Executive Summary

Funded as an exploratory high-risk grant in 2012, the HASTAC-EAGER project yielded corresponding high-payoff results on topics surrounding the use of the online academic network HASTAC. At its core, the project was conducted by two separate postdoctoral fellows that were funded under the grant and who managed to day-to-day research, management, and communication activities for the project. The findings, detailed in a journal paper and three strictly peer-reviewed conference papers, indicate the importance of online academic networks such as HASTAC for improving productivity, collaboration, and interdisciplinary across multiple institutions across the nation. Significant features of the project include its data-driven approach using state-of-the-art data science, its leveraging of advanced visualization to convey the findings, and its use of blog posts as intermediate progress reports.
External Evaluation Report – HASTAC-EAGER (NSF award no. 1243622)

Introduction
This report represents the outcome of an external evaluation of an NSF EAGER grant as part of the HASTAC project (http://www.hastac.org/) titled “EAGER: Assessing the Impact of Technology-Aided Participation and Mentoring on Transformative Interdisciplinary Research” (NSF award #1243622), henceforth referred to as HASTAC-EAGER. The objective of the HASTAC-EAGER grant was to study the academic social network HASTAC.org. According to the NSF award database (http://www.nsf.gov/awardsearch/), the project officially started on August 1, 2012 and is estimated to end on July 31, 2015. It was funded to the amount of $294,793.00 over a total of three years. NSF EAGER (Early-concept Grants for Exploratory Research) grants are implemented to support novel concepts of research that have not been well established.¹ For these reasons, EAGER grants tend to be “high-risk, high-payoff” in nature.

Founded in 2002, HASTAC (“haystack”) is the Humanities, Arts, Sciences and Technology Alliance and Collaboratory, and is an open community of close to 13,000 humanists, artists, social scientists, scientists, and technologists working on topics within the future of learning (numbers from February 2015).² Originally based Stanford University, HASTAC is headquartered at Duke University since 2005, and since Summer 2014 has a second “hub” at City University of New York (CUNY). The project and its website provides an online platform for collaboration across disciplines and scholars, and has so far marshaled more than 12,000 posts, more than 6,000 blog entries, and more than 8,000 comments since 2006. The HASTAC project itself is not the topic of the EAGER grant; rather, the grant studies six years of data from the HASTAC database for the purpose of understanding how to use academic collaboration cyberinfrastructures for meeting the interdisciplinary problems of the future, such as the Grand Challenges identified by the NSF.

The mandate of this external evaluation was “reviewing the work of HASTAC’s EAGER grant—including interim reports, blogs, data, papers, and final reports—and preparing a report that provides a fair and unbiased review.” The overall goal was “assessing the project’s strengths and weaknesses and its overall merit to the academic community.” The external reviewer was chosen to be impartial and unrelated to the project team, yet to have appropriate expertise in the specific area involving the EAGER grant.

In the remainder of this report, we first present the method used in the external evaluation. We then discuss our evaluation for the components research, personnel, training, and validation. We close with our reflections as well as the future outlook of the project and its findings.

Method
Here we outline the biography of the external reviewer, the material used in the evaluation, and the approach employed in reviewing this material.

Reviewer Biography
Niklas Elmqvist is an associate professor in the iSchool (College of Information Studies) at University of Maryland, College Park. He received his Ph.D. in computer science in 2006 from Chalmers University in Gothenburg, Sweden, and has conducted research at Georgia Institute of Technology in Atlanta, GA (Spring 2006), INRIA in Paris, France (Spring 2007), and Microsoft Research in Paris, France (2007-2008). Prior to joining University of Maryland, he was an assistant professor of electrical and computer engineering at Purdue University in West Lafayette, IN. His research area is information visualization, human-computer interaction, and visual analytics. He is the recipient of an NSF CAREER award as well as best paper awards from the IEEE Information Visualization conference, the International Journal of Virtual Reality, and the ASME IDETC/CIE 2013 conference. He is an associate editor of the Information Visualization journal as well as co-editor of the Morgan Claypool Synthesis Lectures on Visualization. His

¹ NSF GPG Chapter II D.2.
² http://www.hastac.org/about
research has been funded by both federal agencies such as NSF and DHS as well as by companies such as Google, NVIDIA, and Microsoft. He is the recipient of the Purdue Student Government Graduate Mentoring Award (2014), the Ruth and Joel Spira Outstanding Teacher Award (2012), and the Purdue ECE Chicago Alumni New Faculty award (2010).

