Open collaboration in the public sector: The case of social coding on GitHub

Ines Mergel

Public Administration and International Affairs, Maxwell School of Citizenship and Public Affairs, Syracuse University, 215 Eggers Hall, Syracuse, NY 13244, United States

A B S T R A C T

Open collaboration has evolved as a new form of innovation creation in the public sector. Government organiza tions are using online platforms to collaborative create or contribute to public sector innovations with the help of external and internal problem solvers. Most recently the U.S. federal government has encouraged agencies to col laboratively create and share open source code on the social coding platform GitHub and allow third parties to share their changes to the code. A community of government employees is using the social coding site GitHub to share open source code for software and website development, distribution of data sets and research results, or to seek input to draft policy documents. Quantitative data extracted from GitHub's application programming interface is used to analyze the collaboration ties between contributors to government repositories and their reuse of digital products developed on GitHub by other government entities in the U.S. federal government. In addition, qualitative interviews with government contributors in this social coding environment provide insights into new forms of co development of open source digital products in the public sector.

1. Introduction

Multiple forms of open collaboration have emerged that enable the U.S. federal government to create innovative solutions to complex pub lic sector problems. Besides open innovation approaches at government organizations, such as NASA (Gustetic, Crusana, Raderb, & Ortegac, 2015), open collaboration also emerged as an alternative practice for the collaborative development of software products in government. In stead of outsourcing the development of IT projects to vendors or con tractors, internally developed code is updated to a social coding website GitHub and external developers are invited to reuse the code or make suggestions for changes to improve the code. Government can take advantage of the willingness of civic co developers to improve existing public goods and lower the reliance on outside contractors. An other purpose of social coding sites is to simply share code that is creat ed in the U.S. federal government for reuse by other government entities at no cost.

The sharing of code on social coding platforms, such as GitHub, al lows government organizations to reuse the code that is once devel oped, make changes and adapt to their own context. In addition, GitHub social networking functionalities increase opportunities for de velopers to request changes of code developed by other and co develop or improve existing code.

In the U.S. federal government, social coding occurs not just for soft ware pieces, such as website elements, or open data platforms,

government research teams also share their data sets, and algorithms, and recently selected agencies have started to co develop policy docu ments in text format on GitHub.

The contributions to existing code as well as the reuse or copying of code for ones own purposes can be traced and interpreted as social net work relationships to derive insights about the collaborative nature of the government GitHub online community. These collaborative net working relationships are one indicator for social coding and observable for members of the platform, however what is less understood so far are the current use of GitHub for government, the reuse of existing code in other agencies, and the reasons why government coders are involved in sharing and contributing to other organizations' software development.

This paper first introduces the concept of open collaboration and dis cusses different approaches used to accomplish open collaboration. Using a multi method approach, data collected from GitHub is used to conduct a quantitative network analysis of indicators for the reuse of code as well as contributions to code posted by the Government GitHub Community, and combined with qualitative interview data with some of the most active coders in the Government GitHub Community. The main research questions this paper answers include: How and why do government agencies engage in social coding and open collaboration approaches as a new form of IT acquisition? And, what are some of the observable outcomes of these interactions?

2. Open collaboration in the public sector

Public sector innovations mostly occur through formal political mandates, a set of highly regulated bureaucratic standard operating

E-mail address: iamergel@maxwell.syr.edu.

procedures to introduce innovations, for example through a formal re quest for proposals from vendors and contractors that then guide the acquisition of new products, or internal improvement processes by frontline workers which then lead to an oftentimes only incremental re design of internal processes or marginal improvements in public service delivery.

In both the public and private sectors, new forms of open collabora tion have emerged in recent years through which innovation occurs outside these trusted and formalized acquisition procedures. New tech nologies allow alternative forms of innovation creation to emerge with out requiring bureaucratic rules and regulations. The benefits that result from these collaborative technologies have the potential to increase the effectiveness of operations in public service, contribute to the transpar ency of innovation creation, and enhance efforts to collaboratively ad dress public problems. However, the use of these technologies inside of government requires the adaptation of the existing acceptable use paradigm for IT tools in the public sector.

Open collaboration is a dynamic process through which (profession al or volunteer) collaborators contribute value to an already existing product or service by adding their own ideas or applying their own knowledge. Forte and Lampe (2013) define open collaboration as "an online environment that (a) supports the collective production of an ar tifact (b) through a technologically mediated collaboration platform (c) that presents low barriers to entry and exit and (d) supports the emergence of persistent but malleable social structures" (p. 536). In many of the following examples an online platform or online interac tions are necessary, however some open collaboration initiatives, such as living labs as part of smart city approaches foster face to face interac tions by inviting innovators to be co located in government spaces.

The underlying processes of open collaboration vary. The most prominent process is crowdsourcing. Citizens report for example pot holes, or missed waste pick up in their neighborhoods through an on line platform, and these collectively created reports are then used to improve government's service delivery (Clark, Brudney, & Jang, 2013). Similarly, voluntary contributions to OpenStreetMap are community driven crowdsourcing processes where citizens initiate the tasks to pro vide a digital contribution to an online mapping project (Budhathoki & Haythornthwaite, 2013). Open collaboration crowdsourcing ap proaches are supported by technology and occur in iterative processes similar to collaborative innovation creation (see for example: Sørensen and Torfing (2012)). Other open collaboration approaches in clude the provision of micro tasks known as citizen science (Gustetic et al., 2015), or co development of solutions for public management problems through open innovation contests such as on the open inno vation platform Challenge.gov (Mergel & Desouza, 2013). These can re sult in the reuse of public sector information in mobile phone apps (Luna Reyes, Bertot, & Mellouli, 2014). In comparison to open innova tion approaches (Saebi & Fossa, 2015), open collaboration is usually not incentivized with monetary prize payments.

