# National mapping of biomedical research resources through voluntary researcher participation - a mixed design study

Jatin Shah, BAMS, PDCR 1;
Guilherme Roberto Zammar, BS 2;
Dimple Rajgor MSc3
Shreyasee Pradhan, MSc 4;
Tam Cam Ha, PhD 5;
Ricardo Pietrobon, MD, PhD 6;

1 Research assistant, Duke NUS Graduate Medical School, Singapore and member of Research on Research group, Duke University, USA
2 Undergraduate Student, Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Paraná, Brazil and member of Research on Research group, Duke University, USA
3 Research fellow, Duke NUS Graduate Medical School, Singapore and member of Research on Research group, Duke University, USA
4 Research fellow, Duke NUS Graduate Medical School, Singapore and member of Research on Research group, Duke University, USA
5 Senior Epidemiologist, National Cancer Centre Singapore, Assistant Professor, DUKE NUS Graduate Medical School Singapore
6 Associate Professor and Vice Chair of Surgery, Duke University, Durham, NC, USA; Associate Professor, Duke NUS Graduate Medical School, Singapore and Director of Research on Research group, Duke University, USA

Corresponding Author:
Ricardo Pietrobon, MD, PhD, MBA
Associate Professor and Vice Chair,
Department of Surgery
Director, Research on Research Group
Duke University Medical Center Box 3094
Durham, NC 27710, USA
rpietro@duke.edu
Telephone: 919-668-2054
Fax: 919-681-8886

# Abstract

## Background

Although the sharing of research resources (RR) information would optimize collaborations and avoid unnecessary redundancies, research stakeholders tend to shun voluntary sharing of data and meta-data about research resources. Therefore this study was conducted to evaluate barriers and incentives contributing to voluntary sharing of research resource information from the bottom-up, i.e., among biomedical researchers.

## Methodology/Principal findings

This project used a mixed design, combining qualitative and quantitative methods and a policy model. We asked research stakeholders to share information related to study databases owned or managed by them and point us to other sources of information (researchers, institutes, organizations). A majority of research stakeholders were concerned about confidentiality, institutional policy and the nature of approval required for sharing information. Majority of respondents who shared requested information delivered it in less than 5-8 weeks, while some pointed to other sources of information.

## Conclusions

Our policy model based on expert opinion and the findings from this study points us towards an overall system of behavior that favors the use of combined top-down and bottom-up policies. A research resource mapping exercise in biomedical research can provide critical information to policy makers of a nation, allowing them to make well-informed and well-planned decisions. Voluntary sharing of research resources is limited by confidentiality, institutional policy and lack of willingness to share information, so indirect methods need to be explored. Research stakeholders can then use this information in a social network format to identify existing resources and build research partnerships, thus leading to economies of scale and reduced duplication.

**Key words**: research resources, mapping, meta data, data sharing, research processes.

## Introduction

With the continued investment in biomedical research, the number of research resources such as databases, experts in multiple fields, and analytical techniques have all grown exponentially. [1,2] Although this growth is inherently an asset, it is also of concern to research policy makers as it is associated with an increase in the degree of disconnect across researchers, who no longer have enough information to optimize collaborations and to avoid unnecessary redundancies. For example, two researchers in complimentary fields may be unknown to each other and therefore miss opportunities for synergistic collaboration. In addition, researchers might initiate data collection for a project when similar data may already be available from another group without them being aware of that. Besides, a research group might plan to buy an expensive laboratory equipment when in fact this same equipment might already be available and under-utilized by another research group next door. In order to avoid duplication and increase cost-efficiency, it is therefore necessary to map research resources in a consistent fashion, represent them in a format that optimizes search, and place this information in a social network that facilitates sharing and optimizes its use. Although description of a mapped research environment represents an ideal scenario, to our knowledge no previous studies have been conducted to evaluate the viability of such a system, particularly on whether voluntary provision of this information by researchers themselves would be a viable alternative.

The closest effort to voluntary provision of information on research resources to date has been attempted through the use of academic social networks. Web-based social network sites are on-line communities of individuals with common interests who share information about their activities. This sharing is provided through a variety of means, including short posts (microblogging), internal email-equivalent messages, and sharing of information such as videos or news. Over the last few years, social networking websites are being used regularly by millions of people around the world, and more recently this phenonemon has extended to scientific communities. Although the level of communication in scientific networks is not as extensive as most popular social networks such as Facebook [3], professional networks such as LinkedIn [4] or scientific networks such as Academia [5] and LabSpaces [6]; they are beginning to enlarge in terms of number of members. These networks have, however, demonstrated a slow growth, which could be explained by most biomedical researchers not necessarily being at the cutting edge in terms of web applications, but could also be related to fears of sharing information about their work and losing their competitive research advantage. [7] To our knowledge however, this potential fear has not been modeled in a systematic fashion that would allow research policy makers to create the appropriate incentives for information sharing.

