Bits and Pieces:
A Revised Look at Metadata eProduction Workflow

Presented by
Bruce D. Rosenblum
CEO
Inera Incorporated

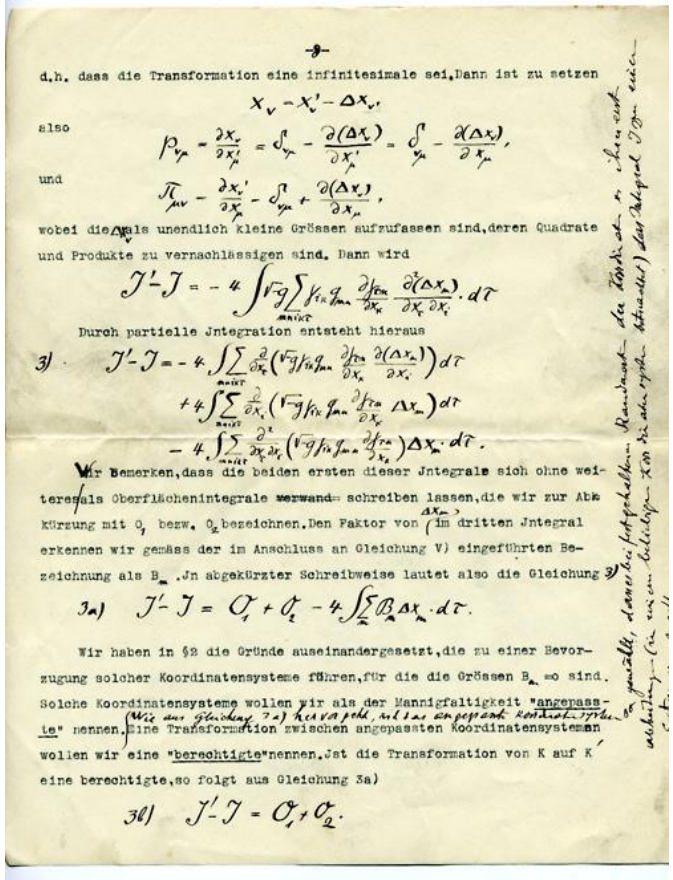
STM International, 5 December 2011

Back In The Olde Days

- **Circa 1990…**

Manuscript Submission

- **Author submitted manuscript**
  - **Typewritten**
  - **… or longhand**
  - **… or both**



-8-

d.h. dass die Transformation eine infinitesimale sei. Dann ist zu setzen

$$x_\nu = x'_\nu - \Delta x_\nu,$$

also

$$p_{\nu\mu} = \frac{\partial x_\nu}{\partial x'_\mu} = \delta_{\nu\mu} - \frac{\partial(\Delta x_\nu)}{\partial x'_\mu} = \delta_{\nu\mu} - \frac{\partial(\Delta x_\nu)}{\partial x_\mu},$$

und

$$\pi_{\mu\nu} = \frac{\partial x'_\nu}{\partial x_\mu} = \delta_{\nu\mu} + \frac{\partial(\Delta x_\nu)}{\partial x_\mu},$$

wobei die Δx_ν als unendlich kleine Grössen aufzufassen sind, deren Quadrate und Produkte zu vernachlässigen sind. Dann wird

$$J' - J = -4\int \sqrt{-g}\sum_{mnik\tau} \gamma_{ik} g_{mn} \frac{\partial \gamma_{\tau m}}{\partial x_k} \frac{\partial^2(\Delta x_m)}{\partial x_\tau \partial x_i} \cdot d\tau$$

Durch partielle Jntegration entsteht hieraus

$$3)\quad J' - J = -4\int \sum_{mnik\tau} \frac{\partial}{\partial x_\tau}\left(\sqrt{-g}\gamma_{ik} g_{mn} \frac{\partial \gamma_{\tau m}}{\partial x_k} \frac{\partial(\Delta x_m)}{\partial x_i}\right) d\tau$$
$$+ 4\int \sum_{mnik\tau} \frac{\partial}{\partial x_i}\left(\sqrt{-g}\gamma_{ik} g_{mn} \frac{\partial \gamma_{\tau m}}{\partial x_k} \Delta x_m\right) d\tau$$
$$- 4\int \sum_{mnik\tau} \frac{\partial^2}{\partial x_\tau \partial x_i}\left(\sqrt{-g}\gamma_{ik} g_{mn} \frac{\partial \gamma_{\tau m}}{\partial x_k}\right) \Delta x_m \cdot d\tau.$$

Wir bemerken, dass die beiden ersten dieser Jntegrale sich ohne weiteres als Oberflächenintegrale ~~verwand~~ schreiben lassen, die wir zur Abkürzung mit O_1 bezw. O_2 bezeichnen. Den Faktor von Δx_m im dritten Jntegral erkennen wir gemäss der im Anschluss an Gleichung V) eingeführten Bezeichnung als B_m. Jn abgekürzter Schreibweise lautet also die Gleichung 3)

$$3a)\quad J' - J = O_1 + O_2 - 4\int \sum_m B_m \Delta x_m \cdot d\tau.$$

Wir haben in §2 die Gründe auseinandergesetzt, die zu einer Bevorzugung solcher Koordinatensysteme führen, für die die Grössen $B_m = 0$ sind. Solche Koordinatensysteme wollen wir als der Mannigfaltigkeit "angepasste" nennen. (Wie aus Gleichung 3a) hervorgeht, ist das angepasste Koordinatensystem so gewählt, dass es bei festgehaltenem Randwert der Koordinaten in ihrer ersten Ableitung (in einem beliebigen Koordinatensystem betrachtet) das Integral J zum einen Eine Transformation zwischen angepassten Koordinatensystemen wollen wir eine "berechtigte" nennen. Jst die Transformation von K auf K' eine berechtigte, so folgt aus Gleichung 3a)

$$3b)\quad J' - J = O_1 + O_2.$$

Manuscript Revisions Workflow

- **Manuscript peer reviewed**
- **Revised manuscript submitted**
- **Add a few bits of metadata to the cover page**

MEMORANDUM

R.J.Reynolds Tobacco Company
Winston-Salem, N.C. 27102

RECEIVED
JUL 22 '82
S. B. W.

F.G. Colby
Associate Director of Scientific Issues
Research Department

RJR

July 16, 1982

Mr. Samuel B. Witt, III

RE: "VITAMIN A" - BOON OR BANE?

SUMMARY

This memorandum attempts to assess, in SOME detail, very "soft" and controversial epidemiological and experimental (animal, in vitro) claims, replete with gaps, that "vitamin A" and/or chemically related substances may be "protective" against several cancers, and particularly lung cancer, in males. In my opinion, the state of the art has TOO many gaps, and therefore does not permit a truly informed judgment on those exceedingly complex issues; neither does it even allow an impeccable design of an economically reasonable study to resolve them. However, speculatively, further assessment of the literature and/or new data may permit a tie-in with the "constitutional hypothesis" and/or my hypothesis of relating smoking and lung cancer to the "weak lung" as a common "cause".

