In our meeting on 6/30/2020, we discussed what to do about metadataID's in the METADATA table that are listed twice, associated with different documents each time. We talked about how in several cases, one doc is probably the field protocols and the other is the lab protocols—they can probably be combined.

Here, I'm listing all cases of the repeated metadataID's, some summary information about what information the documents contain, and what I propose to do/how I'll resolve the issue.

1.
BenthicBacterialProduction.Analysis.20110616

- BenthicBacterialProduction.Analysis.20110616.docx includes information on sample prep and incubations.
- BenthicBacterialProductionTemplate.xlsx is an Excel template for how to enter data/do calculations. It is referred to in the Analysis .docx, though not very conspicuously.

  Proposed solution: Make the reference to the .xlsx document a little clearer (maybe boldface?) and eliminate the .xlsx row from METADATA. Also could edit the .xlsx to include a reference to the protocol described in the .docx.

[DONE]

2.
BenthicPrimaryProduction.Analysis.20110714

- BenthicPrimaryProduction.Analysis.20110714.docx includes information on CO2 concentrations in cores and rate calculations. Links to .xlsx file.
- BenthicPrimaryProductionTemplate.xlsx: see above.

  Proposed solution: Same as 1.

[DONE]

3.
GC5890.CH4.CO2.20110601

- FIDMETH_081111.TXT shows history of changes to the methods, as well as parameters (maybe for the machine used in analysis?)
- GC5890.CH4.CO2.20110601.docx is the experimental protocol. It does not refer to the .TXT file.

  Proposed solution: Add a reference to the .TXT doc in the .docx, and delete the row in METADATA that refers to the .TXT.
  We discussed adding the contents of the .txt file to the .docx in its entirety, instead of just linking the two files, but since the text files have such a particular format, we elected to keep them separate.

[DONE].

4.
GC5890.CH4.CO2.20120618

- CH4_CO2.TXT (similar to 3)
- GC5890.CH4.CO2.20120618.docx (similar to 3)

Proposed solution: Same as 3.
[DONE].

5.
Iso.DIC.20111014

- IRMS with Gas Bench at CEST.docx gives protocol for gas samples.
- DIC Isotope Data.r description at the top says "program to sort through DIC isotope data from the CEST gasbench IRMS. Pull out and average data for each sample". I haven't gone through the script in detail beyond that, but it looks like it's a data processing script, basically.

Proposed solution: Maybe just include a reference to the R script in the .docx? There are no other metadataID's that refer to R scripts (except for 10. below), so I don't think that's a precedent we want to set.
[DONE]

6.
Limno.Sample.20120501

- Limno Field Procedure_Routine_UNDERC 2012.docx describes strictly field methods
- Limno Lab Procedure_Routine_UNDERC 2012.docx describes strictly lab methods, including both before and after lake sampling.

Proposed solution: Combine these into one document (especially important, in my opinion, because the way it's organized now, someone looking to follow the protocol would have to first look at the lab file, then at the field file, and then at the lab file again when they get back). New doc could be called e.g. Limno Procedure_Routine_UNDERC 2012.docx, or Limno Field Lab Procedure_Routine_UNDERC 2012.docx, whatever is clearest.

[DONE]. New doc name is Limno_Procedure_Routine_UNDERC_2012.docx

7.
Limno.Sample.20130416

- Limno Field_Routine_UNDERC 2013.docx strictly field methods
- Limno Lab_Routine_UNDERC 2013.docx strictly lab methods

Proposed solution: Same as 6.

[New aggregate file created and uploaded: Limno_Procedure_Routine_UNDERC_2013.docx. DONE]

8.
Limno.Sample.20160505

- Limno Lab_Routine_UNDERC 2016.docx strictly field methods
- Limno Field_Routine_UNDERC 2016.docx strictly lab methods

Proposed solution: Same as 6. Note additionally that the description in METADATA is empty for this one, whereas the others describe how they differ from previous files, etc. The files themselves don't have any indication of how they differ from previous versions. If desired, I can comb through them and try to spot the differences, and then add that to the METADATA table.

RANDI: Thought I could look through them with context. Here are the differences I see between the most recent (2013) and the current (2016).

-2016 removes the "East and West Long Lake" specification, allowing the 2016 version to be used on all lakes across projects (eg. Olson 5, Monitoring, Long Lake)

-2016 removes collection of DIC Isotopes and pCO2 Isotopes as part of regular sampling.

-2016 changes the method for collecting pCO2 samples. In 2013 they were taken separately; in 2016 they are taken from the same Van Dorn pull as the other gas and water chem samples.

-2016 adds Inlet/Outlet sampling methods.
And there are no changes between 2013 and 2016 for lab methods.

[Deleted old file and replaced it with new combined file: Limno_Procedure_Routine_UNDERC_2016.docx. DONE.]

9.
Methane.Sample.20110601

- Jones Lab Methane Survey 2011_Field Work.docx strictly field methods
- Jones Lab Methane Survey 2011_Lab Work.docx strictly lab methods

Proposed solution: Same as 6.

[Replaced old files with new combined file: Jones_Lab_Methane_Survey_2011.docx. DONE]

10.

MethSurv.Sample.20120522

- MethSurv.Sample.20120522.docx already includes both field methods and before- and after-lake lab methods. I could use this as a model for what the combined docs should look like in 6-9.
- UNDERC_GCdata_Rules_040213.docx gives rules for GC data. Not linked in the .docx above, and although this rules doc lists a couple of affiliated metadataID's, MethSurv.Sample.20120522 is not among them.
- UNDERC_GCdatabaseAppending.R is a script for appending GC data to the database.
- UNDERCgcProcessing.R is a script for processing GC data.

Proposed solution: Refer to the rules .docx in the first .docx [DONE], and refer to the MethSurv.Sample.20120522 metadataID in the rules .docx [DONE], but remove the rules .docx from METADATA. The reason I think we should do this is that the rules .docx already refers to a couple of other metadataID's, both of which are listed in METADATA, but neither of which has a row that points to the rules file. Additionally, remove the R scripts from METADATA--same as 5.

[DONE]

If we're going to add a written reference to UNDERC_GCdata_Rules_040213.docx in MethSurv.Sample.20120522.docx, then we should probably add similar references to the metadata .docx docs for the other two metadataID's referred to in the rules doc. Otherwise, how will people know to go and look at the rules doc? [Added references. Note: I can't find metadataID GC5890.CH4.CO2.20130402: may be outdated?]

11.

Nutrients.20110601

- JonesLab Phosphorus Protocol.docx protocol for phosphorus measurements
- Nitrate or TN Determination with Spec.docx protocol for TN/nitrate measurements

Proposed solution: If I'm understanding correctly, I think this is one case where this metadataID probably should be separated into two different metadataID's in accordance with what's being measured. How I would go about doing this: find tables where this metadataID comes up (likely starting with NUTRIENTS). For each sample, determine whether it's a P or an N sample, and change the metadataID accordingly. If there are any that can't be determined, make a list of those and check them with you. If there are any where the metadataID should refer to a field protocol instead, see if we can find a field protocol doc (I couldn't find one, but maybe you're aware that there's one that this probably refers to?)

[DONE]