The sharing of data and meta-data has been extensively discussed among research policy agencies such as the National Institutes of Health. For example, the NIH Final Statement of Sharing of Research Data states that:

"*NIH reaffirms its support for the concept of data sharing. We believe that data sharing is essential for expedited translation of research results into knowledge, products, and procedures* *to improve human health. The NIH endorses the sharing of final research data to serve these and other important scientific goals. The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers. NIH recognizes that the investigators who collect the data have a legitimate interest in benefiting from their investment of time and effort. We have therefore revised our definition of "the timely release and sharing" to be no later than the acceptance for publication of the main findings from the final data set. NIH continues to expect that the initial investigators may benefit from first and continuing use but not from prolonged exclusive use.*" [8]

These efforts have been reinforced by statements such as former President Bush's signing of the "America COMPETES Act" requiring civilian federal agencies to provide guidelines, policy and procedures to facilitate and optimize the open exchange of data and research between agencies, the public and policymakers. [9] Although these statements are focused on data exchange, similar arguments could be made about the exchange of meta-data, defined as information about data rather than data sets themselves. In favor of the exchange of meta-data, it is plausible that policies enforcing the sharing of meta-data would generate less resistance from researchers since sharing of meta-data would retain their full control over the research resource that demanded their personal effort to be collected [10]. Although the policy mechanisms enforced by the NIH and presidential statements have generated top-down pressure on researchers for the exchange of data and meta-data, at this point it is unclear whether a bottom-up mechanism would be as successful.

Although there has been an increased focus on evidence based policy in recent years [11], apart from issues related to supporting quantitative evidence, implementation issues are not routinely and systematically considered before deciding future course of action.[12] Individual preferences and a conducive social environment represent qualitative factors that are frequently neglected. [11] Although mixed method approaches have been used for health policy formulation [13,14], they do not take into account the unpredictability of a proposed policy intervention. The latter poses a major risk that needs to evaluated for short and long term implications as well as intended-unintended consequences. System dynamics modeling [15] is an effective method that generates a representative model of a complex system and utilizes qualitative, quantitative and historical data as inputs. It is used to simulate policy implications and 'what if' scenarios that guides evidence based policy implementation in various disciplines. Despite the availability and use of such models in health and research policy, to date this method has not been used to guide research policy implementation in relation to research resource management.

We aimed to study the behavior of a research system and simulate its response to top down and bottom up policy scenarios in context of data and meta data exchange. The objective of this study was therefore to use a mixed design to evaluate the barriers and incentives contributing to voluntary sharing of research resource information from the bottom-up, i.e., among biomedical researchers, and using this information to build a model that can allow decision makers in devising future policies toward the sharing of data and meta-data among biomedical researchers.

## Methods

Since the key objective was to build a policy model that allows decision makers to simulate a variety of alternatives and their consequences, our project used a mixed design combining qualitative and quantitative information to better understand the mechanisms underlying voluntary sharing of research resource information. To gather information for the model, we approached research stakeholders working in healthcare organizations or research institutes in Singapore in order to understand their willingness to share information about research resources accessible or available with them. Our choice of Singapore was driven by its policy towards biomedical research innovation as well as its incentives for sharing of biomedical research information within the city state. [16] This study was initially conducted with the aim of generating an operational report and hence we did not obtain written consent. The study was later approved along with the waiver of consent request by the Institutional Review Board at the National University of Singapore so that we could report it as generalizable information through a peer-reviewed publication. The following sections outline the methods used to obtain the qualitative as well as quantitative information, followed by the modeling methods used to combine this information into a policy model.

### Ethnographic study:

### Population/Participants

We followed a convenience sampling strategy [17] for the ethnographic study. The selected stakeholders included biomedical researchers (clinical and translational researchers, department chairs), research administrators, research related IT (information technology) administrators and policy makers overseeing biomedical research activities at the institutional and organizational levels. Several of our stakeholders simultaneously held multiple roles. For the purpose of this study we categorised them into six categories based on their roles in their institutes as follows: 1. Researchers, 2. Researcher cum administrator 3. Administrator 4. Policy maker and 5. IT administrator.

