



Bill Anderson <wlanderson@gmail.com>

Re: [IAO] publication and journal article

10 messages

Larry Hunter <Larry.Hunter@ucdenver.edu>

Fri, Feb 17, 2012 at 8:58 AM

To: jie zheng <zhengj2007@gmail.com>

Cc: information-ontology <information-ontology@googlegroups.com>

I argued 3 years ago that report ought to be obsoleted and, in general, replaced by 'document', which is a parent of 'publication' and 'publication about an investigation'. Alan complained that this would break OBI in some way, and it has been sitting there ever since. I agree with you completely. Get Alan to fix OBI so we can finally close this out.

Larry

On Feb 16, 2012, at 3:26 PM, jie zheng wrote:

> I think journal article is a kind of publication. Current in IAO,
> journal article is a subclass of report. Publication is the sibling of
> report. Two concepts have overlapping. If a report is accepted by a
> publisher, it will be a publication. Also, journal article is a kind
> of publication and a kind of 'publication about an investigation'. Due
> to this multiple inheritance, I think 'publication' and 'publication
> about an investigation' should be defined by N&S.

>

> --

> information-ontology@googlegroups.com

> To change settings, visit

> <http://groups.google.com/group/information-ontology>

--

information-ontology@googlegroups.com

To change settings, visit

<http://groups.google.com/group/information-ontology>

Mathias Brochhausen <mbrochhausen@googlemail.com>

Fri, Feb 17, 2012 at 9:11 AM

To: information-ontology <information-ontology@googlegroups.com>

Hi,

I agree this should really be fixed. The way it is now is monstrous and causes me problems each time I use IAO. I strongly agree with Jie that if

we have stuff like "publication" it needs to be defined classes.

Or we need to stick to represent one partition instead of many (which is not a strategy that seems to have been used in IAO, at all).

If there is anything I can do to help this, let me know.

Best,

Mathias

PS:

There is another open issue regarding "narrative object" and "textual entity" that has already been raised by Larry. That should be resolved, too.

[Quoted text hidden]

Alan Ruttenberg <alanruttenberg@gmail.com>

Fri, Feb 17, 2012 at 10:15 AM

To: Larry Hunter <Larry.Hunter@ucdenver.edu>

Cc: jie zheng <zhengj2007@gmail.com>, information-ontology <information-ontology@googlegroups.com>

On Fri, Feb 17, 2012 at 9:58 AM, Larry Hunter <Larry.Hunter@ucdenver.edu> wrote:

>

> I argued 3 years ago that report ought to be obsoleted and, in general, replaced by 'document', which is a parent of 'publication' and 'publication about an investigation'. Alan complained that this would break OBI in some way, and it has been sitting there ever since. I agree with you completely. Get Alan to fix OBI so we can finally close this out.

It would be Jie that would fix OBI, and it would be Bjoern who would be a stakeholder I would ask for approval to remove report. Reports are defined by an objective (though not logically yet) and are more specific than document. Journal article is too broad for what we work with - I suggest scientific journal article or academic journal article, as these would be more specific (and have different criteria). We don't say anything about peer review, which I also think is wrong. IMO, the asserted hierarchy should be according to the intention/nature of the information, and the inferred hierarchy by processes that are incidental to the content. Publication is a case of the latter, report a case of the former.

Here's what I wrote Jie and forgot to cc.

Is a journal article a journal article before it is published? If after, ok. publication should be a process and publication defined as exactly the specified output of that process.

The editor note on journal article says we should deprecate it in favor of documentation. The definition is better on report, and can be improved given Darren's editor note on report. Scientific reports would also seem to satisfy the def for data item "a data item is an information content entity that is intended to be a truthful statement about something (modulo, e.g., measurement precision or other systematic errors) and is constructed/acquired by a method which reliably tends to produce (approximately) truthful statements."

Responded to Jie but forgot to cc:

So perhaps

Document: Collection of information content entities intended to be understood as a whole

Publication: =def A document that is specified output of some publication process

Report: a document assembled by an author for the purpose of providing information for the audience. A report is the output of a documenting process and has the objective to be consumed by a specific audience. Reports are complete documents or parts of documents.

