NIH ROADMAP MOLECULAR LIBRARIES AND IMAGING PROGRAM

NEEDS ASSESSMENT REPORT

January 2010

Table of Contents

Executive Summary

This report discusses the findings of a needs assessment of the Roadmap Molecular Libraries Program (MLP). The MLP is an integrated set of initiatives, the goal of which is to provide academic researchers high throughput (HTS) screening and chemistry resources to find and develop small molecules that can serve as chemical probes for research. The initiative consists of three main components: (1) a large, shared collection of small molecules, the Small Molecule Repository (SMR); (2) a network of screening and chemistry centers; and (3) a public database of all assay results (PubChem).

The purpose of this needs assessment is to: (1) assess whether the program accomplished its goals during the initial pilot phase (FY2004 to FY2007), and (2) gather feedback from network users and potential users on their level of satisfaction with MLP services. For this assessment, a series of customer satisfaction surveys was administered to current, former or prospective users of the MLP between November 2008 and July 2009.

MLP accomplishments

In general, the SMR, centers and PubChem met their goals for productivity. The SMR grew to 300,000 compounds, PubChem grew to 44,000 users per weekday and the centers produced 68 probes at a decreasing cost.

Satisfaction of network users and potential users

Those surveyed favorably compared the MLP centers to other screening centers. They most commonly listed the compound collection, screening capacity and chemistry support as core features to be maintained or enhanced. They reported that the review process for assay projects and the requirement for assays to be HTS-ready were two main barriers to participation in the program. Many also reported the intellectual property (IP) and data sharing policy as a barrier to participation in the network.

Areas for improvement

Key areas for improvement mentioned were increased support for assay development, a better review process for assay projects and increased chemistry capacity. Users of PubChem suggested improvements in the database to make it more user-friendly. Additional types of outreach were suggested, including a meeting specifically for MLP users and potential users. Some principal investigators (PIs) of non-MLP centers proposed that future funding come from user-fees or grants justified by the production of probes.

Overall, the MLP seems to have met most of its major goals. It received positive feedback from most of those surveyed, although there were some suggestions on how to further increase the effectiveness of the program.

Introduction

The Molecular Libraries Program (MLP)

Small molecules have proven to be extremely important to researchers to examine function at the molecular, cellular, and *in vivo* level in a way that is complementary to genetic studies. Such molecules have also been proven to be valuable for treating diseases, and many medicines marketed today are from this class.

A key challenge is finding a small molecule that is effective at modulating a given biological process or disease state. Currently, researchers must systematically screen tens or hundreds of thousands of small molecules to find a successful match between a chemical and its target. This process is known as high-throughput screening or HTS. The pharmaceutical and biotechnology sectors had developed private HTS capabilities for the purposes of drug development, but similar capabilities had not been widely available in the public sector prior to the establishment of the MLP.

Due in large part to technological advances in chemistry, robotics, and informatics, the MLP has been able to provide these capabilities to the public sector. Through access to large-scale screening centers, a large, shared compound collection and a public database to manage large amounts of diverse data, the public sector now has the ability to more readily conduct HTS experiments. With these capabilities, researchers have been able to identify small molecules and develop them into chemical probes. These probes will lead to new ways of examining the functions of genes and signaling pathways in health and disease.

The NIH also anticipates that the MLP will facilitate the development of new drugs by providing probes that will serve as lead compounds. This should be especially true for rare diseases that may not be attractive for drug development by the private sector.

Components of the MLP

The MLP is composed of several coordinated efforts. The central components of the MLP are a network of screening centers that conducts HTS assays and provides chemistry support for probe development, a central library of small molecules, and a public database on the compounds and results of assays (PubChem). In support of the network, the MLP funds the development of assays, the synthesis of novel compounds, and the development of new instrumentation. In addition, it sponsors a database on published imaging agents, Molecular Imaging and Contrast Agent Database (MICAD).

Screening centers

During the pilot phase, from 2004 through June 30th, 2008, the Molecular Libraries Screening Center Network (MLSCN) was comprised of nine extramural screening centers and one intramural screening center. In July 2008, the MLP began its production phase when five of the ten MLSCN centers and four new centers formed the Molecular Libraries Probe Production Center Network (MLPCN). The function of the centers is to screen submitted assays against the small molecules in the compound collection.[1] The assays are submitted by the research community through two initiatives: the small grant/resource access initiative for HTS-ready assays (R03/X01) and the Fast Track opportunity for rapid entry into the MLSCN.

Although, ideally, the HTS assays are ready to run when assigned to a center, some need further development before implementation. All assays undergo a primary screen to look for initial "hits" (i.e., those compounds that show an effect in the screen). Hits are then confirmed and/or followed up with a number of other assays, sometimes termed "secondary screens" or "confirmatory screens". These tests confirm the specificity of the effect of the hit and its potential for being a useful probe. Some of the assays provide an opportunity to discover different kinds of probes, such as agonists and antagonists, leading to multiple probe projects. After confirmation of the hit, a common first step in characterizing the hits is to perform structure activity relationship (SAR) profiling, in which the chemical structures of the hits and related compounds are compared with their activities against the target to distinguish functional groups that may be critical for activity in the assay. SAR can be done through "analog by catalog" (e.g., by purchasing related compounds), and often reveals the best candidate for follow up chemical modification.

Small Molecular Repository (SMR)

The SMR was designed to acquire, manage, store and distribute the compounds that are screened by the centers. To build the collection, it purchases compounds from companies and receives novel compounds from the laboratories of investigators funded through both the MLP (e.g., the Pilot Scale Libraries initiative) and other NIH programs (e.g., Chemical Methodologies and Library Development [CMLD] centers). The shared collection allows annotation of one "library" of compounds, with the goal of enhancing the utility of the data collected by the centers. The compounds undergo quality control at the repository and are distributed to the centers for screening. Screening centers are expected to deposit novel compounds to the library, increasing the diversity of the collection.

[1] Assays are simple tests which give a signal (e.g., luminescence, color, etc.) when one of the many small molecules screened is having an effect on the target, pathway, etc.

PubChem and MICAD

PubChem was designed to be an open-access database for structures and assay results from screens performed by the MLSCN. In addition, it houses compound information from other sources, including the scientific literature and for-profit institutions. MICAD is an online repository of data on molecular imaging agents for a broad range of imaging modalities (http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/micad/home.html). Its overall goal is to provide a freely accessible, comprehensive electronic resource for information on all *in vivo* imaging agents.

Needs Assessment

The MLP completed a three year pilot phase (FY2004 – FY2007), during which it established the infrastructure for the network, library and databases. The objectives of this report are to inform the MLP Working Group on the status of the network as of the end of the pilot phase (June 30, 2008) and to communicate suggestions from the different stakeholders involved. The MLP Working Group, the Office of Strategic Coordination, and the Evaluation Office worked together to design a study to obtain information to assess the satisfaction with the network of MLP stakeholders. This was accomplished by gathering and analyzing data on user opinions from several surveys, along with general MLP program data detailing the progress of the network. Information collected from the users included current issues that need to be addressed, suggestions for improvement, additional capabilities the network should acquire, and future directions the program should explore.

For this report, several key populations were surveyed to obtain feedback on their satisfaction with the initiative. These groups include: assay providers, compound providers, PubChem users, MICAD users, commercial suppliers of compounds, the PIs of the centers from the MLSCN and the MLPCN, the PIs of screening centers outside the network, and scientific meeting attendees. The experiences of these populations cover the critical components of the initiative, from the development of both assays and the compound library to the dissemination of data with PubChem.

The needs assessment was conducted at this stage to establish a baseline for measuring the impact of changes to the initiative in the production phase. The baseline data includes usage data, implementation measures (e.g., number of compounds in the library), primary outputs data (e.g., number of probes produced by the network), and customer satisfaction data from the users.

Organization of report

This introduction is followed by a methodology section, which explains the methods used for gathering information for the assessment, a results

section, which details the main findings for each component, and a conclusions section, which summarizes and synthesizes the findings.

Methods

The approach to assessing the MLP was to field a series of customer satisfaction surveys and interpret the responses in light of the accomplishments of the program. Customer satisfaction surveys are an effective means of collecting information about the needs of clients, which in this case are the needs of current, former or prospective users of the MLSCN/MLPCN.

Baseline data for the needs assessment were collected during the first year of the production phase from November 2008 through July 2009. MLP staff provided data on accomplishments of the MLSCN during the pilot phase, which ended June 30, 2008. To conduct the customer satisfaction surveys, separate survey instruments were developed for particular groups of individuals. Each group represented a different target population of MLSCN/MLPCN users or non-users. The primary method used to select and obtain information from individuals was via a Web-based survey, although other modes were used, depending on the survey being administered.

All surveys were submitted for approval by the Office of Management and Budget (OMB) under the generic clearance to conduct voluntary customer/partner surveys of the National Library of Medicine (OMB control number 0925-0476). Prior to OMB approval for the surveys, an open Request for Information (RFI) was conducted to ensure the receipt of timely input from both MLSCN users and non–users. RFI respondents self-identified as network users and non-users. Administrative data also were collected by members of the MLP Working Group to supplement the survey and RFI data.

Sampling

The size of the target population varied according to the target group (Figure 1). In most of the surveys with a small target population, a census of individuals was conducted, e.g. all qualified respondents were surveyed. For surveys where a reasonably complete and up-to-date list of potential respondents was available, a list served as the sampling frame. For one survey, respondents self-selected into the sample by responding to broad invitations made at professional conferences. Convenience samples[2] were used for other surveys to select respondents with highly-specific characteristics.

[2] Non-probability samples drawn from that part of the population which is close at hand.

Figure 1: Survey Target Populations		
Survey	*Targeted Number of Respondents*	*Sample Type*
Assay providers[a]	Approx. 141	Census
Commercial suppliers[b]	2	Census
Compound providers[c]	Approx. 38	Census
Meeting attendees[d]	Approx. 15 per conference (2 conf.)	Convenience
MICAD users	Approx. 100	Convenience
PIs of MLPCN[e]	4	Census
PIs of MLSCN[f]	10	Census
PIs of non-MLP centers[g]	Approx. 48	Census
PubChem users	Approx. 100	Quota/Convenience
RFI[h]	Approx. 4000	Convenience

[a]Sent to recipients of HTS-ready R03/X01 awards, but not to Fast Track PIs.

[b]Sent to both commercial companies providing MLSCN probes as research tools.

[c]Sent to Pilot Scale Libraries and Natural Products PIs.

[d]Announced to attendees of sessions of the Experimental Biology meeting (May 2009) and the Endocrine Society meeting (June 2009).

[e]Sent to four new MLPCN center PIs.

[f]Sent to nine extramural MLSCN center PIs, one intramural MLSCN center PI (NIH government employee).

[g]Sent to all known PIs of US and foreign small molecule screening centers (based on Society for Biomolecular Sciences center list).

[h]Sent to MLP site, Roadmap listserv and Society for Biomolecular Sciences listserv.

Recruiting

Participation in the surveys was completely voluntary. If an individual chose not to participate in any of the surveys, this was believed to have no likely impact on their eligibility for or ability to receive future grants, contracts, or services.

Individuals were not offered incentives and did not receive any type of remuneration in return for their participation in their designated survey. All respondents were assured of the confidentiality of their responses.

Potential participants in the surveys were contacted:

- By email invitation
    - The email text contained the URL link for the appropriate survey; or
    - The email text provided information on the Website (with link) where the survey was posted

- By conference invitation
  - Representatives from the MLP Working Group attended professional meetings (conferences) and invited attendees of the conferences
  - A PowerPoint presentation slide with the address of the Website where the survey was posted was shown at the end of the presentation (Appendix A)

- By website invitation
  - A website (http://MLP.nih.gov) contained descriptive text and the URL link for the appropriate survey

Data collection

All information collected was subject to the appropriate privacy protections.

The main mode of administration of the survey was via the Internet/Web (Figure 2). The group of commercial suppliers completed the survey by telephone. Participants were free to answer any number of questions on the survey; thus (n), the number of valid responses[3] to a given question, often varies within the same survey.

Figure 2: Mode of Administration of Survey

Survey	*Mode*
Assay providers	Web
Commercial suppliers	Telephone
Compound providers	Web
Meeting attendees	Web
MICAD users	Web
PIs of MLPCN	Web (paper option)
PIs of MLSCN	Web
PIs of non-MLP centers	Web
PubChem users	Web
RFI	Web

Data collection for the surveys took place between November 2008 and July 2009. The duration and dates of completion of the surveys are listed in Figure 3.

[3] Valid responses are those responses that are both intelligible and relevant to the question asked.