3. The context: social coding on GitHub

The Obama Administration is fostering the adoption of open collab oration approaches in the U.S. federal government. Supported by a set of executive orders, agencies in the U.S. federal government innovate using open source software and new forms of cross agency collabora tion to accelerate the use of open data and improve the adoption of el ements of the Open Government Initiative.[1] Open source code hosting platforms such as GitHub are one part of this open collaboration infra structure, allowing users to share software code and co develop soft ware projects. Prominent examples of GitHub use in government include the data sharing platform Data.gov to facilitate public reuse of

government information, the Congress.gov site that provides a reposito ry of Congressional bills, and the petition platform petitions.whitehouse.gov which allows for public input on the source code with the final code for reuse by others. The software for these sites was either collaborative coded on GitHub for version control pur poses or shared for reuse by other government entities and transparen cy purposes.

The federal government actively encourages open source program mers to collaborate with government to improve software code by allowing programmers to use GitHub's feature called pull requests through which programmers can ask to include modifications into the original code or to alert the original programmer about issues coders noticed when they reuse the software. The platform features include version controls and social alerts that help the initial owner of the code to track other users' activities. Beyond collaboration on software code for application development and website design GitHub allows for open collaboration through data sharing and the co creation of text documents, such as the General Service Administration's social media strategy document. In addition, other innovative uses are slowly emerging, like the use of GitHub for collaborative development of policy drafts or the public display and sharing of draft policies. These ap proaches of government agencies to use GitHub constitute an innova tive approach to open collaboration in the policy making field.

Software code on GitHub is stored in repositories or "repos". Devel opers release their code on the platform and it is then available for col laborative review and suggested revisions. The following graphic in Fig. 1 shows the collaboration process: the initial repository stays un touched and other developers can fork (create a copy of) the repository, comment, ask questions, and suggest changes by sending a pull request. The initial developer will review the changes suggested through the pull request and then decide whether to integrate them into the existing re pository or ignore the change request. The process is characterized by back and forth conversations before issue fixes are eventually integrat ed in the repository. Pull requests include the modified code, the issue that was addressed with the changes, and code comments, of what GitHub calls a 'Living Conversation'. While the majority of repositories are so called single developer repository 'dumps', developers have iden tified that the true value of GitHub is this collaborative and interactive process (see for example: Dabbish, Stuart, Tsay, and Herbsleb (2012)). The following screenshot represents the collaborative coding process. The light line indicates the originally uploaded code and the darker line indicates the copy another developer created, called fork, and the dots indicate revisions as well as comments and responses to the issue:

GitHub users create profile pages similar to other social networking sites that can include their profile picture, full name, organizational affil iation, location, email address, website and organizations on GitHub they belong to. Their user name, date they joined GitHub, as well as the number of followers, who they follow and number of repos they have starred or favorited ("following" is similar to following a user in Twitter) are shown by default. In addition, the profile page includes their public activity, public repositories and a visual representation of their contribution activity (see Fig. 2 for a profile page of a GitHub member):

Developers can be affiliated with organizational accounts. As an ex ample, individual developers who work for government organizations are part of the Government GitHub Community created by GitHub's Ben Balter, a former Presidential Innovation Fellow. In addition, they
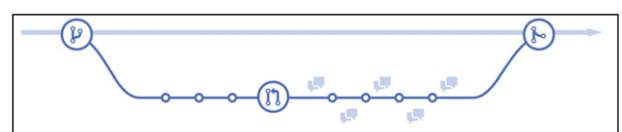
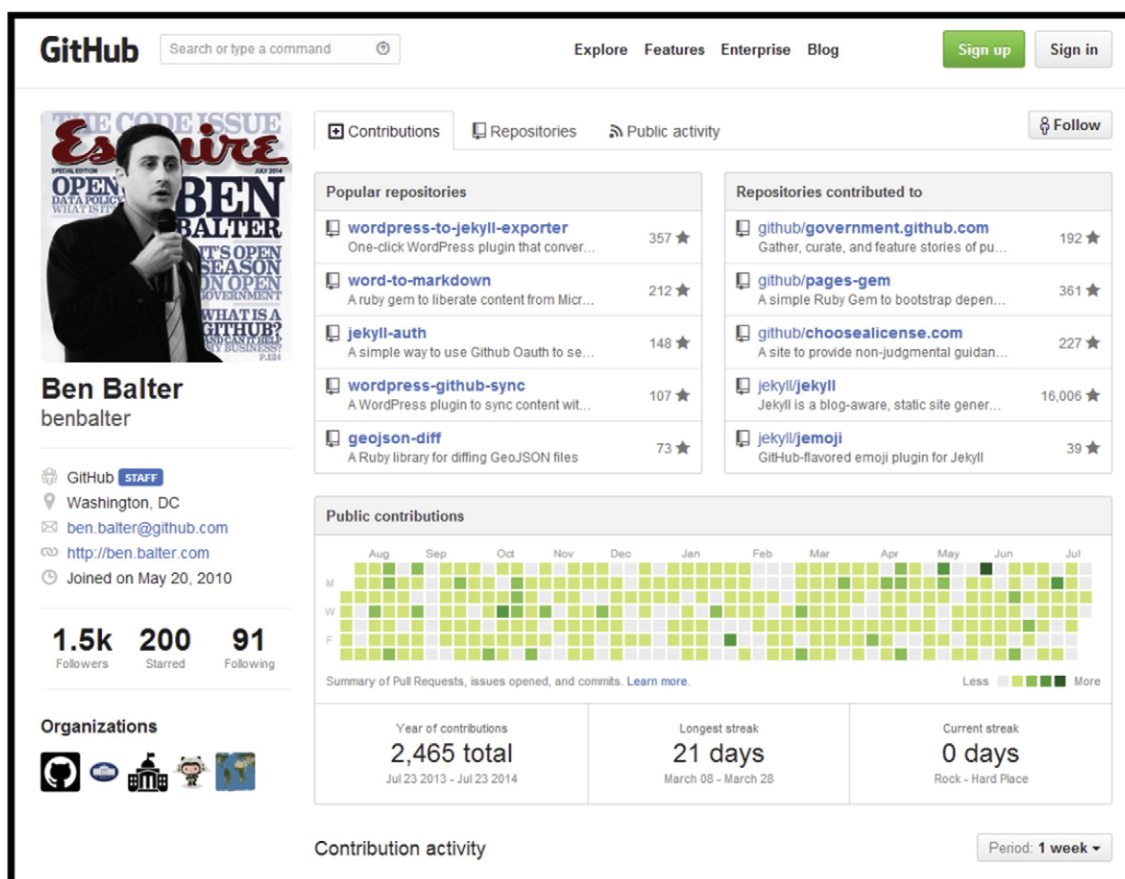**Fig. 1.** Collaborative code review.
Source: GitHub.com.

