Preview of Award 1305190 - Annual Project Report

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1305190
Project Title:	The Neural Engineering Data Consortium: Building Community Resources to Advance Research
PD/PI Name:	Iyad Obeid, Principal Investigator Joseph Picone, Co-Principal Investigator
Recipient Organization:	Temple University
Project/Grant Period:	08/01/2013 - 07/31/2015
Reporting Period:	08/01/2013 - 07/31/2014
Submitting Official (if other than PD\PI):	Joseph Picone Co-Principal Investigator
Submission Date:	07/28/2014
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	Joseph Picone

Accomplishments

*** What are the major goals of the project?**

The major goal of this planning grant is to assess the need within the bioengineering community for an organization devoted to the development of big data resources. Major goals in support of this overarching goal were:

- promote the concept through presentations at major conferences and online surveys
- engage the community to assess needs
- create an advisory board to guide the development of the organization
- develop an organizational plan and estimate of the resources required
- host outreach activities to promote the need and goals of the organization

The original proposal was structured into three phases:

- create an online presence for the organization and promote it within the community
- create an advisory board (Board of Directors) and a preliminary organizational structure
- dissemination of information and collection of feedback

*** What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?**

Major Activities: The first step in creating an online presence for NEDC was to create a web site from which information could be disseminated. The NEDC web site is located at

NEDC and has been operational since the beginning of the project. It was implemented using Drupal which is one of the more popular content management systems in use today. It uses the CorporateClean theme and is structured so that featured events appear in the slideshow at the top of the page. Menus along the top provide access to major categories such as “About Us” and a description of the advisory board. Under the Data tab users can find information about data available for download. We also created a Facebook page (Neural Engineering Data Consortium) and Twitter account (nedcdata). Over the last four months we are averaging about 200 unique visitors a month to the NEDC site and 500 pages viewed. We expect this to increase with the release of the TUH EEG Corpus.

The second step in creating an online presence was to begin conducting a market survey. We created a simple short survey on SurveyMonkey that collects feedback on the types of resources needed and willingness to join a consortium. The survey consists of three required questions and five optional questions. Though many people have promised to take the survey, we have only three responses thus far. We need to increase participation and will be focusing on that in the coming months as we organize a constituents meeting.

A third significant step was taken by organizing a symposium at the first IEEE GlobalSIP Conference held in Austin, Texas in December 2013. Our symposium was titled Advancing Neural Engineering Through Big Data. The symposium consisted of four keynote talks followed by an afternoon poster session. The keynote talks were designed to sample the continuum of potential constituents for data resources. Iyad Obeid began the session with a description of the context and promoting the need for an organization like NEDC. Chris Cieri, Executive Director of the Linguistic Data Consortium (LDC), followed with an overview of their 20-year history. Jack Judy, a former DARPA Program Manager in the bioengineering area, provided a government perspective on the state of research and how data needs to drive this research. Karen Moxon, Associate Director for Research at Drexel University, provided a perspective on how challenging it can be to collect neural engineering data. Over 300 researchers attended the conference, while about 30 of these attended our plenary session. Thirteen papers were presented at the afternoon poster session, which provided an opportunity for the PIs to promote NEDC and its resources. There was considerable interest in the TUH EEG Corpus.

We also began organizing the 2014 IEEE Signal Processing in Medicine and Biology Symposium at Temple University (www.ieeespmb.org/2014). This symposium will have a big data focus and will include a panel session to discuss data and resource needs. We plan to host this symposium at Temple for the next few years since it will provide an excellent forum to promote NEDC. This is a small, focused symposium that should have an attendance of about 100 professionals and include about 40 technical papers and posters. There will be two keynote talks, two lecture sessions and two poster sessions in addition to the panel. We presented a paper on NEDC and the TUH EEG Corpus at the 2013 meeting.

In addition to these activities, we have reached out to several specific organizations to discuss the role NEDC can play in their research. We have created an active collaboration with the Dermatology Department at Temple University’s School of Medicine in the area of characterization of itch. We hope that we can eventually provide reference data for this emerging area of science and are discussing ways an organization like NEDC could facilitate data collection. We have met with University of Pennsylvania and Jefferson Memorial Hospital to discuss similar collaborations. Jefferson is very prominent in the head trauma area and is

expanding its neuroscience focus. They provide an excellent opportunity to develop data resources focusing on the study of head trauma in athletics.