Figure 3: Periods of Survey Data Collection	
Survey	*Period of Activity/Completion Dates*
Assay providers	One month/December 31, 2008
Commercial suppliers	NA[a]/December 22, 2008
Compound providers	One month/January 19, 2009
Meeting attendees	One month/May 23, 2009 (Round 1); One month/July 13, 2009 (Round 2)
PIs of MLPCN	Four months/March 2, 2009
PIs of MLSCN	Four months/March 2, 2009
PIs of non-MLP centers	One month/May 23, 2009
PubChem users	One month/June 8, 2009
RFI	One month/December 21, 2008

[a]Respondents were interviewed by telephone at their convenience.

Follow-up reminders about the surveys were sent to potential respondents between the time the surveys were first released and the survey completion dates. The form, timing, and frequency of these reminders depended on the survey. Reminders were emailed to potential respondents (where email addresses were known).

- Assay providers – Reminded two weeks after survey release date
- Commercial suppliers – Not applicable (surveys conducted at respondents' convenience)
- Compound providers – No reminder within one-month data collection period
- PIs of MLPCN – Reminded after three weeks, one to two times (per individual) within a four-month data collection period
- PIs of MLSCN – Reminded one to five times (per individual) within a four-month data collection period
- PIs of non-MLP centers – Reminded after two weeks
- RFI – No reminder within one-month data collection period

Limitations

- For most of the target groups, a complete census of members of the groups was taken. For groups where a census was not feasible, convenience samples were taken since it was difficult to use a Web survey to draw a representative, probability sample.

- The response rates varied for the surveys. For some of the surveys, the number of responses was less than expected (e.g., MICAD). This precluded drawing valid conclusions from the data provided by the few surveys with low response rates.

- A greater use of mixed-modes (e.g., email, Internet, telephone, and mail/paper surveys) to collect data from diverse population subgroups might have helped increase response rates (i.e., individuals' willingness to respond).

- Web survey response rates can be lower than comparable modes, such as mail[4], although systematic overviews of response rates in Web surveys are not widespread[5].

- In terms of coverage, the sampling frame might need to be adjusted and/or refined for future surveys of groups for which a census is not feasible or desirable (e.g., meeting attendees). This includes a need for better ways to identify and list the individuals who are best-suited to answer the survey questions so that improved representation of the target population is achieved.

Qualitative and Quantitative Analysis

Both quantitative and qualitative methods were used to analyze the survey data. Prior to analysis, the data were cleaned and coded to eliminate non-valid responses and any duplicate cases. Qualitative data were manually and independently coded by two experienced researchers. Any discrepancies between the two sets of coded data were adjudicated jointly by the coders based on a pre-established set of guidelines. Verbatim responses were analyzed for recurrent themes and terminology, and grouped into meaningful categories. Quantitative data were summarized and frequency counts, percentages, means and medians were calculated as appropriate.

[4] Couper, MP, Traugott, MW, and Lamias, MJ. 2001. Web survey design and administration. *Public Opinion Quarterly* 65(2), 230-253; Matsuo, H, McIntyre, KP, Tomazic, T and Katz, B. 2004. The online survey: Its contributions and potential problems. *ASA Proceedings of the Section on Survey Research Methods* 27: 3998-4000.

[5] deLeeuw, ED. 2008. Choosing the method of data collection. In: deLeeuw, ED, Hox, JJ and Dillman, DA, eds. *International Handbook of Survey Methodology.* New York, NY: Lawrence Erlbaum Associates, pp. 113-135.

Results

General Customer Satisfaction

To gauge overall customer satisfaction with the network, surveys were administered to key stakeholders in the MLP: assay providers, compound providers, MLSCN PIs, and MLPCN PIs. Assay providers are investigators that submit assay projects to be screened at the MLSCN. Compound providers submit to the SMR compounds that meet a series of criteria that will ensure their usefulness for HTS. MLSCN PIs are the principal investigators for the centers of the pilot stage of the MLP which developed the screening center capabilities. MLPCN PIs are the principal investigators of the probe production centers which are part of the current Probe Production Phase of the MLP that started in July 2008. In addition, an RFI was published to gather opinions from respondents who self-identified as either non-users of the network or current or former users of the network. Of those RFI respondents who were not current or former users, 13% had submitted a grant to the network, 25% felt they were very familiar with the network, and 50% felt that they were somewhat familiar with the network. Only 8% indicated that they were not very familiar with the network. The surveys and the RFI asked about experiences, challenges, and satisfaction with the network.

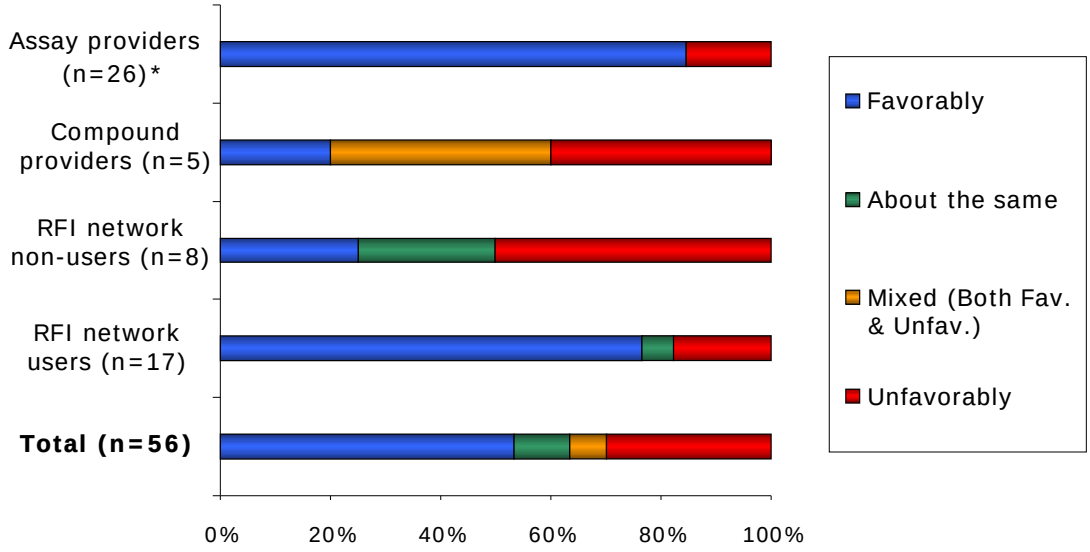
Comparison with other centers

To determine how well the network centers were performing, several groups were asked, “How do your experiences with the MLSCN/MLPCN compare to those at non-MLSCN/MLPCN centers?” (Figure 4). Overall, the respondents made a favorable comparison of the network with other centers (68%, n=56). However, the compound providers and non-users that responded to the RFI had a higher percentage reporting an unfavorable comparison (two of the five and four of the eight, respectively).

Generally, the favorable comparisons included references to the network’s “superior” efficiency and communication. The assay providers also included comments on the superior library, HTS capabilities, and better assay development of the network.

The unfavorable comparisons in all the groups included references to the network being slower than other centers in general and in terms of grant reviews and screen implementation. Two of those that indicated in the response to the RFI that they were non-users also stated that the network had less flexible HTS requirements for assays than other centers. Two compound providers stated that, in addition to being slower than other centers, the network had a more complicated and burdensome process for getting compounds screened.

Figure 4. Responses to "How do your experiences with the MLSCN/MLPCN compare to those at non-MLSCN/MLPCN centers?"



**n* is the number of valid responses to a given question

New publications and grant applications

Part of the MLP mission is to stimulate research through the development and use of its probes. To examine if the MLP was stimulating research among users, both the assay providers and compound providers were surveyed about new publications and grants resulting from their experience with the MLP.

When assay providers were asked, "Did the experience with MLSCN/MLPCN result in new publications?" 43% of the respondents (n=53) indicated that their experience had resulted in new publications. When asked, "Did the experience with MLSCN/MLPCN result in new grants?" 33% (n=53) indicated that their experience had resulted in new grants.

When compound providers were asked, "Did the experience with MLSCN/MLPCN result in new publications?" four of the eight respondents indicated that their experience resulted in new publications. When asked, "Did the experience with MLSCN/MLPCN result in new grants?" one of the eight respondents indicated that their experience resulted in new grants.

In its Assay Development for HTS initiative, the MLP provides small, exploratory (R21) grants to generate HTS assays that may in the future be submitted for screening at an MLP center. The PIs of these grants were not surveyed, but the number of publications produced from their work was determined from NIH progress reports. Forty-four publications were produced from R21s funded in response to the 2004 funding opportunity announcement (FOA), but a decreased number have been produced from grants funded in 2005 - 2007. Similarly, the R21 grants funded in 2004 have

resulted in the most grant applications (86 submissions), with fewer from grants funded more recently.

Continued interest in funding opportunity announcements

It is important for the continued success of the MLP effort that there be continued interest among researchers. To measure the amount of interest among researchers, the number of submitted applications was tracked over time for three important FOAs: HTS-ready Assays, Assay Development for HTS, and Pilot Scale Libraries.

For the HTS-ready Assays FOAs, after an initial 64 applications for the first receipt date, the number of applications stayed close to an average of 38 over the next six receipt dates. For the Assay Development for HTS FOAs, the number of applications submitted was a little over 100 in 2004 and in 2005 before increasing in 2006 with two announcements that received 164 and 132 applications and declining since with 48 applications for the first round of 2008. For the Pilot Scale Libraries FOAs, the number of applications has stayed near an average of 26.5 per year from 2005 through 2008.

Most important core features of network

Survey and RFI respondents were asked to list the three most important core features to maintain or enhance (Figure 5).

Figure 5. Most common response among respondents to, "What are the core features of the MLSCN/MLPCN that must be maintained or enhanced?"

Core feature	***Total (n=75)* %***	*Assay providers (n=31) %*	*Compound providers (n=6) %*	*MLPCN PIs (n=4) %*	*MLSCN PIs (n=8) %*	**RFI network users (n=26) %**
Compound collection	**48**	39	33	75	50	58
Screening	**48**	45	33	50	38	58
Chemistry	**35**	48	17	0	13	35
Network communication	**20**	16	50	75	38	4
Assay development	**16**	23	0	0	13	15
Informatics	**12**	16	0	0	0	15
Public access to data	**5**	3	0	0	38	0

**n* is the number of valid responses to a given question

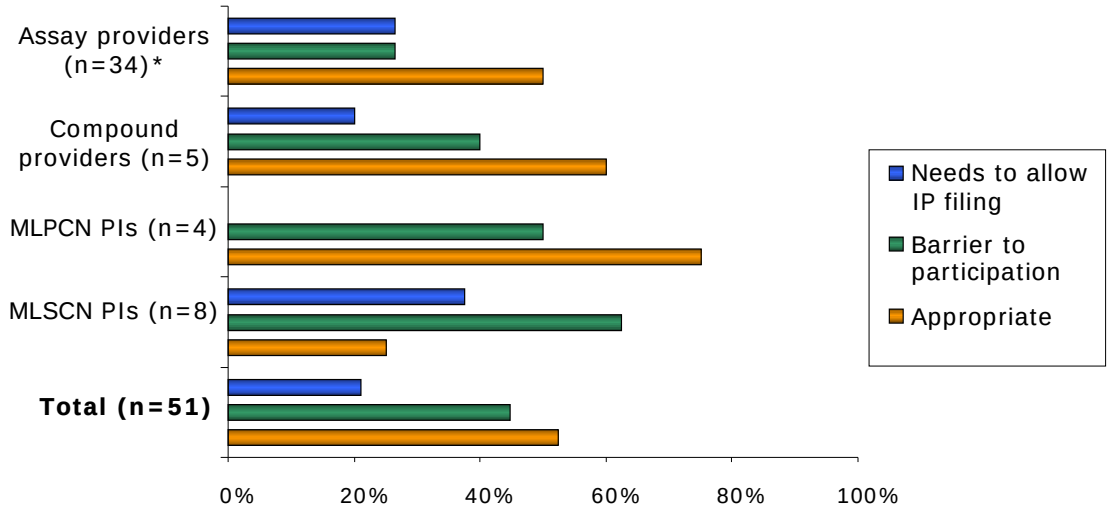
The compound collection and the screening core features were the two most often mentioned (each with 48%, n=75). Chemistry support for pursuing hits and developing probes was the primary feature reported to be important to maintain or enhance among assay providers (48%, n=31). Communication with the SMR on the status of compounds was foremost among compound

providers (three of the six). The MLSCN PIs had a substantial percentage who reported public access to data to be a core feature that needed to be maintained or enhanced (three of the eight).

