



ResourceSync – An Introduction



Todd Carpenter Executive Director, NISO Wolfram Data Summit Thursday, September 6, 2012 With thanks to Herbert Van de Sompel and Robert Sanderson (LANL)

@TAC_NISO Twitter Highlights

- Presenting this afternoon on the ResourceSync project at Wolfram Data Summit #wolframsummit
- I'm pre-tweeing my slides during #rsync presentation. Slides will be posted later today #wolframsummit
- NISO mission develop & maintain technical standards related to information, documentation, discovery & distribution of content #wolframsummit
- Machines don't talk like people do. Then again some people don't talk like other people do, particularly teenagers #wolframsummit
- So where did the ResourceSync project start? #NISO approached OAI about updating the PMH protocol. #wolframsummit
- The #NISO / OAI ResourceSync project was possible through the generous support of the Alfred P. Sloan Foundation. Thank you! #wolframsummit
 - What is RSync trying to solve (1/2): Source Server has resources that change. Destination servers want to leverage some/all of Source #wolframsummit
- What is RSync trying to solve (2/2): How to sync on ongoing basis in near-real-time & at web scale with as little system overhead as poss #wolframsummit
- RSync studied # of existing protocols to determine protocols that best meet needs. Bias against developing new spec from scratch. #wolframsummit
- The goal of ResourceSync is to find the model that most efficiently distributes the content, while limiting the tax on the source system. #wolframsummit
- Very early days in process of standards development. Still in incubation stage. Consensus & adoption phases coming '13 & beyond #wolframsummit
- Draft alpha specification of ResourceSync posted in August, Team meeting in Sept to review comments #wolframsummit



About



Non-profit industry trade association accredited by ANSI

Mission of developing and maintaining technical standards related to information, documentation, discovery and distribution of published materials and media

125+ Member organizations

Volunteer driven organization: 400+ spread out across the world

Represent US interests to ISO in the areas of Information & Documentation

Standards are familiar, even if you don't notice

| 53 | Sweet, The 6, 7, 114-1 |
|--------------|--------------------------------------|
| 37, | T |
| | tags 232, 292 Taylor, Dan 97 |
| | 2.62 |
| 5-6 240-1 | television 35, 132–55 adverts 155 |
| 54-5 | comedy 3, 156 |
| en J. | Long 137, 138 |

| | CONTENTS |
|---|--|
| | VE Annexity-Readerbarg |
| a hora | X Personality/Hoti/Liptimum |
| REI-HE, Called CAD HDP as Wedges, NA STREE-2011 | 8 Actoavietyments |
| | XII And he adver |
| | Z/ Hestalles |
| ava 1998, 3000, 2002 | 1 Addresses |
| Camualson. All rights reserved. | 2 Augusts for cominations for cinematography |
| advantaci to be standbarce with the Copyright, scantaci is assumbarce with the Copyright, scala 1988 | 3 Bateries |
| ention may be reproduced in any material property or straining in any modules by other produces on the wave by a module rady of this publication or used as a part of a | 4 Camera shaka |
| The polyanter or used as a part of a effect the writer permission of the and in accordance with the provisions of | 5 Camera maintenance |
| effect the vertice permission of the and in associations with the postalates of period Patterns in which a postalates of period Patterns in the terminal Agency base that the Council England With 40.0 Design Transform England With 40.0 | 6 Canara threading diagrams |
| and shows constant, begins and the state | 7 Composite cinematography |
| displang in Publication Data | 8 Electricity |
| the Dramangaghers | 9 Equipment check lass |
| as Cataloguing in Publication Data | 10 Expenses |
| to Dranatopopheroll W. Samakov, | 11 rimatoon |
| ese laforenses and index 67 Scharolizana, manuala ato, 1, Tita | 12 rites |
| 0.461 0.0 | 13 Lanses and depth of field tables |
| r | 14 Light and color |
| di Focal Press publicatione Xvev. Totalgress. pom | 15 Mathematics of cinematisgraphy |
| a liner listan by MPG Books U.S. Bodinin, Conwell | 16 special techniques |
| | 17 usas |
| | 18 Venue techniques |
| | - |
| | |

Machines don't talk like people do

Sometimes I wish you had a little green light that would blink on and off like the computer, so I could tell whether you're processing anything or if you've just frozen up.

