Visualizing NNLM

By: Nicole Strayhorn, Associate Fellow 2017-2018

Date: 1/29/18

Project Sponsor: Amanda J. Wilson, National Network Coordinating Office

Abstract

Objective

The goal of this project was to visualize data collected by the National Network of Libraries of Medicine (NNLM) to tell the story of NNLM's outreach activities through visual stories.

Methods

Data was extracted, cleaned, and normalized from NNLM's Training Office and Evaluation Office. Multiple prototypes and dashboards were designed to display the various datasets.

Results

Using Tableau, a heat map analyzing training course needs was created. Also, several dashboards analyzing Network performance measures were designed. A widget was redesigned for the NNLM.gov website incorporating core Network statistics and icons to provide website visitors with a snapshot of NNLM and its activities.

Conclusions

By using graphical representations of data, NNLM can see copious amounts of data in clear, cohesive ways, uncover new insights, and enable them to communicate those insights to others. NNLM also needs to improve their data quality to have better analysis.

Introduction

The National Network of Libraries of Medicine (NNLM) collects extensive data on its members, their activities, and the impact of members' work. This data is compiled and maintained by the NNLM Evaluation Office (NEO) as well as in the NLM Outreach Activities Portal and the NLM External Exhibit System. With the redesign of the NNLM.gov website, the National Network Coordinating Office (NNCO) is exploring ways to leverage NNLM's data for the public through data visualizations and infographics that tell the story of the Network's impact on dissemination of high quality health information to improve the public's health.

Large spreadsheets of data can be difficult to interpret and often prevent organizations from extracting valuable information. Data visualization and infographics offer a better understanding of data and allow discovery of meaningful insights. By building effective visualizations that stand out, NNLM will be able to communicate data-driven messages to its audience with ease and make informed decisions. Therefore, multiple visualizations were developed utilizing NNLM data.

Objective 1: Visualizing Training Course

NNLM develops and teaches training in-person and online courses to diverse audiences nationwide. Visualization will enable the NNLM Training Office to analyze and identify priorities for a new NNLM national training curriculum. NNCO uses it for identifying gaps from the NLM perspective.

Objective 2: Visualizing Quick Facts

NNLM wanted to reinforce a few facts on the NNLM.gov About page with a new vibrant look and feel. These facts include the number of members, outreach activities, and courses offered. The facts are intended to provide a look at NNLM at-a-glance. The original design was text-heavy. NNLM wanted to appeal to a new audience and potential members; to reach their audience in a more meaningful way.

Objective 3: Visualizing Performance Measures

In August 2017, NNLM developed and adopted a list of performance measures to explain their impact in promoting access to biomedical information to improve the public's health. The Network Performance Measures include:

1. Continually build an inclusive, diverse Network through membership management
2. Engage Network members in carrying out the mission of NNLM

3. Employ multiple methods and bidirectional communications channels to inform regional activities
4. Maintain a robust outreach and education program reaching the region's communities and responsive to their needs
5. Contribute to the development and implementation of national and multiregional initiatives - originating from national or local levels - that align with NLM's scope
6. Where possible, appropriate, and reasonable, leverage non-Network partners in the accomplishment of #1-5 above
7. Regularly assess program and project performance in accomplishing aims 1-6 above

This allows transparency, giving NNLM direction and allowing the Network to make informed decisions, improve performance, and increase credibility. Data collected through membership entries, project applications, outreach activities, and evaluations will be used as evidence for each measure. Visualizing this data will allow NNLM to communicate these measures in a visually appealing way.

The final intended deliverables for the project included (1) heat map visualization, (2) infographic for website, and (3) various dashboards of performance metrics. As a result of the project, a new deliverable - (4) analysis of NNLM data quality, was a necessary step in the project.

Methodology

Objective 1: Visualizing Training Course

For the heat map, data was derived from the NNLM Training Office as an Excel spreadsheet in crosstab format. This office compiled a national listing of training topics, levels (e.g. beginner, intermediate, advanced), and intended audience (e.g. librarians, nurses, consumers) they offer annually. The dataset also included each Regional Medical Library's (RMLs) response about what NNLM should be teaching and to whom. Once the data was extracted, the data was cleaned and normalized for analysis and visualizing. After the data was prepared, Tableau was used to visualize the data. Tableau is a business intelligence software tool that supports interactive and visual analysis of data. Throughout the visualization process, I had a few consultations with NNCO and the Training office to present the heat map and receive feedback. Some suggestions included adding specific filters and resizing the format to ensure it would fit on NNLM.gov.