Material
The material used for performing this external evaluation included the following documents:

- Original HASTAC-EAGER proposal (including project summary, project description, references, data management plan, and postdoctoral mentoring plan);
- Final NSF project report (from Spring 2015);
- Academic papers appearing in PLOS ONE (2015) and at the Hawaii International Conference on System Sciences 2015; and
- Publicly available reports, articles, blog posts, data, charts, and other documents from the project website http://www.hastac.org/.

All of the material used for the external evaluation were digital artifacts created as part of the HASTAC-EAGER project. In particular, no interviews were conducted as part of this evaluation because this was not expressed as within the mandate of the evaluation. The external reviewer does not currently have the necessary IRB approval to conduct such interviews. However, because all of the material used in the evaluation are publicly available and are not merged or enhanced in such a way that individuals might be identified, this work does not constitute human subjects research and thus does not require IRB approval.

Approach
The overall approach used in this evaluation involved deep and critical reading of the available material. Documents, reports, and findings were also correlated with external resources—such as digital libraries, conference and journal websites, and data repositories—wherever possible. For each of the sections of this report, we add a special evaluation paragraph whenever the external reviewer has specific feedback.

Research
According to the original proposal as well as the final report, the overall research focus of the HASTAC-EAGER project centered on the following topics (edited and summarized for brevity):

1. New methods to facilitate scholarly collaboration:
   a. Voluntarism and volunteer-based cyberinfrastructures for academic discourse;
   b. Networking and knowledge transfer in online academic spaces;
   c. Mentorship in online academic spaces;
   d. Impact of online networking on young academics compared to unconnected ones;
   e. Impact of online academic spaces on success for underrepresented minorities; and
   f. Factors encouraging or discouraging collaboration.

2. Data analysis of HASTAC collection:
   a. Extraction and analysis of HASTAC.org website database;
   b. Social network analysis of member network;
   c. Content analysis (tags, topics, and documents) on the website; and
   d. Text analysis and mining of textual contents.

3. HASTAC.org engagement and participation:
   a. Ongoing progress reports to invite collaboration and participation; and
   b. Cost-benefit balance to support professional development for academics.
Findings
Summarizing across documents, the major findings of the HASTAC-EAGER project were the following:

- Scholars active in an online academic network were highly productive and collaborative;
- Scholars active in an online academic network were more productive than those not active;
- Academic networks support research and learning yet does not yield higher productivity;
- Interdisciplinary work does not lead to higher productivity;
- Digital humanities derive much from their corresponding humanities discipline;
- HASTAC is an interdisciplinary digital humanities network;
- Contributions to the HASTAC community follows a skewed power-law distribution;
- Contributions to the HASTAC community come mostly from North America; and
- User-assigned tags are very densely connected and clustered.

Evaluation: The amount of quantified findings from the HASTAC-EAGER project is impressive. In particular, the partially confirmed finding that networked scholars are more productive than unconnected scholars can serve as a raison d'être for the entire HASTAC project since it indicates that being active in such a network will give an advantage to young scholars.

Activities
The HASTAC-EAGER final project report describes in great detail a timeline of the activities that happened throughout the duration of the project. This timeline provides excellent insight into the specific events that happened in the project.

Journal Publications
The HASTAC-EAGER project has resulted in one journal publication:

- “How Digital are the Digital Humanities? An Analysis of Two Scholarly Platforms” by Cornelius Puschmann and Marco T. Bastos appeared in the open-access peer-reviewed journal PLOS ONE (impact factor 3.534 in 2013) on February 15, 2015.3 It should be noted that PLOS ONE is different from traditional scientific journals in that it does not use the perceived importance of a paper as a criterion for acceptance or rejection. Instead, PLOS ONE only verifies the scientific rigor of experiments and data analysis, and leaves it to the scientific community to ascertain importance after publication through debate and comments.4 The PLOS ONE website provides tools for such annotations, discussion, ratings, and citation tracking. Given that the paper was posted online only on February this year, it is too early to say much about its perceived impact and importance by the scientific community. However, at the time of this report, the paper has more than 4,000 article views, 3 external comments, and has been mentioned 35 times on Twitter. It has not yet been cited. Nevertheless, these metrics indicate potential for future impact.

Evaluation: The journal output from the HASTAC-EAGER project at the end of the funded period represents an appropriate maturation of the ideas and findings generated by the project.