September 6, 2012

slass*"stbg">clag alt="" height+"1" ssc+"/crocklas/webtg/css/sp/css-imspec/spacet.git" width="s" />c/td>-/tr> Chrystel channersideling's change alter" heighter is neer "Chronical Antion proving can under inder the proving can alter " widther's" drifter to a heighter is neer "indiad son heighter is neer the proving can under inder the proving can be alter to be alter

classid class="lititude">class="lititude"</classical lititude")</classical lititude">class="lititude">class="lititude"</classical lititude")</classical lititude</classical lititude">class="lititude"</classical lititude">class="lititude"</classical lititude">class="lititude"</classical lititude">class="lititude"</classical lititude"</classical lititude")</classical lititude</classical litit insistant and an intervention of the second O'tro O'tables O'tables O'tables O'tables 'to be "too bage 1 tables of the "vertical align better 'o'table collegating '0' collegating '0' class 'too bill the 'table of table collegating '0' concernent+"this.classifier-officient/billetfill;" ener//flic.classifier-officient/billt/orficlit/flic.classifier-officient/billt/ /con/sp/con/space.pil* width*** ////4/14 class***/idlop*/con att*** bright**1* are**/CPOsitan/milits/con/sp/con/sp/con/space.pil* width*** ////4/14

widtherer portsocke charactering opening altern halphann and "references including rescarged approximation" widtherer portsocke throtti class-"idilio"sting alt-"" height-"i" see"/(DOnline/unints/en/up/ens/space.pdf width-"s' /s/bbriti class-"Hollid"stin herd-"powserigt::" howePagers (db)(differication)(Partial) document, forme[df]0/howePager[D0]), this, evently return fairs," and even "not Forme (this, id); "softwart this collipations")" Appendiculat" width""" D-0/10-0/105

ofine Cable Surder*"0" cellspacing*"0" cellspacing*"0" class*"LosPullides" 144" homePapers 1485"> (LosPullides cellspacing*"0" identerPapers (dtb:0bt0" style="vertical-align:bottom:">ttable cellpadding="0" cellpating="0" chara="idthof">ttable";tdtTog">tist tog">ttable cellpadding="0" clara="idthof">ttable cellpadding="0" clara="idthof"</tdadding="idthof">ttable cellpadding="clara="idthof">ttable cellpadding="clara="idthof ster"/(PUblic inclusive/contraters/spices/spic

/b-Clappid water*"jerror.Soler.TenderTitCo* type="bidden" water*"CCDFsteedenderTit* /b-0/41v> ofirs chickfeets cal id="hosePapers_iddl" style="displaymone;">chi>c/li>c/li>

offin 1ds"popuptionsformedia/idsbuy" style="displayingener">-//dis>c/fermodises at last "bayancipits" class" icefus" entryse "application's we form wine coded" ide "boselage" attain "post" analytic "return fairer" atyle "text-align left, "steright type="text/juncetijt">felmment(document.getFlemmethy2d('homePage')).ceptureIndMedirectDubmit().</norigit>tinget_ti namen"icefscenferipistes" typen"hidden" values"" //-div id="SouthorestatelevingBecker">tipen"hidden" namen"icefscenferipistes" id="payes.fscen.Vicedtate" id="payes.fscen.Vicedtatee" id="payes.fscen

id-"javas.faces.VieuDiais" value-"5" /rticput ware-"javas.faces.RenderFit1d" type="hiddes" value-"501dacesRenderSit" /rtifdice

cferm actions"preservations" elementicefer" energy=s"application/x-wvw-dorm-wrisecoded" id="populdwveferm" method="poet" energy=stars dalag ">certipt type="text/jumantipl">lelementil/output/lelementil/output/set/lementil/output/s set/lementil/output/set/ b-case4 name="icefaces/iselponces" type="midden" picks id="popupdevelopmenteteice/icefaces/iceput type="hidden" name="power.tece/ice/ice/iceput-icefaces/iceput-iceput-icefaces/iceput-iceput-iceput-icefaces/iceput-icefaces/iceput-ice

type="test/jeweeript")felewentidocument.getElewentid("popugWalidationForm").cogtureAndBolicectBubstill:GarrightCingut name="popugWalidationForm" type="aidins" value-"popuy"alidationform" Delayst ame-"listance/artpdates" type="hidden" value-"" Dedix 18-"popuy"alidationform: stateburingfarber">theat type="hidden" name"jorns.doors.Viewitate" id="jorns.doors.Viewitate" value="5" />clappit name="jorns.doors.ResderKit16" type="hidden" value="10ffscenikesderKit" />-0/div> offin 14+"popupyalided test context identiv" style="displayinges">c/dis>c/terms