Fig. 2. GitHub profile page.

can be affiliated with a government organization, such as the General Service Administration's 18F developer group (https://github.com/18f). Another social feature includes tagging of users (using @username) in the comments they can leave each other or in the issues they are reporting to a developer's repository. In combination, tagging, following, and commenting functions create a social network of coders and software code.

Contracting out software development is a well established practice in the U.S. federal government. Social coding is a new form of open col laboration for government organizations. The co development of soft ware code has a history outside of government, most known because of the development of the Linux operating system (Crowston, Wei, Howison, & Wiggins, 2012). Previous research has so far looked at indi vidual contributions or specific projects and focused for example on transparency and collaboration of the social coding process or version control (Dabbish et al., 2012), impression formation in online peer pro duction (Marlow, Dabbish, & Herbsleb, 2013), or the individual devel oper relationships and project to project interactions (Thung, Bissyande, Lo, & Jian, 2013).

The core research question for this paper focuses on the social coding aspect of co development and reuse of software code in the public sec tor: How are government agencies collaborating on software code to improve government operations? Specifically, how and why are gov ernment agencies sharing code, who reuses or contributes to existing open source projects shared on GitHub, and what types of code is shared on GitHub?

4. Research design: data collection and analysis

The data collection used to address these questions was designed to include both quantitative and qualitative elements. In a first step, the GitHub website was used to query the repositories created by organiza tions listed as part of the Government GitHub Community. GitHub pro vides limited access to the data its users are creating on the platform as detailed on the following site with information for developers https://developer.github.com/v3/ and therefore creates issues for social science researchers to use data about social interactions (Marres & Weltevrede, 2013). For this research project, a research assistant wrote a Ruby on Rails script to download the data for each government organization on GitHub and assembled them in a database. The data includes the name of the repository, type of coding language used, the number of forks, and the number of closed pull requests, as well as a short descrip tion of the project when available. The data set was then hand coded by one coder to define the type of code in the repository (Codes: software code repositories, data repositories, guidance documents, policy docu ments, rest: unknown).

The data collected from GitHub allowed the identification of those repositories to which other developers contributed code and to derive initial insights about the type of online collaborations in form of the di rection of contributions (i.e., who is contributing to which repositories by sending pull requests to fix issues discovered or added to the code), as well as indicators of reuse of existing government open source content (forking). The quantitative data served as the basis for a social network analysis to trace the inter governmental collaboration network among members of the Government GitHub Community. The quantita tive network data were then used to formulate initial assumptions about social coding among the Government GitHub Community that were later clarified and tested using in depth qualitative interviews with selected members of the U.S. federal Government GitHub Community.

The Government GitHub Community consists of 357 organizations (as of the end of 2014), of which 192 are based in the U.S., 107 are

government organizations from other countries, and 58 identify them selves as civil society organizations working on government projects. Among the U.S. government organizations 126 are identified as federal government organizations. Developers associated with government or ganizations are either fulltime government employees, or contractors with a dot gov. email address. The following Table 1 shows the types of software code the government organizations share.

4.1. Social network construction and analysis

The social coding interactions among government GitHub devel opers can be traced based using a) the contributions to existing code (pull requests), and by) the reuse of code (forks). These two directional interactions on GitHub can be interpreted as social network ing ties and used to create a social network matrix. Reuse of software code defined as forking activities are stored in an nxn matrix, where the ij connection is established by the active forking task (Wasserman & Faust, 1994). This means, that an interaction or tie between two gov ernment actors is established as soon as a coder from agency *i* forked software code from agency *j*. Similarly, for the pull request matrix a con nection was established if a user contributed code to a repository owned by another agency and asked the initial creator to improve the code with their additions (pull requests). The network matrices were con structed using the social network analysis software UCInet and visual ized with NetDraw (Borgatti, Everett, & Freeman, 2002).