Challenges with the network

Several groups were asked, "What are the challenges (or your thoughts) with the intellectual property (IP) and data sharing policy?" While the IP policy was found by the majority of those surveyed to be appropriate as it is (49%, n=51), 35% of respondents reported that the IP policy was a barrier to participation (Figure 6). Portions of the assay providers, compound providers, and the MLSCN PIs all suggested changes that would allow IP filing for those submitting data to PubChem.

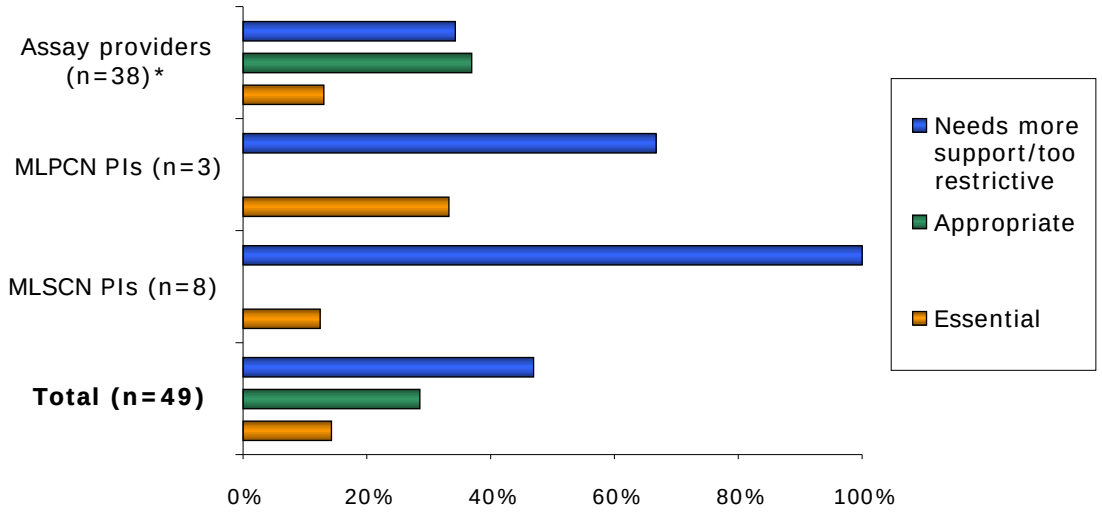
Figure 6. Responses[#] to "What are the challenges (or your thoughts) with the intellectual property (IP) and data sharing policy?"



[#]Respondents were allowed to list multiple challenges or thoughts
**n* is the number of valid responses to a given question

Assays that entered the MLSCN were expected to be "HTS-ready", that is, adaptable to miniaturization and automation. On the issue of the HTS-ready requirement for assay submissions to the network, all of the MLSCN PIs and most (two of the three responses) of the MLPCN PIs that responded to the question indicated a need for more support for HTS development (Figure 7). Thirty-four percent (n=38) of the assay providers reported the need for more support to meet the HTS-ready requirement. Portions of every group indicated that the HTS-ready requirement was essential for the MLP.

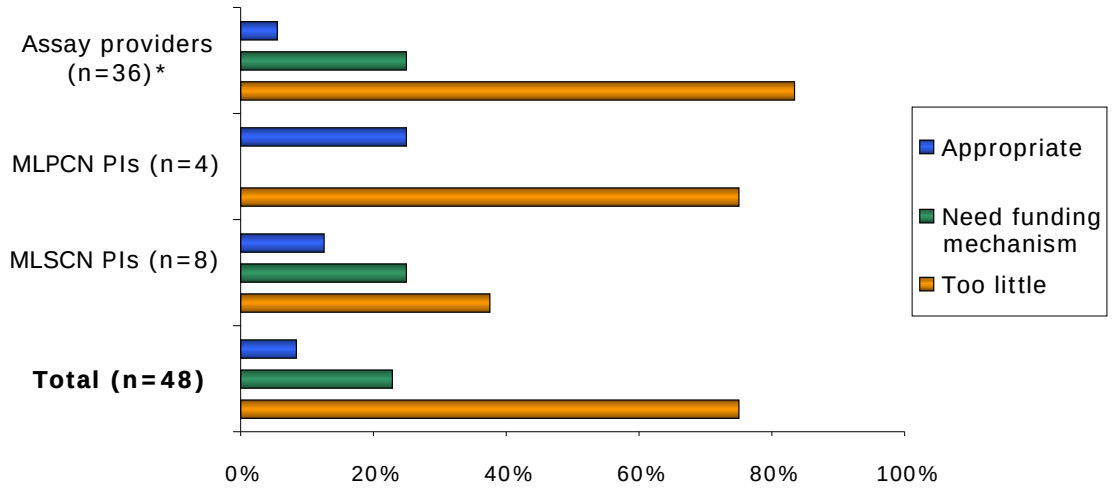
Figure 7. Responses to "What are the challenges (or your thoughts) with the requirement for assays to be HTS-ready or nearly ready?"



**n* is the number of valid responses to a given question

When asked, "What are the challenges (or your thoughts) with the availability of funding for reagent production/secondary screening?" among every group (assay providers, MLSCN PIs, and MLPCN PIs) the most common response was that there was too little funding (Figure 8). In addition, 25% of the assay providers (n=36) and the MLSCN PIs (n=8) indicated a need for a specific funding mechanism for reagent production and secondary screening.

Figure 8. Responses to "What are the challenges (or your thoughts) with the availability of funding for reagent production/secondary screening?"



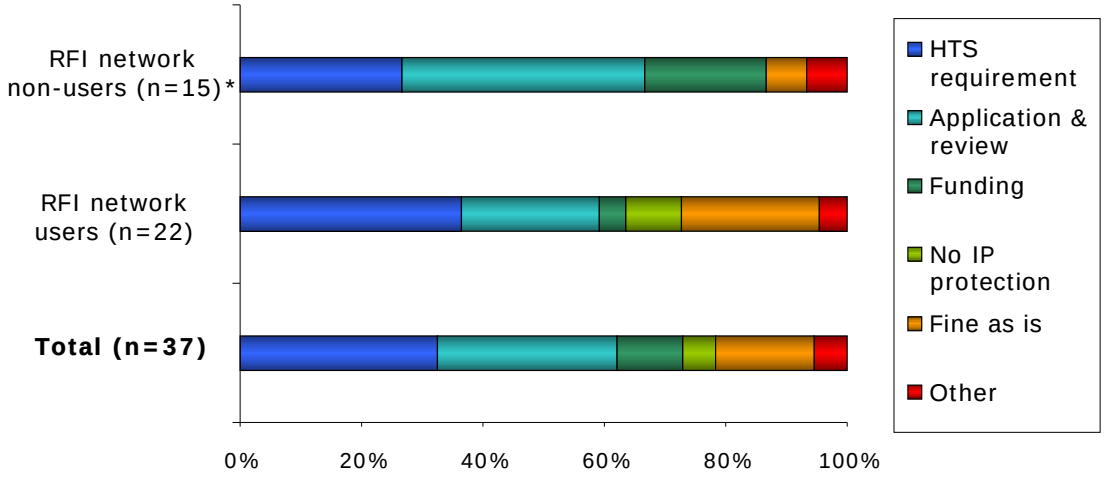
**n* is the number of valid responses to a given question

When network users were asked, “What are the challenges (or your thoughts) with the amount of time and effort required to write an application for access to the MLSCN/MLPCN?” they generally reported that it was appropriate, with all four of the MLPCN PIs, 89% (n=36) of the assay providers, and six of the eight MLSCN PIs indicating it was appropriate.

When asked, “What are the challenges (or your thoughts) with the amount of time from submission of application until results are obtained from screening center?” respondents among the assay providers, MLSCN PIs, and MLPCN PIs had different viewpoints. Almost half (46%, n=35) of assay providers indicated it was “reasonable” to “fast,” and 41% indicated it was slow. For MLSCN PIs, six of the eight respondents indicated it was reasonable, and two indicated it was slow. One of the three MLPCN PIs indicated it was fast and two indicated it was slow.

The responses to the RFI request to “Please describe the greatest challenge(s) for submitting assays or compounds to the MLSCN/MLPCN,” included feedback from both users of the network and non-users (Figure 9). Forty percent of non-users (n=15) identified the application and review process as most challenging, and 27% indicated the HTS-ready requirement for entry to the MLSCN. Thirty-six percent of users (n=22) indicated the HTS-ready requirement and 23% identified the application and review process as one of the biggest and most burdensome challenges.

Figure 9. Responses to RFI statement, “Please describe the greatest challenge(s) for submitting assays or compounds to the MLSCN/MLPCN.”

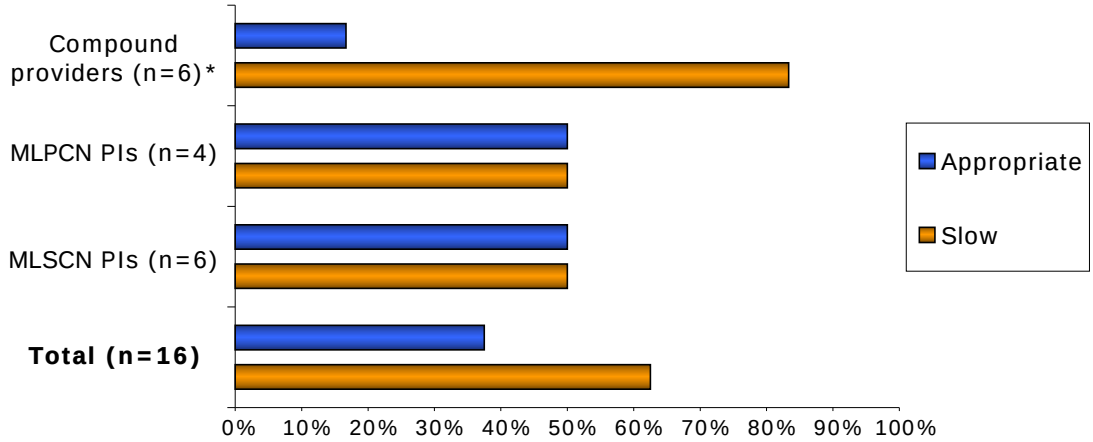


**n* is the number of valid responses to a given question

Compound providers were asked, “What are the challenges (or your thoughts) with the amount of time/compensation to prepare compounds and data files for submission to the compound repository?” Five of the six respondents stated that the time and compensation available for preparing compounds and data files for submission to the SMR was appropriate.

Of the compound providers that responded to the question, "What are the challenges (or your thoughts) with the amount of time from submission of a probe compound until that compound is part of a screening library?" five of the six reported the time to be slow (Figure 10). Half of both the six MLSCN PIs and the four MLPCN PIs felt the time was slow and the other half of each group felt that the time was appropriate.

Figure 10. Responses to "What are the challenges (or your thoughts) with the amount of time from submission of a probe compound until that compound is part of a screening library?"

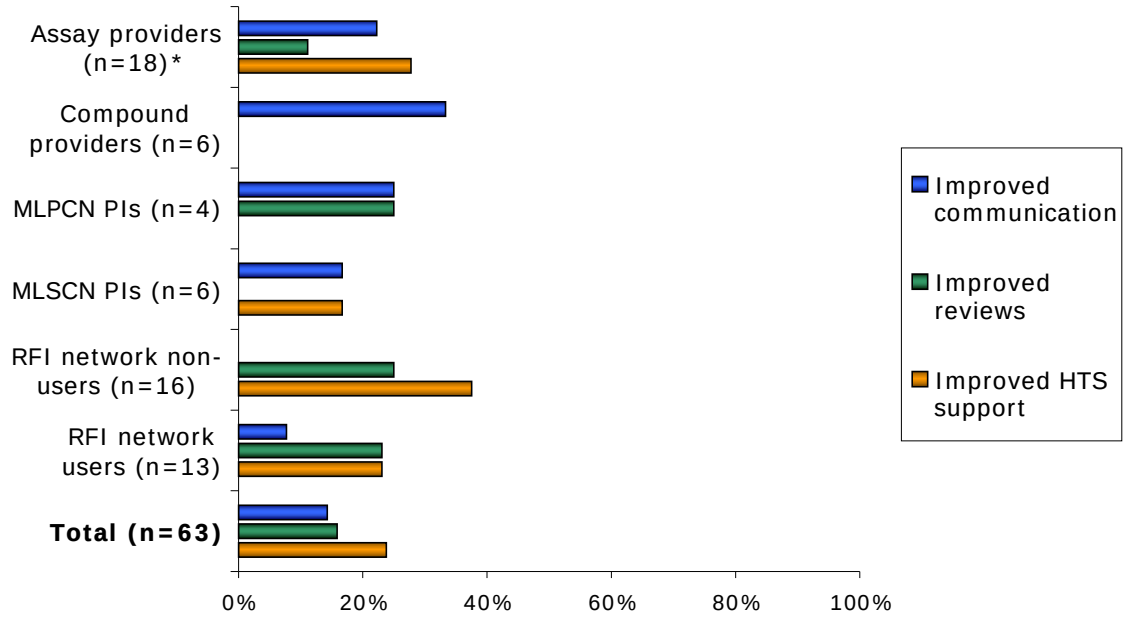


**n* is the number of valid responses to a given question

Recommendations for the network

The three most common suggestions in response to the question, "What specific approaches or concepts would address any of the challenges above?" were: improved support for initial HTS development, improved reviews, and improved communication (Figure 11). The request for more support with HTS development was made by the RFI respondents, both users (23%, n=13) and non-users (38%, n=16), and the HTS-ready assay providers (28%, n=18). Some of the RFI respondents (24%, n=29), assay providers (11%, n=18), and MLPCN PIs (one of the four respondents) suggested improving the review process for assay projects. Suggestions included streamlining the review process, clarifying review criteria, increasing the quality of reviewers and reducing reviewer bias. One assay provider suggested a specific review process for probe development. A need for improved and standardized communication throughout the network was indicated by the assay providers (22%, n=18), compound providers (two of the six respondents), and center PIs (two of the 10 respondents). The assay providers specifically requested a standardized communication plan for centers and assay providers.