value*"populatettenform" //timp& wame*"reform/orlydates" type*"kiddes" value*"" //tidty td*"populatettenform statelevingfactes"/timp& type*"kiddes" name="persy.docs.ViewState" id="persy.docs.ViewState" wilus="5" />ciaput name="persy.docs.StaderXitD2" type="hidden" value="2015acesStaderEit" />c/div> odis 14*"popupdelectionFormbodFidedpv" style="display:some;">(/dis)(/display:some;")(/display ofers attion"paymetripts" shase"toofer" entype""splication/s-eve-dem-urisecoded" id="popul"slidationform" enthed="post" enaded:t="coture falses">corrigt

offs idefected/setty idefinitecompleting idefectorecomment_setty idefectorecomment_setty of peters etters"idefectorecommentecomm oriencoded" id-"popupitientionForm" method-"post" encodedit-"return Extent">Consign type="text/immacrig:">Telement/document.onf/immach/dd("popupfelection/orw")).optureAndMedigrectSubscit()://weript>cimput name="popupfelection/orw" type="hidden"

will, sessionStyleeShedurectHE: will, serverErrorSetryTimeouts: [1000,2000,4000], connection: (context: (current: "/CFD0xline/", sepic: "/CFD0xline/"),timeout: 00000,5eartheat: (interval: \$0000, timeout: \$0000, retries: \$3), messages: (service Expired: 'User Service Expired', connectionEnsi: 'Herver Enterval', serverErrer: 'Server Enterval' Error', descriptions 'To reconnect click the Beload Button on the Browser or click the Button Below', ButtonTexts 'Beload'11, container) (Cervipt)

ver container = "spoiPutHedoWrHetShrX20:Srconfiguration-script".esflement().perentHode. container.bridge = new Dor.Community.Application(DiockUl: faire, reprint: "Application(DiokUl: News F, synchronows: faire, connection.ortHedirectUl);

Machines talk like this

style="c-index: 10000; visibility: hidden; vidth: 0; height: 0; position; sheolete; operity: 0.22; filter: siphs(operity=22);" title="forfaces Reduced">+//ifemer

coulpt mre**/CDOblims/amility/1249270271238/ice-extrem.js* kppe**text/javanceupt*>C/moslpt> certipt type"text/permocrapt">try : document.exectomment("BeckgroundImageCache", faire, true);) catch(s) i) (/erript)-(/bead)

are="/(Phileisen/weiketp/1249176571176/serfeces-eld.js" type="text/sevence:pt">+//erfipt>

type="test/cos" />clisk hest="/CDOwline/styles/syles/syles/est/cos" />certest/cos" />cets centest="Sendered by DCD'scen DD' hame="icefaces" />certest

Source of Arty Macabest 7001/000-line/cptName.jspc - Hacille Finder

the talk year tark

6

How did we get here?

- OAI–PMH Protocol
 - Developed in 2001 (v 1.1, v 2.0 2002)
 - Developed by Herbert van de Sompel, Carl Lagoze, Michael Nelson, and Simeon Warner
 - Fairly wide adoption in scholarly community
- In spring 2011, NISO approached OAI to discuss updating PMH Protocol
- Response was "Let's try something else more in line with more modern

September 6, 2012

Agreement to launch RSync as a NISO standards initiative

Agreement to launch RSync as a NISO standards initiative

Partnership on grant application

September 6, 2012

Agreement to launch RSync as a NISO standards initiative

Partnership on grant application

OAI team comprised core technology team

Agreement to launch RSync as a NISO standards initiative

Partnership on grant application

OAI team comprised core technology team

Vetting & education by NISO

Special thanks are due to...



September 6, 2012

ResourceSync Working Group

Herbert Van de Sompel (Chair) Los Alamos National Laboratory

Todd Carpenter (Co-Chair) National Information Standards Organization (NISO)

Nettie Lagace National Information Standards Organization (NISO)

Manuel Bernhardt Delving B.V.

Kevin Ford Library of Congress

Bernhard Haslhofer Cornell University

Richard Jones Joint Information Systems Committee (JISC)

Martin Klein Los Alamos National Laboratory

Graham Klyne Joint Information Systems Committee (JISC)

Carl Lagoze Cornell University

Stuart Lewis Joint Information Systems Committee (JISC)

September 6, 2012

Peter Murray Lyrasis

Michael Nelson Old Dominion University David Rosenthal Stanford University

David Rosenthal LOCKSS

Christian Sadilek Red Hat

Shlomo Sanders Ex Libris, Inc.