### Research resources (RR) definition

For the purposes of this manuscript, research resource information is defined as information about researchers through their biosketches (education, publications, and funding), information about study databases that belong to each researcher or research group, information about analytical methods with which individual researchers have experience through previous training and/or publications and information about infrastructure and staff available to researchers. To facilitate the communication with our target stakeholders, we formulated standard definitions and data collection formats (SI 1, 2 and 3) for the collection of information on research resources during our study. For the element biomedical study databases, we asked stakeholders to provide us with a general description of their database as well as a data dictionary. General description on a study database corresponds to information like title of database, name and address of contact person, keywords, sampling method, population among other fields. A data dictionary includes information on component variables including category, variable name, code and description on alternative responses, whenever available. For research expert information, we asked stakeholders about their willingness to share their biosketches - a document that contains information such as educational background, work experience, publications, and funding. For analysis methods, we asked them about their willingness to provide information on the various analysis methods they were familiar with, and which they knew from either previous training or publications. We solicited information on analysis methods from researchers only. For infrastructure, we asked stakeholders about their willingness to provide us with information on laboratories and equipment available to them for carrying out their work. Lastly, we collected their impressions on sharing information about staff, including all personnel working under a senior researcher, their qualifications, skillsets and previous experience.

In order to facilitate the collection of research resource information, we developed and deployed two Web-based applications for the collection of information on biomedical databases and experts, the Database of Databases [16] and BioS-Sg [19] respectively. The Database of databases is a web-based repository for meta data related to biomedical study databases. [20] Stakeholders were asked to either enter the information directly into the Database of Databases or to provide us with the information so that it could be entered by our staff. BioS-Sg is a web-based biosketch management application, capturing researcher information on educational background, publications, and funding support in a master biosketch.

## Ethnographic evaluation

To define our sample, we initially prepared a list of major biomedical research institutes and organizations in Singapore containing their name, location, mission, activities, and a general description of their governance structure. From this list we selected a non-random sample that would encompass representatives from major institutes, government agencies as well as representatives of smaller research teams. Each stakeholder was approached by email and phone, followed by face-to-face interviews in cases where they agreed to participate. Out of 30 stakeholders who participated in this study from six institutions, we interviewed 15 of them, while email communication was used with the rest of them. Since ethnographic interview method emphasizes meeting each participant in their own environment [21], we tried to meet stakeholders in their own workplace whenever possible. During each meeting, we initially provided them with a description of our project, followed by a discussion whereby we encouraged them to provide their views, insights and comments about the voluntary sharing of information on biomedical research resources. During these interviews, we followed a semi-structured script, with topics covering their concerns related to sharing of information, additional information that they thought should be collected, their perception about how their peers would respond to sharing information, methods to reduce peer concern and encourage their participation, and their views on how this project could be successfully implemented on a large scale. We also asked stakeholders to (1) share information related to study databases owned or managed by them, (2) point us toward other sources of information (researchers, institutes, organizations) where we would obtain information relevant to our project. All responses were annotated in real time as well as supplemented later by additional notes made about the environment and other aspects perceived by the main field researcher (JS). After each interview we provided stakeholders with a link to the Database of Database and the format (SI 2) they could use to provide us with general information on the study database and the data dictionary. All interviews were followed up through a mix of phone calls and email messages.

All field notes were qualitatively analyzed for emerging themes using the Atlas.ti software (ATLAS.ti GmbH, version 5.2)[22], a desktop based software used to conduct qualitative analysis of textual, graphical, audio or video data. Notes were initially coded, transcribed in a regular text processor, saved as individual files and uploaded to Atlas.ti, with blocks of text representing an idea/concept then tagged individually by the main author (JS) and double checked by another investigator (SP). These tags were then grouped according to their meaning/similarity and identified emerging themes, ultimately mapping them into a network diagram.

## Quantitative evaluation

We determined stakeholder response based on the type of information shared with us. For example: in the form of data dictionaries, case report forms and screenshots of database applications. We also noted responses that pointed to other sources of information. We calculated the time to share research resource information from the date of initiation of the discussion with the research stakeholder to the date when the information was received. To understand the degree of difficulty involved in collecting research resource information, we kept track of the number of stakeholders that had to be approached for obtaining information related to a single study database. In order to determine the documentation involved, we noted the type of documentation required before they could share research resource information. For example we noted whether we had to sign a non disclosure agreement (NDA), provide a one page overview describing the mapping project or both. In cases where information was shared with us, we noted the type of information (data dictionary, CRF, screenshots), corresponding number of files and completeness of shared information. Finally, it was noted whether stakeholders provided information in the requested format and in cases where they did not, we recorded whether they reformatted the information as per our requirements or whether members from our team had to do it.

