Data and Analysis Preservation for RHIC Experiments: the current outlook

M.Potekhin
BNL Nuclear and Particle Physics Software (NPPS) Group
on behalf of Brookhaven National Laboratory

Data & Analysis Preservation Workshop
CERN
November 26th, 2019

Data and Analysis Preservation

- We'll use the term DAP to refer to data **and** analysis preservation
- DAP is commonly described as a union of
  - bit preservation
  - software Infrastructure and application code
  - analysis know-how (knowledge management)
- The goal is to have a reproducible analysis capability over a long period of time
  - also the capability to perform a modified or new analysis within the same framework
  - extension of short-term reproducibility typically required to ensure quality of the analysis
- Funding agencies in the US increasingly require both new and **existing** experiments to plan/to have DAP capabilities
  - the challenge: nobody in the research community has “spare cycles” so implementing DAP is hard or impossible w/o additional resources allocated to it
- In the following, we shall consider some common DAP issues of RHIC experiments but focus mainly on PHENIX as perhaps the most challenging and representative example

RHIC Experiments

- BRAHMS, PHENIX, PHOBOS, STAR... and sPHENIX
- Status in reverse chronological order:
  - sPHENIX - a few years before start of data taking, computing model in development
  - STAR - active, ongoing, large number of publications annually
  - PHENIX - finished data taking in 2016
    - active analysis ongoing with O(10) papers published annually
  - BRAHMS (finished operations in 2006): 2 papers published since 2010 (most recent in 2016)
  - PHOBOS (finished operations in 2005): 5 papers published since 2010 (most recent in 2016)
- For STAR, PHENIX, BRAHMS and PHOBOS, “bit preservation” is in place i.e. the experiment’s data is preserved by the BNL Scientific Data and Computing Center (SDCC) according to current standards
  - but analysis preservation for the two latter experiments will probably be impractical due to entropy - people moved on a long time ago, hence lack of available expertise/critical mass/available effort
- It is interesting to consider sPHENIX, STAR and PHENIX as different points in the lifecycle of an experiment - future, current, and analysing previously collected data

RHIC Experiments (cont'd)

- sPHENIX
  - experienced (but small at this point) software and computing team actively working on the computing model, production and analysis framework
  - ample opportunities to plan ahead and implement a robust DAP plan
  - execution will depend on the effort available for DAP
- STAR
  - a well-coordinated software organization with a broad spectrum of expertise
  - current practices include analysis code review, stringent requirements for analysis notes (e.g. level of detail including technical)
  - experience with containers, demonstrated portability (including DB snapshots)
  - most ingredients in place for DAP
- PHENIX
  - production continues, and many active analyses under way
  - well maintained general infrastructure
    - build and release procedures in place, production jobs run in containers etc
  - ...but there are challenges - see next slide

DAP Challenges in PHENIX

- Software and infrastructure documentation (becoming obsolete)
- Personnel (leaving or migrating to other projects)
  - continuity of knowledge is a problem
  - software effort underfunded, so then -
- Little effort available for active management of the analysis software
  - lacking revision control, packaging and archiving of the **analysis** software developed mainly by individual researchers
- Packaging of certain calibrations/conditions types of data (files outside of DB)
  - also, custom “data artifacts” created in individual analyses e.g. augmented dead or hot/warm channel maps which are hard to trace and reproduce
- More in the backup section...

Organizing DAP in PHENIX

- A DAP Task Force has been created, meetings are held periodically
  - includes documentation and analysis coordinators for PHENIX, members of the Nuclear and Particle Physics Software Group (NPPS)
- An initial DAP plan has been drafted - work in progress
- In the RHIC Science and Technology Review which took place in September 2019 there was strong support for addressing DAP
  - a separate PHENIX review in December, DAP is on the agenda
- Collaboration is/will be seeking support for dedicated effort
  - key to the success of this effort

Software Landscape

- Both STAR and PHENIX originated in the GEANT 3 era
- Simulation tools and geometry description were all based on GEANT 3
  - "starsim" framework in STAR
  - "PISA" framework in PHENIX
- Most tools, production and analyses are however done with ROOT
- In each experiment there is a translation layer/tools for geometry and data
- For true reproducibility GEANT 3 and related software need to be preserved in a working condition
  - should we be able to build this software?
  - or are we content keeping existing binaries/libraries in operation?
  - can the VM/container approach help?
- Of course ROOT itself is not a static piece of software either
  - need to preserve specific versions and libraries
- PHENIX uses the "Fun4All" processing framework developed by the Collaboration
  - to achieve reproducibility, documentation needs to be updated, examples and tutorials refreshed as many analysis macros are using Fun4All machinery
- Last but not least - VMs and containers to preserve the environment, OS and beyond