MEMORANDUM

The U.S. National Research Council's Committee on Diet, Nutrition and Cancer has just published (6/16/82) a major, widely publicized, monograph entitled "Diet, Nutrition, and Cancer", (see especially chapter 9), which was received here yesterday. The thrust of the report is that "cancers of most major sites are influenced by dietary patterns". With reference to the main subject of this memorandum, i.e., the epidemiological data which allegedly show an inverse relationship between the risk of cancers and the consumption of foods that contain "vitamin A" or its chemical relatives, the report alleges: "A growing accumulation of epidemiological evidence indicates that there is an inverse relationship between the risk of cancer and the consumption of foods that contain vitamin A....". But then states: "Most of the data do not show whether the effects are due to carotenoids, to vitamin A itself, or to some other constituent of these foods."

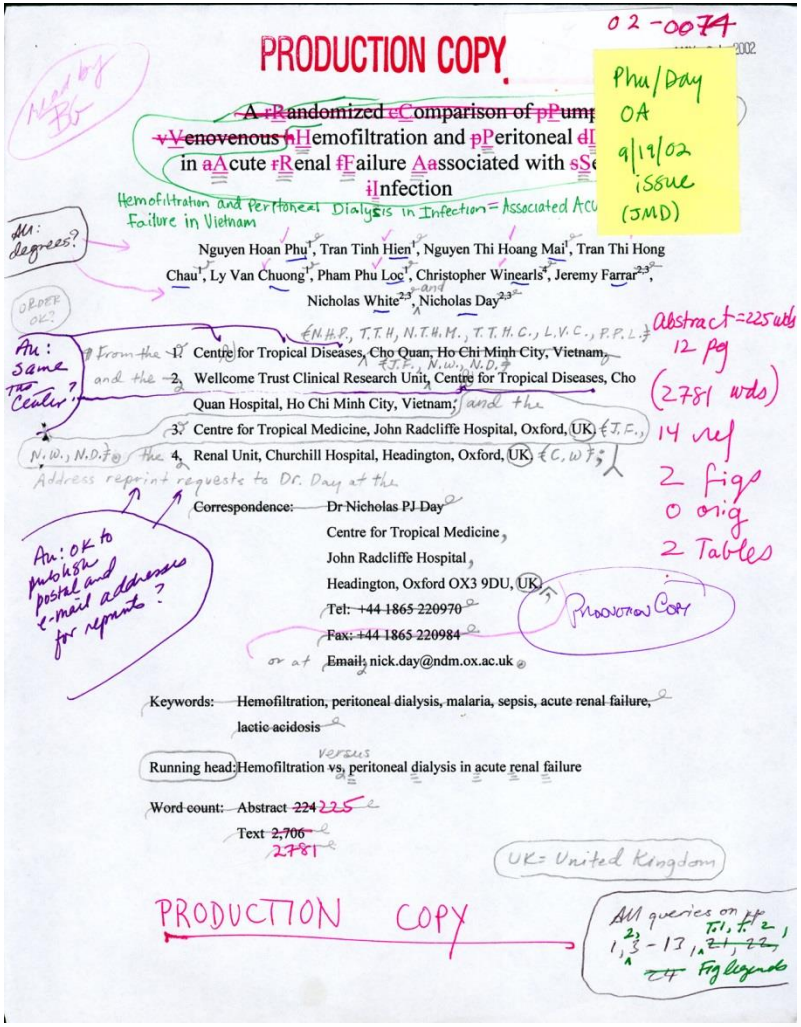
While this assessment is correct as stated, in my view, the report does not adequately assess or does not even mention - and therefore does not come to grips with - MANY other, potentially "confounding", "facts". For example, in man, high levels of vitamin A are allegedly associated with high levels of cholesterol. A number of studies have found that, while high levels of cholesterol and - hardly ever publicized - of vitamin A are associated with high levels of cardiovascular diseases, low levels of cholesterol allegedly are associated with increased cancer incidence, i.e., high cholesterol may be "protective" against cancers.

./..

50574 1611

Manuscript Editing Workflow

- **Edit on paper**



PRODUCTION COPY

A randomized comparison of peritoneal dialysis versus hemofiltration in acute renal failure associated with sepsis

Nguyen Hoan Phu[1], Tran Tinh Hien[1], Nguyen Thi Hoang Mai[1], Tran Thi Hong Chau[1], Ly Van Chuong[1], Pham Phu Loc[1], Christopher Winearls[4], Jeremy Farrar[2,3], Nicholas White[2,3], Nicholas Day[2,3]

1. Centre for Tropical Diseases, Cho Quan, Ho Chi Minh City, Vietnam,
2. Wellcome Trust Clinical Research Unit, Centre for Tropical Diseases, Cho Quan Hospital, Ho Chi Minh City, Vietnam;
3. Centre for Tropical Medicine, John Radcliffe Hospital, Oxford, UK
4. Renal Unit, Churchill Hospital, Headington, Oxford, UK

Correspondence: Dr Nicholas PJ Day
Centre for Tropical Medicine,
John Radcliffe Hospital,
Headington, Oxford OX3 9DU, UK
Tel: +44 1865 220970
Fax: +44 1865 220984
Email: nick.day@ndm.ox.ac.uk

Keywords: Hemofiltration, peritoneal dialysis, malaria, sepsis, acute renal failure, lactic acidosis

Running head: Hemofiltration vs. peritoneal dialysis in acute renal failure

Word count: Abstract 224 225
Text 2,706 2781

Abstract = 225 wds
12 pg
(2781 wds)
14 ref
2 figs
0 orig
2 Tables

UK = United Kingdom

PRODUCTION COPY

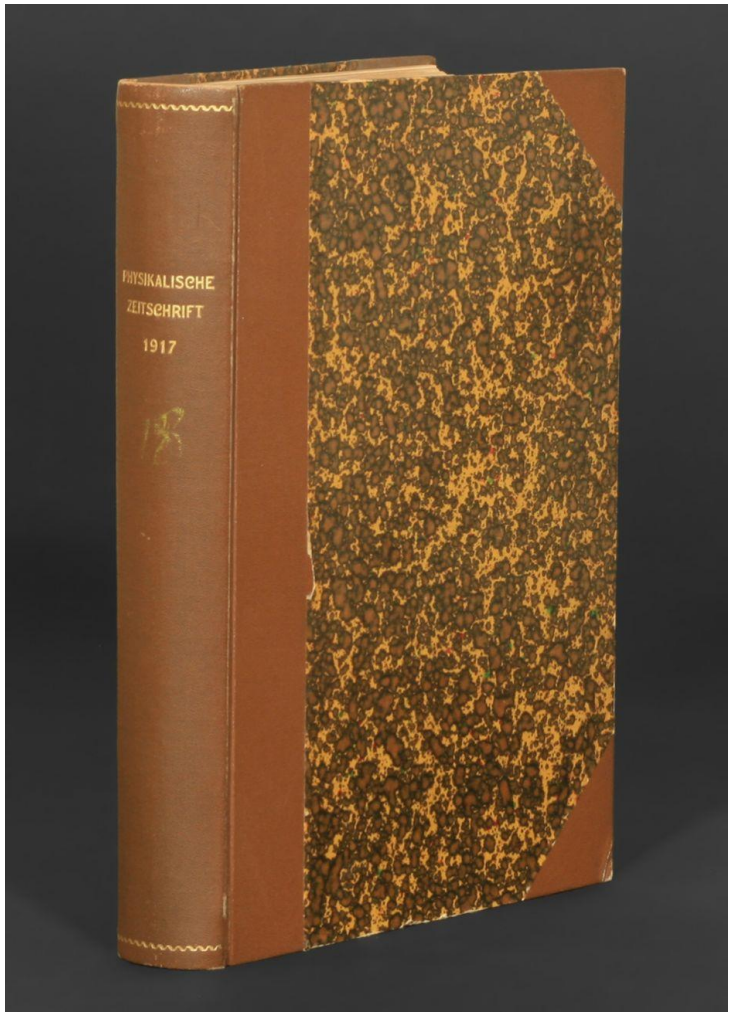
Journal Production Workflow

- **Typeset**
- **Proof**
- **Typeset corrections**
- **Check those copyright forms!**
- **Print**