Figure 11. Responses to "What specific approaches or concepts would address any of the challenges above?"



**n* is the number of valid responses to a given question

Comparison of MLSCN to other screening centers providing service to community

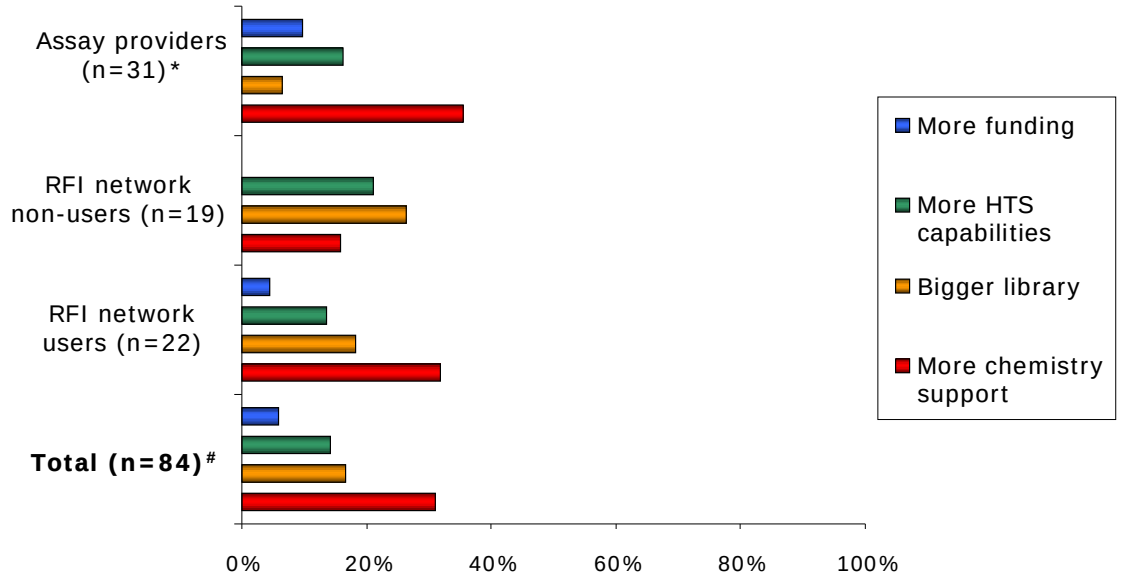
From information gathered from the Society for Biomolecular Sciences (SBS) website and applications to the MLSCN FOA in 2004, it was extrapolated that the number of small molecule (chemical) screening centers providing a service increased from approximately 23 in 2003 to 47 in 2008.

The MLP network is unique among screening centers in the scale and number of screens performed. The network performed the most large-scale screens while other centers performed smaller or fewer screens each year.

Capabilities to add to network

In terms of additional network capabilities desired, the RFI respondents, both network users and non-users, and assay providers mentioned (in descending order of frequency) additional chemistry support, a larger compound library, and more HTS capabilities, and to a lesser extent, more funding (Figure 12).

Figure 12. Responses to statement, "Please describe the capabilities that you would like the network to offer that it does not."



**n* is the number of valid responses to a given question
#Includes compound provider, MLSCN PI, and MLPCN PI responses not shown separately

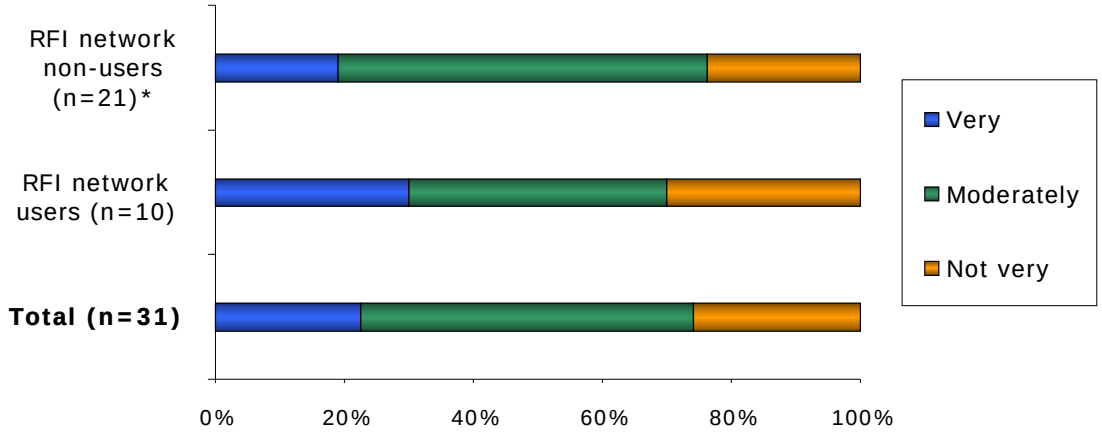
Areas/targets insufficiently explored with small molecules

A wide range of areas of biology that had not been adequately explored with small molecules were identified among all those surveyed and those that responded to the RFI. Few targets were mentioned consistently, with only protein-protein interactions being mentioned by every group that was surveyed (assay providers, compound providers, MLSCN PIs, and MLPCN PIs). Those that responded to the RFI most often indicated no area had been adequately explored.

Satisfaction with PubChem

Most (74%, n=31) of the RFI respondents had a positive response on the user-friendliness of PubChem (Figure 13). In response to the statement in the RFI, "Please comment on the user-friendliness of PubChem", 30% of the respondents that identified themselves as network users (n=10) indicated it was very user-friendly, 40% indicated that is was moderately user-friendly, and 30% indicated that it was not very user-friendly. Similarly, 19% of those that identified themselves as network non-users (n=21) indicated that PubChem was very user-friendly, 57% indicated it was moderately user friendly, and 24% indicated it was not very user-friendly (for additional information, see discussion in the section on Data Dissemination).

Figure 13. Responses to RFI request to "Please comment on the user-friendliness of PubChem."



**n* is the number of valid responses to a given question

Recommended additions to PubChem

Some respondents to the RFI made a single recommendation for additions to PubChem including: links to publications, basic information on compounds, Spotfire and Pipeline Pilot, better search options, FDA approval data, the portion of the library screened for compounds, confirmed therapeutic uses, and activity profiles. There was also one recommendation to remove redundant outputs to make browsing searches easier.

Centers Productivity

Goals and operations of centers

The centers were assigned assays from two sources, the HTS-ready (R03/X01) award and the Fast Track opportunity. The initial goal for the pilot phase (FY2004 – FY2007) was for each center to screen 20 or more assays with up to 100,000 compounds per year. In 2006, the initial goal was modified such that each center was to screen more than 10 assays against 150,000 compounds in 2007 and more than 10 assays against 200,000 compounds in 2008. In 2008, 30% of MLSCN centers screened more than 10 assays against 200K compounds.

The overall number of assays run by the MLSCN was 144 R03/X01 assays and 28 Fast Track assays. On average, it took about five and a half months to obtain primary results. The average number of compounds screened in primary screens was approximately 93,000.

The duration of the screening campaigns, from assay assignment to filing a probe report, ranged from four to 36 months, with an average time of 17 months. These campaigns resulted in 283 probe projects (some screens had multiple probe projects). Of these, 137 chemistry projects progressed through the initial SAR determination. Sixty-eight probe projects, which came from 41 assigned assays, had been completed as of June 30, 2008.

Costs of products

The cost of the products of the centers has gone down as the infrastructure has been built. The cost to screen each well was $15.46 in 2005, and decreased to $1.03 in 2008. As the production of probes increased, the cost per probe decreased. In 2006, when the first few probes were produced, the MLP had expenditures of $24.65 million per probe produced. In 2007 and 2008, expenditures per probe were $1.86 and $2.01 million respectively. In comparison, one of the PIs of a non-MLSCN/MLPCN center mentioned (in an open-ended question on the customer satisfaction survey) that producing a probe for $1-to-2 million would in their case justify funding.

PubChem data quality

From information gathered from the MLP Working Group, during the pilot phase, no secondary assays were submitted to PubChem. PubChem primary screen depositions averaged five to six months following implementation of the assay. The MLP Working Group found the data to be approximately 90% accurate. The inaccuracies were due mostly to errors in data transfer to PubChem.

When the PIs of MLSCN centers were asked in the survey “What is the quantity, timeliness, and accuracy of data submitted to PubChem by your center?” all six respondents gave a positive reply, and all but one rated their center as good or excellent with respect to these characteristics.

Compound Collection

The goals of the library initiative as stated in the 2007 MLP Formal Phase and Transition Proposal were:

1. To acquire a collection of up to 500,000 compounds using rational selection strategies for commercial acquisitions and intensive efforts to obtain novel compounds from non-commercial sources.
2. To implement 100% quality control for purity and identity for incoming compounds, and 10% maintenance quality control assessment of the collection per year to ensure reasonable quality.
3. To store and process compounds using best practices to maintain compound quality.
4. To distribute the compound collection in multiple formats, well arrangements, and plate types to meet the differing needs of the centers within the MLSCN.

Size of the SMR collection

The SMR was established through a contract to Discovery Partners International (DPI) in August 2004. The collection had approximately 100,000 compounds by November 2006 with a projected goal of 300,000 by Fall 2008 and an eventual goal of up to 500,000. At the end of the pilot phase, June 2008, the collection had reached 312,107 compounds. Of those, 281,271 were available for shipment to the centers (i.e., were of sufficient quantity and concentration).

From the open-ended responses to the RFI, fifteen respondents recommended improving the compound collection of the network. These improvements included increasing the size and diversity of the collection and the "continued evaluation of the composition of the NIH molecule library." Six respondents also called for smaller or specific libraries for such compounds as "kinase inhibitors, enzyme inhibitors, etc."

Diversity of the SMR collection

The SMR was designed to include compounds of diverse types including:

- known drugs, toxins, etc.;
- natural products;
- targeted libraries for proteases, kinases, nuclear receptors, etc.; and
- diversity compounds or compounds of unknown activity.

Specific criteria for each type of compound were developed. As of June 2008, the library contained:

- 230,084 diversity compounds of unknown activity;
- 24 Drug Enforcement Agency-registered compounds;

- 9,900 non-commercial compounds;
- 1,536 natural products compounds;
- 1,752 compounds from specialty sets such as approved drugs, failed clinical compounds, toxins; and
- 10,274 compounds from targeted libraries, including 3,428 for kinases and 3,900 for G-protein coupled receptors.

The compound providers were asked, "What specific approaches or concepts would address any of the challenges [with the SMR]?" Three of the eight compound providers indicated the SMR should be novel and gather diverse compounds, and suggested that processing could be shortened by allowing PIs to submit their own quality control data directly to the SMR.

Quality of the SMR collection

The SMR has established 19 standard operating procedures (SOPs) specifically related to operations performed on behalf of the NIH. The SOPs cover a wide range of activities, including final plating and shipments to the centers. The SOPs have been reviewed and updated in 2005, 2006 and 2007.

Incoming compounds are weighed and tested for solubility, identity and purity. The SMR accepts samples with the correct identity and a purity of at least 90%, as determined by liquid chromatography/mass spectrometry or high performance liquid chromatography. In November 2006, the pass rate for new compounds was 88%; in June 2008 it was 82% (322,047/391,170).