Current work in PHENIX aligned with DAP

- Improvements in simulation and analysis software organization are aligned with DAP
- Looking primarily at embedding procedures in a few types of analyses as examples of complex workflows and prime candidates for preservation
  - supported by NPPS
- Capturing user/analysis code in custom repos (used as scratch space for now)
  - making changes to reduce or eliminate extraneous dependencies, cleanup etc
- Creating and/or improving documentation
- Often requires reverse engineering (see next slide)
- The interim goal is to create “self-contained” examples of analysis that are essentially ready for preservation
  - i.e. collections of software, data products, documentation, containers and other components as required
  - does not necessarily need to comply with a strict template as analyses are different

An example of an actual user directory used in analysis

```
check_pdst_vertexfile.root
condor_bb.job
condor_cc.job
condor_hadrons.job
condor.job
condor_jpsi.job
condor_ktomu.job
condor_step3.job
corruptfiles
CorruptFiles_bb.txt
CorruptFiles_cc.txt
CorruptFiles_hadrons.txt
CorruptFiles_jpsi.txt
corruptfiles.txt
Corruptfiles.txt
dcar_bbblack_jpsired_pdst.pdf
dcar_bbblack_jpsired_pdst_ylog.pdf
dcar_bb_from_pythia_logy.pdf
dcar_bb_from_pythia.pdf
dcar_jpsi_from_pythia_logy.pdf
dcar_jpsi_from_pythia.pdf
DeadChannels.dat
diff_mydir_cesardir.txt
DST_embed_bb
DST_embed_bbbar
DST_embed_cc
DST_embed_hadrons
DST_embed_hadrons_3gev
DST_embed_jpsi
DST_embed_ktomu
DST_pythia_hadrons
dst_pythia_hadrons-MB0-0000429896-0000.list
dstpythia.list
dstpythia.root
dstpythia.txt
DST_sim_bb
DST_sim_bbbar
DST_sim_cc
DST_sim_hadrons
DST_sim_hadrons_3gev
DST_sim_hadrons_combined
DST_sim_hadrons_mineXuanSample
DST_sim_jpsi
DST_sim_ktomu
embedpythia.pdf
env.log
env.txt
ERR
event_gen_bb.csh
event_gen_bb_step2.csh
event_gen_bb_step3.csh
event_gen_cc.csh
event_gen.csh
event_gen.csh~
event_gen_hadrons.csh
event_gen_hadrons_step1.csh
event_gen_hadrons_step2.csh
event_gen_hadrons_step3_reass.csh
event_gen_hadrons_step3_xuan.csh
event_gen_jpsi_combined.csh
event_gen_jpsi.csh
event_gen_ktomu.csh
geom.root
geom_run15_v2.root
get_nevents
get_nevents.C
hadd_files.sh
HV
jpsipdst.txt
landau_parameters.txt
libpicodst_object.so
listofcc.txt
listofpythia.txt
LOG
logall.txt
logfiles.txt
logfile.txt
loglog.txt
logout.txt
logs.txt
log.txt
make_picodstobj.C
mc_bjpsi.C
mc_bjpsi_C_ACLiC_dict_rdict.pcm
mc_bjpsi_C.d
mc_bjpsi_C.so
mc_bjpsi.h
merge_picodsts.C
model_bdecay.C
muid_tube_eff_north_Run15pp200.txt
muid_tube_eff_south_Run15pp200.txt
mut.disabledWires.dat
mvcorrputfiles_bb.sh
mvcorruptfiles.sh
mylogfiles.txt
mypythia_cmu_pdst.pdf
mypythia.pdf
mypythia_pdst.pdf
mysimhad.txt
nokeys.txt
Nonexistingfiles.list
Nonexisting.txt
nooffiles.txt
notopen_bb_1.txt
notopen_bb_2.txt
notopen_bb_3.txt
notopen_bb_4.txt
notopen_bb_5.txt
notopen_bb.txt
NotOpenfiles.txt
notopen_hadrons_ajeeta.list
notopen_jpsi_1.txt
notopen_jpsi_2.txt
notopen_jpsi_3.txt
notopen_jpsi_4.txt
notopen_jpsi.txt
notopen.txt
Old
old_pDST_bb
out.log
output_bb_step2.txt
output_bb_step3.txt
output_bb.txt
output_env.txt
PISA
PisaFullfiles.txt
pisa.txt
practice.list
pylist12.dat
#pythia_bb.C#
pythia_bb.C
pythia_bb.C~
pythia_bb_C_ACLiC_dict_rdict.pcm
pythia_bb_C.d
pythia_bb_C.so
pythia_bb.h
pythia_bb.h~
pythia_configuration
pythia_files_bb
pythia_files_bbbar
pythia_files_bb_iter0
pythia_files_bb_list.txt
pythia_files_cc
pythia_files_hadrons
pythia_files_hadrons_3gev
pythia_files_hadrons_combined
pythia_files_hadrons_list.txt
pythia_files_hadrons_mineXuanSample
pythia_files_jpsi
pythia_files_jpsi_combined
pythia_files_jpsi_list.txt
pythia_files_ktomu
pythiafiles.txt
pythia.txt
realdataBG-run15pp_file-forAjeeta.list
realdataBG-run15pp_file-forAjeeta_onefile.list
realdataBG-run15pp_file-forCesar.list
realdataBG-run15pp_file-forXuan.list
realdataBG-run15pp_file.list
realdataBG-run15pp_file_split-forAjeeta1.list
realdataBG-run15pp_file_split-forAjeeta2.list
realdataBG-run15pp_file_split-forAjeeta.list
realdataBG-run15pp_file_split-forCesar1.list
realdataBG-run15pp_file_split-forCesar2.list
realdataBG-run15pp_file_split-forCesar.list
realdataBG-run15pp_file_split-forXuan1.list
realdataBG-run15pp_file_split-forXuan2.list
realdataBG-run15pp_file_split-forXuan.list
realdataBG-run15pp_file_split.list
Run_reassociation.C
run_segments.list
Sim3D++.root
SimFileLists
singlemuon_embed_pdst.root
singlemuon_jpsi_embed_pdst-MB0-0000422070-0001-2.root
split.csh
split_dsts.C
splitfiles
svxPISA.par
temp.txt
totest.list
whencreated.txt
with_n_option.txt
Xuan_allfailedFiles.txt
xuanfilelist.txt
```

