

Report of the NSAC Sub-Committee on Public Access to Research Results

Report on Public Access to Research Results

1 Executive Summary

One aspect of the America COMPETES Reauthorization act is public access to research results, particularly in the forms of scholarly publications and digital data. In response to this, the DOE Office of Science has charged its advisory committees with identifying and assessing the current policies, procedures, and practices for disseminating research results; this report is on the research results in fields that are relevant to the Nuclear Physics program.

Finding 1:

The field of nuclear physics publishes in scholarly journals and uses the publication policies of those journals as well as archives and databases to make its research results available to the public. The results available through these means are the peer-reviewed versions of record (VOR). The VOR represent the ultimate product of the government investment in research and are uniformly available to the public. In most cases this access is free, and in others there is a cost associated with access through the journals themselves. Whenever possible, authors make the VOR available at no cost to anyone who requests them. Measurements in the VOR are often used by others to derive additional physics results.

Finding 2:

Pre-final data in the form of preliminary data, theses, conference presentations, and reports are generally publicly available on pre-print servers (e.g. arXiv and CERN Document Server), conference websites, and published proceedings, and, in some cases, in collaboration talk data-bases. Such results are often disseminated in workshops where through collaborative discussion the results are further understood and developed. In some cases the digital data presented in figures are also made available upon request. Requests from the general public for access to pre-final data are not common.

Finding 3:

Requests for digitized detector signals, processed detector signals, and associated computer codes by others not involved in producing them are in general rare, and because of the complexities in using these data, usually not fulfilled. The knowledge and resources required to utilize these data generally make them useless to persons unfamiliar with the experimental apparatus and the conditions under which the data were collected. There have been exceptions where dissemination of such data was useful, and under these situations these data were provided after publication. There are also situations where scientists may join the collaborations processing the data and then participate in the analysis effort.

Finding 4:

Small focused workshops (such as those at the Institute for Nuclear Theory), summer schools, collaboration meetings, and conferences play a crucial role in disseminating and extending research results. A deeper understanding of both experimental and theoretical nuclear science is enhanced by one-on-one interactions in these settings. The dissemination and sharing of pre-final research at these workshops inspire advances in the field.

Comment 1:

To assure access policies that are sustainable, close collaboration with publishers and other stakeholders is needed. Publishers deliver valuable services to researchers: registering, managing peer review, editing, disseminating and assuring the discoverability and authenticity of research as well as its preservation. Costs incurred in providing such services have to be paid under sustainable business models in order to ensure the continuity of such essential services to the scientific community

Comment 2:

Continued and enhanced use of the online arXiv and nuclear physics databases is an effective way to provide access to the results of publicly-funded research. Frequently the centers hosting the databases provide valuable functionality including evaluation, sorting, fitting, and other analyses of the data. As such, additional resources may be necessary to support these publicly-accessible venues. The availability of these open access sources to research results should be more widely advertised to the public.

Comment 3:

With few exceptions, digitized detector signals, processed detector signals, and associated computer codes are unlikely to be of use beyond that of the immediate collaborations that produced them. Because the data are in such complex and varied forms it would be counterproductive to impose a top-down policy regarding the sharing of them. Individual investigators, collaborations, or laboratories are uniquely able to determine the feasibility of sharing these forms of data. To make this category of data widely available would require significant additional resources with little added benefit.

Comment 4:

Meetings and workshops are essential for maximizing the benefits of research results.

Comment 5:

Presenting research results to the public is an important and recognized responsibility of scientists and the nuclear physics community responds to it through various outreach activities such as Open Houses, Science Saturdays, teacher training, and NSF Highlights open access web-pages, among others.

Examples of best practices in the dissemination of research results and data in nuclear physics:

- The APS journals now provide free access to their published VOR in all public libraries and high schools in the U.S.
- In most cases the author's final manuscript with corrections from peer review is posted on the arXiv, which makes it freely available to anyone with internet access. This is generally allowed by the peer-reviewed journals.
- Some collaborations, usually the larger ones, have formal policies for posting VOR and preliminary results to the arXiv and on collaboration websites, often with associated data tables, and responding for requests to results. This enforces consistency and availability of results.
- National laboratories have public relations and outreach efforts to disseminate research results at a level designed for teachers, schoolchildren, and other members of the general public.
- Small, focused workshops are a very effective way to disseminate research results to interested scientists.

2 Introduction and definitions

In a typical nuclear physics experiment, signals from experimental apparatus are digitized and written to some kind of storage medium, such as magnetic tape or disk. The format of these digitized detector signals varies widely from experiment-to-experiment, and is intended to be readable by analysis software written by the experimenters. To begin the process of converting these digitized signals into physics measurements, analysis software is used to filter out noise, apply calibrations, and convert the signals from binary format into physical variables such as momentum and energy. Further analysis software is then used to identify the signals of interest in the data set. Finally, to interpret the observed signals, complex detector simulation models, also written by the experimentalists, are used and the same analysis applied to the simulated data of the models. A critical final step in the process is the estimation of systematic uncertainties, which relies crucially on detailed knowledge of the experiment. This often involves further calibrations, simulations, and additional measurements. The combination of extracted signals and their estimated uncertainties constitute a final result that is published in a scholarly journal which, often after a formal referee process, becomes the version of record; see figure 1. These measurements are then used by others, who may not be directly involved in the experiment, to test their own hypotheses and generate new ideas.

