Member Nodes - MNDeployment #3748

PPBIO Member Node

2013-05-08 15:07 - Ben Leinfelder

Status:	Operational	**Start date:**	2014-08-29
Priority:	Normal	**Due date:**	
Assignee:	Jing Tao	**% Done:**	100%
Category:		**Estimated time:**	0.00 hour
Target version:	Operational		
Latitude:	-3.17	**MN_Date_Online:**	2016-05-17
Longitude:	-60.00	**Name:**	
MN Description:	A DataOne member node containing mainly Occidental Amazonian datasets	**Logo URL:**	
Base URL:	http://ppbio.inpa.gov.br/	**Date Upcoming:**	
NodeIdentifier:	urn:node:PPBIO	**Date Deprecated:**	
MN Tier:	Tier 1	**Information URL:**	
Software stack:	Metacat	**Version:**	

Description

On IRC, there was someone asking about how KNB members could get their content into DataONE. Turned out David ("dvd") is the new information manager for PPBIO and they may want to set up an independent member node. Right now we have replicated content in the KNB from the PBBIO node (Brazil) but have not released it into DataONE.

I have asked David to send us more details and contact information so we can effectively plan and/or assist in that effort.

Subtasks:

Task # 6175: PPBIO: Planning	**Closed**
Task # 6176: PPBIO: Join DataONE	**Closed**
Task # 6179: PPBIO: Create MN description document	**Closed**
Task # 6182: PPBIO: Developing	**Closed**
Task # 6183: PPBIO: Design, code and component test a new MN implementation	**Closed**
Task # 6184: PPBIO: Local Testing	**Closed**
Task # 6185: PPBIO: Verify that MN passes the Web Tester	**Rejected**
Task # 6186: PPBIO: Verify that MN passes the Replication Tester	**Rejected**
Task # 6187: PPBIO: Testing	**Closed**
Task # 6188: PPBIO: Registration in environment	**Closed**
Task # 6189: PPBIO: Register new Science Metadata formats	**Closed**
Task # 6190: PPBIO: SSL Certificates	**Closed**
Task # 6191: PPBIO: Generate client certificate.	**Closed**
Task # 6192: PPBIO: Verify successful installation of client side certificate	**Closed**
Task # 6193: PPBIO: Verify successful installation of server side certificate	**Closed**
Task # 6194: PPBIO: Register MN	**Closed**
Task # 6195: PPBIO: Set up Node document	**Rejected**
Task # 6196: PPBIO: Set the node status to approved (start synchronization)	**Closed**
Task # 6197: PPBIO: Synchronization	**Closed**
Task # 6198: PPBIO: Set up synchronization of the MN	**Closed**
Task # 6199: PPBIO: Content Review	**Closed**
Task # 6215: PPBIO: Replication testing (if Tier 4)	**Closed**
Task # 6216: PPBIO: Registration in PRODUCTION environment	**Closed**
Task # 6217: PPBIO: Register new Science Metadata formats	**Closed**
Task # 6218: PPBIO: SSL Certificates	**Closed**
Task # 6219: PPBIO: Generate client certificate.	**Closed**

Task # 6220: PPBIO: Verify successful installation of client side certificate	**Closed**
Task # 6221: PPBIO: Verify successful installation of server side certificate	**Closed**
Task # 6222: PPBIO: Register MN	**Closed**
Task # 6223: PPBIO: Set up Node document	**Closed**
Task # 6224: PPBIO: Set the node status to approved (start synchronization)	**Closed**
Task # 6225: PPBIO: Synchronization	**Closed**
Task # 6226: PPBIO: Set up synchronization of the MN	**Closed**
Task # 6227: PPBIO: Content Review	**Closed**
Task # 6228: PPBIO: Verify Science Data	**Closed**
Task # 6229: PPBIO: Verify Science Data content	**Closed**
Task # 6230: PPBIO: Verify that the Science Data is returned with the correct HTTP Cont...	**Closed**
Task # 6231: PPBIO: Verify Science Metadata	**Closed**
Task # 6235: PPBIO: Verify Resource Maps	**Closed**
Task # 6236: PPBIO: Verify Resource Map content	**Closed**
Task # 6237: PPBIO: Verify Resource Map is returned with the correct HTTP Content-Type	**Closed**
Task # 6238: PPBIO: Verify that Resource Maps are correctly processed by CNs.	**Closed**
Task # 6239: PPBIO: Authentication and Authorization	**Closed**
Task # 6240: PPBIO: Science Data access	**Closed**
Task # 6241: PPBIO: Science Metadata access	**Closed**
Task # 6242: PPBIO: Log record access	**Closed**
Task # 6243: PPBIO: Replication testing (if Tier 4)	**Closed**
Task # 6244: PPBIO: Transition to production	**Closed**
Task # 6245: PPBIO: Mutual acceptance	**Closed**
Task # 6246: PPBIO: Verify content available for Current MNs web page	**Closed**
Task # 6247: PPBIO: Create legal documents	**Closed**
Task # 6248: PPBIO: Create news item	**Closed**
Task # 6249: PPBIO: Formal announcement	**Closed**
Task # 7773: Test content in PPBio in production	**Closed**
Task # 7772: Purge and remove mnTestPPBIO from the cn-stage environment	**Closed**
Task # 7808: Please fix ecogrid in PPBio's metadata view in search results	**Closed**
Task # 8493: PBIO: Support search against deprecated identifiers	**Closed**
Task # 8538: PBIO Certificate expiring	**Closed**

