etc) and obligations (*e.g.*, destroying bio-samples and records in case of withdrawal). However, non of these systems use BFO as the upper level ontology and follow the OBO Foundry principles.

Besides the previously known efforts, such as OCRe and UCSD’s permission ontology, two more groups are working on informed consent related web application or eConsent. One is d-act ontology and biobank ontology from Drs. Bill Hogan and Mathias Brochhausen’s groups in UAMS. UAMS’s work is BFO-based, and their work including: 1) OMIABIS (Ontologized MIABIS (OMIABIS) , the Minimum Information About Biobank data Sharing (MIABIS) (ref.); 2) document act ontology (<http://purl.obolibrary.org/obo/iao/d-acts.owl>) (ref.); and 3) OMRSE (Ontology of Medically Related Social Entities) (ref.). We look forward to working with them and the whole community on integrating these ontologies.

Nowedays more and more clinical requirements are fulfilled with electronic formats. We anticipate that the era of electronic consenting is coming soon. To meet this need, the community-based open source ICO will likely become a central component. Different ICO use cases can be developed. For example, since the basic information is recorded in the informed consent forms, it is possible to use the contents specified in the informed consent forms as the linkages for efficient query of specimen stored in biobank. More use cases have been introduced in this article. It is also noted that many of these use cases cannot be easily achieved without collaboration with other ontology teams, software development groups, and domain experts in the clinical and legal regulation departments. However, these use cases will promote us to further develop ICO for its final successful usage in clinical fields.

##### Acknowledgment *(Heading 5)*

We thank Drs. Nicholas H. Steneck and Blake J. Roessler for their valuable discussions and feedback. This research is supported by a University of Michigan MCubed project.

##### References

[1] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S. A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis, "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration," *Nat Biotechnol,* vol. 25, pp. 1251-5, Nov 2007.

[2] P. Grenon and B. Smith, "SNAP and SPAN: Towards Dynamic Spatial Ontology," *Spatial Cognition and Computation,* vol. 4, pp. 69-103, 2004.

[3] B. Smith and W. Ceusters, "Ontological realism: A methodology for coordinated evolution of scientific ontologies," *Applied Ontology,* vol. 5, pp. 139-188, 2010.

[4] R. R. Brinkman, M. Courtot, D. Derom, J. M. Fostel, Y. He, P. Lord, J. Malone, H. Parkinson, B. Peters, P. Rocca-Serra, A. Ruttenberg, S. A. Sansone, L. N. Soldatova, C. J. Stoeckert, Jr., J. A. Turner, J. Zheng, and O. B. I. consortium, "Modeling biomedical experimental processes with OBI," *J Biomed Semantics,* vol. 1 Suppl 1, p. S7, 2010.

[5] J. Zheng, Z. Xiang, C. J. Stoeckert, Jr., and Y. He, "Ontodog: a web-based ontology community view generation tool," *Bioinformatics,* Feb 1 2014.

[6] Z. Xiang, M. Courtot, R. R. Brinkman, A. Ruttenberg, and Y. He, "OntoFox: web-based support for ontology reuse," *BMC Res Notes,* vol. 3, p. 175, 2010.

[7] Z. Xiang, C. Mungall, A. Ruttenberg, and Y. He, "Ontobee: A linked data server and browser for ontology terms," in *The 2nd International Conference on Biomedical Ontologies (ICBO)*, Buffalo, NY, USA, 2011, pp. Pages 279-281 [http://ceur-ws.org/Vol-833/paper48.pdf].

[8] B. Smith, W. Ceusters, B. Klagges, J. Kohler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse, "Relations in biomedical ontologies," *Genome Biol,* vol. 6, p. R46, 2005.

[9] O. Bodenreider, "The Unified Medical Language System (UMLS): integrating biomedical terminology," *Nucleic Acids Res,* vol. 32, pp. D267-70, Jan 1 2004.

[10] I. Sim, S. W. Tu, S. Carini, H. P. Lehmann, B. H. Pollock, M. Peleg, and K. M. Wittkowski, "The Ontology of Clinical Research (OCRe): An informatics foundation for the science of clinical research," *J Biomed Inform,* Nov 13 2013.

[11] A. Grando and R. Schwab, "Building and evaluating an ontology-based tool for reasoning about consent permission," *AMIA Annu Symp Proc,* vol. 2013, pp. 514-23, 2013.