## System dynamics modeling

To summarize the study findings we used a System Dynamics (SD) model. SD [23] focuses on understanding the behavior of complex systems over time. Our model therefore uses stocks and flows to aggregate variables that accumulate and deplete over time such as "shared research resources". In addition, causal arrows were used to create relationships among model elements, thus creating a variety of feedback loops. These feedback loops are largely responsible for the non-linear behavior of our model [24], and can be classified as 'balancing' (which promotes the balance of the system) or 'reinforcing' (which promotes growth of the system). Graphics demonstrating the overall behavior of the system over time were aimed at depicting their overall pattern rather than exact numerical prediction of their outcomes. We used the qualitative responses from the research stakeholders to populate and substantiate our SD model. The preliminary model was discussed by three of the authors (JS, GRZ, RP) and was generated based on an archetypal structure named "limits to success", which represents a behaviour pattern when some elements are limiting the growth of a system [25] As model inputs, we used data from the quantitative analysis of this study and when it was not possible we used subjective values based on discussion and consensus within the researchers, generating the lookups required for the system. The most important equations used in this model can be referred to in the Appendix.

## Results

### Population/Participants

Most of the respondents were either researchers (10/30) or held a dual position as researcher cum administrator (13/30). Very few of them were administrators (2/30), policy makers (1/30), administrator cum policymaker (2/30) or IT administrators (2/30).

### Ethnographic results

Most respondents (stakeholders) understood the underlying concept of research resource mapping and were at least in principle interested in participating and contributing information. Several respondents were able to contextualize this project based on partially similar efforts in the past such as the Singapore researchers database [26]. Despite their interest, respondents quoted a number of roadblocks in sharing information, including lack of time, resource constraints, and institutional policy. Collecting, formatting, uploading and updating information was perceived as a time consuming task. Respondents conveyed that they neither had the time nor considered this a priority. Additionally, considering that such an exercise would involve an immediate and recurring cost, they also cited paucity of resources for supporting such an activity. Resource constraints included lack of support staff to help in activities like gathering, formatting, sharing and updating information needed for our project. They also commented that such activities should be supported and managed by the institution rather than researchers themselves - "I think the institute should fund this." They conveyed that even if they took the onus, they would have to seek approval from their department/institution to enable them to share information and direct resources to support such an activity, which was considered a time consuming and sensitive task. Respondents aware of the dynamic nature of the shared information raised concerns on how and who would update the information They raised questions like: "Who would be incharge of regular updates?" and "What if they were not able to update their own information? Would anyone help them?" We also observed that respondents frequently confused meta data sharing with actual data sharing, thus resulting in concerns regarding confidentiality, approval and intellectual property. Some respondents who confused metadata with data flatly refused to share information. Despite explaining the difference, they were not sure whether they could share metadata. They considered it safer to double check as well as seek approval from their department and institutional management. They also feared whether sharing meta data might be against institutional Intellectual property rights policy. [figure 1]

**Figure 1: Network map based on themes from ethnographic interviews**

Evaluation of the ethnographic notes led to the emergence of themes. Chief amongst them was "understood the concept". Other themes were either associated or originated from this theme. We noted that clarity of concept helped respondents to associate our project with similar past efforts that have attempted to achieve similar goals. Drawing on the experiences and outcomes of the past efforts they questioned the sustainability of our project which, when combined with their concerns about data quality raised questions about the feasibility of the mapping exercise. In some cases respondents were even skeptical about the success of the mapping exercise.

On the other hand, clarity of concept helped them think positively about sharing the requested information (data dictionaries and database description). Despite agreeing to share information, in principle, their decision was constrained on account of the need to seek prior approval from competent authorities. A significant number of respondents were interested in sharing information but were constrained by roadblocks like ownership and institutional norms. In many cases despite clarity of concept and understanding that we were not collecting confidential information, respondents cited operational concerns and the lack of resources to support a research resource mapping exercise. In some cases respondents pointed us to other sources of information.

Nevertheless respondents understood the concept and appreciated the need for such an endeavor, they provided numerous suggestions. These ranged from changes in information fields being collected, information collection methods, possible ways of educating research stakeholders and features for the final social networking website for research stakeholders.

## Quantitative results

Nearly one third of the respondents did not share research resource information (DD/CRF/screenshots) (9/30) and nearly half of the respondents did share it (12/30). A significant number of those who did not share information (8/30), directed us to other sources of information. One respondent initially shared information but later retracted citing the reason that it was against institutional policy. Of the 12 respondents who shared information, (5/30) provided data dictionaries and (6/30) provided case report forms. Only one respondent shared screenshots of their database user interface.