Post-Print Workflow

- **Ship to libraries**
- **Wait for third party indexers**
- **Publish occasional correction or erratum**
- **Publish a very rare retraction**

The Result



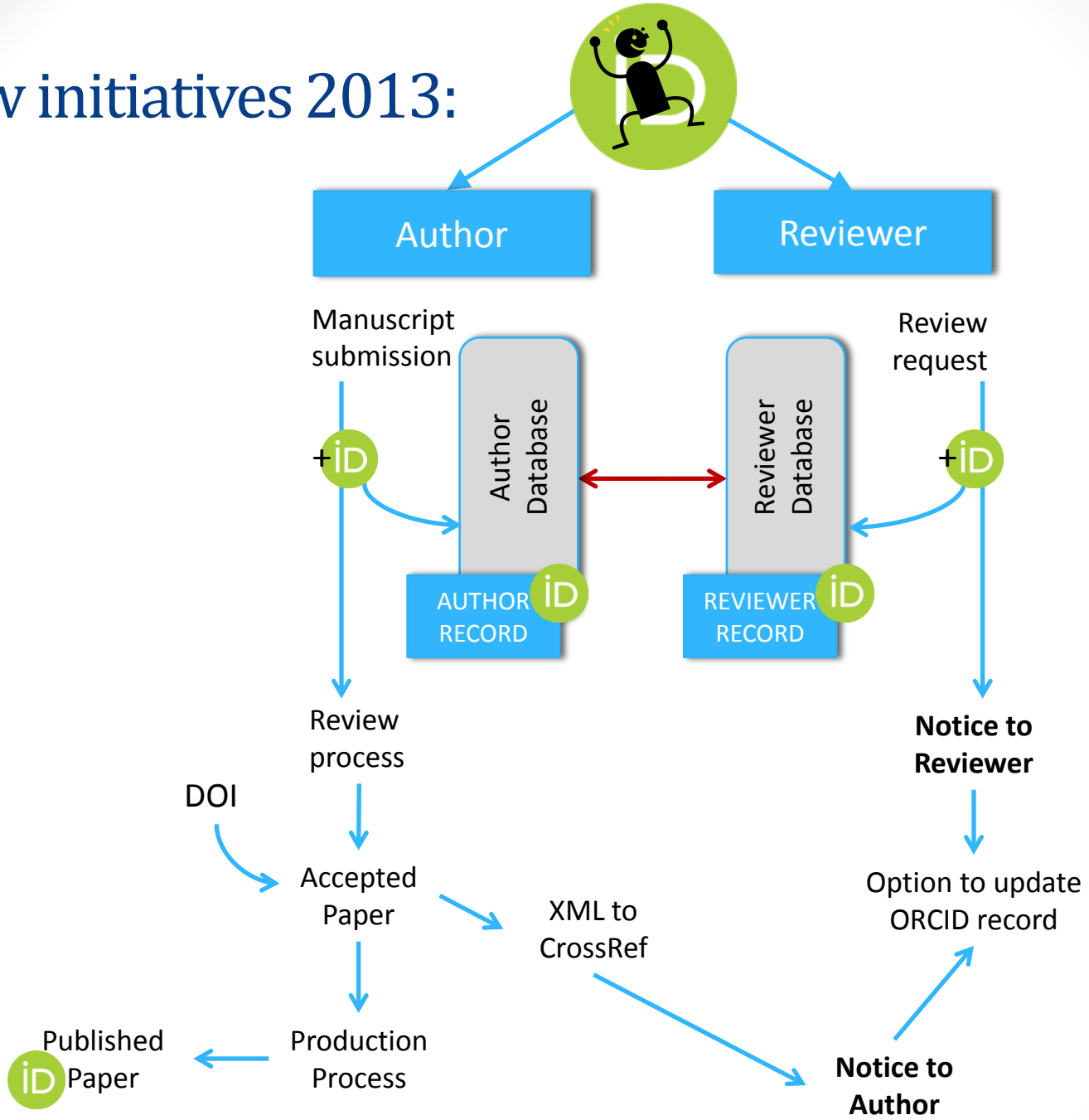
Flash back, Circa 2004

- Manuscript submitted electronically
- Transmittal file accompanies accepted file(s)
- Some non-redundant data copy/pasted or retyped from transmittal to Word
  - Article type
  - Article ID
  - Received, revised, accepted dates
- Electronic file edited & typeset
- Check those copyright forms!
- XML produced at some stage