Knowledge Management (KM)

- **Knowledge management** is key to DAP
- Keeping record of minute (but still crucial) details of the analysis procedures requires a non-trivial extra burden on the researcher
  - getting a quality paper published is the top priority, KM usually takes the back seat
  - ideally needs dedicated support
    - curation of the analysis procedures
- Unless DAP is built into an experiment from day one, proper knowledge management will be the part of DAP which will require most resources
  - also in terms of cost

Knowledge Management - Analysis Notes

- It is a common practice to require **analysis notes** from PWG prior to a decision about publication
- Each note encapsulates the know-how of a particular analysis (at least it's the goal)
  - cross references to other analysis notes in place
- Typically kept in a well-organized archive
  - however there is no standard for such archives across experiments
  - often kept private (a potential issue for DAP)
- Must be complete in terms of the physics content and adhere to specific requirements
  - in particular describe the software and its location and details of its operation
  - in reality rarely done and hard to enforce
- PHENIX has a well-defined note template and policy (although compliance may be an issue)
  - STAR has strict guidelines and policies in place
- The PHENIX app created a while ago to serve as an archive of analysis notes has its limitations:
  - in reality limited to a single PDF document (with links to revisions)
  - search function is pretty basic
  - Custom platform based on PHP… is it durable/maintainable for DAP?

PHENIX Analysis Notes Archive - the query page

PHENIX

Analysis/Technical Notes query form

Search Form

Download Analysis note template here. Use this template to write new analysis note for preliminary requests and final journal publiation
Use this form to search for technical and analysis notes.
Select first the desired note type either analysis notes or techinical notes. Searches may be made by author, title, note number, submission date (year) or by keyword. The default is "All" and it displays the entire list of the selected note type. Either first name or last name can be used for author. The search string for "Search by Submission Date (Year)" should be of the form "yyyy" (e.g. 2012).