The second major task for this project was to create an advisory board and a preliminary organizational structure. Our initial advisory board, referred to as the NEDC Board of Directors, can be found at www.nedcdata.org/drupal/node/11. We carefully selected these members to cover four major constituencies. Dr. Michele Masucci, our Interim VP of Research, represents Temple University and has been highly supportive of this activity. Dr. Chris Cieri represents LDC with whom we have a close working relationship. Building on LDC's best practices is an important part of our plan to mitigate risk. Dr. Emily Caporello represents the government from a sponsorship point of view through her position at DARPA. Dr. Gene Civillico represents government research labs through his affiliation with the FDA. The FDA is a significant piece of the puzzle in bioengineering because they not only influence research directions but also influence how technology is evaluated and certified. They can be a major constituent for NEDC. Finally, Maciej Lazarewicz, MD, PhD, represents industry through his position at Medtronic and his long-term commitment to big data activities. Industrial participation was key to LDC getting off the ground.

We held our initial Board of Directors meeting at Temple on May 1. Some materials from the meeting, including an audio recording of the public portion of the meeting, are available here. The agenda included: (1) Board member introductions, (2) and overview of LDC's operations to ensure everyone was familiar with the scope of the proposed activity, (3) an overview of NEDC as it currently exists, (4) a discussion of the TUH EEG Corpus as a case study in how corpora are developed and disseminated, (5) a discussion of the technical challenges that NEDC can address through its resources (where are the most significant perceived opportunities), and (6) how we can engage the community. The meeting was structured as a half-day meeting including lunch so that people could attend the meeting in a single day trip. One participant due to illness participated remotely via a WebEx session.

The Board was enthusiastic about the creation of this entity but asked many hard questions about plans for funding and sustainability. Chris Cieri's overview of LDC involved an extended discussion of the broad range of activities one faces when releasing data because this was first time most of the Board had witnesses such a detailed brief of all the issues one faces today in developing and releasing data. The appropriateness of the LDC model for bioengineering was discussed at length, and it was acknowledged that the bioengineering community might not have the same focused technology drivers that exist in human language technology. An interesting thread was the discussion of distribution methodologies since bioengineering datasets can be very large. A definite preference was expressed for a Dropbox like mechanism for distribution, so that sites could be easily updated. Overall the Board was very satisfied with the first meeting. We plan to schedule a second meeting around May 1, 2015.

Finally, as part of our outreach effort, we developed a half-day laboratory that introduces high school students to big data concepts. The lab began with teaching students how to record EEG data through an Emotiv EPOC headset. Students were then introduced to Python programming. Students were then shown how to modify a simple pattern recognition program in Python and to add their own algorithm. Twenty-one students participated and about half of them were able to complete the lab in a three-hour timeframe. We plan to repeat this lab at some of the local high schools over the second year of this project. The lab materials are available at WE2.

Specific Objectives:

In the first phase of this project, the goal was to establish a vision and presence for NEDC through activities such as development of a web site, market surveys and involvement in professional conferences.

In the second phase of this project, our goal was to establish an advisory board that would plan an instrumental role in the creation of this organization. Active involvement of the advisory board is critical to the future of the organization.

In the third phase of this project, engagement of the community is the critical step. This includes establishing a clearly articulated need through the collection of letters of commitment, and engagement through the discussion of critical resources that NEDC could provide.

Significant Results:

(1) The NEDC web site is operational and being used to disseminate information.

(2) NEDC has provided a preliminary release of the TUH EEG Corpus that contains EEG data and corresponding physician reports so that the community can provide feedback on the data.

(3) We hosted a special session at IEEE GlobalSIP 2013 titled "Advancing Neural Engineering Through Big Data" that focused on the need for an organization like NEDC by bringing together professionals interested i big data resources in bioengineering.

(4) We are in the planning process for hosting the 2014 IEEE Signal Processing in Medicine and Biology Symposium that will also focus on big data resources. The conference will be held on December 13, 2014.

(5) We have formed an advisory board (http://www.nedcdata.org/drupal/node/11) and held our first planning meeting to discuss NEDC.

(6) We posted an online survey to assess data needs and community interest (http://www.surveymonkey.com/s/DZPJTF3).

(7) We have presented papers at several conferences discussing the vision for NEDC.