Scientific Journal article =def a report consisting primarily of data items that is specified output of a publication process (and hence publication)

Maybe:

Asserted:
Document
Report

Defined:
Publication (inferred to be document)
Scientific Journal Article (asserted document, inferred to be publication)

I have a feeling this needs more thought.

[Quoted text hidden]

Alan Ruttenberg <alanruttenberg@gmail.com>

Fri, Feb 17, 2012 at 10:25 AM

To: Mathias Brochhausen <mbrochhausen@googlemail.com>

Cc: information-ontology <information-ontology@googlegroups.com>

On Fri, Feb 17, 2012 at 10:11 AM, Mathias Brochhausen
<mbrochhausen@googlemail.com> wrote:

> Hi,
>
> I agree this should really be fixed. The way it is now is monstrous and
> causes me problems each time I use IAO. I strongly agree with Jie that if
> we have stuff like "publication" it needs to be defined classes.
> Or we need to stick to represent one partition instead of many (which is
> not a strategy that seems to have been used in IAO, at all).
>
> If there is anything I can do to help this, let me know.
>
> Best,
> Mathias
>
> PS:
> There is another open issue regarding "narrative object" and "textual
> entity" that has already been raised by Larry. That should be resolved,

Narrative object currently says set of propositions. I think that's not right. The thing that makes it narrative is that it is an ordered list of propositions. This distinguishes it from a list of top hit genes. I'd be happy to have it be a subtype of textual entity.

Textual entity exposes a third axis of classification, namely by encoding. So we have classification by:

Processes related to copying/reviewing and other processes that don't change the essentials of what the ICE means
Encoding: textual versus image versus music versus voice versus sign language
Purpose/Objective: to explain something to a certain kind of audience, to be a record of a license/contract - i.e with legal force, to entertain, etc.
Structure: Set of ICEs, list of instructions, combination of objective specifications and plan specifications, ordered by a form, being a set of ICEs that form a discourse structure, etc

What I don't like about removing terms in favor of "document"/"Publication" is that these terms solely capture one aspect of the classification. For document it is that it is a set of ICEs. For publication it is that copies have been made by a certain type of process that includes realization of publishers roles.

Report and Narrative object capture different aspects.

[Quoted text hidden]

jie zheng <zhengj2007@gmail.com>

Fri, Feb 17, 2012 at 10:34 AM

To: information-ontology <information-ontology@googlegroups.com>

I would be happy to fix OBI if needed. OBI has two subclasses of IAO:report and IAO:report is not used in any other places in OBI. I would like to deprecate 'report'. Just like Alan said it is a more specific document. So far, I cannot see the desire to have 'report' than more specific report concepts, such as OBI:'direct submission to IEDB' and OBI:'investigation results report'. If IAO decides to deprecate report, I can send email to OBI developers to check whether it works to move those two terms under IAO:document.

About 'journal article', 'scientific journal article' is what we need in OBI.

I like Alan's proposal about defining 'publication' as the output of publication process. Asserted 'journal article' as 'document' with axiom 'is_output_of' publication process and would be inferred as 'publication'.

Jie

On Feb 17, 11:15 am, Alan Ruttenberg <alanruttenb...@gmail.com> wrote:

[Quoted text hidden]

Larry Hunter <Larry.Hunter@ucdenver.edu>

Fri, Feb 17, 2012 at 10:51 AM

To: Alan Ruttenberg <alanruttenberg@gmail.com>

Cc: jie zheng <zhengj2007@gmail.com>, information-ontology Discuss <information-ontology@googlegroups.com>, bjoern.peters@gmail.com

Comments in line below. Summary: I think there are easy solutions to all of the issues Alan raises, but am also open to a conference call if anyone things we need one.

Larry

On Feb 17, 2012, at 9:15 AM, Alan Ruttenberg wrote:

> On Fri, Feb 17, 2012 at 9:58 AM, Larry Hunter <Larry.Hunter@ucdenver.edu> wrote:

>>

>> I argued 3 years ago that report ought to be obsoleted and, in general, replaced by 'document', which is a parent of 'publication' and 'publication about an investigation'. Alan complained that this would break OBI in some way, and it has been sitting there ever since. I agree with you completely. Get Alan to fix OBI so we can finally close this out.