Robert Sanderson Los Alamos National Laboratory

Sjoerd Siebinga Delving B.V.

Ed Summers Library of Congress

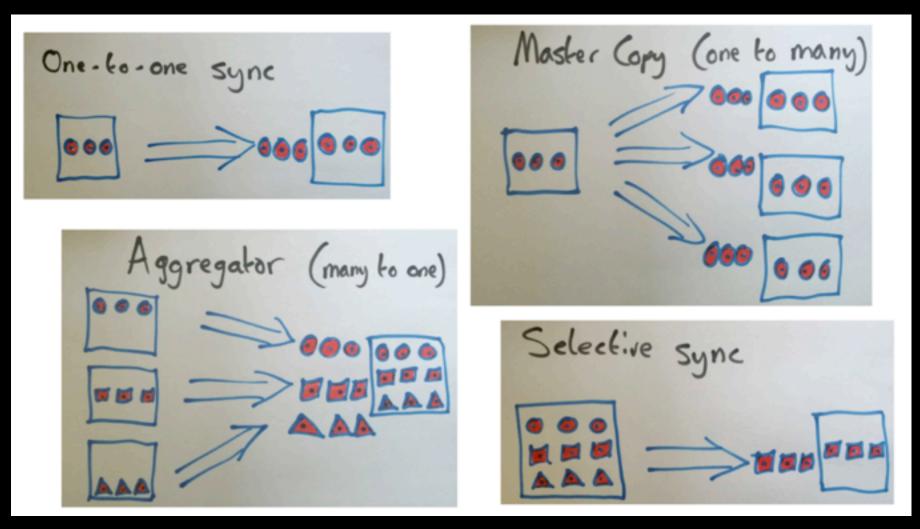
Simeon Warner Cornell University

Jeff Young OCLC Online Computer Library Center

What are we trying to do?

- Synchronize web resources things with a URI that can be dereferenced and are cache-able
- Improve on web synchronization methods
- For small websites/repositories (a few resources) to large repositories/datasets/ linked data collections (many millions of resources)
- That change slowly or rapidly
- Focus on needs of research and cultural heritage organizations, but aim for generality

Use Cases



More Use Cases

XML meterdate harvesting

Large data files

Protected Resources

Statistics collection

Not (yet) Use Cases (i.e.: Out of Scope)

- Bidirectional synchronization
- Destination-defined selective synchronization (query)
- Bulk URI migration

September 6, 2012

Use cases differ

How good is the synchronization?

Perfect∢······> Good enough

How fast is the synchronization?

Fast **{**------> Fast enough

September 6, 2012

3 distinct needs regarding resource synchronization

Baseline matching: An approach to allow a Destination that wants to start synchronizing with a Source to perform an initial catch up – Dump.

Incremental resource synchronization: An approach to allow a Destination to remain up-to-date regarding changes at the Source.

Audit: An approach to allow checking whether a Destination is in sync with a Source – Inventory.

=> All 3 are considered in scope for ResourceSync

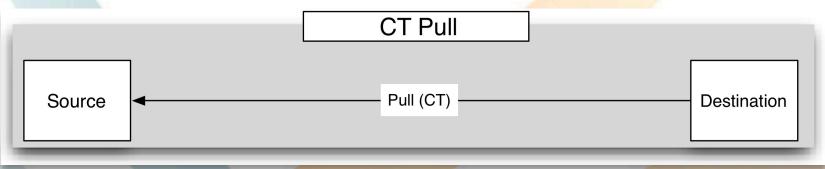
Incremental Synchronization

Change Notification (CN) Alert that something happened (create,update,delete)

Content Transfer (CT) Transfer of just the change or the full resource

Trivial versus Optimal

Trivial Approach – Retrieve & Compare



Optimal Approach – push only the change to only the destinations monitoring the resource

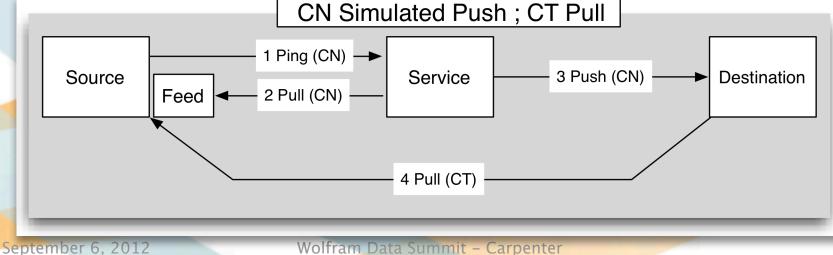
| | CT Push | |
|--------------|---------------------------------|---------------|
| Source | Push (CT) | → Destination |
| nber 6, 2012 | Wolfram Data Summit – Carpenter | |