The resulting social networking relationships were analyzed by computing a specific centrality measure, node degree, that measures the number of individual level ties of each coder created by interacting as outlined above (Freeman, 1979). Node degree centrality can be calcu lated in two ways: 1) outdegree centrality measures the number of nominations an actor makes, and 2) indegree centrality measures the number of nominations from other actors an actor in a network receives (for a similar procedure see for example Krackhardt (1992); or Mislove, Marcon, Gummadi, Druschel, and Bhattacharjee (2007)). Out degree centrality is measured for forking activities and in degree centrality is measured for pull requests to identify the direction of the collaboration activities (see Table 4). Degree centrality can indicate influence or power social actors have over other actors, or access to specific types of information or channels through which a social actor can access re sources in their social network. In this GitHub network the analysis in dicates mostly interest (forking) and contributions (pull requests) to an agency's repository and is in this context interpreted as a form of col laboration. Another interpretation can be that government agencies simply see value in already developed platforms and want to reuse it without investing time and resources in repeating already existing code, or simply learning from the software code (Faraj & Johnson, 2011).

In addition to the quantitative data, insights about the social coding process were collected using a semi structured interview instrument to gain a deeper understanding of how social coding is used in govern ment, how accepted it is, how coders in government are contributing to other repositories, reasons why they are sharing code on GitHub, to what extent they are willing to incorporate pull requests from others, how they deal with security and privacy issues, and what some of the outcomes of these online interactions are. Public managers were

selected based on their membership in an organization listed in the Government GitHub Community. While many federal government or ganizations have created a presence on GitHub and participate in the community, few are active in sharing and reusing code. Therefore, the interview candidates selected include public managers involved in pro jects with the highest numbers of forks and pull requests. This allowed an analytical focus on the actual procedural aspects of social coding, which are at the core of this analysis. Those government agencies that only use GitHub to store code are included in the quantitative data set, but they were not included in the interviews because they would not contribute insights when asked about their social coding perceptions. The interviews with eleven interview partners lasted between 30 and 60 min and included questions about the difference between traditional software code acquisition processes and social coding processes, collab oration with external programmers on GitHub, and cultural and institu tional aspects guiding the adoption of social coding approaches.

The qualitative interviews were recorded with the permission of the interview partners, transcribed verbatim and coded line by line based on the initial review of practitioner reviews and press interviews with government officials guiding the interview questions. This approach is not strictly a theory building attempt, however new concepts emerged out of the conversations across federal agencies and contributed to the social coding network community approach (Glaser & Strauss, 1967).

5. Findings

The results provide insights into an innovative format of software development in government that has not been studied in the public sec tor so far. The social network analysis provides an indicator of how open government practices developed in the U.S. are spreading worldwide. Furthermore, the drivers and barriers for open collaboration among government employees are deduced from the qualitative interviews and provide insights into how open collaboration occurs.

5.1. Finding 1: use of GitHub in government

Government agencies in the U.S. federal government use the social coding platform GitHub for many different purposes. Table 2 provides an overview of the different uses of GitHub. Regulatory agencies such as the Food and Drug Administration or the Environmental Protection Agency are using GitHub to share code that can be easily reused by other agencies with very similar needs. These include simple tools such as website elements or modifications of content management frameworks, such as e.g., Wordpress modules distributed at https://github.com/GSA/data.gov/.

Science and technology agencies have a much wider variety of needs. Research teams tend to share data analysis algorithms from their research papers that they have co developed internally on GitHub

Table 1
Government GitHub Community repositories.

Repositories of the overall Government GitHub Community	7003 repositories
Type of repositories of the federal Government GitHub Community	1437 software code repositories (87.2%) 67 data repositories (4.07%) 44 guidance documents (2.67%) 3 policy documents (0.18%) Rest: unknown

Table 2
GitHub use by agency type.

Type of agency	GitHub use	Examples of government agencies
Regulatory agencies	Software code for web platforms Open data platforms	FDA EPA
Science and technology agencies	Research data and algorithms for scientific papers Technological solutions	NOAA NASA USGS DOI
Health agencies	Algorithms used by research labs Health data sets	NIH
Defense and national security agencies	Geospatial databases Software to merge and integrate intelligence datasets Request-for-information Mobile apps code	DOS NGA USGS

in a private repository. They use the public repository as an open access platform for their research papers to point publishers and readers to the extended version of the algorithms that they are constrained from in cluding in a published paper. They collaborate privately in their research teams on GitHub first, and then release the code to the public, but do not allow active collaborations on their code after publication.

Health related agencies might also have website coding needs, but they mostly publish government health data sets on the platform and use algorithms to automatically update the data, which can then be reused by the other developers or the public. GitHub's version control makes it an ideal platform to share updates to existing datasets at differ ent points in time.

While most of the interview partners reflected on the need to pro tect national security when using GitHub, defense and national security agencies are among the most active and innovative agencies using the platform to share code for a wide range of purposes. For example, the National Geospatial Agency is pushing out code for Request for Information applications that can be used for emergency management purposes across government levels, or to develop mobile apps that can be immediately used by local governments without spending re sources on software develop and immediate interoperability. The fol lowing Table 2 summarizes the types of agencies as well as their main GitHub use:

5.2. Finding 2: collaboration across government

The Government GitHub Community shows an overwhelming con centration on selected actors, among them so called 'civic hackers', such as the nonprofit organization Code for America, or the Open Knowledge Foundation. Most government to government interactions however focus on large scale projects, such as the Project Open Data, the Open Government platforms, or the Department of Veterans Affairs' Blue Button initiative, partly reflecting priorities of the current administration.

The following Table 3 provides an overview of the most active agen cies accepting pull requests from other developers and actively incorpo rating the issues that these developers pointed out in their improvement requests into the existing repository. The cutoff point for the number of closed pull requests was set at 10, thus limiting the subset to only those organizations that have the most interactions with the developer community.