>

> It would be Jie that would fix OBI, and it would be Bjoern who would
> be a stakeholder I would ask for approval to remove report.

OK, I've cc'd Bjoern. Bjoern, any objections to replacing the OBI import of IAO:report with IAO:document?

> Reports

> are defined by an objective (though not logically yet) and are more
> specific than document.

Perhaps we should just make a defined class under 'IAO:document' that is a 'IAO:document about an OBI:investigation' -- would that meet your need? It seems like that fits the narrative definition in the 'IAO:report'.

- > Journal article is too broad for what we work
- > with - I suggest scientific journal article or academic journal
- > article, as these would be more specific (and have different
- > criteria).

'IAO:journal article' should be obsoleted along with report. I think that would be better handled as 'IAO:publication about an OBI:investigation'.

- > We don't say anything about peer review, which I also thing
- > is wrong.

I think there is an easy solution to this: make a subclass of 'IAO:publication' to be 'IAO:peer reviewed publication' (defined as a publication that has been peer reviewed), and have a defined class of 'IAO:peer reviewed publication about an investigation'

- > IMO, the asserted hierarchy should be according to the
- > intention/nature of the information, and the inferred hierarchy by
- > processes that are incidental to the content. Publication is a case of
- > the latter, report a case of the former.

Since we want to obsolete report, the asserted hierarchy would be

IAO:document
IAO:document about an investigation

adding 'published' and 'peer reviewed' via inference to get the inferred hierarchy:

IAO:document
 IAO:document about an investigation
 IAO:published document about an investigation (synonym: publication about an investigation, inferred)
 IAO:peer reviewed published document about an investigation (synonym: peer reviewed
 publication about an investigation, inferred)
 IAO:published document (synonym: publication, definition: document that has been published, inferred)
 IAO:published document about an investigation (synonym: publication about an investigation, inferred)
 IAO:peer reviewed published document (synonym: peer reviewed publication, definition: published
 document that has been peer reviewed, inferred)
 IAO:peer reviewed published document about an investigation (synonym: peer reviewed
 publication about an investigation, inferred)

I'm good with this approach.

- > Here's what I wrote Jie and forgot to cc.
- >
- > Is a journal article a journal article before it is published? If after, ok.
- > publication should be a process and publication defined as exactly the
- > specified output of that process.
- > The editor note on journal article says we should deprecate it in
- > favor of documentation. The definition is better on report, and can be
- > improved given Darren's editor note on report. Scientific reports
- > would also seem to satisfy the def for data item "a data item is an
- > information content entity that is intended to be a truthful statement
- > about something (modulo, e.g., measurement precision or other
- > systematic errors) and is constructed/acquired by a method which
- > reliably tends to produce (approximately) truthful statements."

I think we can just inherit this from OBI:investigation if we define 'IAO:document about an OBI:investigation'

- >
- > Responded to Jie but forgot to cc:
- >
- > So perhaps
- >
- > Document: Collection of information content entities intended to be
- > understood as a whole

OK

- > Publication: =def A document that is specified output of some
- > publication process

OK, although for clarity I would prefer 'IAO:published document' with a synonym 'publication'

- > Report: a document assembled by an author for the purpose of providing
- > information for the audience. A report is the output of a documenting
- > process and has the objective to be consumed by a specific audience.
- > Reports are complete documents or parts of documents.

Yuck. All documents are for the purpose of communicating to an audience (perhaps just one's self later in time). It seems to me that a 'document about an OBI:investigation' handles the output of the documenting process (part of the investigation). If so, there is no need for the 'report' term.

- > Scientific Journal article =def a report consisting primarily of data
- > items that is specified output of a publication process (and hence
- > publication)

Yuck. Many scientific journal articles do not consist primarily of data items. I think we leave trying to define what scientific means to 'OBI:investiation' if we just go with 'published document about an investigation' instead.