(8) We have visited several regional organizations to discuss the need for big data resources and investigate sharing of data.

(9) We have written several follow-on research proposals that attempt to leverage NEDC and would provide some funding to allow NEDC to further develop the TUH EEG Corpus as a proof of concept.

(10) We conducted an outrearch activity by developing and delivering a laboratory on big data applications using EEGs to high school students interested in engineering (Temple's WE2 sumer engineering program).

Key outcomes or Other achievements:

Three relevant observations:

(1) It has been difficult to get people to particpate in the online survey and commit. We have been actively promoting this for 9 months and received many commitments to participate, but follow-through has been low. We are adjusting our strategy to achieve more success. Collection of letters of commitment is critical to the future success of this effort.

(2) Though everyone openly acknowledes the value of big data resources,

researchers are still very protective of their annotated data. We are making progress at generating interest in releasing significant resources collected at other sites, but authors of the most prominent publications involving proprietary data have been less enthusiastic about releasing their data. Equally important, it remains to be seen if people will be willing to be dues-paying members of such a consortium. Responses have been very tepid on this issue.

(3) The lack of computing expertise to process big data may ultimately limit interest in some of these resources. To effectively process the TUH EEG data, for example, extremely efficient parrellelizable code is needed. Therefore, we plan to educate our users on some of the potential pitfalls ahead when trying to manipulate our big data resources.

*** What opportunities for training and professional development has the project provided?**

As part of our outreach activities, we developed a half-day laboratory to introduce high school students to EEG data. This lab was developed in a relatively new programming language, Python, and used consumer EEG systems (Emotiv EPOC). It was a nice opportunity for the PIs to learn more about Python and consumer EEG systems.

*** How have the results been disseminated to communities of interest?**

We actively maintain a web site, www.nedcdata.org, through which we regularly disseminate updates about the project. This web site provides status on recent developments at NEDC. It also hosts a preliminary release of the TUH EEG Corpus and will soon host the production release.

*** What do you plan to do during the next reporting period to accomplish the goals?**

The major goals of the remaining portion of this project are:

- Host a constituents workshop in late 2014/early 2015 (in conjunction with IEEE SPMB 2014).
- Provide a production release of the TUH EEG Corpus to demonstrate in a concrete way the value of such an organization.
- More actively engage the community through laboratory visits and acquire letters of commitment.
- Develop a more concrete plan to attract sustained funding.
- Finalize administrative and business plans for the first five years of the organization.

Products

Books

Book Chapters

Conference Papers and Presentations

I. Obeid, A. Harati and J. Picone (2014). *A Big-Data Approach to Automated EEG Labeling*. Proceedings of Neuroinformatics. Leiden, The Netherlands. Status = AWAITING_PUBLICATION; Acknowledgement of Federal Support = Yes

I. Obeid and J. Picone (2013). *Accelerating the Rate of Technology Development in Bioengineering Through a Common Evaluation Paradigm*. Proceedings of the IEEE EMBS Neural Engineering Conference. San Diego, California, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

I. Obeid and J. Picone (2013). *Advancing Neural Engineering Through Big Data*. IEEE Global Conference on Signal and Information Processing. Austin, Texas, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

I. Obeid and J. Picone (2014). *Big Data in Bioengineering*. IEEE SIgnal Processing in Medicine and Biology

Symposium. Philadelphia, Pennsylvania, USA. Status = AWAITING_PUBLICATION; Acknowledgement of Federal Support = Yes

I. Obeid, A. Harati and J. Picone (2014). *EEG Event Detection Using Big Data*. Proceedings of the 48th Annual Asilomar Conference on Signals, Systems, and Computers. Pacific Grove, California, USA. Status = AWAITING_PUBLICATION; Acknowledgement of Federal Support = Yes

C. Ward, I. Obeid, J. Picone and M. Jacobson (2013). *Leveraging Big Data Resources for Automatic Interpretation of EEGs*. Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium. New York City, New York, USA. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

I. Obeid and J. Picone (2013). *The Neural Engineering Data Consortium: Deja Vu All Over Again*. 20th Anniversary Celebration, Linguistic Data Consortium. Philadelphia, Pennsylvania. Status = PUBLISHED; Acknowledgement of Federal Support = Yes

Inventions

Journals

Licenses

Other Products

Educational aids or Curricula.