Conference Publications
Several conference papers have resulted from the HASTAC-EAGER project. For readers unfamiliar with these research fields, it should be noted that in certain research communities within social media, communication, and human-computer interaction, conferences are peer-reviewed, highly selective, and considered archival. For this reason, conference papers should be regarded “journal-quality” in such

communities, i.e. having the same stature and archival nature as journal articles in other, more traditional, fields. More specifically, here is a summary of the conference papers:

- “Outcompeting Traditional Peers? Scholarly Social Networks and Academic Output” by Marco T. Bastos was presented at the Hawaii International Conference on System Sciences 2015 (HICSS-48) as part of the track “Digital and Social Media,” in the minitrack “Social Networking and Communities.” In its 48th year, the HICSS conference has grown to become a significant venue for a wide range of disciplines in system, data, and computer science. The conference is organized into 10 concentrations, known as tracks, which in turn are split into a total of 100 minitracks.

- “Bridging Structural Holes: Scholarly Collaboration in Online Social Networks” by Marco T. Bastos was presented at the ASE International Conference on Social Computing in August 2014 in Beijing, China. The HASTAC-EAGER final report cites this as a joint IEEE/ASE conference, but it seems like this paper appeared in ASE SocialCom. IEEE SocialCom is a different conference (IEEE SocialCom 2014 took place in Sydney, Australia in December, for one thing), so this appears to be a minor error. IEEE SocialCom is also the more prestigious conference, and is strictly peer-reviewed and highly selective. However, ASE SocialCom appears to also be a highly respected conference with a very impressive 16.6% acceptance rate. Publications are archived by the ACM.

- “Scholarly Collaboration in the HASTAC Social Network” was presented by Marco T. Bastos at the ASE International Conference on Big Data Science and Computing in August 2014 in Beijing, China. Again, as with the ASE SocialCom paper, this one has been tagged as appearing in a joint IEEE/ASE conference, but again the International Conference on Big Data Science and Computing appears to only be affiliated with ASE. Nevertheless, ASE BigDataScience appears to be highly selective, with a 25% acceptance rate for papers. Furthermore, similar to ACM SocialCom, publications in this conference are archived by the ACM.

**Evaluation:** Overall, the number and prominence of the conference publications attributed to the HASTAC-EAGER project is impressive and commendable. The conferences listed here should likely be regarded as selective, yet perhaps a step below equivalent IEEE and ACM—professional societies in the relevant fields—conferences. For example, some HICSS minitracks are only lightly peer-reviewed and have a high acceptance rate. On the other hand, some HICSS tracks are established, reputable, and selective. The “Digital and Social Media” track, while only in its second year in 2015, grew out of the previously established “Digital Media” track and is such a selective track, with ~20% acceptance rate.

**Reports and Blogs Posts**
Beyond the conference and journal papers, the project also resulted in two project reports as well as the final NSF report. These describe specific topics, such as the HASTAC Scholars program—a student community within the HASTAC community—as well as social media within HASTAC.

In addition to these formal reports, the HASTAC-EAGER project also generated a number of blog reports serving as progress reports of the ongoing work. This is also in line with one of the stated goals of the original HASTAC-EAGER proposal. A total of 25 blog posts have been created and can be found on the project website [http://www.hastac.org/collections/eager](http://www.hastac.org/collections/eager). Many of these blog posts also include

---

9. To give some context, the external reviewer has a computer science Ph.D. and publishes in deeply technical scientific communities; this may have an impact on his perceived ranking of conference and journal venues.
visualizations that are used to communicate different aspects of the HASTAC database. Because the external reviewer has specific expertise in this area, we study a few of these blog posts and their visualizations in particular and derive four types used:

- **Text visualizations**: Much of the HASTAC dataset consists of structured or unstructured text, and thus several of the visualizations that have been created use variations of tag clouds,10 hybrid network and text visualizations,11 and variants of topic maps.12 Some of the text corpora were also analyzed using probabilistic topic modeling algorithms.

- **Graph visualizations**: Given the relational and social nature of the HASTAC dataset, several of the visualizations that have been created are appropriately based on graph visualization, mostly node-link diagrams,13 some augmented with edge bundling.14

- **Map visualizations**: The HASTAC.org dataset is geographically distributed—this is one of the benefits of an online academic network such as this, after all—so geospatial visualizations are likely to be of benefit in highlighting this data. Most of such map-based visualizations created appear to be standard maps, often created using Google Maps,15 but some use small-multiple maps to show multiple thematic variables,16 and even cartograms for the distribution of HASTAC members.17

- **Statistical graphics**: Finally, standard statistical graphics—easily accessible in the R system, for example—have also been used to visualize specific abstract datasets. Examples of visualizations used include variations of barcharts and scatterplots.18

**Evaluation**: In general, the approach to visualize the HASTAC data in various ways to finally “see it” from different perspectives is commendable, and served the project very well. Many of the findings discussed in the subsequent conference and journal publications presumably were driven by some of these intermediate visualizations detailing the progress of the project.