Use AN Submission Form to add a new analysis note and TN Submission Form for a technical note.

Type of Note: Analysis Notes

Search by: Select from the list **Search String:**

Search by Author:
(Start typing author's name (last or first) until the desired name appears in the list and then select.)

Run Number: Select from the list

Collision Species/Energy: Select from the list

Physics Working Group: Select from the list

Analysis Type: Select from the list

Search Now Clear Form

PHENIX Analysis Notes - the archive contents

Number	Date	Title	Authors	Key Word	Links
an1425	2019-10-08	Run15 pAu identified pion and anti-proton spectra	Weizhuang Peng, Julia Velkovska	PLHF, p+Al_200GeV, Run-15, identified particles	an1425 draft an1282
an1424	2019-10-08	Run15 pAu identified pion and anti-proton spectra			an1424 draft
an1423	2019-10-08	Run15 pAu identified pion and anti-proton spectra			an1423 draft
an1422	2019-10-08	Run15 pAu identified pion and anti-proton spectra			an1422 draft
an1421	2019-10-08	Run15 pAu identified pion and anti-proton spectra			an1421 draft
an1420	2019-10-08	Jet Analysis in Run 15 p+p Collisions	John Lajoie, Milap Patel, Marzia Rosati, Jonathan Runchey	HHJ, p+p_200GeV, Run-15, jets	an1420 draft
an1419	2019-10-14	Neutral pion R_AA in p+Al, p+Au, d+Au and 3He+Au using combined Run-5, Run-8 and Run-15 \pp reference (PPG202)	Gabor David, Axel Drees, Norbert Novitzky	Heavy Ion, He3+Au_200GeV, p+Au_200GeV, p+Al_200GeV, d+Au_62GeV, d+Au_200GeV, Run-5, Run-8, Run-15, Run-14, single high pT particles, identified particles	an1419.01 draft an1152 an1269 an1270
an1418	2019-10-25	Model calculation of nuclear absorption in J/psi production at backward rapidity in PHENIX	Anthony Frawley	Heavy Ion, p+Al_200GeV, p+p_200GeV, d+Au_200GeV, Run-15, Run-14, quarkonia	an1418.02 draft
an1417	2019-10-22	Low pT Direct Photon Production in Au+Au Collisions at 200 GeV Beam Energy	Gabor David, Axel Drees, Roli Esha, Wenqing Fan, Norbert Novitzky	Photon, PLHF, Heavy Ion, Au+Au_200GeV, Run-14, direct photons	an1417.01 draft
an1414	2019-09-18	Template for PHENIX Analysis Notes	Yasuyuki Akiba, Gabor David		an1414 draft
an1413	2019-08-26	PHENIX Run14, Run15, Run16 PC2/PC3 track matching recalibration	Qiao Xu	PLHF, He3+Au_200GeV, Au+Au_14.6GeV, p+Au_200GeV, p+Al_200GeV, d+Au_62GeV, d+Au_39GeV, d+Au_20GeV, p+p_200GeV, d+Au_200GeV, Au+Au_200GeV, Run-15, Run-14, Run-16	an1413.01 draft
an1412	2019-08-02	K^* production in U+U at $\sqrt{s_{NN}}$ = 192 GeV in Run12	Alexander Berdnikov, Yaroslav Berdnikov, Vladislav Borisov, Dmitry Kotov, Daria Larionova, Iurii Mitrankov	HHJ, PLHF, U+U_193GeV, Run-12, single high pT particles	an1412 draft an065 an1010 an1374 an1401 an1402 an770 an911 an964 PPG148
an1411	2019-08-02	Protons production in Run12 Cu+Au at sqrt(s_NN) = 200 GeV	Alexander Berdnikov, Yaroslav Berdnikov, Dmitry Kotov, Mariia Larionova, Iurii Mitrankov	HHJ, PLHF, Cu+Au_200GeV, Run-12, identified particles	an1411 draft an1074 an1231 an1260 an1374 an683 an814 PPG146
an1410	2019-10-24	Direct, elliptic and triangular flow of π0 → γγ in d+Au collisions at 200 and 62 GeV	Veronica CanoaRoman, Gabor David, Abhay Deshpande, Jaehyeon Do, Axel Drees, Tom Hemmick, Carlos PerezLara	Hadron, Light, PLHF, d+Au_62GeV, d+Au_200GeV, Run-16, correlations, identified particles	an1410.02 draft an1367 an1406 an1407
an1409	2019-06-24	Run14 AuAu EMCal Geometry Tuning	Gabor David, Axel Drees, Wenqing Fan	PLHF, Au+Au_200GeV, Run-14, direct photons	an1409 draft
an1408	2019-06-20	Measurement of subevent cumulant flow in Run15 p+Au and Run 16 d+Au collisions	Ronald Belmont, Qiao Xu	PLHF, p+Au_200GeV, d+Au_62GeV, d+Au_39GeV, d+Au_20GeV, d+Au_200GeV, Run-15, Run-16, correlations	an1408.03 draft an1273 PPG206 PPG221
an1407	2019-05-30	PileUp Rejection Criteria based on BBC	Veronica CanoaRoman, Jaehyeon Do, Carlos PerezLara	Global, Heavy Ion, p+p_200GeV, d+Au_200GeV, Run-15, Run-16, correlations	an1407 draft an1304
an1406	2019-05-27	Q vector calibration	Veronica CanoaRoman, Jaehyeon Do, Carlos PerezLara	Heavy Ion, d+Au_200GeV, Run-16, correlations	an1406 draft
an1405	2019-10-20	Final Results on Double Helicity Asymmetries in Charged Pion Production in Longitudinally Polarized Proton-Proton Collisions at sqrt(s) = 510 GeV	Yuji Goto, Byungsik Hong, Ju Hwan Kang, Sook Hyun Lee, TaeBong Moon, Ralf Seidl, Inseok Yoon	Spin, p+p_510GeV, Run-13, single high pT particles, identified particles, A_LL	an1405.04 draft
an1404	2019-05-08	Measurement and analysis of three-pion HBT correlations for 0-30% Centrality in 200 GeV AuAu collisions	Mate Csanad, Bálint Kurgyis	PLHF, Au+Au_200GeV, Run-10, correlations	an1404.01 draft an1187 an1244 an1288 an920 PPG194
an1403	2019-08-15	J/ψ as a function of p_T in small systems with Yue Hang Leung's Correlated Background: Run15pp and Run15pAu, Run15pAl, Run14^{3}HeAu Centrality	Matthew Durham, Anthony Frawley, Sanghoon Lim, Krista Smith	HHJ, He3+Au_200GeV, p+Al_200GeV, p+p_200GeV, Run-8, Run-15, Run-14, J/psi, quarkonia, lepton pairs	an1403.06 draft an1306 an1354 an1369 an1391
an1402	2019-04-04	K^* production in Cu+Au at $\sqrt{s_{NN}}$ = 200 GeV in Run12	Alexander Berdnikov, Yaroslav Berdnikov, Vladislav Borisov, Dmitry Kotov, Iurii Mitrankov	HHJ, PLHF, Cu+Cu_200GeV, Cu+Au_200GeV, Run-5, Run-12, single high pT particles, phi	an1402.01 draft an065 an1010 an1374 an770 an911 an964 PPG148
				HHJ, PLHF, Au+Au_200GeV,	an1401.01 draft