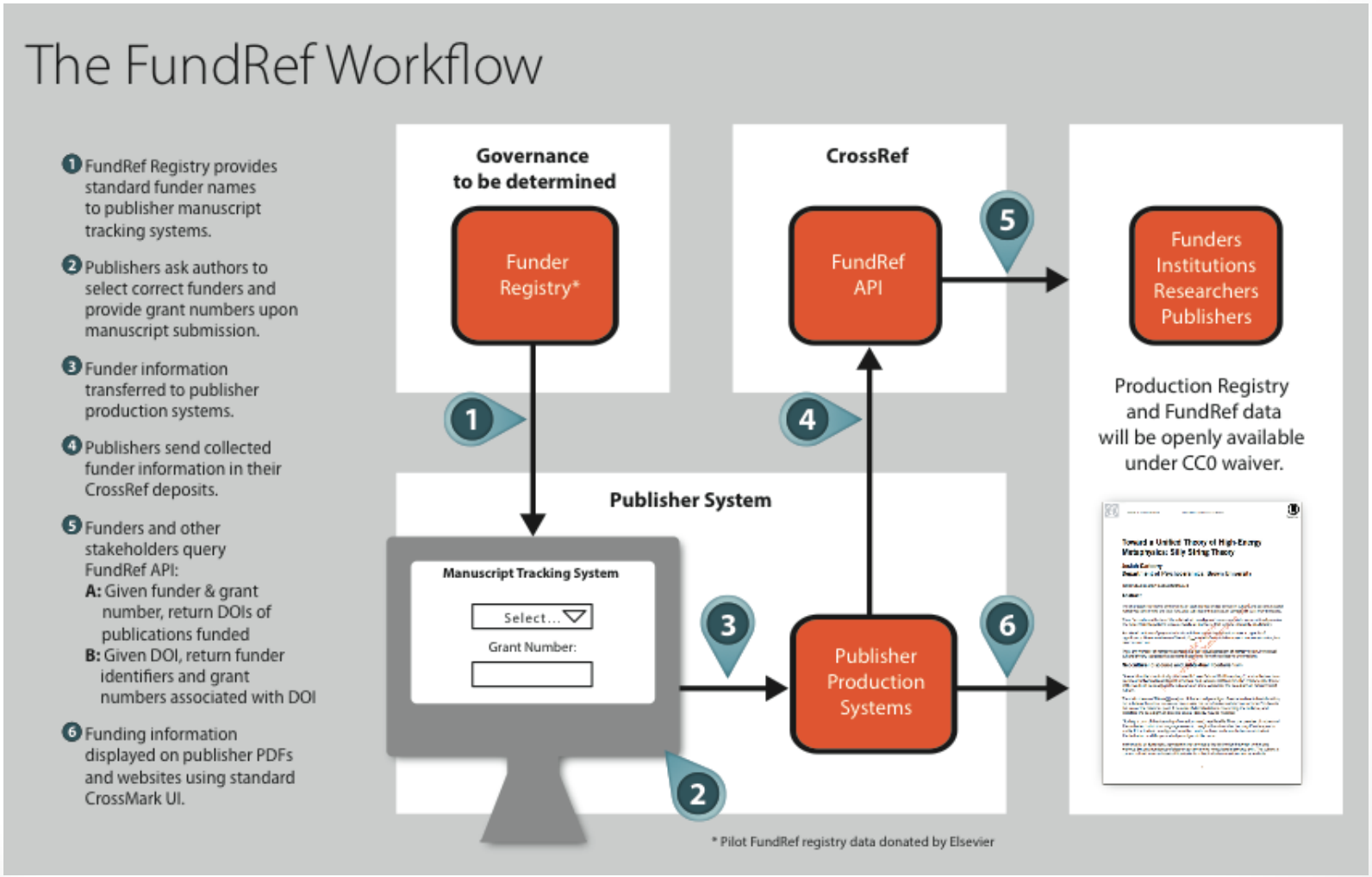
Post-publication, circa 2004

- **Post online**
- **Upload metadata to CrossRef**
- **Ship to libraries**
- **Archive PDF and XML**
- **Publish occasional correction or erratum**
- **Publish rare retraction**

New initiatives 2013:

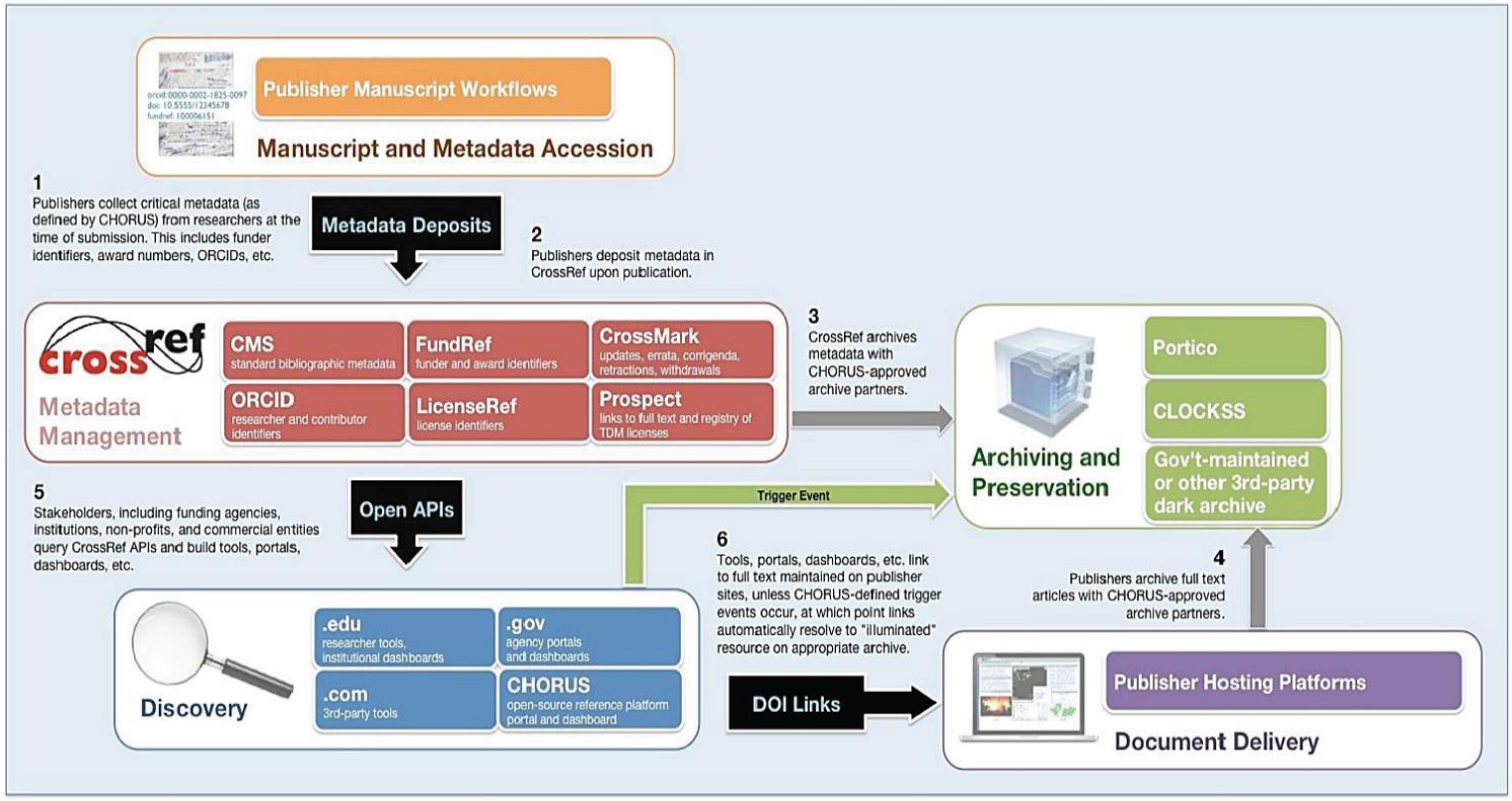


New Initiatives, 2013: FundRef



New Initiatives, 2013: Chorus

CHORUS Conceptual Design



Flash forward, 2014

- Manuscript submitted electronically
- Author fills out reams of online forms (ugh!)
  - Signs up or ORCID
  - Adds ORCID to submission
  - Adds rest of authors and affiliations
    - (maybe; after all it's in the Word file)
  - Adds funding information
    - (maybe; after all it's in the Word file)
  - Conflict of Interest and Disclosure forms
  - Pays author charges

Production, 2014

- Transmittal file accompanies accepted file(s)
- Manuscript file copy edited
- Transmittal and manuscript metadata reconciled
  - Say what?
- Typeset file
- XML produced at some stage
- Check that Creative Commons copyright
- Add enriched metadata and semantics

Post-publication, 2014

- Post online
- Upload metadata to PubMed, CrossRef, others
- Ship print to libraries???
- Comply with funding agency requirements
- Post PDF and XML at one or more archives
  - PMC
  - UK-PMC
  - Portico
- Publish occasional correction or erratum
- Publish occasional retraction