In terms of time involved in sharing research resource information, most respondents (10/12) reverted back with the information (DD/CRF/screenshots) in less than 8 weeks (less than 1 week = 2, 1-4 weeks = 3 , 5-8 weeks = 5). The rest of the respondents (2/12) took more than 13-16 weeks to revert back . Most of the information we collected was received after communication with less than 3 stakeholders for a single data dictionary (Degree 1= 2 respondents, Degree 2 = 4 respondents, Degree 3= 2 respondents). In most cases (8/12), documentation in the form of non disclosure agreement (NDA) or project brief were not required before sharing requested information. Most of the respondents shared only one data dictionary (4/5) or Case report forms for only one database. (4/6). One respondent shared screenshots for the user interface of database application. Very few respondents [1/5] shared data dictionaries in the prescribed format. On similar lines, very few respondents shared complete general description of their databases (3/12). Finally all respondents (12/12) did not provide the information in the requested format and our research team had to reformat it.

## SD Modeling results

Our model describes the behavior of the research system (figure 2) in two hypothetical contexts: voluntary sharing (representing the "bottom up" and named "current" scenario) of information on research resources alone and then the policy model, when mandatory sharing ("top down") works together with voluntary sharing of information. In the first simulation model, the system shows primarily the behaviour of two balancing loops (B1 and B2). This behavior highlights the fact that research stakeholders resist sharing information about research resource due to lack of incentives to share as well as a paucity of resources to support such an effort, showing the low strength of the bottom-up approach alone. The first loop B1 (Balancing-1), represents the arm of voluntary sharing that is affected by resource consumption related to sharing of information on research resources. The second loop, B2 (Balancing-2), represents the arm of time consumption. Both arms culminate in the variable that we name "Limits against sharing", which is influenced too by the "Sharing concerns related to loss of competitive advantage". This "Limits against sharing" archetype represents barriers to the implementation of sharing and determines ultimately the "Stakeholders Resistance to share", and the more resistance they have, less voluntary sharing they do. In this simulation, assuming the fact that the amount of shared research resource is inversely proportional to the money and time consuming, the voluntary sharing among the bottom-up approach shows a discrete growth over time. The behaviour of the system was determined by a number of factors, especially the lookup functions generated by the model inputs previously explained in methods.

In another simulation, the model shows the behaviour of the system when the 'mandatory sharing' works together with voluntary, what we have labelled as the policy model. That policy model is representing the top down approach. Beyond the previous cited loops (B1 and B2), in the Policy model we have one reinforcing loop (R). That loop (R) acts by adding power to the growth of the sharing through a mandatory policy. Another point to noted is the fact that shared research resources complemented by funding support enhances the number and viability of new research projects that can be carried out.

Comparing both hypothetical scenarios ('current' versus 'policy model'), we can see that the limits against sharing decreases over time in both situations, but much more aggressively in the policy model (Figure 3). This behaviour can be explained by obtaining more shared research resources early (see Figure 4), decreasing the money and time needed to share voluntarily. Finally, and more important, is the number of research projects viable due to shared research resources (Figure 5). While in the current scenario, number of research projects grows slowly over time due to the reduced number of shared resources, with the policy model we can see an exponential growth in the number of viable projects.

## Discussion

To the best of our knowledge this is the first study to evaluate the barriers and incentives contributing to voluntary sharing of research resource information from the bottom-up, i.e among biomedical researchers. Based on the ethnographic interviews, we found that although a majority of research stakeholders understood the concept and the need for a research resource mapping exercise, they were concerned about confidentiality, institutional policy and nature of approval required for sharing information for our study. A majority of respondents who shared requested information (meta data and database general description) delivered it in less than 5-8 weeks, while a number of respondents pointed to other sources of information. For a majority of cases, we did not have to provide any documentation to obtain data dictionaries and database general description. The format of almost all the information shared with us was different from what we had requested and our team had to reformat the information received. The results of our SD model simulates the behavior of the research system in response to policy intervention that incorporates the incentive and mandate approach. Taking cue from this model, policy makers and administrators can utilize and simulate several other 'what if scenarios' before making recommendations that have system-wide ramifications.