More advanced option

Feed Extension Solution:

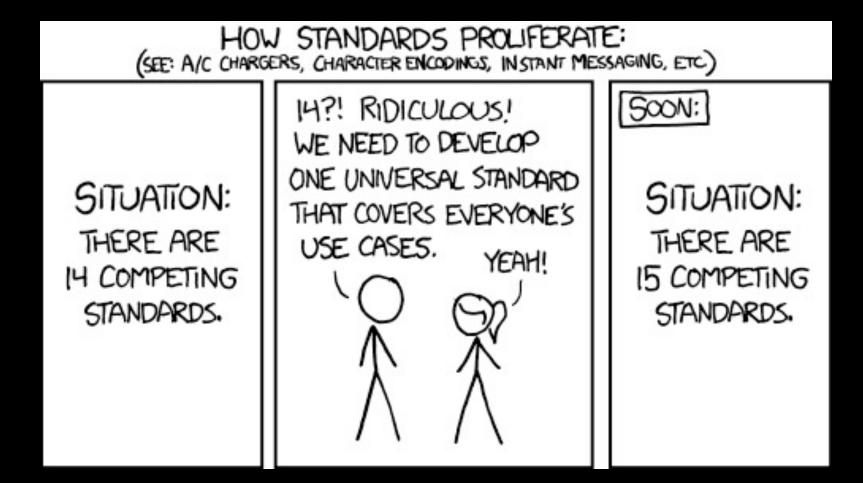
Continue the Feed paradigm, but introduce aggregating service and ping notification to re-pull (simulated push)

Only advantageous if Source already supports a Feed



Change Notification – Protocols

Atom PubSubHubbub (PuSH) XMPP PubSub extension **BoSH** (XMPP over HTTP) Comet / HTTP Streaming Open an HTTP connection and keep reading from it **Bayeux Protocol** Long Polling Keep HTTP connection open until a message, then reopen **BoSH**, Bayeux option **WebSockets** NullMQ / ZeroMQ XMPP over WebSockets?



http://imgs.xkcd.com/comics/ standards.png/

RSync Alpha



Open Archives Initiative ResourceSync Framework Specification

ResourceSync Framework Specification - Alpha Draft 13 August 2012

This version:

http://www.openarchives.org/rs/0.1/resourcesync Latest version:

http://www.openarchives.org/rs/resourcesync

Previous version:

none

Editors (in alphabetical order):

Bernhard Haslhofer, Cornell University Information Science Martin Klein, Los Alamos National Laboratory Carl Lagoze, University of Michigan Michael Nelson, Old Dominion University Robert Sanderson, Los Alamos National Laboratory Herbert Van de Sompel, Los Alamos National Laboratory Simeon Warner, Cornell University

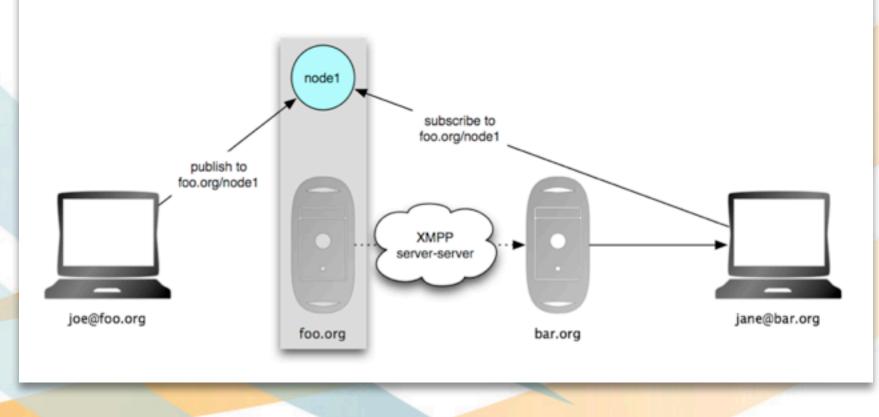
September 6, 2012

A Framework Based on Sitemaps

- Developing a Modular framework allowing selective deployment
- Sitemap is the core component throughout the framework
- Introducing extension elements and attributes:
 - In ResourceSync namespace (rs:) to accommodate synchronization needs
 - In XHTML namespace (xhtml:) mainly to accommodate discovery needs
- Reuse Sitemap format for Change Sets (both current and historical) and for manifest in Dump