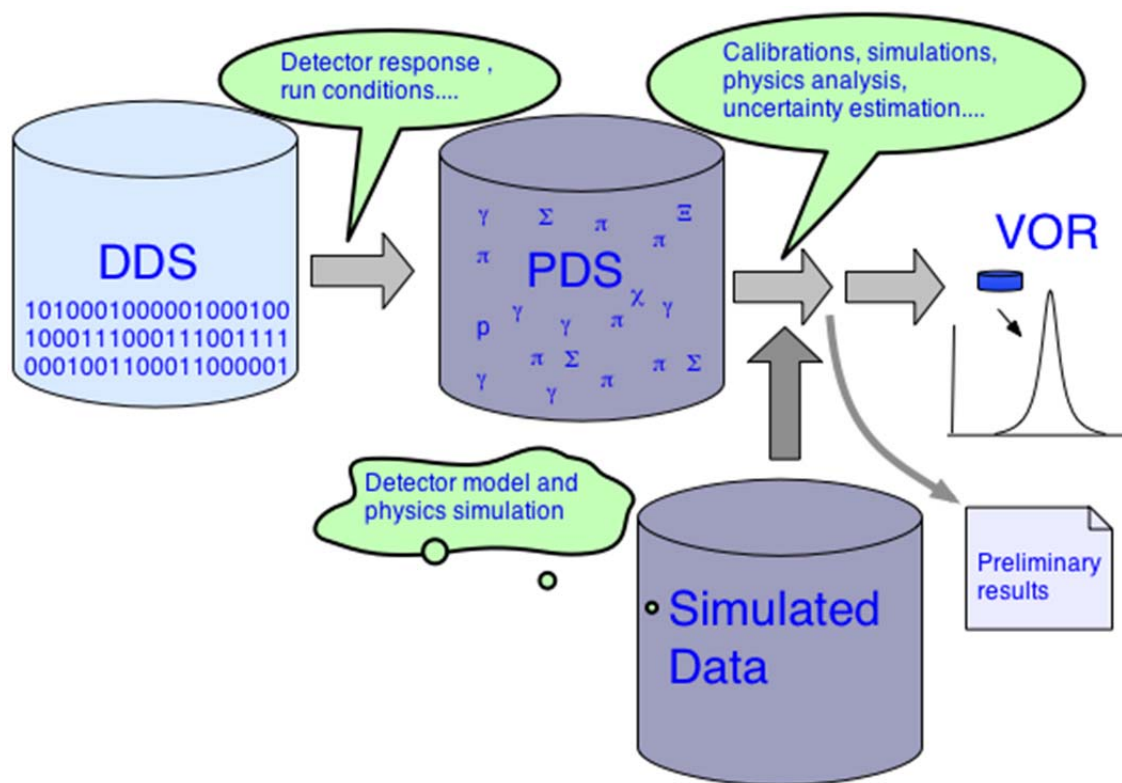


Fig 1: Schematic of research flow from digitized detector signals to published versions of record.

Occasionally the processed detector signals are masked or blinded from the scientists participating in the research to prevent them from consciously or subconsciously affecting or biasing the result. Clinical trials in medical research almost always use some form of blinding to keep scientists from knowing which patients have been given the trial drug and which have been given the placebo; the intent of blinding data is similar in nuclear physics.

During the course of data analysis, experimenters often release preliminary results at conferences and workshops, to inform the community of the status of their work, to exchange ideas, and to get feedback on their work. These preliminary results are typically subject to most of a collaboration's internal review processes. Such preliminary results are not usually intended for use by others to derive further physics results.

Once a collaboration has a final measurement that they would like to publish as a version of record, a scholarly paper is written and passed through the collaboration's full internal review process, then the article is submitted to a refereed scholarly journal and, often simultaneously, to the open access arXiv. Once the article has passed through the referee process of the journal, an updated version that may contain corrections based on the peer review is often posted on the arXiv, replacing the original pre-refereed version. The journal then applies its own copy-editing and formatting procedures to the article before final publication. Gaining access to the copy-edited and formatted version of the article often requires a paid subscription, as the journal holds the copyright to this version. Some journals do make these published versions openly available to public libraries or allow them to be posted on the author's own web pages or laboratory web pages.

In general, the size of the research sets get smaller as the process moves from digitized detector signals to physics results. Table 1 shows the wide range in sizes within and across research sets for a typical experiment or project over one year. In many cases, individual investigators or collaborations maintain the sets of digitized detector signals and processed signals. Some of the larger laboratories have policies and infrastructure for archiving the digitized detector signals and processed signals. More information on archiving can be found in section 5.3 of this report.

| | Digitized Detector Signals | Processed Signals |
|---------------------------------------|----------------------------|----------------------|
| JLab | Up to 100s of Tbytes | Up to 100s of Tbytes |
| RHIC | few Pbytes | few Pbytes |
| LHC | Up to a Pbyte | 100s of Tbytes |
| Neutrinos | 100s of Tbytes | 100s of Tbytes |
| Neutron Physics | up to 10s of Tbytes | up to 100s of Gbytes |
| LE User Facilities | up to few Tbytes | up to few Gbytes |
| Small Groups and Single Investigators | up to few Tbytes | up to few Gbytes |
| Lattice QCD | Total data set = 1 PByte | |

Table 1: Types and Typical Sizes of Nuclear Physics Data Sets for a Typical Experiment or Project Over One Year

Vocabulary:

- DDS = Digitized Detector Signals, often referred to as raw data. Sets of DDS vary in size from a few GBytes, which may be stored on hard drives or memory sticks, to hundreds of PBytes, which require high tech storage systems on clusters of computers or servers operated by the scientists or laboratories
- PDS = Processed Detector Signals; detector responses that have been filtered, calibrated, and processed into physics variables
- Pre-final Research = preliminary results, PhD theses, and conference papers
- Preliminary Results = vetted within a collaboration but not final
- Physics Results = final results of research that usually go into the VOR