Maintained compounds are checked biannually for quality. A sample of 10% of the collection is randomly collected in lots of 200 compounds for quality control via liquid chromatography/mass spectrometry or high performance liquid chromatography. The pass criterion for each sample is the same as for the original deposition: 90% purity or greater. In the summer of 2008, 2,728 of the 2,800 tested compounds passed the criterion for a pass rate of 97%.

Another issue regarding the quality of the compound collection involves certain compounds that are active in a large number of assays. These compounds may have structures that interact non-selectively with biological macromolecules. These structures are known to indiscriminately interact with biological molecules through non-specific interactions and are termed "promiscuous inhibitors"[6]. Promiscuous inhibitors are false positive hits in screens. These compounds waste screening resources and are to be avoided in development of a screening library. The library has at least two compounds that were selected for cherry picking in over 25 assays.

In the open-ended responses among the assay providers, three indicated a need to remove promiscuous compounds from the collection.

[6] McGovern *et al.* 2002. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of Medicinal Chemistry* 45: 1712-1722.

Distribution of the compounds

One of the primary responsibilities of the SMR is to efficiently distribute the collection to the centers both for the primary screens and for re-screening of subsets of positive hits ("cherry picks"). By June 2008, the SMR had distributed 224,819 distinct compounds in a total of 536,825 wells. The total number of cherry-picked compounds distributed was more than 67,000 distinct compounds in 106,365 wells, in response to 407 orders. More than 60,000 cherry-picked compounds were picked more than once, with 45,308 being picked twice and two being picked more than 25 times each. The mean turn around time for cherry pick orders was six days.

For comparison to the model that MLP uses, PIs of non-MLSCN/MLPCN centers were asked, "How do other centers store and distribute their library contents?" The most common way of storage indicated by respondents was "frozen" (61%, n=18). Sixty-one percent of respondents specified that library contents were stored at minus 20 degrees Celsius. Also, sixty-one percent of respondents mentioned storing contents in dimethyl sulfoxide. For distribution, 39% indicated using microplates and 28% said compounds were "distributed directly." Thirty-nine percent indicated that their center did not distribute compounds at all.

Acceptance of compounds into the collection

When the compound providers were asked, "What are the challenges (or your thoughts) with the amount of time from submission of compounds until results are obtained from screening center?" six of the eight respondents indicated there was an issue with the timeliness of compound acceptance into the SMR. They characterized it as "slow," "complicated," and "tedious." Two of the compound providers also indicated that the time it took to get accepted compounds distributed to the centers for screens needed to be faster.

Center satisfaction with SMR

The MLSCN PIs were asked, "What is the quantity, timeliness, and quality of compounds submitted to SMR by your center?" Four of the five MLSCN PIs that responded to the question rated their center as good or excellent.

Six of the seven MLSCN PIs that responded to the question, "What is your center's level of satisfaction with the Small Molecule Repository (SMR)?" indicated a positive level of satisfaction with the compound collection (slightly or very satisfied). Only one PI indicated that he was not satisfied with the compound collection. When asked to specify his answer, he responded that the compound delivery was inconsistent, turnaround time was long, and the chemistry could be improved.

From the open ended responses, four of the MLSCN PIs indicated issues with quality control and assurance of the compounds, such as compound identification and purity. Five of the PIs that responded also reported issues with the time it took to incorporate submitted compounds into the library.

Enrichment of the collection

By June 2008, the SMR had acquired 9,900 non-commercial compounds through a variety of means, including contributions of compounds by:

- the Pilot Scale Libraries initiative of the MLP (5,220 compounds);
- the CMLD centers funded by NIGMS (2,308 compounds);
- other government sources (such as NIDA, NCI and NIEHS) (1,628 compounds); and
- the General Compound Solicitation (339 compounds).

Collection availability

The MLP makes a subset of the library available to researchers at cost. The NIH Clinical Collection contains approximately 450 compounds that have been in clinical trials and have drug-like qualities. Sixty-eight orders for the clinical collection were placed between October 4, 2007 and June 30, 2008.

Compound orders

As an indication of which types of compounds were taken into probe production by the MLSCN, 66,125 commercial compounds were shipped in cherry pick orders. In comparison, 727 non-commercial and 108 natural product compounds were shipped.

Assays

Has the Assay Development for HTS initiative been successful?

The goal of the Assay Development for HTS (R21) initiative is to build a portfolio of scientifically novel and technologically outstanding molecular probe development projects that would be eligible upon completion for entry into the MLSCN. It is open to all areas of biological and biomedical research, and both extramural and intramural investigators. In year one of the R21, $100,000 in direct costs was provided to investigators for the development of promising assay protocols for novel molecular targets, phenotypes and pathways. Assays are developed into a format that can be implemented in HTS, after which projects can enter the MLSCN via a Fast Track process in which the MLP Working Group reviews requests for access to MLSCN resources, using technical criteria of HTS readiness. Acceptance of R21 projects into the MLSCN in year two of the award triggered $25,000 in direct costs to the assay provider, permitting screening and follow up chemistry in collaboration with the center.

Assay Development for HTS applications

NIH records indicate that 670 applications were received up to June, 2008 and 168 were awarded. From those 168 awards, 76 primary HTS assay development assays had been run and 73 pilot screens had been completed.

Figure 14. Assay development for HTS applications from October 2004 to June 2008.

RFA	*Awards*	*Assays developed*	*Pilot screens*	*MLSCN center entry*	*HTS completed*
RM04-012	30	20	20	9	9
RM05-011	39	30	28	10	6
RM06-004	42	23	22	16	4
RM07-001	26	3	3	2	0

Biological or medical areas addressed

The 168 awards included a wide range of biological or medical areas, with cancer and infectious disease being the most common (Figure 15).

Figure 15. Assay Development applications by biological/medical area.

Biological/medical area	*Number of applications*
Aging	6
Arthritis/Muscle/Skin	1
Cancer	47
Dental	2
Developmental Disorders	3
Diabetes/Digestive/Kidney	15
Dietary Supplements	1
Drug Abuse	3
Environmental Health	1
Eye	2
General Medical Science	11
Heart/Lung/Blood	10
Infectious Disease	40
Inflammation	0
Psychiatric Disorder	4
Neurological Disorder	22
Total	**168**

Types of Assay Development for HTS applications

There were 28 phenotypic assays among the Assay Development for HTS applications and 84 of the applications were for an orphan target, which is a target with no known endogenous ligand.

Types of Assay Development for HTS investigators

There were 48 new NIH investigators among the Assay Development for HTS investigators according to self-reporting on the applications. In addition, there were 163 first time providers and five repeat providers among the Assay Development investigators.

Completion of Assay Development for HTS assays

Of those projects that were awarded prior to the May 2008 council, 76 out of 158 or 48% had completed development of the assay.

Of those 76 developed assays, a total of 37 Assay Development for HTS projects (49%) went into the MLSCN by June 30, 2008. Twenty-eight of these went through the Fast Track opportunity and nine others came in through HTS-ready R03/X01 initiatives. For 19 assays, the primary screen had been completed at the MLSCN by June 30, 2008.

Have the HTS-ready R03/X01 initiatives been successful?

The MLP offers HTS-ready assay providers access to the screening centers through three application cycles a year. The specialized small grant mechanism (R03) is used to solicit assays from the research community with up to $25,000 in total costs to defray assay reagents and travel of assay providers to the MLSCN centers. Following a standard application process and peer review, HTS-ready applications are selected based on the significance of the biological target/signaling pathway or cellular phenotype, need for small molecule probes to facilitate research on biological function or disease pathophysiology, HTS-readiness of the assay for implementation, and portfolio balance of assay targets/phenotypes of interest to multiple NIH institutes and centers (ICs). To provide screening opportunities both to MLSCN PIs and to intramural investigators who are not eligible to receive extramural grant support, the MLP spearheaded development of a new mechanism, the X01, which awards access to the screening resource, but no funds. From cycle 2 through cycle 4, only the X01 mechanism was used. From cycles 5 through 7, both the X01 and R03 were used. In 2007, the NIH ruled that NIH intramural investigators were eligible to receive funds from the Common Fund (Roadmap). This development has led the program to discontinue use of the X01 mechanism even though other NIH programs have incorporated it.

Two hundred ninety-two HTS-ready applications were received up to June, 2008. Of those, 144 were awarded. From those 144 awards, the MLP records indicated that 119 had been run and 61 had been completed.

The HTS-ready assay providers were surveyed on several different issues with the MLP. In their responses to some of the open-ended questions, ten assay providers mentioned that more support and assistance was needed with assay development. Related to this issue was a request from eight of the assay providers for enhanced interactions between screening centers and assay providers to allow more input on assay development. Some PIs emphasized a need for “significant NIH intervention with the assay provider and MLPCN participants” and “an opportunity for center staff to partner with assay providers”. Three assay providers mentioned that their contacts with centers were essential for completing the screen. As shown in Figure 7, many assay providers (45%, n=38) indicated the HTS-ready requirement is reasonable; a common response among assay providers was that for the network to function properly, assays should be nearly HTS-ready upon submission. Others (18%) thought that the requirement was not a good idea and that assay providers should consult with HTS facilities and screening centers to discuss the assays beforehand. Some added that there should be more pre-submission help for assay providers to make their assays nearly HTS-ready.

Biological or disease area addressed

The HTS-ready assay projects showed a similar disbursement of areas to that of the Assay Development for HTS projects, with cancer and infectious disease being the most common.

Figure 16. HTS-ready applications by biological/medical area.

Biological/medical area	*Number of applications*
Aging	4
Arthritis/Muscle/Skin	4
Cancer	43
Dental	0
Developmental Disorders	0
Diabetes/Digestive/Kidney	10
Dietary Supplements	0
Drug Abuse	1
Environmental Health	0
Eye	0
General Medical Science	18
Heart/Lung/Blood	11
Infectious Disease	32
Inflammation	2
Psychiatric Disorder	8
Neurological Disorder	11
Total	**144**

Types of HTS-ready applications

There were 53 phenotypic assays among the HTS-ready applications and 20 of the applications were for an orphan target. In addition, there were two profiling assays.

Types of HTS-ready providers

There were 44 new investigators among the HTS-ready providers according to self-reporting on the applications.

Source of HTS-ready providers

The most commonly reported characteristic of the HTS-ready providers was that they were from academic laboratories (Figure 17).

Figure 17. HTS-ready applications by characteristics of providers.

*Characteristic of provider**	*Number of applications*
Academic laboratory	126
Biotech firm	9
First time providers	105
Foreign	4
Government laboratory (including NIH)	7
Repeat providers	17

*One or more characteristics may apply to a given HTS-ready provider.

HTS-ready assays that are completed and resulted in probes

Sixty-one out of 144 or 42% of the HTS-ready assays were completed as of June 2008. Forty-one of those 61 completed assays resulted in 68 new probes.

HTS-ready assays at centers

The percentage of HTS-ready assays from outside the center institution and the percentage of those that are run and completed at the center are shown in Figures 18 and 19, respectively.

Figure 18. Percentage of HTS-ready assays from outside the center institution.

Center institution	*Percentage of assays from outside*
The Burnham Institute	30% (6/20)
Columbia University	70% (7/10)
Emory University	58% (7/12)
National Institutes of Health	67% (18/27)
Scripps Research Institute	53% (8/15)
Southern Research Institute	56%(9/16)
University of Pennsylvania	90% (9/10)
University of Pittsburgh	41% (7/17)
University of New Mexico	10% (1/10)
Vanderbilt University	57% (4/7)
Total	**52% (76/144)**

Figure 19. Percentage of HTS-ready assays from outside the center institution that are run and completed at center.

Center	*Percent run*	*Percent completed*
The Burnham Institute	100% (6/6)	50% (3/6)
Columbia University	29% (2/7)	14% (1/7)
Emory University	100% (7/7)	43% (3/7)
National Institutes of Health	83% (15/18)	67% (12/18)
Scripps Research Institute	75% (6/8)	25% (2/8)
Southern Research Institute	100% (9/9)	22% (2/9)
University of Pennsylvania	56% (5/9)	22% (2/9)
University of Pittsburgh	57% (4/7)	42% (3/7)
University of New Mexico	100% (1/1)	0% (0/1)
Vanderbilt University	75% (3/4)	50% (2/4)
Total	**76% (58/76)**	**39% (30/76)**

Probes

The MLP is designed to leverage a large central collection of small molecules with expertise in HTS and chemistry to produce probes. Potentially, the resulting probes could have broad and important impacts on science. They would provide new tools to researchers to study targets and pathways in novel ways. This broad potential of new probes to accelerate research is one primary reason for the MLP. Although there is often a delay between the funding of biological research and the appearance of products from the research, the MLP has produced probes early in its development.