Despite the fact that innovation has become a global phenomenon, regional centres of excellence continue to be the locus of innovative activities. Regional centres termed as 'innovation hot spots' are seen as building blocks for national innovation capacity as they provide proximity and specialized assets. [27] An overview of innovation and scientific research systems points towards the prevailing imbalance in sourcing, managing and utilizing research resources and calls for a systematic and innovative approach to map, interlink and display a platform easily accessible to all. Providing mapped and interlinked information about existing research resources can benefit the whole spectrum of research stakeholders and trigger regional growth. [28]. Such sharing of critical resource information facilitates the development of highly productive and cost effective research partnerships and networks [29]. Getting a sense of formal and informal regional asset linkages also contributes to a better understanding of the regional economic environment. Resource mapping can thus play a multi-factorial role ranging from identifying and ensuring access to research resources, cultivating new partnerships and relationships, preventing duplication of services and resources, identifying flexible funding strategies, aligning systems and supports for both stake holders (Researchers and Research policy makers) and using data to make informed decisions.

Our findings indicate that voluntary and bottom-up sharing of research resources information may not work due to a variety of reasons encountered in our study. Despite being able to appreciate the concept and need for resource mapping, research stakeholders are limited by the prevailing concerns about data sharing. Concerns about confidentiality, ownership of information and intellectual property rights were frequently cited. Other studies have noted issues such as anonymity and other logistical, technical, legal, and ethical questions [30] affecting data sharing initiatives. Researchers frequently feel threatened by open data sharing not only because their credibility is at stake but also fear losing their competitive advantage. These concerns can be addressed by supportive institutional policies and encouraging trusting relationships among research stakeholders. Previous studies have noted that presence/absence of trust relationships has a direct influence on researchers' willingness to share data. As a result, policy makers need to promote the trust element among research stakeholders through their policies. [31]

In our study, despite providing clarification that we were collecting meta data, respondents continued to have concerns. Fearing institutional or authoritative backlash, research stakeholders considered it safer to double check with management before sharing any kind of information.

Further our findings in terms of the degree of delay and need for documentation while collecting the prescribed information are also indicative of the same. Institutional policy that should ideally be open and supportive to sharing frequently played an opposite role. Intellectual policy and confidentiality may not serve as a hindrance where meta data is involved.

The bottom up approach relies highly on individual research stakeholder decisions and limited resources available to them. They are not only expected to provide the information in a prescribed format but also need to update the shared information periodically. Research stakeholders need to allocate resources to support both these activities, which have long term cost implications. In our study, many research stakeholders did not consider the research resource mapping exercise as their priority or had limited resources to devote for such an exercise and thus were unwilling to allocate time and effort. This can also be deduced from our findings such as 1. a delay in sharing requested information (9-12 weeks in a majority of cases) 2. lack of adherence to the format of information that we had requested (majority of cases in our study) and 3. the fact that most of the data had to be formatted by our team (majority of cases in our study).

At the national level, research policy makers are required to devise and implement policies in response to development challenges posed by global progress and build complementary research capacity. Fragmented and incomplete information about the existing resources, current trends, performance of research domains, related costs, future implications can influence critical decisions. It is common for institutional administrators to see laboratory equipment being duplicated across different groups when a collaboration could lead to substantial economies of scale not only in terms of the equipment itself but also in terms of staff training required to operate the equipment. [29] As an alternative to the 'bottom-up' approach, a 'top-down' approach has its own pros and cons. Devising and mandating a resource sharing national policy could lead to rapid implementation in the short term, but it would need long term commitment to ensure its sustainability. Further, the cost of implementation and maintenance would be significant. Finally updating research resource information regularly would prove to be a major challenge. Chasing busy researchers may not be feasible in the long run. A balanced approach is essential to get research stakeholders on board not only to share but also to update their resource data regularly.

Drawing parallels to our study, The Institute of Medicine, USA conducted 6 case studies on large scale resource sharing to understand the role of institutions in data sharing, barriers for effective sharing and devised solutions for the same. [32] They studied two types of institutional arrangements - the 'repository type' and 'user facilities' models. The committee recommended the need for 1. strong scientific leadership in agencies and research community, 2. core funding, 3. marketing and advertisement, 4. clear guidelines about ownership and access, 5. user fees, 6. clear policies for retaining and discarding data and material, 7. quality control, 8. well defined policies for the function of research and service at the facility and 9. sophisticated information retrieval and transfer systems. Notably, many of these recommendations are similar to our findings. On a final note, the committee also noted that an uniform policy would not be possible on account of the diverse nature of data sharing issues. This resonates with our conclusions, in the sense that there needs to be a balance between mandatory policies and voluntary participation while implementing a resource mapping exercise.

In the absence of popular support and participation, it is difficult to realize the data sharing mandate. Many data sharing initiatives focus on the technical aspects and do not address the need for incentives, training, and promotion of data sharing among key stakeholders [33]. Generating awareness and training biomedical researchers can ensure scalability as well as success of a project. Biomedical researchers should be educated about the benefits of data sharing (i.e.greater visibility and reproducibility of one's own work, increased opportunities for future work and collaboration, advancement of the field) and encouraged to plan for it in designing research projects, just as computer scientists are urged to release software so that empirical results can be easily reproduced and scientific progress can be achieved [34].