- >
- > Maybe:
- >
- > Asserted:
- > Document
- > Report
- >
- > Defined:
- > Publication (inferred to be document)
- > Scientific Journal Article (asserted document, inferred to be publication)
- >
- > I have a feeling this needs more thought.

Agree, see above. :-)

Larry

[Quoted text hidden]

Mathias Brochhausen <mbrochhausen@googlemail.com>

Fri, Feb 17, 2012 at 10:55 AM

To: jie zheng <zhengj2007@gmail.com>

Cc: information-ontology <information-ontology@googlegroups.com>

I absolutely agree with Jie and Alan, here.

Best,
Matt

[Quoted text hidden]

Larry Hunter <Larry.Hunter@ucdenver.edu>

Fri, Feb 17, 2012 at 11:49 AM

To: Alan Ruttenberg <alanruttenberg@gmail.com>

Cc: Mathias Brochhausen <mbrochhausen@googlemail.com>, information-ontology <information-ontology@googlegroups.com>

On Feb 17, 2012, at 9:25 AM, Alan Ruttenberg wrote:

- > Narrative object currently says set of propositions. I think that's
- > not right. The thing that makes it narrative is that it is an ordered
- > list of propositions. This distinguishes it from a list of top hit
- > genes. I'd be happy to have it be a subtype of textual entity.

I agree that narratives have a specific order of the propositions. Since textual entities are linear, I can't really imagine one that was an unordered set of propositions, so I would propose we change the definition ("set" -> "ordered set") rather than subclassing, but if anyone can think of a textual entity that is an unordered set of propositions I would be OK with subclassing.

- > Textual entity exposes a third axis of classification, namely by encoding.
- > So we have classification by:
- >
- > Processes related to copying/reviewing and other processes that don't
- > change the essentials of what the ICE means

That is: 'published', 'peer-reviewed', maybe also 'cited' etc. Agree.

- > Encoding: textual versus image versus music versus voice versus sign language

Oy, this is complicated. My intuitions are that music, voice and sign language are (mostly) not documents. Is an opera a document? My intuition says no. Is the libretto to an opera a document? Yes. Why? Because it is a combination of textual entities and figures. Perhaps we need 'IAO:written document' defined as a "collection of 'IAO:textual entities' and 'IAO:figures' intended to be understood together as a whole".

I'm not sure whether it makes more sense to have "written document" "audio recording" and "audio/video recording" as sibling subclasses of 'IAO:document' or to redefine 'IAO:document' to use the written document definition from above and make "audio recording" and "audio/video recording" etc. siblings of document (distinct subclasses of ICE). I lean towards the latter (recordings as siblings of document, not subclasses).

- > Purpose/Objective: to explain something to a certain kind of audience,
- > to be a record of a license/contract - i.e with legal force, to
- > entertain, etc.

I don't think there are very many distinctions we need to make here. We already have 'directive information entity' and we should probably define legally enforceable documents (we have 'IAO:patent' but its defined as the output of a patent-granting process, not as a legally enforceable entity). I wonder how much we would need to carve up the world in terms of objective or intended audience, though. I have a use case for wanting to distinguish between primary research reports (which I would call published documents about an investigation), review articles, tutorials and textbooks, but I would be cautious about trying to plumb authorial intent (which can be quite different from how a reader uses a document).

- > Structure: Set of ICEs, list of instructions, combination of objective
- > specifications and plan specifications, ordered by a form, being a set
- > of ICEs that form a discourse structure, etc

What are differences that matter here? Don't these all fit into IAO already?

- > What I don't like about removing terms in favor of
- > "document"/"Publication" is that these terms solely capture one aspect
- > of the classification. For document it is that it is a set of ICEs.

- > For publication it is that copies have been made by a certain type of
- > process that includes realization of publishers roles.

Well, I want to remove 'IAO:report' and 'IAO:journal article' because they capture none of these (well).

Document, directive information entity and (peer-reviewed) publication are the main distinctions used by biomedical researchers, at least on their CVs. Speaking of which, we should also think about where "grant proposal" would go as an ICE. Seems to be a subclass of document that could be defined as an input to a grant-giving process, and one of its components would be an 'IAO:study design'. I am fine with most of the other ones you raised as well, or are you concerned that there are yet further dimensions along which we need to classify ICEs?