We developed a half-day laboratory introducing high school students to big data, machine-learning and Python programming. The code is available from our web site. This was part of our outreach activities. The lab was offered on July 8, 2014 to 22 female high school students participating in our Women in Engineering Summer Program. In the second year of the project, we plan to take this on the road and visit local high schools.

In the lab, students learn how to acquire EEG data from an Emotiv EPOC headset. They can view the data on a computer. We then teach them how to do a little Python programming. They end the lab by developing a simple algorithm and inserting it into a pattern recogntion system and measure performance on a brainwave recognition task.

The lab was designed to stimulate interest in big data, machine learning and programming. We also explain the activity is part of a funded effort by NSF and explain the importance of getting involved in research as undergraduates.

Other Publications

Patents

Technologies or Techniques

Thesis/Dissertations

Websites

NEDC Data and Resource Survey
http://www.surveymonkey.com/s/DZPJTF3

A brief survey to assess community needs.

The Neural Engineering Data Consortium
http://www.nedcdata.org

This is the primary portal from which users can access our resources. We currently post updates about the project and the release of our first corpus.

Participants/Organizations

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Obeid, Iyad	PD/PI	1
Picone, Joseph	Co PD/PI	1

Full details of individuals who have worked on the project:

Iyad Obeid
Email: iobeid@temple.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Promoted the organization through participation in several conferences. Created the NEDC web site.

Funding Support: None.

International Collaboration: No
International Travel: No

Joseph Picone
Email: joseph.picone@gmail.com
Most Senior Project Role: Co PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Organized a symposium focused on big data. Provided management and leadership for the overall project.

Funding Support: None.

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?
Nothing to report.

Have other collaborators or contacts been involved? No

Impacts

What is the impact on the development of the principal discipline(s) of the project?

The focus of the planning grant is to study the demand for an organization devoted to data resources. There is no doubt that the resources delivered by such an organization will have a major impact of science and technology. Thus far this project has increased the awareness of the scope and extent of such an organization. Researchers in the bioengineering field, and many of the existing sites delivering resources, are now more aware of the operational challenges including intellectual property issues.

What is the impact on other disciplines?

Evaluation-driven research and/or open competitions is a model that is slowly spreading into many science disciplines. NEDC's immediate focus has been neuroscience data only because of its access to such resources. As more constituents are identified, we expect to broaden the scope of our resources into many related bioengineering disciplines. We hope NEDC can successfully promote the concept of data-driven research on big data into many areas of bioengineering.

What is the impact on the development of human resources?

This project is provided an excellent opportunity for the PIs to gain exposure in a new research community. It is particularly valuable for Dr. Obeid since it has increased his visibility in a leadership role in the bioengneering community.

What is the impact on physical resources that form infrastructure?

Though we did not develop the TUH EEG data on this project, its release through the NEDC web site is a major contribution to the field. We are in negotiations with two other sites for the release of their data and expect that to happen in the second year of the project. These resources are helpful in demonstrating the impact an organization like NEDC can have.

What is the impact on institutional resources that form infrastructure?

The Senior Vice-Provost for Research at Temple University is a member of our Advisory Board and has expressed support of the effort. We are not at a stage where institutional resources are needed, but planning is underway in the event we are successful at attracting significant additional funding.

What is the impact on information resources that form infrastructure?

The main product of this project are by definition information resources (data available from the NEDC web site). These are expected to have a major impact on bioengineering research. Initially our focus has been EEG data to leverage a significant on-going research project at Temple University.

What is the impact on technology transfer?

Not relevant. However, the TUH EEG Corpus, available from the NEDC web site, is being used by the PIs for a commercialization effort that is being jointly funded by the University City Science Center and Temple University. The data is playing a critical role in the development of machine learning technology.

What is the impact on society beyond science and technology?

NEDC will not have such a broad reach. However, the technology it will enable will have a significant impact in the healthcare industry. For example, the TUH EEG Corpus will enable a new generation of technology to be developed that automatically interprets EEGs. Previous attempts at this have failed due to the lack of big data resources. The bioengineering technology enabled by the resources an organization like NEDC can offer will have profound impacts on medicine and health.

Changes/Problems

Changes in approach and reason for change

Nothing to report.

Actual or Anticipated problems or delays and actions or plans to resolve them

We requested an extension of the grant to take advantage of opportunities to increase community engagement by hosting a regional conference and presenting at additional conferences.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.