- VOR = version of record, usually peer-reviewed
- Published Research = final research results that go into the peer-reviewed and copy-edited VOR
- Open Access = access to research via the internet in such a way that the material is free for all to read and use
- Green Access = access of the author's final version and in some cases the VOR in an institutional repository or in a subject repository such as the arXiv. This is one of the open access policies of the APS, AIP, and many other scholarly publishers.
- Delayed Open Access = only subscription access for a limited period (typically one year), then the manuscript is freely available
- Golden Access = through a Creative Commons license, someone or some institution (e.g. the author or the author's home institution) pays to make the article freely available upon publication

3 Methodology

Nuclear physics spans a wide range of research from the essence of the quark-gluon plasma to nucleon and nuclear structure to nuclear astrophysics and is studied by individuals as well as by small groups and large collaborations. Not surprisingly, the policies, practices, and procedures for access to research results vary widely throughout the nuclear physics community and a fairly large sub-committee was needed to respond to this charge. Each member of the subcommittee was responsible for collecting information on the accessibility of research results in a sub-field of nuclear physics or laboratory associated with a sub-field. Most committee members used a questionnaire to collect information, see section 8. Committee members, areas of responsibility, and notes are listed below:

- Jefferson Lab (JLab): Curtis Meyer (Carnegie Mellon University) Information was obtained from all four halls at JLab. Halls A, B, and D are, for the most part, hall-wide collaborations, while Hall C supports individual experiments without a global collaboration. Hall D is currently being built as part of the 12-GeV upgrade, so the Hall D information is only a plan and has not yet been implemented, as there are currently no data.
- The Large Hadron Collider (LHC): Julia Velkovska (Vanderbilt University) Information was obtained through contact with the collaborations' publication committees, physics coordinators, and through publicly available collaboration by-laws and other written policy documents.
- Neutrinos: Josh Klein (University of Pennsylvania) collected information from members of the neutrino experiments.
- Neutron Physics: Fred Wietfeldt (Tulane University) collected information from representatives of the neutron community.
- Low Energy User Facilities: Michael Thoennessen (Michigan State University and the National Superconducting Cyclotron Laboratory) collected information from the low-energy nuclear physics facilities.
- Small groups and single investigators: Mark Riley (Florida State University) and Carl Brune (Ohio University) collected information from a number of small groups and single investigators.
- Relativistic Heavy Ion Collider (RHIC): Helen Caines (Yale University) Spokespersons of the four RHIC experiments were contacted and responses were provided by STAR and BRAHMS. An informal discussion with PHENIX and a review of its publication policy (a link to this policy is in the appendix) provided answers that were essentially the same as those from STAR and BRAHMS.

- Theory: Scott Pratt (Michigan State University) A set of five questions (see section 8) was sent out to six members of the theory community representing a cross section of nuclear physics, in terms of research area and seniority.
- Lattice QCD: Paul Mackenzie (Fermi National Accelerator Laboratory) The USQCD Collaboration, which includes almost all US lattice gauge theorists in the US, is an organization for the purpose of creating and deploying hardware and software infrastructure for lattice QCD. As the Chair of USQCD, Paul Mackenzie provided the information used in this report.

Most of the work on this charge was done through email and conference calls. The committee met in person once to formulate the executive summary.

4 Open access policies of journals, archives, and data bases commonly used in nuclear physics

The primary means of disseminating research results in the field of nuclear physics is through publication of these results in scholarly journals. The results are also distributed through electronic archives and public databases, following the policies of the scientific journals. These published results are the peer-reviewed VOR and are the ultimate representation of the government investment in the underlying research. These published results are uniformly available to the public, usually for free. However, in some cases there is a cost associated with accessing the scientific journals themselves. Whenever possible, authors make the VOR available at no cost to anyone who requests them.

4.1 Major journals for Nuclear Physics

The most common professional journals in which nuclear physics research results are published are those of the American Physical Society (APS: Physical Review Letters, Physical Review A – E, ...), the American Institute of Physics (AIP: Journal of Applied Physics, Journal of Mathematical Physics, Review of Scientific Instruments, ...), Elsevier (Nuclear Instruments and Methods in Physics Research, Nuclear Physics A – B, Physics Letters B, ...), the Institute of Physics (Journal of Physics G), and the European Journal of Physics. Generally, all journal articles are accessible online through the publisher's website with a paid subscription, and frequently on the author's institutional website or other electronic archives. Prior to publication most publishers require the copyright of an article be transferred from the author to the publisher, and the publisher retains responsibility for the VOR. In spite of the copyright transfer, most publishers allow authors to post VOR on their own website and their institutional website. Most publishers even allow authors to post their final manuscripts with corrections from peer review on the freely accessible arXiv. Indeed, the APS fully embraced the arXiv in 1997.

Many publishers are adopting or experimenting with dissemination mechanisms that allow open and free access to their articles such as delayed open access or golden access in which the author or the author's institution pays to make the article freely available upon publication. At this time none of the journals typically used by nuclear physicists use delayed open access but golden access is offered by most of these publishers. An example of a golden access mechanism is the one initiated by APS in January 2011 with its journal Physical Review X (PRX) and expanded to most of its other journals in February 2011 in which authors pay an article-processing charge whereby their accepted manuscripts

will be openly accessible. PRX and a few other APS journals are fully accessible through this mechanism. For the other APS journals this mechanism is in addition to traditional subscription-funded publication; authors may choose one or the other for their accepted papers. It is too early at this time to know the level to which authors of nuclear physics articles in PRC are supporting golden access. Another dissemination model being tried by the high energy physics community is the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP³). In this model, HEP funding agencies and libraries would form a consortium in which participating agencies and libraries voluntarily provide publishing costs, and publishers would make the electronic versions of their journals openly accessible. The SCOAP³ project has been soliciting funding pledges for four years and has reached only 70% of the 2007 budget goal. CERN has made first results of the LHC, both from high-energy and nuclear physics, available to scholarly publishers through a Creative Commons License in which CERN will maintain copyright to the VOR. CERN pays the publisher an Open Access Article Processing Charge for this, and it is transparent to the authors. Funds collected by publishers through subscriptions and golden access fees cover the costs of administering the peer review process, editing and type-setting the VOR, as well as archiving the VOR.

