

FORCE11 - DCIP Executive Meeting

November 4, 2015 08:00 AM Pacific Time

Attendees: Tim Clark, Jeffrey Grethe, Stephanie Hagstrom, Maryann Martone, Jo McEntyre, Joan Starr

Regrets: Martin Fenner, Carole Goble, Simon Hodson

AGENDA

1. Identifier minting
 - a. N2T approach - John Kunze
 - i. John could not be on the call; we will defer presentation to next week.
 - b. We need to be able to mint identifiers
 - i. Many biomedical repositories do not have globally resolvable identifiers, they just have accessions
 - c. We had talked about identifiers.org.
 - i. It is not as fully supported as a complete service – it is in transition from a research project.
 - ii. It is not as secure as DOIs in terms of its maturity.
 - iii. We want journals to make data citable so we don't want identifier schemas to disappear.
 1. We need a URL and a curie type mechanism to make data citable.
 2. We probably don't need to get overly obsessed with identifiers at this time.
 - iv. Does the European Nucleotide Archive (ENA) have a globally resolvable identifier?
 1. They have an accession number
 - v. Could a curie-like approach be a possible step 1 for this pilot?
 1. It is a tried and trusted current method
 2. We could take a similar approach as ORCID – they can take a PMID, a curie type item, a DOI or URL
 - a. Jo will find out if ORCID has written about this approach – would be useful as a reference
 - i. **Action:** Maryann or Joan will try to follow up with Laurel Haak at the ORCID meeting to ask their guidelines
 - b. It would be beneficial for the pilot to align with current methods and to explain our reasoning
 - vi. The real opportunity with this pilot is to discuss with publishers what the difficulties right now with citing data
 1. This pilot is attempting to handle too many variables
 - a. We need to define what the pilot will do and make sure that it is a measurable outcome

2. The pilot is really about the publishers but we did propose that we would lay a path for the repositories to take steps to achieving JDCP support (through levels – bronze, gold, etc.)
 - a. The level achievements may not be very constructive – could alienate repositories that only receive a bronze rank
 - i. Our task is not to judge the repositories but offer them the opportunity to have their data appropriately cited by implementing our approach
 - b. RRIDs posed two simple questions: “how many papers are using this particular resource?” and “are the resources used identifiable by the authors?”
 - i. This allowed the project to achieve its original aims
 3. What publishers and repositories have signed up?
 - a. Elsevier, PLOS, BioMed Central, Nature, Faculty of 1000, IOS Press
 - i. One journal from each of these publishers would be sufficient
 4. **Action:** Reduce pilot complexity and establish achievable goals
2. Repositories list
- a. w/ BioSharing & publishers; selecting repositories to work with
 - i. BioSharing is working with publishers to recognize certain repositories as the right ones to use to deposit data
 1. Tim suggested to Susanna that it would be useful to have a merged list that identifies a repository and what publisher’s list it’s in (a union of the publishers’ lists) – how should we deploy this resource?
 - a. It would be useful to have a common list that people are not obliged to use but exists as a possible resource
 - b. Should consider what each publisher’s criteria is for adding repositories to their lists
 - c. **Action:** Tim will send this list to Jo to use during her workshop
 - d. Members of Susanna’s group have joined the DCIP working group (including Susanna herself), how do we want to handle the addition of new members?
 - i. **Action:** we need to freeze working group membership except through invitation
 - b. Regarding Protein Data Bank’s citation practices
 - i. Chun-Nan Hsu (UCSD, bioCADDIE Pilot Project 3.2 Co-Leader) performed citation analysis of PDB and encountered many problems trying to figure out how many times PDB has been cited and what data sets were cited
 1. One of the types of use cases we could serve, is his

2. It took Chun-Nan four months to perform a citation analysis which resulted in being incomplete
3. Nature Scientific Data article
 - a. **Action:** Tim will review it once more to make sure our repository interaction aligns with today's discussion
 - i. After reviewing it, Tim will send it out for a last call for revisions
4. Group lead transition through December
 - a. Tim is on medical leave for surgery until the end of 2015
 - b. Maryann and Joan will lead the group in the interim