Considerations for Invenio

- Know-how can be preserved in the form of collections of documents, data and software
  - analysis notes (see above) - hopefully with step-by-step instructions
  - associated analysis code (in repos or archives e.g. tarballs)
  - containers
- Software infrastructure
  - VMs, CVMFS
  - conditions and calibrations type data stored in files
- As mentioned above, PHENIX has a custom archive system for analysis notes. There is also a common tool for document management in the US - the “DocDB” developed at FNAL in early 2000s, deployed at FNAL, BNL and other labs
  - some useful functionality...but getting old and hard to develop and maintain (Perl)
  - no full text search
- Need a modern and durable tool for DAP, store collections of related objects
- Migration to a modern platform like Invenio RDM would be welcome
  - will have to wait till late 2020 due to the current timeline of the project

Invenio@BNL

- The SDCC is working with CERN and 10 other multidisciplinary and commercial institutions to build a research data management platform named InvenioRDM
  - Intended to be completed by mid 2020
- Two BNL scientific communities are developing and using Invenio V3 base custom applications hosted by SDCC (full cycle from dev to deployment)
  - open science repository designed to manage/harvest material science related records
  - deployment of Federated Identity to access resources
    - CILogon, Federated ID (InCommon / COmanage) support for authentication using Invenio's OAuth2 native module.
- Nonproliferation and National Security Department
  - the production environment required the deployment of isolated resources, 2FA DUO capable for authentication using Keycloak and Invenio's Oauth native module.