**Software**

Besides the visualizations and analyses that were run on the HASTAC dataset, the HASTAC-EAGER project also resulted in an R package for social media search and analysis called SocialMediaMineR. The package has been made available as Open Source under a GPL license on CRAN,19 and allows for querying data about URLs using social media systems such as Facebook, LinkedIn, Pinterest, Reddit, Twitter, and StumbleUpon using a unified R interface. Presumably, this approach allows for social media researchers to quickly and easily write R code that subsequently analyzes the queried data.

---


19 [http://cran.r-project.org/web/packages/SocialMediaMineR/](http://cran.r-project.org/web/packages/SocialMediaMineR/)

---
Evaluation: It is impressive that the project has also created reusable software beyond its original goals of merely analyzing the HASTAC dataset. The SocialMediaMineR package appears highly useful and is likely to be of interest to any social media researcher looking for an easy way to query social media data. However, while the package has been made on CRAN (the Comprehensive R Archive Network), there appears to have been no updates to it since its original publication in June 2014.

Personnel
Led by PI Davidson, the personnel involved in the HASTAC-EAGER project involved two postdoctoral scholars and two technical support staff members from the HASTAC project.

Investigators
The sole investigator for the HASTAC-EAGER project is PI Cathy N. Davidson. PI Davidson is a Distinguished Professor and Director of the Futures Initiative at the Graduate Center of the City University of New York (CUNY). She is also the Ruth F. DeVarney Professor Emerita of English at Duke University and was previously the John Hope Franklin Humanities Institute Professor of Interdisciplinary Studies at Duke. Her research is on digital humanities, technology, collaboration, cognition, learning, and the digital age; she has published more than 20 books on such topics.

Funded Personnel
The primary part of the HASTAC-EAGER funding appears to have been used for funding a postdoctoral fellow to conduct the research in the project. The position of HASTAC postdoctoral fellow has been filled by two different people throughout the duration of the project:

- **David Sparks** was employed as a postdoctoral fellow as part of the HASTAC-EAGER grant from January 2013 to August 2013. He has a Ph.D. from Duke University in political science from 2012 and conducts research on U.S. politics, social network analysis, and statistics. According to his website, he is a statistical consultant for the Boston Celtics (Sep. 2012).
- **Marco Toledo Bastos** was employed as a postdoctoral fellow as part of the HASTAC-EAGER grant from October 2013 until December 2014. Since 2015 he is now a postdoctoral researcher at University of California at Davis in the Department of Environmental Science and Policy (http://www.des.ucdavis.edu/), where he leads the social media analysis portion of the Knowledge Networks Project. Dr. Bastos has a Ph.D. in communication and works on connections between mainstream and social media. His personal website lists 17 selected papers on topics within or closely related to the topics covered by the HASTAC-EAGER project. It appears as if the present project had a positive impact on his continuing career and skills.

Evaluation: The choice of driving the majority of the work in HASTAC-EAGER project through a postdoctoral scholar appears to have been a good decision given the success of the project. Postdoctoral scholars already have existing skills and knowledge of research, and are thus often more self-starting than graduate students. This is evident from the scholarly output resulting from the project.

Changes
The most significant personnel change throughout the project appears to have been postdoctoral scholar David Sparks leaving the project in August 2013, being replaced by Marco Bastos in October 2013.

In addition, PI Davidson moved from Duke University to the City University of New York (CUNY) in July 2014. As part of this move, HASTAC@CUNY was created, directed by PI Davidson.

---

20 https://dsparks.wordpress.com/
21 http://marco.toledobastos.com/
**Evaluation:** The two-month gap between the two postdocs suggests one of several options: that Dr. Sparks left the project abruptly, that finding a suitable replacement was challenging and time-consuming, or that the hiring schedule had to be adjusted due to Dr. Bastos’ time schedule. Nevertheless, based on the timeline of work, the productivity and results from the project does not appear to have suffered. Similarly, the PI moving from Duke to CUNY does not appear to have had much impact on the project; it is not even mentioned in the timeline in the final project report.

**Professional Development**
The two postdoctoral scholars that were employed gained experience as researchers and scholars, as data scientists, as nonprofit members, and as communicators. Technical skills acquired included data extraction, analysis, and management as well as visualization. Furthermore, the postdocs were also able to gain experience in writing blog posts, white papers, and academic papers. In addition, they were also trained in scientific computing and software engineering in building R packages for analysis and visualization.