Objective 2: Visualizing Quick Facts

For the Quick Facts, a prototype of a new widget was designed using MS PowerPoint. Freely available icons with no copyright restrictions were collected from the web to be included in the new design. Five different mockups were developed. Throughout the design process, feedback was received from NNCO and the NNLM Web Services Office (NWSO) on the widget design. NWSO made the final design edits and added the new widget design to NNLM.gov mid-January 2018.

Objective 3: Visualizing Performance Measures

Data was extracted from both NNLM.gov and the NNLM Staff Intranet that is compiled and maintained by NEO. The NEO collect data on members, activities, training courses, training evaluation, outreach projects, professional development, exhibits, participant information, and technology. After extracting the data, it is cleaned and normalized for analysis and visualizing. Tableau was used to visualize the data and create multiple dashboards. Throughout this process, I joined meetings, had several consultations with NNCO to gain a better understanding of the data collected for performance metrics and receive feedback about the dashboards that were created. Some of the meetings included:
- Attended NNCO's weekly staff meetings to discuss my work and stay abreast of issues.
- Attended meeting with NEO and Amanda Wilson to discuss performance metrics.

- Attended NNCO meeting to discuss draft of performance metrics document and provided input.
- Attended NNLM Membership Working Group meeting to discuss Membership Profile dashboards I created and get feedback.

Our meetings opened new conversations about data quality which was affecting my analysis.

Results

Deliverable 1: Visualizing Training Course

To visualize training course needs, a heat map (Figure 1) was created to show variance across multiple variables including training courses, audience type, and the count of regional medical libraries that identified a courses topic (e.g. data management) for a specific audience (e.g. academic libraries) as a priority.

Figure 1: NNLM Identified Training Course Needs Year 3 Heat Map
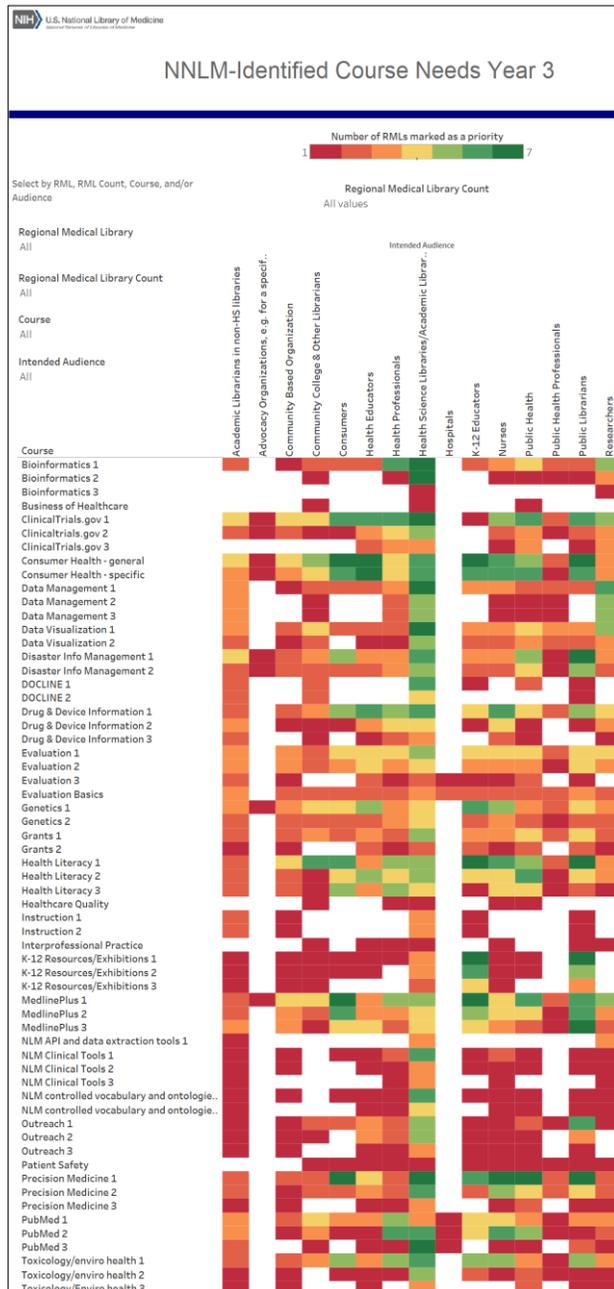
Deliverable 2: Visualizing Quick Facts