History

#1 - 2013-05-15 20:19 - Ben Leinfelder

- Description updated

From David:
I'm having trouble in contact but "chris" said x.509 certs are already on way to us.
Peld is like a subproject from PPBio. Peld data comes just from long term research.
We host both projects on INPA here on Manaus.

#2 - 2013-05-24 14:44 - Chris Jones

In terms of certificates, I told David we would be creating DataONE-signed certs. We haven't done so yet.

#3 - 2013-05-24 18:12 - Matthew Jones

- Target version set to Deploy by end of Y4Q4

- Start date set to 2013-05-24

- Due date set to 2013-07-31

#4 - 2013-05-24 18:13 - Matthew Jones

From Flávia Pezzini, regarding the relationship between PPBIO and PELD member nodes:

"Our plan is the same, today PPBio is replicating with PELD server, and PELD is connected with KNB. We want PPBio to become a member node and PELD data replicated to it and not all of PELD data will be available on DataONE. I´ve asked David to cc me in the conversations. Hope everything works out fine! "

#5 - 2013-07-11 03:31 - Debora Drucker

PPBio Node is working on having the MN Tier 4 Node operational by Sep/2013. Server is being transfered to LNCC (Petropolis/RJ) and deployment in staging environment is planned to take place in august.

#6 - 2013-08-01 19:50 - Bruce Wilson

- Target version changed from Deploy by end of Y4Q4 to Deploy by end of Y5Q2

- Due date changed from 2013-07-31 to 2014-01-31

#7 - 2013-08-05 15:59 - Laura Moyers

- Longitude set to -60.00

- Latitude set to -3.17

#8 - 2013-09-06 18:45 - Matthew Jones

Update from Debora Drucker on Sept. 4:

PPBio´s server content was successfuly copied to a new server at LNCC. We are now at the stage of exchanging certificates. LNCC aparantly is able to generate thrid parties certificates. Peld server will be kept at Inpa and PPBio server will start to operate from LNCC. Instead of configuring replication between PPBio-KNB, we could configure the new server to replicate to DataONE. The idea is to keep replication between PELD a nd PPBio servers. We´re having a conference between the brazilians involved next monday.

#9 - 2013-10-08 18:30 - Bruce Wilson

- Status changed from New to Planning

#10 - 2013-10-08 18:33 - Bruce Wilson

- Base URL set to http://ppbio.inpa.gov.br/

- MN Tier set to Tier 4

- Software stack set to Metacat

Node to be hosted at Laboratório Nacional de Computação Científica (LNCC, in Rio de Janeiro). Telecon on 2013-10-08. Currently have a running Metacat instance at LNCC. Need to address client certificates.

#11 - 2013-11-06 03:30 - Matthew Jones

Debora indicated that content was successfully copied to the LNCC Metacat node, and is now available at quickly let you know that finally ppbio content was replicated at lncc http://ppbio-amoc.lncc.br/knb/style/skins/ppbio/ .

Need to contact them to determine whether they are ready for next steps in configuring the node.

#12 - 2014-01-17 15:40 - Bruce Wilson

- Target version changed from Deploy by end of Y5Q2 to Deploy by end of Y5Q3

- Due date changed from 2014-01-31 to 2014-04-30

#13 - 2014-04-14 13:57 - Laura Moyers

- Due date changed from 2014-04-30 to 2014-07-31

- Target version changed from Deploy by end of Y5Q3 to Deploy by end of Y5Q4

#14 - 2014-04-22 18:00 - Matthew Jones

Update as of 2014-04-22: Discussed the status of the PPBIO installation with Debora Drucker, who indicated that they have decided to change plans in terms of where the node will be installed. Although they may eventually succeed in getting a node configured and installed at LLNC, they have decided that it would be more effective to get their existing Metacat installation that is running at INPA in Manaus configured as a DataONE node. Then at a later date they might add another node at LLNC and configure it to replicate data, but that would be decided later. The current plan is to move the current PPBio Metacat instance onto new hardware that they purchased and is running at INPA, and then configure that to be a Member Node. As Ben and I already helped get the certificates and other software running for them on PPBIO, that move should be straightforward. Debora and David will contact Ben and I if they need assitance with the technical server move and configuration, and Debora will update this ticket with a planned timeline once she discusses it with others at INPA.