InvenioRDM: a turn-key open source research data management platform

Lars Holm Nielsen Apr 29, 2019 Invenio

CERN has partnered with 10 multidisciplinary institutions and companies to build a turn-key open source research data management platform called InvenioRDM, and grow a diverse community to sustain the platform.

The InvenioRDM project is funded by the CERN Knowledge Transfer Fund, as well as all the participating partners, including:

- Brookhaven National Laboratory (US)
- Caltech Library (US)

Extending Invenio support at BNL

- Evaluation in sPHENIX as a replacement for DocDB
  - hoping for long-term support of Invenio and derived products
  - working on a set of requirements
  - while InvenioRDM is still in development, looking for alternatives such as creating an interim Invenio V3 based application, discussions ongoing with various BNL groups
  - SDCC is helping to coordinate this effort
- Evaluation for the BNL EIC working group
  - ties into the EIC Collaborative Tools effort (NPPS)
  - may have to wait for InvenioRDM….. Interim solutions?

Declarative Analysis Tools

- In recent years there is growing interest in declarative analysis tools i.e. description of data and analysis workflow in a declarative as opposed to procedural language
  - declarations of workflows and data transformations as opposed to functions called in the event loop
- Interesting materials presented at recent CHEP, among others:
  - CMS effort (REANA) - yadage, CWL
  - “Fast HEP” toolkit
- Modular, compact descriptions of complex analyses
- May facilitate CI
- Looks promising for reusing analysis tools in the context of DAP
- RECAST: a powerful platform, worth studying

Declarative Analysis Tools cont'd

- Concern 1: prioritization of effort. If the software is ready to be wrapped into a declarative platform, it must be already in a very good shape. Are we there yet?
  - depends on the situation in a given experiment
- Concern 2: what is the learning curve? Does it fit into the limited resources allocated to DAP?
  - CMS is using REANA
  - for an experiment just starting on this path this may not be realistic
  - use in PHENIX currently in doubt due to limited resources

The role of BNL SDCC

BROOKHAVEN NATIONAL LABORATORY | Scientific Data and Computing Center

- BNL SDCC is Tier-0 for all of the RHIC experiments
  - Provides critical services including bit preservation and others related to DAP
  - Tens of PB of data preserved
  - According to the US DOE mandate to preserve RHIC data

- SDCC has a versatile team of experts that help solve problems outside of core production activities (cf. Invenio)

PHENIX - BNL as the host site

- In recent years BNL SDCC serves as the single fully functional center of the PHENIX computing infrastructure. At present virtually all of the stages of data processing and analysis (including simulation) take place at BNL
- A variety of data products necessary for operation of PHENIX software (databases, files containing conditions/calibrations data etc) are stored and managed by SDCC
  - complete removal of dependencies would be very costly
- For that reason, any DAP plan in PHENIX needs to assume that future reuse and analysis of the data will take place at BNL

Summary of the current DAP plan in PHENIX

- DAP Elements included in the current plan:
  - bit preservation
  - configuration/Conditions data management at the analysis level (TBD)
  - knowledge Management (KM): information gathering + documentation
  - a DAP Metadata system for KM (Invenio the current prime candidate)
  - software (containers, CVMFS for provisioning) + reference data for validation
- Prioritization of the data for DAP based on statistics, stability and quality
  - start with “golden datasets”
- Time profile and distribution of the effort
  - with people leaving the experiment the expertise is dissipating fast
  - there is a window of opportunity to capture knowledge and it must be done **now**
  - resources would be spent most efficiently within 1-2 years from now, whereas in 3+ years the return on investment will be much less

RHIC DAP effort and Community Collaboration

- What can we at BNL learn/borrow/reuse?
  - General guidance, best practices
  - Tools e.g.
    - Help/collaboration with Invenio RDM
    - Can the CAP system be ported/adapted for use outside of CERN?
- How can we contribute to DAP in the community?
  - N.B. BNL is already a part of the effort on Invenio RDM
  - Invenio-based applications
  - Possible contributions to other common tools (e.g. testing)
  - Sharing our experience of implementing DAP in a complex environment