Since July 2010 every public library and since February 2011 every high school library in the US has been able to register for access to all APS Physical Review articles without any associated fees or charges, making the articles freely available even for printing to users while in the libraries and high schools. As of April 2011, 573 libraries and 103 high schools had registered for this access. In the first 5 months of 2011, 0.028% of the downloads from the APS site were from public libraries and high schools.

We note that the materials and data published in articles are not included in the copyright and there is an expectation on the part of the publishers, codified in journals' instructions to authors, that these be made available upon request to anyone requesting them. This is indeed the practice of all nuclear physicists contacted for this report. Some journals offer the functionality in their online articles of embedding tabulated data and other materials into articles as supplementary material. This option is not widely used by nuclear physicists nor is there an accepted format standard.

4.2 Fully open access archives

The arXiv (<http://arxiv.org/>) is a fully open access archive of e-prints in physics, mathematics, non-linear science, computer science, quantitative biology, quantitative finance, and statistics hosted by Cornell University with mirror sites around the world. Laboratory-based open access archives also exist but the arXiv is the most widely used open access archive for nuclear physics research results. Peer review is not administered by the arXiv nor is it required for submission, however submissions must conform to Cornell University academic standards. In addition to archiving authors' final manuscripts, the arXiv and the CERN Document Server also contain pre-prints of research articles, and both have useful search capabilities.

The arXiv is an example of Green Access in which neither the author nor the reader pays for access to the research results. The practice of Green Access is not sustainable if it leads to readers or institutions cancelling their journal subscriptions. This form of access does not account for the costs of the arXiv itself and Cornell has begun an effort to expand funding sources for arXiv to ensure its stability and continued development. At this point support is voluntary and may not be sufficient.

4.3 Databases

The nuclear physics community often submits research results to freely accessible databases such as the National Nuclear Data Center (NNDC, <http://www.nndc.bnl.gov/>), The Durham HepData Project (HEPDATA, <http://durpdg.dur.ac.uk/>), the Data Analysis Center (<http://gwdac.phys.gwu.edu/>), the Joint Institute for Nuclear Astrophysics (JINA, <http://www.jinaweb.org/index.html>) and others. These databases are accessed millions of times each year.

Programs like the U.S. Nuclear Data Program also provide an important service through critical independent evaluation efforts. Results from many different VOR, and thereby from many different experiment approaches, are compared and collectively analyzed to extract recommended values for physical values such as nuclear ground state properties, excited level properties, decay and radiation properties, and reaction cross section values. Recommended values of physical parameters with robust uncertainties are provided in easily searchable databases and provide an important resource for the basic science community as well as applied areas of research, from astrophysics to nuclear energy. These evaluation databases are freely available on the web and provide significant added value beyond the original VOR.

5 Access to research results

Most nuclear physicists have developed practices regarding access to their research results. This has been done as individuals, or through the collaborations to which they belong or via the laboratories where they work. In some cases, formal policies and procedures have also been developed. Requests for VOR are rare – perhaps because they are readily available through journal websites and collaboration or lab websites (see subsections below). Requests for data in VOR in numerical form with error bars are common and almost always honored, either by the person contacted or by another member of the collaboration. Usually this is done by pointing the requester to the relevant webpage or database entry.

5.1 Practices, policies, and procedures regarding access to the VOR

For the most part, the nuclear physics community provides open access to final research results through the practices of the journals in which it publishes that research and the use of openly accessible archives and databases. Accessibility to this level of research is the same for the general public as it is for scientific peers. Requests by the public for final research results range from none to rare and when received, such requests are honored. Additionally, final data are frequently submitted to openly accessible data bases. It is worth noting that with most final research being openly accessible on-line, it is nearly impossible for authors to know how many times members of the public or peers had viewed the research as most on-line sites do not currently indicate viewing rates.

When considering to which journal a paper should be submitted, the nuclear physics community usually targets the highest impact journals while at the same time being cognizant of the access policies of specific journals. Thus most often the journals selected are APS, AIP, ... (see section 4.1) which are high in impact factor and openly accessible. While this is indeed the practice, there does not appear to be any formal policy in the nuclear physics community regarding journal selection for publications.

Posting of authors' final manuscripts in the openly accessible arXiv ranges from some individual physicists having this practice to it being universal for the larger groups and collaborations. RHIC and LHC collaborations have formal policies requiring these manuscripts be posted in the arXiv or their own web-pages. When the arXiv is used, physicists and collaborations often post their research onto the arXiv when it is submitted to a journal. Upon acceptance for publication by the journal, the arXiv is then updated with the final manuscript that contains any corrections from peer review.

Making data contained in VOR, such as data that are displayed in a figure or referred to as part of a calibration procedure, accessible to other physicists who will use those data in their own analyses can increase the potential research benefit of the data. As noted in the previous section, the electronic format of APS journals allows authors to embed tables of data with error bars to the figures which show that data or to relevant points of the text, as supplemental material. Only the LHC collaborations have policies supporting the use of the embedding tool. To date, other nuclear physicists only occasionally use this tool. However, data and tables of data are routinely made accessible through a variety of other methods. These methods include submitting data to the NNDC database when requested by the NNDC, posting numerical data to the HEPDATA Durham database, and posting tabulated data to experiment-specific publicly accessible web sites, generally hosted by the larger laboratories.