The NNLM at a Glance widget is aimed at providing quick facts about the Network to casual observers, NNLM staff needing to update presentations, and funders. The original widget design was text heavy in tiny font. NNCO asked me to review and provide input on how to make the widget more visually appealing and engaging. Instead of bombarding users with text, it was decided to let them explore the widget using alluring graphics, and discover the content on their own.

Figure 2: Original Widget Design

Figure 3: One of the Widget Prototypes

Figure 4: Final Widget Design Accepted and Implemented

NNLM at a Glance

QUICK FACTS
NNLM has 6787 members

Since 5/1/16 NNLM has:

- Provided 1644 educational activities
- Engaged 16850 participants in training
- Held exhibits at 205 events:
  - Shared information with 15,277 visitors
  - Provided 3,713 demonstrations of NLM products

TRAINING BY RESOURCE

Last Updated 10/2/17

NNLM at a Glance
6,787
Members

December 2017

8 Regions

5 Offices

XX,XXX
Courses Taught

X,XXX
Local Community Projects

More information

NNLM at a Glance
(MAY 1, 2016 - PRESENT)

6,898
MEMBERS

70,961
TRAINING PARTICIPANTS

2,140
EDUCATIONAL ACTIVITIES

4,467
NLM RESOURCE DEMOS

updated Jan 9, 2018

More Information

Deliverable 3: Visualizing Performance Measures

In August 2017, NNLM adopted new Network Performance Measures to better tell the story of activities and partnerships. NNCO wanted to develop dashboards that visualize NNLM data according to these measures for transparency and accountability. One of the performance measures is to continuously build an inclusive, diverse Network through membership management. A dashboard was created to look at NNLM membership. Membership was analyzed spatially, by organization type, and numerically by funding cycle.