**External Collaborations**
The HASTAC-EAGER personnel organized a one-day workshop in May 2014 at Duke University on the impact of scholarly networks and cyberinfrastructures on cross-discipline, multi-institutional research. This workshop attracted more than 40 participants discussing these topics.

**Discussion**
Web-based cyberinfrastructures for academic communities are becoming commonplace, and multiple such systems now exist for specific disciplines. Furthermore, many general-purpose social media platforms, such as Twitter and Facebook, are routinely used for academic communities as well. Taken together, all of these platforms constitute “competitors” of sorts to the HASTAC community and website. However, it is important to realize that at the time the HASTAC project was founded in 2002, the concept of social media was still nascent and the online social network platforms that are ubiquitous today simply did not exist. The table below is by no means exhaustive—an exhaustive survey is outside the scope of this external evaluation, as is an in-depth review of the HASTAC project itself—but includes several representative social media platforms to put the importance of the HASTAC project in context.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Founded</th>
<th>Purpose</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>UseNet</td>
<td>1980</td>
<td>Newsgroup discussions</td>
<td>General</td>
</tr>
<tr>
<td>Open Diary</td>
<td>1998</td>
<td>Social blogging service</td>
<td>General</td>
</tr>
<tr>
<td>LiveJournal</td>
<td>1999</td>
<td>Social blogging service</td>
<td>General</td>
</tr>
<tr>
<td>HASTAC</td>
<td>2002</td>
<td>Social networking service for academics</td>
<td>Academic</td>
</tr>
<tr>
<td>nanoHUB</td>
<td>2002</td>
<td>Social networking service for academics</td>
<td>Academic</td>
</tr>
<tr>
<td>myspace</td>
<td>2003</td>
<td>Social networking service</td>
<td>General</td>
</tr>
<tr>
<td>Facebook</td>
<td>2004</td>
<td>Social networking service</td>
<td>General</td>
</tr>
<tr>
<td>Twitter</td>
<td>2006</td>
<td>Social microblogging service</td>
<td>General</td>
</tr>
<tr>
<td>Academia.edu</td>
<td>2008</td>
<td>Social networking for academics</td>
<td>Academic</td>
</tr>
<tr>
<td>ResearchGate</td>
<td>2008</td>
<td>Social networking for scientists</td>
<td>Academic</td>
</tr>
<tr>
<td>DIA2</td>
<td>2011</td>
<td>Social networking for STEM education</td>
<td>Academic</td>
</tr>
</tbody>
</table>

For this reason, the mission of the HASTAC-EAGER project—e.g., to study six years of HASTAC collaboration data, including posts, comments, and blog entries—is an important and unique opportunity to gain insight into the formative stages of the use of the Internet as a communication and collaboration medium for academic discourse and learning. The results gained from it are thus of particular interest.

The traditional model for scientific projects funded by the NSF is to focus on academic publications as the sole metric of success. While the HASTAC-EAGER project certainly has pursued such publications in a
significant way—one journal paper and three conference papers are listed as being derived from the project—the project has also pursued an alternative to scientific publication by publishing a total of 25 blog posts to its existing HASTAC website. While this idea certainly is not unique to the HASTAC-EAGER project, it is intriguing in that it challenges the traditional academic publishing model, and it would be interesting to learn what the project team thinks of the practice. On the positive side, this approach would presumably increase the visibility of the project, facilitate engagement with the audience, and thus ultimately lead to higher impact in the community. However, in terms of negatives, blog posts are difficult to quantify when assessing someone’s academic qualifications, lead to important results being disseminated before they are appropriately published in an archival venue, and require additional effort to create.

EAGER (Early-concept Grants for Exploratory Research) grants\(^2^2\) are awarded by the National Science Foundation to projects in their early stages that are untested but potentially transformative, representing what the NSF calls “high-risk high-payoff” research. The nature of the HASTAC-EAGER project clearly falls within this category, and the numerous findings, reports, and publications arising from it is a testament to the high payoff. Given this success, it would be customary (even if by no means necessary) for the project to continue, possibly by applying for larger-scale, core program funding from the NSF. Areas of possible future funding include (1) exploring the collaboration hub method pioneered by the HASTAC project and evaluating its institutional impact, (2) applying the HASTAC model to other field such as social studies, data sciences, and computation and information science, and (3) investigating the use of the HASTAC model and system for online hybrid learning communities and pedagogical applications.

\(^{22}\) Ibid 1.