Jefferson Lab, which hosts experiments and collaborations that vary greatly in size, exemplifies the general practice of how access to the numerical results contained in VOR is handled. For the Hall A and C experiments in which the number of data points is small, the numerical results are normally published as a table in the paper. In Hall B, where much larger data sets are produced, there is a requirement that the numerical data be inserted into a completely publicly accessible database run by the collaboration. It is also typical to make the data available to databases such as Durham and SAID. When someone in the scientific community asks for published data, the requests are quickly honored, either by providing the data, or sending a pointer to where the data could be obtained.

Research from neutrino experiments falls broadly into two classes: those that relate to measurements of neutrino physics and those that use neutrinos as astrophysical (or even, recently, geophysical) probes. Generally speaking, the research associated with measurements of neutrino physics tends to require detailed analyses that require associated detector and physics simulations to interpret. Research for astrophysical measurements can be simpler, and use processed detector signals such as the times of all candidate events observed in a detector. It is common practice to honor requests for such 'time series' data, but only after the scientists producing the time series have published their own studies of these data.

For neutrino physics measurements, experimental collaborations provide enough information in the physics results contained in the VOR that members of the neutrino community or even the general public can fit these results under any model or hypothesis they choose, and many publications by theorists have done exactly this. The data points themselves are often provided within the VOR in a table, but in cases where this does not happen, the collaborations either make them generally available on an open collaboration web page, or directly to individuals who request them. As analyses have become more sophisticated, additional information such as correlation matrices or χ^2 maps, which cannot easily fit within a journal article, have also been provided by request or on web pages.

The LHC and RHIC collaborations have policies for access to numerical data used for figures (plots, histograms, etc.) published in VOR. These policies complement the open access publication of articles and aim to reach the largest possible community without barriers. All experiments maintain web pages

with information on what research is accessible and links to it. The RHIC experiments post the VOR, copies of the individual figures, and tables of the data in the figures on their own websites. The LHC experiments go further by posting final manuscripts with corrections from peer review to the CERN Document Server and the arXiv, and submitting the numeric data to the HEPDATA Durham database (<http://durpdg.dur.ac.uk/>). The HEPDATA system is well structured, well maintained and has a long tradition in collecting important High Energy Physics data. It is also used by the Nuclear physics community. The ALICE collaboration has a policy of sending the data to HEPDATA concurrently with the submission of the article to the peer-reviewed journal, while ATLAS and CMS submit the data when the paper is published. The HEPDATA has search capability, providing easier access to the data. It is currently often linked to SPIRES, and in the near future, it will be linked to the INSPIRE database, which has detailed records of the publications, complete with links to the journal server, laboratory document servers, arXiv entry, references, citations, and HEPDATA. INSPIRE is jointly developed by CERN, DESY, Fermilab and SLAC and is presently available in a beta version.

The practices of the nuclear theory community are consistent with those of nuclear experimentalists in that the arXiv is in wide use and has led to greatly enhanced dissemination of ideas and both published and preliminary results. By posting data tables corresponding to published figures, several experimental collaborations have made the task of collecting results for comparison with theory far easier than in previous times.

USQCD does not have a policy on submitting VOR to the arXiv, but it is the universal practice. There is no policy on selecting which journal to submit VOR to, although APS journals are the predominant choice. USQCD considers the inclusion of numerical values of plotted points with statistical and systematic uncertainties to be part of its “standard best practices”, but it does not have a formal policy about this. Although there is no general policy for handling requests from others in the scientific community for USQCD VOR, it would be very unusual for research in VOR not to be made freely available. The major ensembles of lattice gauge configurations, which represent the results of a large fraction of the USQCD computer usage, may be freely downloaded from the web by anybody.

5.2 Practices, policies, and procedures regarding access to pre-final versions of research and codes

We are not aware of any nuclear physicist or collaboration having received requests from the public for pre-final research. However, the RHIC and LHC experiments have been asked on numerous occasions for copies of their event displays, but it is thought that this was for “art” purposes rather than science.

Many of the larger collaborations at Jefferson Lab and all of the RHIC and LHC collaborations have policies that define the conditions under which research that is not yet at the final stage may be presented at conferences and included in proceedings. This preliminary research has typically undergone most of the analysis deemed necessary by the collaboration. To the best assessment of the collaboration, preliminary data are considered correct but, because these data are not yet final, the data may change slightly prior to final publication. Conference proceedings and pre-prints of publications and student dissertations are frequently posted on collaboration web pages, the arXiv, or the CERN Document Server.

With few exceptions all the experimenters and collaborations felt that it was not possible to make processed detector signals or digitized detector signals available to non-collaborators, due to the complex nature of the corrections that have to be applied to the DDS. Without the detailed knowledge of the exact experimental setup and run conditions for each dataset, analysis of the data would likely lead to misleading/incorrect results. Also, the size of the DDS sets is prohibitive for general dissemination. The practice of blinding DDS from the scientists analyzing it to prevent bias clearly prohibits making the DDS accessible outside the collaboration.

There was an expectation that when pre-final research is made available to others outside the originating group, the scientists receiving the pre-final research would wait to publish their findings until after the originating group had published theirs. Proper use of the pre-final research is expected to be non-trivial to outside peers due to the complexity and uniqueness of the experiments. Both SNO and Super-Kamiokande have provided 'binned' time series data publicly, but only after the collaborations had completed their own publications of results from these data sets themselves. In some cases this meant postponing the availability of a data set even when a request was made from members of the neutrino community. There have been a few instances with Gammasphere experiments where groups that were not part of the measurement asked for access to the data at a later time, typically after the collaboration had completed its publications from the experiment. The management of the ATLAS facility at the Argonne National Laboratory encourages this sharing of data and simply tries to ensure that from the start there is an understanding about how proper credit will be given to the group(s) that carried out the original measurement. There does not appear to be an explicit policy regarding this.