Our suggestions do have limitations because a real social network might have different incentives and therefore results maybe different from what we perceive. In addition, a convenience sampling method has its own limitations. It has been argued that the results of a study using convenience samples are seldom generalizable. [35] Despite these facts, exploratory, clinical and illustrative situations can be adequate justifications for the use of convenience sampling. [36] Our study was largely exploratory and thus qualifies for its use.

A research resource mapping exercise in biomedical research can provide critical information to policy makers of a nation, allowing them to make well-informed and well-planned decisions. Implementation of such an exercise is time and cost intensive and needs to be planned after considering all involved stakeholders and their perspectives. In the absence of a systems approach it might fail to yield the expected benefits. Since voluntary disclosure of research resources is often limited by confidentiality, institutional policy and lack of willingness to share information, indirect methods like text mining of public domain data need to be explored. Research stakeholders can then use this information in a social network format to identify existing resources and build research partnerships thereby leading to economies of scale and reduced duplication. [37]

## Acknowledgement

We would like to acknowledge team "Research on Research" for the Literature matrix platform [38], Manuscript templates [39] and Database of databases[20], Duke University Health System, NC, USA.

## References

1. Rosenthal E T (2006) Robert Weinberg Uses AACR Award Presentation as Bully Pulpit to Help Secure Future of Scientific Discovery. Oncology Times. 28(9):15-16.

2. Sinha AM, Hripcsak G, Markatou M (2009) Large datasets in biomedicine: a discussion of salient analytic issues. The Journal of the American Medical Informatics Association. 16(6):759-67.

3. Facebook. Available from: http://www.facebook.com/.

4. Linkedin.  Available from: http://www.linkedin.com/.

5. Academia. Available from: http://www.academia.edu/.

6. Labspaces. Available from: http://www.labspaces.net/.

7. Cheung K-H, Yip KY, Townsend JP, Scotch M (2008) HCLS 2.0/3.0: Health care and life sciences data mashup using Web 2.0/3.0. Journal of Biomedical Informatics 41: 694-705.

8. Final NIH statement on sharing research data, (2003) in *NOT-OD-03-032, National Institutes of Health (NIH)*. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

9. United States. Congress. House of Representatives.(2007) *The America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act (COMPETES)*. http://science.house.gov/legislation/leg_highlights_detail.aspx?NewsID=1938

10. NCRR's 2009-2013 Strategic Plan, (2007) *National Center for Research Resources - NCRR* http://www.ncrr.nih.gov/strategic_plan/

11. Black, Donald (2001) Evidence based policy: proceed with care. British Medical Journal. 323: 275-279

12.Elmore, R.F. (1979/1980). Backward Mapping: Implementation Research and Policy Decisions. Political Science Quarterly, 94, 601–616

13.Barbour RS: The case for combining qualitative and quantitative approaches in health services research. Journal of Health Services Research and Policy 1999, 4(1):39-43.

14. Johnstone PL: Mixed methods, mixed methodology health services research in practice. Qualitative Health Research 2004,14(2):259-271.

15. Sterman, John D. (2000): Business Dynamics: System Thinking and Modeling for a Complex World. Irwin/McGraw Hill.

16. *Databases of database*. Available from: http://tools.researchonresearch.org/dodsg/index.php?service=simplesearch.

17 Sinha G (2009) Singapore's Science Bet. Science.

18. Cohen L, Manion L, Morrison KRB (2003) Research methods in . *Routledge*

19. *BIOS Sg*.   Available from: https://tools.researchonresearch.org/biossg/servlet/Controller.

20. de Carvalho, E. C. A., Batilana, A. P., Simkins, J., Martins, H., Shah, J., Rajgor, D., Shah, A., et al. (2010). Application description and policy model in collaborative environment for sharing of information on epidemiological and clinical research data sets. *PloS One*, *5*(2), e9314

21 Emerson R, Fretz R, Shaw L (1995) Writing Ethnographic Fieldnotes.(Chicago, University Of Chicago Press).

22. Atlas.ti software. Available at http://www.atlasti.com/

23. Sterman J (2006) *Business Dynamics* (McGraw-Hill College).

24. Doyle J, Radzicki M, WS T (1998) Measuring Change in Mental Models of Dynamic Systems: An Exploratory Study System Dynamics Review.