The most active organizations include the U.S. Geological Service's (part of the Department of Interior) repositories with 1908 closed pull requests, and the Consumer Financial Protection Bureau (CFPB). This in dicates that most of the activities occur in areas of national priority such as national security (hazards, earthquakes, etc.), priorities set by the Digital Services team at the White House, or individual Members of Con gress, such as Senator Warren's CFPB project for consumer protection. Other projects include web platforms such as the open data platform, the FCC's e Regulations platform, the Department of Veterans Affairs' Blue Button initiative to share veterans' health records within the VA's medical system, and the White House's petitions platform (see Table 3 for an overview of repositories by agency and number of pull requests).

5.3. Finding 3: reusing existing code (forking content)

In GitHub's version control language, developers are forking content of other developers by creating a copy of the initial repository. With a single click, the platform saves a copy of the whole repository including issue reports, versions, and comments in the second developer's page and the initial developer sees who has copied the repository. By modify ing the fork, a contributor can send a pull requests to the initial owner and ask them to incorporate the changes, thus merging the fork back with the master branch. Or the code in the original repository can be reused and substantially modified for another purpose, developing into an independent branch.

Table 3
Most collaborative agencies (as of 2014).

Organization name	Repository name(s)	# of closed pull requests
Department of the Interior	National Geothermal Data System, Center for Integrated Data Analysis, United States Geological Survey, Hazards, Earthquake events, Water quality	1908
Consumer Financial Protection Bureau	API for Home Mortgage Disclosure Act Data, eRegulations API	1518
Department of Defense	Advanced Distributed Learning (ADL), virtual-world-framework, ERDC Computational Mechanics, National Geospatial-Intelligence Agency	1202
National Aeronautics and Space Administration	Sunpy, NeoGeographyToolkit, Vision Workbench, US Integrated Ocean Observing System	917
GSA	Project Open Data, 28F, Presidential Innovation Fellows, GSA	440
Department of Energy	National Renewable Energy Laboratory, kbase, Joint Bio Energy Institute, Los Alamos National Lab, High Performance Computing	419
United States Congress	Library of Congress	98
Department of Veterans Affairs	Blue Button initiative, department-of-veterans-affairs	98
The White House	Petitions.whitehouse.gov, API standards, mobile apps	61
Department of Health and Human Services	National Cancer Informatics Program, National Institutes of Health — National Center for Biotechnology Information, Food and Drug Administration — Privacy Policy	53
Smithsonian Institution	Smithsonian	17
U.S. Peace Corps	Volunteers in the field tracker, Malaria App, Ushahidi crowd map	10
Information Sharing Environment (ISE.gov)	Project Interoperability	10

Across the U.S. federal government forked repositories are relatively common, as the 'fork network' in Fig. 3 shows. Projects such as the VA's Blue Button platform, the OpenFDA platform, CFPB's eRegulations plat form and many projects created by GSA's 18F and Presidential Innova tion Fellows are frequently forked by members of other agencies. The network diagram shows the direction of forks (the arrow head points to the organization forking content from another organization). The size of the nodes indicate the number of forks created by a specific organization:

What the overall forking collaboration network cannot tell is why public managers are copying each other's repositories and how (or whether) they actually reuse the content. Network narratives solicited from the interviews with public managers indicate that there are a wide variety of reasons why public managers use the forking function in GitHub. Generally, forking repositories is much more common than submitting pull requests with requests for changes of an existing repos itory, in part reflecting the single click ability to fork a repository using the GitHub interface. The process indicates that once developed plat forms are easily sharable across government, as one of the interview partners says: "Anytime we can fork right off the stuff, we try to because it allows us the possibility to submit something back to them, so that has helped them."

Another public manager observed that the reuse of repositories has become an important mechanism in the federal government: "There is a big trend to open source entire platforms on GitHub. So for instance, we did that with the FDA, where we released open.fda.gov, and all the code underneath was publicly released. That's good from a transparency aspect, and that's good from a public outreach aspect, and that's good from getting contributions back to us to help us improve our own thing aspect."

Fig. 3. Network composition indicating number of forked content using degree centrality.

However there are technological barriers to this model of copying and reusing each other's work:

"No one is really going to take that entire project and copy it line for line. Where the real utility comes is open sourcing really small pack ages that do one particular function. A really good example of this is: I released a library that allowed an agency to provide what's called an OMB bureau code, and it would translate that a Treasury budget line item code. It's basically just a simple utility to convert one iden tifier to another identifier; that's all it does. And I think actually those types of utilities are more successful, because it allows an agency to say, I have this specific function that the program already needs, and rather than writing it myself, I can just take this tool that already ex ists and plug it in. You find that the most successful kind of forked projects are projects that do small kind of building blocks, rather than try and redo everything."

At one side of the continuum, public managers fork repositories with the intent to contribute to the improvement of an existing platform, or as one public manager points out: "GitHub is just a much easier, more nimble in a platform to work on, and more responsive to needs. And it's also where a lot of the community is already. We just made an orga nization for our Blue Button group and started contributing, and sure enough, people who are not working for VA or government in any way, have been forking the projects and making contributions."

Forking one's own repositories can also lead to a new form of trans parency that might lead to increased collaboration. One public manager explained the process: "So by me forking and taking a side path there, and making changes that others can see, we had the opportunity to re view those before we actually pushed those changes and [before] they were live, and real people were seeing these changes. It became a

good staging area for us to do that. Forking gives you the opportunity for other people to make distributive changes, and then to review those before bringing it back into the primary repository."