While no policy seems to exist regarding accessibility of DDS, it appears that almost no nuclear physicist or collaboration would release DDS to either the public or peer scientists, as proper use of the pre-final research would be non-trivial to outside peers and the public due to the complexity and uniqueness of the experiments. Only one case was reported in which processed detector signals were shared with outside scientists. This was where the collaboration made absolute neutron flux measurements that were shared with other scientists for diagnostic purposes when requested.

One avenue for scientific peers to gain access to pre-final versions of research is to join the collaboration producing the DDS and processed research. When new members join a collaboration, they are expected to fulfill the responsibilities of that collaboration, which often include data taking duties and other service related tasks, as well as abide by the collaboration's publication policies. This may make it impossible in practice for a single member of the public. For Hall-D at JLab there is the intent to make processed detector signals available at a four-vector level, but none of the details have been worked out, and this may turn out to be rather difficult.

In general there is no policy or practice of making internal analysis notes accessible to those outside the collaborations. In lattice gauge theory, internal notes are generally very informal and not useful outside the collaboration, although the major ensembles of lattice gauge configurations may be freely downloaded from the web by anybody. USQCD does not track downloads, but based on publications it appears that tens of thousands of configurations have been downloaded by 10-20 different projects. JLab internal notes are an important part of the documentation process of an analysis but it was generally felt that they represented an intermediate result and would not be publicly available. In some cases these could be obtained directly from the authors, and in others they would require collaboration approval to release them.

The National Superconducting Cyclotron Laboratory (NSCL) has a policy that investigators are expected to share with other researchers the primary data, samples, physical collections, and other supporting materials created or gathered in the course work under NSF grants. This sharing should be at no more than an incremental cost and should occur within a reasonable time. The NSCL also has a pre-print server to aid in dissemination of research but there does not appear to be an official policy requiring its use or that of any other eprint server. NSCL policy is that a researcher requesting access to pre-final research shall contact the Spokesperson for access. Requests to NSCL for access to material and analyses related to published research will be forwarded to the Spokesperson of the pertinent experiment. Collaboration on interpretation and publication of research data is encouraged.

There are several types of data in lattice QCD. Lattice gauge configurations and the quark propagators derived from them are very costly to compute. Ensembles of gauge configurations are of moderate size (up to 10s of Tbytes), while quark propagator data sets can be very large (up to 100s of Tbytes). It is straightforward to reuse gauge configurations for many different purposes. A standard data format has been created for them by the International Lattice Data Grid to facilitate sharing. It is the policy of USQCD that gauge configurations created by member collaborations are to be shared within USQCD immediately, and that they be made publicly available via the International Lattice Data Grid within six months of the first publication arising from them. Quark propagators are tailored much more closely to specific physics projects, so sharing is less universal. There is no general policy and sharing is coordinated by projects with allied interests. There has been some propagator sharing among all of the major groups in nuclear lattice gauge theory. The smaller, more highly processed detector signals arising from analysis of quark propagators are highly specific to individual projects and have no well-defined format. Sharing of such data requires collaborating with the proprietor of the data, which occasionally happens. USQCD devotes considerable resources to creating community codes and libraries for lattice QCD. These are made publicly available on the web to groups in the US and around the world.

In addition to the nuclear lattice effort, many other members of the nuclear theory community are involved in large-scale computation projects, such as those sponsored by DOE's SciDAC program (Scientific Discovery through Advanced Computing), to simulate supernova explosions, to simulate collisions at RHIC or to calculate nuclear properties and reactions. In addition there are shared data such as tables of nuclear masses or nuclear wave functions. The deliverables from these collaborative projects often include complex codes that are a valuable product of the government investment. Such codes are usually made available for download or are published in Computer Physics Communications. Smaller collaborations or individual theorists also generate codes, which are often published or posted on-line, usually at web pages maintained by the individual groups. There are a few locations where multiple efforts are posted, e.g., Purdue maintains a repository for transport codes for heavy ion collisions at https://karman.physics.purdue.edu/OSCAR/index.php/Main_Page. Access to these codes varies widely. Many, especially the larger and more sophisticated projects, are freely available for download, while others can be acquired by contacting the authors. Only a few codes are treated as proprietary, and often that distinction is lifted after the authors have had the opportunity to publish several papers. There is significant work involved in making a code openly available and the limited work force available for development means that maintaining codes and documentation for broad public access is not always feasible. In addition, most codes tend to be under constant development and may be designed to run on specific hardware.

Often the best way to facilitate the use of large-scale codes within the community is for developers to visit interested users and assist them with getting started, or to organize workshops where developers and users interact (such as those run by the Institute for Nuclear Theory). Such workshops have been

very effective, and will continue to be necessary to fully exploit the intellectual investment represented by these efforts. In this way, best ideas and practices better propagate throughout the communities. For example, such workshops have been used to provide detailed comparison between competing transport codes used for modeling heavy ion collisions for RHIC.

5.3 Archiving

The long-term archival beyond the date of publication of data not included in the VOR or submitted to the databases is difficult for several reasons. The DDS and PDS are usually stored on hard disks or magnetic tape. These media have a finite lifetime. In addition, obsolescence of the media readers, computer codes, and/or operating systems may be an issue. It is also necessary that auxiliary information about experiments, such as contained in laboratory notebooks, remain available in order to interpret the DDS or PDS. Finally, the inevitable loss of key people limits the usefulness of archived DDS or PDS. Much of the analysis expertise required to interpret these data resides in students and postdocs, who are in short-term positions. The collaborations themselves also have finite lifetimes.