Number and type of probes developed

From the first 161 assigned projects, 41 grant projects produced probes. Those 41 grant projects produced a total of 68 probes in the pilot phase. An analysis of the probe reports by the MLP Working Group revealed that of the 68 probes produced, 47 were discovered in biological assays active against a specific target while the other 21 were discovered in phenotypic assays with an unknown mechanism of action.

Quality and uniqueness of probes

Of the 68 probes that have been produced, 25 are for novel targets or activities. Of the remaining probes that are not for novel targets, 32 had characteristics that were improved from previous compounds. Seventeen had improved selectivity and 12 had improved potency. Most of the probes had a conventional mechanism of action such as competitive inhibition while three had novel mechanisms of action.

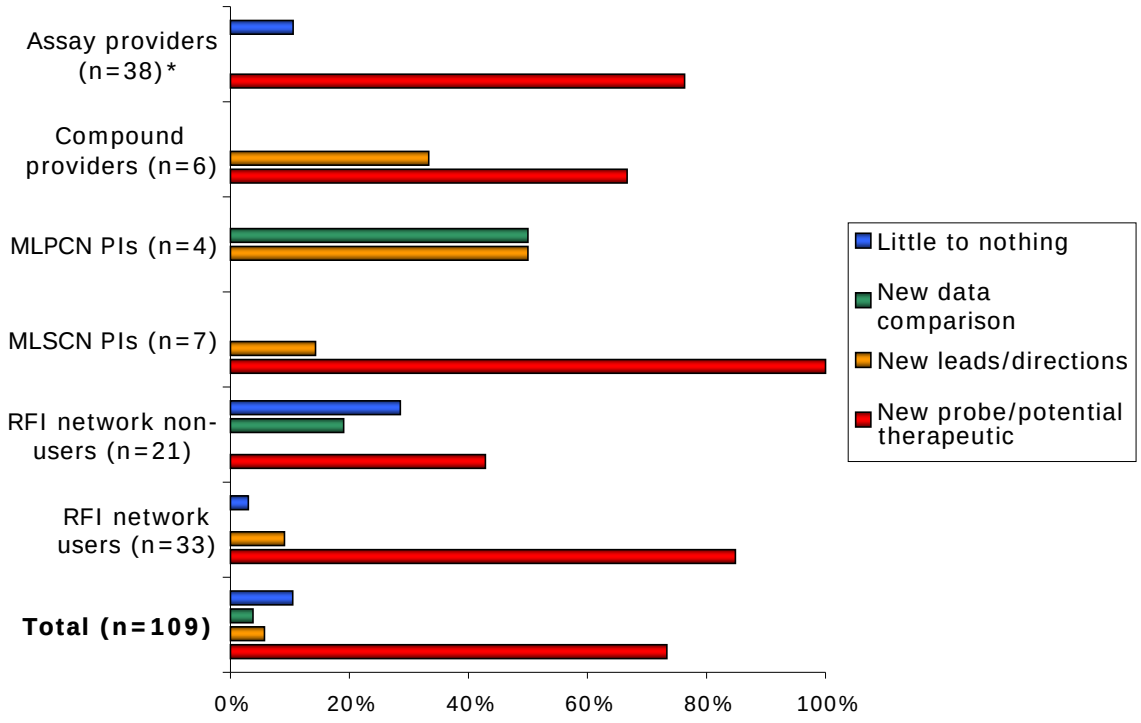
All five of the MLSCN PIs that responded to the question, “If produced probes are not for novel targets or activities, are the produced probes better than previously existing compounds?” indicated that the probes produced were better than previous compounds.

When asked, “Do the produced probes represent a novel mechanism or approach for regulating a pathway or target?” all six of the MLSCN PIs who responded gave an affirmative response.

Application of probes to research

The RFI and surveys of the assay providers, compound providers, and network PIs asked, "What is the current or potential application of MLSCN/MLPCN data and probes to your research?" The most common response (71%, n=109) was that there was a new probe or a potential therapeutic for their respective research interest (Figure 20). A portion of the non-users that responded to the RFI (19%, n=21) and two of the four MLSCN PIs that responded indicated that the data allowed for new comparisons between their findings and those of the network. Another portion of the non-users that responded to the RFI (29%, n=21) indicated that the data were of little or no use.

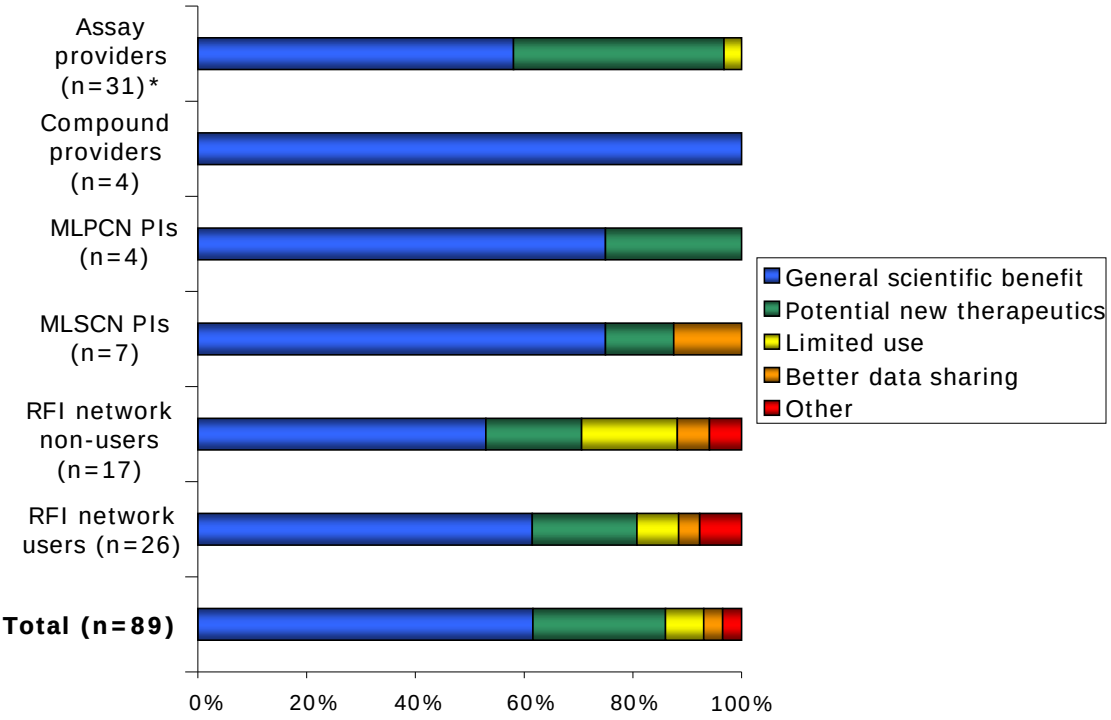
Figure 20. Responses to "What is the current or potential application of MLSCN/MLPCN data and probes to your research?"



**n* is the number of valid responses to a given question

Those that responded to the RFI along with the assay providers, compound providers, and network PIs were asked, "What is the current or potential application of MLSCN/MLPCN data and probes to the broader research community?" The most common response among all the groups (63%, n=89) was that there was a general scientific benefit of the MLP probes or data (Figure 21). Five of the six different groups also indicated the potential for new therapeutics with 25% of the total respondents. The non-users that responded to the RFI were the most likely of all the groups (three of the 17 respondents) to indicate limited use of the data and probes of the network.

Figure 21. Responses to "What is the current or potential application of MLSCN/MLPCN data and probes to the broader research community?"



**n* is the number of valid responses to a given question

Availability of and demand for probes

Having the probes commercially available to all researchers is critical for maximizing their impact on the research of the general researcher community.

By June 30, 2008, two commercial suppliers had decided to offer MLSCN-discovered probes to their customers. The commercial suppliers were asked, "What number of produced probes is available to the broader user community?" The two commercial suppliers surveyed indicated that they make a total of three MLSCN probes available to the broader community.

The commercial suppliers were asked, ""How many requests have there been for commercially available probes?" The two commercial suppliers surveyed indicated that a total of 130 requests had been made for the three probes.

The MLP Working Group conducted a search on PubChem to determine how many of the 68 MLSCN probes were obtained from commercial sources prior to their designation as probes and found that half were from these sources.

The PIs of MLSCN centers were asked, "How many requests have there been for commercially available probes?" The four that responded indicated a sum

(aggregate) total of 131 requests with a mean of 33 requests per center.

The PIs of MLSCN centers were asked, “What number of compounds/probes identified through your center, as part of the MLSCN/MLPCN, have been requested by users?” The four that responded indicated a sum total of 531 requests with a mean of 133 requests.

Advancement of probes

The PIs of MLSCN centers were asked, “What number of probes have been advanced after leaving the MLSCN?” The six that responded indicated a sum total of 38 probes being advanced with a mean of about six per center. The compound providers were also asked this question. They indicated in three valid responses that zero probes have been advanced after leaving the network.

Publications, grants and patents resulting from probes

The PIs of MLSCN centers were asked, “How many publications came from the probes: Directly by your center?” From the seven valid responses, a sum total of 209 publications were reported with an average of 30 per center and a range of four to 50.

The PIs of MLSCN centers were asked, “How many new grant proposals emerged from the program (e.g., for further optimization or use of the probes)?” From the six valid responses, a sum total of 43 new grant proposals were reported with an average of seven per center.

The PIs of MLSCN centers were asked, “How many new licenses and patents emerging from the program have been applied for/granted?” From the seven valid responses, a sum total of 26 licenses and patents had been applied for with an average of four per center.

Data Dissemination

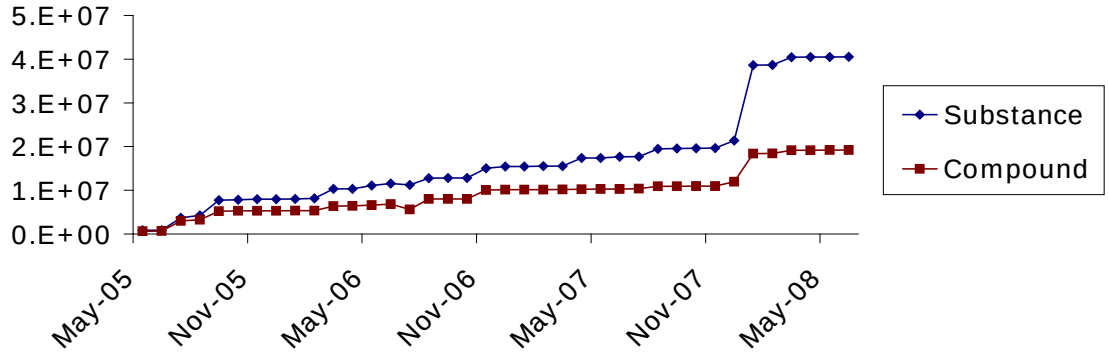
Effectiveness of data dissemination

Data dissemination is a critical component of the MLP. The MLP has two databases dedicated to collecting, storing, and disseminating knowledge about compounds and probes. Both are housed at the NIH National Center for Biotechnology Information (NCBI). PubChem serves as the database of chemical structures and activities and includes assay results from the MLSCN. MICAD is the database for imaging probes published in the scientific literature.

Development of PubChem infrastructure

The PubChem database was launched by NCBI in September 2004 with 600,000 chemical structures. In May 2005, it opened its automatic, web-based deposition system to public data from outside the NIH. By November 2006, PubChem involved 16 PhD-level computational scientists and PubChem data had been integrated into gene, protein, and literature databases of the NLM. From that time forward, the goals for PubChem were to "shift focus from start up infrastructure to increasing capacity to manage large volumes of data including data QC and maintenance, further integration with NCBI gene and protein databases, and the development and provision of data mining tools." By June 2008, PubChem had grown to include 40 million substances and 19 million compounds (Figure 22).

Figure 22. Growth in PubChem substances and compounds.



Improvements in the retrieval and browsing tools for PubChem

PubChem has added several new tools since its inception. For example in 2007, PubChem added a cross reference for protein targets and created an

assay-to-protein target link in Entrez. It provided a web-based tool for clustering chemical structures based on 2-D similarity, and developed a bioactivity summary tool that provides an overview of biological testing results and activity profiling for chosen compounds. Also in 2007, it added a structure-activity analysis tool that classifies compounds and assays simultaneously based on similarities of chemical structure, activity profiles and target.