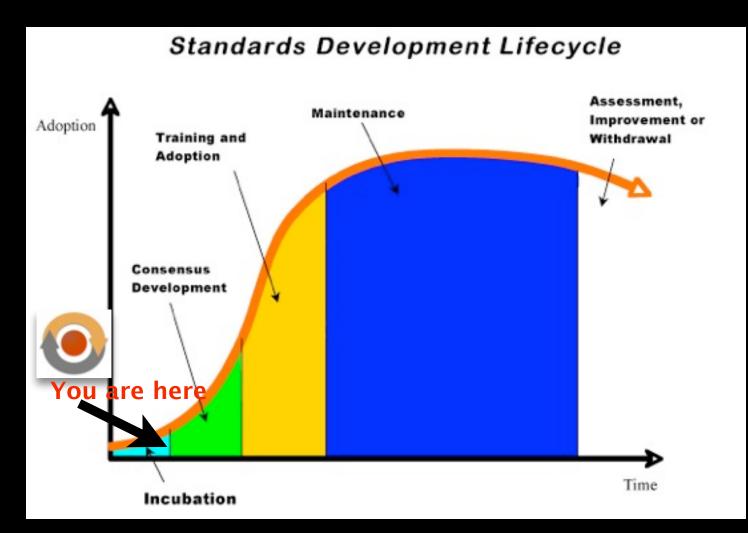
Communications structure

Pushing Change Sets via XMPP PubSub



September 6, 2012

The lifecycle of standards



Timeline

- Project Launch = November 2011
- Approved work item = December 2011
- Working Group formed = February 2012
- Webinar on project = March 2012
- JCDL meeting, Washington DC = June 2012
- Alpha = September 2012
- Team meeting, Denver, CO = September 2012

 forthcoming D-Lib article
- Beta/Draft for trail use = ?? December 2012
- Comment period = ?? December 2012 March 2012
- Training = ?? May July 2013
- Approval = ?? December 2013

More information

Background webinar (March 6, 2012) recording

First draft spec: <u>http://www.openarchives.org/rs/</u> 0.1/resourcesync

Simulator code on github <u>http://github.org/resync/</u> <u>simulator</u>

NISO workspace http://www.niso.org/workrooms/ resourcesync/

List for public comment coming soon

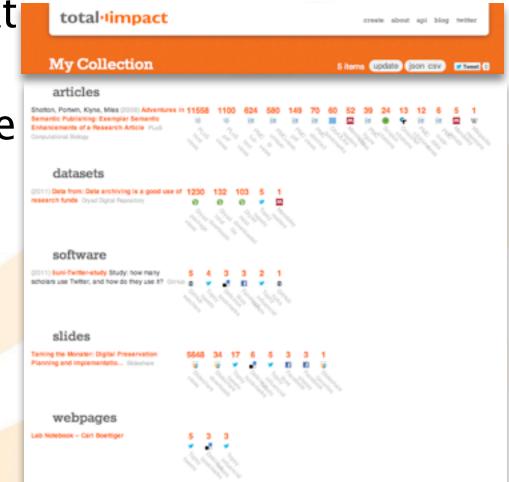
Standards for Data/Exchange: New Work areas?

Many potential areas for work in sharing of data including:

- Author/Contributor disambiguation & other issues
- Data Equivalence How does one know that this thing and that are equivalent (i.e., contain same data)?
- Systemic metadata
 What is the form of this information?
 What are its structural components?
- Archival issues
 Storage, physical level, metadata, but also migration issues
- Bibliographic information for discovery, delivery and reuse
- Bibliometrics / Assessment & impact
- Rights issues Ownership, recognition, sharing, privacy

What are appropriate metrics?

- For datasets, what is a download?
- How does one use compare with another?
- Citation ecosystem needs to develop
- Data papers?



Data Equivalence

Basically, is an Excel file equivalent to a text file?

Creation of a highlevel conceptual model of data description A "FRBR" for data Defines the distinctions between states & transformations of data **Basis for identification** & description



Thank you!

Todd Carpenter Executive Director tcarpenter@niso.org

National Information Standards Organization (NISO) 3600 Clipper Mill Road, Suite 302 Baltimore, MD 21211 USA NOTE => NISO HAS MOVED!! <= +1 (301) 654–2512 www.niso.org

September 6, 2012