In many cases, data archiving is left up to the individual investigators or collaborations. Some of the larger laboratories have policies and infrastructure in place for data archiving. For example, the NSCL maintains an archival copy of raw data for two years after the completion of the measurement; after that archiving is the responsibility of the investigators. Jefferson Laboratory has maintained archival copies of nearly all DDS that have been collected. This archive is actively managed to maintain it on fresh modern media. The CERN experiments have formed task forces that have been charged to prepare policies and a plan for the long term data preservation and access, in coordination with CERN management. Common policies will be in place for the proton-proton and the nucleus-nucleus collision data. A series of workshops on this topic have been conducted by the high-energy community. Information about the most recent one can be accessed from the CERN Document Server: <http://cdsweb.cern.ch/record/1352488> .

For smaller experiments, past experience indicates that it is unlikely that the DDS or PDS will be used after the publication of the VOR. Typically, 5-10 years after publication it is no longer possible to utilize archived DDS or PDS, for the reasons stated above. In many cases, it appears that maintaining long-term access to older data in a useful fashion is not cost effective. For the large collaborations and facilities, the long-term archiving of DDS and PDS is more likely to be practiced, due to the high cost of obtaining these data and the availability of the infrastructure to do so.

6 Access to the general public through outreach

As mentioned above, the research results in nuclear physics are the final papers published in refereed journals as the VOR. Many of these results are available to the general public without any access restrictions on preprint servers. Almost all publishers allow researchers to post versions of the VOR on their websites and distribute these versions freely for non-commercial purposes. In general the VOR are written for experts and colleagues in the field. Typically the general public and peer scientists outside the group generating the VOR do not have the knowledge, expertise, or tools to make use of DDS or even PDS. Thus, an important aspect of making research results accessible to the general public is to present the data and results in a format that is understandable to the general public. This is an important and recognized responsibility in the nuclear physics community as demonstrated in a working

group white paper for the most recent NSAC Long Range Plan

http://science.energy.gov/~media/np/nsac/pdf/docs/2007_lrpwp_education.pdf

The National Laboratories invest considerable effort and resources into disseminating the latest research results and making them accessible to the public. Examples include RHIC at Brookhaven (<http://www.bnl.gov/rhic/physics.asp>), JLAB (<http://education.jlab.org/>), LBL (<http://www.lbl.gov/nsd/resources/education.html>) and the NSCL (<http://www.nscl.msu.edu/outreach>). These sites explain the science of the respective facilities, publish press releases, give links to publications and further detailed information of the research groups, and offer tours to the general public. Additionally, specific attention is given to educating K-12 students and offering teaching and learning tips and tools for their respective teachers.

The NSF actively solicits research highlights from its Principal Investigators that are written for the general public and posted on openly accessible webpages <http://www.research.gov/seeinnovation>. A new NSF policy also requires that when an award is completed, the PI must submit a report summarizing the sponsored research for the public; these reports are also openly accessible at: <http://research.gov>.

All CERN experiments maintain web-pages aimed at the general public, which include collections of photographs, videos, virtual tours, newsletters, press releases, children's corners, and education links: <http://aliceinfo.cern.ch/Public/Welcome.html> <http://cms.web.cern.ch/cms/> <http://www.atlas.ch/> Such materials are also available with search capabilities in a variety of languages through the CERN Document Server, which presently contains 56,139 items in its Multimedia and Outreach collection (<http://cdsweb.cern.ch/collection/Multimedia%20%26%20Outreach?ln=en>).

Individual researchers and smaller research groups often utilize public relations media outlets of their respective universities and colleges to present their research to the public. In addition, many researchers make a special effort to inform the public and educate K-12 students by giving public lectures (for example at Science Cafés <http://www.sciencecafes.org/>, or Saturday Morning Physics <http://cyclotron.tamu.edu/smp/>).

Education and outreach are essential components in making the research results accessible to the general public. Only if the public understands the motivation for the research efforts will it appreciate and understand the results.

7 Closing remarks

For both theoretical and experimental endeavors, the final results of all nuclear physics research programs are peer-reviewed journal articles based on the research. These published works become the version of record (VOR) of the research and of the government investment that made the research possible. Nuclear physics VOR are widely accessible through scholarly journals in paper and electronic formats. While access to the online archives and databases is in general free of charge, the scholarly journals have subscription charges. Scientists have access to most scholarly journals through the subscriptions paid by their institutions; the public can access those journals by visiting research libraries holding subscriptions to those journals, such as public universities. Journals published by the APS can be accessible at every public library and high school in the United States. Although journals hold the copyright on published VOR, most journal publishers allow authors to post VOR on their own website, their institutional website, and to even post their final manuscripts with corrections from peer review on the freely accessible arXiv. Additionally, authors make the VOR available at no cost to anyone who requests them. There do not appear to be any significant obstacles to VOR accessibility.

Formal policies regarding access to pre-final research in the form of preliminary analysis, theses, conference presentations, and reports only exist in the larger collaborations. However, it is common practice throughout the nuclear physics community to post pre-final research on freely accessible pre-print servers (e.g. arXiv and CERN Document Server), conference websites, and, in some cases, in collaboration talk data bases. Requests for access to preliminary research are frequent and are often granted. Requests for digitized detector signals, processed detector signals, and associated computer codes by others not involved in producing them are in general rare, and because of the complexities in using these data, usually not fulfilled. There are also situations where scientists may join the collaborations processing the data and then participate in the analysis effort.

A common practice in the nuclear physics community, to increase the potential benefit of research results, is the participation in focused workshops, summer schools, collaboration meetings, and conferences. The dissemination and sharing of pre-final and final research at these workshops lead to advances in the field.