Forking is oftentimes done for simple learning purposes. As public managers gain more experience with GitHub and start to use it for ver sion control, they copy existing code to better understand how another agency has designed, for example, code for a mobile app. Another rea son is to learn from existing code as one manager said: "I'd say [forking happens] a moderate amount. The primary purposes of forking are to review the repositories, and experiment with them yourself, to make a suggested improvement to the original, or to take and then build upon for your own uses."

One public manager also noticed that software code once developed in government could contribute to economic development outside gov ernment: "Folks have adopted some of these tools for use in other EMR systems, their own startups."

5.4. Finding 4: reasons for sending pull requests and contributing to other agencies' code

Contributions to already existing repositories through sending pull requests constitute another form of interaction in the social coding pro cess. Once a repository is forked, developers have the ability to make changes, point out issues, and send their changes back to the original re pository owner with the request to integrate the changes. These can be small modifications, such as typos or simple improvements to the code, to more expansive requests customizing the initial repository for requestor's specific agency needs.

External developers, both outside of government or from other agencies, can examine the forked code and identify potential improve ments. The pull requests are then carefully vetted internally by the

initial agency before an authorized developer accepts them. One public manager described the internal decision making process about pull re quests as follows:

"We had an opportunity to look over those [requested changes] and have long discussions about what sort of things to do, and whether things will change in the future that might affect some of the things he's building. We worked out those issues within the pull request it self. GitHub is a nice platform because you can actually comment even on individual lines and say: line number 120, this change will affect something else, or we can do this another way, or this will be an issue. So we can have a pretty robust conversation right in that request. And then when it was ready, somebody who had the rights to the main repository would go and merge those changes in. We have accepted a real contribution. It provides useful functionality."

However, pull requests or improvements to the existing repository are not always substantial in nature. As a matter of fact, people point out issues they observe or only ask for minor changes that will improve the overall quality of a repository. For example: "People are interested and they say, hey, I saw that you weren't coding this field as an integer, but it should be a D. So I changed the code to include as a D. Then you click confirm, and those changes then appear in your repository as well."

The quality of pull requests submitted by developers is carefully evaluated and agencies do not commit every pull request submitted to them. Similarly to other open source communities, the content is evalu ated as well as the developers reputation (see for example: Budhathoki and Haythornthwaite (2013); Butler, Sproull, Kiesler, and Kraut (2013)). One public manager explains:

"There are definitely pull requests we reject. For instance, a lot of times people are adding features that we don't want to have, arguing things that we don't think are relevant to one of the projects. Other times, people might be inexperienced developers, and they write code that may not perform very well, or may be a little messy. In which case, we may say, can you rewrite this, or if they are unable or unwilling to rewrite it, we may say, this is a good feature, we plan on developing it ourselves; we're going to implement in a different way that you proposed, so we'll reject that full request."

In summary, the following Fig. 4 shows the pull request interactions among the U.S. federal government agencies. Similar to forks, organiza tions such as the 18F, GSA, Presidential Innovation Fellows, and the Pro ject Open Data receive the most pull request (highlighted as larger nodes in the network). At the periphery, simple collaborations between pairs of agencies are developed with a low number of pull requests (see Fig. 4):

6. Discussion

The analysis of the social coding process in the U.S. federal govern ment shows how developers collaborate on software code across agen cies. Many agencies reuse code already developed in other parts of government to learn from existing innovations and as a result there is less direct contribution to the existing code. Reusing and adapting code for their own needs is a relative passive process through which public managers copy code, and then potentially internally modify it for their own needs. However, these incremental innovations to existing code are not measureable when an agency does not contribute back to the original code.

Working with the code and then discovering issues, either errors or pointing out opportunities for improvement, is a much more active form of social coding or open collaboration, but less frequently observ able in the U.S. federal government at this stage of the adoption of social coding practices. Nevertheless, this paper showed that agencies are aware of technological innovations developed throughout the overall system of government and are willing to either reuse these innovations, participate in the improvement of other agencies' code to create a public good, or simply learn from the code. The contribution to the existing lit erature is an understanding of the current use of social coding on one specific platform, GitHub, and among the government community.
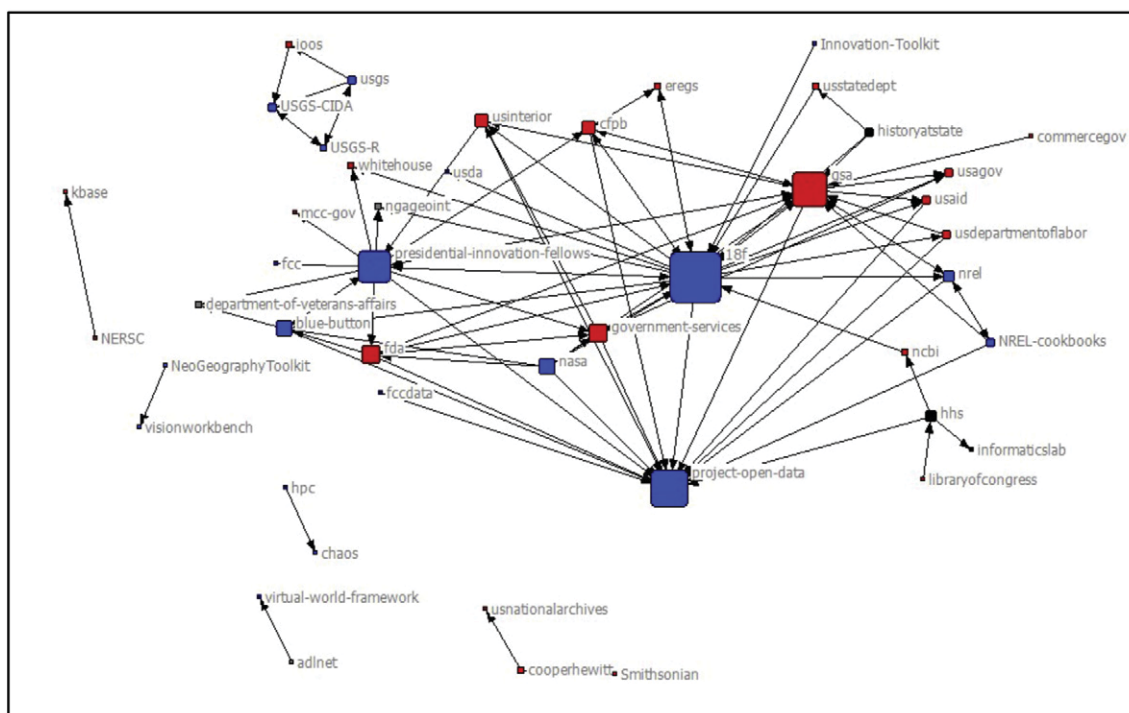