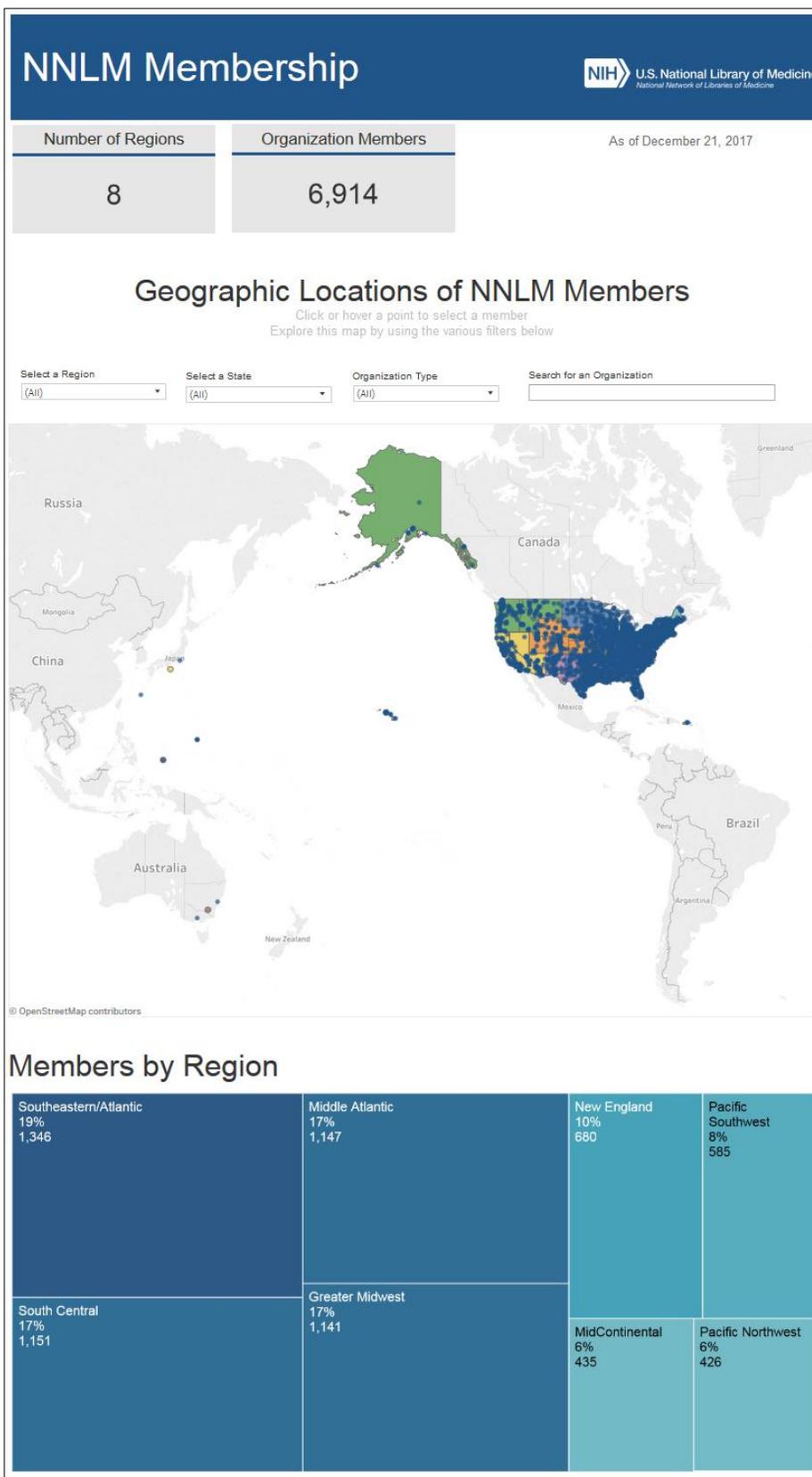*Figure 5: NNLM Membership Profile Dashboard Part 1*