Disseminating both research results and information about the research to the general public is an important part of the nuclear physics program. In addition to individual scientists and groups of scientists giving public talks, hosting open houses at their laboratories, and providing K through 12 teacher training, all major labs have active outreach programs.

8 Appendices

8.1 Questionnaire to researchers and research collaborations in experimental nuclear physics

- 1) Do you have a policy of all publications and proceedings being submitted to arXiv as well as the appropriate journal? Lacking a policy, is this your normal practice?
- 2) Do you have a policy on how your target journals are selected? Aside from a policy, do you have a bias towards certain journals over others?
- 3) In publishing results, do you have a policy on disseminating the numerical values of your plotted points along with their statistical and systematic errors? E.g., a table in the paper, web-based distribution, or public databases such as DURHAM?
- 4) Do you have a policy for handling requests from others in the scientific community for your published results? If you do not have a policy, how is this handled?
- 5) Do you have a policy for making internal analysis notes available to the public? How do you handle requests for these notes?
- 6) Do you make your data available to scientists outside your collaboration?
 - 6a) If yes, at what level of processing are the data available?
 - 6b) What requirements come with accessing the data?
 - 6c) How long are the data available?
 - 6d) What if any requirements would be placed on publication of this data?
 - 6e) How often do scientists outside your collaboration request access to your data?
- 7) If you were to make data available outside your collaboration, what would be required for someone outside your collaboration to be able to utilize said data?
- 8) Do you make your data available to the general public?
 - 8a) If yes, at what level of processing are the data available?
 - 8b) What requirements come with accessing the data?
 - 8c) How long are the data available?
 - 8d) What if any requirements would be placed on publication of this data?
 - 8e) How often do you get requests from the general public for access to your data?

8.2 Questionnaire to theorists

- 1) For your sub-area, can you give one or two examples of large-scale codes or programs that are used by multiple members of the community? Are these codes posted for download by anyone? If not, do you understand the rationale?
- 2) Do you feel that when you see numerical results in your sub-area, you can generally believe that these are "reproducible"? For example, can you go back years later and find the code associated with the publication? (assuming the theory is not so well described that the details are irrelevant)
- 3) Do you find it easy to access experimental data in the form needed for theoretical analysis or comparison? If you don't generally access experimental data, feel free to speculate regarding the issues encountered from other people in your sub-area. In particular, comment about whether the data are sufficiently complete (errors, etc.) and in a useful format.
- 4) Do you often run into situations where published papers are difficult to access due to your lack of journal subscriptions?
- 5) Do you have any insight or recommendations for how the theory community might better provide information to the public outside the nuclear science community?

8.3 Sample Policies

Large collaborations usually have formal publication policies for producing, handling, and disseminating research results. These policies contain detailed instructions for handling VOR publications, preliminary analyses and figures for talks, reports, Ph.D. theses, and data requests from the media and general public. Without such formality it would be difficult for a large number of collaborators to behave consistently and coherently. The risk of releasing an erroneous or incomplete result is greatly reduced. The following are some examples:

http://www.phenix.bnl.gov/phenix/WWW/docs/pub_policy/current/pub_policy.pdf

http://aliceinfo.cern.ch/secure/system/files/documents/editorial_board/ALICE_Publication_Policy_V6.pdf

http://www.jlab.org/~obrien/CLASdocs/bylaws.htm#Data_Release

8.4 List of Acronyms

A number of acronyms have been used throughout this report. We list and briefly define all of them here for convenience.

- AIP: The American Institute of Physics.
- ALICE: One of the LHC experiments.
- APS: The American Physical Society.
- arXiv: An open-access system for distributing publications.
- ATLAS: Name of two distinct research sites
 - One of the LHC experiments.
 - A low-energy nuclear physics facility at the Argonne National Lab.

- BRAHMS: One of the RHIC experiments.
- CERN: The European organization for nuclear research located in Switzerland.
- CMS: One of the LHC Experiments.
- DDS: Digitized detector signal, a term defined in this report.
- DESY: The Deutsche Electron-Synchrotron Lab located in Germany.
- DOE: The Department of Energy.
- Fermilab: The Fermi National Accelerator Laboratory.
- Gammasphere: A low-energy nuclear physics facility at the Argonne National Lab.
- Gbyte: A unit of data corresponding to roughly a billion bytes.
- Hall-A ... Hall-D: The four experimental halls at Jefferson Lab.
- HEPDATA: The Durham high-energy physics data project.
- INSPIRE: A particle physics literature database that is an upgrade to SPIRES.
- INT: The Institute for Nuclear Theory.
- JLab: The Thomas Jefferson National Accelerator Laboratory.
- LHC: The Large Hadron Collider.
- NNDC: The National Nuclear Data Center.
- NSAC: The Nuclear Science Advisory Council.
- NSCL: The National Superconducting Cyclotron Facility.
- NSF: The National Science Foundation.
- Pbytes: A unit of data storage corresponding to roughly a million Gbytes.
- PDS: Processed detector signals, a term defined in this report.
- PHENIX: One of the RHIC experiments.
- PRC: The Physical Review C, an APS journal.
- PRX: The Physical Review X, an open-access APS journal.
- QCD: Quantum Chromodynamics.
- RHIC: The Relativistic Heavy Ion Collider.
- SciDAC: Scientific Discovery through Advanced Computing
- SLAC: The Stanford Linear Accelerator Center.
- SNO: The Solar Neutrino Experiment.
- SPIRES: A high-energy physics literature database.
- STAR: One of the RHIC experiments.
- Super-Kamiokande: A large underground neutrino experiment in Japan.
- Tbytes: A unit of data storage corresponding to roughly a thousand Gbytes.
- USQCD: The Lattice QCD collaboration of U.S. users.
- VOR: Version of record, a term defined in this report.