Fig. 4. Network composition indicating number of pull requests received. Node size indicates number of ties.

The traditional software development process in the public sector is reduced to two main actors: either in house programmers who adapt existing code or external contractors who are hired through a formal grant or contract and a highly bureaucratic acquisition process. Discovering the social coding process on GitHub in combination with qualitative interview data with those government employees actively involved in social coding processes provides a deeper understanding of the process and the reasons why government organizations are engaged in open collaboration processes outside the standard innovation acquisition process in government. The existing literature mostly focused on the procedural aspects, here insights emerge about the strategic and operational choices and the reasons why government organizations use social coding and open collaboration approaches. Depending on the context of each agency the purpose varies from sharing approaches of whole platforms to encourage reuse, increases of transparency of existing data sets or algorithms used in publications from government labs, or the attempt to support national security aspects and improvement of (for example emergency management) practices.

Innovations tend to occur among certain actors, such as the U.S. White House, Digital Services or 18F, with less innovation (actual code uploads and major revisions to existing projects) contributed from other government actors. Therefore, I suggest interpreting the early social coding practices as innovative forms of open collaboration in government with a positive perception indicated in form of the reuse of the code by other agencies. At this early stage it is unclear whether the outcomes of the innovations for those agencies that are reusing the code are positive or to what extent code is actually reused for new projects that didn't exist in an agency before. An indirect outcome for U.S. federal government is an increased transparency of government operations, similar to what Dabbish et al. (2012) suggested.

This paper contributes to our understanding of the innovation process in the specific setting in the context of government and additional research is necessary to better understand the organizational change processes, innovation adoption and institutionalization processes in agencies other than the U.S. federal government. How does a reuse of code lead to open government outside the U.S.? Does it save costs for international governments? Does the reuse of the open data platform in other countries also lead to release of more open data and the resulting transparency or economic development opportunities that are currently observable in the U.S.? Questions also remain open about the scaling up of social coding initiatives, as well as how collaborative coding can be routinized and officially institutionalized in government.

While not necessarily the main purpose, the results can also be interpreted through the lens of the open government movement. Collaboration on and reuse of software code, especially code produced by the Project Open Data or text documents that provide strategies, best practices or step by step guides, indicate several different issues: First, they show a general interest in the open government initiative and highlight which other government organizations are interested in the already existing code. Second, it also shows that those who are interested in improving existing code are more likely to actually reuse the whole platform or parts of the platform for their organizations, indicating the spread of elements of the open government initiative. Third, the direction of reuse (forking) indicates that open government platforms developed in the U.S. federal government are spreading to the state and local government level. Earlier research has indicated that open government initiatives or innovations are starting on the local government level, where citizens are directly involved in Open Government Initiative projects. This research project shows that open data software code is reused by local governments in the U.S., which reduces the need to innovate or invest resources to replicate already existing efforts that other government organizations have already invested in.

Appendix A

Table 4
Outdegree & indegree centrality for forked code.

	Outdegree	Indegree
usgs	94.000	2.000
18f	60.000	36.000
presidential-innovation-fellows	50.000	11.000
eregs	34.000	15.000
cfpb	27.000	44.000
USGS-R	23.000	31.000
gsa	21.000	54.000
government-services	18.000	4.000
USGS-CIDA	13.000	91.000
nasa	8.000	0.000
hhs	8.000	7.000
fda	7.000	6.000
historyatstate	5.000	0.000
Smithsonian	5.000	0.000
visionworkbench	5.000	4.000
fccdata	4.000	1.000
NeoGeographyToolkit	4.000	5.000
usinterior	4.000	3.000
hpc	3.000	0.000
nrel	3.000	2.000
commercegov	3.000	0.000
NREL-cookbooks	3.000	0.000
usaid	3.000	2.000
blue-button	3.000	9.000
NERSC	3.000	0.000
ncbi	3.000	2.000
usdepartmentoflabor	3.000	0.000
libraryofcongress	3.000	0.000
HHSIDEAlab	2.000	0.000
ioos	2.000	6.000
opengovplatform	2.000	0.000
informaticslab	1.000	1.000
kbase	1.000	3.000
fcc	1.000	12.000
adlnet	1.000	0.000
usstatedept	1.000	1.000
arcticlcc	1.000	0.000

Table 5
Outdegree and indegree centrality for pull requests sent to repositories for updates.