New Initiatives…

- **ORCID**
- **FundRef**
- **CrossRef Metadata Services**
- **CHORUS**
- **Etc.**

…Rely on New Metadata…

- **Authenticated ORCID**
  - **Ideally for each author**
- **Funding information**
  - **With institutional ID**
- **License information**
  - **With correct attributes**
- **Updated article versions**
  - **CrossMark or other records**

…And More Complex Workflows

- **More data collected from authors**
- **More data integration required by publishers**
- **More metadata sent to indexes**
- **More full text deposits to archives**

Basic Rules of Metadata

- **Enter once, enter right**
- **Don't:**
  - **do manually what you can do automatically**
  - **burden researchers with administrative tasks**
  - **copy/paste**
- **Do**
  - **validate early and often**
  - **single-source metadata**
  - **re-optimize the workflow**

Data Synchronization

- **Remember this bullet:**
  - **“Transmittal and manuscript metadata reconciled”**

- **Problems**
  - **Author reconciliation**
  - **ORCID reconciliation**
  - **Funding reconciliation**
  - **License reconciliation**

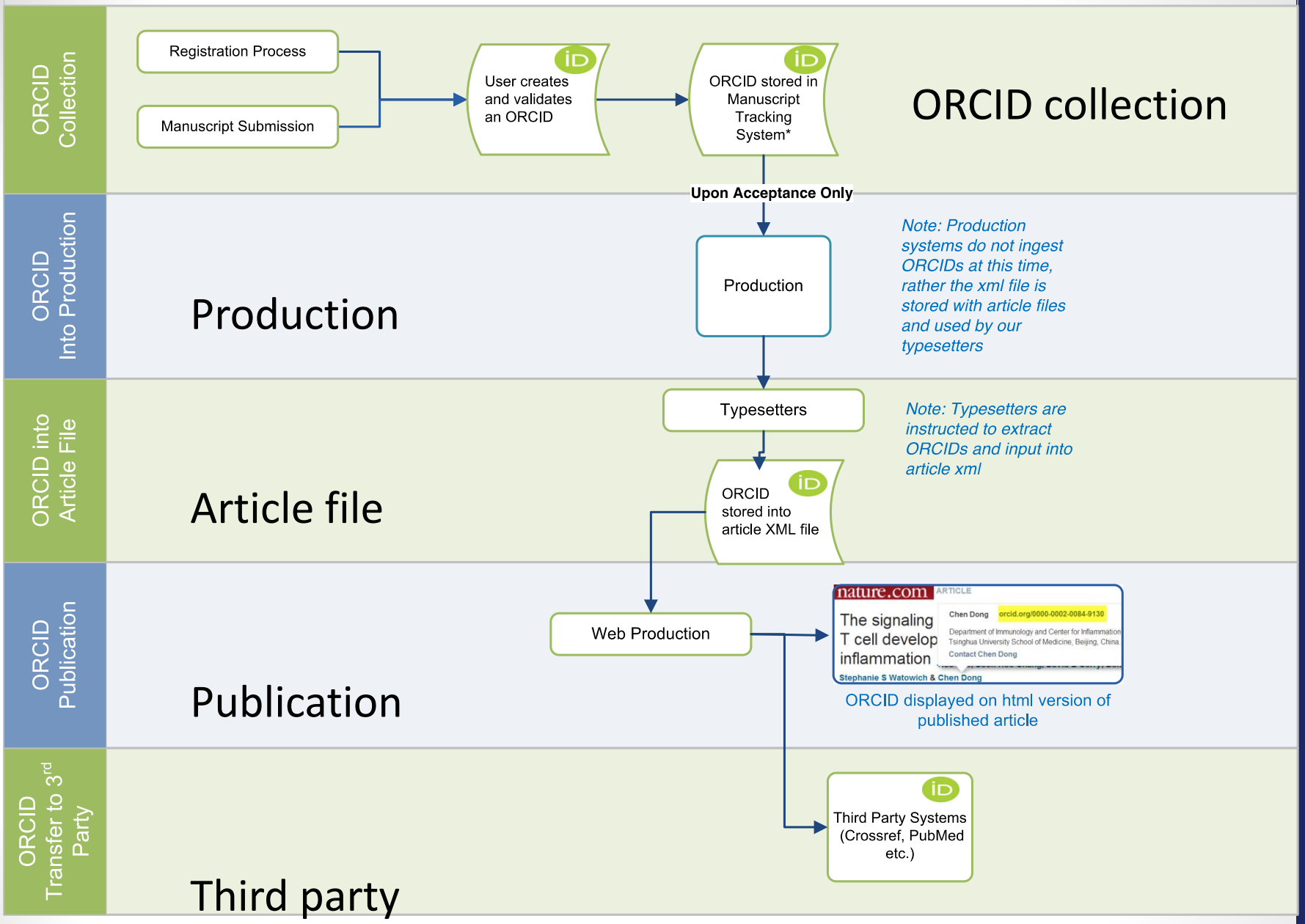
Author Reconciliation

- **Problem 1**
  - **Submitting author doesn't enter all authors in submission system**
- **Problem 2**
  - **Author list changes during revisions**
  - **Submitting author doesn't update submission system**
- **Solution**
  - **Rely on manuscript author list**