Members by Organization Type

No Type 28% 1,958

Public Library 18% 1,225

Health Sciences Library 11% 777

Other 5% 320

Hospital 20% 1,352

Academic Institution 16% 1,097

Community Based 5% 314

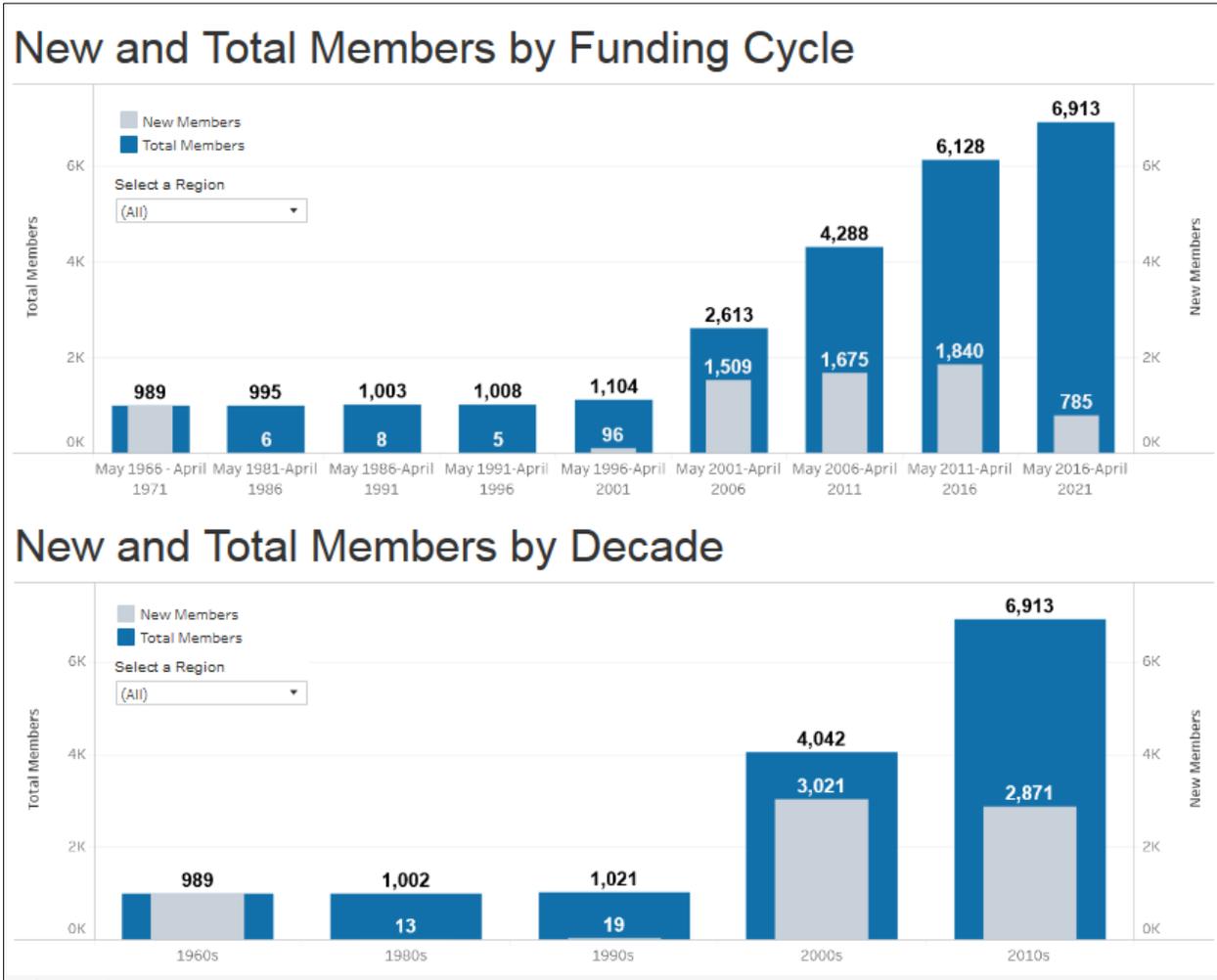Public Health

Community College 4% 272

K-12 2%

Government Agency 4% 243

Association 2%

Regional Members by Organization Type

Organization Type	Greater Midwest	MidContinental	Middle Atlantic	New England	Pacific Northwest	Pacific Southwest	South Central	Southeastern/Atlantic
No Type	1	364		1	1	308	316	967
Academic Institution	267	13	215	195	94	75	128	110
Association	37	1	51	14	8	1	10	24
Clinic/Other Health Care	117		100	32	35	3	86	12
Commercial	17		10	15	10	3	8	4
Community Based	25	2	53	38	84	6	87	19
Community College	83	7	32	34	43	8	43	22
Faith-based	146		26	2	9	1	37	11
Government Agency	38	6	13	19	51	15	58	43
Health Sciences Library	375	10	55	59	35	157	60	26
Hospital	448	12	300	130	91	107	206	58
K-12	6	5	61	28	10	3	30	4
Other	48	2	152	26	25	6	33	28
Public Health Organizati..	84	4	54	12	21	5	31	17
Public Library	205	26	346	247	44	58	232	67

Figure 7: NNLM Membership Profile Dashboard Part 3

New and Total Members by Funding Cycle

New Members
Total Members

Select a Region
(All)

Funding Cycle	Total Members	New Members
May 1966 - April 1971	989	
May 1981-April 1986	995	6
May 1986-April 1991	1,003	8
May 1991-April 1996	1,008	5
May 1996-April 2001	1,104	96
May 2001-April 2006	2,613	1,509
May 2006-April 2011	4,288	1,675
May 2011-April 2016	6,128	1,840
May 2016-April 2021	6,913	785

New and Total Members by Decade

New Members
Total Members

Select a Region
(All)

Decade	Total Members	New Members
1960s	989	
1980s	1,002	13
1990s	1,021	19
2000s	4,042	3,021
2010s	6,913	2,871

Discussion

Deliverable 1: Visualizing Training Course

In January 2018, NNCO used the heat map to review RML input compared to NLM priorities. They could identify areas of interest and importance for NLM that were not reflected in the RML priorities. NNCO's input into the NNLM National Training Curriculum fills the gaps identified by the heat map visualization.