	Outdegree	Indegree
usgs	1449.000	2.000
cfpb	629.000	387.000
eregs	385.000	638.000
USGS-R	371.000	71.000
18f	208.000	94.000
presidential-innovation-fellows	182.000	30.000
GSA	75.000	125.000
government-services	42.000	8.000
nasa	26.000	0.000
blue-button	19.000	31.000
fda	15.000	37.000
USGS-CIDA	15.000	1758.000
NREL-cookbooks	12.000	2.000
adlnet	12.000	0.000
usinterior	7.000	3.000
nrel	7.000	12.000
hpc	7.000	0.000
historyatstate	5.000	0.000
hhs	4.000	1.000
usdepartmentoflabor	3.000	1.000
commercegov	3.000	0.000
usstatedept	2.000	1.000
NeoGeographyToolkit	2.000	0.000
NERSC	2.000	0.000
Smithsonian	2.000	0.000
usaid	1.000	2.000
cooperhewitt	1.000	2.000
libraryofcongress	1.000	0.000

Table 5 (*continued*)

	Outdegree	Indegree
fccdata	1.000	0.000
ioos	1.000	5.000
ncbi	1.000	1.000
Innovation-Toolkit	1.000	0.000
department-of-veterans-affairs	0.000	5.000
informaticslab	0.000	1.000
usda	0.000	1.000
fcc	0.000	3.000
usagov	0.000	3.000
chaos	0.000	7.000
usnationalarchives	0.000	1.000
mcc-gov	0.000	2.000
whitehouse	0.000	2.000
kbase	0.000	2.000
ngageoint	0.000	24.000
project-open-data	0.000	215.000
virtual-world-framework	0.000	12.000
visionworkbench	0.000	2.000

References

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for windows: Software for social network analysis.* Harvard, MA: Analytic Technologies.

Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for open collaboration crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist, 57*(5), 548–575.

Butler, B. S., Sproull, L., Kiesler, S., & Kraut, R. (2013). Community effort in online communities: Who does the work and why? In S. P. Weisband (Ed.), *Leadership at a distance: Research in technologically-supported work* (pp. 171–193). Taylor & Francis.

Clark, B. J., Brudney, J., & Jang, S. -G. (2013). Coproduction of government services and the new information technology: Investigating the distributional biases. *Public Administration Review, 73*(5), 687–701.

Crowston, K., Wei, K., Howison, J., & Wiggins, A. (2012). Free/Libre open-source software development: What we know and what we do not know. *ACM Computing Surveys (CSUR), 44*(2) (n.p.).

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: Transparency and collabroation in an open software repository. *Paper presented at the CSCW'12* (February, 11–15, 2012). (Seattle, Washington).

Faraj, S., & Johnson, S. L. (2011). Network exchange patterns in online communities. *Organization Science, 22*(6), 1464–1480. http://dx.doi.org/10.1287/orsc.1100.0600.

Forte, A., & Lampe, C. (2013). Defining, understanding, and supporting open collaboration lessons from the literature. *American Behavioral Scientist, 57*(5), 535–547 Retrieved from http://www.andreaforte.net/ForteLampe-Submitted.pdf.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215–239.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Chicago: Aldine Publishing Company.

Gustetic, J. L., Crusana, J., Raderb, S., & Ortegac, S. (2015). Outcome-driven open innovation approaches at NASA. *Space Policy.* http://dx.doi.org/10.1016/j.spacepol.2015.06.002.

Krackhardt, D. (1992). The strength of strong ties: The importance of philos in an organization. In N. Nohria, & R. Eccles (Eds.), *Networks and organizations: Structure, form, and action* (pp. 216–239). Boston, MA: Harvard Business School Press.

Luna-Reyes, L. F., Bertot, J. C., & Mellouli, S. (2014). Open government, open data and digital government. *Government Information Quarterly, 31*(1), 4–5.

Marlow, J., Dabbish, L., & Herbsleb, J. (2013). Impression formation in online peer production: Activity traces and personal profiles in GitHub. *Paper presented at the CSCW'13* (February 23–27, 2013). (San Antonio, Texas).

Marres, N., & Weltevrede, E. (2013). Scraping the social? Issues in live social research. *Journal of Cultural Economics, 6*(3), 313–335.

Mergel, I., & Desouza, K. (2013). Implementing open innovation in the public sector: The case of Challenge.gov. *Public Administration Review, 73*(6), 882–890.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. *Paper presented at the IMC '07 proceedings of the 7th ACM SIGCOMM conference on Internet measurement.* NY: New York.

Saebi, T., & Fossa, N. J. (2015). Business models for open innovation: Matching heterogeneous open innovation strategies with business model dimensions. *Research Policy.* http://dx.doi.org/10.1016/j.emj.2014.11.002.

Sørensen, E., & Torfing, J. (2012). Collaborative innovation in the public sector. The Innovation Journal: The Public Sector Innovation Journal, 17(1), 1–14. Retrieved from http://www.innovation.cc/volumes-issues/intro_eva_sorensen_torfing_17v1i1.pdf.

Thung, F., Bissyande, T. F., Lo, D., & Jian, L. (2013). Network structure of social coding in GitHub. *Paper presented at the CSMR — 17th European conference on software maintenance and reengineering* (Genova).

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge: Cambridge University Press.

Ines Mergel is Associate Professor of Public Administration and International Affairs at the Syracuse University's Maxwell School of Citizenship and Public Affairs. Her research focuses on government technology innovations, especially social networking services in the public sector.