When asked, "Have there been improvements in the retrieval and browsing tools for PubChem?" the majority of the assay providers (n=10), compound providers (n=4), MLPCN PIs (n=4) and MLSCN PIs (n=3) who responded to the question indicated that there had been improvements made to PubChem (71%, 75%, 57% and 100%, respectively).

When invited to give an open-ended response, six of the eight compound providers who responded indicated that PubChem needed further improvement or development. Two of these compound providers mentioned specific issues with queries and data outputs, including problems with truncation of results and additional information on summaries. Three of the nine MLSCN PIs indicated that PubChem was an important or powerful tool for scientists.

Recommendations for additions to PubChem were made by some (27%, n=30) of the RFI respondents who provided information on this topic. These included specific add-ons such as "Spotfire" and "Pipeline Pilot," as well as the need for additional data fields for information on FDA-approved compounds, commercial sources, and basic aspects of compounds. One respondent to the RFI who was a network user specifically mentioned adding direct links to publications and also suggested removing redundant outputs.

When asked to "List up to three software tools or features you would like to see added to PubChem," eighteen PubChem users made suggestions. Four indicated a substructure search, three mentioned additional structure display software, and two reported each of the following: ChemAxon applets, benchmarking for QSAR, and more "name" information.

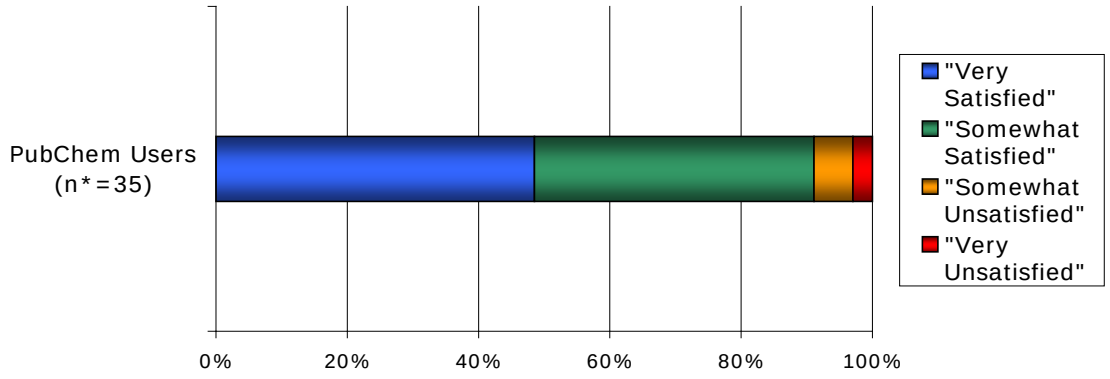
Experiences and satisfaction using PubChem

Both network users and non-users responded to the RFI request to "Please comment on your experiences using PubChem." Of the four valid responses from network users, three indicated that their experience was positive and one indicated a difficult experience. Of the 16 relevant responses from network non-users, 44% indicated a positive experience, 25% indicated a neutral experience, and 31% indicated a difficult experience.

The PubChem users were asked the question, "What is your level of satisfaction with PubChem?" Forty-nine percent (n=35) indicated that they were "very satisfied" with the user-friendliness of PubChem, 43% indicated

they were “somewhat satisfied,” six percent indicated they were “somewhat unsatisfied,” and three percent indicated they were “very unsatisfied.”

Figure 23. Responses to “What is your level of satisfaction with user-friendliness of PubChem?”



**n* is the number of valid responses to a given question

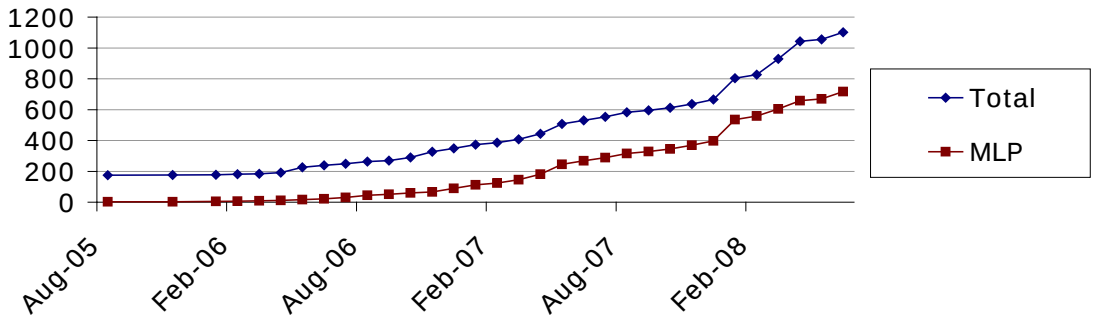
Related data based on open-ended survey responses by MLSCN PIs (n=9) showed that while PubChem was reported by three PIs to be an important or powerful tool for scientists in general, six commented that PubChem needed additional work to be effective. Examples included improving user-friendliness and having data that were consistent, standardized and better organized.

PubChem and data sharing was also discussed in most (three of the four) open-ended responses made by the MLPCN PIs. Two of the PIs indicated difficulty with browsing the data due to inconsistency with the data and difficulty uploading and downloading data. Two also indicated that the large scale data sharing helped efforts by improving secondary screening and identifying promiscuous compounds.

Another indication of the utility of PubChem is the number of citations in PubMed. By June 2008, 40 publications in PubMed mentioned PubChem in their abstracts, suggesting the research used PubChem data. Of these, only seven were from the MLP, including four from PubChem staff. Another indication is that the MLP Working Group reported that PubChem training sessions have been well attended whether held at NCBI or at ML steering committee meetings.

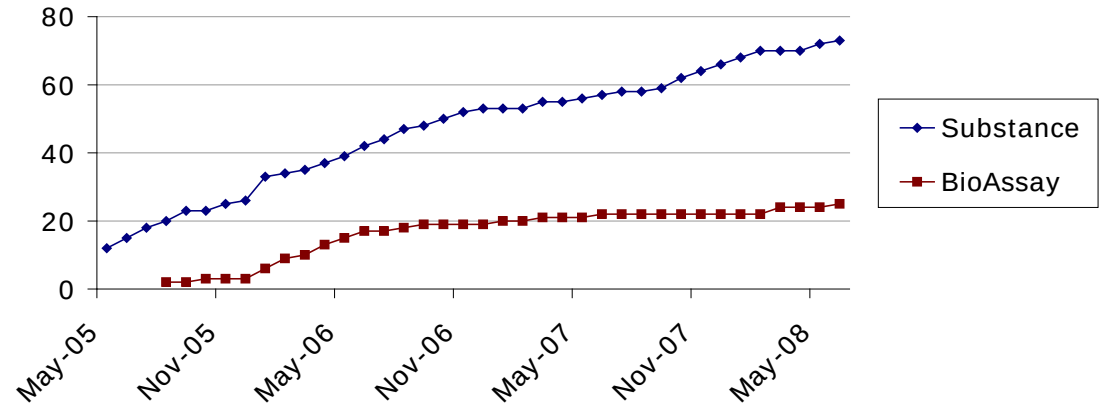
Perhaps the most direct indications of the utility of PubChem are the numbers of depositions and hits on the website. The number of assay depositions grew from 380 to 700 in the 18 months prior to June 2008 (Figure 24).

Figure 24. Growth in PubChem bioassays.



The number of different institutions depositing data on substances increased from 15 in June 2005 to 57 in June 2007 to over 70 in June 2008 (Figure 25). Most of these are non-MLP institutions since the MLSCN centers numbered only ten. The number of different institutions depositing bioassay data grew from 17 in June 2006 to 22 in June 2008.

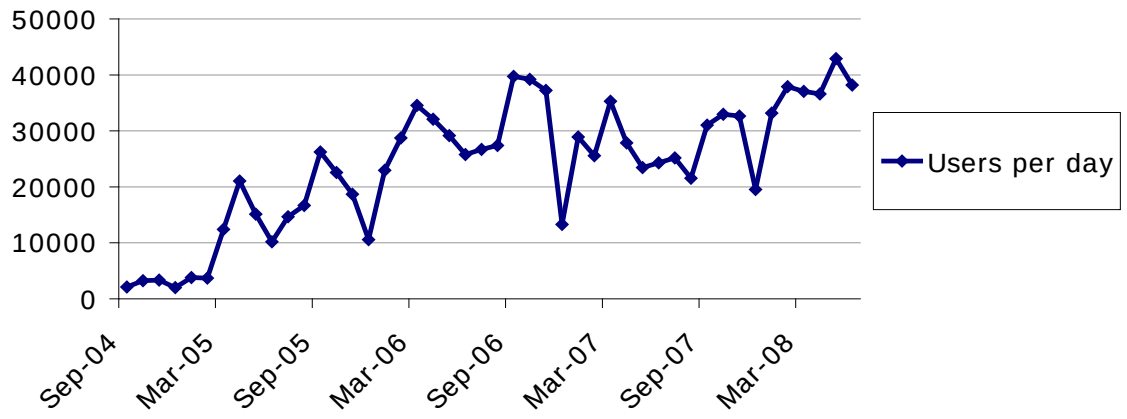
Figure 25. Growth in PubChem contributing organizations.



Finally, the number of PubChem users per weekday has increased from approximately 10,000 in June 2005 to 30,000 in June 2006 to 44,000 in June 2008 (Figure 26).

When PubChem users were asked, “Are you familiar with the Molecular Libraries Screening Center Network/Molecular Libraries Probe Production Centers Network (MLSCN/MLPCN)”, 26 out of the 35 respondents said they were not.

Figure 26. Growth in PubChem users per day.



Development of MICAD infrastructure

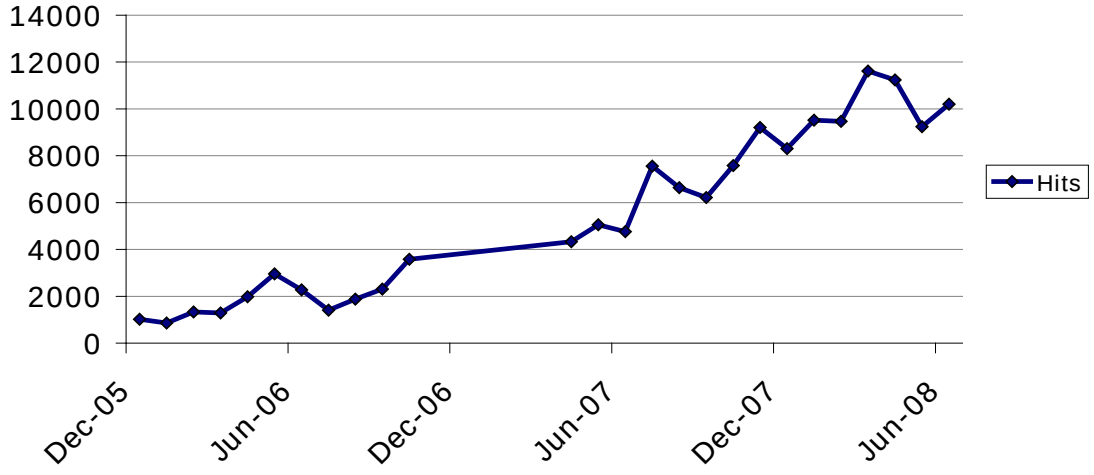
In September 2004, the first curator was hired for the MLP imaging agent database, MICAD. After a year devoted to developing the format and infrastructure, the database was launched in September 2005. Entries, which are mini-review chapters on individual imaging agents, are prepared by four curators and two part-time editors.

Improvements and satisfaction with MICAD

An indication of the value of MICAD is the number of chapter depositions. The 2007 MLP Formal Phase and Transition Proposal stated that one metric of success for MICAD is to add 200 curator-prepared entries each year beginning in 2008. By June 2006, MICAD contained approximately 136 curator-prepared entries with reference links to MEDLINE and related NCBI resources, including PubChem. In July 2008, it contained 500 entries, suggesting that an average of 182 entries per year were made between 2006 and 2008. All of the depositions are from non-ML investigators or MICAD staff.

Another stated metric of success is a measurable increase in usage of the database, as determined by an increase in “hits” to the website. The number of hits has been steadily increasing, with approximately 2,000 in June 2006, 4,000 in June 2007 and 10,000 in June 2008 (Figure 27).

Figure 27. MICAD usage measured by “hits” per month.



Outreach

From the inception of the MLP in 2004, it has provided new and unique capabilities to the academic researcher. Since the new technologies it provides to academic researchers are traditionally found in industry, outreach is important for informing them about these new opportunities. Outreach has been conducted by both NIH staff and center staff to advertise the MLP to the community at large and generate interest in its efforts. This outreach has consisted mostly of presentations but additional types of outreach have been conducted.