Deliverable 2: Visualizing Quick Facts

In January 2018, NNCO, NEO and NWSO - inspired by recommendations and prototype - developed and refined a final widget (Figure 4). This widget went live on NNLM.gov week of January 16th. NNCO is currently soliciting additional "core metrics" to add to the widget. In the future, "More Information" will link to NNLM dashboards.

Deliverable 3: Visualizing Performance Measures

NNLM's new Membership Working Group uses these dashboards to inform objectives and goals for the membership directory and new tools needed for membership management. These visualizations are also going to be used in NNLM grant progress reports as part of required templates issued by NNCO. The dashboards also highlighted the need for changes in membership directory permissions to improve data quality.

Deliverable 4: Data Review

While the visualizations revealed a wealth of information, the data cleaning process did shed light on data quality issues. There are multiple inconsistencies in NNLM's data. One possible root cause is weak data stewardship. The volume of data owned by NNLM is acquired and aggregated from multiple channels including employees, data migrations, system errors, and multiple users. Unfortunately, this opens data for various issues with quality such as accuracy, consistency, completeness, and validity. For example, multiple non-NNLM employees produce NNLM's member directory, however different people sometimes incorrectly populate a field without even realizing it. NNLM has not been responsive to errors or changes in membership information causing issues for analysis. Furthermore, a review of the data (Table 1) was conducted to identify key data issues to fix and improve for future analysis.

Table 1: NNLM Membership Data Review

Data	Type of Issue	Reasoning
Member Enrollment Dates	Accuracy	Past data migration caused some inaccuracies that need to be verified

Member Addresses	Consistency Accuracy Validity Completeness	About 14 percent of the members were not geographically located due to various address issues including: ● PO Box used as street addresses, cannot be geocoded ● Armed forces did not have complete addresses and utilize international addresses that do not follow the US address structure ● Missing information including numbers, streets, cities, states, and zip codes ● Misspellings in addresses
Duplicate Organizations	Consistency	Some organizations have multiple records in the membership directory, inflating numbers
No region assigned	Completeness	3 international organizations
No Organization Type	Completeness	28 percent of the members need to be assigned an organization type for better analysis

Recommendations

1. NNLM needs to improve data quality and data accuracy.
   a. Review and clean data on a scheduled basis
   b. Utilize a controlled vocabulary
      i. An example would be to use same region names for every spreadsheet; Currently they vary in every spreadsheet.
   c. Create, maintain, and use a data dictionary
   d. Be vigilant about missing data
2. NNLM should consider hiring one full time Data Manager or Data Analyst for NNCO or NEO.
   a. Data Integrity and Management 35% of time
      i. To manage and maintain data systems
      ii. Perform validation checks of data to identify errors and omissions
      iii. To automate data workflows to reduce staff burden in activities that support manual processes
      iv. Work with every RML to ensure one staff member in that region is cleaning records in the Membership Directory
      v. Develop scripts for automating data cleanup by using R or python

b. Data Analytics 35% of time
    i. Provide analysis and visualizations when necessary for internal and external stakeholders
    ii. Communicate relevant data insights through reports, presentations, and visual aids
    iii. Be familiar with tools such as Open Refine or MS Excel for data cleaning and Tableau or comparable data visualization tool for analysis and visualization
c. Collaboration 15% of time
    i. Participates in staff meetings involving NNCO, NEO, NTO, and NWSO
    ii. Make recommendations on performance measures
    iii. Communicate with NEO and NWSO about data changes or requirements
    iv. Make recommendations for software, hardware, and data storage upgrades
    v. Work with Data Coordinators on NNLM RD3: Resources for Data-Driven Discovery
d. Training 10% of time
    i. Design and deliver workshops and trainings on visualization tools and techniques for NNCO and the RMLS
    ii. Deliver training on
e. Documentation 5% of time
    i. Create and maintain documentation on technology, tools, and best practices for data management

Acknowledgements