Larry

[Quoted text hidden]

Jonathan A Rees <rees@mumble.net>

Sat, Feb 18, 2012 at 3:25 PM

To: Alan Ruttenberg <alanruttenberg@gmail.com>

Cc: Larry Hunter <Larry.Hunter@ucdenver.edu>, jie zheng <zhengj2007@gmail.com>, information-ontology <information-ontology@googlegroups.com>

On Fri, Feb 17, 2012 at 11:15 AM, Alan Ruttenberg

<alanruttenberg@gmail.com> wrote:

- > We don't say anything about peer review, which I also thing
- > is wrong.

Good luck getting your curators to check whether any given article has ever been peer reviewed - or even to be able to find out, if they do want to know. E.g. tell me whether the articles published in the journal Sociobiology are peer reviewed. I suspect such a distinction with either be unused or used with error (i.e. without meaning).

Jonathan

[Quoted text hidden]

Bjoern Peters <bpeters@liai.org>

Sun, Feb 19, 2012 at 4:14 PM

To: Larry Hunter <Larry.Hunter@ucdenver.edu>

Cc: Mathias Brochhausen <mbrochhausen@googlemail.com>, information-ontology <information-ontology@googlegroups.com>, Alan Ruttenberg <alanruttenberg@gmail.com>

First of all, I absolutely agree that report <-> document (and all of IAO) needs work. No, I am not okay with obsoleting 'report'. Rather I would like to tackle a broader re-organization of IAO along the lines of different axes of classification that Alan mentioned in his last email, and that we have postponed for too long. Here are some semi-coherent thoughts. I am all for having a phone call.

The original intention of 'report' when we were still doing IAO in OBI was to capture all the ICEs through which a scientist 'reports' results from an investigation, be it a journal article, database submission, progress report to funding agency, conference poster, scivee video, etc. This seemed to fit fine as a subclass under 'document' and is distinguished from other documents such as e.g. a plan specification which is not about something that has happened already. The classification was solely based on 'intent'. Shortening the label from 'report of results' to 'report' and the overall confusing state of IAO have not helped to make this clear.

I agree with Larry's point that "All documents are for the purpose of communicating to an audience (perhaps just one's self later in time)". In fact, I would argue that this is true for all ICEs. Can we capture that in the ICE definition, and ideally associate a given ICE with creator(s) and intended consumer(s) through logical relations? If not, what ICEs are in the scope of IAO that do not meet these criteria?

As for the different axes of classification, I believe we need all that Alan has identified. Repeated

- 1) Processes related to copying/reviewing and other processes that don't change the essentials of what the ICE

means

2) Encoding: textual versus image versus music versus voice versus sign language

3) Purpose/Objective: to explain something to a certain kind of audience, to be a record of a license/contract - i.e. with legal force, to entertain, etc.

4) Structure: Set of ICEs, list of instructions, combination of objective specifications and plan specifications, ordered by a form, being a set
of ICEs that form a discourse structure, etc

"Report of investigation results" has a clear purpose / objective, and no limitations on the other axis.

The present definition of "document" is "A collection of information content entities intended to be understood together as a whole". This definition clearly distinguishes a 'document' from ICEs that are intrinsically parts of documents, such as a 'title', but there is no distinction along any of the 4 axes above. However, it seems that (following Larry's previous email) the label 'document' does not fit for e.g. an opera or me talking to someone. There is an expectation that a document is at least somewhat enduring over time, which can be captured by limiting the types of encodings. I was expecting a sound recording to be a child of document and sibling of 'written document'.

- Bjoern

[Quoted text hidden]

--

--

Bjoern Peters
Assistant Professor
La Jolla Institute for Allergy and Immunology
9420 Athena Circle
La Jolla, CA 92037, USA
Tel: [858/752-6914](tel:8587526914)
Fax: [858/752-6987](tel:8587526987)
<http://www.liai.org/pages/faculty-peters>

[Quoted text hidden]