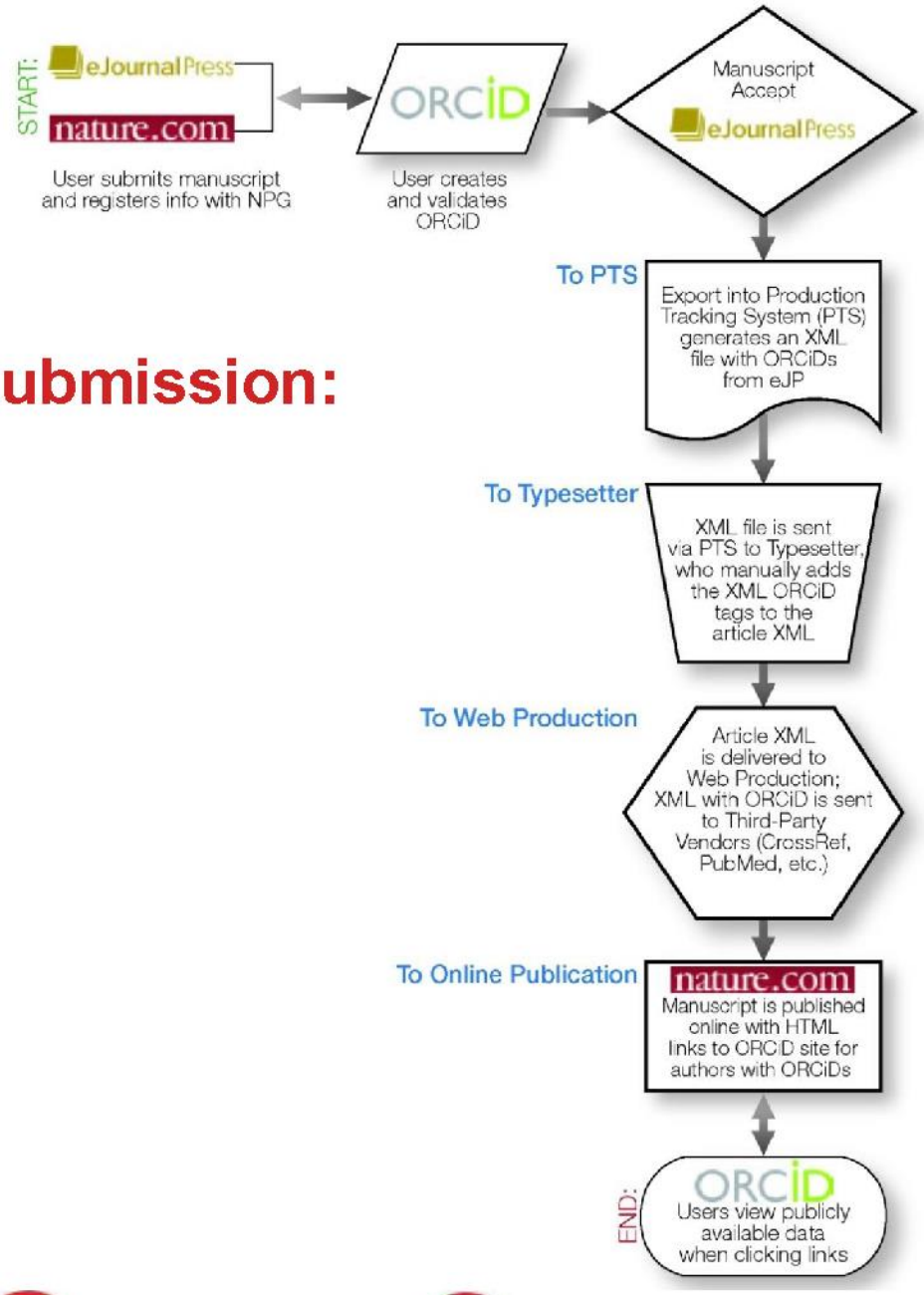
ORCID Issues

- Imagine every author with an ORCID
  - Robust tracking of research funding and publication

- How do achieve and ORCID for every author?
  - Request from submitting author?
    - May burden submitting author
  - Request for each author in paper?
    - May slow submission, especially with lots of authors
    - ORCIDs must be authenticated
  - Look up in production?
    - No name/affiliation lookup in ORCID today
    - ORCIDs must be authenticated

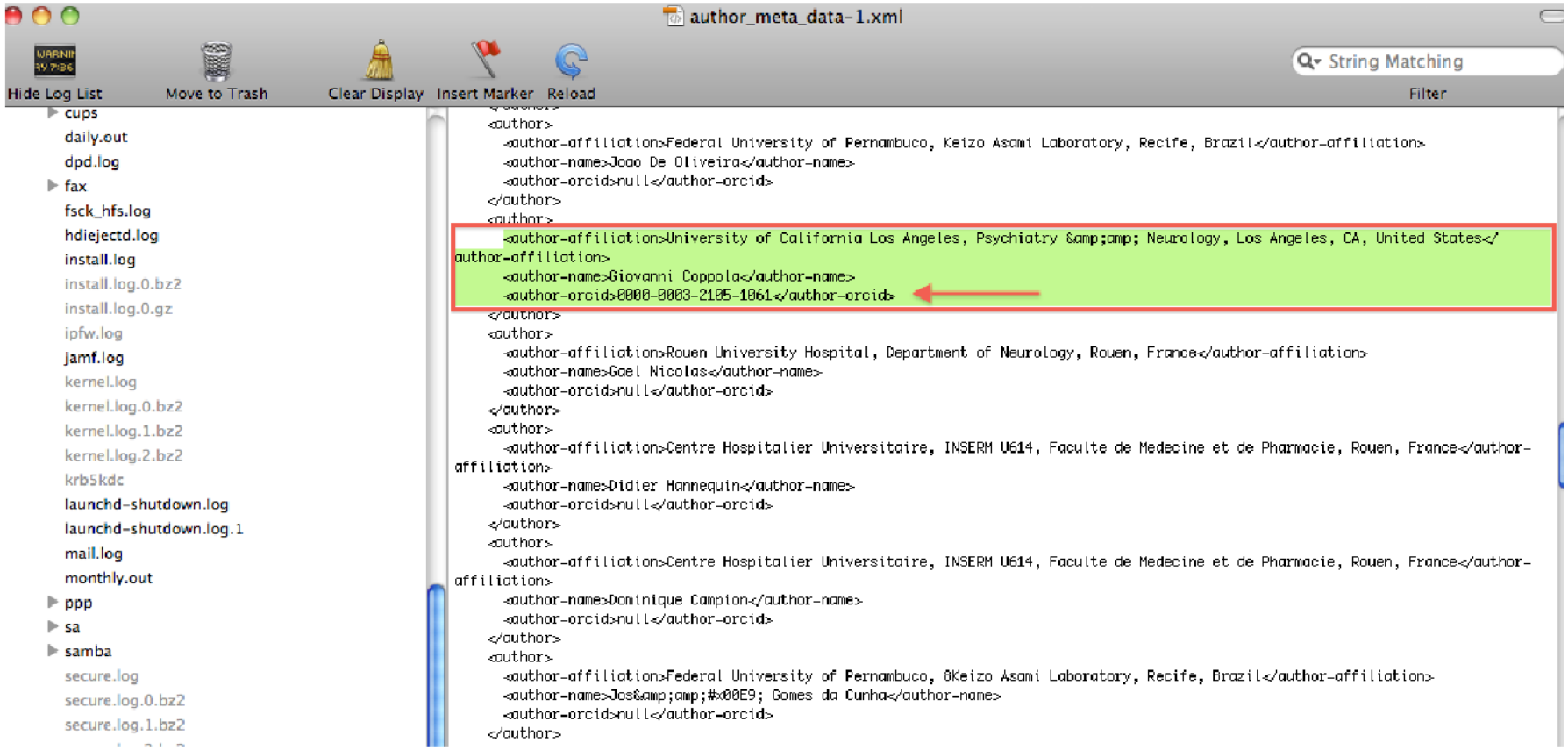

ORCID Collection
Registration Process
Manuscript Submission
User creates and validates an ORCID
ORCID stored in Manuscript Tracking System*
ORCID collection
Upon Acceptance Only
ORCID Into Production
Production
Production
Note: Production systems do not ingest ORCIDs at this time, rather the xml file is stored with article files and used by our typesetters
ORCID into Article File
Typesetters
Note: Typesetters are instructed to extract ORCIDs and input into article xml
Article file
ORCID stored into article XML file
ORCID Publication
Web Production
nature.com ARTICLE
The signaling T cell develop inflammation
Chen Dong orcid.org/0000-0002-0084-9130
Department of Immunology and Center for Inflammation Tsinghua University School of Medicine, Beijing, China
Contact Chen Dong
Stephanie S Watowich & Chen Dong
ORCID displayed on html version of published article
Publication
ORCID Transfer to 3rd Party
Third Party Systems (Crossref, PubMed etc.)
Third party