25. Wolstenholme, E 2003 Towards the definition and the use of core set or archetypal structures in System Dynamics, *System Dynamic Review*, 19:7-26

26.*Singapore researchers database*. Available from: http://researchers.a- star.edu.sg

27. US Economic Development Administration (2005), Measuring Regional Innovation. A Guidebook for Conducting Regional Innovation Assessments. Council on Competitiveness, U.S

28. Competitiveness, C.o. (2006) Asset Mapping Roadmap:A Guide to Assessing Regional Development Resources. *US Department of Labor,Employment and Training Administration.*

29. Grindley, P, Mowery DC, Silverman B (1994) SEMATECH and collaborative research: Lessons in the design of high-technology consortia. *Journal of Policy Analysis and Management* 13(4): 723-758.

30. Zink A, Silman AJ (2008) Ethical and legal constraints on data sharing between countries in multinational epidemiological studies in Europe report from a joint workshop of the European League Against Rheumatism standing committee on epidemiology with the "AutoCure" project. *Annals of the Rheumatic Diseases* 67: 1041-1043.

31. Beaulieu JA, (2003) Accessing and sharing data from publicly funded research. Trust building and data sharing - an exploration of research practices, technologies and policies. *The Virtual Knowledge Studio for the Humanities and Social Sciences VKS.*

32. Research, Institute of Medicine (U.S.). Committee on Resource Sharing in Biomedical (1996) Resource sharing in biomedical research, eds *Berns KI, Bond EC, & Manning FJ* (National Academies Press, Washington DC, USA).

33. Piwowar HA, Becich MJ, Bilofsky H, Crowley RS (2008) Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers.*PLoS Med* 5: e183.

34. Pedersen T (2008) Empiricism Is Not a Matter of Faith. *Computational linguistics* 34: 465-470.

35. Courneya KS, Friedenreich CM (1999) Physical exercise and quality of life following cancer diagnosis: a literature review. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine* 21(2): 171-179.

36. Ferber R (1977) Research by Convenience. *The Journal of Consumer Research* 4: 57-58.

37. Margins, National Research Council (U.S.). Ad Hoc Panel to Investigate the Geological and Geophysical Research Needs and Problems of Continental. 1979. *Continental margins*. National Academy of Sciences.

38. Pietrobon R, et al. (2004) A suite of web applications to streamline the interdisciplinary collaboration in secondary data analyses. *BMC Med Res Methodol.* 4:29

39. Shah J, Shah A, Pietrobon R. (2009) Scientific writing of novice researchers: what difficulties and encouragements do they encounter? *Acad Med.* 84(4):511-516.

**Figure Legend:**

Figure 1: Network map based on themes from ethnographic interviews

Figure 2: Causal loop diagram focusing on the processes involved in sharing research resource information

Figure 3: Limits against sharing

Figure 4: Shared Research Resources

Figure 5: Research projects viable due to funding and shared resources

Figure 6: example database general description

**Table Legend:**

Example data dictionary

# Appendix

```
Mandatory Sharing=
   Research projects viable due to funding and shared resources/2

************************************************************
Resource consumption = WITH LOOKUP (
   Shared Research Resources*2,
      ([(0,0)-(10,10)],(0.0611621,9.7807),(0.642202,7.36842),(2.38532,5),(4.15902
,3.07018),(6.42202,1.88596),(8.40979,1.14035),(9.90826,0.263158) ))

************************************************************
Time consumption = WITH LOOKUP (
   Shared Research Resources*3,
      ([(0,0)-(10,10)],(0.122324,9.82456),(0.152905,8.94737),(0.550459,5.70175)
,(1.34557,3.42105),(2.99694,2.10526),(5.04587,1.22807),(7.24771,0.701754),(
9.93884,0.175439) ))

************************************************************
Sharing concerns related to loss of competitive advantage=
   11

************************************************************
Voluntary Sharing= WITH LOOKUP (
   (Stakeholders Resistance to share research resources),
      ([(0,0)-(100,100)],(0.917431,97.807),(11.0092,75.8772),(24.7706,45.1754),
(56.5749,21.4912),(88.685,8.77193),(100,1) ))

************************************************************
Mandatory Sharing=
   Research projects viable due to funding and shared resources/2

************************************************************
Shared Research Resources= INTEG (
   Sharing,
      10)

************************************************************
Research Projects=
   Shared Research Resources*Funding/30
Units: 1/year

************************************************************
Research projects viable due to funding and shared resources= INTEG (
```

```
  Research Projects/10,
    10)

**********************************************************************
Research projects viable due to funding and shared resources= INTEG (
  Research Projects/10,
    10)
```