#15 - 2014-05-05 20:25 - Debora Drucker

David, Bill, Livia, Flavia and I discussed and our plan is to:
- Have server at INPA configured to become a Member Node by May 30th
- Testing until June 20th
- Register server in production - June 20th
- MN operational / MN deployment announcement - June 27th

#16 - 2014-05-15 17:43 - Bruce Wilson

- Status changed from Planning to Testing

#17 - 2014-08-06 18:30 - Laura Moyers

- Longitude changed from -60.00 to 60.00

- Due date changed from 2014-07-31 to 2014-10-31

- Target version changed from Deploy by end of Y5Q4 to Deploy by end of Y1Q1

- MN Description set to A DataOne member node containing mainly Occidental Amazonian datasets

- NodeIdentifier set to urn:node:PPBIO

MN Description taken from the stage-2 Node Document list - probably needs to be revised both here and in the Node Doc for production.

#18 - 2015-01-23 16:21 - Laura Moyers

- Target version changed from Deploy by end of Y1Q1 to Deploy by end of Y1Q2

#19 - 2015-03-10 13:14 - Laura Moyers

- Assignee changed from Ben Leinfelder to Mark Servilla

- Target version changed from Deploy by end of Y1Q2 to Deploy by end of Y1Q3

PPBio is upgrading their metacat. Probably unrelated, we're seeing an incomplete harvest of content in stage-2.

#20 - 2015-05-18 14:21 - Laura Moyers

- Target version changed from Deploy by end of Y1Q3 to Deploy by end of Y1Q4

David indicated that their content is ready to move to production, but they have some hardware (new server) and software (certificate renewal) issues to address first. Move out to Q4 (30 Sept).

#22 - 2015-09-16 16:27 - Laura Moyers

- Target version changed from Deploy by end of Y1Q4 to Deploy by end of Y2Q1

PPBio has a new person on staff, Tim Vincent, who will be working with this.

Tim Vincent tim.in.manaus@gmail.com

#23 - 2015-10-22 01:06 - Laura Moyers

Tim says they have successfully updated their Metacat but they still need to test out uploading with Morpho and fix a few things. Next step is to set up the PPBio MN in stage (not stage-2).

#24 - 2015-11-11 15:22 - Laura Moyers

- Target version changed from Deploy by end of Y2Q1 to Deploy by end of Y2Q2

#25 - 2015-12-09 19:14 - Laura Moyers

Tim Vincent met with Jing, MarkS and Laura today to run through the next steps. Tim hopes to be able to configure the MN tomorrow and start synchronization. The content is the same as it was when David testing in stage-2, so we anticipate few if any content-related errors.

#26 - 2016-03-21 15:21 - Laura Moyers

- Target version changed from Deploy by end of Y2Q2 to Deploy by end of Y2Q3

#27 - 2016-05-17 16:23 - Laura Moyers

- Status changed from Testing to Operational

- Target version changed from Deploy by end of Y2Q3 to Operational

- MN Tier changed from Tier 4 to Tier 1

PPBio is currently operating as a Tier 1 MN. There was some discussion about operating at Tier 4 and in fact some testing was done for Tier 4 functionality, but it was decided to go to production as Tier 1 at first.

#28 - 2016-06-02 15:59 - Laura Moyers

- MN_Date_Online set to 2016-05-17

#29 - 2017-09-07 00:10 - Laura Moyers

Email conversation between Tim and Mark 8/23/17:

Tim:
I would appreciate your feedback on the following....

The datasets that we upload to Metacat become available in 3 locations..
1) our metacat server
2) DataOne and
3) SiBBr (www.sibbr.gov.br/)

all with the same ID, but not, obviously, the same URL. Using the ID PPBioAmOc.171.1, in Google search, for example, does not generate a result.

Mark:

Google's search engine bots will only index content that it can access and if there is no robot.txt file that requests Google not to index the content. Google can traverse web forms, but only if the values can be selected (e.g., drop down lists, radio buttons, but not typed in).

Tim:

In Brasil, there is an online database of curricula called lattes (http://lattes.cnpq.br/), where it would be useful to put the ID (like a doi) and enable the dataset to be easily found, but what is the best way to this without listing 3 URLs (maybe one of the servers is offline), if a google for the ID doesn't find the dataset?