ORCIDs in submission: overview



4: ADD DATA TO MANUSCRIPT XML

Once the manuscript is edited, the Author Meta Data XML file, containing author names, affiliations and ORCID IDs, is assigned to the typesetters who will eventually create the manuscript XML. Currently our typesetters obtain the ORCID IDs manually from that file.



nature publishing group

Author/ORCID Reconciliation

- **Authors: Manuscript file**
- **ORCIDs: Transmittal file**
- **Today's Reality: Copy/Paste**

- **Better solution: Electronic reconciliation of transmittal and manuscript**
  - **Compare number of authors**
  - **Compare author names**
  - **Integrate ORCIDs**
    - **Automatically if authors match**
    - **Manual reconcile on match failure**

FundRef

- More organizations (funders, publishers) want to track funding
- Add to your XML, and CrossRef deposits
  - JATS 1.1 adds institution ID
  - http://www.crossref.org/fundref/index.html
  - http://www.crossref.org/fundref/fundref_registry.html
  - See also tools at http://labs.crossref.org/

Funding Information

- Old way
  - Author added to acknowledgement paragraph
  - Few publishers tagged funding
- New way
  - Author adds to acknowledgement paragraph
  - Author selects funder in submission system
    - Controlled vocabulary (CrossRef via Elsevier)
  - More publishers tag funding for FundRef

Data Synchronization

- **Remember this bullet:**
  - **“Transmittal and manuscript metadata reconciled”**

- **Problems**
  - **Authors skip funding in submission process**
    - **AIP: 20% of authors do not add**
  - **Funder name in acknowledgement doesn’t use controlled vocabulary**
  - **Which funder is used in XML: transmittal or manuscript?**

Sample Acknowledgements

- This work was funded by an NIH cooperative agreement from the National Heart, Lung, and Blood Institute (grants U01HL077821, U01HL077826, U01HL077823). National Center for Research Resources (grant UL1RR025752), now at the National Center for Advancing Translational Sciences, provided laboratory testing
- This material was based on work supported by the Department of Veterans Affairs, Veterans Health Administration, Rehabilitation Research and Development Service, grant F4096R to Milton V. Icenogle, MD, Cardiology Section, Veterans Integrated Service Network 18, Albuquerque, New Mexico.
- This study was funded by Australian National Health and Medical Research Council (NHMRC) project grant 455209
- This work was supported by the German Israel Foundation under Grant No. 1–2038.1114.07, the Israel Science Foundation under Grant No. 1380021, the Deborah Foundation, the Poznanski Foundation, and MAFAT

Funding Issues

- **Today's reality**
  - **No reconciliation**
  - **Manual reconciliation at copy editing**
  - **Report unknown funders to CrossRef: fundref.registry@crossref.org**

Funder Reconciliation

- **Better solution: Electronic reconciliation of transmittal and manuscript**
  - **Compare transmittal funder with manuscript ack**
  - **Tag funding agencies automatically if funders match**
  - **Manually reconcile if no match or no transmittal funders**
- **But…**
  - **Still may need manual tagging of grant numbers**
  - **Still need copy-editor review**
    - **All funders in manuscript tagged**
    - **All funders in transmittal found in manuscript**
    - **All funding information tagged correctly**

License Information

- Old way
  - Checked copyright forms
  - Printed copyright statement
    - Publisher
    - Author
    - Government
- New way
  - Author selects traditional or OA option
  - Author's institution requires specific OA option
  - Reconciled license statement placed in XML

LicenseRef

License Information

```
<custom_metadata>
  <program name="fundref">
    <assertion name="fundgroup">
      <assertion name="funder_name">                    "Office of Naval Research"
        <assertion name="funder_identifier">http://dx.doi.org/10.13039/100000006</assertion>
      </assertion>
      <assertion name="award_number">N00014-11-1-0313</assertion>
    </assertion>
  </program>
  <program name="AccessIndicators">
    <license_ref applies_to="vor">http://creativecommons.org/licenses/by/3.0/deed.en_US</license_ref>
  </program>
</custom_metadata>
</crossmark>
<doi_data>
  <doi>10.1063/1.4802237</doi>
  <timestamp>20130419184426</timestamp>
  <resource>http://link.aip.org/link/APPLAB/v102/i15/p153303/s1&amp;Agg=doi</resource>
  <collection property="text-mining" setbyID="aip">
    <item>
      <resource content_version="vor">http://link.aip.org/link/applab/v102/i15/p153303/pdf/CHORUS</resource>
    </item>
  </collection>
</doi_data>
</journal_article>
</journal>
</crossref>
</doi_record>
```

License Issues

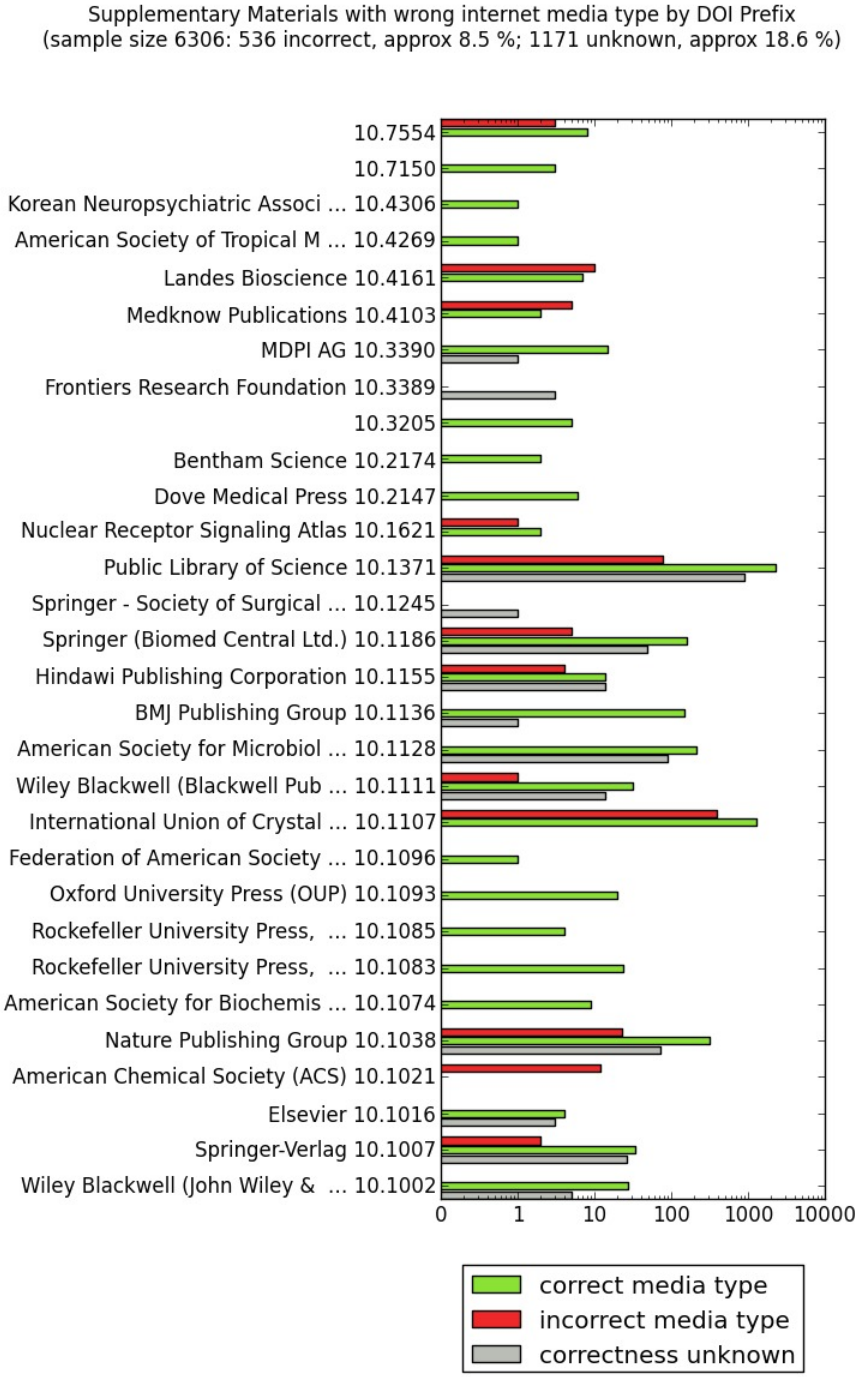
- JATS-Con 2013 paper
  - Inconsistent XML as a Barrier to Reuse of Open Access Content
  - http://www.ncbi.nlm.nih.gov/books/NBK159964/
  - Reuse of PMC media by Wikimedia Commons
- License metadata
  - Machine-readable for automated content re-use
  - Attribute and text must agree
- Supplementary Material format
  - Media type needs to match file format