NIH staff outreach

Outreach has consisted of thirteen different members of the MLP Working Group giving a total of 66 presentations through June 30, 2008. Several were at international meetings and eight of these were abroad. MLP Working Group members also organized two mini-symposia and a panel discussion at a large meeting. The MLP website generated over 13,000 hits in the 12 months prior to June 2008.

Center outreach

The PIs of the MLSCN centers were asked, “How effective has your center been at outreach to the user community (e.g. number of hits on center websites; number of center staff presentations)?” The five PIs that responded indicated that for each center anywhere from one to 10 presentations were given per month.

In addition, the issue of outreach was brought up as a concern by the PIs of the MLSCN and MLPCN centers. Two of MLSCN PIs recommended additional outreach activities to educate participants or potential participants on what HTS and probes are and developing HTS compatible assays.

All four of the MLPCN PIs recommended additional outreach activities to educate participants or potential participants on the network including: information on submitting compounds, HTS development, the significance of hits, and patenting issues.

Respondents to the RFI also mentioned the issue of outreach. In the open-ended responses, three of those that responded to the RFI indicated that there was a need for educational outreach for scientists with suggestions of an “assay development short course or boot camp” and a meeting between network and non-network scientists.

Sustainability

The MLP has been approved for ten years of support from the Common Fund/Roadmap through 2013. The 2007 MLP Formal Phase and Transition Proposal described a number of reasonable approaches to obtaining long-term support for the different components of the MLP when they transition out of Roadmap funding. It recommended that the centers and the compound collection should receive multi-IC support due to the widespread use and expense of these components. The other components including PubChem and MICAD were suggested for individual IC support. While some changes have been made in response to the Mid-Course Review, the specifics of the transition of the MLP from Roadmap support to IC support have yet to be defined.

Changes made to the program

At the time of the writing of this report (January 2010), several changes had been made to improve the ML program based on feedback from the Mid-Course Review and from lessons learned in the pilot phase of the program. In the production phase, the MLP is focusing on more difficult problems and providing novel resources for research investigators. It is encouraging the development and screening of phenotypic assays for which the target activity may not be known and has increased the chemistry capacity of the network, allowing academic researchers to develop more useful probes from initial hits.

To increase synergy within the network, changes were made in several initiatives. The cheminformatics efforts were refocused to more closely complement the efforts of the centers and a workshop was held to address the need for bioinformatics tools to mine chemical biology data in PubChem. The chemical diversity of the compound library continues to be enhanced in the SMR through *de novo* library synthesis initiatives and by acquisition of previously characterized, diverse libraries. The centers are now required to deposit to the SMR probes they generate as well as a cluster of analogs around each probe compound to be part of the screening collection. The MLP created an NIH Clinical Collection set of 450 chemicals that have a history of use in human clinical trials. The SMR supplies the set at cost to any researcher. The SMR has also implemented some process improvements to improve acceptance time for non-commercial compounds.

Several changes have been made in the FOAs. The FOA for the cheminformatics research centers was not re-issued and funding for the exploratory cheminformatics centers ended in 2008. The ADME/Tox initiative completed its objectives within the ML program and transitioned to investigator-initiated research supported by multiple ICs in 2008. In addition,

two of the imaging initiatives transitioned to support by multiple ICs at the end of FY 2008. The reissued instrumentation FOA was written with the expectation that instrumentation developed within the ML program would be tested in the centers when feasible.

To facilitate probe production, the number of screening centers was reduced while the size of centers with proven outstanding probe development capacity was increased. Three types of screening centers were funded: comprehensive centers that would aid assay providers in developing assays, implement the assays at ultra high-throughput levels, perform informatics and follow up hits with medicinal chemistry; specialized screening centers that would run high-value but difficult (usually lower throughput) assays; specialized chemistry centers focused on chemistry to follow up on hits from the specialized screening centers and also add to the overall chemistry capacity of the network. Two critical needs of the community were addressed by these changes: increased capacity to perform both phenotypic assays and chemistry, which is often rate-limiting in probe development. Furthermore, these changes are intended to enhance synergy among the network centers. At the same time, to ensure the MLPCN centers remain at the forefront of the chemical genomics field, the FOA for the probe production centers (RM08-005) included a request for aims that would significantly advance the field of chemical genomics.

The MLP has expanded opportunities for access to the screening facilities of the centers. In October 2007, the Fast Track opportunity was extended to any NIH grantee with a high-throughput assay. To extend the expertise of the screening centers to potential assay providers, in September 2008, the MLP issued a notice describing how potential assay providers can obtain technical assistance from the MLPCN centers to facilitate submission of HTS-ready assay applications to the MLP.

Management changes

Several management changes have been made to better facilitate participation and improve efficiency of the program. These changes in network management include implementation of: a Chemical Probe Development Plan (CPDP) when HTS-ready or Fast Track assays are assigned to a center for implementation; development of an internet-based Common Assay Reporting System (CARS) to track progress of assay implementation, HTS, chemical probe development, data deposition to PubChem, and probe reports within individual centers and across the network; the Chemical Coordination Committee to coordinate and fully utilize chemistry resources across the network to increase the production of probes. The IP and Data Sharing Policy was updated to allow for a six-month embargo on probe reports to allow for publications by the centers, assay providers, and chemists.

To coordinate the changes during the production phase, the MLP established a Coordination Working Group to: identify best practices for network integration of center operations and communication; assist in the resolution of network problems; provide tracking of activities of the MLPCN through the Common Assay Reporting System (CARS); maintain CARS and a MLPCN website that will provide up-to-date information on the progress of the network towards meeting its milestones and goals, including a list of new network probes and notable data depositions to PubChem; provide monthly updates to the MLP Working Group on network progress; and oversee assignment of assays from the research community to the MLPCN centers.

In addition, the MLP Working Group established a new process for review and approval of probes by NIH staff. To standardize reporting of the biological and chemical properties of probes generated by the MLPCN, a new probe report form was developed. In addition, the MLP hired a new program director to oversee the integrity of both probe reports and the accepted probes.

Lessons from other large programs

There are several examples of innovative processes in similar programs that the MLPCN could adopt to increase its value. The Roadmap Technology Centers for Networks and Pathways program has R01 spoke grants that leverage the expertise of the centers. The MLPCN could have similar R01s to leverage its expertise in screening and chemistry. The Human Microbiome Program has a data coordination center that does what its name implies. Considering the large amounts of diverse data within MLSCN, it may be very beneficial to create a similar center. The National Cancer Institute-supported Mouse Models of Human Cancer Consortium makes mice freely available to PIs. The MLPCN could make its probes freely available or if budgetary realities make this infeasible, for a fee; commercial vendors are already making a few of the probes available to investigators. Some programs encourage their centers to form independent collaborations, and the MLPCN has started to encourage the chemistry centers to collaborate with investigators who have validated hits from HTS assays screened outside the MLSCN in order to make medicinal chemistry expertise more widely accessible. The MLP is beginning to bring chemistry projects in from investigators who already have validated hits in assays performed outside of the MLPCN.

Service models for the future

There are several possible models for MLP to consider in the future transition from Roadmap funding.

A fee-for-service model might be more cost-effective while providing a greater incentive for the screening centers to work as a network. This strategy could be combined with an effort to make single use plates of the

library accessible to qualified screening centers outside the present network. The center would become “certified” to receive plates after agreeing to deposit all data in PubChem, and the certification could last for one or two years.

Another possibility is the model used by the NIH large scale sequencing centers which form a network of highly specialized expertise in DNA sequencing. In the NIH large scale sequencing centers, the funding of sequencing capacity is done separately from the selection of new sequencing projects. The sequencing centers compete for funding based on technical merit and then their projects are decided upon by NIH staff, following advice from working groups composed of members of the research community. The future MLPCN could operate in the same way: centers could compete for funding based on technical merit and then the assay targets could be decided by NIH staff, based on input from the scientific community.

To get an outside perspective from those with relevant expertise, PIs of non-network screening centers were asked, “In 2014, the MLPCN will transition from the ten-year start-up funding providing by the NIH Roadmap. What are the business models for other large service or resource programs that would be a good fit for sustaining the MLPCN after Roadmap funding ends?” The PIs gave in depth responses with several suggestions for the future but did indicate that there would be difficulty in making the transition and finding a sustainable model.

In terms of funding, the PIs were split on how to fund the centers and network. Seven PIs suggested using grants and donations from partnerships or collaborations to fund the centers, while six others suggested using fees for services or access (e.g., to PubChem) or licenses to developed probes to fund the centers and network.

Six of the PIs suggested that fees or justification for grants could come from the products of the network if they were of a high quality, including probes with commercial or high health impact potential, standardized data, and a fully refined library of compounds.

In terms of the structure of the network, three of the PIs suggested moving to a “staff-assisted” model, where scientists wishing to conduct a screen are provided facilities and consulting by the network but conduct the screens themselves. One respondent mentioned that, while this model features less automation, it offers more flexibility and can accommodate many screening projects simultaneously.

Conclusions

In the first phase of the MLP, the MLSCN showed dramatic growth and began to produce outputs in the form of probes and raw data. Generally, the program was found to be on pace with most of its major goals. While the program received high marks from the majority of those surveyed, there were some areas that could be improved that may further increase the effectiveness of the program.

MLP accomplishments

- Besides probes, the raw data from screens can be considered an output of the program. The centers largely met their goals when the emphasis was placed on performing larger screens less frequently.
- Ninety-two new NIH investigators were funded through the Assay Development for HTS and HTS-ready FOAs.
- Of the 76 Assay Development for HTS projects that had completed assays, a total of 37 (49%) met criteria for entry into the MLSCN by June 30, 2008.
- Sixty-one HTS-ready assays were completed as of June 2008. Forty-one of those 61 completed assays resulted in a total of 68 new probes.
- Of the 68 probes produced, 25 were for novel targets or activities and 32 of the remaining had specific improved characteristics.
- By 2007, the cost per probe had been reduced to $2.0 million.
- There have been over 650 requests for MLSCN probes.
- Since August 2004, the SMR has steadily increased its number of compounds to greater than 300,000.
- PubChem has continued to grow, containing data on 18 million compounds in June, 2008.
- MICAD has steadily grown throughout the pilot phase in the number of both users and mini-reviews.

Stakeholder views

- In the surveys, the MLSCN compared well with other centers. The majority of those asked to compare it with other centers said it was better in some way. Specifically, the MLSCN performed its functions better with a larger library, better HTS capabilities, and better assay development. While the MLSCN was cited for being slower than other centers, this may be an unavoidable reality due to the size of the network and library.
- The compound collection and the screening capabilities of the centers were identified as the two most important features of the network to maintain or enhance. The medicinal chemistry capabilities were also frequently mentioned.

- The requirement that all incoming assays be HTS-ready or nearly ready is an important part of the high volume screening capacity of the network. The majority of those surveyed indicated that there needs to be more support for developing HTS assays to fulfill this requirement including more consistent communication with the centers during development of the assays and increased financial support.
- Reagent production and secondary screening is reported as important for following through with screens and producing probes. Lack of funding for reagent production and secondary screening was a sizeable concern among those surveyed.
- The assay providers indicated that the review process for assay applications could be simplified and enhanced.
- The compound providers indicated that the process for submitting compounds to the SMR could be simplified.
- Those scientists that were not participating in the network identified the assay application and review process as the most significant barrier to using the network.
- Those with experience with the network most commonly indicated a need for more chemistry support for development of probes.
- Assay providers, compound providers and MLSCN PIs suggested that the IP policy be altered to allow patent filing.
- Quality assurance of the compounds in the SMR is an important issue. Some of the PIs of the centers suggested measures may be needed to improve identity and purity of the compounds within the library.
- There were numerous suggestions among PubChem users for ways to improve the functionality and user-friendliness of PubChem.
- Outreach is important for increasing participation and interest in the MLP. Some of the users of the MLP that replied to the RFI suggested increasing outreach through training sessions and meetings specifically for MLP-related research, with network and non-network scientists attending.
- Future models for the MLP are an important issue for sustaining the efforts of the program. The PIs of non-MLP centers had varied suggestions for a future model of the network that included funding between grants/donations and fees and moving to a more flexible “staff-assisted” approach.

Suggestions for future data collection

- To simplify and accelerate data analysis, it might be helpful to reduce the number of open ended questions and replace these with closed response questions.
- To help make the surveys more effective, it might be helpful to pretest/pilot test instruments before these are finalized.
- To aid data collection from the ML Working Group, consider incorporating data gathering into the routine operations of the MLP.