Mark:

The best approach is the one you mention - the use of DOIs. Short of that, I would recommend using the DataONE resolve method on the coordinating node, which will display all of the locations for the object identified by the PID in the DataONE federation, albeit in XML. For PPBIO science metadata (i.e., EML), this would be your Metacat and the CN. For example, resolve for "david.3.2" returns the following:

curl -s -X GET https://cn.dataone.org/cn/v2/resolve/david.3.2
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>

david.3.2

urn:node:PPBIO
 https://ppbiodata.inpa.gov.br/metacat/d1/mn
v1
v2
 https://ppbiodata.inpa.gov.br/metacat/d1/mn/v2/object/david.3.2

urn:node:CN
 https://cn.dataone.org/cn
v1
v2
 https://cn.dataone.org/cn/v2/object/david.3.2

/ns2:objectLocationList

So maybe providing the URL to the CN resolve for each object PID would work? Since SiBBR is unknown to DataONE, it will not be in the list (QED).

Tim:

Even if I copy and paste the citation from DataOne, google doesn't find the dataset.

Mark:

I don't know if Google indexes DataONE - probably a simple question to the right DataONE brain trust will provide an answer.

Tim:

For the moment perhaps we will have to list the 3 URLs or just our server or yours...SiBBr is not as reliable....we await funds for EZID, but in the meantime what are your thoughts?

Mark:

Not to rain on your EZID/DOI parade, but EZID is no longer accepting contracts for DOI registration outside of the University of California system. They are referring existing and new requests to DataCite. Having said that, DataONE is planning on filling this role in the near future - we'll keep you informed on this point.

#30 - 2018-01-09 21:09 - Dave Vieglais

- Sprint set to Ongoing Operational

#31 - 2018-01-16 20:28 - Amy Forrester

email bouncing: Bill Magnusson bill@inpa.gov.br
1/16: query Tim Vincent tim.in.manaus@gmail.com

#32 - 2018-03-09 17:17 - Amy Forrester

2/28/18: {Tim vincent}: I recently uploaded the following to our Metacat, but I don't see it on DataOne.
https://ppbiodata.inpa.gov.br/metacatui/#view/PPBioAmOc.221.3

> {reply}everything is working as expected. The culprit seems to be an unusually large backlog of work. We just need you to hang tight a little longer and all should settle out.

3/9/18: {Tim Vincent}: I don't know if you can answer this, or if I need to submit a ticket, but when a dataset is updated on Metacat the number changes from, say, PPBioMA.28.8 to PPBio.28.10.

Now, Metacat picks this up and advises that there is a newer version and links to the newer version. However, DataOne doesn't seem to be doing this.

This means that a researcher who submits a DataOne linlk to their work in an artcle, will later have to request an ammendment which is somewhat inconvenient.

Can you shed some light on this?

> {Dave} Not sure that they mean. If you look at PPBioMA.28.8 in the search page: https://search.dataone.org/index.html#view/PPBioMA.28.8 there is a clear statement at the top of the page that a newer version is available.

> {Chris} If PPBio is inserting content into their Metacat with Morpho (which I think they are, but would need to confirm) - it uses the identifier scheme with the revision like scope.id.rev. Morpho uses the old Metacat API to insert content, not the DataONE API, and so Metacat sees that and will auto-generate a SystemMetadata document for the new content. For updates, it also sets the obsoletes/obsoletedBy fields in the sysmeta. DataONE indexes that relationship when it harvests the new content. So yeah, I'm not entirely understanding the issue, but perhaps it's because of a harvesting delay? Maybe they're not seeing it right away?

> {Jing} Metacat itself will not give the warning . I know morpho will give, perhaps MetacatUI will give you as well.

- {reply} Are you still experiencing harvesting delays? If you look at PPBioMA.28.8 in the search page:
  https://search.dataone.org/index.html#view/PPBioMA.28.8 there is a clear statement at the top of the page that a newer version is available.

#33 - 2018-03-12 15:07 - Amy Forrester

3/12/18: {Tim vincent}: Two issues
1. Searching for "PPBioMA.28.8" did not return any results, however searching for the current revision "PPBioMA.28.10" did return the expected record. **We will look into supporting search against deprecated identifiers.**
2. Replication: *It is important that our actual data is on another server, because here in Brazil, at INPA, the service infrastructure is not good.*
* PBIO replication policy set to: replicationAllowed set to "false" in the system metadata

   Sent information re: setting value in the system metadata

#34 - 2018-03-19 16:06 - Amy Forrester

3/19/18: sent follow-up email

   {reply from Tim Vincent} The replication got a bit complicated (threw up an error possibly related to a certificate) so I registered with slack to get a better line of communication with Matt, but I am really busy with something else for a few days.

   I will return to this towards the end of the week.

#35 - 2018-03-26 17:05 - Amy Forrester

- Longitude changed from 60.00 to -60.00

#36 - 2018-08-31 15:20 - Mark Servilla

- Assignee changed from Mark Servilla to Jing Tao