Inconsistent License

- **Be careful to keep license information consistent:**
  - **<license license-type="open-access" xlink:href="http://creativecommons.org/licenses/by/2.5/"> <license-p>Re-use of this article is permitted in accordance with the Creative Commons Deed, Attribution 2.5, which does not permit commercial exploitation.</license-p> </license>**
  - **CC BY: "This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials." (http://creativecommons.org/licenses/)**

- **Supplementary media files with the wrong media type, by DOI prefix, from a sample of articles indexed in the first week of May 2013. Unknown refers to MS Office files, between the different formats of which we did not distinguish. (Source: http://www.ncbi.nlm.nih.gov/books/NBK159964/)**

- **Inconsistent information makes text & data mining or content reuse difficult or impossible**

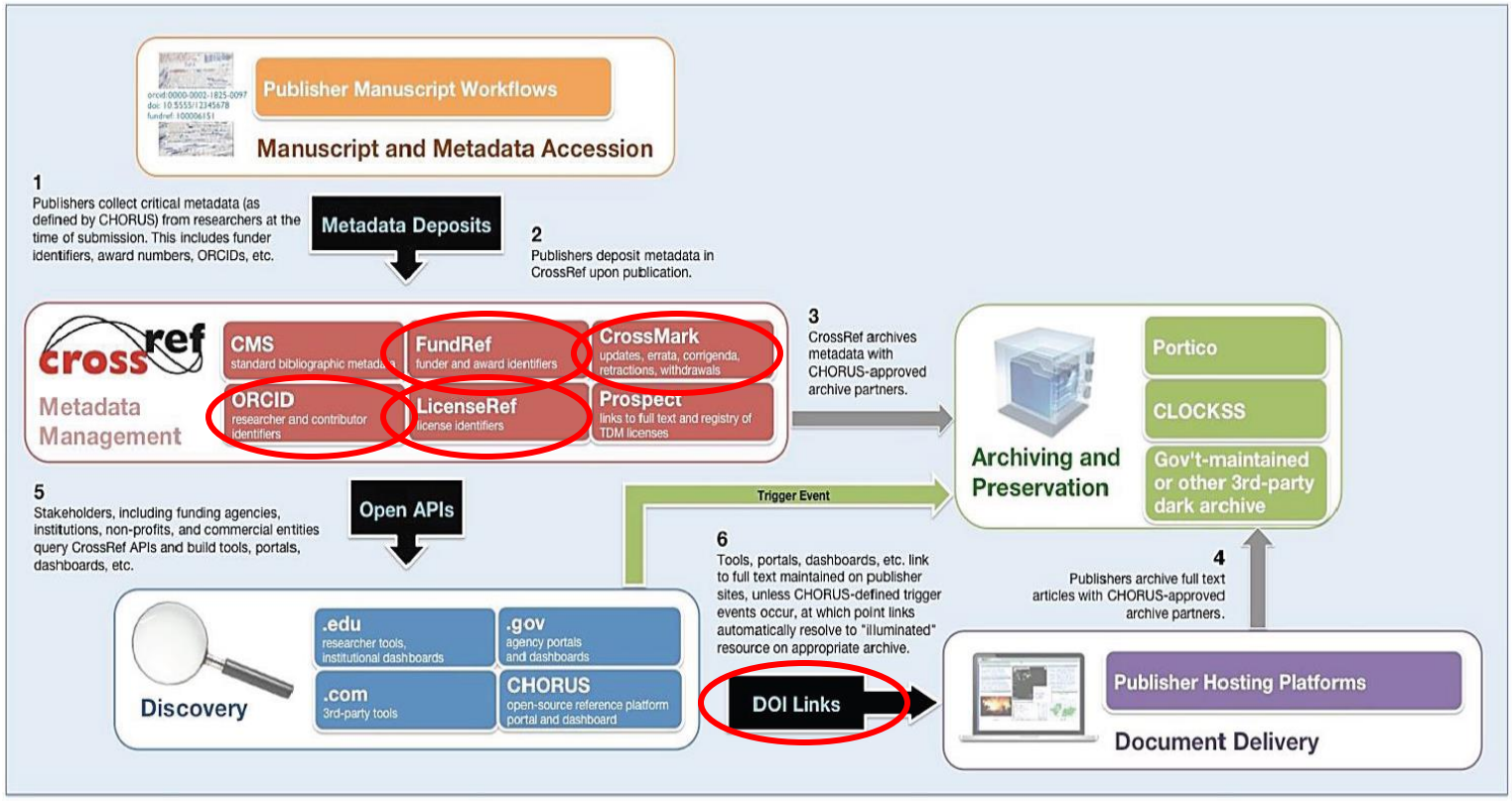


Supplementary Material

- **Recommended Practices for Online Supplemental Journal Article Materials**
  - **NISO RP-15-2013**
  - **http://www.niso.org/workrooms/supplemental**

CHORUS Requires New Metadata

CHORUS Conceptual Design



Next Steps

- Review enhanced metadata requirements
- Review workflows to meet requirements
- Implement reconciliation and QA workflow steps
- Deposit new metadata to CrossRef
  - ORCID
  - FundRef
  - LicenseRef
  - JATS abstracts for text and data mining

Conclusions

- Publishers no longer “throw issues over the wall”
- Scholarly publishing is more integrated then ever
- Integration drives new initiatives
- More integration requires more metadata
- Metadata reconciliation and accuracy is vital for new initiatives to succeed!
- Review and revise your metadata workflow

Questions?

Bruce Rosenblum

CEO

Inera Inc.

brosenblum@inera.com

617-932-1932